WORLD CLIMATE PROGRAMME

DATA • APPLICATION • IMPACT • RESEARCH

551.501.6 : 551.589 (03)
GUI

WORLD CLIMATE DATA PROGRAMME

GUIDELINES ON THE QUALITY CONTROL OF SURFACE CLIMATOLOGICAL DATA

(Prepared by P.F. Abbott (U.K.) as Rapporteur in
the WMO Commission for Climatology)

May 1986

WCP - 85

BIBLIOTHEQUE OMM WMO LIBRARY

WMO/TD-No. 111

WORLD METEOROLOGICAL ORGANIZATION

01-5354

OMM
WMO
LIBRARY

The WCP consists of four major components implemented by WMO in conjunction with other international organizations:

The World Climate Data Programme (WCDP)
The World Climate Applications Programme (WCAP)
The World Climate Impact Studies Programme (WCIP)
The World Climate Research Programme (WCRP)

WCP 85
TD 111

GUIDELINES ON THE QUALITY CONTROL OF SURFACE CLIMATOLOGICAL DATA

TABLE OF CONTENTS

Page

1. INTRODUCTION 1

2. GENERAL CONSIDERATIONS 1

2.1 The Data before Quality Control 1

2.1.1 Identity 1
2.1.2 Data Format 2
2.1.3 Order 2
2.1.4 Duplicate Records 2
2.1.5 Incorrect Units of Measurement 3

3. BASIC QUALITY CONTROL 3

3.1 Syntax 3
3.2 Internal Consistency 3
3.3 Impossible Values 4
3.4 Extremes 4
3.5 Outliers 4

4. USING REDUNDANT INFORMATION 4

4.1 Functionally Related Parameters 4
4.2 Check Totals 5

5. TIME CONSISTENCY CHECK 5

5.1 Change from Previous Value 5
5.2 Linear Interpolation for Frequently Observed Data . . . 6
5.3 Interpolation for Less Frequently Observed Data 6
5.4 Conclusions and Recommendations 6

6. AREAL QUALITY CONTROL 7

6.1 "Buddy Checking". 7
6.2 Areal Quality Control of Small Regions 7
6.3 Matrix Estimation 9
6.4 Comparison with Changes at Near Neighbours 9
6.5 A Multivariate Statistical Method 11
6.6 Conclusions and Recommendations 12

7. TREATMENT OF SPECIAL PARAMETERS 13

7.1 Areal Quality Control of Rainfall 13
7.2 Areal Quality Control of Sunshine 16
7.3 Areal Quality Control of Pressure 16
7.4 Areal Quality Control of Wind 17

BIBLIOTHEQUE OMM WMO LIBRARY

TABLE OF CONTENTS (continued)

Page

8. HOMOGENEITY OF DATA 18

8.1 Change of Instrument 18
8.2 Instrument Error 19
8.3 Sudden Changes of Site or Exposure 19
8.4 Gradual Changes of Site or Exposure 20
8.5 Other Factors 21
8.6 Conclusions and Recommendations 21

9. TESTS FOR HOMOGENEITY 21

9.1 Introduction 21
9.2 General Principles 22
9.3 Correlation 22
9.4 Comparison with Near Neighbours 23
9.5 Comparison with Other Parameters 23
9.6 Running Means 24
9.7 Detection of Change by Examining Accumulated Totals . . 25

9.7.1 General Principle 25
9.7.2 Cumulative Sums 25
9.7.3 Maronna and Yohai Method 27

9.8 Difference between Successive Values 28
9.9 Significance of Change 29

9.9.1 Student's t-Test 29
9.9.2 F-ratio Test 30
9.9.3 Analysis of Variance 31
9.9.4 Principal Component Analysis 33

9.10 Conclusions and Recommendations 34

10. DISTRIBUTION-FREE TESTS 34

10.1 Ranking 34
10.2 Tests for Change in Mean Value 34

10.2.1 Wilcoxon's Test 35
10.2.2 Mann-Witney Form of the Wilcoxon Test 35

10.3 Test For Dispersion Differences 36
10.4 Tests For Distribution Differences 37

10.4.1 Kolmogorov-Smirnov Test 37
10.4.2 Wald-Wolfowitz Runs Test 38

10.5 Conclusions and Recommendations 38

TABLE OF CONTENTS (continued)

Page

References 40

List of Tables

Table 1 Mean deviation, standard deviation and expected range of the change to some meteorological parameters after t hours 43

Table 2 Mean deviation, standard deviation and expected range of the difference between the observed value and that estimated by linear interpolation from values t hours before and after 44

Table 3 Aitkin interpolation from values t hours before and after . 45

Table 4 Mean deviation $|\chi|$, standard deviation (σ) and expected range e of the difference between the observed value and that estimated from 9 other neighbouring stations 46

Table 5 Mean maximum temperature at Penkridge and Elmdon and derived statistics 47

Table 6 Accumulation of Elmdon (x_i) and Penkridge (y_i) data required for the Maronna and Yohai method 48

Table 7 Tabulation for Von Neumann ratio on Elmdon/Penkridge differences 49

Table 8 Basic statistics for the Penkridge/Elmdon differences . . . 50

List of Figures

Fig. 1a Annual mean of daily maximum temperature at Penkridge (⊙) and the running mean of these data averaged over 4 years(Δ) 51

Fig. 1b Difference between the annual mean of daily maximum temperature at Elmdon and Penkridge (⊙) and the running mean of these averaged over 4 years(Δ) 51

Fig. 2 Accumulation of the departure from the mean of maximum temperature at Penkridge 1962-1980 52

Fig. 3 Accumulated differences between the Elmdon and Penkridge mean of daily maximum temperature 1962-1980 53

TABLE OF CONTENTS (Continued)

		Page
Fig. 4	Accumulated ratio of the wind speed at Greenock to the wind speed at Abbottsinch October 1978 to April 1979	54
Fig. 5	Residual error of the Penkridge monthly mean of daily maximum (a) and minimum (b) temperature estimated using Spackman's principal component model	55
Fig. 6	Total hours of sunshine in S. England on 16 May 1981	56

List of appendices

Appendix I Basic quality Control Checks for Hourly Data in the FM 12-VII SYNOP Code.

Appendix IIa Basic Quality Control Checks for Daily Data.

Appendix IIb Quality Control of Miscellaneous Daily Parameters.

Appendix III Typical Errors Found in Sunshine Data Recorded by a Campbell-Stokes Recorder.

GUIDELINES ON THE QUALITY CONTROL OF SURFACE CLIMATOLOGICAL DATA

1. INTRODUCTION

This paper describes various methods which can be used for the quality control of meteorological data. A range of methods are covered, from very basic checks for correct coding, to time and areal consistency. Homogeneity of the data is dealt with separately.

In general quality control is concerned with the detection of isolated errors in the data and is normally done soon (within a few months) after the creation of the data. On the other hand tests for homogeneity are concerned with detecting bias or drift which may have taken place over a long period of time; the detection normally requires a careful study of large amounts of data which can be done only at a much later stage, often after many years.

Most of the methods described have been used on meteorological data in the United Kingdom, and some indication is given of the merits of each method based on this experience.

2. GENERAL CONSIDERATIONS

Errors in data have many causes, e.g., defective instruments, misreading, incorrect coding, transcription and transmission. The detection of these errors requires some knowledge of the type and cause of these errors in order to assist in the design of suitable detection methods.

What must be appreciated is that not all errors can be detected at the quality control stage; for example, a copying mistake causing an error of 0.1°C in temperature would normally be impossible to detect if the original document is not available. The best that can be done is to establish confidence limits within which we can be reasonably certain that the difference between an estimate and the measured value does not exceed a given threshold. Statistics provide methods whereby these limits can be defined within any required probability level. In this paper the limit is set to the 99% probability level, i.e., on average 1 value in 100 will be outside these limits, suggesting that probably it is in error when in fact it is correct (called an error of the First Kind). In practice, other than purely statistical considerations, such as available manpower, may determine these limits, so that the number of data values flagged as erroneous are within manageable proportions, avoiding many more queries than are likely to be really in error or which can be corrected reasonably.

2.1 The Data before Quality Control

2.1.1 Identity

An important consideration is "can the data be identified?" If the type, date, time and location of the data are not included, the data will normally be worthless. Occasionally it may be possible to identify data when they are embedded amongst other messages which can be recognized, such as observations received via the global telecommunications system.

2.1.2 Data Format

When meteorological data are coded into synoptic messages the message must have a defined format. This is particularly the case with, for example, the WMO, 1982 SYNOP code where most groups are preceded by an indicator digit and the observation divided up into sections separated by the section number repeated three times. Before examining the meteorological data contained in the message it is advisable to check the format of the message, otherwise incorrect group indicators and incorrect or missing section separators may cause data to be rejected. Checks should be made within each section to ensure that the indicators to the five-figure groups do not decrease as one proceeds through the message except in Section 3 of SYNOP or in Section 5; in other sections, if they do then this suggests that either the groups have been incorrectly numbered or that the three-digit section separator is missing.

2.1.3 Order

The data may be available only as manuscript or in some machineable form, e.g., punched Hollerith cards, paper tape, magnetic tapes or discs, or optical discs. Whatever its form, it will have been collected in such a way that it will be in some kind of order, e.g.

- Climatological order - all values for one station in time order
- Synoptic order - all values from many stations for one time

Since this may not be the order which is finally required, before undertaking the considerable task of reorganising the data into the required sequence, it is advisable to take advantage of the existing arrangement first. For example, it is much easier to carry out areal quality control, which requires comparison with neighbouring stations, while the data are in synoptic order. When this process has been completed the data may be rearranged into climatological order thus making it much easier to carry out time sequence checks.

2.1.4 Duplicate Records

Duplicate records can occur in a number of ways, often depending upon the source of the data. If the source is via the telecommunications system, then incorrect station numbers which just happen to be the same as another station will either lead to two records with the same station number but with different data or to one record replacing the other. Depending upon the order of replacement the final record may or may not be correct. When the source is punched cards then duplicates can occur because either the record was accidentally re-keyed or a mistake was made and the record re-keyed but the first attempt not removed. In the first case the records are likely to be the same whereas in the second they will not be identical.

Duplicates in the final archive can be found fairly easily by sorting the data by station number and time, so that all identical records cluster together, after which they can be removed. This method will also detect those occasions when the same data are assigned to two different stations, though in

this case the sort must be done on some of the meteorological parameters in the data record, e.g., temperature, wind, rain and sunshine. Although it is possible that one of these parameters might be the same it is extremely unlikely that all of them will be the same, strongly suggesting that duplication has occurred.

Replacement is much more difficult to detect without referring back to the observing station (not always possible) though areal quality control may detect this type of error (see Section 6).

2.1.5 Incorrect Units of Measurement

If data have been collected from several sources then the units of measurement can be different, e.g., temperature in whole degrees Celsius, tenths of a degree Celsius or in degrees Fahrenheit. Sometimes these units are indicated by a code, for example i_w in the WMO 1982 SYNOP code, and usually very large errors will result if this code does not indicate correctly the actual units of the data reported.

One way of overcoming this problem is to record the units expected separately. Then the reported code can be queried if it does not agree with this expectation. If this is not possible, the sequence checks described in Chapter 5 or Sections 3.4 and 3.5 will usually detect most of these errors.

3. BASIC QUALITY CONTROL

3.1 Syntax

Many data are coded by following a well defined set of rules. These same rules can be incorporated into computer programmes for checking the data. Experience with the processing of United Kingdom and GATE* data shows that most errors seem to occur in the reporting of cloud. This occurs probably because the rules for coding cloud are the most complicated or because cloud is not usually a measured parameter and depends very much upon the knowledge which the observer has of clouds and of the appropriate coding procedures.

3.2 Internal Consistency

Since within one observation the parameters are somewhat interrelated, a simple check comprises making sure that the observation is consistent within itself. Examples are: snow with temperatures correspondingly low; drifting snow with a report of strong wind; fog with poor visibility.

---

* GATE - Global Atmospheric Research Programme (GARP) Atlantic Tropical Experiment.

3.3 Impossible Values

Certain parameters are constrained within absolute limits, e.g., wind direction between 0 and 360° degrees, cloud amount between 0 and 8/8, or 9 for sky obscured. A failure of a test for these limits means that it is absolutely certain that the value is incorrect.

Tests for syntax, internal consistency and impossible values are listed in Appendix I for hourly data and in Appendix II for daily climatological data. The tests in Appendix I follow the work of V.V. Fillippov, (1968), Hoflich, Meissner and Hoffman (1975) and some unpublished work, particularly by D.E. Parker, but have been revised to correspond with the 1982 WMO SYNOP code.

3.4 Extremes

Most parameters have known climatological limits depending upon the locality. Here it is advisable to set the test within these limits so that the extreme value will be queried and checked fairly soon after it occurs.

Since there is usually a seasonal variation in climatological variables, the test can be improved by applying separate climatological range checks for each month of the year.

3.5 Outliers

A simple method for detecting wild values is to first rank the data then trim off the upper and lower 25%. From the remaining central 50% determine the mean, the standard deviation σ from the half range, and w (half the difference between the lowest and highest value in the trimmed set). Provided the data are distributed normally,

$$\hat{\sigma} = 1.4826\ w$$

All the data can now be checked to see if they are within acceptable limits, e.g., the limits for a significance level of 1% would be $\pm$ 2.576σ and for 0.1% would be $\pm$ 3.291σ.

4. USING REDUNDANT INFORMATION

To make the most efficient use of storage, derived values determined from the primary data may not be stored, but if they are an integral part of the original data then they can be used as checks.

4.1 Functionally Related Parameters

Humidity parameters, such as relative humidity, vapour pressure and dewpoint are usually derived from the dry bulb and wet bulb temperature. Assuming that the humidity parameters are correct then they can be used to check that the dry and wet bulb temperatures have been entered correctly. However, in practice, the main source of error occurs in the hand computation of relative humidity, vapour pressure and dewpoint. Thus, where both dry bulb and wet bulb temperatures are available it is best to concentrate effort on checking these, using the sequence checks described in Section 5, and then

recomputing these humidity parameters; a simple job for a computer if the data are available in a suitable form. Even with SYNOP's, where only dry bulb and dewpoint temperatures are reported, this is probably still the best approach.

4.2 Check Totals

This is a very powerful check for detecting copying errors and is often used for rainfall and sunshine. If the daily values do not add up to the monthly total then either the total is wrong or at least one of the daily values is wrong.

For hourly rainfall and sunshine it is best if they are entered on a grid showing hours along rows and days down columns with both row and column totals. If the data including totals are then keyed into a computer, the computed totals can be checked against the entered totals and all rows and columns found in error flagged for checking.

5. TIME CONSISTENCY CHECK

These tests compare values with those measured before and/or after the event and are only suitable for parameters which are highly correlated in time, e.g., temperature, pressure and wind.

5.1 Change from Previous Value

The change in value in one time step from t = -t to t = 0 should not exceed specified limits, i.e.,

$$\Delta = x_o - x_{-t}$$

$$\text{then} \quad |\Delta| \leqq e \quad .$$

Values of e for various parameters are given in Table 1 for data measured at 1 to 3 hour intervals and also for those measured only once a day.

This method can be extended to compare the change in one time step to either the change at near neighbours (see 6.4) or the change at other levels. For example a scheme suitable for checking soil temperature (T) measured at several levels could be

$$\Delta T_\ell = T_o - T_{-t} \qquad \text{at level } \ell.$$

$$\text{If} \quad |\Delta T_\ell| \leqq e, \text{ the data are acceptable.}$$

Values of e at various levels are given in Table 1.

If the data fail this test then before rejecting the value compare the change ΔT_ℓ with the change ΔT_L at another level L. The change should be of the same sign provided ℓ and L are fairly close,

i.e., $\Delta T_\ell \times \Delta T_L \geqq 0$

and $|\Delta T_\ell| \leqq (e_\ell/e_L)\ |\Delta T_L|$

5.2 Linear Interpolation for Frequently Observed Data

The value estimated by linear interpolation between values measured t hours before and after,

$$\hat{x}_o = (x_{-t} + x_{+t})/2$$

can be used to check the measured value x_o, the check being

$$|x_o - \hat{x}_o| \leqq e.$$

Values of e are given in Table 2 for values measured at 1, 3 and 24 hour intervals.

5.3 Interpolation for Less Frequently Observed Data

When data are unequally spaced in time or observed less frequently (e.g., every 6 hours) than necessary to resolve the time variation of the parameter being analyzed, it is best to use non-linear interpolation methods, e.g., Aitken's interpolation method, see Table 3.

Good estimates can often be obtained by attempting to fit the smaller signal arising from either the diurnal or day-to-day variation. If the diurnal variation is small then linear interpolation between values measured at fixed hourly intervals may still give satisfactory results. On the other hand if the diurnal variation is large but the day-to-day changes small, then linear interpolation can be used to estimate a value from the values measured at the same time on the day before and after.

5.4 Conclusions and Recommendations

Although the change from a previous value (method 5.1) is probably the most widely used, the linear interpolation method (5.2) is usually better. Since there is considerable persistence in wind direction these methods can be used, but with caution, because quite genuine and sudden changes occur with, for example, the passage of fronts, squalls.

For parameters observed at 3 or 6 hour intervals, a non-linear interpolation method such as Aitken's should be used though the other methods can be effective if used sensibly (see 5.3).

None of the methods described in this section are suitable for rainfall, sunshine, cloud amount or visibility because quite sudden changes, particularly at certain times of the year, are a normal occurrence.

6. AREAL QUALITY CONTROL

If there are stations nearby then these can be used as independent checks. The effectivenesss of areal quality control will depend upon the density of the observing network and how well a parameter correlates in space. Maximum temperature is highly correlated in space and so the network used for quality control can be fairly sparse, whereas minimum temperature is less correlated because of topographical effects. There can be large spatial variation in rainfall and sunshine data which therefore require a dense network if they are to be resolved correctly.

Areal quality control usually requires the processing of quite large quantities of data; thus, good data organization becomes important. Usually the job can be reduced to manageable proportions by having the data ordered synoptically (see 2.1) and, of course, a computer with adequate direct access storage (magnetic disc) will make the task relatively easy.

6.1 "Buddy Checking"

This method compares the data from a station with one or more neighbours chosen to be climatologically similar. Although a straight comparison can be made, it is better to compare the departures from the station mean and thus remove the real and quantifiable difference between the two stations.

The following steps are involved:

(i) For station j compute the monthly mean $\bar{x}_j$,

(ii) On day i compute departure of the daily value from the monthly mean,

$$x_{ij} - \bar{x}_j = \Delta_{ij}$$

(iii) Compute the difference in departures at station j and station j-1 (or j+1),

$$\Delta = |\Delta_{ij} - \Delta_{i,j-1}| \quad .$$

A query should be made if $\Delta \geqq 2.5°C$ for temperature

$\geqq 4.0$ hours for sunshine

$\geqq 5.0°C$ for depression of grass min. below air min. temperature.

6.2 Areal Quality Control of Small Regions

For areal quality control, the country is usually divided into regions which experience has shown to be climatologically similar. (Discriminant analysis can be used to indicate likely groupings). First the monthly mean for each station in an area is computed. Then for each station the departure of each daily value from its own station mean is compared with the corresponding average departure computed from the remaining stations in the area.

The steps for an area containing n stations are:

(i) Compute the monthly mean $\bar{x}_j$ for each station j.

(ii) For day i compute the mean and variance of the daily departures for all stations except the station k under test:

$$\bar{x}_i = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq k}}^{j=n} (x_{ij} - \bar{x}_j)$$

$$\sigma_i^2 = \frac{1}{n-2} \sum_{\substack{j=1 \\ j \neq k}}^{j=n} (x_{ij} - \bar{x}_j)^2$$

(iii) Compute the departure on day i for station k:

$$x'_{ik} = x_{ik} - \bar{x}_k$$

(iv) Compute the difference between the departure at the station and the mean departure of all the other stations in the area:

$$\Delta = |x'_{ik} - \bar{x}_i| \quad .$$

An error is signalled if $\Delta \geqq e$ or $\Delta \geqq 2\sigma_i$, where e = 2.5°C for maximum temperature, 3.0°C for minimum temperature, and 4.0 hours for daily sunshine total.

This method has been described by Sneyers (1972) and used in the United Kingdom where each area contains between 6 and 14 stations. This method is quite good for checking maximum, minimum and grass minimum temperatures, though because of the small sample size, problems can arise when there is more than one value in error in an area. The method is not as good for sunshine because of the higher variability of this parameter in the UK, and is not used at all for rainfall. The areas were determined empirically mostly from consideration of the areal variation of temperature; the areas would almost certainly be different if they had been determined from rainfall data.

6.3 Matrix Estimation

Meteorological observations can be tabulated to form a matrix of dimensions time x stations:

Time →

Stations ↓

$$\begin{matrix} x_{11} & x_{21} & x_{31} & - & - \\ x_{12} & x_{22} & x_{32} & - & - \\ x_{13} & x_{23} & x_{33} & - & - \\ x_{14} & x_{24} & x_{34} & - & - \end{matrix}$$

At any one time, it would be expected that the stations all have values of similar magnitude. For example, in a polar airstream all stations would have relatively low temperatures. Likewise at any one station it would be expected that all values have a similar magnitude typifying that station; e.g., some stations would nearly always be colder than other stations in the system.

If this is accepted then x_{ij} can be estimated by

$$\hat{x}_{ij} = \sum^{j} x_{ij} \cdot \sum^{i} x_{ij} \div \sum^{i} \sum^{j} x_{ij} \quad ,$$

i.e., Column total x Row total ÷ Grand total.

An error is signalled if $|\hat{x}_{ij} - x_{ij}| > e$.

Statistics of this difference are given in Table 4 for a network of 10 stations at three different network densities.

6.4 Comparison with Changes at Near Neighbours

This method uses the principle that a change at a station should be similar in magnitude and sign to that at near neighbours. Ideally the station under test should be at the centre of the network of neighbouring stations.

Consider a system with two neighbours A, B, (ideally one on each side of the query station Q) with measurements at time t = 0 and t = t. The arrangements can be illustated schematically:

Value of element at time t = o	t = t	Location
A_o	A_t	Comparison station or neighbour
Q_o	Q_t	Query or station under test
B_o	B_t	Comparison station or neighbour

The change which occurred at each station between time t = 0 and t is :

$$\Delta A = A_t - A_o \qquad \Delta Q = Q_t - Q_o \qquad \Delta B = B_t - B_o \quad .$$

The average change at the neighbours is

$$\bar{\Delta}_{AB} = (\Delta A + \Delta B)/2 \ ,$$

and this is then compared with the change ΔQ and a query raised if

$$|\Delta Q - \bar{\Delta}_{AB}| \geqq e \ .$$

Buss (1977) suggests that for daily maximum and minimum temperatures e should be $(\Delta A - \Delta B)/2$ or 3°C, whichever is the greater.

Another version of "buddy checking" proposed by Buss is to compare the maximum temperature at neighbours with that at the query station after making an altitude correction,

$$\lambda = [h_Q - (h_A + h_B)/2]/100 \quad °C \ ,$$

where h_Q, h_A, h_B are the heights (metres) of the query and the neighbour stations. Applying this correction

$$\overline{AB} - (Q + \lambda) > 5°C \ ,$$

where $\overline{AB} = (A + B)/2$, or alternatively

$$|A - (Q + \lambda)| > 7°C \quad \text{or} \quad |B - (Q + \lambda)| > 7°C \ .$$

An estimate of the value at the query station can be made using the initial value and the average change at the neighbours, i.e.,

$$\hat{Q}_t = Q_o + \bar{\Delta}_{AB} \quad ,$$

or from the average value at the neighbours using the altitude correction

$$\hat{Q}_t = \overline{A_t B_t} - \lambda \quad .$$

Buss combined these to make the estimate, i.e.,

$$\hat{Q}_t = \frac{1}{2} [(Q_o + \bar{\Delta}_{AB}) + (\overline{A_t B_t} - \lambda)] \quad .$$

6.5 A Multivariate Statistical Method

Spackman (1979) has developed a method using Principal Component Analysis with Varimax rotation to provide a set of 15 Factor Scores for each station. These factors are fitted to the data on a particular day by linear regression:

$$x(s,d) = a_0(d) + a_1(d)\, f_1(s) + a_2(d)\, f_2(s) + \ldots\, a_n(d)\, f_n(s) + r(s,d),$$

where

$\hat{x}(s,d)$ is the estimate of $x(s,d)$ at station s on day d,

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

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

$r(s,d)$ is the residual error which is unexplained by regression, i.e.,

$$r(s,d) = x(s,d) - \hat{x}(s,d) \quad .$$

The residual will arise from:

- inadequacy of the factor model to fit the data on a particular day
- instrumental errors
- observer errors
- data processing errors including rounding and truncation.

In the United Kingdom the mean residual $\bar{r}(s,d)$ of maximum and minimum temperature is usually small ($\simeq$ 0.3°C). The standard deviation $\sigma(r(s,d))$ of the residuals is approximately equal to 1.0°C (range 0.5 to 1.9°C) for minimum temperature.

The standard deviation increases from south to north and some stations, not necessarily near neighbours, have very similar characteristics, e.g., the islands in the Irish Sea. Consequently for quality control purposes the country is divided into 10 districts plus 3 groups:

(i) Orkney, Shetland and N. Scotland (Caithness),

(ii) Isle of Man (Irish Sea),

(iii) Channel Islands (English Channel),

and mean residuals and standard deviation of residuals averaged for each of these areas. These average values are used to check the data. An observation is deemed suspect if the standardized residual

$$\left| \frac{r(s,d) - \bar{r}(s,d)}{\sigma(r(s,d))} \right| > 3.25 \quad .$$

A computer programme plots a suspect value at the centre of a map covering an area of approximately 100 km x 100 km along with any other data values in the area. The programme also prints out the queried value, the residual and computed estimate and a suggested reason for the error (index, reset, $\pm$ 5 or $\pm$ 10 error).

Values of the standardized residual of $\leqq$ 3.5 are usually acceptable; $\geqq$ 4.5 are usually in error. The limit of 3.25 detects most (but not all) unacceptable values.

Of the queried values approximately 50% will probably be acceptable. The technique works better for temperature and sunshine than for rainfall, which it is sometimes unable to fit adequately using a network of approximately 600 stations ($\simeq$ 1/10th of the complete rainfall station network).

A step by step procedure has not been described for this method because a considerable amount of computation and an advanced knowledge of statistics is required. The method is only practical where a large computer is available along with good statistical computer programmes such as supplied by the BMDP (Biomedical Computer Programme) package.

6.6 Conclusions and Recommendations

Areal quality control is an essential part of the checking process to ensure that the data have spatial consistency. The most successful methods make comparisons with more than one station and place the query station at the centre of a network which is not too sparse, or objectively select comparison stations, not necessarily close by, which have similar characteristics.

7. TREATMENT OF SPECIAL PARAMETERS

The large spatial variation of some parameters requires a dense network of observations if essential details are to be resolved adequately. Other problems arise because the distribution of some parameters are not normal, making it difficult to use the usual statistical tests which assume a normal distribution. Usually this difficulty can be overcome by transforming the data by some operation such as using the logarithm or reciprocal of the data value.

Rainfall, sunshine and wind have some of these characteristics and are dealt with in this section.

7.1 Areal Quality Control of Rainfall

The method described by Shearman (1975) is used in the United Kingdom where the network is fairly dense (approximately 6000 stations).

The rainfall at a station is compared with an estimate determined from neighbouring stations which, ideally, should be evenly distributed around the query station. This is achieved by considering each octant around the query station up to a distance of 25 km. The nearest suitable station is selected from each sector; if 8 stations are not available (e.g., at a coastal site), then other stations are selected to make up the eight. The rainfall amounts, r_i, are converted to percentages of the annual average: $R_i = 100r_i/\bar{r}_i$.

The steps are then:

(i) Compute the mean and standard deviation σ of these percentages. Since these values will be used eventually to estimate a value at the query station, only those values within the empirical limit (mean $\pm$ 1.75σ) are used to estimate the mean. This limit is probably too restrictive in showery conditions when there are likely to be genuine widely different measurements. Thus, an additional test can be applied: If the present weather code from any station in the area in the last 24 hours is greater than or equal to 70 (snow, showers or thunderstorms) then all the values are used.

(ii) Using the n stations which are within these acceptable limits, compute an estimate $\hat{R}$ from the ratios at the neighbours R_i, (i = 1,2,....n), weighted by the inverse distance D_i squared

$$\hat{R} = \frac{\sum R_i/D_i^2}{\sum 1/D_i^2} \quad .$$

After converting the percentage of annual rainfall, R, back to actual rainfall, r, an error is signalled if

$$|r - \hat{r}| > 2.5 \text{ mm}, \quad \text{or}$$

$$|r - \hat{r}| > C\,\sigma/\sqrt{n}, \quad \text{where}$$

$C = 2$ for daily rainfall amounts

$C = 4$ for monthly rainfall amounts.

Some of the reasons for incorrect rainfall measurements that are known and can be allowed for are:

(a) Wrong days - whole month

Sometimes the observer will credit rainfall to the wrong day for the whole month leading to a record out of phase by 1 or 2 days. This can be checked by correlating the estimated rain with the measured rain lagged -2, -1, 0, +1, +2 days and reallocating the rainfall backwards or forwards the appropriate number of days as determined from the maximum correlation coefficient, provided it exceeds 0.95.

(b) Wrong days - part month

This usually occurs because the observer has entered too many or too few zeros during a dry spell. The same procedure as in (a) is used but the day shifting is limited to a period of rain preceded and followed by two or more dry days.

(c) Indicated accumulations

If an observer is unable to make rainfall measurements every day, e.g. at weekends or because he is unable to reach the site, then rainfall will accumulate and an honest observer will indicate this on his rainfall returns.

If the rainfall accumulation over n days is r', then it can be reapportioned to each day i using the relationship

$$r_i = r' \frac{\hat{r}_i}{\sum^{n} \hat{r}_i},$$

where $\hat{r}_i$ is the estimated rainfall on day i, and $\sum^{n} \hat{r}_i$ is the accumulation over n days estimated from the neighbours.

(d) Unindicated accumulations

If accumulations have not been indicated by the observer but are suspected because the difference, d, between the observed value and that estimated from neighbours is substantially greater than zero ($d > 2.5$ mm), then it is sometimes possible to detect this by searching backwards up to 10 days accepting deficits ($d < 0$) caused by an indicated rainfall of zero and stopping at the first day with measurable rain ($r > 0$). The accumulation can then be reapportioned as in (c).

(e) Transposed values

Values on consecutive days may be transposed either by clerical or other error and can be detected by looking for differences, d, of similar magnitude between the observed and estimated values on consecutive days but opposite in sign, i.e.

$$d_i \simeq - d_{i+1} \quad .$$

If after reallocating the values the differences satisfy the conditions

$$d \leqq 2.5 \text{ mm} \quad \text{and}$$

$$d \leqq C\, \sigma/\sqrt{n} \quad ,$$

then the reallocation can be accepted.

In practice this method is only effective in clearly defined cases, suggesting that the conditions should be made more restrictive or used only to warn that values may have been transposed.

(f) Incorrect time of observation

The normal time for measuring rainfall amounts in the U.K. is 0900 GMT. If for some reason the rainfall is measured an hour or so later, during which time appreciable rain has fallen, then the measured value will be in excess and the value the following day, if measured at the correct time, will be in deficit. A procedure the same as that described in (e) can be used but with caution.

An alternative simple method for the quality control of rainfall data is to plot the values on a map, draw isohyets and then inspect the result. Obviously the method will be more effective if backup information such as carefully analyzed synoptic weather maps and satellite or radar pictures are available.

Rainfall can be very variable in space and time particularly in unstable weather conditions; this has been extremely well demonstrated by Hatch (1976) who showed a set of maps drawn from an increased network of rainfall stations. The sparse network gave no indication of an extremely intense storm which was resolved by the dense network.

7.2 Areal Quality Control of Sunshine

A method very similar to that for the areal quality control of rainfall can be used except that it is best to use ratios of hours of actual sunshine to the length of day (ratio $\leqq 1$) for the quality control of daily sunshine totals. For monthly sunshine the ratio of total actual sunshine to the climatological monthly average may be used.

Alternatively, the values can be plotted on a map (Fig. 6) and analyzed by hand, ideally supported with synoptic weather maps, satellite pictures where available, and a good knowledge of the climatology of sunshine.

Apart from the many types of errors which can occur in sunshine data (some typical errors found for sunshine recorded by a Campbell-Stokes sunshine recorder are described in Appendix III) the measurement can be low because of some obstruction. These low values cannot be compared with neighbouring measurements without making some adjustment. If the elevation of the horizon can be measured then the percentage obstruction can be calculated and used to adjust the data. Alternatively if the horizon is unknown, which is often the case with historic data, then the percentage obstruction can be estimated from the ratio of the maximum sunshine ever recorded in a particular month over many years to the astronomical length of day for that month.

7.3 Areal Quality Control of Pressure

The geostrophic wind relation can be used to check pressure differences between stations (except near the equator), i.e.,

$$\Delta_P = 2\rho\omega \sin\Phi \, V_g \, \Delta x \quad ,$$

where

Δ_P is the pressure differences (mb) between two stations Δx(km) apart,

V_g the geostrophic wind,

Φ latitude,

ρ air density $\simeq 0.00129$ g.cm^{-3},

ω earth's angular velocity $= 2\pi/24$ hour^{-1} .

Since the surface wind V_o is usually the measured value available, this can be substituted by assuming that

$$V_g/V_o \simeq 2.5 \text{ for land stations in the United Kingdom.}$$

If the surface wind is inclined $\theta°$ to the line joining the two stations then:

$$\Delta_P = a \sin\Phi \, V_o \sin\theta \, \Delta x \quad ,$$

where

$a = 0.00453$ when V_o is measured in knots and Δx in nautical miles

$a = 0.00475$ when V_o is measured in m/s and Δx in kilometres.

If Δ_{Po} is the observed pressure difference then the quality control check is

$$|\Delta_P - \Delta_{Po}| < e \quad ,$$

where e is made proportional to the distance between the stations:

$$e = 0.003\Delta x + 0.8 \text{ mb.}$$

This method can be improved by making the ratio V_g/V_o a function of the time of day and the time of year.

7.4 Areal Quality Control of Wind

Apart from basic range checks and time sequence checks the wind can be estimated from the pressure gradient using the geostrophic relationship,

$$V_g = \frac{\delta p}{\delta x} / (2\omega\rho \sin\Phi) \quad ,$$

either by a method similar to that described under pressure or by first fitting a surface to the pressure field. However, if the area to be fitted is quite large and the pressure distribution complex a high order polynominal will be required to obtain a good fit.

Normally such high powered techniques are not thought to be worthwhile, it being more important to detect bias, which is fairly common in wind instruments. Wind vanes can become loose or shift due to maltreatment or storm damage. Speed recorders may suffer a zero-shift or a scale change.

A method has been developed by Bryant (1979) for detecting these defects. The method is essentially a form of "buddy checking" whereby a pair (or pairs) of stations are first analyzed to establish normals. Then data for each month are compared with these normals as described by the following method:

(i) Establish normals between two stations A and B by analyzing at least 1 year of hourly wind data as follows:

(a) For each 20° sector 's' at one station determine the average direction difference between the two stations and their corresponding variance:

$$\bar{\Delta}d_s = \frac{1}{n} \sum (d_A - d_B)_s \qquad \sigma(d)_s^2 = \frac{1}{n-1} \sum (\Delta d_s - \bar{\Delta}d_s)^2$$

(b) Combine these sector values to form the overall values

$$\bar{\Delta}d = \frac{1}{s} \sum_s \bar{\Delta}d_s \qquad \sigma(d)^2 = \frac{1}{s} \sum \sigma(d)_s^2$$

Note that σ^2 is not the total variance (which would include the variation of the sector means which are assumed to be real and predictable).

(ii) For the sample month repeat the above calculations to determine the sample difference $\bar{\Delta}'d$, then test

$$|\bar{\Delta}d - \bar{\Delta}'d| < e \quad .$$

Trials have established that e is approximately equal to $\sigma(d)$.

A similar method can be used for wind speed except that the ratio of the speed at station A to that at station B should be used, instead of wind speed differences.

Obviously this method will only work if the airstream flowing over A is also flowing over B. To help ensure that only such cases are considered the method is applied only to those occasions when the wind is greater than 7 kn. A further refinement is to consider only winds during the daylight hours when it is more likely that the airstream at A and B will be similar; however this drastically reduces the amount of data and therefore is not usually adopted.

8. HOMOGENEITY OF DATA

In practice it is almost impossible to achieve homogenous data. All sorts of factors, some unavoidable, will introduce inhomogeneity into the data. The effect can take the form of either a bias, trend or oscillation.

This section does not describe all the possible factors which may cause inhomogeneity in the data but rather provides a description of some of the hazards known to the author so that they may serve as a warning to the user of the problems which can arise.

8.1 Change of Instrument

All instruments have limited sensitivity, so that they are either unable to respond to changes which occur above certain frequencies, introducing averaging, or to respond immediately to sudden changes (lag).

Consequently a change of type of instrument will nearly always cause a sudden irregularity in the data. For example, if an anemometer is replaced by another more sensitive instrument which is more responsive (lower lag and averaging time) and with a lower stopping speed, this may be expected to alter:

(i) Level of gusts,

(ii) Mean speed at low speeds,

(iii) Number of calms recorded,

and thus appear to have altered the climatology.

8.2 Instrument Error

Even when an instrument is read correctly the value may be wrong because the calibration of the instrument was imperfect; this can be for a number of reasons:

(i) Response assumed to be perfectly linear;

(ii) Calibration wrong;

(iii) Calibration method imperfect - this may not be appreciated at the time of calibration. Examples where this might occur are when the instrument is exposed or mounted differently from that actually used in the field, or where assumptions are made which later prove to be invalid;

(iv) Calibration drift - this may be gradual or sudden. Gradual changes would arise if, for example, friction was allowed to steadily increase due to lack of maintenance or arise because of aging, as can happen in electric resistance thermometers unless some compensating method is used. Sudden changes arising from accidental damage or maltreatment are fairly common and can also be due to shock or release of stress, as can happen with aneroid barometer pressure capsules.

8.3 Sudden Changes of Site or Exposure

If an instrument is moved, either horizontally or vertically, it will almost certainly cause a sudden change to the measurement. Even quite small movements to new positions within an instrument enclosure will have altered relative distances to nearby objects and the effect these have on the instrument. Larger movements will result in the instrument having a different exposure to wind, rain and sunshine which, in turn, will affect the measurement. Changes in height above ground of any instrument e.g., raingauge lip above ground, anemometer height, Stevenson Screen height, in fact the height of any sensor, will cause substantial alterations to measurement. Some changes which may affect measurement are described below.

Changes to the Stevenson Screen:

(i) Change in size of screen;

(ii) Change in Louvre arrangement or the floor, altering ventilation;

(iii) Change of position of instrument inside the screen. There may be substantial temperature gradients from the sides to the centre of the screen in certain conditions;

(iv) Change of structure on which the screen stands. This is noticeable if the structure is made much more solid or is boarded in any way;

(v) Change of soil surface on which the screen stands e.g., dug up, allowed to become unkempt with long grass, or rubbish dumped on the ground below the screen;

(vi) Change in the colour of the screen, particularly the common case of it becoming dirty or the paint deteriorating, does not seem to cause serious discrepancies;

(vii) Changes to instrument mounting. Whether an instrument is mounted inside some enclosed box, on top of a pole or on a side arm to a pole or pylon, will have some effect upon the measurement. For example, the ideal mounting for an anemometer is at the top of a single slender pole; anemometers mounted on side arms to poles or, worse, lattice pylon structures, are seriously affected by these structures, altering both the average level and variability of the measurement;

(viii) Changes to protective shields on instruments, such as radiation shields to thermometers or shields protecting anemometer rotors from accidental damage.

8.4 Gradual Changes of Site or Exposure

Many of the changes already described can cause gradual change, though there are other more obviously gradual changes which may introduce a trend into the data. These are described below:

(i) Gradual settling of soil which will affect temperature gradients near the surface, drainage, and thermal conductivity;

(ii) Gradual growth of trees and shrubs which can produce two opposite effects, i.e. shelter or increased turbulence, depending upon the distance between the trees and the sensor. For example, shelter will reduce wind speed but may increase rainfall catch if the shelter is not too close. Increased turbulence will alter the variability of most measurements and may also alter the average magnitude, particularly if the turbulence contains large eddies;

(iii) Urbanization, which may also cause shelter or increased turbulence, produces other side effects such as heat output (urban heat island effect) and pollution, affecting visibility and possibly sunshine and radiation;

(iv) The gradual appearance (1-5 years) of nearby man-made structures, e.g., very large buildings, motorways, and reservoirs.

8.5 Other Factors

The value that a sensor measures can be presented to the user in several ways, e.g., graphic, counter, recorded and stored electronically or read directly against some scale. Each method may have defects due to friction, extraneous vibration, parallax, and sensitivity of the indicator to environmental conditions, e.g., temperature, humidity, and ventilation. Thus, even if the sensor remains the same and only the method of indication is altered, this can introduce inhomogeneities into the data.

Apart from changes to sensors, indicators and site, there can be changes to supporting systems which will also affect the measurement, e.g., wires becoming kinked, stretched, or corroded; ventilation or vacuum systems developing blockages or leaks and moving parts requiring cleaning and lubricating.

8.6 Conclusions and Recommendations

There are clearly many pitfalls to changing or moving instruments. As some of these will be unavoidable certain precautions should be taken to lessen the worst effects. Good records should be kept about the site and exposure and any changes noted. All maintenance, painting and repairs should be recorded along with any instrument changes, even when the replacement is of the same type. Also the site should be visited regularly so that independent assessments can be made and instruments checked against reference instruments brought to the site especially for this purpose. Whenever an instrument is replaced by one of a different type, the new and old instruments should be run simultaneously, if possible for at least two years, so that any differences can be quantified.

9. TESTS FOR HOMOGENEITY

9.1 Introduction

Inhomogeneity will either take the form of bias, drift or oscillation, or some combination of these. The aim of tests for homogeneity is to indicate when such an irregularity has been introduced into the data and, if possible, assess what form the irregularity takes and whether or not this is significant.

Throughout this section reference will be made to one particular set of data in order to illustrate the various methods. The data are the annual means of daily maximum temperature measured at Penkridge (the query station P) and Elmdon (the good quality comparison station E). The two sites are located near Birmingham, 37 km apart at approximately 100 m altitude.

The purpose of most of the analyses is to show that the Penkridge data are not homogeneous, and methods are described for locating when the changes occurred and their nature and significance.

9.2 General Principles

The level of the change being sought is often much smaller than the level of noise and genuine signal in the data, and most methods which aim to remove the unwanted signal will only confuse the investigation. Commonly used methods for removing part of the signal from the data are: computation of some sort of running mean or computation of the differences between successive values, comparison with some other parameter and comparison with the same parameter but measured at a near neighbour. The choice of method will often depend upon what other data are available.

Practically all the methods which use comparison with other data make the assumption that good correlation between the data exists even though the degree of correlation may not have been discussed.

9.3 Correlation

Where the variable y_i under test is known to be correlated with some other parameter x_i then a relationship can be established and used to make an estimate $\hat{y}_i$ of the test data and the residual error $\hat{y}_i - y_i$ examined for any unusual trend.

Usually a very simple linear relationship is assumed, i.e.,

$$\hat{y}_i = ax_i + b \quad .$$

For temperature, pressure and wind direction the assumption that $a = 1$ is often made, then

$$b = \bar{y} - \bar{x} \quad .$$

For sunshine, rainfall and wind speed it is often assumed that $b = 0$, then

$$a = \bar{y} / \bar{x} \quad .$$

Where these assumptions are not appropriate then a and b can be determined from the least squares estimates

$$a = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_1^n (x_i - \bar{x})^2} \quad ,$$

$$b = \bar{y} - a\bar{x} \quad ,$$

and the correlation coefficient r is

$$r = \frac{\sum_{1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\{\sum_{1}^{n}(x_i - \bar{x})^2 \sum_{1}^{n}(y_i - \bar{y})^2\}^{1/2}} .$$

Great care is needed when using these methods because high correlations do not necessarily lead to meaningful results. For example total sunshine appears to be highly correlated with either the maximum temperature or the diurnal range. However, most of this correlation arises because both sunshine and temperature are high in summer and low in winter. If this effect is removed by only studying a single month over many years, then the correlation decreases and, for example, becomes insignificant in the United Kingdom during the winter months (the correlation coefficient decreasing from ≃ 0.9 in summer to less than 0.4 in winter, when advection becomes a relatively important factor in determining the air temperature). To avoid these problems it is usual to compare the annual mean value in each year.

9.4 Comparison with Near Neighbours

One or more stations are selected which are similarly situated, not too far away and having similar climatological characteristics. These neighbours can be selected by choosing stations having the highest correlation as determined from data measured at the test station which are not thought to be suspect. Usually the correlation coefficient is required to be high (> 0.95). Both the differences and ratios of the neighbour to the test station are computed for each year. This method removes the most significant climatological signals (diurnal, monthly and annual variations) but still leaves, apart from noise, effects due to exposure and height.

For the Penkridge data, Elmdon was selected as the comparison station and the differences between the annual mean maximum temperatures computed (Table 5, Column 5). These differences or their running mean, see (9.6), can be plotted against time and then inspected to see if any bias or trend can be detected. Figure 1b suggests that the data before 1975 have a lower mean than after 1975.

9.5 Comparison with Other Parameters

Where a suitable neighbouring station is not available, if relationships exist between the test parameters and other parameters measured at the station, it is sometimes possible to use these to test if the record is homogeneous. Sneyers (1957) demonstrated some successful examples of this method by comparing, e.g. sunshine duration vs. diurnal range of temperature (maximum - minimum) and vs. total cloud amount.

Sometimes relationships involving more than one parameter will prove useful; for example, the minimum air temperature can be estimated from the maximum air temperature and dewpoint measured the previous afternoon using McKenzie's formula:

$$T_{min} = \frac{1}{2} (T_{max} + T_{dew}) - k \quad ,$$

where k depends upon average cloud cover and wind speed during the night.

Air temperature measured at one level will be correlated with air temperature measured at another level, i.e.,

$$\frac{T_1}{T_2} = a \log \frac{h_1}{h_2} \quad ,$$

and this relationship can be improved by allowing "a" to vary with cloud cover, insolation and wind (Pasquill, 1974). A similar method can be used for wind.

Correlations between different parameters measured at the same location are usually less than the correlation of the same parameter measured at near neighbours; consequently this method is generally not as good as the method of 9.4.

9.6 Running Means

A time series x_t, t = 1 to n, can be averaged over a time interval m (m << n) either as overlapping means (t = 1 to m, 2 to m + 1, ...) or as non-overlapping means (t = 1 to m, m + 1 to 2 m, ...) and the resulting smoothed plot examined for significant departures.

This simple form of averaging in which each value has equal weight is not entirely safe. Spurious periodicities of between 2 and 3 times the averaging period 'm' can be introduced by the method, and maxima (or minima) in the smoothed data may not coincide with actual maxima in the original data (Reynolds, 1978). To overcome these problems Burroughs (1978) advises using binomial weights. (The weights are determined from the averaging period m.)

As an example, running means can be computed for the Penkridge annual mean of daily maximum temperatures (Table 5, Column 2). The binomial coefficients for an averaging period of 4 years (m = 4) are 1, 3, 3, 1 and the corresponding weights 1/8, 3/8, 3/8, 1/8 (The denominator is the sum of the coefficients.), e.g., the weighted mean for the first four values, 1962-1965, are

$$1/8 \times 11.62 + 3/8 \times 11.62 + 3/8 \times 12.65 + 1/8 \times 12.17 = 12.07,$$

which should be plotted midway between 1962 and 1965. These running means have been computed (Table 5, Column 4) and then plotted in Figure 1a.

Running means cannot resolve the time of change to less than m except by making further sub-divisions. Also, this method will only remove the higher frequencies.

9.7 Detection of Change by Examining Accumulated Totals

9.7.1 General Principle

A time sequence of a parameter x_t can be summed and the accumulated total

$$\sum_t x_t$$

at time t plotted, arranging the scale of the plot so that the slope is approximately at 45°.

The slope of this plot when t = N is

$$\frac{1}{N}\sum_{1}^{N} x_t = \bar{x} \quad ;$$

however, if the data between t_1 and t_2 are defective by a constant amount (bias) c then the slope is

$$\frac{1}{n}\sum_{t_1}^{t_2} (x_t + c) = \bar{x} + c \quad .$$

If the defect is due to scale, say by a factor f, then the slope of the accumulated plot will be

$$\frac{1}{n}\sum_{t_1}^{t_2} f\, x_t = f\, \bar{x} \; .$$

9.7.2 Cumulative Sums

Buishand (1981), accumulated the sum of the deviation from the mean, i.e.,

$$S_t = \sum_{i=1}^{t} (x_i - \bar{x}) \qquad i = 1, 2 \ldots t \quad .$$

If the mean is stationary, then S_t should fluctuate about zero. However, if there is any bias whereby the mean changes from μ to $\mu + \Delta$ at the point t = m, then before this point x_i will be mostly less than $\bar{x}$ (assuming Δ is positive) and S_t will become increasingly negative. After the point t = m the x_i will tend to be greater than $\bar{x}$ and S_t will become increasingly positive before returning to zero at t = n, i.e.,

$$S_o = 0 \quad \text{and } S_n = 0.$$

The statistic which is normally used to measure any bias in the mean is the <u>greatest range</u> of S_t, i.e.,

$$R = \text{Max } (S_t) - \text{Min } (S_t), \qquad 0 \leqq t \leqq n,$$

or alternatively the <u>standardized range</u>

$$R_s = R/\sigma_x \quad .$$

This method was used on the Penkridge data, Fig. 2, where

$$R = 0.98 - (-2.33) = 3.31$$

$$R_s = 3.31/0.50 = 6.62 \quad .$$

Statistics of R or R_s are not generally available in standard statistical tables, but percentage points of R_s for various values of n can be obtained from Wallis & O'Connell (1973) or Buishand (1982).

The advantage of the method is that it does not require other data; however a significant part of the variation of the residual $x_i - \bar{x}$ will be genuine climatology making it difficult to interpret the plot of S_t against t. This problem can be overcome by making comparisons with an estimate $\hat{x}_i$ determined from one or more neighbours, then either accumulating x_i and plotting this against the accumulation of $\hat{x}_i$ (known as double-mass curve analysis, Kohler, 1949), or better still accumulating the residual $x_i - \hat{x}_i$ against t (Craddock, 1979). The latter method is more sensitive than the double mass curve method, giving a greater change of slope for the same bias.

Applying this method to the Penkridge data, Elmdon data were used as an estimate and the difference E-P accumulated (Table 5, Column 8) and plotted against the year (Fig. 3). The data clearly separate into two sets: up to 1974 the slope or average difference was 0.304, and thereafter it was 0.435.

An example of an error due to scale occurred at Greenock, on the west coast of Scotland, when a service engineer reported on 22 February 1979 that an anemometer had been reading too high. Although the engineer corrected the fault, the time when it had first started was unknown except that it was probably after his previous visit on 25 November 1978. To determine when the fault occurred the wind at Greenock was compared with the wind at the nearest station of Abbottsinch, 25 km inland. Because the area is very hilly only those occasions were used when the airstream was likely to be the same over both stations, i.e., when the wind was in line with both stations, the time was between 1100h and 1500h and the speed greater than 10 kt, when there is likely to be good turbulent mixing.

The ratio of the wind speed at Greenock to the wind speed at Abbottsinch was accumulated and, since the wind did not always satisfy the above conditions, the accumulated total was plotted against successive events (Fig. 4). The plot shows quite clearly that the fault occurred between 6 and 8 December 1978 and confirms that it was corrected on 22 February 1979. The plot also shows that the Greenock anemometer was <u>over-reading by as much as 42%</u>.

These methods cannot indicate which set of data, if any, are correct but will show if and when any inhomogeneity occurred, usually within two or three plotted points. A linear plot will only be obtained if all the major signal has been removed, leaving a residue consisting of a constant factor plus noise.

9.7.3 Maronna & Yohai Method

A good method for detecting both the time and magnitude of change has been developed by Maronna & Yohai (1978). This method is an extension of the double mass curve analysis and has been used successfully by Potter (1981) on rainfall data. The analysis will be made easier if either a small computer or programmable calculator is available. The method requires that each series of data must be distributed normally, be serially independent and be stationary (except for a shift in the mean of the series under test). Assuming that we have a known set of n data values, of which y are to be compared with another set x known to be homogeneous, then the steps are:

(i) Form a new series of cumulative sums

$$X_i = \frac{1}{i} \sum_{j=1}^{i} x_j \qquad\qquad Y_i = \frac{1}{i} \sum_{j=1}^{i} y_j$$

then

$$\bar{X} = X_n \qquad\qquad \bar{Y} = Y_n \quad ;$$

(ii) Compute the variance and covariance sums

$$S_x = \sum_{j=1}^{n} (x_j - \bar{X})^2 \qquad\qquad S_y = \sum_{j=1}^{n} (y_j - \bar{Y})^2$$

$$S_{xy} = \sum_{j=1}^{n} (x_j - \bar{X})(y_j - \bar{Y}) \quad ;$$

(iii) Then for each value in the series compute

$$F_i = S_x - (X_i - \bar{X})^2 \; n \; i/(n-i) \;, \qquad i < n$$

$$D_i = [S_x(\bar{Y} - Y_i) - S_{xy}(\bar{X} - X_i)] \; n/[(n - i)F_i]$$

$$T_i = [i(n - i)D_i^2 \; F_i]/(S_xS_y - S_{xy}^2) \quad ;$$

(iv) Note the value of i when T_i is a maximum - this indicates the year before the mean change. The value of D_i at this point is an estimate of the amount of change.

The results from applying this method to the Penkridge (y) and Elmdon (x) data are shown in Table 6. The maximum value of T_i occurs when i = 13, corresponding to 1975 when the estimated change D_{13} = 0.12. This agrees well with estimates made in section 9.7.2. Critical values of T have been produced by Maronna & Yohai and extended by Potter, and these indicate that T_{max} = 5.6 is not really significant ($P \simeq 0.19$).

9.8 Difference between Successive Values

The low frequency signal (climatology) can be removed by taking the difference between successive values. A method which makes use of this is the von Neumann ratio test:

$$N = \frac{\sum_{t=1}^{n-1} (x_t - x_{t+1})^2}{\sum_{t=1}^{n} (x_t - \bar{x})^2} .$$

It can be shown that for large n,

$$N = 2(1 - r),$$

where r is the lag-one autocorrelation coefficient.

If the mean is stationary and the data serially independent then the expected value of N is 2. However, if the mean increases from $\bar{x}$ to $\bar{x} + \Delta$ at the point t = m, then, since

$$r = \frac{1}{n} \sum_{t=1}^{n} (x_t - \bar{x})(x_{t+1} - \bar{x})/\sigma_x^2 ,$$

before t = m, on average, both x_t and x_{t+1} will be less than the mean $\bar{x}$ and the product $(x_t - \bar{x})(x_{t+1} - \bar{x})$ will be positive. Similarly after t = m, on average, both x_t and x_{t+1} will be greater than $\bar{x}$, again giving a positive product and hence $r > 0$ and $N < 2$. Buishand has shown that under these conditions the expected value of N is

$$\varepsilon(N) \simeq 2/[1 + m(n - m)\Delta^2/n\,\sigma_x^2] ,$$

and for $n \geqq 20$, N is distributed approximately normally with variance

$$\sigma_N = 2\sqrt{\frac{(n - 2)}{(n - 1)(n + 1)}} .$$

The von Neumann ratio was computed for the Penkridge data and gave N = 1.12. Even though n is slightly less than 20 the normal approximation was assumed, i.e.,

$$\sigma_N = 2\sqrt{\frac{(19 - 2)}{(19 - 1)(19 + 1)}} = 0.43 .$$

Therefore, at the 5% level of significance the expected value of N (one tail test) is

$$\varepsilon(N) = 2 - 1.645\sigma_N = 1.285,$$

suggesting that, perhaps, the data are not homogeneous.

If there is a trend in the data, as is the case with the Penkridge data which increased during the period 1962 to 1976 (Fig. 1a), then this also causes the von Neumann ratio to be less than 2. Since this trend was genuine it should be removed by computing the von Neumann ratio on the Elmdon/ Penkridge differences (Table 7). Doing this gives N = 1.61 which is not significant (P = 18%).

9.9 Significance of Change

Once a change has been diagnosed it is then necessary to establish whether the change is significant. The following statistical tests which are described in many texbooks on statistics are designed specifically for this purpose. These tests all require that the data are distributed normally (sometimes referred to as "parametric tests"). When this assumption is not acceptable then it is sometimes possible to make the data normal by some operation such as taking the reciprocal or logarithm of the data. If the data are normal then the methods under "Distribution-Free Tests", Chapter 10, should be used.

9.9.1 Student's t-Test

For two sets of uncorrelated data x_1 and x_2, the Student's t-test can be used to establish the level of significance of the difference between the two sets. The population means μ_1 and μ_2 and the population variances σ_1^2 and σ_2^2 are usually unknown and initially assumed unequal. First compute the sample means and variances $\bar{x}_1$, $\hat{\sigma}_1^2$ and $\bar{x}_2$, $\hat{\sigma}_2^2$. If the variances are not significantly different (established by using the F-ratio test, see Section 9.9.2), as is usually the case, the variances can be pooled as

$$\hat{\sigma}^2 = \frac{(n_1 - 1)\hat{\sigma}_1^2 + (n_2 - 1)\hat{\sigma}_2^2}{n_1 + n_2 - 2}$$

and Student's t computed by

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}\ (1/n_1 + 1/n_2)^{1/2}},$$

with $n_1 + n_2 - 2$ degrees of freedom.

The most common hypothesis is "is $\bar{x}_1$ different from $\bar{x}_2$", i.e. $\bar{x}_1$ may be higher or lower than $\bar{x}_2$, and extreme differences may belong to either tail of the distribution and thus require a "two tailed test" to establish significance.

Example:

Test if the differences between the Elmdon and Penkridge data for the period 1962-1974 are significantly different from those for the period 1975-1980. (Basic statistics for these two samples are given in Table 8.)

Since the sample variances are not significantly different at the 5% level (see Section 9.9.2) the pooled variance is

$$\sigma^2 = \frac{12 \times 0.009476 + 5 \times 0.005910}{13 + 6 - 2} = \frac{0.143262}{17} = 0.008427 \quad .$$

$$\text{Student's } t = \frac{0.435 - 0.304}{[0.008427\ (1/13 + 1/6)]^{1/2}} = \frac{0.131}{0.0453} = 2.89 \quad ,$$

With 17 degrees of freedom, this is significant at the 1% level.

Note 1: If t had been computed for the 18 possible sets 1962 and 1963-80, 1962-63 and 1964-80, etc., and the maximum value used to indicate when and if a significant change had occurred, then a level of significance of 1% is the best that could be obtained in 18 trials and the true probability is 18% (1% times 18 trials) (Reynolds, 1978).

Note 2: If the purpose of the investigation had been to establish directly whether the Penkridge data were significantly different from the Elmdon data, then the above test would not be appropriate because the two sets of data are correlated. The solution in this situation is to test the hypothesis that the difference between the two sets should be zero.

9.9.2 F-Ratio Test

This test provides a means of establishing whether the variances of two populations are significantly different. The ratio F is given by:

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \qquad (\sigma_1^2 > \sigma_2^2) \quad ,$$

where $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are the sample variances computed from n_1 and n_2 values respectively. Since the variances are computed from the sample data the degrees of freedom of the two variances are

$$\nu_1 = n_1 - 1$$
$$\nu_2 = n_2 - 1$$

respectively.

Obviously if the variances are equal F = 1, but due to random sampling the ratio can vary from 1 and the sample variances may still not be significantly different. Statistical tables are available giving the limit of the F-ratio for various probability levels. The test is a one tail test establishing whether $\hat{\sigma}_1^2$ is significantly greater than $\hat{\sigma}_2^2$.

Applying this test to the Penkridge/Elmdon differences, the variances for the two samples are computed (Table 8); then

$$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{0.009476}{0.005910} = 1.60 \quad .$$

From statistical tables at the 5% probability level and degrees of freedom $\nu_1 = 12$ and $\nu_2 = 5$, the ratio F can be as great as 4.68 without there being any significant difference between the sample variances, so we conclude that $\hat{\sigma}_1^2$ is not significantly different from $\hat{\sigma}_2^2$.

For most meteorological data large samples are normally required to establish with confidence whether the variability has altered.

9.9.3 Analysis of Variance

This is a powerful method for testing whether two or more sets of data conform to some model. A simple model which is often used is the "fixed effect model":

$$x_{ij} = L + d_j + z_{ij},$$

where: x_{ij} is the ith value in the jth set.

L is the general level common to all observations

d_j is the departure from L in the jth set and is fixed for all values in the jth set.

z_{ij} is the random error unexplained by the model and is distributed normally with zero mean and variance σ^2, called the residual variance.

The null hypothesis that the poplulation means of each set are equal is the same as saying that the population means are all equal to L and that the d_j are all zero.

For a full treatment of analysis of variance the reader should consult a textbook on statistics. The analysis for this model involves computing three sums of squares:

(i) The total variance with N-1 degrees of freedom:

$$S = \sum^{j} \sum^{n_j} (x_{ij} - \bar{x})^2 = \sum^{j} \sum^{n_j} x_{ij}^2 - \frac{T^2}{N} \quad ,$$

where the jth set contains n_j values and $\sum^{j} n_j = N$, and $\bar{x}$ is the overall mean of all N values, i.e.,

$$N\bar{x} = \sum^{j} \sum^{n_j} x_{ij} = T \quad .$$

(ii) The variation between sets with $\nu_1 = j - 1$ degrees of freedom:

$$S_1 = \sum^{j} n_j(\bar{x}_j - \bar{x})^2 = \sum^{j} \frac{T_j^2}{n_j} - \frac{T^2}{N} \quad ,$$

where $\bar{x}_j = T_j/n_j$, the mean of n_j values in the jth set.

(iii) The residual variance with $\nu_2 = \sum^{j} (n_j - 1) = N - j$ degrees of freedom:

$$S_2 = \sum^{j} \sum^{n_j} (x_{ij} - \bar{x})^2 = \sum^{j} \sum^{n_j} x_{ij}^2 - \sum^{j} \frac{T_j^2}{n_j} \quad .$$

Note that $S = S_1 + S_2$ i.e., the total variability is the sum of the variability of the individual set means about the overall mean plus the variability of the individual values about their own set mean. If there are significant differences between the set means then their variability will exceed the residual variances.

The analysis is normally presented in the form of a table.

Source	Sum Squares	Degrees of Freedom	Mean Squares	F-ratios
Between Sets	S_1	$\nu_1 = j - 1$	$M_1 = \frac{S_1}{\nu_1}$	$F = \frac{M_1}{M_2}$
Within Sets (residual)	S_2	$\nu_2 = N - j$	$M_2 = \frac{S_2}{\nu_2}$	
Total	S	$N - 1$		

Using this method on the Penkridge/Elmdon differences where there are only two sets ($j = 2$), before and after 1975 (use Table 5, column 5, with sums from Table 8):

$$\sum_j \sum_{n_j} x_{ij}^2 = 2.4788 \quad ,$$

$$\frac{T^2}{N} = \frac{6.56}{19} = 2.2649 \quad ,$$

$$\sum_j \frac{T_j^2}{n_j} = \frac{3.95^2}{13} + \frac{2.61^2}{6} = 2.3355 \quad .$$

The analysis of variance table is

Source	Sum Squares	Degrees of Freedom	Mean Squares	F-ratios
Between Set 1 and Set 2	0.0706	$\nu_1 = 2 - 1 = 1$	0.0706	8.40
Within Sets (residual)	0.1433	$\nu_2 = 19 - 2 = 17$	0.0084	
Total	0.2139	18		

With this value of F the differences between the two sets is significant at the 1% probability level.

Analysis of variance using the fixed effect model with only two sets is exactly equivalent to the Student's t test described in Section 9.9.1. The advantage of analysis of variance methods is that they provide powerful techniques for analysing many sets of data and allowing for various factors using a suitably designed model.

9.9.4 Principal Component Analysis

The principal component model proposed by Spackman can be used to estimate a station's value from other stations in the area, i.e.,

$$\hat{x}(s,d) = a_0(d) + a_1(d)\, f_1(s) + a_2(d)\, f_2(s) + \ldots$$

The residual difference between this estimate and the observed value,

$$r(s,d) = x(s,d) - \hat{x}(s,d),$$

is usually fairly small (less than 0.05°C for temperature and 0.5 hr for sunshine). Sudden changes in the general level of the residual would suggest that some inhomogeneity has been introduced into the data.

The residuals of the monthly mean of the daily maximum temperatures have been plotted in Fig. 5a and indicate that Penkridge became colder by approximately 0.2°C relative to other stations in August 1976. This change and the amount agrees well with the simpler method described earlier; however, the methods do not agree on the date of change. This kind of discrepancy is to be expected in variable data, where it is very difficult to be precise about when a change occurred unless there exists an appropriate reason. For Penkridge, a search of the station history file did not reveal any obvious reason for the observed change. However, since the investigation of the minimum temperature, Fig. 5b, also indicated a similar but larger change (this time in May 1975), it is probably safe to conclude that Penkridge became colder relative to neighbouring stations around 1975/76.

9.10 Conclusions and Recommendations

When trying to establish whether data are homogeneous, the conclusion will be much more sound if it has been based on adequate information. For this reason as much relevant data as possible should be analyzed. The comparison with the same parameter measured at a well chosen near neighbour will nearly always be better than the comparison with some other parameter, even if measured at the same station. Similarly, the conclusions made from making comparisons with more than one station, e.g., Spackman's method, should carry more weight than the deduction that could be made from a comparison with only one neighbour station. If one can demonstrate a significant difference by independent tests, e.g., a comparison with near neighbours and a comparison with another parameter, then the resulting conclusion is much more likely to be correct.

10. DISTRIBUTION-FREE TESTS

In recent years a lot of interest has been given to statistical tests which do not make any assumption about the distribution of the data. The methods described here are all rank tests which have the additional advantage of requiring very little calculation.

Like the parametric tests described in Section 9.9, a statistical requirement is that the two sets are uncorrelated. For the Penkridge/Elmdon data this was achieved by examining the difference in value between the two stations (see Section 9.4).

10.1 Ranking

The data are numbered according to their arithmetic size, the lowest value having rank 1 and the highest of N numbers having rank N. If several values have equal rank then each value should be assigned the average rank, e.g., if there are 3 values of rank 12 then the next value will be ranked 15 and the average rank of the 3 ties is (12 + 13 + 14)/3 = 13. The ranks of the Penkridge/Elmdon differences are give in column 6 of Table 5.

10.2 Tests for Change in Mean Value

The tests described in this section have been designed to establish whether there is a difference in mean value (also referred to as bias, "mean slippage" or change in location, level, or central tendency).

10.2.1 Wilcoxon's Test

This is a very easy test to apply, simply being the sum of the ranks r_i in the smaller set

$$W = \sum^{n_1} r_i \quad , \qquad n_1 \leqq n_2 \quad .$$

The expected value of W is

$$\varepsilon(W) = \frac{1}{2} n_1 \ (N + 1) \qquad N = n_1 + n_2 \quad .$$

If W_α is the lower 100α per cent value then the upper 100α per cent value is

$$2\varepsilon(W) - W_\alpha.$$

Tables of W_α are available (Biometrica, 1970) for small samples $n_1 \leqq 25$. For larger samples the distribution of W is approximately normal with variance

$$\sigma^2(W) = \frac{1}{12} n_1 n_2 (N + 1).$$

Applying this test to the Penkridge/Elmdon differences, the sum of the ranks for the 1975-1980 set are $n_1 = 6$ and $n_2 = 13$. Then

$$W = 19 + 18 + 9 + 16 + 15 + 12 = 89,$$

and

$$\varepsilon(W) = 6(19 + 1)/2 = 60.$$

From tables, W_α ($\alpha = 0.01$) = 31. Therefore, at the 1% probability level the upper limit is

$$2\varepsilon(W) - W_\alpha = 2 \times 60 - 31 = 89,$$

i.e., the value of W = 89 computed from the data is significant at the 1% probability level.

10.2.2 Mann-Whitney Form of the Wilcoxon Test

This method is equivalent to the Wilcoxon test with similar power for detecting location differences between two sets but does not require the data to be ranked. Each value in the first sample of n_1 values is compared with each value in the second sample of n_2 values, yielding $n_1 n_2$ comparisons. The number of occasions are counted when the value in the first sample is greater (or lower) than each value in the second sample. The mean count should be $n_1 n_2/2$ and any significant departure from this will indicate bias.

Applying this method to the Penkridge/Elmdon difference, the total number of times each of the 6 values in the 1975/1980 sample are less than the 13 values in the 1962-1974 sample is:

$$U = 0 + 0 + 5 + 1 + 1 + 3 = 10.$$

This value is much less than the expected value of 6 x 13/2 = 39, and tables show this to be significant at the 1% level.

The Wilcoxon test and the Mann-Whitney form of this test are both very efficient tests with power similar to the Student's t-test even with small samples.

10.3 Test for Dispersion Differences

Unfortunately, distribution-free tests for dispersion using ranking methods tend to be sensitive to location differences. A suggested method for overcoming this is to make the set means more nearly equal. This can be achieved by first computing for each set the differences from their own median value and then ranking these joint values before applying David's test, which is less sensitive to location differences. David's statistic is

$$V = \frac{1}{n-1} \sum^{n} (r_i - \bar{r})^2,$$

where r_i is the rank of the ith value in the set of n_1 values and r the mean. The other set contains n_2 values and $N = n_1 + n_2$. The expected value of V is

$$\varepsilon(V) = N(N + 1)/12 \quad ,$$

with variance

$$\sigma^2(V) = n_2\ N(N + 1)[3(N + 1)(n_1 + 1) - N\ n_1]/[360\ n_1(n_1 - 1)] \quad .$$

Applying this method to the Penkridge/Elmdon difference:

Median of set 1 (n_1 = 13) = 0.29

Median of set 2 (n_2 = 6) = 0.43.

The differences from these medians are computed for each set (Table 5, Column 9), then the 19 values are ranked (Table 5, Column 10).

From the first set:

$\bar{r} = 134/13 = 10.31$

$$V = \frac{1}{12} \{(19 - \bar{r})^2 + (3.5 - \bar{r})^2 \ . \ . \ . \ (11.5 - \bar{r})^2\} = 31.65$$

$\varepsilon(V) = (19 \times 20) \div 12 = 31.67$

$\sigma^2(V) = 6 \times 19 \times 20 \times (3 \times 20 \times 14 - 19 \times 13) \div (360 \times 13 \times 12) = 24.07$

Standardized variate is $\dfrac{V - \varepsilon(V)}{\sigma(V)} = \dfrac{31.65 - 31.67}{\sqrt{24.07}} = -\ 0.004.$

Assuming that r is distributed normally this is not significant, i.e., the dispersion of the two sets is not significantly different.

10.4 Tests for Distribution Differences

These tests are concerned with any kinds of differences in the distribution of two sets including differences in location, dispersion, skewness, etc.

10.4.1 Kolmogorov-Smirnov Test

For this test the maximum difference (D) between the relative cumulative frequency of each set is used. To make sure that the maximum difference is obtained the data should be divided into classes as small as possible. Alternatively, if the data have already been ranked, then an easier method is to compute the difference between the accumulations at each rank, i.e., each value from set 1 will add $1/n_1$ to the difference and each value from set 2 will subtract $1/n_2$ from the difference.

For small samples (both n_1 and $n_2 < 25$) Pearson and Hartley provide tables of $n_1 n_2 D$ for various probability levels.

If the sample is larger (n_1 or $n_2 \geqq 25$) an approximation can be used to determine the upper significance level of D. At the 5% probability level, this is

$$1.36\ [(n_1 + n_2)/n_1 n_2]^{1/2} \quad .$$

A more general approximation for any sample size is to use the chi-squared test, where

$$\chi^2 = 4\ D^2\ \frac{n_1 n_2}{n_1 + n_2}$$

with 2 degrees of freedom, but for small samples this approximation is conservative.

Applying the Kolmogorov-Smirnov test to the Penkridge/Elmdon differences (x in Table 5), since the data are ranked the difference between the accumulated relative frequencies will increase by 1/13 for every value from set 1 and decrease by 1/6 for every value from set 2, and the maximum difference can be quickly determined by tabulating the data as follows:

Set 1	.07	.24	.25	.26	.28	.29	.29	.30		.35	.37	.38	.40			.47		
Set 2									.33			.38		.42	.44		.50	.54
Difference: Thirteenths	1	2	3	4	5	6	7	8	8	9	10	11	12	12	12	13	13	13
-Sixths									1	1	1	2	2	3	4	4	5	6
								*										

The maximum difference D is 0.615 at * (8/13 - 0/6), hence

$$n_1 n_2 D = 6 \times 13 \times 0.615 \simeq 48.$$

From tables this is only significant at the 10% level (two tail test). The chi-squared approximation,

$$\chi^2 = 4 \times 0.615^2 \times 6 \times 13/(6 + 13) = 6.21,$$

also suggests that for a two tailed test there is no significant difference between the two distributions. For small samples, D has to be quite large before a significant difference can be detected safely.

10.4.2 The Wald-Wolfowitz Runs Test

This test was orginally used to test for randomness in a binary sequence of digits. To apply this method to the Penkridge/Elmdon differences the joint set of $n_1 + n_2$ values are ranked (Table 5, Column 6) and then labelled A if they belong to the set of n_1 values, or labelled B if they belong to the set of n_2 values. The number of runs (sequences of all A's and all B's) are then counted, i.e.,

Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Set	A	A	A	A	A	A	A	A	B	A	A	B	A	A	B	B	A	B	B
Run				1					2		3	4		5		6	7		8

For small samples (n_1 or $n_2 < 25$) statistical tables should be used. In the above example the number of runs would have to be less than 6 for any significant difference at the 5% probability level. For larger samples (both n_1 and $n_2 > 25$) the distribution of the number of runs is approximately normal and the expected value is

$$\frac{2n_1n_2}{N} + 1 \quad ,$$

with variance

$$\frac{2n_1n_2\ (2n_1n_2 - N)}{N^2(N - 1)}$$

10.5 Conclusions and Recommendations

With both the Kolmogorov-Smirnov test and the Wald-Wolfowitz test it has not been possible to detect any significant differences in the distribution of each of the sets from the Penkridge/Elmdon data, primarily because the size of the samples is so small. The same problem occurs with tests for differences in dispersion. However, the Wilcoxon test and the Mann-Whitney form of it are very powerful at detecting location differences even with small samples.

The tests described in Chapters 9 and 10 provide means of establishing whether inhomogeneities have been introduced into the data. Great care is needed to make sure that the correct conclusions have been made. To ensure this there are several important points worth noting:

(a) The tests must be valid for the circumstances. For example, the Student's-t test requires that the two sets of data should be uncorrelated;

(b) Additional evidence to support the conclusion should be obtained. This can be achieved by carrying out tests with other, independent data, e.g., comparison with a different parameter measured at the same site. Alternatively an examination of the station history file might provide evidence of a site or instrument change;

(c) If the only evidence for a change is the data itself, then one cannot be as certain as when additional independent evidence is available. Those methods which start by assuming that a change, if any, could be anywhere in the data , e.g., von Neumann ratio, Maronna & Yohai or Student's t (note a), indicate the possibility of a change in the Penkridge data (18 possibilities) with a level of significance of only 18%. However, if the test starts with the knowledge that a change occurred in 1975, with this fact gained from independent evidence, then the test, e.g., Student's t or Wilcoxon, is only concerned with measuring the difference between the data before and after 1975 with a level of significance of 1%;

(d) When some operation has been carried out on the data prior to analysis, it is important that the effect of this is properly appreciated, i.e., has some of the high or low frequency been removed? Does the residue from the operation still contain any climatological signal due to exposure or height? etc.? These considerations may help to explain the result obtained;

(e) If one is doubtful about a particular technique, then try it out either on a set of data which is known to be homogeneous or a set of artificially produced data with known statistical properties. The test should be carried out first on data without any bias or trend and then on the data after these have been introduced deliberately. These trials will then provide a good background for making sensible conclusions.

The tests which have just been described provide a means of establishing whether inhomogeneities have been introduced into the data. Used in reverse the methods also provide a means of making the data homogeneous. However, because there is no absolute standard to work with they cannot provide "correct" data. If it is necessary to homogenize the data for some particular purpose, the original archive data should not be destroyed and replaced by these homogenized data, as future work may produce superior methods which will need the original data.

References

Brooks, C.E.P. and Carruthers, N., 1953. Handbook of Statistical Methods in Meteorology, London, HMSO.

Bryant, G.W., 1979. Archiving and quality control of climatological data. Met. Mag, 108, p. 309.

Bryant, G.W., 1979. Quality control of anemograph data: Met Mag, 108, p. 260.

Buishand, T.A., 1982. The analysis of homogeneity of long-term rainfall records in the Netherlands. Wetenschappelijk Rapport 81-7, De Bilt.

Buishand, T.A., 1982. Some methods of testing the homogeneity of rainfall records.

Burroughs, W.I., 1978. On running means and meteorological cycles. Weather 33 p. 101.

Buss, D.W., 1977. New quality control procedures for climatological data. Atmospheric Environment Service, Canada.

Craddock, J.M., 1979. Method of comparing annual rainfall records for climatic purposes: Weather 34, p. 332.

David, F.N., 1956. A note on Wilcoxon's and allied tests. Biometrika, 43, p. 485.

Dixon, W.J. and Brown, M.B., (editors), 1977. Biomedical Computer Programs. P series. Univ. of Calif. Press.

Done, A.L., 1980. Monthly mean minimum temperature 1959-1977: Coefficients of a set of factors and station residuals. Meteorological Office Internal Report, Bracknell.

Essenwanger, D.M., 1970. Analytical procedures for the quality control of meteorological data. Boston, Mass. Amer. Met. Soc. Met. Monograph, Vol. 11, No. 33.

Fillipov, V.V., 1968. Quality control of meteorologival data. WMO World Weather Watch Planning Report No. 26.

Froberg, C.E., 1966. Introduction to numerical analysis. Addison-Wesley.

Hatch, D.J., 1976. The estimation of local rainfall by simple gauging. Weather, 31, p. 79.

Heino, R., 1972. Climatological Data processing in Finland. Tech. Report No. 5, Finnish Meteor. Institute.

Hoflich, O., Meissner, H.H. and Hoffman, L., 1975. Description of a computer program for checking marine meteorological observations of merchant vessels. Deutscher Wetterdienst, Hamburg.

Kohler, M.A., 1949. Double mass analysis for testing the consistency of records and for making adjustments. Bull. Amer. Meteor. Soc. p. 188

Kolmogorov, A., 1941. Confidence limits for an unknown distribution function. Ann. Math. Statist., 10, p. 275.

Mann, H.B. and Witney, D.R., 1947. On a test whether one of two random variables is statistically larger than the other. Ann. Math. Statist., 18, p. 50.

McKenzie, F., 1943. A method of estimating night minimum temperature. SDTM No. 68, Meteorological Office, Bracknell.

Maronna, R. and Yohai, V.J., 1978. A bivariate test for the detection of a systematic change in Mean. J. Amer. Statis. Assoc. 73 p. 640.

Parker, D.E., 1976. Quality control of the GATE synoptic scale subprogram data centre. Final data set surface data. Meteorological Office Internal Report, Bracknell.

Pasquill, F., 1974. Atmospheric diffusion. Ellis Harwood Ltd.

Pearson, E.S. and Hartley, H.O. (Editors), 1970. Biometrika Tables for Statisticians. Cambridge University Press.

Potter, K.W., 1981. Illustration of a new test for detecting a shift in mean in precipitation series. Mon. Weather. Review, 109, p. 2040.

Reynolds, G., 1978. Two statistical heresies. Weather 33 p. 74.

Sarson, P.B., Variation of the difference between two earth temperature. Scientific Paper No. 14., HMSO, London.

Shearman, R.J., 1975. Computer quality control of daily and monthly rainfall data. Met. Mag. 104, p. 102.

Siegel, S., 1956. Nonparametric statistics for the behaviourial sciences. McGraw-Hill New York.

Sneyers, R., 1957. Sur la determination de l'homogeneite des series climatologiques, J. Scientifique de la Meteorologie VII-28 p. 359.

Sneyers, R., 1972. Sur une methode statistique de detection des observations aberrantes du reseau thermometrique belge. Arch. Met. Geoph. Bickl, Ser. B, 20 p. 269

Sneyers, R., 1975. Sur l'analyse statistique des series d'observations. WMO Technical Note No. 143.

Spackman, E.A., 1979. A multivariate analysis of temperature within the U.K. climatological network. Meteorological Office Internal Report, Bracknell.

Spackman, E.A. and Singleton, F., 1983. Recent developments in the quality control of climatological data. Met. Mag. Vol. III, p. 301.

Thom, H.C.S., 1966. Some methods of climatological analysis. WMO Technical Note No. 181.

Wald, A., 1950. Statistical decision functions. Wiley, New York.

Wallis, J.R. and O'Connell, P.E., 1973. Firm reservoir yield - How reliable are historic hydrological records? Hydrol. Sci. Bull. XVIII p. 347.

Walther, S. and Elbashon, D., 1974. Routine climatological verification practice in Israel. Israel Meteorological Service, Series C, No. 30.

Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics, 1, p. 80.

Wilcoxon, F., 1947. Probability tables for individual comparisons by ranking methods. Biometrics, 3, p. 119.

Wilcoxon, F., 1966. Climatic change. Technical Note No. 79. World Meteorological Organization, Geneva.

Table 1. Mean deviation, standard deviation and expected range of the change to some meteorological parameters after t hours.

Hourly Parameters		t = 1 hour			t = 3 hours			Daily Parameters		t = 1 day		
		$\lvert x \rvert$	σ	e	$\lvert x \rvert$	σ	e			$\lvert x \rvert$	σ	e
Dry bulb temperature	°C	0.6	0.8	2.0	1.2	1.7	4.5	Maximum temperature	°C	1.9	2.5	6.5
Wet bulb temperature	°C	0.4	0.6	1.5	0.9	1.3	3.5	Minimum temperature	°C	2.3	3.0	7.5
Dew point temperature	°C	0.6	0.9	2.3	1.0	1.5	4.0	Grass minimum temperature	°C	3.3	4.2	11.0
Pressure	mb	0.4	0.5	1.3	1.1	1.5	4.0	Concrete minimum temperature	°C	2.8	3.5	9.0
Wind direction (10 min)	°True	10.0	15	40	17	28	70	Earth temperature depth 10 cm	°C	1.2	1.5	4.0
Wind speed (10 min)	Kn	2.0	2.5	7	3.5	4.5	12	Earth temperature depth 10 cm	°C	0.7	1.0	2.5
Mean direction (1 hour)	°True	7.5	12.5	30				Earth temperature depth 30 cm	°C	0.3	0.4	1.0
Mean speed (1 hour)	Kn	1.5	1.9	5				Earth temperature depth 50 cm	°C	0.2	0.25	0.7
								Earth temperature depth 100 cm	°C	0.1	0.15	0.4
								Mean wind speed (24 hour)	Kn	3	3.5	9.0

$$\bar{x} = \frac{1}{n}\sum(x_t - x_0) \qquad |x| = \frac{1}{n}\sum|x_t - x_0| \qquad \sigma^2 = \frac{1}{n-1}\left[\sum(x_t - x_0)^2 - n\bar{x}^2\right]$$

Table 2. Mean deviation, standard deviation and expected range of the difference between the observed value and that estimated by Linear Interpolation from values t hours before and after.

Hourly Parameters		t = 1 hour			t = 3 hours			Daily Parameters		t = 1 day		
		$\lvert x \rvert$	σ	e	$\lvert x \rvert$	σ	e			$\lvert x \rvert$	σ	e
Dry bulb temperature	°C	0.3	0.4	1.0	0.8	1.1	2.8	Maximum temperature	°C	1.5	1.9	5.0
Wet bulb temperature	°C	0.2	0.4	1.0	0.6	0.8	2.1	Minimum temperature	°C	1.8	2.3	6.0
Dew point temperature	°C	0.5	0.7	1.8	0.7	1.0	2.6	Grass minimum temperature	°C	2.8	3.5	9.0
Pressure	mb	0.2	0.2	0.5	0.4	0.6	1.5	Concrete minimum temperature	°C	2.3	2.9	7.5
Wind direction (10 min)	°True	12	30	75	25	45	115	Earth temperature depth 10 cm	°C	0.9	1.2	3.0
Wind speed (10 min)	Kn	1.5	2.0	5.0	2	3	8	Earth temperature depth 20 cm	°C	0.5	0.7	2.0
Mean direction (1 hour)	°True	10	27	70	-	-	-	Earth temperature depth 30 cm	°C	0.2	0.25	0.7
Mean speed (1 hour)	Kn	1.0	1.3	3.5	-	-	-	Earth temperature depth 50 cm	°C	0.1	0.15	0.4
								Earth temperature depth 100 cm	°C	0.05	0.08	0.2
								Mean wind (24 hour)	Kn	2	2.8	7

Table 3 Aitkin Interpolation from values t hours before and after

Parameter	t = 3 hours			t = 6 hours		
	$\lvert x \rvert$	σ	e	$\lvert x \rvert$	σ	e
Dry bulb temperature °C	0.7	1.0	2.6	1.3	1.8	4.6
Wet bulb temperature °C	0.5	0.7	1.8	1.0	1.4	3.6
Dew point temperature °C	0.9	1.2	3.1	1.1	1.5	3.9
Pressure mb	0.3	0.5	1.3	0.8	1.2	3.1

See Table 1 for a definition of $\lvert x \rvert$, σ and e

Table 4 Mean deviation $|x|$, standard deviation σ and expected range e of the difference between the observed value and that estimated from 9 other neighbouring stations.

District Area (km) Density (stations per (100 km)^2)	NW London 75 x 70 19.05			SE England 150 x 140 4.76			S England 240 x 310 1.34		
Parameter	$\|x\|$	σ	e	$\|x\|$	σ	e	$\|x\|$	σ	e
Maximum temperature °C	0.4	0.5	1.3	0.5	0.6	1.5	0.6	0.8	2.1
Minimum temperature °C	0.6	0.8	2.1	0.8	1.0	2.6	0.9	1.1	2.8
Grass minimum temperature °C	1.8	2.3	5.9	2.0	2.5	6.4	2.2	2.7	7.0
Mean wind °C	1.0	1.2	3.1	1.0	1.3	3.3	1.1	1.4	3.6
Earth temperature depth 30 cm °C	0.2	0.3	0.8	0.2	0.3	0.8	0.2	0.3	0.8
Earth temperature depth 100 cm °C	0.2	0.2	0.5	0.2	0.2	0.5	0.2	0.2	0.5

Table 5 Mean Maximum Temperature at Penkridge and Elmdon and derived statistics

Year	Mean Maximum °C Penkridge P	Elmdon E	Running Mean of P	Difference (E-P) x	Rank of x r	Running Mean of x	Accumulated sum of x $\sum x$	Difference from set median m x-m	Rank of (x-m)
1962	11.62	12.09		0.47	17		0.47	0.18	19
1963	11.62	11.86		0.24	2		0.71	-0.05	3.5
			12.07			0.345			
1964	12.65	13.05		0.40	14		1.11	0.11	17.5
			12.33			0.351			
1965	12.17	12.54		0.37	11		1.48	0.08	15
			12.50			0.322			
1966	12.58	12.84		0.26	4		1.74	-0.03	6
			12.71			0.261			
1967	13.06	13.35		0.29	7		2.03	0	9.5
			12.75			0.215			
1968	12.59	12.66		0.07	1		2.10	-0.22	1
			12.66			0.249			
1969	12.48	12.86		0.38	13		2.48	0.09	16
			12.77			0.319			
1970	12.99	13.34		0.35	10		2.83	0.06	13
			12.90			0.319			
1971	13.16	13.45		0.29	6		3.12	0	9.5
			12.81			0.281			
1972	12.28	12.53		0.25	3		3.37	-0.04	5
			12.78			0.272			
1973	13.13	13.41		0.28	5		3.65	-0.01	7.5
			12.95			0.316			
1974	12.84	13.14		0.30	8		3.95	0.01	11.5
			13.14			0.412			
1975	13.38	13.92		0.54	19		4.49	0.11	17.5
			13.18			0.469			
1976	13.35	13.85		0.50	18		4.99	0.07	14
			12.87			0.434			
1977	12.39	12.72		0.33	9		5.32	-0.10	2
			12.47			0.404			
1978	12.39	12.83		0.44	16		5.76	0.01	11.5
			12.29			0.411			
1979	12.07	12.49		0.42	15		6.18	-0.01	7.5
1980	12.55	12.93		0.38	12		6.56	-0.05	3.5

Table 6 Accumulation of Elmdon (x_i) and Penkridge (y_i) data required for the Maronna and Yohai method.

YEAR	i	x_i	y_i	X_i	Y_i	F_i	D_i	T_i
1962	1	12.09	11.62	12.090	11.620	4.462	-0.220	3.850
1963	2	11.86	11.62	11.975	11.620	3.143	-0.131	1.802
1964	3	13.05	12.65	12.333	11.963	3.913	-0.100	1.842
1965	4	12.54	12.17	12.385	12.015	3.664	-0.108	2.517
1966	5	12.84	12.58	12.476	12.128	3.764	-0.060	0.938
1967	6	13.35	13.06	12.622	12.283	4.336	-0.023	0.176
1968	7	12.66	12.59	12.627	12.327	4.140	0.051	0.904
1969	8	12.86	12.48	12.656	12.346	4.112	0.038	0.527
1970	9	13.34	12.99	12.732	12.418	4.486	0.039	0.619
1971	10	13.45	13.16	12.804	12.492	4.834	0.056	1.372
1972	11	12.53	12.28	12.779	12.473	4.548	0.079	2.448
1973	12	13.41	13.13	12.832	12.527	4.842	0.100	4.064
1974	13	13.14	12.84	12.855	12.552	4.930	0.121	5.590
1975	14	13.92	13.38	12.931	12.611	5.211	0.091	3.016
1976	15	13.85	13.35	12.993	12.660	5.027	0.078	1.839
1977	16	12.72	12.39	12.976	12.643	5.096	0.097	2.296
1978	17	12.83	12.39	12.967	12.628	5.106	0.079	1.075
1979	18	12.49	12.07	12.941	12.597	5.224	0.037	0.130
1980	19	12.93	12.55	12.940	12.595			

Table 7 Tabulation for Von Neumann Ratio on Elmdon/Penkridge difference

	Year	Difference (E-P) x	Successive Difference $x_t - x_{t+1}$	Departure from mean $x_t - \bar{x}$
	1962	0.47		0.1247
			0.23	
	1963	0.24		-0.1053
			-0.16	
	1964	0.40		0.0547
			0.03	
	1965	0.37		0.0247
			0.11	
	1966	0.26		-0.0853
			-0.03	
	1967	0.29		-0.0553
			0.22	
	1968	0.07		-0.2753
			-0.31	
	1969	0.38		0.0347
			0.03	
	1970	0.35		0.0047
			0.06	
	1971	0.29		-0.0553
			0.04	
	1972	0.25		-0.0953
			-0.03	
	1973	0.28		-0.0653
			-0.02	
	1974	0.30		-0.0453
			-0.24	
	1975	0.54		0.1947
			0.04	
	1976	0.50		0.1547
			0.17	
	1977	0.33		-0.0153
			-0.11	
	1978	0.44		0.0947
			0.02	
	1979	0.42		0.0747
			0.04	
	1980	0.38		0.0347
Sum		6.56	0.09	0.0
Sum Square		0.3453	0.3449	0.2139

Table 8 Basic statistics for the Penkridge/Elmdon differences

	n	Σx	Σx^2	$\bar{x}$	σ^2	Median
Sample 1 1962-1974	13	3.95	1.3139	0.304	.009476	0.29
Sample 2 1975-1980	6	2.61	1.1649	0.435	.005910	0.43
Total 1962-1980	19	6.56	2.4788			0.35

Figure 1a Annual mean of daily maximum temperature at Penkridge (⊙) and the running mean of these data averaged over 4 years (△).

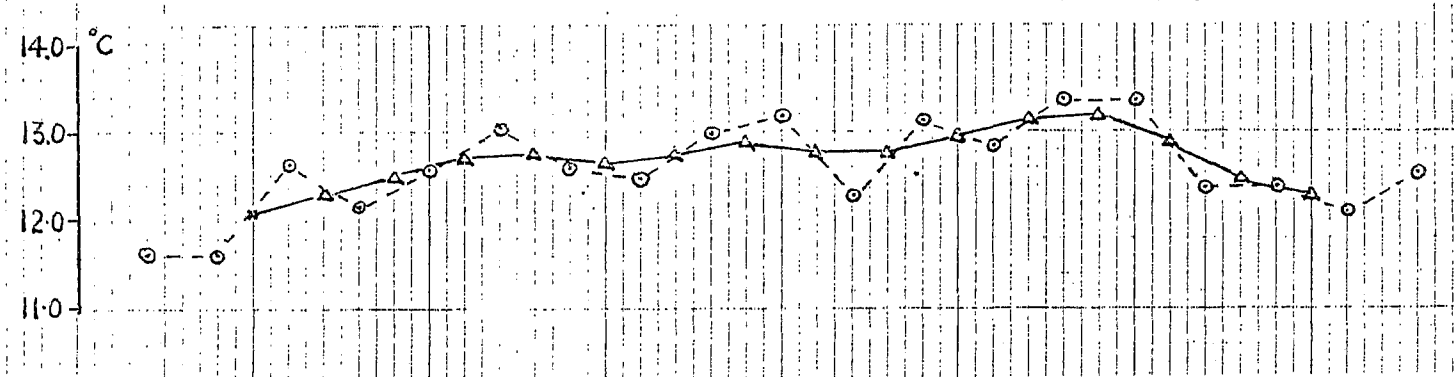
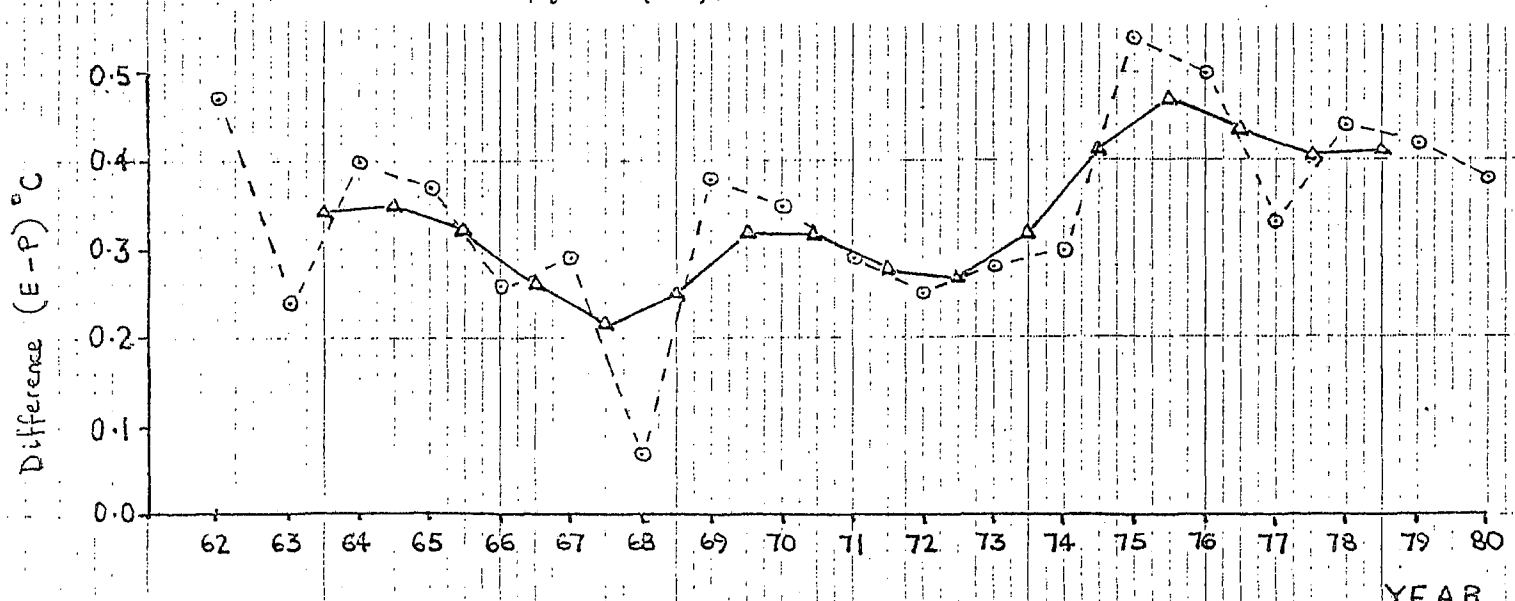


Figure 1b Difference between the annual mean of daily maximum temperature at Elmdon and Penkridge (⊙) and the running mean of these averaged over 4 years (△).



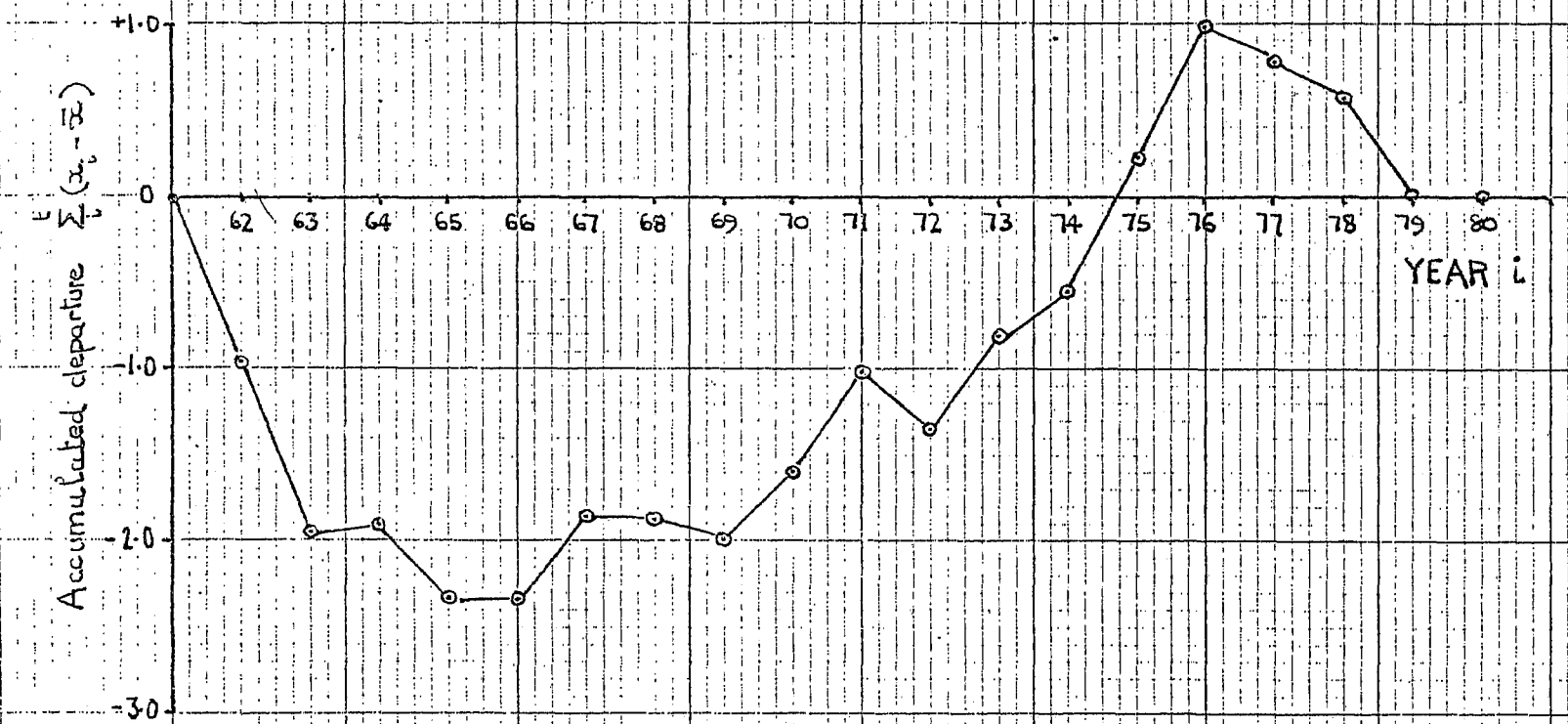


Figure 2 Accumulation of the departure from the mean of maximum temperature at Penkridge 1962-1980.

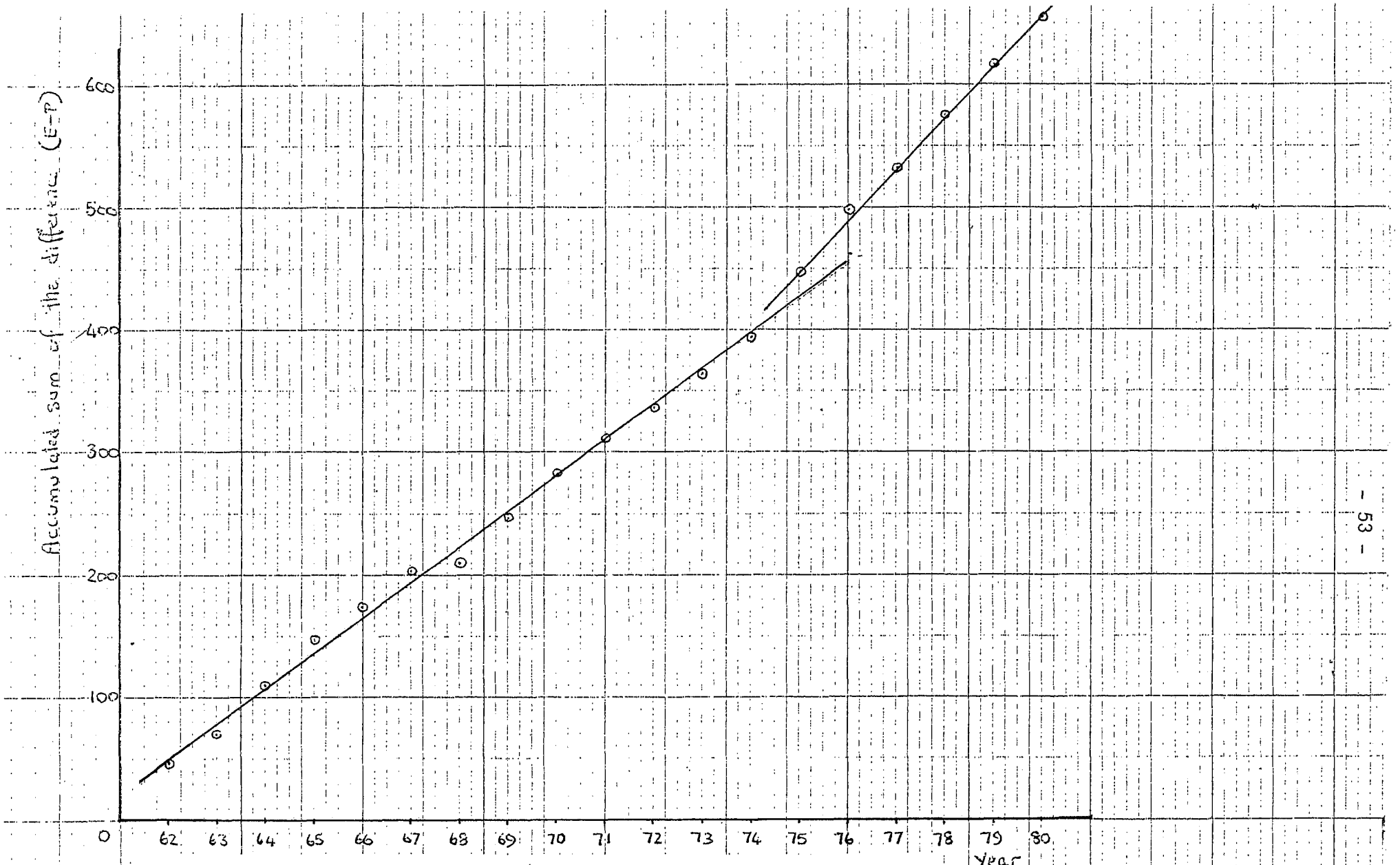


Figure 3. Accumulated differences between the Elmdon and Penkridge mean of daily maximum temperature 1962-1980.

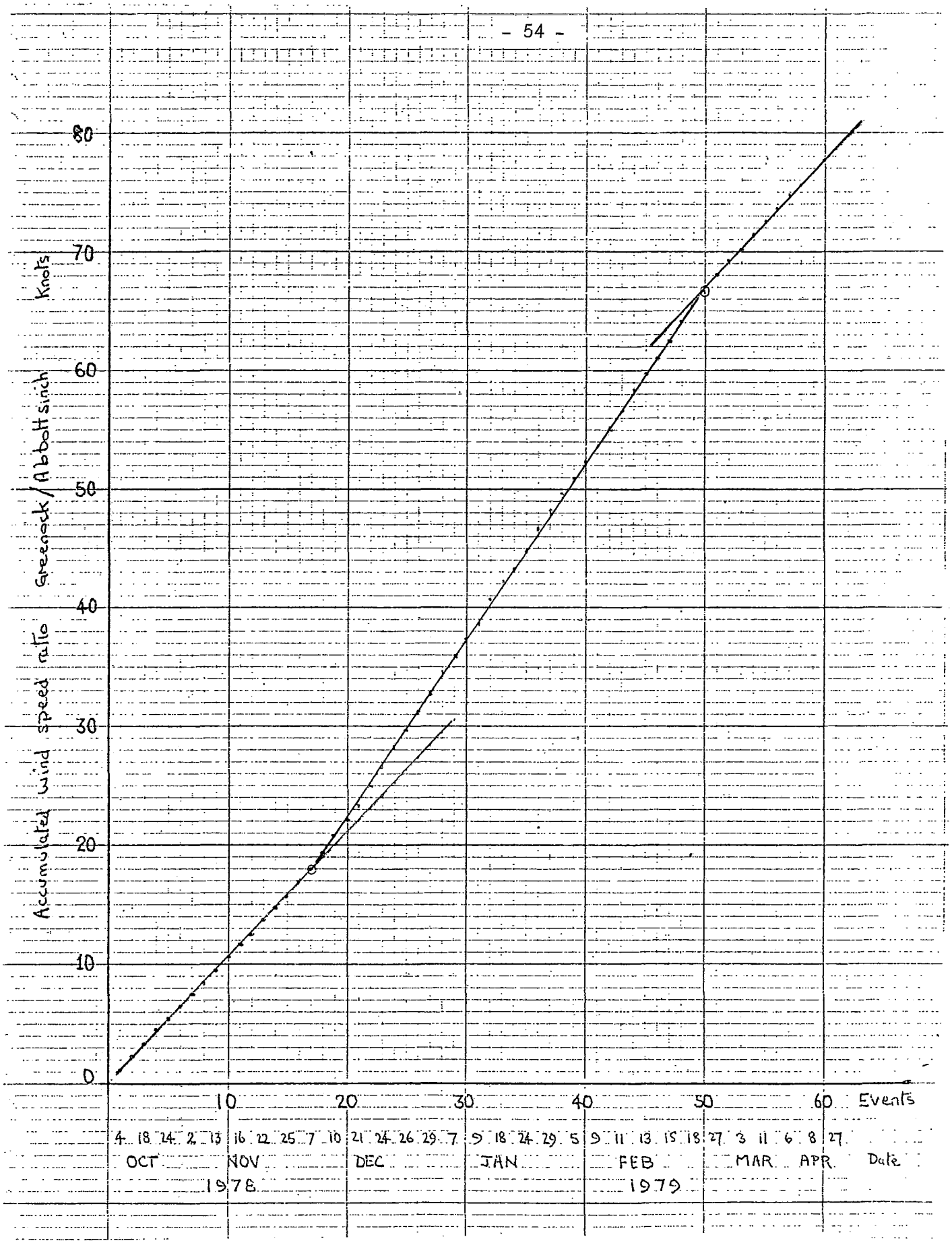


Figure 4. Accumulated ratio of the wind speed at Greenock to the wind speed at Abbottsinch October 1978 to April 1979.

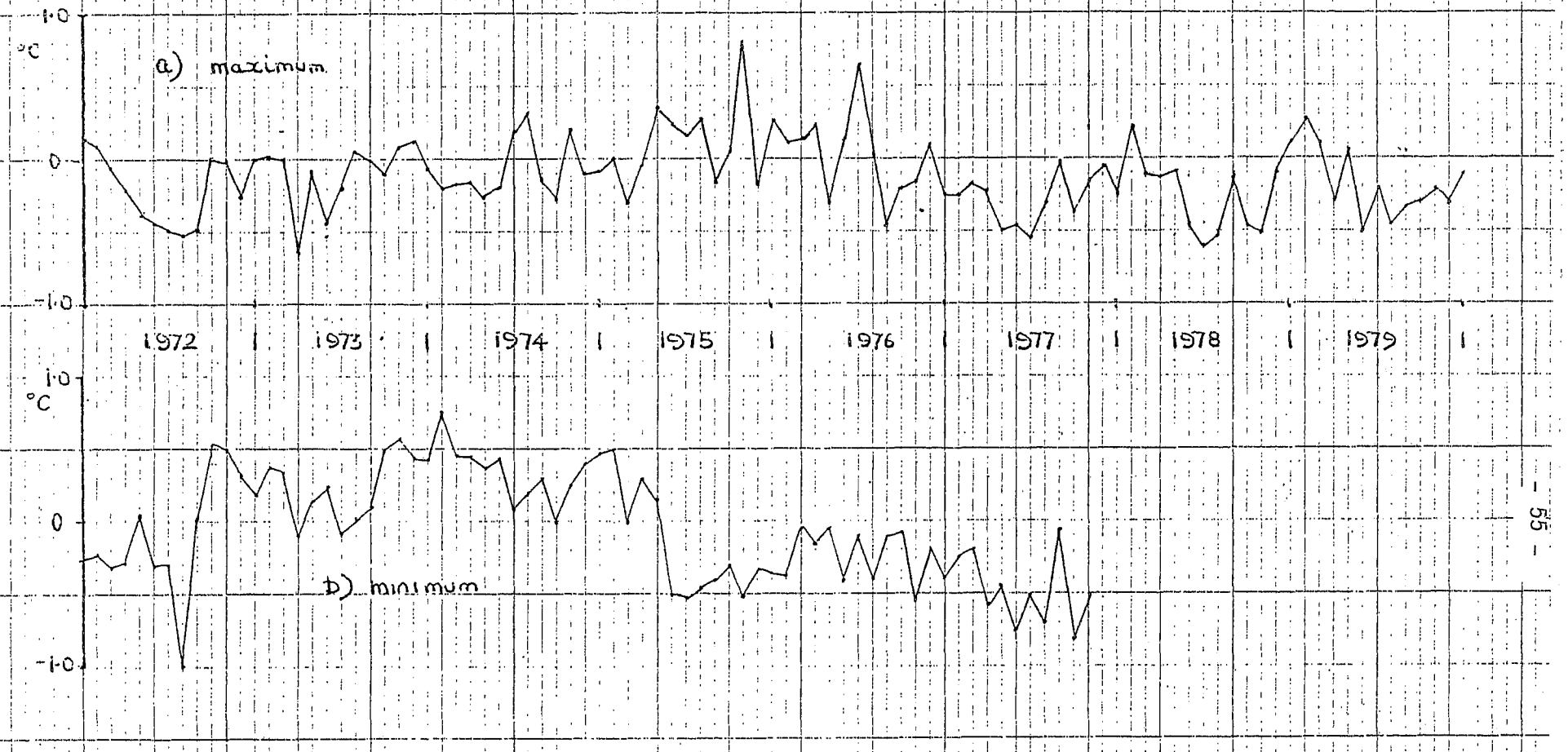


Figure 5. Residual error of the Penkridge monthly mean of daily maximum (a) and minimum (b) temperature estimated using Spackman's Principal Component model.

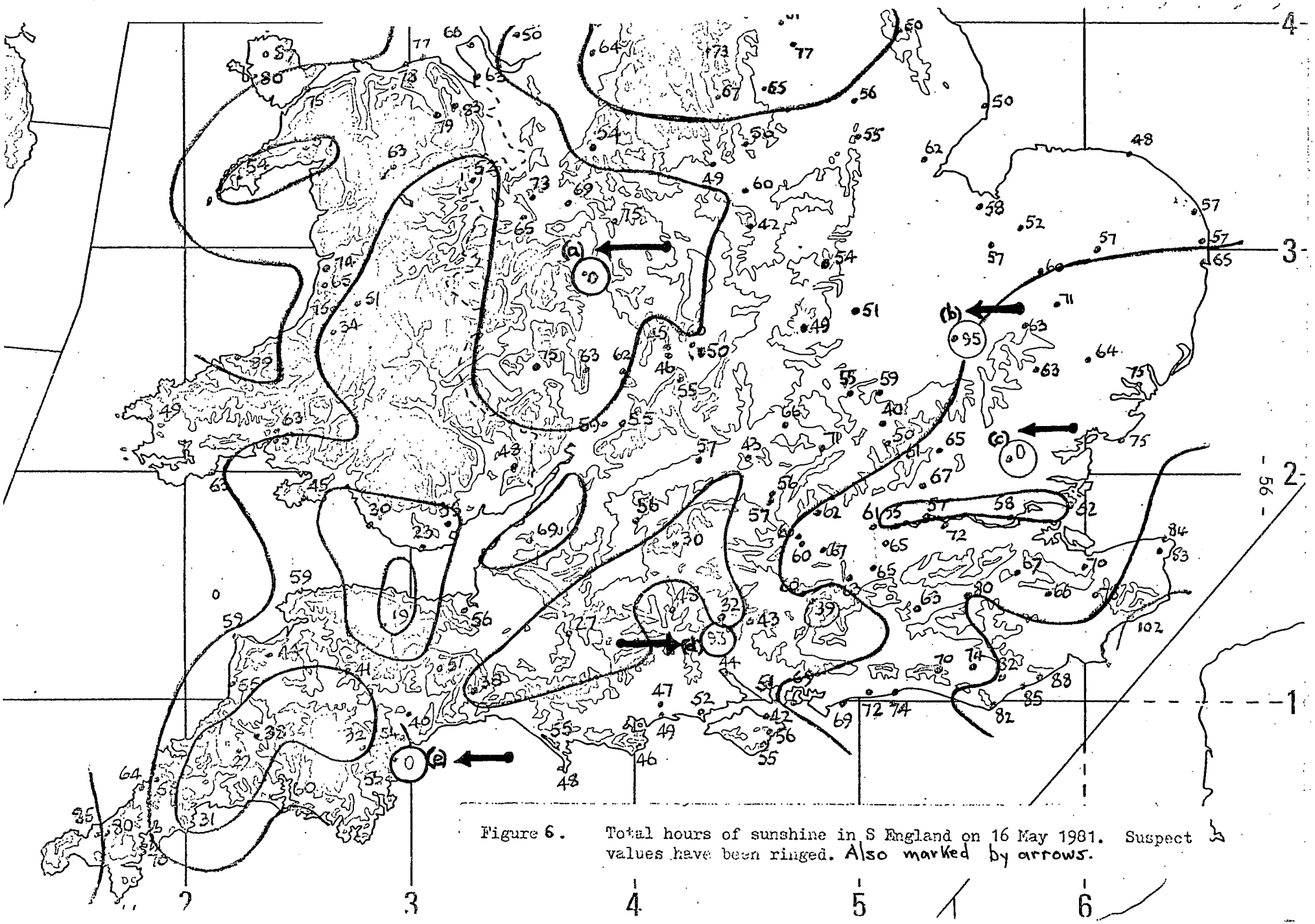


Figure 6. Total hours of sunshine in S England on 16 May 1981. Suspect values have been ringed. Also marked by arrows.

BASIC QUALITY CONTROL CHECKS FOR HOURLY DATA IN THE FM 12-VII SYNOP CODE

A value is queried if any of the following checks are true:

Date/time dd/mm/yy hh

hh hour outside the range 00 to 23
dd day outside the range 1 to 'Number of days in the month'
mm month outside the range 1 to 12

Wind ddff

Units of wind speed and whether the wind is estimated or measured is indicated by i_w

Range check	$0 > dd > 36$ and $dd \neq 99$ (variable)	
	$0 > ff > 35$ m/sec^{-1}	$i_w = 0$ or 1
	$0 > ff > 70$ kn	$i_w = 3$ or 4
check for 'calm'	$dd = 00$ and $ff \neq 00$	
	$dd \neq 00$ and $ff = 00$	
check 'light and Variable'	$dd = 99$ and $ff \geq 4$ m.sec^{-1}	$i_w = 0$ or 1
	$dd = 99$ and $ff \geq 8$ kn	$i_w = 3$ or 4
Blowing snow or sand	$ff < 05$ m.sec^{-1} ($i_w = 0$ or 1) or $ff < 10$ kn ($i_w = 3$ or 4) and $ww = 07$ or 30 to 39	

VV > 0.5 km and ww = 33 to 35, 39, 54, 55, 74 or 75
VV ≥ 1.0 km and ww = 41 to 49, 52, 53, 57, 59, 69, 72, 73, 82, 86 or 94
VV > 4.0 km and ww = 64, 65, 97, 98 or 99
VV > 6.0 km and ww = 40
VV > 7.0 km and ww = 04 to 10, 30, 31, 32, 38, 62, 63, 67, 76, 81, 84, 88, 99 or 92
VV < 1.0 km and ww ≠ 04, 05, 06, 11, 12, 17 or 30 to 99
VV < 5.0 km and ww = 15

<u>Past Weather</u> W_1W_2

W_1 and W_2 both use the same code. Only reported if i_x = 1 or 4.
Range check: W_1 and W_2 not between 0 and 9.

$W_1 < W_2$
W_1 or W_2 = 5 to 8 and RRR=0 (or i_R = 3)
W_1 or W_2 = 0 to 4 or 9 and RRR > 0 (i_R = 1 or 2)
W_1 or W_2 = 7 and TT > 5 during last 3 hours before 0300, 0900, 1500, 2100 GMT
W_1 or W_2 = 7 and TT > 5 during last 6 hours before 0000, 0600, 1200, 1800 GMT

Present Weather ww

Reported in the 7 group provided $i_x = 1$ or 4.

Freezing weather inconsistent with temperature TT

ww = 24, 36 to 39, 48, 49, 56, 57 or 66 to 79	and	$TT \geqslant 5$
ww = 83 to 88, 93 or 94	and	$TT \geqslant 7$
ww = 22, 23 or 26	and	$TT \geqslant 10$
ww = 21, 25, 50 to 55, 58 to 65, 80 to 82, 91 or 92	and	$TT \leqslant -5$

Fog or mist inconsistent with Dew Point depression DP = TT - TD

ww = 42 to 49	and	$DP > 0.5^{\circ}C$
ww = 10 to 12, 40 or 41	and	$DP > 1.0^{\circ}C$

Fog, mist or snow inconsistent with visibility VV

See 'Visibility'

Present weather inconsistent with total cloud cover N

ww = 00 to 16, 18 to 29, 40, 42, 44, 46 or 48	and N = 9
ww = 14 to 17, 19, 50 to 75 or 77 to 99	and N = 0
ww = 43, 45, 47 or 49	and $N \neq 9$

Present weather inconsistent with measured rainfall RR

ww = 20 to 27, 51, 53, 55, 61, 63, 65, 71, 73, 75	and RR = 0
ww = 63, 65, 73, 75	and $RR \leqslant$ Trace

Present weather inconsistent with visibility VV - see 'Visibility'

Total Cloud Cover N

$0 > N > 9$

$N = 0$ and $N_h \neq 0$ or N_h missing

$N = 9$ and $N_h \neq 9$

$N = 9$ and h or hh reported (h = 0 to 9)

$N = 9$ and C_H reported

$N < 8$ and $C_H = 7$

$N = 9$ and $VV \geqslant 1000$ m

$N = 9$ and C_L, C_M and/or C_H reported

$N = 0$ and C_L, C_M and/or C_H reported ($\neq 0$)

$N = 8$ and $C_L = C_M = 0$ and $C_H = 0, 1, 4, 5, 6, 8, 9$

$N < N_H$ Amount of low cloud cannot be greater than the total cover

Checks if significant cloud layers (8 groups) are reported:

$N = 1$ to 8 and significant cloud groups not reported

$N = 9$ and second or third significant cloud layer reported

$N = 9$ and first significant cloud layer not reporting 'sky obscure' or 'vertical visibility'

Amount of low cloud N_h

$0 > N_h > 9$

$N_h = 0$ and $C_L > 0$ or $C_M = 1$ or 2

$N_h \neq 0$ and $C_L = C_M = 0$

$N_h = 0$ and h = 1 to 8

$N_h = 9$ and h = 1 to 8

$N_h = 8$ and $C_L = 1$ to 9 and $C_M = 0$ to 9

$N_h = 9$ and $C_M = 0$ to 9

$N_h = 9$ and $C_L = 1$ to 9

$N_h = 8$ or 9 and $C_H = 0$ to 9

Amount of low cloud N_h continued.

If significant cloud layers (8 groups) are reported

N_h = 1 to 8 and $C \neq 3$ to 9
N_h = 9 and C reported
N_h = 1 to 8 and N_s = 9 in the first 8 group
$N_h < N_s$ in any 8 group with C = 5 to 9

Height of Lowest cloud h

If h is not reported then hh should be reported in a 9 group

$0 > h > 9$ (h can be missing, denoted by /)
$h \leq 8$ and $C_L = 0$
h = 1 to 8 and N = 0 or 9
h = 1 to 8 and N_h = 0 or 9
h not reported and N_h = 1 to 8
h not reported and C_L = 1 to 9

Temperature TT and Dew Point TD

Range check: $-50 > TT > +50$ Worldwide
$-70 > TD > +35$

$TT < TD$

For sequence checks see chapter 5 and Tables 1-3 for comparison with maximum and minimum temperature see Appendix IIa.

Height of Lowest cloud hh

If hh is not reported in a 9 group then h should be reported

$0 > hh > 89$
$hh < 56$ and $C_L = 0$
$hh > 56$ and $C_L = 1$ to 9
$hh > 00$ and $N = 9$
$hh \geqslant 00$ and $N = 0$

These height checks are for temperate regions, for polar and tropical regions see the table at the end of this appendix.

Type of Low Cloud C_L

$0 > C_L > 9$
C_L missing and $N \neq 9$
$C_L = 0$ and C_M missing (obscure)
$C_L = 0$ and $C_M = 0$ and C_H missing
$C_L = C_M = C_H = 0$ and $N_h \neq 0$ or $N \neq 0$
$C_L = 0$ and $C_M \neq 2$ and $h < 7$ (1500 m)
$C_L = C_M = 0$ and $h \leqslant 7$ (1500 m)

If significant cloud layers (8 groups) reported

$C_L > 0$ and first 8 group missing
$C_L = 0$ and $C = 6$ to 9
$C_L > 0$ and $C \neq 6$ to 9 in first 8 group
$C_L = 0$ and $C_M > 0$ and $C \neq 3$ to 5 in first 8 group
$C_L = C_M = 0$ and $C_H > 0$ and $C \neq 0$ to 2 in first 8 group
$C_L = 5$ and $C = 8$
$C_L \neq 3$ or 9 and $C = 9$
$C_L = C_M = C_H = 0$ and at least one 8 group reported

Type of Medium Cloud C_M

$0 > C_M > 9$
C_M missing (obscure) and $N_h \leq 4/8$

If significant cloud layers (8 groups) reported:

C_M = 1 to 9 and first 8 group missing
C_M = 1 to 9 and C = 0 to 2 in first 8 group
C_M = 0 and C = 3 to 5 in any 8 group
C_M = 1 and first medium cloud 8 group has $N_s = 8$ and $C \neq 4$
C_M = 2 and first medium cloud 8 group has $N_s = 8$ and $C \neq 4$ or 5
C_M = 3 to 6 or 8 or 9 and first medium cloud 8 group has $N_s = 8$ and $C \neq 3$
C_M = 0 and $N_s = 9$ in first 8 group

Type of High Cloud C_H

$0 > C_H > 9$
C_H obscure (missing) and $N < 4$ and N_h reported

If significant cloud layers (8 groups) reported:

C_H = 0 and C = 0 to 2 in any 8 group
C_H = 4, 5, 6, 8 or 9 and first high cloud 8 group has $N_s = 8$.
C_H = 7 and 8 group having C = 2 with $N_s < 8$
C_H = 1 to 4 with high cloud 8 groups with $C \neq 0$
C_H = 5 or 6 with high cloud 8 groups with $C \neq 0$ or 2
C_H = 7 or 8 with high cloud 8 group with $C \neq 2$
C_H = 9 with high cloud 8 group with $C \neq 1$

Significant Cloud (8 groups) $8N_sCh_sh_s$

The height checks are for temperate regions, for polar and tropical regions see the table at the end of this section.

The following checks are common to all 8 groups

Check height h_sh_s agree with type C

$C = 1, 2$ and $h_sh_s < 66$ (4800 m)
$C = 3, 4$ and $h_sh_s < 57$ (2100 m)
$C = 5$ and $h_sh_s > 73$ (6900 m)
$C = 6$ to 9 and $h_sh_s > 56$ (1800 m)

Check significant cloud type C agrees with C_L, C_M and C_H

$C = 0, 1, 2$ and $C_H = 0$ or missing (obscured)
$C = 3, 4, 5$ and $C_M = 0$ or missing (obscured)
$C = 6, 7, 8$ and $C_L = 0$ or missing (obscured)
$C = 9$ and $C_L \neq 3$ or 9

Check significant cloud height h_sh_s agrees with C_L, C_M and C_H

$h_sh_s > 73$ (6900 m) and $C_H = 0$
$56 \leqslant h_sh_s \leqslant 73$ (between 1800 m and 6900) and $C_M = 0$
$h_sh_s < 56$ (1800 m) with $N_s \geqslant 8$ in the same 8 group and C_M reported ($C_M \neq 2$)
$h_sh_s < 66$ (4800 m) with $N_s \geqslant 8$ in the same 8 group and C_H reported ($C_H = 1$ to 9)

Check significant cloud type C and amount N_s agree with C_L, C_M, C_H

$C = 2$ and $N_s = 8$ and $C_H \neq 7$
$C = 2$ and $N_s \neq 8$ and $C_H = 7$
$C = 0$ to 2 and C_H not reported
$C = 3$ and C_M not reported
$C = 6$ to 9 and C_L not reported

Check significant cloud amount N_s

$1 \geqslant N_s \geqslant 9$
$N_s > N$
$N_s = 8$ and $N \neq 8$
$N_s = 9$ and $C = 0$ to 9 in the same 8 group
$N_s = 8$ with $C = 3$ to 9 in the same 8 group and $C_H = 0$ to 9
$N_s = 8$ with $C = 6$ to 9 in the same 8 group and $C_M = 0$ to 9
$N_s = 8$ or 9 and subsequent 8 groups occur

Additional checks on 8 groups:

1st 8 group - $8N_1C_1h_1h_1$

$0 > N_1 > 9$
$N_1 \geqslant 8$ and a second or third 8 group reported
$N_1 = 9$ and $C_L = 1$ to 9
$N_1 = 9$ and $C_M = 1$ to 9
$N_1 = 9$ and $C_H = 0$ to 9
$h_1h_1 > 57$ (2100 m) and $N_h > 0$ or $C_L > 0$
$h_1h_1 \leqslant 57$ (2100 m) and $N_h = 0$
$h_1h_1 > 73$ (6900 m) with $N_1 > 0$ and $C_M = 1$ to 9
$N_1 = 8$ and $C_1 \neq 6$ when $C_L = 4$ or 5

$N_1 = 8$ and $C \neq 7$ when $C_L = 6$
$N_1 = 8$ and $C_1 \neq 7$ or 8 when $C_L = 7$
$N_1 = 8$ and $C_1 \neq 6$ or 8 when $C_L = 8$

2nd 8 group - $8N_2C_2h_2h_2$

$N_2 < 3$ except when $C_2 = 9$
$h_2h_2 \leqslant h_1h_1$
$h_2h_2 \leqslant 56$ (1800 m) and $N_2 > N_h$

3rd 8 group - $8N_3C_3h_3h_3$

$N_2 < 5$ except when $C_3 = 9$
$N_3 > N$
N_3 reported and either N_1 or $N_2 = 8$ or 9
$h_3h_3 \leqslant 56$ (1800 m) and $N_3 > N_h$
$h_3h_3 \leqslant h_2h_2$

Height range for different cloud types for each Region.

Region		Low C_L	Middle C_M	High C_H
Polar	ft	0-6500	6500-13000	10000-25000
	km	0-2	2-4	3-8
	h_sh_s	00-56	56-63	60-75
Temperature	ft	0-6500	6500-23000	16500-45000
	km	0-2	2-7	5-13
	h_sh_s	00-56	56-73	66-83
Tropical	ft	0-6500	6500-25000	20000-60000
	km	0-2	2-8	6-18
	h_sh_s	00-56	56-75	70-86

Pressure PPPP or Height hhh

PPPP can be the pressure corrected either to mean sea level (4 group) or to the height of the observing station (3 group) or to some standard height. hhh is the height (geopotential metres) above mean sea level of some standard pressure surface (sually 850 mb, 700 mb and 500 mb, indicated by a_3 in the 4 group).

If a 4 group is reported then check the second digit a_3 to establish whether PPPP or hhh coded:

a_3 = 0 or 9	PPPP reported
a_3 = 1, 5, 7 or 8	hhh reported
a_3 = 2, 3, 4 or 6	invalid

Range check: $930 > PPPP > 1050$

Extreme Limits for Tropical Regions given in Table 7

For station level pressure a tropical climatological normal P_n has been defined by Parker as

$$P_n \text{ (mb)} = 1013.0 - 0.1135h + 4.5 \times 10^{-6} h^2$$

where h is the station elevation in metres.

Check that the tendency p_o (or h_o) agrees with the value P_o (or H_o) now and P_t (or H_t), t hours before

$$\left.\begin{array}{l} |p_o - (P_o - P_t)| > 1 \text{ mb} \\ |h_o - (H_o - H_t)| > 20 \text{ m} \end{array}\right) \quad t = 3 \text{ or } 24 \text{ hours}$$

Table 7 Extreme climatological limits for Pressure and Height in Tropical Regions

Indicator	Parameter		Area A	B	C	D
4 group	Mean sea level pressure (mb)	max	1020	1024	1024	1020
		min	1000	1000	980	995
3 group	Station Level pressure (% of normal pressure P_n)	max	100.7	101.2	101.2	100.7
		min	98.4	98.4	97.0	97.9
	1000 gpm pressure (mb)	max	910	915	915	910
		min	890	890	870	885
	2000 gpm pressure (mb)	max	810	815	815	810
		min	790	790	770	785
	3000 gpm pressure (mb)	max	720	725	725	720
		min	700	700	680	695
	4000 gpm pressure (mb)	max	637	642	642	637
		min	621	621	605	617
4 group						
$a_3 = 1$	1000 mb height (gpm)	max	200	240	240	200
		min	0	0	-200	-50
$a_3 = 8$	850 mb height (gpm)	max	1600	1640	1640	1600
		min	1400	1400	1200	1350
$a_3 = 7$	700 mb height (gpm)	max	3200	3250	3250	3200
		min	3000	3000	2800	2950

This table was taken from Parker, 1976 and was designed to reveal blatant errors in GATE data. The tropical regions are:

A. 10°S - 25°N except areas occupied by regions B, C and D.

B. 15°S - 10°S and a region in the Atlantic 10°N - 25°N between 15°W - 50°W.

C. 10°N - 25°N between 50°W and 180°W.

D. 00° - 25°N between 30°E and 70°E.

BASIC QUALITY CONTROL CHECKS FOR DAILY DATA

A query is raised if any of the following checks are true.

Maximum Temperature TX

Range check: $-40 > TX > +45$ Worldwide

$TX < TN$

$TX > Max(TT) + 3$ Max (TT) is the maximum dry bulb temperature measured at each hour.

$TX > Max(TT) + 5$ Max (TT) is the maximum dry bulb temperature measured at every 3 hours.

Minimum Temperature TN

Range check: $-50 > TN > +30$ Worldwide

$TN > TX$

$TN < Min(TT) - 3$ Min (TT) is the minimum dry bulb temperature measured at each hour.

$TN < Min(TT) - 5$ Min (TT) is the minimum dry bulb temperature measured at every 3 hours.

Grass Minimum Temperature TG

Range check: $-55 > TG > +25$

$TG > TN$ Except when fog or rime.

$TG > TN + 1$ and $ww = 41$ to 49

$TN - TG > 7.5$ Amount of low and/or medium cloud $\leq 2/8$

$TN - TG > 3.0$ Amount of low and/or medium cloud = 8/8

Rainfall Total RRR

Range check $0 > RRR > 250$ mm

The period over which the rainfall total refers is given by t_R. Normally $t_R = 1$ to 4 (6 to 24 hours) and if during this period hourly observations have been made at the station then the following checks against present weather code 'ww' can be made

RRR missing (dry) and ww = 20 to 27 or ww ≥ 50

RRR = 0 not allowed in the code for RRR

RRR > 0 (rain including Trace) and ww = 00 to 19 or 28 to 40

RRR = 001 to 989 and ww ≠ 53, 54, 55, 57, 59, 62 to 65, 67, 69, 72 to 75, 81, 82, 84, 86, 88, 90, 92, 94 to 97, 99

RRR = 010 to 989 and ww ≠ 59, 64, 65, 67, 69, 74, 75, 81, 82, 84, 85, 90, 92 94, 97, 99

Snow Depth sss

sss = 001 to 996 and E' = 0, 1 or 5

sss = 997 to 998 and E' = 3, 4, or 7 to 9

sss reported and E' not reported

State of Ground E (without snow or ice) and E' (with snow or ice)

E reported with E' (E and E' are mutually exclusive)

E reported with sss (snow lying)

E = 0 or 9 and RR > 10 mm

E = 4 or 5 and TN > 5

E = 4 or 5 and TG > 0

E' reported and TN > 5

E' reported and TG > 0

E' = 0, 1 or 5 and sss = 001 to 996

E' = 3, 4 or 7 to 9 and sss = 998

Sunshine Total SSS

SSS < 0 and SSS ≥ Theoretical length of the day appropriate to the month and latitude.

QUALITY CONTROL OF MISCELLANEOUS DAILY PARAMETERS

Apart from the usual parameters dealt with in Appendix IIa some countries also record other parameters and these are discussed below.

Most of the tests are based upon measurements made only in the United Kingdom.

Concrete Minimum TC (minimum temperature on the surface of a 2.5 cm thick concrete (60 x 60 cm) slab measured by a mercury in glass minimum thermometer).

TC < TG + 1	In winter (October to February) an average 2 occasions per month occur when the concrete minimum is slightly less than the grass minimum.
TC > TN + 1	and month April to September (summer)
TC > TN + 2	and month October to March (Winter)

Reports of 'Day of a Weather Event' during the past 24 hours

These events can be indicated by setting the value to 1 if the event occurred or 0 if the event did not occur.

Day of Snow DS

DS = 1 and RR = 0
DS = 1 and TN > 5 (TN = minimum temperature)

If hourly data are also available then check against present weather code 'ww'.

DS = 0 and ww = 22, 23, 26, 68 to 79, 83 to 88, 93 to 95 or 97

Day of hail DH

DH = 1 and RR = 0
DH = 0 and RR > 0

TYPICAL ERRORS FOUND IN SUNSHINE DATA

Recorded by a Campbell-Stokes recorder

1. A value of zero is returned when it is almost certain that there was sunshine. This can arise for several reasons;

   a. Glass sphere has been removed (stolen, broken): Fig. 6 value (c);

   b. Sunshine card lost, stolen or seriously damaged;

   c. Failure to change the card, later substituting a blank card: Fig. 1, value (a);

   d. Incorrect card appropriate to the time of year fitted or instrument not set to the correct latitude so that the sun is not focused on the card.

2. The number of observed hours of sunshine may be reduced due to some obstruction, which may be temporary, e.g., a large truck parked very near the recorder, semi-permanent because the sphere is dirty, (e.g., bird droppings), or permanent because of a high horizon caused by tall buildings, trees or mountains. Reductions due to this last factor can be allowed for by measuring the elevation of the horizon as seen from the sunshine recorder and working out the percentage obstruction. Alternatively if the horizon obstruction is unknown it can be estimated for a particular month from the ratio of the maximum sunshine observed many years to the astronomical length of day. The percentage obstruction is not used to correct the sunshine but used to help make reasonable comparisons with neighbouring stations.

3. Sunshine reduced because the sunshine card has not been correctly centred in the recorder or the axis of the sunshine record is not pointing due south so that the sunshine burn runs off the end of the card.

4. Sunshine reduced because the sunshine recorder is not level. This causes the burn to be inclined at an angle to the axis of the card. If the sunshine is then measured relative to the axis of the card instead of along the length of the burn, it will cause a slight discrepancy.

5. Incorrect apportionment of sunshine. This can occur when the sunshine card is changed at a fixed time during daylight. If the sunshine card is changed at 0900, the sunshine burn before 0900 must be assigned to "tomorrow". Similarly for a sunshine card changed at 1800 the sunshine burn after 1800 must be assigned to "yesterday" (Fig. 6 value (e)).

6. Sunshine total assigned to the wrong day. This can arise due to some clerical error when copying the sunshine values or because the sunshine cards have been muddled. This latter problem can be avoided by dating the cards with an ink which does not run when wet before it is loaded into the recorder. If the cards have been muddled then it is sometimes possible to sort them out by examining the cards and noting the position of the burnt trace which changes from day to day particularly near the equinoxes at high latitudes (Fig. 6, value (b)).

7. Clerical errors - Several types of errors can occur when copying data. The error (d) in Fig. 6 was a transcript error and should have been 39 instead of 93.