LONDON, METEOROLOGICAL OFFICE.

Met.O.3 Technical Note No.3.

A multivariate analysis of temperature within the UK climatological network.
By SPACKMAN,E.A.

London,Met.Off.,Met.O.3 Tech.Note No.3, 1979,31cm.Pp.21,45 pls.18 Refs.Abs.p.1.

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

# MET O 3 TECHNICAL NOTE NO 3

## A Multivariate Analysis of Temperature within the UK Climatological Network

by

E A Spackman

November 1979

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

## Contents

1. <u>Summary</u>

This paper describes the work of the Climatological Research Group of the Climatological Services Branch of the Meteorological Office and presents results from one of the investigations currently in progress. It shows how daily values of a climatological parameter, daily minimum temperature, may be represented by a small number of Empirical Orthogonal Functions (using Factor Analysis) and gives examples of their use for dividing the UK into regions, assessing the quality and representativeness of observations and drawing climatological maps of 30-year averages.

2. <u>The Climatological Research Group</u>

The Group was set up within the Climatological Services Branch in 1975 to study the basic principles determining local climate with the following general terms of reference:

a) To study the factors which influence the several elements which make up the climate of a particular locality (e.g. location, orography etc).

b) To assess quantitatively, using appropriate statistical methods, the representativeness of an observing station and to specify the network of stations required to estimate the climate of a particular locality within specified limits in any part of the United Kingdom. The network may need to include non-standard sites.

c) To define a map of the United Kingdom showing climatological districts each having a relatively homogeneous or readily definable climate.

d) To study statistical methods for the analysis of climatological data and for the presentation of information to users in the most effective way (e.g. extreme values, combined frequencies of two or more elements, duration of specified conditions).

It was decided that to be effective the group should be insulated from day-to-day enquiries. Since the group started in August 1975 it has concentrated on the following topics:

i) Investigation of the errors involved in interpolation (both 'linear' and 'optimum') of temperature, sunshine and strong wind. Most of the work so far has used data from the topographically simple area of East Anglia. The results of studies on maximum and minimum temperature, and sunshine have already been published by Hopkins (1977) and work is continuing on a study using strong winds.

ii) A study to determine the optimum number of years of data on which to base an estimate of mean temperature for the future. This is a so-called 'Optimum Averaging Period' problem. As a result of previous studies WMO currently recommends an averaging period of 30 years and the present work was undertaken to see whether it is possible to give more specific guidance for the United Kingdom. Since there are all too few stations with suitable long-period homogeneous records it was also desirable to provide information on the minimum acceptable averaging period. Smith (1978) has prepared a note of the findings.

iii) An analysis of Variance and Covariance on monthly mean temperature data at a small sample set of 48 stations specifically chosen in an attempt to identify and quantify the effect of altitude on temperature. The results so far, which had to be confined to a limited amount of data (6 years) suggest that a considerably longer period of data is required because of the large variability from year to year and the unrepresentative nature of the period chosen (1972 to 1977). The effect of other variables (e.g. distance from coast, latitude etc) will also be studied.

iv) The determination of simple spatial characteristic patterns of individual climatic variables across the U.K. The principal aim has been to provide patterns which can be used to divide the U.K. into climatic regions. They may also be used for other purposes: to identify physical effects which are important in determining the climate at a point, to assess the probable errors in measurement of climatic variables, to assess the representativeness of values from a particular station, to draw climatological maps etc.

The remainder of this paper will concentrate specifically on the last topic. It begins to provide answers to some aspects of the questions raised in all four of the terms of reference but particularly in (a) and (c).

## 3. Introduction to work on Climatic Regions

In order to begin to study the variation of climate across the United Kingdom in an objective manner it was decided to concentrate on each climatological variable separately and to develop a technique which could fairly easily be used for them all - air temperature, sunshine, rain, humidity, earth temperature etc. The initial aim was to determine for daily minimum temperature a set of regions covering the United Kingdom which each have a relatively homogeneous or readily definable climate.

It was decided to use Multivariate Statistical Methods and this was also an opportunity to gain familiarity in their use in a field in which they have not previously been applied. The programs are now readily available in many statistical packages. The particular set used in the analyses below were the Biomedical Computer Programs (BMDP) which are described by Dixon (1975).

The method adopted was to reduce the data to manageable quantities and obtain a set of characteristic patterns using a Factor Analysis and then to group stations with a Cluster Analysis. Finally a set of Discriminant Functions were obtained, using a Discriminant Analysis, which were evaluated to classify each point of a 10 km grid covering the UK into a specific region.

### 3.1 Techniques

#### 3.1.1 Factor Analysis

Factor Analysis is a technique which aims at reproducing the correlation or covariance matrix of a set of 'variables' measured on many 'cases' from the knowledge of a small number of 'factors'. Thus a formidable volume of data may be reduced to manageable proportions. The factors identify modes of variation in the data which, it is hoped, may also point to the underlying physical causes. The particular form of Factor Analysis used is Principal Component Analysis followed by an orthogonal rotation on the first few components.

The aim of Principal Component Analysis (PCA) is to attempt to assess the structure of variables within a particular set independently of any relationship which they may have to variables outside the set. PCA of a set of m original variables $X_i$ produces m new variables called 'principal components' denoted by $C_i$ where

$$C_i = b_{i1}X_1 + b_{i2}X_2 + \dots + b_{im}X_m$$

The coefficients b are chosen subject to the conditions:

1. Successive components have the largest possible variance.

2. All pairs of components are uncorrelated.

3. The squares of the coefficients involved in any one component sum to unity (Normalisation).

The values of b are found by computing the 'eigenvectors' of the covariance or correlation matrix and the proportion of variance 'accounted for' by each component is derived from the 'eigenvalues'. It is possible to consider just a few of the components and perform a rotation on them by relaxing one or more of the above conditions to obtain what is usually termed a simpler structure. An orthogonal rotation which maximises the variance of loadings within columns of the factor loadings matrix is often used and referred to as a VARIMAX rotation. Within each factor this normally produces only large or small loadings - a structure which usually simplifies the interpretation of the components. The final result may be expressed by

$$X_{ij} = a_{i1}f_{1j} + a_{i2}\,f_{2j} + \ldots\ldots + a_{ip}f_{pj} + r_{ij} \;\ldots\ldots \qquad 1.$$

where $f_{1j}, f_{2j} \ldots f_{pj}$ are factor scores at each station (j)

$a_{i1}, a_{i2} \ldots a_{ip}$ are factor loadings for each factor

$r_{ij}$ are error or residual terms accounting for the unexplained variance. (including error and specific variance)

$X_{ij}$ is the original variable for day i at station j

A decision has to be taken on the number of components that will be retained for rotation. There is no theoretical answer and in this application the following principles were considered:

a) The components rotated should explain a large amount of variance of the original data.

b) Only a small number of factors are desirable for use in subsequent analyses (e.g. clustering, discrimination and regression).

c) The components used should not extend into the region where the explained variance decreases approximately as log variance through to the high order components, as these components explain 'noise' in the observations.

Principal Components Analysis is frequently selected in initial studies of multivariate data and there are now very many examples of its use in the literature. One of the early applications to meteorological data was by Grimmer (1963) who studied monthly mean values of temperature over Europe and the East Atlantic and more recently Craddock and Flood (1969) have applied the method to representation of 500 mb geopotential surfaces. In the particular context of climatological data, Steiner (1965) used mean values of variables for 1931-60 at 67 stations in the USA and White and Lindley (1976) have performed analyses with 18 variables at Moor House CLIMAT station using time series data both of daily values and values meaned over months or seasons. Gregory (1975) has also introduced the idea of Factor Analysis for the study of regional patterns of long term rainfall.

The mathematical solution of Principal Components Analysis (and Factor Analysis) is now discussed in many standard statistical texts (e.g. Harman (1967), Morrison (1967), King (1969) and Kendall (1975)) and will not be repeated here. The BMDP4M program was used to perform the analysis.

### 3.1.2 Cluster Analysis

Clustering is a process of classifying members of a set into groups such that the members of each group have similar characteristics as opposed to the alternative hypothesis that they are unstructured. The process of clustering is discussed in standard texts (e.g. Kendall (1975)) and Everitt (1974) discusses many of the possible algorithms which may be used.

The factor scores at each station are independent variables in p dimensions (when there are p factors). They can be 'clustered' to identify natural groupings of stations. In this work the BMDP2M program was used. It is an agglomerative clustering algorithm in which each station is first regarded as a single group. The 'distance apart' of each group is defined by

$$d_{ij}^2 = \sum_{k=1}^{p} (f_{ik} - f_{jk})^2$$

where $d_{ij}$ is the distance between stations (or groups) i and j in p-dimension factor space

and $f_{ik}$ is the factor score for station i on factor k.

The two nearest stations (or groups) are combined to form one group and values of the factors for the group are computed as the mean over all stations in the group. The process may be repeated until all stations form one group, but may be stopped at any stage when the required degree of detail is achieved and the groups may then be identified.

### 3.1.3 Discriminant Analysis

Given a set of stations which can 'a priori' be divided into groups and a set of variables upon which this division may be made, a set of Discriminant Functions can be determined using a Discriminant Analysis. New stations (or points) can be classified by evaluating the functions using values of the variables at the stations (or points) to determine the probability with which the station or point belongs to the group. Essentially the process defines a set of Discriminant Functions, linear in the variables, which are used to classify the stations into groups. The Functions are derived so that the pooled within-group variance is minimised with respect to the between-group variance. The process is described in standard textbooks such as Kendall (1975) and the BMDP7M program was used. This program also evaluates the discriminant functions and gives: a) the probability with which each station belongs to each group and b) a standardised distance (the Mahalanobis Distance) in the factor space which is the distance of each station to the centroid of each group. This information is provided both for stations used in the discrimination process and others that are not used providing that measurements or estimates are available for each variable.

## 3.2 Data

Daily values of minimum temperature, nominally measured at 0900 GMT for 1973 to 1977, were used in the analysis. Figures 1(a) and 1(b) show the location of most of the stations from which the data were used; some stations are omitted from the figure as otherwise they would obscure information already plotted. Figures 2(a) and 2(b) give the percentage of observations that were available at each station. Most stations provide observations each day and those with fewer than 99% of possible observations either have data missing for substantial periods when instruments were not functioning properly or the station commenced or ceased operation in the period.

Missing values at the stations were estimated by taking the mean of all values in the county for the day and adjusting it by the annual average difference between station value and county value. Estimates were only made for stations at which there were at least 330 observations (300 for 1977) in the year. After the estimation of missing values there were about 670 stations each year with a complete set of daily values.

In the Factor Analysis values were used at 3-day intervals in order to minimise the effects of serial correlation and to provide a set of data for which the dependence between the variables has been minimised. Also the actual values used were anomalies of the temperature at each station with respect to the mean temperature over all stations on each day.

## 3.3 Results

### 3.3.1 Characteristic Patterns

Maps of the 'factor scores' of daily minimum temperature are shown in figures 3(a) to 3(o). They have been obtained by performing an initial factor analysis on the covariance matrix using temperature data separately for each year. Figure 4 shows the variance explained by each component prior to rotation for 1977. It can be seen that the first few components explain most of the variance, and beyond component 20 it decreases approximately linearly as log variance. It was decided to rotate 15 components which explained 84.8%. The distribution of variance within each component is similar each year - as shown by the values given in table 1. A final 'factor analysis' was performed using each of the 15 initial factors obtained for each of the years 1973, 1974, ... 1977. A summary of the variance explained is also given in table 1 in the column headed '1973/77'. This time the first 13 components each explain a very similar amount of variance suggesting that the factors are generally common to all years. Again it was decided to retain 15 components for rotation.

Only 469 stations could be used in the final factor analysis (the rest were missing too much data in at least one year for temperatures to be estimated). Missing factor scores at the stations were estimated by linear regression on the final factor scores for those years for which initial factors had been obtained. Values or estimates of the factor scores were available at 679 stations.

The maps of the factor scores were drawn automatically by computer. To do this values were first interpolated from station points to points on a 10 km grid using a 'single variable analysis' package developed in the Central Forecasting Branch for the 'analysis' of scalar variables. The method of use of the package is described by Hall, revised Forrester and Golding (1978). At each grid point a weighted average of the variable is obtained in a similar fashion to that employed in the operational analysis of geopotential fields as described by Flood (1977). Isopleths of the values at the grid points were then drawn using a microfilm plotter. For technical reasons, the interpolation of factor scores to grid points has been performed separately on the two areas indicated in Figure 5. The isopleths of the final factor scores shown in Figures 3(a) to 3(o) have been subjectively adjusted, where necessary, to match the northern and southern areas.

Several factors were found to correlate well with altitude. Table 2 gives the coefficients of linear regression of the factor scores on altitude for each of the 10 'climatological districts' of the UK and for all stations together. (The climatological districts are shown in figure 5). It was decided that as factors 1, 4, 9, 10, 12 and 13 were significantly correlated with altitude in most districts they should be reduced to sea level so that:

i) Simple maps would be produced with the complex features of topography removed and

ii) detail could be introduced when required based on the known topography over the UK.

The coefficients obtained over all stations together were used for reduction to sea level. Factor 2 was not reduced to sea level as the pattern was clearly dependent on the distance from coast. Its correlation with altitude arises from the fact that altitude and distance from coast are correlated.

In order to show that the detailed topography of the UK is not represented by the distribution of existing climatological stations an analysis of altitudes using only those stations included in the final Factor Analyses is shown in figure 6. This should be compared with figure 7 which is a map of the actual altitude drawn from values on a 10 km grid which have been meaned over a 10 km x 10 km area. Over much of England the general representation of topography is quite good although the detail is poor. Over Scotland, because of the sparse network and unrepresentative location of many stations, even the general representation is poor.

The spatial smoothness of each factor (after reduction to sea level for factors 1, 4, 9, 10, 12 and 13) can be assessed from the root mean square differences between station factor scores and values interpolated by the single variable analysis given in table 3 for those stations with complete data for 1973-77. It can be seen that values range from 0.09 for the smoothest patterns to 0.44 for the roughest. This variation is only partly explained by the scale of variation of each pattern within each factor. The greater part is due to influences that cannot be spatially represented on a 10 km scale (e.g. slope, aspect etc) or to features that are characteristic of individual station sites.

In the maps in figures 3(a) to 3(o) the factor scores (x 10) are drawn with continuous lines for positive values and pecked lines for negative values. The regions of contrasting sign and the shape of the isopleths are important - not the actual sign of the values.
Factor 1 represents an altitude variation superimposed on a latitude variation - a 'latitude' factor. Factor 2 shows a contrast, particularly over England and Wales, between inland and coastal regions. This is a contrast which is particularly marked on radiation nights under clear skies and anticyclonic conditions and may be interpreted as a 'radiation' factor. Factor 3 is broadly a contrast between a) the North of England (in particular the Irish Sea coast) and b) North Scotland and Southern England - an 'Irish Sea' factor. Factor 4 is an altitude variation superimposed on a west south-west to east north-east contrast - a 'longitude' factor. It is also possible to identify specific features of many of the remaining factors: factor 6 shows a contrast across the Pennines, factor 7 is a coastal/inland contrast for North Scotland, factor 9 is specific to East Anglia, factor 10 shows a contrast between the coastal regions of the Moray Firth and Firth of Forth and the Lowlands with the rest of the northern parts of the UK., factor 12 is an east to west contrast over Scotland and factors 13 and 15 seem to explain some of the individual responses of stations - particularly over England and Wales.

There is no reason why the maps should explain the causes of the variation of minimum temperature; they are just one way of describing the variation as 'characteristic patterns'. Nevertheless they do support intuitive ideas and previous studies of the climate of the UK in which latitude, distance from coast, altitude and specific effects of topography have been found to be important.

Preliminary investigations, using solely 1976 data, showed that many of the factors were strongly associated with particular synoptic patterns. This was revealed by evaluating the mean sea level pressure pattern for each factor for those days on which the factor had a high loading.

Factor 1 occurred with westerly flow typical of average conditions and factor 2 under anticyclonic conditions. Factor 3 was important with a strongish north-westerly airflow and factor 4 with northerlies or southerlies (depending on the sign of the factor loading coefficient). The other factors corresponded to the position of the anticyclone or ridge conditions affecting particular regions of the UK.

The factors determined above are one of a set of Empirical Orthogonal Functions; this set gives an optimum representation of the data. The factors may be used in several ways, particularly for regression, but the principal reason for their derivation was to enable a set of Climatic Regions to be objectively defined. Their use for the assessment of the quality and representativeness of observations, and for drawing climatological maps will also be shown

### 3.3.2 Climatic Regions

Climatic Regions applicable to values of daily temperature have many uses including

- selection of groups of stations with similar characteristics for use in areal quality control
- guidance in answering climatological enquiries
- understanding the underlying causes of the variation of temperature
- definition of areas for forecast purposes.

Figure 8 shows the areas currently used for areal quality control. They have been defined using experience gained over many years and have been chosen for use with both minimum and maximum temperatures. Also, each area was drawn so that it contained not more than 14 stations. Some stations are treated as belonging to two areas. The aim of this work is to determine regions objectively and to try to optimise the location of boundaries between regions. The process has been tackled in two steps - the first to find groups of stations and the second to locate the boundaries.

Cluster Analysis was used to group stations. For convenience the analysis was performed separately for the two overlapping areas shown in figures 1(a) and 1(b) and was interrupted at a point where each group could not be sensibly subdivided any further. Figures 9(a) and 9(b) show the results of the clustering. Each station is identified by its group number (0 to 41 in the northern area and 50 to 98 in the southern area) but note that a further 53 stations in the northern area and 78 in the southern area are not shown as they would otherwise obscure values already plotted. Most groups were chosen with at least 5 stations in the group but some exceptions were made for obviously distinct groups such as group 41/98 (Moor House and Widdy Bank Fell in the high Pennines). Some stations do not cluster with any other station and are identified by **.

A set of lines have been drawn which help to identify the groups. These lines were objectively determined by a method described below but at this stage they are not intended to identify stations into their precise groups. It can easily be seen that there would be great difficulty in subjectively determining lines to mark zones of transition from one group to another. Most of the groups are geographically coherent and cluster stations which one would expect to have similar daily climates (such as groups 60 or 61 along the southeast coast, group 76 for London, group 53 for most of Cornwall). One group which is not geographically coherent is group 55 - Grendon Underwood, Medmenham, Easthampstead, Marlborough and Lacock. This is a group of stations often regarded as 'frost hollows'. Santon Downham, in East Anglia, a frost hollow mentioned by Manley (1970) as having almost twice as many air frosts as nearby stations, does not cluster nor do Elmstone and Bastreet which are also frost hollows.

There are many alternative ways of performing a cluster analysis of the stations either by adopting alternative algorithms or by weighting the factors in some way or other. The present approach, which gives equal weights to the factors produces a set of regions which have a similar daily climate. It is also possible to weight each factor in proportion to the variance it explains over a given period. The regions would then describe climatic zones corresponding to that period. This was done in preliminary work using only 8 factors with data for 1976. The result showing the UK broken down into only a few regions is shown in figure 10. In this case the lines have been drawn subjectively to delineate the regions. The thicker lines delineate the country into just 7 regions and distinctly show a coastal region, an inland region of England, a south Scotland/N Ireland region, a Scottish Highland region and a N Scotland coastal region. The regionalisation is not the same when the factors are equally weighted.

The next stage is to classify new stations or points. This was done by evaluating a set of Discriminant Functions determined by performing a Discriminant Analysis on the factor scores with the stations grouped according to the results of the Cluster Analysis. Each point on a 10 km grid covering the UK (but treating north and south areas separately) has thus been classified and figures 11(a) and 11(b) show isopleths of the probability regions for each group. Only 3 isopleths have been drawn for each group - 50% (continuous line), 95% (pecked) and 99.9% (dotted). The 50% lines provide an estimate of the position of the boundaries between groups that could not be drawn using only the knowledge of the group to which each station belongs. The distance apart of the 50% and 95% lines indicates the width of the transition zone from group to group whilst the 99.9% line gives an idea of the coherence of each group. The map gives a picture of groups appropriate to locations which are representative of an area of 10 km x 10 km or so and not to individual locations with unrepresentative characteristics.

The output of the Discriminant Analysis indicated that a few stations had been wrongly classified - these stations were reclassifed. Some other stations for which the probabilities indicated that they might equally lie in more than one region were not used for classification purposes since each station could only be used in one region.

Many of the regions shown in this classification seem to have a close relation with the orography. A North Highland Region (38) is split in two by a Great Glen Region (36) and is more or less surrounded to north and west by the same region. Region 10 approximately defines the Tay Valley whilst there are regions corresponding to the Tweed Valley (15) and The Vale of York (20/58). There are narrow coastal regions around many parts of England and Wales and the high ground of Dartmoor and Exmoor is defined. London is a region (76) as are the Cheshire Plain (62) and Salisbury Plain (74).

The Mahalanobis Distance can be used to identify which parts of the country are a long way (in the factor space) from any of these groups. Figure 12 shows the minimum Mahalanobis Distance (multiplied by ten) to the centre of the nearest group. Larger values indicate when new groups need to be defined or where more stations need to be located so that groups can be defined. The figure, in which pecked values indicate a distance greater than 5, shows that there are problem areas in the north of Cornwall, Somerset, Dorset, parts of Wales and generally over many areas of Scotland. Santon Downham is also strongly marked in this figure and since there is comparatively no shortage of stations in East Anglia it shows how unrepresentative are the values of some of the factor scores at Santon Downham

### 3.3.3 Quality Assessment

This section shows how statistics based on the factor scores may be used to assess the general quality of observations at each station.

The distribution of the residuals from equation 1 may be studied both with respect to day and to station in order to determine how well the factor model fits the observed data on given days or at particular stations. The residuals include error contributions from at least the following sources

- the inadequacy of the factor model in representing the peculiarity of a station site or the temperature variation on a given day.
- instrumental errors
- observer errors
- data processing errors.

Since the model accounts for a large proportion of the variance of the data it is worthwhile studying the residuals to identify stations with large values. All days of 1974 were used in a regression of minimum temperature against the factor scores for 1973/77 using the BMDP6M multivariate linear regression program. In order to use as many stations as possible the estimated values of missing data were used in the regression, but residuals from these values were not used in compiling statistics for each station. The standard deviation of the residuals ranged from 0.5 degC to 1.9 degC with a mean over all 627 stations of 0.97 degC. A similar analysis using maximum temperatures (regressed on factors for 1973/77 determined from maximum temperatures) gives values ranging from 0.4 degC to 1.6 degC with a mean value of 0.79 degC over 598 stations. The distribution of the standard deviations is shown in figure 13; it can be seen that the distribution is skew and that for most stations values lie between 0.6 and 1.4 degC for minima and between 0.5 and 1.1 degC for maxima. Values for stations with small residuals are given in table 4(a) and those with large residuals in table 4(b). In general values are large (or small) for both minima and maxima at the same station except for 5258 KEW, 5592 EASThampstead and 5418 FERNhurst.

The KEW values are from a psychrometric instrument (one of only three stations using this method of measurement in these data) and EASThampstead has the characteristics of a 'frost-hollow' under some (not yet specified) conditions. It can also be seen that several Meteorological Office stations are shown in the list with small residuals but none are in the list with large values.

In order to give a further glimpse into the variation of the residuals two further tables are provided. Table 4(c) gives values for the WMO Reference Stations and table 4(d) for stations in England and Wales considered by the Climatological Services Branch as 'frost-hollows'.

Preliminary investigation of the general distribution of the standard deviation of residuals suggests that:

- residuals in the north tend to be larger than in the south (though this may merely reflect the station density)
- known frost-hollows generally have large residuals ($\gtrsim$1.1 degC)
- Meteorological Office stations tend to have small residuals
- stations with long period records tend also to have smaller residuals

Studies to confirm these possibilities (in particular with respect to the type of station, authority and length of record) are in progress.

Values of the mean residuals (which are also given in tables 4(a) to 4(d)) are generally small (i.e. less than 0.3 degC) though larger values have been found for some stations with incomplete records over the five years and for some relatively isolated places such as The Scilly Islands and stations in the Orkneys and Shetlands.

Some stations with large residuals provide very severe problems during quality control processes (e.g. USK) whilst others have already closed (e.g. HAVErford West, HOUGhall, MACKworth) or are no longer received (e.g. CEINwys, WREXham), and the quality of returns from others leaves much to be desired (e.g. FERNhurst, HAMPton Loade, HARRow Weald). Another problem is that many of those stations with large residuals also have incomplete records both for 1974 and for the period 1973/77.

The above statistics show how well values at a station may be estimated using the observed day to day values from all stations in the network. It is also possible to assess how well stations fit spatially by studying the differences between the observed factor scores and the values interpolated by the single variable analysis. The analysis was tuned to fit observations closely yet retain reasonably smooth patterns; nevertheless,not all factor scores are fitted well. Table 5 gives all stations with residuals greater than 1.0 or five times the root mean square deviation when compared with the residuals over all stations in the area. Some stations are particularly poorly fitted which indicates that they behave very peculiarly under conditions favourable for the factor. Some of these stations are regarded as 'frost-hollows' (see table 4(d)); it can be seen that many do not fit in at least one factor - particularly factor 13.

The LONDon Weather Centre (urban roof site) and CHELtenham (urban site) are other examples of stations which do not fit - in this case because they are too large in factors 3 and 6 respectively. Values of the analysis residuals can clearly be used to identify stations with unusual or un-representative properties worth further investigation.

It is interesting to compare the standard deviation of the station residuals with values of estimates of r.m.s. errors of linear interpolation given by Hopkins (1977) in a study of interpolation errors for East Anglia. Approximate values taken from his paper are given in table 6 for two station spacings representing dense and sparse data areas. Values are certainly of the same general level although Hopkins' values for maximum temperature are a bit smaller. Together these studies suggest that the random variation of daily extreme temperatures measured in a Stevenson Screen may have an r.m.s. variation of about 0.5 degC for minimum temperature and about 0.4 degC for maximum temperature. However, many combinations of site and observer cannot achieve such small values.

All the information found above indicates how well (or badly) data from each station can be fitted and values using recent data should prove very useful in highlighting stations that need further investigation. Large deviations of factor scores from the analysed patterns would indicate stations which are unrepresentative under the conditions when the factor is important, whilst poor statistical values in estimating daily temperatures might indicate poor observing practices.

Quality assessments from this type of work may help to identify which stations need more frequent inspections and which stations may be regarded as not representative of wider areas. It should also be possible, when several years data have been fully processed, to identify stations where changes have affected the homogeneity of records.

3.3.4 Climatological Maps

Climatological maps are usually drawn using long period means of the order of 30 years. Detail on such maps is often generated by using stations with shorter periods and adjusting their values to the long period by comparison with nearby long period stations. It is possible to determine coefficients for the characteristic patterns which are appropriate to a different period from that for which they were derived provided there are stations with sufficient data operating in both periods. The map obtained by combining the factors with these coefficients then has absolute values corresponding to the chosen period and relative detail corresponding to the period used to derive the factors. This can be done because the factors are an efficient and optimum representation of daily values and ipso facto of long period means.

The 1941/70 temperature averages published by the Meteorological Office (1975) were based on data from about 280 stations whereas the characteristic patterns formed above are based on data from 679 stations. Appropriate coefficients for each factor were obtained by linearly regressing monthly mean values of the 1941/70 minimum temperatures on the factor scores for 1973/77. Only 233 stations with values both of 1941/70 averages and 1973/77 factor scores could be used to determine the coefficients and some of these with substantially less than 30 years of data were given reduced weight in the regression. An example, the January map, obtained by this process, is shown in figure 14(a). Values which are representative on a scale of about 10 to 20 km have been reduced to sea level using a lapse rate of 5 degC per 1000 metres. The map may be compared with the corresponding published average map reproduced in figure 14(b). It can be seen that the values, in general, differ by less than 0.5 degC. The fit, from the regression on the 233 stations that could be used, is indicated by the following statistics:

- Standard deviation of residuals: about 0.25 degC
- squared multiple correlation of the observed temperatures with the factor scores: 0.96
- all residuals are approximately normally distributed with values between -0.65 degC and +0.66 degC except for Scilly (1.08 degC)

and Lizard (1.30 degC).

There are, however, several differences in the patterns. The largest occur:

a) over southwest England where the main gradient is over Somerset and Dorset rather than parallel to the coast of Devon,

b) over West Wales where there is virtually no gradient parallel to the coast,

c) over North Wales where the temperature minimum is very much reduced,

d) over Northwest Scotland where there is much less gradient parallel to the coast and

e) over Central Scotland where values are a little warmer.

It may be noted that in producing the official maps

i) There are virtually no inland stations in Devon and Somerset,

ii) The detail along the west coast of Wales is mainly based on 4 shorter period stations all near Aberystwyth of which only Aberystwyth exceeds 1.7 degC.

iii) In West Wales isopleths are drawn to fit Haverfordwest which is both a 'frost-hollow' and on the one year sample of data used above is poorly fitted by the factor score maps.

Comparisons are especially difficult to make because

- in the areas mentioned above there are few stations with a full 30 years data

- in Central and North Scotland many stations are in valleys which are not necessarily typical of a 10 x 10 km local area

- a lapse rate of 5 degC per 1000 metres is assumed for reduction to sea level. Values other than this would give different results.

Nevertheless the objective map is quite reasonable. For the sample shown it does however suggest that there is much less influence on temperature due to distance from the west coast than is shown on the published map.

It is very easy to produce the map corresponding to actual topography. The map corresponding to the topography of figure 7 is shown in figure 14(c). This is very appealing in that it gives a true impression of the actual temperature distribution and could of course be easily redrawn using any suitable scale of topography. The objective charts are a little rough but it would be easy to smooth them mathematically.

As more and more data, both raw and processed are handled by computer, the production of climatological maps by entirely objective techniques needs to be investigated so that the best products can be obtained which make the best use of all available data. Objective techniques offer several advantages:

- maps can easily be produced at any scale with the appropriate spatial smoothing of altitude

- the maps may be corrected by any specified lapse rate

- values can be supplied in computer compatible form

- the methods of analysis can take into account the quality of observations at each station

- the shape and detail of maps may be appropriate to a short period with a lot of data whilst the absolute values correspond to a much longer period of data from many fewer stations.

The main disadvantage is that patterns of the long period may be poorly reproduced if the short period is anomalous though the general level of temperature is unaffected.

If further tests on such techniques as these are successful it might be possible to draw maps of parameters which otherwise prove very difficult due to the small number of actual observations. For instance, the number of hours with temperature above freezing may be well represented by factors of maximum and minimum temperature. Such a chart could then be drawn with the detail of a network of about 650 stations although hourly temperatures are measured at fewer than 150 stations.

## 4. Conclusion

This paper has shown how it is possible to use Multivariate Statistical Methods to derive characteristic patterns of climatological variables across the UK and to use them to divide objectively the country into regions. It has also shown how the methods may give statistics which provide answers to questions which were hitherto almost unanswerable - for instance about the errors and representativeness of observations. Finally the characteristic patterns were used to produce objective versions of climatological maps which are usually produced entirely subjectively.

## Acknowledgements

The work in this paper has made considerable use of many existing computer routines. W Mills has been responsible for obtaining the data from climatological archive data sets and also for using standard Meteorological Office software in a suitable form for the production of microfilm from which many of the charts in this paper have been produced directly. B Golding (Central Forecasting Branch) has given advice on using the Single Variable Analysis program and has also supplied specialised versions of the isopleth drawing routines. The work owes much to the initial investigations of J Hopkins and would not have been attempted without a statistical package such as the BMDP programs.

## References


CRADDOCK, J.M. (1965). The Statistician, 15, No. 2, p. 143. A Meteorological Application of Principal Component Analysis.

CRADDOCK, J.M. and FLOOD, C.R. (1969). Quart.J.R. Met.S. 95, p.576. Eigenvectors for representing the 500 mb geopotential surface over the northern hemisphere.

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

EVERITT, B. (1974). Heinemann. Cluster Analysis.

FLOOD, C.R. (1977). Meteorological Office Met O 2b Tech Note No. 41. The Meteorological Office operational objective analysis system.

GREGORY, S. (1975). Weather, 30, No. 9, p.276. On the delimitation of regional patterns of recent climatic fluctuations.

GRIMMER, M. (1963). Quart.J.R.Met.S. 89, p.395. The space filtering of monthly surface temperature anomaly data in terms of pattern using empirical orthogonal functions.

HALL, D.C. revised FORRESTER, D.A. and GOLDING, B. (1978). Meteorological Office, Met O 2b Tech Note. No. 46. A Single Variable Analysis Program Package.

HARMAN, H.H. (1967). Univ. of Chicago Press. Modern Factor Analysis.

HOPKINS, J. (1977). Met.Mag 106, p.278. The spatial variability of Daily Temperatures and Sunshine over Uniform Terrain.

KENDALL, Sir M (1975). Chas Griffen & Co. Ltd. Multivariate Analysis.

KING, L.J. (1969). Prentice-Hall. Statistical Analysis in Geography.

MANLEY, G. (1970). World Survey of Climatology, Vol 5, p.81. The climates of the British Isles.

METEOROLOGICAL OFFICE (1975). Climat.Memo.No.73. Maps of Mean and Extreme Temperature over the United Kingdom 1941-70.

MORRISON, D.F. (1967). McGraw-Hill. Multivariate Statistical Methods.

SMITH, S.G. (1978). Meteorological Office. Met O 3 Tech Note No. 1. Long term Optimum Averaging Period for temperatures in the United Kingdom.

STEINER, D. (1965). Tijds.Kon.Nederl.Aardrijks Gemoot 82, No.4, p.329. A Multivariat Statistical approach to Climatic Regionalization and Classification.

WHITE, E.J. and LINDLEY, D.K. (1976). J.Envir.Manage. 4, p.161. The reduction of climatological data for ecological purposes: a preliminary analysis.

TABLE 1

Variance explained by first 15 components of Principal Component Analysis

(Daily Minimum Temperature)

Values for 1973, .... 1977 are $(deg.C)^2$ and for 1973/77 are $(standard\ scores)^2$. Cumulative percentage is given in brackets.

| Component | Year 1973 | 1974 | 1975 | 1976 | 1977 | 1973/77 |
|---|---|---|---|---|---|---|
| 1 | 260.9(39.0) | 247.9(45.6) | 261.3(38.7) | 234.6(38.5) | 264.8(45.0) | 5.29 (7.1) |
| 2 | 132.2(58.8) | 74.3(59.3) | 111.3(55.2) | 94.2(54.0) | 87.6(59.8) | 5.14(14.1) |
| 3 | 36.5(64.2) | 28.6(64.6) | 47.8(62.3) | 35.1(59.8) | 39.1(66.5) | 4.95(20.7) |
| 4 | 31.3(68.9) | 20.1(68.3) | 29.8(66.7) | 33.7(65.3) | 23.1(70.4) | 4.85(27.2) |
| 5 | 19.4(71.8) | 15.4(71.1) | 22.9(70.1) | 22.6(69.0) | 17.8(73.4) | 4.72(33.6) |
| 6 | 15.2(74.1) | 13.6(73.6) | 20.4(73.2) | 19.1(72.1) | 13.2(75.7) | 4.64(39.8) |
| 7 | 11.0(75.7) | 10.1(75.4) | 13.4(75.2) | 16.2(74.8) | 8.8(77.2) | 4.36(45.7) |
| 8 | 9.1(77.1) | 7.4(76.8) | 10.5(76.7) | 10.9(76.6) | 8.5(78.6) | 4.25(51.4) |
| 9 | 8.6(78.4) | 6.8(78.1) | 9.3(78.1) | 8.4(78.0) | 7.7(79.9) | 4.10(57.0) |
| 10 | 7.3(79.5) | 6.2(79.2) | 8.2(79.3) | 7.4(79.2) | 6.0(81.0) | 3.91(62.2) |
| 11 | 6.1(80.4) | 5.7(80.2) | 6.5(80.3) | 6.2(80.2) | 5.4(81.9) | 3.82(67.4) |
| 12 | 5.5(81.2) | 4.7(81.1) | 6.3(81.2) | 5.7(81.1) | 4.9(82.7) | 3.59(72.2) |
| 13 | 4.5(81.9) | 4.2(81.9) | 5.7(82.1) | 5.4(82.0) | 4.7(83.5) | 3.37(76.7) |
| 14 | 4.4(82.5) | 3.6(82.5) | 5.4(82.9) | 4.5(82.8) | 4.0(84.2) | 2.84(80.6) |
| 15 | 3.9(83.1) | 3.5(83.2) | 4.6(83.5) | 4.1(83.4) | 3.8(84.3) | 2.29(83.6) |

TABLE 2:

## Relationship between Factor Score and Altitude

(Values are coefficients of linear regression of factor score on altitude. Units: metres$^{-1}$ x 1000)

| Factor | District 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | All stations |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $-3.67^{xx}$ | $-3.46^{xx}$ | $-5.21^{xx}$ | $-1.77^{xx}$ | $-4.24^{xx}$ | $-5.79^{xx}$ | $-3.74^{xx}$ | $-3.68^{xx}$ | $-3.98^{xx}$ | $-3.64^{xx}$ | $-4.34^{xx}$ |
| 2 | $-4.23^{xx}$ | $-4.11^{xx}$ | $-3.52^{x}$ | | | $-7.72^{xx}$ | $-7.17^{xx}$ | $-4.71^{xx}$ | $-3.83^{xx}$ | $-2.12^{xx}$ | $-3.85^{xx}$ |
| 3 | $-2.44^{x}$ | | | | | | $-2.16^{xx}$ | $-4.31^{xx}$ | $-2.25^{xx}$ | $-4.72^{xx}$ | |
| 4 | | $-2.79^{xx}$ | $-3.88^{xx}$ | $-6.09^{xx}$ | | $-6.71^{xx}$ | | $-3.00^{xx}$ | $-2.43^{xx}$ | $-2.37^{xx}$ | $-3.05^{xx}$ |
| 5 | | | | | $2.26^{xx}$ | $4.22^{xx}$ | | | | $2.38^{x}$ | $2.89^{x}$ |
| 6 | $2.19^{xx}$ | | $3.51^{x}$ | | $3.31^{xx}$ | | $5.91^{xx}$ | | | | |
| 7 | $-6.42^{x}$ | | | | $2.65^{x}$ | | | $2.18^{xx}$ | | $4.04^{xx}$ | |
| 8 | | | $6.55^{xx}$ | | | | $3.40^{xx}$ | | | $2.59^{xx}$ | |
| 9 | $3.92^{xx}$ | $4.55^{xx}$ | | | $7.13^{xx}$ | | $5.95^{xx}$ | $4.27^{xx}$ | $5.11^{xx}$ | $6.18^{xx}$ | $3.74^{xx}$ |
| 10 | | | $7.30^{xx}$ | $6.70^{xx}$ | $3.04^{x}$ | $6.72^{xx}$ | | | $4.07^{xx}$ | | $2.71^{x}$ |
| 11 | | | | | | | | | | | |
| 12 | | $4.58^{xx}$ | $5.46^{xx}$ | $4.56^{xx}$ | $3.88^{xx}$ | $4.92^{xx}$ | | $3.50^{xx}$ | $2.01^{xx}$ | $3.89^{xx}$ | (1.97) |
| 13 | $5.48^{xx}$ | $4.14^{xx}$ | $3.41^{x}$ | $1.25^{x}$ | $9.81^{xx}$ | $5.44^{xx}$ | $2.51^{xx}$ | | $4.61^{xx}$ | $3.94^{xx}$ | $3.85^{xx}$ |
| 14 | | | | | $4.44^{xx}$ | | $2.49^{xx}$ | | | | |
| 15 | | | $-3.16^{xx}$ | $-1.06^{xx}$ | | $-8.57^{xx}$ | | | | | |
| No. of Stations | 45 | 83 | 45 | 49 | 80 | 95 | 66 | 55 | 98 | 63 | 677 |

Note: Significance levels of coefficients are indicated: $^{xx}$ for 1% and $^{x}$ for 5%

Values are omitted or given in brackets if coefficient is not significant at 5% level.

TABLE 3

Root mean square differences between station factor scores and values interpolated by single variable analysis. Values are standard scores

| Factor | North Area | South Area |
|---|---|---|
| 1 | 0,11 | 0.09 |
| 2 | 0,22 | 0.27 |
| 3 | 0,21 | 0.25 |
| 4 | 0,11 | 0.12 |
| 5 | 0,20 | 0.16 |
| 6 | 0,18 | 0,29 |
| 7 | 0,38 | 0.22 |
| 8 | 0,22 | 0.23 |
| 9 | 0,22 | 0.23 |
| 10 | 0,32 | 0.38 |
| 11 | 0,15 | 0.27 |
| 12 | 0,27 | 0.24 |
| 13 | 0.19 | 0.44 |
| 14 | 0,26 | 0.16 |
| 15 | 0,20 | 0.34 |

TABLE 4(a):

## List of stations with small residuals for 1974

Values are given of the standard deviation ( $\sigma(r)$) and mean ($\bar{r}$) of the residuals for both Minimum and Maximum Temperatures. The rank number is included. (MO indicates Meteorological Office station)

| DCNN Station Name | | Minimum | | | Maximum | | |
|---|---|---|---|---|---|---|---|
| | | $\sigma(r)$ | $\bar{r}$ | rank on sd | $\sigma(r)$ | $\bar{r}$ | rank on sd |
| 5576 READing | | 0.58 | +0.05 | 3 | 0.41 | +0.01 | 1 |
| 4406 EDGBaston | | 0.55 | +0.05 | 1 | 0.53 | -0.07 | 23 |
| 4522 OXFOrd | | 0.57 | +0.07 | 2 | 0.48 | +0.01 | 7 |
| 5131 HAMPton | | 0.61 | -0.00 | 10 | 0.43 | -0.07 | 2 |
| 5258 KEW (Psych) | MO | 0.94 | +0.10 | 320 | 0.44 | +0.04 | 3 |
| 3537 ROTHamsted | | 0.59 | +0.06 | 4 | 0.49 | -0.06 | 11 |
| 5113 HEAThrow | MO | 0.63 | +0.06 | 18 | 0.44 | -0.03 | 4 |
| 5592 EASThampstead | MO | 0.89 | -0.02 | 247 | 0.46 | +0.01 | 5 |
| 5642 SOUThampton (M.Park) | MO | 0.60 | +0.06 | 5 | 0.75 | +0.09 | 279 |
| 3534 GARSton | | 0.62 | +0.07 | 16 | 0.47 | -0.11 | 6 |
| 4206 WATNall | MO | 0.60 | +0.02 | 6 | 0.59 | -0.15 | 65 |
| 8881 EXMOuth | | 0.60 | +0.00 | 7 | 0.72 | -0 05 | 247 |
| 5588 HURLey | | 0.80 | +0.07 | 142 | 0.48 | +0.02 | 8 |
| 4061 SHEFfield | | 0.61 | +0.10 | 8 | 0.77 | -0.03 | 303 |
| 5259 KEW (N.W.S) | MO | 0.67 | +0.03 | 29 | 0.49 | +0.06 | 9 |
| 5863 LARKhill | MO | 0.61 | +0.16 | 9 | 0.58 | +0.02 | 42 |
| 9142 ALDErgrove | MO | 0.73 | +0.04 | 66 | 0.49 | +0.12 | 10 |
| 1646 EDINburgh (R.O.) | | 0.62 | -0.15 | 11 | 0.65 | -0.01 | 121 |
| 5670 SOUThsea | | 0.62 | +0.00 | 12 | 0.58 | -0.06 | 47 |
| 4447 ELMDon | MO | 0.73 | -0.07 | 69 | 0.50 | +0.02 | 12 |
| 5694 ALICe Holt | | 0.67 | +0.09 | 30 | 0.50 | +0.03 | 13 |
| 4398 WITTering | MO | 0.62 | -0.01 | 13 | 0.58 | -0.06 | 50 |
| 3374 WYTOn | MO | 0.62 | +0.04 | 14 | 0.58 | -0.03 | 45 |
| 9347 LOUGhgall | | 0.86 | +0.12 | 222 | 0.50 | -0.10 | 14 |
| 5575 SHINfield | | 0.65 | -0.04 | 21 | 0.52 | +0.20 | 15 |
| 4043 HUDDersfield Oakes | | 0.62 | +0.08 | 15 | 0.79 | -0.18 | 328 |

TABLE 4(b):

## List of stations with large residuals for 1974

Values are given of the standard deviation ($\sigma(r)$) and mean ($\bar{r}$) of the residuals for both Minimum and Maximum Temperatures. The rank number is included.

| DCNN Station Name | | Minimum | | | Maximum | | |
|---|---|---|---|---|---|---|---|
| | | $\sigma(r)$ | $\bar{r}$ | rank | $\sigma(r)$ | $\bar{r}$ | rank |
| 7229 GRIZedale | | 1.88 | -0.29 | 627 | 1.11 | +0.05 | 577 |
| 8555 USK | * | 1.83 | -0.41 | 626 | 1.04 | -0.15 | 554 |
| 1411 ABERfoyle | * | 1.73 | -0.18 | 625 | 0.99 | +0.11 | 535 |
| 7657 RUTHin | | 1.68 | -0.18 | 624 | 1.15 | +0.11 | 584 |
| 7905 CEINws | ** | 1.67 | -0.17 | 623 | 1.08 | +0.18 | 569 |
| 5157 HARRow Weald | | 1.61 | +0.02 | 622 | 1.39 | -1.53 | 596 |
| 8752 WINFrith | | 1.58 | -0.18 | 621 | 0.84 | +0.09 | 400 |
| 8153 HAVErford West | | 1.57 | -0.37 | 620 | - | - | - |
| 4162 MACKworth | | 1.55 | -0.02 | 618 | 1.57 | +0.16 | 598 |
| 5418 FERNhurst | | 1.55 | -0.65 | 619 | 0.57 | +0.01 | 41 |
| 4792 HAMPton | | 1.54 | -0.08 | 617 | 0.92 | +0.43 | 492 |
| 2378 FILEy | | 1.06 | +0.21 | 440 | 1.54 | +0.10 | 597 |
| 6747 DUNDeugh | | 1.48 | +0.01 | 616 | 0.87 | +0.05 | 429 |
| 2163 HOUGhall | | 1.47 | -0.28 | 615 | 0.91 | +0.21 | 468 |
| 0088 BALTasound | | 1.46 | +0.19 | 614 | 1.03 | +0.54 | 553 |
| 0104 SULE Skerry | | 1.46 | -0.04 | 613 | - | - | - |
| 7178 GREAt Dun Fell | | - | - | - | 1.34 | -1.98 | 595 |
| 0575 FORT Augustus | | 1.31 | -0.22 | 590 | 1.30 | -0.12 | 594 |
| 4217 MANSfield | * | 1.26 | -0.04 | 574 | 1.30 | -0.09 | 593 |
| 0343 CASSley | | 1.26 | -0.60 | 571 | 1.29 | -0.01 | 592 |
| 2494 SKEGness | * | 1.02 | -0.17 | 398 | 1.27 | +0.17 | 591 |
| 3064 CROMer | | 0.99 | +0.34 | 374 | 1.25 | -0.71 | 590 |
| 5827 LACOck | | 1.07 | +0.06 | 451 | 1.22 | +0.21 | 589 |
| 6896 NEWTon Stewart | | 1.21 | +0.21 | 588 | 1.46 | +0.07 | 612 |
| 9970 JERSey (St Helier) | | 0.93 | +0.06 | 305 | 1.19 | +0.16 | 587 |
| 7408 RHYL | | 1.29 | +0.18 | 584 | 1.18 | +0.20 | 586 |
| 9241 KILKeel | | 1.16 | +0.06 | 515 | 1.18 | +0.28 | 585 |

Note: * indicates stations which report with precision of 0.5 deg.C

** indicates station which reports with precision of 1.0 deg.C

TABLE 4(c)

## List of WMO Reference Stations and Statistics for 1974

Values are given of the standard deviation ($\sigma(r)$) and mean ($\bar{r}$) of the residuals for both Minimum and Maximum Temperatures. The rank number is included.

| DCNN Station Name | Minimum | | | Maximum | | |
|---|---|---|---|---|---|---|
| | $\sigma(r)$ | $\bar{r}$ | rank | $\sigma(r)$ | $\bar{r}$ | rank |
| 0043 LERWick (Psych) | 1.12 | +0.11 | 486 | 0,85 | +0,23 | 409 |
| 0044 LERWick (S Screen) | 1.09 | +0.17 | 466 | 0,86 | +0,31 | 418 |
| 0425 STORnoway | 1.24 | +0.14 | 567 | 0,89 | +0,06 | 450 |
| 1577 LEUChars | 1.08 | -0.18 | 457 | 0,70 | +0,06 | 198 |
| 2245 LEEMing | 0.83 | +0.09 | 175 | 0,81 | -0,03 | 352 |
| 3127 HONIngton | 0.66 | -0.03 | 25 | 0,61 | +0,05 | 78 |
| 4406 EDGBaston | 0.55 | +0.05 | 1 | 0,53 | -0,07 | 23 |
| 5113 HEAThrow | 0.63 | +0.06 | 18 | 0,44 | -0,03 | 4 |
| 5258 KEW(Psych) | 0.94 | +0.10 | 320 | 0,44 | +0,04 | 3 |
| 5259 KEW (N.W.S) | 0.67 | +0.03 | 27 | 0,49 | +0,06 | 9 |
| 6677 ESKDalemuir (Psych) | 1.07 | +0.04 | 446 | 0,63 | -0.04 | 90 |
| 6679 ESKDalemuir (S Screen) | 1.02 | +0.08 | 404 | 0,71 | +0,08 | 221 |
| 7377 RINGway | 0.71 | +0.15 | 51 | 0,64 | +0.02 | 111 |
| 7511 VALLey | 0.90 | -0.10 | 259 | 0.88 | -0.01 | 437 |
| 8812 PLYMouth Hoe | 0.75 | -0.09 | 87 | 0,69 | +0,07 | 186 |
| 9142 ALDErgrove | 0.73 | +0.12 | 66 | 0,49 | +0.04 | 10 |

TABLE 4(d):

List of 'frost hollows' (England and Wales) and statistics for 1974

Values are given of the standard deviation ($\sigma(r)$) and mean ($\bar{r}$) of the residuals for both Minimum and Maximum Temperatures. The rank number is included.

| DCNN Station Name | Minimum | | | Maximum | | |
|---|---|---|---|---|---|---|
| | $\sigma(r)$ | $\bar{r}$ | rank | $\sigma(r)$ | $\bar{r}$ | rank |
| 2163 HOUGhall | 1.47 | -0.28 | 615 | 0.91 | +0.21 | 468 |
| 2273 PICKering | 1.13 | +0.08 | 495 | 0.82 | +0.04 | 370 |
| 2425 LINColn | 1.20 | -0.08 | 540 | 0.60 | -0.03 | 73 |
| 3031 SANTon Downham | 1.17 | -0.16 | 526 | 0.70 | +0.07 | 192 |
| 4237 WARSop | 1.09 | +0.13 | 467 | 0.65 | +0.06 | 119 |
| 4341 CALDecott | 0.98 | -0.06 | 366 | 0.58 | -0.07 | 46 |
| 4561 MEDMenham | 1.13 | +0.00 | 496 | 0.54 | +0.02 | 29 |
| 4567 GRENdon Underwood | 0.89 | +0.03 | 252 | 0.53 | -0.14 | 25 |
| 4787 NEWPort | 0.99 | +0.04 | 372 | 0.59 | +0.01 | 61 |
| 4886 PRESton Wynne | 1.40 | -0.31 | 602 | 0.83 | -0.04 | 388 |
| 5255 MICKleham | 1.22 | -0.20 | 555 | 0.61 | +0.00 | 82 |
| 5656 WINChester | 1.05 | -0.13 | 422 | 1.06 | -0.03 | 565 |
| 5827 LACOck | 1.07 | +0.06 | 451 | 1.22 | +0.21 | 589 |
| 5877 MARLborough | 0.90 | +0.20 | 260 | 0.52 | -0.01 | 16 |
| 7228 CARTmel | 1.22 | +0.13 | 556 | 0.76 | +0.07 | 292 |
| 7229 GRIZedale | 1.88 | -0.29 | 627 | 1.11 | +0.05 | 577 |
| 7623 ALWEn | 1.21 | -0.03 | 546 | 0.72 | -0.02 | 238 |
| 7665 LOGGerheads | 1.16 | +0.01 | 522 | 0.78 | -0.02 | 320 |
| 7884 CORWen | 1.31 | -0.09 | 591 | 0.89 | +0.03 | 451 |
| 8068 GOGErddan | 1.05 | -0.01 | 423 | 0.68 | +0.15 | 169 |
| 8153 HAVErford West | 1.57 | -0.37 | 620 | - | - | - |
| 8235 CARMarthen | 1.21 | +0.19 | 547 | 0.84 | +0.04 | 398 |
| 8555 USK | 1.83 | -0.41 | 626 | 1.04 | -0.15 | 554 |
| 8853 TOTNes | 1.21 | +0.09 | 548 | - | - | - |

TABLE 5:

## Stations whose factors fit poorly with the analysis

Values are given of residuals ('observed' - 'analysed') greater than 1 or 5 x rms
(Note: * indicates station with less than 3 years data).

| South Region | | | | North Region | | | |
|---|---|---|---|---|---|---|---|
| **Factor 1** | | | | | | | |
| (5 x rms = 0.45) | | | | (5 x rms = 0.55) | | | |
| 7905 | CELNws | -0.58 | | | | | |
| 5157 | HARRow Weald | -0.57 | | | | | |
| 5046 | LONDon W.C. | +0.52 | | | | | |
| **Factor 2** | | | | | | | |
| (5 x rms = 1.35) | | | | (5 x rms = 1.10) | | | |
| 7623 | ALWEn | -1.46 | | 6478 | CARNwath | -1.25 | |
| 8153 | HAVErford West | -1.15 | * | | | | |
| **Factor 3** | | | | | | | |
| (5 x rms = 1.25) | | | | (5 x rms = 1.05) | | | |
| 5683 | BUTSer (Hillhampton) | -1.52 | * | 9162 | DIVIs Mountain | -1.50 | |
| 8153 | HAVErford West | -1.06 | * | 6276 | EARLs Hill | -1.38 | |
| 5656 | WINChester | +1.09 | * | (6457 | SALSburgh | -0.96) | * |
| 5046 | LONDon W.C. | +1.32 | * | | | | |
| **Factor 4** | | | | | | | |
| (5 x rms = 0.60) | | | | (5 x rms = 0.55) | | | |
| 8153 | HAVErford West | -1.29 | * | | | | |
| 5805 | TROWbridge | -0.61 | * | | | | |
| **Factor 5** | | | | | | | |
| (5 x rms = 0.80) | | | | (5 x rms = 1.00) | | | |
| 5387 | ELMStone | -1.16 | | 9134 | LISNafillan | -1.07 | |
| 5418 | FERNhurst | -0.88 | | 9162 | DIVIs Mountain | +1.03 | |
| 5490 | HASTings | +0.94 | | | | | |
| 8768 | SHAFtesbury | +1.26 | * | | | | |
| **Factor 6** | | | | | | | |
| (5 x rms = 1.45) | | | | (5 x rms = 0.90) | | | |
| 5387 | ELMStone | -1.63 | | 9434 | LISLap Fell | +1.01 | * |
| 7912 | MOEL Cynnedd | -1.54 | | 2257 | MARTon | +1.06 | * |
| 4967 | CHELtenham | +1.44 | | 1993 | SOURhope | +1.66 | * |

## Factor 7

| (5 x rms = 1.20) | | | | (5 x rms = 1.90) | | | |
|---|---|---|---|---|---|---|---|
| 7604 | BETWs-y-Coed | -1.94 | * | 0584 | LAGGanlia | -2.48 | |
| 5646 | SPARsholt | -1.00 | * | 6064 | KNAPdale Forest | -1.01 | * |
| 8073 | SWYDdffynnon | +1.16 | * | 9162 | DIVIs Mountain | +1.08 | |
| | | | | 0482 | FORTrose | +1.13 | |
| | | | | 6457 | SALSburgh | +1.36 | * |
| | | | | 1103 | CLAShnoir | +1.72 | * |
| | | | | 0499 | TARBatness | +2.22 | |
| | | | | 0449 | KNOCkanrock | +2.81 | |

## Factor 8

| (5 x rms = 1.25) | | | | (5 x rms = 1.10) | | | |
|---|---|---|---|---|---|---|---|
| 5387 | ELMStone | -1.48 | | 1411 | ABERfoyle | -1.64 | * |
| | | | | 1993 | SOURhope | +2.53 | * |

## Factor 9

| (5 x rms = 1.15) | | | | (5 x rms = 1.10) | | | |
|---|---|---|---|---|---|---|---|
| 5337 | CHATham South | -1.23 | * | 4057 | HARRogate | -1.62 | |
| 8768 | SHAFtesbury | +1.52 | * | | | | |

## Factor 10

| (5 x rms = 1.90) | | | | (5 x rms = 1.60) | | | |
|---|---|---|---|---|---|---|---|
| 2425 | LINColn | -3.33 | | 2273 | PICKering | -2.46 | |
| 2273 | PICKering | -2.26 | | 9025 | LOUGhermore Forest | -1.69 | |
| 8853 | TOTNes | -2.01 | * | 9091 | TRAAd Point | -1.13 | * |
| 8986 | BASTreet | -1.48 | | 6896 | NEWTon Stewart | +1.07 | * |

## Factor 11

| (5 x rms = 1.35) | | | | (5 x rms = 0.75) | | | |
|---|---|---|---|---|---|---|---|
| 4146 | CHATsworth | -1.87 | * | 1462 | GLENeagles | -1.14 | |
| 5258 | KEW (Psych) | -1.84 | | 1411 | ABERfoyle | +1.14 | * |
| 8153 | HAVErford West | -1.13 | * | | | | |
| 7866 | BALA | -1.00 | | | | | |
| 5297 | ADDIngton | +1.02 | * | | | | |
| 7678 | BWLChgwyn | +1.08 | | | | | |
| 8009 | ABERporth | +2.22 | | | | | |

## Factor 12

| (5 x rms = 1.20) | | | | (5 x rms = 1.35) | | | |
|---|---|---|---|---|---|---|---|
| 8235 | CARMarthen | -1.20 | | 1042 | GRANtown-on-Spey | -1.54 | |
| 5426 | NORTh Heath | +1.31 | | | | | |

## Factor 13

(5 x rms = 2.20)

| | | | |
|---|---|---|---|
| 5877 | MARLborough | -2.53 | |
| 5387 | ELMStone | -2.45 | |
| 3031 | SANTon Downham | -2.39 | |
| 8633 | NETTlescombe | -1.29 | |
| 8153 | HAVErford West | -1.18 | * |
| 5258 | KEW (Psych) | -1.10 | |
| 5805 | TROWbridge | -1.07 | * |
| 4958 | INNSworth | -1.03 | * |
| 4146 | CHATsworth | -1.02 | * |
| 5592 | EASThampstead | -1.00 | |
| 8078 | LLETy-evan-Hen | +1.02 | |
| 4364 | MOULton Park | +1.17 | * |
| 8632 | NETTlescombe (Birds Hill) | +1.24 | |
| 8073 | SWYDdffy | +1.29 | * |
| 4995 | LITTle Rissington | +1.99 | |

(5 x rms = 0.95)

| | | | |
|---|---|---|---|
| 1303 | FETTercastle | -1.32 | * |
| 6896 | NEWTon Stewart | -1.17 | * |
| 1478 | ASHIntully Castle | -0.99 | |
| 2260 | YORK Heslington | +0.99 | |

## Factor 14

(5 x rms = 0.80)

| | | | |
|---|---|---|---|
| 5683 | DUTSer (Hillhampton) | -1.19 | * |
| 8312 | CENArth | -0.99 | * |

(5 x rms = 1.30)

| | | |
|---|---|---|
| 6287 | STIRling (Batterflats) | -1.13 |
| 1478 | ASHIntully Castle | +1.54 |
| 6276 | EARLs Hill | +1.74 |

## Factor 15

(5 x rms = 1.80)

| | | | |
|---|---|---|---|
| 5046 | LONDon W.C. | -1.45 | |
| 5297 | ADDIngton | -1.26 | * |
| 5393 | ST.Margarets Bay | -1.22 | * |
| 5392 | DOVEr (RMS) | -1.05 | |
| 5399 | MARGate | -1.03 | |
| 5687 | LONG Sutton | -1.02 | |
| 5365 | THROwley | -1.01 | * |
| 4465 | COVEntry Airport | +1.03 | |
| 5380 | FOLKestone | +1.03 | |
| 8964 | ST.Austel (Bethel) | +1.18 | * |
| 5323 | HADLow College | +1.71 | |

(5 x rms = 1.00)

| | | | |
|---|---|---|---|
| 9190 | KILRoot | -1.33 | * |
| 9188 | LARNe | -1.02 | * |

TABLE 6:

Estimates of rms errors of interpolation (degC) of daily temperatures by linear interpolation - taken from Hopkins (1977)

| | Minimum | | Maximum | |
|---|---|---|---|---|
| Station spacing: | 25 km | 75 km | 25 km | 75 km |
| Network: | | | | |
| - Best possible | 0.60 | 0.75 | 0,35 | 0.45 (summer)<br>0.55 (winter) |
| - Typical | 0.70 | 1.00 | 0,60 | 0.65 (summer)<br>0.75 (winter) |

# Figures

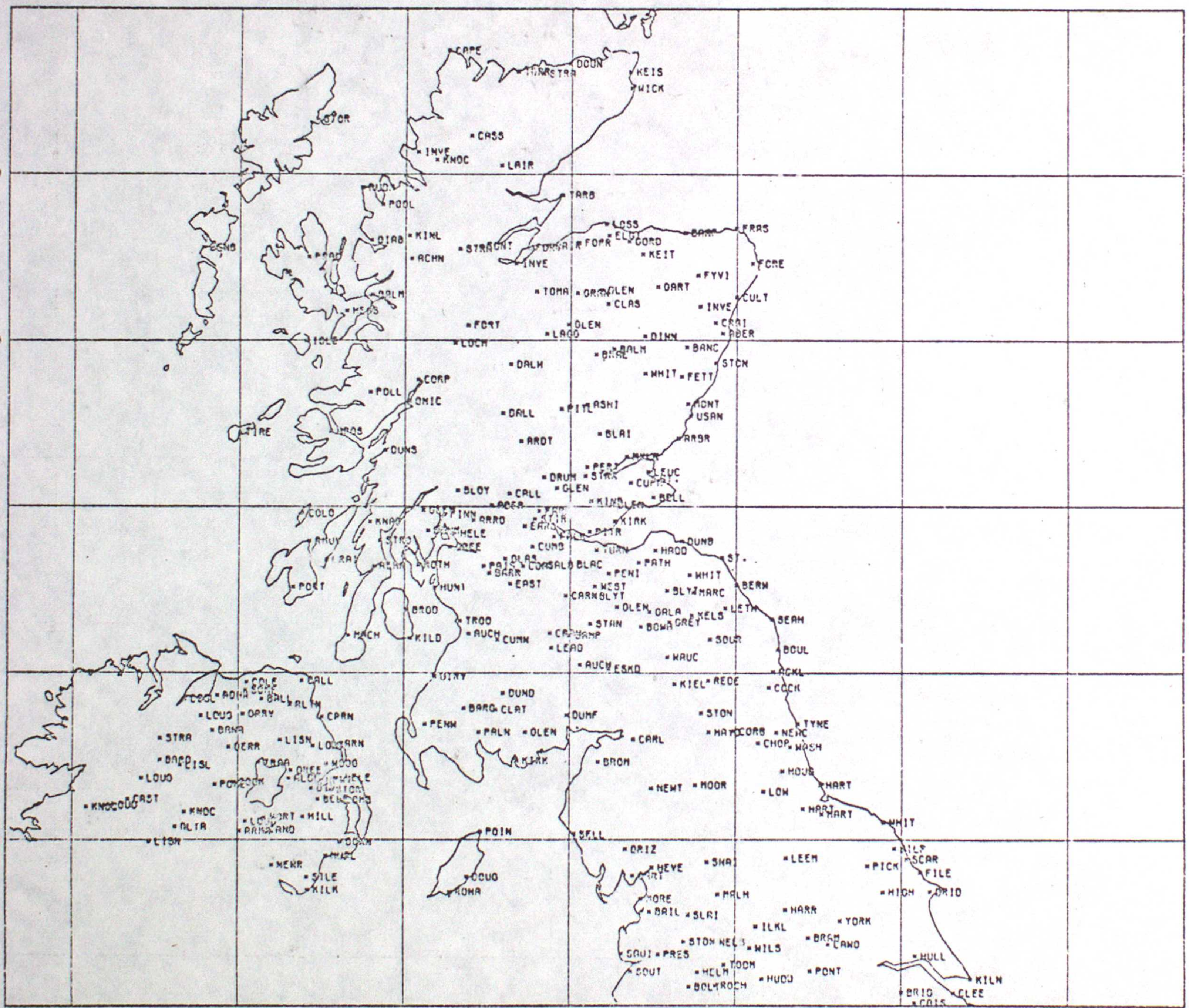
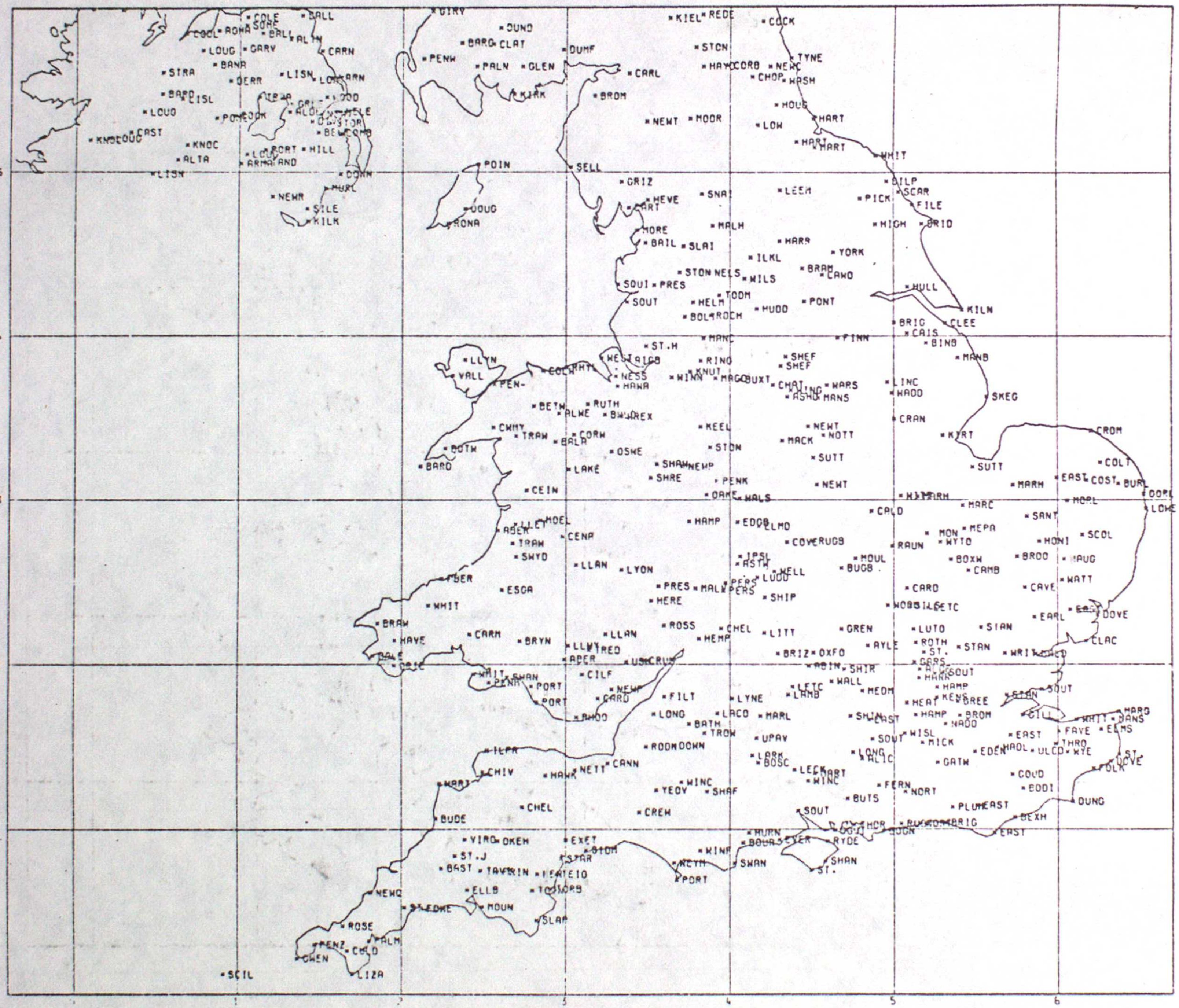



Fig. 1(a)

Location of stations for north of U.K.

The first 4 letters of each station name are given




## Fig. 1(b)

Location of stations for south of U.K.

The first 4 letters of each station name are given

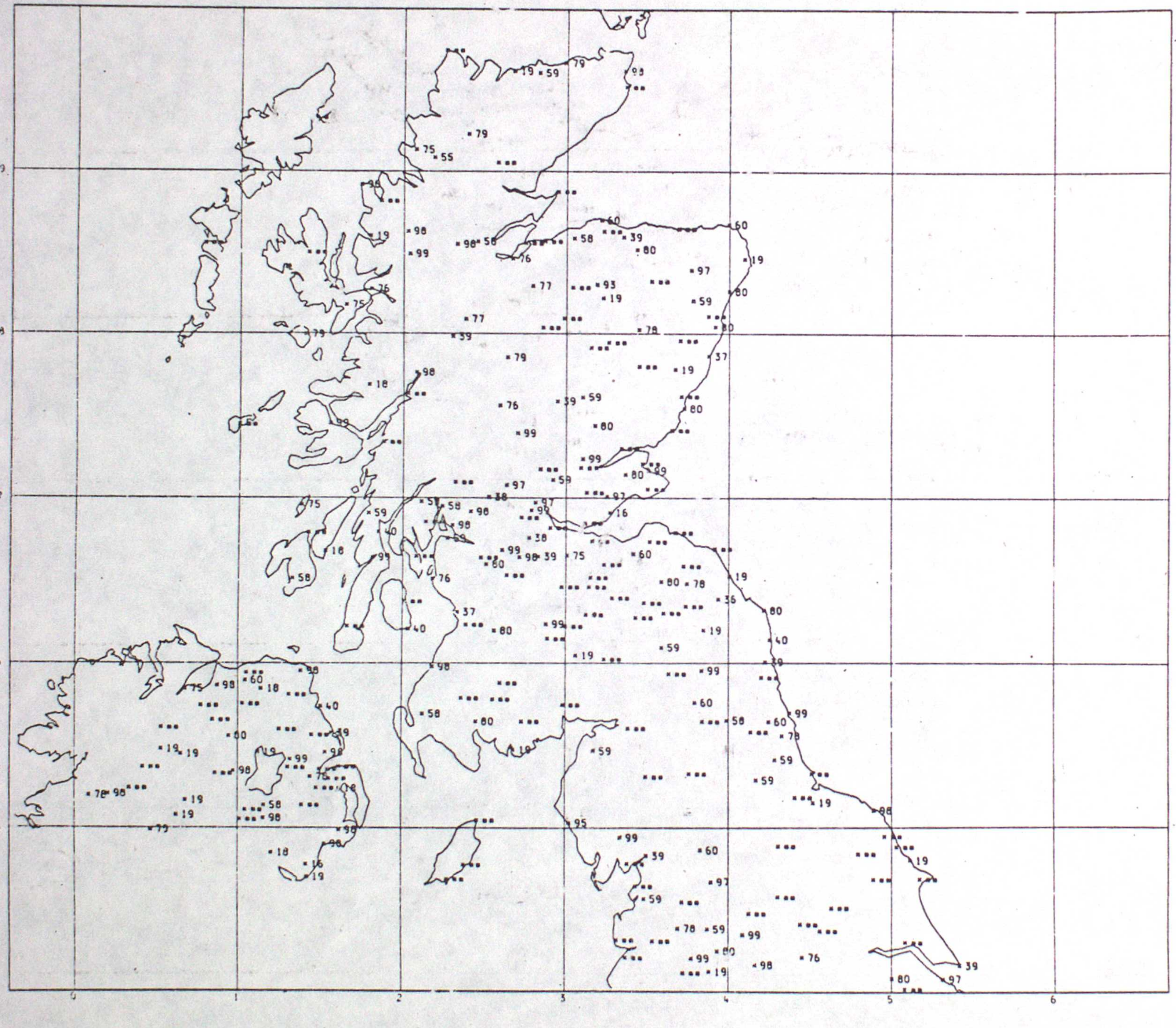


Fig. 2(a)

Percentage of possible observations of Daily Minimum Temperature for north of U.K. - 1973-77

** = 100%

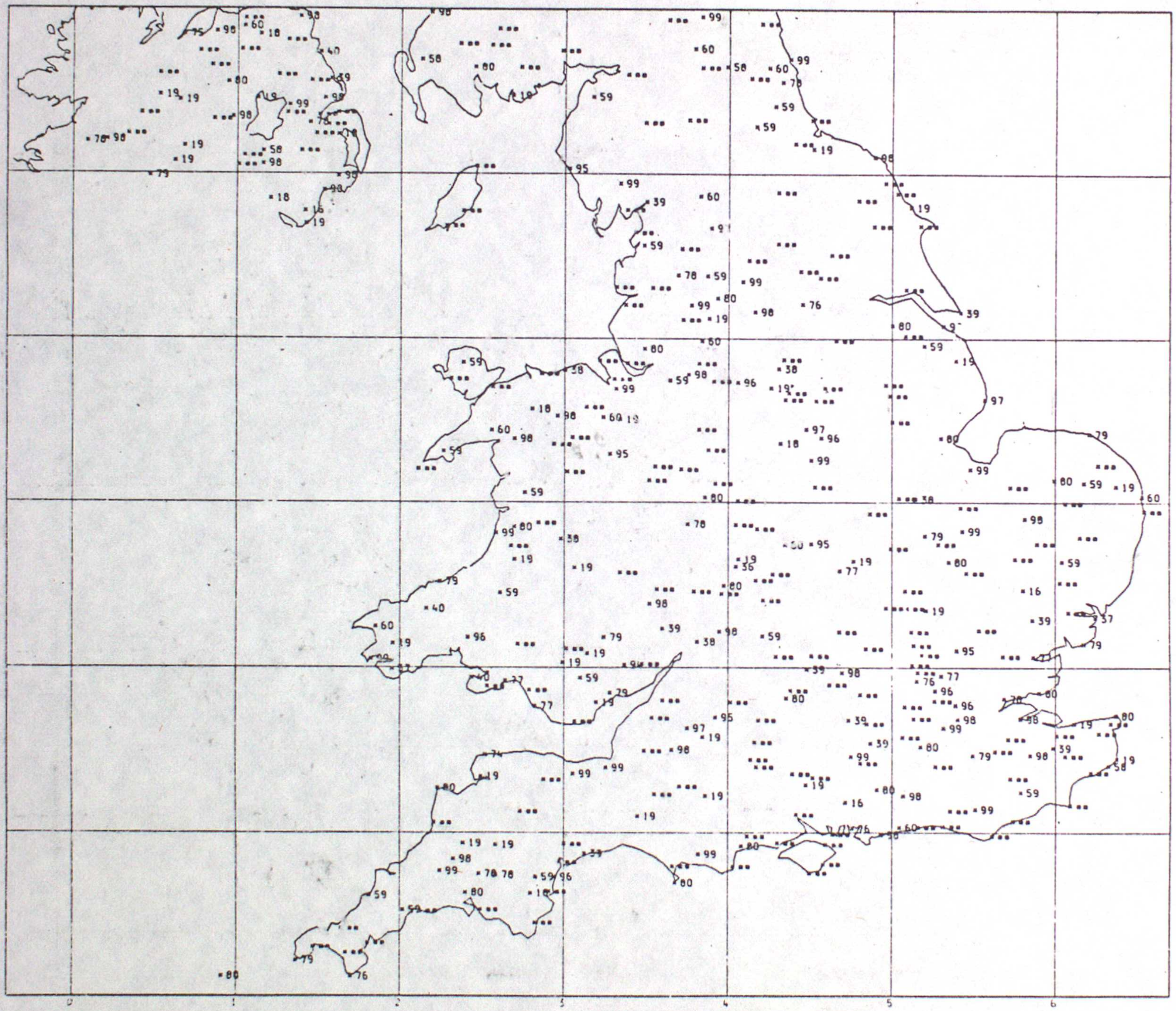


Fig. 2(b)

Percentage of possible observations of Daily Minimum Temperature for south of U.K. - 1973-77

** = 100%

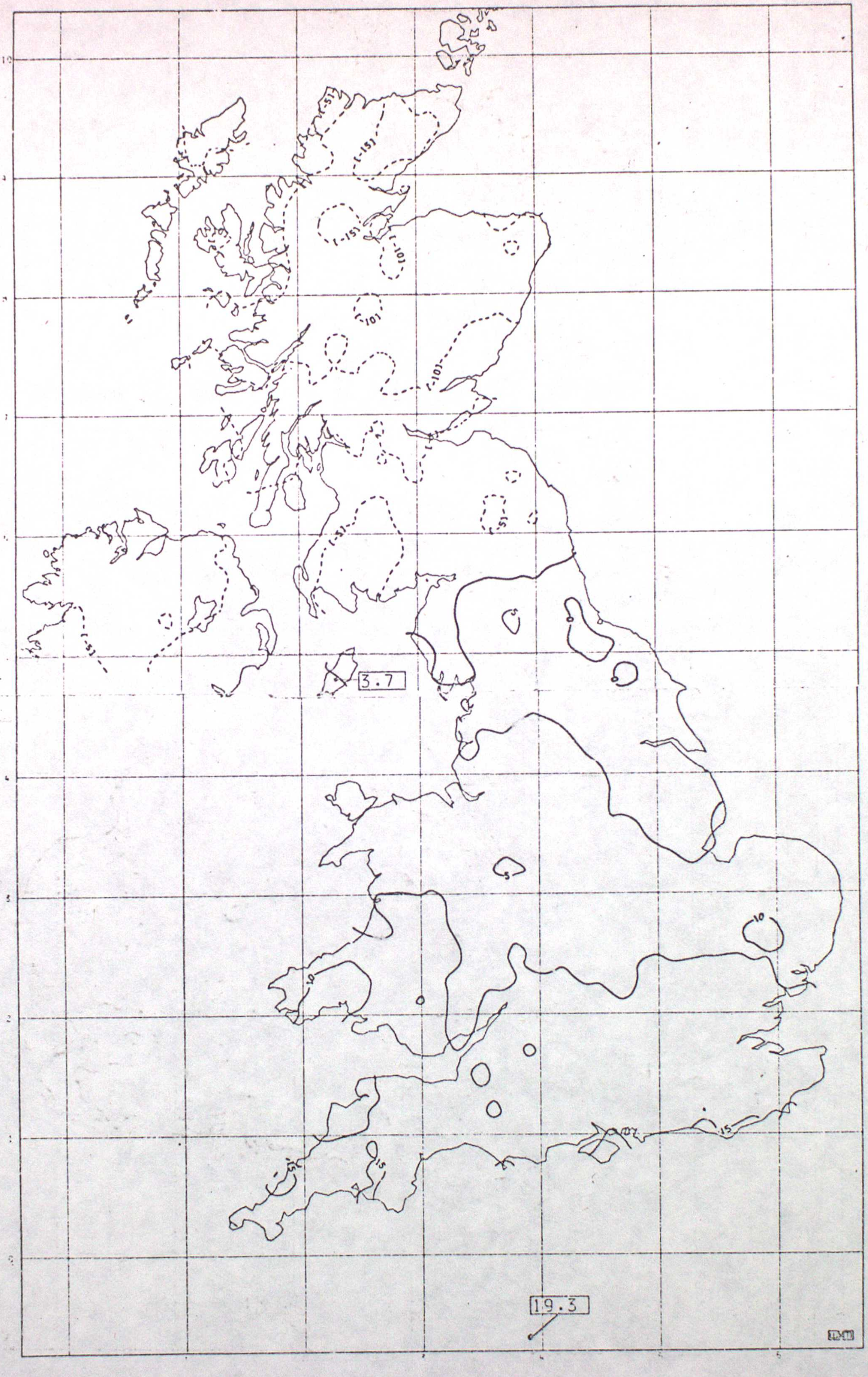



Fig. 3(a) Daily Minimum Temperature 1973-77 - Factor 1
(Reduced to m.s.l. using factor score lapse of +0.00434 per metre)

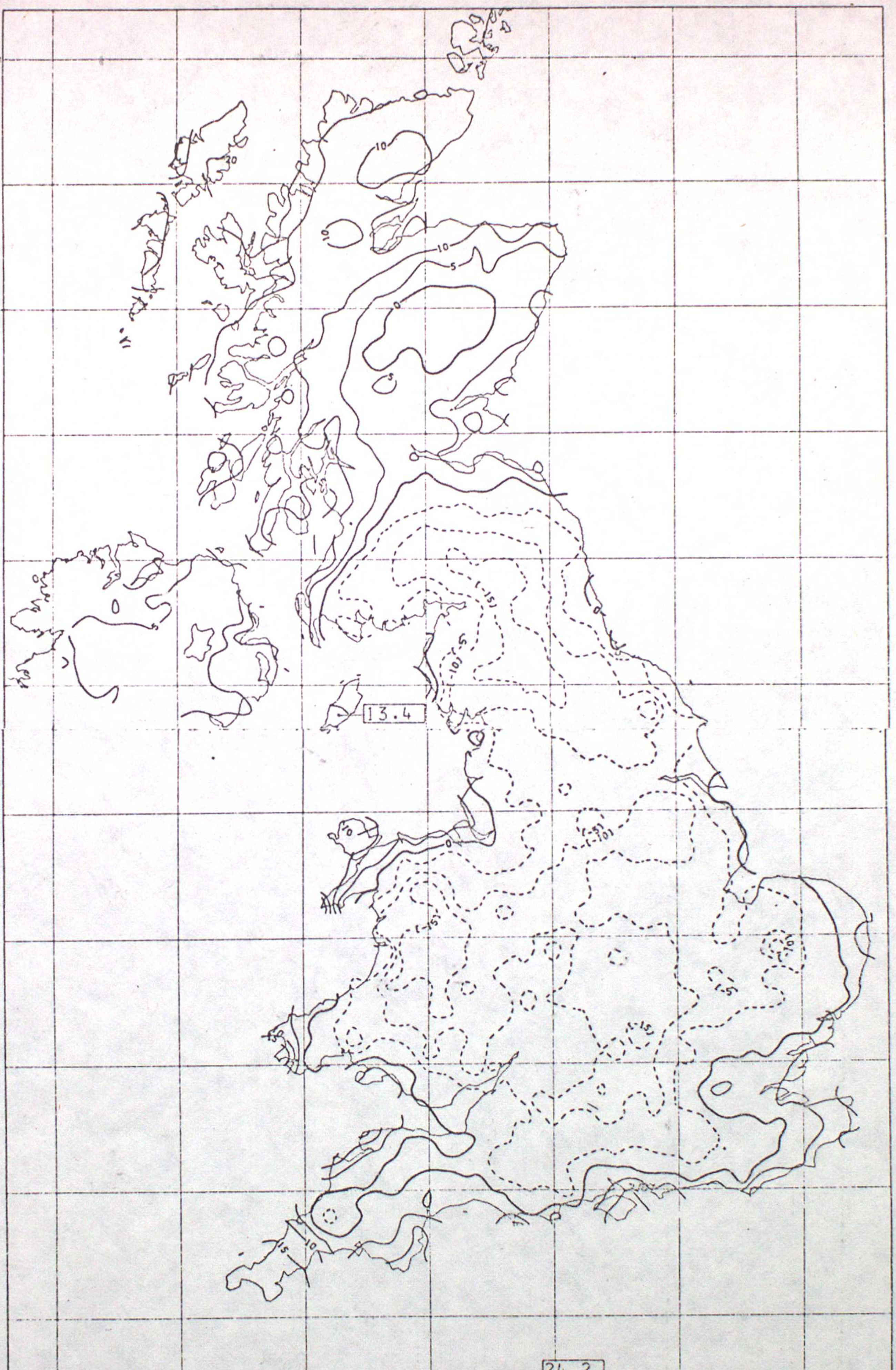



Fig. 3(b) Daily Minimum Temperature 1973-77 - Factor 2

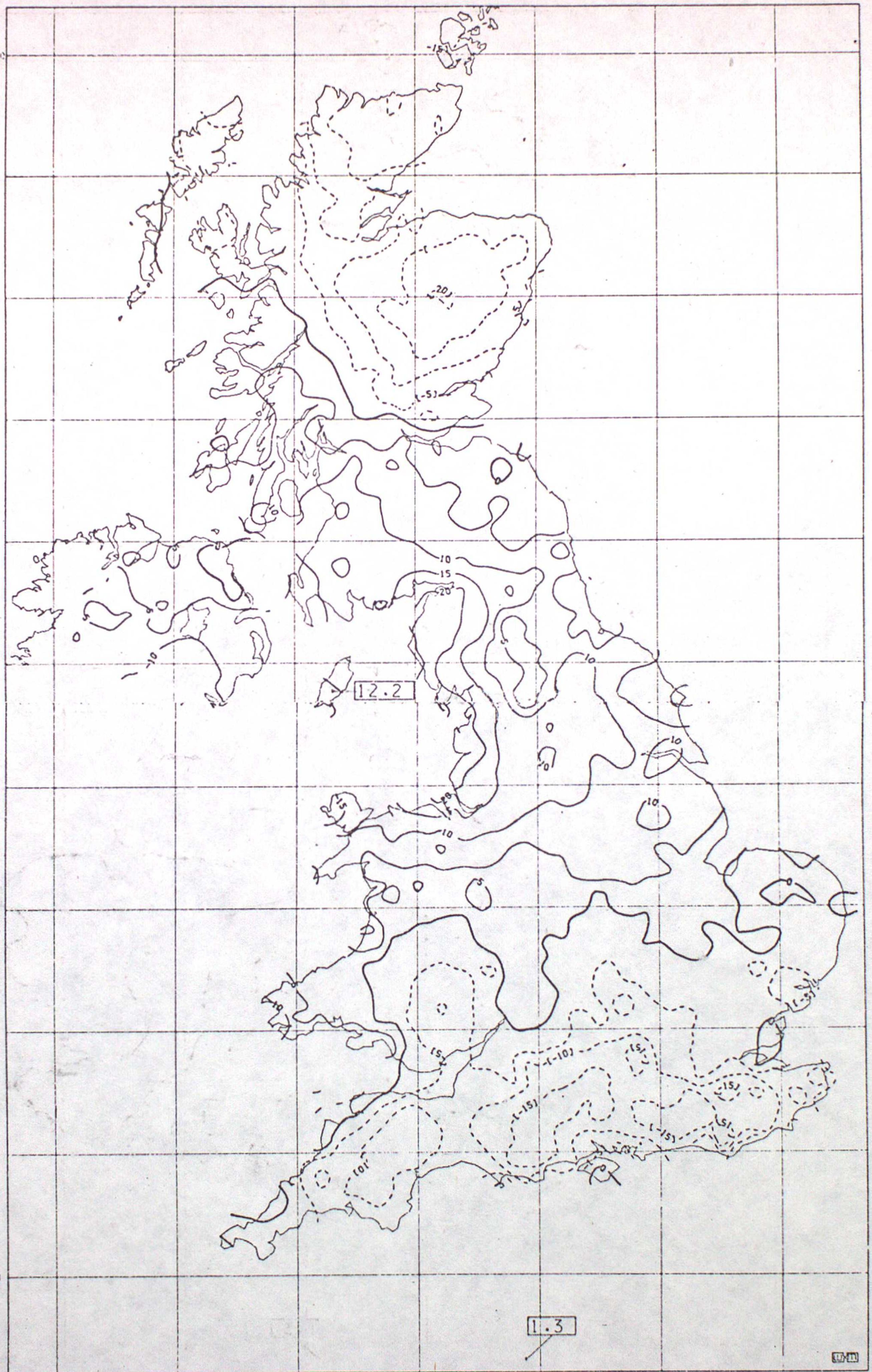



Fig. 3(c) Daily Minimum Temperature 1973-77 - Factor 3

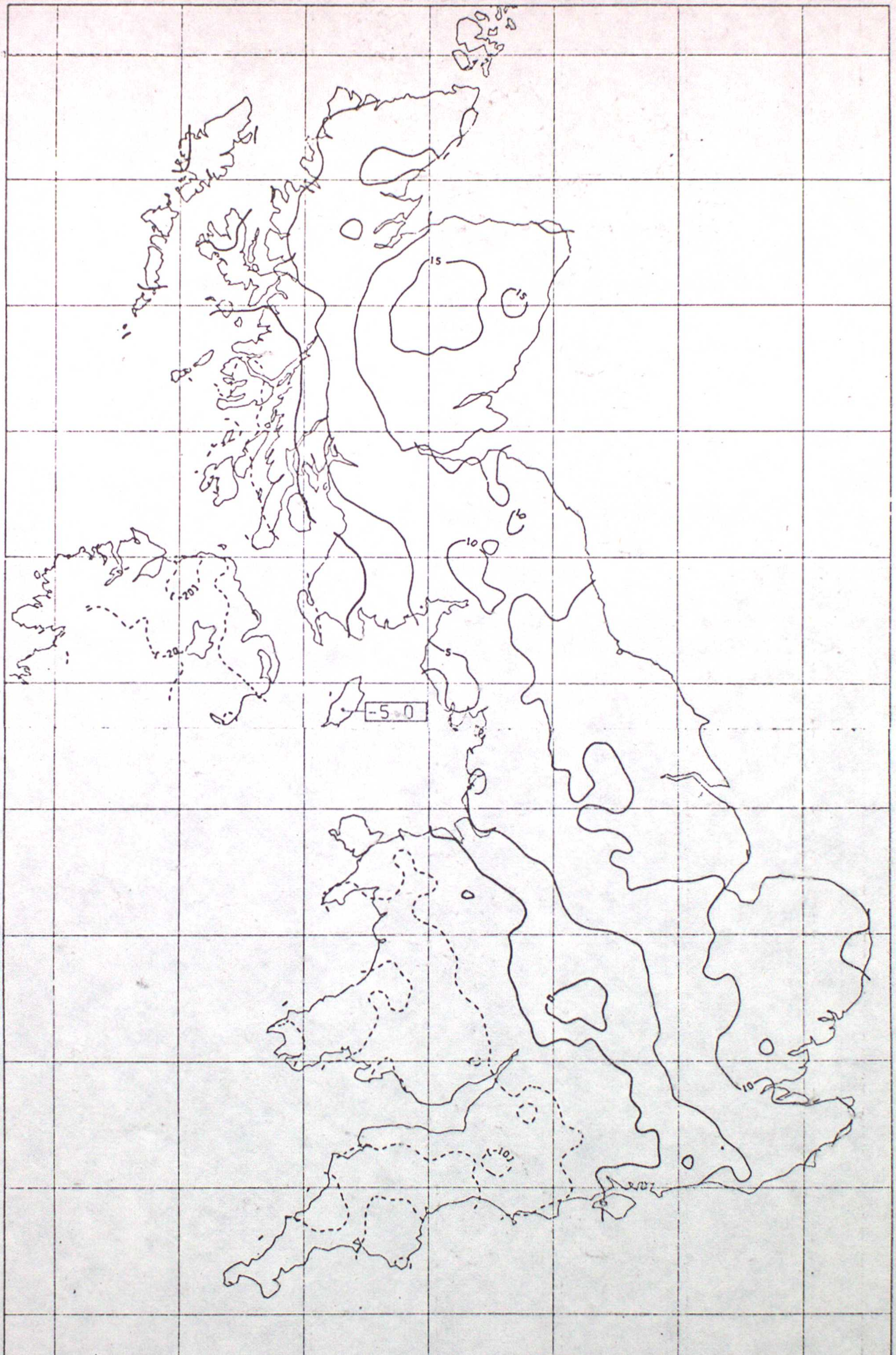



Fig. 3(d) Daily Minimum Temperature 1973-77 - Factor 4
(Reduced to m.s.l. using factor score lapse of +0.00305 per metre)

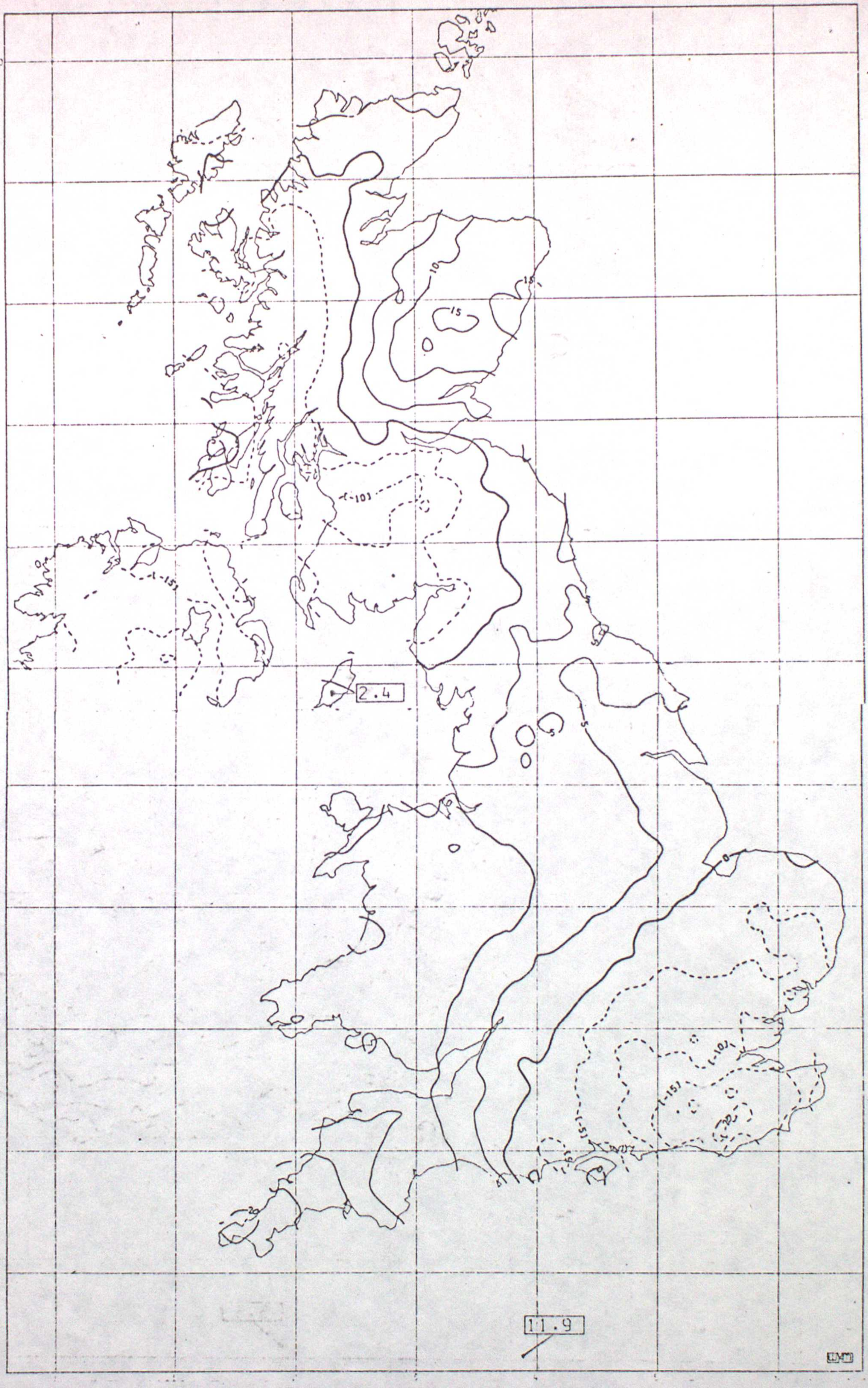



Fig. 3(e) Daily Minimum Temperature 1973-77 - Factor 5

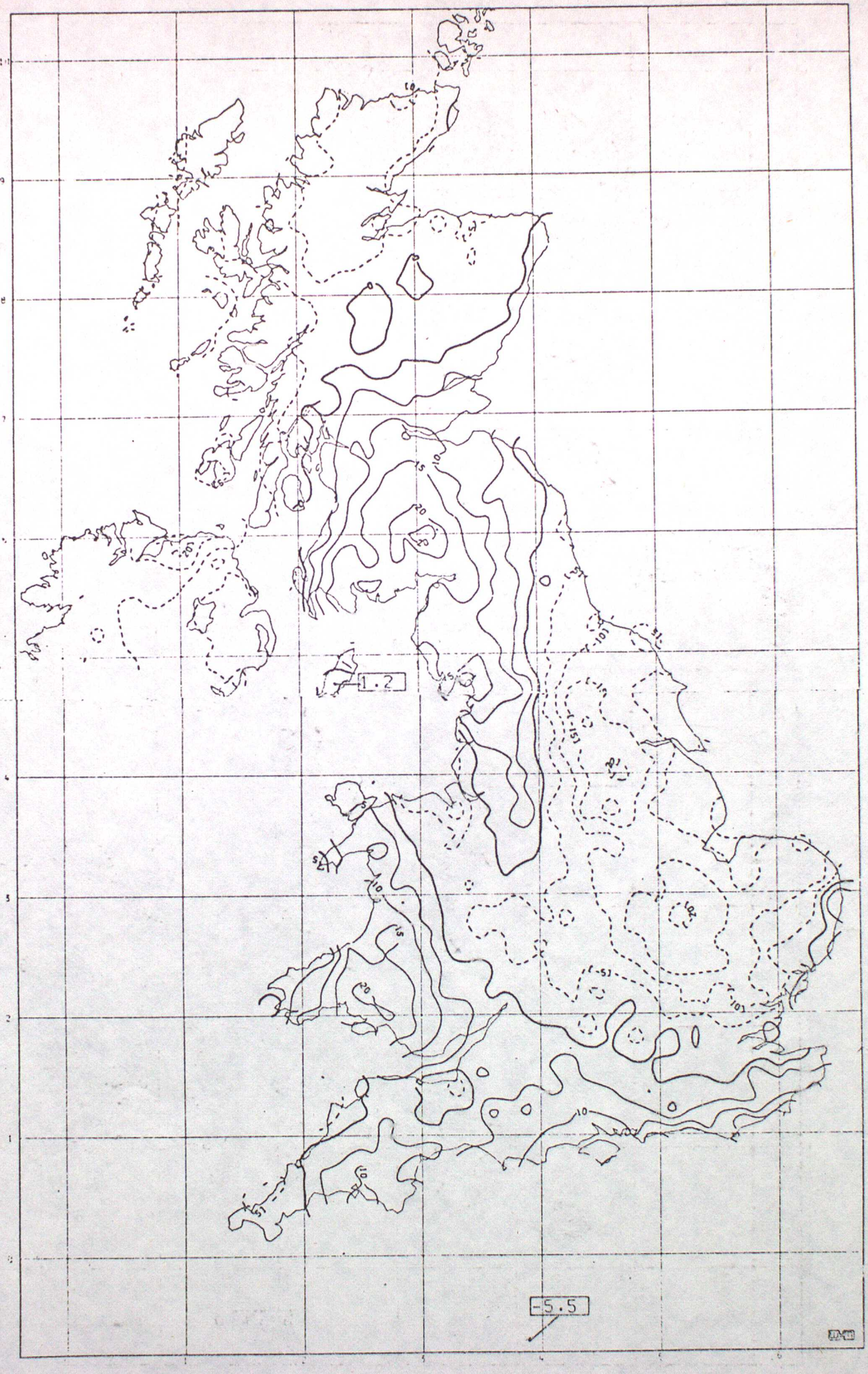



Fig. 3(f) Daily Minimum Temperature 1973-77 - Factor 6

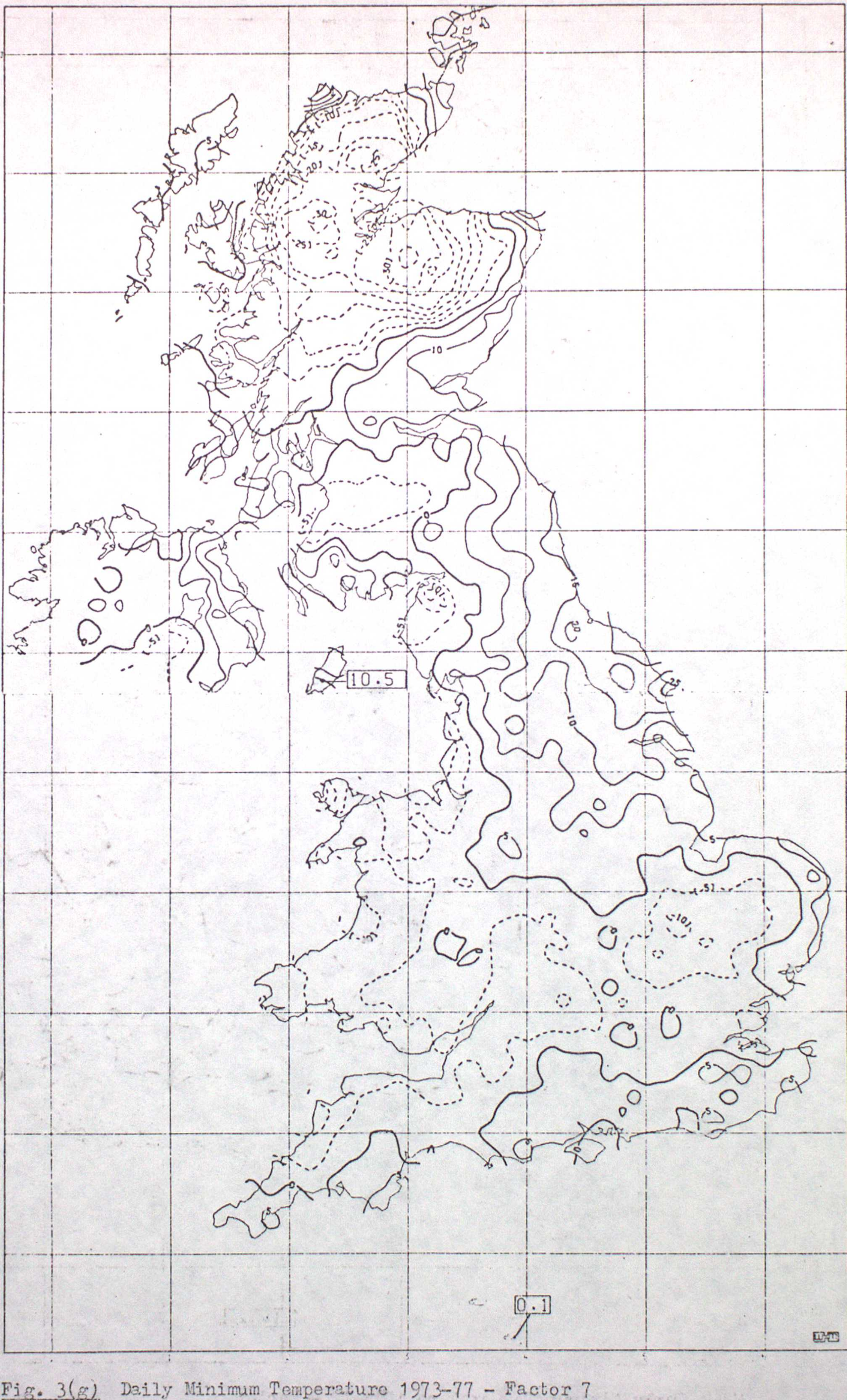



Fig. 3(g) Daily Minimum Temperature 1973-77 - Factor 7

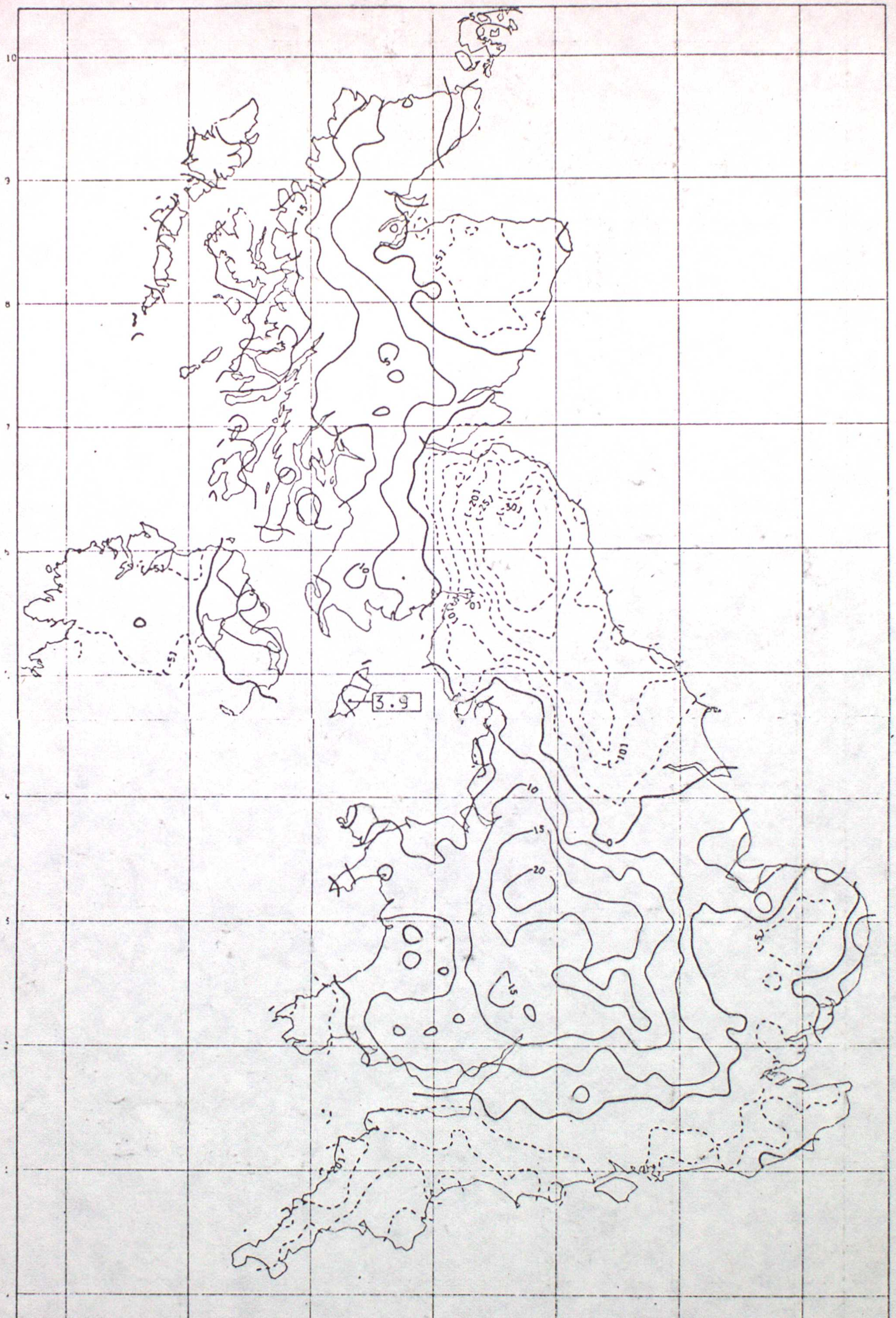



Fig. 3(h) Daily Minimum Temperature 1973-77 - Factor 8

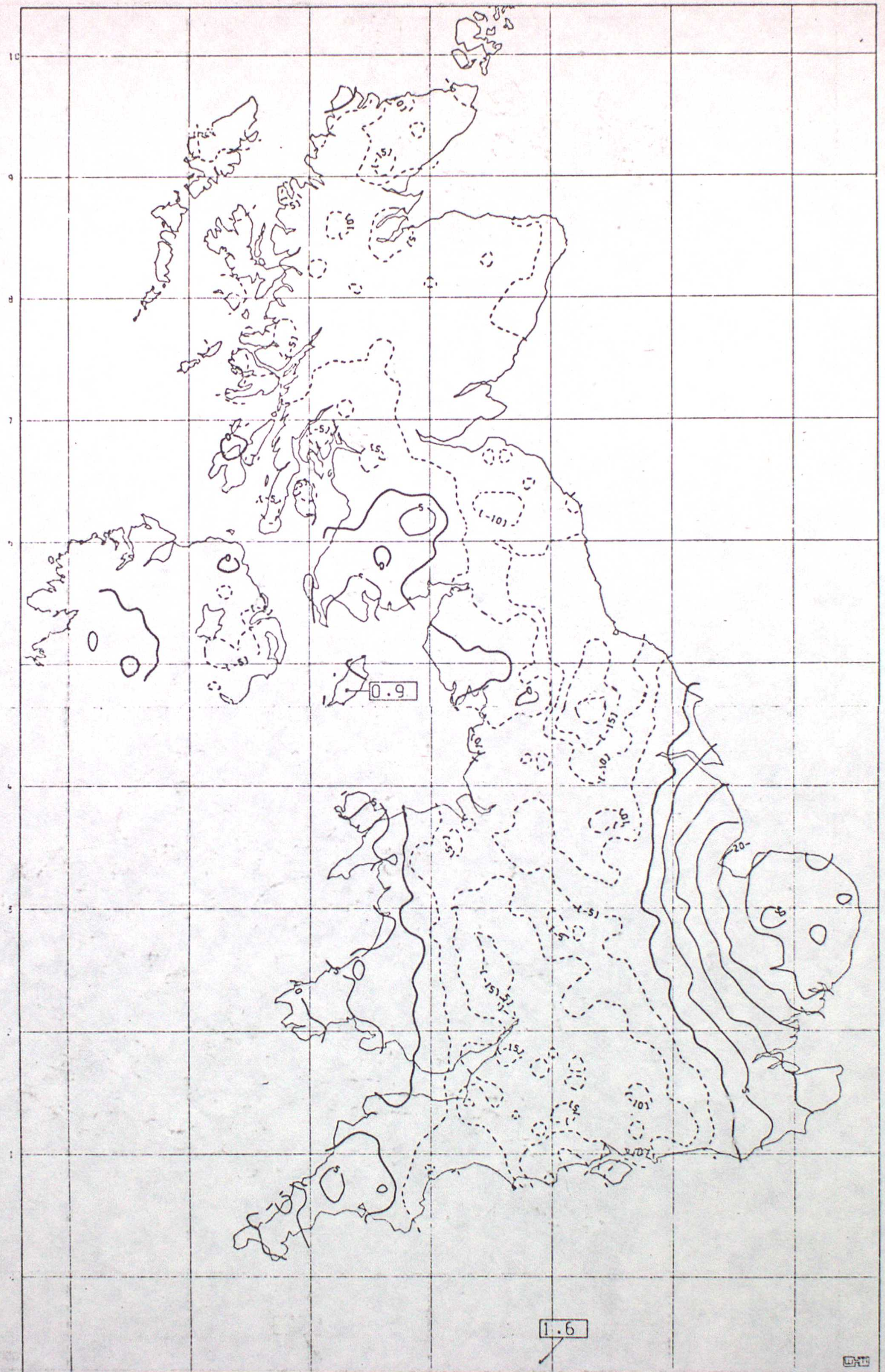



Fig. 3(i) Daily Minimum Temperature 1973-77 - Factor 9
(Reduced to m.s.l. using factor score lapse of -0.00374 per metre)

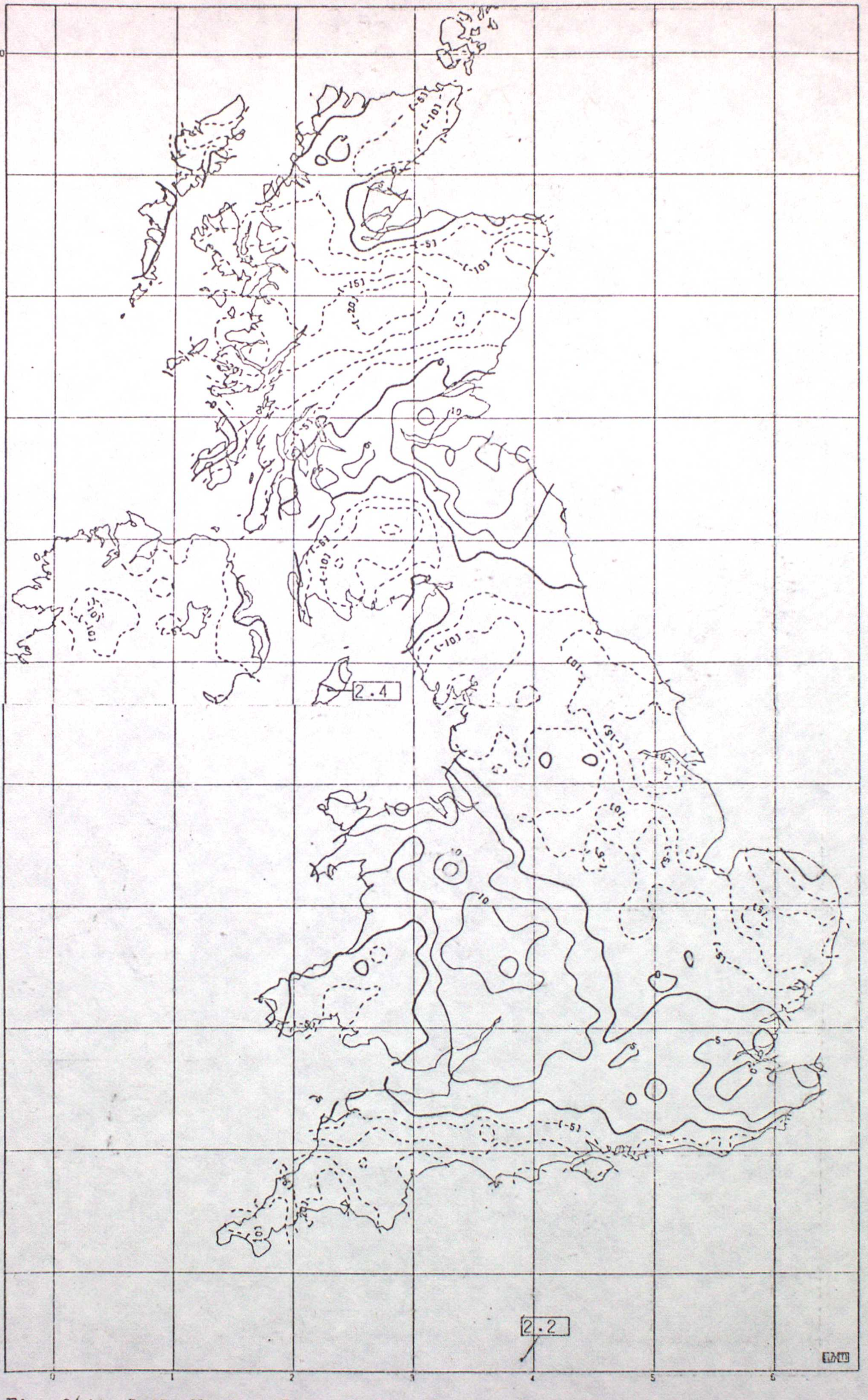



Fig. 3(j) Daily Minimum Temperature 1973-77 - Factor 10
(Reduced to m.s.l. using factor score lapse of -0.00271 per metre)

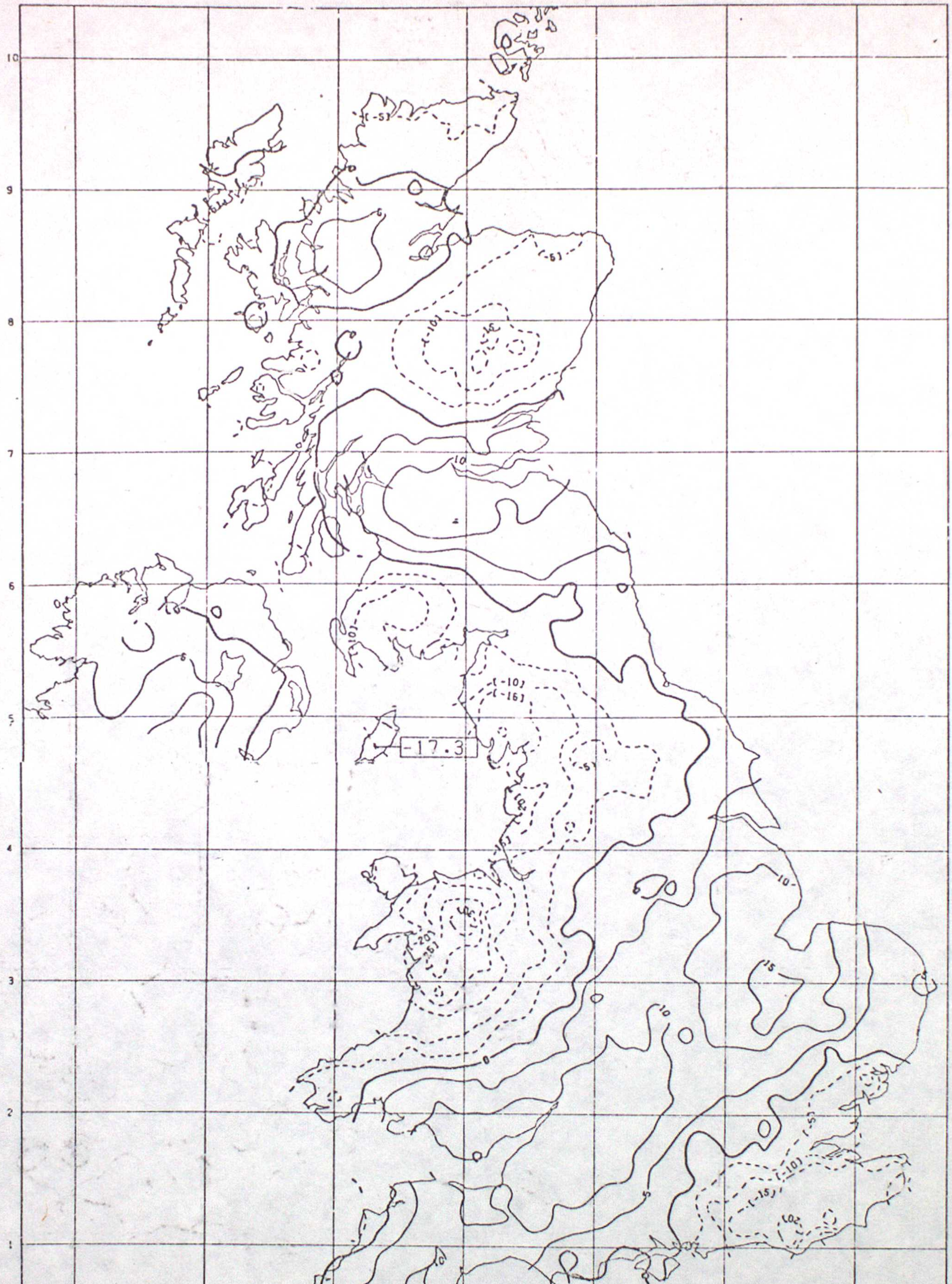



Fig. 3(k) Daily Minimum Temperature 1973-77 - Factor 11


Fig. 3(1) Daily Minimum Temperature 1973-77 - Factor 12
(Reduced to m.s.l. using factor score lapse of -0.00197 per metre)

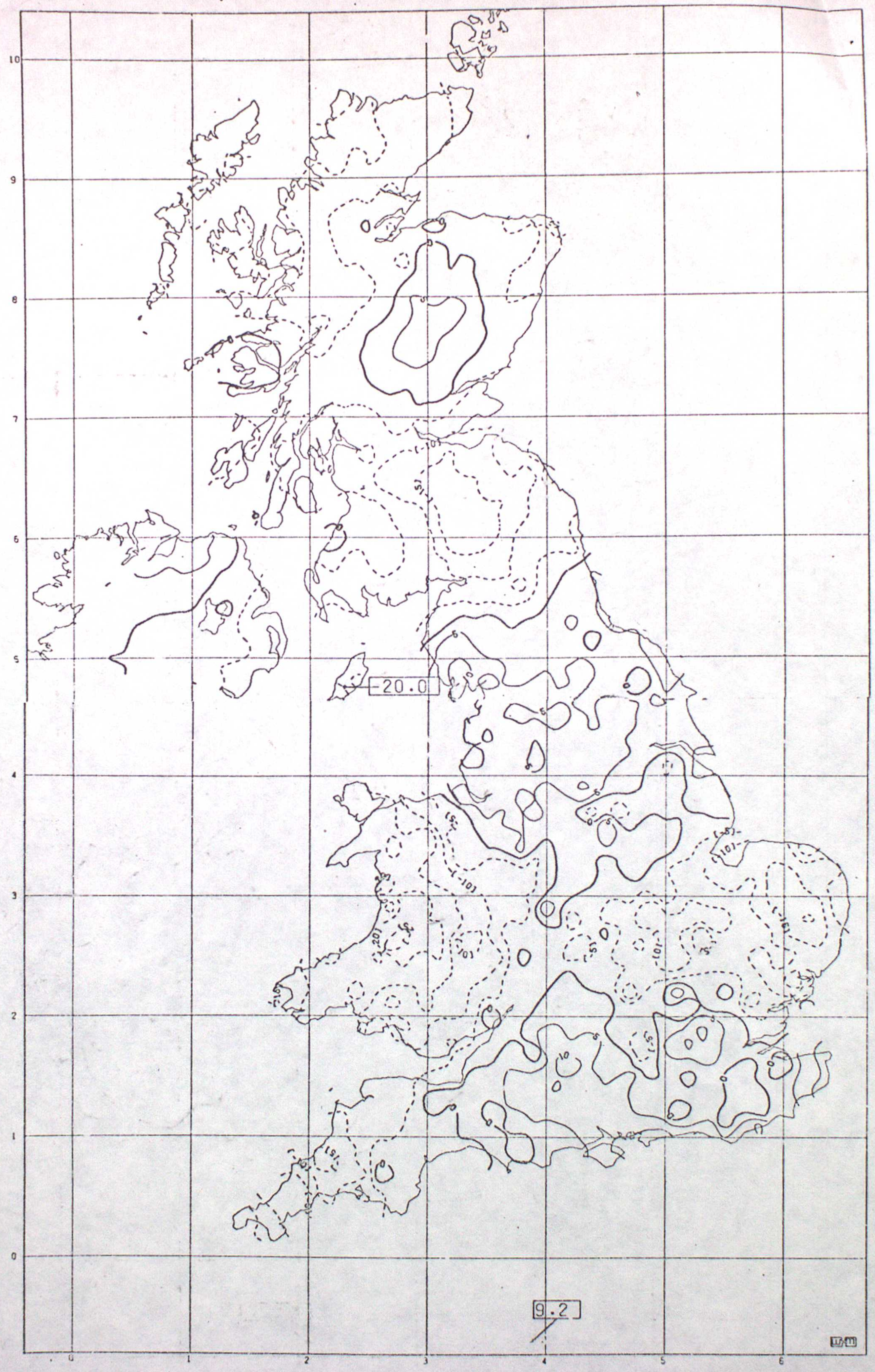



Fig. 3(m) Daily Minimum Temperature 1973-77 - Factor 13
(Reduced to m.s.l. using factor score lapse of -0.00385 per metre)

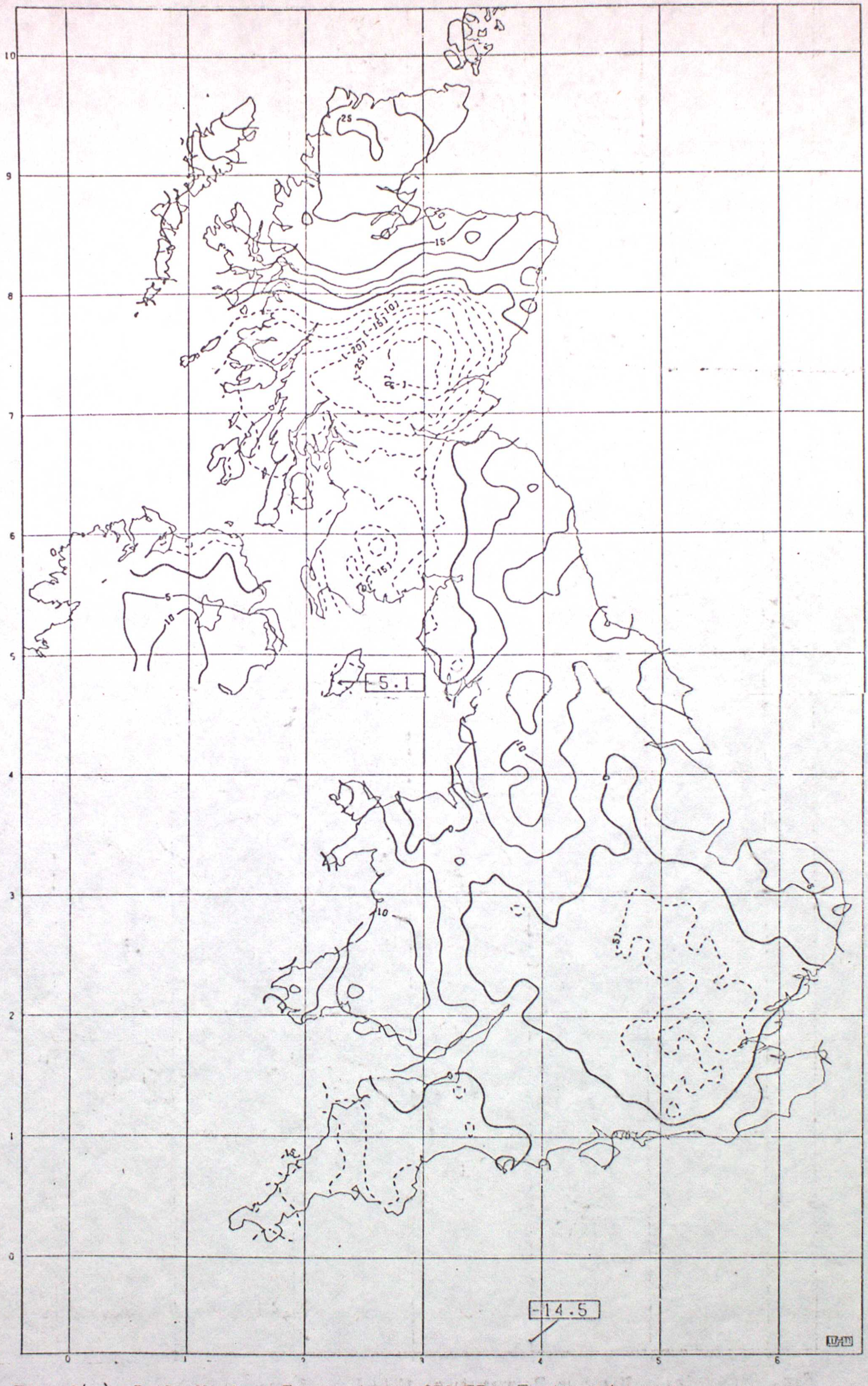



Fig. 3(n) Daily Minimum Temperature 1973-77 - Factor 14

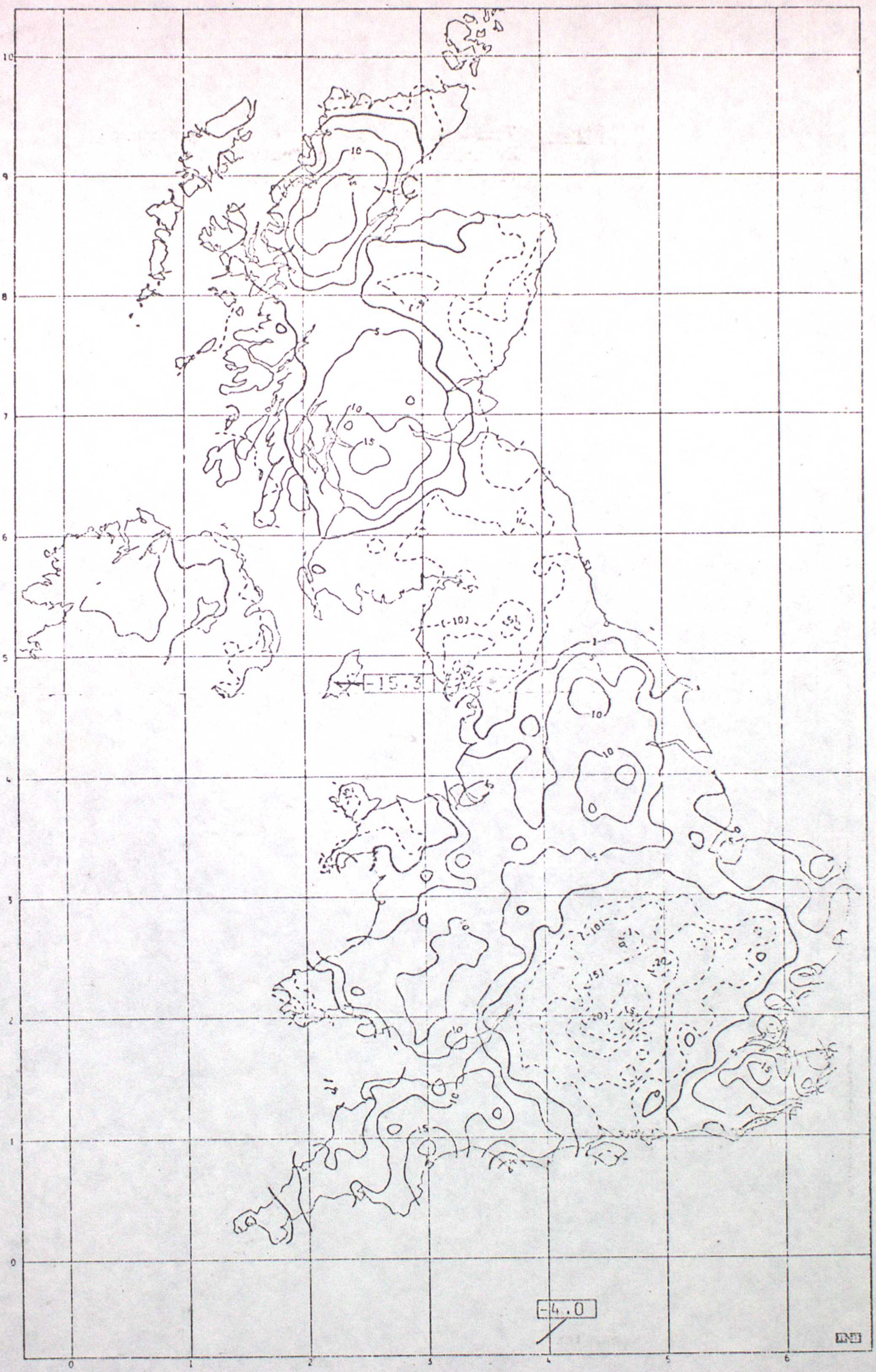



Fig. 3(o) Daily Minimum Temperature 1973-77 - Factor 15

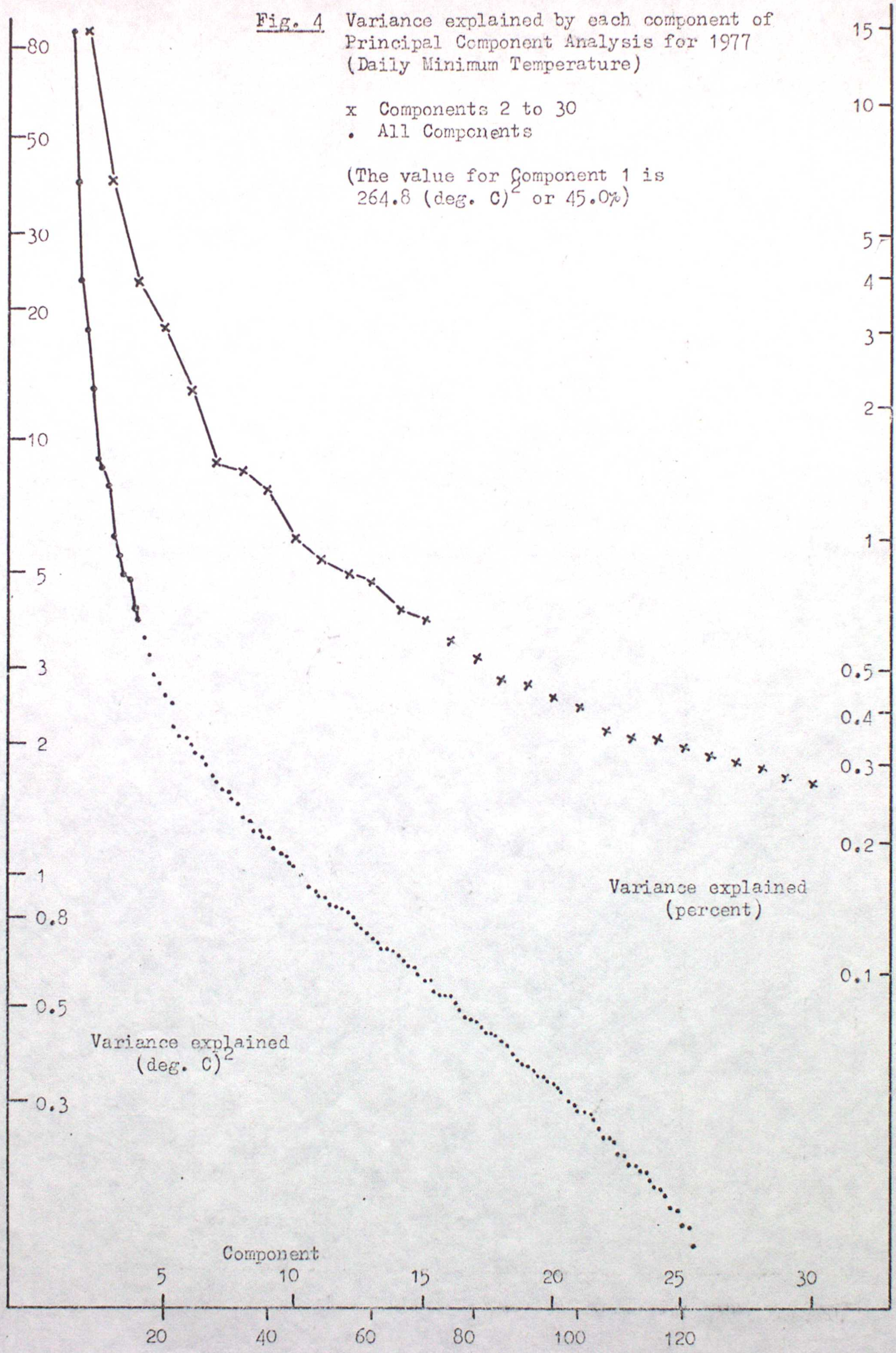


Fig. 4 Variance explained by each component of Principal Component Analysis for 1977 (Daily Minimum Temperature)
x Components 2 to 30
. All Components
(The value for Component 1 is 264.8 (deg. C)2 or 45.0%)
80
50
30
20
10
5
3
2
1
0.8
0.5
0.3
15
10
5
4
3
2
1
0.5
0.4
0.3
0.2
0.1
Variance explained (percent)
Variance explained (deg. C)2
Component
5
10
15
20
25
30
20
40
60
80
100
120

This page is intentionally left blank

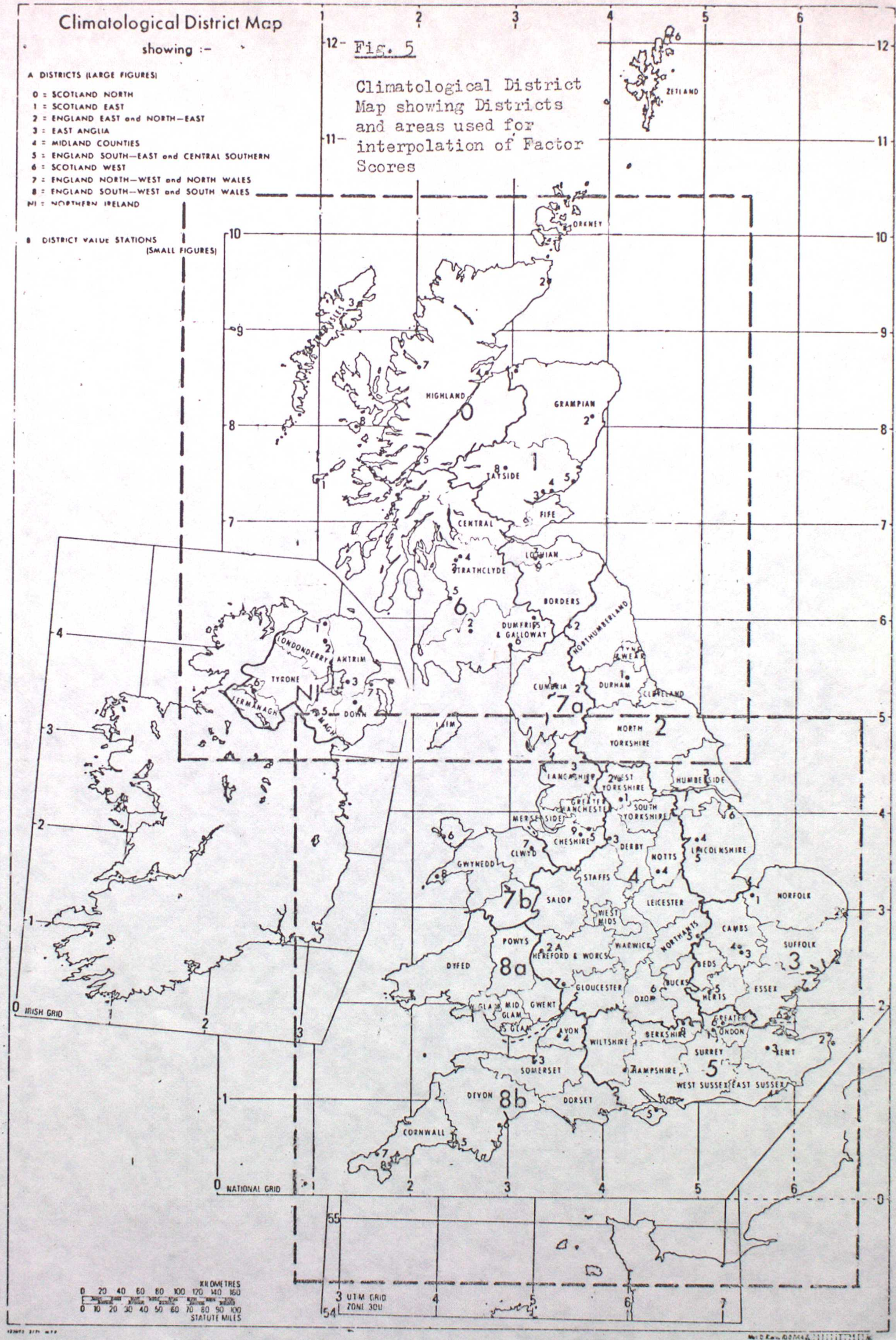



Fig. 5

Climatological District Map showing Districts and areas used for interpolation of Factor Scores

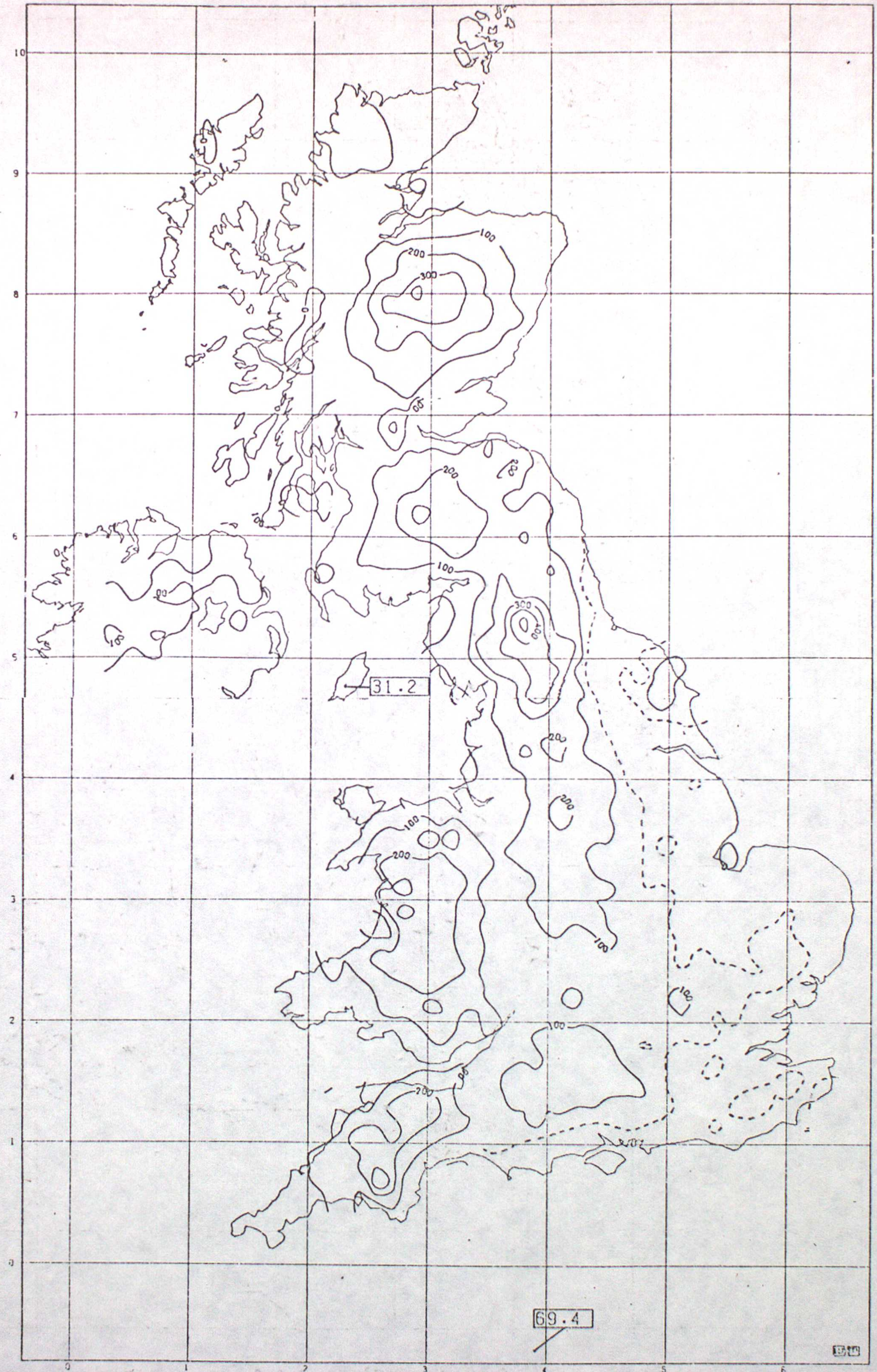



Fig. 6 Topography of U.K. interpolated from values at Climatological Stations (Contours are at 100 metre intervals. 50 metre contour - over south and east of England only - is shown as a pecked line)

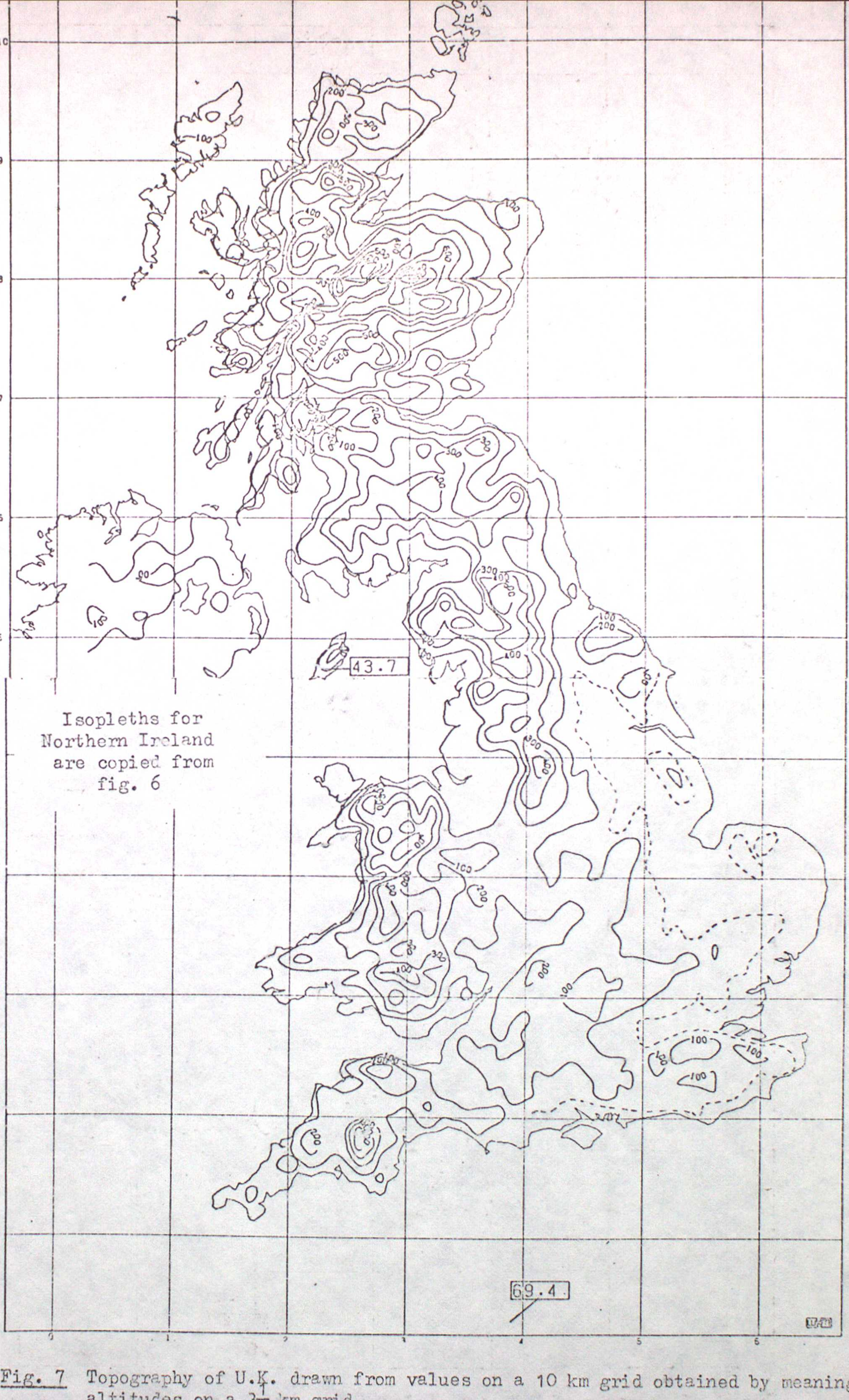
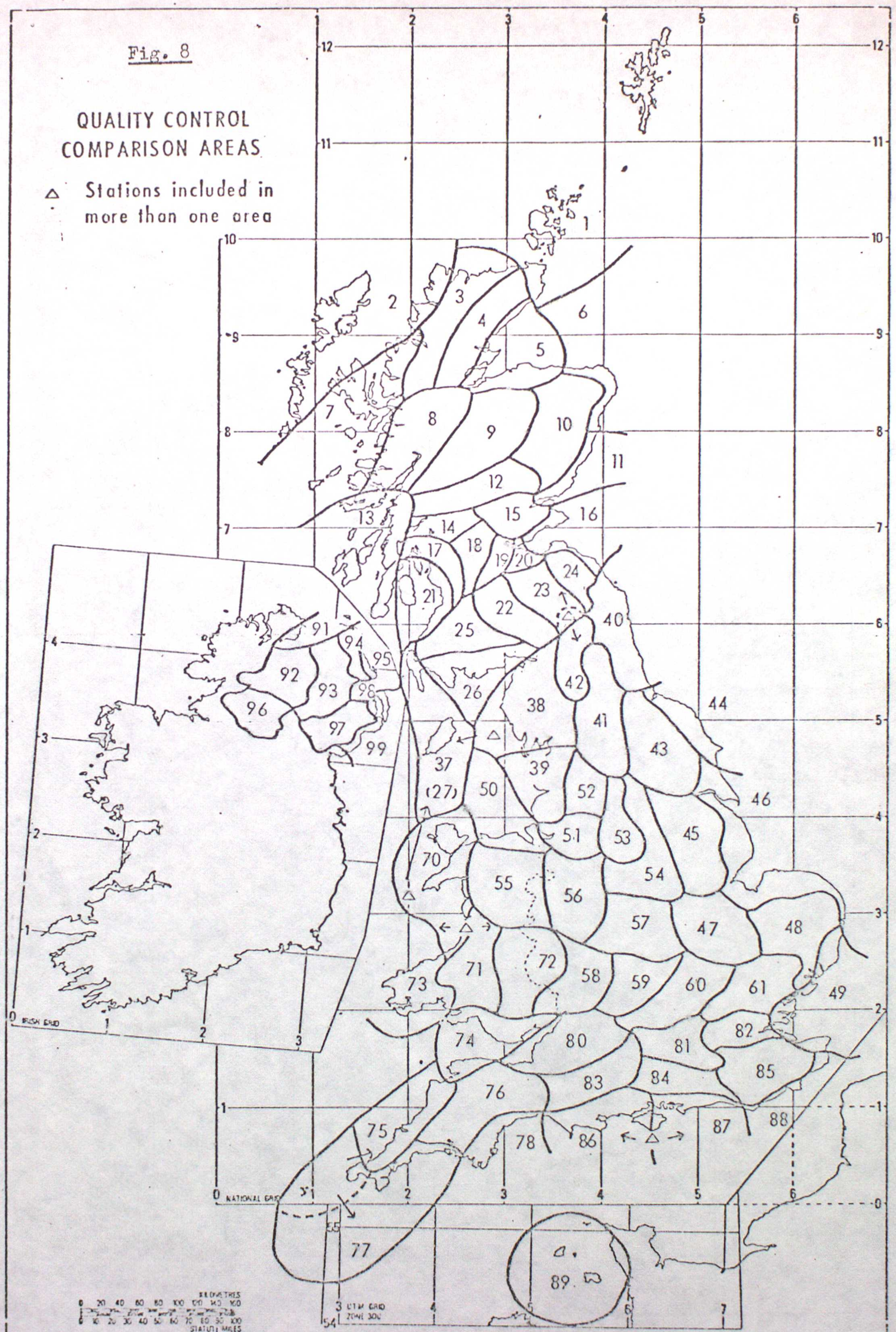
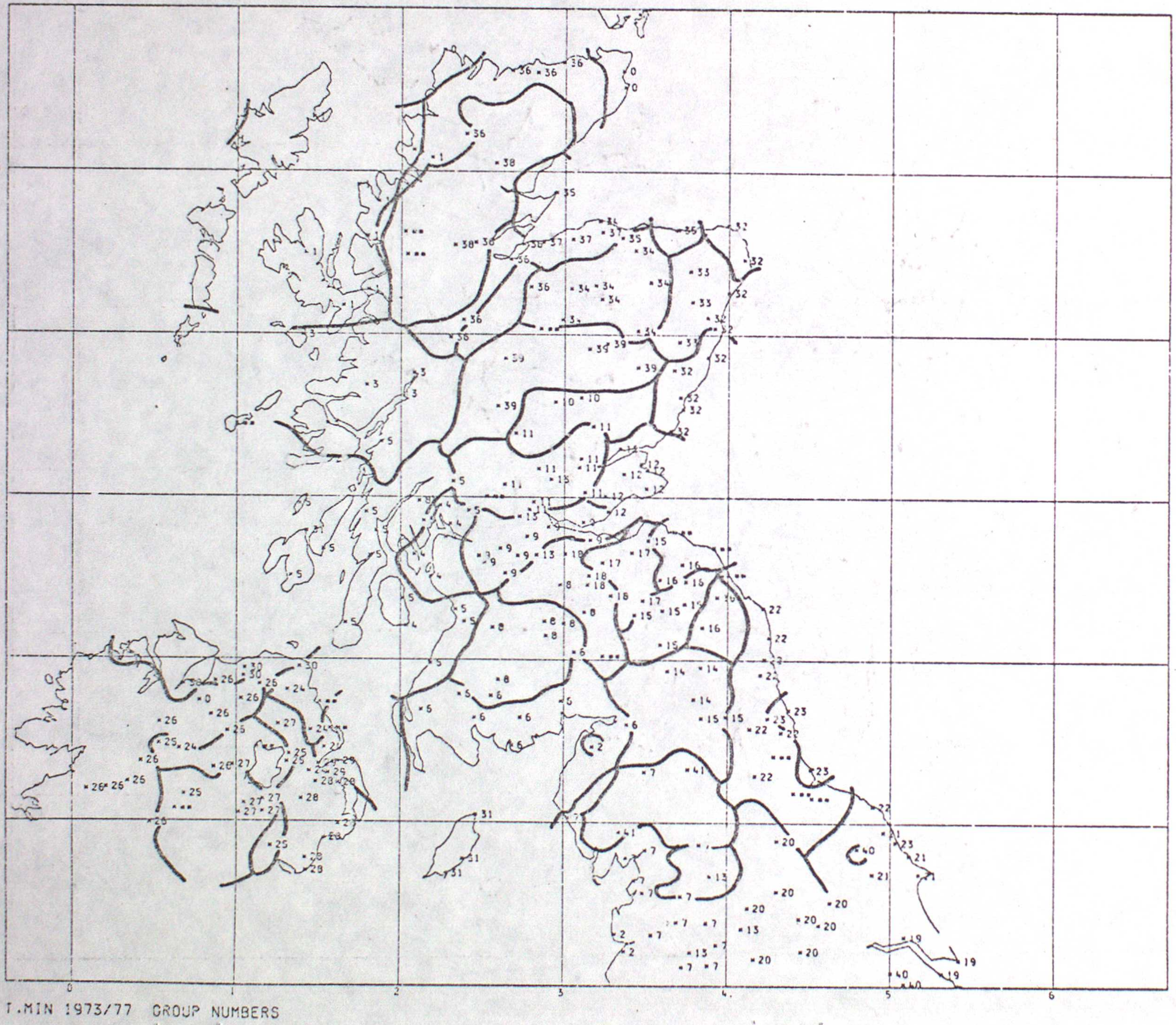



Fig. 7 Topography of U.K. drawn from values on a 10 km grid obtained by meaning altitudes on a $3\frac{1}{3}$ km grid.
(Contours are at 100 metre intervals. 50 metre contour - over south and east of England only - is shown as a pecked line)

Fig. 8


QUALITY CONTROL
COMPARISON AREAS
△ Stations included in more than one area
IRISH GRID
NATIONAL GRID
U.T.M. GRID ZONE 30U
KILOMETRES
STATUTE MILES




## Fig. 9(a)

Group numbers for north of U.K. from Cluster Analysis of unweighted Factor Scores of 15 Factors for Daily Minimum Temperature - 1973-77

The lines indicating the regions are simplified from those in fig. 11(a)

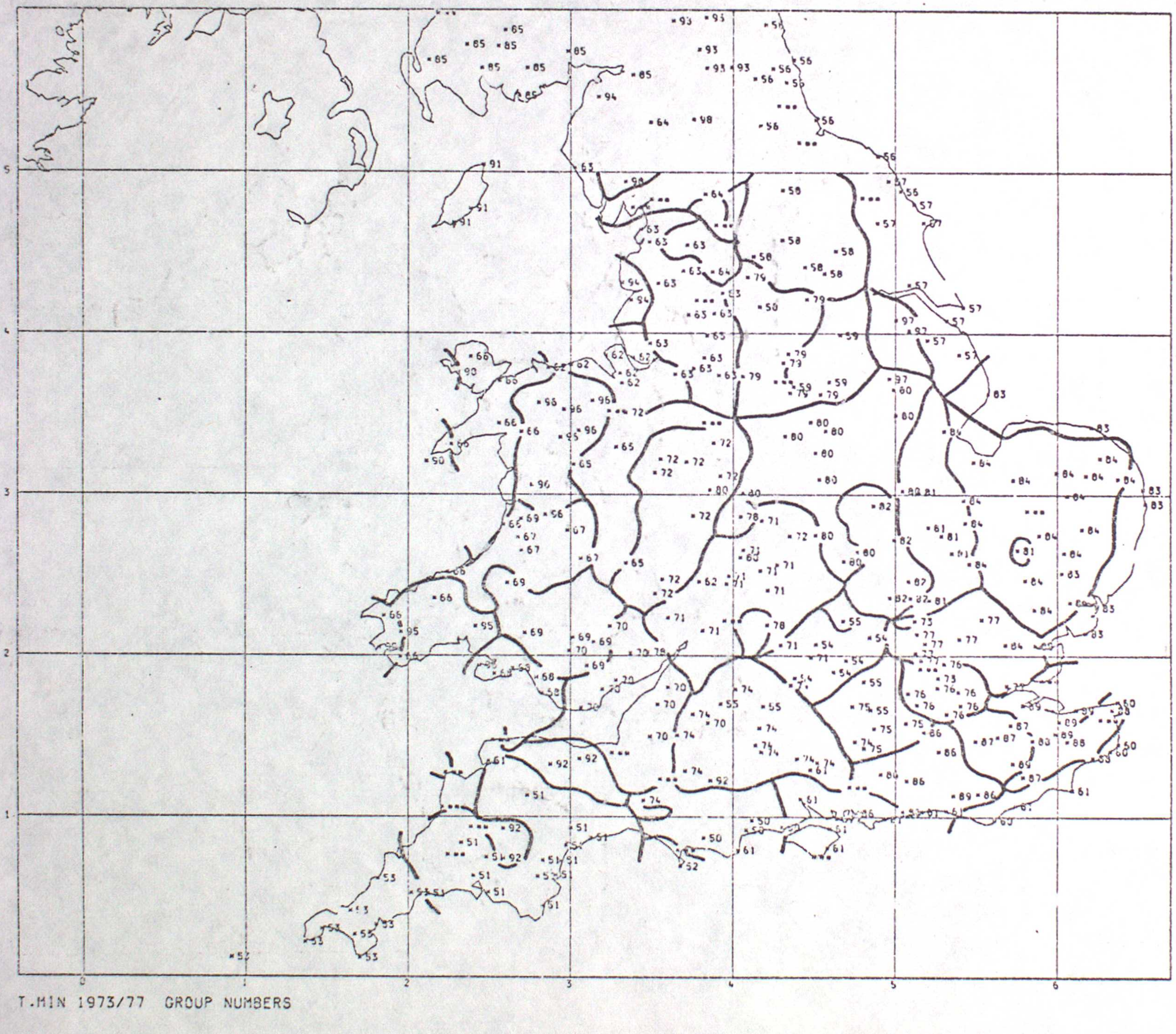



Fig. 9(b)

Group numbers for south of U.K. from Cluster Analysis of unweighted Factor Scores of 15 Factors for Daily Minimum Temperature - 1973-77

The lines indicating the regions are simplified from those in fig. 11(b)

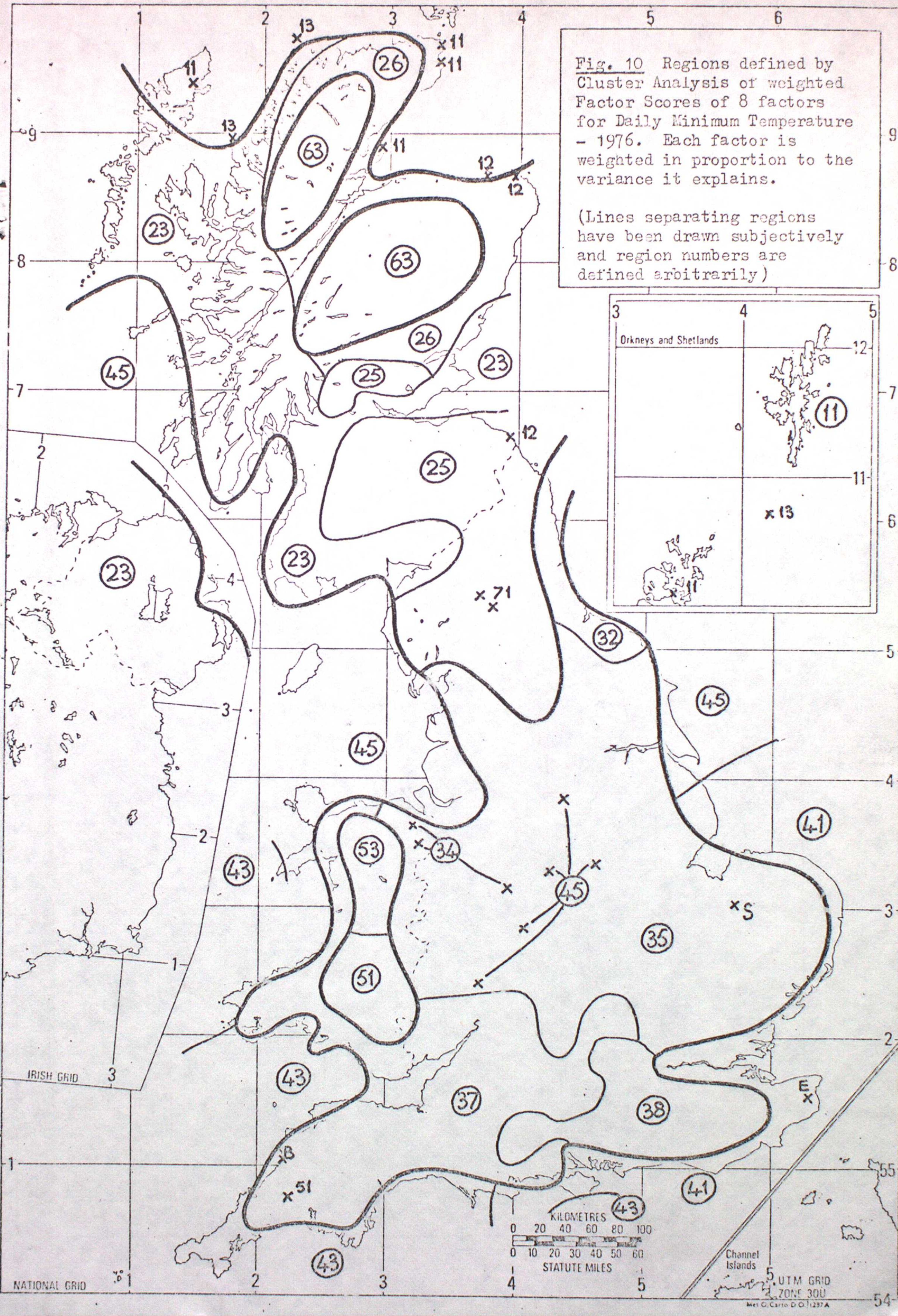



Fig. 10 Regions defined by Cluster Analysis of weighted Factor Scores of 8 factors for Daily Minimum Temperature - 1976. Each factor is weighted in proportion to the variance it explains.

(Lines separating regions have been drawn subjectively and region numbers are defined arbitrarily)

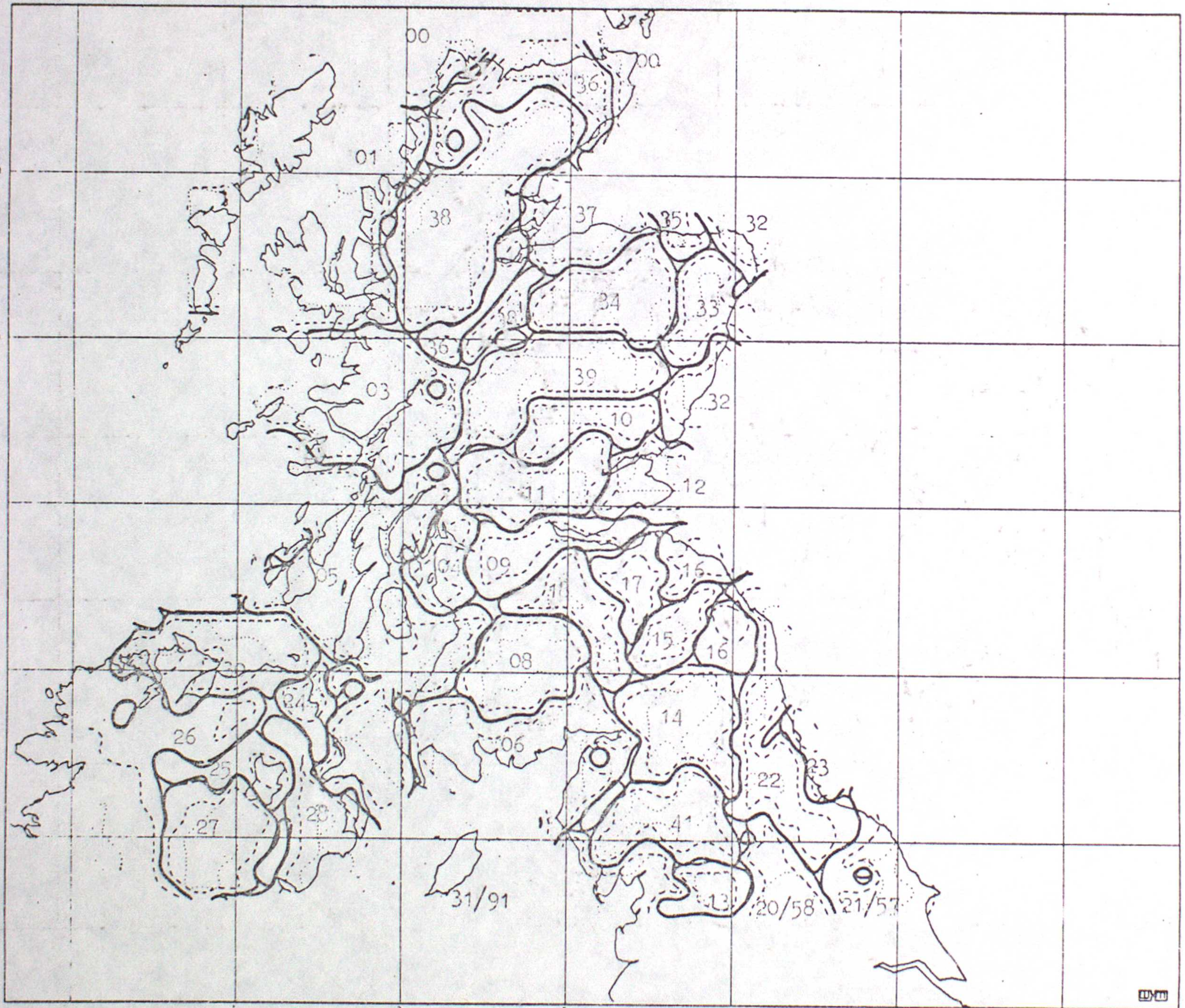



Fig. 11(a)

Regions defined by Discriminant Analysis for north of U.K. using Factor Scores of 15 Factors for Daily Minimum Temperature - 1973-77

The lines are isopleths of belonging to particular groups

—— = 50%
---- = 95%
...... = 99.9%

Numbers identifying the groups are the same as those in fig. 9(a)

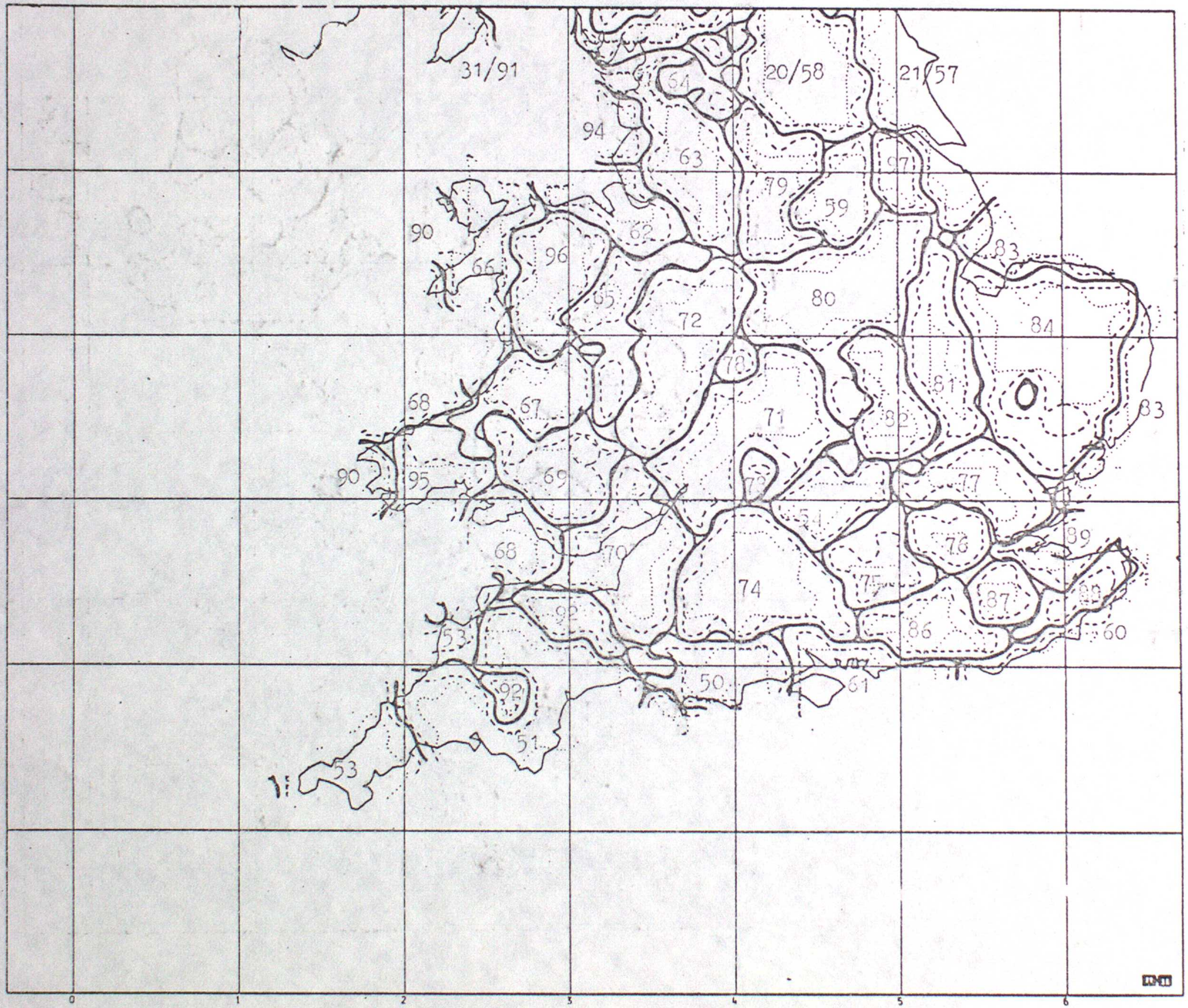



Fig. 11(b)

Regions defined by Discriminant Analysis for south of U.K. using Factor Scores of 15 Factors for Daily Minimum Temperature - 1973-77

The lines are isopleths of belonging to particular groups

—— = 50%
---- = 95%
...... = 99.9%

Numbers identifying the groups are the same as those in fig. 9(b)


Fig. 12 Minimum value of Mahalanobis Distance to any group using groups defined in fig. 9(a) and fig. 9(b) for Daily Minimum Temperature - 1973-77 (Values are standardised distances x 10)

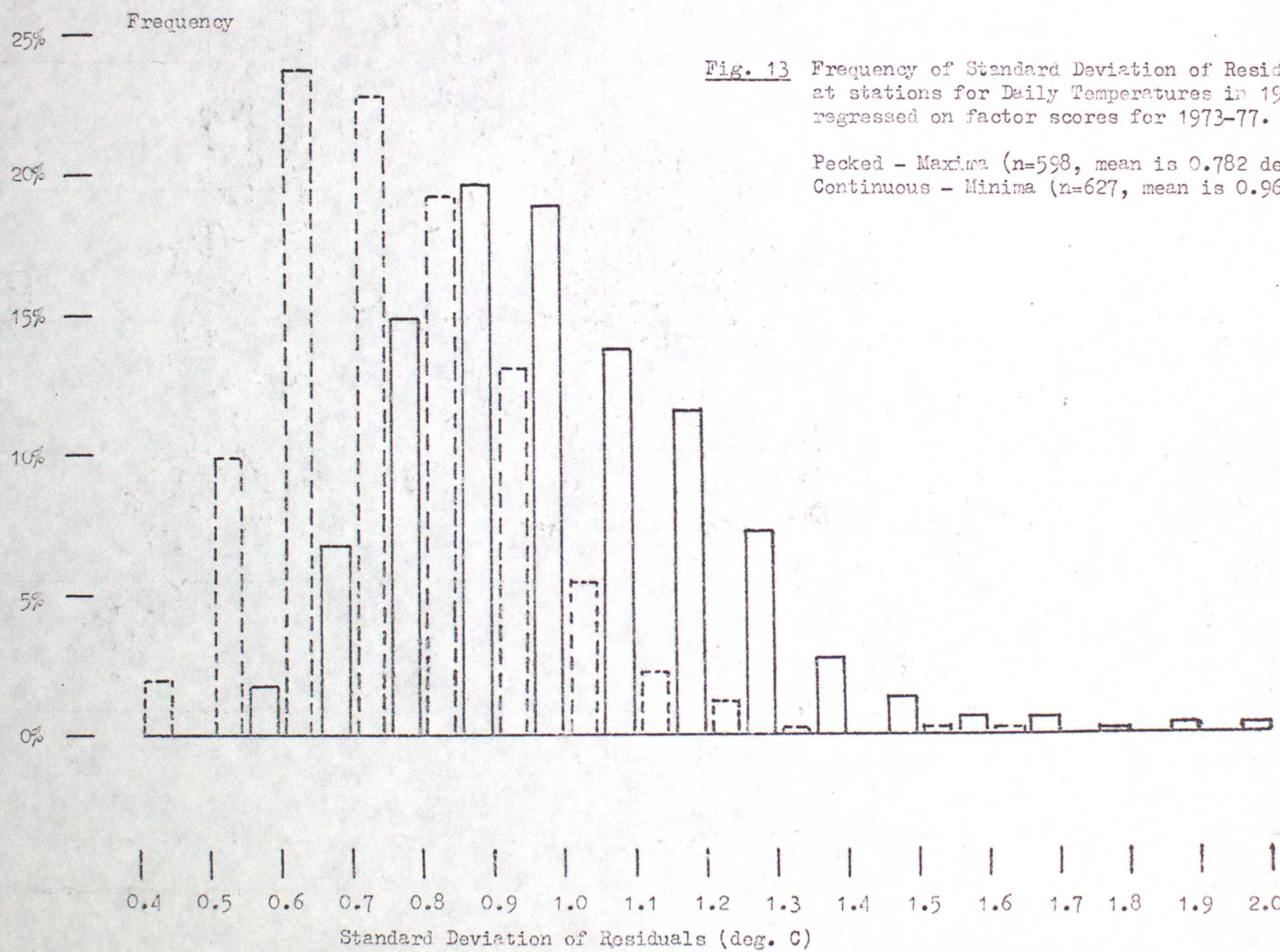



Fig. 13 Frequency of Standard Deviation of Residuals at stations for Daily Temperatures in 1974 regressed on factor scores for 1973-77.

Pecked - Maxima (n=598, mean is 0.782 deg. C)
Continuous - Minima (n=627, mean is 0.965 deg. C)

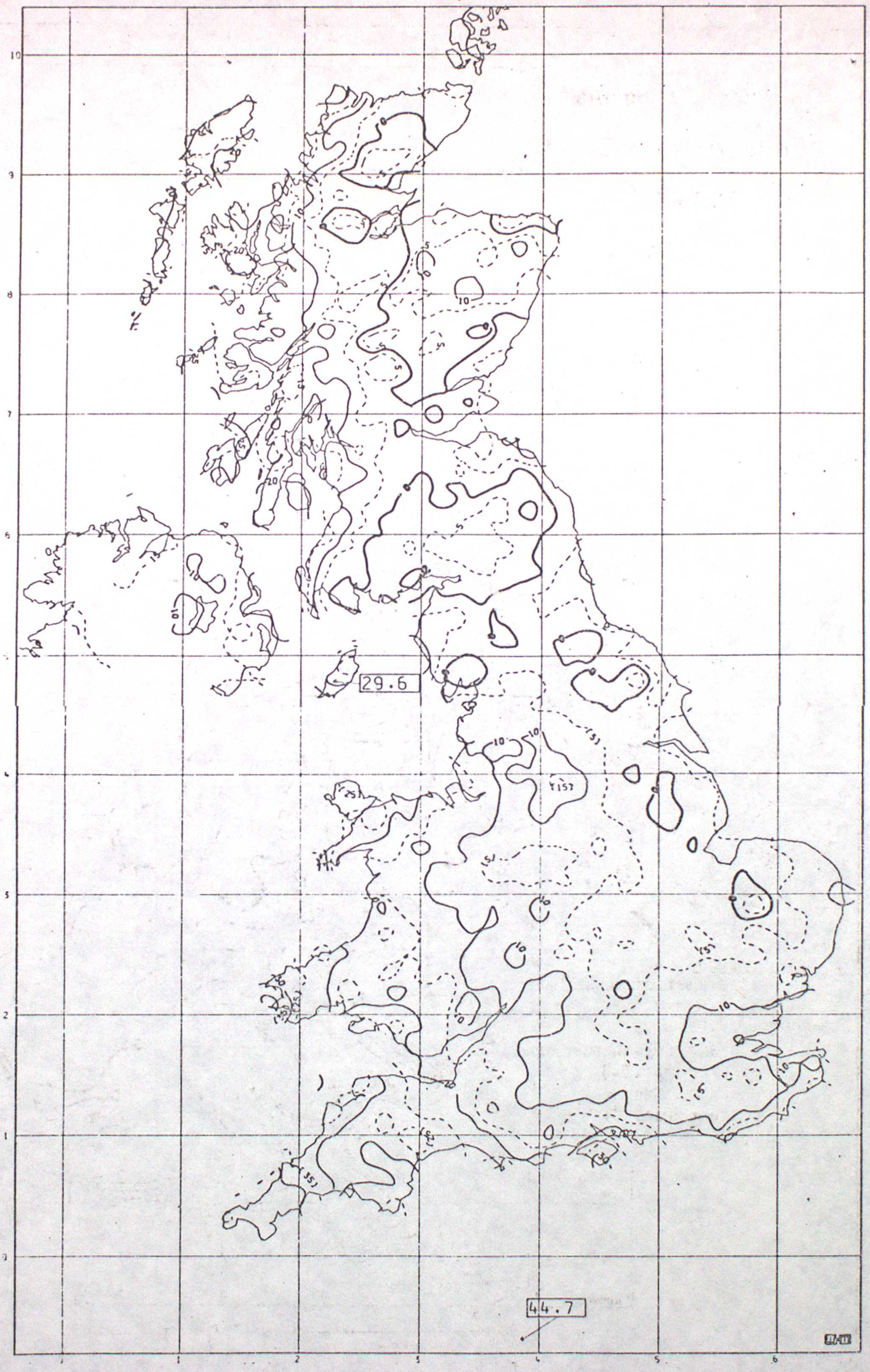



Fig. 14(a) Climatological Normal for January (1941–70) reduced to m.s.l. derived from 15 Factors of Daily Minimum Temperature for 1973–77 (Values are deg.C x 10 and are reduced to m.s.l. using a lapse rate of 5 deg.C per 1000 metres)

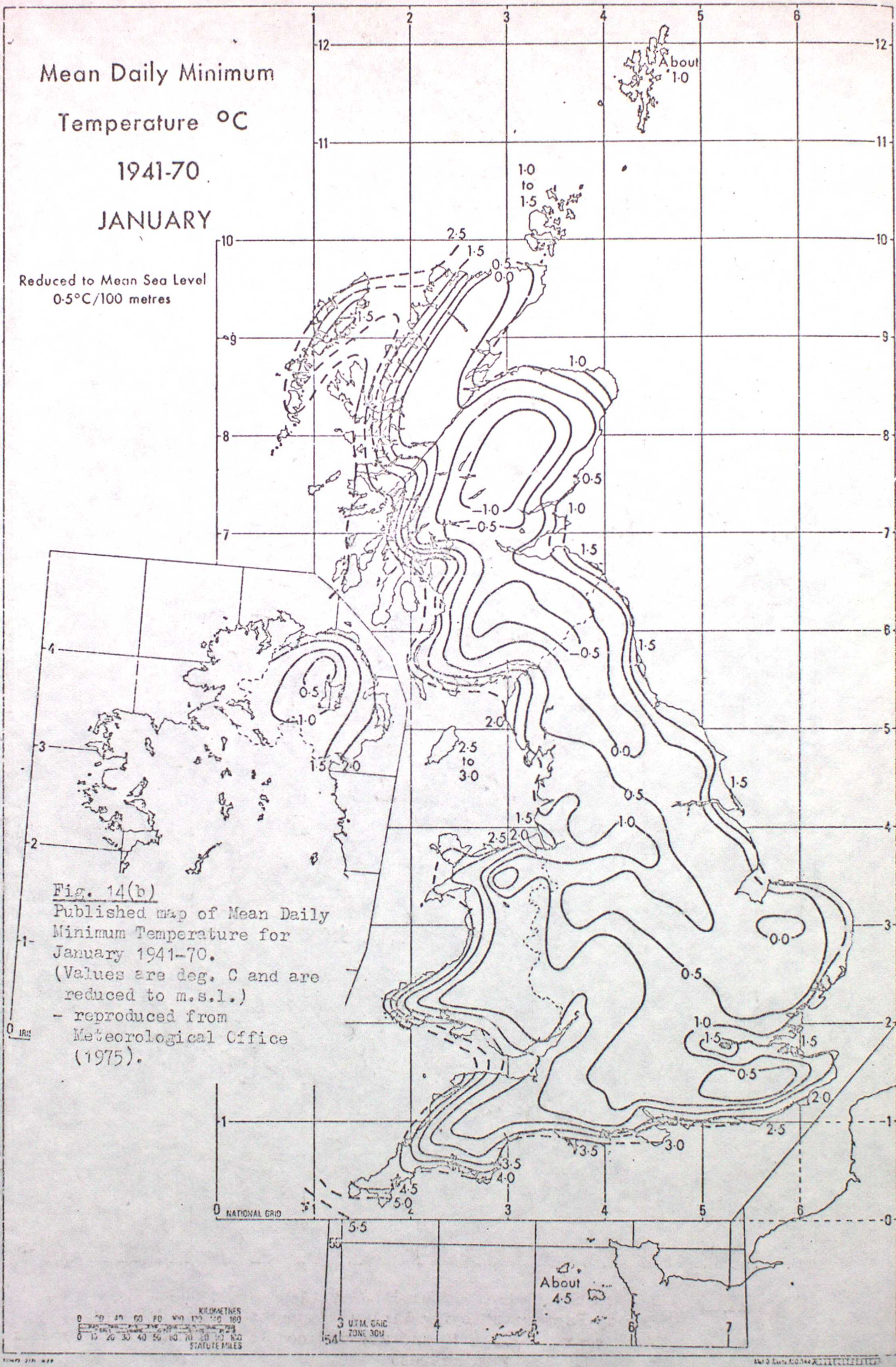



Fig. 14(b)
Published map of Mean Daily Minimum Temperature for January 1941–70.
(Values are deg. C and are reduced to m.s.l.)
– reproduced from Meteorological Office (1975).

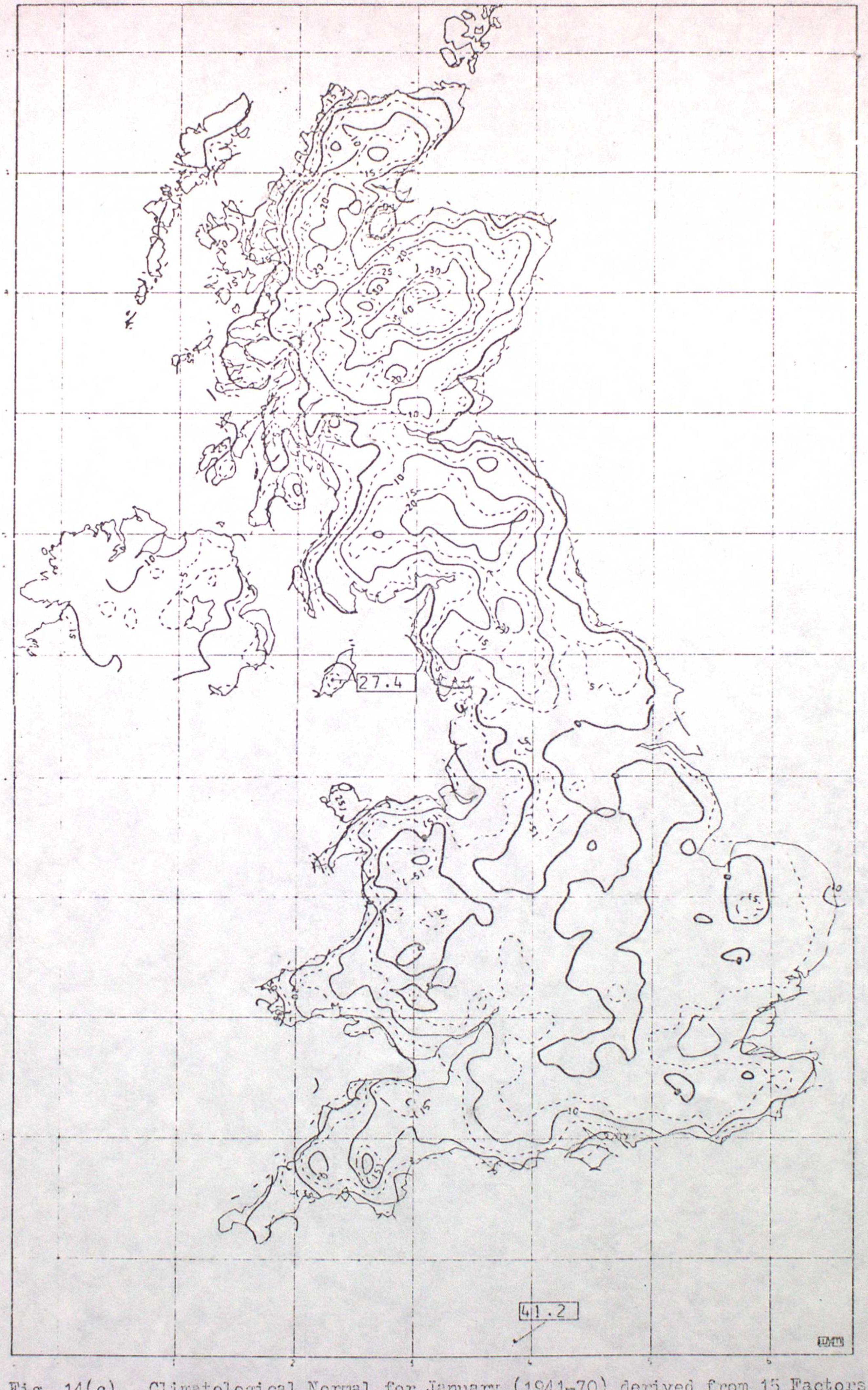



Fig. 14(c) Climatological Normal for January (1941-70) derived from 15 Factors of Daily Minimum Temperature for 1973-77 (not reduced to m.s.l.) (Values are deg. C x 10)