SOME STUDIES OF RAINFALL VARIATIONS OVER THE BRITISH ISLES

Thesis submitted by

JOHN ANDREW BLAIR-FISH

for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF EDINBURGH

1976

ABSTRACT

This investigation of mainly Scottish rainfall is in two parts: studies of space variations (first part) and of time variations (second part). Matrices of mean monthly rainfall for networks of stations in Central Scotland are decomposed using T-mode eigenvector analysis to identify common spatial variations; and daily rainfall values for selected "pure" synoptic situations, and also annual values over a gauge network in the Solway region, are regressed with physical parameters. The most important set of eigenvector spatial multipliers, describing over 95 per cent of the variance of the matrix, is used to interpolate between stations and "predict" mean monthly rainfall for new sites. Multiple linear regression relationships between rainfall on the one hand and altitude and distance from South and West coasts on the other, are compared for different cases. The validity of a linear approximation to rainfall variations in relation to physical parameters is discussed, using values of regression and correlation parameters and station regression residuals.

In the second part of the thesis, aspects of rainfall time-series, in particular the persistence of spells of wet and dry days and periodicities in annual and monthly series, are investigated. The simple Markov model and various modifications to it are used to describe both the distribution of spells of all lengths, and also of those greater than a specified length. Two further models, relating respectively to persistence patterns and to the occurrence of rare events, are also discussed.

The methods of filtering, and power spectrum, of time-series

are employed to identify periodicities; the results for different stations are compared to help assess the significance of individual results and also to distinguish variations common to "West" and "East" stations respectively. S-mode eigenvector analysis is also used to derive common time variations in different rainfall records. The results of filtering and power spectrum analysis of individual rainfall records and of eigenvector multiplier time series are compared with those of similar analyses of P and C atmospheric circulation indices (which describe the frequency of weather types effecting most rainfall); relationships between oscillations and the above indices are also investigated using cross spectrum analysis. The most important results of these time series analyses are the confirmation of the presence of an oscillation of period close to 2.0 years which occurs in series of P and C indices and rainfall; also, the revelation of a significant 3.1 year periodicity in C index and "East" station rainfall.

PREFACE

This thesis was written by John Andrew Blair-Fish using results of research work carried out by him in the Department of Meteorology, University of Edinburgh with a National Environment Research Council studentship.

ACKNOWLEDGEMENTS

I should like to thank Dr. D. H. McIntosh, my Supervisor and Head of the Meteorology Department, University of Edinburgh for the opportunity to pursue this research and his guidance in the pursuit of the work; also, Dr. A. S. Thom, Dr. K. J. Weston, Dr. I. Edwards, Dr. C. N. Duncan, Dr. A. W. Daly and Dr. J. C. Curran of the same Department for their advice, constructive criticism, and encouragement.

I should also like to express my gratitude to members of the Meteorological Office for useful discussions and for supplying the appropriate data, and to members of the Climatic Research Unit, East Anglia University for their suggestions on time-series analysis of climatic data; to Mrs. L. R. Edwards for her accurate typing of the manuscript; and finally, to the National Environment Research Council for financial support during the three years of the work.

CONTENTS

LIST OF CONTENTS OF CHAPTERS 1-5 AND APPENDICES

Page

Page

TABLES

FIGURES

REFERENCES

D.P.M. = "The Distribution of Precipitation in Mountainous Area", W.M.O. Tech. Note No. 326.

Alam, F.C.K. (1972): "Distribution of precipitation in mountainous areas of West Pakistan". D.P.M. Vol. II, pp 296 - 306.

Bleasdale, A. (1963): "Estimation of standard period averages for stations with incomplete data". Hyd. Mem. London, No. 5.

Bleasdale, A. and Chan, Y.K. (1972): "Orographic influences on the distribution of precipitation". D.P.M. Vol. II, pp 322-334.

Browning, K.A., Hill, C.W., and Pardoe, E.F. (1974): "The structure and mechanism of precipitation and the effect of orography in a winter-time warm sector". Q.J.R. Met. S., Vol. 100, pp 309-330.

Browning, K.A., Hill, C.W., and Pardoe, E.F. (1975): "The nature of orographic rain at winter-time cold fronts". Q.J.R. Met. S., Vol. 101, pp 333-353.

Brunt, D.C. (1925): "Periodicities in European Weather". Phil. Trans. Roy. Soc. A., Vol 225, pp 247-302.

Chatfield, C. (1966): "Wet and dry spells". Weather, Vol. 21, pp 308-311.

Chuan, C.K. and Lockwood, J.C. (1974): "An assessment of topographic controls in the distribution of rainfall in the Central Pennines". Met. Mag., Vol. 103, pp 275-288.

Dee Weather Radar Project (1974): "The use of a radar network for the measurement and quantitative forecasting of precipitation". Water Resources Board.

Douglas, C.K.M., and Glasspool, J.C. (1947): "Heavy orographic rainfall in the British Isles". Q.J.R. Met. S., Vol. 73, pp 111-142.

English, E.J. (1973): "An objective method of calculating areal rainfall". Met. Mag., Vol. 102, pp 292-297.

Effroymsen, M.A. (1960): "Multiple regression analysis". Mathematical methods for digital computers. Part V, 17. Edited by A. Ralston and H.S. Wilf, Wiley.

Fisher, Sir R.A., and Yates, F.C. (1963): "Statistical tables for biological, agriculture and medical research". Oliver and Boyd, 6th Edition, London.

Fritts, H.C. (1971): "Multivariate techniques for calibrating time series to reconstruct anomalies in paleoclimate". Int. Symp. on Probability and Statistics in the Atmospheric Sciences, pp 59-63.

Gabriel, K.R. and Neumann, J.C. (1962): "A Markov chain model for daily rainfall occurrence at Tel Aviv". Q.J.R. Met. S., Vol. 88, pp 90-95.

Gray, B.M. (1975): "Medium term fluctuations of rainfall and temperature in South East England". Water Resources Board.

Green, J.R.C. (1970): "A generalised probability model for sequences of wet and dry days". Mon. Weather Rev. Vol. 98, pp 238-241.

Hinich, M.J., and Clay, C.S. (1968): "An application of the discrete Fourier transform in estimates of power spectra, coherence, and bispectra of geophysical data". Rev. Geophysics, Vol. 6, pp 345-363.

Hirose, M., and Kutzbach, J.F. (1969): "An alternative method of eigenvector computation". J. Applied Met., Vol. 8, p 701.

Jenkins, G.M., and Watts, D.G. (1968): "Spectral analysis and it applications". Holden-Day, Sans Francisco.

Jenkinson, A.F. (1955): "Frequency distributions of the annual maximum (or minimum) value of meteorological elements". Q.J.R. Met. S., Vol. 81, pp 158-197.

Kohler, M.A. (1949): "On the use of the double-mass analysis for testing the consistency of meteorological records and for making required adjustments". Bull. Amer. Met. S., Vol. 30, pp 188-189.

Kutzbach, J.F. (1967): "Empirical eigenvectors of sea-level pressure, surface temperature, and precipitation complexes over North America". J. Applied Met., Vol. 6, pp 791-802.

Lamb, H.H. (1950): "Types of weather around the year in the British Isles : annual trends, seasonal structure of the year, singularities". Q.J.R. Met. S., Vol. 76, pp 393-438.

Lamb, H.H. (1972a): "British Isles weather types and a register of the daily sequence of circulation patterns 1861-1971". Geophys. Mem. London, No. 117.

Lamb, H.H. (1972b): "Climate present, past, and future. Vol. 1. Fundamentals and climate now". Methuen, London.

Lawrence, E.N. (1957): "Estimation of the frequency of runs of dry days". Met. Mag., Vol. 86, pp 225-269, pp 301-303.

Lefevre, R. (1972): "Aspect de la pluviometrie dans la region du Mont Cameron". D.P.M. Vol. II, pp 290-306.

Le Marche, V.C., and Fritts, H.C. (1971): "Anomaly patterns of climate over Western United States, 1700-1930, derived from principal component analysis of tree ring data". Mon. Weather Rev., Vol. 99, pp 138-142.

Manley, G. (1974): "Central England temperatures, monthly means 1659 to 1973". Q.J.R. Met. S., Vol. 100, pp 389-405.

Mitchell, J.M., et al (1966): "Climatic Change". W.M.O. Tech. Note No. 79.

Mossman, R.C. (1896): "The Meteorology of Edinburgh". R. Grant, Edinburgh.

Murray, R., and Lewis, R.P.W. (1966): "Some aspects of synoptic climatology of the British Isles as measured by simple indices". Met. Mag., Vol. 95, pp 193-203.

Murray, R., and Benwell, P.R. (1970): "PSCM indices in synoptic climatology and long-range forecasting". Met. Mag., Vol. 99, pp 232-245.

Peck, E.L. (1972): "Relation of orographic winter precipitation to meteorological parameters". D.P.M. Vol. II, pp 234-243.

Pedersen, K. (1972): "Orographic influences on the distribution of precipitation". D.P.M. Vol. II, pp 161-171.

Rayner, J.N. (1971): "An introduction to spectral analysis". Pion, London.

Rodda, J.C. (1967): "The rainfall measurement problem". Geochemistry, Precipitation, Evaporation, Soil Moisture, Hydrometry. General Assembley of Bern. Sept-Oct 1967, pp 215-230.

Rodda, J.C. (1970): "On the question of rainfall measurement and representativeness". Int. Ass. Scientific Hydrology, pp 174-185.

Ross, G.H. (1975): "Maximum Entropy Spectral Analysis". Occasional Note, Met. O2175/2, Geophysical Fluid Dynamics Laboratory, Meteorological Office, Bracknell, Berks.

Salter, P.M. (1972): "Areal rainfall analysis by computer". D.P.M., Vol. II, pp 49-510.

Schove, D.J. (1950): "Climatic fluctuations in Europe and the Atlantic". Q.J.R. Met. S., Vol. 76, pp 147-164.

Shapiro, R.C. (1975): "The variance spectrum of monthly mean Central England temperatures". Q.J.R. Met. S., Vol. 101, pp 679-681.

Stidd, C.K. (1967): "The use of eigenvectors for climatic estimates". J. Applied Met., Vol. 6, pp 255-264.

Storr, D.P. and Fergusson, H.L. (1972): "The distribution of precipitation in some mountainous Canadian watersheds". D.P.M., Vol. II, pp 243-268.

Tyson, P.D., Dyer, T.G.J., and Mamtse, M.N. (1975): "Secular changes in South African rainfall, 1880-1972". Q.J.R. Met. S., Vol. 101, pp 817-835.

Whitmore, J.S. (1972): "The variance of mean annual rainfall with altitude and locality in South Africa as determined by curvilinear analysis". D.P.M., Vol. I, pp 188-201.

Williams, C.B. (1952): "Sequences of wet and dry days considered in relation to the logarithmic series". Q.J.R. Met. S., Vol. 78, pp 91-96.

Williamson, E. and Bretherton, M.H. (1964): "Tables of the logarithmic series distribution". Math. Stat. Annal, Vol. 35, pp 284-298.

Yap, W.C. (1973): "The persistence of wet and dry days in Sunger Bulop". Met. Mag., Vol. 102, pp 240-245.

CHAPTER 1

1. INTRODUCTION TO THE STATISTICAL ANALYSIS OF RAINFALL VARIATIONS

1.1 Purpose

This thesis which investigates in a statistical manner, basic variations of mainly Scottish rainfall in time and space, falls into two parts. The first part investigates rainfall space variations using eigenvector analysis of mean monthly rainfall data and regression analysis between daily rainfall and physical parameters. Inherent patterns in rainfall distribution are distinguished which may be used to answer some of the problems caused by the inadequacy of distribution of rainfall measuring stations, perhaps also to facilitate the interpolation between rainfall gauges in other situations.

The second part of the thesis investigates the non-randomness of rainfall time series, in particular the persistence of spells of wet and dry days, and the periodicities in annual and monthly rainfall time series. The underlying aim is to find, if possible, statistical properties which are of predictive value.

Rainfall is notoriously the most difficult meteorological element to measure and to predict. The attempt is here made to apply statistical techniques to extract the maximum amount of information in relation to its space and time variations. The process of analysis of rainfall data may be seen as an iterative one, derived spatial and temporal patterns being used to analyse new data.

1.2 Data

Availability of data has played a part in the selection of data for study. Spatial variations are examined for two areas in Scotland whose rainfall has not previously been extensively studied. These two area - a belt across Central Scotland from the Clyde to the Forth, and the area surrounding the Solway Firth - have relatively dense gauge networks: the latter region experiences rainfall from approaching systems which are almost unmodified by their path over land.

The data used to study the persistence of dry spells relate to various parts of the British Isles. Days are described as either "wet" or "dry" and a spell of dry (or wet) days of length r is defined as a spell of r such days. The analysis attempts to model the frequency of the occurrence of such spells by Markov persistence and modifications to it. Regional variations in spell distribution and model parameters are investigated, using 40 years of data from each station.

Monthly and annual time series for 11 stations, distributed over the whole of Scotland, which have available data for at least 80 years and which do not suffer from severe inhomogeneities, or gaps in records, are investigated for non-randomness. Stations are classed as "West" or "East" and results are compared within each group, between groups, and also with the results of applying similar analysis to circulation indices. These indices were first derived by Murray and Lewis (1966) from Lamb's (1950) daily classification and synoptic types for the period 1861 to 1971, and were subsequently revised by Murray and Benwell (1970) in the light of Lamb's (1972a) reclassification.

The P (progressive) index is a measure of the difference in frequency between days of progressive and blocked synoptic types, positive values indicating a predominance of the former type. Similarly, the S (Southerly) index measures the difference in frequency between days with Southerly and Northerly circulation. The C (cyclonic) index is positive when cyclonic days predominate over anticyclonic days. The M (meridional) index measures the frequency of meridional types. The P and C indices describe the frequencies of weather types particularly associated with rainfall.

Murray and Benwell (1970), in their regression analysis between rainfall and circulation indices showed that rainfall over England and Wales was very significantly correlated with C, while over Scotland the correlations between P and rainfall were the more marked. The Scottish rainfall data seems therefore to be related to the progression of rain-bearing synoptic systems to a greater extent than that of England and Wales. Within Scotland itself regional variations are perhaps to be expected with "West" station rainfall showing the closer relationship with P, and "East" station rainfall with C. The extent to which this is true is investigated in Chapter 5.

1.3 Problems in the statistical analysis of rainfall variations

1.3.1 Lack of data in space and the assessment of gauge representativeness

The chief problem which hinders the successful statistical analysis of rainfall variations is, as already mentioned, the lack of adequate data in time and space. In the first half of the thesis, gauge networks are required of sufficient density to permit the derivation of spatial variations in mean monthly rainfall using eigenvector analysis, and to assess the regression relationship between sets of daily rainfall values and physical parameters. On a larger scale, analysis of non-random variations in rainfall requires rainfall stations with such separation that similar temporal variations are found in several records. As common variations found in several gauge records over periods of years are more significant than those found in a single gauge record, it is desirable to have gauges sufficiently close together that they may be expected to show similar variations. Also gaps in individual records may then be filled by using readings from comparable neighbouring gauges.

Even when data are available with the required separation, it is desirable to ascertain the representativeness of a given gauge to the rainfall falling over the surrounding area. The difficulty of precise rainfall measurement and the differences that may arise between gauges separated by very small distances mean that representativeness is difficult to assess.

The geographical position of a gauge, its exposure, the height and aspect of the particular land on which it is sited, affects its instantaneous reading, and to a lesser extent, its monthly or annual reading. Over a period of a month or a year random variations are averaged out and the gauge reading is determined by systematic effects, and in particular by the position of the gauge relative to topography and to the prevailing wind.

Various methods have been employed in an attempt to assess the representativeness of a gauge to a given area. Rhodda (1967,

1970) calculated correlation coefficients between a series of station gauge values. These coefficients in general decreased as station separation increased but the decrease was not uniform as other factors such as topography, gauge exposure, and prevailing wind direction must be considered. It was possible to perform further correlation analysis between gauge correlation coefficients (derived from a series of measurements at each gauge) and the separation between gauges. Attempts to correlate the gauge correlation coefficients with other physical parameters describing the relative positions of the gauges, e.g. the difference in height between the gauges did not produce significant results.

While the assessment of gauge representativeness is not considered further below, it should be remembered when studying the results of analyses in this thesis that a given gauge network may not necessarily be representative of possible rainfall spatial variations.

1.3.2 Lack of data in time

In the regression analyses between physical parameters and rainfall of a "pure" type, continuous or hourly gauges are, ideally, required to isolate the rainfall of a particular sub-system from that of other sub-systems. The distribution of such gauges is very sparse in most regions. Usually there are only two or three autographic gauges within an area of 1000 km^2. Recently, magnetic tape recorders (MTER) have been introduced to measure rainfall continuously as a supplement to the traditional Dynes recorders. However, it is only in the area of the Dee Weather Radar Project that a comprehensive network of MTER recorders has been installed. Daily rainfall gauge records can be used to describe rainfall of a pure type if most of the rainfall occuring on that day results from the passage of a well-defined subsystem. Most daily rainfall figures reflect several different synoptic influences, and even in cases where days may appear to be of a pure type it cannot be certain that all the rainfall over the area under study on that day results from a particular subsystem. A similar doubt relates to the regression equations which define the rainfall variations. Thus the lack of data over very short periods tends to hinder the isolation of rainfall distributions resulting from a particular subsystem.

When analysing spells of dry days, sufficient data are required to measure the frequencies of rarer long spells and to "predict" their future occurrence. The task of predicting rare long spells is facilitated by long continuous records from a given station as does the existence of gauges with similar spell-frequency distributions. Records of sufficient length (at least 30 years) are often not readily available from several sites in a given area.

In the case of studies of rainfall time variations, continuous homogeneous records of sufficient length to assess the significance of a given rainfall variation are required. In order to study low frequency oscillations of period, say, 20 years, at least 80 years of data are required. It is important to know if such periodicities, or any apparent trends, occur continuously throughout a given record, and whether the phase of the periodicities is constant throughout the record.

In the case of short time series an apparent trend (trend being defined as the monotonic increase or decrease of members of a series) may reflect a long-term periodicity. While a parameter may increase in an almost linear fashion over the period of observation, it may subsequently decrease in value. Apparent trends in meteorological records over a few decades may reflect oscillations of period of the order of a century.

1.3.3 Discontinuities and inhomogeneities

The ideal of having continuous homogeneous rainfall records available to study variations in time series, and the possibility of filling some gaps in records by the use of neighbouring gauges with similar records, have already been mentioned. The problems raised by the gauge inhomogeneity, namely the assessment of where the inhomogeneities in the gauge record occurred, and of how these inhomogeneities may be allowed for, are not easy to solve. Any change in the site, height, or local exposure of the gauge will mean that values from the two sites are not strictly comparable. In order to make the values comparable an overlap of several years will be necessary. It may however be permissible, where no overlap of data exists, to ignore site changes if the apparent change in mean gauge value is less than five per cent, the likely systematic error of the gauge.

Changes of observer may lead to erroneous variations in gauge records. If each observer systematically over- or underestimates readings then true variations remain intact but if the observer subsequently changes the systematic errors may show as a slippage in the mean value of the record, or in the case of smoothed or filtered data, as an apparent trend.

To reveal inhomogeneities in series, subsections of a record may be compared with its long-term mean using Cramer's test and anamalous values for a section of a given record may be compared with corresponding values of other records; or Kohler's (1949) test may be used to compare directly records from different gauges. The former method is used in the analysis in Chapter 5.

1.3.4 The statistical significance of results of time series

While the statistical significance of the results of eigenvector analysis and of regression analysis can be gauged from relative values of eigenvalues and values of correlation coefficients, the significance of non-random variations in time series is less easy to assess. This further problem needs to be borne in mind throughout the analysis of time-series.

In the analysis of the persistence of spells of wet and dry days, model parameters for five of the seven models used are adjusted to give the best fit to the data under consideration. The "predicted" distributions of spell frequencies according to a given model is then compared with the observed distribution using the χ^2 test, frequencies of long spells being pooled. The efficiency of χ^2 test in assessing the goodness of fit is less with (infrequent) longer spells than with (frequent) shorter spells, while model parameters are themselves much influenced by shorter spells.

Thus "predictions" of the occurrence of long spells and the precise assessment of probabilities of a given day continuing for a further day are subject to uncertainty.

In the analysis of monthly and annual time series, the effects of a given non-random variation, whether a periodicity, trend, or form of persistence, must be isolated from those of all other variations before its significance can be tested.

Analysis of raw data by power spectrum analysis will reveal effects of periodicities, trend, and persistence. In order to investigate a particular periodicity, its effect on spectral estimates must be isolated from those of other non-random variations. Raw data may be "detrended" and effects of persistence may be accounted for by plotting a "persistence only" spectrum for comparison with the actual spectrum, if the form of persistence can be determined. Persistence is defined as the ability of successive members of a series to remember their antecedent value or values. With the simplest form of persistence - Markov persistence of first order - the value of each member of the series is determined linearly from the previous member only and serial correlation coefficients at lag r are given by the r th power of that at lag one. The spectrum of Markov persistence, known as the "red noise" spectrum, has a decay of power from high to low frequencies, similar to the exponential decay of serial correlation coefficients in the correlogram, the exact shape of the spectrum being easily determined from the spectral window and the "lag one" correlation coefficient. Other forms of persistence and their effects are less easy to determine and model.

Leakage of power from spectral peaks, interference effects between peaks, and harmonics of other peaks are effects which all hinder the isolation of a given periodicity and the assessment of its significance.

1.4 Methods

1.4.1 T-mode eigenvector analysis

In Chapter 2 T-mode eigenvector analysis is used to derive inherent patterns of mean monthly rainfall data of Central Scotland and of the Solway region. The analysis proceeds on similar lines to Stidd's analysis of Nevada mean monthly rainfall. Independent temporal eigenvectors with associated independent spatial eigenvector multipliers are constructed. The most significant eigenvector and its set of multipliers are used to "predict" mean monthly rainfall at new sites, spatial components which are "smoothed" for topography by linear regression being interpolated to new sites.

Predictions made for a series of stations with short-period data are compared with estimates of long-term means made from their short-period records using the method described by Bleasdale (1963). In this latter method ratios of short-period to standard period means are determined for long-term gauges surrounding a short-period gauge. These ratios are interpolated to short-period sites and are used to adjust the short-period means to standard periods.

1.4.2 Regression between daily rainfall and physical parameters

In Chapter 3 regression relationships are derived between the physical parameters of altitude, distance from the South coast, distance from the West coast, and daily rainfall values or mean annual rainfall. The days are chosen such that rainfall from one particular system, e.g. a front or warm sector could be described as being of a "pure" type. While numerous attempts have been made to correlate mean monthly and annual rainfall with altitude, none has been made, to the author's knowledge, for daily rainfall.

Bleasdale and Chan (1972) correlated mean annual rainfall at various sites over the British Isles with station altitude and in their discussion of observed relationships suggested that an attempt should be made in future work to perform regression analyses for weather types of a pure form which would be likely to produce heavy orographic rainfall. These analyses were to be compared with those in which the rainfall v. altitude regression relationships was expected to be absent or reversed, e.g. thunderstorms. It is in answer to these suggestions that regression analyses of daily rainfall values with physical parameters for various pure synoptic types is performed below.

It should be noted that the relationship between rainfall and topography is already used to interpolate between rainfall gauges indirectly. Mean annual rainfall isohyets are drawn using topographical maps as an overlay to determine the shape of the isohyets where no gauge data are available. In the assessment of areal rainfall for specific days, daily rainfall values are expressed as a percentage of mean annual rainfall and are objectively interpolated to grid points (Salter 1972, English 1973), the mean annual rainfall field effectively expressing the variation of rainfall with topography.

1.4.3 Models of persistence for wet and dry days

The Markov model, in which serial correlation coefficients at lag r are given by those at lag one ρ_1 say, raised to the r th power ρ_1^r, was the first model used to describe the persistence of wet and dry spells (Gabriel and Reumann 1949). Two other models were suggested which modified the exponential decay of serial correlation coefficients by a factor dependent on the lag: William's (1952) log model and Green's (1970) modified log model in which serial correlation coefficients became $\frac{\rho_1^r}{r}$ and $\frac{\rho_1^r}{r + a}$ respectively. Another modification to the Markov model proposed by Yap (1973) incorporated Markov persistence but the actual value of ρ_1 was determined by the spell length, being constant within a given spell.

Besides these four models, two other models proposed by Lawrence (1957) to describe the distribution of dry spells are used in Chapter 4. These models do not incorporate persistence of a predetermined form as do those above by their model parameters. The "natural persistence" model compares the probabilities of a spell of a given length continuing another day at different stations and expresses these probabilities as the sum of an area and a station-dependent term. The "Jenkinson probability" model attempts to predict the frequency of long spells using return period probability curves determined by the mean annual spell length, its standard deviation, and its two-year standard deviation.

These seven models are used in Appendix 1 and in Chapter 4, to describe frequencies of spells of any length and also those greater than a given length.

1.4.4 Methods applied in time series analysis

General medium-term variations are investigated using decadal means and low-pass filtering of data. Specific periodicities in monthly and annual rainfall are sought using power spectrum analysis. Methods of spectral analysis which give stable results and which have been applied extensively to geophysical data are used. These are the Blackman-Tukey autocovariance method and the fast Fourier transform (FFT) method applied in the manner suggested by Rayner (1971) namely with filtering of data prior to analysis and summation of estimates into bands. Other methods of spectral analysis, such as the maximum entropy method (Ross 1975), and their application in locating low frequency periodicities are discussed.

Results of these analyses are compared among themselves in an attempt to distinguish common variations in "East" and "West" stations and in circulation indices similarly analysed. The relationship between oscillations in rainfall and P and C indices is further investigated using cross spectrum analysis. The Blackman-Tukey cross covariance approach and the FFT method, similar to the autocovariance and FFT methods of power spectrum analysis, are used to compute cospectra and quadrature spectra from which coherence and phase estimates between rainfall and circulation indices are produced.

S-mode eigenvector analysis is used to construct independent rainfall time-series which are sets of eigenvector multipliers. These series describe common time variations present in each individual series to an extent determined by the value of the eigenvector space element. An attempt is made to correlate eigenvector space patterns with station position and station mean annual rainfall in order to see which influences on climate are described by a given eigenvector. The eigenvector multiplier time are also analysed in a similar way to those of individual rainfall series using low-pass filtering, power spectrum analysis, and cross spectrum analysis with P and C indices. Results are compared with those of individual series in an attempt to label time series as describing variations of "West", "East", or all stations. The presence of a variation in an eigenvector time series as well as an individual series increases its apparent significance.

CHAPTER 2

2. THE ANALYSIS OF MEAN MONTHLY RAINFALL USING EIGENVECTORS

2.1 Introduction

The purpose of this Chapter is to explain how eigenvector analysis is performed on a time by space matrix to distinguish independent patterns with temporal and spatial components, and to show the chief uses and advantages of eigenvector analysis. The technique is illustrated by the decomposition of matrices of mean monthly rainfall for several stations in Scotland, in a similar manner to Stidd's analysis of Nevada rainfall (1967).

In the analysis of several rainfall patterns or mean rainfall patterns occuring over a period of time, it is desirable to isolate independent spatial patterns which are present to a greater or lesser extent in each of the original spatial distributions. It is useful to derive further coefficients explaining to what extent each derived spatial pattern is prevalent in each of the original spatial distributions. Eigenvector analysis constructs a series of individual spatial patterns each with an associated time series. These time series are also independent of one another.

In the case of mean monthly rainfall for twelve months over a given network of ST stations, the analysis expresses each station's rainfall for a given month as the sum of twelve terms. Each term consists of a value for that month common to each station multiplied by a spatial coefficient dependent on that station. Each pattern, consisting of twelve one-month time elements and ST space elements, accounts for a certain proportion of the variance in the original matrix and some of the patterns will be more important than others. The advantages of eigenvectors to decompose time by space matrices over other methods are:

1. Each eigenvector pattern's temporal and spatial components are independent. This is demonstrated by orthogonality.
2. Eigenvectors do not assume a particular functional relationship between values of data and are derived irrespective of the spacing of data points. The accuracy and usefulness of spatial patterns determined by other mathematical functions is dependent on the spacing of data points and the form of the functions.
3. Usually only a few such patterns are needed to describe most of the variance in the original time by space matrix. In cases considered only one of a possible twelve eigenvectors is needed to describe 99 per cent of the variance in the original matrix.

The uses of eigenvector analysis illustrated below are:

1. A small amount of data may express the total variance present in a time by space matrix.
2. The eigenvector patterns may bring into focus inherent patterns in the original matrix, and describe the effects of known influences on climate.
3. Eigenvector patterns may be used to interpolate between data points and "predict" mean monthly rainfall. This exercise is facilitated by the reduction in the redundancy in data and hence in the amount of data needed for

interpolation.

4. These spatial patterns describing types of rainfall distribution may be of some use as an interpolation aid when analysing daily rainfall patterns.

5. When analysing time series of length about a century for a few stations, basic time series may be constructed using eigenvector analysis which describe simply the climate fluctuations in these series. Extrapolation of these series may be of predictive value.

The method of eigenvector analysis and uses 1 to 3 are demonstrated in this Chapter.

Some other uses of eigenvectors in meteorology are given by Craddock and Flood (1970) and Craddock and Colgate (1974). Eigenvector analyses of mean sea level pressure over the Northern Hemisphere may be used to identify and correct errors in data; to predict monthly pressure fields using, as predictors, eigenvector coefficients derived from daily values of preceding months; and to select situations whose sequels can be used for prediction by matching daily eigenvector coefficients. States of upper airflow of predictive value are also classified from eigenvector coefficients of daily values of 500 mb thickness.

Craddock, J.M., and Flood, C.R. (1970): "Eigenvectors for representing the 500 mb geopotential surface over the Northern Hemisphere". Q.J.R. Met. S., Vol. 95, pp. 576-593.

Craddock, J.M., and Colgate, M.G.C. (1974): "The use of eigenvectors for smoothing and prediction". Journal Inst. Maths. and its applications, 12, pp. 152-160.

$_1$ contains the n_1 most significant columns of E.

$$A = EM \qquad (2.1a)$$

$$A = E_1 M_1 \qquad (2.1b)$$

To perform the analysis A is multiplied by its transpose A^1 to form a 12 by 12 square symmetric matrix which describes the relationship between rainfall in different months (equation 2.2) and eigenvectors X_j of B with eigenvalues λ_j are then found (equation 2.3).

$$B = A A^1 \qquad (2.2)$$

$$B X_j = \lambda_j X_j \qquad j = 1,2,.., 12 \qquad (2.3)$$

There are twelve eigenvectors each with twelve elements. Each eigenvector describes a temporal variation in rainfall, each element in the eigenvector representing a single month. The X_j eigenvectors are orthogonal and they can be arranged to be normalised to 1 (equation 2.4)

$$X_j^1 X_j = \delta_{ij} \qquad (2.4)$$

where δ_{ij} is the Kronecker delta function.

The proportion of the variance in the original matrix which is explained by a given eigenvector X_j is determined by the ratio of the associated eigenvalue λ_j to the sum of all the eigenvalues in equation 2.5.

$$\sigma^2 (X_j)) = \frac{\lambda_j}{\sum_1^n \lambda_j} \qquad (2.5)$$

The variance described by the first n_1 eigenvectors is the ratio of the sum of the first n_1 eigenvalues, arranged in decreasing order of significance, to the sum of all the n eigenvalues as in equation 2.6. The larger the eigenvalue, the greater the proportion of variance described by the eigenvector.

$$\sum_{1}^{n_1} \sigma^2 (X_j) = \frac{\sum_{1}^{n_1} \lambda_j}{\sum_{1}^{n} \lambda_j} \qquad (2.6)$$

As the first n_1 eigenvectors which describe at least 95 per cent of the variance are of interest, n_1 is chosen so that the ratio in equation 2.6 is 0.95 or greater.

From the twelve X_j eigenvectors and the first n_1 eigenvectors, matrices E and E_1 are formed as in equation 2.7a and 2.7b.

$$E = \left[X_1 , X_2 \quad \ldots \quad X_{12} \right] \qquad (2.7a)$$

$$E = \left[X_1 , X_2 \quad \ldots \quad X_{n_1} \right] \text{ where } n_1 < 12 \qquad (2.7b)$$

M and M_1 in equations 2.1 may then be found by multiplying the transposes of E and E_1, E^1 and $E_1^{\,1}$, by A as in equations 2.8. $E\,E^1 = \Gamma$ as the eigenvectors are orthonormal.

$$E^1 A = E^1 E M = M \qquad (2.8a)$$

$$E_1^{\,1} A = M_1 \qquad (2.8b)$$

Since E is a 12 by 12 matrix and A a 12 by ST matrix, M will be a 12 by ST matrix and similarly M_1 will be an n_1 by ST matrix. Thus M and M_1 consist of 12 and n_1 spatial patterns with one element per station. Each eigenvector X_j, representing a temporal variation, will have an associated spatial variation with ST elements. To describe the variance of A fully, 12 patterns are required but n_1 patterns may be sufficient to explain all the variance of significance.

Sets of eigenvector multipliers, MR_j and MR_k, which are rows of M and are derived from eigenvectors X_j and X_k as in equation 2.9a, are also orthogonal to one another as in equation 2.9b.

$$MR_j = X_j'A \qquad (2.9a)$$

$$\begin{aligned} MR_j\, MR_k' &= X_j'A(X_k'A)' \qquad (2.9b) \\ &= X_j'AA'X_k \\ &= 0 \text{ if } j \neq k \\ &= \lambda_j \text{ if } j = k \end{aligned} \Bigg\} \text{ from equations 2.2 and 2.3}$$

2.3 The advantages of eigenvector analysis over other methods

The original matrix decomposition into space and time components can alternatively be performed using spherical harmonics or Tschebychev polynomials to describe the spatial variations. With such methods the number of patterns needed to describe a given proportion of the variance in the original matrix is usually larger than the number of required eigenvectors. While no comparison is made below between eigenvector analysis and decomposition using other methods such as spherical harmonics, the fact that one or two eigenvectors out of twelve can describe 99.9 per cent of the variance in the original matrix suffices to illustrate this point.

When methods other than eigenvector analysis are used to decompose a time by space matrix, the spatial patterns are dependent on the properties of the mathematical function used. The form of the pattern is predetermined by the function and is dependent on the spacing of the data points. For eigenvector analysis the relative positions of the data points have no effect in the determination of the eigenvectors. This is a definite advantage, especially in the analysis of rainfall distributions where the network of stations is irregular.

Spherical harmonics or Tschebychev polynomials may be used to construct orthogonal patterns which describe the series of mean monthly rainfall distributions. However, the associated temporal patterns will not necessarily be independent or orthogonal, unlike the eigenvector patterns.

Because eigenvectors are orthogonal in both temporal and spatial components, eigenvector analysis may proceed either using a time by time symmetric matrix (T-mode) or a space by space symmetric matrix (S-mode). For S-mode the symmetric matrix B is constructed as A^1A instead of AA^1 for T-mode. Results of computation of eigenvectors by the two different methods have been shown by Hirose and Kutzbach (1969) to be mathematically identical apart from different normalisation. The choice of T or S mode depends on the dimension of the original matrix. As it is computationally faster to derive eigenvectors from small matrices, S-mode is used where the space dimension is smaller than the time dimension and vice versa.

S-mode is used later to analyse eleven annual rainfall series of length eighty-four years. Eleven eigenvectors with eleven space components are derived, from which eleven sets of eighty-four multipliers are further produced. It is these sets of multipliers with one element of each year, which form basic rainfall time series describing common temporal variations. The eleven eigenvector elements describe to what extent each of the derived series determine the station's rainfall.

Thus eigenvector analysis has the advantage that as each set of eigenvector multipliers as well as each eigenvector is independent, there is no distinction in the way spatial and temporal components are treated and analysis may proceed by S or T-mode according to the dimensions of the matrix.

2.4 The choice of the type of data, real or anomaly

Eigenvector analysis may be performed on real data, anomaly data, or anomaly data normalised such that the total variance of each column is unity. In the case of mean monthly rainfall data, the mean of each month may be subtracted from each station to construct anomaly data, the departures of each station from the all-station mean monthly value. Such values for each station for a given month may further be squared and summed, and each monthly value divided by this value. This will produce normalised anomalies and ensure the all-station monthly variance is unity, as well as the all-station mean being zero as in the case of anomaly data. The diagonal elements of the symmetric matrix will then be unity, and the off diagonal elements will have values in the range 0 to 1.

Kutzbach (1967) in his discussion of eigenvector analysis of a combined matrix of the three climatic variables of rainfall, temperature, and pressure for eighty stations describes the differences between using real data, anomaly data, or normalised anomaly data. The symmetric matrix A above will be a cross-product matrix, a covariance matrix, or a correlation matrix and the first eigenvector multipliers will have the closest resemblance to observed fields, anomaly fields, or normalised anomaly fields according to the three types of data used. In Kutzbach's analysis normalised variations were required as the three climatic parameters had different means and variances. When variables are normalised, each variable at each point in the data field is of equal importance in determining the eigenvector patterns.

In the present analysis real data are used to analyse mean monthly rainfall for two areas of Scotland. The first eigenvector from the 12 by 12 symmetric cross product matrix shows the annual cycle of rainfall variation while the isopleths of the first eigenvector multipliers reflect mean annual rainfall variations. For the region of Central Scotland, anomaly data are also analysed. In this case, the first eigenvector and its multipliers describe the annual cycle of rainfall variability between stations and the mean annual rainfall anomaly for each station respectively. Normalised anomaly data produce similar patterns to anomaly data and are not discussed further.

It is the eigenvector multipliers of anomaly data which are used to interpolate between rainfall station values. Mean monthly anomaly rainfall values are estimated for new sites by multiplying the first eigenvector by the interpolated eigenvector multiplier. Hence, monthly rainfall may be "predicted" from this result and the all-station mean monthly rainfall value. The anomaly eigenvectors

Table 2.1 Data used in eigenvector analysis

Area	Period	No. of Stations	Type of Station	Station Network	Source	Data used in the analysis	Results		
							Eigenvectors	Multiplier	Discussion
Solway (i)	1916-50	69	long-term		Hydrological Mem. No. 26	Real	Table 2.2 Fig. 2.3		Section 2.6
(ii)	1941-70	207	short and long-term	" x " Fig. 2.1	Met. Office Archive tapes	Real	Table 2.2 Fig. 2.3	Fig. 2.10	Section 2.6
Central Scotland (i)	1916-50	(209 (((82 (	short-term long-term	" . " Fig. 2.2 " x " Fig. 2.2	Hydrological Mem. No. 26 and No. 32	Real	Table 2.2 Fig. 2.3	Fig. 2.6 and smoothed Fig. 2.7	Section 2.6
(ii)	1916-50	82	long-term	" . " Fig. 2.2	Hydrological Mem. Nos. 26 and 32	Real	-- Similar to (i) --		
(iii)	1916-50	82	long-term	" . " Fig. 2.2	Hydrological Mem. Nos. 26 and 32	Anomaly	Fig. 2.11	Fig. 2.12 and smoothed Fig. 2.13	Section 2.7
(iv)	1916-50	82	long-term	" . " Fig. 2.2	Hydrological Mem. Nos. 26 and 32	Normalised Anomaly	-- Similar to (iii) --		

and their multipliers will explain more of the space variations than real eigenvector multipliers as their determination will be more sensitive to small variations.

2.5 The analysis of "real" mean monthly rainfall

2.5.1 Data and the use of short-period gauges

A summary of the data used for the eigenvector analysis appears in Table 2.1. Short-term data stations were previously extended to standard periods using the procedure involving the interpolation of ratios between short-period and standard period records to short-period gauges described by Bleasdale (Hydrological Memorandum No. 5) and mentioned in section 1.3. For Central Scotland the first eigenvector derived from 82 long-term stations and their multipliers were compared with those derived from the complete network of 82 long-term stations and 209 short-term stations. The first eigenvectors describing 99 per cent of the variance in each data set are very similar (Table 2.2) and the multipliers for the 82 long-term stations have similar values in both analyses. Thus the use of short-term records extended to standard periods by the procedure described by Bleasdale was justified as it did not affect inherent patterns already revealed to be present by eigenvector analysis. The interpolation and extrapolation processes involved in this latter procedure are essentially linear so that the 209 values have some linear dependence on the 82 long-term stations. The eigenvector analysis, on the other hand, does not necessarily involve any linear dependence between station values.

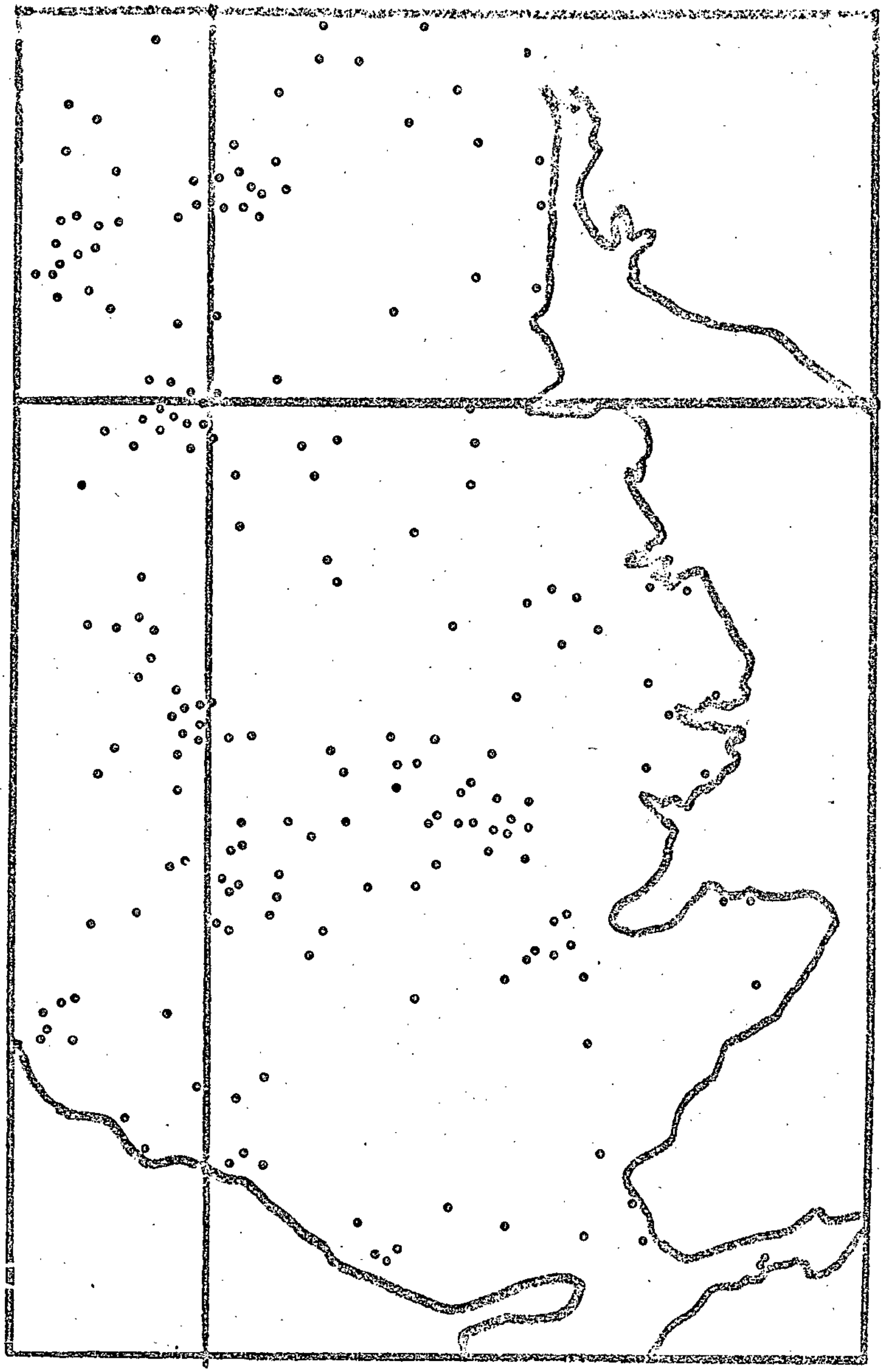

Figure 2.1 Solway 1941-70 mean monthly rainfall station network

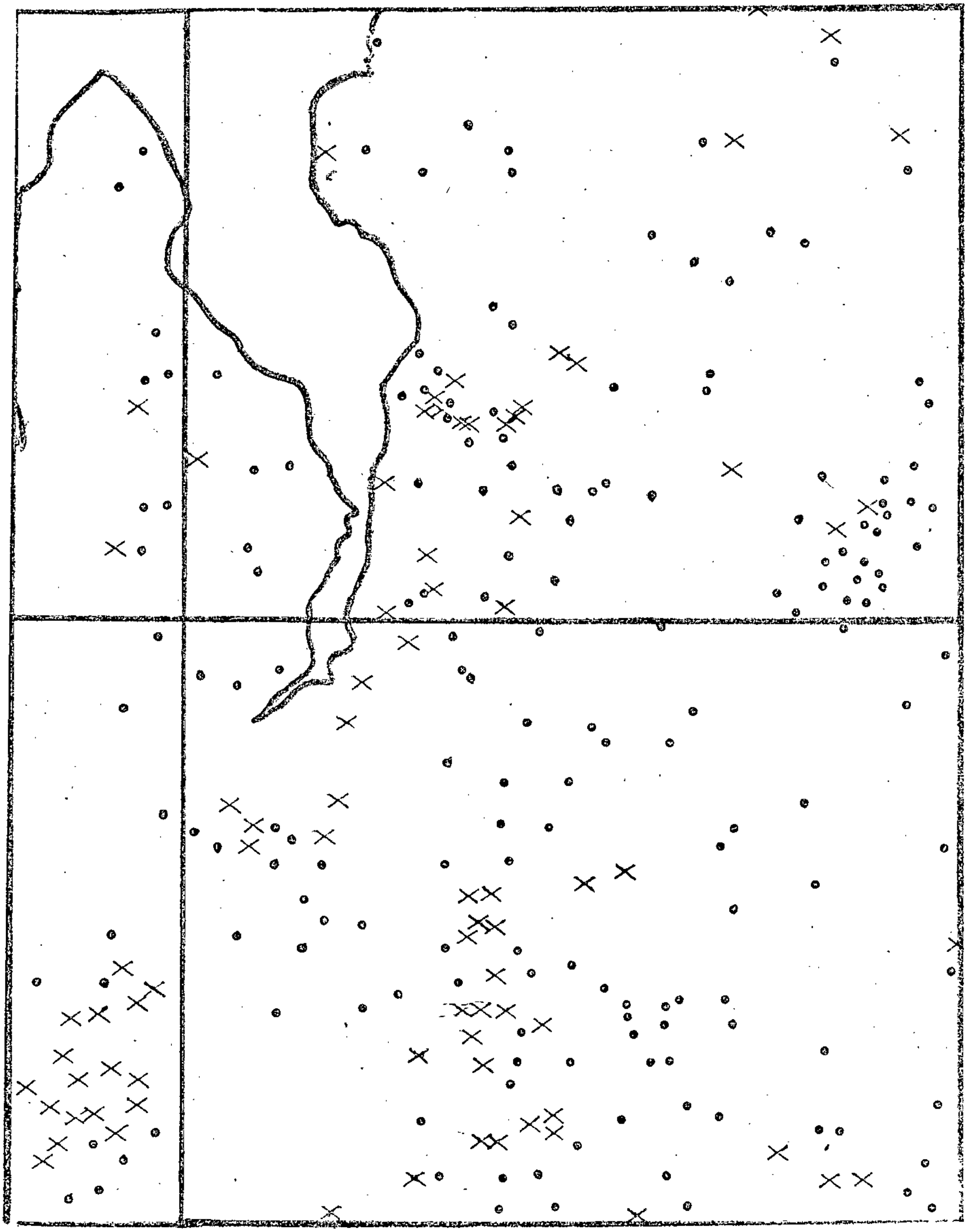

Figure 2.2 Central Scotland 1916-50 mean monthly rainfall station network

(X - long-term stations; • - short-term stations)

Table 2.2 Mean Monthly Rainfall Eigenvectors - Scotland

	Central Scotland 1916-50				Solway 1916-50		Solway 1941-70	
	1st Eigenvector		2nd Eigenvector					
	292 stations	81 stations	292 stations	81 stations	1st Eigenvector	2nd Eigenvector	1st Eigenvector	2nd Eigenvector
January	0.389	0.391	0.351	0.331	0.399	0.229	0.337	0.220
February	0.258	0.260	0.193	0.241	0.246	0.179	0.233	0.210
March	0.213	0.215	0.302	-0.004	0.213	0.018	0.218	0.052
April	0.200	0.199	-0.389	-0.031	0.195	-0.037	0.214	-0.015
May	0.211	0.206	-0.311	-0.305	0.203	-0.141	0.216	-0.213
June	0.199	0.193	-0.225	-0.244	0.198	-0.249	0.205	-0.376
July	0.259	0.249	-0.489	-0.501	0.255	-0.538	0.247	-0.434
August	0.289	0.284	-0.470	-0.465	0.280	-0.576	0.296	-0.465
September	0.295	0.291	-0.144	-0.154	0.296	-0.086	0.352	-0.143
October	0.375	0.382	0.140	-0.141	0.370	0.163	0.345	0.149
November	0.337	0.336	0.113	0.092	0.343	0.232	0.350	0.147
December	0.348	0.353	0.421	0.416	0.361	0.357	0.370	0.499
% variance explained	99.6	99.3	0.4	0.6	99.4	0.1	99.9	0.1

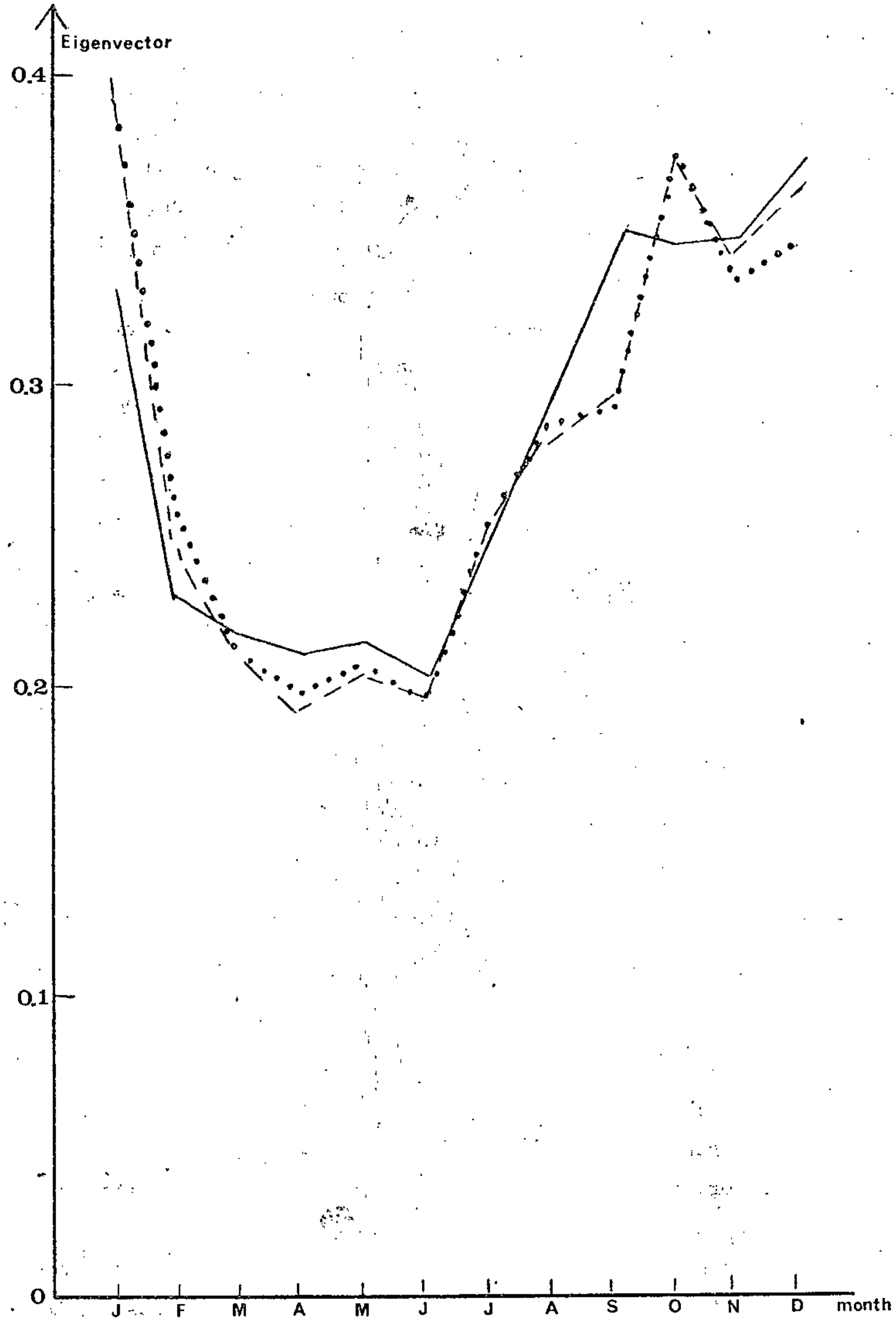

Figure 2.3 First eigenvectors, mean monthly rainfall Scotland

(—— 1914-70 Solway; --- 1916-50 Solway;
··· 1916-50 Central Scotland)

2.5.2 Results

In Table 2.2 the amount of variance explained by each eigenvector is determined by the ratio of the first eigenvector to the sum of the twelve eigenvalues. The first eigenvector explains most of the variance and clearly reflects the annual cycle in rainfall values (see also Fig. 2.3). The 1916-50 first eigenvectors for the different areas are almost identical, which reflects overlap of data between areas and the similarity between their climates.

The eigenvectors have maxima in January and October, and a minimum value over the months April, May and June. The 1941-70 first eigenvector for Solway varies more uniformly being derived from a denser network of stations, and has a large value from September to January, and a small value from February to June.

The multipliers of the first eigenvector illustrated in Figs. 2.6 and 2.10 should be compared with the annual rainfall maps Figs. 2.5 and 2.9, and the topographic maps Figs. 2.4 and 2.8. The resemblance between mean annual rainfall isohyets and first eigenvector multiplier isopleths is strong. Both are closely related to topography. The variation of rainfall with distance from the West coast for Central Scotland and from the South and West coasts for the Solway region can also be seen.

It should be remarked at this point that while mean annual rainfall and first eigenvector multipliers, and all-station mean monthly rainfall and first eigenvectors, are very similar, the two first eigenvector patterns filter noise out of the data in a different way to simple averaging. Also, while in the case considered, eleven out of twelve eigenvectors were insignificant and could be

Figure 2.4 Topography Central Scotland

(Land over 800 feet hatched)

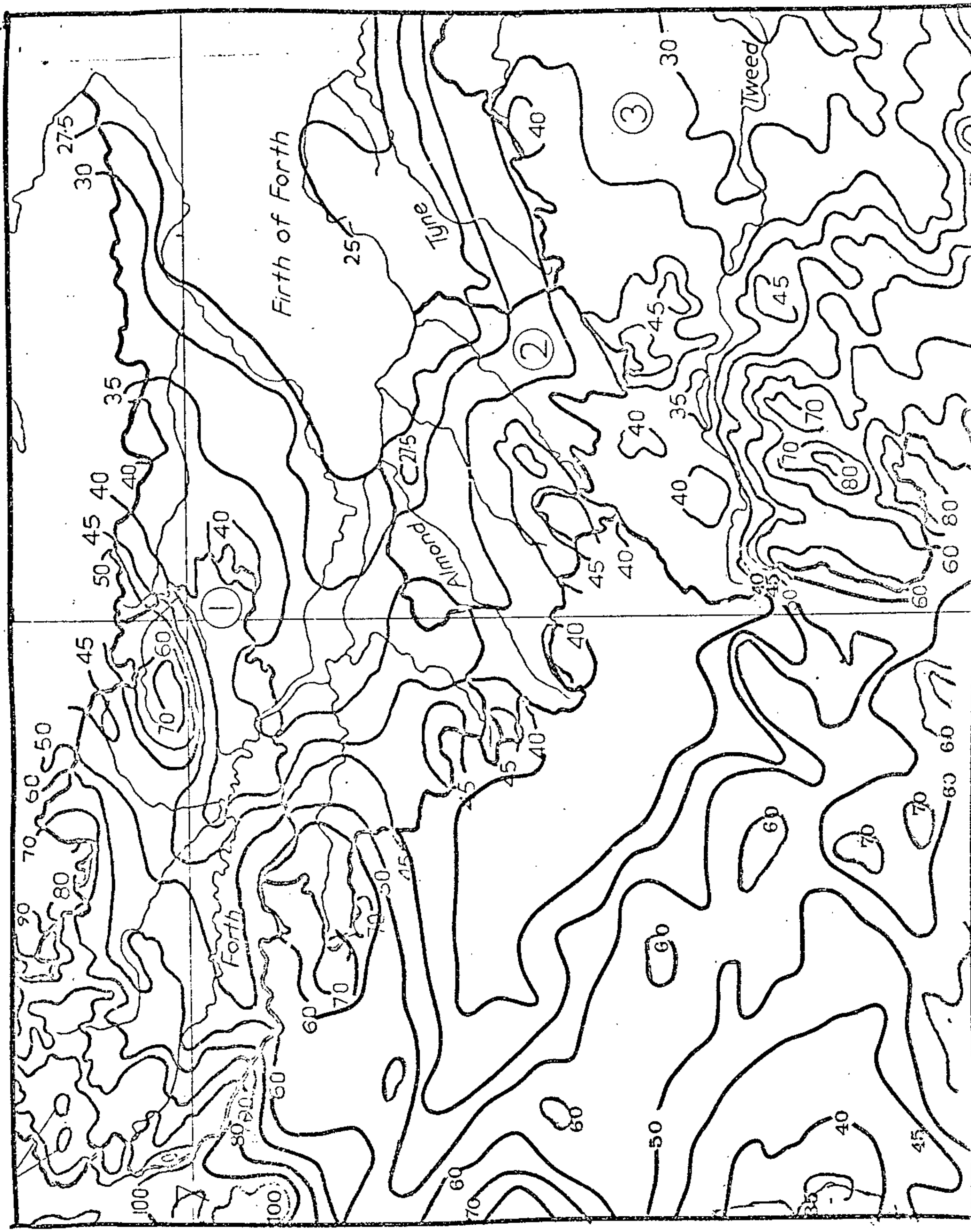

Figure 2.5 Mean annual 1916-50 rainfall Central Scotland (information from Hydr. Mem. Nos. 26 and 32)

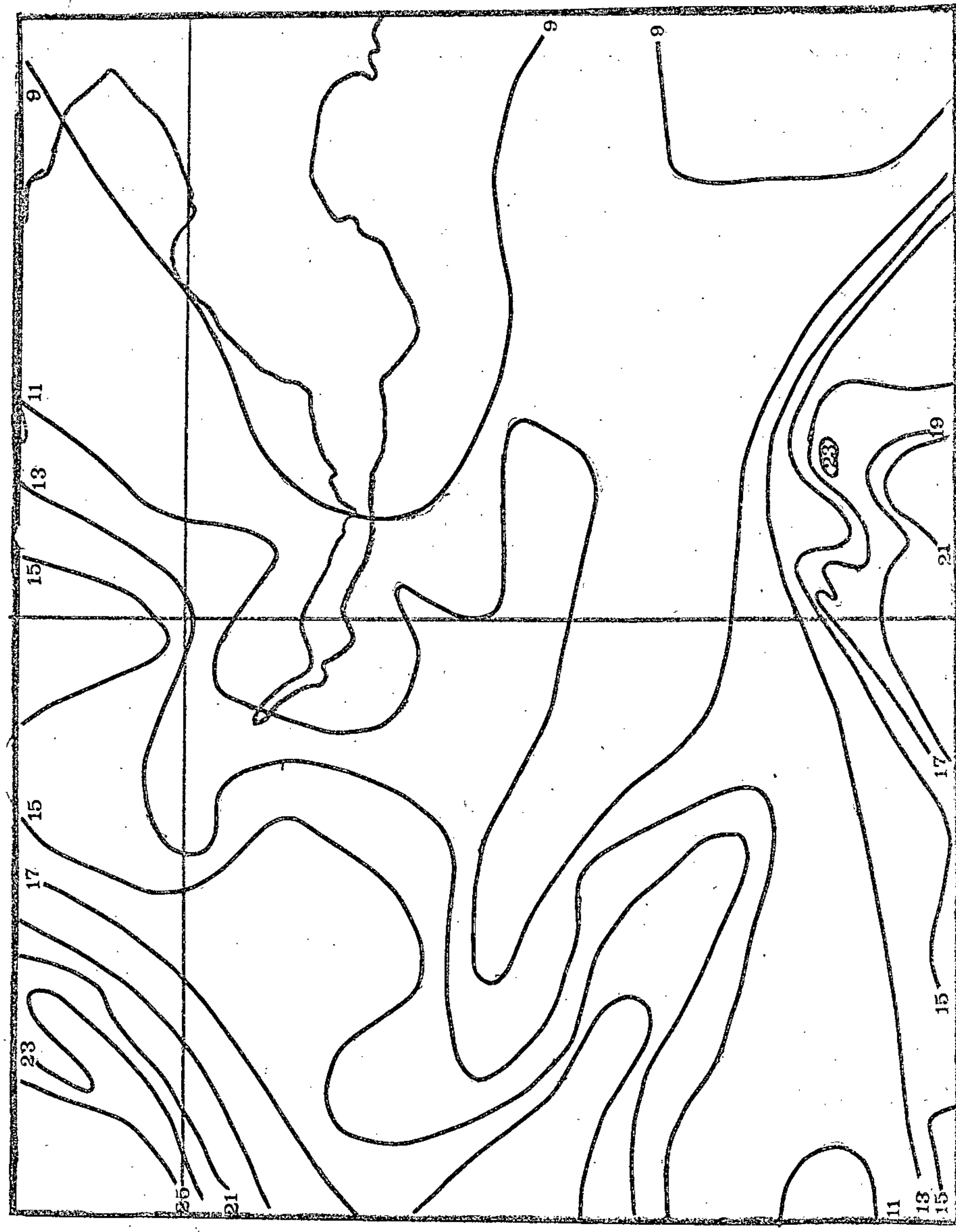

Figure 2.6 First eigenvector mulitplier, mean monthly rainfall Central Scotland

Figure 2.7 First eigenvector multipliers, mean monthly rainfall Central Scotland smoothed for topography

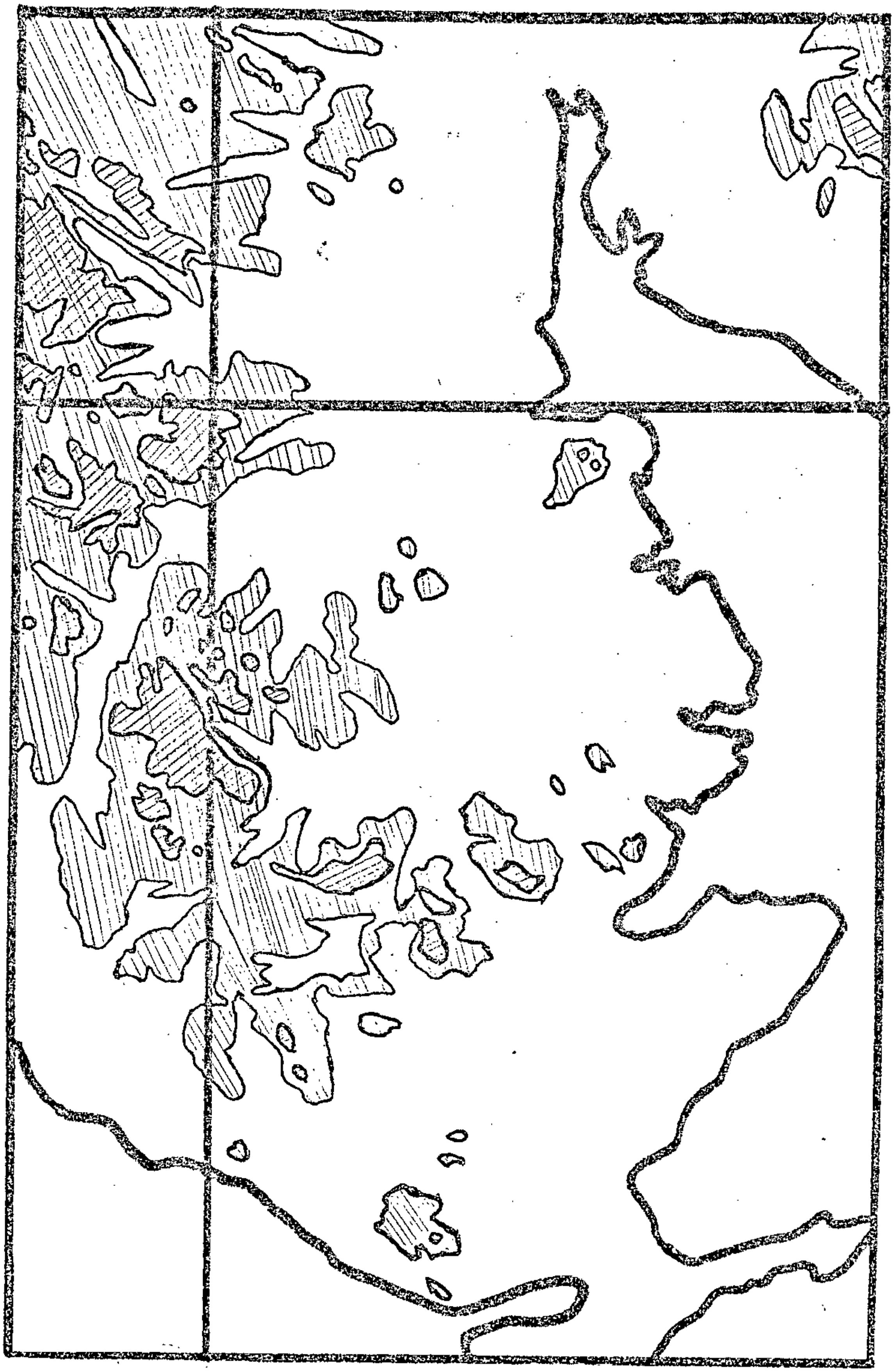

Figure 2.8 Topography Solway Region

(Land over 800 foot hatched; land over 1400 foot double-hatched)

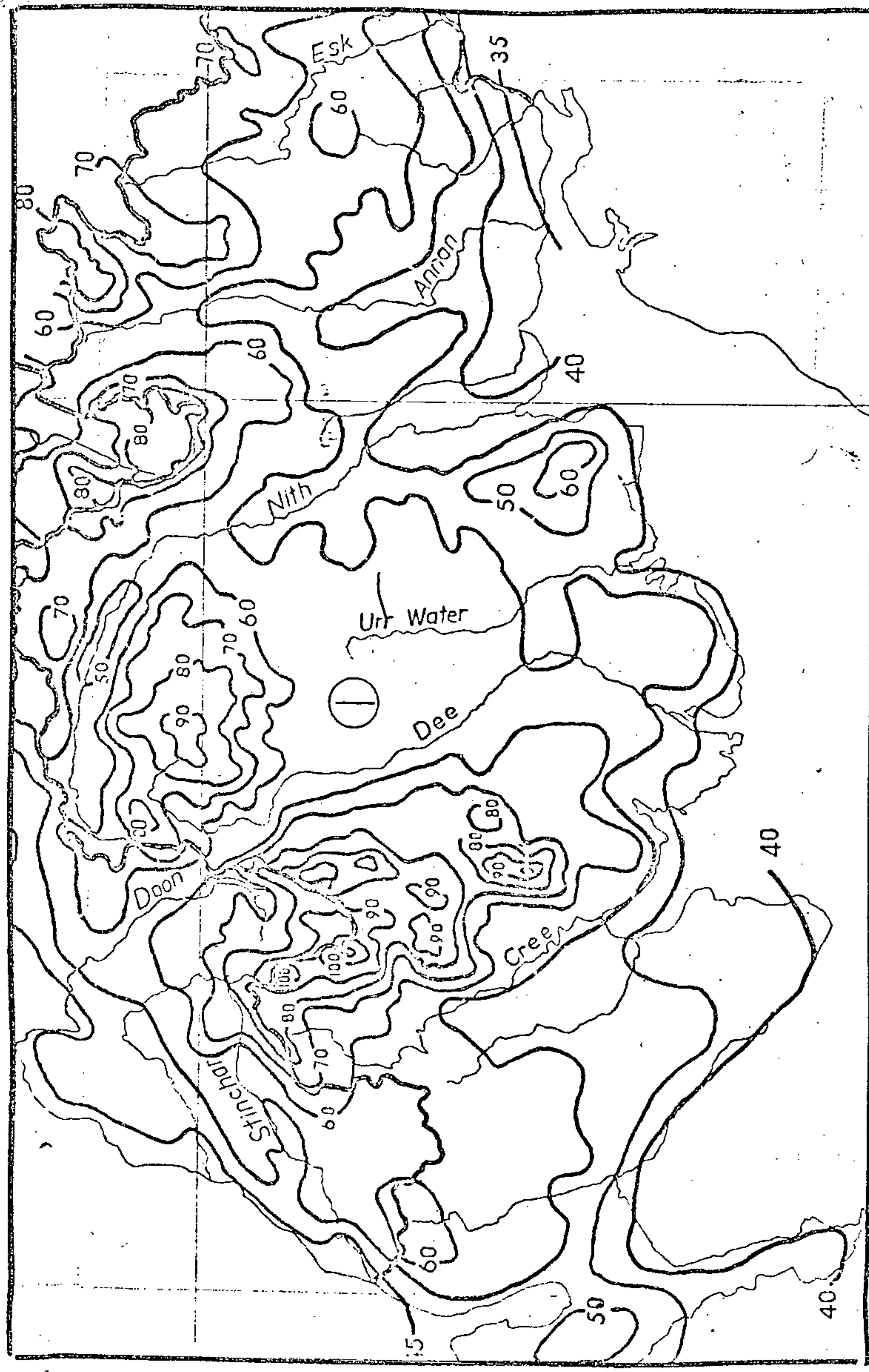

Figure 2.9 Mean annual 1916-50 rainfall Solway region (information from Hydr. Mem. No. 26)

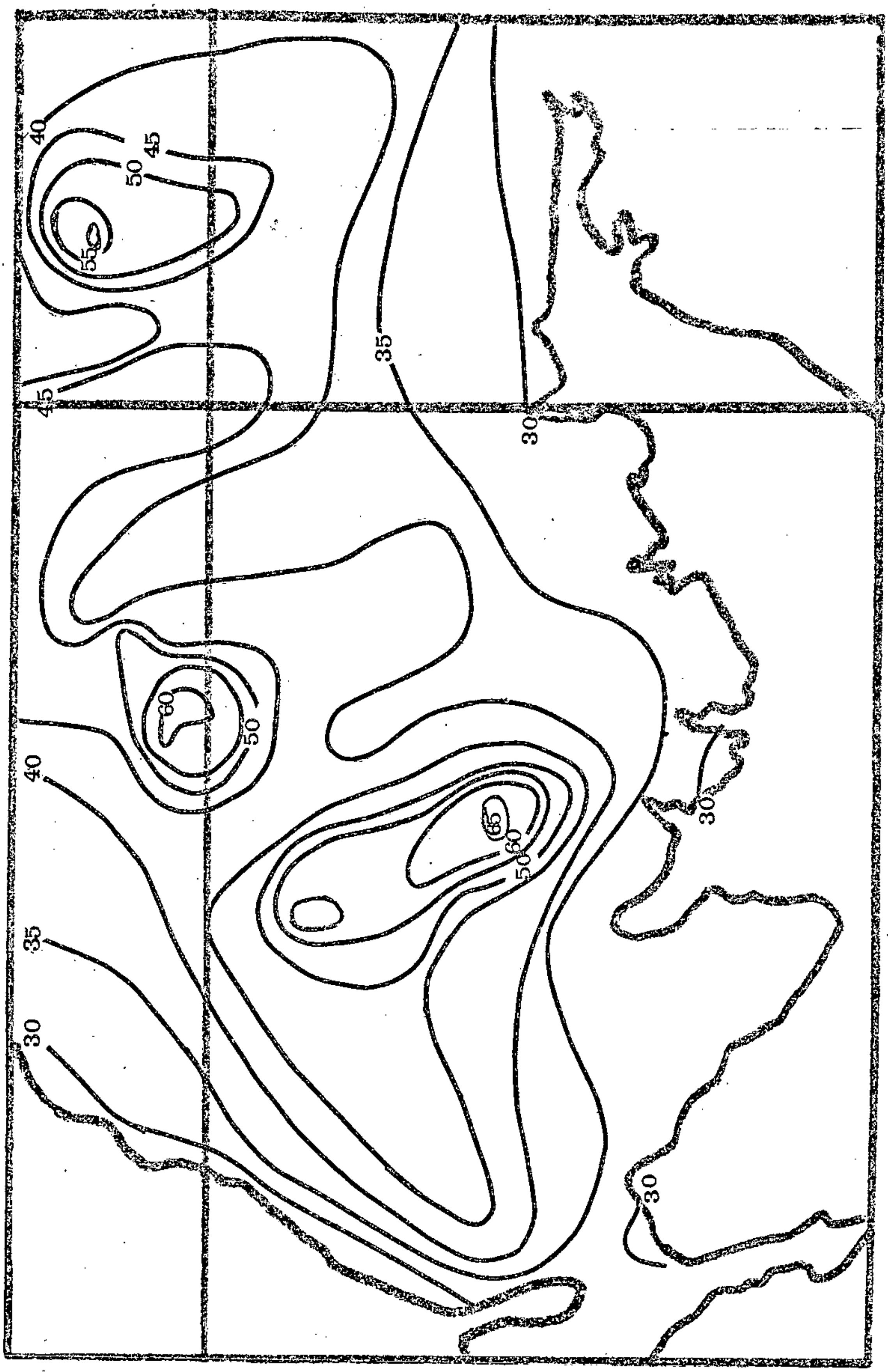

Figure 2.10 First eigenvector multiplier, mean monthly rainfall 1941-70 Solway region

considered as noise, in some other cases the second eigenvector (and perhaps others) might be significant. In these cases, eigenvector analysis would have isolated more information contained in the time-space matrix than simple averaging.

2.5.3 Regression analysis of first eigenvector multipliers with altitude

In order to further investigate the rainfall variations described by the eigenvector multipliers, regression analysis between first eigenvector multipliers and altitude were carried out. Table 2.3 contains a summary of the regression analyses, with correlation coefficients r, and regression parameters a and b in the regression equation 2.10 where y is eigenvector multiplier and x station height.

$$y = a x + b \qquad (2.10)$$

Table 2.3 Summary of regression analyses between first eigenvector multiplier and altitude

Area	No. of stations	Type	r	a	b
Solway 1916-50	69	long-term	0.819	11.76	0.0067
Solway 1931-70	207	long- and short-term	no direct information on station altitude		
Central Scotland	291	long- and short-term	0.576	10.50	0.0071
	82	long-term	0.374	11.28	0.0077

The three correlations are significant at the one per cent level using values tabulated by Fisher and Yates (1963).

The Central Scotland multiplier residuals, the differences between actual eigenvector multipliers and multipliers predicted by equation 2.10 are shown in Figure 2.7. These show that most of the rainfall variations apart from the effects of topography can be explained by the direction of the prevailing wind. The non-linearity of rainfall variation with altitude is shown by extra rainfall and positive residuals for stations exposed to the prevailing wind, and negative residuals for sheltered stations or stations in the lee of hills. The hills in the North West of the region lying in the prevailing wind have positive anomalies. The area in the centre of the figure at the crossing of the grid lines and enclosed by the 0 contour arises from the Ochil Hills releasing more rainfall than that expected by the regression relationship.

In sheltered areas the linear smoothing of multipliers for topography overcompensates for the increase in rainfall and hence multiplier value with altitude (in the exposed areas considered above it undercompensates). The Moorfoot Hills and the Pentlands lie within the area enclosed by the -6 contour in Figure 2.7. These hills are surrounded to the South and West by other ranges of hills, and are thus sheltered from the prevailing wind. Thus the increase in rainfall with altitude for these hills is much smaller than that in the North West of the region and rather smaller than that in the South of the region. The ranges of hills in the South are in their turn less exposed than those in the North-West as they are sheltered by further ranges of hills to the South and West.

In both Figures 2.6 and 2.7 the values of the multipliers decrease round the Forth Estuary, the contours tending to follow the outline of the estuary instead of lying nearly North-South at right angles to the prevailing wind direction. The areas around the estuary experience a sheltering effect on all sides except from the Easterly quarter; it is from this direction that about half of all the heavy falls of rain occur (Mossman 1896). Falls from this Easterly direction make an important contribution to rainfall, despite the relative infrequency of winds from this quarter.

2.6 Analysis of annual mean monthly rainfall and its use in interpolation

Analysis of the anomaly data for the 82 Central Scotland long-term stations produce a first eigenvector describing temporal variation in their deviation about the 82 station monthly mean (Fig. 2.11). The values decrease from a large value in winter (October to January) to a small value in summer (April to September) with a slight drop in November compared to the other winter months. This shows that rainfall has a higher variability between stations in winter than in summer. The first eigenvector accounts for 99.2 per cent of the variance in anomaly data.

The derived first eigenvector multipliers (Fig. 2.12) and the residuals from their regression with altitude (Fig. 2.13) produce patterns similar to those for real data (Fig. 2.6 and 2.7). However, as the isopleths now describe deviation of station rainfall rather than mean station rainfall, the positive anomalies in the North-West and South of the region and the negative anomalies around the Forth Estuary appear more pronounced. The regression

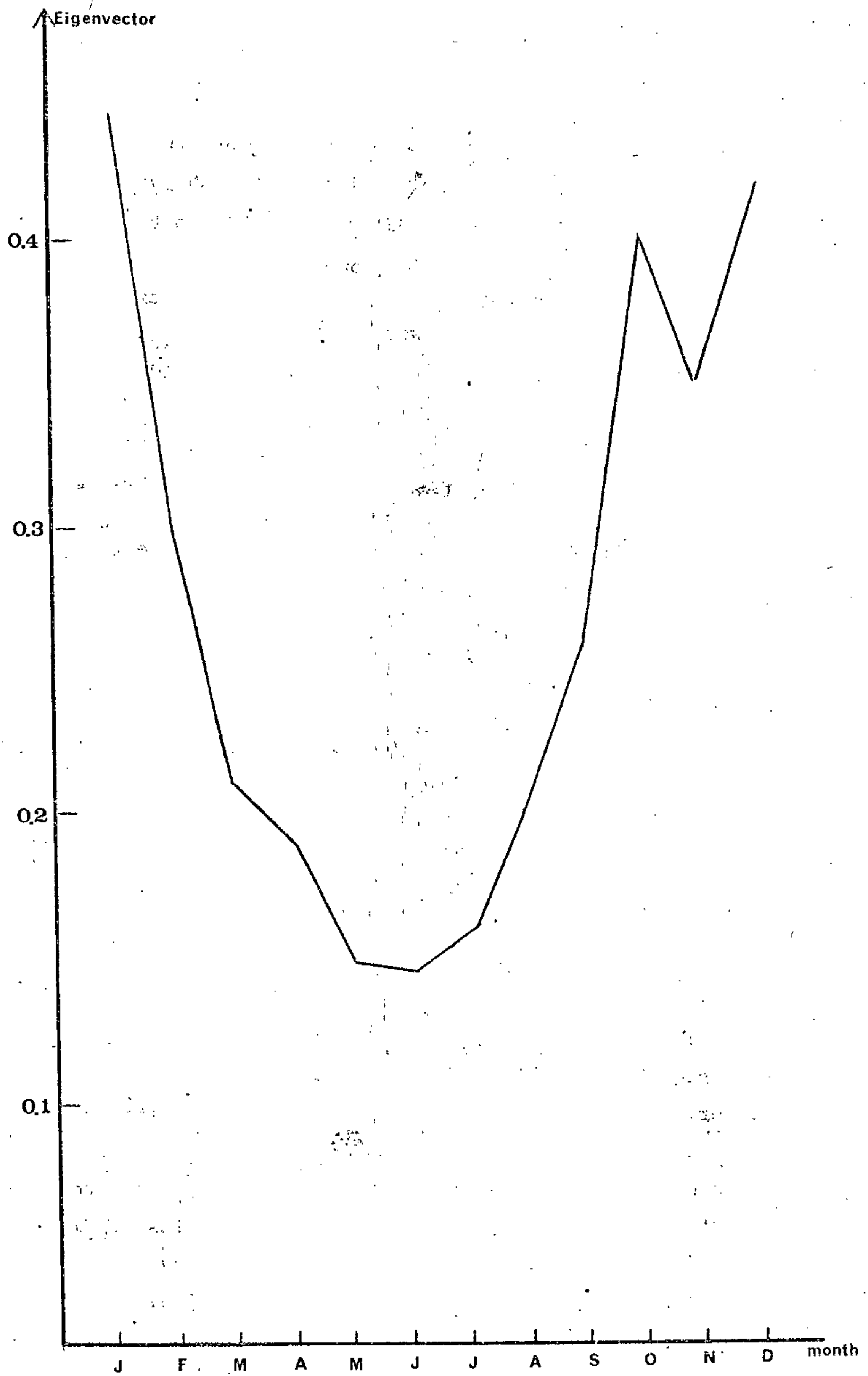

Figure 2.11 First eigenvector; anomaly monthly rainfall Central Scotland

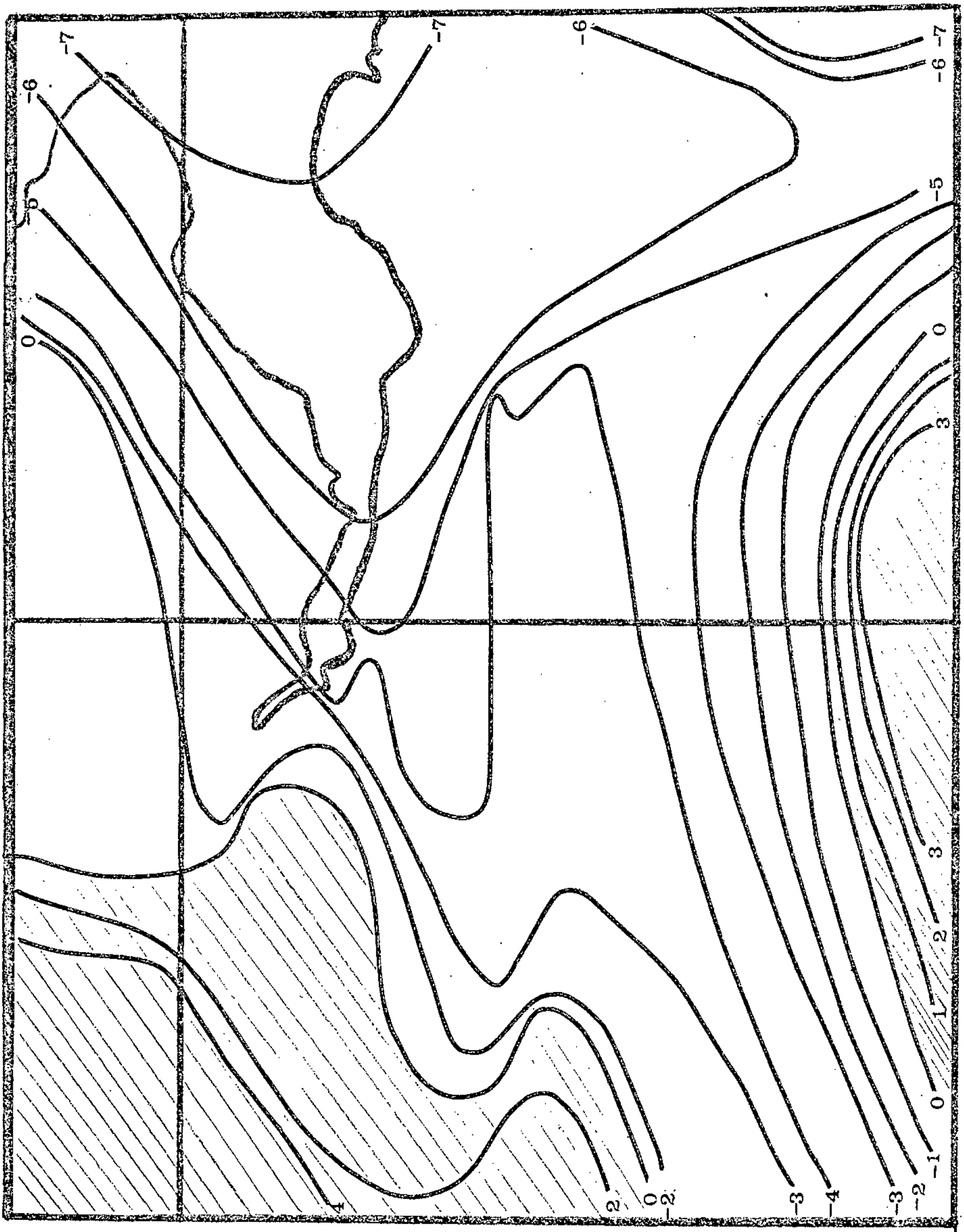

Figure 2.12 First eigenvector multipliers, anomaly monthly rainfall Central Scotland

Figure 2.13 First eigenvector multipliers, anomaly monthly rainfall Central Scotland smoothed for topography

parameters in equation 2.10 are $a = 3.87$, $b = 0.0088$ with a correlation coefficient of 0.41, which is significant at the one per cent level.

The smoothed anomaly multipliers when used with the all-station mean monthly rainfall and the regression relationship between multipliers and altitude, provide the maximum amount of information concerning mean monthly rainfall. Five sets of information, all-station mean monthly rainfall, monthly eigenvector, eigenvector multiplier-altitude regression relationship, station altitude, and eigenvector multiplier residuals are used to describe the mean monthly rainfall of a station. To "predict" mean monthly rainfall at a new site two values are required - the height of the site and the interpolated smoothed multiplier value.

As the anomaly smoothed multipliers have less overall variability between stations than the unsmoothed multipliers, or those of real data, it is easier to interpolate multiplier values to a new site. As many as possible of the smoothed isopleths were drawn in Figure 2.13, and values were interpolated at the sites of 209 short-term stations not used in this analysis. From these values, and using the other four sets of information listed above, twelve mean monthly rainfall values were predicted for each of these 209 stations.

These values were compared with those given in the Hydrological Memorandum, produced as described above by the interpolation extrapolation procedure from short-term records using surrounding long-term stations. The largest differences between the two sets of values occurred in areas where the smoothed multipliers were drawn close together (hence increasing the subjectivity involved in the interpolation) and in areas with a sparse distribution of readings.

In mountainous areas, rainfall can vary considerably over small distances between successive hills. Even when some of these variations have been removed by the regression analysis with altitude, the interpolation remains subjective. The variations over short distances which cannot be accounted for by a few gauges are further discussed in Chapter Three where daily rainfall values are regressed with altitude.

The standard deviations of the 202 short-term stations about their means for each month, produced by this method, were compared with the standard deviations of the original 202 values about their means. This choice of method takes account of the fact that monthly rainfall anomalies were used in the eigenvector interpolation procedure. The sets of mean monthly 202 station means do not differ significantly (Table 2.3) though those interpolated from the eigenvector method tend to be larger, perhaps as a result of the large number of gauges lying in the positive residual area in Figure 2.11. The total variance between rainfall stations explained using the eigenvector interpolation procedure is about 70 per cent of that in the extrapolation-interpolation procedure. It may be assumed that the latter values are accurate to five per cent and represent the best estimate of mean monthly rainfall for these stations. Considering the sparcity of the gauge network, the accuracy of the predictions using the eigenvector technique demonstrates that this method of extending a network of a series of mean data values is viable.

Table 2.3 Long-term means and standard deviations of 209 short-term stations (i) extrapolated by the standard procedure, (ii) interpolated from eigenvector analysis

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
(i) mean	5.60	3.81	3.16	2.94	3.18	2.96	3.89	4.39	4.35	5.56	4.91	5.00
standard deviation	7.82	3.62	1.81	1.49	0.92	0.89	1.09	1.70	2.77	6.43	4.80	7.05
(ii) mean	5.87	4.00	3.29	3.06	3.27	3.05	3.99	4.51	4.51	5.80	5.12	5.26
standard deviation	6.89	3.09	1.56	1.27	0.78	0.74	0.88	1.40	2.34	5.50	4.09	6.01

$$\frac{\sum \text{standard deviation (ii)}}{\sum \text{standard deviation (i)}} = 0.726$$

2.7 Summary and other possible uses of eigenvectors

In this Chapter the technique of eigenvector analysis has been explained. Its advantages over other types of analysis have been outlined. These arise from the orthogonality of each set of time and space components and from the fact that no assumptions are made about the underlying distribution represented by the original time-space matrix.

The way in which eigenvector analysis may "highlight" underlying distributions in the original matrix has been demonstrated. Mean and anomaly patterns of rainfall have been constructed with time and space components which filter out noise in data in a different manner to simple averaging. The manner in which eigenvector multiplier fields reflect rainfall variations associated with topography and the direction of the prevailing wind has been demonstrated.

Eigenvector analysis gives a clear reduction in the amount of data needed to represent the variations described by a space-time matrix. As only one eigenvector appears to be significant above, two column matrices with 12 and ST elements, where ST is the number of stations used, represent most of the information about rainfall variability described in the original 12 by ST matrix.

The use of eigenvector analysis to "predict" mean monthly rainfall at a new site has also been shown to given reasonable results. The sparseness of the original 82 gauge network considered above is overcome by removing noise present in the data by means of eigenvector analysis. Variations between gauges are further reduced by smoothing some of the variations in eigenvector multipliers due to topography using a linear regression relationship between multipliers and station altitude.

These uses of eigenvectors have previously been demonstrated by Stidd on Nevada mean monthly rainfall (1967). The further consideration of the derived space pattern as a "background" field for interpolation between readings in a specific situation has not been previously considered and is discussed in Chapter 3.

The use of eigenvectors to analyse time-series from several stations for common variations is considered in Chapter 5. Time series derived from eigenvector analysis of meteorological parameters such as pressure anomalies (Fritts 1971) or precipitation values (Le Marche and Fritts 1971) observed over periods of decades at several stations have also been compared with those derived from the analysis of comparable tree-ring data. By correlating the sets of eigenvectors derived respectively from climatic data and from tree-rings over a common period of time, it has been possible to extend climatic records retrospectively into periods for which only tree-ring data are available. While this particular technique is not discussed further in this thesis, it does demonstrate that eigenvector analyses of time-series from several sites do describe real variations present in the data which may be of predictive value.

It has been pointed out that the analysis on mean monthly data could be extended to monthly data of individual years. Eigenvectors and their multipliers derived from month-by-station matrices could then be compared. Each analysis might be expected to reveal more than the single significant eigenvector found in the case of mean values, and common types of variation might be revealed in eigenvectors of a given number for different years.

CHAPTER 3

3. REGRESSION STUDIES BETWEEN RAINFALL AND PHYSICAL PARAMETERS

3.1 Introduction

In this chapter multiple regression analyses between the physical parameters altitude or effective altitude, distance to West coast, distance to South coast, and daily rainfall, are used to isolate the influences on the rainfall of a network of stations for particular days of "pure" synoptic type. The effects of topography and distance from the sea on individual station rainfall are expressed by a linear regression relationship. The validity of the approximation of a linear increase of rainfall with altitude is discussed using the results of these analyses.

3.2 General points on the choice of rainfall regression parameters

The following points should be considered when performing regression analyses between rainfall and physical parameters:

1. Any physical parameter in a regression equation may describe more than one effect of a station's position on rainfall. The task of isolating different influences on rainfall is not straight forward. In the regression analyses below, altitude varies with distance of a station from both coasts as hills rise away from the sea. Thus it

is not possible to distinguish effects of topography from those of distance from the sea in this case.

2. While it is possible to represent the position of a station by several different sets of parameters, the set which is most suitable in a given case will depend on the area studied. Regression parameters obtained for particular areas (and synoptic situations) must be viewed with caution if used to "predict" rainfall in neighbouring areas.

3. The actual station height is not always the best parameter to describe the height of the gauge. A parameter describing the height of the land surrounding the gauge whose rainfall the gauge represents, often gives best results in regression analyses. The manner of computing effective height, and the area whose height should be assessed, depend on the physical shape and relative height of the hills surrounding the station. Storr and Ferguson, in their analysis of monthly rainfall, used effective height assessed over nine 5-km grid points. Chuan and Lockwood (1974) in their analysis of mean annual and seasonal rainfall of the Western Pennines used mean height assessed over circles, drawn at intervals of 1-km from the station, with four points at each interval together with the station height. They found that mean altitude over an 8-km radius gave the best correlations. It is demonstrated below that this parameter is not suited to describe the effective height of Scottish stations.

4. A regression relationship between rainfall and altitude assumes that rainfall increases with height in a particular way, usually linearly. Rain shadow effects are not accounted for. It is found in some cases (using other methods) that rainfall ceases to increase with height on the windward side of a mountain above a certain height (Lefreve, 1972; Alam, 1972) and in others that rainfall increases beyond the physical peak on the lee-side, (Storr and Ferguson, 1972). Results of linear regression analyses should be investigated to see if they reveal such effects.

5. The amount of rainfall falling in an area is generally affected by the topography of neighbouring hills. Besides the effect of moisture being precipitated on neighbouring hills, the system may be itself dynamically modified. Storr and Ferguson (1972) used the parameters of distance to barrier, barrier-height, and shield effect to describe the influences of neighbouring hills in the regression analysis of monthly rainfall. The "barrier" was the highest elevation upwind of the station along the direction of the prevailing wind. The shield effect described the total effect of all the neighbouring hills by summation of the barrier height and the other local barriers along the prevailing wind direction. Alam (1972) used a parameter of barrier height to label members of a series of rainfall v. elevation curves.

6. Parameters such as gauge exposure (the angle in radians in which there is no topographic feature higher than the gauge), maximum rise (the range of height between the highest

and lowest points within a given radius) used by Chuan and Lockwood, or aspect (the direction of the mountain slope on which the gauge lies relative to the axis of the mountain), used by Alam, may be used as secondary regression parameters. While they take account of the detailed position of the gauge, they are difficult to assess and usually only take account of a small proportion of the total variation between gauges.

3.3 The selection of data for regression of daily rainfall

3.3.1 The area

The points mentioned above were borne in mind in the selection of the area and of the physical parameters to study the use of regression analyses to describe daily rainfall variations. The problems of lack of data mentioned in Chapter 1, the inadequate network of daily gauge records in remoter parts of the British Isles and the necessity to have data of "pure" synoptic type, were also considered in the choice. The area around the Solway Firth was chosen as systems approaching this area are little affected by the topography of neighbouring regions; the only sheltering effects occur to the North and East, from which directions few systems bring intense rainfall. Thus the problems of modification of a pure system by topography and of the parameterisation of the effects of airflow over neighbouring hills were not present in this region. The rainfall gauge network for this region (Figure 3.1) was also denser than that for similar areas in the Western Highlands exposed to approaching systems.

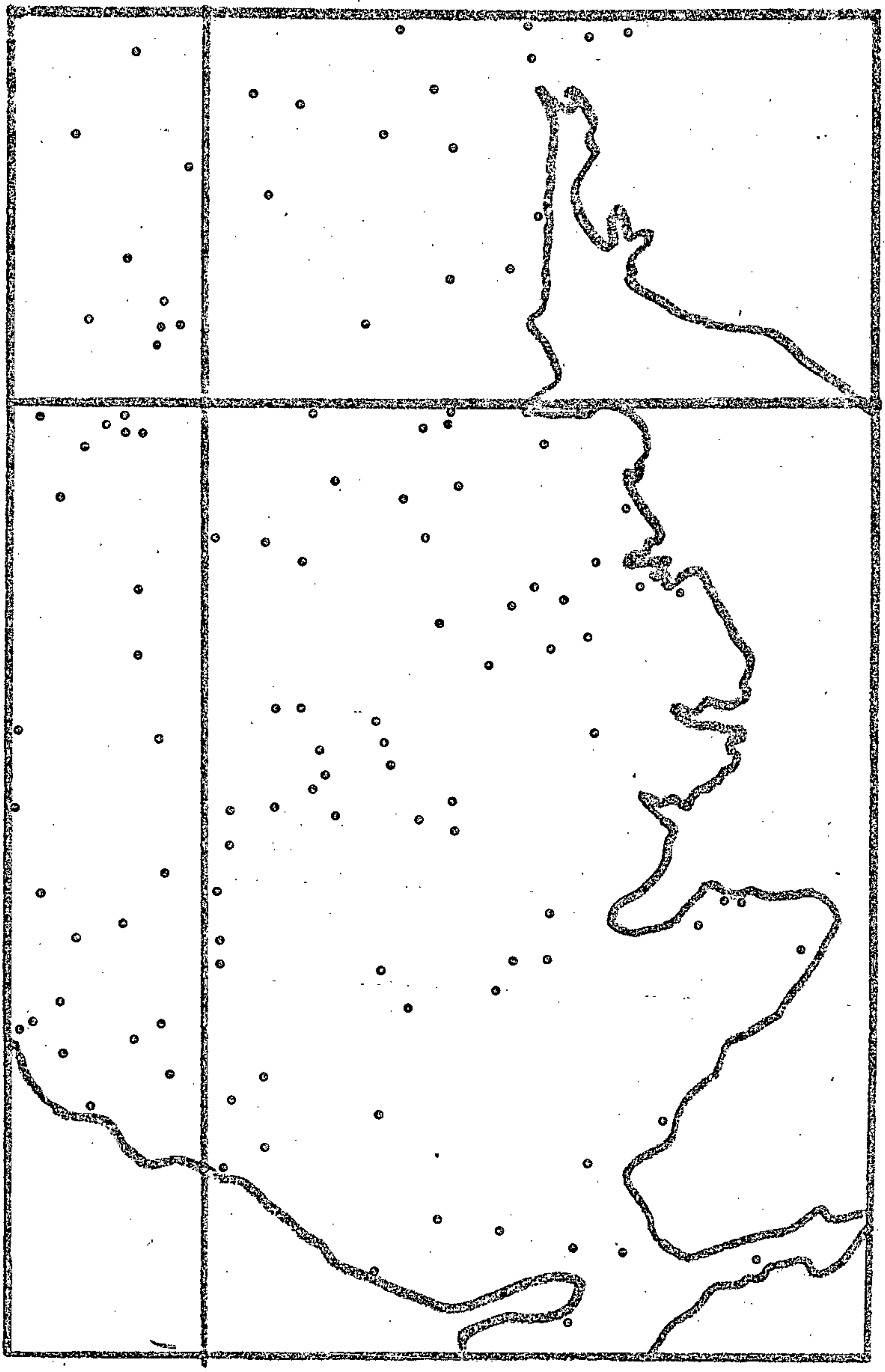

Figure 3.1 Solway daily station network

Of the 110 gauges about 80 had records available for any particular day.

The topography of the area is illustrated in Figure 2.8. The Southern Uplands rise to the North and East from the Solway Firth. The hills are more rounded than the Pennines, considered in Chuan and Lockwood's analysis, but less rounded than those of the Scottish Highlands.

3.3.2 Physical parameters

Distance from the South and West coasts, d_s and d_w, gauge height h or effective height, $\bar{h}$; were considered to be the most important parameters relating to rainfall variations. Secondary parameters, such as gauge exposure and aspect, were not considered, being difficult to assess from the study of a detailed topographical map, and having been found in other analyses often to be of little significance. The effective gauge height parameter was expected to describe some of the variations which would be described by secondary parameters.

The effective height parameter was found by averaging the height of the station and those of the four surrounding $3\frac{1}{3}$ km. grid points. This method of assessment was chosen because the $3\frac{1}{3}$ km. grid of topography was readily available, and it was an easier method of assessment than Chuan and Lockwood's method of assessment of mean altitude over 3 km. using 13 spot heights. The difference between the $3\frac{1}{3}$ km. grid height and height over 3 km. radius was not expected to be significant for the area studied.

The fact that height of land over the 3 km. surrounding the gauge was more appropriate to this case than height over 8 km. was suggested by the results in Table 3.1, where heights over 8 km. for Central Scotland and for the Pennines are given. For Talla Lins Foot, Victoria Lodge, and Stobbo Castle, the mean height is larger than spot height but there is no corresponding increase in mean annual rainfall, showing that the hills near the distance of 8 km. from the gauge do not increase its mean annual rainfall. The Edinburgh gauges, those at Blackford Hill and Astley Ainslie Hospital, are situated 600 metres apart, and are located in an area with topography broadly similar to that of the Solway region. The two gauges have a 3.6 per cent difference in mean annual rainfall while the difference in their 8 km. mean height is 11 per cent and that in spot height is -60 per cent. These variations show that an 8 km. mean height is not an appropriate parameter to describe the effective height of gauges in Central Scotland with respect to rainfall. Land lying within a smaller distance of a gauge than this determines the effective gauge height. The 8 km. radius mean height is more appropriate to the Pennines, which are less rounded and slightly lower than the hills of Southern and Central Scotland.

While secondary parameters such as gauge exposure and aspect were not considered in the regression analyses, and while it was hoped that effective height would describe gauge position better than would spot height, the height parameters, h or $\bar{h}$, were themselves correlated with d_w and d_s (Table 3.4). The method of stepwise regression ensured that as each variable was added the effects it described were not already explicitly represented in the regression relationship. However, because of the inter-relationship between parameters, the increase in the square of the

Table 3.1 8 km radius mean height, spot height and mean annual rainfall - 1. Edinburgh Area
2. West Pennines

Station	Grid Reference	spot height ft.	mean height ft.	mean annual rainfall ins. 1916 - 50
1.				
Ochil Hills Hospital	30977076	800	748	47.49
Uphall No. 8	30246708	577	460	34.75
Middleton Hall	30616716	350	392	33.84
Harperig	31026613	900	962	40.63
Edinburgh Blackford Hills	32596706	441	269	27.53
Astley Ainslie Hospital	32516713	270	304	28.53
Fairmilehead Waterworks	32496683	590	405	31.18
Liberton	32736690	407	384	28.75
Glen Cottage	32236635	739	908	38.12
Glencorse Filters	32256631	638	815	36.04
Martyr's Cross	32296623	750	882	37.24
Gladhouse Res.	32996544	915	1016	37.67
Roseberry	33086570	750	786	33.67
North Berwick	35556853	51	62	25.69
West Calder Addiewell	30016626	620	704	37.69
Talla Lins Foot	31336203	966	1602	61.04
Victoria Lodge	31066231	900	1421	50.14

/...

Table 3.1 continued

Station	Grid Reference	spot height ft.	mean height ft.	mean annual rainfall ins. 1916 - 50
Carden Knowes	35776372	300	558	29.58
Stobo Castle	31796368	594	1100	38.50
Floors Castle	37076345	195	316	25.95
Spittal Tower	35876182	425	516	31.12
2.				
Swineshaw Moor	40104008	1340	890	48.57
Black Clough	41273984	1643	1430	59.16
Pikenage	40984001	926	1375	50.85
Upper Headon	40984035	1717	1369	69.83

Table 3.2 Description of Case Studies

Case	Date	Weather Type	Wind	Remarks
A	05.08.71	Thundery low		
B	29.11.72	Cold front	W	Cold front from NW. Some waving occurred.
C	12.02.73	Showery low	NW	Low remained to the East and gave heavy snow showers.
D	10.12.73	Cold front	SW veering NW	Front arrived from NW
E	10.01.74	Occlusion	SW	Occlusion moved NE
F	17.01.74	Warm sector	W	
G	04.09.74	Occlusion	W	Occlusion moved NE
H	12.09.74	Cold front	S	Cold front moved NE. Some development.
I	20.12.74	Warm sector	W	
J	21.12.74	Warm front	SW	Warm front moved up from South and then retreated.
K	16.02.75	Occlusion	W	Occlusion moved E, and developed a wave. The warm sector on wave moved N.

multiple correlation coefficient, and thus in the amount of variance described, at steps 2 and 3 was small (Table 3.3).

3.3.3 The "pure type" days

The cases used for the multiple regression studies were chosen such that each distribution was of one particular type and that half of the gauges had a reading of over 10 mm, and some (about 10) had readings over 20 mm. These days (Table 3.2) included winter-time cold fronts (cases B and D) and winter-time warm sectors (cases F and I), the types of system Browning, Hill, and Pardoe (1974, 1975) studied for the effects of topography on rainfall. A summer-time cold front (case H) was included for comparison with cases B and D. The showery low with its associated North Westerly airstream was notable for the heavy orographic rainfall it produced to the North of the area.

Occluded fronts (cases E, G and K) of varying complexity were also investigated to see to what extent rainfall from more complicated systems may be related to physical parameters, the case K being particularly complex due to the development of a wave. The thundery low (case A) is included as a case where rainfall, though heavy, was definitely not of an essentially orographic nature.

All these systems could be classed as producing intense rainfall due to a front, a non-frontal depression, a warm sector, or an individual convective storm. The mean annual rainfall distribution was similarly analysed from comparison.

3.4 The method of multiple regression analysis

A stepwise multiple analysis regression was carried out using the Biomedial Computer Program supplied by the University of California. At each step in the analysis, the variable which gives the greatest reduction in the variance of the dependent variable is added. This variable may also be described as the one with the highest partial correlation coefficient with the dependent variable which has been partialed on the variables already added. The "P value" of a dependent variable, which is defined as the square of the ratio of the regression coefficient to its standard deviation, measures the relative importance of a regression parameter in describing the variance of the dependent variable. The variables added may be restricted to those of given significance by specifying a controlling P value for variables not to be entered into the regression equation. At each step the variable with highest P value will be added provided this value is larger than the control value. A detailed discussion of the method is given in the BMD manual (Dixon 1968) and in Effroymsen (1960).

In the regression equation:

$$y = a + bx_1 + cx_2 + dx_3 \qquad (3.1)$$

y is used to "predict" rainfall in mm. from x_1, altitude or mean altitude in metres, x_2 distance to the South coast in km, and x_3 distance to the West coast in km. x_1, x_2 and x_3 will be referred to as h or $\bar{h}$, d_s and d_w respectively in the analysis below.

Analysis of "transformed" rainfall, daily rainfall expressed as a percentage of mean annual rainfall and first eigenvector multiplier, were also carried out to determine the extent to which rainfall distributions of specific cases might be described in terms of these mean distributions. The mean annual rainfall and first eigenvector multiplier had to be interpolated from Figures 2.9 and 2.10 for daily stations with no long-term means.

In order to examine the validity of the linear assumptions in the analyses, regression residuals, i.e. differences between actual rainfall values and those predicted by equation 3.1, were plotted and isopleths drawn. The particular effects of increased rainfall on exposed slopes and of rain shadows were looked for in the patterns of the isopleths. These derived regression parameters and residuals were expected to provide answers as to the validity of the linear assumption mentioned in Point 3. above.

3.5 Discussion of results

3.5.1 Real rainfall data regression analyses

A summary of the steps in the regression analysis of real rainfall data appears in Table 3.3. The addition of extra variables beyond the first step did not increase the amount of variance in rainfall explained, the increase in R^2 being small. The correlation matrix between mean annual rainfall and physical parameters appears in Table 3.4; this illustrates that "independent" variables are in fact interrelated and implies limited usefulness of more than one such variable.

Table 3.3 Summary of steps in multiple regression analysis using (i) mean altitude, (ii) spot altitude. R - multiple correlation coefficient, R^2, increase in R^2 at each step, P value to enter variable.

Case	Date	Step No.	Variable added	R	R^2	Increase in R^2	P	No. of variables added
(i)	annual mean	1	$\bar{h}$	0.763	0.581	0.581	150.1	1
		2	d_w	0.768	0.590	0.009	2.3	2
		3	d_s	0.777	0.604	0.013	3.6	3
(ii)	annual mean	1	h	0.776	0.603	0.603	163.8	1
		2	d_w	0.784	0.615	0.013	3.5	2
		3	d_s	0.788	0.621	0.005	1.5	3
A(i)	05.08.71	1	d_s	0.433	0.188	0.188	22.7	1
		2	$\bar{h}$	0.521	0.272	0.084	11.1	2
		3	d_w	0.533	0.285	0.013	1.8	3
(ii)	05.08.71	1	d_s	0.433	0.188	0.188	22.7	1
		2	d_w	0.507	0.257	0.069	9.0	2
		3	h	0.512	0.262	0.006	0.7	3
B(i)	29.11.72	1	d_s	0.466	0.217	0.217	25.0	1
		2	$\bar{h}$	0.481	0.231	0.014	1.6	2
		3	d_w	0.520	0.270	0.039	4.8	3
(ii)		1	d_s	0.466	0.217	0.217	25.0	1
		2	d_w	0.484	0.234	0.017	1.9	2
		3	h	0.525	0.276	0.042	5.1	3
C(i)	12.02.73	1	d_w	0.520	0.271	0.271	33.4	1
		2	$\bar{h}$	0.556	0.309	0.039	5.0	2

/...

Table 3.3 continued

Case	Date	Step No.	Variable added	R	R^2	Increase in R^2	P	No. of variables added
C(i)	12.02.73	3	d_s	0.575	0.330	0.021	2.8	3
(ii)		1	d_w	0.520	0.271	0.271	33.4	1
		2	d_s	0.537	0.287	0.018	22.3	2
		3	h	0.550	0.303	0.014	18.2	3
D(i)	10.12.73	1	$\bar{h}$	0.576	0.332	0.332	45.2	1
		2	d_w	0.603	0.364	0.032	4.5	2
		3	d_s	0.639	0.409	0.045	6.8	3
(ii)		1	h	0.517	0.267	0.267	33.1	1
		2	d_w	0.558	0.311	0.044	5.8	2
		3	d_s	0.609	0.371	0.062	8.5	3
E(i)	10.01.74	1	$\bar{h}$	0.719	0.518	0.518	99.8	1
		2	d_s	0.742	0.551	0.033	5.8	2
		3	d_w	0.744	0.553	0.027	0.5	3
(ii)		1	h	0.626	0.391	0.391	57.8	1
		2	d_s	0.668	0.446	0.055	9.2	2
		3	d_w	0.682	0.465	0.018	3.1	3
F(i)	17.01.74	1	d_w	0.660	0.435	0.435	72.4	1
		2	d_s	0.718	0.516	0.080	15.4	2
		3	$\bar{h}$	0.720	0.518	0.002	0.4	3
(ii)		1	d_w	0.660	0.435	0.435	72.4	1
		2	d_s	0.718	0.516	0.080	15.8	2
		3	h	0.727	0.528	0.012	2.4	3
G(i)	04.09.74	1	$\bar{h}$	0.306	0.093	0.093	9.0	1

/...

Table 3.3 continued

Case	Date	Step No.	Variable added	R	R^2	Increase in R^2	P	No. of variables added
G(i)	04.09.74	2	d_w	0.377	0.142	0.049	4.9	2
		3	d_s	0.399	0.159	0.017	1.7	3
(ii)		1	d_w	0.321	0.103	0.103	10.0	1
		2	h	0.405	0.164	0.061	6.3	2
		3	d_s	0.428	0.183	0.019	2.0	3
H(i)	12.09.74	1	d_w	0.186	0.035	0.035	3.2	1
		2	$\bar{h}$	0.323	0.104	0.070	6.9	2
		3	d_s	0.324	0.105	0.001	0.1	3
(ii)		1	d_w	0.186	0.035	0.035	3.2	1
		2	h	0.224	0.050	0.016	1.5	2
		3	d_s	0.241	0.058	0.008	0.7	3
I(i)	20.12.74	1	$\bar{h}$	0.578	0.335	0.335	44.2	1
		2	d_s	0.665	0.442	0.108	16.8	2
		3	d_w	0.710	0.505	0.062	10.8	3
(ii)		1	h	0.593	0.352	0.352	47.8	1
		2	d_s	0.693	0.480	0.127	21.3	2
		3	d_w	0.727	0.528	0.048	8.8	3
J(i)	21.12.74	1	d_s	0.612	0.374	0.374	52.6	1
		2	d_w	0.805	0.648	0.273	6.7	2
		3	$\bar{h}$	0.825	0.681	0.033	8.9	3
(ii)		1	d_s	0.612	0.374	0.374	52.6	1
		2	d_w	0.805	0.648	0.273	6.7	2
		3	h	0.826	0.682	0.035	9.4	3

/...

Table 3.3 continued

Case	Date	Step No.	Variable added	R	R^2	Increase in R^2	P	No. of variables added
K(i)	16.02.75	1	$\bar{h}$	0.295	0.087	0.087	8.0	1
		2	d_w	0.337	0.114	0.026	2.5	2
		3	d_s	0.340	0.114	0.035	9.4	3
(ii)		1	h	0.231	0.053	0.053	4.8	1
		2	d_w	0.272	0.074	0.020	1.8	2

Table 3.4 Correlation matrix between mean annual rainfall and physical parameters

	$\bar{h}$	d_s	d_w	mean annual rainfall
$\bar{h}$	1.000	0.540	0.092	0.763
d_s		1.000	-0.335	0.333
d_w			1.000	0.011
mean annual rainfall				1.000

In Table 3.5, it is shown that mean annual rainfall has higher correlation coefficients with physical parameters than rainfall on individual days. This in turn shows that mean annual rainfall is more clearly related to topography that is that of individual days; and may imply that the daily rainfall data do not reflect entirely "pure type" rainfall.

The parameter effective altitude did not produce significantly better correlations with rainfall than spot altitude, and in the case of mean annual rainfall and cases C, D, E and K, the use of spot altitude produced higher correlation coefficients.

Altitude was the most significant regression parameter for six cases. Multiple correlation coefficients were higher for all winter-time cases (except the developing occlusion, case K) than for summer-time systems. As case K developed a wave, it was not in some senses a pure type. Cases K and H, both of which were developing systems, produced the lowest multiple correlation

Table 3.5 Summary of regression analyses of daily rainfall using (i) mean altitude, (ii) spot altitude. R - multiple correlation coefficient, a - regression constant, regression coefficients between rainfall and, b - altitude h or $\bar{h}$, c - distance to West coast d_w, d - distance to South coast d_s, and F - ratio of variance. * and ** denote significance at five per cent and one per cent levels respectively.

Case	Date	R	a	b	c	d	F	order variables entered
(i)	mean annual rainfall	0.777	1155.9	2.1549	-2.5900	-1.1057	53.9**	$\bar{h}, d_w, d_s$
(ii)		0.788	1162.3	2.3956	-2.441	-0.0689	57.8**	h, d_w, d_s
A(i)	05.08.71	0.534**	8.60	0.0127	0.0584	-0.0782	12.7*	$d_s, \bar{h}, d_w$
(ii)		0.512**	8.70	0.0064	0.0862	-0.0720	11.4*	d_s, d_w, h
B(i)	29.11.72	0.520**	15.97	0.0166	-0.1002	0.0780	10.9*	$d_s, \bar{h}, d_w$
(ii)		0.525**	15.98	0.0198	-0.1027	0.0799	11.2*	d_s, h, d_w
C(i)	12.02.73	0.575**	8.85	0.0127	0.0848	-0.0248	14.5*	$\bar{h}, d_w, d_s$
(ii)		0.550**	8.91	0.103	0.0983	-0.0203	12.7*	h, d_w, d_s
D(i)	10.12.73	0.639**	4.82	0.0224	0.1146	0.0461	20.5*	$\bar{h}, d_w, d_s$
(ii)		0.609**	4.90	0.0201	0.1321	0.0526	17.5*	h, d_w, d_s

/...

Table 3.5 continued

Case	Date	R	a	b	c	d	F	order variables entered
E(i)	10.01.74	0.744**	6.94	0.0569	0.0360	0.0620	37.6**	$\bar{h},d_s,d_w$
(ii)		0.682**	6.59	0.0516	0.0915	0.0883	26.3**	$\bar{h},d_s,d_w$
F(i)	17.01.74	0.720**	12.09	0.0083	0.4348	-0.1504	33.0**	$d_w,d_s,\bar{h}$
(ii)		0.727**	12.22	0.0223	0.3923	-0.1597	34.3**	d_w,d_s,h
G(i)	04.09.74	0.399**	13.19	0.0204	-0.0551	0.0261	5.4	$\bar{h},d_w,d_s$
(ii)		0.428**	13.33	0.0257	-0.0659	0.0272	6.3	h,d_w,d_s
H(i)	12.09.74	0.324*	22.91	0.0359	-0.2341	0.0159	3.4	$d_w,\bar{h},d_s$
(ii)		0.241	23.09	0.0162	-0.1573	0.0415	1.8	d_w,h,d_s
I(i)	20.12.74	0.710**	-1.20	0.0130	0.0655	0.0530	29.2**	$\bar{h},d_s,d_w$
(ii)		0.726**	-1.09	0.0163	0.0582	0.0537	32.1**	h,d_s,d_w
J(i)	21.12.74	0.825**	-6.98	0.0129	0.1574	0.1600	61.1**	$d_s,d_w,\bar{h}$
(ii)		0.826**	-6.83	0.0151	0.1536	0.1611	61.6	$d_s,d_w,\bar{h}$
K(i)	16.02.75	0.338*	15.15	0.0300	-0.0963	-0.0065	3.53	$\bar{h},d_w,d_s$
(ii)		0.272	15.47	0.0283	-0.826			h,d_w

coefficients.

Of the two cases, H and F, where d_s was the most important regression parameter, the rapid decrease of rainfall in case H with increasing distance from the South coast could be explained by strong Southerly winds. In case F, on the other hand, the increase in rainfall away from the South coast probably reflected a variation of rainfall and topography; d_s and $\bar{h}$ are significantly interrelated (Table 3.4). Some rainfall for case F, a winter-time warm sector, was expected to be of an intense orographic nature.

The two systems, B and J, for which d_s was the most significant regression parameter, were accompanied by strong West winds. In case B rainfall increased from the West, and to a lesser extent from the North, as the front developed in its motion eastwards. System A, the thundery low, originated in the West and became less active as it moved eastwards.

The highest of the regression coefficients b, describing the rate of increase of rainfall with altitude, occurred for the winter-time occlusion case E, with a value twice as large as any other values of b. If the cases studied are representative of intense systems crossing the area, this result would imply that the heaviest rainfall of an orographic nature occurs for winter-time occlusions. The highest values of c and d occurred in case J where the regression constant a is negative.

There would seem to be few generalisations to make concerning regression coefficients from Table 3.5. Cases B and D, were similar in synoptic type with fronts following similar paths across the area from the North West, but the subsequent development in case B meant that the regression parameters were different.

Table 3.6 Summary of regression analyses of daily rainfall expressed as (1) a percentage of mean annual rainfall, (2) a percentage of eigenvector field. Multiple correlation coefficient R, regression constant a, regression coefficient between rainfall and: b - mean altitude $\bar{h}$, c - distance to West coast d_w, d - distance to South coast d_s, and F ratio of variance. * and ** denotes significance at five per cent and one per cent levels respectively.

Case	Date	R	a	b	c	d	F	order variables entered
A(i)	05.08.71	0.504**	0.75		0.006	-0.006	16.3*	d_s, d_w
(ii)		0.507**	26.4	-0.0081	0.232	-0.203	11.0*	$d_s, d_w, \bar{h}$
B(i)	29.11.72	0.647**	1.44	-0.0008	-0.007	0.007	21.1*	$d_s, d_w, \bar{h}$
(ii)		0.677**	50.4	-0.0402	-0.228	0.262	24.8*	$d_s, \bar{h}, d_w$
C(i)	12.02.72	0.522**	0.72	-0.0001	0.009	-0.001	11.0*	$d_w, d_s, \bar{h}$
(ii)		0.554**	25.0	0.0132	0.331	-0.004	13.0*	$d_w, \bar{h}, d_s$
D(i)	10.12.73	0.577**	0.04	0.0003	0.001	0.000	15.0*	$d_w, d_s, \bar{h}$
(ii)		0.609**	14.9	0.0079	0.381	0.152	17.7*	$d_w, d_s, \bar{h}$
E(i)	10.01.74	0.560**	0.66	0.0022	0.006	0.006	13.8*	$\bar{h}, d_s, d_w$
(ii)		0.556**	23.8	0.0059	0.218	0.217	13.6*	$\bar{h}, d_s, d_w$

/...

Table 3.6 continued

Case	Date	R	a	b	c	d	F	order variables entered
F(i)	17.01.74	0.809**	0.92	-0.0016	0.041	-0.010	58.0**	$d_w, d_s, \bar{h}$
(ii)		0.805**	32.1	-0.0050	1.444	-0.352	56.4**	$d_w, d_s, \bar{h}$
G(i)	04.09.74	0.319*	1.15		-0.0003	0.003	4.9	$d_s, d_w, \bar{h}$
(ii)		0.324*	40.1	-0.0065	-0.0967	0.101	3.3	d_s, d_w
H(i)	12.09.74	0.329*	1.89	0.0009	-0.0174	0.003	3.5	$d_w, d_s, \bar{h}$
(ii)		0.346*	66.2	0.0159	-0.0580	0.117	4.0	$d_w, d_s, \bar{h}$
I(i)	20.09.74	0.691**	-0.03	0.0055	0.0055	0.005	26.3**	$d_s, d_w, \bar{h}$
(ii)		0.697**	-1.4	0.0071	0.1949	0.163	27.0**	$d_s, d_w, \bar{h}$
J(i)	21.09.74	0.828**	-0.04	0.0001	0.0125	0.012	62.4**	$d_s, d_w, \bar{h}$
(ii)		0.840**	14.2	-0.0033	0.0433	0.426	68.7**	$d_s, d_w, \bar{h}$
K(i)	16.02.75	0.194	1.23	0.0009	-0.0061	0.000	1.1	$d_w, \bar{h}, d_s$
(ii)		0.195	41.3	0.0222	-0.1798	0.043	1.1	$d_s, d_w, \bar{h}$

3.5.2 Transformed rainfall data regression relationships

In order to investigate to what extent variations in daily rainfall can be represented by those of mean annual rainfall or the first eigenvector multipliers of Chapter 2, regression analyses of transformed data were performed as discussed above and the results given in Table 3.6 were compared with those of the real data given in Table 3.4. Reductions in values of multiple correlation coefficients are small, showing that the background fields do not account for much of the variance in daily gauge values. For cases B and F multiple correlation coefficients are larger for transformed data, thus showing that in some respects rainfall may be classed as abnormal for these cases. Case B, being a developing cold front, might be expected to show abnormal variations on other grounds.

The biggest reductions in correlation coefficients using transformed data occur for the cases E and K, and are of order 0.1. The reductions in size of correlation coefficients are nearly the same in both cases, showing that the first eigenvector multipliers of real mean monthly rainfall data describe the same variations as those in mean annual rainfall data.

Though the reductions in correlation coefficients are not large, altitude is a considerably less important parameter when transformed data are used. This shows that the principal variations in daily rainfall values which can be described by mean annual rainfall are those due to topography.

3.5.3 Rainfall regression residuals

In order to observe which variations in rainfall due to topography can be accounted for by linear regression analyses, charts of rainfall regression residual isopleths, Figures 3.2 to 3.9 were studied with reference to the topographical map, Figure 2.8, and to the synoptic conditions of each case, Table 3.2. In Figure 3.2 isopleths of mean annual rainfall residuals are plotted at intervals of 10 mm, while in Figures 3.2 to 3.9 residuals for cases A, B, C, F, G and H are plotted at intervals of 5 mm. Areas of positive anomaly (observed rainfall larger than regression prediction) are hatched. The placing of the isopleths was somewhat subjective in regions where the gauge distribution was sparse.

Figure 3.2 provides the most information as to the validity of the assumption of a linear relationship between rainfall and topography. It is mean annual rainfall which has the highest correlation coefficient with altitude and thus best describes rainfall variations with topography. In Figure 3.2, positive anomalies occur on the sides of hills exposed to the prevailing wind. This is particularly noticeable for the ranges of hills in the area marked X, and to a lesser extent for that marked Y. Rainshadows and negative anomalies are seen to the lee of these hills, though not necessarily immediately in their lee.

The existence of large positive and negative anomalies, especially those lying near the coast lines, implies that the additional information explained by the parameters d_w and d_s compared to $\bar{h}$ alone, in the mean annual rainfall analysis, is small. The size of these anomalies, together with the small increases in R^2 in Table 3.3 for steps 2 and 3, show that the use of d_w and d_s

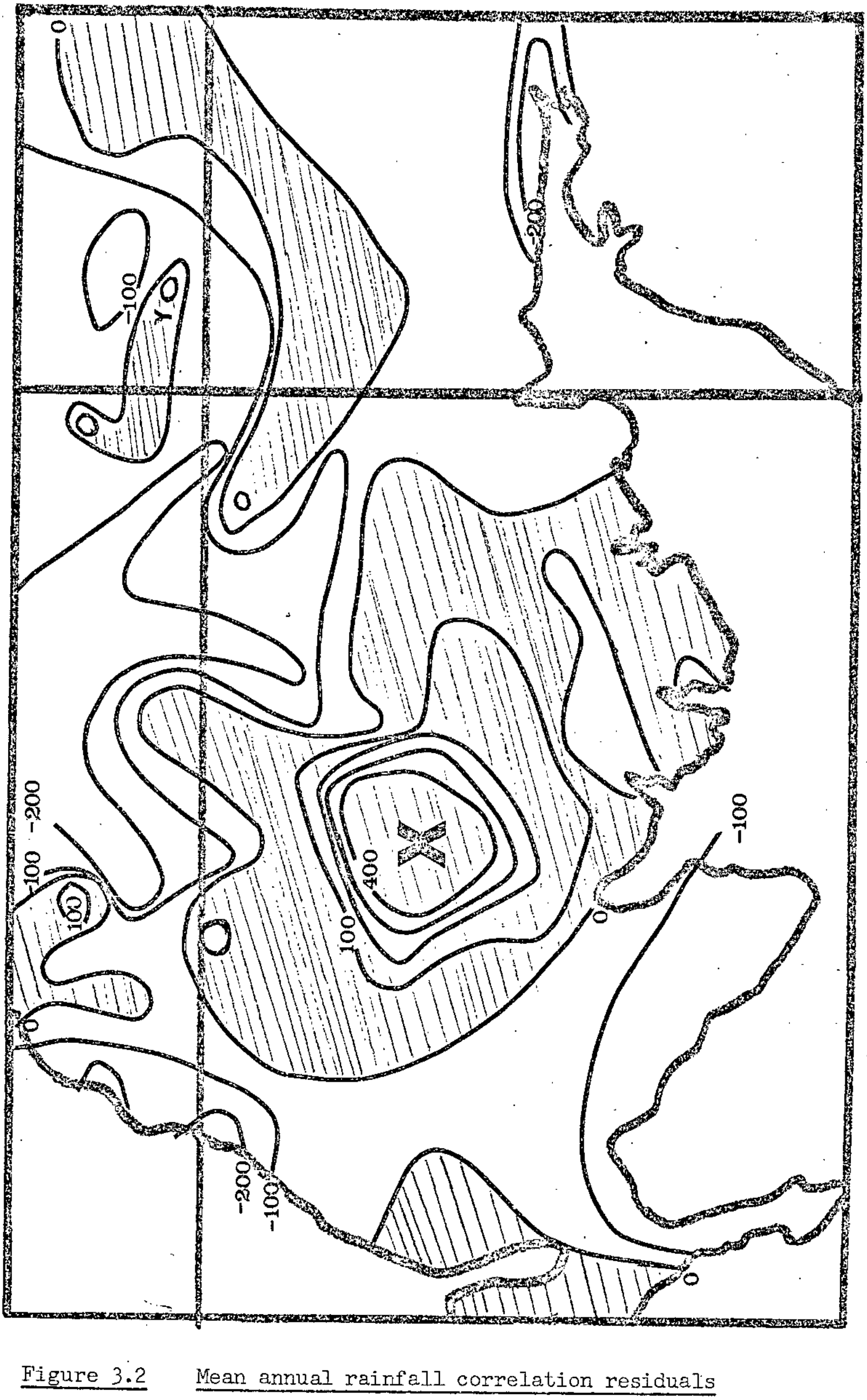

Figure 3.2 Mean annual rainfall correlation residuals

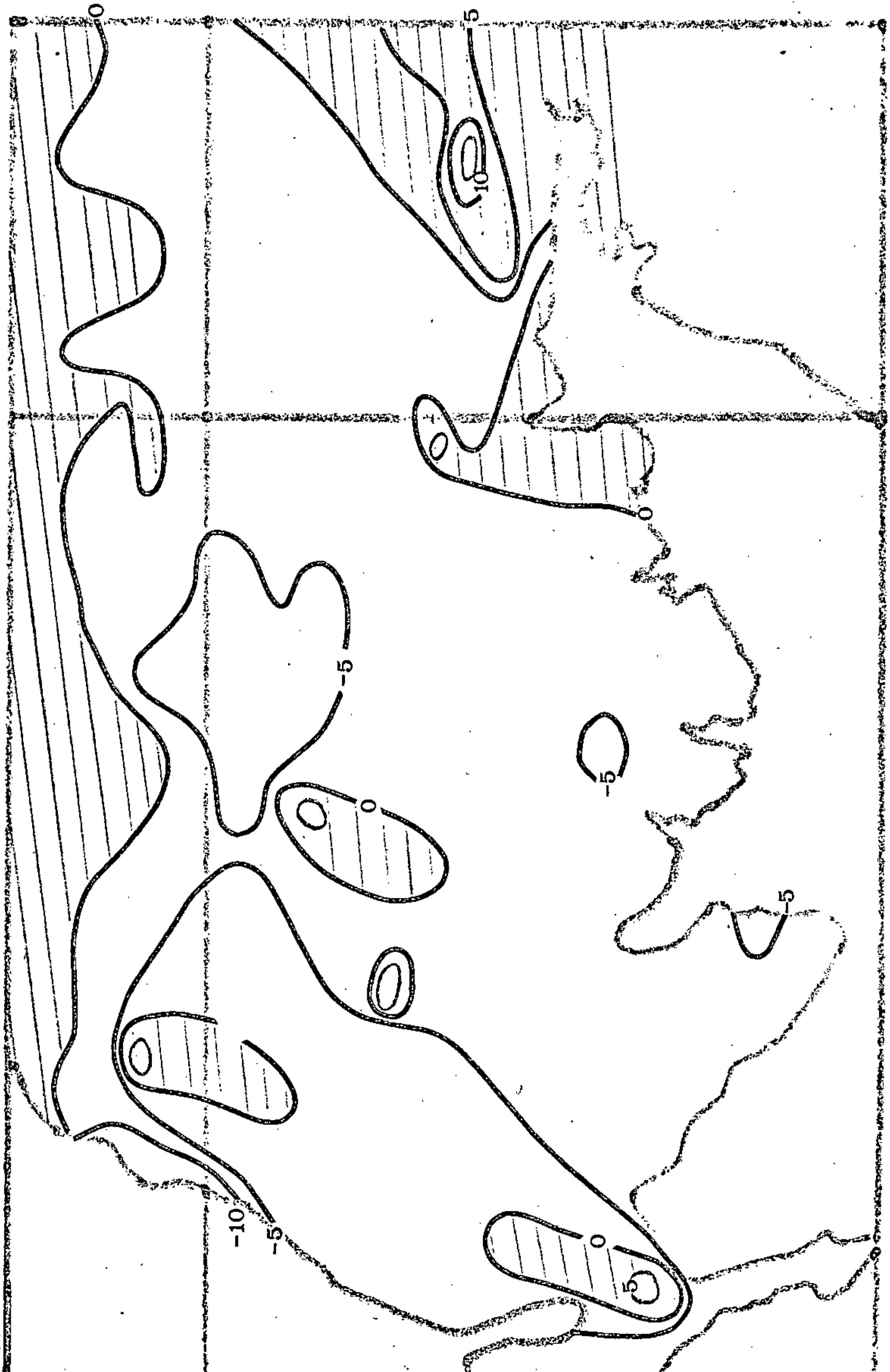

Figure 3.3 Correlation residuals, case A, 5th August 1971 (Thundery Low)

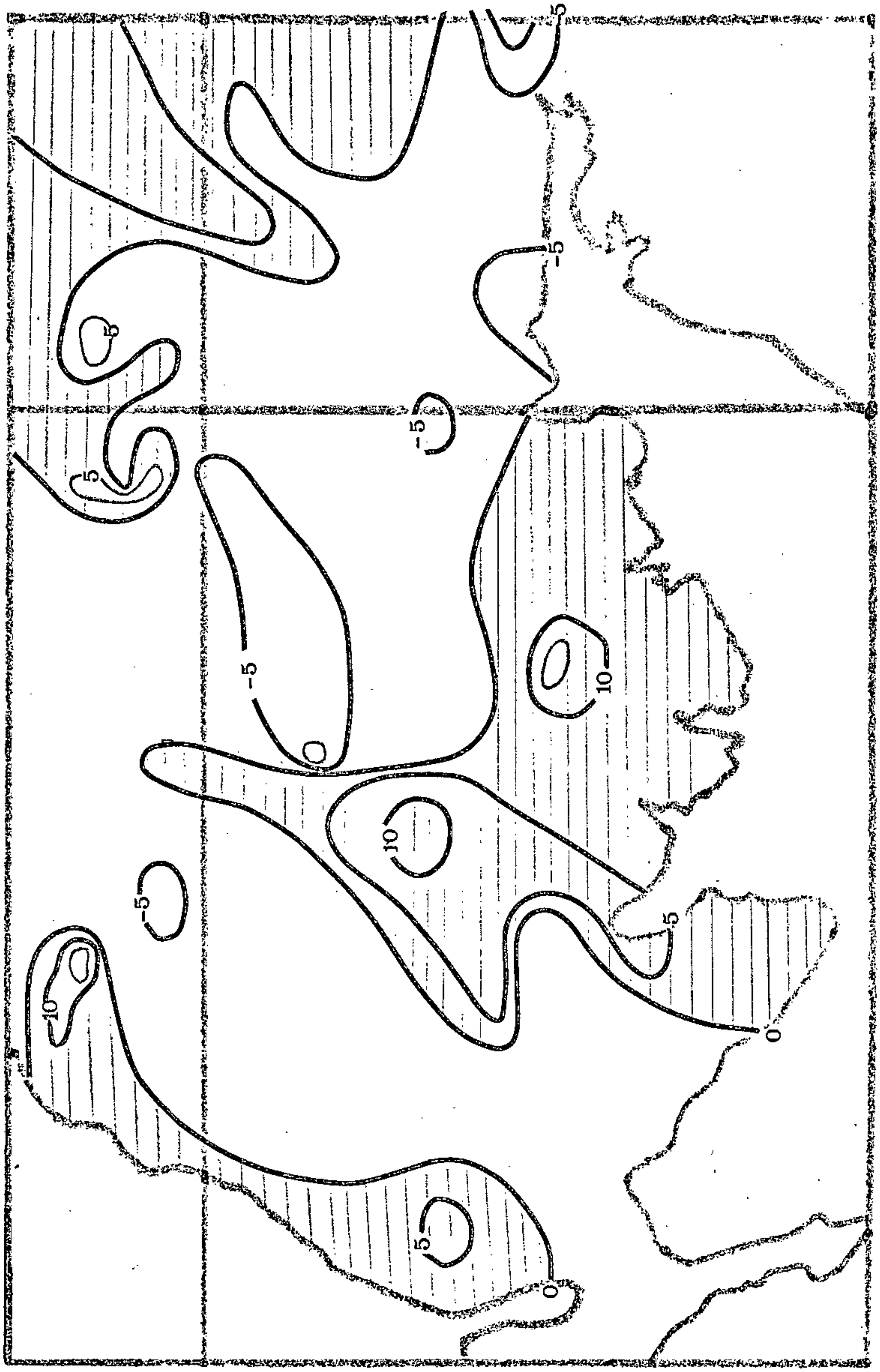

Figure 3.4 Correlation residuals, case B, 29th November 1972 (Cold Front)

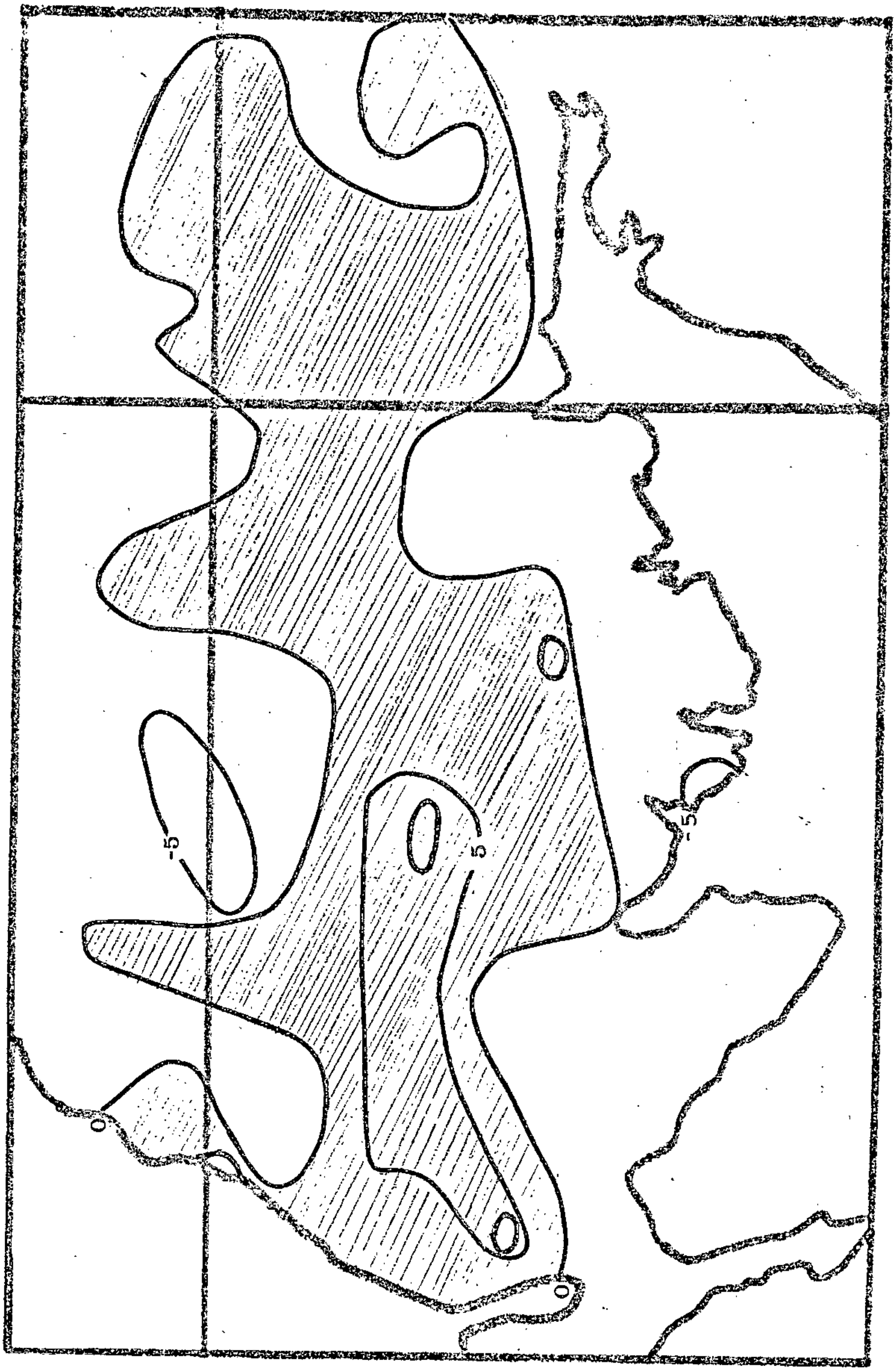

Figure 3.5 Correlation residuals, case C, 12th February, 1973 (Showery Low)

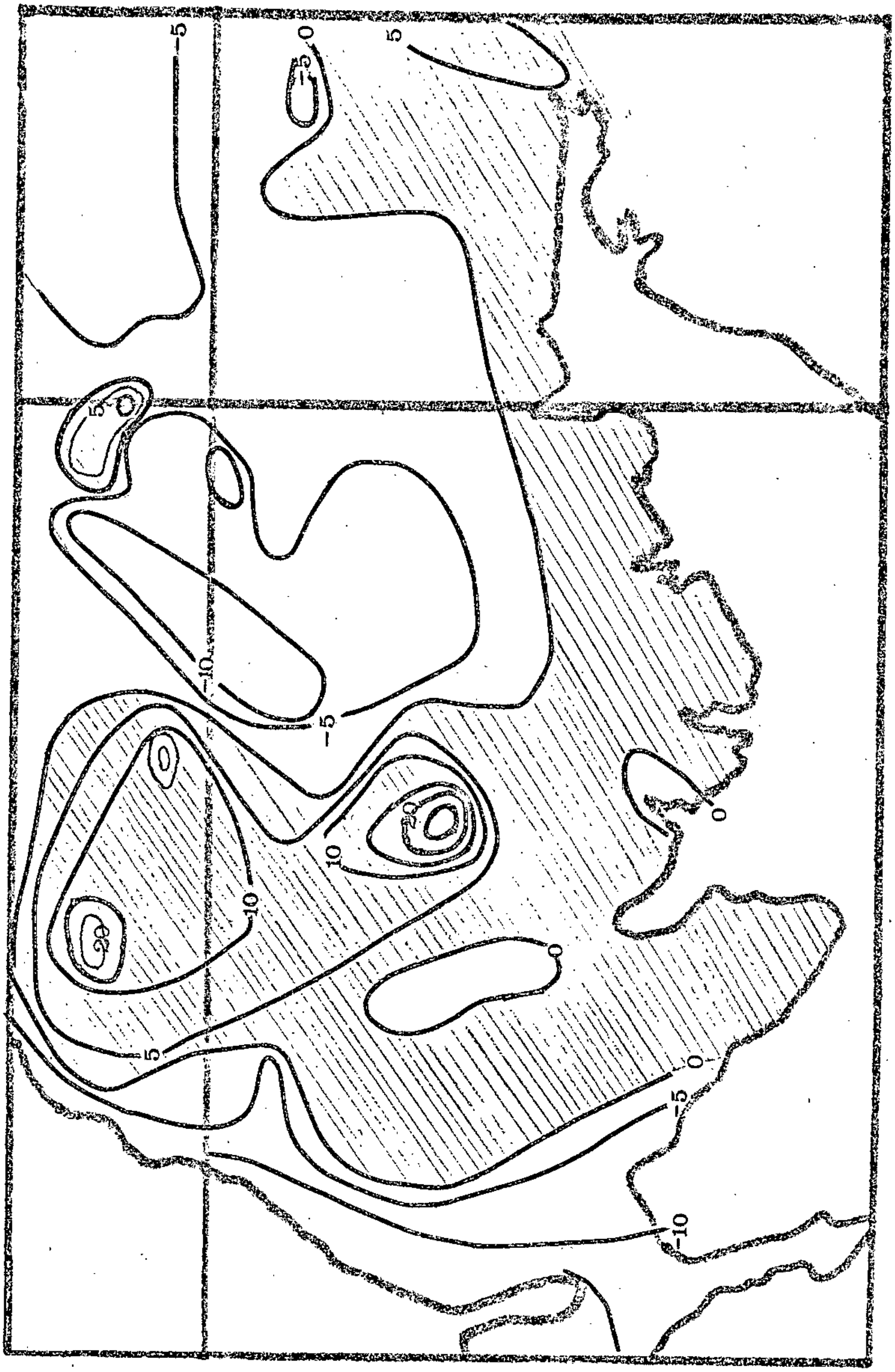

Figure 3.6 Correlation residuals, case F, 17th January 1974 (Warm Sector)

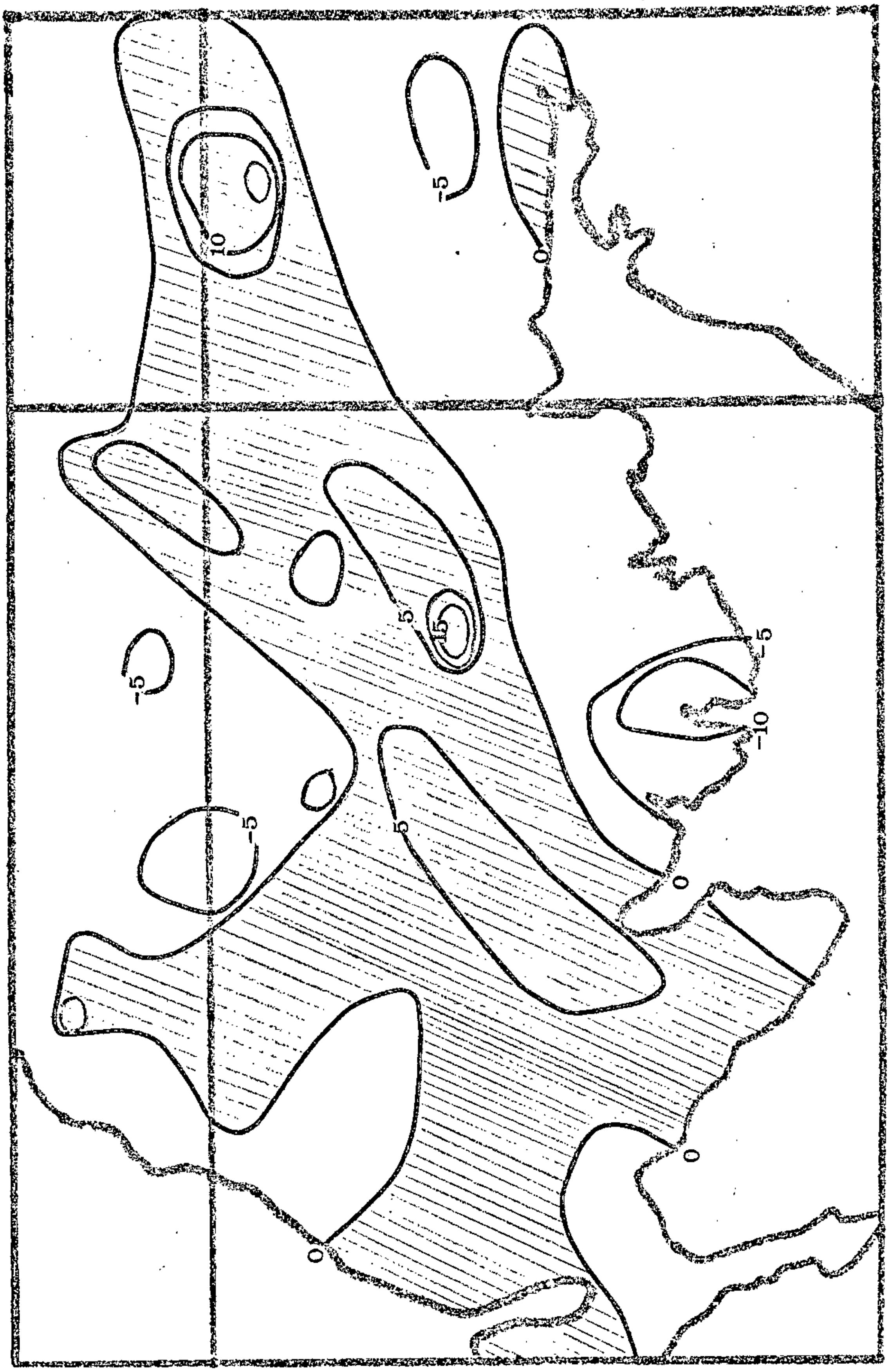

Figure 3.7 Correlation residuals, case G, 4th September 1974 (Occlusion)

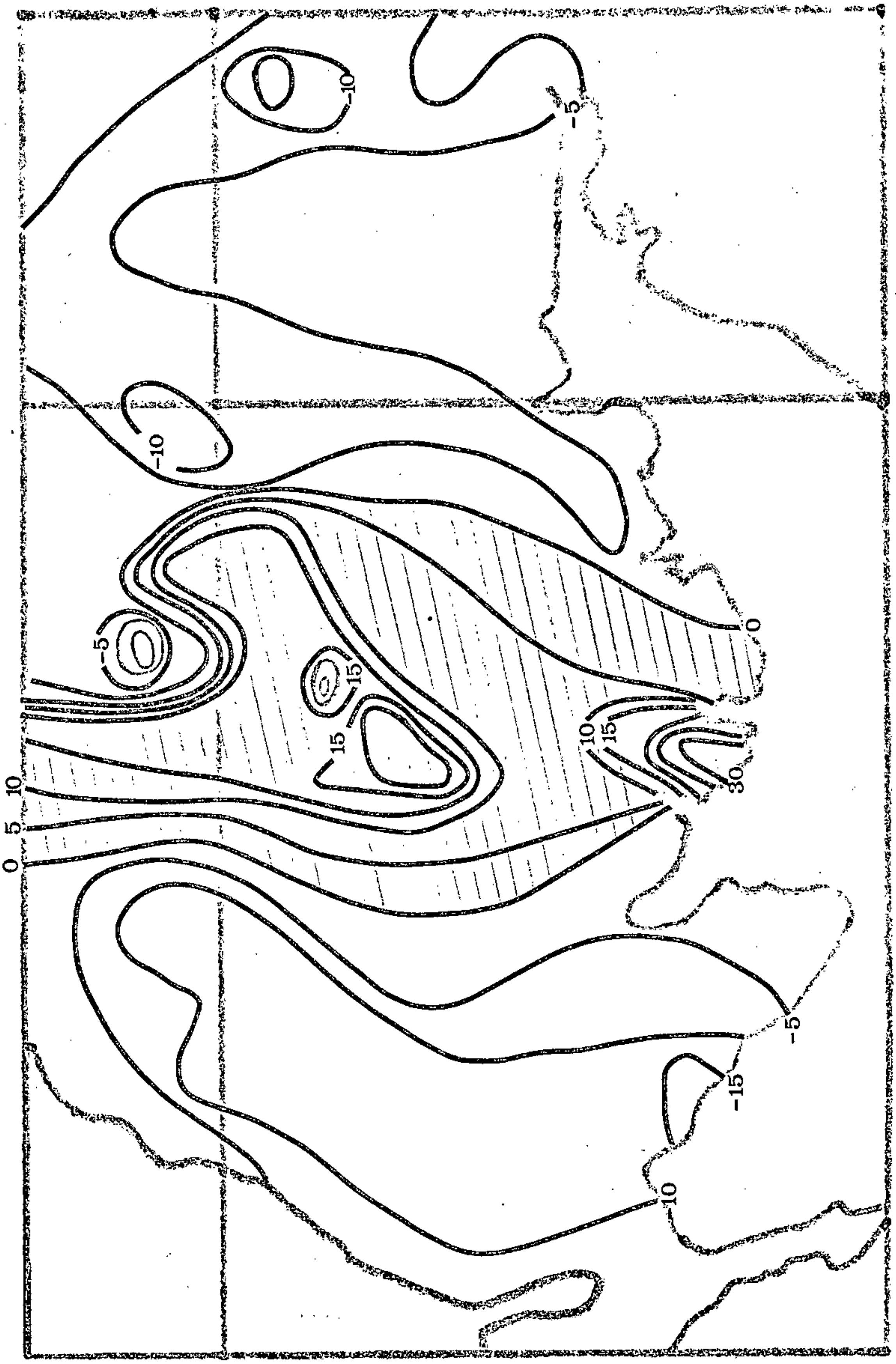

Figure 3.8 Correlation residuals, case H, 12th September 1974 (Cold Front)

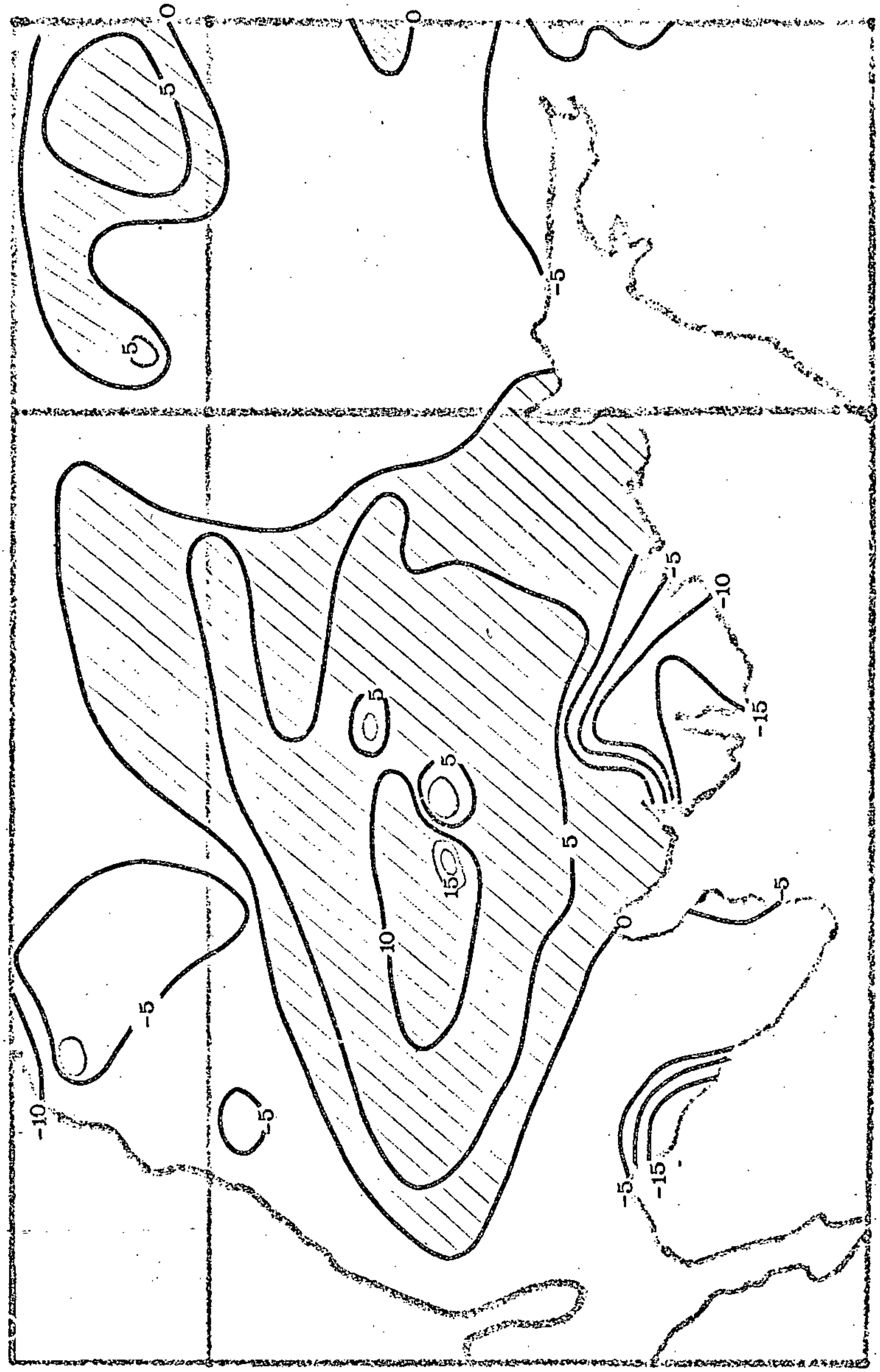

Figure 3.9 Correlation residuals, case K, 16th February 1975 (Occlusion)

did not significantly increase the variance described by the regression relationship.

An area of positive anomaly near area X also occurs in Figure 3.4 (case B), Figure 3.5 (case C), Figure 3.6 (case F), and Figure 3.9 (case K). The position of this anomaly and the shape of the isopleths is dependent on wind direction. There are also anomalies near area Y of Figure 3.2 in Figure 3.4 (case B), Figure 3.6 (case F), and Figure 3.9 (case K). Positive and negative anomalies occur close together as air flows over a small but exposed range of hills in the area Y. The shape of these isopleths is again determined by the direction and strength of the wind.

In case H, d_s, the most important physical parameter, measured the decrease of rainfall, predicted by the regression equation, from the effect of southerly winds. The regression residuals showed that most of the variations between gauges occurred along a central North-South band through the area X of Figure 3.2. Positive anomalies of varying size along this band showed that rainfall was under-predicted, apart from the negative anomalies in the North due to the rain-shadow effect of the area X hills in the path of the southerly winds. Over the rest of the region, there were negative anomalies with small variations between gauges in residuals, showing that rainfall was over-predicted.

Residuals from case A (Figure 3.2), C (Figure 3.4), and G(Figure 3.7) bear the least resemblance to those of mean annual rainfall. These cases were rare types of intense rainfall distribution - case A a thunderstorm, case C a showery low with an associated North-westerly airstream, and case G a cold front with a southerly wind. However, residuals of case F, which appeared by comparison of correlation coefficients of transformed and untransformed data to have an abnormal rainfall distribution, have similarities to those of mean annual rainfall. Common configurations of isopleths can be recognised in both diagrams with their positions shifted between diagrams.

While regression coefficients for individual cases are different, there are common residual isopleth patterns in mean annual rainfall and the daily cases, whose detailed shape and position for a particular case are dependent on wind velocity. The underlying topography producing the rainfall is of course the same in each case, and an approximate description of this is provided by the station heights. Some other features of topography which affect rainfall distribution may be seen by a comparison of regression residuals and the topographical map.

The residuals in themselves do not express much about the dynamics of the system producing the rainfall apart from the wind velocity. Residual patterns, Figures 3.3 to 3.9 cannot be recognised as describing rainfall of a particular type. The dynamical development of the systems under study, in cases B and K, as they crossed the region, cannot be gauged from the residual patterns.

3.6 Conclusions on regression studies

Daily rainfall values for systems of a "pure" type can thus be regressed with physical parameters, altitude usually being the most significant parameter in such analyses. Correlation coefficients are usually lower than those for regression of mean annual rainfall, which describes the most common variation between rainfall and topography which is present to a greater or lesser extent in most of the individual cases.

Only one physical parameter, usually altitude, seems to be of significance in multiple regression analysis. The distance of each station from the coast line where the "system originated", described above by d_w or d_s, is the alternative primary regression parameter. The use of an effective station height, assessed over a $3\frac{1}{3}$ km. grid, instead of station height does not in general significantly increase the values of the correlation coefficients.

The rates of increase of rainfall with altitude as revealed by regression coefficients varied between systems but the rate appeared to be highest for a winter-time occlusion. Winter-time systems, which are in general more intense than those of summer, had the higher correlation coefficients; developing systems had poorer correlations than well-defined systems.

The question raised above in point 3 as to the validity of an assumption of a linear variation between rainfall and altitude was answered by values of the correlation coefficients and the regression residual patterns. The exposure of mountain slopes to the flow of moist air would seem to explain the pattern of regression residuals when they are studied with reference to the topographical map and synoptic conditions. However, the regression analyses or residual patterns did not directly reveal the different effects of the various synoptic conditions apart from those due to wind velocity.

It is not entirely certain from the results of the analyses that daily rainfall values themselves describe rainfall of a pure type as correlation coefficients are smaller than for the cases of mean annual rainfall. Studies of rainfall records over short periods from a network of autographic gauges would enable the isolation of effects of a particular system and perhaps the decomposition of rainfall from different sections of the system. Different stages of development of the system as it effects rainfall could perhaps be distinguished.

It may however be concluded that regression methods are unsuited to analysis of rainfall variations inherent in daily values owing to the very small frequency of sufficiently pure types.

CHAPTER 4

4. MODELS OF THE DISTRIBUTION OF SPELLS OF WET AND DRY DAYS

4.1 Introduction

When time series are analysed in meteorology they are often tested to determine the extent to which their persistence may be described as Markov persistence, and how accurately members of the series can be determined from the Markov process. This process assumes that the probability of the occurrence of an event in a given time interval depends only on its occurrence in a previous equal time interval. In particular this model has been used extensively to describe the distribution of spells of wet and dry days (e.g. Chatfield 1966, Gabriel and Neumann 1962). In this context the model implies that the probability of any particular day being wet or dry depends only on the character of the previous day. A full discussion of the application of this model and three other models - William's (1952) log model, Green's (1970) modified log model, and Yap's (1973) modified geometric model - is given in a previous paper (Blair-Fish, 1975 - see Appendix 1).

The advantage of the Markov model over other models is that only one parameter - the probability of a dry day following a dry day (and that of a wet day following a wet day) - is needed to predict the nature of a given day from that of the previous day. The simple log model also uses only one basic parameter, but, in this case, the probability is weighted by a factor determined by the number of previous days of a given type.

Long wet and dry spells are usually the least well predicted by the models. Such spells occur infrequently and are given little weight when model parameters are calculated. For most practical purposes it is long dry spells (and sometimes wet spells) which are of most interest. In order to emphasise the importance of long spells further investigations have been carried out using the same data as in the previous paper. There (as in most work on the subject) models were used primarily to describe spells of specific length (i.e. the probability of a spell lasting an exact number of days) rather than 'cumulative spells' (the probability of spells of at least a given length). When spells of a given length and greater are summed to give cumulative spells, some smoothing of data occurs and it is usually easier to fit models to this type of spell; cumulative spells are probably also the more useful for planning purposes. For the Markov model it will be shown that the change in spell description affects only the normalisation of parameters.

In the further investigations of spells below, the Markov, log, and modified geometric models are fitted to spells of length greater than five days. The "Jenkinson probability" and "natural persistence" models used by Lawrence (1957) are also discussed and some of their uses and limitations are demonstrated.

4.2 Discussion of the models

4.2.1 The Markov model

The probabilities of spells of length 1, 2, 3 .... r days form a geometric series $q, q^2, q^3, \ldots q^r$ and these probabilities are normalised so that the total probability of a spell of any length is unity. The normalisation constant is then $\frac{1-q}{q}$ and the number of spells of length r is $N\,q^r\,(\frac{1-q}{q})$ where N is the total number of spells. The number of spells of length r or greater (cumulative spells) is $\frac{1-q}{q} N \sum_r q^x$ i.e. $N_q{}^r$.

This model may also be applied only to spells greater than a minimum length in order to emphasise longer, less frequent spells and because Markov persistence may not be applicable to shorter spells. If only spells greater than x days are considered, and N_x is the number of spells of length at least x days, the number of spells of length r and at least length r will be $N_x(1-q)q^{r-x-1}$ and $N_x q^{r-x-1}$ respectively.

To fit the model the mean spell length (or the mean continuation of a spell beyond a minimum number of days) is calculated i.e. the ratio of the total number of days of given type T.D to the total number of spells N_x. In terms of model parameters:

$$TD = N_x \left(\frac{1-q}{q}\right)\sum q^r\, r = \frac{N_x}{1-q} \qquad (4\cdot1)$$

The mean spell length is thus $\frac{1}{1-q}$.

If any spell data can be described by the Markov model, a plot of the number of spells of given length r, n_r (or spells of length at least r, N_r) against r will produce a straight line on semi-log paper as can be seen in equation 4.2.

$$\log n_r = \log q^{r-1}(1-q)N_x = (r-1)\log q + \log N_x + \log(1-q) \qquad (4.2a)$$

$$\log N_r = \log q^r N_x = (r-1) \log q + \log N_x \qquad (4.2b)$$

The intercept ofonthe line the x-axis will be greater in the case of 4.2b, there being more spells of at least a given length than of an exact length. If a Markov model is constructed for spells greater than x days, where x = 3 in Lawrence's work and x = 5 in the work below, then the straight line will be drawn for r = x onwards.

4.2.2 The log model

In this model the probabilities of spells lasting exactly 1,2,3 .... r days are proportional to $q, \frac{q^2}{2}, \frac{q^3}{3} \ldots \frac{q^r}{r}$ and the number of spells of length r is $N \left(\frac{-1}{\log(1-q)}\right) \frac{q^r}{r}$ where N is again the total number of spells. The number of spells lasting at least r days is given by:

$$\frac{-N}{\log(1-q)} \sum_{r}^{\infty} \frac{q^r}{r} = \frac{N}{\log(1-q)} \left(\sum_{1}^{r-1} \frac{q^r}{r} - 1\right) \qquad (4.3)$$

The total number of days of a given type is given by:

$$\frac{-N}{\log(1-q)} \sum_{1}^{\infty} \frac{q^r}{r} \times r = \frac{-N}{\log(1-q)} \frac{q}{1-q} \qquad (4.4)$$

The mean spell length (as defined above) is again used to fit the model to spell data. q is found from this mean length, $\frac{-1}{\log(1-q)} \frac{q}{1-q}$, by a recursive process or by a graphical method. Results are published in tables prepared by Williamson and Bretherton (1964).

The ratio of the number of spells lasting r + 1 to r days, F_r, is given by $\frac{r}{r+1}q$. This ratio increases as r increases; for the Markov model it has the constant value q. Thus the persistence of a spell increases with spell length if the distribution is described by the log model, while it is constant for the Markov model. The log model cannot be simply applied to data which does not include the more frequent short spells, as the probability of any spell of length r is defined for all values of r according to a predetermined pattern. Also as the persistence factor F(r) depends strongly on r, the model is only applicable to data for spells of specific length and not to those of cumulative spell data.

Green proposed a modification to the log model such that the probabilities of spells of length 1,2,3 ... r days were proportional to $\frac{q}{1+a}, \frac{q^2}{2+a}, \frac{q^3}{3+a} \ldots \frac{q^r}{r+a}$ where a is a modifying parameter ($0 < a < \infty$). For the simple log model a = 0 and for the Markov model $a = \infty$. This model attempts to explain more fully the variation in persistence in spells of different length. F_r increases with spell length more slowly than for the log model and equals $\frac{r+a}{r+a+1}$. While Green successfully fitted this model to most spell data, the model was found in the previous work (Blair-Fish 1975) to be more relevant to wet spells than to dry spells which are well described by the simple log model.

The modified log model is difficult to fit to data, requiring successive adjustments to values of q and a after an initial guess has been made to their values from the shape of the distribution (which may be initially plotted on semi-log paper). If the number of steps required for the computation is large, a correspondingly large amount of computer processing time is required. The physical significance of the parameter a is not obvious, also, differences in a between stations found in cases where the model has been applied do not follow a clear pattern. For these reasons, and in particular because of the success of simpler models, the modified log model has not been used in this further work.

4.2.3 Yap's modified geometric model

Here the probability, p, of a spell lasting a further day is assumed constant within a spell of given length but to vary with spell length. Two parameters, a and b, are used to determine these probabilities such that p is assumed to be a beta variate. In Yap's paper and in Appendix 1, it is shown that the probability of a spell lasting another day is $\frac{b}{a+b}$ and F(r), the ratio of the probability of spell length r+1 to that of spell length r, is $\frac{a+r-3}{a+b+r-1}$. F(r), the measure of persistence, increases with r and tends to 1.

The model is fitted, using the mean spell length (as defined above) and the mean square spell length. The latter parameter is the sum of the number of days of length r times r^2 divided by the total number of spells. The factorial moments for the distribution, u_1^1, the mean spell length, and u_2^1, the difference between mean square length and mean spell length, are related to a and b as in equations 4.5 and 4.6.

$$b = \frac{2u_1^1 \ (u_1^1 - 1) - 2u_2^1}{2u_1^1 \ (u_1^1 - 1) \ u_2^1} \qquad (4.5)$$

$$a = (u_1^1 - 1)\ (b - 1) \qquad (4.6)$$

The modified geometric model may be applied to data in which shorter spells are omitted. In this case the first few days, (the first four in cases below), of each spell are ignored and the probability of a spell continuing is assumed constant from the next (fifth) day onwards.

4.2.4 Lawrence's "natural persistence" model

In this model, the actual probabilities of spells continuing at least another day are considered. The ratio of the probability of a spell lasting at least r days to one lasting at least r+1 days is calculated for $r > 3$, say, for each station. Then if N_3 is the total number of spells lasting at least 3 days, the number of spells lasting at least 3, 4, 5, 6 ... days can be expressed as the series: N_3, $N_3\ C_4$, $N_3\ C_4\ C_5$, $N_3\ C_4\ C_5\ C_6$ ... Lawrence found that when C_4, C_5, C_6, .. were plotted for each station, the variations of C_r with r were similar for different stations; he was thus able to generalise variations in persistence as measured by C_r.

For a dense network of stations, a series of charts of N_3, C_4, C_5, C_6 ... may be plotted and these parameters may be interpolated to new stations. If the variations of C_r with r are similar at all stations in an area, values of C_4, C_5, C_6 ... may be averaged and the number of spells lasting at least 3, 4, 5, 6, ... days may be expressed as: N_3, $N_3(\overline{C_4} + x)$, $\overline{N_3}(\overline{C_4} + x)(\overline{C_5} + x)$, $N_3(\overline{C_4} + x)(\overline{C_5} + x)(\overline{C_6} + x)$ ... where $\overline{C_4}$, $\overline{C_5}$, and $\overline{C_6}$ are the mean areal values of C_4, C_5 and C_6. It may also be possible to replace N_3 by $\overline{N_3}$, the mean areal value of N_3 and hence express the series of probabilities as: $\overline{N_3}$, $\overline{N_3}(\overline{C_4} + x)$, $\overline{N_3}(\overline{C_4} + x)(\overline{C_5} + x)$, $\overline{N_3}(\overline{C_4} + x)(\overline{C_5} + x)(\overline{C_6} + x)$. x, the parameter modifying the mean value of C_r to that of a particular station, is determined by the relative number of days occurring at each station of the type considered. If C_1 is the ratio of the number of dry days to the total number of days considered for a station, and $\overline{C_1}$ is the mean value of this ratio, x is the difference between C_1 and $\overline{C_1}$.

4.2.5 The "Jenkinson probability" model

In this model the maximum spell length occurring in each year is extracted from the data. The mean μ, the standard deviation σ_1, and the two-year standard deviation σ_2 of the maximum annual spell length are calculated. To find the latter parameter, the maximum run lengths are ranked in order from smallest to largest. The m th member of the series is given the weight 2m + 1 and the standard deviation is then found, with the m th member assigned the frequency 2m + 1.

The parameter $\frac{\sigma_1}{\sigma_2}$ is then calculated and a parameter $R(\frac{\sigma_1}{\sigma_2}, y)$ is used to determine the return period of a spell lasting at least D days. y is the probability function $-\log_e \log_e \frac{1}{p}$ where p is the probability. The return period, t, for annual maximum spell length increases with y, the relationship being

$$\frac{1}{t} = 1 - \exp(-y) \qquad (4.7)$$

The curve of y against D is given by

$$D = \mu + R\left(\frac{\sigma_1}{\sigma_2}, y\right)\sigma_1$$

Tables of R are given in Jenkinson (1955) and Lawrence (1957), the latter of whom applied the method to describe spells for individual months of the year. The curve is approximately a straight line: for $\frac{\sigma_1}{\sigma_2} = 1$ it is an exact straight line, while for $\frac{\sigma_1}{\sigma_2} < (>)\ 1$ the curve is concave downwards (upwards) and D has a lower (upper) bound. In principle, D has no upper limit (but has a lower limit as a spell must last at least one day). It is thus expected that $\frac{\sigma_1}{\sigma_2} < 1$, as is usually the case.

The mean annual frequency of runs of length D or more, G_D, can be determined from the fact that $G_D \sim \exp(-y)$. From the y against D curves, annnual frequencies may be obtained; frequencies for periods of l years are obtained by multiplying G_D by l.

The Jenkinson probability model is most suited to calculating the frequency of spells of length around the mean annual spell length as it is this length which is used to determine the distribution.

4.3 Application of the models to eight stations

For many practical purposes, e.g. farming and water resource management, the incidence of long dry spells is of special interest.

In the previous work (Blair-Fish, 1975) both dry and wet spells were considered. The Markov and modified geometric models are considered below for spells of length give days and greater for the original eight stations of Edinburgh, York, Whitby, Cum Dyli, Oxford, Falmouth, March and Edgbaston. The use of the Jenkinson probability model in relation to long dry spells is also considered.

In order to provide more reliable estimates of long (infrequent) spells the data for each of these stations were not subdivided into months. In comparison, Lawrence (1957) considered spell frequencies for individual months; while his method provided little information on long dry spells for individual months and stations, the dense network of stations with similar climates yielded several estimates (not however, strictly independent) of the monthly distribution of spells.

4.4 The use of the "natural persistence" model and of persistence patterns

The relatively sparse network of stations used in this study does not permit extensive interpolation of parameters between the eight stations. No clear variations in model parameters emerge in the various analyses. Lawrence was able to chart the model parameters and interpolate between stations in a dense network of stations in Southern and Eastern England. These interpolations included values of C_r in the "natural persistence" model.

The values of C_r calculated for each of the eight stations are displayed in Figure 4.1 for $r \geq 6$ together with the values of N_5, the number of spells lasting at least five days. The grouping

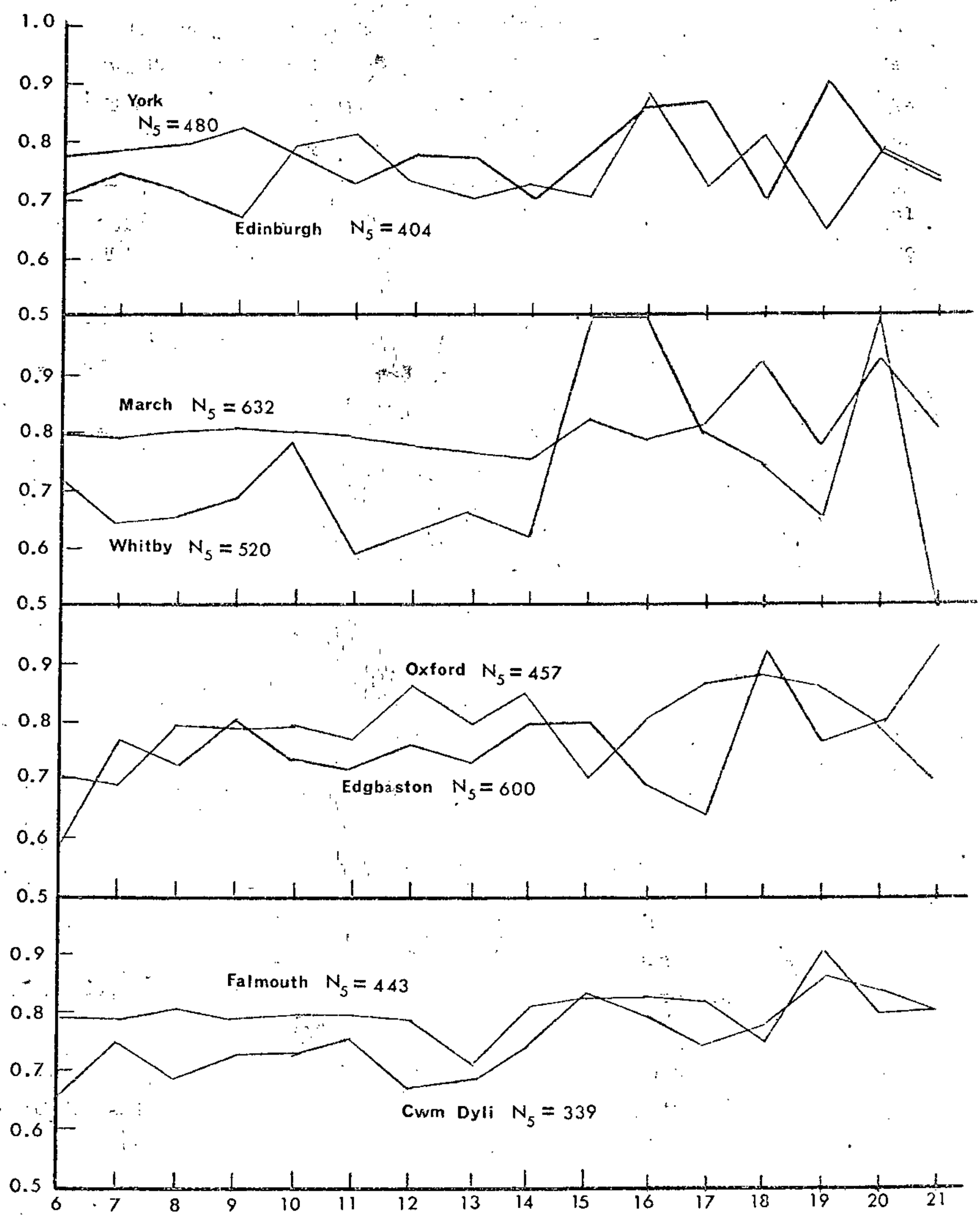

Figure 4.1 Natural persistence patterns (C_r v r)

into four sets of two stations in the figure is made from the similarities in climate of, and the geographical positions of, the stations. The two "wet" stations in the West of the country, Cum Dyli and Falmouth, show the greatest similarities.

Lawrence summarised the variations of C_r with r for his network of stations. Persistence was defined as an increase of C_r with r. A more obvious definition would be to define an increase in C_r with r as an increase in persistence, i.e. the "cumulative" persistence factor for r days is defined as the probability of a spell lasting at least another day beyond r days. (This is the definition used herein). C_r is approximately constant for eight stations from 6 to 16 days. For spells of length greater than 16 days oscillations of C_r about its mean value increase in amplitude, partly as a result of the paucity of such spells. This is especially noticeable in the case of Whitby where only 22 years of continuous data were available.

4.5 Results - the Markov model and modified geometric models

The persistence factor, C_r, is constant for the Markov model. The Markov model when applied to dry spells of length five days or greater, was found to produce a significant fit to seven out of the eight sets of data (see Table 4.1). The modified geometric model also fitted these data, with a significantly better fit than the Markov model in most cases. Results for the log model applied to the complete distribution of spell lengths also appear in Table 4.1. These latter fits are in general less good than those for either of the former models.

Table 4.1 Summary of model parameters (dry spell data)

STATION	ALL DATA LOG MODEL		SPELLS AT LEAST 5 DAYS LENGTH					JENKINSON PROBABILITY MODEL			
			MARKOV MODEL		MODIFIED GEOMETRIC MODEL						
	q	$P(\chi^2)$	q	$P(\chi^2)$	a	b	$P(\chi^2)$	μ	σ_1	$\frac{\sigma_1}{\sigma_2}$	$P(\chi^2)$ (spell length $>\mu$)
EDINBURGH	0.82	0.40	0.73	0.40	22.9	9.3	0.40	13.12	4.26	0.96	0.001
YORK	0.86	0.30	0.78	0.10	37.3	11.5	0.60	17.85	7.01	0.84	0.001
WHITBY	0.80	0.30	0.68	0.60	59.7	28.4	0.40	11.86	4.45	0.96	0.50
CWM DYLI	0.83	0.20	0.73	0.80	271.9	100.5	0.70	13.72	4.62	0.94	0.20
OXFORD	0.86	0.05	0.77	0.04	19.4	6.7	0.20	18.95	7.40	0.96	0.60
FALMOUTH	0.87	0.30	0.80	0.95	141.4	37.2	0.95	17.80	6.27	1.03	0.10
MARCH	(0.87) (modified) (a=0.337)	0.40	0.79	0.50	89.5	23.3	0.95	20.75	7.09	0.90	0.001
EDGBASTON	0.85	0.02	0.75	0.50	44.4	15.6	0.20	15.83	6.23	0.97	0.50

For none of the above models were the fits to the tail-end of the distribution tested in detail. In applying the χ^2 test spells whose expected frequency were less than five, were pooled into categories before the test was applied. The expected frequency of long spells of a given range of length were then compared with observed values, while for shorter spells frequencies of spells of given length were tested.

In applying the χ^2 test, the modified geometric model had one fewer degrees of freedom than the Markov or log model as the mean square spell length was used as an additional variable to define the model parameters a and b of the former distribution.

4.6 Results - an alternative application of the Markov and modified geometric models

An alternative approach to applying the Markov and modified geometric models is to sum the number of spells lasting at least 5, 6 or 7 days and to use these values as direct input data to determine model parameters (i.e. to consider spells of cumulative length in place of spells of exact length). The frequency of long spells obviously falls off less rapidly for this distribution than for that of spells of specific length, though some grouping of spells of different lengths is still needed in order to apply the χ^2 test to the tail-end of the distribution. Such grouping amounts to double integration of spell frequency against spell length and is physically not very meaningful. Also it is difficult, in the case of the modified geometric model, to sum the number of spells at the tail-end of the distribution when this number decreases slowly with spell length and testing of the fit at this end of the

Table 4.2 Model parameters using the alternative approach (cumulative spells)

	MARKOV		MODIFIED GEOMETRIC		
	q_1	$P(\ ^2)$	a_1	b_1	$P(\chi^2)$
EDINBURGH	0.73	0.40	201.2	73.4	0.99
YORK	0.80	0.001	46.9	12.1	0.90
WHITBY	0.69	0.70	35.2	16.6	0.90
CWM DYLI	0.74	0.40	36.6	13.6	0.99
OXFORD	0.81	0.001	43.9	11.5	0.70
FALMOUTH	0.80	0.99	145.2	37.1	0.99
MARCH	0.81	0.80	106.7	26.4	0.99
EDGBASTON	0.76	0.05	27.9	9.9	0.80

It has been pointed out that application of the χ^2 text in this instance is invalid owing to the lack of independence implicit in cumulative data.

distribution is therefore difficult. (For the Markov Model the tail-end summation amounts to summing a geometric series.)

In processing the new form of input data, running the same programs to fit the models as previously, mean cumulative spell length beyond five days, and mean square cumulative spell length beyond five days are implied variables which determine model parameters. The first (second) parameter is the ratio of the sum of the number of dry days occurring beyond a wet days times $r(r^2)$ divided by the total number of dry days.

In terms of the new Markov probability, q_1, and N_5, the number of spells lasting at least five days, the number of spells lasting at least r days is $N_5(1-q_1)q_1^{\ r-1}$. Using the original approach this number was $N_5 q^{r-1}$. Thus values of q_1 will be larger than values of q (compare Tables 4.1 and 4.2).

The fits of the Markov model for this approach are as good as for the original approach, as might be expected since the new approach only amounts to a different normalisation. The fits using the modified geometric model to the cumulative data are very good. Indeed, the values of χ^2 are such as to give rise to doubt as to the validity of the approach.. At any rate, the original model was formulated for a random variate p with constant value within a given spell determined by the spell length. The distinction between different runs (spells) and the meaning of p becomes confused in this approach.

4.7 Results - the "Jenkinson probability" model

Values of μ, σ_1, and $\frac{\sigma_1}{\sigma_2}$ are given in Table 4.1. The ratio

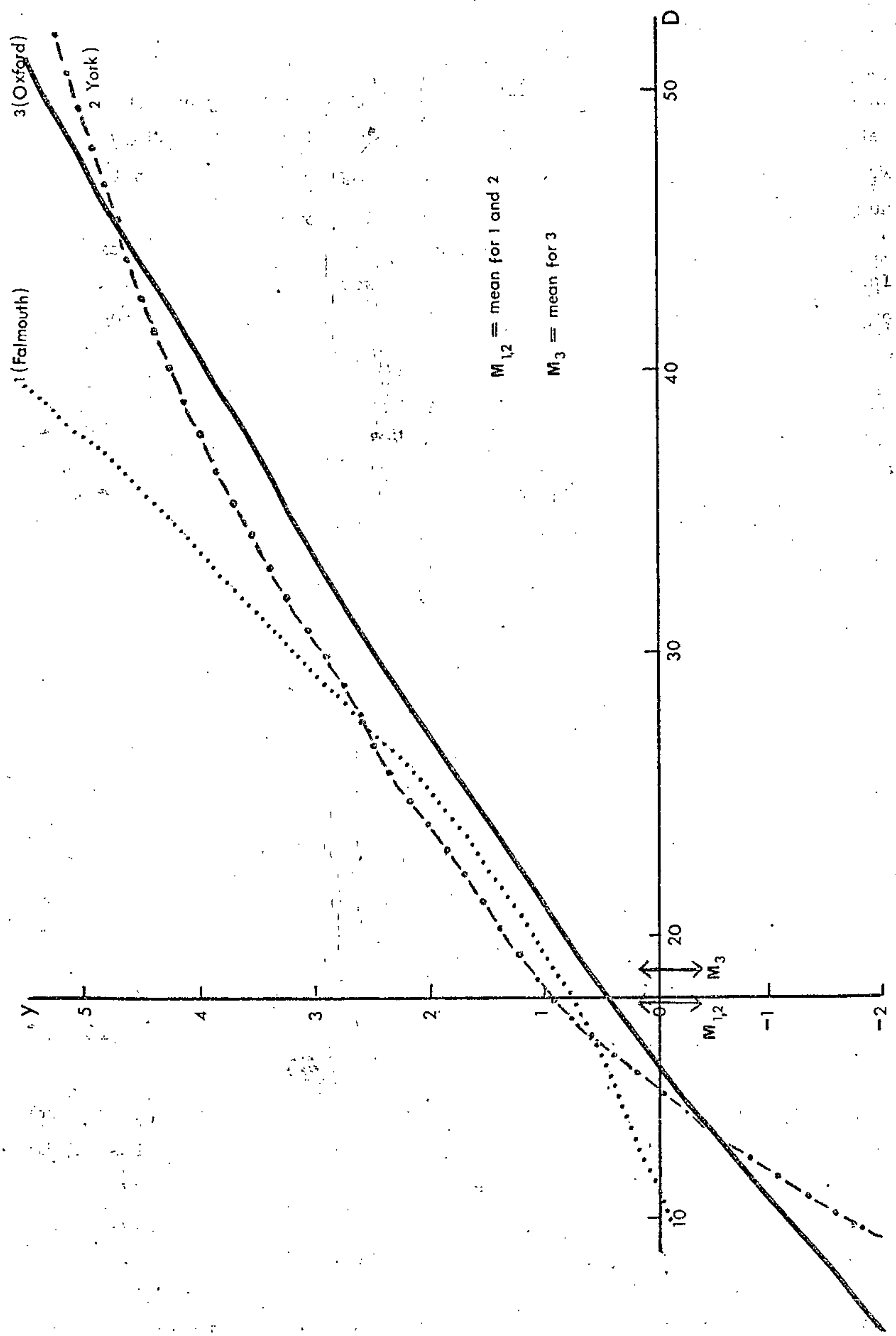

Figure 4.2 Jenkinson Probability Curves (y v D) for York, Oxford, and Falmouth dry spells

$\frac{\sigma_1}{\sigma_2}$ is seen to be < 1 in seven out of eight cases; this implies that of a plot of y against D is concave downwards (gradient positive) with a lower bound to D. This in turn means that the persistence of a spell increases in length over the range of D considered (spells greater in length than the mean annual length. In the Lawrence terminology persistence is positive.) Curves of y against D for York, Oxford and Falmouth are given in Figure 4.2; these are based on values of μ, σ_1, and $\frac{\sigma_1}{\sigma_2}$ in Table 4.2 and $R(\frac{\sigma_1}{\sigma_2}, y)$ in Lawrence.

The values of the expected frequencies of spells of length D or greater, $G_D \times l$, where $l = 40$ for forty years of data, were computed and compared with the observed frequencies using a χ^2 test. Some pooling of data was necessary for long spells (though cumulative frequencies are greater than those of spells of specific lengths). Results of the χ^2 test appear in Table 4.1. From estimates of frequencies made from the models, and with the number of rarer spells pooled into categories, the number of spells estimated or observed in each category was of the same order of magnitude and hence was given the same weight in the test. In the light of these considerations, this model was considered to produce good estimates of frequencies of long spells for five out of eight cases considered.

4.8 Tests of the models on new data

1975 was notable for its exceptionally dry summer in various areas of the British Isles. Rainfall spell data, extracted from records of the University of Edinburgh, Meteorology Department's

Table 4.3 Tests of model parameters on new data

	EDINBURGH 1974-1975					GREENWICH 1921-22		
Spell Length	Observed frequency	Cumulative frequency	Markov cumulative frequency	Modified geometric cum. freq.	Jenkinson probability frequency	Observed frequency	Cumulative frequency	Markov cumulative frequency $q = 0.8$
1	20	55				48	112	
2	13	35				17	64	
3	6	22				13	47	
4	6	16				8	34	
5	4	10	10.1	9.8		8	26	26.0
6	2	6	7.1	7.2		1	18	20.8
7	0	6	5.3	5.2		1	17	16.6
8	0	6	3.9	3.8		3	16	13.3
9	1	4	2.8	2.8		2	13	10.6
10	1	3	2.1	2.1	1.4	5	11	8.5
11	0	2	1.5	1.5	1.0	0	6	6.8
12	0	2	1.1	1.1	0.7	0	6	5.4
13	0	2	0.8	0.8	0.5	0	6	4.4
14	0	2	0.6	0.6	0.4	0	6	3.5
15	0	2	0.4	0.4	0.2	1	6	2.8
16	0	2	0.3	0.3	0.2	2	5	2.2
17	0	2	0.2	0.2	0.1	1	3	1.7
18	1	2	0.2	0.2	0.1	1	2	1.4
19	1	1	0.1	0.1	0.1	0	1	1.1
20	0	0	0.1	0.1	0.1	0	1	0.9 $P(\chi^2) = 0.30$

gauge for the period 20th September 1974 to 19th September 1975, were compared with "predictions" of dry spell frequencies made from the 1931-70 Blackford Hill Observatory gauge, using the Markov and Jenkinson models, as shown in Table 4.3. Results obtained for cumulative data from the Markov model are seen to provide acceptable predictions of spells of length 5 to 11 days. The spell of length 19 days is only predicted to occur once in 7.7 years using the Markov model and once in 10.4 years using the Jenkinson model. The latter does not appear to give better estimates of longer spells than the former in this case.

Two successive years of data, 1921-22, for Greenwich were also analysed and the actual and cumulative distributions of observed dry spells there are also given in Table 4.3. The cumulative spell frequencies for length 5, 6, 7 ... days were compared with those based on the Markov parameter, q_1, interpolated from values at neighbouring stations. A value of $q_1 = 0.8$, interpolated from Edgbaston, March and Oxford, and the observed value of 26 occurrences of spell length five days or greater were used to "predict" the occurrence of spells of length at least 6, 7 ... days. A value of $P(\chi^2)$ of 0.30 was obtained when predictions were compared with the observed distribution. This showed that the probability of a spell lasting a further day can be predicted from a Markov parameter interpolated from surrounding stations.

4.9 Conclusions

The Markov model has been shown to be of considerable value in predicting frequencies of dry spells of length five days or greater. Successful application of the model to subsections of a distribution have been shown to emphasise its value. Its simplicity and ease of application, using basically one parameter, ensure that it will continue to be used in the analysis of spell data and in many other analyses of time series.

The Jenkinson model has been shown to be of some value in relation to rare long spells though its complexities may discourage its application. The modified geometric model, log model, and modified log model can also be reasonably fitted to spell data but are more difficult to apply than the Markov model - the modified log model particularly so. The physical significance of the latter model is also not clear.

Considerable similarities were observed in the statistics of the two wet stations - Cwm Dyli and Falmouth; to a lesser extent also between those of Edinburgh and York, of Whitby and March, and of Edgbaston and Oxford. The recognition of similar spell distribution patterns at different stations, whatever their separation, may be of use in types of climatic analysis other than that of spells.

However, it should be pointed out that none of the models, of whatever complexity, adequately describes the incidence of very long spells, though the Markov and Jenkinson models may provide some guidelines concerning their occurrence.

CHAPTER 5

5. TIME SERIES OF SCOTTISH RAINFALL AND BRITISH ISLES CIRCULATION INDICES

5.1 Introduction

In the previous chapter it was shown that a Markov process describes well the distribution of spells of length five days or greater. Persistence, the simplest form of which is Markov persistence, is one form of non-randomness present in time series, others being trend and periodicities. In this chapter rainfall records from twelve stations are tested for homogeneity and their non-random elements are investigated using decadal means, filtering techniques, power spectrum analysis, and eigenvector analysis. Results from different stations are compared and stations are classed "West" or "East". As mentioned in Chapter 1, the presence of any trend or periodicity in several records considerably increases its significance.

Common variations in "West" and "East" stations are compared with those of circulation indices revealed in the results of similar analyses. Cross spectra between rainfall and circulation indices are also computed to investigate the relationship between oscillations in rainfall and indices. The P and C indices measure the frequencies of progressive and cyclonic types of circulation which produce most rainfall. The extent to which P variations relate more closely to the rainfall of "West" stations and C variations to that of "East" stations, as might be expected from synoptic experience and from Murray and Benwell's (1970) correlation analyses between monthly rainfall and indices is investigated.

Periodicities which are apparent in Scottish data are compared with those of Brunt's (1925) analysis of European weather, with Gray's (1975) analysis of S.E. England rainfall and temperature, and with results catalogued by Lamb (1972b). Periods of generally low or high rainfall values covering several decades are also compared with periods with a predominance of a given circulation type, as given by Lamb's (1972a) classification according to the frequency of Westerly type over several decades and by Schove's (1950) classification of circulation types over Europe and the North Atlantic.

5.2 The data and its apparent inhomogeneities

Monthly and annual values of circulation indices and of rainfall are analysed using the techniques described in detail in Section 5.3. The station positions are shown in Figure 5.1 and the periods of data used are given in Table 5.1

In Table 5.1 there is overlap for the Edinburgh records from different sites only in 1896. Readings at Blackford Hill and at other sites to the south of Edinburgh towards the Pentlands receive more rainfall than Charlotte Square and other sites in the town. Most of the pre-1896 records are in the town. Comparison of long-term means for the periods 1785 to 1896 and 1896 to 1973 show the higher yield of the latter record at the more exposed Blackford Hill site. In 1896 itself the town gauge of Charlotte Square gave 599.1 mm while the Blackford Hill gauge gave 616.5 mm.

At Loch Leven Sluices for the overlap period of 1933 to 1944,

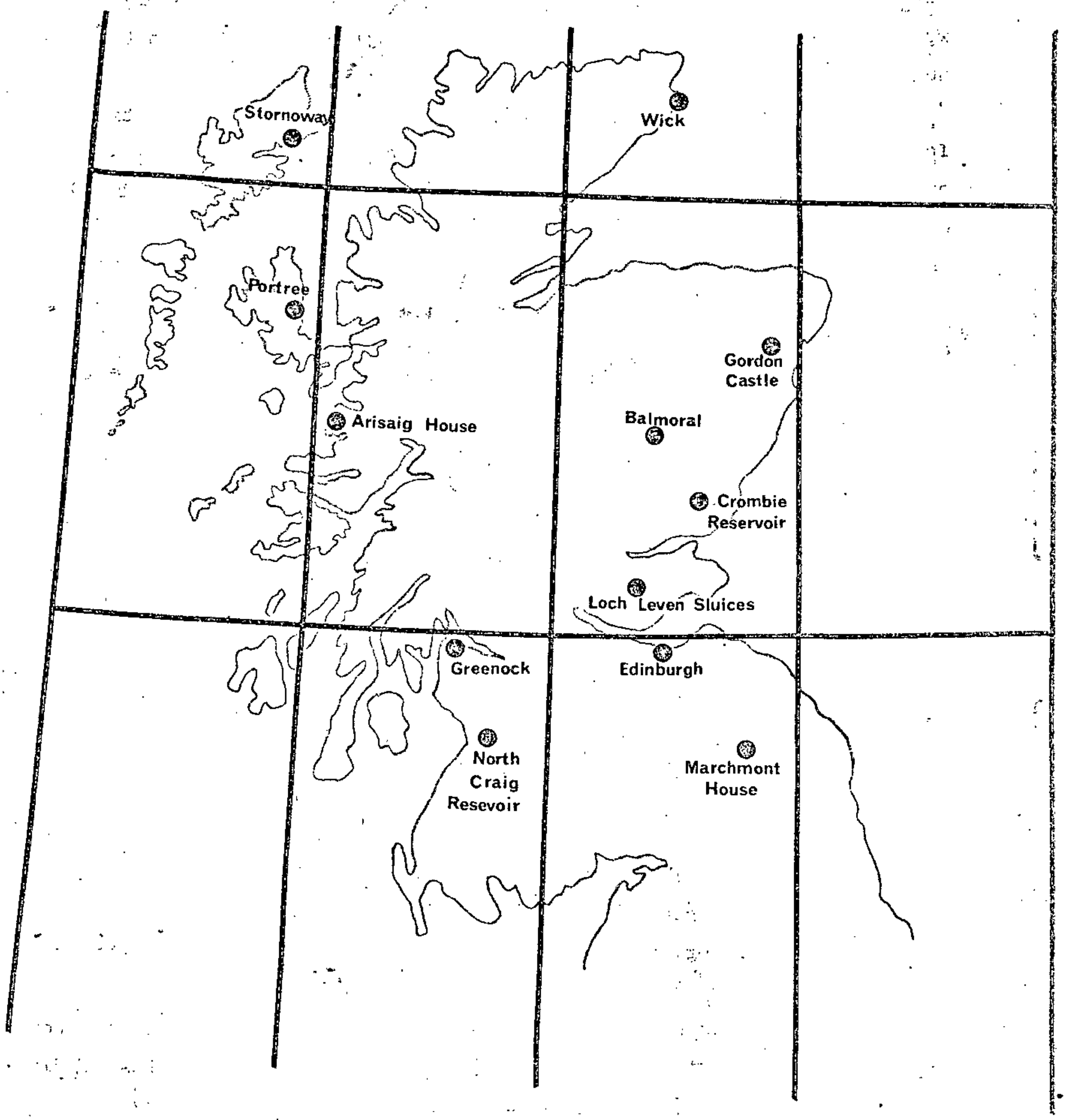

Figure 5.1 Station network for studies of secular changes

the 1877 to 1944 gauge had a mean value of 872.4 mm (standard deviation 26.1 mm) and the 1931 to 1971 gauge 921.3 mm (standard deviation 27.8 mm). As the readings from the original gauge were not homogeneous in themselves, it proved difficult to construct one continuous homogeneous record from readings at the two sites. In fact, records from the first site to 1933 and the second site from 1933 were considered as one record.

For none of the stations was it considered possible to homogenise the rainfall records prior to further analysis, so as to allow for changes of gauge site or exposure since sufficient overlap was not available to compare records from different sites. An attempt was made to identify inhomogeneities within a record, using Cramer's test (see 5.3.1), and any apparently anomalous decadal mean value for a given station was then compared with corresponding decadal means for other stations. The locations of the gauges (Figure 5.1) did not allow direct comparison of gauge records for relative inhomogeneities using one of Kohler's (1949) test. Neither was any gauge-record considered to be of such homogeneity that it could be used as a basic comparison gauge. (Kohler's test takes one of two forms: (i) sets of annual values are plotted on semi-log paper and the natural tendency for precipitation amounts to bear a constant ratio between locations then appears as a constant difference; (ii) series of annual totals are plotted against each other in the form of cumulative sums, plotted points tending to fall on straight lines for records of relative homogeneity.)

The stations in Table 5.1 were subsequently divided into "East" and "West" stations with the first seven classed as "East"

Table 5.1 Summary of Scottish rainfall data for studies of secular changes

Record	Period	Sub-Period	Site	Comments
Marchmont House	1867-1973			Minor changes of exposure.
Edinburgh	1785-1973	1770-1805	Unknown	Observer, Mr. Adie) C.F. Mossman (1896)
		1805-1821	Various	
		1822-1855	Unknown	Observer, Mr. Adie)
		1856-1896	Charlotte Square	
		1896-1973	Blackford Hill	
Loch Leven Sluices	1842-1971	1842-1871		Slight changes of site in 1871 with 1 per cént change in mean value.
		1871-1944		
		1933-1971		
Crombie Res.	1975-1973			
Balmoral	1882-1973			Minor site changes in May 1955, June 1967.

/...

Table 5.1 continued

Record	Period	Sub-Period	Site	Comments
Gordon Castle	1865-1973			
Wick	1877-1973	1877-1941	Coastguard Station	
		1941-1945	Airfield	
		1945-1973	Airfield	Improved site.
Stornaway	1877-1973	1876-1930	Town	
		1931	Coast	More exposed than town.
		1932-1936	Town	
		1937-1973	Coast	
Arisaig House	1890-1973			Minor site changes in May 1955.
Portree	1900-1905			
	1910-1973			Minor site changes in 1936.
Greenock	1878-1973			
North Craig Res.	1880-1973			

stations. The assignment of Wick to the East category was somewhat arbitrary, a fact which emerges in the comparison of the results of the analyses of the records of different stations made in sections 5.3 and 5.4.

5.3 Methods of analysis

5.3.1 Decadal means of annual rainfall

Decadal means were first calculated for each station and compared with its overall mean using Cramer's test. The latter compares the mean value of a subrecord of n values, x_k, with its overall mean, $\bar{x}$, using its overall standard deviation S. The test defines a statistical t_p as in equation 5.1 which is distributed as Student's t with N-2 degrees of freedom (Mitchell et al, 1966).

$$t_p = \left(\frac{n\,(N-2)}{N - k(1 + t_k^{2})} \right)^{\frac{1}{2}} t_k \qquad (5.1)$$

where $$t_k = \frac{x_k - x}{S}$$

Decadal means were tested against long-term means for each station. Those decades which had a mean value significantly different from the long-term mean, as determined by the 95 per cent of the two-tailed t-test, were compared with decadal means of other stations in order to isolate anomalously wet or dry decades which might suggest inhomogeneities in individual records.

5.3.2 "Low-pass" filtering

Decadal means only give a rough guide to the overall variations in rainfall. Ten year running means, for example, provide a continuous estimate of mean values. However, running means suffer from various pit-falls; in particular, some long-period oscillations may be exaggerated and shifted in phase. If the values which are meaned are weighted, spurious periodicities may be partially eliminated and peaks and troughs in a record may be located more accurately.

Weighted nine-year running means were applied to the annual rainfall records, in an attempt to filter out high-frequency variations. The filter weights were determined using binomial coefficients, as suggested by Mitchell et al (1966). Ordinates of the Gaussian probability curve are represented by the binomial coefficients C_k in equation 5.3 and are used to determine the filter weights, w_i, which are normalised as in equation 5.4

$$C_k = \frac{m!}{k!\,(m-k)!} \qquad (5.3)$$

$$\sum_{-n}^{+n} w_i = 1 \qquad (5.4)$$

A filtered time series is thus produced with each member $\bar{x}_t$ being produced from the 2n + 1 values of the original series x_t.

$$\bar{x}_t = \sum_{-n}^{+n} W_i \, x_{t+i} \qquad (5.5)$$

$$R_f = \cos^m \pi f \qquad (5.6)$$

The frequency response of the binomial filter R_f, is expressed by equation 5.6. m in equation 5.6 is determined so that R_f falls to 0.5 for oscillations of period equivalent to six times the standard deviation, $\frac{\sqrt{m}}{2}$, of the binomial distribution. For the records analysed, the desired condition was to suppress effects of oscillations of period less than ten years so that the filter response to such oscillations was less than 0.5 i.e. $6\frac{\sqrt{m}}{2}$ was to be about 10 and m = 12 to the nearest integer. The weights W_j normalised to 1 were hence determined as:

$$W_o = 0.22, \quad W_{\pm 1} = 0.20, \quad W_{\pm 2} = 0.12,$$

$$W_{\pm 3} = 0.05, \quad W_{\pm 4} = 0.02 \qquad (5.7)$$

Low-pass filtered time series should reveal the presence of any low frequency oscillation in the data and it should be possible to obtain an indication of the phase of such an oscillation from peaks and troughs in the filtered record. The filtered record will also reveal those variations which are present only in sections of a record, a property which will not be revealed by spectral analysis.

5.3.3 Power spectrum analysis

In order to investigate periodicities present in the rainfall series, power spectra were computed using (i) the Blackman - Tukey autocovariance approach, (ii) the Fast Fourier transform. As remarked in Chapter 1, the power spectrum has the property of revealing trend, persistence, and periodicities. The effects of trend may be removed (see below) but the effect of persistence in determining the underlying shape and form of the power spectrum, against which the significance of the periodicities must be tested, may be difficult to determine. Simple "red noise" spectra may be plotted and used to test for the presence of periodicities if persistence is of a simple Markov type. Other forms of persistence are more difficult to model and their spectra are not easily computed. The spectra below, where persistence did not seem to be very prevalent, were treated as white noise spectra with possible periodicities, the significance of which was investigated.

5.3.3.1. The Blackman - Tukey autocovariance approach (ACV)

The Biomedical Computing Porgram (BMD) written by W. J. Dixon et al (1968) at the University of California was used. Autocovariances $R_x(p\Delta t)$ at lag p; where Δt is the time interval between observations, were calculated (equation 5.8) and detrended (equation 5.9) by a least squares method.

$$R_x(p\Delta t) = \frac{1}{n-p} \sum_{q=1}^{n-p} x_q \, x_{q+p} \qquad p=0,1,2..m \tag{5.8}$$

$$A_x(p\Delta t) = R_x(p\Delta t) - \beta - \alpha i \qquad i=0,1,..n-1 \tag{5.9}$$

where $$\beta = \sum_{0}^{n} \frac{x_i \,(2i - n + 1)}{\frac{(n-1)\, n\, (n+1)}{6}}$$

$$\alpha = \frac{\bar{x} - \beta\,(n-1)}{2}$$

and $\bar{x}$ is the mean of x_i.

Raw spectral estimates $P_x(u)$ are obtained at frequencies $\frac{u}{2\Delta t}$ in equation 5.10 and are then smoothed by "hanning" as in equations 5.11.

$$P_x(u) = \frac{2\Delta t}{\pi} \sum_{o}^{m} \varepsilon_p \, A_x(P\Delta t) \cos \frac{uP\pi}{m} \quad p=0,1,..m \qquad (5.10)$$

where
$$\varepsilon_p = 1 \qquad 0 < p < m$$
$$= \frac{1}{2} \qquad p = 0, m$$

$$SP_x(0) = 0.54\ P_x(0) + 0.46\ P_x(1) \qquad (5.11a)$$

$$SP_x(u) = 0.23\ P_x(u-1)+0.54\ P_x(u)+0.23\ P_x(u+1) \qquad (5.11b)$$

$$SP_x(m) = 0.54\ P_x(m) + 0.46\ P_x\ (m-1) \qquad (5.11c)$$

5.3.3.2 The Fast Fourier transform method (FFT)

The analysis proceeds in the manner suggested by Rayner (1972). In this method series of observations are increased to length n from D observations so that n has a value 2^l where l is an integer. Series are first meaned, and detrended using a least squares method, i.e. linear regression is performed between members of the series and time and members are then adjusted so that their mean value is zero and they have no linear trend. A cosine bell filter function h(j) in equation 5.12 is applied to the time series and this is equivalent to applying the "hanning function" H(f) to spectral estimates as in equations 5.11 above.

$$h(j) = \tfrac{1}{2}\left(1 - \cos\left(\frac{\pi j}{G}\right)\right) \qquad 0 \leq j \leq G$$

$$= 1 \qquad G \leq j \leq D-G$$

$$= \tfrac{1}{2}\left(1 - \cos \frac{\pi(D-j)}{G}\right) \qquad D \leq G \leq j \leq D \tag{5.12}$$

where $G \sim \frac{D}{10}$ and j, the interval between observations, is equivalent to $p\Delta t$ above.

Filtered series are then extended to length 2^1 by the addition of zeros and Fast Fourier sine and cosine coefficients are calculated at frequencies $\left(\frac{k}{2\pi\Delta t}\right)$ apart where $0 < k < \frac{n}{2}$ in equation 5.13

$$a_k = \frac{2}{n} \sum_{o}^{n-1} h_j \, x_j \cos \frac{2\pi j k}{n} \tag{5.13a}$$

$$b_k = \frac{2}{n} \sum_{o}^{n-1} h_j \, x_j \sin \frac{2\pi j k}{n} \tag{5.13b}$$

The power spectrum, the spectrum of variance, is then calculated by squaring these estimates and dividing their sum by two, except in the case of k = 0 and $\frac{n}{2}$ where b_k is zero. The spectra are further smoothed to increase the stability of the estimates by summing estimates into non-overlapping bands of width five as in equation 5.14, at frequencies $\frac{u}{2\pi\Delta t}$

$$SP_x(0) = a^2(0) + \sum_{1}^{2} \frac{a^2(k) + b^2(k)}{2} \tag{5.14a}$$

$$SP_x(u) = \sum_{5u-2}^{5u+2} \frac{a^2(k) + b^2(k)}{2} \tag{5.14b}$$

$$SP_x(m) = \sum_{\frac{n}{2}-2}^{\frac{n}{2}-1} \frac{a^2(k) + b^2(k)}{2} + a^2(\tfrac{n}{2}) \qquad (5.14c)$$

Alternative band widths of summation to that of five may be used. Averaging over wider bands produces more stable estimates but with smaller resolution.

An alternative method of increasing the stability of the spectral estimates is to average estimates at a given frequency obtained from different samples of a given record.

5.3.3.3 Confidence limits of spectral estimates

In order to increase the number of estimates at low frequencies, the lags used in the ACV approach may be increased; alternatively, in the case of the FFT method the bandwidth of summation may be reduced. The confidence limits of each spectral estimate will thereby be reduced at the expense of higher resolution.

If several samples of a given spectrum are considered each sample spectrum can be assumed to be distributed above the value of the population estimate as $\frac{\chi^2}{\nu}$ where ν is the number of degrees of freedom of each sample estimate. If the sample variance, σ^2, is an unbiased estimate of the population spectrum, $\hat{\sigma}^2$, then for 90 per cent of the time σ^2 will be defined by the limits in equation 5.15 (Rayner 1971).

$$\frac{\nu\hat{\sigma}^2}{\chi^2(\nu,\ 0.05)} \leqslant \sigma^2 \leqslant \frac{\nu\hat{\sigma}^2}{\chi^2(\nu,\ 0.95)} \qquad (5.15)$$

In order to estimate the significance of a particular peak in a single sample spectrum, its spectral value must be tested against that of the local continuum. The assessment of the appropriate value of the latter is somewhat subjective and is dependent on the other forms of non-randomness present. If the series has been detrended, only effects of persistence, which are assumed to be insignificant in the spectra below, and those of other periodicities need be considered in determining the shape of the continuum. Leakage of power from spectral peaks to neighbouring estimates may occur (see Hinich and Clay 1968) and can complicate the assessment of the value of the local continuum. Interference between peaks in the spectra may also affect certain estimates.

Sample spectral peaks may be tested against the value of $\frac{\chi^2(\nu,\ 0.05)}{\nu}$ times the value of the determined local continuum if the oscillation corresponding to the peaks can be expected on a priori grounds and the sample spectrum can be assumed to belong to a population of spectra with a similar peak. If the peak does not correspond to a wavelength which is noteworthy in previous studies, more stringent tests need to be applied. The probability of a peak occurring in one spectrum must be related to the joint probability of its occurrence in m spectra as in equation 5.16.

$$q_t = 1 - (1 - q_d)^m$$

$$\text{i.e.} \quad q_d = \frac{q_t}{m} \qquad (5.16)$$

$$q_d = \frac{\chi^2(\nu, 0.05)}{m} \qquad (5.17)$$

The inequalities in equation 5.15 are thus altered by the factor m and can thus be used to assess the significance of peaks in individual spectra. For the Blackmann-Tukey autocovariance approach, ν is given by equation 5.18, and for the ACV method (with non-overlapping bands) by equation 5.19

$$\nu = \frac{2n}{m} - \frac{1}{2} \quad \left(= 5.5 \text{ if } m = \frac{n}{3}\right) \qquad (5.18)$$

$$\nu = \frac{W\,(D-G)}{\frac{n}{2}} \quad \left(= \frac{10(D-G)}{n} \text{ for } W = 5\right) \qquad (5.19)$$

5.3.3.4 The choice of method FFT v ACV

The FFT method is computationally faster than the ACV method and the amount of leakage of power from peaks in the spectra to neighbouring peaks has been shown to be less for the FFT method (Hinich and Clay, 1968). However, the overall stability of a given number of estimates, as defined by confidence limits in equation 5.15, is usually greater with the ACV method. The confidence limits of ACV estimates widen further compared to ACV estimates with the number of data points available falling below a value of 2^1 as can be seen in equation 5.18.

Using the ACV method, the maximum lag that can be used to compute power spectra is considered to be equal to one third of the number of data points, and for such cases the number of degrees of freedom of each estimate is 5.5. In the case of annual power spectra of indices, 111 years of data were used with the maximum permissible lag of 37 to produce 37 estimates. To produce approximately the same number of estimates using the FFT method, summation of raw estimates over bands of width two would produce 32 estimates with 3.4 degrees of freedom. The variance of such estimates would be unacceptably high, and in the case of series of length 80 or 90 years it would be even higher.

For monthly data of length 85 years, D = 1020 and FFT estimates can be summed over bands of width five to produce 102 estimates with 8.9 degrees of freedom. With the ACV method the maximum number of points which can be analysed, and the maximum lag are 1000 and 199 respectively, these limits being set by the program used. For the monthly data, 984 data points were analysed using the ACV method with a lag of 199 to produce 199 estimates with 9.8 degrees of freedom. The over-all confidence limits of each set of estimates will be approximately the same but individual FFT estimates, being fewer in number, will have greater stability. The resolution of ACV estimates will be greater due to the larger number of estimates.

5.3.3.5 Difficulties in the application of power spectrum analysis

The chief difficulties in the use of power spectra to find periodicities may be summarised as:

(1) The modelling of the effects of persistence which cannot be approximately described as Markov persistence.

(2) The assessment of the amount of leakage of power from spectral peaks to neighbouring estimates.

(3) The recognition of peaks in the power spectrum which arise from interference between other periodicities revealed in the spectrum.

(4) The recognition of peaks which represent harmonics of other periodicities.

(5) For high frequency peaks a decision as to how much power has been aliased from frequencies not resolved by the analysis.

(6) The assessment of the exact frequency of a periodicity (since spectral estimates are made for frequency bands).

(7) The assessment of the statistical significance of periodicities revealed by the spectral peaks.

The first three and the last of these difficulties have already been mentioned. The fifth becomes important when effects of the quasi-biennial oscillation are investigated below in annual spectra where the Nyquist period is two years.

5.3.4 Cross spectrum analysis

5.3.4.1 Cross covariance approach

Cross spectra were calculated in order to investigate the relationship between periodicities which were apparent in rainfall and circulation index time series. Cross spectrum analysis proceeded in a similar fashion using the BMD program and cross-covariance approach as in power spectrum analysis. Cross-covariances are computed between series x and y (equation 5.19) and detrended (equation 5.20) in a similar way to autocovariances:

$$R_{xy}(p\Delta t) = \frac{1}{n-p} \sum_{1}^{n-p} x_q \, y_{q+p} \qquad p=0,1,2..m \qquad (5.19a)$$

$$R_{xy}(-p\Delta t) = \frac{1}{n-p} \sum_{1}^{n-p} x_{q+p} \, y_q \qquad p=0,1,2..m \qquad (5.19b)$$

$$A_{xy}(p\Delta t) = R_{xy}(p\Delta t) - \beta^1 - \alpha^1 t$$

$$\text{where} \quad i = 0,1,..n-1 \qquad (5.20)$$

β^1 and α^1 are defined for xy in a similar way to β and α for x in equation 5.7. The cospectrum $C_{xy}(u)$ and the quadrature spectrum $Q_{xy}(u)$, are given in equations 5.21 and 5.22 in a similar way to the power spectrum.

$$C_{xy}(u) = \frac{\Delta t}{\pi} \sum_{0}^{m} \varepsilon_p \left(A_{xy}(p\Delta t) + A_{xy}(-p\Delta t)\right) \cos \frac{pu\pi}{m} \qquad (5.21)$$

$$Q_{xy}(u) = \frac{\Delta t}{\pi} \sum_{0}^{m} \varepsilon_p \left(A_{xy}(p\Delta t) - A_{xy}(-p\Delta t)\right) \sin \frac{pu\pi}{m} \qquad (5.22),$$

where $u = 0,1,2...m$ and $\varepsilon_p = \frac{1}{2} \quad p = 0,m$

$$\varepsilon_p = 1 \quad 0 < p < m.$$

Unlike autocovariances, cross-covariances are not symmetrical about lag 0, and are therefore calculated for both positive and negative lags. The sine transform of their difference (the quadrature spectrum) in addition to the cosine transform of their sum (the cospectrum) are then computed. The cospectrum describes the relationship between the two series considered exactly in or out of phase, while the quadrature spectrum considers the relationship at lag one quarter of a cycle. The cospectrum and quadrature spectrum are subsequently smoothed by "hanning", as in equation 5.11, to produce smoothed estimates $SC_{xy}(u)$, $SQ_{xy}(u)$.

The complete relationship between the two series at a given frequency $\frac{u}{2m\Delta t}$ is measured by the coherence square COHSQ(u), which is analogous to the correlation coefficient, r_{xy}, and by the phase of the cross spectrum $\Phi_{xy}(u)$.

$$\mathrm{COHSQ}(u) = \sqrt{\frac{(SC_{xy}(u))^2 + (SQ_{xy}(u))^2}{SP_x(u)\ SP_y(u)}} \qquad (5.23)$$

$$\Phi_{xy}(u) = \tan^{-1} \frac{SQ_{xy}(u)}{SC_{xy}(u)} \qquad (5.24)$$

Estimates of coherence and phase are inversely related but the actual values of the coherence square is dependent on the power spectrum estimates $SP_x(u)$, $SP_y(u)$.

5.3.4.2 The FFT method

With this method analysis proceeds on similar lines to the computation of the FFT power spectra. Series are detrended, filtered using the Tukey cosine bell of equation 5.12, and zeros added to produce a series of length n which equals a value of 2^{l}. Fourier cosine $a_x(k)$, $a_y(k)$, and sine $b_x(k)$, $b_y(k)$ transforms are then calculated using the FFT. From these coefficients the raw cospectrum and quadrature spectrum are calculated as in equations 5.25 and 5.26 and are then summed into bands of width five as in equations 5.27 and 5.28. Fourier sine coefficients for harmonics 0 and $\frac{n}{2}$ are zero and this is taken account of in equations 5.27 and 5.28.

$$C_{xy}(k) = \frac{a_x(k)\, a_y(k) + b_x(k)\, b_y(k)}{2} \tag{5.25}$$

$$Q_{xy}(k) = \frac{a_x(k)\, b_y(k) - a_y(k)\, b_x(k)}{2} \tag{5.26}$$

$$SC_{xy}(0) = 2C_{xy}(0) + \sum_{1}^{2} C_{xy}(k) \tag{5.27a}$$

$$SC_{xy}(u) = \sum_{5u-2}^{5u+2} C_{xy}(k) \tag{5.27b}$$

$$SC_{xy}(m) = \sum_{\frac{n}{2}-2}^{\frac{n}{2}-1} C_{xy}(k) + 2C_{xy}\left(\frac{n}{2}\right) \tag{5.27c}$$

$$SQ_{xy}(0) = Q_{xy}(k) \tag{5.28a}$$

$$SQ_{xy}(u) = \sum_{5u-2}^{5u+2} Q_{xy}(k) \tag{5.28b}$$

$$SQ_{xy}(m) = \sum_{\frac{n}{2}-2}^{\frac{n}{2}-1} Q_{xy}(k) \tag{5.28c}$$

5.3.4.3. Confidence limits of coherence square and phase

Jenkins and Watts (1968) show that ACT(u) defined by equation 5.29 is distributed normally with variance $\frac{1}{\nu}$ where ν is the number of degrees of freedom of each estimate.

$$ACT(u) = \text{arc tanh} \; | COHSQ(u) |$$

$$= \frac{1}{2} \ln \left(\frac{1 + 1 \sqrt{COHSQ(u)}|}{1 - 1 \sqrt{COHSQ(u)}|} \right) \qquad (5.29)$$

$$\text{var } ACT(u) = \frac{1}{\nu} \qquad (5.30)$$

The modulus sign arises since the actual sign of $\sqrt{COHSQ(u)}$ is determined by the values of the cospectrum and quadrature spectrum. For a given probability level, confidence limits $g[\%]$ may be placed on ACT(u) as in equation 5.31 and extreme values of ACT(u) at these limits may be transformed to values of coherence square. For a completely incoherent pair of series the average coherence square determined from the variance of ACT(u) would be $\frac{2}{\nu}$

$$ACT(u) \pm g[\%] \left(\frac{1}{\;} \right)^{\frac{1}{2}} \qquad (5.31)$$

$\tan\Phi_{xy}(u)$ is also distributed normally with variance given by equation 5.32 and the confidence limits of $\tan\Phi_{xy}$ are defined by equation 5.33.

$$\text{var } \tan\Phi_{xy}(u) \simeq \sec^4\Phi_{xy}(u) \frac{1}{\nu} \left(\frac{1}{COHSQ_{xy}(u)} - 1 \right) \qquad (5.32)$$

$$\tan\Phi_{xy}(u) \pm g[\%] \left\{ \sec^4\Phi_{xy}(u) \frac{1}{\nu} \left(\frac{1}{COHSQ_{xy}(u)} - 1 \right) \right\} \quad (5.33)$$

In equation 5.32 it can be seen that the variance of the phase angle is dependent on the value of the coherence square, and is very dependent on the value of the phase angle on account of the $\sec^4\Phi_{xy}$ factor. Confidence limits on phase angles are a minimum for Φ_{xy} equal to 0 or π, and a maximum, encompassing the complete range of phase angles for Φ_{xy} equal to $\frac{\pi}{2}$ or $\frac{3\pi}{2}$.

As the expression in equation 5.30 is complicated, 90 per cent confidence limits on Φ_{xy} for certain high values of coherence square were tabulated in Table 5.2 for $\nu = 5.5$ as in annual cross spectra and in Table 5.3 for $\nu = 9$ as in monthly cross spectra. $g[90\%]$ in equations above will have the usual value of 1.645 as determined by the tables of the normal distribution.

In Tables 5.2 and 5.3 only values of Φ_{xy} in the range 0 to $\frac{\pi}{2}$ have been considered due to the symmetry properties of the expression 5.32. Values of phase angles between π and $\frac{\pi}{2}$, π and $\frac{3\pi}{2}$, 2π and $\frac{3\pi}{2}$ have similar confidence limits to angles in the range 0 to $\frac{\pi}{2}$. Angles in the table are expressed in radians and in fractions of a circle.

Table 5.4 gives 90 per cent confidence limits on coherence square for annual and monthly cross spectra, together with ranges of phase angles of interest when considering in-phase and anti-phase relationships between oscillations confidence limits of a coherence estimate rise as values of coherence fall. When investigating cross spectra between series, the primary use of coherence is to know the probability that oscillations of given frequency in two series are completely coherent (COHSQ = 1) or incoherent. This probability depends both on the actual values of coherence and their confidence limits. Average values of coherence square for completely incoherent series are also given and it is useful to compare the lower confidence limits of a coherence estimate with such a value when assessing the relative significance of a given coherence estimate.

From coherence values given in Table 5.4 it was decided that a value of 0.85 or greater in the case of annual series, and 0.80 or greater in the case of monthly series, could be considered as significant coherent estimates, i.e. that oscillations of a given frequency in two series would be related for such estimates. Smaller values of coherence, 0.75 in the case of annual series and 0.70 in the case of monthly series, could also be regarded as significant if they occurred in several cross spectra.

The phase relationship between significantly coherent oscillations and the confidence limits of the phase angle were then investigated. The ranges of phase angle which could describe an in-phase or anti-phase relationship, between significantly coherent oscillations were determined from the confidence limits of phase angles given in Tables 5.2 and 5.3. These ranges are given in Table 5.4 for various values of coherence.

When coherence values increase, confidence limits on phase angles decrease. The ranges of values of Φ_{xy} representative of a possible exactly in-phase relationship become smaller as coherence values increase, while the ranges of Φ_{xy} representative of a probable nearly in-phase relationship become larger.

In Table 5.4 the choice of values of significant coherence

Table 5.2 90 per cent confidence limits on phase angle $0 - \frac{\pi}{2}$ for annual rainfall and indices cross spectra

$\Phi_{xy}(u)$ in radians, $\Phi_{xy}'(u)$ in fractions of a circle. LL, UL upper and lower limits of phase angle for given coherence values

Φ_{xy}	Φ_{xy}'	COHSQ = 0.85		COHSQ = 0.90		COHSQ = 0.95	
		LLΦ_{xy}'	ULΦ_{xy}'	LLΦ_{xy}'	ULΦ_{xy}'	LLΦ_{xy}'	ULΦ_{xy}'
0.000	0.000	-0.046	0.046	-0.037	0.037	-0.025	0.025
0.063	0.010	-0.036	0.055	-0.027	0.046	-0,016	0.035
0.126	0.020	-0.027	0.064	-0.018	0.056	-0.006	0.045
0.188	0.030	-0.018	0.073	-0.008	0.065	0.004	0.055
0.314	0.050	-0.000	0.092	0.011	0.084	0.023	0.074
0.628	0.100	0.043	0.138	0.056	0.131	0.071	0.123
0.942	0.150	0.077	0.183	0.097	0.178	0.118	0.171
1.257	0.200	-0.008	0.224	0.089	0.222	0.151	0.217
1.571	0.250	-0.250	0.250	-0.250	0.250	-0.250	0.250

Table 5.3 90 per cent confidence limits on phase angles
$0 - \frac{\pi}{2}$ for monthly rainfall and index cross spectra

$\Phi_{xy}(u)$ in radians, $\Phi_{xy}'(u)$ in fractions of a circle. LL, UL upper and lower limits of phase angle for given coherence values

Φ_{xy}	Φ_{xy}'	COHSQ = 0.70		COHSQ = 0.80		COHSQ = 0.90	
		$LL\Phi_{xy}'$	$UL\Phi_{xy}'$	$LL\Phi_{xy}'$	$UL\Phi_{xy}'$	$LL\Phi_{xy}'$	$UL\Phi_{xy}'$
0.000	0.000	-0.055	0.055	-0.043	0.043	-0.029	0.029
0.063	0.010	-0.046	0.064	-0.034	0.052	-0.019	0.039
0.126	0.020	-0.038	0.073	-0.024	0.061	-0.010	0.048
0.188	0.030	-0.029	0.082	-0.015	0.071	0.010	0.058
0.314	0.050	-0.012	0.100	0.003	0.089	0.019	0.077
0.628	0.100	0.027	0.144	0.047	0.136	0.067	0.126
0.942	0.150	0.051	0.188	0.083	0.181	0.112	0.173
1.257	0.200	-0.098	0.227	0.029	0.224	0.136	0.219
1.571	0.250	-0.250	0.250	-0.250	0.250	-0.250	0.250

Table 5.4 Confidence limits on coherence square and important ranges of phase angles for annual and monthly cross spectra

actual value COHSQ	ANNUAL RAINFALL lower limit COHSQ	upper limit COHSQ	Φ_{xy}' ranges of interest 1.		2.		MONTHLY RAINFALL lower limit COHSQ	upper limit COHSQ	Φ_{xy}' ranges of interest 1.		2.	
0.90	0.65	0.97	0.97	0.03	0.93	0.07	0.70	0.96	0.98	0.02	0.90	0.10
			0.47	0.53	0.43	0.57			0.48	0.52	0.40	0.60
0.85	0.51	0.96	0.96	0.04	0.94	0.06	0.59	0.92				
			0.46	0.56	0.44	0.56						
0.80	0.37	0.95					0.49	0.92	0.96	0.04	0.92	0.08
									0.46	0.54	0.42	0.58
0.75	0.30	0.94					0.40	0.90				
0.70							0.32	0.89	0.94	0.06	0.93	0.07
									0.44	0.56	0.43	0.57
completely incoherent series		0.36						0.22				

estimates and of phase estimates which may be considered to represent in-phase or nearly in-phase relationships is subjective. While the variance of coherence and phase estimates may be expressed by equations 5.30 and 5.32, the physical significance of the estimates and their variance can only be determined by the investigator.

5.3.5 Eigenvector analysis

Eigenvector analysis was carried out as described in Chapter 2 on an 11 station by 84 year matrix using S-mode analyses. Real data and normalised deviation data matrices were multiplied by their transposes to produce 11 by 11 symmetric cross product and correlation matrices respectively. In the latter case the 84 year station means were subtracted from each station values and the 84 values were then normalised so that the sum of squares of the 84 values was unity.

11 eigenvectors with 11 space elements, one per station, were computed and sets of eigenvector multipliers were produced by the matrix multiplication of each eigenvector by the original matrix. Each set of eigenvector multipliers was considered as a basic time series and the relative extent to which each series described time variations at a given station was determined by the eigenvector space element. The overall significance of a given eigenvector and its set of multipliers in describing rainfall variations expressed in the original matrix was determined by the eigenvalue.

The sets of significant eigenvector multipliers were also considered as time series and the results of analysis of multiplier series and of individual series were compared. The 11 eigenvector space elements were first compared with station mean annual rainfall and its standard deviation, and with station position, in order to see if a particular set of eigenvector multipliers was of more relevance to particular stations. The analysis of the sets of eigenvector multipliers proceeded in the same way as for individual rainfall series using the "low-pass" filter, the power spectrum, and the cross spectra between the series and those of circulation indices. The comparison of the results of analysis of individual and eigenvector multiplier series attempted to identify multiplier series as describing variations of "East" or "West" stations.

Each eigenvector and its set of multipliers is independent of other eigenvectors and their sets of multipliers as demonstrated in Chapter 2. Eigenvector multiplier time series may therefore be expected to show particular time variations, periodicities of definable frequency or definite trends, rather than the sum of several variations.

5.4 Results of analysis of annual series

5.4.1 Decadal means

The values of decadal means of "East" Scottish rainfall stations appear in Table 5.5 and of "West" stations in Table 5.6, those values which are significantly above or below long-term means (according to Cramer's test) being underlined. Decadal means expressed as percentages of long-term means and averaged over "East" and "West" stations are plotted in Figure 5.2. General variations in the tables and figure can be compared with those of circulation indices (Table 5.7), and anomalously large or small values may be

Table 5.5 Decadal means "East" rainfall stations

	EDINBURGH		LOCH LEVEN SLUICES		MARCHMONT HOUSE		CROMBIE RES.		BALMORAL		GORDON CASTLE		WICK	
	mean	stand dev.	mean	stand dev.	mean	stand dev.	mean	stand dev.	mean	stand dev.	mean	stand dev.	mean	stand dev.
Period	1784 -	1896	1842 -	1973	1868 -	1973	1875 -	1973	1882 -	1973	1866 -	1973	1877 -	1973
Overall	651.9	120.4	908.0	144.3	816.4	141.5	894.3	155.4	843.5	122.3	751.8	103.1	763.8	91.9
	1896 -	1973												
	669.4	120.6												
1791-00	666.5	164.0												
1801-10	580.1	123.0												
1811-20	635.1	102.0												
1821-30	669.9	138.9												
1831-40	646.9	100.4												
1841-50	614.7	116.5	918.2	142.2										
1851-60	656.1	108.5	888.5	162.1										
1961-70	688.6	96.6	885.4	126.0										
1871-80	764.4	136.4	978.2	39.1	1011.4	186.2					833.1	117.9		
1881-90	616.7	97.8	904.2	129.5	857.5	102.6	929.6	155.2	866.6	162.8	689.6	72.2	726.0	78.8

/...

Table 5.5 continued

	EDINBURGH		LOCH LEVEN SLUICES		MARCHMONT HOUSE		CROMBIE RES.		BALMORAL		GORDON CASTLE		WICK	
	mean	stand dev.	mean	stand dev.	mean	stand dev.	mean	stand dev.	mean	stand dev.	mean	stand dev.	mean	stand dev.
1891-00	629.9	81.3	927.9	127.3	834.6	120.1	916.7	127.8	828.5	100.6	$\underline{829.8}$	$\underline{63.1}$	784.5	96.6
1901-10	646.4	130.8	908.6	173.7	788.2	126.2	886.5	107.9	829.1	133.6	760.8	94.7	794.7	78.7
1911-20	649.0	130.2	873.0	146.8	823.5	122.4	843.0	158.0	886.5	107.2	752.6	113.3	736.6	71.0
1921-30	723.3	110.2	931.9	127.8	876.6	99.8	954.0	136.4	865.9	148.1	745.1	106.0	774.8	94.1
1931-40	685.1	74.4	874.8	99.8	778.8	81.8	897.1	78.2	862.6	87.4	719.7	72.7	751.1	76.6
1941-50	690.3	98.1	952.8	112.0	764.5	85.6	885.2	106.2	792.5	83.1	779.0	74.7	757.0	79.0
1951-60	654.3	130.2	935.5	153.7	726.2	125.2	846.3	176.8	848.1	127.0	759.3	84.8	806.6	65.5
1961-70	673.1	102.1	915.4	95.0	782.1	106.7	851.4	122.9	860.6	74.2	727.2	76.5	803.3	94.2

Underlined values differ significantly from long-term means as determined by Cramer's test.

Table 5.6 "West" Coast stations decadal means

Records are given in m.m.

	NORTH CRAIG RESEVOIR		GREENOCK		ARISAIG HOUSE		PORTREE		STORNOWAY	
	mean	stand dev.	mean	stand dev.	mean	stand dev.	mean	stand dev.	mean	stand dev.
Period	1881 - 1973		1878 - 1973		1890 - 1973		1900 - 1904, 1910 - 1973		1876 - 1970	
Overall	1125.7	173.2	1533.2	229.1	1603.5	211.1	1708.9	273.6	1189.4	178.3
1881-90	924.3	77.0	1518.5	210.4					1217.3	105.9
1891-00	1099.8	11.9	1557.6	156.3	1625.6	164.8			1331.9	254.9
1901-10	1134.6	153.9	1591.6	240.0	1577.3	142.0			1258.2	155.2
1911-20	1061.2	109.2	1626.1	211.2	1620.5	116.1	1515.9	177.5	1296.2	89.3
1921-30	1217.4	148.8	1700.1	212.9	1684.6	165.6	1832.4	156.7	1279.6	81.1
1931-40	1171.4	155.7	1445.8	225.9	1568.4	269.0	1743.2	340.4	1155.5	155.6
1941-50	1240.5	177.8	1585.1	249.3	1554.2	266.7	1884.9	314.5	1116.4	127.6
1951-60	1127.0	166.1	1434.2	183.2	1628.9	222.2	1709.9	188.0	1076.3	113.1
1961-70	1193.8	130.6	1505.5	173.2	1579.4	225.0	1736.9	248.2	1088.3	142.1

Underlined values differ significantly from long-term means, by Cramer's test.

Table 5.7 Decadal means - circulation indices

	P		S		C		M	
	mean	s.d.	mean	s.d	mean	s.d	mean	s.d
1861-70	6.62	8.18	-2.96	2.23	-6.11	4.03	13.36	1.75
1871-80	2.94	6.73	-1.32	2.61	-2.13	6.11	14.00	2.76
1881-90	3.08	6.54	-0.33	3.22	-4.84	3.92	14.66	1.41
1891-00	2.36	5.12	-1.62	2.25	-6.89	3.38	13.97	2.25
1901-10	4.77	6.47	-1.18	2.86	-5.13	3.71	14.14	2.27
1911-20	6.57	7.71	-0.91	4.19	-3.73	4.76	11.72	8.62
1921-30	9.70	5.32	-0.65	3.00	-2.40	4.84	15.40	8.42
1931-40	3.04	5.55	-1.65	1.80	-4.61	2.75	14.40	5.44
1941-50	5.88	7.02	-0.67	2.80	-6.52	3.90	14.13	4.79
1951-60	2.19	7.03	-2.61	2.91	-4.90	5.93	14.22	1.79
1961-70	0.75	7.48	-2.58	2.23	-3.00	3.34	15.43	9.07
Overall	4.36	7.18	-1.50	2.93	-4.57	4.60	14.13	5.34

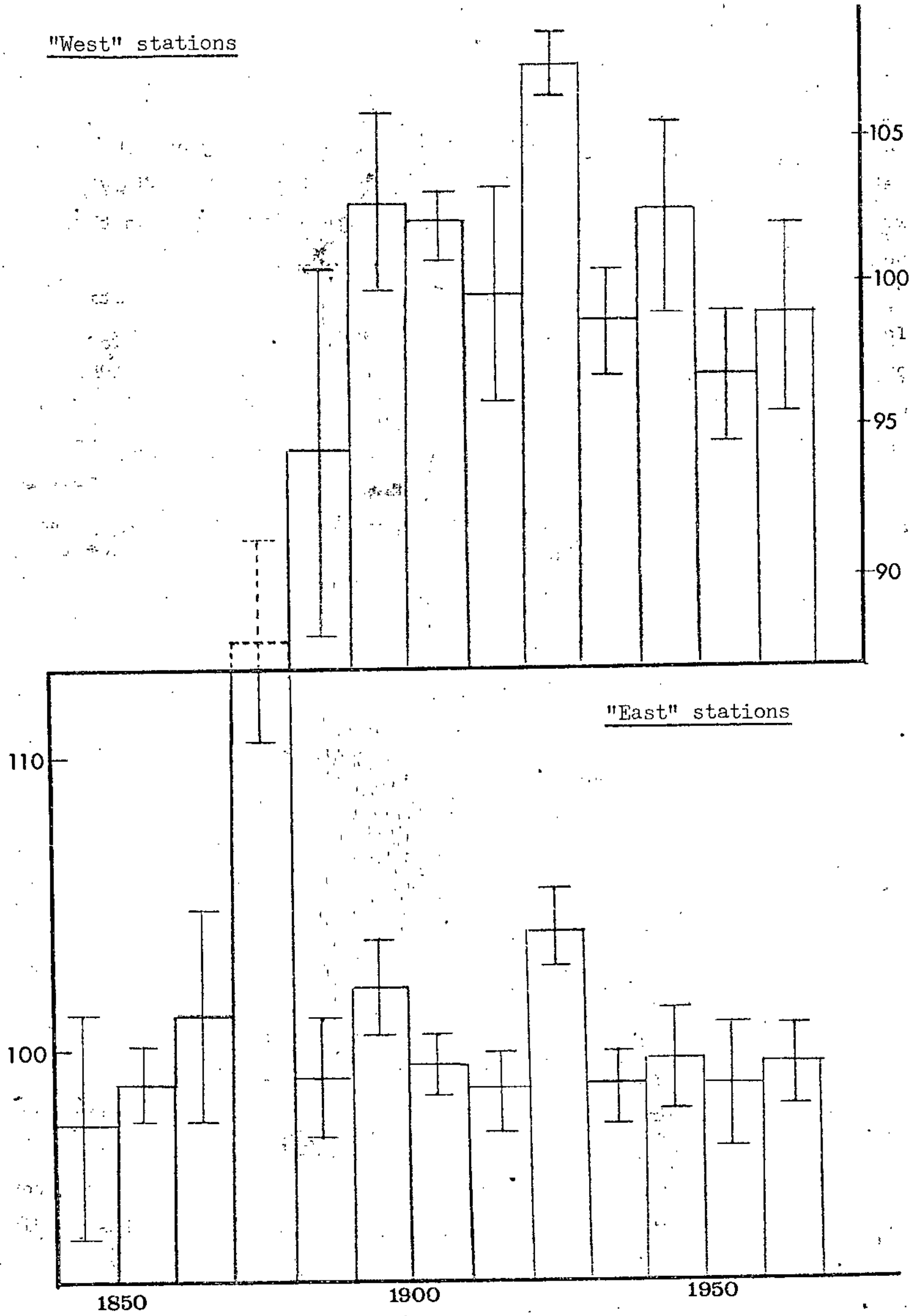

Figure 5.2 Decadal means averaged for (a) "West" stations, (b) "East" stations.

further considered in the light of these latter variations.

A notable feature of the Stornoway record (absent elsewhere) is the low values from the 1930's onwards. This is attributed to the discontinuity introduced by a change of site made in 1937.

At all stations the decade 1921 to 1930 is one of high rainfall and is also one of high P and C index. Heavier rainfall may be affected by a greater frequency of progressive and cyclonic weather types as measured by the P and C indices. The 1871 to 1880 decade is also particularly wet for "East" stations and is one of high C index. Only these two decades can be distinguished as wet for sets of stations whose high rainfall can be linked to high index values. General variations in "East" and "West" stations as expressed by decadal means do not correspond closely to those of C and P, though such relationships emerge for higher frequency variations in the analyses below.

The remaining anomalous values of the wet 1890's for Gordon Castle, and wet 1940's for North Craig Reservoir and Portree, and the dry 1910's for Portree remain unrelated to general variations in rainfall, and could, as in the case of Stornoway, reflect inhomogeneities in the gauge record. Portree, with two anomalous values, is particularly suspect.

5.4.2.1 "Low-pass" filtered rainfall and circulation indices

The results of applying the binomial filter in section 5.3(b) to circulation indices and rainfall appear in Figures 5.3 to 5.6. The filter failed to remove all the high frequency variations in rainfall as there were large differences between individ-

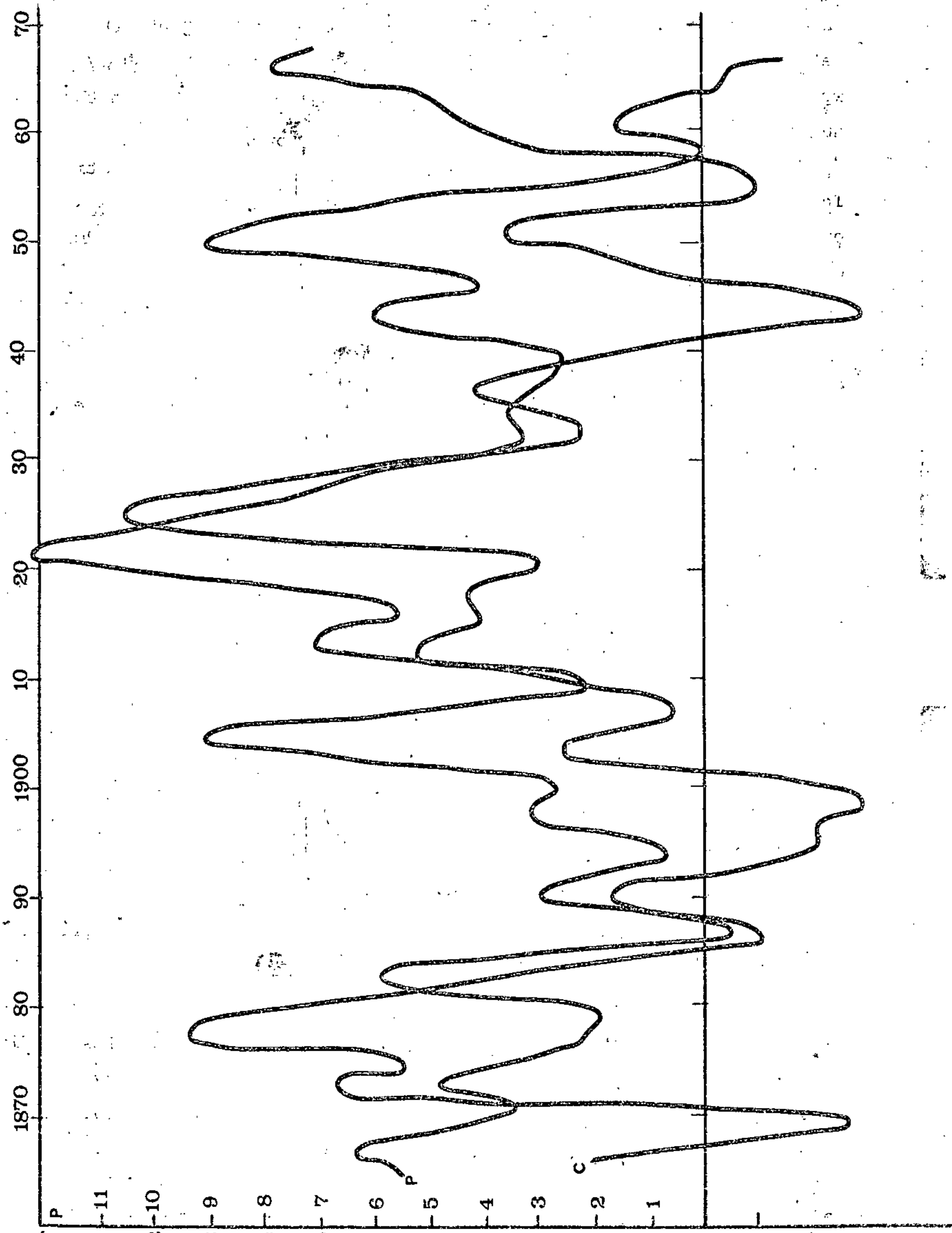

Figure 5.3 "Low-pass" filtered annual P and C indices

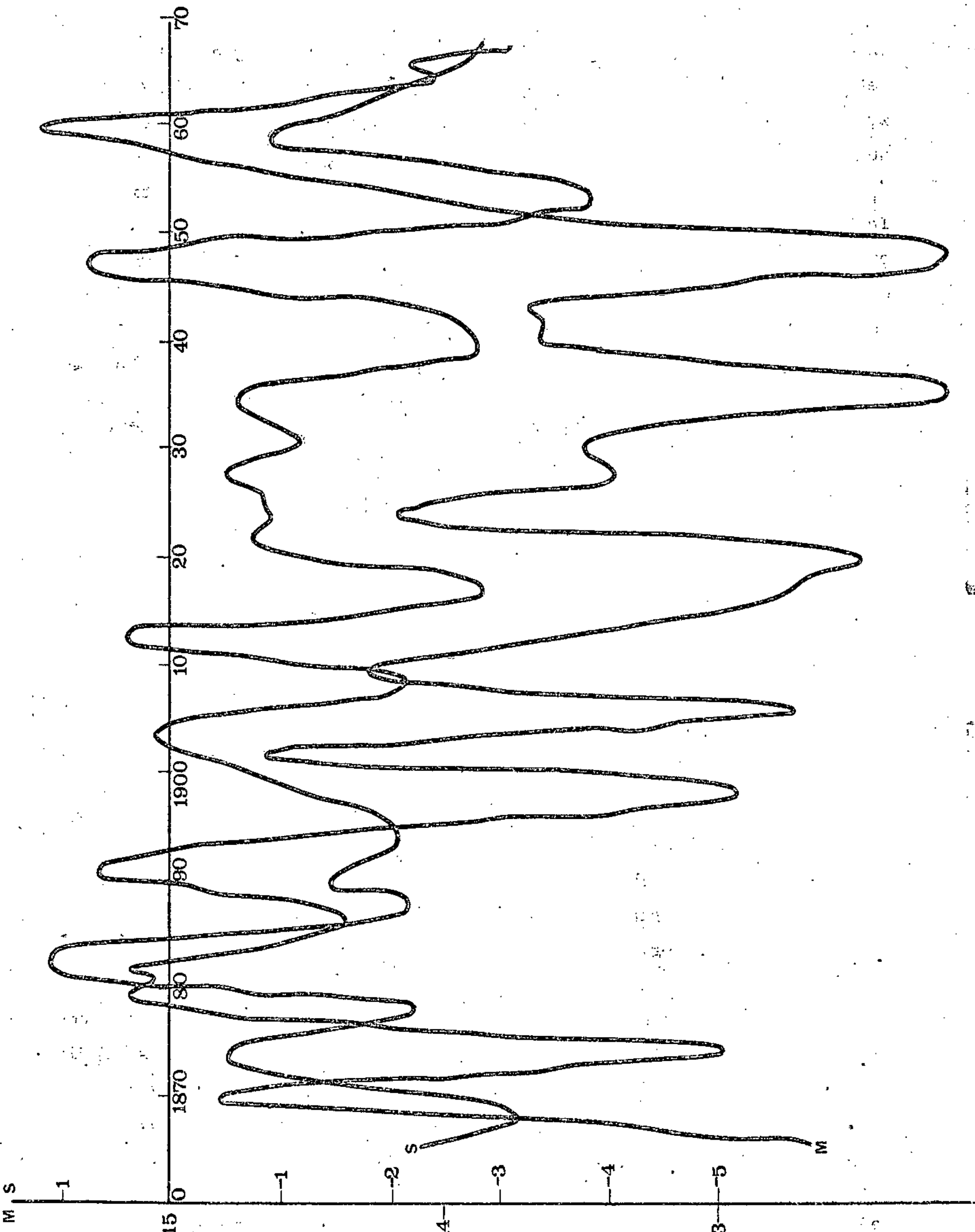

Figure 5.4 "Low-pass" filtered annual S and M indices

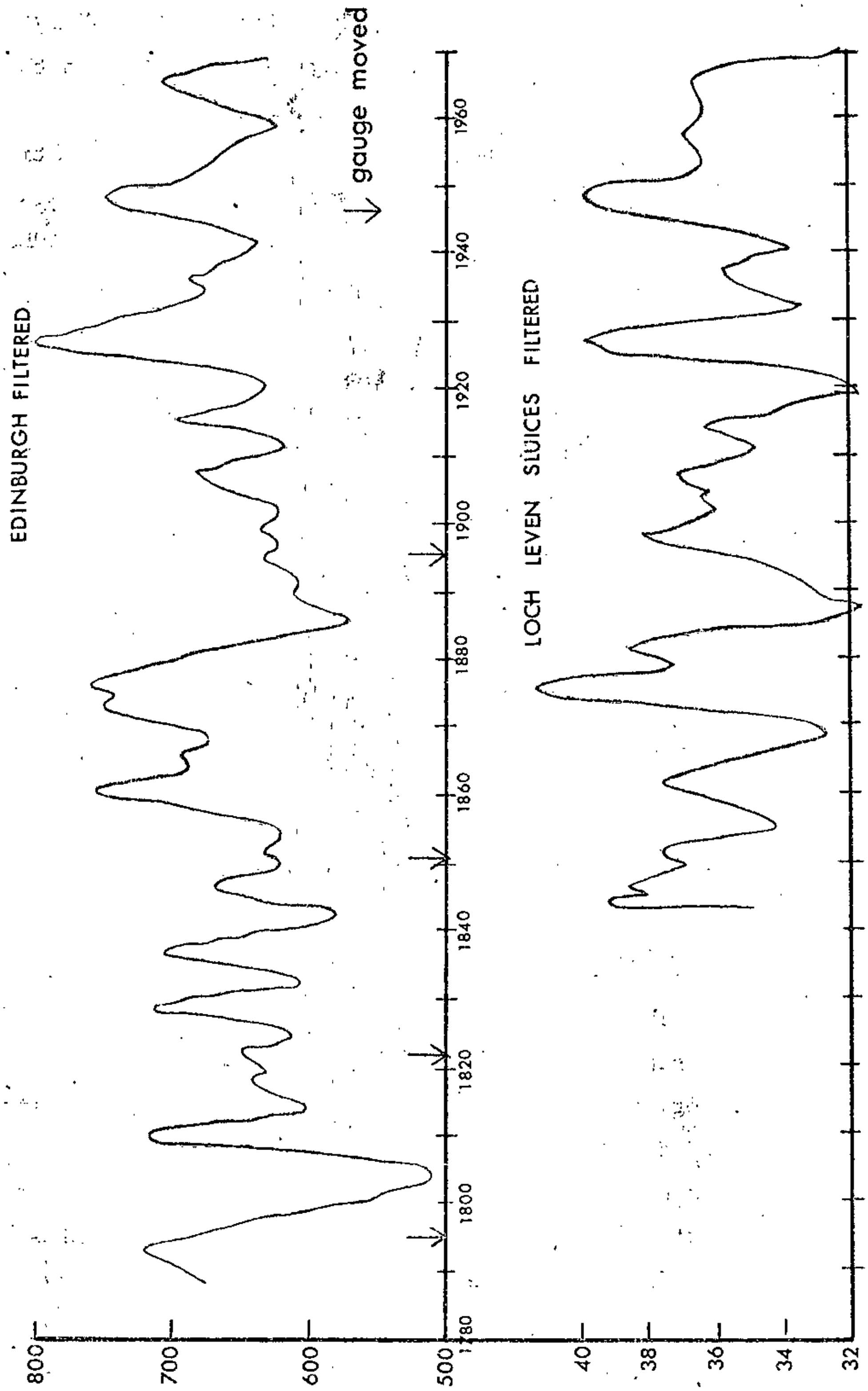

Figure 5.5 "Low-pass" filtered annual rainfall - Edinburgh and Loch Leven Sluices

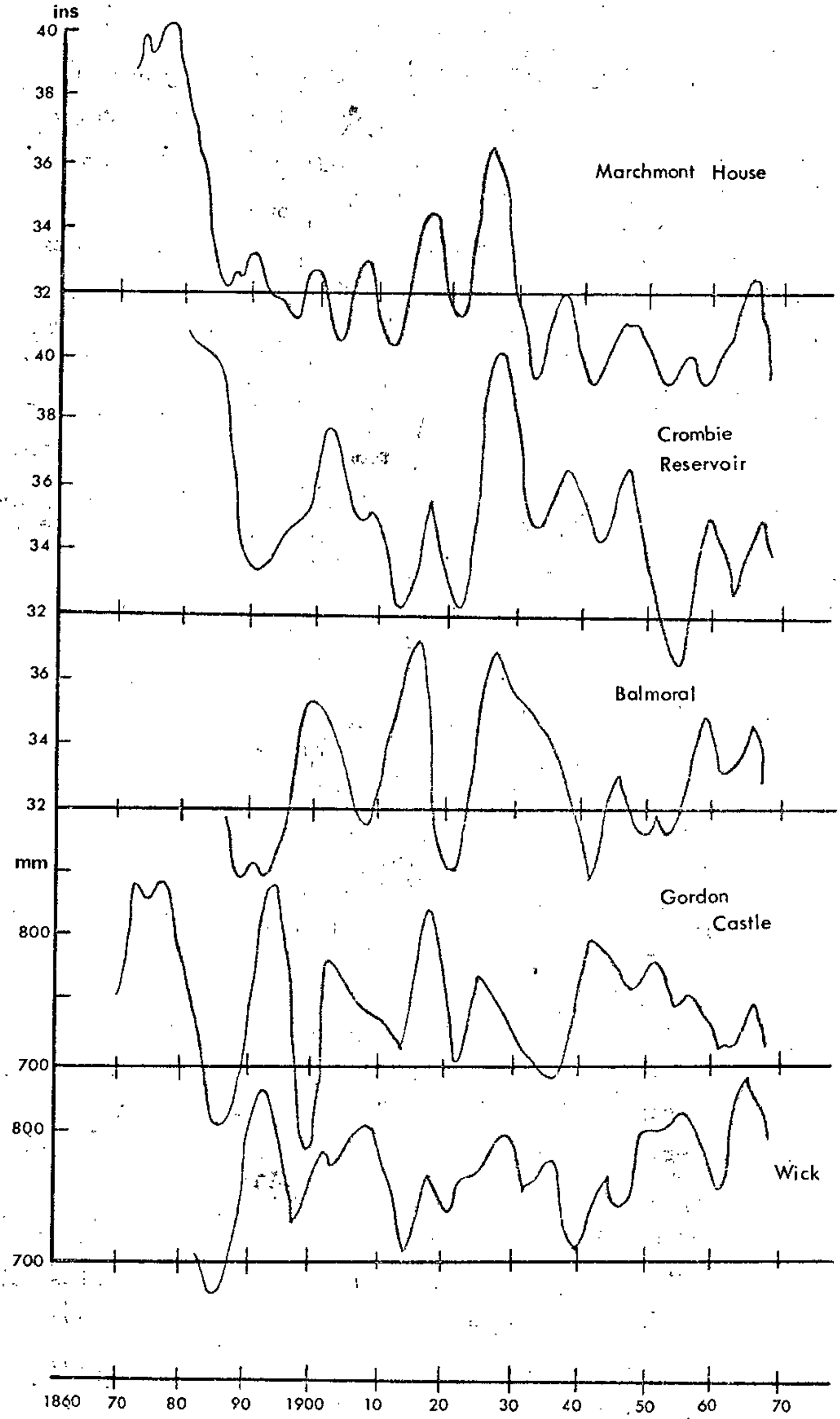

Figure 5.6 "Low-pass" filtered annual rainfall - Marchmont House, Crombie Reservoir, Balmoral, Gordon Castle, Wick.

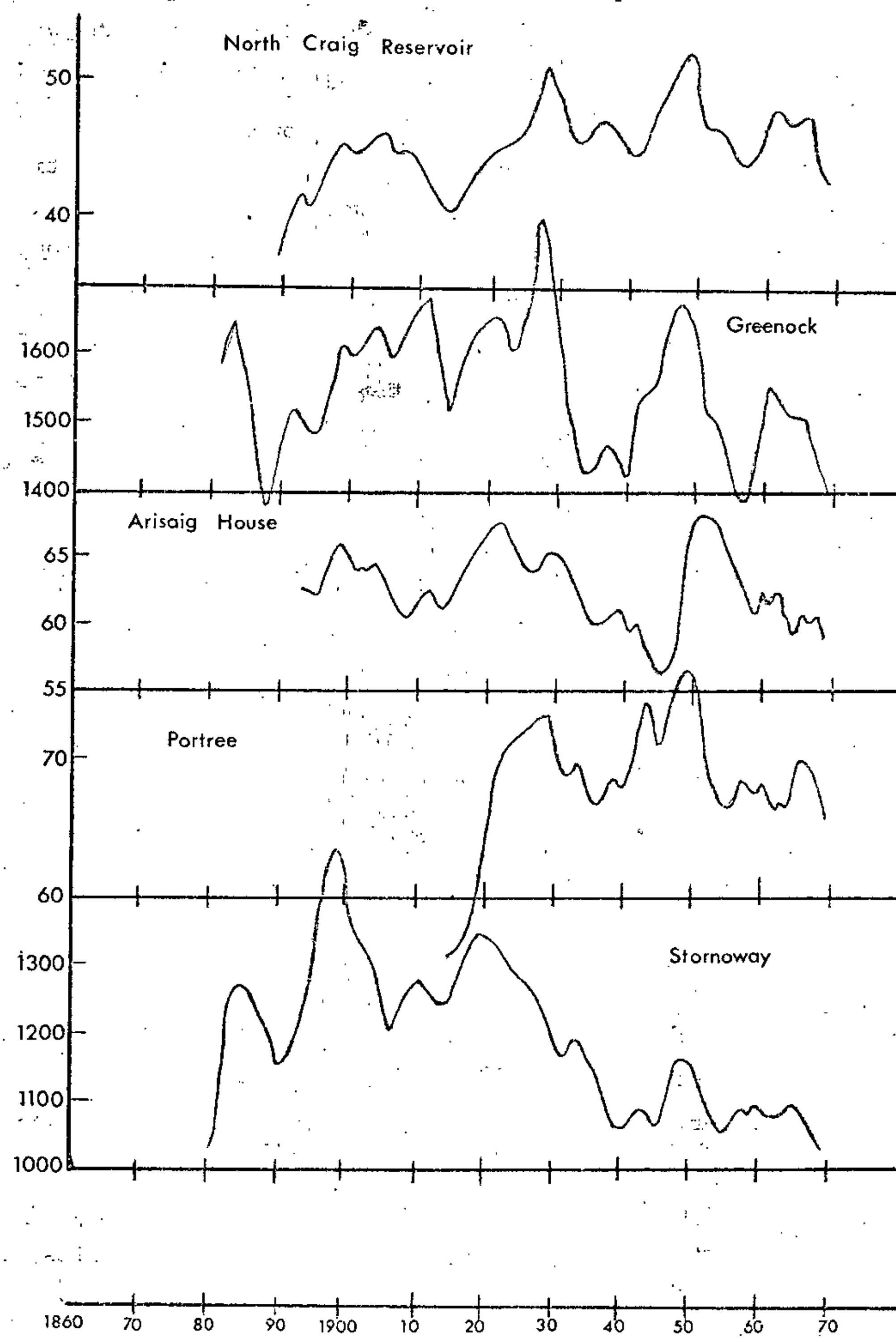

Figure 5.7 "Low-pass" filtered annual rainfall - "West" stations

ual years. Four years of data were lost at each end of the record as nine terms were required to compute each filtered value. It is of some interest to note that changes of site in the Edinburgh gauge in 1822, 1851, and 1899 can be recognised from a distinguishable secondary maximum value in the filtered data in Figure 5.5. The effect of the site change in Stornoway's record in 1937 is also apparent as in the decadal mean values.

Since nine years of data contibute to each member of the filtered series and since correlations between members of each series are larger than for unfiltered series, it is sufficient to compare rough positions of peaks and troughs in different series to identify wet and dry periods and possible common variations. For "East" stations maxima occur around the late 1870s, the late 1920s, and the late 1940s (apart from Gordon Castle in the 1940s) and minor peaks occur in 1917 and 1967. There are corresponding peaks in C index in 1876, 1925, and 1967.

The late 1920s and 1940s peaks are also present in "West" stations. There is also a peak in 1883 for Greenock and 1885 for Stornoway. The late 1940s peak and the 1880s peak correspond to a 1950Ppeak and a minor 1950 C peak, and an 1883 P peak respectively. Both the early 1920s peaks in P and the late 1920s peak in C describe the general increase in circulation strength in the 1920s which affected heavy rainfall. P and C can be seen in Figure 5.3 to have some other similar variations with P leading C.

Minima in rainfall data occur in the late 1880s, early 1910s, and the early 1940s at all stations, and in the early 1920s and the early 1960s at "East" stations. Corresponding minima occur in C around 1886 and 1943, and in P around 1887 and 1940. There

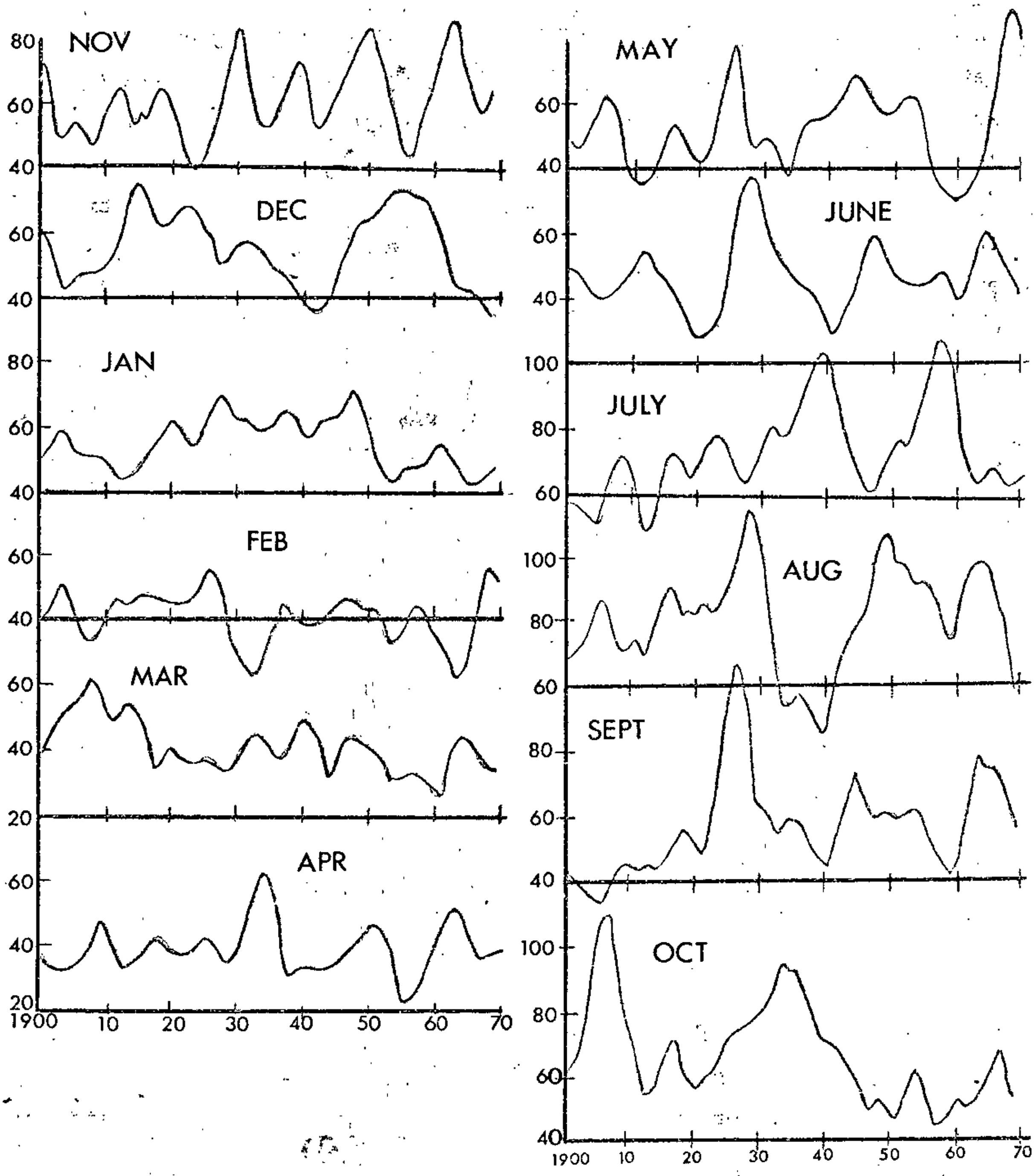

Figure 5.8 "Low-pass" filtered rainfall series of calendar months (different years) - Edinburgh Blackford Hills

are also 1869 minima in C index and Loch Leven rainfall, and an 1869 subsidiary minimum at Edinburgh. (Only the latter two stations have records for this early period.)

There are no peaks or troughs in S and M indices which can be related to those of rainfall. C index tends to give a better indication of rainfall variations than P especially for "East" stations though the two indices are themselves related. In some cases P variations are similar to those of "West" stations, when the same is not true to C variations.

No recurring oscillation of definite period stands out for a given record or for common subsections of several different records. The power spectrum analyses below investigate the presence of oscillations both of periods greater than ten years, which are not apparent in these filtered data, and of periods less than ten years whose effect has been removed by the filtering of data in this section.

5.4.2.2. Some results of "low-pass" filtering of rainfall series of calendar months (different years)

In Figure 5.8 graphs of filtered values of rainfall for each calendar month from 1896 onwards are shown for the Blackford Hill-Edinburgh gauge. Rainfall values are largest in July and August and show greater variability in August, September and October. As a result, the filtered rainfall curves for August and September show the closest resemblance to the annual rainfall curve in Figure 5.4. This perhaps trivial but not obvious point can be seen in the analyses of monthly records of other stations - the variations in rainfall for the months of August and September bear the closest resemblance to those of annual rainfall.

5.4.3 Power spectra

A list of the length of data, n years, and maximum lag, m years used for the computation of power spectra by the Blackman-Tukey autocovariance method is given in Table 5.8 for rainfall stations and circulation indices. The longest rainfall records were of the same length as the 1861 to 1971 records of circulation indices. Series of index spectra were calculated for each of the different lengths of rainfall station records but as these spectra were similar, only those for length 111 years are shown here. Records of index of the same length as those of rainfall records and their derived power spectra were used in 5.3.4. to calculate cross spectra and coherence estimates.

Three of the peaks listed in Table 5.8 are significant at the five per cent level for a "population spectrum"; these peaks would be significant if expected on a priori grounds. The Crombie Reservoir peak which is underlined, is significant at the five per cent level for a "sample spectrum". As the former three peaks were not 'predicted' and are present in only one spectrum, they cannot be considered to be of importance. The 3.1 year period peak appears in other spectra at similar periods and is especially noticeable in C index and in the rainfall of Edinburgh, Loch Leven Sluices, and Marchmont House. As the number of lags and length of record used were different in each case, and as spectral estimates are for frequency bands, the presence of peaks such as those at neighbouring frequencies in different series is sufficient to signify the presence of a particular oscillation in several series.

Table 5.8 Data used in the computation of annual rainfall and circulation index power spectra

Station	Period	No. of years n	Maximum lag m	Results	Period of significant peaks (years)
P index	1861-1971	111	37	Fig. 5.9	7.4
C index	1861-1971	111	37	Fig. 5.9	
S index	1861-1971	111	37	Fig. 5.10	
M index	1861-1971	111	37	Fig. 5.10	
Edinburgh	1861-1971	111	37	Fig. 5.9	12.3
Loch Leven Sluices	1861-1971	111	37	Fig. 5.9	
Marchmont House	1867-1971	105	35	Fig. 5.9	
Crombie Reservoir	1875-1971	97	32	Fig. 5.9	3.1
Gordon Castle	1866-1971	106	35	Fig. 5.9	
Balmoral	1882-1971	90	30	Fig. 5.9	15.0
Wick	1877-1971	95	32	Fig. 5.10	
Ansaig House	1890-1971	82	27	Fig. 5.10	
Portree	1910-1971	62	21	Fig. 5.10	
Greenock	1878-1971	94	31	Fig. 5.10	
North Craig Reservoir	1880-1971	92	30	Fig. 5.10	

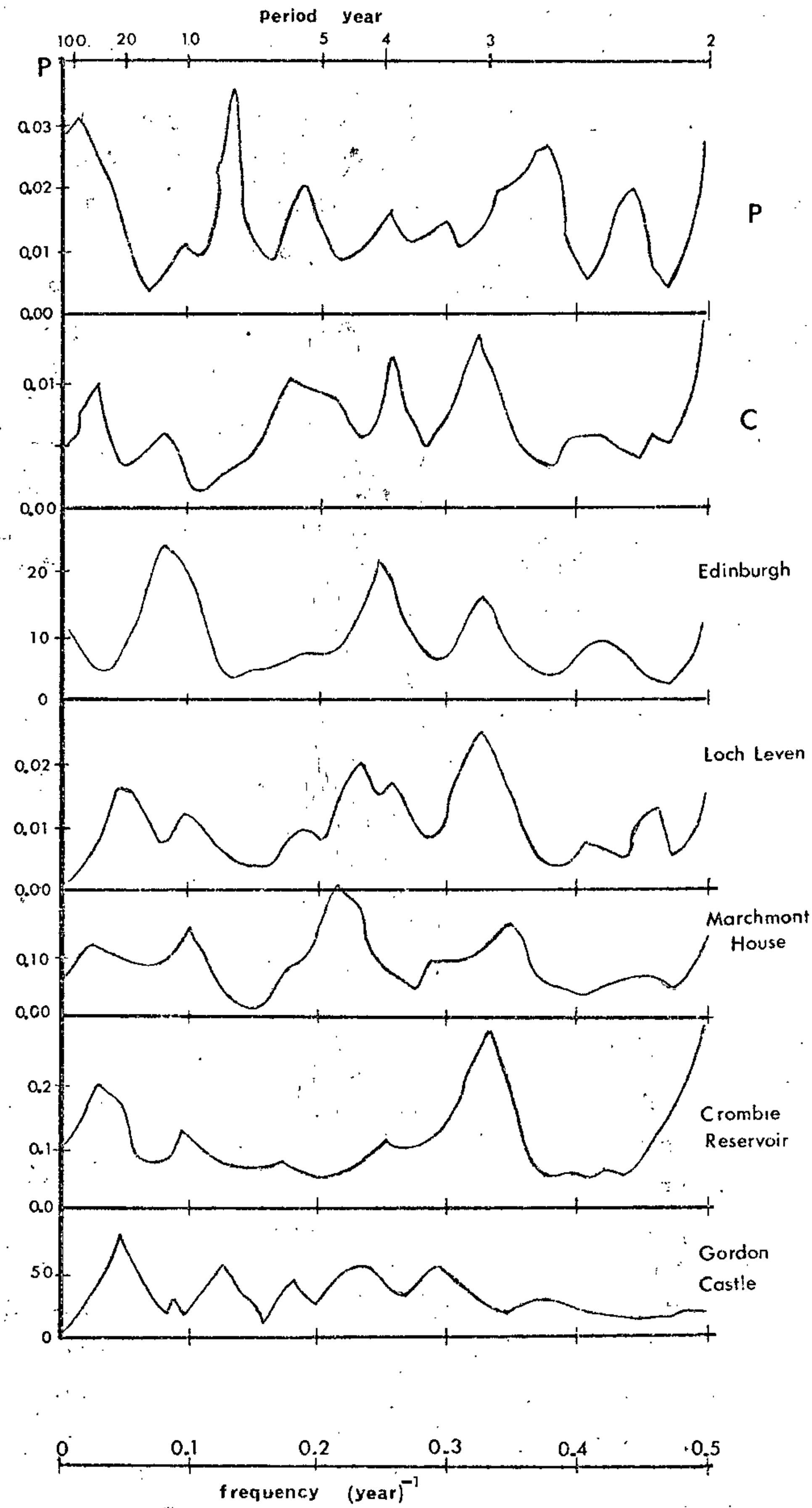

Figure 5.9 Power spectra - P and C annual indices, Edinburgh, Loch Leven Sluices, Marchmont House, Crombie Reservoir, and Gordon Castle rainfall

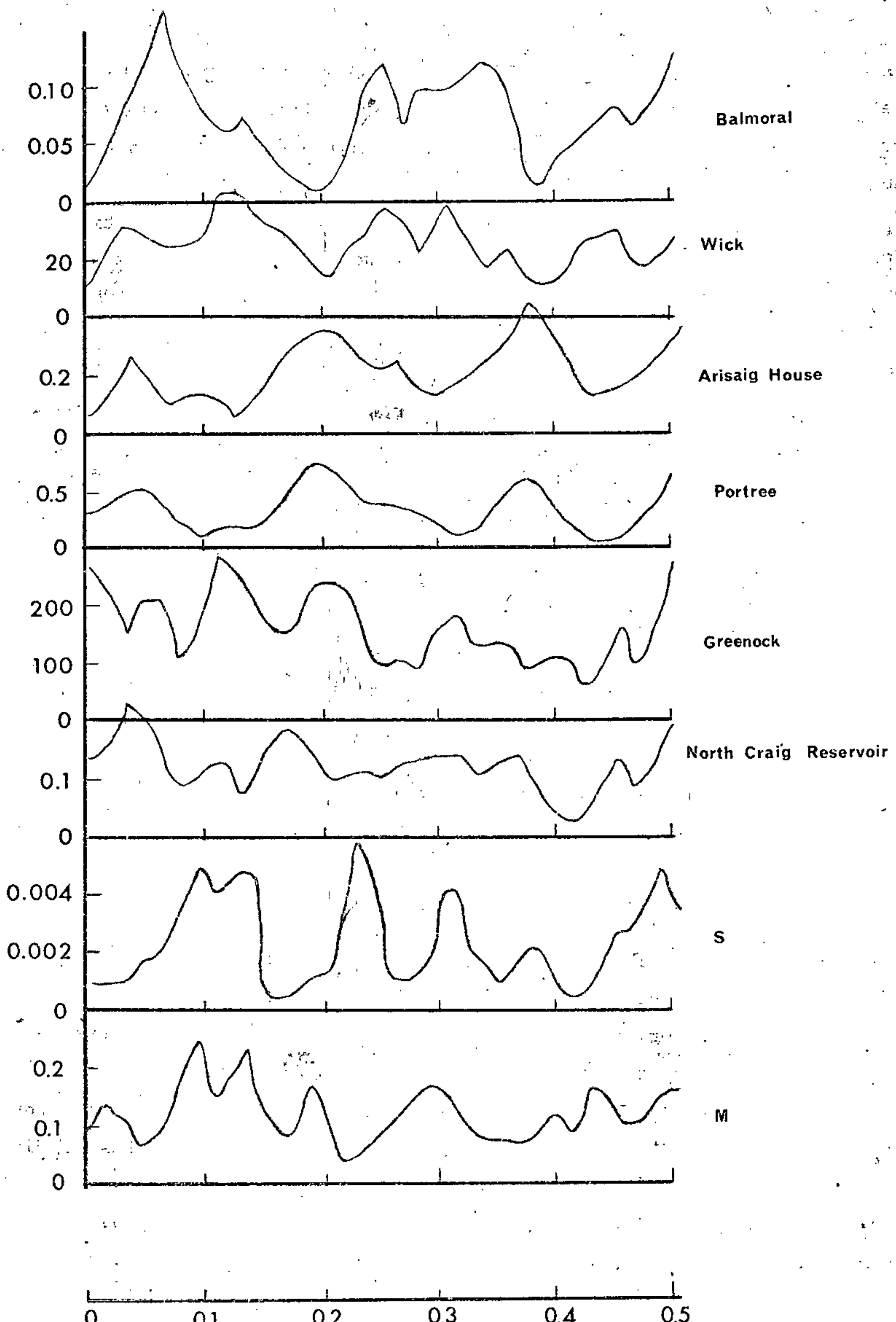

Figure 5.10 Power spectra - Balmoral, Wick, Arisaig House, Portree, Greenock, North Craig Reservoir, annual rainfall series and S and M indices

There is considerable evidence of the presence of a quasi-biennial oscillation in the data though aliasing of power from non resolvable high frequencies mentioned above in point 5, complicates the assessment of the significance and precise frequency of peaks at and near the Nyquist frequency. All the spectra show rising power towards two years, and in the case of P index, C index, and Crombie Reservoir this feature is prominent. The effects of aliasing in increasing the power at the two year period cannot be easily assessed. High frequency oscillations are studied in more detail below using monthly data since the significance and precise frequency of the quasi-biennial oscillation can be more easily assessed in monthly data.

Low frequency peaks are not easily interpreted due to the difficulty of locating the exact frequency of the peaks (point 6 above) and to possible effects of persistence (point 1). Uncertainties in locations of low frequency peaks imply large uncertainties in wavelength, the reciprocal of frequency. Persistence which may be present in the P index makes the distinction of the effects of a possible 74 year periodicity from those of persistence subjective. The persistence does not appear from the shape of the spectra to be of a simple Markov type.

In these spectra peaks can be seen which may be lower harmonics of high frequency peaks (as mentioned in point 4) or may be the result of interference between peaks (point 5). If a 2.0 year periodicity is present in C and Balmoral, the four year peak in these spectra probably represent lower harmonics of it. The 2.0 year P peak (frequency 0.5 cycles per year) could interfere with the 2.6 - 2.7 year peak to produce the observed 7.1 - 7.5

year peak (0.14 - 0.13), and the 2.0 year peak in C with the 3.1 year peak (0.32) to produce the observed 5.7 year peak (0.18). The 12.3 year peak (0.08) in Edinburgh could be a beat between the 3.1 years (0.32) and 4.1 year (0.24) periodicities. Part of the broad peak in S from 6 to 10 years (0.14 - 0.10) may result from interference between peaks at 2.03 years (0.49) and 4.1 years (0.24).

The most importance results in this section are the detection of a significant oscillation of period around three years, and of the well-known quasi-biennial oscillation. There are no definite low frequency peaks revealed by this analysis such as an 11 or 22 year sunspot cycle. The high frequency oscillations are studied in section 5.5 in greater detail using monthly data.

5.4.4. Cross spectra annual rainfall and circulation indices

In order to correlate the oscillations in circulation indices and rainfall, cross spectra between rainfall records and indices were calculated in section 5.3.4 above. High coherence values together with the phase angle between cross spectral estimates at given periods appear in Tables 5.9, 5.10 and 5.11; examples of coherence and phase spectra are given in Figure 5.12.

Cross spectra between P and C indices were also calculated but are not shown. The only significant relationship to emerge was for a period of 2.0 years, where the coherence square was 0.98, and the phase angle 0.99 of a circle. Sharply rising power towards two years exists in the power spectra of both C and P, and it seems probable that a quasi-biennial oscillation is present in both P and

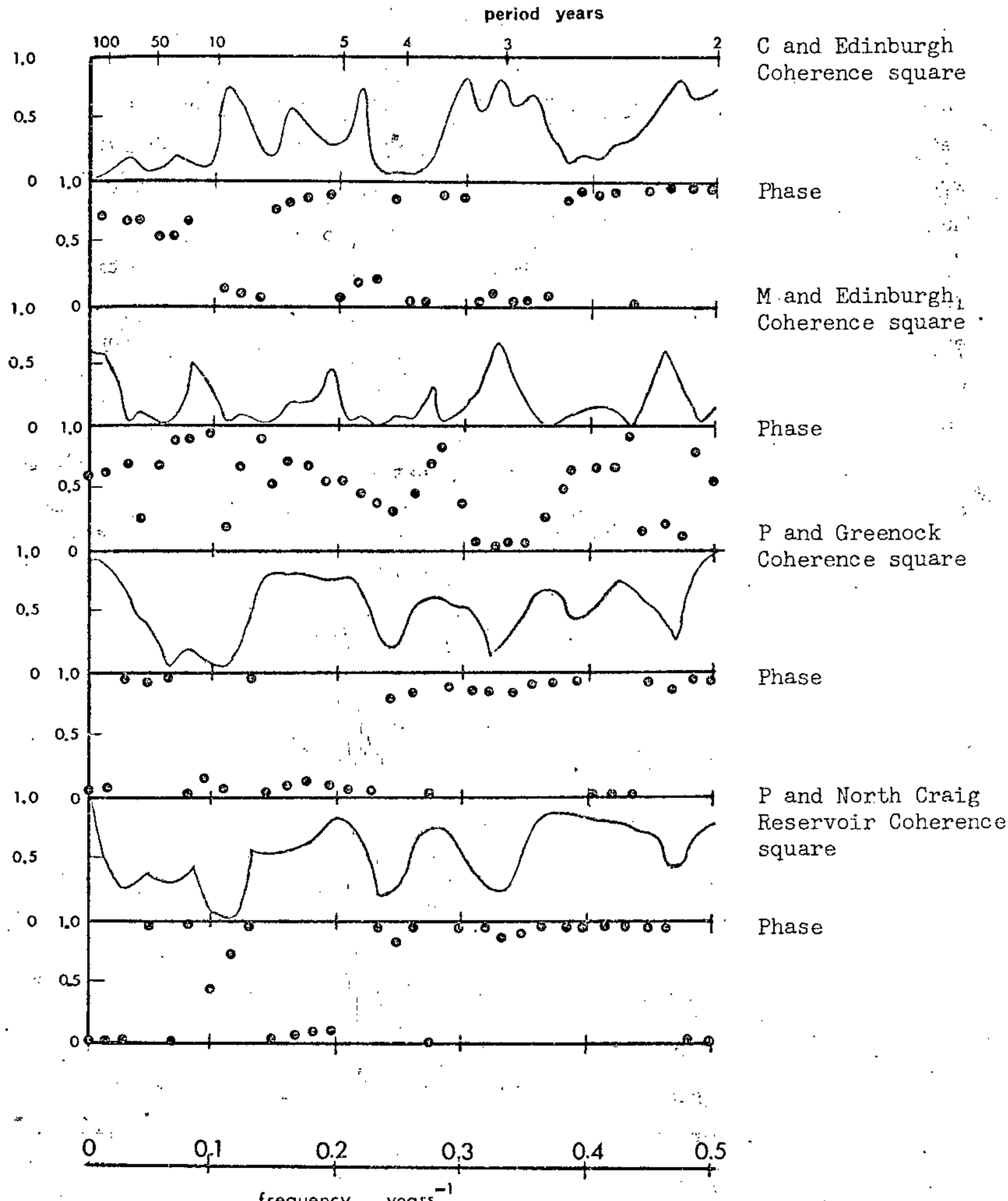

Figure 5.11 Coherence and phase spectra between annual rainfall and circulation indices (examples)

Table 5.9 Coherence values > 0.70 between C and "East" annual rainfall

Period T (years), coherence square COHSQ, phase of C relative to rainfall in fractions of a circle $\bar{\Phi}_{xy}'$																				
EDINBURGH			LOCH LEVEN SLUICES			MARCHMONT HOUSE			CROMBIE RESERVOIR			BALMORAL			GORDON CASTLE			GREENOCK		
T	COHSQ	$\bar{\Phi}_{xy}'$	T	COHSQ	$\bar{\Phi}_{xy}'$	T	COHSQ	$\bar{\Phi}_{xy}'$	T	COHSQ	$\bar{\Phi}_{xy}'$	T	COHSQ	$\bar{\Phi}_{xy}'$	T	COHSQ	$\bar{\Phi}_{xy}'$	T	COHSQ	$\bar{\Phi}_{xy}'$
									∞	1.00	0.82	∞	1.00	0.97	∞	1.00	0.38			
						23.3	0.83	0.02				60.0	0.82	0.98						
						17.5	0.88	0.03												
						14.0	0.89	0.95												
									6.41	0.79	0.18									
4.63	0.72	0.18							4.56	0.77	0.03									
			4.35	0.82	0.98	4.37	0.82	0.00	4.25	0.73	0.97									
			3.91	0.73	0.09													3.88	0.82	0.05
						3.50	0.96	0.88				3.53	0.70	0.90						
3.37	0.79	0.84							3.36	0.83	0.92							3.44	0.71	0.70
						3.19	0.76	0.06	3.20	0.92	0.01	3.15	0.91	0.05						
3.09	0.81	0.08	3.09	0.86	0.07	3.04	0.90	0.05	3.05	0.96	0.04	3.00	0.86	0.05	3.05	0.88	0.09			

/...

Table 5.9 continued

EDINBURGH			LOCH LEVEN SLUICES			MARCHMONT HOUSE			CROMBIE RESERVOIR			BALMORAL			GORDON CASTLE			GREENOCK		
T	COHSQ	$\bar{\Phi}_{xy}'$	T	COHSQ	$\bar{\Phi}_{xy}'$	T	COHSQ	$\bar{\Phi}_{xy}'$	T	COHSQ	$\bar{\Phi}_{xy}'$	T	COHSQ	$\bar{\Phi}_{xy}'$	T	COHSQ	$\bar{\Phi}_{xy}'$	T	COHSQ	$\bar{\Phi}_{xy}'$
			2.96	0.79	0.07	2.91	0.78	0.08	2.91	0.84	0.06				2.91	0.95	0.04			
			2.85	0.74	0.13				2.78	0.75	0.08									
									2.46	0.71	0.95									
									2.37	0.79	0.95				2.41	0.72	0.94			
			2.31	0.73	0.97	2.33	0.82	0.00	2.28	0.79	0.96				2.33	0.71	0.93			
			2.24	0.86	0.95	2.26	0.83	0.03	2.20	0.71	0.04									
2.12	0.82	0.93							2.14	0.77	0.08	2.14	0.75	0.03						
			2.05	0.73	0.01	2.06	0.81	0.00				2.07	0.87	0.03						
2.00	0.71	0.99	2.00	0.98	0.00	2.00	0.87	0.99	2.00	0.85	0.01	2.00	0.93	0.01				2.00	0.91	0.99

C with the same phase. Aliasing of power in the cross spectra, as well as in the power spectra, may enhance these estimates to some extent.

The results of cross spectrum analysis below between P or C and rainfall show that in general there is no relationship between low frequency oscillations as might be expected from results of low-pass filtering and power spectrum analysis of individual series. Oscillations in C or P around two or three years may however be related to those of "East" and "West" stations respectively.

High coherence values between C and rainfall records occur around two and three years. The coherence and phase estimates between C and Lock Leven Sluices, Marchmont House, Crombie Reservoir, Balmoral, and Greenock definitely suggest that the two year periodicities in C and these rainfall records are related and are in phase. Aliasing of power may again increase cross spectrum estimates as well as power spectrum estimates at and near the two year period.

Coherence peaks between C and rainfall occur at similar frequencies to power spectrum peaks near a period of three years suggesting that the three year periodicities in C and rainfall are related. Significant coherence peaks occur near three years between C and Marchmont House, Crombie Reservoir, Balmoral, and Gordon Castle and almost significant peaks between C and Loch Leven Sluices. The C index periodicity has an average phase lead of 0.05 of a circle which amounts to 0.15 years or two months.

There are other coherence peaks between C and rainfall records but they do not occur near enough to peaks in individual power

Table 5.10 Coherence values > 0.70 between P and "West" annual rainfall

Period T, (years), coherence square, COHSQ, phase of P relative to rainfall in fractions of a circle, Φ_{xy}'														
NORTH CRAIG RESERVOIR			GREENOCK			ARISAIG HOUSE			PORTREE			WICK		
T	COHSQ	Φ_{xy}'	T	COHSQ	Φ_{xy}'	T	COHSQ	Φ_{xy}'	T	COHSQ	Φ_{xy}'	T	COHSQ	Φ_{xy}'
∞	0.96	0.03	∞	0.91	0.01	∞	1.00	0.79						
			62.1	0.88	0.07									
			6.90	0.83	0.02									
			6.21	0.79	0.11									
						6.00	0.77	0.06						
5.46	0.74	0.11	5.65	0.79	0.14	5.40	0.74	0.07						
5.00	0.84	0.10	5.18	0.76	0.11	5.00	0.76	0.04						
4.60	0.72	0.03	4.76	0.80	0.07									
												4.27	0.87	0.19
												4.00	0.82	0.15
3.53	0.77	0.00												
						3.00	0.82	0.72	3.00	0.75	0.98			
						2.84	0.86	0.95	2.81	0.92	0.96			

/...

Table 5.10 continued

NORTH CRAIG RESERVOIR			GREENOCK			ARISAIG HOUSE			PORTREE			WICK		
T	COHSQ	Φ_{xy}'	T	COHSQ	Φ_{xy}'	T	COHSQ	Φ_{xy}'	T	COHSQ	Φ_{xy}'	T	COHSQ	Φ_{xy}'
2.72	0.87	0.96	2.69	0.87	0.94	2.70	0.90	0.98	2.62	0.92	0.98			
2.61	0.85	0.96				2.57	0.88	0.01						
2.50	0.83	0.96				2.46	0.83	0.03	2.47	0.76	0.00			
2.40	0.80	0.99	2.39	0.79	0.06									
2.31	0.76	0.97	2.30	0.72	0.01							2.28	0.76	0.33
2.22	0.70	0.96										2.21	0.70	0.31
						2.16	0.77	0.94						
						2.07	0.78	0.97						
						2.00	0.83	0.98	2.00	0.71	0.97			

spectra for them to describe a relationship between periodicities which are apparent in individual series. For example, the coherence peaks at 4.3 years between C and Loch Leven Sluices, and C and Marchmont House occur near minor power spectrum peaks in Loch Leven Sluices and Marchmont House but the C spectrum has a trough at this period.

High coherence estimates occur between P and "West" rainfall stations at 2.7 years. Of the records with such coherence with P, those of Arisaig House and Portree have power spectrum peaks at this period while Greenock and North Craig Reservoir have flat spectra around this period. As the P spectrum itself has a peak at this period it may be considered that there may be a 2.7 year periodicity present in P and "West" rainfall records. The oscillation in P index lags behind that in the rainfall series by one to two months in the case of North Craig Reservoir and Greenock, and by an insignificant amount compared to the confidence limits of the phase angle in the case of Arisaig House and Portree.

Coherence values were investigated for possible relationships between oscillations in P and "West" rainfall of period between five and ten years, near the 7.4 year P power spectrum peak. It was not certain whether this latter peak was produced by the same effects as produced the 8.9 year Greenock peak as coherence estimates were not significant. The coherence square of 0.83 and phase angle of 0.02 of a circle at 6.9 years may express effects of some common minor disturbance in these series.

While "East" stations tend to have variations consistent with those of C, and "West" stations with those of P, Wick does not fall into either category. Neither, as in the case of the other stations,

Table 5.11 Coherence peaks between S or M and annual rainfall

Period in years (coherence square, phase in fractions of a circle of index relative to rainfall)

EDINBURGH	S	∞ (1.00,0.98) 74.1(0.93,0.97							
	M								3.09(0.78,0.04)
LOCH LEVEN SLUICES	S	∞ (1.00,0.57)	24.7(0.85,0.06)	12.3(0.73,0.82)		7.41(0.86,0.84)	4.63(0.73,0.95)		
	M	∞ (1.00,0.75) 74.1(0.75,0.28)			8.19(0.81,0.31)				3.09(0.75,0.07)
MARCHMONT HOUSE	S					7.00(0.76,0.00)		3.89(0.78,0.94)	
	M							3.89(0.89,0.74)	
CROMBIE RESERVOIR	S								
	M				8.00(0.71,0.17)				2.46(0.85,0.69) 2.13(0.72,0.35)
BALMORAL	S					6.00(0.70,0.24)			
	M								2.22(0.72,0.28) 2.14(0.80,0.27)
GORDON CASTLE	S		23.4(0.82,0.27)						
	M								
WICK	S						4.93(0.80,0.00)		
	M							3.56(0.74,0.11)	
PORTREE	S	41.8(0.96,0.08)							
	M								

/...

Table 5.11 continued

ARISAIG HOUSE	S				2.77(0.77,0.81) 2.57(0.72,0.78)
	M		18.0(0.70,0.35)		
GREENOCK	S	(0.96,0.04) 62.1(0.90,0.00)			
	M	(0.93,0.53)		4.43(0.75,0.60)	
NORTH CRAIG RESERVOIR	S	59.9(0.73,0.16) 3.33(0.76,0.14) 2.00(0.72,0.07)			2.62(0.75,0.76)
	M				

are there any significant coherence estimates between Wick and S or M which clearly correspond to power spectra peaks. There is a coherence peak at 4.9 years between S and Wick, but neighbouring power spectrum peaks occur at 4.4 years in Wick and 4.0 years in S.

5.4.5 Eigenvector analysis

5.4.5.1 The analyses

In order to try further to correlate variations in different time series and to isolate common periodicities and trends, eigenvector analysis was performed to derive time series describing common variations in individual records; these derived series were subsequently analysed in a similar way to the original.

The first two eigenvectors of real data and the first three eigenvectors of normalised deviation data, using S-mode analysis as described in section 5.3.4 appear in Table 5.12. The first three normalised deviations were considered to be of importance from the relative values of the eigenvalues. Together they explain 67 per cent of the variance in the anomaly data.

The first real eigenvector accounts for most of the variance in the real data, and its elements have a 0.99 correlation with station mean annual rainfall. Thus it does not increase knowledge about rainfall spatial variations. The derived time series of eigenvector multipliers represent mean annual rainfall variations over the eleven stations. These results are similar to those of Chapter 2 where the analysis of a real mean monthly rainfall matrix produced a first eigenvector describing the annual cycle in mean monthly rainfall and an associated set of multipliers describing the spatial variations in mean annual rainfall.

Table 5.12 Eigenvectors (S-mode) of Scottish rainfall 1890-1973

	Real data		Normalised deviation data			Mean annual rainfall	Annual rainfall standard deviation
	first	second	first	second	third		
North Craig Res.	0.328	0.135	0.356	0.291	-0.010	1125.7	173.1
Greenock	0.443	0.258	0.373	0.307	-0.163	1153.3	229.1
Arisaig House	0.458	0.418	0.277	0.443	-0.097	1603.5	211.0
Stornoway	0.341	0.298	0.249	0.381	0.178	1189.4	178.3
Wick	0.220	-0.071	0.300	0.076	-0.547	763.8	91.9
Gordon Castle	0.212	-0.325	0.210	-0.295	-0.561	751.7	103.1
Balmoral	0.240	-0.249	0.336	-0.195	0.081	843.5	122.1
Crombie Res.	0.253	-0.481	0.302	-0.410	0.207	894.3	155.5
Loch Leven Sluices	0.261	-0.326	0.387	-0.235	0.248	908.0	144.3
Marchmont House	0.188	-0.091	0.013	0.015	-0.448	669.4	120.5
	0.228	-0.365	0.329	-0.355	0.083	816.6	141.5
Percentage variance	98.7	0.5	33.7	23.8	9.5		
Cumulative percentage variance	98.7	99.2	33.7	57.5	67.0		

The normalised deviation eigenvectors, on the other hand, do not appear to be correlated to station mean annual rainfall or its standard deviation. The second eigenvector does seem to show some effect of station position, having positive elements for "West" stations and negative elements for "East" stations (apart from Wick and Edinburgh which have small positive and very small positive values respectively). Arisaig House, the station which might be expected to show the greatest effect of a West coast position, has the largest positive element, while Crombie Reservoir, whose position is farthest East, has the largest negative element. Crombie Reservoir also appears in the analyses above to have rainfall time variations which are least similar to those of "West" stations and most similar to those of C index. Using similar criteria, Stornoway and Marchmont House might be judged to be the second most Westerly and Easterly stations respectively, (a fact which seems to be reflected in their second eigenvector elements). Thus the second eigenvector seems at first sight to show an inherent property in the normalised deviation matrix which can be related to station position in relation to prevailing wind, while other eigenvectors do not appear immediately to reflect known effects on rainfall.

5.4.5.2 "Low-pass" filtered eigenvector multipliers

The first set of real eigenvector multipliers (1E) and the first three sets of normalised deviation eigenvector multipliers (1ON, 2ON, 3ON), considered as time series of length 84 years, were filtered using the filter of 5.3.2. The filtered data appear in Figure 5.12. 1E shows the mean annual rainfall variations

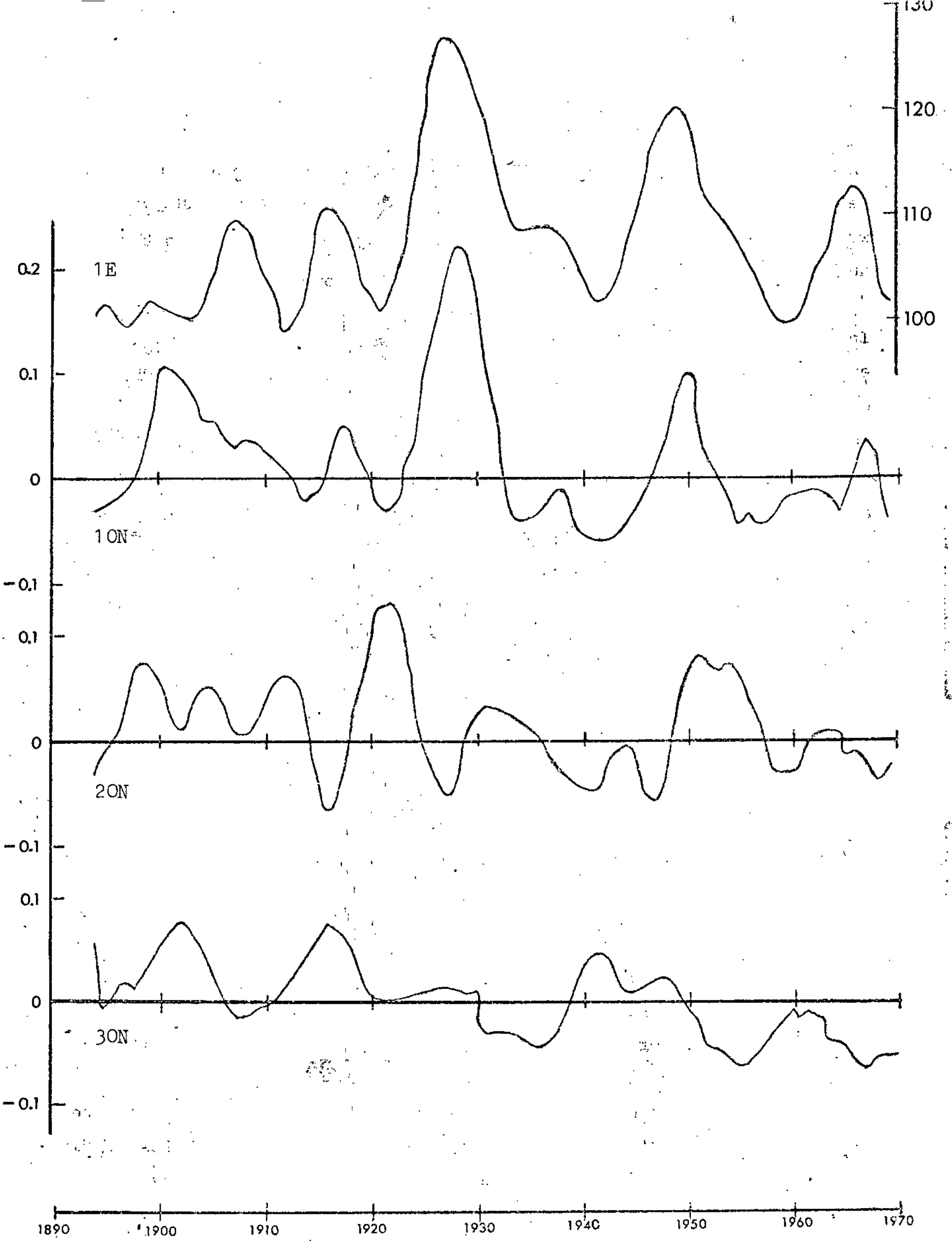

Figure 5.12 "Low-pass" filtered eigenvector multipliers

with peaks in 1907, 1916, 1925, 1949 and 1966, and troughs in 1912, 1921, 1942 and 1960. These features, apart from the 1907 peak occur to a greater or lesser extent in all stations above in Figures 5.5, 5.6 and 5.7 especially the "East" stations. 1ON resembles 1E apart from over the first 20 years where its major peak occurs in 1910. Both 1E and 1ON curves have similarities to that of C (Figure 5.3) with C leading each series in the early part of the record and lagging in the latter part. While the peak in C in 1913 appears to correspond to peaks in 1ON and 1E in 1917, and the peak in C in 1925 to those in 1ON and 1E in 1927, the 1952 C peak occurs after the 1949 1ON and 1E peaks.

The filtered 2ON series resembles the filtered P series (Figure 5.3) with common peaks in 1897, 1905, 1912, 1922 and 1950, and common troughs in 1893, 1909, 1917, 1940, 1947 and 1959. 3ON shows trend and its rather flat peaks and troughs do not appear to represent those of individual rainfall records.

Thus time series of eigenvector multipliers show some common variations in individual rainfall series which are also present in circulation indices. As there are more "East" than "West" stations used in the eigenvector analyses the most significant variations revealed by 1ON and 1E correspond most closely to those of "East" stations. 2ON shows some of the common variations found in "West" stations and P.

5.4.5.3 Power spectra of eigenvector multipliers

Figure 5.13 shows eigenvector multiplier power spectra together with those of P and C, each calculated from 84 years of data to a maximum lag of 28 using the ACV method. Peaks signifi-

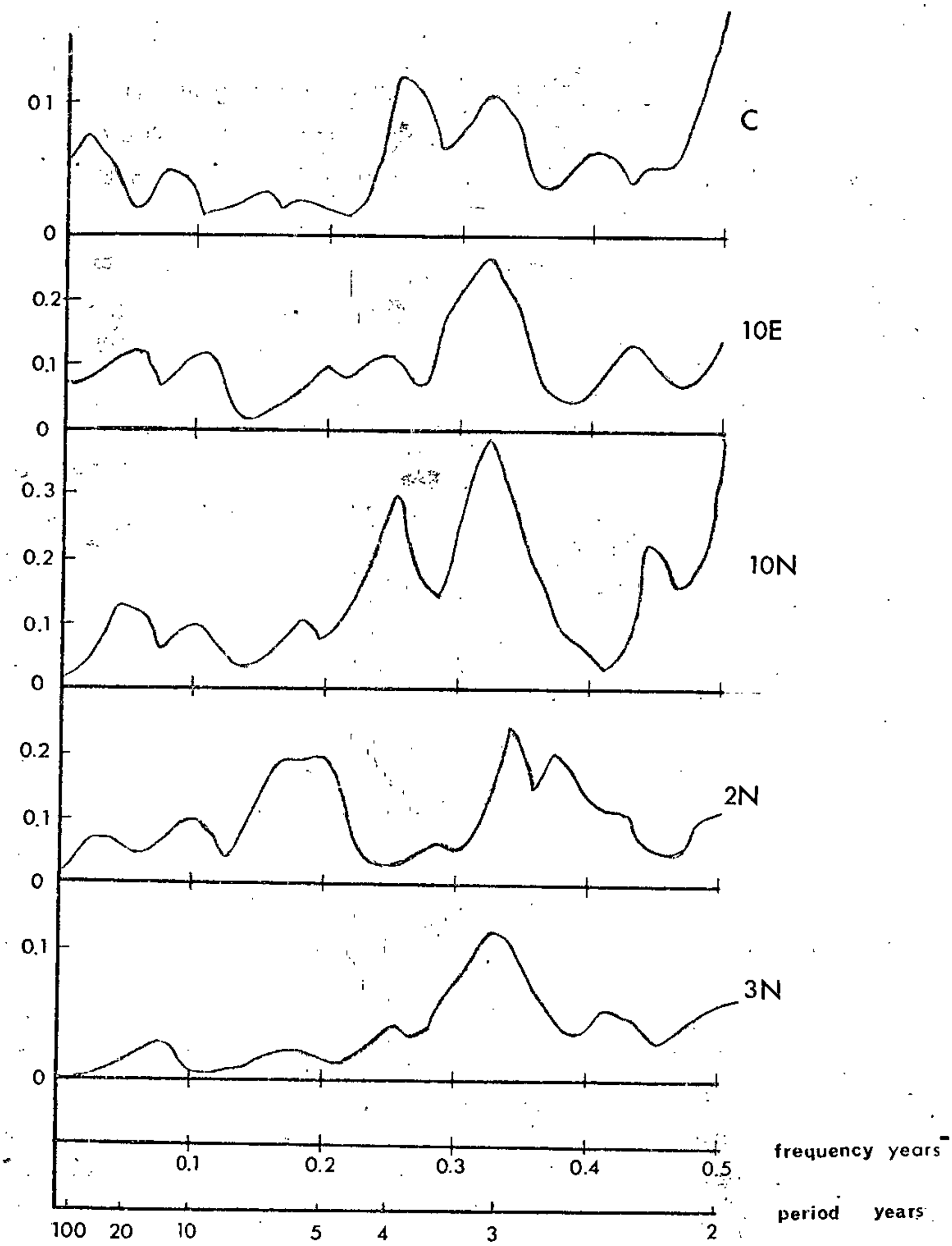

Figure 5.13 Power spectra - P, C, and eigenvector multiplier series

cant at the five per cent level for a population of similar spectra occur near three years: at 3.1 years in 1E, 10N, 30N which corresponds to the period of the oscillation in C, most "East" stations and Wick and Greenock; and at 2.9 years in 20N with a subsidiary peak at 2.7 years which corresponds to a 2.7 year peak in P, Arisaig House, and Portree.

The presence of the quasi-biennial oscillation in the rainfall data is confirmed in the spectra of 10N though some of the rising power may again be a result of aliasing of power from non-resolvable high frequencies. This oscillation is also present in P and C. The peak at four years in 10N is probably a harmonic of this two year cycle.

The broad 5 to 6 year peak in 20N is not close enough in period of the 7.9 year P peak to describe the effect of a common variation. The broad peak however does describe some features of "West" station power spectra and hence some of "West" rainfall variations. There is a similar broad peak in Greenock, Portree and Arisaig House, and a sharp peak at 5.8 years in North Craig Reservoir.

Thus the suggestions that 10N and 1E describe common rainfall variations especially those of "East" station rainfall, and that 20N describes "West" station rainfall is confirmed. Further investigation of the relationship between these series and P and C index series is made using cross spectrum analyses.

5.4.5.4 Cross spectra - P and C, and eigenvector multipliers

Coherence and phase spectra between C and 10E, 10N, 20N,

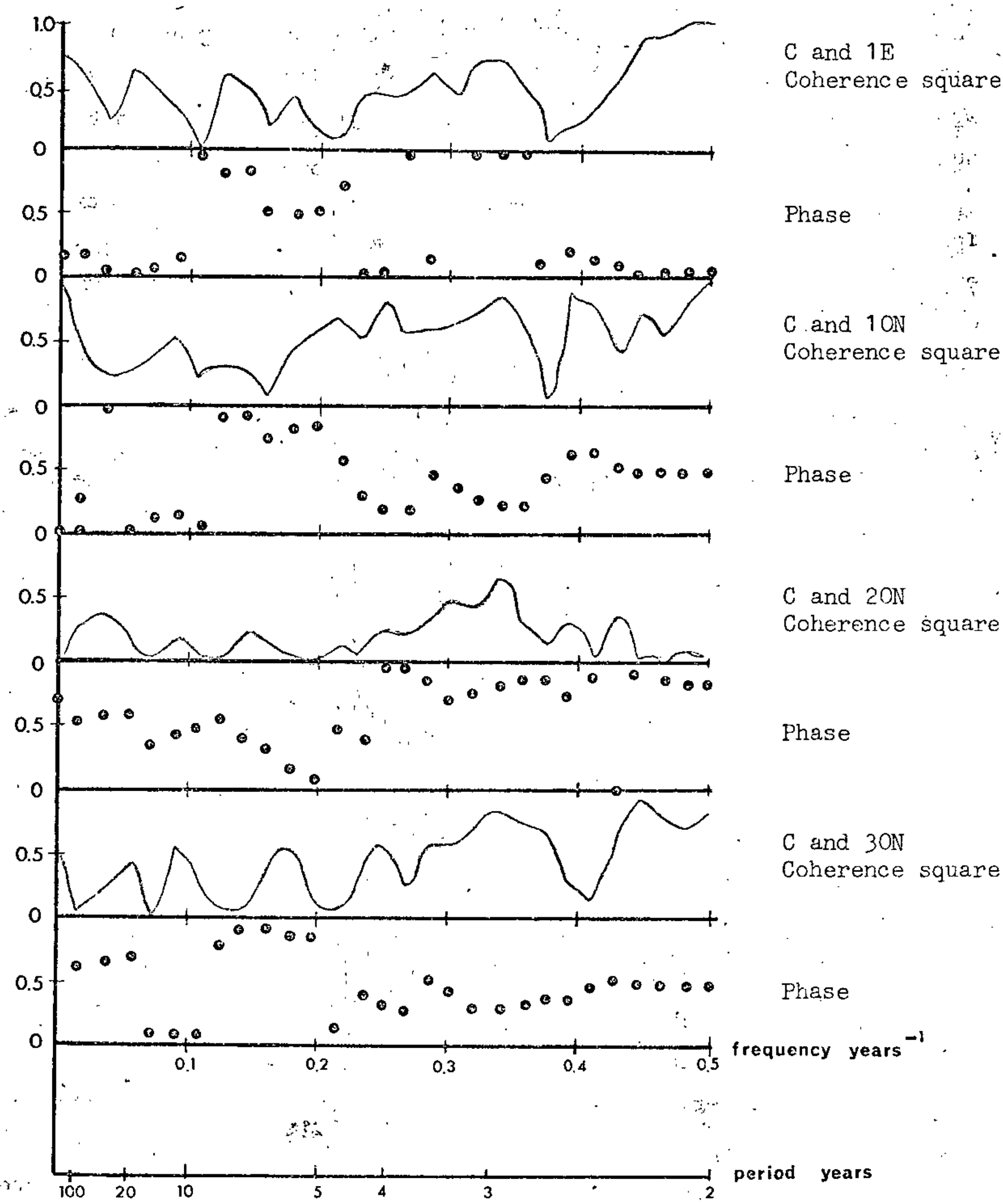

Figure 5.14 Coherence and phase spectra between C and eigenvector multiplier series

and 30N appear in Figure 5.14. Confidence limits on coherence and phase are as given for annual spectra in 5.3.4.3 and these limits are used to assess the significance of these estimates.

The quasi-biennial oscillation which is present in 10N and is possibly present in 1E (Figure 5.13) has a very high coherence value with that of C. The phase angle and its confidence limits suggest that 1E and C are exactly in phase while 10N and C are in anti-phase. Since aliasing of power may again affect cross spectrum estimates in addition to power spectrum estimates near the Nyquist frequency, the exact significance of these coherence estimates and the precise frequency of the quasi-biennial oscillation cannot be determined.

The oscillations found at 3.1 years in C and 10N have significant coherence with one another though the phase relationship between the two oscillations which is one of quadrature, has large uncertainties. The phase lead of 10 months in the C oscillation lies between 0 and 18 months at 90 per cent confidence limits. Neighbouring coherence peaks between C and 10N at 2.9 and 2.5 years also have associated large uncertainties in the phase angle of cross spectrum estimates. The other coherence peaks between C and 10N, which occur at four years, with an uncertain phase relationship between the oscillations, and at zero frequencies, appear to be of little importance.

The 20N and C coherence spectrum shows no common variations in 20N and C while that of 30N and C has high values at and near the three-year peak in their power spectra, near the troughs in power spectra at 2.2 years, and at 2.0 years. The phase lead of C over 30N for the three-year period, which is one of primary interest, is about one year but could lie between 11 and 17 months at 90 per cent confidence limits.

There are few significant values in the coherence spectra between P index and eigenvector multipliers, these estimates not being shown here. The quasi-biennial oscillation, present in P, produces significant coherence with 10N (COHSQ = 0.90, Φ_{xy}' = 0.49) with an anti-phase relationship and almost significant coherence with 1E (COHSQ = 0.83, Φ_{xy}' = 0.98) with an almost in-phase relationship. These phase relationships are the same as those between C and 10N, and C and 1E.

30N and P have high coherence only at zero frequency (COHSQ = 0.99, Φ_{xy}' = 0.97). The only relationship between P and 20N which is apparent in coherence estimates occurs at 5.1 years (COHSQ = 0.86, Φ_{xy}' = 0.15). There are no coherence peaks at 2.9 or 2.7 years where spectral peaks occurred in 20N and P. There is a broad peak in 20N over the range of periods five to six years which describes some features in individual power spectra of "West" stations as mentioned above. However, as the coherence estimates for the range of periods five to six years are only just significant, and as the phase angle at 5.1 years has large uncertainties ($0.08 < \Phi_{xy}' < 0.18$ at 90 percent confidence limits), the use of these coherence estimates to describe the relationship between P and "West" rainfall variations will be limited.

Thus cross spectrum analyses between eigenvector multipliers and circulation indices provide further means of identifying the two sets of first eigenvector multipliers as describing common time variations in all rainfall series, especially those in "East" stations and those present in C. However, these analyses provide little further information about the second and third sets of eigenvector multipliers and their relationships to P and C.

5.5 Analysis of monthly rainfall time series

5.5.1 Power spectrum analysis

In order to investigate further high frequency oscillations in rainfall and circulation indices, power spectra of monthly rainfall and indices were calculated using the Blackman-Tukey autocovariance approach and the fast Fourier transform method. Data series of 984 values for 1890-1971 with a lag of values being determined by the limits of the BMD program of a maximum of 1000 data points and a maximum lag of 199. Series of 1020 data points for 1886-1971 were extended to 1024 by addition of zeros and their fast Fourier transforms calculated, in the case of Arisaig House where data were not available prior to 1890, 984 points were used.

The 199 spectral estimates produced by the ACV method and the 102 FFT estimates produced by the summation over elementary bands of width five are plotted in Figure 5.15 for P and C indices, and in Figure 5.16 for typical "East" and "West" stations, Crombie Reservoir and North Craig Reservoir.

As discussed in section 5.3.3 the number of degrees of freedom of each set of estimates is approximately the same though a given FFT estimate will have smaller confidence limits than a given ACV estimate as there are fewer FFT estimates. The FFT spectra being plotted from fewer estimates appear to be smoother than ACV spectra, but sets of spectra are very similar. Minor peaks in FFT spectra are in most cases less pronounced than those in ACV spectra, and

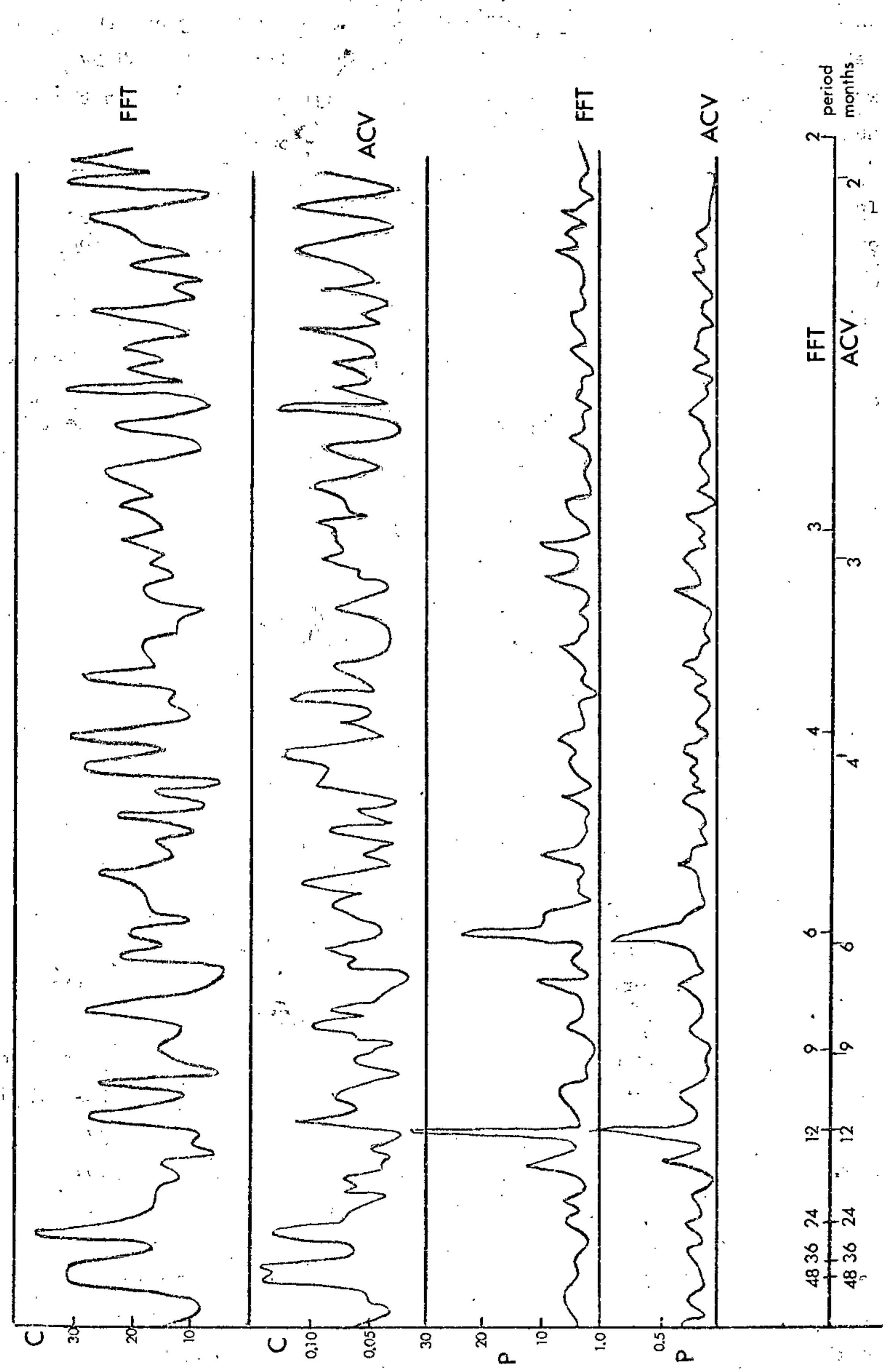

Figure 5.15 Power spectra - monthly C and P indices using FFT and ACV methods

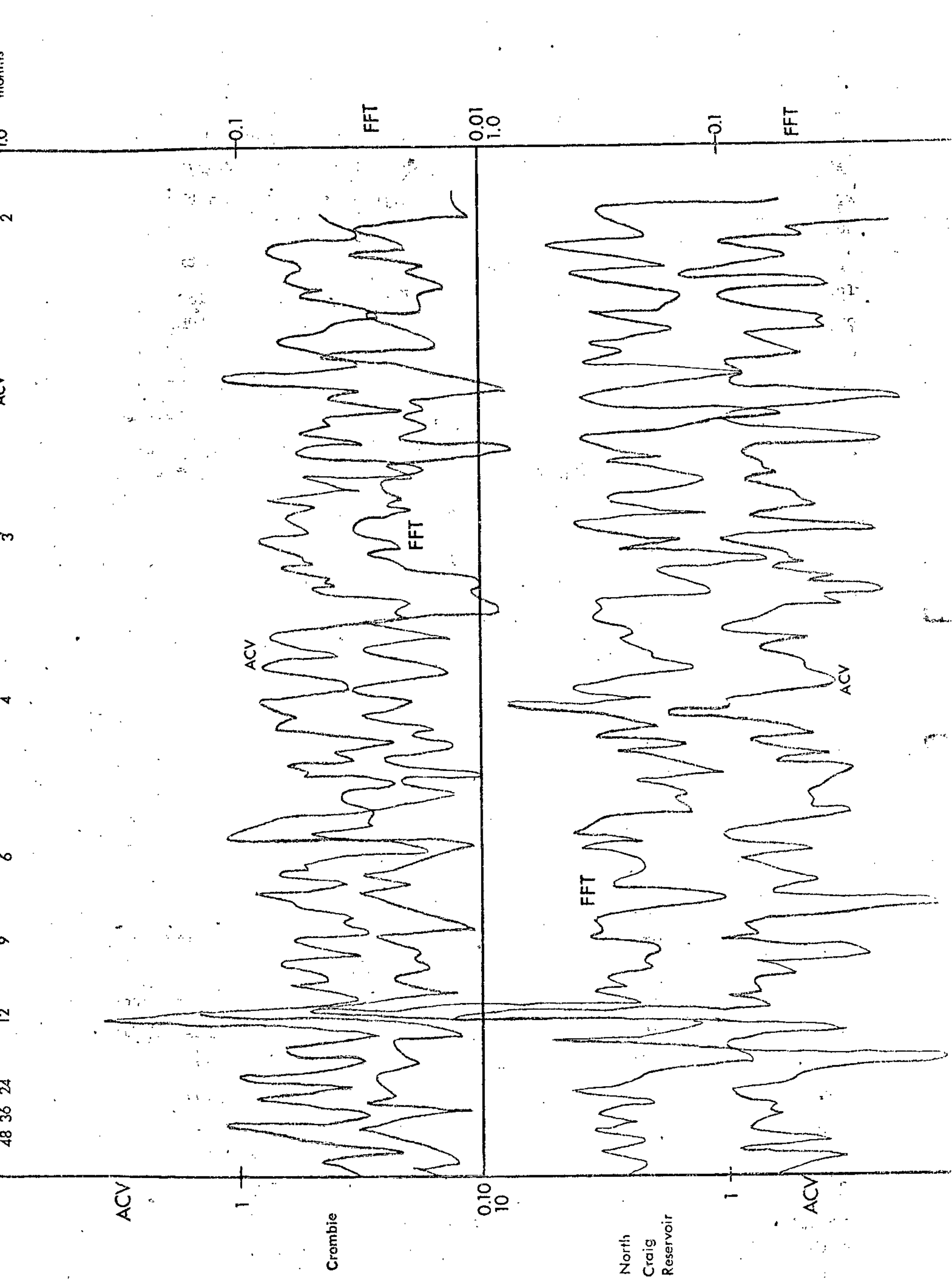

Figure 5.16 Power spectra - Crombie Reservoir, and North Craig Reservoir using FFT and ACV methods

some minor peaks are more clearly identifiable as side bands of major peaks. However, the extent to which real effects as well as noise are smoothed out in the FFT spectra compared to the ACV spectra is not immediately obvious. Most of the peaks in either spectra are in any case of little significance when tested using the $\frac{\chi^2}{\nu}$ distribution as in section 5.3.

The principal peaks of period greater than 11 months are listed in Table 5.13. All the rainfall spectra and the P index spectra have peaks which are significant for individual spectra at 12 months. In the case of P, and of some rainfall stations, the asymmetry of the annual rainfall cycle effects a peak at six months. The C index does not have a 12 month peak but an 11 month peak which may describe the effect of an annual cycle.

The 14.7 and 14.4 month peaks, some of which are significant at the five per cent level for a population spectrum, could be side-bands of the major 12 month peaks. A $14\frac{2}{3}$ month peak was however found in Brunt's (1925) analysis of Edinburgh rainfall and was the most significant periodicity in 10 out of the 12 European rainfall records, for the period 1760 to 1925, which he examined. Brunt also found oscillations of period 18 months in London and Edinburgh rainfall which is also observed in Edinburgh and three other records listed in Table 5.12. Brunt's Edinburgh data could be considered to belong to the same population of data as that studied here. His results are therefore mentioned at this point as they increase the significance of the results listed in Table 5.12. Other periodicities in this frequency range found in other meteorological records are discussed in Section 5.5.

In Table 5.13 the quasi-biennial oscillation and the three

Table 5.13 Major peaks in power spectra of monthly rainfall and circulation indices

Periods are given in months - 1. using the autocovariance method, 2. using fast Fourier transform method. Significant values at five per cent level of a population spectrum are underlined, and those at five per cent level of a sample spectrum are doubly underlined.

Marchmont House	1					36.2						14.7		$\underline{\underline{12.1}}$	
	2					39.4						14.4		$\underline{\underline{11.9}}$	
Edinburgh	1					36.2					18.9	14.7		$\underline{\underline{12.1}}$	
	2					39.4					18.3	14.4		$\underline{\underline{11.9}}$	
Loch Leven Sluices	1				49.7	$\underline{36.2}$					18.9			$\underline{\underline{12.1}}$	
	2		64.0											$\underline{\underline{11.9}}$	
Crombie Reservoir	1					36.2			23.4		18.9	14.7		$\underline{\underline{12.1}}$	
	2					39.4								$\underline{\underline{11.9}}$	
Balmoral	1					36.2			23.4			14.7		$\underline{\underline{12.1}}$	
	2					39.4						14.7		$\underline{\underline{11.9}}$	
Gordon Castle	1					36.2			23.4		18.9	14.7		$\underline{\underline{12.1}}$	
	2					39.4			22.2				13.4	$\underline{\underline{11.9}}$	
Wick	1	79.6						28.4	22.1		18.9	14.7		$\underline{\underline{12.1}}$	
	2											14.4		$\underline{\underline{11.9}}$	
Stornoway	1			$\underline{56.8}$				$\underline{28.4}$				$\underline{14.7}$		$\underline{\underline{12.1}}$	
	2							$\underline{28.4}$				$\underline{14.4}$		$\underline{\underline{11.9}}$	

Table 5.13 continued

Arisaig House	1		66.4				30.6		23.4			14.2		12.1	
	2									20.1		14.4		11.9	
Greenock	1		66.4			36.2				19.9		14.2		12.1	
	2		64.0							20.1		14.4		11.9	
North Craig Reservoir	1		66.3			36.2			23.4			14.7		12.1	
	2								22.2			14.4		11.9	
P	1													12.1	
	2											14.4		11.9	
C	1					36.2			23.4						11.1
	2					39.4			22.2			14.4			11.2

year oscillation observed in annual power spectra produce the most significant peaks other than that of 12 months. Both these periodicities have significant values at the five per cent level for a population spectrum in the case of the C spectrum. In other spectra the oscillation of period around three years appears to be more common than the one of period around two years. The fact that the quasi-biennial oscillation is less apparent in monthly spectra than annual spectra and that the peaks occur at 22 to 23 months in monthly spectra, shows that some aliasing of power occurred near two years in the annual spectra.

Of the lower frequency peaks in the monthly spectra, those of Arisaig House and Greenock at 64 to 66 months correspond to their broad annual spectrum peak at 5.0 to 5.5 years, while that of Stornoway at 56 months represents a harmonic of its 28 month oscillation. Stornoway's spectra is not in itself similar to those of other series, partly on account of its inhomogeneities.

5.5.2 Cross spectrum analysis

Coherence peaks and phase angles of cross spectrum estimates calculated using the CCV and FFT methods are given in Table 5.14 for cross spectra between C and monthly rainfall and in Table 5.15 for those between P and monthly rainfall. Confidence limits of coherence and phase estimates are given in Tables 5.3 and 5.4 respectively.

The quasi-biennial oscillation in C produces significant coherence values with "East" rainfall series even in cases where the oscillation is not apparent in individual power spectra. The phase angles of the cross spectrum estimates suggest that the oscillations in C and rainfall could be in phase, except in the case of Wick and

Table 5.14 Cross spectra coherence peaks between C and monthly rainfall

Period (coherence, phase of rainfall oscillation relative to C index in fractions of a circle)

Marchmont House	1		26.6(0.71,0.97) 24.8(0.87,0.01) 23.4(0.94,0.03)			11.1(0.70,0.98)	
	2		25.0(0.90,0.99) 22.3(0.92,0.03)				
Edinburgh	1		26.6(0.72,0.96) 24.8(0.88,0.97) 23.4(0.76,0.98)				
	2		25.0(0.85,0.96)				
Loch Leven Sluices	1						
	2		25.0(0.85,0.97)			-	10.1(0.86,0.89)
Crombie Reservoir	1	36.1(0.81,0.98)	23.4(0.72,0.00) 22.1(0.73,0.99)	18.9(0.70,0.99) 18.1(0.73,0.00)	12.8(0.78,0.03)	11.1(0.74,0.95) 10.2(0.80,0.90) 9.5(0.74,0.89)	
	2						10.1(0.89,0.91)
Balmoral	1		24.8(0.73,0.01) 23.4(0.76,0.02)	18.9(0.74,0.95) 18.1(0.80,0.93)			
	2		25.0(0.78,0.02)	16.8(0.86,0.04)			
Gordon Castle	1		23.4(0.71,0.95)				
	2		25.0(0.71,0.91)				
Wick	1				12.8(0.73,0.01)		
	2		25.0(0.74,0.87)				
Stornoway	1				12.8(0.70,0.03)		
	2					11.9(0.88,0.08)	

Table 5.15 Cross spectra coherence peaks between P and monthly rainfall

Period (coherence, phase of rainfall oscillation relative to C index in fractions of a circle)
1. CCV, 2. FFT methods

Marchmont House	1					12.1(0.87,0.93)	
	2					11.9(0.87,0.94)	
Edinburgh	1					12.1(0.90,0.88)	
	2					11.9(0.87,0.89)	
Loch Leven Sluices	1					12.1(0.93,0.99) 11.7(0.82,0.00)	
	2					11.9(0.95,0.99)	
Crombie Res.	1					12.1(0.83,0.97)	
	2					11.9(0.82,0.97)	
Balmoral	1					12.1(0.92,0.03)	
	2					11.9(0.90,0.03)	
Gordon Castle	1					12.1(0.87,0.90)	
	2					11.9(0.82,0.90)	
Wick	1					12.1(0.84,0.07)	
	2					11.9(0.84,0.02)	
Stornoway	1	36.2(0.75,0.08) 33.3(0.76,0.04)	26.6(0.72,0.04)		14.7(0.87,0.00) 14.2(0.81,0.00)	12.1(0.92,0.08) 11.7(0.82,0.08)	11.1(0.82,0.03) 10.7(0.80,0.06)
	2		28.4(0.88,0.05)	22.2(0.82,0.96)			11.9(0.88,0.08)

/...

Table 5.15 continued

Arisaig House	1	36.2(0.86,0.03) 33.3(0.89,0.03)	24.9(0.86,0.96) 23.4(0.81,0.97)	18.9(0.84,0.00) 18.1(0.88,0.98)	14.7(0.84,0.00) 14.2(0.94,0.00)	12.4(0.83,0.06) 12.1(0.93,0.01) 11.7(0.86,0.01)	
	2					11.9(0.93,0.03)	
Greenock	1		26.6(0.74,0.04) 24.9(0.83,0.99)			12.1(0.91,0.09) 11.7(0.83,0.09)	11.1(0.77,0.01)
	2					11.9(0.90,0.08)	
North Craig Reservoir	1	33.3(0.75,0.05)	22.1(0.74,0.03)	18.1(0.84,0.88)		12.4(0.83,0.99) 12.1(0.92,0.00) 11.7(0.83,0.02)	
	2			18.3(0.84,0.89)		11.9(0.92,0.02)	

Gordon Castle whose coherence values with C are also the least significant of the "East" stations.

Oscillations of period around three years which are of significance in the monthly power spectra of C and Loch Leven Sluices, only produce significant coherence estimates between C and Crombie Reservoir. There is also high coherence between C and Crombie Reservoir for periods around 11 months, and Crombie Reservoir seems to have the highest overall coherence with C. This was also the case in the analysis of annual time series.

In the case of P, there are significant coherence values between the annual oscillations in P and in every rainfall station series. The phase lags between P and rainfall vary with station, but have small confidence limits, of less than one month in most cases. "East" stations tend to lag behind P, and "West" stations to lead P. Edinburgh has the longest lag behind P, while Balmoral has an abnormal phase lead for an "East" station. Of the West stations, Stornoway has the longest phase lead over P. The variation in phase relationship between the annual cycle in P and that in rainfall series reflects synoptic experience with progressive systems travelling from West to East.

The lower frequency coherence peaks between P and "West" stations do not in general correspond to peaks in individual power spectra and probably reflect the overall close relationship between variations in P and "West" rainfall.

5.6 The relation of these results to other work

5.6.1 High frequency oscillations - the two and three year periodicities

Comparison has already been made between the results of Brunt's analysis of European weather elements and those of Scottish rainfall. Brunt's significant $14\frac{2}{3}$ month periodicity and his 18 month periodicity were also found in the data analysed above. Gray (1975) lists coincident peaks in the FFT power spectra of Kew pressure, South East England rainfall, Central England temperature, and C index using monthly data, and these peaks may be compared with those of the Scottish rainfall data. All four of Gray's spectra have peaks at 3.36, 3.13 and 1.89 years, i.e. 40.3, 36.6, and 22.2 months. The latter peak corresponds to that in Scottish monthly rainfall data. There is also some evidence in monthly spectra above of two peaks near three years but in the annual spectra only the peak at 3.1 years occurs. It would seem doubtful that in Gray's analysis there are two independent peaks near three years, and probable that one of the 3.36 and 3.13 year peaks is a side-band of the other peak.

Gray also found peaks at 2.96, 2.76, and 2.20 years in all the spectra except that of Central England temperature, and at 1.23 years in all the spectra except that of Kew pressure. The 14 to 15 month Scottish rainfall peaks seem to correspond to those at 1.23 years. There is also evidence in annual spectra of peaks in P and West stations at 2.7 years, and at 2.7 and 2.9 years in the second eigenvector multiplier time series which seems to describe "West" rainfall variations. These peaks may be related to Gray's 2.76 and 2.96

year peaks.

Of the high frequency periodicities found in meteorological parameters, the quasi-biennial oscillations is in a large number of cases the most significant. Shapiro (1975) gives an analysis of Manley's (1975) Central England temperature series which was used by Gray in the above analysis and has been analysed by other workers. Shapiro detected a small but significant peak at 25.5 months in the power spectra of Manley's raw data as well as a very significant annual peak and a six month peak. Data filtered by the use of a 12 month running mean again revealed a 25.5 month peak when the spectrum was compared with a red noise background spectrum. The frequency of the peak in Gray's analysis of the same data was at 22.7 months, showing that different methods of analysis produce slightly different results and the exact frequency of a significant oscillation can be difficult to determine.

Lamb (1972b) gives a survey of periodicities revealed in the analysis of meteorological elements and related data. Of the high frequency oscillations that with period around two years, the quasi-biennial oscillation, occurs most often and is of most significance, though it is less significant than very low frequency oscillations of period 100 years or more. Lamb also discusses the presence of the quasi-biennial oscillation in the winds of the lower stratosphere which reverse direction from West to East. Wind circulation derived from pressure anomalies and the frequency of blocking types also show this oscillation. Oscillations occurring in the Northern and Southern Hemisphere have been shown to be linked by the successful correlation of pressure anomalies in the two Hemispheres.

The quasi-biennial oscillation in circulation indices referred to above may be linked to those of pressure anomalies describing the circulation over the Northern Hemisphere, and indirectly linked to the winds in the equitorial stratosphere. The P and C indices in their turn describe the frequency of weather types affecting rainfall. The C index describes the frequency of cyclonic circulation which produces a large proportion of rainfall at Scottish stations, especially "East" stations. Thus the quasi-biennial oscillation present in rainfall over the British Isles, circulation over the British Isles, circulation over the Northern Hemisphere, and winds in the stratosphere may be linked and correlated. On the other hand, the three-year rainfall oscillation can only be related to that of C and is not found extensively in other meteorological time series.

5.6.2 General variations in circulation and rainfall over 100 years

No medium or low frequency oscillations which were of definite significance were found in the rainfall or circulation indices series above, though similar variations were found in filtered records. Rainfall records can also be divided into epochs which correspond to those in which a certain type of circulation predominates. Lamb (1972a) summarises the over-all variation in the frequency of Westerly and blocking types over the period 1861 to 1971 into the periods 1861 to 1874 and 1900 to 1954 with a marked prevalence of Westerly types, and 1875 to 1899 and 1955 to 1971 with a marked prevalence of blocking types. Lamb's frequency of Westerly types corresponds to the P index, and his epochs can be seen in the decadal means of P index in Table 5.7 and filtered P index in Figure 5.3. These epochs can also be identified in C index which shows similar variations to the P index with a lag of about five years.

Individual Scottish rainfall records exist only prior to 1880 for "East" stations, and the period between 1871 to 1880 can be identified as one of high "East" rainfall and C index. The first half of the 20th century can be recognised as being relatively wet in most filtered records in Figures 5.5 to 5.7 and in the decadal means of "West" stations. This period is one of high P and C indices associated with the prevalence of Westerly types. The exact period of the wet epoch varies between stations. Thus Lamb's classification of periods of years according to the predominance of a circulation type can be recognised in rainfall records.

Schove (1950) studied variations in temperature, rainfall, and wind for the period 1875 to 1925 using overlapping 30-year periods; variations in mean values for these sub-periods may be compared to those of Scottish rainfall and circulation indices. Schove suggested that rainfall and temperature anomalies should be discussed in terms of an "area" term depending on the pressure anomaly and a "local" term depending on the wind anomaly. These anomaly terms have similarities to the C and P indices, and would be of particular importance in the determination of rainfall of sheltered stations and of stations exposed to the prevailing wind, respectively.

Schove's description of different climatic phases in terms of wind and pressure anomalies expresses rainfall for overlapping 30-

year periods as a percentage of the 1901 to 1930 mean for the "East" sites of York and Edinburgh, the "area" sites of Greenwich and Oxford, and for all available "England" sites. The "East" sites have similar variations to those of "East" Scottish rainfall as described by groups of decadal means in Figure 5.2 or in sections of filtered data records in Figures 5.5 and 5.6. Available "West" Scottish rainfall for the period 1880 to 1930, shows similar variations to those of "England" data. Individual decadal means for different sets of data as against 30-year means do not however show similar variations. The only similarities between Schove "England" decadal means for periods between 1850 and 1940 and those of Scottish data are the very wet decades 1871 to 1880 and 1921 to 1930.

5.6.3 Other methods of spectral analysis to resolve low frequencies

While common variations were found in the filtered records, and to a lesser extent in decadal means, no common low frequency oscillations were identified in the power spectrum analyses using the above methods. Neither were individual lower frequency peaks of significance in themselves. An increase in spectral resolution might provide further information about low frequency oscillations. Using the spectral methods of the Blackman-Tukey autocovariance method and of summation of FFT estimates of filtered data above, the number of final estimates could be increased to produce high resolution by increasing the maximum lag, or decreasing the bandwidth of summation respectively. However, the variance of such estimates would be unacceptably high.

Recently new methods have been suggested which produce a greater number of low frequency estimates, and some of these methods have been shown to give stable results. Gray (1975) suggested a perturbation technique to increase the stability of raw Fourier transform estimates providing additional information at low frequencies. 'Ordinary' Fourier transforms are calculated for a given harmonics at slightly different frequencies by changing the length of data used and estimates for a given harmonic are averaged over a series of perturbation. As frequency shifts for each perturbation are dependent on the harmonic number, being smallest for the lower harmonics, the number of perturbation estimates which may be averaged is largest for low harmonics and low frequencies. This method is thus suited to low frequencies and has been shown to give stable results.

However, this approach cannot be used in cross spectrum analysis and is also unsuited to short series, such as those of annual rainfall data above. In the case of monthly rainfall series it did not seem likely that further information could be obtained at low frequencies; raw fast Fourier components did not suggest the presence of low frequency peaks which could be investigated further by use of 'ordinary' Fourier transforms and the perturbation technique.

In the last few years the maximum entropy method (MEM) which has high resolution at low frequencies, has been applied to short time series of physical data. MEM is "data adaptive" in that the method of filtering and smoothing of the estimates is determined by the noise characteristics of the series under study, and is not of predetermined form as in other methods discussed above. Spectral estimates produced by MEM have a higher variance than those of other methods but their significantly higher resolution is capable of providing information on periodicities of wavelength equal to the length of the data. A full discussion of the method and results of numerical experiments on synthetic time series using a maximum entropy routine is given in Ross (1975).

The statistical significance of MEM estimates has not been extensively considered, and the high variance of such estimates suggested that MEM spectral peaks should be related to the results of other analyses before being accepted as revealing oscillations in the data.[1]

5.7 Summary

In this chapter general variations in rainfall and circulation indices have been investigated using decadal means and "low-pass" filtering of the data, and the presence of periodicities has also been investigated using power spectrum analysis. The relationship between oscillations in circulation indices and rainfall has also been investigated using cross spectrum analysis. An attempt has been made with some success, using eigenvector analysis, to describe common temporal variations by a series of independent time series whose relative presence in a given rainfall record can be defined.

Most of the rainfall records appear to be homogeneous when their decadal means are compared with their long-term means, and

---

Note 1: A subsequent analysis of the Scottish annual rainfall series using MEM, produced similar results to those using the Blackman-Tukey autocovariance approach in section 5.4.3 (see Appendix 2).

with those of other stations. They do not show trend to any great degree, though the third eigenvector time series, which describes some nine per cent of the variance in the annual data does show trend. Persistence in the time series as revealed in the power spectra is neither extensive nor of a simple Markov type. In most cases the power spectrum may be described as being that of white noise with some disturbances.

Of these disturbances it is only those which appear in the power spectra of annual data at high frequencies which are of significance, i.e. high frequency periodicities are the only non-random element in rainfall time series.

While common variations exist in low-pass filtered records, decadal means, and longer sub-periods of the records, there are no common definable medium or low frequency oscillations revealed in the power spectra. The extent to which high frequency periodicities present in the power spectra represent real processes is not easy to determine. The annual cycle and the quasi-biennial oscillation could be expected on a priori grounds. The significance of the 3.1 year periodicity in C and rainfall, expecially in "East" Scottish rainfall and the first eigenvector time series, and the significance of the coherence estimates and in-phase relationship between the 3.1 year periodicity in C and the rainfall series, is sufficient to suggest that a 3.1 year periodicity represents a real process affecting C index and rainfall.

The overall relationship between indices and rainfall has been demonstrated in cross spectrum analysis. "West" stations tend to have high coherence with P, and "East" stations with C, though variations in P and C are to some extent related. The annual cycle in P is also closely related to those of all rainfall records, leading those of "East" stations and lagging those of "West" stations by less than one month in either case.

In future work, further investigations of low frequency variations could be made using other methods of spectral analysis though in the absence of fresh data it does not seem likely that further inherent patterns of rainfall series can be revealed. It would seem from this analysis that the two- and three-year periodicities are the most important non-random elements present in Scottish rainfall time series and in circulation indices. In order to investigate their use in "prediction" of future rainfall, "band-pass" filters would have to be applied to monthly rainfall records to eliminate effects of all oscillations other than the two- or three-year periodicity. The phase of each periodicity in each record could then be determined.

APPENDIX 1

AN INVESTIGATION INTO SPELLS OF WET AND DRY DAYS BY REGION AND SEASON FOR GREAT BRITAIN

Reproduced from the

METEOROLOGICAL MAGAZINE

Vol. 104, 1975, pp. 360–375

AN INVESTIGATION INTO SPELLS OF WET AND DRY DAYS BY REGION AND SEASON FOR GREAT BRITAIN

By J. A. BLAIR-FISH

551.577.22(410)

AN INVESTIGATION INTO SPELLS OF WET AND DRY DAYS BY REGION AND SEASON FOR GREAT BRITAIN

By J. A. BLAIR-FISH
(Meteorology Department, University of Edinburgh)

Summary. Distribution of wet and dry spells are considered in relation to Markov, simple logarithmic, modified logarithmic, and modified geometric models explained in this paper. Data from eight stations distributed over the British Isles show that a simple logarithmic model can usually describe dry-spell data while the modified logarithmic and geometric models describe wet-spell data. The variations in model parameters do not correlate well with region. Data considered by season for Oxford show that, on average, autumn and winter dry spells there are shorter than dry spells in spring and summer, while winter wet spells are slightly longer than those in the other seasons; these variations determine the model parameters.

Introduction. Many investigations have been conducted into the distributions of sequences of wet and dry days. The most popular model used to fit the distributions of spell lengths has been the simple Markov model which assumes that the probability of any particular day being wet or dry depends only on the character of the previous day (for instance Chatfield,[1] Gabriel and Neumann[2]). Williams[3] first suggested a logarithmic series as a fit to sequences of wet and dry days and this model has since been applied to other data (Cooke,[4] Chatfield[1]). Green[5] proposed a modified logarithmic model of which the simple logarithmic and Markov models are special cases. This model, which used two parameters, satisfactorily fitted 33 out of 36 cases collected by Green and others; these included observations of duration of rainstorms, and of intervals between them, collected by Weiss.[6] Yap[7] proposed a modification to the simple Markov model in which the probability parameter (of a wet or dry day being followed by a similar day) was a variable, though a constant within any given spell length.

New data are here investigated in relation to seasonal and regional variations. Distributions from eight stations are compared. Those for Oxford are further divided into four seasons and examined in more detail.

The models. The probabilities of spells of length 1, 2, 3 . . . r wet or dry days are defined by the various models as follows:

Model 1 : Markov Chain Model

$q, q^2, \ldots q^r$, with normalizing constant $\dfrac{1-q}{q}$.

Model 2 : Williams's logarithmic model

$q, q^2/2, q^3/3, \ldots q^r/r$, with normalizing constant $\dfrac{1}{\log(1-q)}$.

Model 3 : Green's modified logarithmic model

$\dfrac{q}{1+a}, \dfrac{q^2}{2+a}, \ldots \dfrac{q^r}{r+a}$ with $0 \leqslant a \leqslant \infty$ and the normalizing constant determined by the requirement $\displaystyle\sum_r \frac{q^r}{r+a} = 1$. In each case the normalizing constant ensures that the total probability is unity. In order to fit models 1 and 2 from data the mean spell length is used (i.e. the total number of wet (or dry) days divided by the total number of wet (or dry) spells) to find q.

For model 1, mean spell length $= \displaystyle\sum_r \frac{1-q}{q} q^r r = \frac{1}{1-q}$.

For model 2, mean spell length $= \displaystyle\sum_r \frac{-1}{\log(1-q)} \frac{r\ q^r}{r} = \frac{-q}{\log(1-q)(1-q)}$.

For model 1, q can be found directly from the mean spell length; for model 2 it is found by a recursive process or from tables published by Williamson and Bretherton.[8]

It may be noticed that models 1 and 2 are special cases of model 3 for $a = \infty$ and 0 respectively.

To fit model 3 the method of minimum chi-square is used. We let q approach 0 from 1 and let a approach 0 from some value greater than, say, 6 in successive steps; the distribution for given a, q is tested for fit at each step by the chi-square test. The parameters a and q are altered each time the chi-square value falls as compared with the values of a, q for previous smallest values of chi-square. In applying the chi-square tests, spells of length greater than a certain value (about 15) are grouped together into one category. The program stops when chi-square falls below a certain value determined by the number of categories; the a and q values for the minimum chi-square value are taken as best fit values.

For model 4 we assume that the probability of a dry or wet spell is p, where p is a random variable having a constant value within any one run, but different values in different runs (as Yap[7]); p is assumed to be a random variate

$$f(p) = \frac{p^{a-1}\,(1-p)^{b-1}}{B\,(a,\,b)}$$

where a, b are constants of the distribution and $B(a,b)$ is the Beta function.

The probability of a run of days is given by

$$P(r) = \frac{1}{B(a,b)} \int_0^1 (1-p)\, p^{r-1}\,(1-p)^{b-1}\,p^{a-1}\,dp,$$

where $(1-p)$ is a normalizing factor and p^{r-1} arises from $r-1$ days following the first wet (or dry) day. Then

$$P(r) = \frac{B\,(a+r-1,\,b+1)}{B\,(a,\,b)},$$

$$P(1) = \frac{b}{a+b}, \text{ and } r \geqslant 2,$$

$$P(r) = \frac{a+r-2}{a+b+r-1}\,P\,(r-1),$$

where we have used the definitions

$$B\,(x, y) = \frac{(x-1)!\,(y-1)!}{(x+y-1)!} = \int_0^1 p^{x-1}\,(1-p)^{y-1}\,dp.$$

To fit the model we take factorial moments about the origin U'_1, U'_2 for the first two moments, i.e.

$$U'_1 = \int_0^1 f\,(p)\,(1-p) \sum_0^\infty p^{r-1}\,r\,dp$$

$$= \frac{1}{B\,(a,\,b)} \int_0^1 \frac{1}{(1-p)^2}\,(1-p)\,p^{a-1}\,(1-p)^{b-1}\,dp$$

$$= \frac{B\,(a,\,b+1)}{B\,(a,\,b)} = \frac{a+b-1}{b-1}$$

$$U'_2 = \int_0^1 \mathrm{f}(p)\ (1 - p) \sum_0^\infty p^{r-1}\, r(r-1)\ \mathrm{d}p$$

$$= \frac{1}{B\,(a,\, b)} \int_0^1 p^{a-1}\,(1-p)^{b-1} \frac{2p}{(1-p)^2}\, \mathrm{d}p,$$

$$= \frac{2}{B\,(a,\, b)}\ B(a+1,\, b-2),$$

$$= \frac{2a\,(a+b-1)}{(b-1)\ (b-2)}.$$

Then $b = \dfrac{2U'_1\,(U'_1 - 1) - 2U'_2}{2U'_1\,(U'_1 - 1)\ U'_2}$,

$a = (U'_1 - 1)\ (b - 1)$.

U'_1 is equated to mean spell length and U'_2 to the difference between mean-square spell length and mean spell length.

Persistence. As a measure of persistence we may use the ratio of the probability of spell length $(r + 1)$ to spell length, r, $\mathrm{F}(r)$ say. For models 1, 2 and 3, $\mathrm{F}(r) = P(r+1)/P(r) = q\ \ ((r+a)/(r+a+1))$. For the general case of model 3, $0 \leqslant a \leqslant \infty$; models 1 and 2 are special cases of model 3 for $a = \infty$ and 0 respectively. $\mathrm{F}(r)$ is constant for model 1 for all r and equals q. In general, $\mathrm{F}(r)$ increases with spell length r and with model parameter a; its rate of increase decreases as r increases and $\mathrm{F}(r)$ tends to q in the limit.

For model 4, $\mathrm{F}(r) = P(r+1)/P(r) = (a + r - 3)/(a + b + r - 1)$ and the measure of persistence increases with spell length, tending to 1 for large r.

Model fitting for eight stations. The data used were for 40-year periods: 1921–60 for York, Cwm Dyli (North Wales), Oxford, Falmouth, March and Edgbaston, 1931–70 for Edinburgh; for Whitby, the shorter period 1921–42 was used. Difficulty was experienced in finding stations with long-term continuous rainfall records with a constant threshold for recording rainfall; threshold values were 0·01 in for all stations apart from Edinburgh and Edgbaston with 0·2 mm. These data are given in Appendix I (dry spells), Appendix II (wet spells); graphs of spell length distribution for Edinburgh, Falmouth, Cwm Dyli and March are illustrated as representative examples in Figures 1 to 6.

The chi-square test was used to test the fit of models 1 to 4 to the observed distribution with an acceptance level of $P(\chi^2) \geqslant 0{\cdot}05$. For dry spells the logarithmic model fitted the data for all stations except March and Edgbaston. For March the modified logarithmic model fitted the data for small a ($a = 0$ for the simple logarithmic model); for Edgbaston no model fitted the dry-spell data. Neither the Markov nor the modified geometric models produced distributions to fit any of the dry-spell data. The parameter q did not show any systematic variation among the stations.

For wet spells the modified geometric and modified logarithmic models fitted most data. The exceptions were Cwm Dyli for the geometric model,

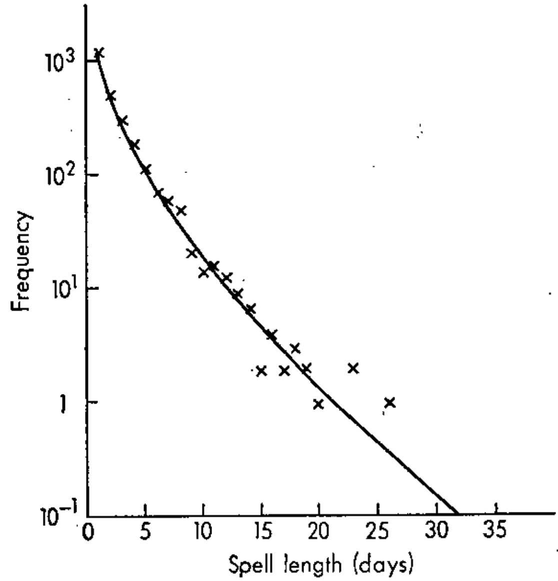

FIGURE 1—DRY SPELLS AT EDINBURGH, LOGARITHMIC MODEL $q = 0{\cdot}82$.

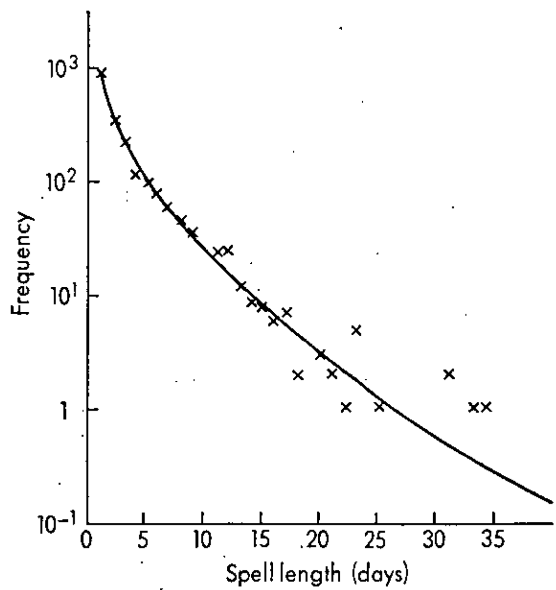

FIGURE 2—DRY SPELLS AT FALMOUTH, LOGARITHMIC MODEL $q = 0{\cdot}87$.

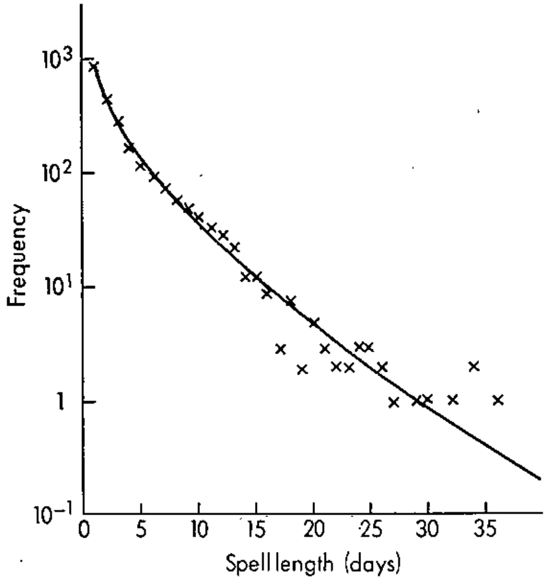

FIGURE 3—DRY SPELLS AT MARCH, MODIFIED LOGARITHMIC MODEL
$q = 0{\cdot}87$, $a = 0{\cdot}34$.

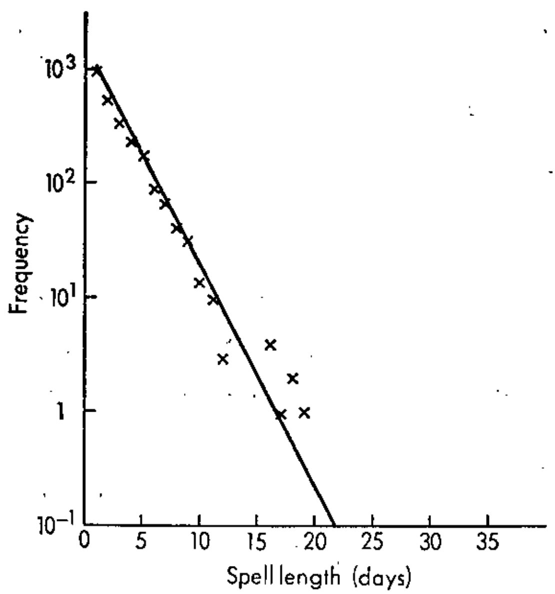

FIGURE 4—WET SPELLS AT EDINBURGH, MARKOV MODEL
$q = 0{\cdot}64$.

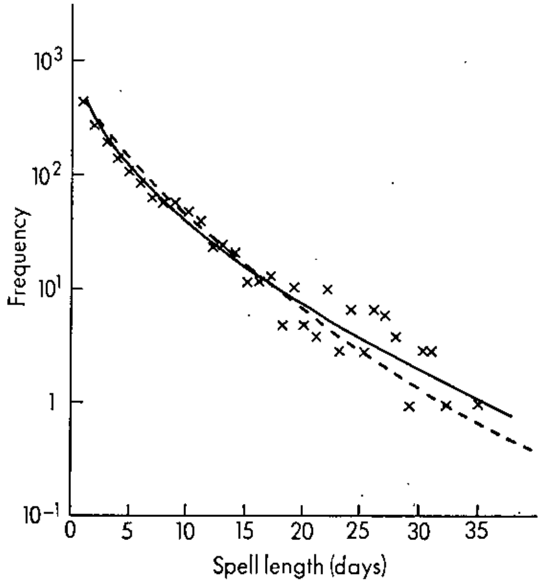

FIGURE 5—WET SPELLS AT CWM DYLI
——— modified logarithmic model, $q = 0{\cdot}90$, $a = 1{\cdot}18$.
- - - - modified geometric model.

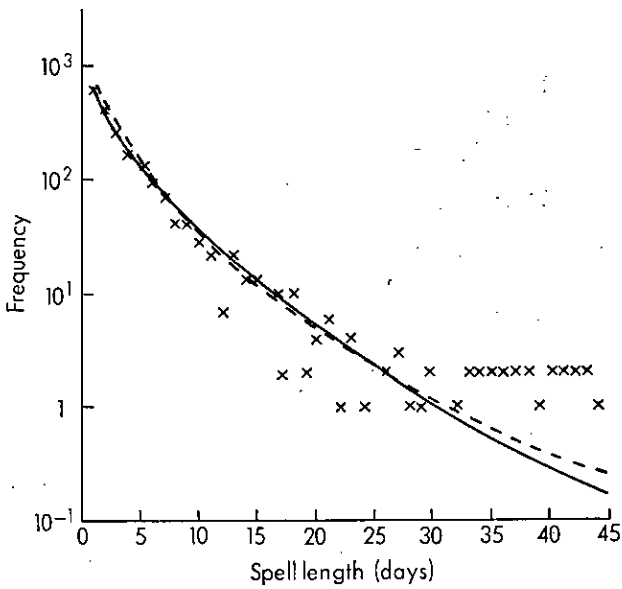

FIGURE 6—WET SPELLS AT FALMOUTH
——— modified logarithmic model, $q = 0{\cdot}88$, $a = 0{\cdot}92$.
- - - - modified geometric model.

Falmouth for the modified logarithmic model and Edinburgh fitted by neither modified model. However, a simple Markov model, which is a special case of both modified models, fitted the Edinburgh data. The modified logarithmic model usually produced a slightly better fit than the modified geometric model, though differences were only apparent for longer, less-frequent spells.

Variations of the parameters a and q for wet spells did not correlate well with region or with mean annual rainfall. Cwm Dyli (mean annual rainfall 140·46 in (3567·68 mm) for the period 1916–50) and Falmouth (43·00 in, 1092·20 mm), the two wettest stations, had slightly higher values of q than the other stations and showed slightly greater persistence of wet spells. The other stations had mean annual rainfall in decreasing order as follows: Edgbaston 30·70 in (779·78 mm), Edinburgh 27·53 in (699·26 mm), Whitby 25·66 in (651·76 mm), York 24·70 in (627·38 mm), and March 23·07 in (585·98 mm).

An examination was made of the effect of a change of threshold for the two stations which recorded in millimetres. It was found that with a threshold value of 0·1 mm the Edgbaston dry-spell data fitted a logarithmic model (no fit found for 0·2 mm), and that Edinburgh wet-spell data fitted both modified models (a Markov fit found for 0·2 mm). For Edgbaston wet-spell data and for Edinburgh dry-spell data the change of threshold was found to cause only a slight change in the model parameters.

Comparative persistence of wet and dry spells. Using models 1 to 3 the values of the measure of persistence $F(r)$ were compared for wet and dry spells for each station. For March, $F(r)$ was larger for all dry spells than for wet spells. For Edinburgh, York, Edgbaston and Oxford, $F(r)$ was larger for dry spells of length greater than two days; for Whitby, $F(r)$ was larger for dry spells longer than five days. For Edinburgh, where a Markov model produced a best fit to wet-spell data, $F(r)$ was of course constant. For the wetter stations, Cwm Dyli and Falmouth, $F(r)$ was larger for wet spells for all r. Thus we infer that dry spells are more persistent at 'dry' stations and wet spells are more persistent at 'wet' stations. For intermediate stations wet spells are more persistent for short spells only. The variations probably reflect the passage of synoptic features. Anticyclones tend to build up slowly over two or three days and last for longer periods than do individual depressions. For 'wet' stations effects of minor disturbances are greater than at other stations and wet spells tend to be more persistent than dry spells. However, analysis of spell data does not distinguish the effect of individual disturbances; a long wet spell may result from several successive depressions.

Seasonal variations (see Appendices III and IV and Figures 7–9). The Oxford data for 1852–1970 were divided into four seasons—winter (December to February), spring (March to May), summer (June to August) and autumn (September to November), the divisions between seasons being taken at the end of a spell. Each seasonal set of data was tested for the distribution of spells according to the above model. Dry spells again fitted the log model and wet spells the modified geometric model. For dry spells the mean spell lengths were similar for autumn and winter (2·875 and 2·921 days) and for spring and summer (3·498 and 3·344); the corresponding values of the parameters q in the logarithmic model were 0·84 for autumn and winter, 0·88 for spring and 0·87 for summer. For wet spells it was found that mean spell lengths

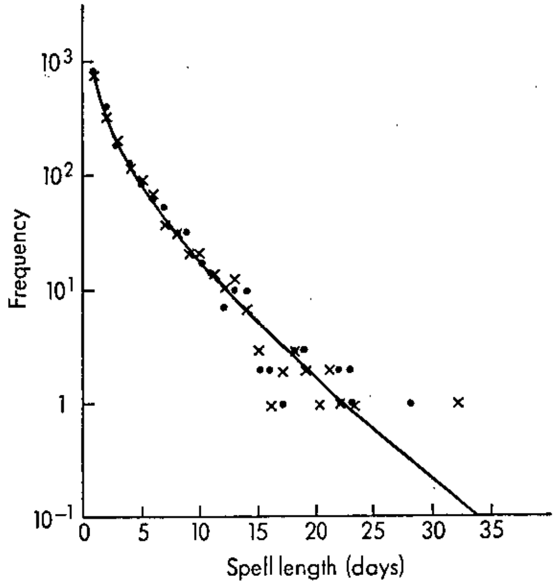

FIGURE 7—DRY SPELLS AT OXFORD, LOGARITHMIC MODEL
● Autumn × Winter

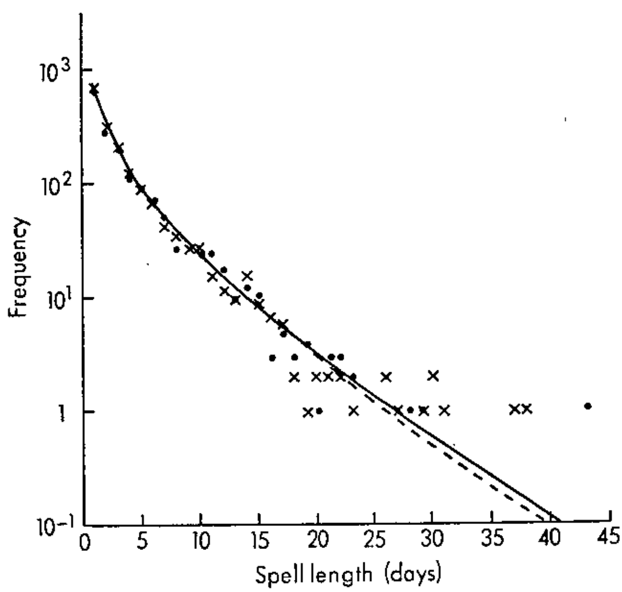

FIGURE 8—DRY SPELLS AT OXFORD, LOGARITHMIC MODEL
Continuous line and dots refer to spring; $q = 0{\cdot}88$.
Pecked line and crosses refer to summer; $q = 0{\cdot}87$.

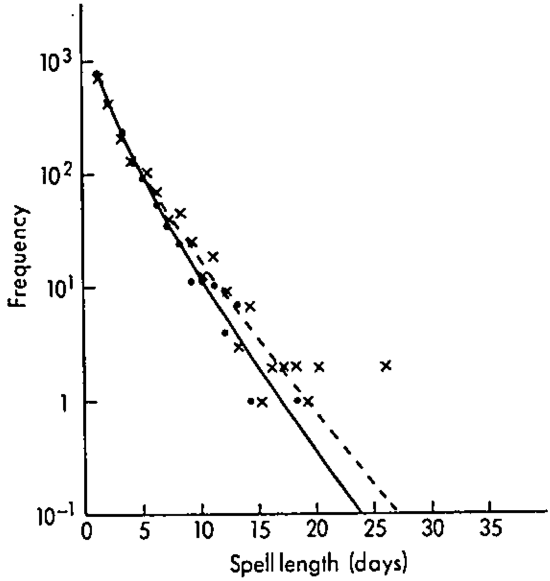

FIGURE 9—WET SPELLS AT OXFORD, MODIFIED LOGARITHMIC MODEL
Continuous line and dots refer to summer; q = 0·87, a = 1·09.
Pecked line and crosses refer to winter; q = 0·78, a = 1·59.

decreased from winter (2·932) to summer (2·621) with spring (2·783) and autumn (2·750) having similar lengths. The parameters a and q which produced the best fit to spring and autumn wet spells also produced a good (but not best) fit to summer wet spells (see table below).

	a	q	$P(\chi^2)$
Spring	2·075	0·75	0·50
Autumn	2·075	0·75	0·05
Summer	2·075	0·75	0·05
	1·087	0·75	0·30
Winter	1·581	0·78	0·10

Cumulative distributions. As regards extremes, a model which describes cumulative spell distributions, i.e. the number of spells of length greater than a specified value, may be of more practical value than one describing individual spells. For this reason the spell data were also considered cumulatively. For dry spells only the modified logarithmic model was found to fit the cumulative data and that at only four out of the eight stations; the parameters a and q of the model were 0·87 and 1·09 respectively for York and Oxford, 0·81 and 2·07 for Cwm Dyli, and 0·87 and 2·07 for March. On the other hand it was found that none of the models fitted cumulative wet-spell data. It was usually the rarer long spells which failed to fit the models for cumulative data since after the summation of data their relative weight in the fit was decreased.

Conclusions. We may agree with Green's conclusion that the modified logarithmic model (of which Markov's model and the simple logarithmic model are special cases) fits most spell data. As a first approximation, we may say that the simple logarithmic model fits dry-spell data with q about 0·85; for wet spells, the modified logarithmic model fits most data, with a about 2 or 3 and q about 0·7 or 0·8 for other stations. The modified geometric model also fits most wet-spell data. The models give only a rough guide to the occurrence of infrequent long spells.

Seasonally, dry spells are slightly longer for spring and summer than for autumn and winter, one q-value for each half of the year being sufficient to describe the data. For wet spells, different values of a and q are needed for longer winter spells than those for other seasons.

REFERENCES

1. CHATFIELD, C.; Wet and dry spells. *Weather, London*, **21**, 1966, pp. 308–310.
2. GABRIEL, K. R. and NEUMANN, J.; A Markov chain model for daily rainfall occurrence. *Q J R Met Soc, London*, **88**, 1962, pp. 90–95.
3. WILLIAMS, C. B.; Sequences of wet and dry days considered in relation to the logarithmic series. *Q J R Met Soc, London*, **78**, 1952, pp. 91–96.
4. COOKE, D. S.; The duration of wet and dry spells at Moncton, New Brunswick. *Q J R Met Soc, London*, **79**, 1953, pp. 536–538.
5. GREEN, J. R.; A generalized probability model for sequence of wet and dry days. *Mon Weath Rev, Washington*, **98**, 1970, pp. 238–241.
6. WEISS, L. L.; Sequences of wet or dry days described by a Markov chain probability model. *Mon Weath Rev, Washington*, **92**, 1964, pp. 169–176.
7. YAP, W. C.; The persistence of wet and dry spells in Sungei Buloh, Selangor, *Met Mag, London*, **102**, 1973, pp. 240–245.
8. WILLIAMSON, E. and BRETHERTON, M. M.; Tables of the logarithmic series distribution. *Ann Math Statist, Ann Arbor*, No. 35, 1964, pp. 284–298.

APPENDIX I—DRY-SPELL FREQUENCIES

	Edinburgh		York		Whitby		Cwm Dyli		Oxford		Falmouth		March		Edgbaston		Edgbaston*	
Spell length	Observed	Log. model	Observed	Log. model	Observed	Log. model	Observed	Log. model	Observed	Log. model	Observed	Log. model	Observed	Log. model	Observed	Log. model	Observed	Log. model
days																		
1	1241	1280·9	1004	1033·4	662	689·3	886	862·6	1004	1034·8	904	868·9	944	951·0	1039	1068·1	1045	1076·9
2	524	524·7	476	444·4	287	276·8	340	359·3	476	444·5	358	377·9	487	476·7	460	456·9	459	454·0
3	311	286·6	228	254·8	154	148·2	170	199·6	261	254·5	216	219·2	319	291·7	261	260·7	269	255·2
4	194	176·1	197	164·4	92	89·3	120	124·7	141	164·0	116	143·0	179	196·4	154	167·3	135	161·4
5	120	115·4	107	113·1	57	57·4	84	83·1	131	112·7	92	99·5	128	139·6	141	114·5	142	108·9
6	74	78·8	82	81·1	51	38 4	65	57·7	99	80·7	73	72·1	164	102·9	81	81·6	81	76·5
7	61	55·4	60	59·8	32	26·4	59	41·2	46	59·4	53	53·8	79	78·0	73	59·9	63	55·3
8	50	39·7	40	45·0	19	18·6	36	30·0	38	44·6	47	40·9	61	59·9	39	44·8	36	40·8
9	21	28·9	43	34·4	9	13·3	26	22·3	39	34·1	36	31·6	52	46·8	43	34·1	37	30·6
10	15	21·3	31	26·6	13	9·6	17	16·7	26	26·4	29	24·8	43	37·0	34	26·3	28	23·2
11	17	15·9	24	20·8	7	7·0	17	12·6	22	20·6	24	19·6	36	29·5	23	26·4	17	17·8
12	14	11·9	19	16·4	4	5·2	11	9·7	15	16·2	25	15·6	30	23·7	18	16·0	13	13·7
13	9	9·0	11	13·0	3	3·8	6	7·4	9	12·9	12	12·5	24	19·2	10	12·7	9	10·7
14	7	6·9	10	10·4	0	2·8	3	5·7	15	10·3	9	10·1	13	15·6	8	10·1	5	8·4
15	2	5·2	5	8·4	0	2·1	3	4·5	7	8·2	8	8·2	13	12·8	10	8·0	8	6·6
16	4	4·0	4	6·7	1	1·6	3	3·5	4	6·6	6	6·7	9	10·5	8	6·4	7	5·2
17	2	3·1	5	5·5	1	1·2	2	2·7	3	5·4	7	5·5	3	8·7	1	5·2	3	4·1
18	3	2·4	2	4·4	1	0·9	1	2·2	3	4·3	2	4·5	8	7·2	3	4·2	2	3·3
19	2	1·9	4	3·6	0	0·7	1	1·7	4	3·5	4	3·7	2	5·9	2	3·4	1	2·6
20	1	1·5	4	2·9	1	0·5	1	1·3	1	2·9	3	3·1	5	4·9	2	2·8	2	2·1
21	0	1·1	1	2·4	0	0·4	0	1·1	2	2·4	2	2·5	3	4·1	0	2·3	0	1·7
22	0	0·9	1	2·0	0	0·3	0	0·8	3	1·9	1	2·1	2	3·4	3	1·8	2	1·4
23	2	0·7	1	1·6	1	0·2	1	0·7	1	1·6	5	1·8	2	2·9	0	1·5	0	1·1
24	0	0·5	0	1·3	0	0·1	1	0·5	0	1·3	0	1·5	3	2·4	0	1·2	0	0·9
25	0	0·4	1	1·1	0	0·1	0	0·4	1	1·1	1	1·2	3	2·0	0	1·0	0	0·7
26	1	0·3	2	0·9	0	0·1	0	0·3	1	0·9	0	1·0	2	1·7	0	0·8	0	0·6
27	0	0·3	0	0·8	0	0·1	0	0·3	0	0·7	0	0·9	1	1·4	0	0·7	0	0·5
28	0	0·2	0	0·6	0	0·1	0	0·2	2	0·6	0	0·7	0	1·2	0	0·6	0	0·4
29	0	0·2	1	0·5	0	0	1	0·2	2	0·5	0	0·6	1	1·0	0	0·5	0	0·3
30	0	0·1	0	0·4	0	0	0	0·1	1	0·4	0	0·5	1	0·9	1	0·4	0	0·3
31	0	0·1	1	0·4	0	0	0	0·1	1	0·3	2	0·4	0	0·7	0	0·3	0	0 2
32	0	0·1	1	0·3			0	0·1	0	0·3	0	0·4	1	0·6	1	0·3	1	0·2
33	0	0·1	0	0·3					0	0·2	1	0·3	0	0·5	0	0·2	0	0·1
34	0	0·1	2	0·2					0	0·2	1	0·3	2	0·4	1	0·2	1	0·1
35			0	0·2					0	0·2	0	0·2	0	0·4	0	0·1	0	0·1
36									0	0·1	0	0·2	1	0·3	0	0·1	0	0·1
37									1	0·1	0	0·1	0	0·2	0	0·1	0	0·1
38									0	0·1	0	0·1	0	0·2	0	0·1	0	0·1
39									0	0·1	0	0·1	0	0·2	0	0·1	0	0·1
40									0	0·1	0	0·1	0	0·2	0	0·1	0	0·1
χ^2		15·7		20·0		13·3		20·9		29·9		20·1		18·7		31·5		27·9
$P(\chi^2)$		0·40		0·30		0·30		0·20		0·05		0·30		0·40		0·02		0·05
q		0·82		0·86		0·803		0·833		0·859		0·870		0·875		0·85		0·84
a														0·337				

Observed denotes observed frequency; Log. denotes expected frequency (logarithmic model).
* When threshold decreased from 0·2 mm to 0·1 mm.

APPENDIX II—WET-SPELL FREQUENCIES

Spell length	Edinburgh Observed	Edinburgh Expected frequency (Markov)	York Observed	York Modified log.	York Modified geometric	Whitby Observed	Whitby Modified log.	Whitby Modified geometric	Cwm Dyli Observed	Cwm Dyli Modified log.	Cwm Dyli Modified geometric
days											
1	1003	950·8	907	895·5	882·9	523	489·4	492·8	477	495·6	413·9
2	570	612·7	520	504·2	531·2	293	302·6	309·1	322	308·6	313·2
3	336	394·9	305	307·9	327·9	182	193·7	197·3	212	213·2	239·2
4	251	254·4	204	197·5	207·2	121	127·0	127·9	153	156·3	184·2
5	195	164·0	136	130·8	133·7	74	84·8	84·2	127	119·0	143·0
6	96	105·7	92	88·7	87·9	62	57·4	56·2	95	93·0	111·9
7	68	68·1	64	61·2	58·8	40	39·4	38·0	68	74·1	88·1
8	43	43·9	41	42·8	40·0	39	27·2	26·0	63	60·0	69·9
9	33	28·3	25	30·3	27·6	12	19·0	18·0	62	49·1	55·7
10	14	18·2	26	21·6	19·3	11	13·3	12·6	51	40·6	44·7
11	10	11·7	11	15·6	13·7	13	9·3	8·9	41	33·9	36·1
12	3	7·6	9	11·3	9·8	7	6·6	6·4	25	28·4	29·3
13	11	4·9	2	8·2	7·1	4	4·7	4·6	26	24·0	23·9
14	3	3·1	6	6·0	5·2	4	3·4	3·3	24	20·4	19·6
15	0	2·0	6	4·4	3·9	1	2·4	2·4	12	17·3	16·1
16	4	1·3	2	3·2	2·9	3	1·7	1·8	13	14·8	13·3
17	1	0·8	2	2·4	2·2	3	1·3	1·3	14	12·7	11·1
18	2	0·5	4	1·8	1·6	1	0·9	1·0	5	11·0	9·2
19	1	0·3	3	1·3	1·3	1	0·7	0·8	11	9·5	7·7
20	0	0·2	0	1·0	1·0	1	0·5	0·6	5	8·2	6·5
21	0	0·1	0	0·7	0·7	0	0·3	0·4	4	7·1	5·4
22	0	0·1	1	0·6	0·6	0	0·2	0·3	11	6·2	4·6
23	0	0·1	0	0·4	0·5				3	5·4	3·9
24	0	0·0	0	0·3	0·3				7	4·7	3·7
25	0	0·0	1	0·2	0·2				3	4·1	2·8
26	0	0·0	1	0·2	0·2				7	3·6	2·5
27	1	0·0							6	3·1	2·1
28									4	2·8	1·8
29									1	2·4	1·6
30									3	2·1	1·4
31									3	1·9	1·2
32									1	1·6	1·0
33									0	1·4	0·9
34									0	1·3	0·8
35									1	1·1	0·7
36									0	1·0	0·6
37											
38											
39											
40											
41											
42											
43											
44											
χ^2		22·2		10·4	12·0		14·8	16·8		24·7	49·4
$P(\chi^2)$		0·05		0·30			0·20	0·10		0·30	0·001
q		0·64		0·78			0·76			0·90	
a				1·58			3·43			1·18	

APPENDIX II *continued*

Oxford Observed	Oxford Modified log.	Oxford Modified geometric	Falmouth Observed	Falmouth Modified log.	Falmouth Modified geometric	March Observed	March Modified log.	March Modified geometric	Edgbaston Observed	Edgbaston Modified log.	Edgbaston Modified geometric
884	873·1	865·0	628	660·8	628·9	1328	1270·3	1294·3	884	902·0	871·2
532	525·2	528·6	427	362·7	409·5	577	584·6	615·3	554	519·5	536·9
310	328·8	330·1	264	251·1	276·5	283	295·4	304·6	311	321·4	338·2
197	211·6	210·3	164	176·2	192·4	176	157·6	156·3	219	207·9	217·2
131	139 0	138 6	137	129·0	137·4	83	87·1	82·9	154	138·5	142·1
96	92·7	89·0	98	97·2	100·3	58	49·4	45·3	90	94·4	94·5
55	62·6	60·3	68	74·8	74·7	19	28·5	25·4	59	65·4	63·8
50	42·7	41·0	43	58·5	56·6	10	16·7	14·6	48	45·8	43·8
39	29·4	28·2	41	46·3	43·6	13	9·9	8·6	32	32·5	30·4
14	20·3	19·7	28	37·1	34·0	6	5·9	5·1	25	23·2	21·4
14	14·2	13·9	22	29·9	27·0	4	3·6	3·1	14	16·7	15·2
11	9·9	9·9	7	24·3	21·5	1	2·2	2·0	11	12·1	10·9
9	7·0	7·1	22	19·9	17·3	2	1·3	1·2	9	8·8	7·9
6	4·9	5·2	13	16·3	14·1	0	0·8	0·8	6	6·5	5·8
2	3·5	3·8	13	13·5	11·6	0	0·5	0·5	4	4·7	4·3
1	2·5	2·8	10	11·2	9·6	1	0·3	0·3	5	3·5	3·2
1	1·8	2·1	2	9·3	8·0	0	0·2	0·2	3	2·6	2·4
3	1·3	1·6	10	7·8	6·8	0	0·1	0·1	2	1·9	1·8
2	0·9	1·2	2	6·5	5·7	0	0·1	0·1	1	1·4	1·4
1	0·7	0·9	4	5·4	4·8				0	1·1	1·1
1	0·5	0·7	6	4·6	4·1				1	0·8	0·8
0	0·3	0·5	1	3·9	3·5				0	0·6	0·6
1	0·2	0·4	4	3·2	3·0				1	0·4	0·5
0	0·2	0·3	1	2·7	2·6				0	0·3	0·4
			0	2·3	2·3				1	0·2	0·3
			2	2·0	2·0				0	0·2	0·2
			3	1·7	1·8				1	0·1	0·2
			1	1·4	1·6				0	0·1	0·2
			1	1·2	1·4						
			2	1·0	1·2						
			1	0·9	1·1						
			1	0·7	1·0						
			2	0·6	0·9						
			2	0·5	0·8						
			2	0·5	0·7						
			2	0·4	0·6						
			2	0·3	0·6						
			2	0·3	0·6						
			1	0·3	0·5						
			2	0·2	0·4						
			2	0·2	0·4						
			2	0·2	0·3						
			2	0·2	0·3						
			1	0·1	0·3						
	11·2	12·7		50·2	21·5		14·1	1·61		4·7	5·6
	0·60	0·40		0·01	0·20		0·10	0·05		0·98	0·98
	0·75			0·88			0·66			0·78	
	3·05			0·92			1·35			1·81	

APPENDIX III—SEASONAL DRY-SPELL FREQUENCIES, OXFORD

Spell length *days*	Winter Observed	Winter Log. model	Spring Observed	Spring Log. model	Summer Observed	Summer Log. model	Autumn Observed	Autumn Log. model
1	831	828·5	685	713·5	758	769·3	827	869·3
2	331	349·9	301	314·7	327	336·1	408	365·6
3	205	197·1	204	185·0	207	195·8	191	205·0
4	119	124·9	118	122·4	130	128·3	131	129·3
5	94	84·4	98	86·4	97	89·7	87	87·0
6	70	59·4	76	63·5	71	65·3	62	61·0
7	38	43·0	54	48·0	44	48·9	54	44·0
8	33	31·8	28	37·0	37	37·4	32	32·3
9	21	23·9	31	29·0	29	29·0	34	24·2
10	22	18·2	26	23·1	28	22·8	18	18·3
11	14	13·9	27	18·5	16	18·1	15	14·0
12	10	10·8	19	14·9	12	14·5	7	10·8
13	13	8·4	10	12·2	10	11·7	10	8·4
14	6	6·6	13	10·0	16	9·5	10	6·5
15	3	5·2	11	8·8	9	7·8	2	5·1
16	1	4·1	3	6·8	7	6·3	2	4·0
17	2	3·3	5	5·6	5	5·2	1	3·2
18	3	2·6	3	4·7	2	4·3	3	2·5
19	2	2·1	4	3·9	1	3·6	3	2·0
20	1	1·7	1	3·3	2	3·0	1	1·6
21	2	1·4	3	2·8	2	2·5	0	1·3
22	1	1·1	3	2·3	2	2·1	1	1·0
23	2	0·9	2	2·0	1	1·7	1	0·8
24	1	0·7	0	1·7	0	1·4	0	0·7
25	1	0·6	0	1·4	0	1·2	0	0·5
26	0	0·5	0	1·2	2	1·0	0	0·4
27	0	0·4	0	1·0	1	0·9	0	0·4
28	0	0·3	1	0·9	0	0·7	1	0·3
29	0	0·2	1	0·7	1	0·6	0	0·2
30	0	0·2	0	0·6	2	0·5	0	0·2
31	0	0·1	1	0·5	1	0·4	0	0·2
32	1	0·1	0	0·5	0	0·4	0	0·1
33	0	0·1	0	0·4	0	0·3	0	0·1
34	0	0·1	0	0·3	0	0·3	0	0·1
35	0	0·1	0	0·3	0	0·2	0	0·1
χ^2		20·34		23·4		12·1		11·5
$P(\chi^2)$		0·70		0·15		0·70		0·10
q		0·84		0·88		0·87		0·84

APPENDIX IV—SEASONAL WET-SPELL FREQUENCIES, OXFORD

Spell length *days*	Winter Observed	Winter Modified log.	Winter Modified geometric	Spring Observed	Spring Modified log.	Spring Modified geometric	Summer Observed	Summer Modified log.	Summer Modified geometric	Autumn Observed	Autumn Modified log.	Autumn Modified geometric
1	718	697·1	680·4	680	682·3	660·9	757	816·7	734·3	762	746·6	744·8
2	409	392·6	408·9	393	386·0	392·3	404	414·2	424·9	434	422·5	435·9
3	212	239·7	252·2	211	232·4	238·5	245	234·6	251·9	241	254·5	262·0
4	142	153·7	159·8	144	145·6	148·2	150	141·4	151·9	174	159·4	161·4
5	105	101·9	103·3	90	93·8	94·0	98	88·6	93·5	84	102·7	101·7
6	69	69·1	68·2	61	61·6	60·7	56	57·1	58·6	68	67·5	65·3
7	40	47·7	45·8	44	41·1	39·8	38	37·5	37·3	46	45·0	42·8
8	46	33·4	31·3	39	27·8	26·6	26	25·0	24·2	21	30·4	28·5
9	26	23·6	21·7	23	19·0	18·0	12	16·9	15·8	28	20·8	19·3
10	13	16·9	15·3	13	13·0	12·3	12	11·5	10·5	11	14·3	13·2
11	19	12·1	10·9	4	9·0	8·6	11	7·9	7·1	13	9·9	9·2
12	9	8·8	7·9	7	6·3	6·0	4	5·5	4·8	5	6·9	6·5
13	3	6·4	5·7	3	4·4	4·3	7	3·8	3·3	10	4·8	4·6
14	7	4·7	4·2	3	3·1	3·0	1	2·7	2·3	3	3·4	3·3
15	1	3·4	3·1	1	2·2	2·2	0	1·9	1·6	1	2·4	2·4
16	2	2·5	2·4	0	1·6	1·6	2	1·3	1·2	2	1·7	1·8
17	2	1·9	1·8	1	1·1	1·2	2	0·9	0·8	0	1·2	1·3
18	2	1·4	1·4	2	0·8	0·9	1	0·7	0·6	2	0·9	1·0
19	1	1·0	1·0	0	0·6	0·7	0	0·5	0·4	1	0·6	0·7
20	2	0·8	0·8	0	0·4	0·5	0	0·4	0·3	1	0·4	0·6
21	0	0·6	0·6	1	0·3	0·4	0	0·3	0·2	1	0·3	0·4
22	0	0·4	0·5	0	0·2	0·3	0	0·2	0·2	0	0·2	0·3
23	0	0·3	0·4	1	0·1	0·2	0	0·2	0·1	0	0·2	0·3
24	0	0·2	0·3	1	0·1	0·2	0	0·1	0·1	0	0·1	0·2
25	0	0·2	0·2	0	0·1	0·1	0	0·1	0·1	0	0·1	0·2
26	1	0·1	0·2	0	0·1	0·1	0	0·1	0·1			
χ^2		19·9	29·4		11·3	15·2		13·4	10·7		19·4	21·4
$P(\chi^2)$		0·10	0·01		0·50	0·20		0·30	0·50		0·05	0·05
q		0·78			0·75			0·75			0·75	
a		1·58			2·075			1·087			2·075	

APPENDIX 2

MAXIMUM ENTROPY POWER SPECTRA OF P AND C CIRCULATION INDICES AND SCOTTISH ANNUAL RAINFALL TIME-SERIES

MAXIMUM ENTROPY POWER SPECTRA OF CIRCULATION INDICES AND SCOTTISH ANNUAL RAINFALL

Introduction

Investigations further to the work of section 5.3.3, which used the Blackman-Tukey autocovariance method (ACV) of power spectrum analysis of Scottish rainfall and circulation indices, were carried out using maximum entropy method (MEM) power spectra to locate more accurately the frequency of spectral peaks already found and to further look for low frequency peaks. This method was brought to the notice of the author after the investigations reported in Chapter 5 had been carried out.

The MEM is described in detail by Ross (1975) who carried out numerical experiments on synthetic time series using this method and also the fast Fourier transform method (FFT). MEM was developed by Burg (1967) and was also proposed by Parzen (1969) who derived it from autoregressive modelling. The MEM is "data adaptive" in that the window function is not defined as in the ACV and FFT methods but is implicitly altered by the data being processed to suit the noise characteristics of the signal. It is exactly equivalent to modelling by autoregressive decomposition (see Ulrych and Bishop 1975) and presents the same problem of the best choice of the AR order.

The spectral estimate produced by MEM has a higher variance than that of most other methods, but this is more than compensated for by its significantly higher resolution and its capacity to resolve wavelengths comparable to the length of the data sample

(Parzen 1969, Ulrych 1972, Chen and Stegen 1974).

Method

The Blackman-Tukey autocovariance method of spectral analysis computes the Fourier transform, $P_x(f)$, of the autocorrelation function $R_x(p)$ for $p < m$ where $m < n$, the number of data points (see equation 5.10). For an infinite series the Fourier transform of the autocorrelation function, $PS_x(f)$, would be an exact estimate of the power spectrum as in equation A2.1.

$$PS_x(f) = \Delta t \sum_{-\infty}^{+\infty} R_x(p) \exp - 2\pi i p \Delta t \qquad \text{(A2.1)}$$

In order to allow for the finite length of the series and the truncation of the autocorrelation function for $p > n$, only the first m autocorrelations are used and raw spectral estimates are smoothed. The number of autocorrelation values, m, and the spectral window used are determined by the investigator; in the above analysis $m = \frac{n}{3}$ and "Hanning" function is applied to the raw data as in equation 5.11. The spectral windows tend to place information which may not exist into the data and to make assumptions about variations in members of the series outside the range of available data. Spectral estimates for short-length data will be of limited use and will be in error for wavelengths comparable to the length of the data.

A method which recognises the lack of information outside the length of the series and makes maximum use of the autocorrelations available is desirable. Loss of information is gain of entropy which may be defined as

$$h = \frac{1}{4f_N} \int_{-f_N}^{+f_N} \ln PS_x(f)\, df \qquad (A2.2)$$

where f_N is the Nyquist frequency.

This expression may be maximised with respect to unknown autocorrelations which is equivalent to maximising the lack of information about them. Thus for a series of n values

$$\frac{\partial h}{\partial R_x(p)} = 0 \qquad p \geqslant n$$

$$= \int_{-f_N}^{+f_N} \frac{\exp - 2\pi i f p \Delta t}{PS_x(f)} \qquad (A2.3)$$

which implies the Fourier series truncates and

$$\frac{1}{PS_x(f)} = \frac{1}{2f_N} \sum_{-n+1}^{-n-1} C(p) \exp(-2\pi i f p \Delta t) \qquad (A2.4)$$

with $C(p) = C(-p)^*$ to ensure that $PS_x(f)$ is real. $PS_x(f)$ can thus be expressed by

$$PS_x(f) = \frac{2f_N}{\sum_{-n+1}^{n-1} C(p) \exp(-2\pi i f p \Delta t)} \qquad (A2.5)$$

$C(p)$ must be chosen so that $PS_x(f)$ gives the unknown autocorrelations

$$R_x(p) = \int_{-f_N}^{+f_N} PS_x(f) \exp\,(2\pi\, i\, f\, p \Delta t)\, df \qquad (A2.6)$$

since the autocorrelations are the inverse transform of the power spectrum.

These conditions lead to the filtering of the data by means of a prediction error filter which whitens the series. If a time-series $x(q)$ has a Fourier transform $X(w)$ and $H(w)$ is the transform of the filter which whitens $x(q)$ then

$$X(w)\, H(w) = K^2 \quad \text{where K is a constant} \qquad (A2.7)$$

or $$X(w) = \frac{K^2}{H(w)^2}$$

which is equivalent to A2.5.

An estimator of the power spectrum can then be shown to be

$$PS_x(f) = \frac{\overline{P}_{m+1}}{\left| 1 + \sum_{j=1}^{m} a(j,m) \exp -2\pi\, i\, f\, j \Delta t \right|^2} \qquad (A2.8)$$

where $\overline{P}_{m+1}$ is the mean output of the $m+1$ point prediction error filter whose first coefficient is unity (see Ross 1975). The prediction error filter coefficients, $a(j,m)$ where $0 < j \leqslant m$ are given by the matrix equation

$$\begin{vmatrix} R_x(0) & R_x(1) & \cdots & R_x(m) \\ R_x(1) & R_x(0) & & R_x(m-1) \\ \vdots & \vdots & & \vdots \\ R_x(m) & & & \end{vmatrix} \begin{vmatrix} 1 \\ a(1,m) \\ \vdots \\ a(m,m) \end{vmatrix} = \begin{vmatrix} \overline{P}_{m+1} \\ 0 \\ \\ 0 \end{vmatrix} \qquad (A2.9)$$

In order to calculate autocorrelation coefficients $R_x(0), R_x(1) \ldots R_x(n)$ without making assumptions about the series $x(q)$ outside the data range, a recursive method of estimating prediction error coefficients is used. For $m = 0$, $R_x(0) = \overline{P}_1$ and $\overline{P}_1$ is determined directly from the data:

$$\overline{P}_1 = \frac{1}{n} \sum_{1}^{n} x(q)^2 \qquad \text{(A2.10)}$$

The $m+1$ point prediction error filter is calculated from the m-point prediction filter by A2.9 and by minimising the power output from the $m+1$ point filter with respect to $a(m,m)$. The mean power output is determined by running the filter over the data in both forward and backward directions, thus ensuring that $a(m,m)$ does not exceed unity.

For m=2 this minimisation of power implies

$$\frac{\partial}{\partial a(1,1)} \left[\sum_{1}^{n-1} \left(x(q) + a(1,1)x(q-1) \right)^2 + \left(x(q)+a(1,1)x(q+1) \right)^2 \right] = 0 \qquad \text{(A2.11)}$$

which in turn produces the result

$$a(1,1) = -2 \sum_{1}^{n-1} \frac{x(q)\, x(q+1)}{x(q)^2 + x(q+1)^2} \qquad \text{(A2.12)}$$

Equation A2.9 gives

$$\begin{vmatrix} R_x(0) & R_x(1) \\ R_x(1) & R_x(0) \end{vmatrix} \left| \begin{pmatrix} 1 \\ 0 \end{pmatrix} + a(1,1) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right| = \begin{vmatrix} \overline{P}_2 \\ 0 \end{vmatrix} \qquad \text{(A2.13)}$$

Thus

$$R_x(1) = - a(1,1)\ \bar{P}_1 \tag{A2.14}$$

$$\bar{P}_2 = \bar{P}_1\ (1 - a(1,1)^2) \tag{A2.15}$$

In general the minimisation of power output from the m+1 point filter implies that

$$a(m,m) = -2 \sum_{1}^{n-m+1} \frac{E^f_{q+m,\,m-1}\ E^b_{q,\,m-1}}{(E^f_{q+m,\,m-1})^2 + E^f_{q,\,m-1})^2} \tag{A2.16}$$

where $E^f_{q,m}$ and $E^b_{q,m}$ are forward and backward error series defined by

$$E^f_{q,m} = x(q)+a(1,m)x(q-1) + a(2,m)x(q-2)\ldots+a(m,m)x(q-m) \tag{A2.17a}$$

$$E^b_{q,m} = x(q)+a(1,m)x(q+1) + a(2,m)x(q+2)\ldots+a(m,m)x(q+m) \tag{A2.17b}$$

By definition

$$a(0,m) = 1 \tag{A2.18}$$

Equation A2.9 enables the calculation of the remaining coefficients of the (m+1)th order filter, its mean power output, and the estimate of the m th order autocorrelation coefficient as in equations A2.19, A2.20 and A2.21.

$$a(j,m) = a(j,m-1) + a(m,m)\, a(m-j,\, m-1) \qquad \text{(A2.19)}$$

$$\overline{P}_{m+1} = (1 - a(m,m))\overline{P}_m \qquad \text{(A2.20)}$$

$$R_x(m) = -\sum_{1}^{m} a(j,m)\, R_x(m-j) \qquad \text{(A2.21)}$$

From the error coefficients and mean power output, spectral estimates may be obtained as in equation A2.8 .

Stability of MEM estimates : number of filter points

The MEM spectral estimates are somewhat dependent on the number of filter points used for their computation. As the number of points is increased, spectral peaks become more pronounced and shift in frequency. Side bands of major peaks also develop for a large number of filter points. The MEM does not possess a criterion for stability in the choice of the number of points to compute the "best" estimate which will reduce the noise to a minimum.

Akaike (1969) has introduced a statistic called the Final Prediction Error, FPE. FPE is defined as the mean square prediction error:

$$FPE = \frac{1}{n} \sum_{1}^{n} (x_1(q) - x_2(q))^2 \qquad \text{(A2.22)}$$

where $$x_1(q) = x(q) - \sum_{1}^{n} x(q) \qquad \text{(A2.23)}$$

and $$x_2(q) = \sum_{j=1}^{m} a(j,m)\, x_1(q-j) \qquad \text{(A2.24)}$$

The equation A2.22, describes the unresolvable statistical deviation of the series from the true autoregressive components. FPE is not constant with order. For a series generated by an autoregressive process of order m, spectral estimates of order greater than m can be obtained from the equations above by recursion. As the order is increased above m, the statistical deviation of the components from the true components become larger.

Akaike proposed that if an estimate of FPE could be found, the order at which it becomes a minimum would be the best estimate of the true order. He laid the theoretical basis for taking the estimate of FPE for order m as a factor of the error power output, $\overline{P}(m)$, as in equation A2.25 if the mean has not been removed from the series.

$$FPE(m) = \frac{n+m}{n-m} \overline{P}(m) \qquad (A2.25)$$

MEM does not yet have a consistent statistical test for the evaluation of significance of spectral estimates. MEM has a high resolution but the relative importance and significance of spectral peaks cannot readily be deduced from their amplitudes.

Results

Results of applying the MEM method to the annual series of circulation indices and Scottish rainfall analysed in Chapter 5 appear in Table A2.1. The number of filter points for the best estimate was chosen using the Final Prediction Error estimate from the mean power output as in equation A2.25. Spectral peaks obtained by the Blackman-Tukey autocovariance approach are given

Table A2.1 Power spectra of circulation indices and Scottish rainfall using: 1. MEM 2. Blackman-Tukey ACV

Station	No. of data points	Log/ No. of filter points	Period of peaks (years)									
1 P index	111	9					7.4				2.7	2.0
2	111	37					7.1				2.7	2.0
1 C index	111	7						5.1		3.1		2.0
2	111	37								3.1		2.0
1 Edinburgh	78	6								3.4		
2	76	26			10.5					3.2		
1 Loch Leven Sluices	132	10							4.4	3.0		
2	111	37							4.4	3.1		
1 Marchmont House	106	10							4.4	2.9		
2	105	35							4.7	3.0		
1 Crombie Reservoir	99	7								3.2		2.0
2	97	32	32.0							3.1		2.0

/...

Table A2.1 continued

Station	No. of data points	Log/ No. of filter points	Period of peaks (years)									
1 Gordon Castle	108	11			9.6				4.8	3.2		
2	106	35	26.0			8.0						
1 Balmoral	92	10			10.2				4.1	3.0		
2	90	30		15.0					4.0	3.0		
1 Wick	97	8					6.7			3.5		
2	95	32			10.8							
1 Arisaig House	84	8						5.2			2.7	
2	82	27						5.1			2.7	2.0
1 Portree	62	10							4.2		2.6	
2	62	21						5.1			2.6	
1 Greenock	96	7						5.7		3.1		2.0
2	94	31				8.9			4.8			2.0
1 North Craig Reservoir	93	7						5.6		3.1		2.0
2	92	30	30.0							3.1		2.0

for comparison.

Results from the two methods tend to agree and there are no consistent low frequency peaks in MEM spectra. The presence of the 2.0 year peak in P, C, and six rainfall stations is confirmed by MEM and this peak appears to be the most pronounced in these spectra. A peak around 3.0 years in C, and at all stations except Wick, Arisaig House and Portree, which occurred in ACV spectra, is also present in MEM spectra. There is also some suggestion of a recurring peak between four and five years in both MEM and ACV rainfall spectra though peaks do not occur for the same period in different records.

Thus MEM spectra confirm the results found at high frequencies in ACV spectra; they also indicate no significant low frequency oscillations in rainfall.

References

Akaike, M. (1969): "Fitting autoregressive models for prediction". Annals of the Institute of Statistical Mathematics. Vol. 21 pp 243-247.

Burg, J.P. (1967): Papers presented at the 37th and 40th Annual International Meetings at Oklahoma City, Oklahoma, and New Orleans, Louisiana.

Chen, W.Y. and Stegen, G.R. (1974): "Experiments with Maximum Entropy Power Spectra of Sinusoids". J. Geophys. Res. Vol. 79, pp 3019-3022.

Parzen, E. (1969): "Multiple Time Series Modelling" in "Multivariate Analysis", Vol. II edited by P.R. Krishnaiah, N.Y. Academic Press.

Ross, G.H. (1975): "Maximum Entropy Spectral Analysis". Met. Office Occasional Note, Met. 021.

Ulrych, T.J. (1972): "Maximum Entropy Power Spectrum of Truncated Sinusoids". J. Geophys. Res. Vol. 77, pp 1396-1400.

Ulrych, T.J. and Bishop, T.N. (1975): "Maximum Entropy Spectral Analysis and Autoregressive Decomposition". Rev. Geoph. and Sp. Phys., Vol. 13, pp 183-200.

Smylie, D.E., Clarke, G.K.C. and Ulriych, T.J. (1973): "Analysis of Irregularities in the Earth's Rotation" from "Methods in Computational Physics". Vol. 13, edited by Bolt, Alder and Fernbach, Academic Press, N.Y.