7332-1001

IHLIB
TEXTBASE

INSTITUTE OF HYDROLOGY

Report No 40 May 1977

METHODS FOR EVALUATING THE U.K. RAINGAUGE NETWORK

by

P E O'Connell, M A Beran,
R J Gurney, D A Jones and R J Moore

ABSTRACT

This report outlines some of the basic properties of rainfall and elementary network design. This is followed by a literature review on raingauge network design and a review of user requirements. The methodology for both direct and indirect (i.e. rainfall/runoff modelling) evaluation of the UK network is described, followed by suggestions for modifications to the existing network to meet current demands more efficiently.

PREFACE

On 26 March 1976, a seminar was held at the Meteorological Office's College at Shinfield Park, Reading, on 'Meteorological Office Services (Other than Forecasts) for Water Management'. During the course of this meeting a lively discussion took place on the usefulness of rainfall data to the Water Authorities who have the major burden of collecting it, and to the Meteorological Office who have their own internal requirements. As a result of this discussion, Dr J C Rodda and Mr S F White of the Department of the Environment raised with the National Water Council the need for the Water Industry to make its requirements for rainfall data known to the Meteorological Office. As a result, the Water Data Unit proposed that an investigation should be carried out to determine whether or not the existing raingauge network satisfied user requirements. A specification for the investigation was agreed in May 1976 between the Institute of Hydrology and the Water Data Unit after the initial reaction of other interested parties had been tested. Funding for the project was agreed upon in July 1976; work began on 1 August 1976. The work described in Section 6 of the report was authorized in January 1977 and the report completed in May 1977.

The study divided into two main parts. The first dealt with study definition and with obtaining the views of the many rainfall data users on the adequacy of the network for their purposes. This phase of the project was handled by Mr M A Beran and was essentially completed in December 1976. The project was taken over by Dr P E O'Connell in October 1976 who was responsible for the analytical content of the report. Other major contributors were Dr D A Jones (statistician), Dr R J Gurney (spatial analyst) and Mr R J Moore (hydrologist). Computational and other assistance was provided by Mrs A E Sekulin, Mr F A K Farquharson, Mr M W Venn, Miss H Dracos, Mrs J French, Mrs P B Moore and Mrs H Brimacombe.

ACKNOWLEDGEMENTS

Mr C Folland and Mr J Tate of the Meteorological Office were carrying out a parallel study of the UK raingauge network for the Meteorological Office's own purposes and provided much of the information for this report. This has been included, with their permission, in Sections 4.5, 4.6 and 7.2.3. The authors of this report are grateful for this contribution and for the several exchanges of views on various aspects of the study. Useful discussion also took place with Dr J C Rodda of the Water Data Unit, who also advised at the study specification stage and provided some references. Mr D Fiddes of the Department of the Environment provided several useful forums for discussing the progress of the study. Numerous individuals from the organisations listed in Appendix B expended time and effort in providing the required information; for this the authors of the report are extremely grateful.

CONTENTS

1. INTRODUCTION

1.1 General

A raingauge network may be regarded as an integral part of an information system, the basic functions of which are the collection, processing and dissemination of rainfall data. These three basic functions are undertaken in response to the demands for various levels of information placed on the system by users. In general, it will not be economically feasible to design an information system such that the level of information required by every user is made available. Ideally, the design should then provide an equitable basis on which the costs and benefits associated with the design of the system could be allocated to the various users. In the case of rainfall data, however, it is usually not possible to quantify the benefits accruing to users from various levels of information and more empirical procedures have to be adopted.

In designing a raingauge network, attention has traditionally been confined to those aspects of the design and operation of the network that are concerned with data collection. The network provides the necessary information to achieve a statistical description of the rainfall process in time and space. On the basis of this description, rainfall quantities not directly measurable by the network can be estimated, and associated measures of error specified which will primarily be a function of network density. In principle, then, the network density required to meet stated error criteria can be determined when these criteria can be specified explicitly by users.

The specification of error criteria by users is not, however, a straightforward procedure. Where rainfall data form the direct basis of some decision, then the user can probably state an error criterion with confidence, as he will hopefully be able to evaluate the consequences of various degrees of error for the outcome of his decision. Methods of network design which have as their basis an error criterion stated explicitly for an estimated rainfall quantity will be referred to as direct methods in this report. Where rainfall represents but one input to a complex decision-making process, a model of the process is ideally required whereby the sensitivity of a decision and its outcome (measured in economic or other terms) to the level of error in the rainfall input could be evaluated. This is rarely practicable; by analysing the nature of the decision, an attempt is made to infer empirically what a tolerable error level is for the rainfall input. Sometimes, an error criterion can be stated explicitly for a variable which can be related functionally - or through a simulation process - to rainfall e.g. streamflow or soil moisture deficit. For example, by using lumped and spatially distributed models of the rainfall-runoff process, it should be possible to infer what the required raingauge configuration and density is to achieve a desired level of accuracy in modelling streamflow. Such an approach will be referred to as an indirect method of network design in this report.

The question of whether or not a network provides sufficiently accurate estimates of rainfall quantities is, as already noted, one which may be addressed within a statistical framework. In this report, the term 'accuracy' will be used in association with a statistical estimate, and is quite distinct from the issue of how close a measured rainfall quantity is to true rainfall. This latter question is obviously an important one as far as network design is concerned, in that it is governed primarily by such factors as the type of gauge used to measure rainfall and its siting. In this report, such aspects of network design are considered as being of an 'operational' nature, and are not considered; for a comprehensive review of the subject, the reader is referred to Rodda (1971). However, in carrying out a statistical analysis of the UK raingauge network in this report, account is taken of the fact that measurement errors may exist in recorded rainfall quantities. Other closely related aspects of network design, such as the processing and dissemination of rainfall data, are not considered.

The main intention in Section 1 of this report is to outline some of the basic physical and statistical properties of rainfall, and to introduce the reader to some of the basic notions underlying network design at an elementary level. This is followed in Section 2 by a review of much of the vast literature that exists on raingauge network design. In doing this, a representative selection of papers are reviewed in detail, and the remainder are included in Appendix A as a classified bibliography. In Section 3, a review of user requirements for rainfall data in the United Kingdom is presented. This information was acquired by approaching many of those organisations whose activities were thought to involve significant use of rainfall data; a list of these organisations is presented in Appendix B. For the purposes of network design, user requirements are reduced to stated error criteria for estimated rainfall quantities; details of the many and varied uses of rainfall in the United Kingdom are also presented in Appendix B.

Sections 4, 5 and 6 contain most of the methodology whereby an evaluation of the UK raingauge network may be carried out. This was accomplished by selecting two areas in the East and North of England, and applying the techniques presented in Sections 4 and 5 in these areas. Section 4 deals with the characterization of rainfall as a spatial correlation process, while Section 5 shows how measures of the accuracy of estimation of various rainfall quantities can be derived therefrom. The subsidiary problem of the likely effect of climatic change on network density has also been considered. The procedures described in Sections 4 and 5 essentially represent direct methods of network design. In Section 6, some rainfall-runoff modelling is undertaken to illustrate how indirect methods of network design might be applied. Section 7 then proceeds to carry out an evaluation of the existing network for the areas considered, and suggests how modifications to the existing network could be achieved to meet existing demands more efficiently. Some conclusions and suggestions for further work are finally presented.

As far as possible, the consistent usage of symbols throughout the report has been maintained. However, this was not possible in Section 6 without rendering the terminology inconsistent with that already well established in the literature. Thus, the use of symbols in Section 6 is internally consistent, but not necessarily so with other sections of the report. In view of the large number of diagrams presented in the report, these have not been presented at the relevant points in the text, but have been placed collectively at the end of each section.

1.2 The United Kingdom raingauge network

1.2.1 Storage gauges

The most extensive and dense raingauge network in the UK is of daily-read storage raingauges. There are approximately 7000 of these gauges conforming to a Meteorological Office standard of construction, observation practice and site: of these 4500 are sited in England, 700 in Wales, 1500 in Scotland and 300 in Northern Ireland. In England and Wales 3600 are owned by Water Authorities, 1400 by private observers and 300 are at Meteorological Office climatological stations. Some gauges have an extra large container and can be used for weekly or monthly measurements. This is necessary in remote sites where a daily visit is impractical; about one half of the 1500 Scottish stations are of this type.

The network has grown from very few gauges in 1860 to 3000 in 1900 and 7000 in 1977, mainly in response to specific needs, and its density is very variable; Figure 1.1 shows the mean distance between raingauges in mainland Britain and indicates clearly the influence of population density (see Section 7.2.3). Areas of special interest in some river headwaters used for water supply may also be identified as having a higher raingauge density. There is much less variation in density in Northern Ireland because the network is more recent, and has been more systematically planned.

The Meteorological Office collects and scrutinizes all data from the daily raingauge network. Doubtful values are sometimes checked with the observer and corrected if necessary before the data are archived. The Meteorological Office also arranges for site inspections, ideally biennially, although the growth of the network has far outstripped the staff capacity for these reviews. Arrangements are now being made for observer authorities to undertake their own inspections.

G.J. Symons established the British Rainfall Organisation in 1861; annual rainfall totals, and increasingly also monthly totals, have been published since. Daily and monthly totals for all stations which report to the Meteorological Office have also been archived on magnetic tape since 1961; these tapes, the "British Rainfall" tapes, have been used in this study. The data have been rigorously quality controlled, and are available within two years of the end of the observation period.

Monthly data for selected stations with long records are also available from the Meteorological Office on magnetic tape. Daily and monthly forms which were prepared by the original observers when the "British Rainfall" volumes of the time were compiled are also stored in the Meteorological Office archives. The data used in Section 4.3 were obtained from this source.

1.2.2 Continuously recording gauges

The network of autographic or continuously recording raingauges is relatively sparse and there are no current standards for instrumentation or observational practice. There are 1170 gauges in all, of which 830 are sited in England (60% are in the Midlands and South-East); 200 are in Wales, 90 in Scotland and 50 in Northern Ireland. A map of their locations has been published by the Water Data Unit (1975). The Meteorological Office operates about 100 of these. There is no planned system of quality control and archiving of the data from autographic gauges as has been developed for the daily-read network, although a catalogue giving site details is available from the Institute of Hydrology.

Organisations with research interests, such as the Institute of Hydrology and Bristol University, maintain networks of autographic gauges and digitise the results. The autographic gauges of the Meteorological Office and most other organisations record on paper charts, but most records have not been *digitised* and their data are thus not readily available for computational purposes. Some digitised records are available for North Surrey, and for the Cardington and Winchcombe experiments, all from the Meteorological Office. Digitised radar data are available for several areas, most notably for the Dee catchment.

1.3 A short physical description of the rainfall process

Precipitation occurs when large droplets of water are formed, so large that they may not be held in suspension in the atmosphere. These large droplets may be created either by coalescence or by accretion of smaller droplets onto particles in the ice phase. There is some controversy about the relative importance of each mechanism (see Mason, 1971), but both require a high density of water droplets in the atmosphere with a range of diameters. These droplets are produced by adiabatically cooling a parcel of air below its dew point, usually as a result of upward movement. Although the interaction of physical processes which occur when precipitation is formed is complicated and not fully understood, methods have been derived (e.g. Betts, 1971; Mason and Jonas, 1974) for estimating the duration and spatial extent of rising air parcels producing clouds; these authors show that the characteristic duration of these rising air parcels is about 30 minutes, and their mean diameter is about 1 km. Only under certain circumstances, such as in cumulo-nimbus clouds or depressions, can the vertical movement of an air parcel be prolonged sufficiently to allow precipitating droplets to form.

Systems containing air parcels whose upward movement results in precipitation may vary greatly in areal extent. They may be small scale, as with cumulo-nimbus clouds, or large scale, as with depressions, although in many cases a large-scale system may contain discrete smaller ones. The rainfalls measured at two sites under the same system will be more alike than falls recorded under different rainfall producing systems. Heavy falls also tend to be associated with rapid vertical movement within smaller systems (cumulo-nimbus clouds) and lighter falls with the slower vertical movement of larger systems (depressions).

Depressions cross the British Isles typically in about 10 hours, often with little modification to their structure. It follows that rainfall at widely separated places may be related. The measured falls may be related even if the separation of the measurement points is larger than the instantaneous diameter of the rain-producing system, if the measurement sites are within the envelope within which a system has precipitated during the course of its movement.

The study of rainfalls associated with any one rainfall producing system, whether a cumulo-nimbus cloud or a depression, necessitates the use of raingauges which accumulate rainfall over time intervals no greater than the life of the system. Autographic and daily gauges show the patterns produced by individual storms, but in the British Isles monthly gauges almost always show the aggregate pattern from several storms.

Parcels of air, whether in isolated cells, squall lines, or depressions, may be lifted vertically not only by convection but by topography, which may cause a parcel of air to be lifted if it is being advected. This effect usually intensifies already existing rain-producing cells, and results in highland areas having greater rainfall than lowland areas. The greater the change in altitude, the greater is the increase in rainfall; the longer the time interval over which rainfall is accumulated, the closer the relation between rainfall and topography.

Two storms separated in time by an interval longer than the time taken to dissipate the kinetic energy of the first may give falls that may still be related, because many of the features which produced the first storm may also cause the second (e.g. sea surface temperature anomalies), and because a storm may interact with its environment, thereby making a subsequent storm more or less likely (e.g. by modifying the relative vorticity field or by cooling the ground through precipitation). Further, because individual storm systems occur within the global atmospheric circulation, the time between independent events may be long. Smagorinsky (1967) considered that the smallest time for independence between events is at least three weeks; nevertheless, for the purposes of this study, measurements of rainfall events have been taken as independent if their separation in time is greater than the kinetic energy dissipation time of mid-latitude depressions, effectively about six days.

1.4 Statistical properties of rainfall

1.4.1 Descriptive statistics

Table 1.1 gives means and standard deviations of daily, monthly and annual precipitation totals for twelve stations in Northern and Eastern England; the regions thus referred to are shown in Figure 1.2. *Table 1.1 shows that* mean rainfall in Northern England is higher than that in Eastern England for almost all cases; the standard deviations are also higher. The greater values are from sites at higher altitudes, as shown particularly by station 77790, which is at a considerably higher altitude than the other stations shown, on the Pennines west of Wakefield.

Depressions travel from the west, and may be "rained out" before they reach Eastern England. The North of England receives proportionately more rain caused by depressions because it is nearer to the Atlantic Ocean than Eastern England. Depressions are more common and give more rain in winter, and so the North of England receives more rain in that season. In Eastern England, however, convection from isolated cells is more common: here the rainfall is greater in summer, when these cells occur more frequently. These seasonal effects are only slight, but are illustrated by two stations shown in Figure 1.3.

Table 1.1 Means and standard deviations of rainfall totals for daily, monthly and annual values, 1961-1974, Eastern and Northern England (mm)

Station Number	Daily mean	Daily S.D.	Monthly mean	Monthly S.D.	Annual mean	Annual S.D.
			EASTERN ENGLAND			
151238	1.35	2.74	50.0	22.9	600.0	89.9
156677	1.45	2.78	40.0	18.4	480.0	90.6
171992	1.62	3.19	53.8	24.2	645.6	112.7
175514	1.22	2.22	50.8	26.0	609.6	86.0
181126	1.50	2.98	48.0	25.8	576.0	80.1
191591	1.36	3.01	46.9	25.3	562.8	73.4
			NORTHERN ENGLAND			
17260	1.65	3.57	57.5	23.4	690.0	120.6
32189	1.55	3.11	51.8	26.5	621.6	96.6
37225	1.66	3.11	53.0	28.5	636.0	95.8
77790	3.16	6.23	106.1	40.6	1273.2	186.7
108956	1.85	3.27	71.0	36.1	852.0	131.6
127979	1.23	2.24	50.8	28.2	609.6	104.0

Footnote: Daily values were for days selected at regular 20 day intervals throughout the period

Table 1.1 shows that the standard deviation of daily values is about twice the mean; for monthly and annual values, however, the standard deviation is less than the mean. This illustrates the fact that as the sampling interval is decreased, so the increasing occurrence of zero values causes the frequency distribution to become more skew. Thus about half daily rainfall totals are zero, although this proportion varies with the area concerned; the time series of hourly rainfall totals contains many more zeroes. Extreme skewness resulting from a high proportion of zeroes may preclude the straightforward application of commonly used statistical methods, although devices such as variate transforms may sometimes assist.

The skewness of sample rainfall distributions is illustrated in Figures 1.4 - 1.9, which give histograms of daily, monthly and annual totals of rainfall at two stations, one each in Eastern and Northern England. Four categories of daily rainfall are shown:

(i) Days selected at regular 20 day intervals throughout the period.

(ii) Days selected when the mean daily rainfall at twelve stations within each area was above 2 mm.

(iii) As for (ii), but with a threshold of 5 mm.

(iv) As for (ii), but with a threshold of 10 mm.

For (ii), (iii) and (iv), no day was selected if a day within the previous five days had also been selected. This reduces the effect of serial correlation.

Figures 1.4 - 1.7 show that skewness is large for all four daily categories, but becomes smaller as the threshold for acceptance as a "rain-day" is increased. Monthly rainfall totals are also positively skewed (Figure 1.8), but annual totals are less so (Figures 1.9).

The histograms of Figure 1.4 - 1.9 also point to some differences between the rainfall regimes of Eastern and Northern England. The stations shown have similar means for all sets of data, yet histograms for the stations in Eastern England generally show evidence of higher values. This is caused by the greater proportion of rain produced by isolated convection cells, which tend to give larger totals than the rain from depressions predominant in Northern England.

1.4.2 Correlation properties of rainfall in time and space

A brief physical description of the rainfall process was given in Section 1.3, particularly in relation to rainfall-producing mechanisms within the U.K. As raingauge network design requires a statistical approach in which statistical models are used to describe the rainfall process in time and space, some knowledge of the physical mechanisms producing rainfall can assist in structuring such models. If rainfall is aggregated over longer periods such as months or years, then a high degree of spatial 'relatedness' will be much less. This will be particularly true of rainfall generated by convective cells which have a limited spatial coverage. Again, for shorter time intervals, rainfall at a point will, however, tend to be related to rainfall during preceding intervals at the same point; as the intervals over which rainfall is aggregated increases, then this 'relatedness' in time will tend to disappear, there being no obvious physical basis for its existence. However, the presence of an annual cycle in monthly values does mean that for example, December rainfall in one year will be related to November rainfall; however, deviations from the respective mean rainfalls for both months will, in general, not be related.

There are two possible statistical approaches to describing the temporal and spatial 'relatedness' characterizing the rainfall process. One approach seeks to establish a description directly in the time-space domain, and has as its basis serial and spatial correlation. An alternative approach is to seek a representation in the frequency domain through spectral analysis; however, for the purposes of network design, the former approach offers much more flexibility and results can be more readily interpreted. Accordingly, a time-space correlation approach will be adopted in this report.

The next question to be considered is what suitable measures of rainfall correlation in space and time can be adopted. Little choice exists here, as the great proportion of statistical theory relies on the use of the linear correlation coefficient. Consider two points in space with measurements of rainfall at concurrent time points denoted as X_t and Y_t; then the population correlation coefficient between X_t and Y_t is defined as

$$\rho(X_t, Y_t) = \frac{E\{(X_t-\mu_X)(Y_t-\mu_Y)\}}{\sqrt{\{E(X_t-\mu_X)^2 E(Y_t-\mu_Y)^2\}}} \quad .... (1.1)$$

where μ_X and μ_Y are the means of X_t and Y_t, respectively; this is sometimes referred to as the lag-zero cross correlation coefficient, and is invariably employed when characterizing spatial correlation in rainfall. Similar measures can also be defined between X_t and $Y_{t+\tau}$, $\tau = \pm1, \pm2, \pm3, ...$ In practice, $\rho(X_t, Y_t)$ and similar measures will be estimated from the available sets of observations on the processes X_t and Y_t. The correlation coefficient can also be used to describe temporal correlation in either of the time series X_t and Y_t; for the series X_t, the lag τ autocorrelation coefficient is defined as

$$\rho(X_t, X_{t-\tau}) = \frac{E\{X_t-\mu_X\}\{X_{t-\tau}-\mu_X\}}{E(X_t-\mu_X)^2} \quad .... (1.2)$$

Given an area with rainfall measurements at a set of P gauges, it is possible to compute $P(P+1)/2$ different values of $\rho(X_t, Y_t)$; for moderate areas within the U.K., P may well be in the order of 500-1000, so clearly this is a rather unmanageable description. A much more viable approach is to adopt one or more 'key' or 'central' stations within the area, and calculate estimates of $\rho(X_t, Y_{t,i})$, $i = 1,2,...,P$ where X_t denotes rainfall at the key station, $Y_{t,i}$ denotes the rainfall at station i and P is the total number of stations considered. These point correlation values may then be viewed as estimates of the ordinates of a surface which is theoretically smooth and continuous over two dimensional space, with the correlation between the central station and any surrounding point being described in terms of the coordinates of the two points. This leads to the concept of a spatial correlation function; in the case of rainfall, for reasons discussed in Section 1.3, this correlation function will decay with distance from the central station, the rate of decay depending upon the time interval of the rainfall process. Assuming for the moment that the true rainfall can be measured within the area, then the correlation at zero distance from the central station would obviously be unity. If it assumed further that the rate of decay of correlation with distance is the same in all directions from the central station then the contours of the spatial correlation function will be circular as illustrated in Figure 1.10(i), with the spacing between the contours defined by the rate of decay of correlation with distance. Such a correlation function can be described completely by a one dimensional representation of correlation with distance, i.e. all cross sections of the correlation surface are the same (Figure 1.10(ii)).

However, in practice, it is usually found that spatial correlation functions do not have the same decay rate in different directions; under such circumstances a correlation function is said to be anisotropic. Accordingly, a two dimensional description of correlation as a function of distance and direction is required. An illustration of an anisotropic correlation function is given in Figure 1.11.

It would obviously be convenient if one correlation function with one set of parameters could be used to describe an area; however as an area gets larger and as the relief becomes more variable, this may not be possible. If the parameters governing the decay rate and anisotrophy of the correlation function are found to vary from one central station to another, then the rainfall process is said to be spatially non-stationary or non-homogeneous with respect to the spatial correlation structure. Various levels of non-stationarity can obtain within an area: for example, (i) the mean (ii) the variance (iii) the parameters of the correlation function may vary spatially. It may be necessary to invoke one or more of these non-stationarity assumptions to provide an adequate statistical description of the rainfall process; this will be particularly true in hilly and mountainous areas. For a pair of central stations, illustrations of correlation functions for rainfall processes which are (a) isotropic and non-stationary, (b) anisotropic and non-stationary are given in Figures 1.12 and 1.13, respectively. Stationarity and isotropy are discussed further in Sections 2 and 4.

In order to establish the form of the correlation function for an area and estimate its parameters, estimates of the correlation between one or more central stations and surrounding stations need to be derived from the available data, and a correlation function fitted to the sample correlations. Here, it is necessary to reconsider the assumption made earlier that the correlation at zero distance is unity. Due to the fact that recorded point rainfall is subject to measurement error and is affected by local microclimatic irregularities, estimated correlation functions may suggest that the correlation at zero distance is less than unity. This point will be discussed further in Section 4, where procedures for estimating correlations and fitting correlation functions to daily, monthly and annual rainfall data from two areas in the U.K. are considered, and in Section 5, where its implications for network design are considered.

1.4.3 Derivation of network accuracy

Invariably, the main demands placed on a raingauge network are either for the provision of estimates of rainfall during a particular interval at a point or over an area. In the absence of a gauge at the point of interest, it is clear that some procedure must be established for estimating rainfall at the ungauged point. It is convenient that rainfall is spatially correlated, as this forms the basis from which point and areal estimates may be obtained. The reliability of such estimates will in general depend on distances from surrounding gauges, the level of correlation obtaining between rainfall at points separated by such distances, the magnitude of measurement errors at the gauged points, and the estimation procedure used.

The derivation of a point estimate involves an interpolation procedure whereby a set of weights is applied to the rainfall at the surrounding gauges. Various procedures can be used for deriving the set of weights; among such procedures only one will be optimal in the sense that the error variance of the estimated quantity will be minimal among all possible linear estimators. However, the derivation of the optimal estimators will, in general, involve more computational effort than simple procedures. Various point and areal estimation procedures are derived and discussed in Section 5.

In order to illustrate the principles of interpolation a simple example will be considered here. It is assumed that an estimate of rainfall is required at a point which lies at the midpoint of a line of length l connecting two adjacent gauges with measured rainfalls X_1 and X_2; stationarity and isotropy are assumed over the distances considered. The relative error variance or relative mean square error of an estimate may be defined as

$$\xi = \frac{E}{\sigma^2} \qquad \dots. \ (1.3)$$

with E, the error variance (m.s.e.) defined as

$$E = E\{(\hat{Y} - Y)^2\} \qquad \dots. \ (1.4)$$

where $\hat{Y}$ is the estimated value, Y is the true value and σ^2 is the variance of the rainfall process. Assuming for the moment that the measurement error at the two gauged points is zero, then the weights to be applied will be equal.

If the means at the two gauges are assumed known and equal to μ, then the most obvious interpolation procedure would be to estimate $\hat{Y}$ as

$$\hat{Y} = \mu + b_1(X_1 - \mu) + b_2(X_2 - \mu) \qquad \dots. \ (1.5)$$

with $b_1 = b_2 = 0.5$. This corresponds to simple linear interpolation and the corresponding relative m.s.e. of $\hat{Y}$ may be shown to be

$$\xi = \frac{3}{2} + \frac{1}{2}\rho(l) - 2\,\rho(\tfrac{l}{2}) \qquad \dots. \ (1.6)$$

where $\rho(l)$ and $\rho(\frac{l}{2})$ are the correlations between rainfall at distances l and $l/2$, respectively. If, for example, $\rho(\frac{l}{2}) = 0.90$ and $\rho(l) = 0.84$ then $\xi = 0.12$.

However, the application of weights of 0.5 is arbitrary; if it is required to use weights which give minimum m.s.e. then these weights may be shown to be

$$b_1^* = b_2^* = \frac{\rho(\frac{l}{2})}{1 + \rho(l)} \qquad \dots. \ (1.7)$$

which for the example quoted gives $b_1^* = b_2^* = 0.489$. The resulting minimum mean square error is

$$\xi_{opt} = 1 - \frac{2\rho^2(\frac{l}{2})}{1 + \rho(l)} \qquad \dots. \ (1.8)$$

which for the current example gives $\xi_{opt} = 0.1196$. There is little difference in this case between linear and optimal estimators; this will not always be the case. If $\rho(\frac{l}{2})$ and $\rho(l)$ are reduced to 0.8 and 0.65 respectively, then the linear and optimal relative mean square errors and 0.225 and 0.2242, thus illustrating the sensitivity of interpolation error to the level of spatial correlation.

If measurement error exists at the two gauges, and the relative variance of this error, $\eta = \sigma_\epsilon^2/\sigma^2$, is the same at both points, then equations (1.6) - (1.8) become

$$\xi = \frac{3}{2} + \frac{1}{2}\rho(l) - 2\rho(\tfrac{l}{2}) + \frac{1}{2}\eta \qquad \ldots\ (1.9)$$

$$b_1^* = b_2^* = \frac{\rho(\frac{l}{2})}{1 + \eta + \rho(l)} \qquad \ldots\ (1.10)$$

$$\xi_{opt} = 1 - \frac{2\rho^2(\frac{l}{2})}{1 + \eta + \rho(l)} \qquad \ldots\ (1.11)$$

The quantity η may be estimated as $(1 - \hat{\rho}(0)/\hat{\rho}(0)$ where $\hat{\rho}(0)$ denotes the estimated correlation at zero distance, as discussed in Section 1.4.2. If $\rho(0) = 0.95$, $\rho(\frac{l}{2}) = 0.90$ and $\rho(l) = 0.84$, then $\xi = 0.146$ and $\xi_{opt} = 0.144$, which illustrates that measurement error decreases interpolation accuracy. It should be noted here that $\rho(\frac{l}{2})$ and $\rho(l)$ refer to the correlation function of 'true' rainfall and that $\rho(0)$ relates to measured rainfall. It may be shown that measurement error imposes a theoretical limit on the accuracy which may be obtained from a given number of gauges, no matter how close the gauges are to the interpolation point.

The purpose of the foregoing short discussion and example has been to illustrate how spatial correlation can be used as a basis for interpolating point rainfall; this is also true of areal rainfall estimation. Provided a well defined analytical procedure is adopted, it is possible to derive expressions giving accuracies for the estimated quantities. The application of more general interpolation procedures under more general conditions is considered in Section 5.

1.6 Summary

In this introduction, the general problem of raingauge network design has been considered, and the approach adopted in this report outlined. In a brief review of the United Kingdom raingauge network, daily-read storage gauges have been distinguished from recording or autographic gauges. Data from the former are collected and transmitted to the Meteorological Office, where they are quality controlled and computer archived. For the much less sparse network of autographic gauges, no planned system of quality control and archiving exists.

Some physical properties of the rainfall process have been described with particular reference to rainfall producing mechanisms in the U.K. This provides information relevant to the structuring of statistical models of rainfall in time and space. Some basic statistics of rainfall in the U.K. have been considered through reference to some stations in the East and North of England. The use of correlation to describe the statistical structure of rainfall in time and space has then been considered in outline, followed by a brief illustration of how measures of network accuracy can be derived on the basis of spatial correlation.

Figure 1.1 Mean spacing of daily-read storage raingauges, 1977, in kms.

Figure 1.2 Areas of Eastern and Northern England used in this report

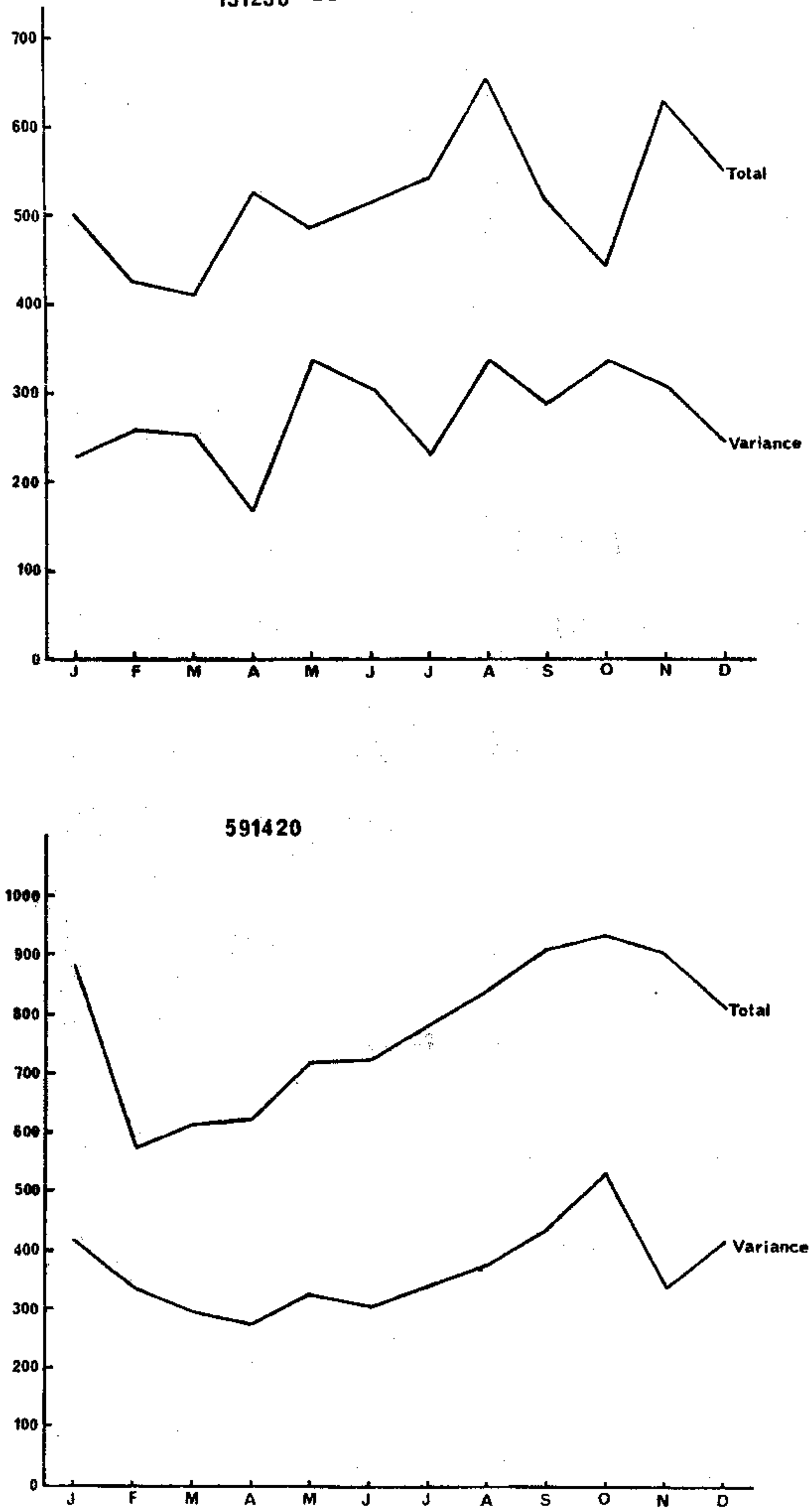


Figure 1.3 Variation with season of mean monthly totals and variances for two gauges in Eastern England (151238) and Northern England (591420)

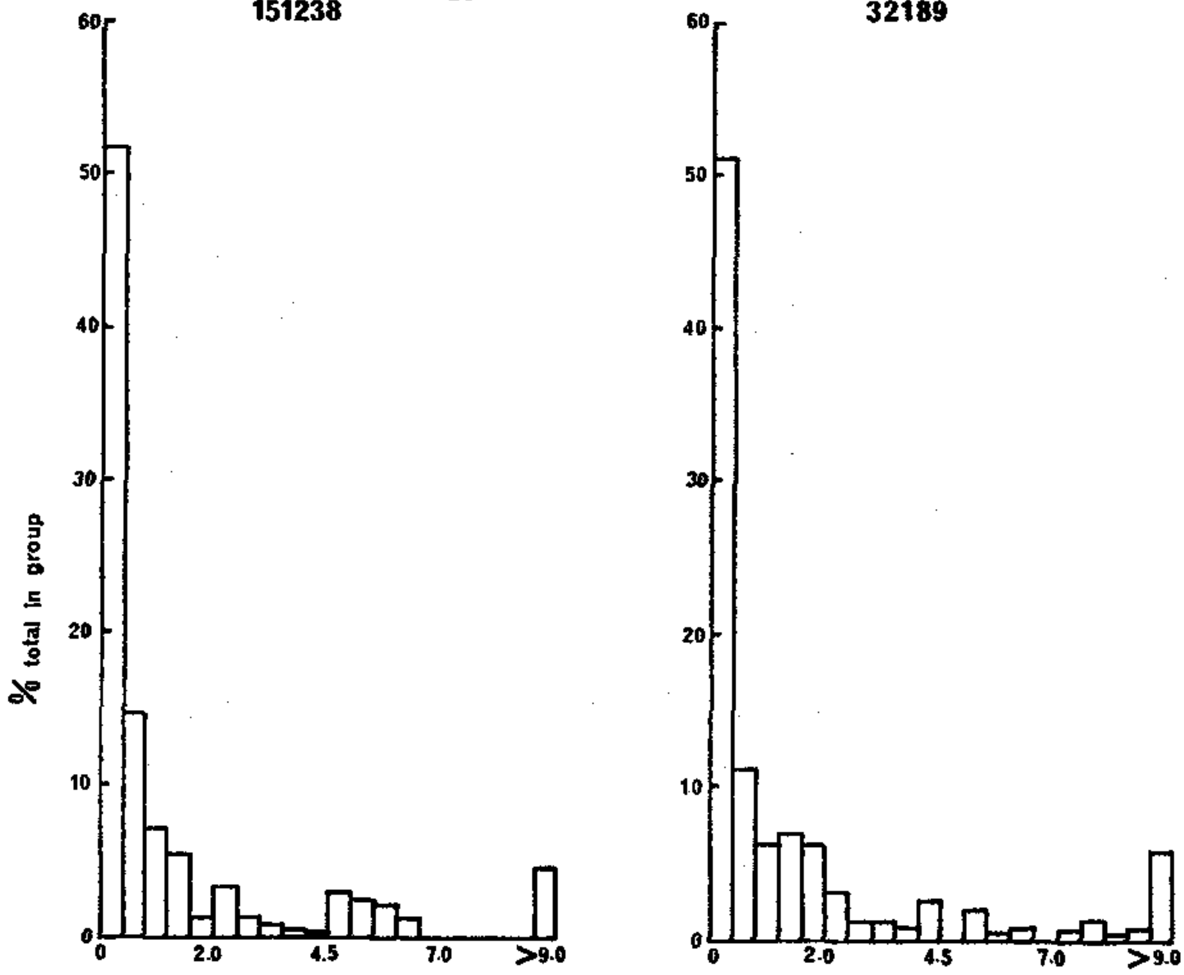


Figure 1.4 Histograms of every 20th daily rainfall total for two gauges in Eastern England (151238) and Northern England (32189)

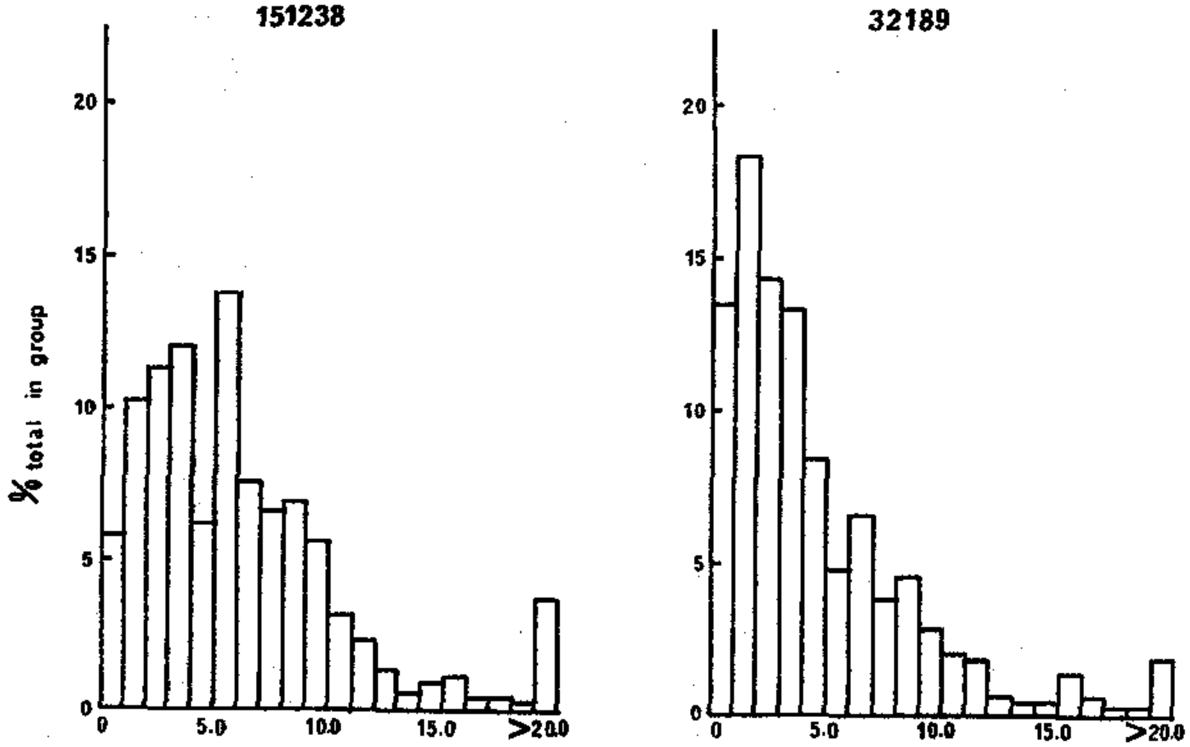


Figure 1.5 Histograms of daily rainfall totals for two gauges in Eastern England (151238) and Northern England (32189), for days with area mean ⩾ 2 mm

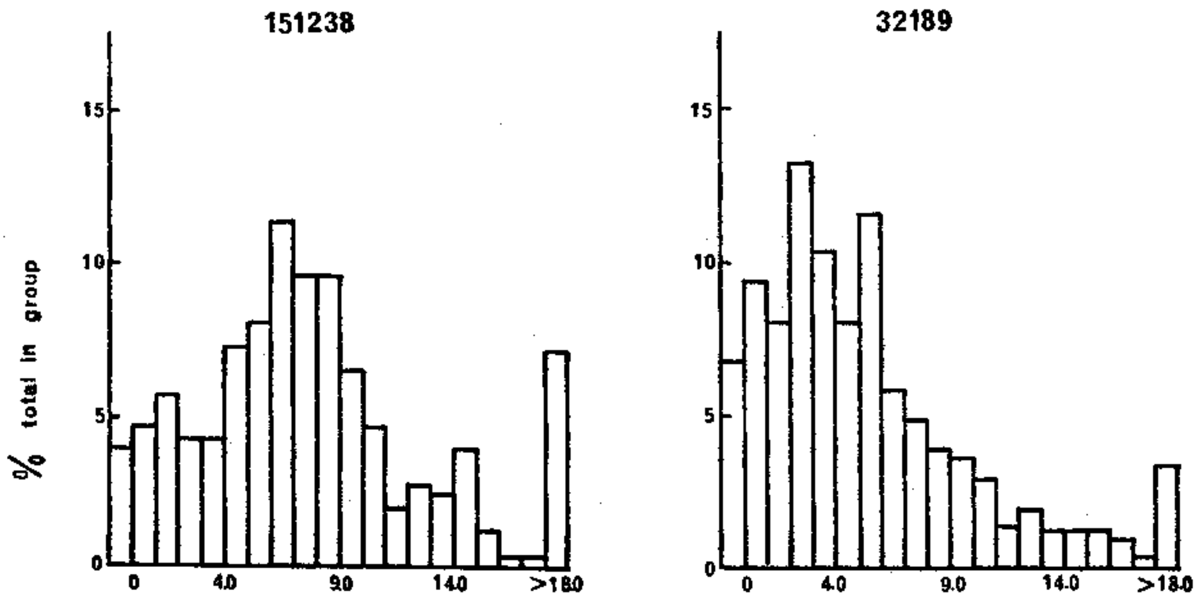


Figure 1.6 Histograms of daily rainfall totals for two gauges in Eastern England (151238) and Northern England (32189), for days with area mean ≥ 5 mm.

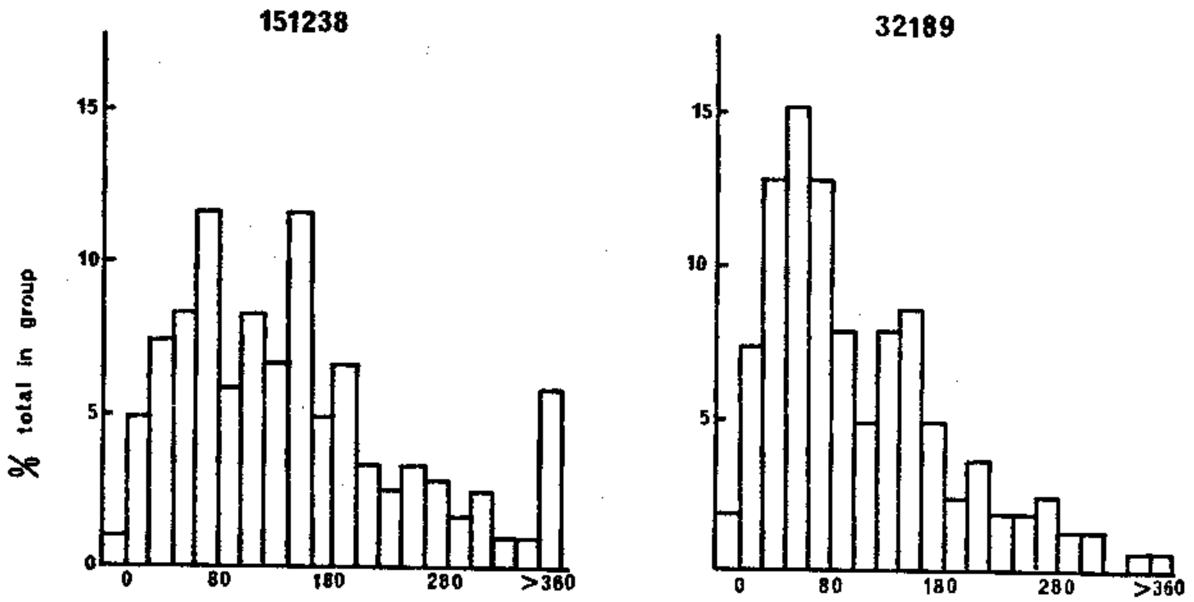
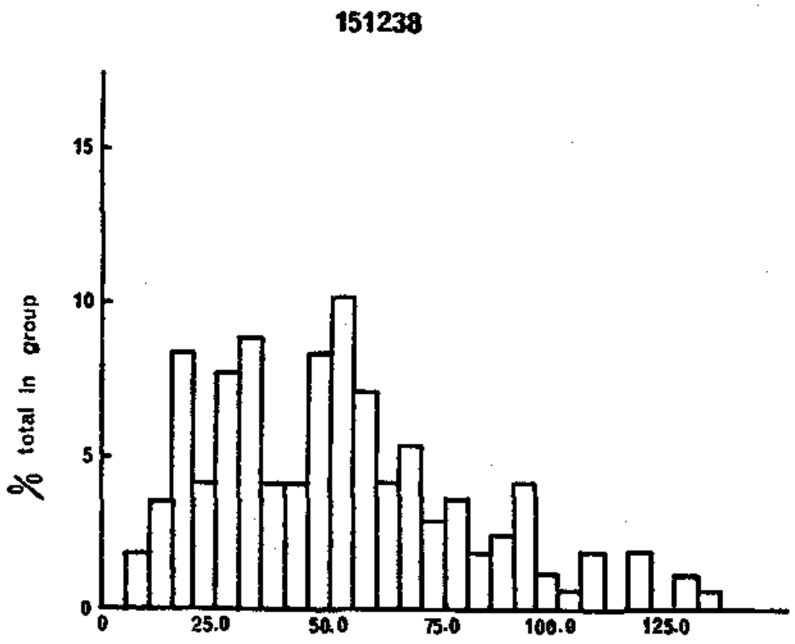


Figure 1.7 Histograms of daily rainfall totals for two gauges in Eastern England (151238) and Northern England (32189), for days with area mean ≥ 10 mm.



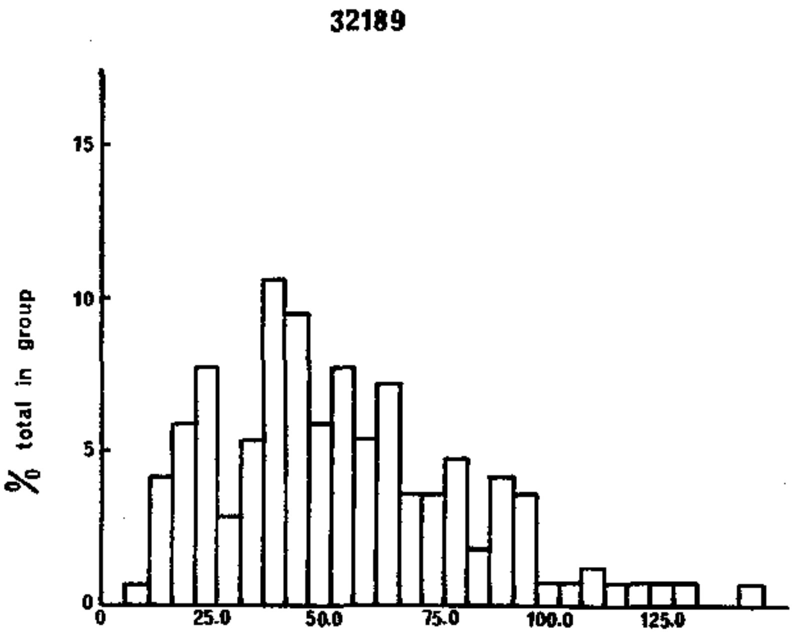


Figure 1.8 Histograms of monthly rainfall totals for two gauges in Eastern England (151238) and Northern England (32189)

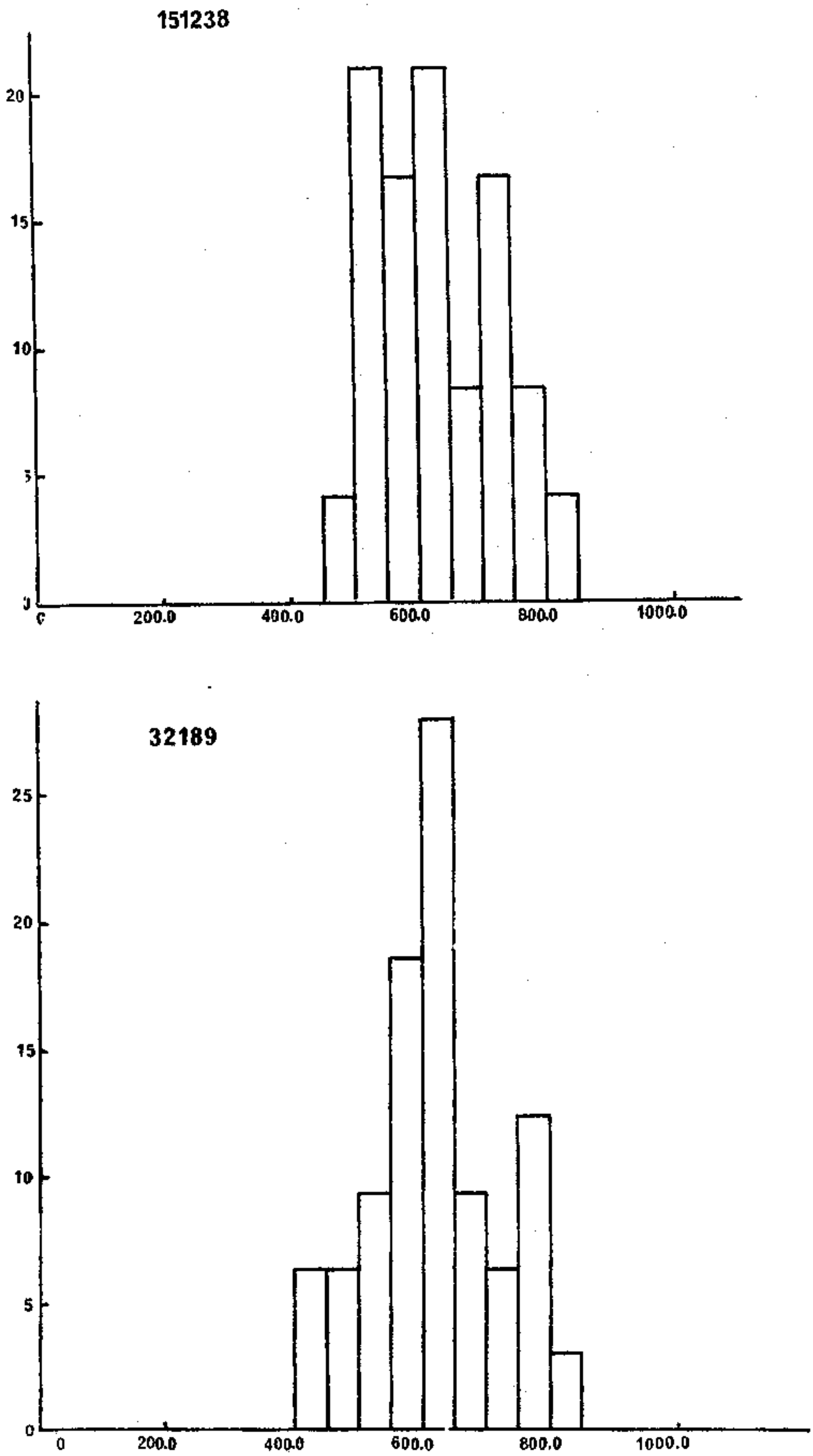
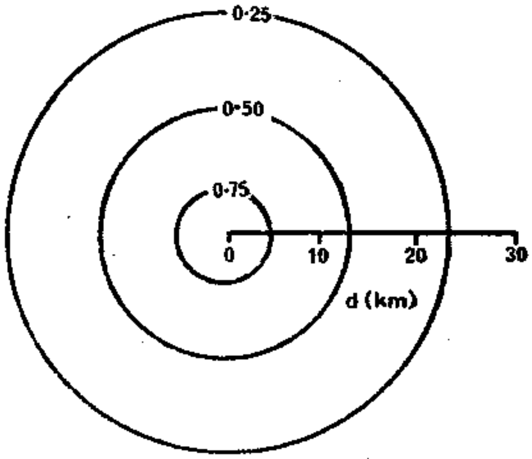
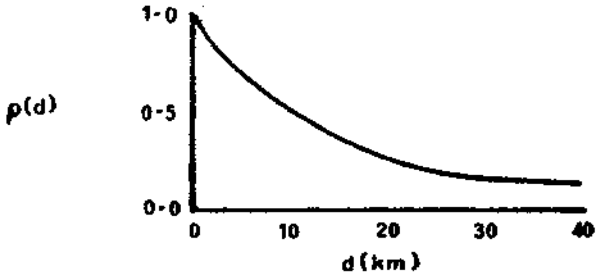


Figure 1.9 Histograms of annual rainfall totals for two gauges in Eastern England (151238) and Northern England (32189)



(i)



(ii)

Figure 1.10 A two dimensional (i) and one dimensional (ii) representation of a circular (isotropic) spatial correlation function

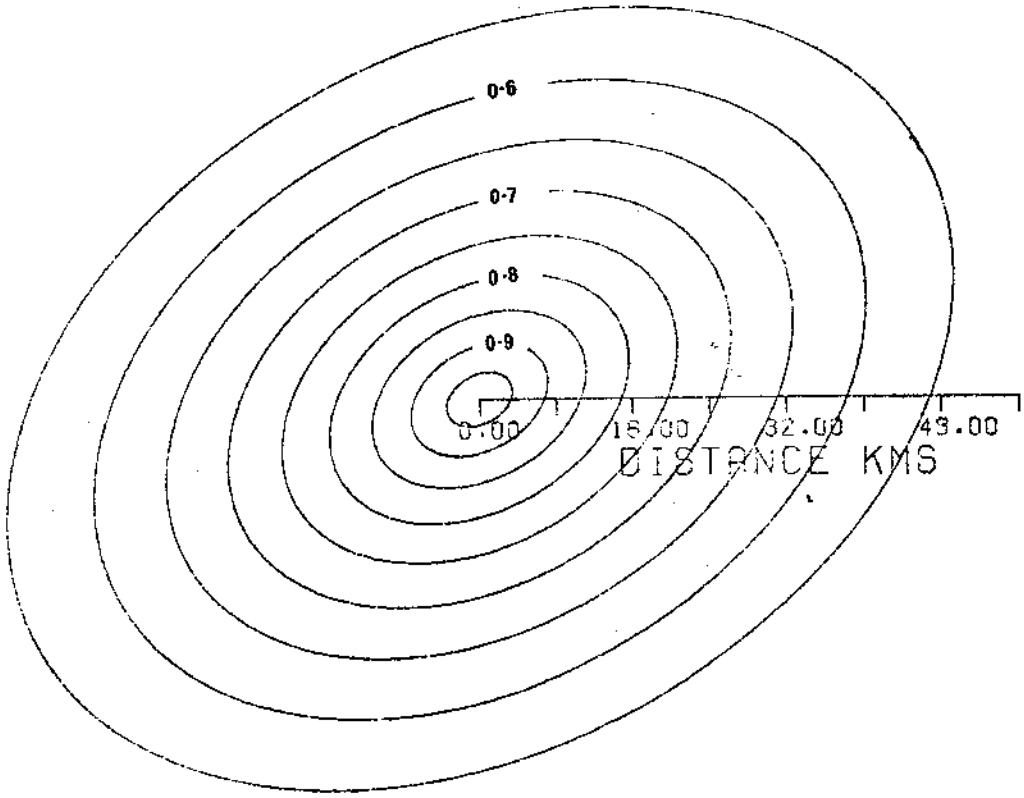


Figure 1.11 A two dimensional representation of an isotropic spatial correlation function.

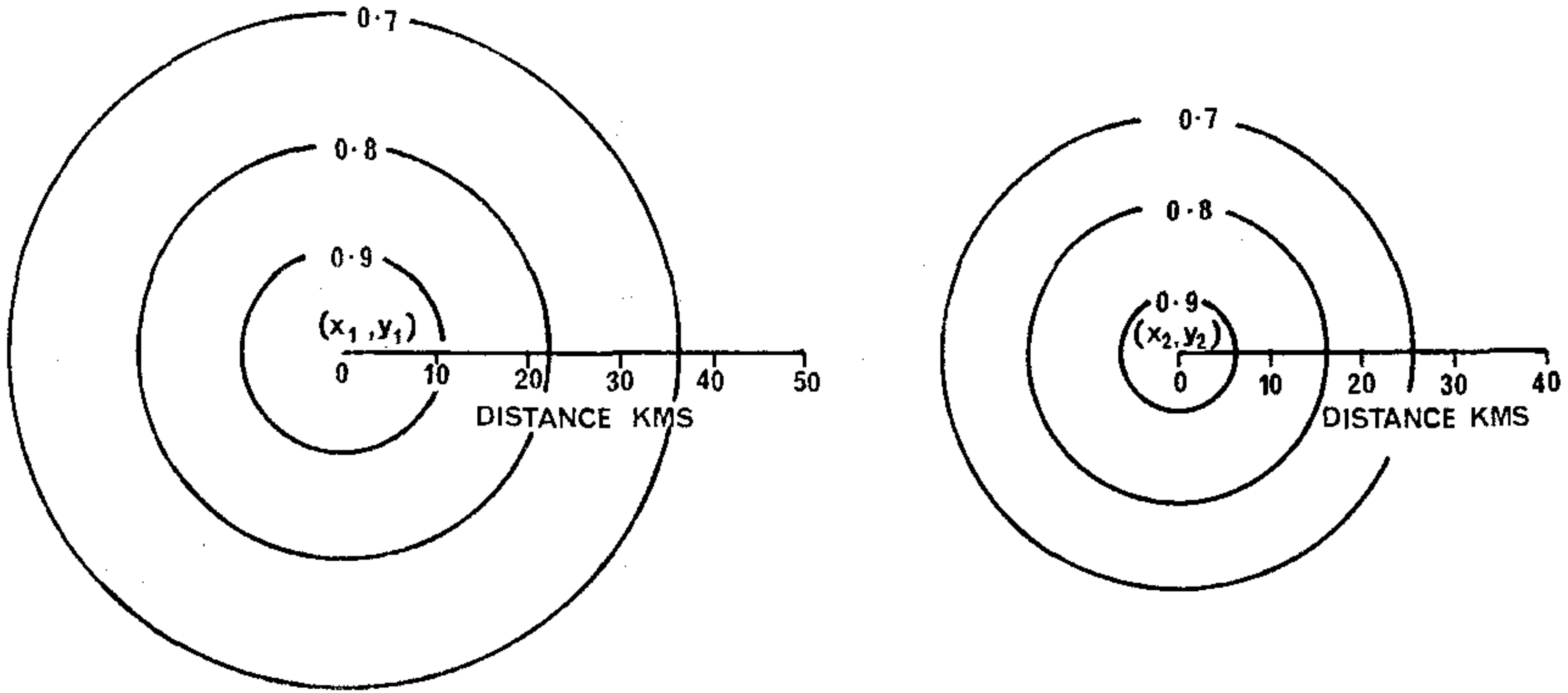


Figure 1.12 A two dimensional representation of an isotropic non-stationary spatial correlation function for two central stations (x_1, y_1) and (x_2, y_2)

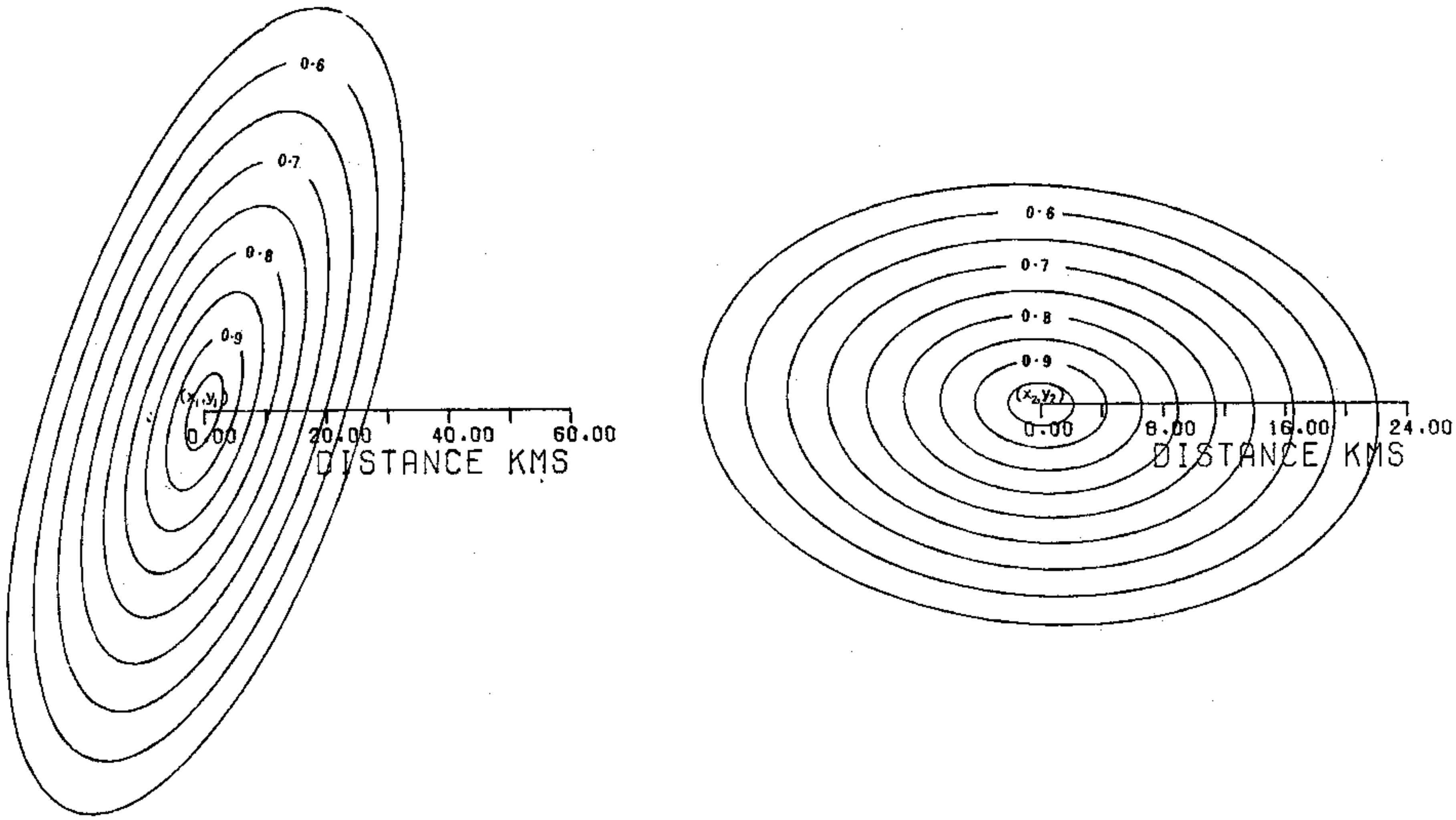


Figure 1.13 A two dimensional representation of an anisotropic non-stationary correlation function for two central stations (x_1, y_1) and (x_2, y_2).

2. RAINGAUGE NETWORK DESIGN : A LITERATURE REVIEW

2.1 Introduction

Many ways of designing rainfall networks have been described in the literature. These have all required the postulation of some criterion of performance, such as that the standard error of daily rainfall at any point is less than some given value; not all criteria or methods have been objective, and many so-called objective methods have used subjectively-chosen criteria. Some research workers, indeed, consider the involvement of subjective judgement to be necessary; thus, Kohler (1958), in a review of network design criteria, stated that "studies directed towards determining cost-to-benefit ratio or error versus network density are somewhat academic". The difficulty of siting raingauges has been considered by some, including Kohler, to be more serious than that of establishing the number to be deployed.

Rodda (1969) summarises much of the work on network design. He reports a useful distinction between three levels of network, used by the U.S. National Water Data System. These levels can be adopted for classifying precipitation networks, although their original purpose was to classify surface water networks.

Level I networks are used to acquire information for national planning purposes, to give gross estimates of water resources, to provide surveillance of major storms and to provide a national data bank.

Level II networks supplement Level I networks in particular basins or regimes, giving extra information for local planning.

Level III networks gather information for particular operational purposes for local water management.

The networks designed for each level need not be the same; indeed, the total network of raingauges in an area may contain components from all three levels.

Methods of raingauge network design may be divided into direct and indirect methods. Direct methods use some criterion calculated from observed rainfall totals such as the standard error of an estimate of areal mean rainfall; indirect methods, on the other hand, use some criterion such as the accuracy with which streamflow is predicted using a model, thus being dependent not only on rainfall but also on a description of the rainfall-runoff process. Some criteria that have been used for network design are considered in Section 2.2; direct methods of network design are considered in Section 2.3, and indirect methods in Section 2.4.

2.2 Some criteria used for network design

Most methods of network design used in the UK have been based at least in part on subjective criteria. The standard most widely used in the UK is that recommended by Bleasdale (1965), and given in Table 2.1. This table is based on the experience of Meteorological Office workers, particularly those who compile "British Rainfall", and represents the number of gauges which they feel gives a satisfactory representation of the rainfall in a given area.

Table 2.1 Minimum number of raingauges for monthly rainfall estimates

Area (km^2)	Number of monthly gauges
26	2
260	6
1300	12
2600	15
5200	20
7800	24

Clemesha-Smith et al (1937) recommended a standard for experimental and upland catchments, as reproduced in Table 2.2. This standard recommends a denser network than that given in Table 2.1, but Bleasdale considers that this is commensurate with the greater accuracy needed in special areas.

Table 2.2 Minimum number of raingauges needed in reservoired moorland catchments

Area (km^2)	Number of daily gauges	Number of monthly gauges	Total
2	1	2	3
4	2	4	6
20	3	7	10
40	4	11	15
81	5	15	20
121	6	19	25
162	8	22	30

The criteria adopted in both cases are that the raingauge distributions appear "reasonable" when plotted on a map. In effect, it is assumed that rainfall is related to topography, so that when isopleths of rainfall are compared by eye with contours, the agreement is close. However, it is not always certain that a simple relationship with altitude should be adopted, and, further, there is no objective measure of whether the agreement is reasonable. A similar approach is to compare isopleths drawn using a dense raingauge network with those produced using a subset of gauges. However, comparisons between maps are subjective; furthermore, this method is highly dependent on the interpolation technique used when drawing the isopleths, particularly if the sampling density is low.

Most workers, including Clemesha-Smith et al (1937) and Bleasdale (1965) recommend that precipitation should be sampled both spatially and altitudinally. This approach has been followed in the design of many experimental networks, for instance those described by Clarke et al (1973). Shaw (1965) considered that even this simple criterion could not be adopted, because of the difficulty of siting gauges, particularly on higher ground, and instead adopted a method where one raingauge was sited in each 26 km^2 in a 252 km^2 area when an operational raingauge network was designed in Devon.

Most network design-criteria have been evolved for networks to measure monthly or annual precipitation totals: less work has been done for shorter period totals, or for other precipitation statistics. This partly reflects demand and partly the difficulties noted in Section 1 of treating statistics of short period rainfalls. McCullagh (1975) described a method which was used to design a network of telemetering raingauges in the Trent valley. The method will be described in Section 2.3, but the criterion adopted for acceptable accuracy was that the standard error of a daily total at any point should be less than 4 mm. The number of gauges recommended was 53, in a 10436 km^2 area, or 1 per 197 km^2. The siting of these gauges was also discussed, several different sites and densities being examined in order to find the configuration that gave the lowest standard error of an estimate of daily areal rainfall, subject also to the criterion for point rainfall given above.

Radar has been considered by some workers as a method of reducing the network of conventional raingauges. Harrold et al (1973) state that radar is equivalent to between 1 and 4 gauges per 100 km^2, depending on the type of rainfall, for hourly rainfall rates. They use a mean absolute error criterion for making this assessment, as discussed in Harrold et al (1974). The use of radar to measure rainfall is discussed more fully in Section 3.4.

2.3 Direct methods of raingauge network analysis and design

2.3.1 Introduction

Many direct methods of network design consider the precision of interpolated rainfall estimates, and it is convenient to divide interpolation methods into two categories, local and global. Consider the rainfall interpolated at point A a distance d_B from gauge B, and d_C from gauge C. The value at A is statistically dependent on the values at both B and C, and in the absence of other information may be taken as the weighted sum of the values at B and at C. For local estimation the weights w_B and w_C applied to the values at gauges B and C depend on the distance d_B and d_C, whereas for global estimation the weights w_B and w_C are independent of d_B and d_C. The division into local and global methods is useful, even for techniques not based on interpolation.

2.3.2 Local fitting techniques

2.3.2.1 Outline

Only a few raingauges are needed in an area if the rainfall totals recorded at these raingauges are closely related; the converse is also true. Two measures of the relation between rainfalls at two gauges are:

(i) *Spatial correlation*: This has been used by Gandin (1963), amongst others, for assessing the relationships between pairs of gauges. If (X_t, Y_t) $(t = 1, 2, \dots N)$ are a set of measurements of rainfall over the same interval of time (t) at two stations, the usual estimates of the variances, covariances and correlation are

$$S_X^2 = \frac{1}{N-1} \sum_{t=1}^{N} (X_t-\bar{X})^2 \quad , \qquad S_Y^2 = \frac{1}{N-1} \sum_{t=1}^{N} (Y_t-\bar{Y})^2$$

$$S_{XY} = \frac{1}{N-1} \sum_{t=1}^{N} (X_t-\bar{X})(Y_t-\bar{Y}) \quad , \qquad r_{XY} = S_{XY}/(S_X S_Y)$$

where $\bar{X} = \frac{1}{N} \sum_{t=1}^{N} X_t$, $\bar{Y} = \frac{1}{N} \sum_{t=1}^{N} Y_t$ are estimates of the means at the two stations.

As discussed in Section 4.3, there are problems with using these estimates of variances, covariances and correlation because the set of measurements (X_t, Y_t) are not independent over time, and because rainfall totals for short periods are not normally distributed (Section 1.3.1).

(ii) Spatial variogram: This has been used by Matheron (1971), amongst others; Gandin (1963) also discusses the variogram, calling it the structural function. It is estimated as

$$\gamma_{XY} = \frac{1}{2N} \sum_{t=1}^{N} (Y_t-X_t)^2$$

If the variances S_X^2 and S_Y^2 are equal, then this can be reduced to the estimate of the correlation given above. The usefulness of this statistic will be illustrated later, but it may be considered asthe variance of the increments of a spatial process.

An estimate of the unknown value at a point can be found from values observed at other points. This can best be done by estimating the statistical relationship between the unknown value and the observed values, then weighting the observed values accordingly. The method of doing this using spatial correlation will be described and then differences when the variogram is important will be indicated.

The correlation between an unmeasured and a measured value may be estimated using the correlations between the values at measured points. It has already been noted that correlation varies with distance, and it is then a conceptually simple idea to estimate this relationship.

Plots of estimates of sample correlations against distance may be used to suggest a relationship between correlation and distance. A function is used to characterise this relationship, with parameters estimated using the sample correlations between measured points. Several functions have been fitted, most having an exponential decay with increasing distance. Gandin (1963; 1970) fitted a circular correlation function that is invariant with direction; for example, if $\rho(d)$ is the correlation at a fixed distance d and a, b are constant for each direction

$$\rho(d) = a + (1-a)\exp(-bd)$$

Rodriguez-Iturbe and Mejia (1973) used a modified Bessel function:

$$\rho(d) = bdK(bd)$$

where b is again constant for each direction; this was because Whittle (1954) could not associate an exponential correlation function with any simple physical mechanism. However, there was little difference between the derived network deviations using either correlation function.

Given a satisfactory relationship between correlation and distance, the value (Y) at an ungauged point may be estimated as a weighted sum of the values $(X_1,\ldots X_p)$ at neighbouring gauged points. If $(\hat{Y})$ is a linear estimator of (Y), and $(X_1,\ldots X_p)$ are known for the same time interval, then

$$\hat{Y} = a + b X_1 + b_2X_2 + b_3X_3 + \ldots\ldots + b_pX_p$$

where a_1, b_1, $\ldots b_p$ are known constants.

The accuracy of this estimator $\hat{Y}$ is characterised by its mean square error,

$$\mathrm{mse}(\hat{Y}) = E\{(\hat{Y}-Y)^2\}$$

It is also often required that $\hat{Y}$ be an unbiased estimator, i.e.:

$$\mathrm{bias}\ (\hat{Y}) = E\{(Y-Y)\} = 0$$

The values $a, b_1,\ldots b_p$ are found so that mse $(\hat{Y})$ is a minimum, and that the estimate $\hat{Y}$ is unbiased. Details of the method for calculating the weights is given in Section 5.2.1, which shows that, for estimation of the mean square error, it is necessary either (i) to estimate long term means or (ii) to assume they are unknown but equal. Schlatter (1975) considers that neither of the methods (i) and (ii) above is satisfactory, and recommends a check on the estimate of mean square error by omitting measurements from a few raingauges when estimating correlograms and weights; values are then interpolated to the raingauge sites not used in the analysis, and the estimated mean square error compared with the calculated mean square error. However, Sharon (1972) was doubtful whether a single mean square error criterion could be used to plan networks at the three levels mentioned in Section 2.1; he even doubted the usefulness of a mean square error criterion.

Gandin (1963) and others have assumed that the parameters of the correlation function do not vary in space. However, two types of variation may occur: first, the parameters may be direction dependent; second, the parameters of the function may vary from place to place. If the correlation may be assumed to be stationary under translations then it is possible to fit a direction-dependent function (Whittle, 1954; Schlatter, 1975). If the parameters are not constant from place to place, then it may be possible to allow for this variation if it is smooth and not discontinuous.

There are other assumptions of stationarity: namely, that variances are equal for all stations, and that means are either equal and unknown, or unequal but known, for all points. Matheron (1971) considers these assumptions too strict, and instead recommends the use of the variogram in a technique he terms Kriging (after D S Krige, a South African mining engineer who first suggested the technique). In Kriging only the spatial covariants of the process are assumed stationary; fitting a function to the variogram and assigning weights to the measurement points is similar to the correlation-based methods already described. Matheron (1971) uses simple functions to fit to the variogram, often fitting by eye.

All work based on these techniques has used a linear estimator of weights. However, as shown by Matheron (1975) non-linear estimators may be used, and these should improve the accuracy of estimates.

2.3.2.2 Applications of local fitting techniques

The use of local fitting techniques for network planning has been considerable. Three applications of the planning of precipitation networks will be described here as examples.

Cislerova and Hutchinson (1974) applied optimal interpolation for the redesign of the raingauge network of Zambia, using annual rainfall totals. Standard errors of 10% and 15% were chosen to represent the upper limits of accuracy of interpolation. The maximum admissible distances between gauges were then calculated, using a simplified version of optimal interpolation that calculated the standard error of interpolation only at the mid-point of a line segment midway between neighbouring stations. A mid-point of a line segment is assumed to have the greatest standard error of interpolation of any point between two gauges.

Delhomme and Delfiner (1974) considered two separate data sets : one consisted of annual rainfall totals, and the other of totals for thirteen storm events in a catchment in Eastern Chad. For each data set, the standard error of interpolation was found at each point on a fine grid by Kriging. It was assumed that a new gauge would be sited at the position (B) which had the largest standard error. The gain in precision was estimated for each data set, by simulating the exercise of estimating standard errors assuming an extra gauge was established at position (B). The greatest standard error of interpolation was noted, and the gain calculated by comparison with the initial estimates. The reduction in the standard error of areal estimates was also found. No criteria of the reduction in standard error necessary before it was decided to establish a new gauge were stated in the paper.

A similar approach was adopted by McCullagh (1975) except that he did not calculate the reduction in standard error given by the establishment of a new gauge. He used the map of standard errors of interpolation to decide if a new gauge should be established. Using daily totals he adopted a criterion that a point should have a minimum standard error of interpolation of 4 mm before a new gauge was established.

2.3.3 Global fitting techniques

2.3.3.1 Outline

Global fitting techniques are more diverse than local fitting techniques and do not use a common method; they are most useful when the standard error of estimate of areal rainfall is needed, as they give an identical estimate of the standard error for all points. Most global fitting techniques which have been used for raingauge network planning have used the correlation between precipitation records for two time periods to estimate the standard error of areal rainfall totals. Some further techniques have been based on spectral analysis. These two will be described in turn.

2.3.2 Correlation techniques: outline

These techniques are based on the work of Nicks (1965) and of Sutcliffe (1966). Sutcliffe (1966) estimated the increase in the standard error of mean areal rainfall when a gauge was removed. He separated spatial variations of annual rainfall into a persistent pattern, a systematic error and a random error.

Thus,

$$X_{it} = P_{it} (1 + a + b_t) + \varepsilon_{it}$$

where x_{it} is the measured rainfall at time t (t = 1,....,N) at site i(i = 1,....,P), P_{it} represents the persistent element of the true rainfall, a and b_t are constants representing persistent errors, and ε_{it} is a random error. A linear relationship is assumed between the long-term true areal mean rainfall $X^*_{..}$ and the persistent element P_{it}. The persistent errors b_i and the relationships between long-term point means and areal means are assumed to be independent between gauges. The correlation coefficient $r(X_{i1}, X_{i2})$ then depends mainly on the error terms ε_{i1}, ε_{i2}, and it is possible to find the standard error of the estimate (X_{i1}, X_{i2}). This result may be used to assess the accuracy of a network, and of a particular gauge, if records exist for two independent time periods to allow the assessment of $r(X_{i1}, X_{i2})$.

2.3.3.3 Correlation methods : examples

Herbst and Shaw (1969) used Sutcliffe's method to assess the network in two catchments in South Devon, and three other catchments in southern Britain, when estimating monthly areal rainfall. The results obtained varied very widely. It was suggested that the densities estimated using the method varied because different types of rainfall occurred in different catchments. A disadvantage of the method is that very long records for several gauges are needed in order to give stable estimates of long-term areal rainfall for two independent time periods.

Nicks (1965) applied a Student's t-test to compare reduced and full networks of gauges, using data from only one time period collected from an instrumented network in Oklahoma (158 gauges in a 2900 km^2 area). Daily totals of rain $\geq$ 0.01 inches at any one gauge over a 3-year term were used, giving a total of 411 days. Mean daily rainfall was calculated for each day, using the full network, and several reduced networks. Areal means for each reduced network were compared with those for the full network using Student's t. Only one network, that with 5 gauges, showed significant differences from the full network of 158 gauges; however, as the areal mean values for the networks are not independent these results should be treated with caution.

Stephenson's (1968) analysis, similar to Nicks (1965), used daily and monthly total rainfalls in Somerset. The results for daily totals were inconclusive, because of the extremely non-normal sample frequency distributions. For monthly data, he found that at least 30 gauges were needed to assess mean areal monthly rainfall in a 5200 km^2 area.

2.3.3.4 Spectral methods

Eagleson (1967) suggests that a spectral representation of the spatial rainfall pattern is useful, if it allows a clear identification of a frequency above which there is very little power; this may then be taken as the Nyquist frequency, thereby providing an upper bound for the spacing of gauges. The grid spacing would be chosen as twice the Nyquist frequency. However, sample spectra of rainfall do not exhibit any clear cut-off, certainly for any time period of integration greater than one minute. Ord and Rees (1976) consider that spectra could be produced using radar data, but this has not been done for network planning. As the definition of a Nyquist frequency is rather arbitrary, this method appears subjective if applied to network planning.

A similar approach was adopted by Hershfield (1966), using the spatial correlogram rather than the spatial spectrum. He calculated a mean correlogram for fifteen storm events in fifteen catchments. The mean distance to the 0.9 correlation level was noted and used to give a gauge spacing. This correlation level was arbitrarily chosen, and no account was taken of anisotropy in the correlogram. Variations in spacing were noted for storm events with different return periods.

2.4 Indirect methods of raingauge network analysis and design

2.4.1 Introduction

Indirect methods of network design seek to optimise some criterion such as the error in prediction of streamflow or soil moisture deficit, or in some other derived quantity required by the user. If direct methods of network design are rejected, two approaches are possible : the first requires the use of a model relating rainfall to runoff (or some other hydrological variable) whilst the second uses a cost-benefit analysis to compare the costs of alternative networks with benefits accruing from their implementation. These will be considered in turn. Most network design using indirect methods has used criteria based on streamflow prediction; this is therefore given greater emphasis in this section, although there is no reason why other user requirements cannot be considered similarly.

2.4.2 Rainfall-runoff models

Distributed rainfall-runoff models offer the best hope of studying the effects of spatial variations in precipitation on runoff, but most are still under development and have not been used extensively for network design purposes. Usually, lumped models have been used, where in effect the number of gauges required to satisfy the model is that which gives a stable estimate of areal rainfall. One example of each type will be described.

Richards (1975) used a lumped model developed by Dickinson and Douglas (1972) to study the effect of network reduction on goodness of fit and streamflow prediction. Three catchments were used in modelling, with areas of 197 km^2, 1.5 km^2 and 19 km^2, and model parameters were fitted by least squares to three-hourly streamflows. Two catchments had only one autographic raingauge each, while the third had three autographic gauges. Areal mean rainfalls were then calculated for each three hour period, using Thiessen polygons. These areal mean rainfalls were calculated for five gauge configurations in each of the three catchments. For each network configuration, model parameters were estimated, together with the coefficient of determination (R^2) derived from observed and fitted three-hour streamflows. Few differences were found in the values of R^2 for different configurations. This could be the result of inadequacies of the model and of the fact that the model was recalibrated for each network configuration. Thus, the effects of modelling errors and inadequate input definition on R^2 would be difficult to separate. The method of distributing the daily totals using only a few autographic gauges may have reduced the effect of gauge configuration.

Bras and Rodriguez-Iturbe (1976) used a non-linear, spatially-distributed model based on the kinematic wave approximation to the St. Venant shallow water flow equations. Rainfall was represented as a first-order Markov process, giving realisations (continuous both in time and space) that were converted into point totals over discrete time intervals. To estimate the covariance structure of the rainfall process, it was necessary to assume (i) that measurement errors at different points were independent, (ii) that the process was stationary over time, and (iii) that measurement errors and rainfall intensities were independent. Estimates of the covariance structure were made (for all points) using a Kalman filter technique, from those points at which measurements were taken. The use of a Kalman filter allows the mean square error of estimation to be obtained, independently of the actual observations. In the prediction of runoff, infiltration and evaporation is ignored, and so the only uncertainty allowed in estimating the runoff comes from the stochastic nature of the rainfall.

The approach was tested on an idealised basin of 212 km^2. A time increment of one hour was used; a family of storms with a mean velocity of 32 $km\ h^{-1}$, a mean duration of 6 hours, and a mean maximum fall of 38 $mm\ h^{-1}$ was modelled. The hourly rainfall intensities were assumed to obey a single exponential correlation function. They conclude that peak estimate mean square errors are reduced when gauges are located in the upstream reaches of a catchment, but this increases the estimation error of the rising stage of a storm hydrograph. Station location was found to be more important than the number of stations, but it was found also that the mean square error was considerably reduced when up to eight (the maximum number) stations were added to the measurement network. It was found that the falling limb of a storm hydrograph could be estimated accurately using only a few gauges, but many more gauges were needed to estimate the rising stage and peak discharge with concomitant accuracy.

There are some limitations in this method. The assumptions are very restrictive, particularly in assuming no infiltration. This considerably increases the estimate of mean square error. The results are dependent both on the description of the spatial structure of the rainfall and on the rainfall-runoff model, although the authors noted that perturbing the parameters of the correlation function describing the rainfall process had little effect on the results. These also refer only to one type of storm, and would have to be generalised before any network design could be implemented. The numerical complexity of the method leads to high computational costs. The authors do not suggest an acceptable mean square error of streamflow prediction, so this problem is unresolved; nevertheless, the method offers some physical basis for network design, and allows assessment of the contribution from individual gauges to the modelled streamflow.

2.4.3 Decision theory models

Models using decision theory relate the costs of a measurement network to the benefits accruing from it. This is often performed by simulation; usually the benefits from only one use are considered because of the difficulty of assessing benefits from many of the uses of raingauge networks. This approach illustrated by an example where flood warning benefits have been assessed.

Grayman and Eagleson (1971) simulated relationships between the costs of network provision and the benefits of flood warning. Both radar and raingauge systems were evaluated in several stages: first, a model was designed to estimate the covariance structure of rainstorms; second, various sampling schemes were tested, to find their effect on the sample covariance structure; third, a distributed rainfall-runoff model was used to convert the sample rainfall values to runoff hydrographs. Each of these stages involved a simulation study, with many storms, measuring systems and catchments, one combination of which was subject to an economic analysis, using decision theory. Many assumptions and simplifications were made in this work. In particular, the assessments of costs and benefits and the relationships between variables were simplified. In only one case study was a full economic analysis performed. Other case studies relating raingauge networks to runoff were similar to those of Bras and Rodriguez-Iturbe (1976: see also 2.4.2). In the case study including an economic analysis, an area of 16637 km^2 in Pennsylvania was taken. Radar was shown to have only a very small additional benefit compared with raingauges alone, and a network of one gauge per 492 km^2 was recommended.

A network design approach with its basis in decision theory has been developed by Moss and Karlinger (1974) and Moss (1975). While the method has been evolved primarily for streamflow network design, its significance warrants mention here. Attention is focussed on a measure of the information provided by a streamflow network, "equivalent years of record" (Hardison, 1969) and a methodology is evolved whereby a set of network designs can be generated which satisfy accuracy criteria specified a priori for various streamflow parameters. Sample statistics of streamflow will inevitably differ from population statistics; consequently, incorrect decisions about whether or not a network meets an accuracy requirement may be made when sample statistics are used in lieu of population statistics. The strength of the approach developed by Moss and Karlinger (1974) is that full account is taken of this uncertainty in generating feasible network designs by utilizing Bayesian decision theory. If suitable economic cost-benefit relationships are available, an optimal design can finally be chosen.

2.5 Conclusions

The most flexible direct method of network planning is provided by a local fitting technique, if the assumptions of the method can be met. Methods using the correlogram to describe spatial variations are more closely based on widely-used statistics than those using the variogram.

Indirect approaches have been described through a few examples, all of which are specific in that they relate only to one use of rainfall information. No study has been made of the effect of model choice, or of using different assumptions when fitting the model. Potentially, methods using decision theory could be generalised, but the difficulty of relating rainfall to all its uses through models, and then assessing the economic benefits of the predictions, is very great. Thus, indirect approaches offer considerable promise for the design of special purpose networks, but at present are too specific for use in designing a general purpose network.

3. USER REQUIREMENTS FOR RAINFALL DATA IN THE UNITED KINGDOM

3.1 Introduction

This chapter discussed user requirements for rainfall data. These were established by asking various organisations to describe the role played by rainfall data in their operations. Presenting this information in Section 3.3 serves to demonstrate to data collectors the vast range of tasks in which their data are employed, and also helps to identify a core of uses on which the adequacy of the existing network may be judged and on which recommendations for revision of the network may be based. Section 3.2 gives the method of describing accuracy requirements and Section 3.3 describes briefly rainfall usage (a fuller description is included in Appendix B. Section 3.4 discusses the possible future role of radar in rainfall measurement; Section 3.5 summarises the requirements which may be used to assess the existing network.

3.2 Purpose of approaching data users

A sample of organisations from the water industry and other important sectors was approached by the Institute of Hydrology and each visited organisation was asked to specify its requirements in terms of a tolerable error, either proportional (e.g. 10%) or absolute (e.g. 5 mm). A similar cooperative enquiry was conducted by the Meteorological Office who circularised their own branches, plus agro-meteorological representatives and urban hydrologists. By considering the consequences to the data user of a more serious error the stated figure was assessed as that error required not to be exceeded on either 68%, 95% or 98% of occasions corresponding to either 1, 2 or 2.33 standard errors.

Users were also asked to specify whether they required point or areal rainfall estimates and, if the latter, over what area; they were also asked the time interval for the rainfall totals they needed (hour, day, etc). These aspects, spatial and temporal resolutions, were more easily answered, usually being evident from the nature of the application.

With this information, and the methods presented in Sections 5 and 6, it is possible to judge and, in some cases, test whether the existing network meets the users' specification. It is also possible to design an optimal network catering for any individual use. However, this is not the purpose of this investigation; a national network as defined in Section 1 should provide a sufficient density of raingauges to permit all major users to interpolate rainfall values to sites or regions of interest.

Many of the organisations approached expressed views on the total system of collection, quality control, archiving and dissemination of rainfall data. Views were expressed regarding instrumentation, regularity of central reporting, raingauge location problems, time of reading (9 to 9 or midnight to midnight), extra gauges to compensate for lost records and rainfall quality as well as quantity. While all these subsidiary questions are of paramount importance in the ultimate network design this report takes the view that they relate more to the operation of the network, and focusses attention on those error and resolution matters that impinge directly on the network density problem.

3.3 Organisations approached

The following sections give a brief outline of the uses of rainfall data by various sectors. Appendix B lists all the organisations who assisted in acquiring the necessary information and also gives a much fuller description of individual uses and requirements.

3.3.1 Water industry

This industry is taken here to include all the Water Authorities in England and Wales and corresponding organisations in Scotland and Northern Ireland as well as Central Government departments and research organisations that service the Authorities. The Water Industry has a special place in that, not only is it a major data user, but it is also the main data collecting agency. The data from many raingauges nominally in private hands are initially collected by the Water Authorities before being transmitted to the national archive.

Every Water Authority, representative Scottish Purification Boards and Regional Councils, and the Northern Ireland Department of the Environment were visited. Their uses may be for operational purposes, using real-time data for a specified point or catchment, or else for a planning purpose which demands either retrospective data or rainfall statistics.

The calendar month is the most common time interval for those operational and planning tasks that require data as a time sequence. It is the interval used in water balance studies of all kinds, e.g. resource studies, ground-water recharge, gauging station checks, and modelling. Monthly totals are required for areas corresponding to useful surface or groundwater resource units, perhaps typically 100 km^2 but sometimes as small as 10 km^2. A commonly quoted error criterion is ± 10% which is not to be exceeded on more than 2% of occasions, i.e. corresponding to 2.33 standard errors. Totals over shorter time intervals, 5 days or occasionally 1 day, are sometimes useful in 'fine tuning' of operating rules both in water resource and quality schemes. However, these are more likely to relate to dry low flow periods. In some types of study absolute accuracy of rainfall data is required and a network of ground level gauges may be called for to calibrate the standard Mk II gauge in almost universal use.

Flood analysis and warning, especially in urban areas, requires short duration rainfall data obtained from gauges reporting or telemetering their data to some control centre. Because their location is totally determined by this specialised use these gauges cannot be regarded as part of the national network, although their location and spacing is an important local problem.

For the analysis of past flood events it is necessary not only to know the total volume of rain, for which the daily data normally suffice, but also the timing and variation of intensity with time within the storm. This latter requirement calls for autographic raingauges at a sufficient density to sample storm events. It is not thought to be practicable or necessary to space such raingauges in accordance with the known rapidly declining correlation distance relationship (Rodda et al, 1976). This indicates that a practically unattainable gauge spacing would be required to achieve an interpolation accuracy of similar order to measurement accuracy. In any case hydrological models, even those that permit a spatially varying input, introduce considerable errors of their own in converting input to output, so network demands cannot be based on the time and space input capabilities of the model, only on the sensitivity of model output to various levels of description of the rainfall input. This is the approach adopted in Section 6.

A separate approach might be to base the spacing of autographic raingauges upon some such criterion as that, in rural areas, 50% of storm cells should be sampled and, in urban areas, 95% should be sampled. This has not been developed further in this report but would require data on the distribution of storm cell size, track speed and direction. Some such information has recently become available in Shearman (1977).

Rainfall statistics are required for some planning purposes. The most important statistic is the mean annual rainfall averaged over some standard period. Its importance is due to its central role in very many informal decisions. For example, when obtaining a rapid assessment of the rainfall regime at a particular site it is very common simply to compare that site with another of known performance on the same isohyet. Other common uses for annual average rainfall are in standardising other more complex statistics such as quantiles of short duration rainfall or accumulated deficits, and also in describing storm event isohyets using the isopercental method. The assumption that more complex statistics can be standardised to smooth results and aid mapping has had the effect of imposing the spatial detail of annual average rainfall on the spatial pattern of the other statistics. Because of sampling error in the derivation of the statistics, especially those related to extremes of rainfall surplus and deficit, it is not possible without very long records to test the hypothesis that annual average rainfall is an adequate standardising factor. On the other hand the maps so produced seem to serve the purposes of the users of the statistics. Thus in this study it has been assumed that the network density needed for rainfall statistics is the same as that for annual average rainfall.

3.3.2 Civil engineering, consultancy, building and construction

Consulting engineers operating in the water field were found to have generally similar rainfall data requirements to the Water Industry. However, being outside the industry they do not have immediate access to the data and are more dependent upon the national archive.

A brief survey of the construction engineers' requirements revealed that apart from their considerable dependence on accurate weather forecasts they had little continuing or systematic requirements for data. Large sites have found it necessary to keep rainfall records to confirm such matters as time lost due to rain. Indices of the suitability of site conditions for plant operation or building activity could be constructed using rainfall statistics. Generalised statistics on time lost due to rain had been assembled and were in use for planning work programmes. Sites in different parts of the country have been compared by reference to rainfall intensity statistics.

The choice of wall cladding and roof construction materials and such factors as drain gully, soakaway and gutter sizes are assisted by the availability of rainfall statistics. Special purpose indices have been developed by building researchers and have been mapped; these enable average site conditions to be predicted and wall and roof materials to be chosen to provide adequate protection. Because this use of rainfall data is as part of an index, and thus only partially related to the necessary protection standards, a limited raingauge network suffices. There is one specialised requirement for information on driving rain which necessitates special purpose-designed gauges. Snow is an important factor in roof design and although the raingauge network is used also to record the water equivalent of snow the data only partly satisfy the requirements of roof and other design work. Thus proper consideration of this large topic would need to include other methods of recording snow depth and it has not been included in this study.

3.3.3 Agriculture

Agriculture is by far the largest industry with a direct interest in rainfall data. Although much of the use is informal the scale of the more quantitative use can be realised from the fact that up to 11000 trials or experiments take place each year, many of them requiring parallel climatic information. The rainfall interest is rooted in the critical role of rainfall in determining plant and animal growth and health.

The structure of the industry is such that operational decisions are made about rainfall sensitive activities, such as irrigation or assessing disease risk, at a regional level by the Agricultural Development Advisory Service Officers using a limited network of reporting stations. In Scotland and Northern Ireland however, Colleges of Agriculture undertake some of these advisory roles. The required time interval for rainfall data is very variable depending on the particular purpose, totals over one to three days being common. Data is normally required for a single point, but, except for experimental purposes, accuracy requirements are not high. Other climatic variables such as evaporation are of course very important, thus calling for a network of weather stations.

For planning purposes rainfall data are essential for deciding total irrigation requirement and suitability of a range of agricultural practices such as animal housing standards, and studying the year to year variations in disease development and crop yield. The agricultural requirement for rainfall data is not limited to lowland Britain: the needs of forestry, liver fluke forecasting, and bracken burning all call for a coverage of upland regions.

In common with the water industry the agriculture sector is interested in soil moisture deficit (SMD) both operationally and as statistical information for planning.

3.3.4 Meteorology

The Meteorological Office currently processes the data from some 7,000 standard raingauges. Internally its interest in the data is for validating forecasts and developing new forecasting models, producing SMD maps and answering enquiries. In the future forecasting models will be based upon a 25 km grid hopefully enabling a 6 hour rainfall forecast to be reported and checked. An accuracy requirement of 0.5 mm (1 standard error) has been quoted for this purpose when the rainfall forecast exceeds 10 mm. The Enquiry Branch of the Meteorological Office answers a large variety of enquiries including those covered in the following two sections. An accuracy requirement of 5 mm for daily rainfall both for point and areal estimates has been quoted by the Enquiry Branch as being sufficient to meet these enquiries.

Miscellaneous services provided by the Meteorological Office concerning instruments and observational practice perform a very relevant function to network operation but do not themselves place demands for data upon the raingauge network.

Some meteorological and climatic research is carried out at other centres and quite commonly calls for a dense network over randomly selected parts of the country. The Institute of Hydrology, for example, has a copy of the National Archive of daily rainfall data which it uses in a variety of statistical studies. A recent one required a density of one gauge per 30 km^2.

3.3.5 Public utilities

Public utilities like the Central Electricity Generating Board and the North of Scotland Hydro-Electric Board use rainfall data for demand forecasting, pollution monitoring and identifying problems with generating sets. Telecommunication is affected by intense rainfall and both retrospective and real-time data are used for route planning and monitoring by the Post Office.

3.3.6 Law, insurance, health

The police in their capacity of providing emergency relief, traffic management, and occasionally in forensic work, also use rainfall data. It is possible to insure against rainfall so statistics are required for premium calculation and a nearby observer and gauge has to be available. Rainfall is an important agent in depositing pollutants from the atmosphere. Several organisations are involved in developing instrumentation for the biological and chemical analyses of rainfall, and also the calculation of mass balances. This latter aspect requires rainfall depth data but as yet, because rainfall quality data is available over such a sparse network, this is not a stringent requirement. It is foreseen that networks for the measurement of rainfall quality and quantity will be needed adjacent to major sources of polluting materials.

3.4 Use of radar

3.4.1 Introduction

The Central Water Planning Unit is at present working on a project to evaluate the benefits of radar to Water Authorities (Water Resources Board; Bussell, 1976). Collaborators in this study include several Water Authorities, the Water Research Centre, the Meteorological Office and the Royal Signals and Radar Establishment. Radar has already been used as a research tool in the study of cloud behaviour by the latter two organisations, and by the Appleton Laboratory and some universities. It has also already been applied to particular hydrological tasks both in the UK (the River Dee Research Program) and in the USA. However, little has been published about its potential in providing general purpose data such as that provided by the raingauge network for non-specialised users, i.e. a digital archive of small-area rainfall depths. The potential of radar for this more general application is reviewed here and is based upon discussions with all the establishments mentioned above, and on the literature concerned with the accuracy of rainfall estimated by radar.

3.4.2 Current radar experience

The main impetus for quantitative rainfall measurement using radar has arisen in flood warning where the accuracy requirement, although perhaps more stringent than for the meteorological application referred to above, has still not been high. This is because the methods for forecasting streamflow use models with spatially lumped inputs and the models are themselves a source of further uncertainty. On the other hand radar has many benefits which it would be quite impracticable to reproduce using telemetering raingauges. These include the blanket coverage of all catchments - not just a few selected ones - and also the visual immediacy of the information, which could possibly be used to give a forecast for a useful time ahead. These advantages and others are well described in various CWPU papers and papers by the 'Operations systems group' of the Dee Weather Radar Project, which has also been concerned with comparative costs of radar and conventional equipment. Problems associated with operating and interpreting the radar data have been described in an extensive literature (e.g. Huff (1967), Wilson (1970), Harrold et al (1974), Brandes (1975), Hill et al (1977)).

The performance of radar relative to raingauges over an area has been tested using the Dee Weather Radar. The radar reflectivity, after conversion to rain depth using a small number of calibrating raingauges, was compared with the data from a dense network of up to 76 raingauges. Because of radar's blanket coverage the problem of interpolation does not arise, and thus it is necessary only to consider the comparative accuracy of a radar estimate and an estimate based upon interpolating between raingauges.

Differences between rainfall estimates derived from radar and raingauges for various durations and areas have been derived as part of the Dee research program. The results are quite impressive and indicate strongly that radar has a use in a system dedicated to flood warning.

3.4.3 Further evaluation work

It is felt that because the Dee results (Collier, 1975) were obtained for a specific purpose the following extra factors would need to be taken into account before a fair comparison with a general purpose network can be made.

a. The spatial pattern of the radar reflectivity was used to aid the interpolation of the isohyetal pattern of the 76 raingauge network and it is felt that this would pre-dispose towards a close agreement, especially when considering smaller areas than the catchments mostly considered in the Dee Research program. Some more comparisons should be made using an independent method of areal weighting.

b. Radar and raingauge comparisons were invariably carried out for particular events. It is felt that the extension of the comparisons to cover data for clock hours or calendar days and months would be desirable. The existing comparison was made largely over durations when rain was known to occur and, while this is valuable, many network uses require a daily, monthly or annual time scale. Indications from work in the USA (Wilson, 1975) are that a mean absolute error of between 5% and 10% only may be possible for monthly data even for very large areas.

c. Although the radar reflectivity was calibrated on a point comparison with a check gauge, conversion factors are currently only worked out for specified catchments. If radar were to be substituted for the raingauge network, factors for converting radar reflectivity to rain depth would have to be calculated for every point. While it is not envisaged that a calibrating raingauge would be needed for each element of the radar image the present method of calibration would need revision in terms of the density of calibrating gauges and the frequency of calibration.

d. The Dee study indicated that drift between the radar measuring level and the ground surface due to wind was not an important factor for the purposes investigated. However, further studies will probably be necessary to improve performance down to small areas, such as are appropriate in many urban hydrological and agricultural applications.

e. Intuitively it might be thought that radar performance in flatter terrain might be better than that over mountainous country such as the Upper Dee; however, this is not necessarily so. The treatment of areas not seen by the radar clearly has to be considered as it will involve merging the data from radar and other sources. Further experimental sites such as that at Clee Hill will add to the understanding of location criteria.

f. At present radar data are not readily available to general purpose users and if it were to become so then the details of how the information could be digitised, archived, and, most importantly, disseminated would have to be worked out. This would have to include such considerations as a skeleton back-up network of conventional raingauges to infill data during periods when a radar installation is not working.

g. A more comparable measure of error than 'mean absolute error': $\bar{Y} = \frac{1}{N} \Sigma |(\text{Obs} - \text{Est})/\text{Obs}|$ should be adopted. If the distribution of differences is normal then about 40% of observation estimate pairs would differ by more than one absolute error. The type of error criterion that users have in mind when an accuracy requirement such as ± 10% is quoted is that say not more than 1 in 20 or not more than 1 in 100 observations depart from a standard. The following table gives equivalent error values between the two schemes.

Table 3.1 <u>Conversion of mean absolute error to standard error</u>

Mean absolute error ($\bar{Y}$)	Error expressed as probability of greater standardised difference (Obs - Est)/Obs		
	32% (1 s.e.)	5% (2 s.e.)	1% (2.6 s.e.)
5%	6.3%	12.5%	16.3%
10%	12.5%	25.1%	32.6%
15%	18.8%	37.6%	48.9%
20%	25.1%	50.1%	65.2%
25%	31.3%	62.7%	81.4%
30%	37.6%	75.2%	97.7%

Footnote: This table assumes that the standardised differences are normally distributed. Thus 1 mean absolute error is equivalent to $\sqrt{2/\pi}$ = .798 standard deviations. If for example a mean absolute error (mae) of 10% is quoted then 32% of errors will exceed 12.5%, 5% will exceed 25.1% and 1% will exceed 32.6%. This implies that if a user accuracy requirement of ± 16.3% is specified with the intention that it should not be exceeded more than 1% of the time the corresponding mae would need to be 5%. A reduction to 10% would require an mae of less than 5%.

It is clear from a comparison of the quoted accuracy of radar estimates with the stated requirements of users of rainfall data that radar as presently used and tested is capable of meeting some, but not all, of their needs.

3.4.4 Conclusions

It is concluded from this review of radar capability that radar is a useful tool for those purposes where the blanket coverage, immediate availability and visual display that it provides are important. Basin management for purposes of river regulation and flood warning is perhaps the best example of these but other important uses such as the analysis of past floods both on rural and larger urban catchments, agricultural advice on irrigation, spraying and cultivation, preparation of SMD maps, many types of general enquiry, and telecommunication routing would all stand to gain from the use of radar to augment the network.

On the other hand pending the further research outlined in paragraphs (a) to (g) above, radar measured rainfall is not at present a proven substitute for standard gauges for those network purposes where there is a more stringent accuracy requirement. It is therefore proposed in this report to note the potential of radar rainfall measurement but otherwise to consider the raingauge network only on the basis of the performance of conventional rainfall measuring equipment.

3.5 Criteria for network design

3.5.1 Summary of user requirements for sequential data

Following a review of operational and planning uses of rainfall data a large variety of different requirements have been quoted. A set of predominant requirements has been identified and these are to be used in evaluating the present network and could be used for redesigning it.

The requirements are quoted in terms of area and time interval accuracy of estimated rainfall, and an error criterion expressed as percentage of observations in which the stated error bounds should be met, equivalent to 1, 2 or 2.33 standard errors. Table 3.2 shows these for the various uses.

In preparing Table 3.2 only the more critical users have been listed and some users whose requirements are mirrored by other organisations have not been mentioned separately. As already stated the frequency with which the permitted error should not be exceeded has been subjectively assessed. The flood analysis use has also been subjectively assessed. The flood analysis use has also been subjectively assessed in terms of frequency of measurement of storm cells. As has been explained in Section 3.3.1, this is because the actual manner in which the data are interpolated is itself subjective and not readily amenable to interpretation in the usual way without a considerable research effort. Unfortunately time did not permit further work on storm cell statistics.

3.5.2 Summary of user requirements for rainfall statistics

Items marked * in Table 3.2 are for statistics of rainfall. One cannot equate the requirements for these statistics with those for the basic element, for example 2 day rainfall, from which the statistics are developed. In fact the true variability is not known - maps are drawn based upon the existing network in the belief that this is sufficiently dense to represent its variability. In practice the annual average rainfall map is very commonly used to interpolate many rainfall statistics and although, as explained in Section 3.3.1, it is difficult to test the validity of this method, the resultant maps appear to be satisfactory.

It is recommended that the network density sufficient for point assessment of annual average rainfall to ± 25 mm at 98% of sites should be considered as sufficient also for the interpolation of such statistics. This is, of course, contingent on there being a sufficient length of data at the sites to calculate the point values of the statistics within tolerable limits.

Table 3.2 Summary of rainfall requirements

Use	Time	Space	Permitted error	With what frequency
Water balance	1 month	50 km^2	10%	98%
Flood volumes	1 day	10 km^2	20%	95%
Antecedent flood conditions	5 day	10 km^2	20%	95%
Rural flood analysis	Life of cell	Convective cell size	- -	Measurement on 50%
Urban flood analysis	Life of cell	Convective cell size	- -	Measurement on 95%
Flood design *	0.5 hr to 2 day	Point	10%	95%
Urban flood analysis *	5 min	1 km^2	10% or 0.5 mm	98%
Building interruption *	$\frac{1}{2}$ hour	Point	20%	95%
Irrigation	2 days	Point	5%	98%
Seed germination	$\frac{1}{2}$ month	Point	5 mm	98%
SMD map	1 day	20 km x 20 km	5%	95%
Forecast validation (rain > 10 mm)	12 hours	50 km x 50 km	0.5 mm	68%
Long range forecast	$\frac{1}{2}$ month	10^4 km^2	5%	95%
General enquiries (areal)	1 day	100 km^2	5 mm	98%
General enquiries (point)	1 day	Point	5 mm	98%
General enquiries (average)	1 year	Point	25 mm	95%
Telecommunications *	10 sec	Lines 1 km to 60 km	20%	95%

4. SPATIAL CORRELATION ANALYSIS

4.1 Introduction

Essentially rainfall is measured at individual points : how good a record this provides of the actual precipitation at neighbouring ungauged points is determined by how closely related the rainfalls are at different points in the area. One measure of this "relatedness" is given by the spatial correlation structure of the measured rainfalls. In this section the application of correlation analysis to two areas in the United Kingdom is discussed. Section 4.2 gives a brief description of the data used in this analysis. The estimation and fitting of correlation functions is discussed in Sections 4.3 and 4.4 while Sections 4.5 and 4.6 give a description of some related work by the Meteorological Office.

4.2 Rainfall data

4.2.1 General

The type of rainfall information readily available was outlined briefly in Section 1.2. Obviously, for an extensive data analysis, computer compatible data are desirable; for U.K. rainfall these are provided by the "British Rainfall" tapes and the MARAIN tapes produced by the Meteorological Office.

The "British Rainfall" tapes currently give daily and monthly totals of rainfall for all daily and monthly gauges in the U.K. for the period 1961-1975. The data for recent years have been rigorously quality controlled, both manually and by computer (Shearman, 1977). These tapes provided the data for most of the work described in this report. The MARAIN tapes give monthly rainfall totals for about 400 selected stations in the U.K. These stations were selected for their reliability and their length of record. However, according to Craddock (personal communication), the period which had a rainfall regime least like the present regime, and for which reliable records are available, preceded the period covered by the MARAIN tapes. Data were therefore extracted from the manuscript records of the British Rainfall Organisation for the period 1875-1890. Eastern England was selected, there being many gauges operating at that time in that area. However, the British Rainfall Organisation archives are only now being sorted, and so the data for only 38 stations were accessible. The data for these 38 stations were coded onto computer cards, and used in the analysis described in Sections 4.3 and 4.4.

4.2.2 Data extracted for present study

Each "British Rainfall" tape contains one year's rainfall data for the whole of the U.K. Thus to access the records for one station for the period 1961-1975 involves loading 15 tapes. While this is not in itself difficult, the amount of information on each tape means that it is not practicable to keep more than one or two complete years of data in computer main memory at any one time. Thus it is necessary to write programmes not only to retrieve the data stored on tape but also to select certain data of interest to be used in later analyses. Even if data for one area only is selected, with, for example, 600 raingauges, the size of the associated data files can quickly become very large after fifteen years of information have been included.

Monthly totals and daily totals (in five categories as described in Section 1.4.1) were extracted from the British Rainfall tapes for two areas, Eastern and Northern England (Figure 1.2). It was found that for some years the monthly totals were not identical to the sum of the daily totals because of differences in the quality control applied to daily and monthly values. For this study monthly totals were calculated by accumulating daily totals. There are many gauges with short records only: in order to reduce the size of the data files, only those stations which had more than 48 months of data during the period 1961-75 were selected. Five categories of daily data were used. Daily rainfall values are serially correlated and therefore the data must be selected to satisfy the underlying assumptions of independence, since serial correlation makes any spatial correlation analysis much more complex. For four of the selected categories of daily rainfall, data were extracted such that serial independence was guaranteed, while for the fifth category serial dependence was accounted for. As clear from Section 3, different users require different types of data : four of the five categories were chosen to examine how some of these requirements could be met. The categorization was achieved using twelve selected stations within each area. If the mean rainfall for the twelve stations in an area for each day was above a set threshold, then the records for all stations in that area on that day were included, as long as no day within the previous five days had also been included. This last proviso was set to allow for the effect of serial correlation. The thresholds chosen were over 2 mm mean rainfall, over 5 mm mean rainfall, and over 10 mm mean rainfall. A fourth category of data was chosen which allowed for both wet and dry days : for this, every twentieth day in each year was selected starting with the 20th of January. The fifth category was designed to examine whether serial correlation among the daily values might be used to improve the accuracy of interpolation procedures: to this purpose, using a 5 mm threshold as in the second category above, the data for each rain-day and the previous two days were extracted.

Annual totals were also compiled from monthly data. Those stations with less than ten years of record were excluded. There were 374 stations in all used in the analysis of Eastern England annual totals, compared with 672 for daily and monthly analyses. 704 stations were used to provide annual data for the North of England, compared with 1153 for daily and monthly analyses.

4.3 The estimation of sample correlations

4.3.1 The principle of correlation estimation

If rainfall is measured at a number of points in an area, a natural way of estimating the rainfall at an ungauged point is to take a weighted average of the observed values. Similarly, a weighted average of the observed rainfalls at a number of points would provide an estimate of the average rainfall over the area. In order to calculate the accuracies of such weighted averages it is necessary to know the covariances between the amounts of rain recorded at different points: in general these will be different for different pairs of points. Direct estimates of these covariances are not available for all pairs of points since this would require measurements of rainfall at every point in the area. It is therefore to be hoped that the covariance structure has certain simplifying features which might be used to obtain those covariances not directly estimable. For instance it is reasonable to assume that the correlation (and covariance) function should be a continuous (smooth) function over space. A further possibility might be that the correlation is a function only of the relative positions of the two points (e.g. the distance between them) and not of their absolute spatial coordinates, i.e. the process is spatially stationary.

A first step in investigating the covariance structure of rainfall is to calculate the covariances between measurements at gauged points. As described in the previous section the data sets considered consisted of 672 stations in the area denoted as East England and 1153 in North England. From these it would be possible to estimate $\frac{1}{2}$ x 672 x 673 = 226,128 and $\frac{1}{2}$ x 1153 x 1154 = 665,281 covariances between different pairs of stations in each area - clearly unmanageable numbers. To overcome this it was decided to estimate the covariance between a particular "central" station and all other stations in the region: this was done for several such central stations within each region to provide a reasonably representative subset with which to investigate the correlation structure.

Given a set of measurements (X_t, Y_t) $(t=1,\ldots N)$ of rainfall over the same interval (t) of time at two stations, the usual estimates of the variances, covariances and correlation are

$$S_X^2 = \frac{1}{N-1} \sum_{t=1}^{N} (X_t - \bar{X})^2 \quad , \quad S_Y^2 = \frac{1}{N-1} \sum_{t=1}^{N} (Y_t - \bar{Y})^2$$

$$S_{XY} = \frac{1}{N-1} \sum_{t=1}^{N} (X_t - \bar{X})(Y_t - \bar{Y}) \quad , \quad r = S_{XY}/(S_X S_Y) \qquad \ldots. \qquad (4.1)$$

where $\bar{X} = \frac{1}{N} \sum_{t=1}^{N} X_t$, $\bar{Y} = \frac{1}{N} \sum_{t=1}^{N} Y_t$ are estimates of the means at the two stations. Here it is assumed that the measurements at different time-points are statistically independent: under such circumstances the above formulae provide unbiased estimates of the variances and covariances of the measurements no matter what their distribution is.

If a particular statistical distribution for the data can be assumed then other, better, estimates can be formed. For instance plotting the data suggests that the square roots of the monthly total rainfalls have an approximate joint normal distribution. The relations for the higher moments of normal distributions then suggest, as estimates of the means, variances and covariances of the monthly totals,

$$\hat{\mu}_X = \hat{\mu}_1^2 + \hat{\sigma}_1^2 \quad , \quad \hat{\mu}_Y = \hat{\mu}_2^2 + \hat{\sigma}_2^2 \quad ,$$

$$\hat{\sigma}_X^2 = 2\hat{\sigma}_1^2(\hat{\sigma}_1^2 + 2\hat{\mu}_1^2), \qquad \hat{\sigma}_Y^2 = 2\hat{\sigma}_2^2(\hat{\sigma}_2^2 + 2\hat{\mu}_2^2), \qquad \ldots. \qquad (4.2)$$

$$\hat{\sigma}_{XY} = 2\hat{\sigma}_{12}(\hat{\sigma}_{12} + 2\hat{\mu}_1\hat{\mu}_2),$$

where $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, $\hat{\sigma}_{12}$ are estimates of the means, variances and covariances of the square roots of the monthly totals. A trial run using such estimates suggested that there was indeed slightly less scatter among the estimates of correlations since the effect of the large values from the upper tail of the distribution is reduced. However estimates found via expressions (4.2) require considerably more computation than the simple estimates in equations (4.1) and therefore were not used in the rest of the study. Moreover these estimates depend on a distributional assumption which has not been fully checked.

The central stations were chosen from amongst those that had a full set of data for the fourteen years 1961–1974. However, for the majority of stations in the data set there were periods of missing data. In these cases the correlations were calculated using the estimators given in equation (4.1) with (X_t, Y_t) covering only the pairs of values for which both X_t and Y_t were recorded. This is not the only possible procedure but it has the advantage of being simple: a disadvantage is that it can result in inconsistent values of the correlation coefficient.

An alternative would be to use a maximum likelihood method which utilises all the recorded rainfall data. This would require specific distributional assumptions and the full version of this procedure has the following unfortunate feature (unless special assumptions are made): if observations cover N time periods at P stations, even if there is only one missing value (P–1 stations with N observations and 1 station with N–1), then the estimate of the covariance between a station having missing data and any other station eventually becomes more variable as the number of stations, P, increases (Morrison, 1971; Moran, 1974). An attempt was made to use a maximum likelihood procedure with a limit to the number of stations used in calculating any one correlation; this was not pursued as the method appeared to need a great deal more investigation than time allowed. In using either the simple approach or the maximum likelihood method, it is assumed that whether or not an item of data is missing is not related to the rainfall magnitude actually occurring: if experience showed that this were not the case, e.g. if missing values tended to be days of high rainfall, then this fact could in principle be incorporated in a maximum likelihood method.

The estimates used here are based on very few statistical assumptions; if further assumptions can be made then much better estimates of the distributional parameters can be obtained. The validity of such assumptions merits careful prior investigation, as they may be at variance with the basic physical properties of the process. The following subsections report the results of applying the estimates described above to various classifications of data.

The circle about each central station was divided into eight octants and these were numbered consecutively as in Figure 4.1; this diagram also shows the symbols used for plotting the corresponding correlations in the diagrams that follow.

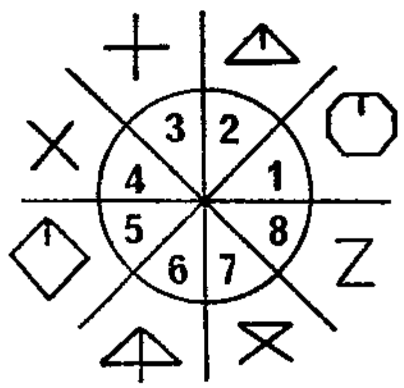


Figure 4.1: Symbols used for plotting different orientations from central stations.

4.3.2 Correlations between annual totals

For the data set consisting of total annual rainfalls the estimators given in equation (4.1) were calculated using only concurrent pairs of observations. Figures 4.2 - 4.5 shows plots of correlation against distance from 2 central stations in each of the two areas. The correlations are plotted with different symbols depending on the orientation of the line from the central station to the other station as described above. Only the estimates of correlation based on 10 or more pairs of observations are plotted. The curves drawn through the points are explained later.

4.3.3 Correlations between monthly totals

The treatment of the data consisting of monthly totals is now described. It might be of interest to assume a different mean, variance and covariance structure for the rainfall for each different calendar month of the year and on this assumption estimates for the means and covariances for each month can be calculated using equations (4.1). However these estimates would be based on only 14 values from the years 1961-74 and would therefore have a large sampling variation. A compromise procedure is to treat the mean rainfalls for different months as seasonal and to assume constant variances and covariances throughout the year. Estimates of the variances and covariances are then obtained by averaging the estimates for each month:

$$S_X^2(\mathrm{AVE}) = \frac{1}{12}\{S_X^2(\mathrm{JAN}) + S_X^2(\mathrm{FEB}) + \dots \},$$

$$S_{XY}(\mathrm{AVE}) = \frac{1}{12}\{S_{XY}(\mathrm{JAN}) + S_{XY}(\mathrm{FEB}) + \dots \}, \qquad \dots\dots \quad (4.3)$$

and

$$r(\mathrm{AVE}) = S_{XY}(\mathrm{AVE})/\{S_X(\mathrm{AVE})S_Y(\mathrm{AVE})\}$$

is the corresponding estimate of the correlation. If the variances and covariances are not constant as assumed these estimates still provide reasonably representative values for the year as a whole. A technical point is that, whereas the rainfall totals in adjacent months may be statistically dependent, the estimators given in equation (4.1) are applicable to the data for separate months since, being twelve months apart, they may reasonably be assumed independent. There has not been time to investigate seasonality more fully. A difference in spatial correlation structure over the year would result from differing predominant rainfall regimes.

The estimates of correlation for the monthly data obtained in the above fashion are plotted in Figures 4.6 - 4.9. Due to the large number of points which would otherwise appear only the correlations corresponding to stations in alternate octants (Figure 4.1) are shown. The lines fitted through the sample correlations are explained in Section 4.4.

In Figures 4.10 - 4.11 the corresponding correlation plots are given for the data for stations in East Anglia for the 16 years 1875-1890. Comparison with Figures 4.6 - 4.7 shows that there is no marked difference between the correlation structures existing in 1875-1890 and more recent times.

4.3.4 Correlations between daily totals

As indicated in Section 4.2, four different classifications of daily data were analysed and the graphs illustrating the spatial correlation in these sets of data are presented in Figures 4.12 - 4.27. It may be noted that the correlation between the rainfall at points decays faster with distance for days classified as having an average rainfall of over 5 mm than for days having average rainfall of over 2 mm, and faster still for days having more than 10 mm average rainfall. This is presumably because high rainfall is produced by relatively localised events.

The fifth set of data consisted of days on which an average rainfall of over 5 mm occurred together with the previous two days data, no matter what rain fell on those days. This data set (which extended over 8 years) was used to estimate the correlations between the rainfalls in the area on different days. Figure 4.28 shows the correlation between the rainfall at a central station on a 'rain-day' and the rainfalls at other stations on the same day. In Figure 4.29 the correlation between the rainfall at a central station on a 'rain-day' and the rainfalls on the previous day at other stations is shown: the corresponding plot for a two day time difference is shown in Figure 4.30. The correlations between the rainfall on the day before the 'rain-day' at the central station and the rainfalls at other stations on the previous day are shown in Figure 4.31. The appearance of Figure 4.31b in particular warrants some remark: there are a number of large positive values among the sample correlations at moderately large distances. It is felt that these may be caused by the discrete component at zero of the distribution of rainfalls which is not entirely removed by the threshold condition. The purpose of the analysis of this set of data was to determine whether the rainfall recorded on previous days could be used to improve the accuracy of interpolation procedures. The results indicate that the extent of such improvement is likely to be small although this conclusion may simply reflect the transient nature of moderately heavy rainfall.

4.4 Fitting of correlation functions

4.4.1 Choice of a correlation function

The purpose of fitting a function to the estimated correlations between rainfalls at different stations is to give values not only for the correlations between a central station and ungauged points but also between pairs of ungauged points: to be able to do this some assumptions are needed about the form of the correlation function $\rho(\underset{\sim}{x};\underset{\sim}{y}) = \rho(x_1,x_2;y_1,y_2)$ which gives the correlation between the rainfalls at points with coordinates $\underset{\sim}{x} = (x_1, x_2)$ and $\underset{\sim}{y} = (y_1,y_2)$.

The plots of the correlations against distance from the central stations suggest that a possible function that would represent the decay of correlation at a distance d in a fixed direction is

$$\rho(d) = a + (1 - a) \exp(- bd) \qquad \ldots. \qquad (4.4)$$

where a and b are constant for each direction. The plots also suggest that the rate of decay parameter might be different for different directions: accordingly the following direction-dependent function is suggested. For two points with coordinates $x = (x_1,x_2)$, $y = (y_1,y_2)$, and u_1 and u_2 given by

$$u_1 = x_1 - y_1, \; u_2 = x_2 - y_2$$

the correlation function is given by

$$\rho(\underset{\sim}{x};\ \underset{\sim}{y}) = \rho(u_1,u_2) = a + (1-a)\exp\left[-b\{(u_1+c_1u_2)^2 + c_2{u^2}_2\}^{\frac{1}{2}}\right] \quad \ldots \quad (4.5)$$

where a, b, c_1, c_2 are constants to be determined. This family of functions represents surfaces which decay exponentially at different rates in different directions but to the same level. Lines of equal correlation will appear as concentric elipses. Restrictions on the constants are

$$a \leqslant 1 \quad ; \quad b \geqslant 0 \quad ; \quad c_2 \geqslant 0.$$

It may be noted that the function (4.5) has the property of stationarity under translation, i.e. the correlation between two points depends on their relative positions and not on their absolute coordinates: this allows the correlation between ungauged points to be estimated.

4.4.2 Parameter estimation

A particular member of the family of functions (4.5) may be selected to represent the spatial correlation decay from a central station in the following way. Fisher's z-transformation (Kendall & Stuart, 1969, p.390),

$$z(r) = \tfrac{1}{2}\log_e\left(\frac{1+r}{1-r}\right),$$

of the sample correlation r has the property that, if r is based on a sample of n independent pairs of Normally distributed values, z(r) is approximately Normally distributed with mean $\frac{1}{2}\log_e(1+\rho)/(1-\rho))$ (ρ bring the true correlation) and variance $(n-3)^{-1}$ if ρ is small and n moderately large. While these assumptions do not hold in this case it is still reasonable to fit the parameters a,b,c_1,c_2 of the correlation function by minimising

$$g(a,b,c_1,c_2) = \sum_{i=1}^{P} (n_i-k)\ (z_i-f_i)^2 \quad \ldots\ldots \quad (4.6)$$

where

$$z_i = \tfrac{1}{2}\log_e\left(\frac{1+r_i}{1-r_i}\right) \qquad (i = 1, \ldots P)$$

$$f_i = \tfrac{1}{2}\log_e\left(\frac{1+\rho_i}{1-\rho_i}\right) \qquad (i = 1, \ldots P)$$

$$\rho_i = \rho(x_{1i}-x_1^*, x_{2i}-x_2^*) \qquad (i = 1, \ldots P)$$

and where r_i is the sample correlation between the i'th station at coordinates (x_{1i},x_{2i}) and the central station at coordinates (x_1^*, x_2^*) based on n_i pairs of values, P being the number of stations considered. The number k was taken as 3 for the daily and yearly totals and as 36 for the monthly totals. The objective function to be minimised is not obtained rigorously from statistical assumptions: equation (4.6) can be regarded simply as one way of giving higher weight to those correlations based on more observed values. The usual statistics derived from such a fitting operation, such as standard errors and correlations of estimates, have no meaning in this case since, for example, the r_i (i=1, ... P) are dependent amongst themselves. It is clearly possible to extend the above procedure to use data from several central stations but this has not been done; rather the correlation function is fitted separately for each central station to give an indication of whether the same correlation function applies to each area.

The problem of minimising the function (4.6) is that of minimising a weighted sum of squares : many mathematical subroutine libraries contain procedures for handling such minimisations. It is therefore fairly simple to fit correlation functions by the above method. The functions fitted at the central stations are shown in Figures 4.2 - 4.27 together with the sample correlations: here a cross-section of the function is drawn for each of the octants for which correlations are plotted, the cross-section being along the line through the middle of the octant concerned.

Consideration of Figures 4.2 - 4.27 suggests that, while the family of functions used provides a good fit at some of the stations, there is considerable room for improvement at others. Firstly, for the North of England data set there is a marked effect associated with the Pennine Range - in Figures 4.8a and 4.8b the correlations of the central station with stations in the 4th and 5th octants show a rapid drop and then a rise with distance. Other central stations, on both sides of the Pennines, also show a similar drop - this drop usually being just on the opposite side of the range from the central stations. It is really only reasonable to consider the function in equation (4.5) in a topographically homogeneous region. The second deficiency of the fitted functions is at short distances. For example, Figure 4.7 suggests that a better fit might be obtained by a function not necessarily approaching the value 1 at zero distances. Such effects could be due to either or both of two causes:

(i) meteorological phenomena affecting only very small areas (compared to the distances between raingauges),

or (ii) inaccuracies in the measurements taken at the raingauges.

However, for other central stations in the same region (Figure 4.6) the correlation plots do indicate a correlation approaching 1 at zero distances - clearly this could be because some raingauges are more subject to measurement errors than others. In any case there is a lack of stations very close together to distinguish between a rapid drop and a discontinuity at the origin of the correlation function. In judging visually the fit of the function to the correlations it should be remembered that the estimates of correlations at different places are not independent and thus the points plotted corresponding to neighbouring raingauges will tend to lie on the same side of the true correlation.

In the Section 4.5 some similar work by the Meteorological Office is described. They have fitted circular correlation functions to a large number of small regions contained within a larger area: the resulting fitted functions can be used to investigate the spatial homogeneity of the correlation structure. The results presented in Section 7.2.3 indicate that there are considerable differences locally in the rates of decay of correlation over the area of Northern England investigated.

It is felt that the correlation functions fitted here provide a reasonable description of the decay of correlation with distance bearing in mind that topographical features also have a large effect on the correlations. In Section 5 the functions fitted are used to show how measures of the accuracy of a raingauge network can be obtained when the correlations between all points in an area were assumed known. In the next section a short discussion is given of the plots of correlations in Figures 4.2-4.27.

4.4.3 Discussion

The main emphasis of Sections 4 and 5 of this report is to show how the accuracy of interpolated estimates of rainfall can be computed, and the plots of sample correlations and the functions fitted to them are given simply to illustrate the application of the methods suggested. Nonetheless, it is possible to make some comparisons among the plots of correlations given in Figures 4.2 - 4.27.

It has already been noted that, for the different categories of daily data, the correlation functions are such that the correlations of the "over 10 mm rainfall" data decay fastest with distance and those for arbitrary days decay slowest. Here "rate of decay" may be taken as equivalent to the distance to an 0.9 correlation level on the fitted function. The correlations for the monthly data decay more slowly than the correlations between rainfall data for arbitrary days but the plots suggest that correlations for the yearly totals decay at the same rate, or slightly faster, than those for monthly totals.

The spatial correlation functions fitted are such that correlations may decay at different rates in different directions from the central point but must decay to the same level. These asymptotic values can best be interpreted as giving the correlations for moderately large distances (distances of the same order as those in the plots): it is not to be expected that correlation would remain constant with distance indefinitely. The asymptotic correlation levels for the different categories of data are roughly at the same general level of about 0.5 - 0.7. For the data for days of moderate and heavy rainfall the asymptotic correlation levels drop to nearly zero. It could be argued that having different rates of decay with different directions about a central station has the physical interpretation of there being a prevailing wind direction over the area: gauges within the path of the same storm components would receive roughly the same rainfall.

Compared with the differences between the correlation functions fitted to stations within the same region there is little difference between the fitted correlations for the two areas "Eastern" and "Northern" England. The difference between the function for the same region could be due either to sampling errors - it has not been possible to test this statistically - or to a real spatial variation in the correlation properties of rainfall.

4.5 Localised Analyses

The Meteorological Office (M.O.) has also undertaken an analysis of the correlation structure of rainfall. For this work they have considered a number of relatively small areas, each containing 30 or so raingauges, and have estimated the correlations between all pairs of stations within each area.

When determining the interpolation accuracy available from a particular network (see Section 5) the correlation between rainfalls at different points needs to be known only for distances up to about 20 - 25 kms in practice. Therefore the M.O. investigation proceeded as follows, using daily rainfall data for the period 1969 to 1974 extracted from the Meteorological Office Rainfall archive. The period 1969-1974 was chosen because previous data are less reliable and because the final quality control of the 1975 and 1976 data was not complete at the time of the exercise. Firstly data for all stations situated within 25 km of a given point were extracted and any gauges for which more than 365 days of data were missing were removed from the analysis. Only "wet-days" were chosen, a wet-day being one on which the mean rainfall over the 25 km radius area exceeded 2.0 mm. To eliminate serial correlation problems two days were skipped after each accepted wet-day before continuing the chronological search. The correlations between the recorded rainfalls at every pair of stations were estimated and these sample correlation coefficients were plotted against intergauge distance. An example of such a plot is shown in Figure 4.32. This and many other similar diagrams suggested that, for such short distances, a simple straight line could be used to represent the correlation decay with distance. The straight line shown in Figure 4.32 was obtained using the following procedure. The Fisher z-transformations of each of the sample correlations were first calculated and these transformed values were averaged over discrete distance bands of width 3 kms. The averaged values were transformed back to correlation coefficients using the inverse transformation and a straight line was then fitted by least squares to the resulting points, taking the averaged correlation at the mean distance of each band.

Clearly in adopting the above procedure it has been assumed that the correlation function is spatially homogeneous and isotropic within the area under consideration. The method allows for the existence of measurement errors by not forcing the fitted correlation function to pass through 1 at zero distance. Let $\rho(0)$ be the intercept of the fitted function on the y-axis, then an estimate of the average (measurement) noise index over the area is given by

$$\bar{\lambda}^2 = (1 - \hat{\rho}(0))/\hat{\rho}(0) \qquad \ldots \qquad (4.7)$$

where the noise index is the ratio of the variance of measurement errors to the variance of the true rainfall. (See also Sections 1.4.3 and 5.2.2). The estimate of the correlation function of the underlying rainfall is then

$$\hat{\rho}^*(d) = (1 + \bar{\lambda}^2)\hat{\rho}(d) = \frac{\hat{\rho}(d)}{\hat{\rho}(0)} \quad (d > 0) \qquad \ldots \qquad (4.8)$$

where $\hat{\rho}(d)$ is the estimate of the observed correlation function.

4.6 Eigenvector analysis

4.6.1 The method

Treatments of principal component or eigenvector analysis can be found in a number of standard texts, such as that of Kendall and Stuart (1968, p285). A spatial correlation analysis, which does not assume that noise indices at different stations are the same, can be carried out using this technique. Here a brief summary of the method is given.

The observed rainfall data at P gauges over N time periods is used to calculate a P x P matrix of correlations of observations at the gauges: from this a set of eigenvalues $E_1, \ldots E_p$ ($E_1 \geq E_2 \geq \ldots \geq E_p \geq 0$) and the corresponding eigenvectors, $\underset{\sim}{v}_1, \underset{\sim}{v}_2, \ldots, \underset{\sim}{v}_p$, are computed using a standard computer programme. The first and second principal components are $\underset{\sim}{v}_1$ and $\underset{\sim}{v}_2$ respectively, and so forth. Here

$$v_j = (v_{j1}, v_{j2}, \ldots v_{jp})^T. \qquad \ldots \qquad (4.9)$$

The amount of the total original variance represented by any one eigenvector is proportional to its associated eigenvalue and

$$E_1 + E_2 + \dots E_p = P. \qquad \dots\dots \qquad (4.10)$$

The eigenvectors are the set of vectors which enable the largest part of the original variance of the data to be represented in the most economical manner: they have the properties

$$\underset{\sim}{v}_i^T \underset{\sim}{v}_i = \sum_{k=1}^{P} v_{ik}^2 = 1, \ \underset{\sim}{v}_i^T \underset{\sim}{v}_j = \sum_{k=1}^{P} v_{ik} v_{jk} = 0 \ (i \neq j). \qquad \dots\dots \qquad (4.11)$$

The original data matrix, $\underset{\sim}{X} = \{X_{ij}\}$, where X_{ij} is the observation at time i recorded on gauge j, has the representation

$$X_{ij} = \sum_{k=1}^{P} v_{ki} \, c_{kj} \quad (i = 1, \dots P; \ j = 1, \dots N) \qquad \dots\dots \qquad (4.12)$$

where

$$c_{ij} = \sum_{k=1}^{P} v_{ik} \, X_{kj} \quad (i = 1, \dots P; \ j = 1, \dots N). \qquad \dots\dots \qquad (4.13)$$

By comparison with equation (4.12), the reconstruction $X_{ij}^{(q)}$ of X_{ij} using the principal q eigenvectors is given by

$$X_{ij}^{(q)} = \sum_{k=1}^{q} v_{ki} \, c_{kj} \quad (i = 1, \dots P; \ j = 1, \dots N). \qquad \dots\dots \qquad (4.14)$$

In this way the rainfall values can be reconstituted to include increasing amounts of the original variance.

The total variance of the observed rainfall will be made up partly by the underlying rainfall patterns and partly by random noise, this being assumed spatially and temporally uncorrelated. Thus, if the eigenvectors representing the underlying rainfall can be identified (i.e. the principal components of variation), the remaining eigenvectors can be used to describe the noise. The supposed underlying rainfall can be generated using equation (4.14) and the associated correlation structure can be estimated in the usual manner.

A noise index for each individual gauge can be calculated from the variance of the differences between the original data and the reconstituted data. This would be particularly useful for identifying "singular" gauges, i.e. those that differ widely from the general rainfall pattern. The mean value, $\bar{\lambda}^2$, of the noise index over the area being analysed can be estimated by

$$\bar{\lambda}^2 \simeq (P - \sum_{k=1}^{q} E_k) / \sum_{k=1}^{q} E_k. \qquad \dots\dots \qquad (4.15)$$

4.6.2 An application

The eigenvector analysis described above has been applied to an area near Northampton using 18 stations, and daily data for 77 'wet' days. The 'wet' days were selected from the 1976 data set using only those days when the mean rainfall recorded by these gauges exceeded 2 mm. Figures 4.33 - 4.38 show the first 6 eigenvectors plotted at their associated gauge locations. The percentages of the total original variance represented by each of the first six eigenvectors were 74.3%, 10.8%, 4.6%, 3.1%, 1.4%, 1.2%.

To illustrate the procedure of filtering noise from the original data, the rainfall has been reconstituted using successively the first 3, 4, 5 and 6 eigenvectors. These reconstituted data have been used to produce a simple illustrative correlation pattern by correlating the data at each gauge with one near the centre of the area (gauge 159796 (Ravensthorpe)) and plotting the correlations on their respective gauge locations. The original correlation pattern is shown in Figure 4.39 and the correlation patterns obtained on reconstituting the data using the first 3, 4, 5 and 6 eigenvectors are shown in Figures 4.40 - 4.43. It can be seen that the correlation pattern of the reconstituted data approaches that of the original data as more eigenvectors are included. The individual noise indices for the stations are plotted in Figure 4.44; these were derived on the assumption that only the first six eigenvectors represented the underlying rainfall pattern.

The difficult and not yet completely resolved problem is to identify the significant eigenvectors. Methods used to date have been mainly subjective involving manual contouring of the plotted eigenvector terms. If the pattern of these terms is apparently coherent then the eigenvector is said to be significant. Another method suggested involves plotting the logarithm of the eigenvalue against the eigenvector number. The theory is that those eigenvectors representing noise will lie on a straight line and any associated with true rainfall will deviate significantly from this straight line. In practice this test has proved too insensitive for practical application.

The method of principal components is based essentially on the representation of the data by equation (4.12): whether there is a physical basis for such a representation needs clarification.

4.7 Summary

Using selected sets of annual, monthly and daily rainfall data in two areas in Eastern and Northern England, spatial correlation analyses were carried out. One approach involved estimating sample correlations, and fitting spatial correlation functions to these, for selected central stations in each area; in doing this, it was felt that there was not sufficient statistical evidence of measurement error or noise in the observed rainfall to warrant special recognition in the analysis. The second approach adopted by the Meteorological Office consisted of a similar correlation analysis based on correlations between all pairs of stations within small areas; this approach allows an assessment of stationarity over a large area and assumes that measurement error might exist.

The use of the fitted correlation functions in estimating the accuracy with which point interpolation and areal averaging may be carried out is illustrated in Section 5.

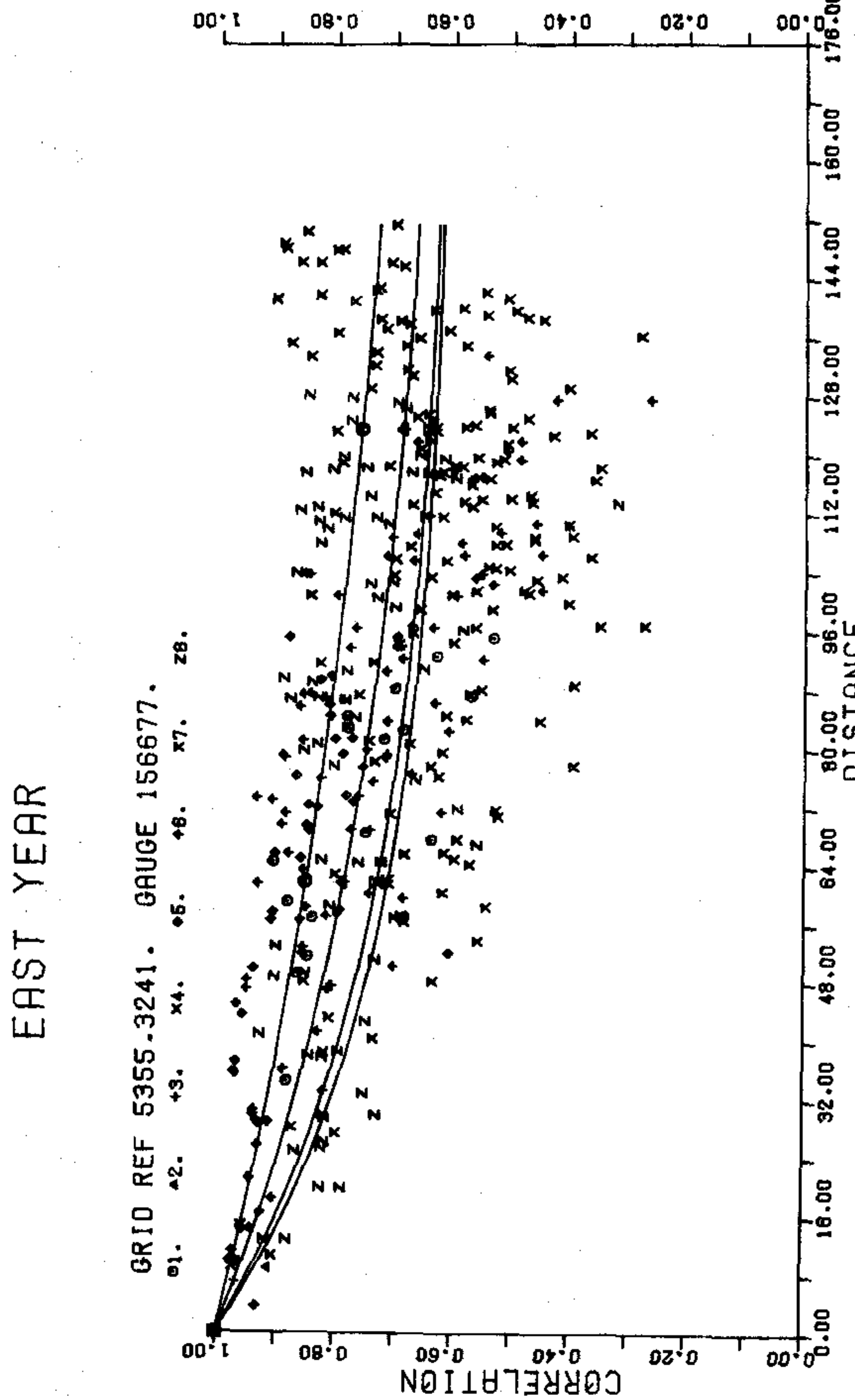


Figure 4.2 : Sample and fitted correlation functions.
Eastern England: yearly totals (1961-74).

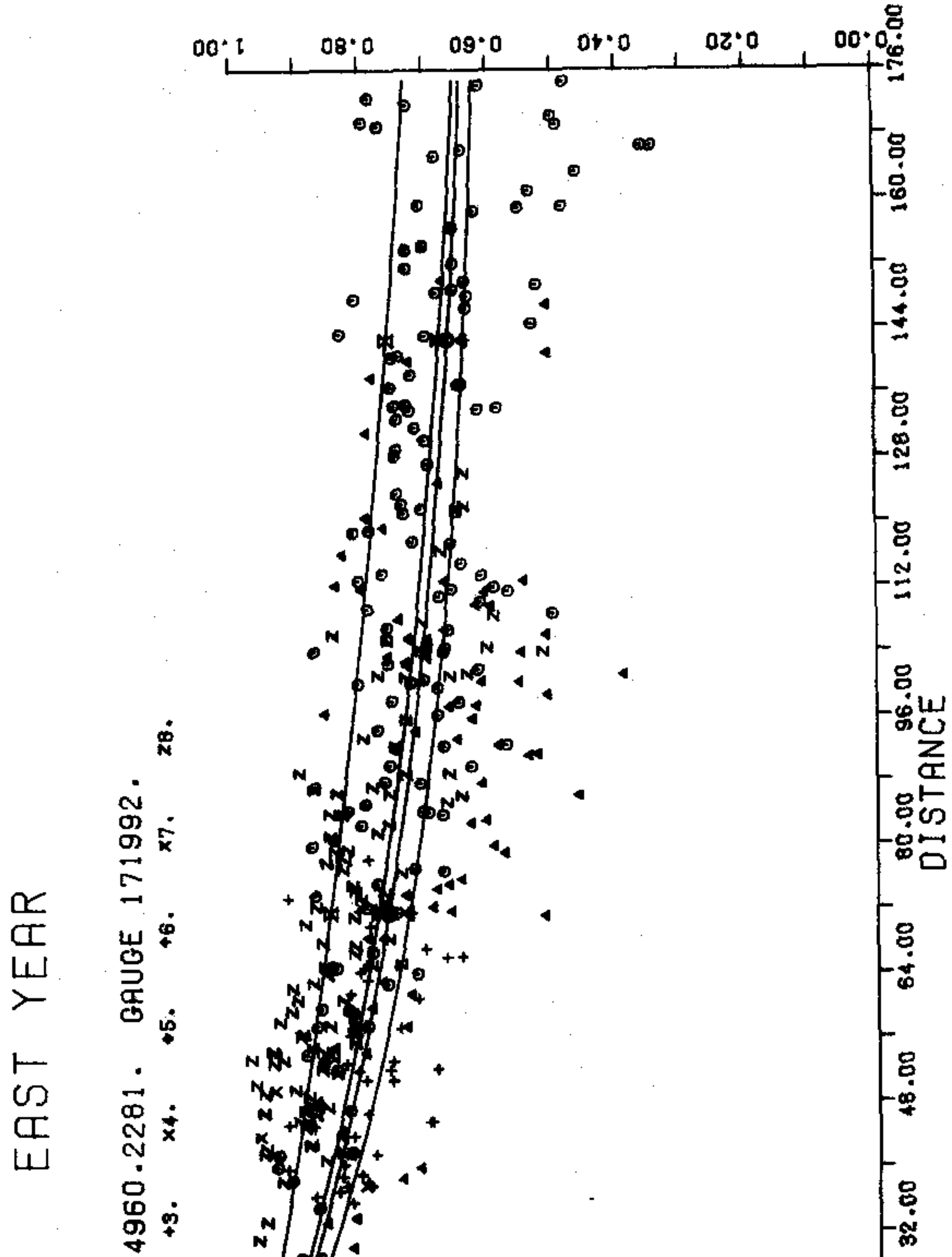


Figure 4.3 : Sample and fitted correlation functions.
Eastern England: yearly totals (1961-74).

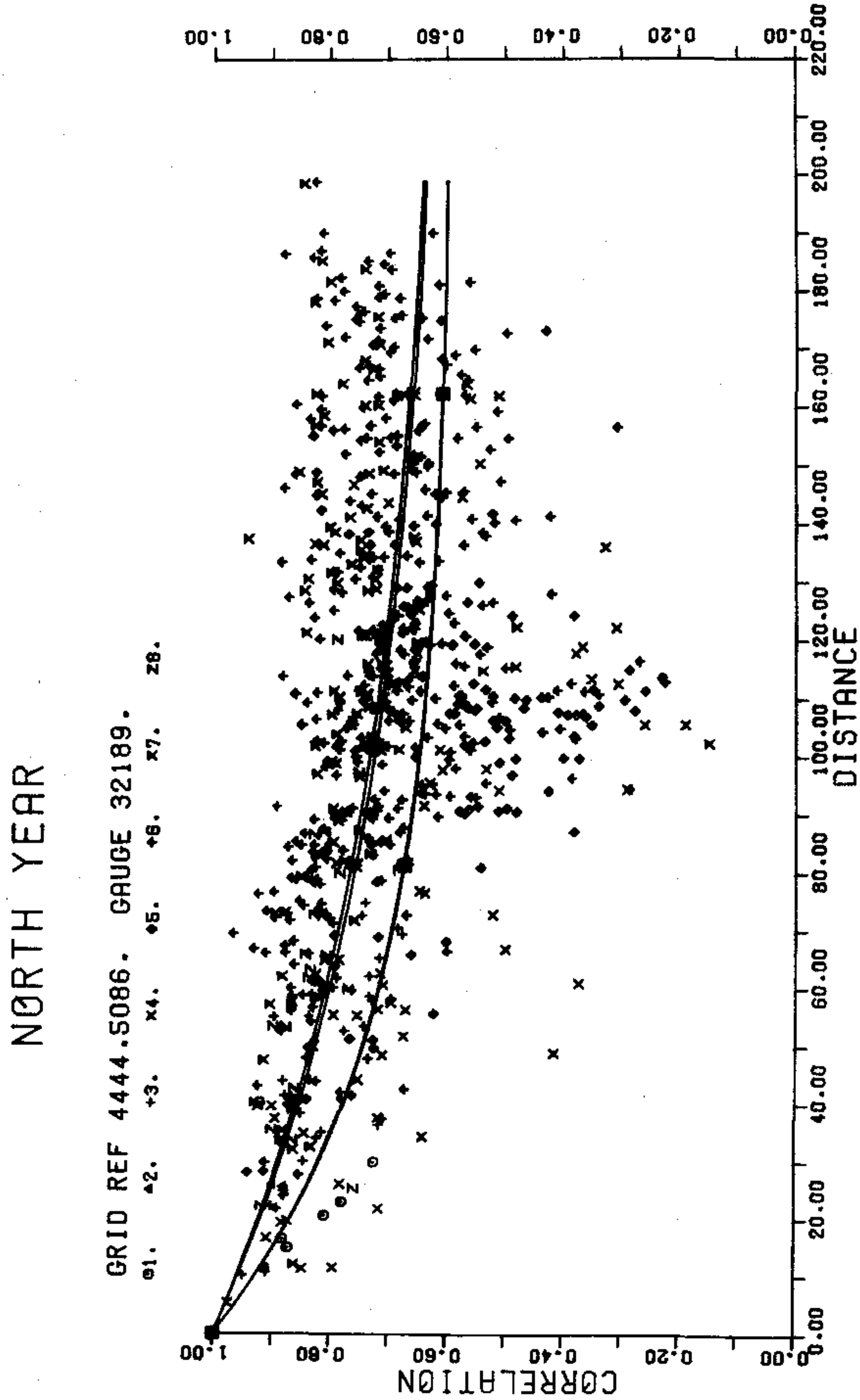


Figure 4.4 : Sample and fitted correlation functions.
Northern England: yearly totals (1961-74).

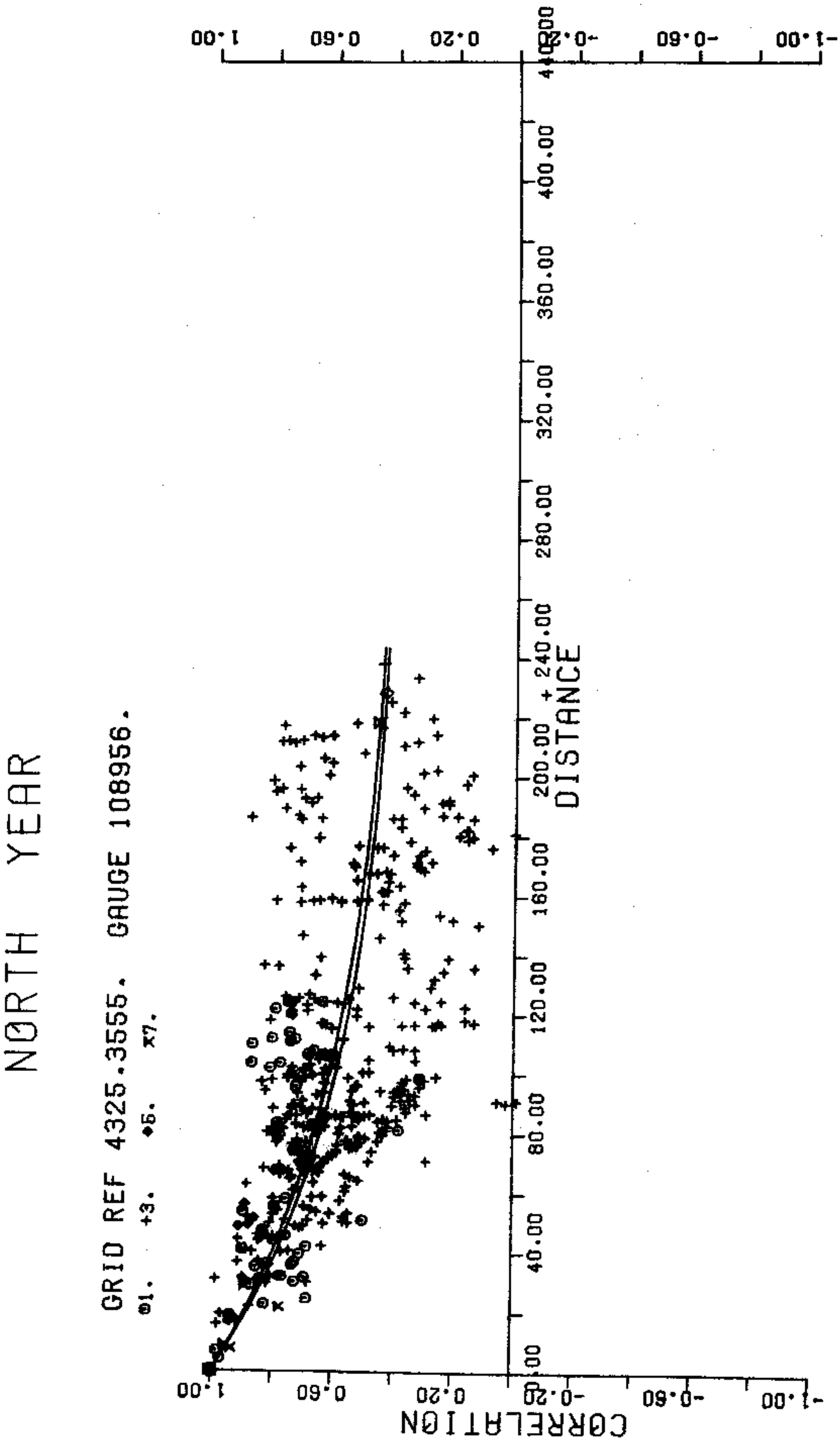


Figure 4.5a : Sample and fitted correlation functions.
Northern England: yearly totals (1961-74).

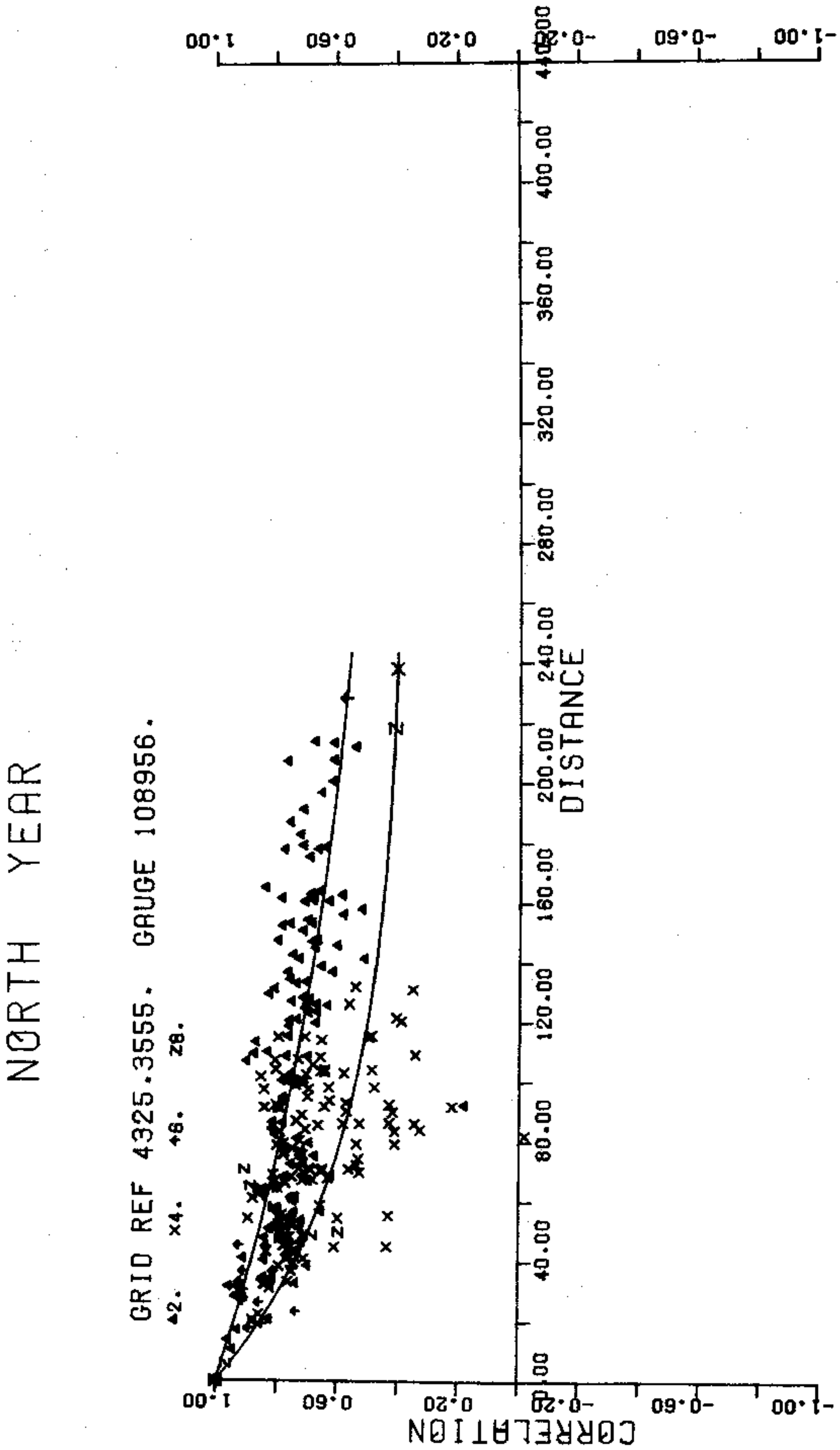


Figure 4.5b : Sample and fitted correlation functions.
Northern England: yearly totals (1961-74).

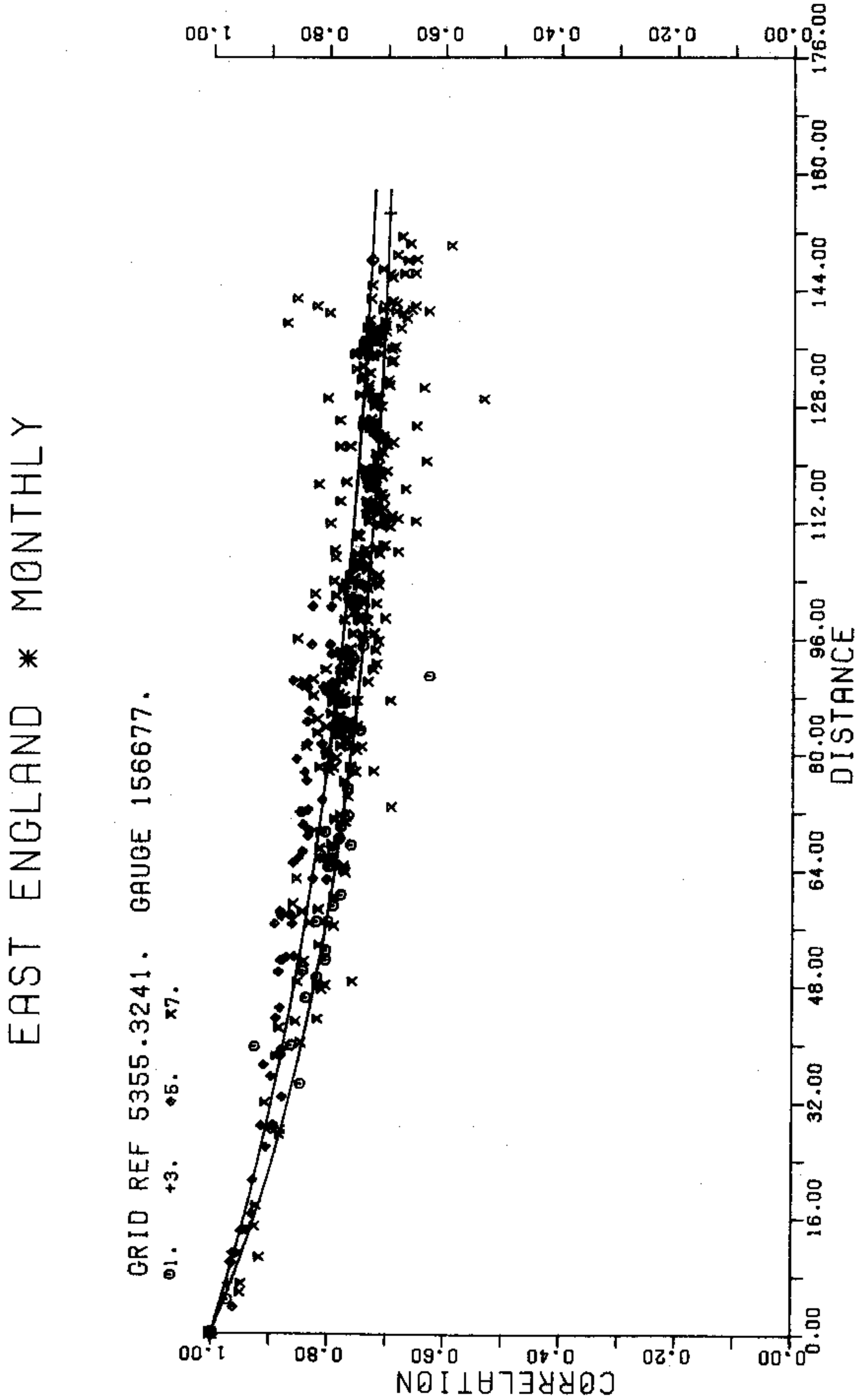


Figure 4.6a : Sample and fitted correlation functions.
Eastern England: monthly totals (1961-74).

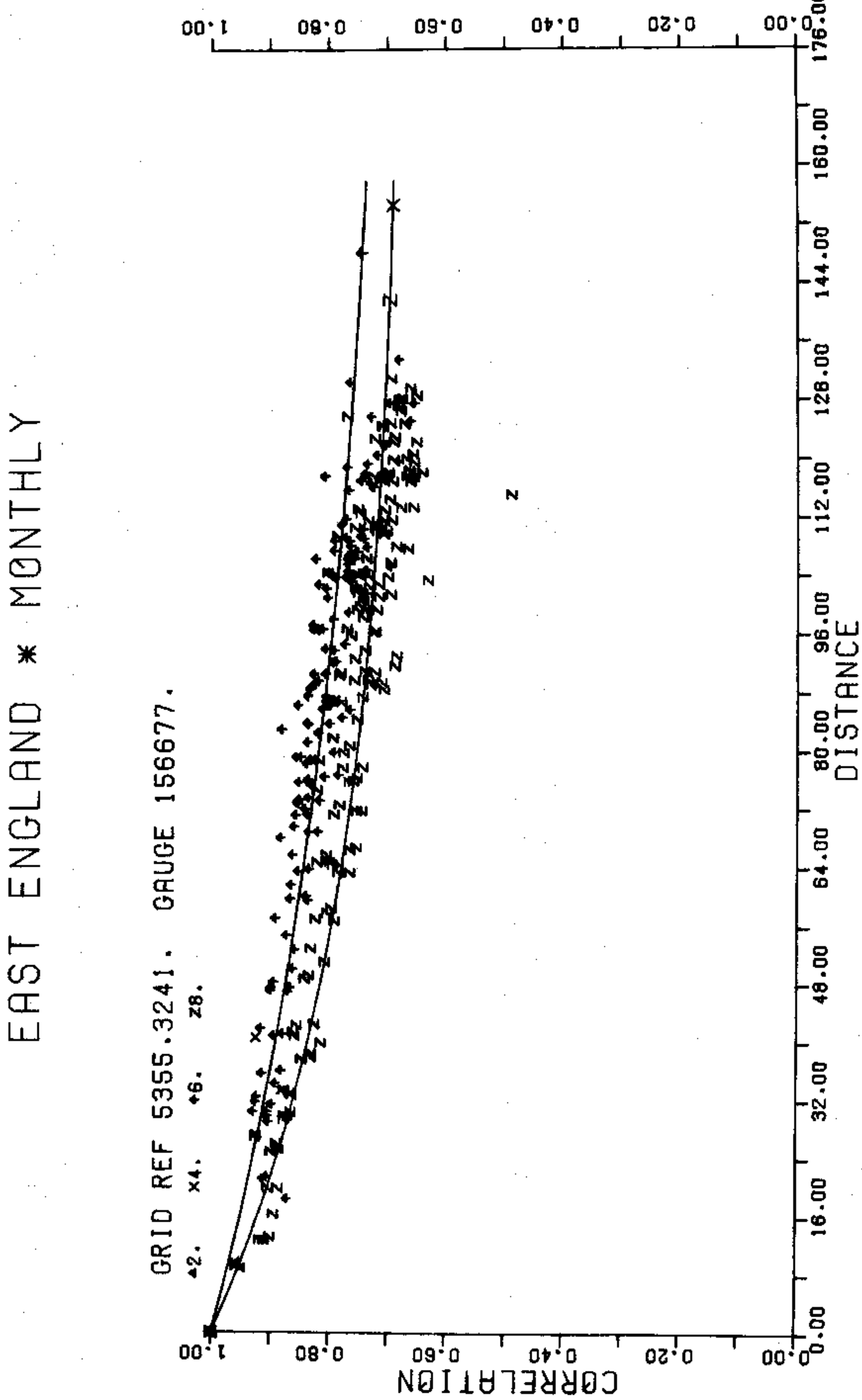


Figure 4.6b : Sample and fitted correlation functions.
Eastern England: monthly totals (1961-74).

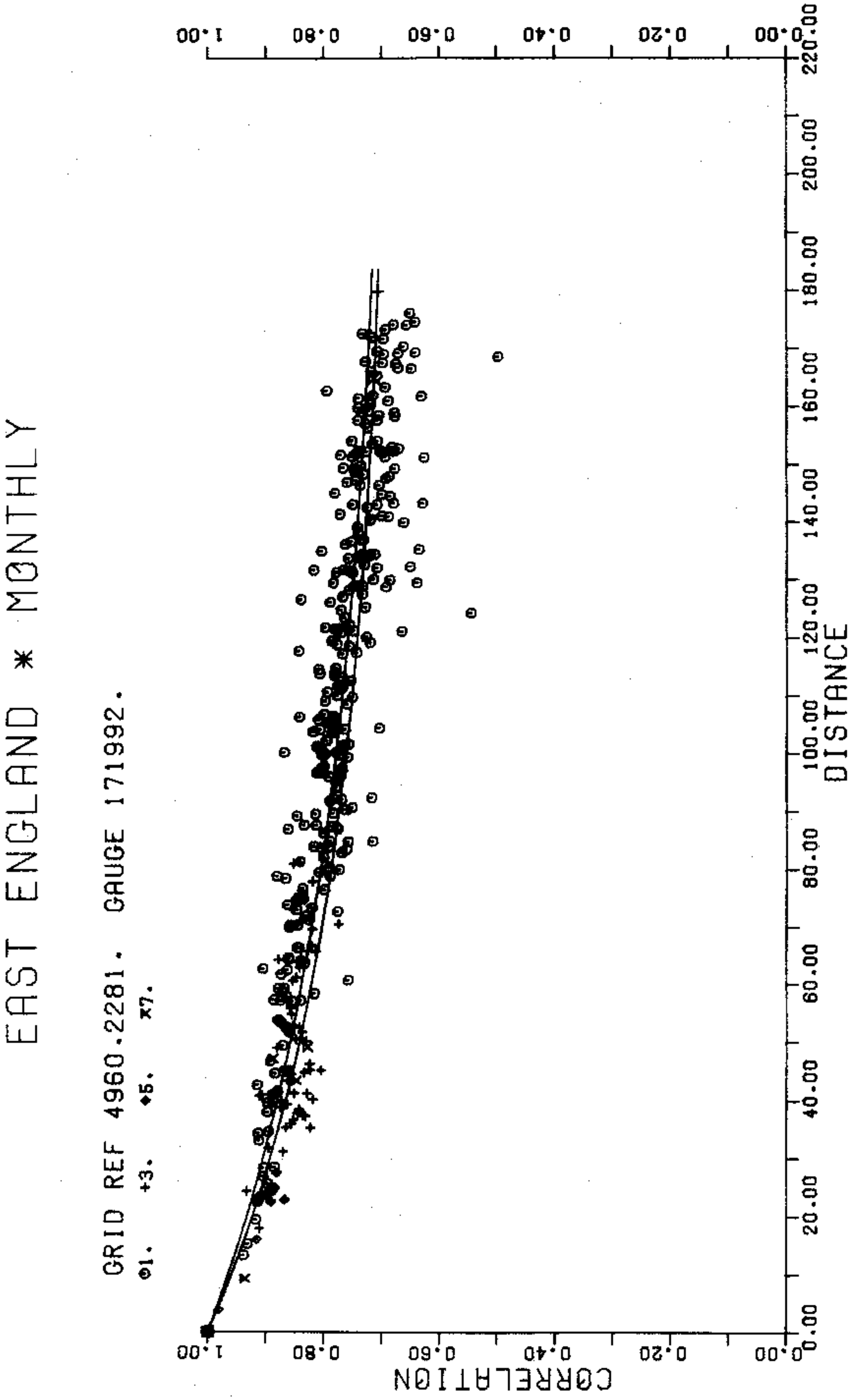


Figure 4.7a : Sample and fitted correlation functions.
Eastern England: monthly totals (1961-74).

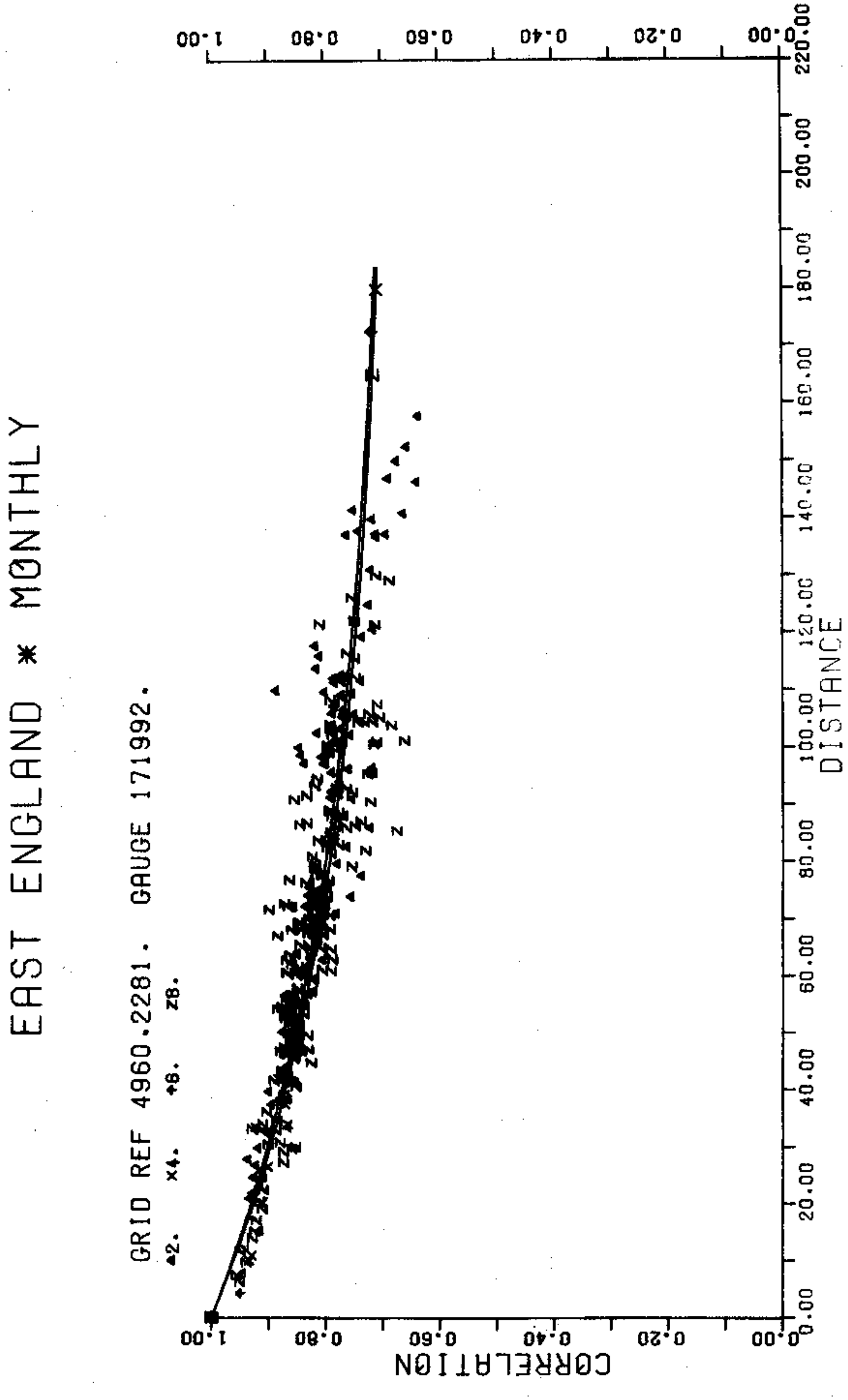


Figure 4.7b : Sample and fitted correlation functions.
Eastern England: monthly totals (1961-74).

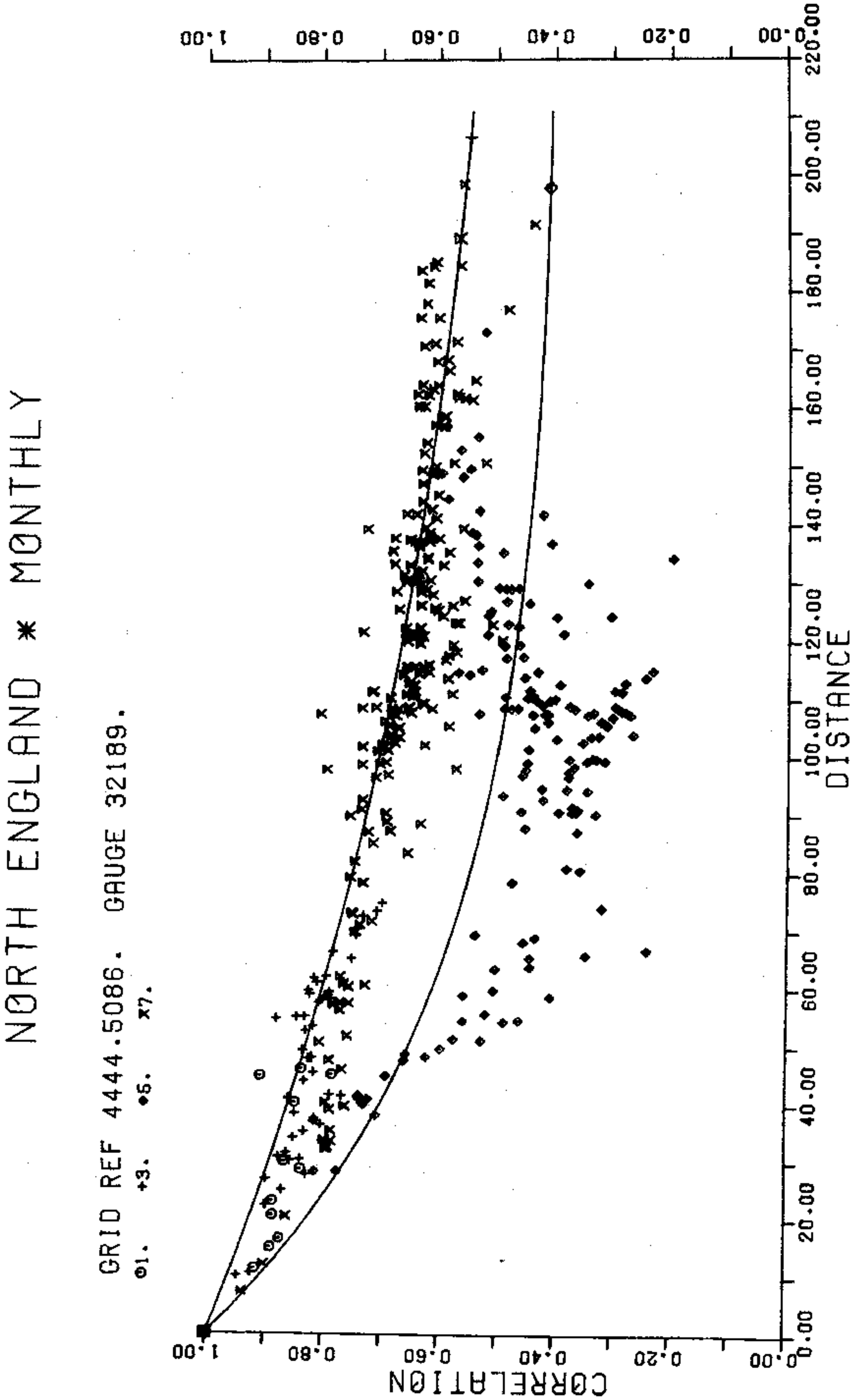


Figure 4.8a : Sample and fitted correlation functions .
Northern England: monthly totals (1961-74).

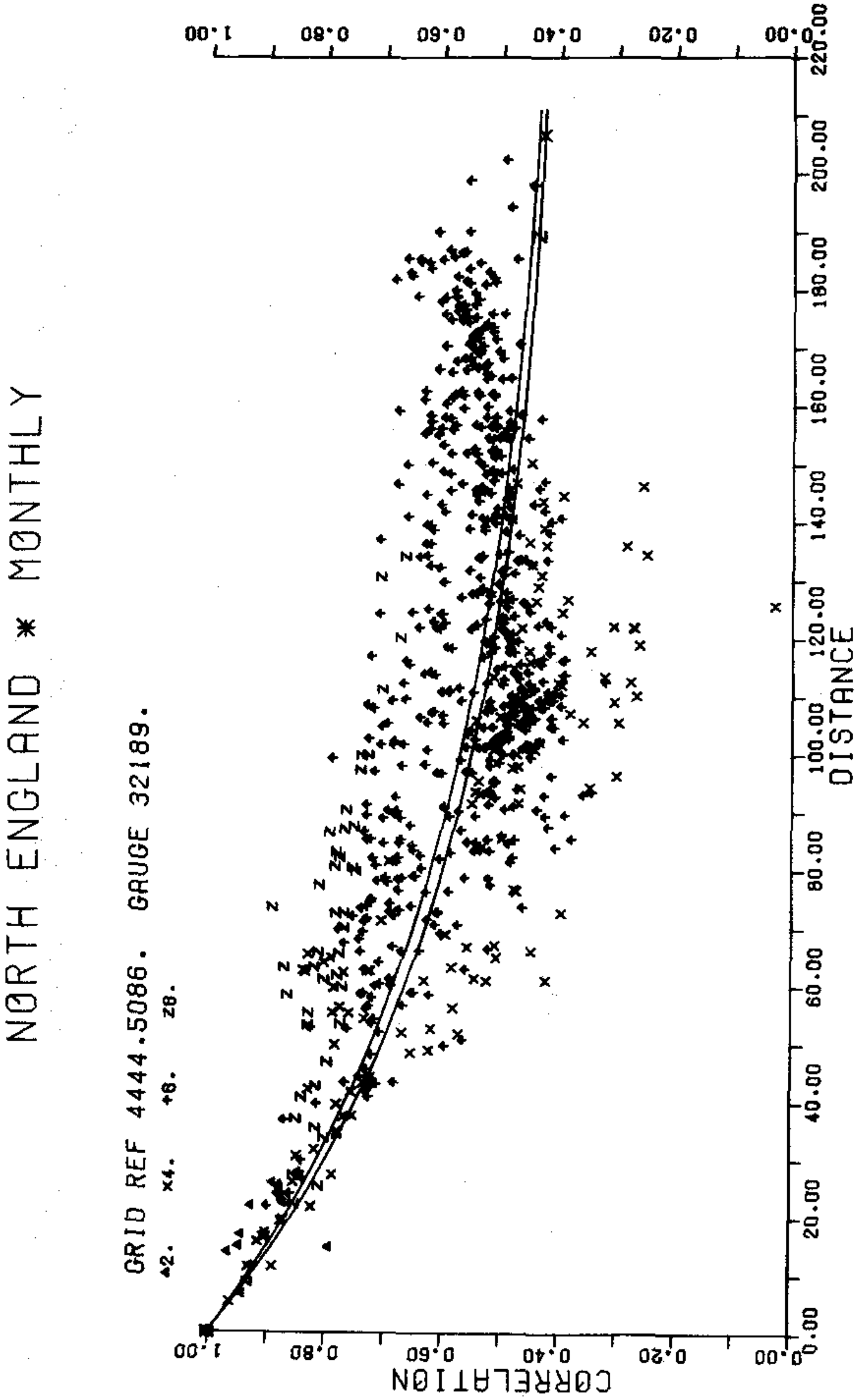


Figure 4.8b : Sample and fitted correlation functions.
Northern England: monthly totals (1961-74).

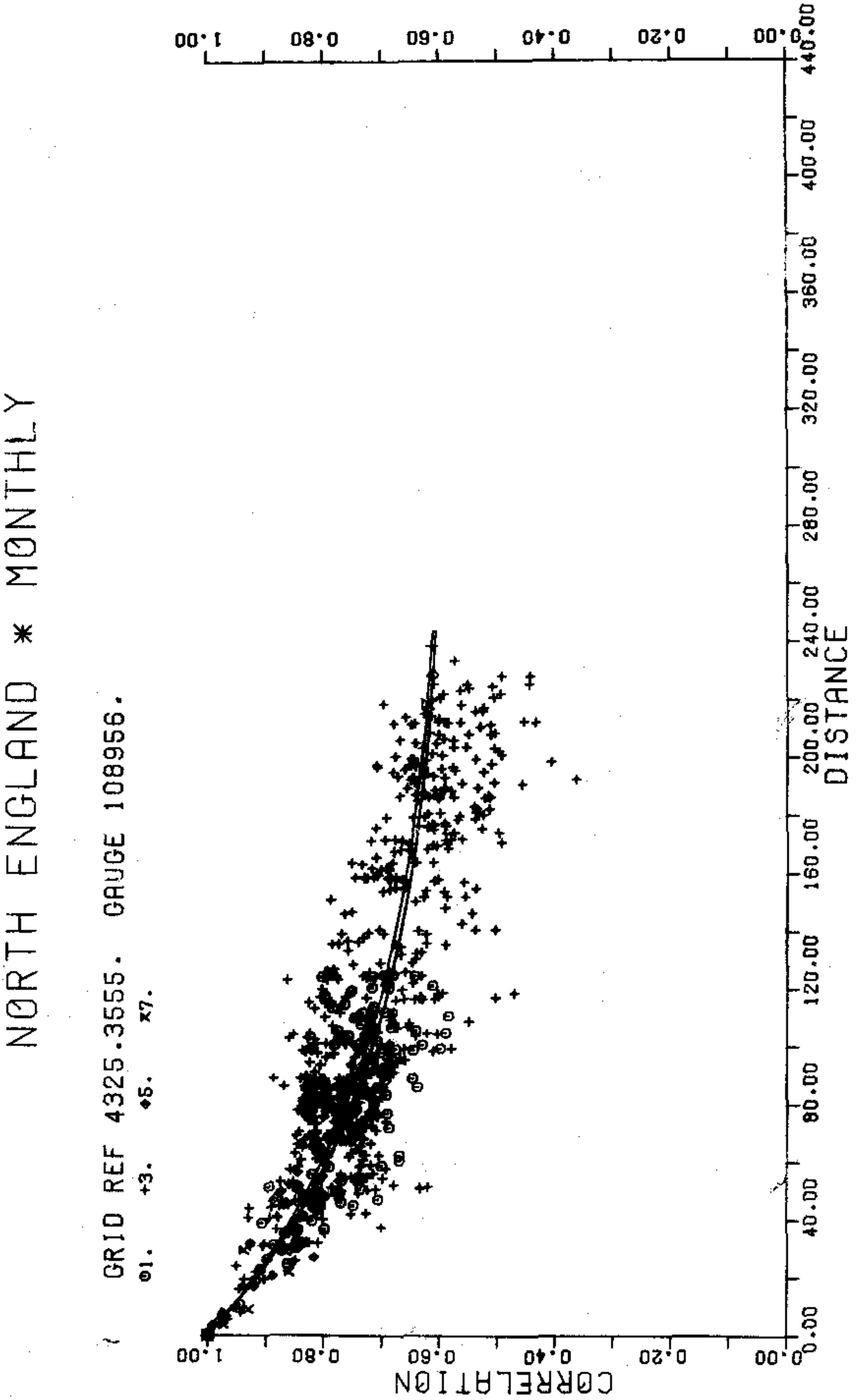


Figure 4.9a : Sample and fitted correlation functions.
Northern England : monthly totals (1961-74).

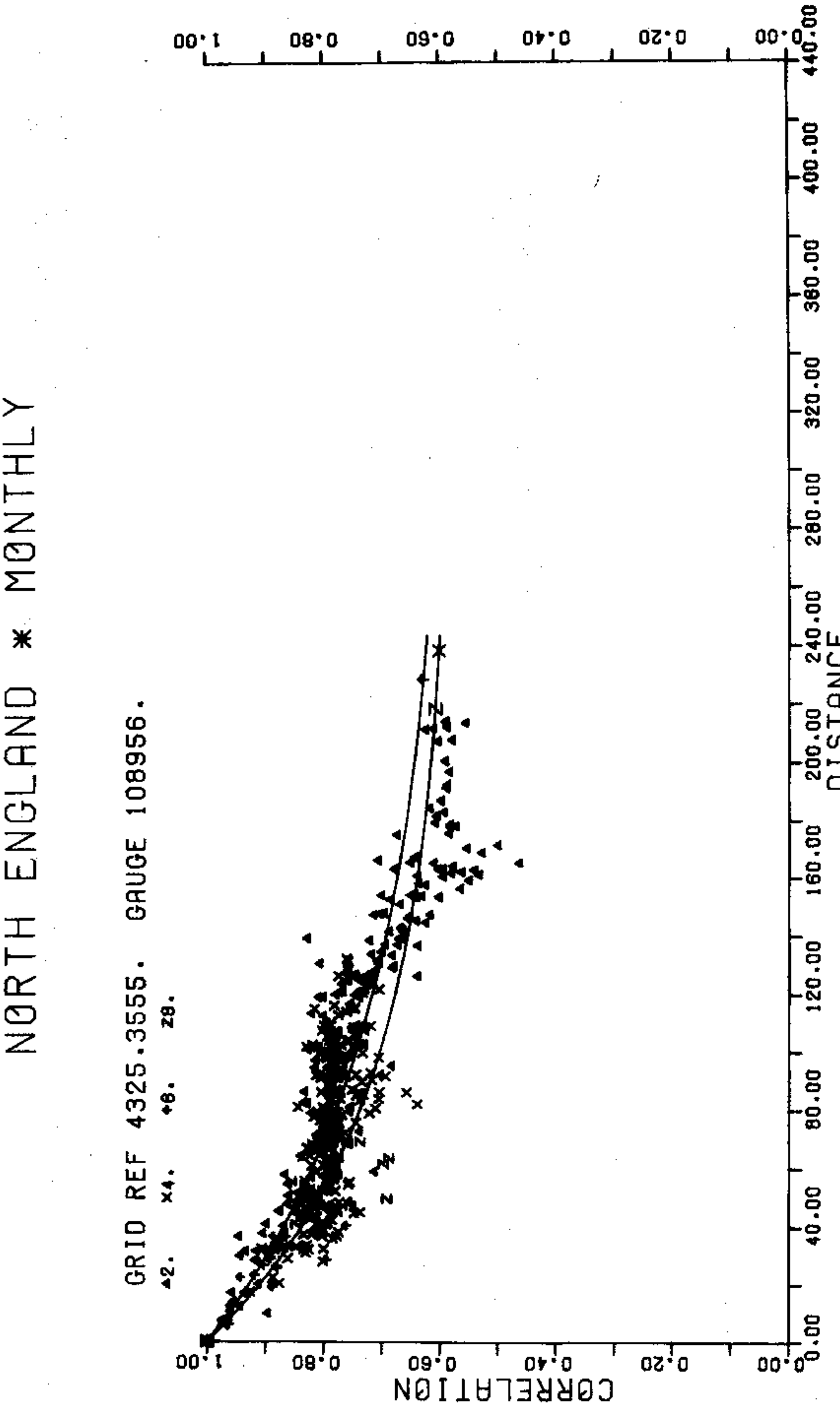


Figure 4.9b : Sample and fitted correlation functions.
Northern England: monthly totals (1961-74).

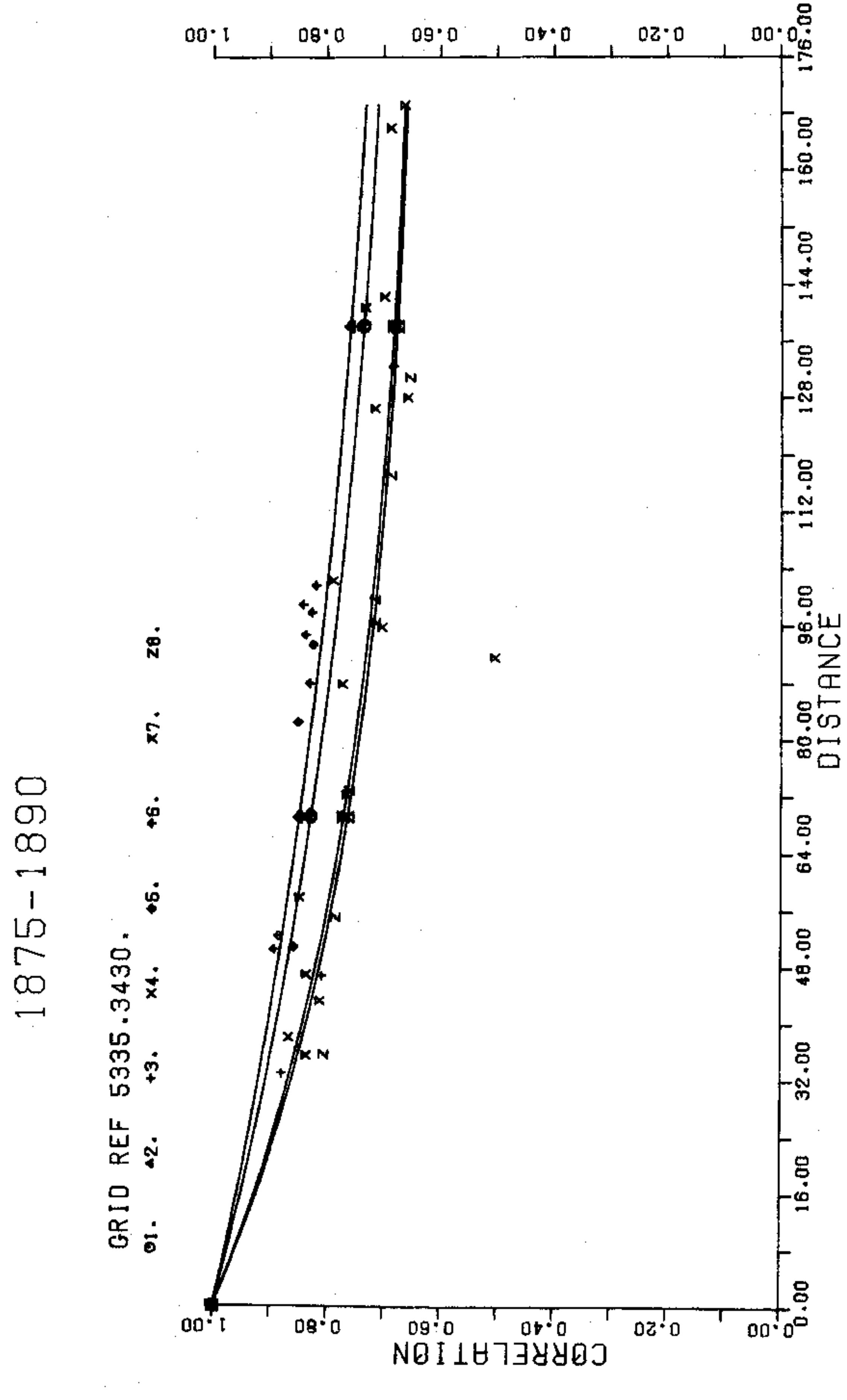


Figure 4.10 : Sample and fitted correlation functions.
Eastern England: monthly totals (1875-90).

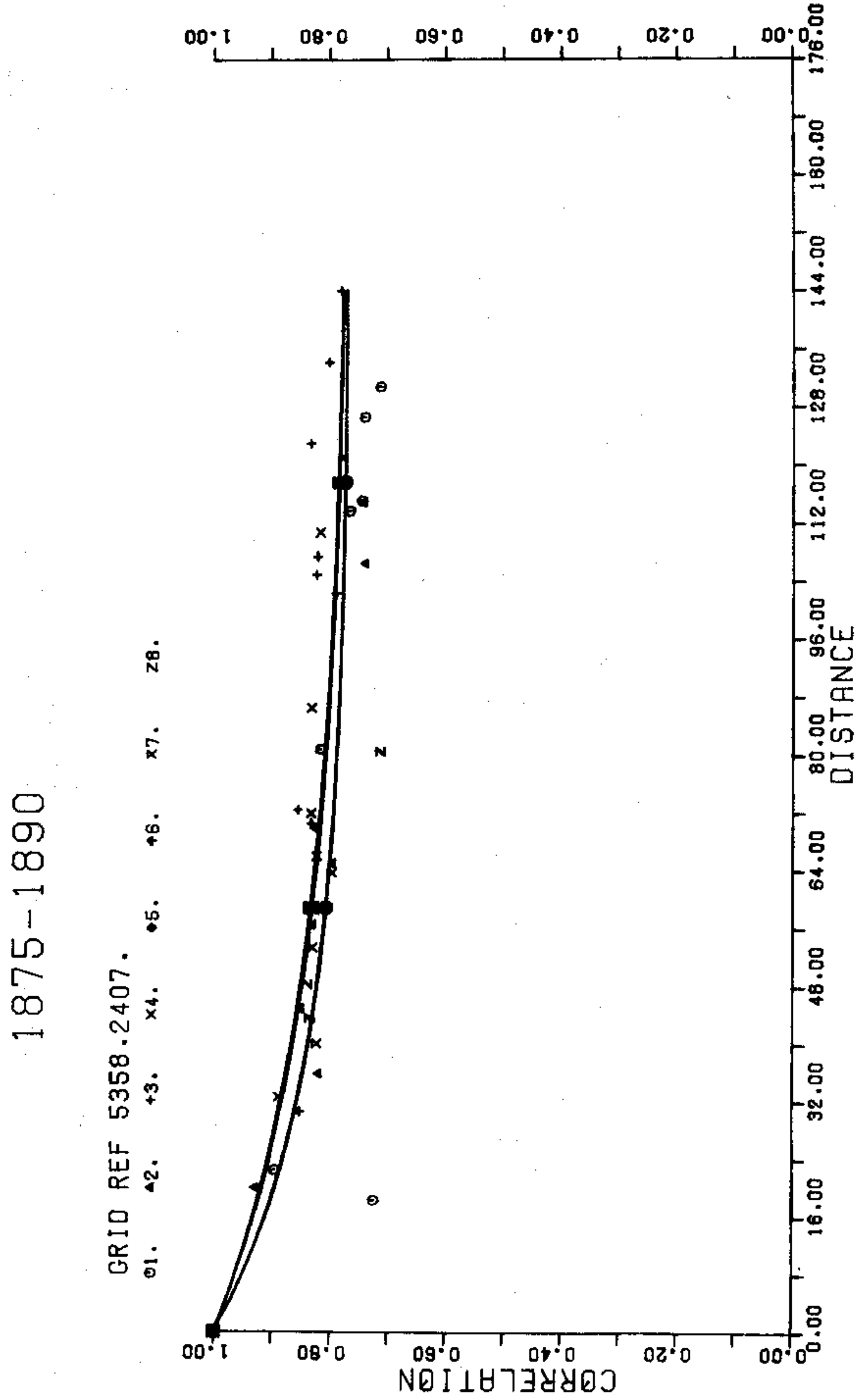


Figure 4.11 : Sample and fitted correlation functions.
Eastern England: monthly totals (1875-90).

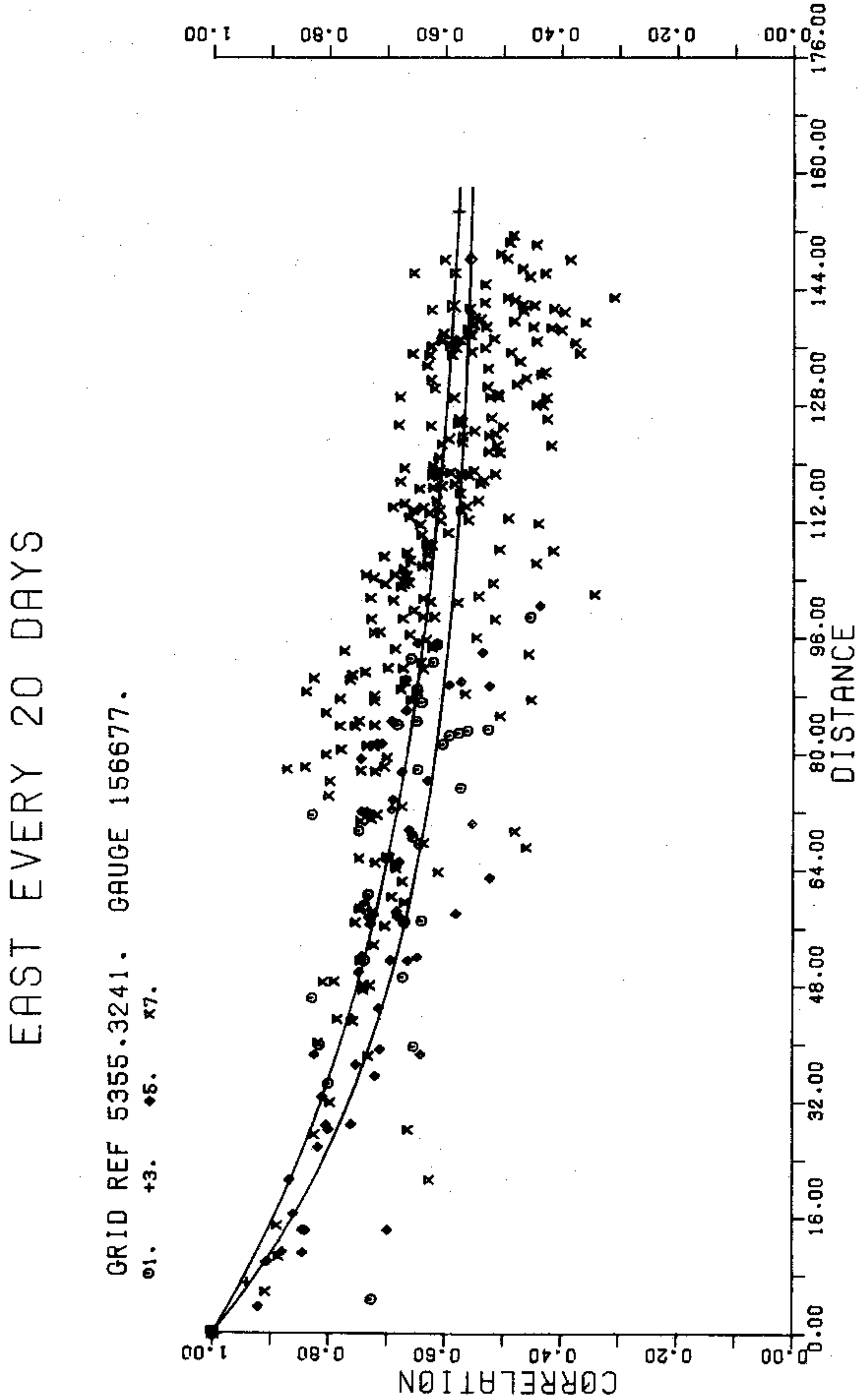


Figure 4.12a : Sample and fitted correlation functions.
Eastern England : daily rainfall totals.

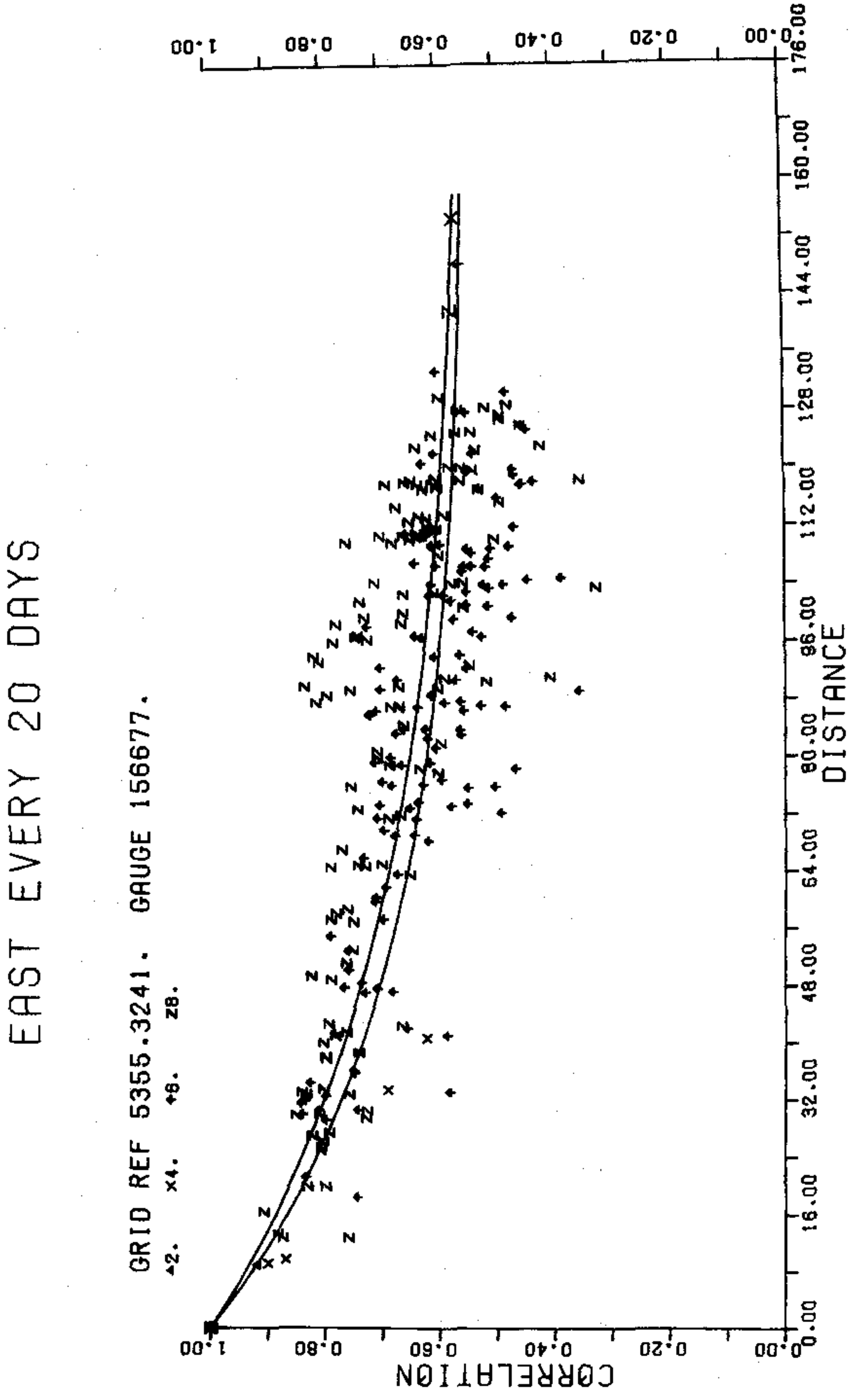


Figure 4.12b : Sample and fitted correlation functions.
Eastern England : daily rainfall totals.

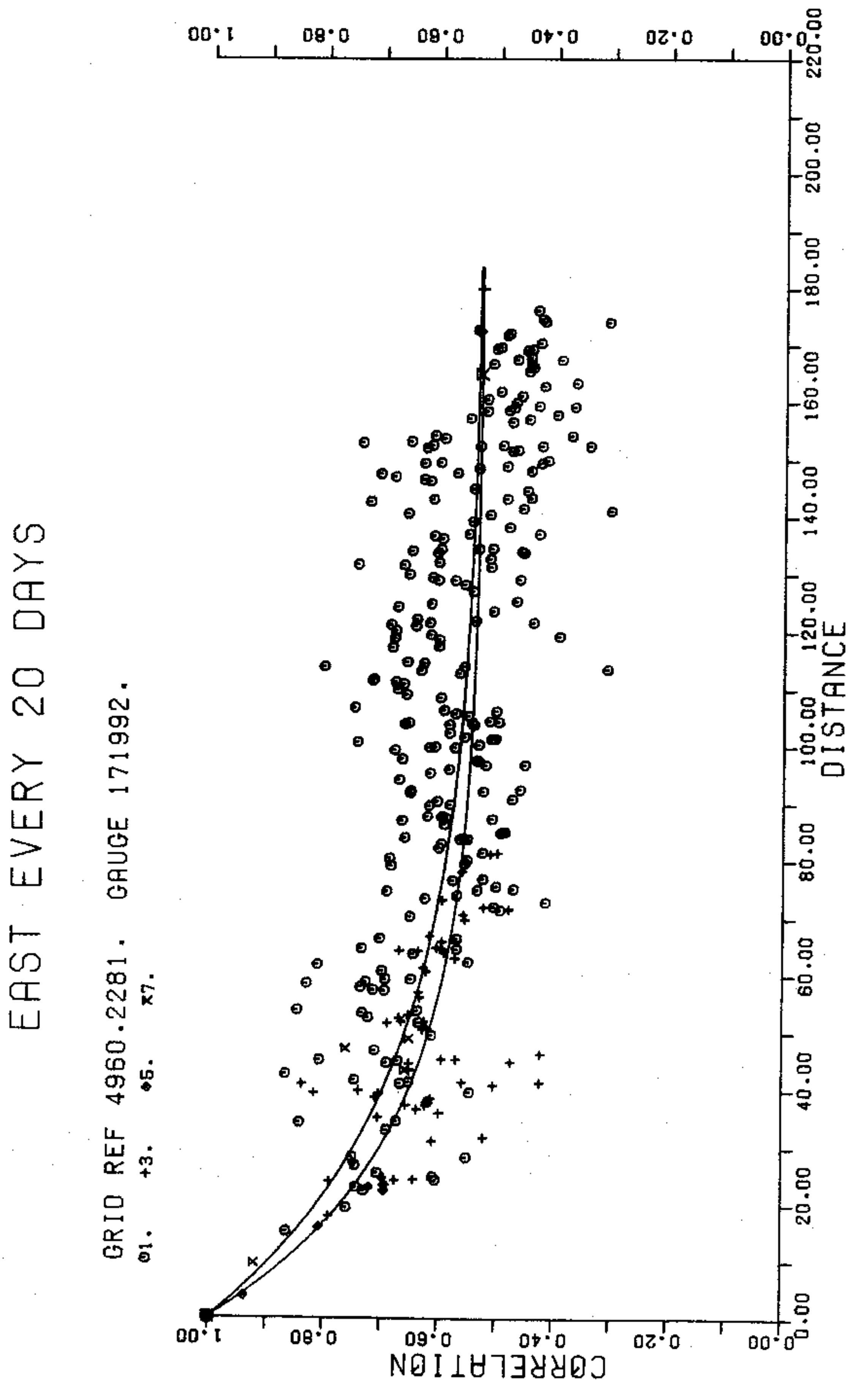


Figure 4.13a : Sample and fitted correlation functions.
Eastern England : daily rainfall totals.

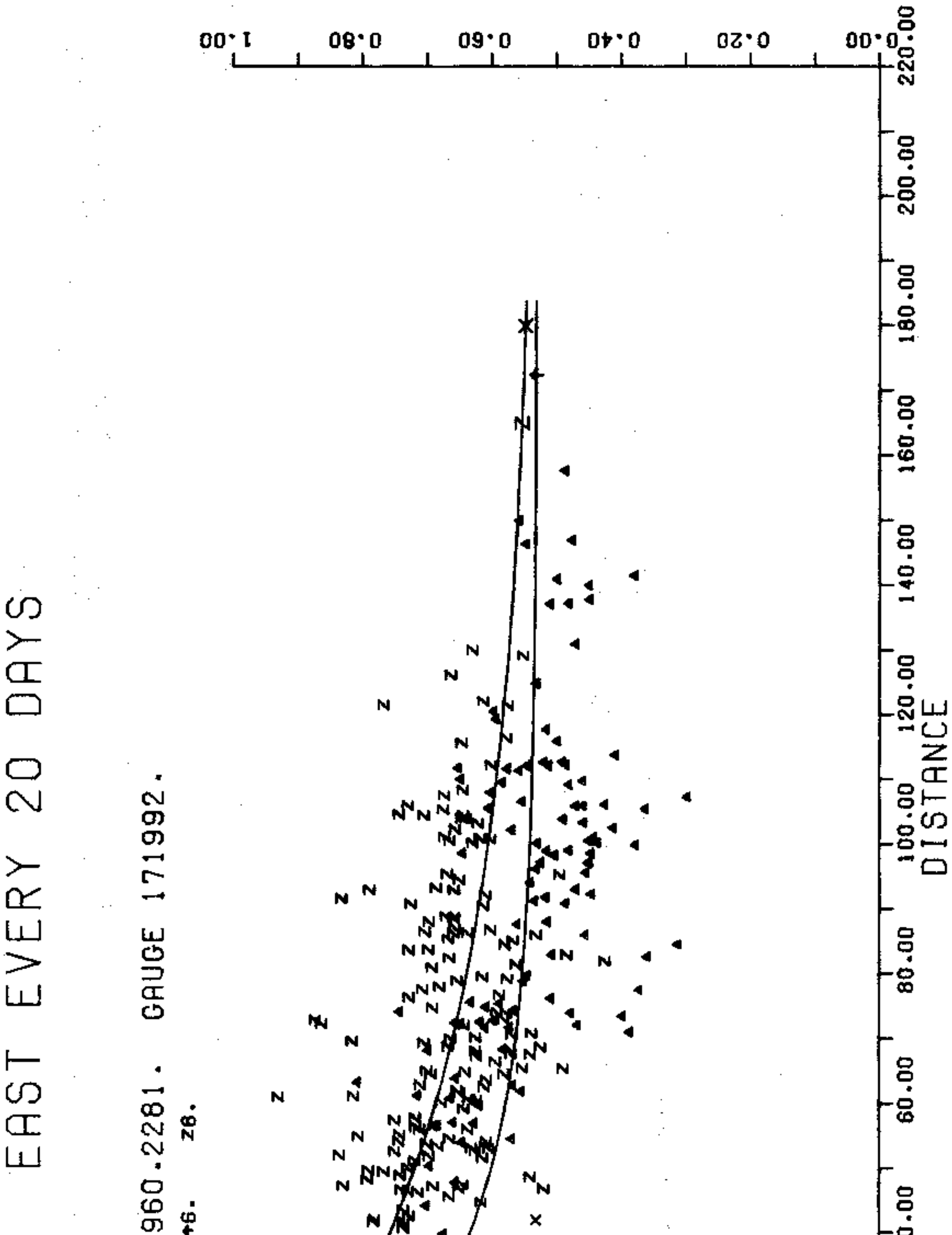


Figure 4.13b : Sample and fitted correlation functions.
Eastern England : daily rainfall totals.

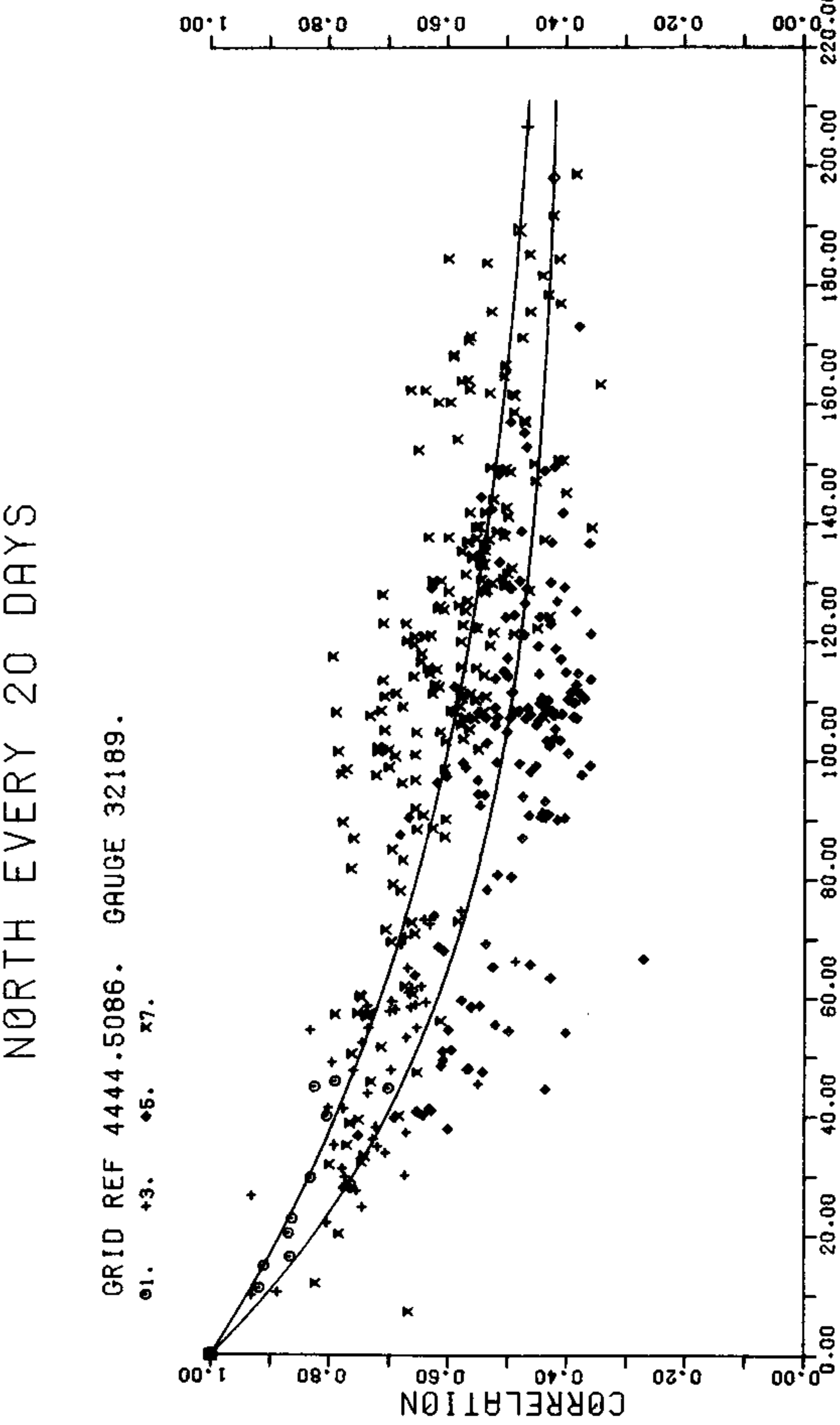


Figure 4.14a : Sample and fitted correlation functions.
Northern England : daily rainfall totals.

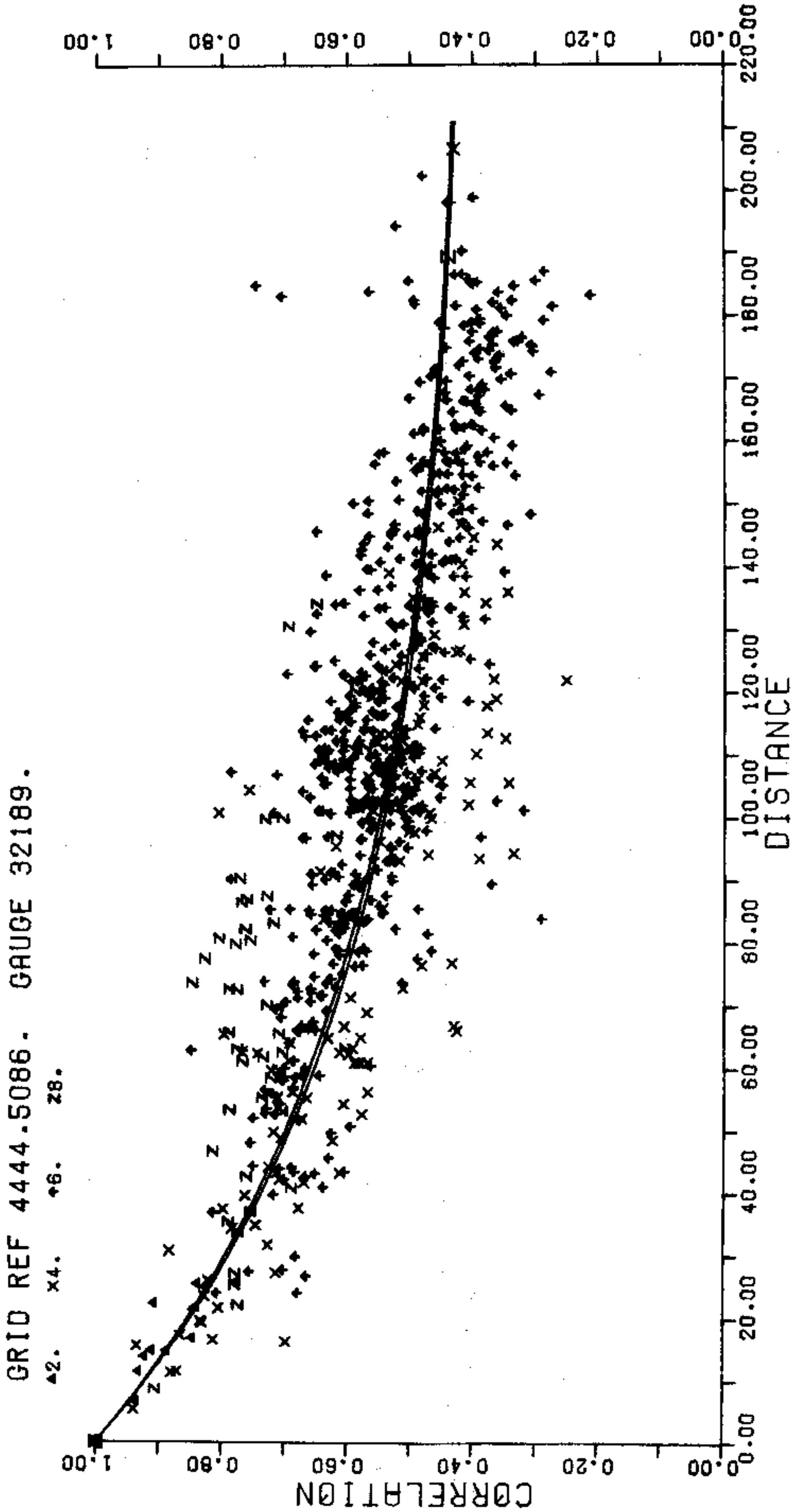


Figure 4.14b : Sample and fitted correlation functions.
Northern England : daily rainfall totals.

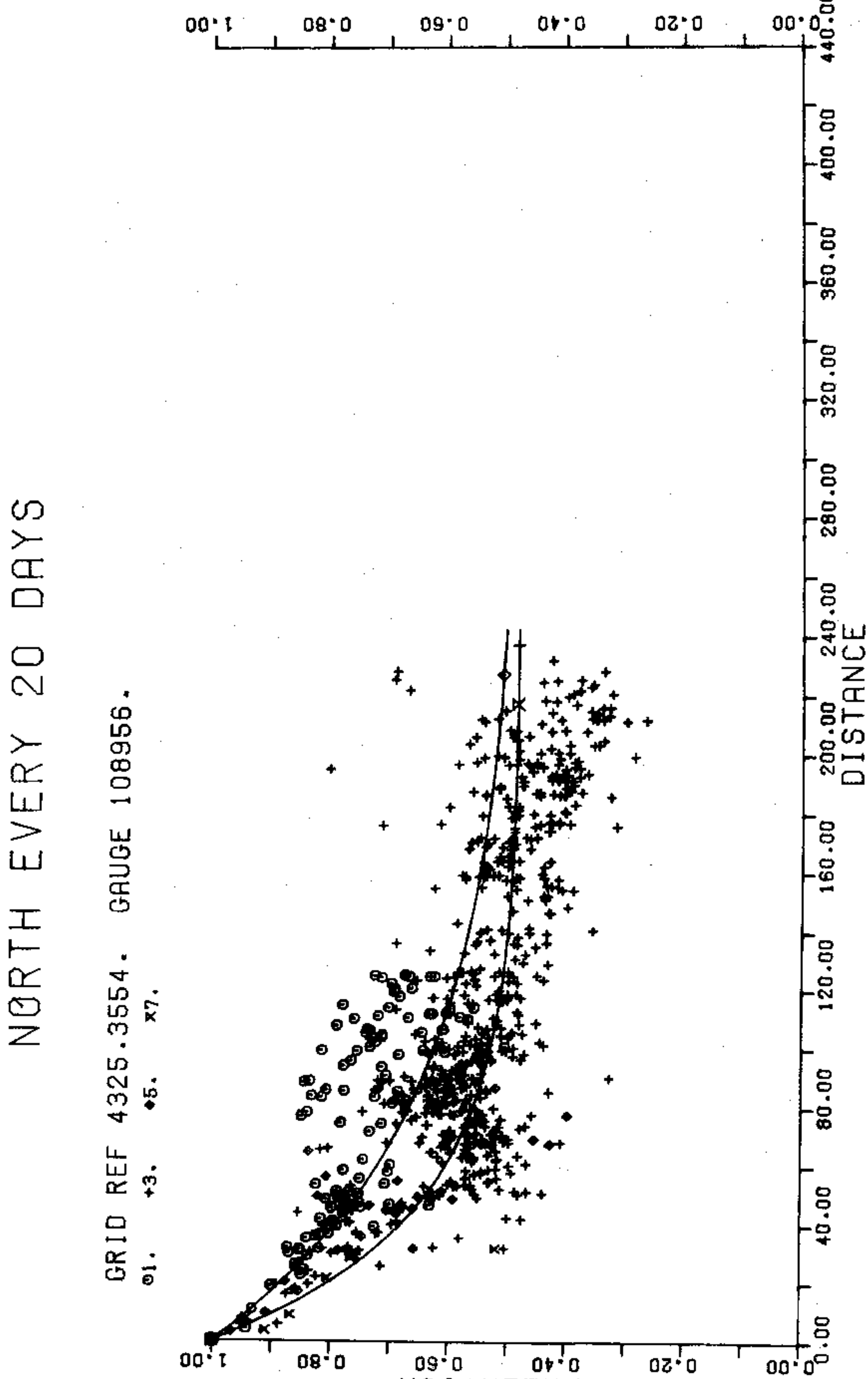


Figure 4.15a : Sample and fitted correlation functions.
Northern England : daily rainfall totals.

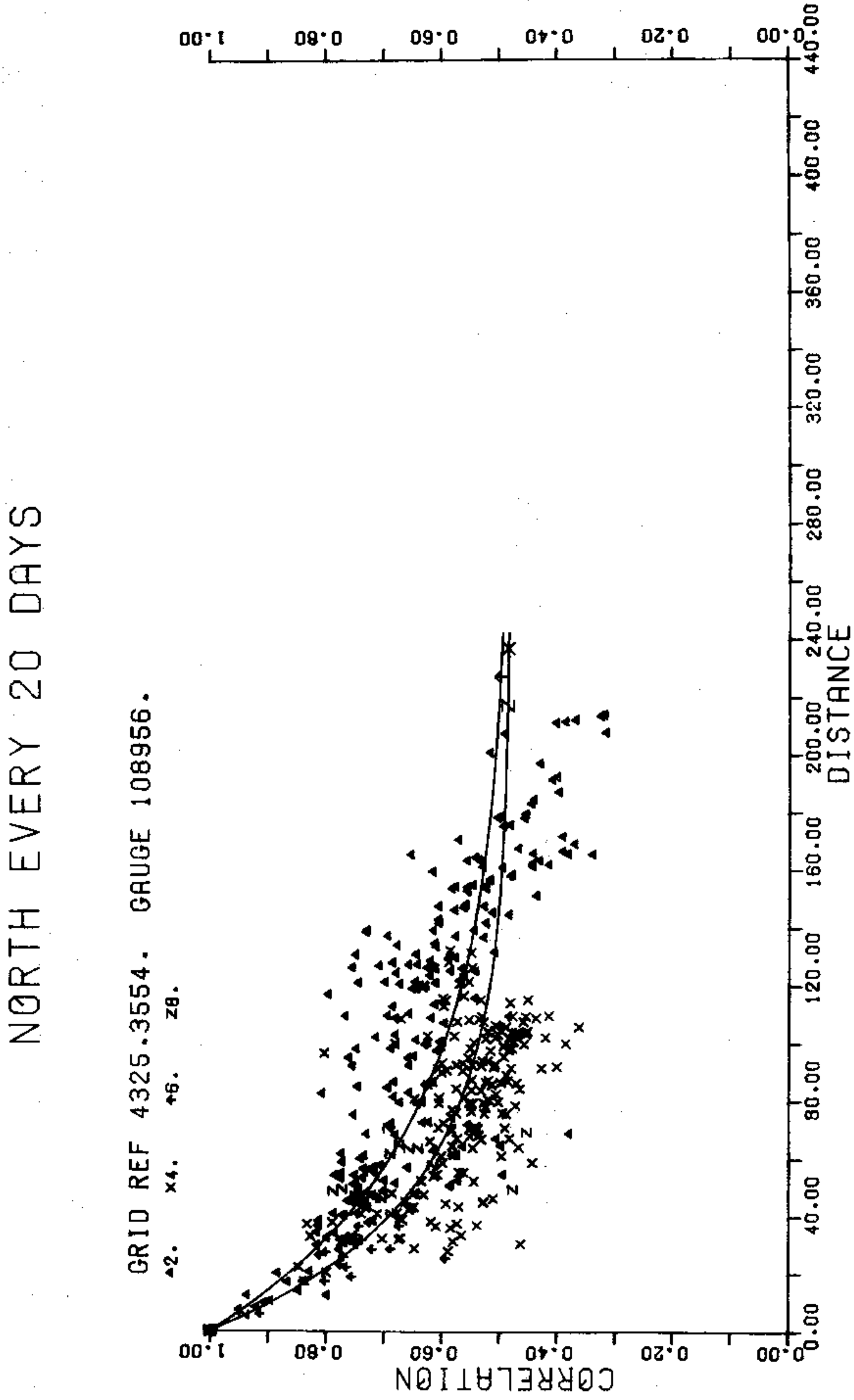


Figure 4.15b : Sample and fitted correlation functions.
Northern England : daily rainfall totals.

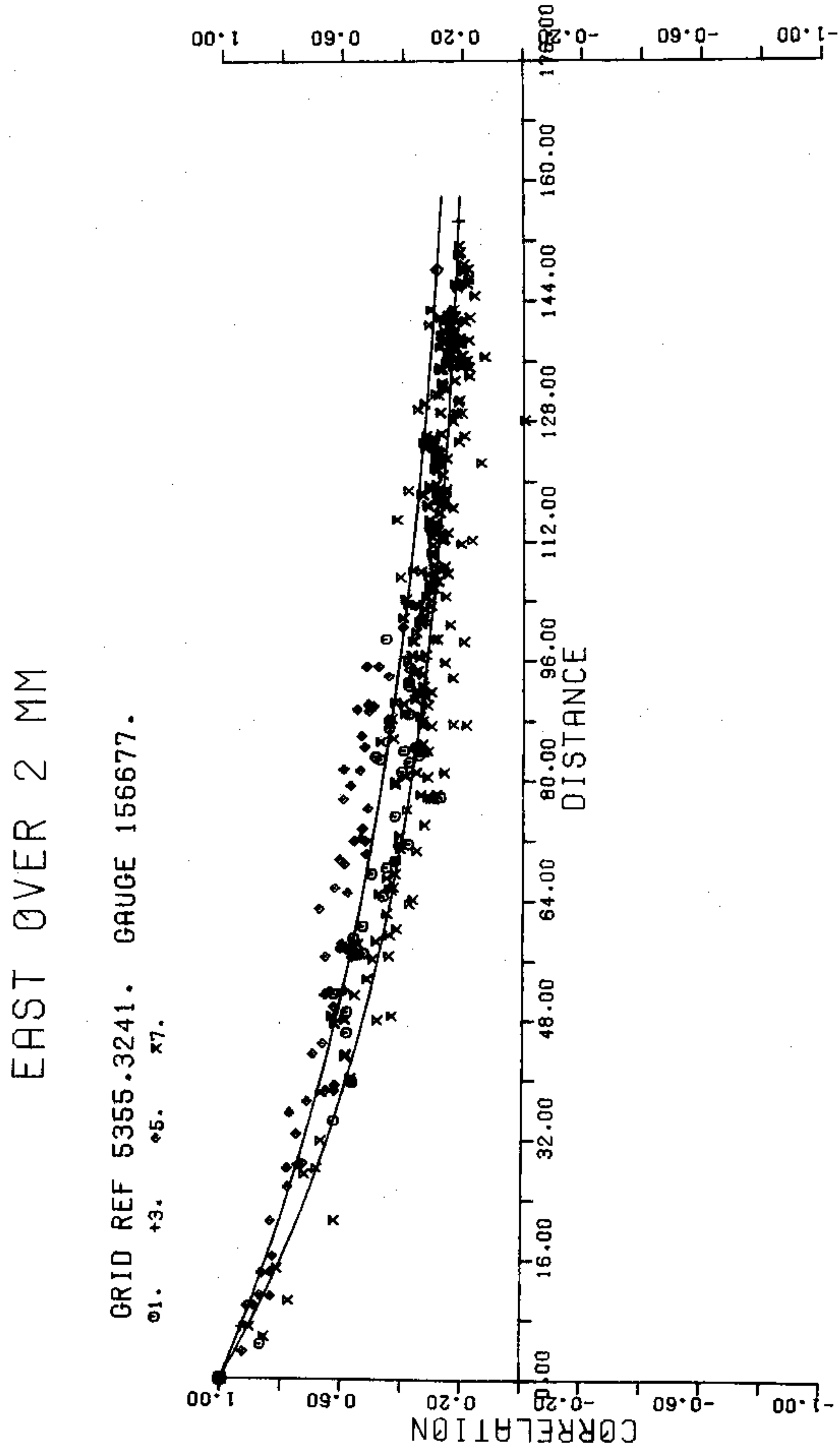


Figure 4.16a : Sample and fitted correlation functions.
Eastern England : days with rainfall over 2 mm.

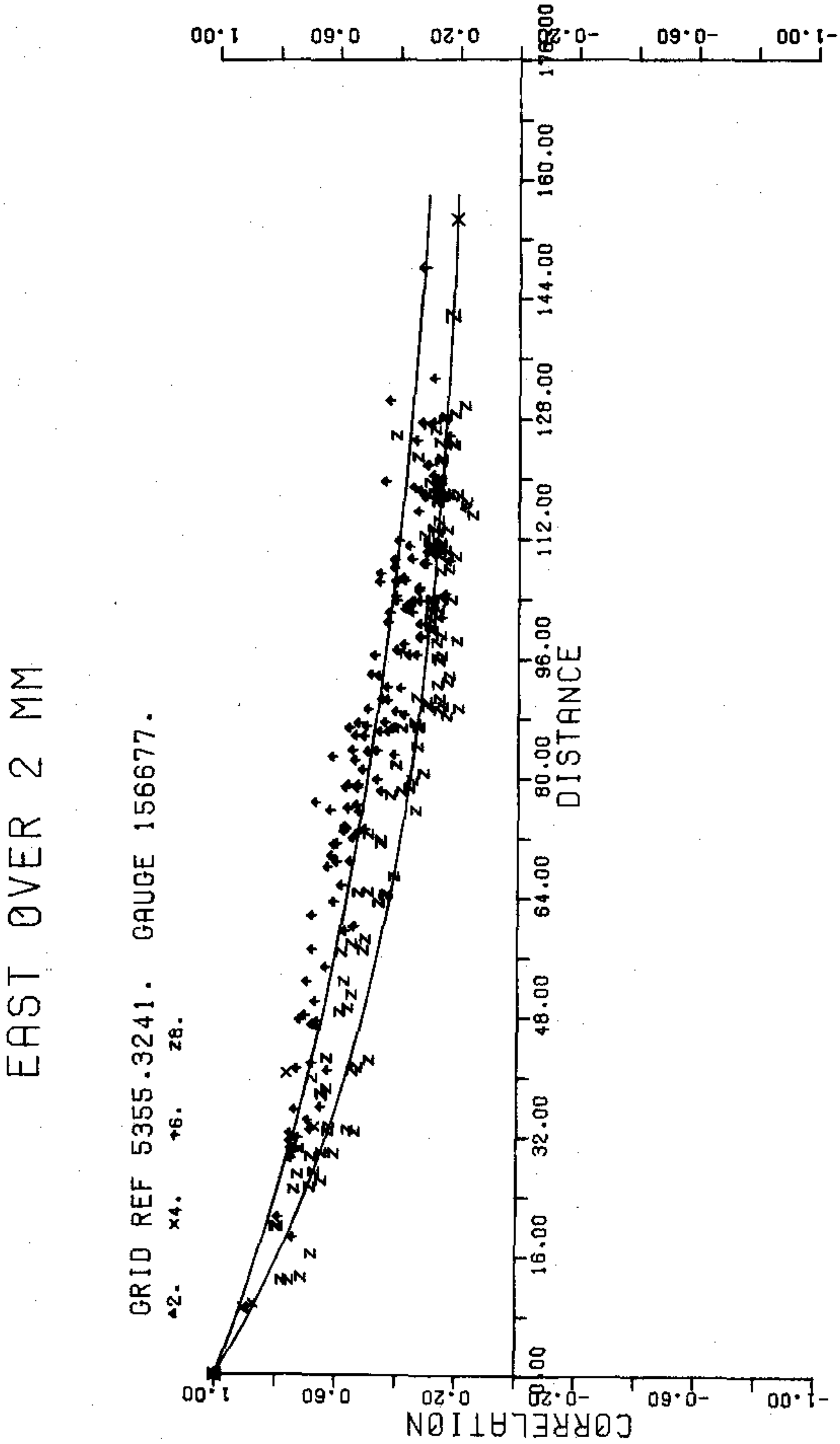


Figure 4.16b : Sample and fitted correlation functions.
Eastern England : days with rainfall over 2 mm.

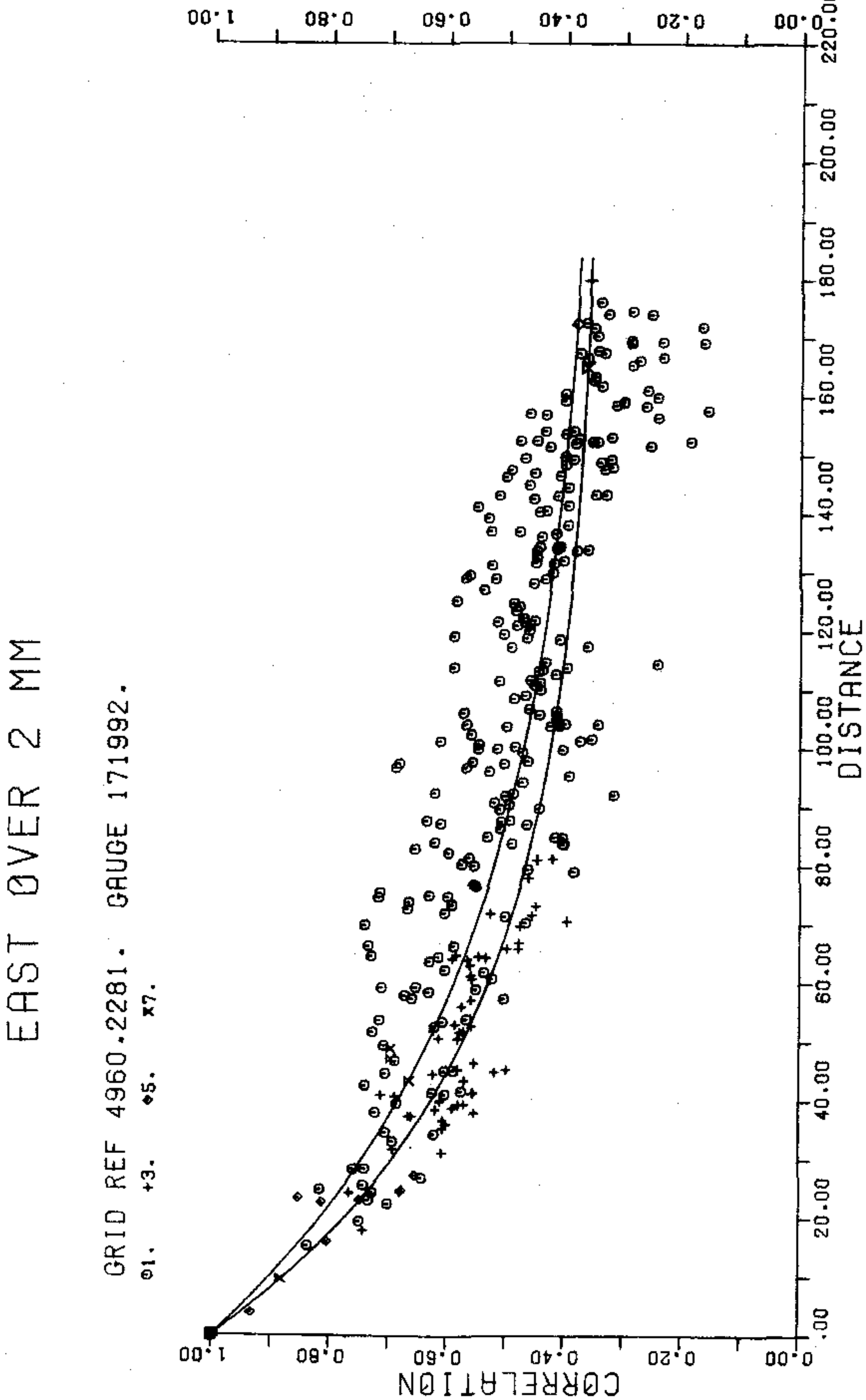


Figure 4.17a : Sample and fitted correlation functions.
Eastern England : days with rainfall over 2 mm.

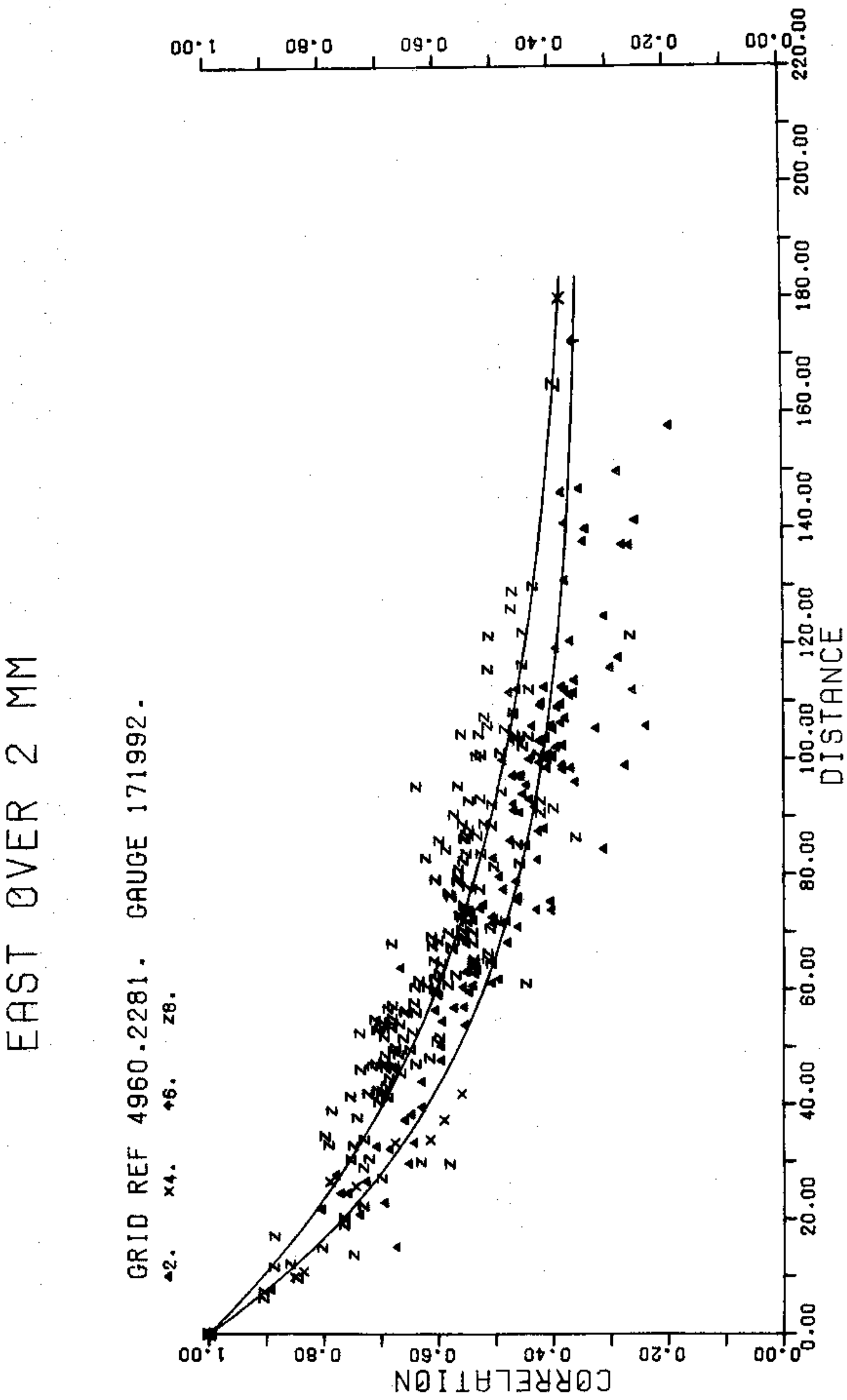


Figure 4.17b : Sample and fitted correlation functions.
Eastern England : days with rainfall over 2mm.

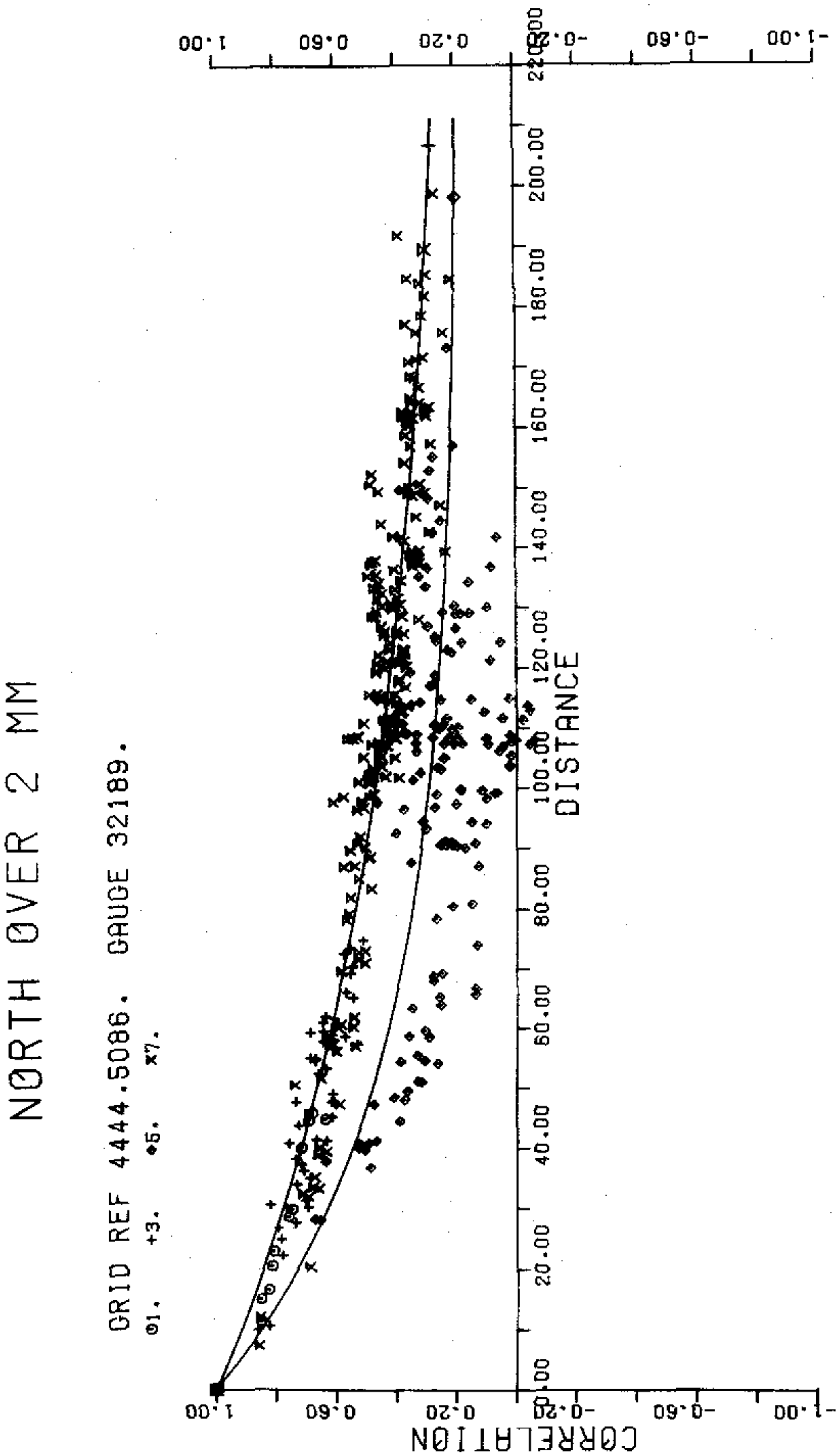


Figure 4.18a : Sample and fitted correlation functions.
Northern England : days with rainfall over 2 mm.

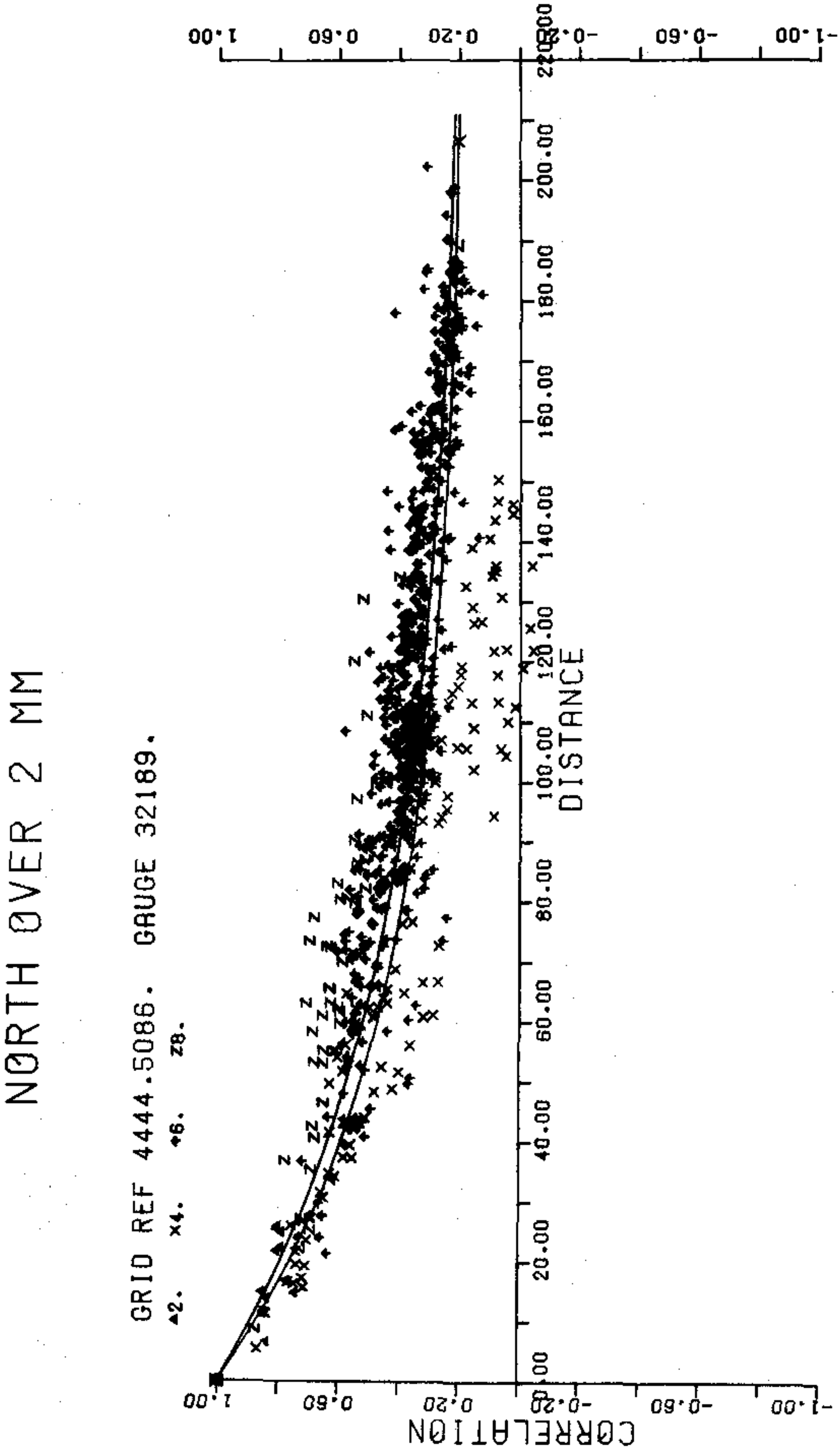


Figure 4.18b : Sample and fitted correlation functions.
Northern England : days with rainfall over 2 mm.

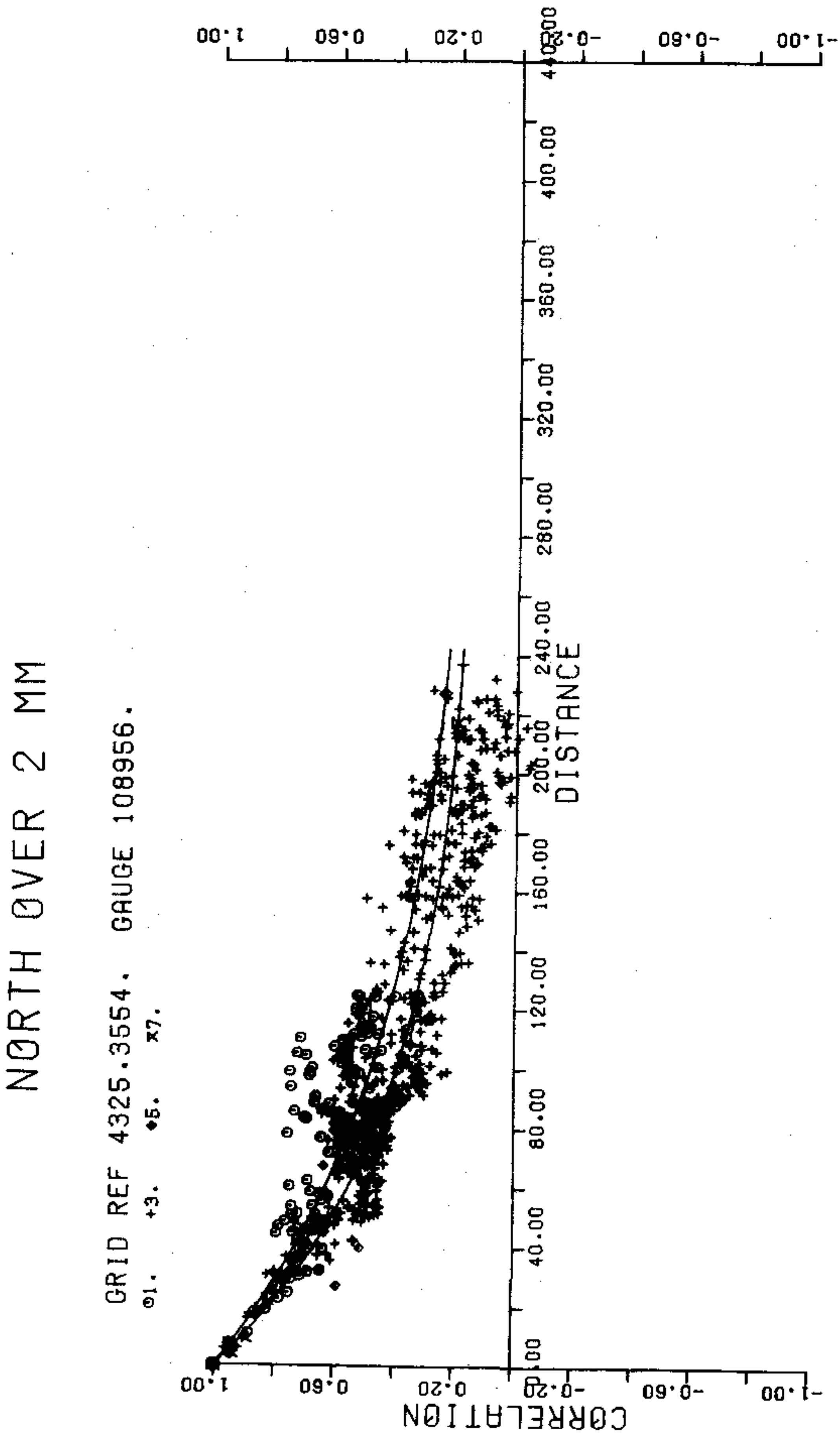


Figure 4.19a : Sample and fitted correlation functions.
Northern England : days with rainfall over 2 mm.

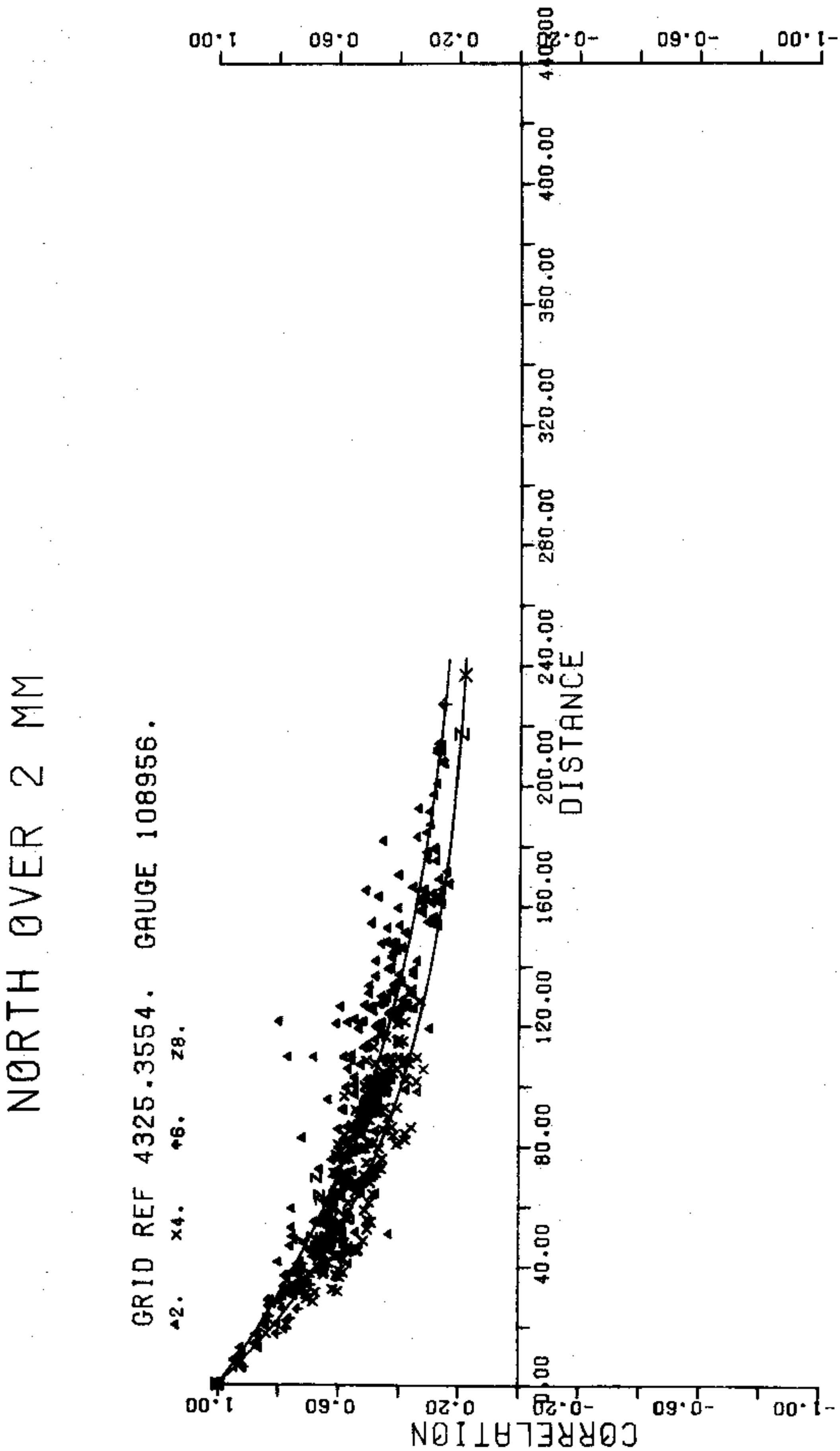


Figure 4.19b : Sample and fitted correlation functions.
Northern England : days with rainfall over 2 mm.

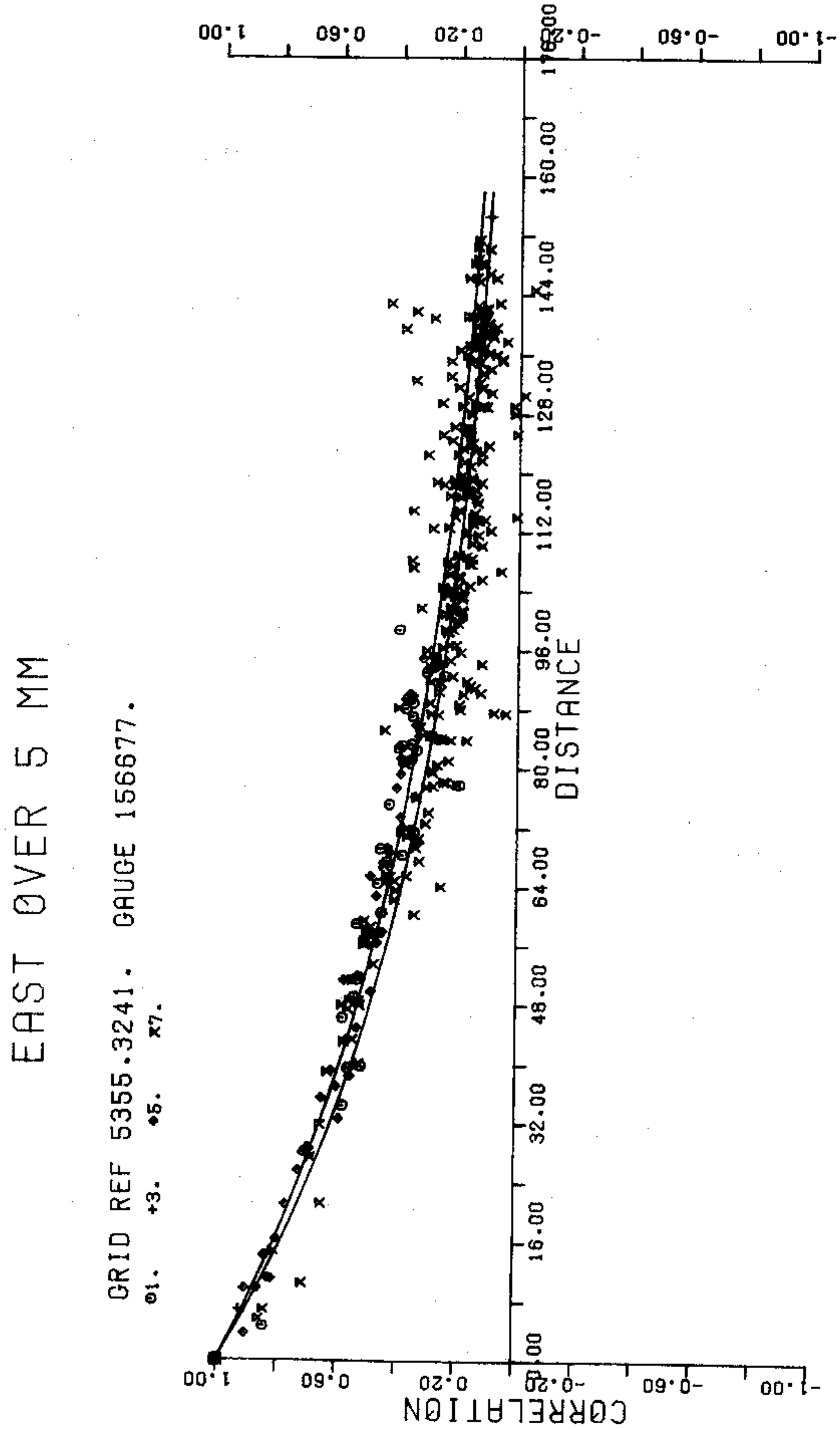


Figure 4.20a : Sample and fitted correlation functions.
Eastern England : days with rainfall over 5 mm.

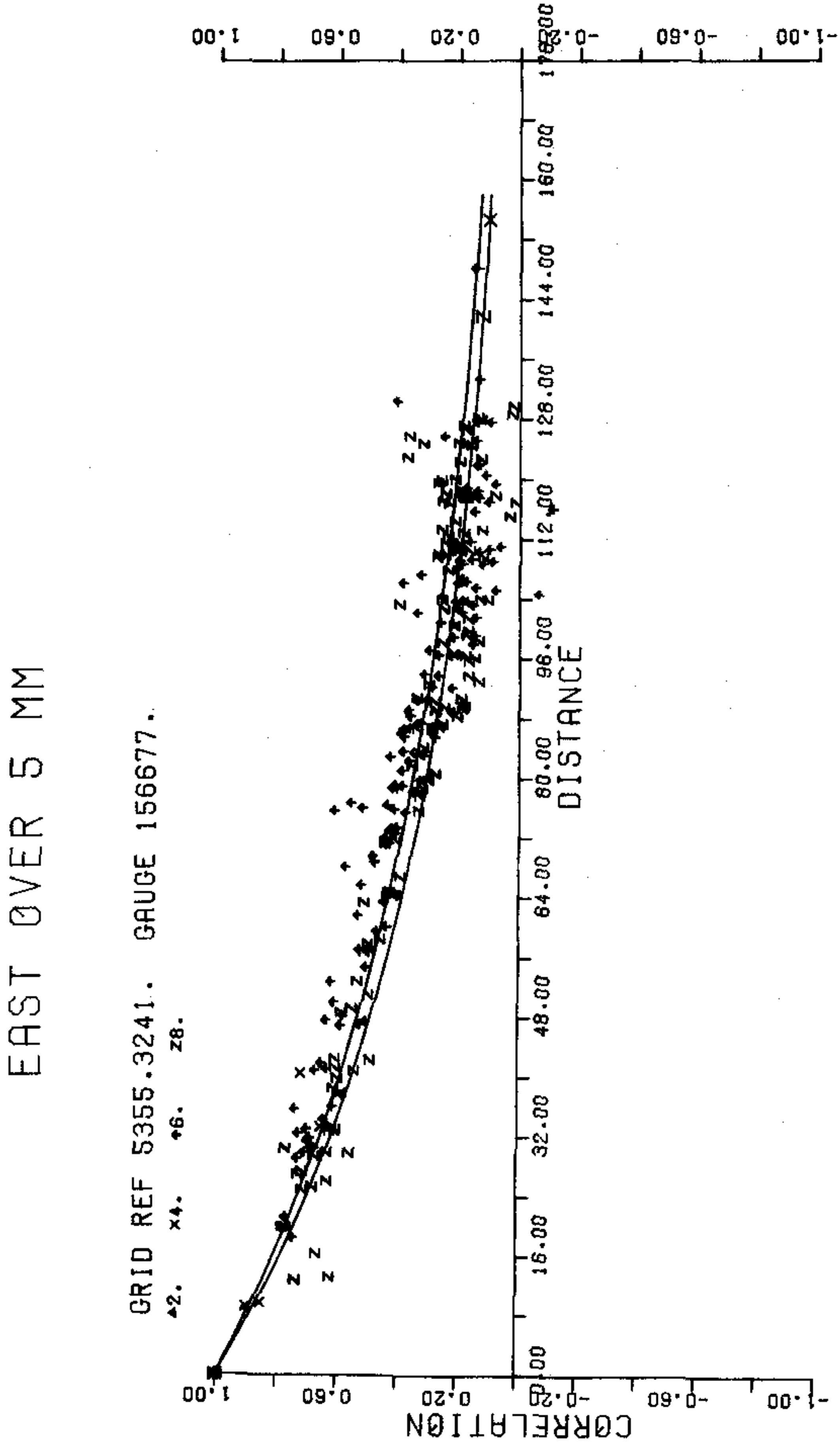


Figure 4.20b : Sample and fitted correlation functions.
Eastern England : days with rainfall over 5 mm.

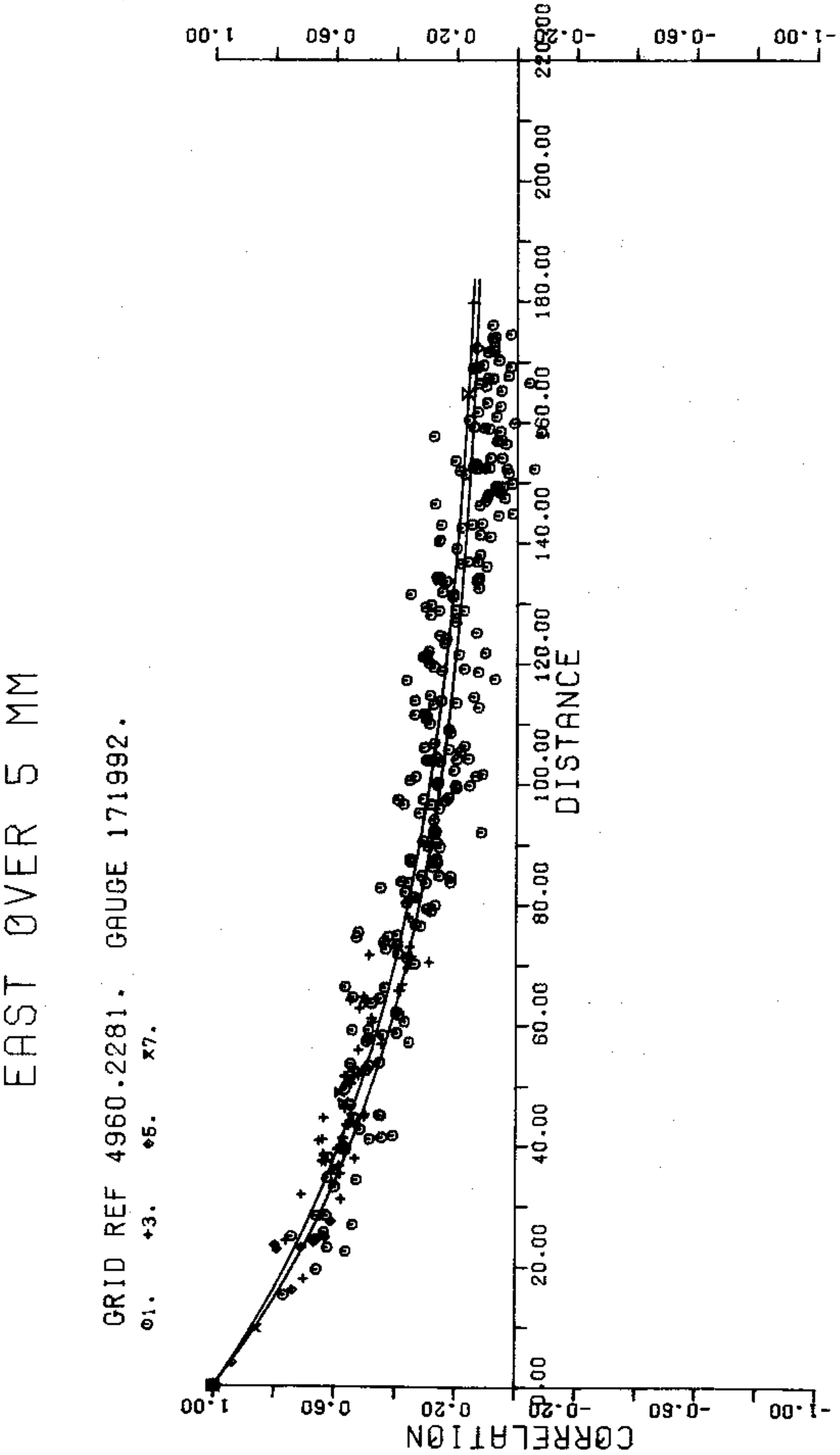


Figure 4.21a : Sample and fitted correlation functions.
Eastern England : days with rainfall over 5mm.

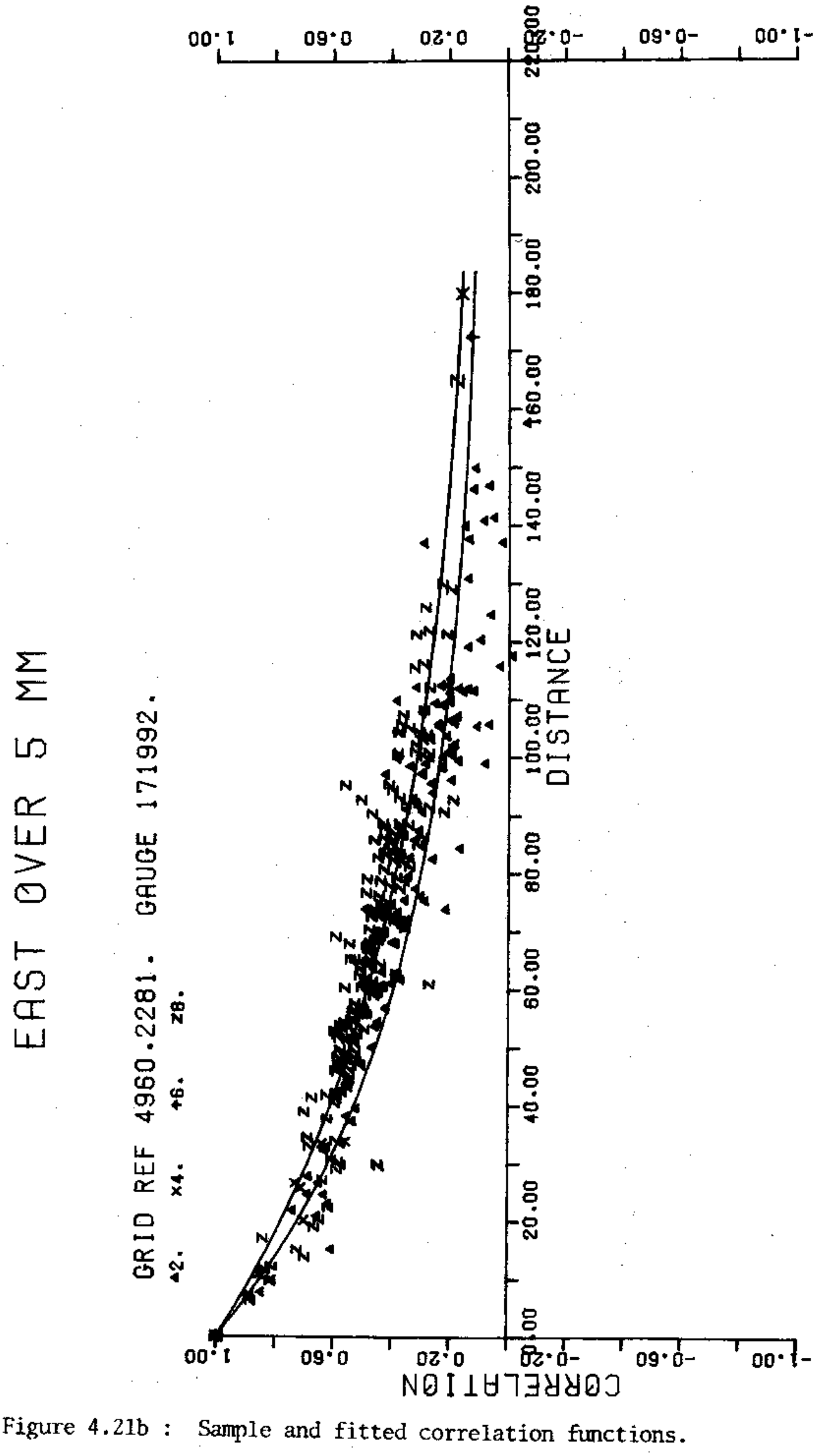


Figure 4.21b : Sample and fitted correlation functions.
Eastern England : days with rainfall over 5 mm.

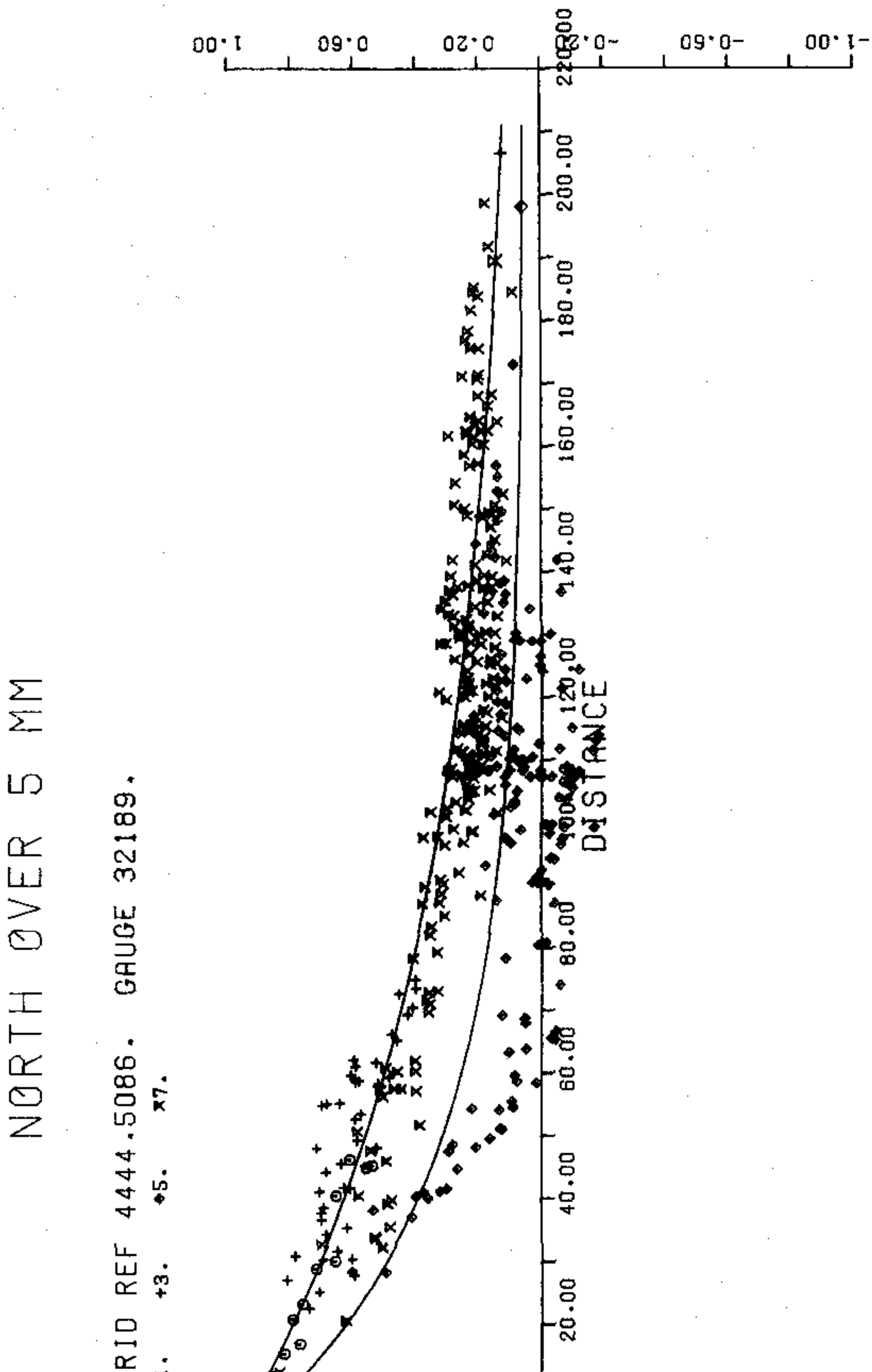


Figure 4.22a : Sample and fitted correlation functions.
Northern England : days with rainfall over 5 mm.

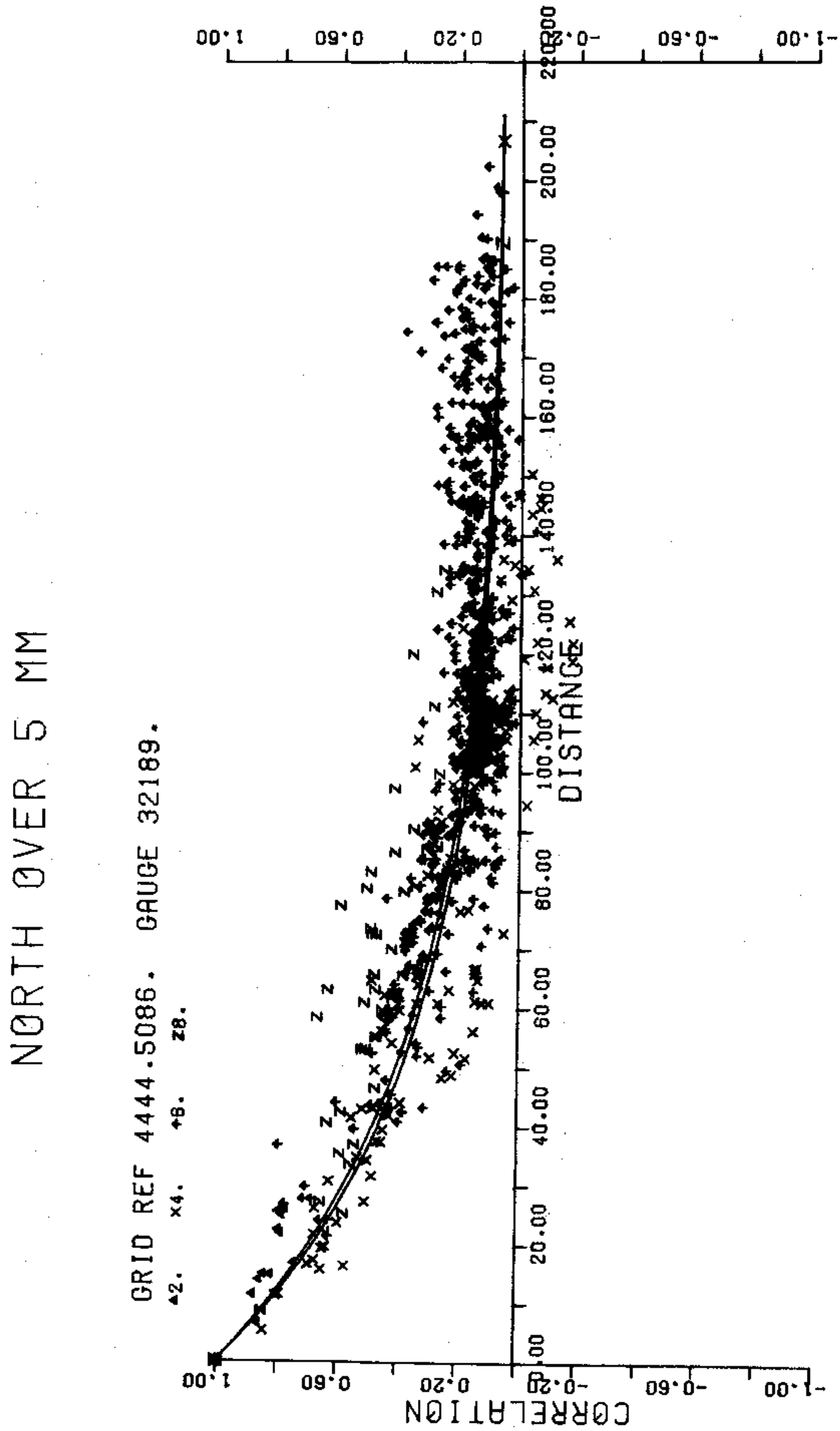


Figure 4.22b : Sample and fitted correlation functions.
Northern England : days with rainfall over 5 mm.

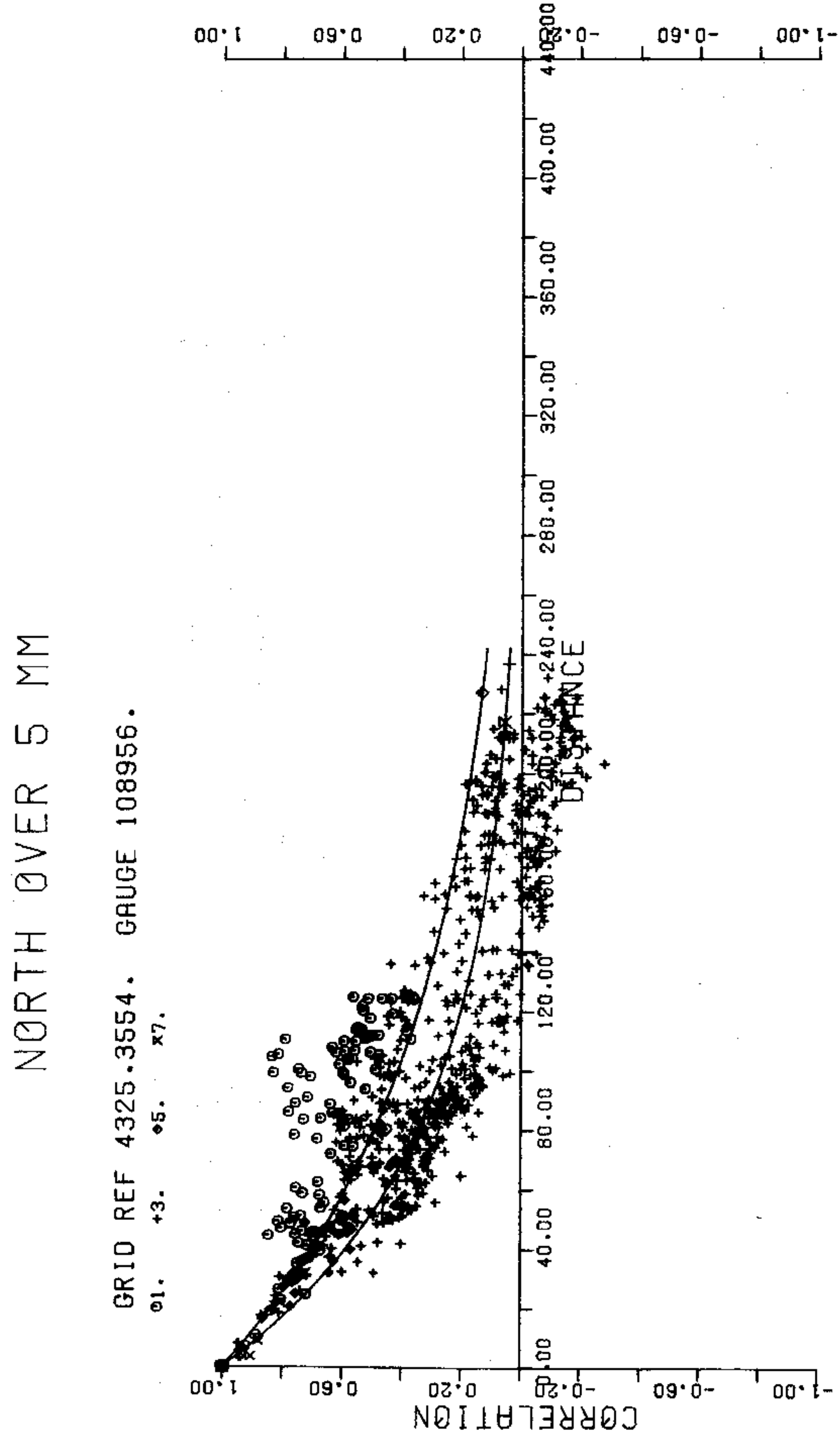


Figure 4.23a : Sample and fitted correlation functions.
Northern England : days with rainfall over 5 mm.

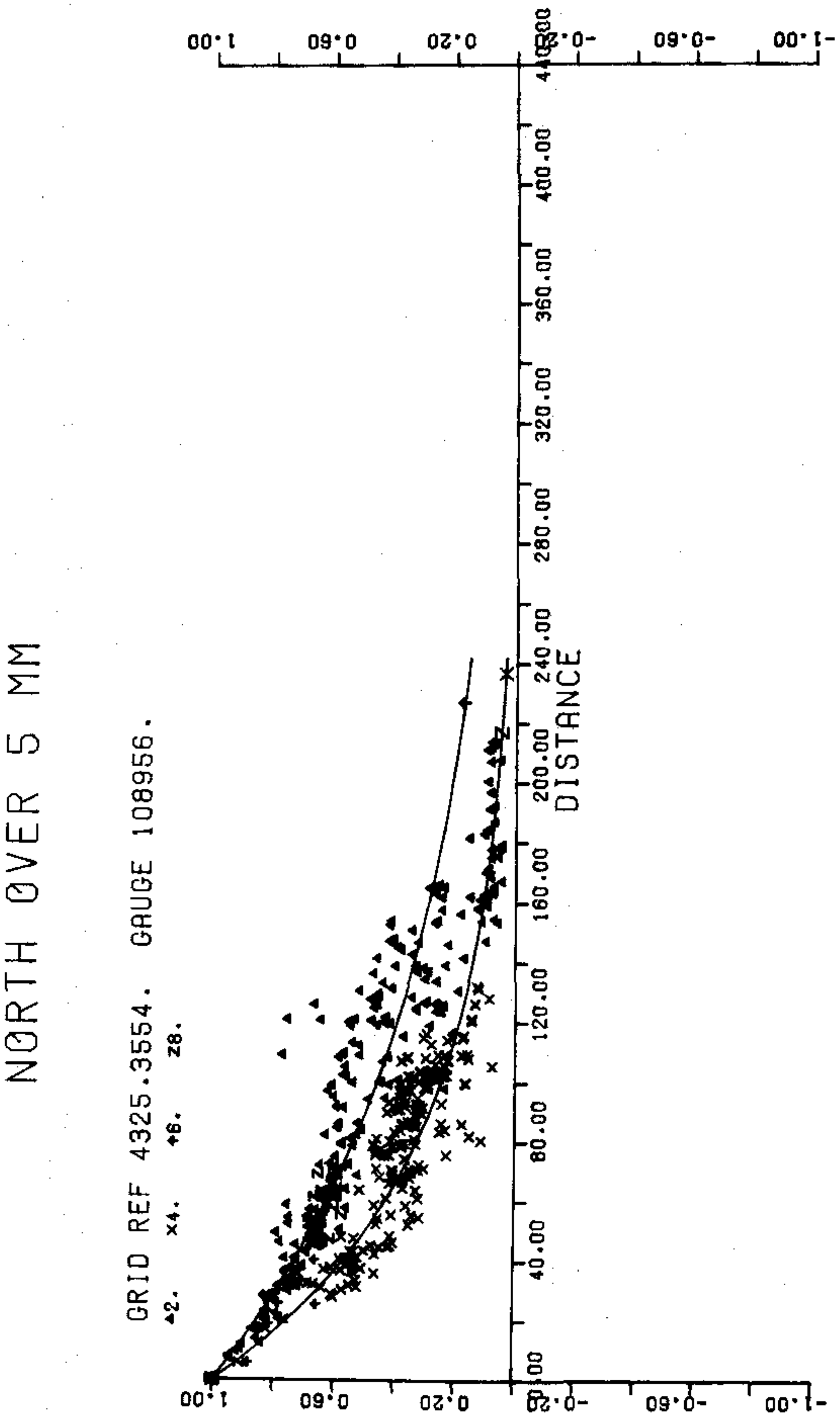


Figure 4.23b : Sample and fitted correlation functions.
Northern England : days with rainfall over 5mm.

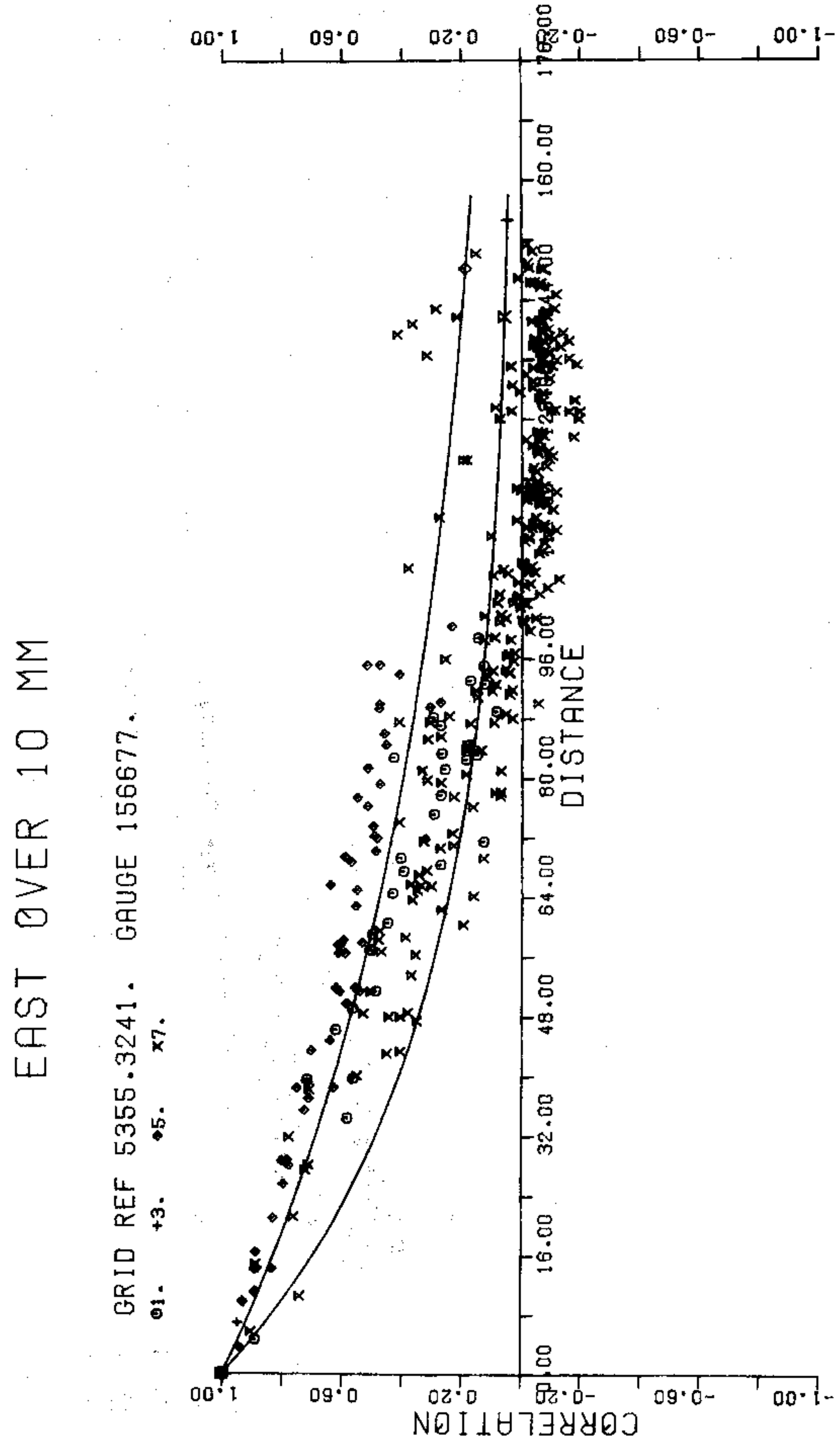


Figure 4.24a : Sample and fitted correlation functions.
Eastern England : days with rainfall over 10 mm.

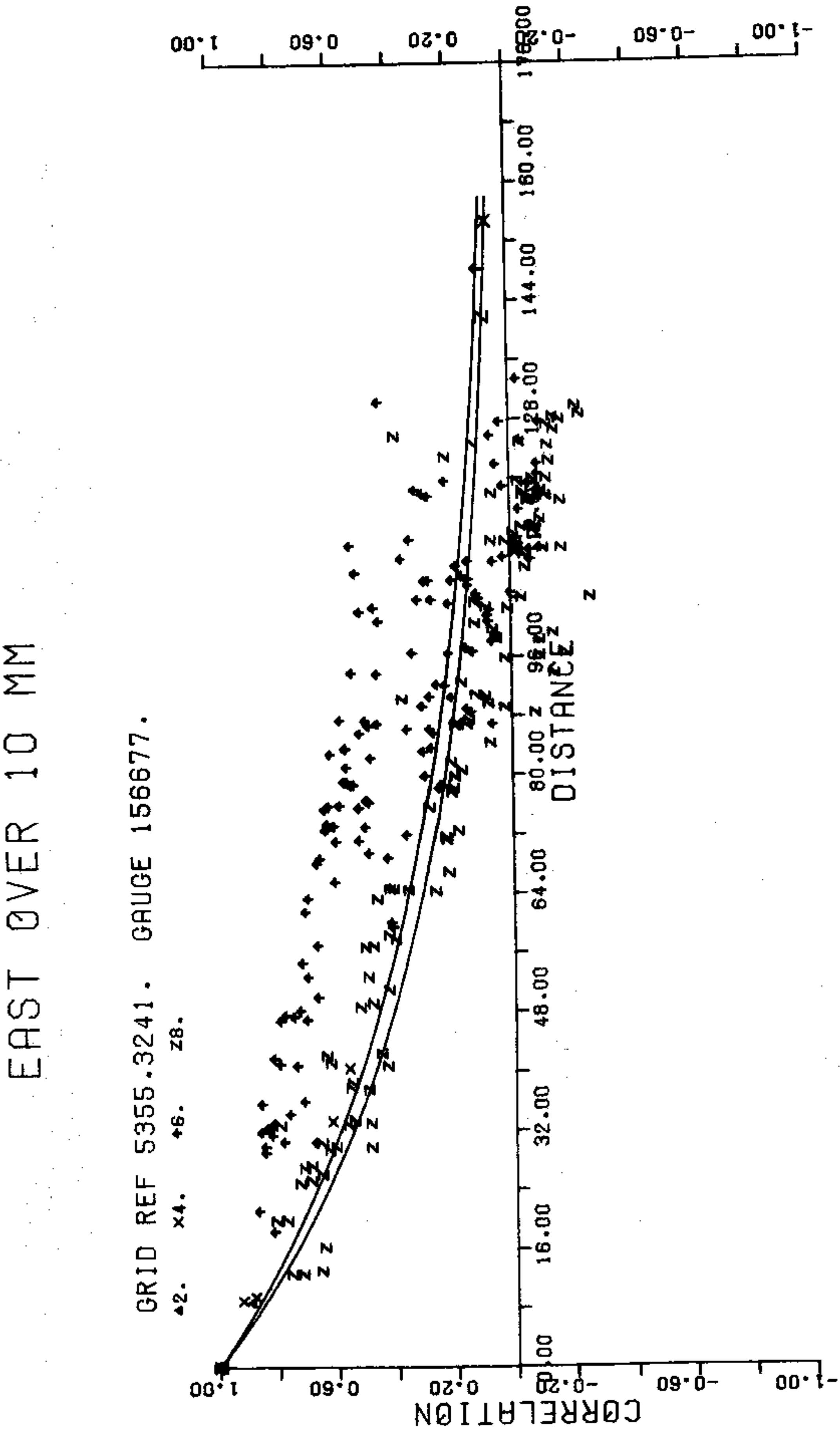


Figure 4.24b: Sample and fitted correlation functions.
Eastern England : days with rainfall over 10 mm.

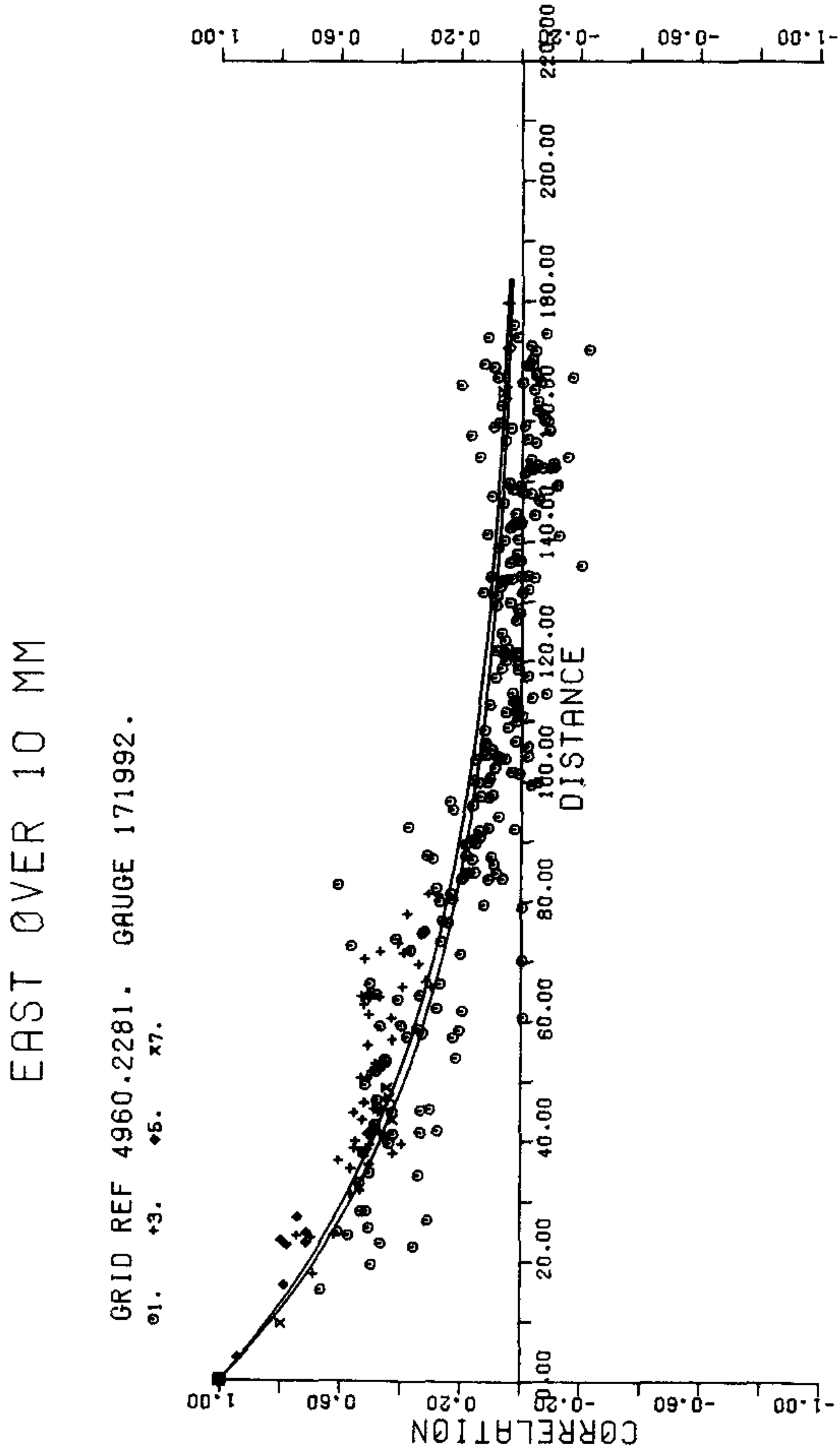


Figure 4.25a : Sample and fitted correlation functions.
Eastern England : days with rainfall over 10 mm.

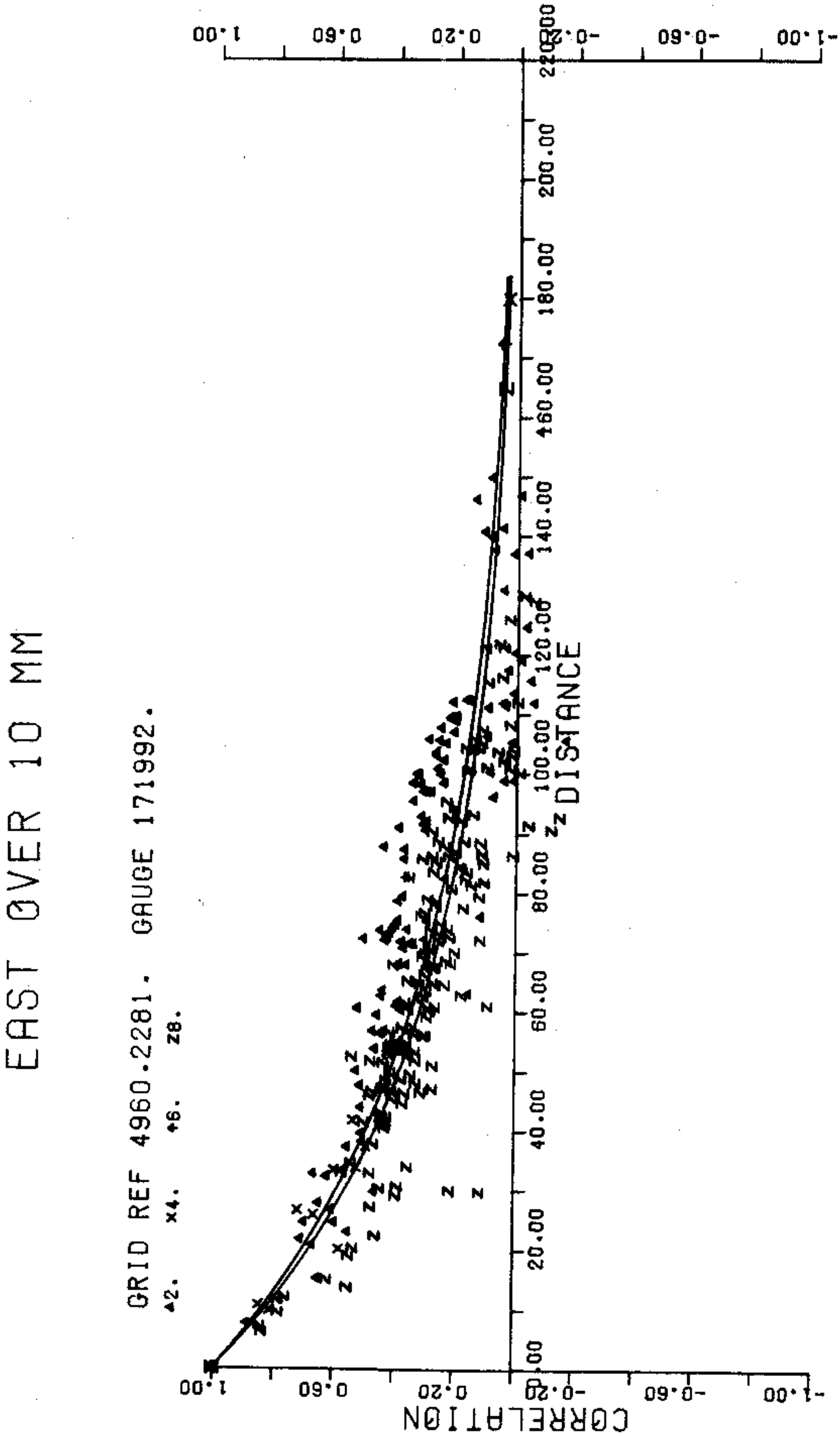


Figure 4.25b : Sample and fitted correlation functions.
Eastern England : days with rainfall over 10 mm.

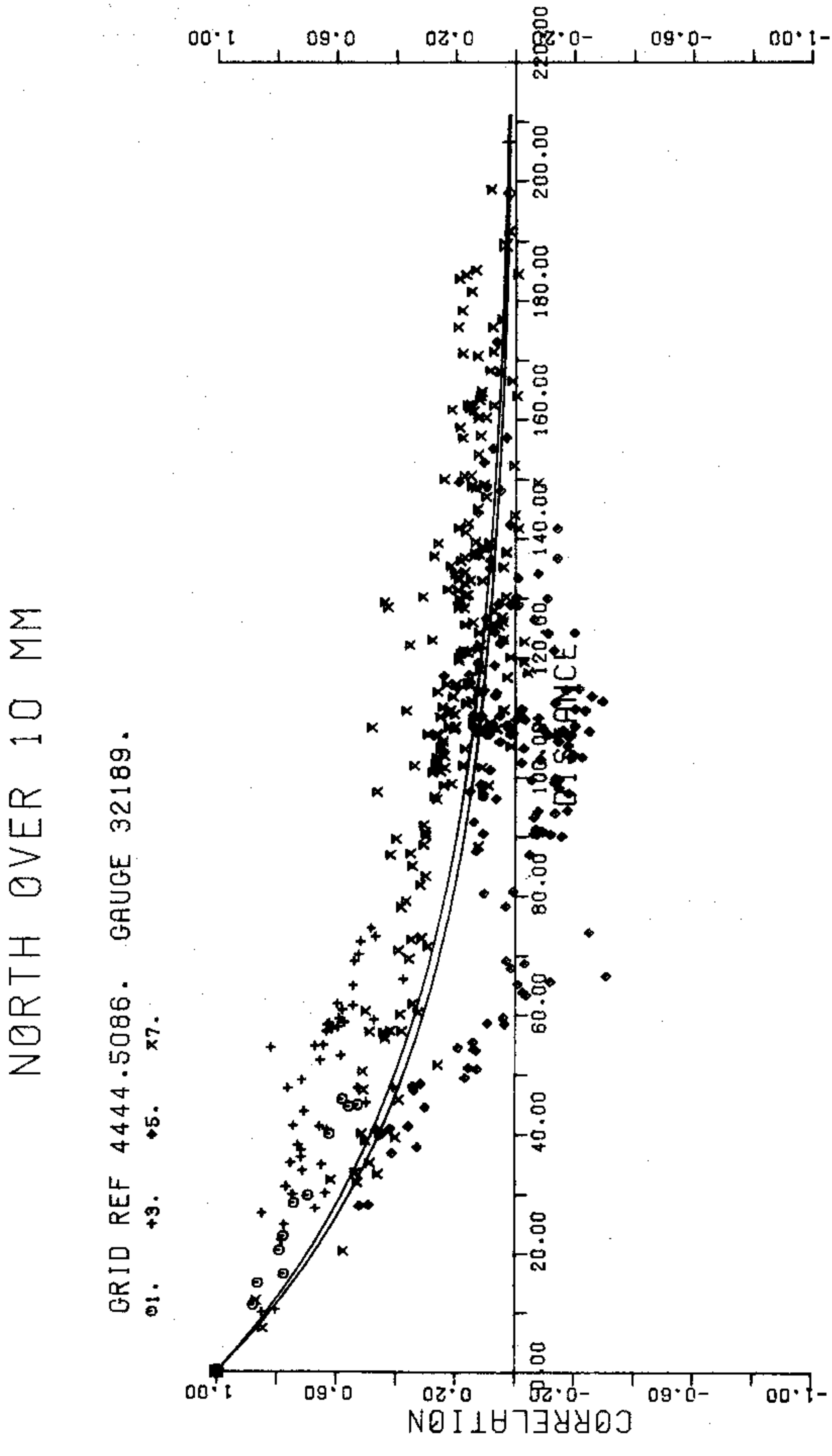


Figure 4.26a : Sample and fitted correlation functions.
Northern England : days with rainfall over 10 mm.

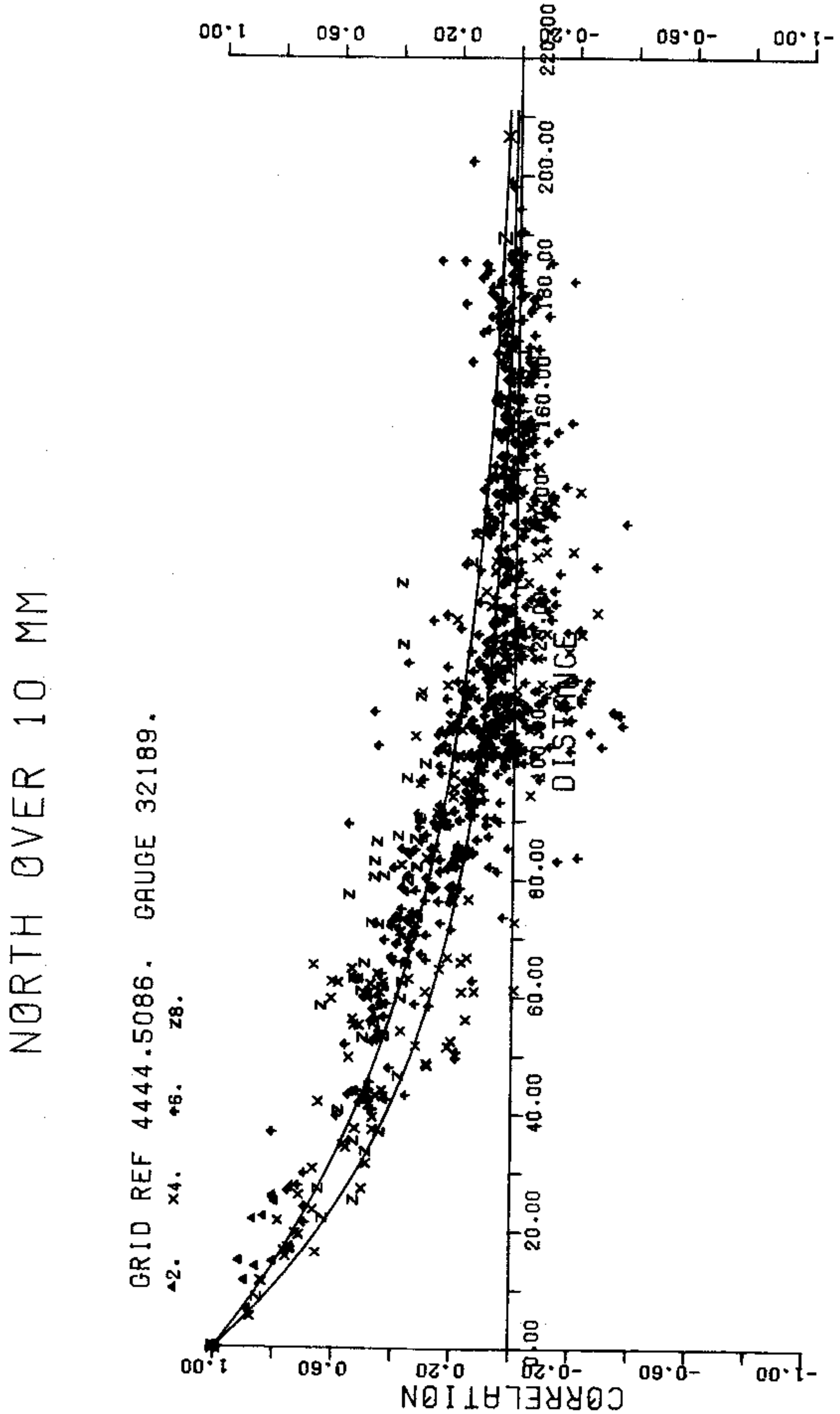


Figure 4.26b : Sample and fitted correlation functions.
Northern England : days with rainfall over 10 mm.

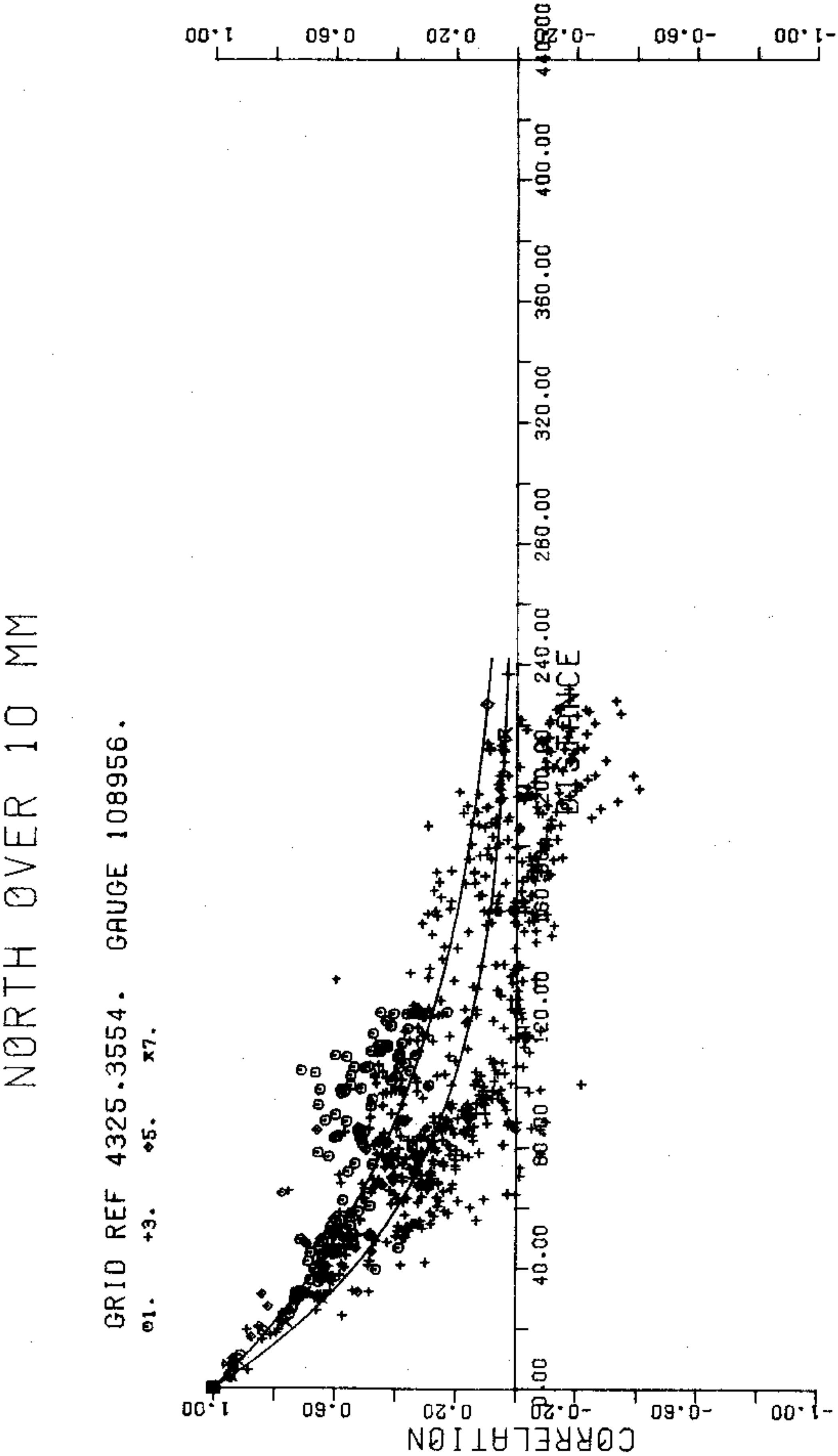


Figure 4.27a : Sample and fitted correlation functions.
Northern England : days with rainfall over 10 mm.

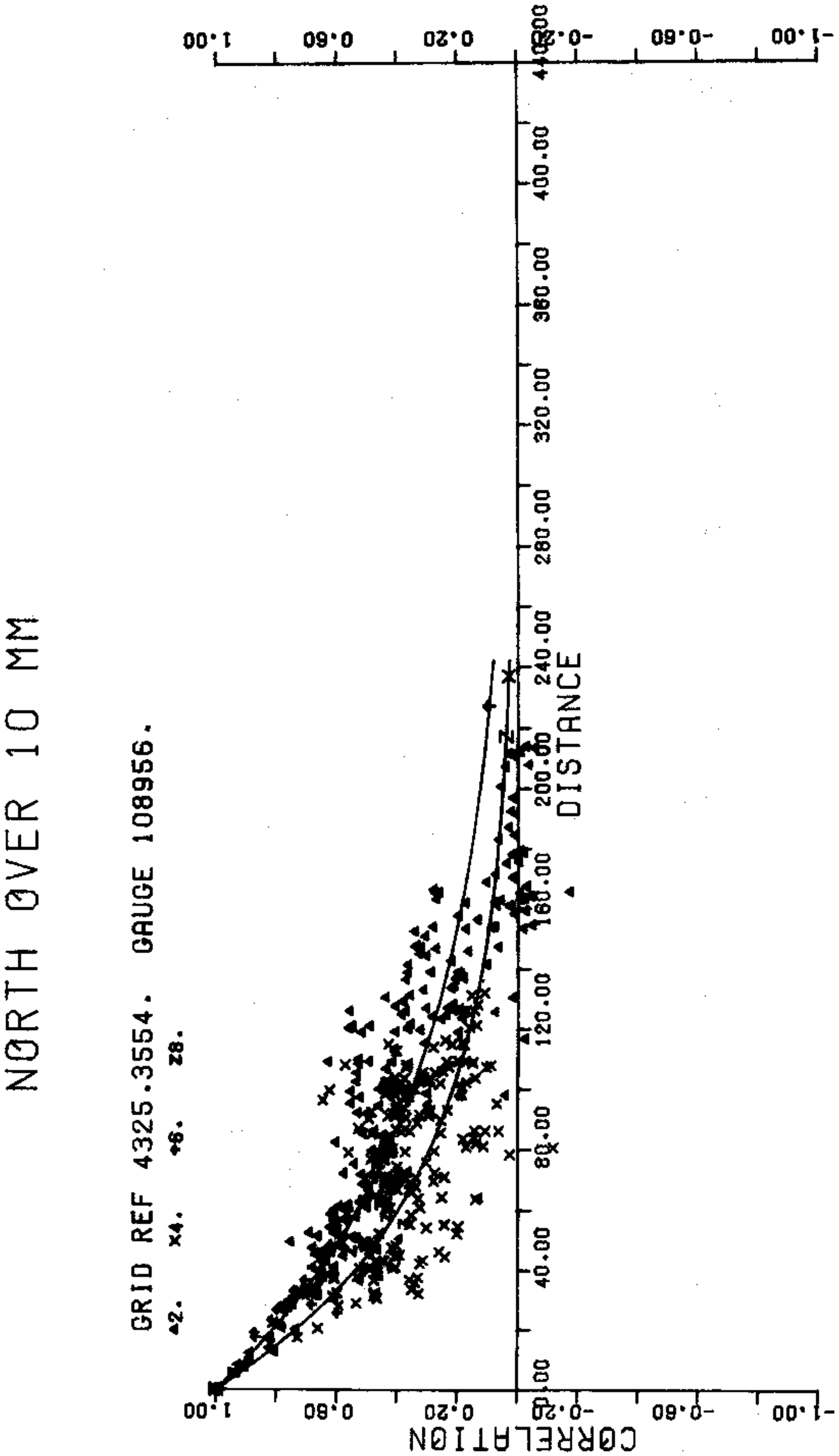


Figure 4.27b : Sample and fitted correlation functions.
Northern England : days with rainfall over 10 mm.

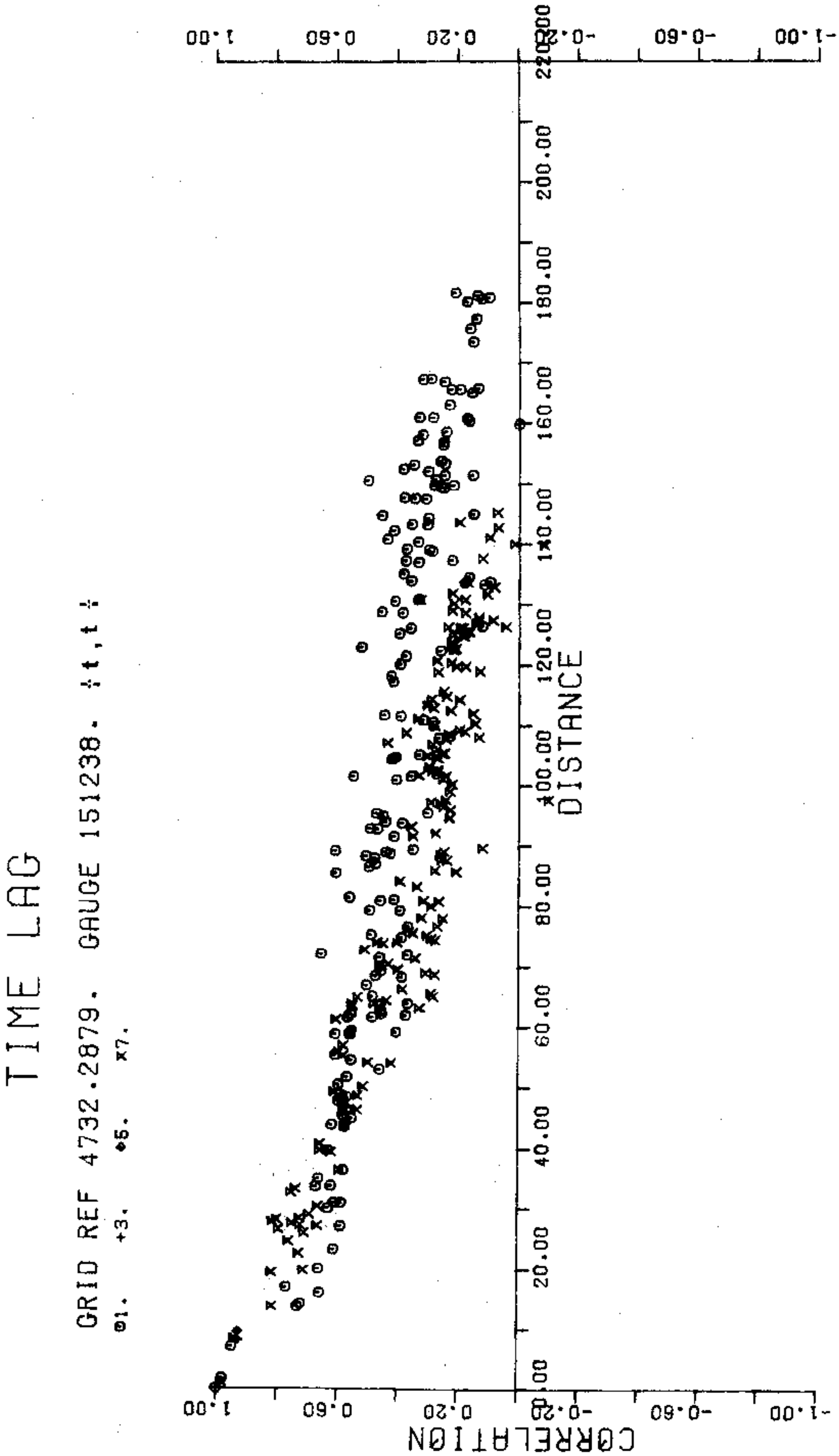


Figure 4.28a : Sample correlation functions.
Eastern England: correlation between rainfalls at stations on the same day (rainfall over 5 mm).

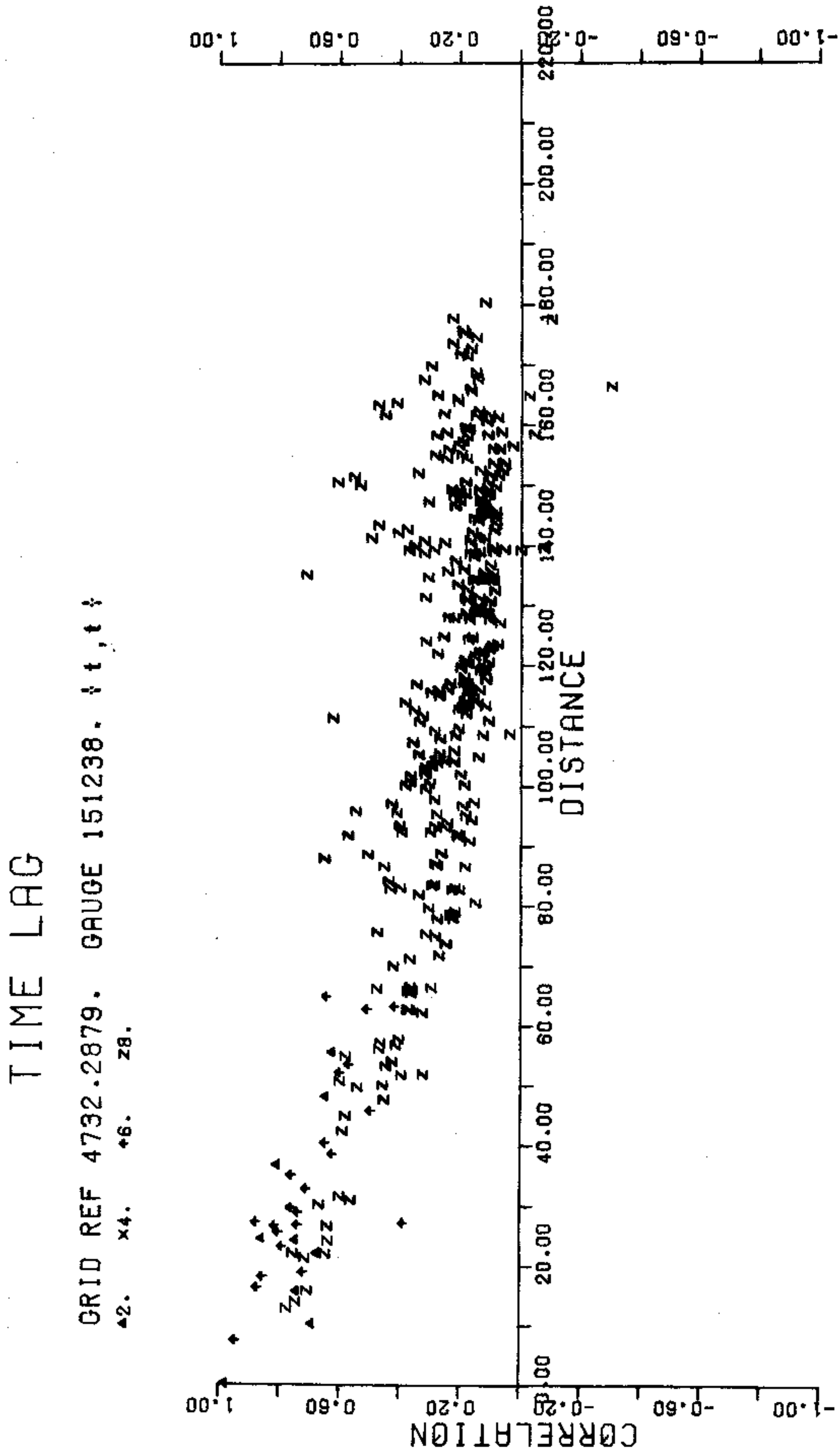


Figure 4.28b : Sample correlation functions.
Eastern England: correlation between rainfalls at stations on the same day (rainfall over 5 mm).

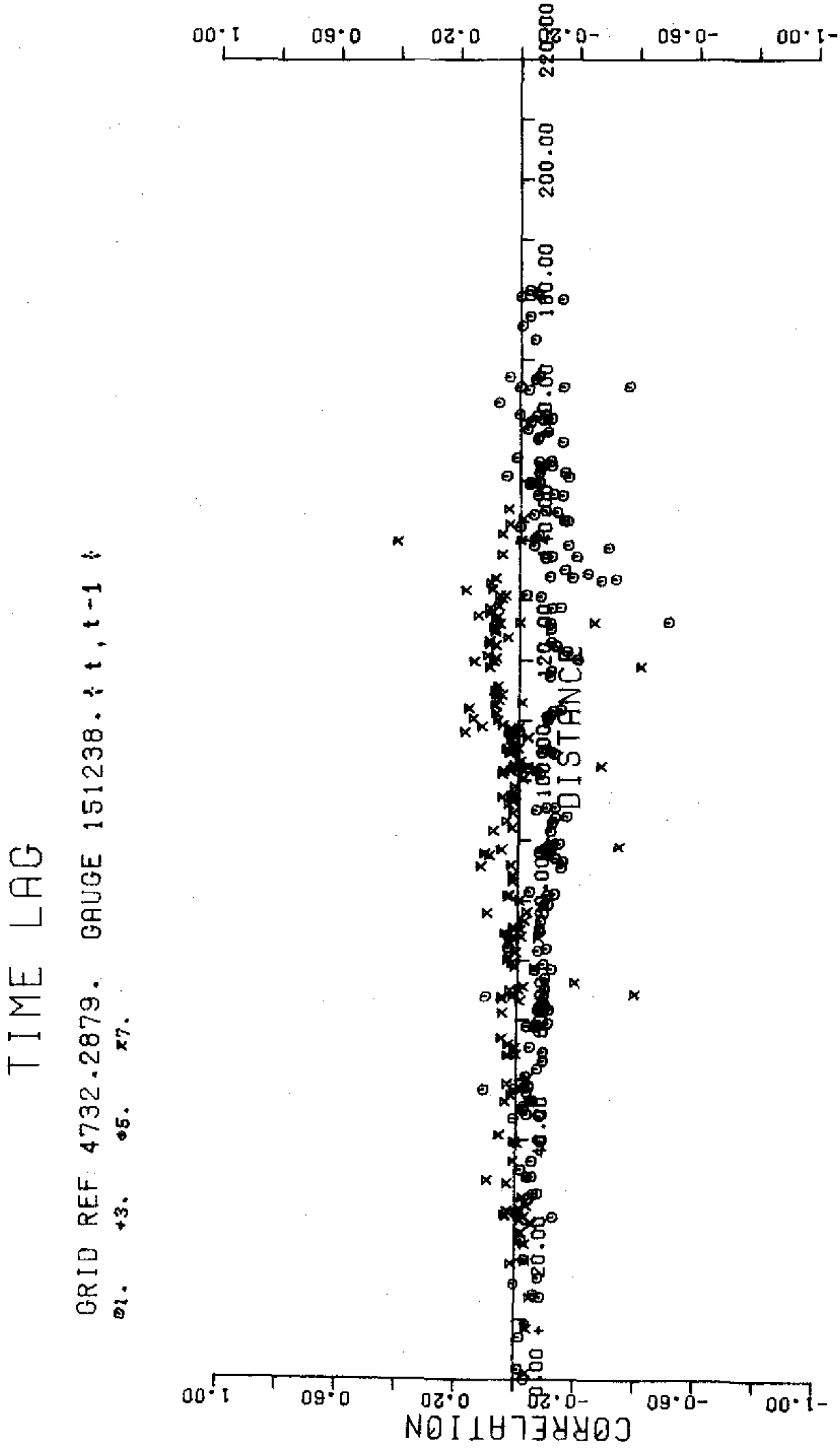


Figure 4.29a : Sample correlation functions. Eastern England: Correlation between rainfall (on day with rain over 5 mm) with *previous day's total at other stations.*

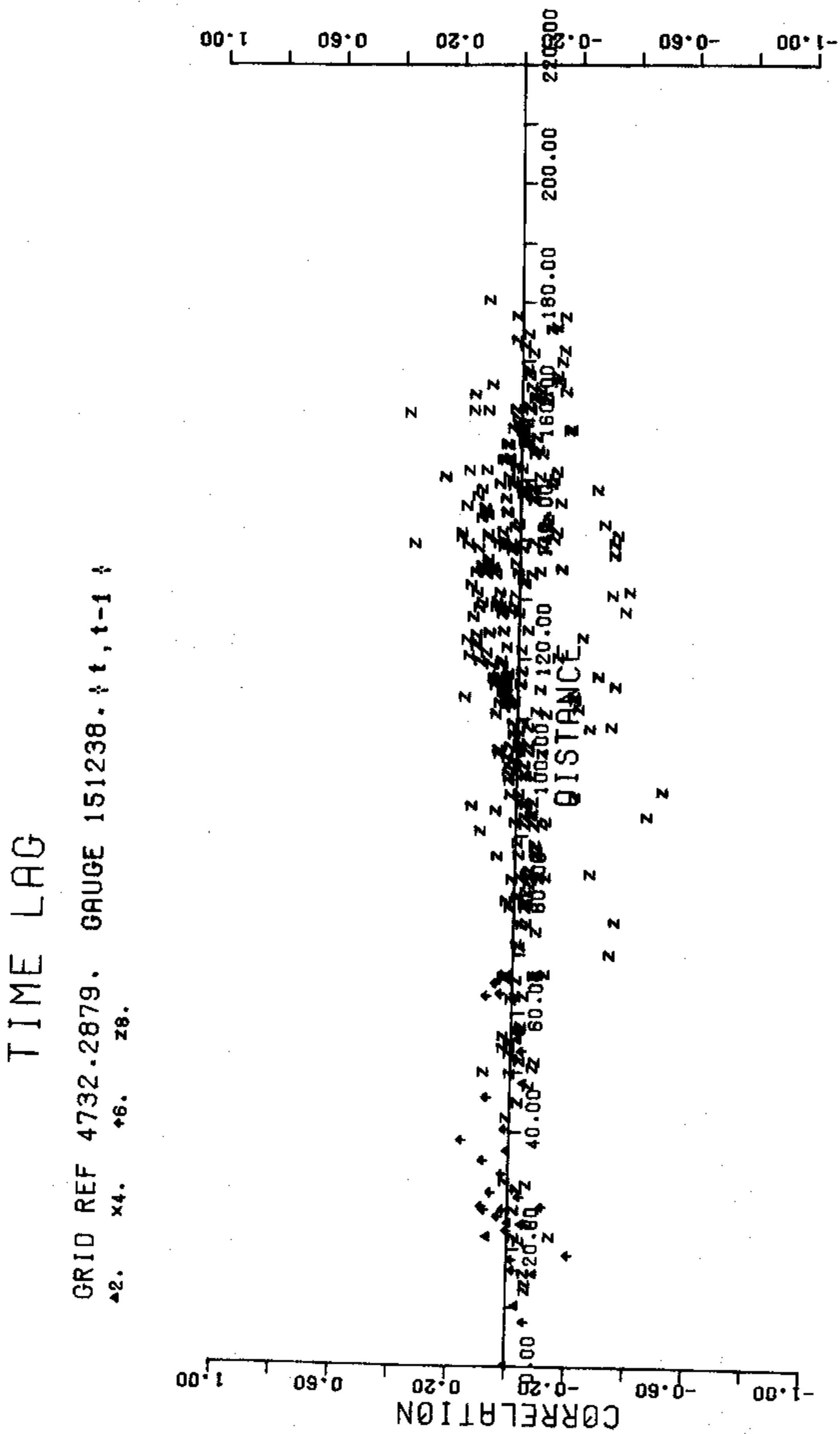


Figure 4.29b : Sample correlation functions. Eastern England: Correlation between rainfall (on day with rain over 5 mm) with previous day's total at other stations.

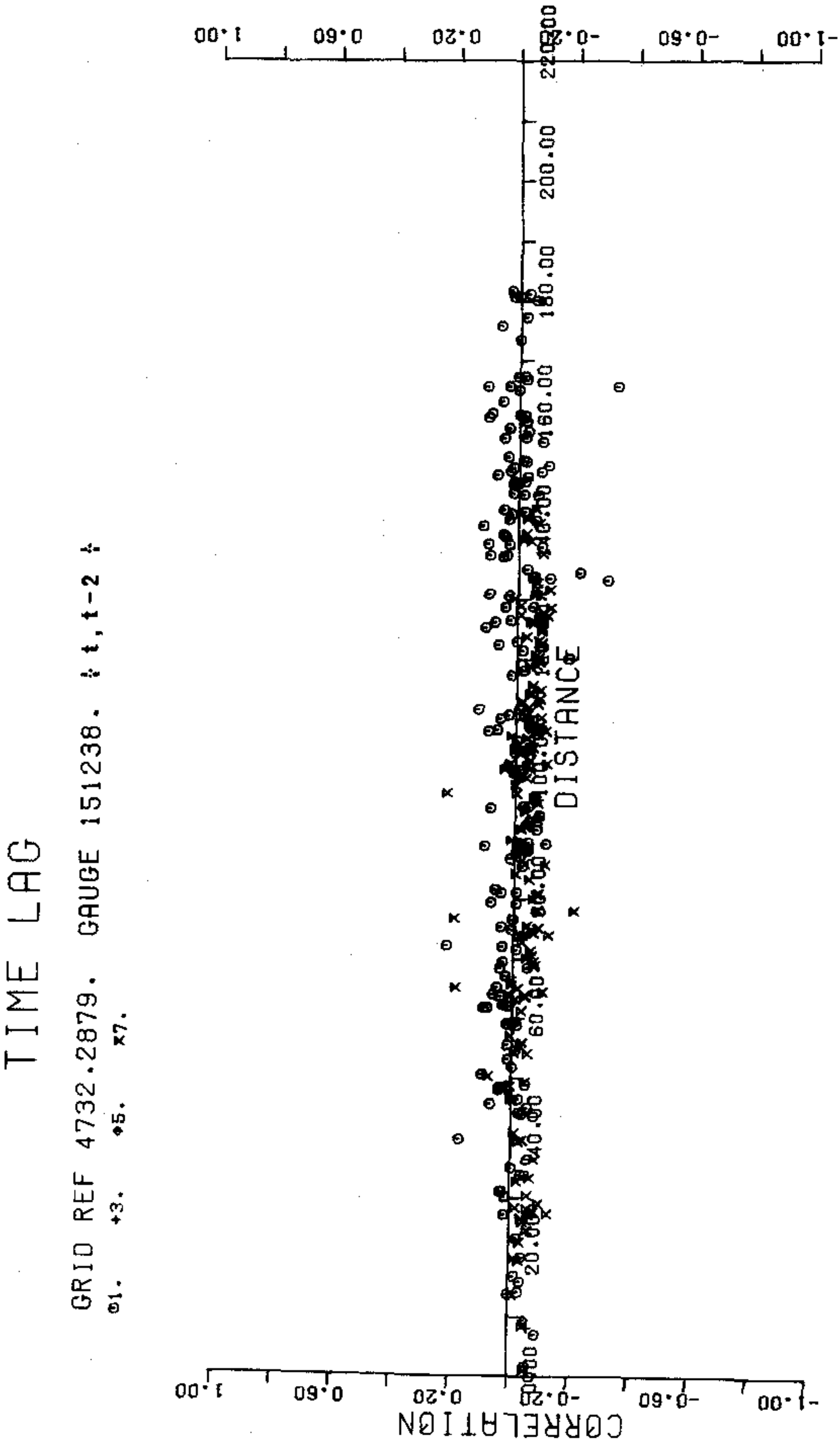


Figure 4.30a : Sample correlation functions. Eastern England: Correlation between rainfall (on day with rain over 5 mm) with rainfall two days before at other stations.

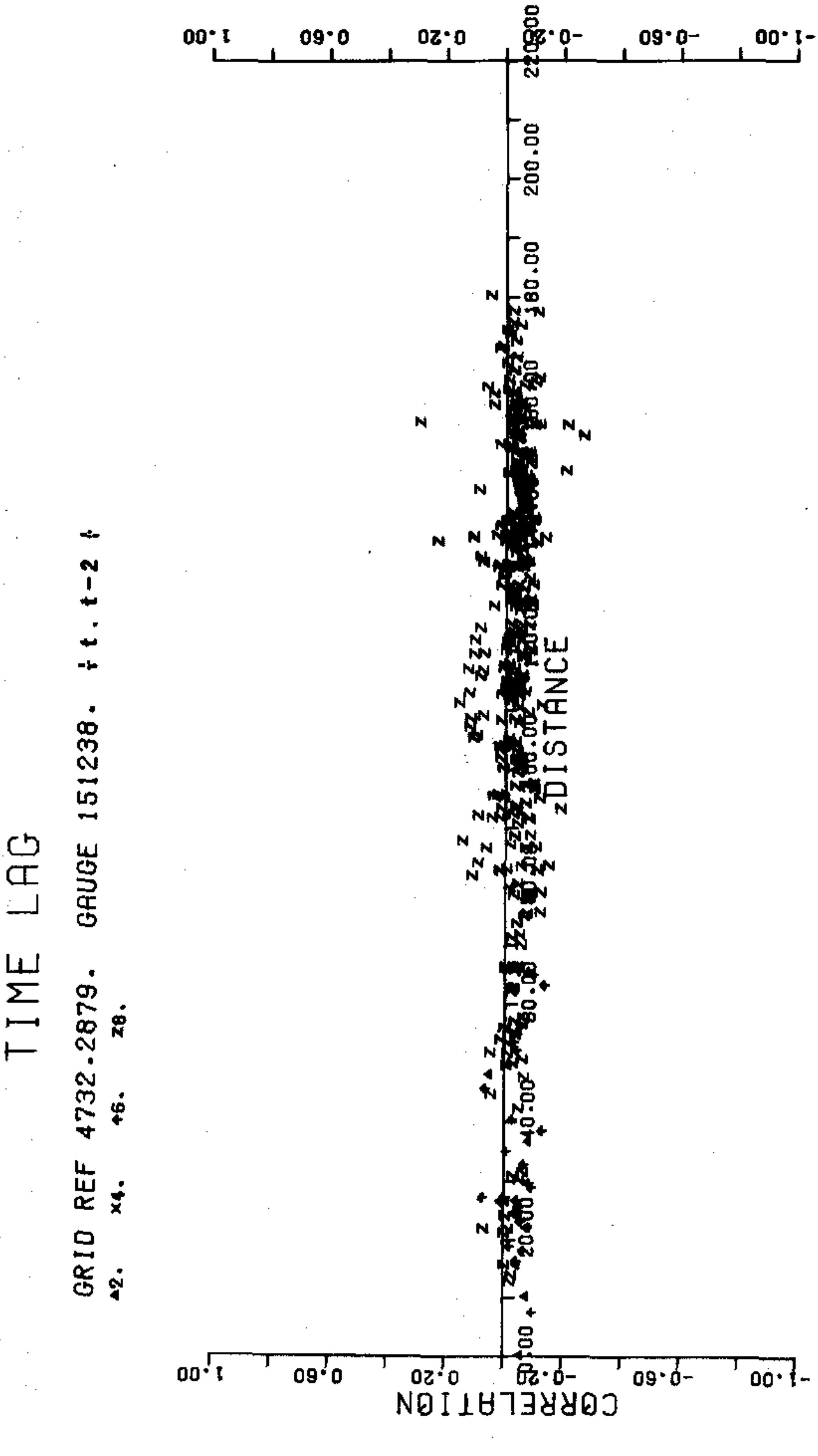


Figure 4.30b : Sample correlation functions. Eastern England: Correlation between rainfall (on day with rain over 5 mm) with rainfall two days before at other stations.

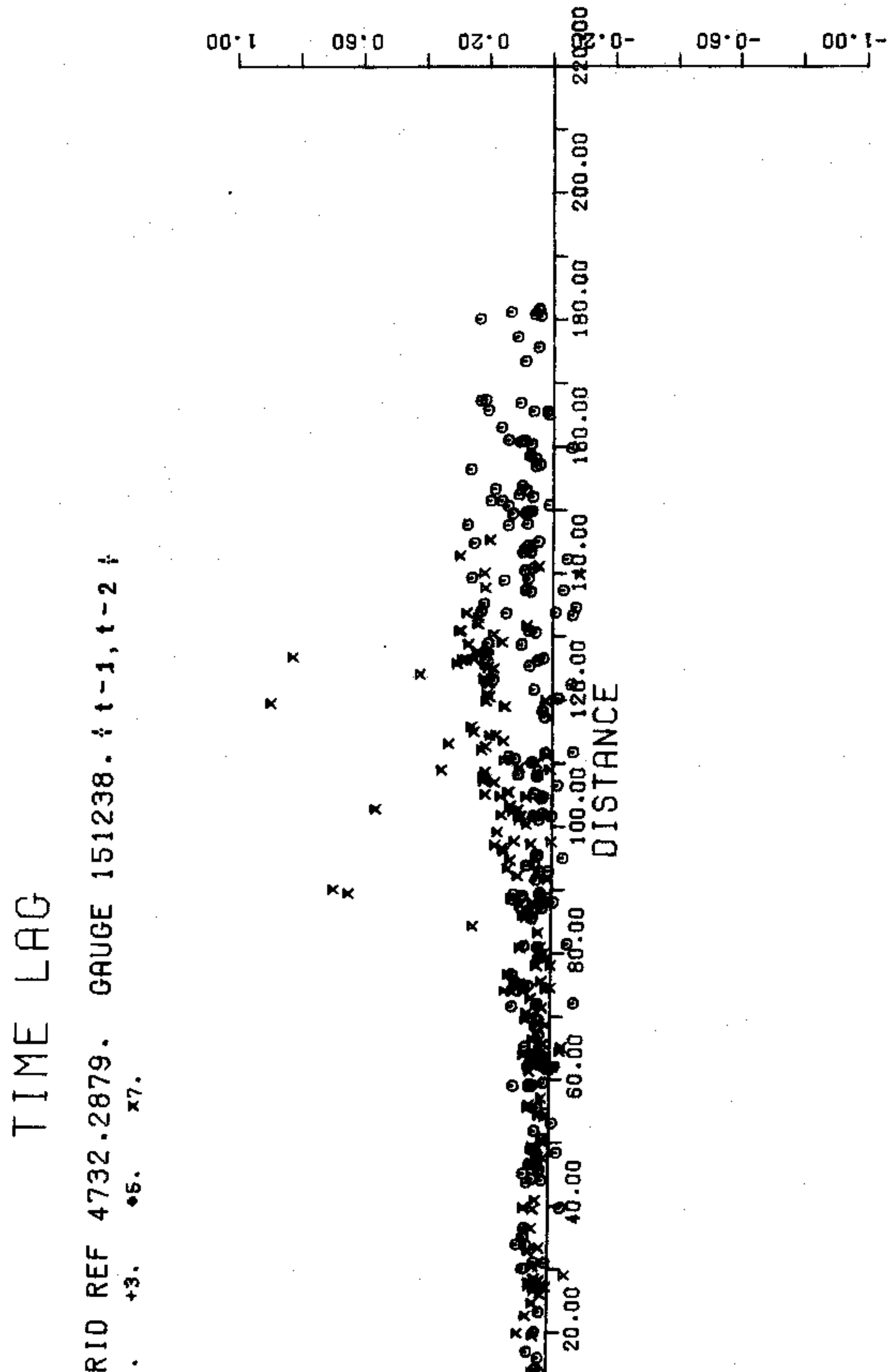


Figure 4.31a : Sample correlation functions. Eastern England: Correlation between rainfalls on the preceding day and two days preceding a rainfall of over 5 mm.

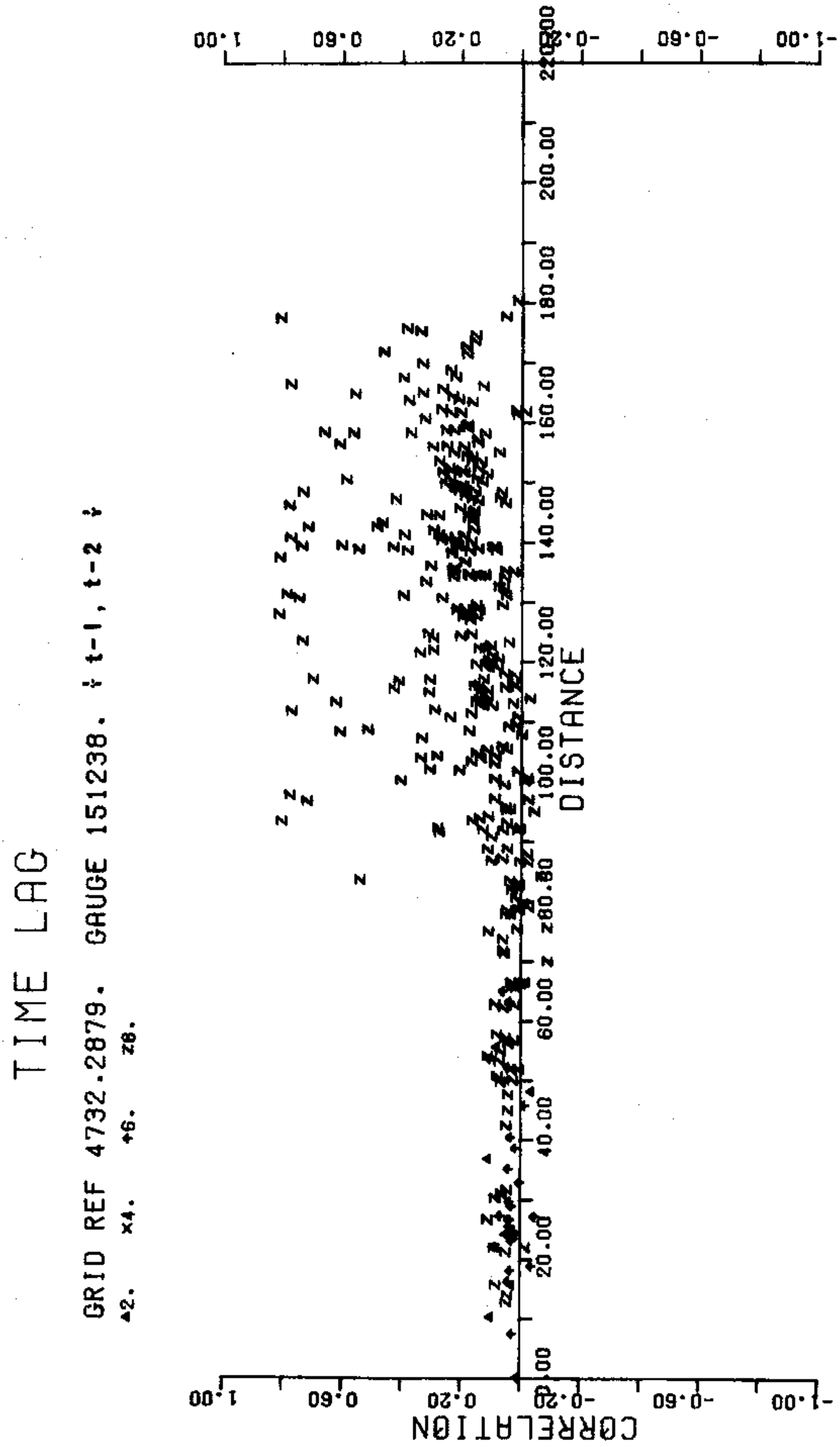
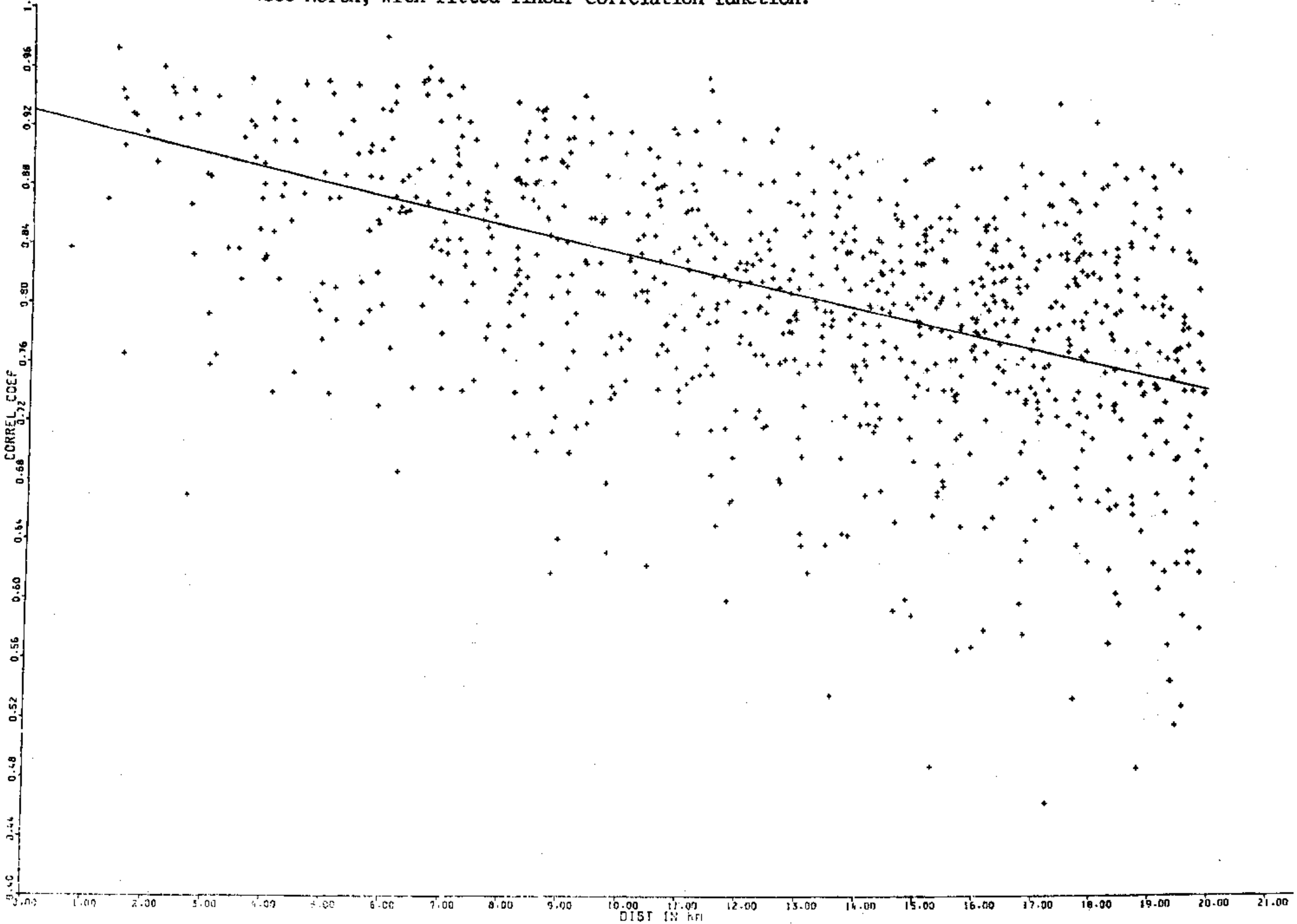


Figure 4.31b : Sample correlation functions. Eastern England: Correlation between rainfalls on the preceding day and two days preceding a rainfall of over 5 mm.

Fig. 4.32: Sample correlations between all pairs of stations in 25 km radius circle about 4139 East, 4306 North, with fitted linear correlation function.



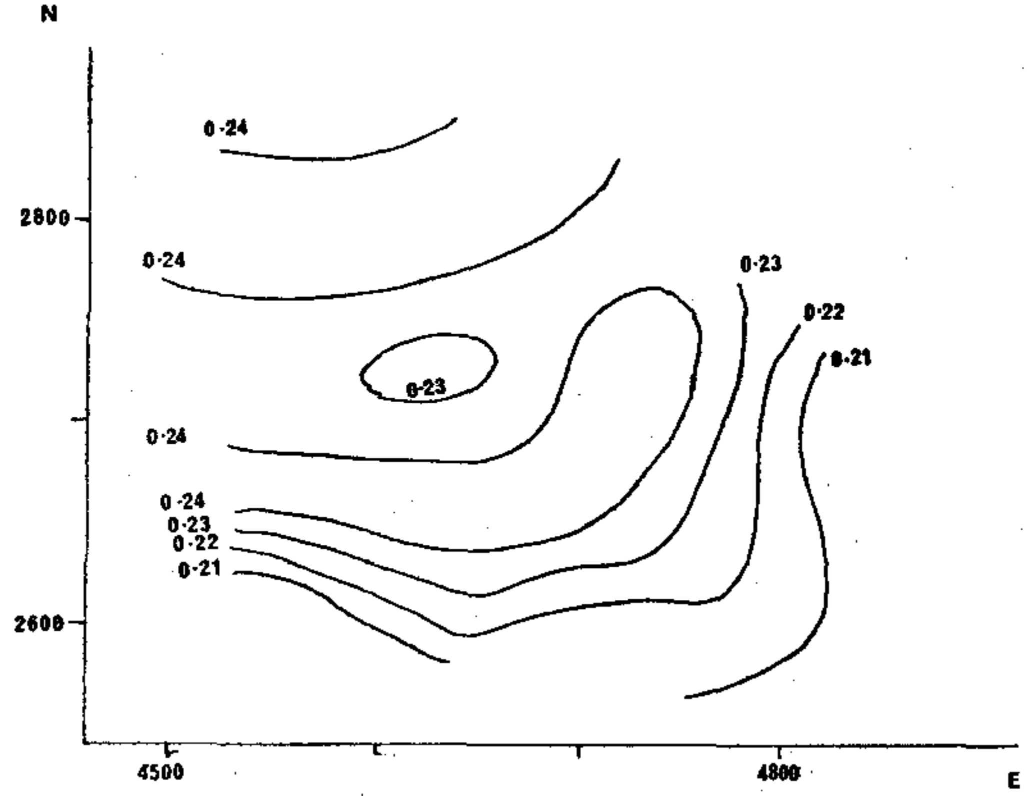


Fig. 4.33: Map of first eigenvector

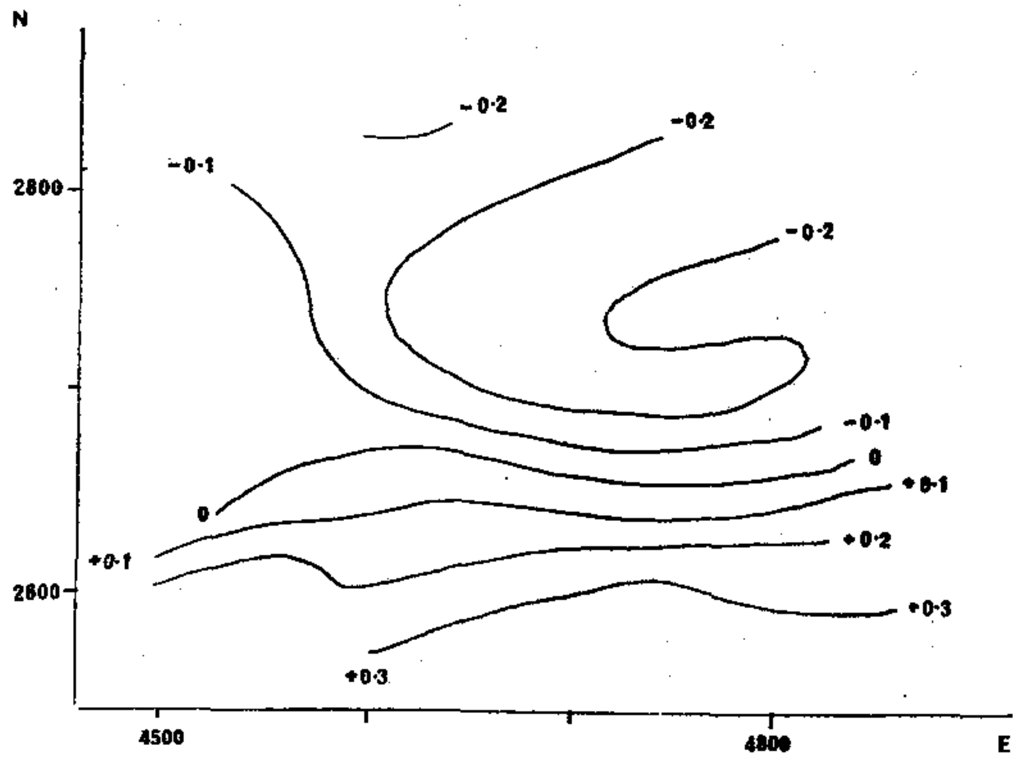


Fig. 4.34: Map of second eigenvector

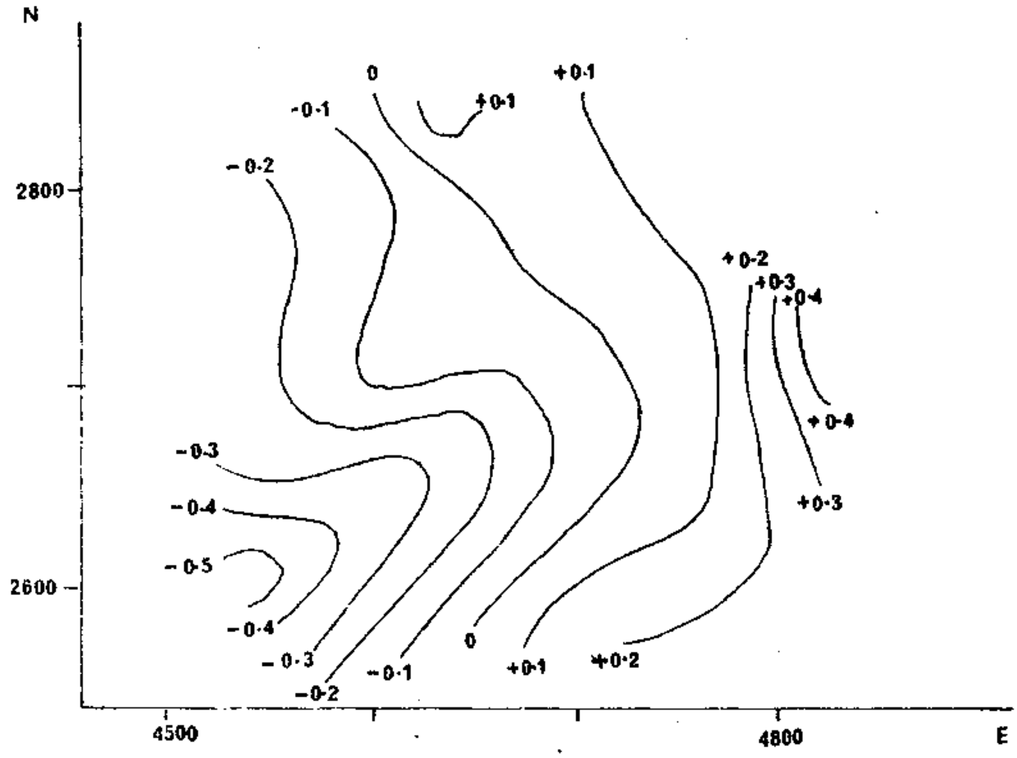


Fig. 4.35: Map of third eigenvector

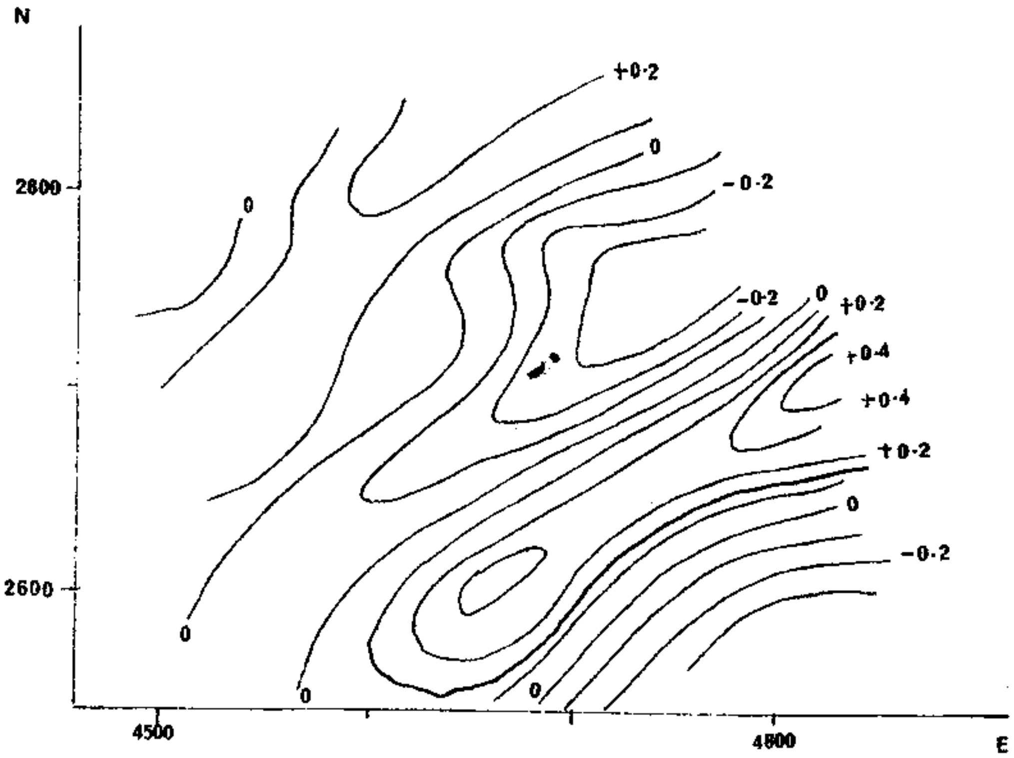


Fig. 4.36: Map of fourth eigenvector

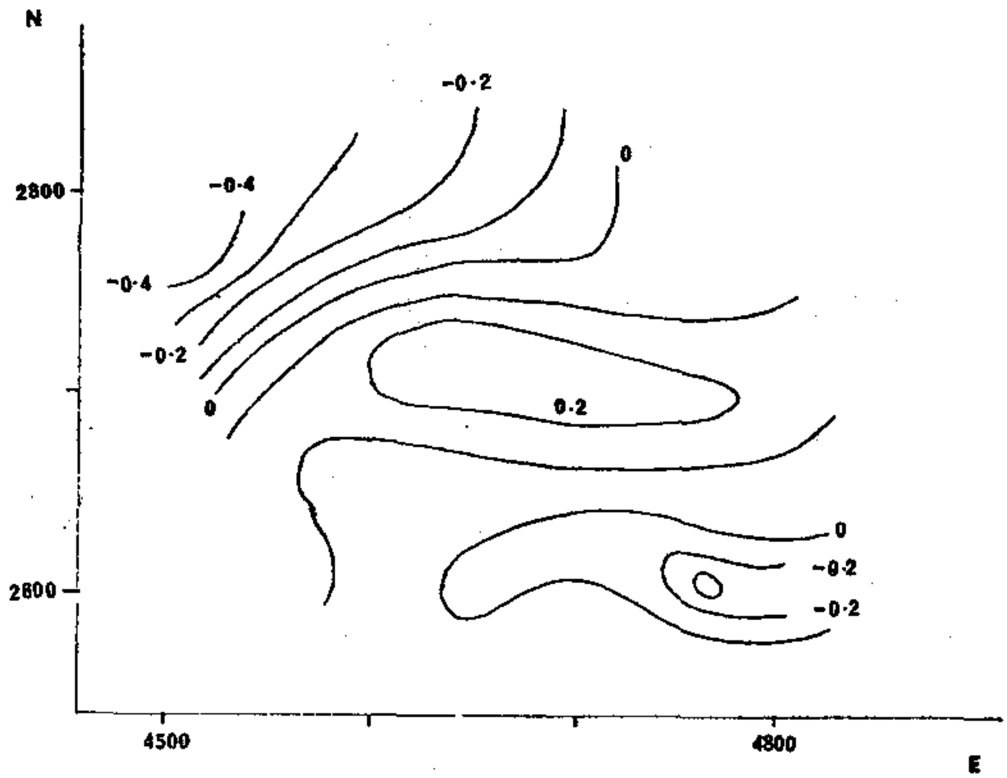


Fig. 4.37: Map of fifth eigenvector

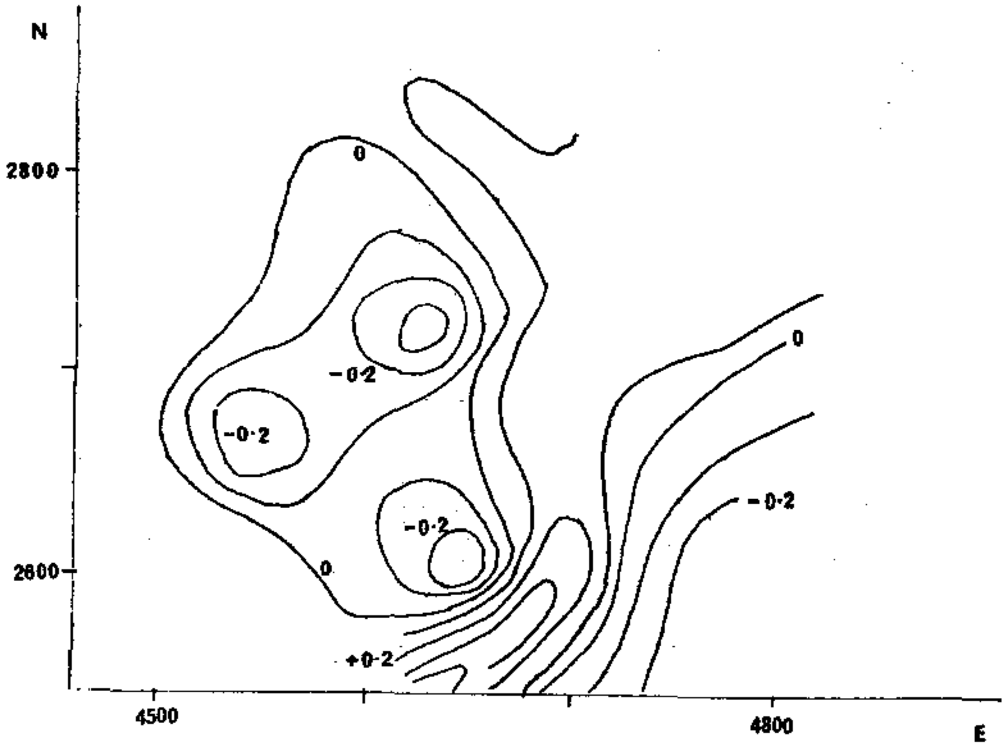


Fig. 4.38: Map of sixth eigenvector

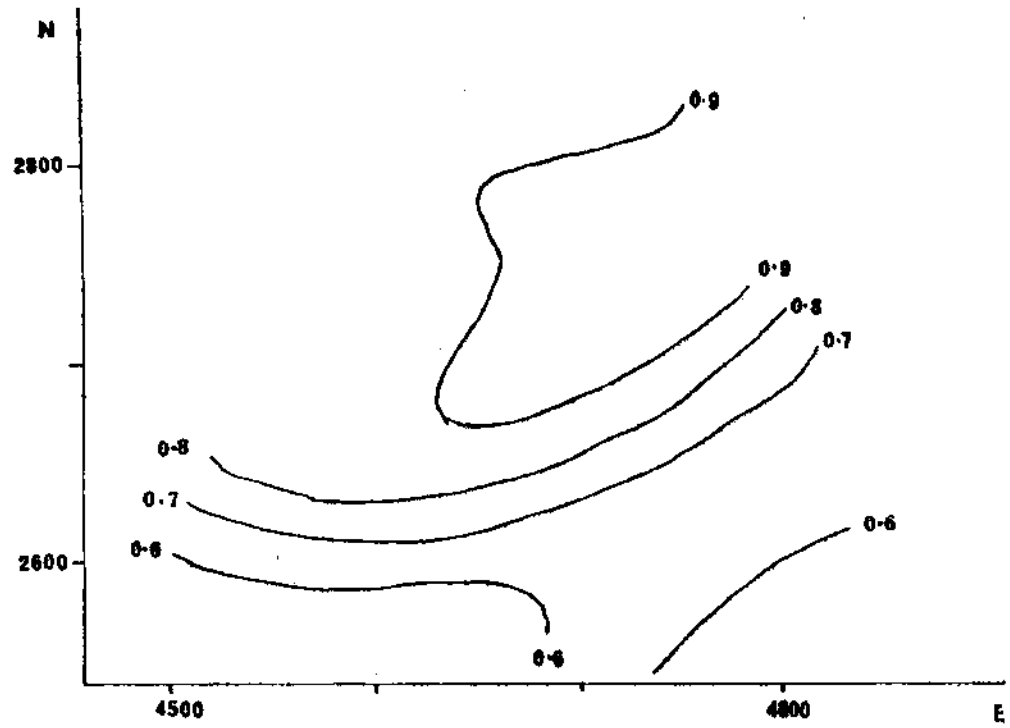


Fig. 4.39: Correlation map compiled using original data.

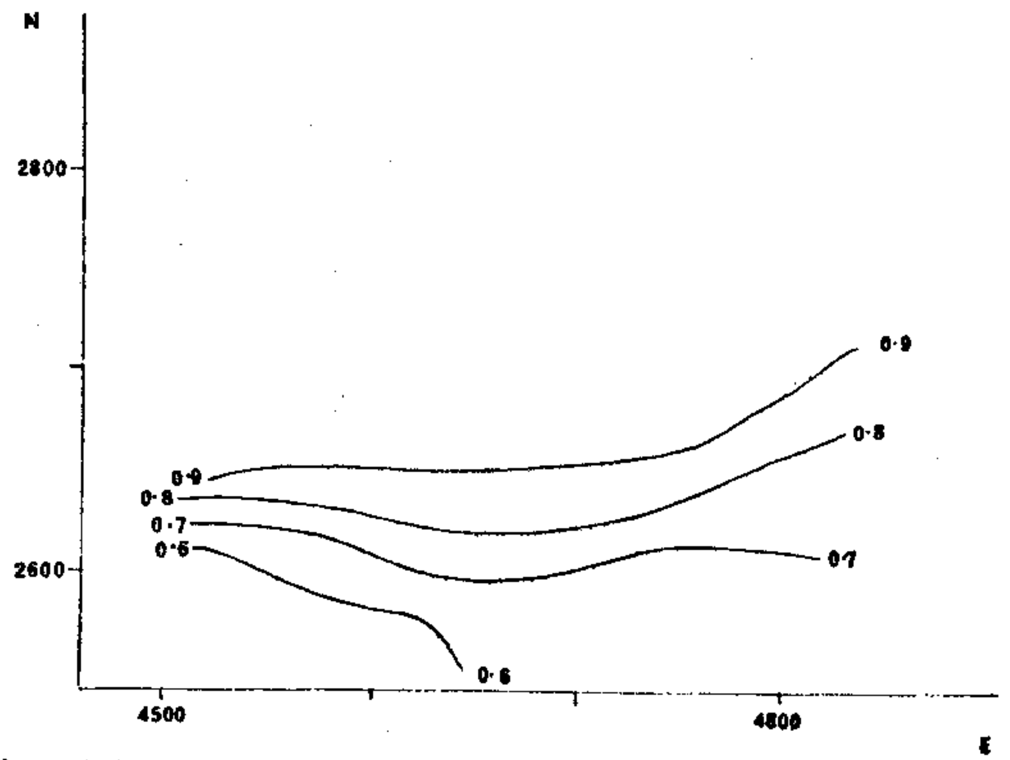


Fig. 4.40: Correlation map from reconstructed data using first three eigenvectors.

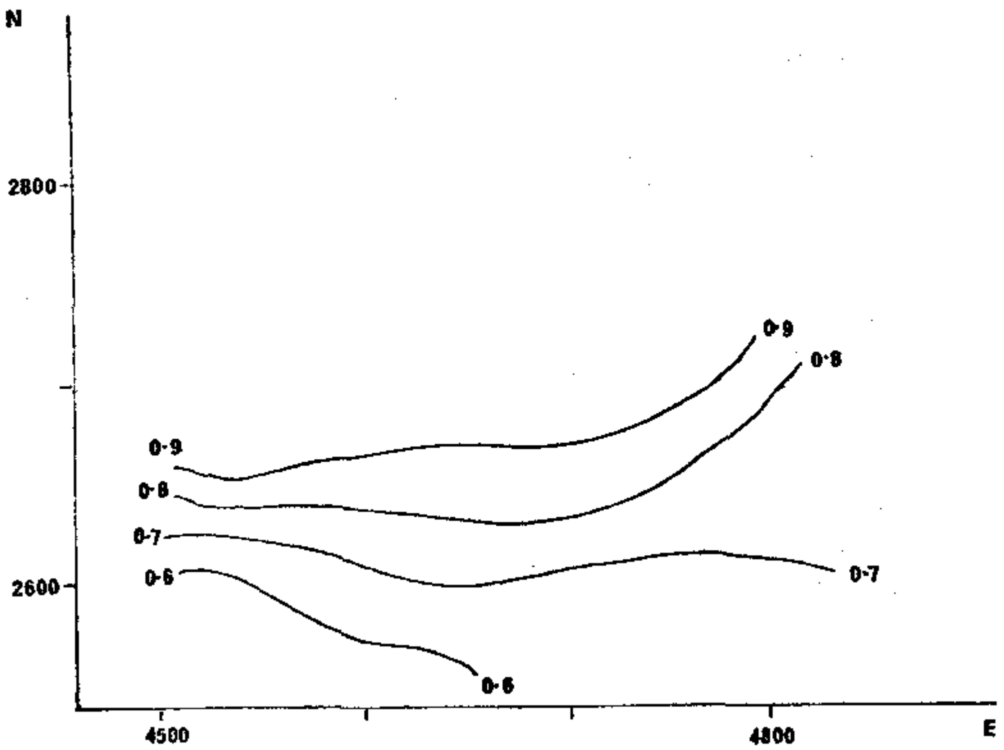


Fig. 4.41: Correlation map from reconstructed data using first four eigenvectors

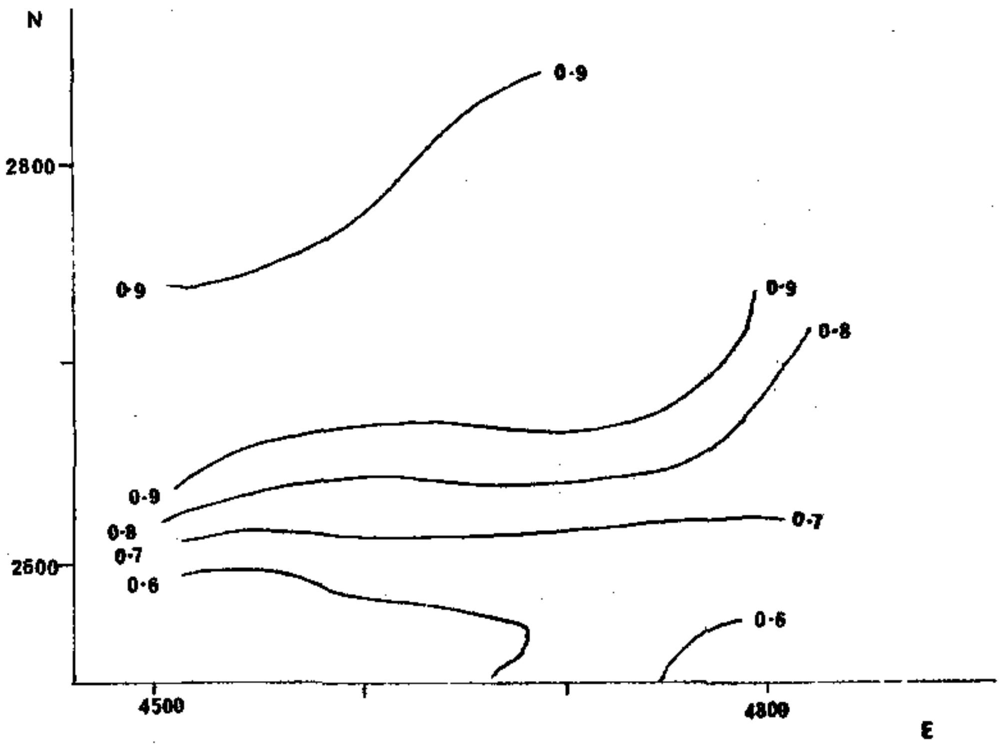


Fig. 4.42: Correlation map from recontructed data using first five eigenvectors.

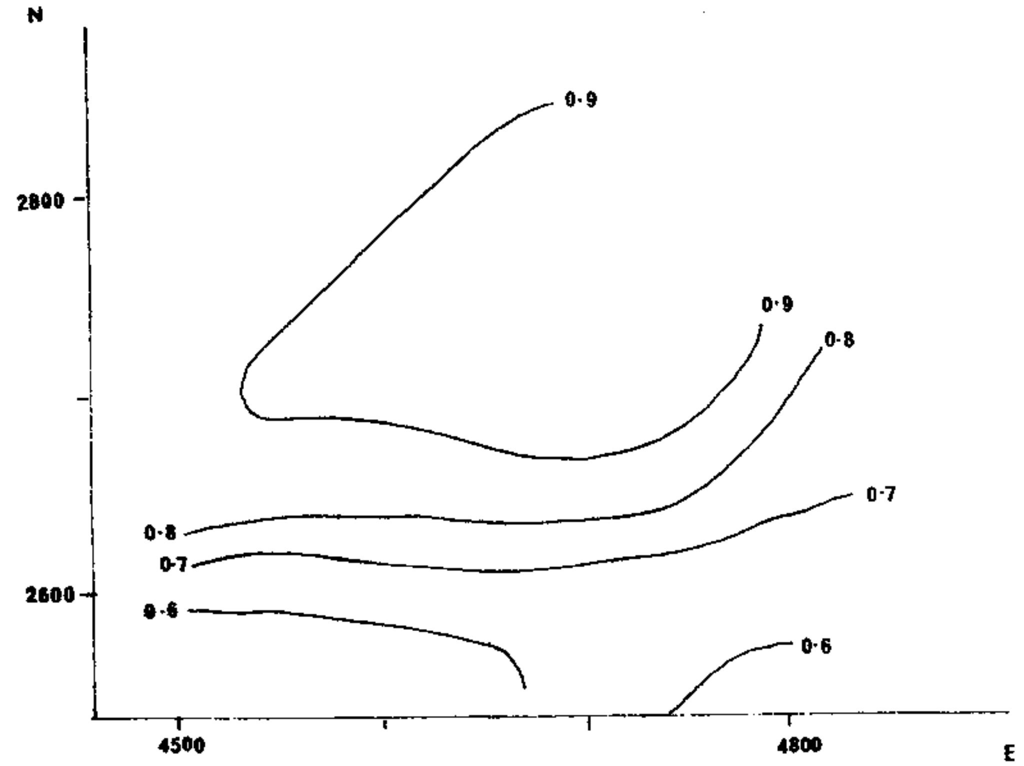


Fig. 4.43: Correlation map from reconstructed data using first six eigenvectors.

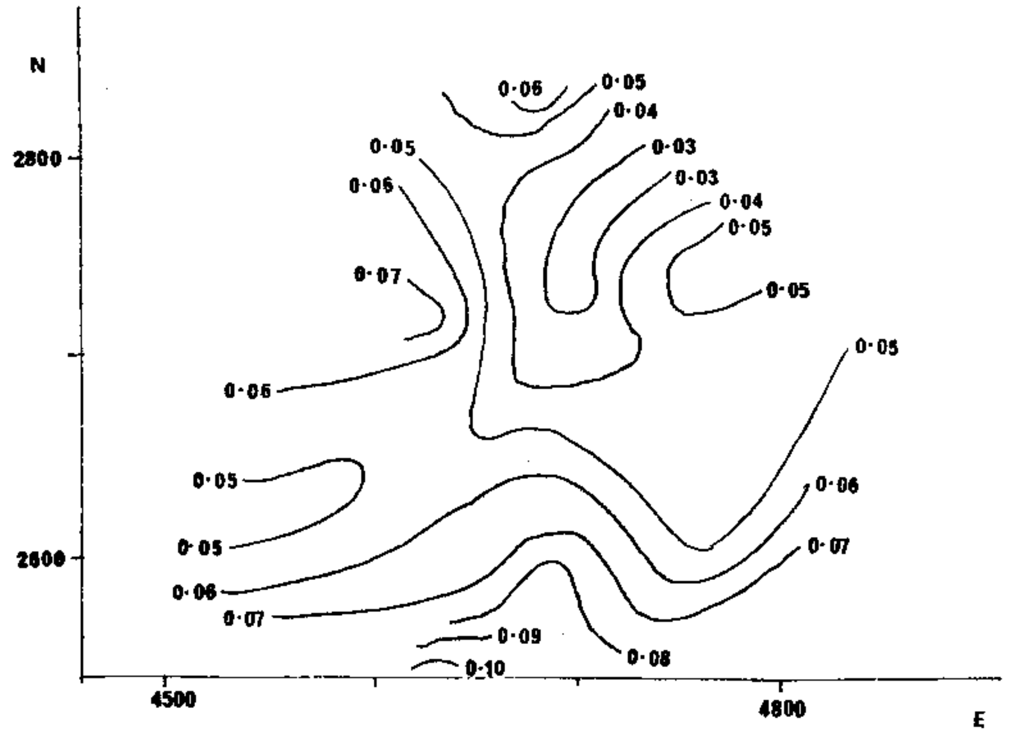


Fig. 4.44: The noise index at individual stations; constructed assuming that only the first six eigenvectors contain significant data.

5. METHODOLOGY OF RAINGAUGE NETWORK EVALUATION: DIRECT METHODS

5.1 Introduction

A review of user requirements (Section 3) for rainfall data in the UK has shown that the main requirements are for the provision of estimates of rainfall during a particular interval either at a point or in the immediate vicinity of a point or else over a moderately large area (0.5 to 10^4 km^2, say). For these uses, measures of the adequacy of a particular network can be obtained fairly directly; these are considered in Section 5.2 for point interpolation and Section 5.3 for areal averaging.

5.2 Point Interpolation

5.2.1 General Case

Measurements of the rain falling during particular intervals are available at a large number of points distributed over the country (approximately 1 every 35 km^2 for daily totals). When a value for the rainfall at a particular point which is not coincident with, nor sufficiently near, a gauged point is needed, an estimate of this rainfall may be constructed from the measurements taken at the existing neighbouring stations. The accuracy of this procedure obviously depends on the estimator that is used.

Clearly such an interpolation may need to be performed in a wide variety of circumstances - either when high powered computers are on hand or when such sophistication is unavailable or unwarranted. To accommodate the latter situation, attention is concentrated on simple estimates: as a first step only linear estimates are considered. If measurements $(X_1, \ldots . X_p)$ are available at P gauges and an estimate of Y, the rainfall at a particular point over the same time interval, is required then the linear estimates Y of Y are of the form

$$\hat{Y} = a + b_1X_1 + b_2X_2 + \ldots . . b_pX_p \qquad \ldots . \qquad (5.1)$$

where a, b_1, ...b_p have known values. The accuracy of such an estimator may be quantified by its mean square error (mse)

$$mse(\hat{Y}) = E\{(\hat{Y}-Y)^2\}$$

which measures how close the estimator is to the required value. A related measure is the bias

$$bias\ (\hat{Y}) = E\{(\hat{Y}-Y)\}$$

and, although this is not necessary, it is often required that the estimator be unbiased:

$$bias\ (\hat{Y}) = E\{\hat{Y}-Y\} = 0.$$

The estimators considered here will be unbiased.

Suppose that the long-term average rainfalls (for the interval being considered) are known to be $u_1, \ldots u_p$ at each of the gauged sites and u_Y at the ungauged site. Then estimators $\hat{Y}$ of the form

$$\hat{Y} = \mu_Y + b_1(X_1-\mu_1) + \ldots + b_p(X_p-\mu_p) \qquad \ldots . \qquad (5.2)$$

are unbiased for any constants b_1, ... b_p and it remains to choose these constants. The mean square error of the estimator (5.2) is

$$\text{mse}\ (\hat{Y}) = \text{var}\ (Y) - 2\{b_1\text{cov}(X_1,Y) + \dots b_p\text{cov}(X_p,Y)\} + \{b_1{}^2\text{var}(X_1) + 2b_1b_2\text{cov}(X_1,X_2) + \dots\dots b_p{}^2\text{var}(X_p)\}$$

In matrix notation this is

$$\text{mse}\ (\hat{Y}) = \sigma_{YY} - 2\underset{\sim}{b}^T\underset{\sim}{\sigma}_{XY} + \underset{\sim}{b}^T\underset{\sim}{\sigma}_{XX}\underset{\sim}{b} \quad \dots \quad (5.3)$$

where σ_{YY} = var Y, $\underset{\sim}{\sigma}_{XY} = (\sigma_{X_1Y}, \sigma_{X_2Y} \dots \sigma_{X_pY})^T$,

$\underset{\sim}{\sigma}_{XX} = \{\sigma_{X_iX_j}\}$

and where $\sigma_{X_iY} = \text{cov}(X_i,Y)$, $\sigma_{X_iX_j} = \text{cov}(X_i,X_j)$.

For any particular choice of constants equation (5.3) gives the mean square error of the estimator under the assumptions made. The vector $\underset{\sim}{b}$ may be chosen to minimise mse $(\hat{Y})$ given by equation (5.3) and this leads to the coefficients b* of the optimal linear estimator

$$\hat{Y}^* = \mu_Y + b_1^*(X_1 - \mu_1) + \dots b_p^*(X_p - \mu_p), \quad \dots \quad (5.4)$$

defined by

$$\underset{\sim}{b}^* = \underset{\sim}{\sigma}_{XX}{}^{-1}\underset{\sim}{\sigma}_{XY} \quad \dots \quad (5.5)$$

The corresponding value of the minimum mean square error is

$$\text{mse}\ (\hat{Y}^*) = \sigma_{YY} - \underset{\sim}{\sigma}_{XY}{}^T\underset{\sim}{\sigma}_{XX}{}^{-1}\underset{\sim}{\sigma}_{XY} \quad \dots \quad (5.6)$$

It is clear that to calculate the coefficients $\underset{\sim}{b}^*$ the covariances $\text{cov}(X_i,Y)$ $\text{cov}(X_i,X_j)$ $(i,j = 1, \dots P)$ must be known. If these covariances are known only approximately (i.e. estimated from data) then (5.5) may still be used: the mean square error of such an estimator will then be larger than that given by (5.6) but hopefully not by a large amount. The assumption that values for the long-term means u_1, ...u_p and u_Y are known exactly does not hold in practice.

One procedure is then to assume that the long-term average rainfall at all places is constant - this may be a reasonable assumption over suitably homogeneous regions. In this case, denoting $\mu_1 = \dots = \mu_p = \mu_Y = \mu$, the estimator (5.2) becomes

$$\hat{Y} = \mu + b_1(X_1-\mu) + \dots + b_p(X_p-\mu)$$

$$= \Sigma b_iX_i + \mu(1 - \Sigma b_i) \quad \dots \quad (5.7)$$

If is known then the above considerations apply and (5.6) is the mean square error of the optimal estimator.

If there is no information about μ then the only estimators of the type (5.7) that can be calculated are of the form

$$\hat{Y} = \Sigma b_iX_i \quad \dots \quad (5.8)$$

and for which

$$b_1 + b_2 + \dots + b_p = 1. \qquad \dots \qquad (5.9)$$

Under this condition equation (5.3) for the mean square error still holds and the optimal estimator among this class can be found by minimising the mse subject to the constraint (5.9). This gives (by, for example, the method of Lagrange multipliers) the vector $\underset{\sim}{b}^{**}$ minimising the mean square error

$$\underset{\sim}{b}^{**} = \underset{\sim}{\sigma}_{XX}^{-1}(\underset{\sim}{\sigma}_{XY} + \theta\underset{\sim}{1}) \qquad \dots \qquad (5.10)$$

where $\underset{\sim}{1}$ denotes a vector of ones and

$$\theta = (1 - \underset{\sim}{1}^T \underset{\sim}{\sigma}_{XX}^{-1} \underset{\sim}{\sigma}_{XY})/(\underset{\sim}{1}^T \underset{\sim}{\sigma}_{XX}^{-1} \underset{\sim}{1}).$$

The value of the mean square error of the estimator

$$\hat{Y}^{**} = \underset{\sim}{b}^{**T}\underset{\sim}{X} \qquad \dots \qquad (5.11)$$

obtained using these coefficients is

$$\text{mse}(\hat{Y}^{**}) = \sigma_{YY} - \underset{\sim}{\sigma}_{XY} \underset{\sim}{\sigma}_{XY}^{-1} \underset{\sim}{\sigma}_{XY} + (1 - \underset{\sim}{1}^T \underset{\sim}{\sigma}_{XX}^{-1} \underset{\sim}{\sigma}_{XY})^2/(\underset{\sim}{1}^T \underset{\sim}{\sigma}_{XX}^{-1} \underset{\sim}{1}) \qquad \dots \qquad (5.12)$$

and clearly this is always larger than (5.6). Once again the covariances between the sites need to be known in order to calculate $\underset{\sim}{b}^{**}$ by equation (5.10); however, as before, reasonable estimators could be obtained by using estimated values of the covariances. It may be noted that using an estimator $\hat{Y} = \Sigma b_i X_i$ with $\Sigma b_i = 1$ has the attractive property that if the rainfalls at each of the gauged points are equal then this same value is produced as the interpolated value: the more general estimator (5.7) would produce a value shifted towards the long-term mean.

In the numerical work presented later two assumptions are made. These are that the variances of the measured rainfalls at all the stations are the same and that, effectively, the rainfall is measured without error. With these assumptions the equations (5.3–5.5) can be rewritten in terms of the correlation matrix $\underset{\sim}{R}_{XX}$, the correlation vector $\underset{\sim}{r}_{XY}$ and the common variance σ^2. Thus the mean square error of the estimator (5.2) is given by

$$\text{mse}(\hat{Y}) = \sigma^2\{1 - 2\underset{\sim}{b}^T\underset{\sim}{r}_{XY} + \underset{\sim}{b}^T\underset{\sim}{R}_{XX}\underset{\sim}{b}\} \qquad \dots \qquad (5.13)$$

and the vector $\underset{\sim}{b}^*$ minimising this is

$$\underset{\sim}{b}^* = \underset{\sim}{R}_{XX}^{-1} \underset{\sim}{r}_{XY}, \qquad \dots \qquad (5.14)$$

the minimum value being

$$\text{mse}(\hat{Y}^*) = \sigma^2\{1 - \underset{\sim}{r}_{XY}^T \underset{\sim}{R}_{XX}^{-1} \underset{\sim}{r}_{XY}\}. \qquad \dots \qquad (5.15)$$

Similar formulae can be written for optimal estimators under the restriction $\Sigma b_i = 1$. The correlations $\underset{\sim}{R}_{XX}$, $\underset{\sim}{r}_{XY}$ are taken to be given by the spatially stationary correlation functions fitted in Section 4 to individual central stations.

To summarise three interpolators may be distinguished. Firstly, that of equation (5.2),

$$\hat{Y} = \mu_Y + \underset{\sim}{b}^T (\underset{\sim}{X} - \underset{\sim}{\mu})$$

may be used if the long term means are known. Here $\underset{\sim}{b}$ is an arbitrary vector of coefficients. The second interpolator, equation (5.4),

$$\hat{Y}^* = \mu_Y + \underset{\sim}{b}^{*T}(\underset{\sim}{X} - \underset{\sim}{\mu})$$

chooses $\underset{\sim}{b} = b^*$ to minimise the interpolation mse. The third interpolator, equation (5.11),

$$\hat{Y}^{**} = \underset{\sim}{b}^{**T}\underset{\sim}{X}$$

is applicable when the long-term means at points in the area are equal but unknown and the coefficients $\underset{\sim}{b}^{**}$ are chosen to minimise the mean square error subject to the condition $\underset{\sim}{b}^{**T}\, \underset{\sim}{1} = 1$ which ensures that the interpolator is unbiased. A further possibility would be to assume that estimates of known variance of the means $\underset{\sim}{\mu}$, μ_Y are available: then the mean square error of an estimator of an analogous form to (5.2) can be calculated and a minimisation of this would lead to an estimator optimal in this situation.

5.2.2 Effect of measurement errors

If measurement errors exist then formulae (5.1)-(5.15) still apply with the small change that the quantity to be estimated, Y, is the true rainfall at the point rather than a quantity measured by a raingauge: the quantities X_i are the recorded measurements at particular sites. If the true rainfalls at these sites are $Y_i (i = 1,\dots P)$ then a reasonable assumption is that

$$X_i = Y_i + \varepsilon_i$$

where ε_i $(i = 1, \dots P)$ are uncorrelated amongst themselves and uncorrelated with the true rainfalls Y, $Y_j (j = 1, \dots P)$. Then, in the notation of expression (5.3), the covariances required in formulae (5.3-5.12) are given by

$$\sigma_{X_i Y} = \sigma_{Y_i Y},$$

$$\sigma_{X_i X_j} = \sigma_{Y_i Y_j} \quad (i \neq j),$$

$$\sigma_{X_i X_i} = \sigma_{Y_i Y_i} + \sigma_{\varepsilon_i \varepsilon_i} .$$

If the variances of rainfall and of measurement errors are constants (σ_{YY}, $\sigma_{\varepsilon\varepsilon}$ respectively) and if

$$\underset{\sim}{R}_{YY} = \{\mathrm{corr}(Y_i, Y_j)\}$$

and

$$\underset{\sim}{r}_Y = \{\mathrm{corr}(Y_1, Y), \dots, \mathrm{corr}(Y_P, Y)\}^T$$

are arrays of correlations of the true rainfall at pairs of points, then the formulae corresponding to equations (5.13)-(5.15) are, with $\eta = \sigma_{\varepsilon\varepsilon}/\sigma_{YY}$,

$$\mathrm{mse}\ (\hat{Y}) = \sigma_{YY}\{1 - 2\underset{\sim}{b}^T r_Y + \underset{\sim}{b}^T(\underset{\sim}{R}_{YY} + \eta \underset{\sim}{I})\underset{\sim}{b}\} \quad \dots \quad (5.16)$$

for the mean square error of an estimator $\hat{Y}$ of the true rainfall and

$$\underset{\sim}{b}^* = (\underset{\sim}{R}_{YY} + \eta \underset{\sim}{I})^{-1}\, \underset{\sim}{r}_Y \quad \dots \quad (5.17)$$

for the coefficients of the optimal linear estimator $\hat{Y}^* = \mu_Y + \underset{\sim}{b}^T(\underset{\sim}{X} - \underset{\sim}{\mu})$ which has mean square error

$$\text{mse}(\hat{Y}^*) = \sigma_{YY}\{1 - \underset{\sim}{r}_Y^T \, (\underset{\sim}{R}_{YY} + \eta \underset{\sim}{I})^{-1} \, \underset{\sim}{r}_Y\}. \quad \ldots. \quad (5.18)$$

Under these circumstances mse $(\hat{Y}^*) \geq \frac{1}{p}\sigma_{\varepsilon\varepsilon}$. This means that if a fixed number of gauges are to be used then there is a limit to the accuracy with which rainfall can be interpolated no matter how close the gauges are sited - essentially because, even at a gauged site, the rainfall is only known to within the measurement error. If measurement errors do exist it is possible to derive a better estimate of the rainfall at a gauged point than the raingauge measurement at that site.

This model incorporating "measurement errors" is covered by the theory given in Section 5.2.1, the only distinction being that the estimator Y is constructed to estimate the true rainfall at a point rather than the value that would be obtained by measuring with a raingauge. Equations (5.16)-(5.18) are equivalent to equations (37)-(39) given by Gandin (1970). The optimal coefficients $\underset{\sim}{b}^{**}$ under the restriction $\underset{\sim}{b}^{**T}\underset{\sim}{1} = 1$ may also be obtained to cover the case when the mean rainfalls over the area are assumed constant but unknown.

5.2.3 Some simple cases

Equation (5.3) gives the accuracy of any particular linear estimator of the form (5.2) while (5.6) and (5.12) give the minimum mse that can be achieved with estimators based on particular assumptions about the means. If optimal estimators are used the accuracy of interpolation to a fixed point cannot be decreased by including extra gauged points in the set of points on which the interpolation is to be based. This is because the best estimator $\hat{Y} = b_1X_1 + \ldots + b_pX_p + b_{p+1}X_{p+1}$ must do at least as well as the best estimator $\hat{Y} = b_1X_1 + \ldots + b_pX_p$. Within limits the accuracy of interpolation increases as points further away are included in the interpolation scheme. Therefore the "accuracy of a network" as defined by the mean square error of interpolation depends on

(i) the number of stations used in the interpolation formula,

(ii) the form of the interpolation estimator - possibly suboptimal,

(iii) the position of the point to which interpolation is to be made, relative to the gauged sites.

Clearly (i) may vary considerably depending on the availability of computer resources: in the following, interpolation formulae based on measurements at three stations are considered since such interpolations are fairly simple to carry out. If only three stations are to be used there is a simple and intuitively reasonable interpolation procedure which fits a plane through the values observed at the three points and then uses the value given by the plane as the estimated value at intermediate points. This is a linear interpolator and is given by

$$\hat{Y} = b_1X_1 + b_2X_2 + b_3X_3$$

where

$$b_1 = \{(v_2-q_2^*)\ (q_1^*-w_1) + (w_2-q_2^*)\ (v_1-q_1^*)\}c \quad ,$$

$$b_2 = \{(w_2-q_2^*)\ (q_1^*-u_1) + (u_2-q_2^*)\ (w_1-q_1^*)\}c \quad ,$$

$$b_3 = \{(u_2-q_2^*)\ (q_1^*-v_1) + (v_2-q_2^*)\ (u_1-q_1^*)\}c \quad ,$$

$$c = \{v_2(u_1-w_1) + w_2(v_1-u_1) + u_2(w_1-v_1)\}^{-1} \quad ,$$

$$b_1 + b_2 + b_3 = 1, \qquad \ldots . \qquad (5.19)$$

and where (u_1,u_2), (v_1,v_2), (w_1,w_2) are the cartesian coordinates of the gauged points and (q_1^*,q_2^*) are the coordinates of the interpolated point. This will be called the simple linear interpolator. Beside this simple interpolator, optimal interpolators are also considered in the two situations where long term average means are assumed known and where those are unknown but assumed equal. Contours of interpolation accuracy for these three estimators are given in Figures 5.1 - 5.3. Here it is assumed that the means at the three stations are equal and that the variance of the measurements is the same at all sites. The function contoured is the fraction, f, of the corresponding standard deviation that remains as the root-mean-square-error of interpolation error

$$f = \{\text{mse}(\hat{Y})/\sigma_{YY}\}^{\frac{1}{2}} \qquad \ldots . \qquad (5.20)$$

Thus on the first diagram a point on the 0.05 contour has an error of interpolation of 0.05 times the standard deviation of the rainfall at an individual point. The assumption made about the correlations between points in this area is that the correlations are a function of distance only, namely

$$\rho(d) = 0.7 + 0.3\ \exp\{-0.012d\}.$$

This function, which is used solely to compare properties of the three estimators, is very similar to functions fitted in Eastern England to correlations of monthly totals at different sites. The coordinates in the figures are in kilometres. Examination of the diagrams shows that, inside the triangle formed by the three measurement sites, there is very little difference in the performance of the three estimators but that the simple linear estimator becomes markedly worse as the side of the triangle is crossed moving away from the centre. The two optimal estimators are always very close in performance, there being no visible difference in the diagrams until well outside the triangle; the estimator assuming the means known is always slightly better than that assuming the means equal but unknown. Other correlation functions and other choices of the positions of the 'gauged stations' lead to similar conclusions.

It is of interest to consider the interaction of non-circular correlation functions with the orientation of the triangle of stations. Figures 5.4 - 5.7 show the corresponding contour plots for four different orientations of a moderately extreme triangle with the correlation function taken to be that which was fitted to a station in Eastern England for monthly totals. Here the contours are of interpolation error using an optimal estimator for which the means are assumed equal but unknown. The maximum error of interpolation inside the triangle varies between 0.23 and 0.30 of the standard deviation of measurements at individual stations, which is about 27 mm for monthly total rainfalls (mean monthly total 50 mm). Thus interpolation from stations in this configuration would be with a worst root mean square error of 6 to 8 millimetres. However the length of the triangle is over 30 kms - somewhat large for typical areas of the UK. The worst orientation is when the length of the triangle lies in the direction of most rapidly decreasing correlation.

The accuracy with which an interpolation can be made depends very much on the distance of the interpolation point from the gauged points and, for a triangle of measured sites, the interpolation error is at its largest at the 'centre' of the triangle and on the sides of the triangle. Since, in a real situation, there is always the possibility of considering measurements made at other neighbouring sites when attempting to interpolate to a point near the edge of triangle, it is convenient to take as the measure of accuracy of a trio of stations the interpolation error at the 'centre' of the triangle which will here be taken as the centroid or 'centre of gravity' of the triangle. This procedure avoids making a search for the maximum or locally maximum interpolation error - often the maximum lies on a side of the triangle and there may be no local maximum near the 'centre'.

One way to study the effect of different network densities on interpolation error is to choose a particular shape of triangle - the triangle formed by the hypothetical gauged points - and to consider this triangle at different scales. Thus similar triangles of different sizes are considered and for each of these and for several correlation functions the interpolation error at the centre of the triangle is found. The results of this procedure are shown in Figures 5.9 - 5.22. Here two shapes of triangle have been chosen, an equilateral triangle and a right-angled triangle, and these are shown in Figure 5.8. The 'size' of the triangles is taken to mean the length of the E-W side of the triangles and the configuration of the right-angled triangle has been chosen so that the areas of two triangles of equal 'size' are also equal. The orientation of the triangles was as shown in Figure 5.8. Thus for example Figure 5.9 shows graphs of the accuracy of the simple linear interpolator (as a fraction of the point standard deviation) to the centre of an equilateral triangle against the size of the triangle. The four lines correspond to the four correlation functions fitted in Section 4 to the sample correlations for yearly totals (2 stations in Eastern England and 2 in Northern England).

The correlation functions fitted in Section 4 to the various categories of data have been used to derive the accuracies of interpolation (Figures 5.9 - 5.22) corresponding to these categories. In discussing the plots of correlation obtained in Section 4.4 it was noted that the rates of decay of correlation with distance varied with the category of data: this is reflected here in the graphs of accuracy obtained from the fitted correlation functions. Thus, for example, interpolation can be performed more accurately for yearly and monthly data than for data for arbitrary days and the type of data for which interpolation is least accurate is heavy rainfall. The graphs of accuracy obtained from the functions fitted to correlations of monthly rainfall for the years 1961-1974 and for 1875-1890 are very similar, again corresponding to the similarity of the fitted functions.

The graphs for interpolation in equilateral and right-angled triangles are also very similar for the same categories of data, the only exception being those for yearly totals (Figures 5.9 and 5.16). Here there is a large spread between the lines corresponding to different stations for the right-angled triangle but much less spread for the equilateral triangle. This could be a result of the anisotropic correlation functions fitted: these were further from circularity for the yearly data than for the other categories but were based on fitting to far fewer points.

5.2.4 Non-linear Estimators

The following remarks apply also the areal averaging problems described in Section 5.3. Throughout this section it is assumed that linear estimators of point rainfall (and of areal average rainfall) are used. If all functions of the measured rainfalls X_i at a set of sites $(i=1,\ldots P)$ are admitted as possible estimators of the required quantity, Y, then the best estimator, in the sense of minimising the mean square error, is the conditional expectation.

$$\hat{Y}(X_1 \ldots X_p) = E\{Y \quad X_1, X_2, \ldots X_p\} \qquad \ldots . \qquad (5.21)$$

This is a linear function of $X_1 \ldots X_p$ when the joint distribution of $(Y, X_1, \ldots X_p)$ is multivariate normal but for most other distributions $E(Y \mid X_1, \ldots X_p)$ is non-linear. The estimators considered in the rest of this section have been linear estimators and therefore, since rainfall distributions are known not to be normal, it is to be expected that better estimators can be found by allowing them to be non-linear in the observations. The advantage of linear estimators is that their mean square error depends only on the second-order moments (covariances) of the distributions and these can be estimated fairly simply. In order to calculate the best non-linear estimator, the joint distribution needs to be fully known: this would require a great deal of investigation. The accuracy of the best non-linear estimator is given by the conditional variance,

$$\text{mse}(\hat{Y}) = \text{var}(Y \mid X_1, \ldots X_p) \qquad \ldots . \qquad (5.22)$$

and will in general depend on the observed values at the gauged sites. Thus in some rainfall situations it may be possible to interpolate more accurately than in others. The measures of accuracy of the linear estimators apparently do not depend on the observed values $X_1, \ldots X_p$: this is because the mean square error is averaged over the distribution of observed values.

Linear estimators are used in this work because of their great simplicity and because of the extremely large amount of work necessary to investigate the distributional properties of rainfall. Nevertheless it is felt that, especially in the case of daily total rainfalls and totals over shorter periods, a great deal would be gained from the use of non-linear interpolators after detailed consideration of the joint distributions.

5.3 Areal Averages

Much of the discussion in section 5.2 for point interpolation holds also for the problem of areal averages. Once again only linear estimators $\hat{Y}$ of the areal average rainfall Y are considered

$$\hat{Y} = a + b_1 X_1 + \ldots + b_p X_p . \qquad \ldots . \qquad (5.23)$$

The mean square error of an unbiased estimate of this form is again given by (5.3) where now, for example, $\text{cov}(X_i, Y)$ is the covariance of the rainfall at site i with the average rainfall over the area under consideration in the same interval. If $\gamma(\underset{\sim}{u}, \underset{\sim}{v})$ represents the covariance between the rainfalls at two points with coordinates $\underset{\sim}{u} = (u_1, u_2)$, $\underset{\sim}{v} = (v_1, v_2)$ then

$$\text{cov}(X_i, Y) = \frac{1}{A} \int_A \gamma(\underset{\sim}{x}_i, \underset{\sim}{u}) d\underset{\sim}{u} \qquad \ldots . \qquad (5.24)$$

$$\text{var}(Y) = \frac{1}{A^2} \int_A \int_A \gamma(\underset{\sim}{u}, \underset{\sim}{v}) \, d\underset{\sim}{u} \cdot d\underset{\sim}{v} \qquad \ldots . \qquad (5.25)$$

where x_i is the coordinate vector of the i'th station and A is the physical area of the region under consideration and the integrals, denoted $\int_A$, are two dimensional integrals over the region. The mean square error of any linear estimate of areal average rainfall can then be calculated from the covariance between rainfalls at pairs of points. Here it is assumed that the raingauge measurements are made without error.

Once again, it is possible to construct optimal estimators of the areal average under the assumption that the long-term average rainfalls are known at each point or under the assumption that these have a constant but unknown value. In practice, for a P gauge estimator, either would require the evaluation of a 4-dimensional integral, P 2-dimensional integrals and the inversion of a P x P symmetric matrix in order to calculate the coefficients of the linear estimator. Further the accuracy of an estimator clearly depends on several factors, namely the size of the area and the number and positions of the gauges: it would be very difficult to investigate fully the effect of all these factors and therefore very simple forms for both the raingauge network and the estimator to be used have been assumed.

The gauges are assumed to be sited on a square grid with a square area allocated to each gauge. The area over which the average rainfall is required is also assumed to be square and composed of an integral number of these gauge-squares (Figure 5.23). The estimate used for the areal average rainfall is simply the average of the measurements at the gauges within the area: for this very special arrangement of gauges this estimate is also that which would be obtained by the method of Thiessen polygons. For this particular estimator

$$\hat{Y} = \frac{1}{P} \sum_{i=1}^{P} X_i \qquad \ldots\ldots \quad (5.26)$$

the mean square error is, assuming that the long-term average rainfall at different sites is constant over the area,

$$\text{mse } (\hat{Y}) = \frac{1}{P^2} \sum_{i=1}^{P} \sum_{j=1}^{P} \gamma(\underset{\sim}{X}_i,\underset{\sim}{X}_j) - \frac{2}{PA} \sum_{i=1}^{P} \int_A \gamma(\underset{\sim}{X}_i,\underset{\sim}{u})d\underset{\sim}{u}$$

$$+ \frac{1}{A^2} \int_A \int_A \gamma(\underset{\sim}{u},\underset{\sim}{v})d\underset{\sim}{u}\, d\underset{\sim}{v} \qquad \ldots\ldots \quad (5.27)$$

The assumption of a square grid and square area of interest greatly reduces the amount of numerical calculations needed to evaluate the integrals if the further assumption that the covariance function is circular is made (Bras & Rodriguez-Iturbe, 1976). Because, in some sense, all directions are taken into account in the expression (5.27) for the error of areal averaging, it was felt that there would be no great difference in the results of using circular and non-circular correlation functions and, because this allowed the calculation of the integrals, a circular correlation, $\rho(d) = a + (1-a) \exp(-bd)$, was used. The parameters of this function were fitted in the manner described in Section 4.

With the restrictions imposed above, the accuracy of the estimate of areal average rainfall depends on

(i) the size of the area under consideration

and

(ii) the distance between the rainguages.

Again the measure of accuracy is taken as the ratio of the root mean square error of the estimator to the standard deviation of the rianfall at a point.

$$f = mse(Y)\sigma_{XX}^{\frac{1}{2}} \qquad \text{.............(5.28)}$$

This measure of accuracy is plotted in Figures 5.24-5.27. Here it is assumed that a fixed square area is under cosideration and the accuracy is plotted for a number of different densities of rainguages: the points on this line are joined by straight lines for convenience only since the requirement that the square area is made up of intergral number of guage-squares is satisfied only for particular spacings of the grid. In theory it would be possible to evaluate the accuracy of a linear estimate of a real average rainfall for any area based on any particular configuration of rainguages.However the evaluation of the necessary intergrals impose a practical constraint on what can be undertaken. The simple situation considered here at least provides an indication of the effect of changing the density ofrainguages and of considering areas of different sizes. It can be seen from Figures 5.24-5.27, that as the area over which the average rainfall is required increases, the mean square error of the estimator decreases if the rainguage spacing remains fixed.

5.4 Summary

Three ways of interpolating rainfall to an unguaged point have been discussed.These interpolators are similar, being weighted linear combinations of observed values at guaged points. The difference between the interpolators lies in the way the coefficients in the linear combination are chosen: for two of the interpolators the coefficients are chosen to be the best possible according to the assumptions that can be made about the long-term properties of rainfall at places within the region. The coefficients of the third interpolator would usually be chosen intuitively to give high weight to nearby observations. Expressions for the accuracy of the interpolators have been given.

The different interpolators have been applied to the case of interpolation from the three guaged methods of interpolation have been produced. These show that, provided the point to which interpoation is made is within the traingle formed by the three guages, there is little difference in performance between the interpolators: theis is of course only true if the assupmtions under which the interpolators are constructed actually hold. Some brif consideration has also been given to the effect of the triangle of guages if the correlation function about a point is anisotropic. Taking as an example the simple linear interpolator, it has been shown how the accuracy of interpolation obtainable from atriangle of fixed shape is related to its size: graphs of this relationship have been presented.

In the discussion of the accuracy of estimates of areal average rainfall a particular regular arrangement of the gauges within the area under consideration was assumed for computational expediency. This allowed the accuracy of a simple estimator to be related to the density of gauges within the region and to the size of the area for which the average rainfall is required.

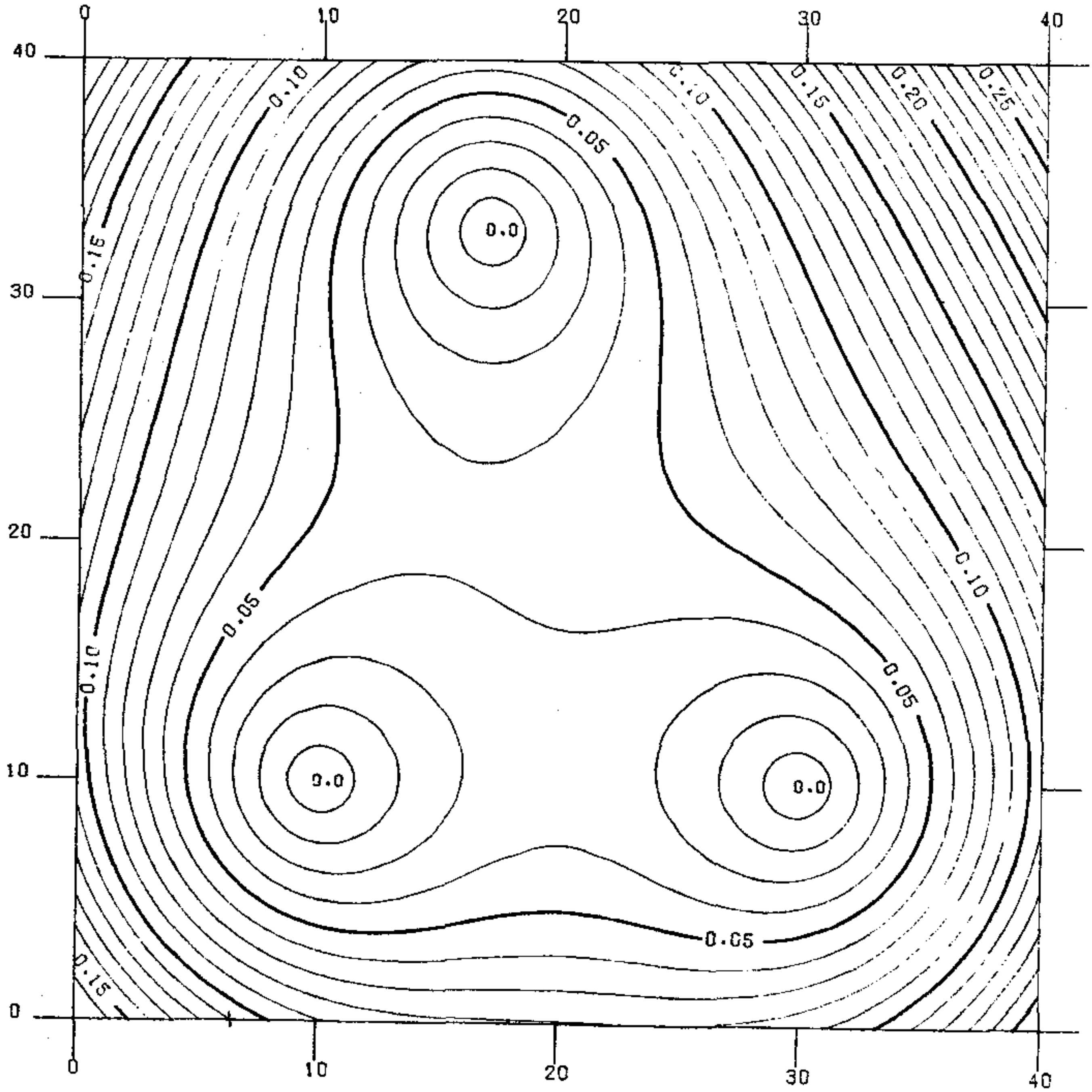


Figure 5.1 : Contours of interpolation error from three sites to neighbouring points using simple linear interpolation

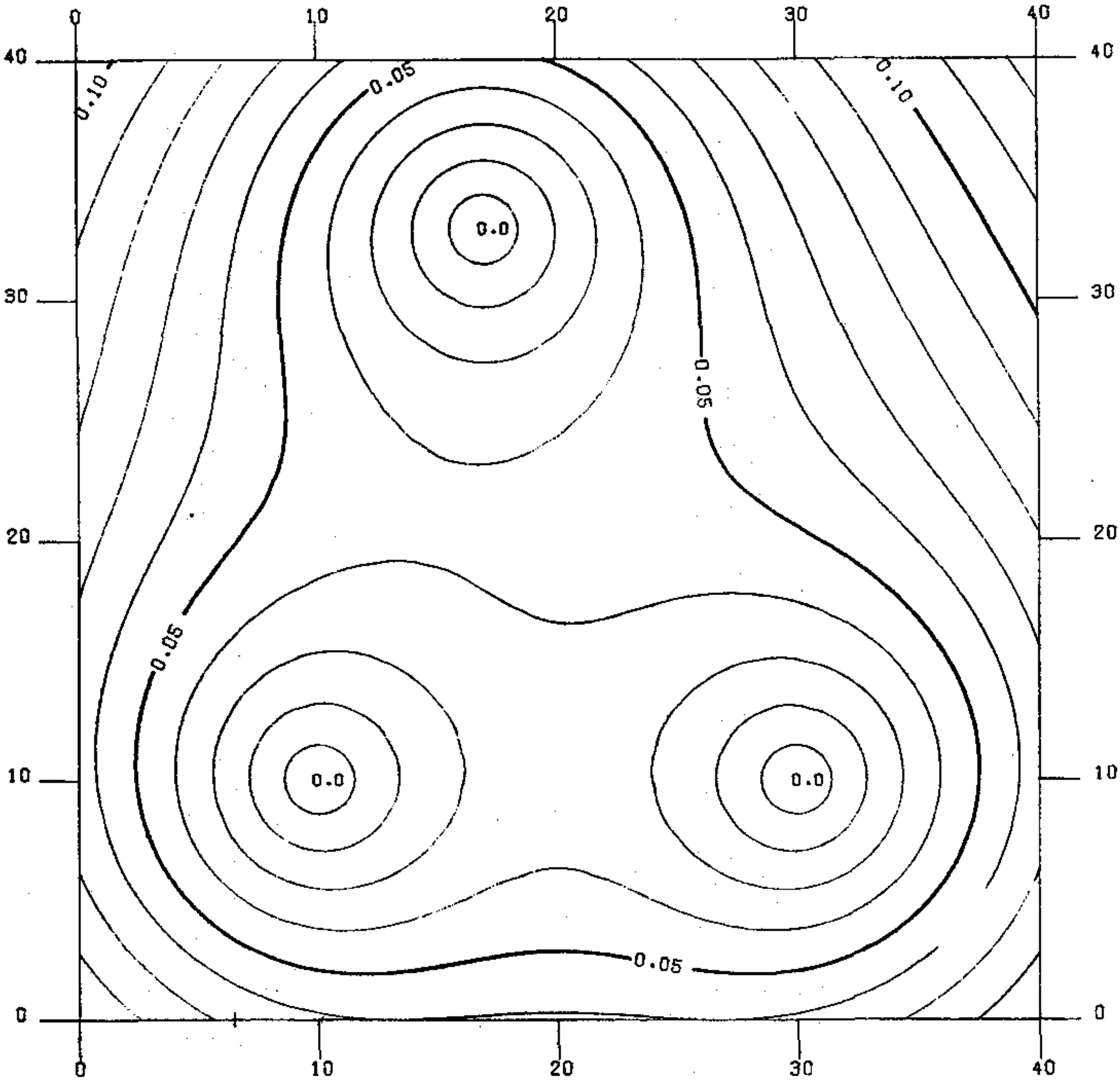


Figure 5.2 : Contours of interpolation error from three sites to neighbouring points using optimal interpolation assuming that mean rainfalls are known.

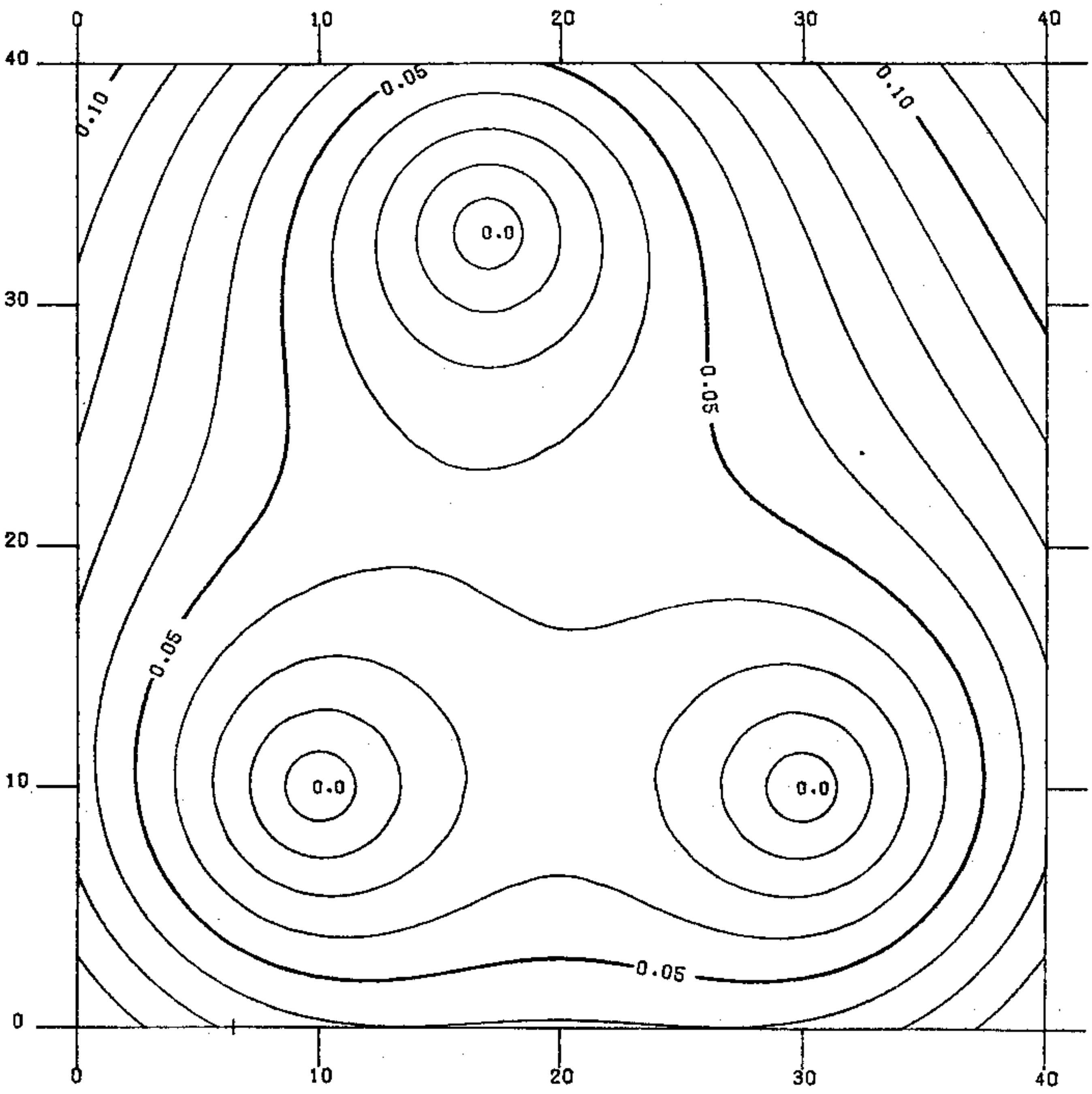


Figure 5.3 : Contours of interpolation error from three sites to neighbouring points using optimal interpolation assuming that mean rainfalls are constant but unknown.

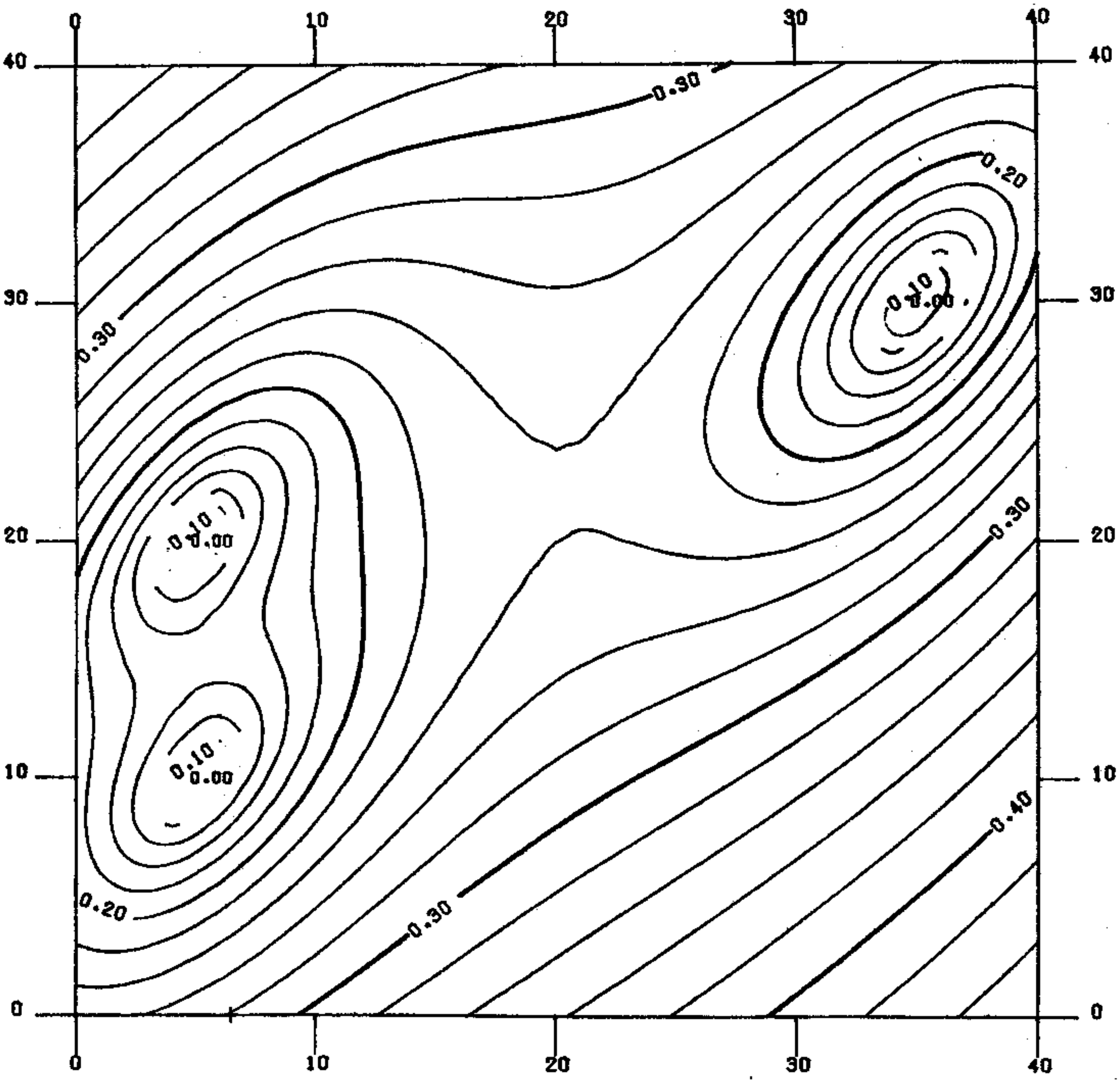


Figure 5.4 : Interpolation error for an anistropic correlation function : optimal interpolation with means equal but unknown. (orientation 1).

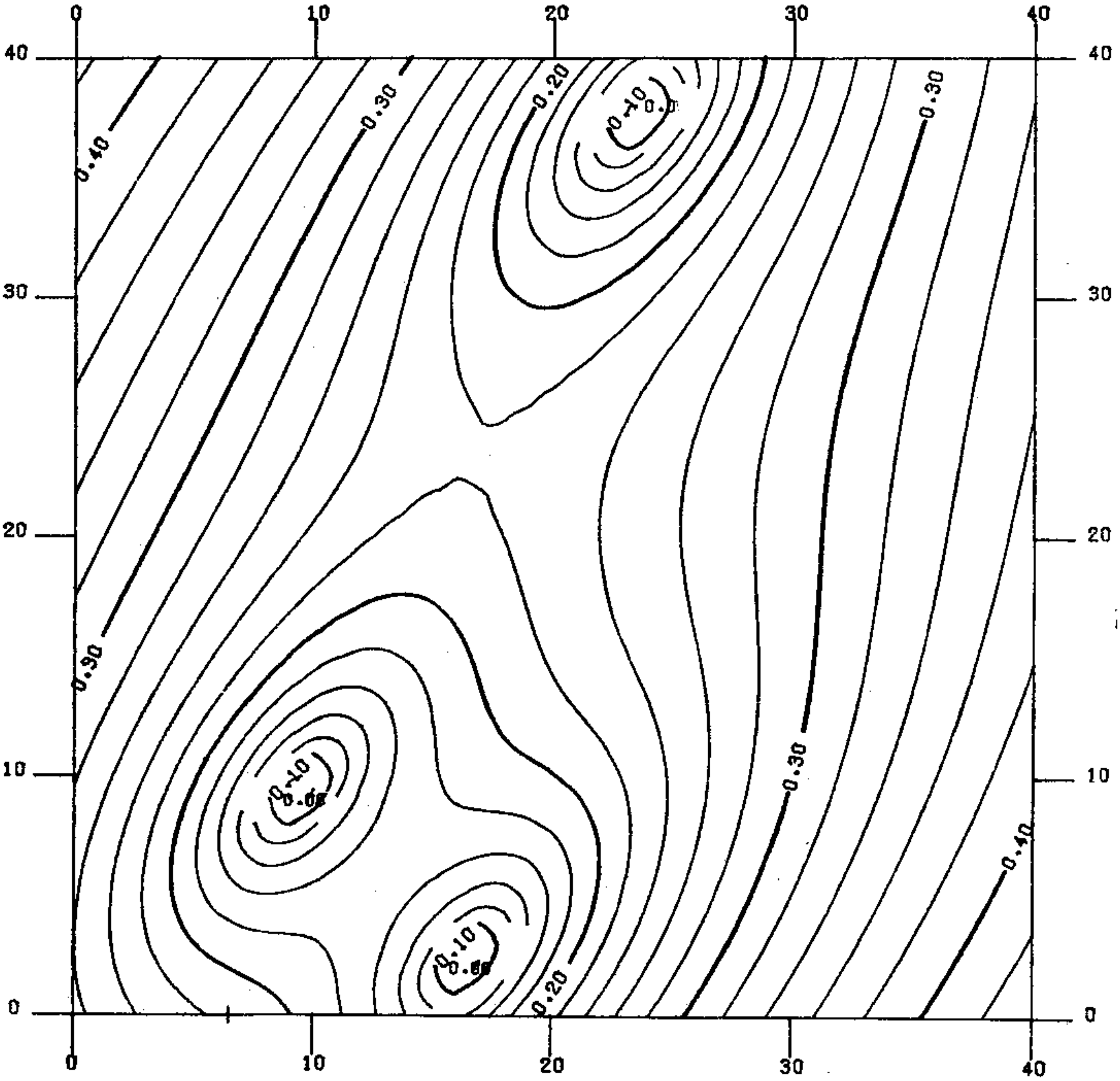


Figure 5.5 : Interpolation error for an anisotropic correlation function: optimal interpolation with means equal but unknown. (orientation 2)

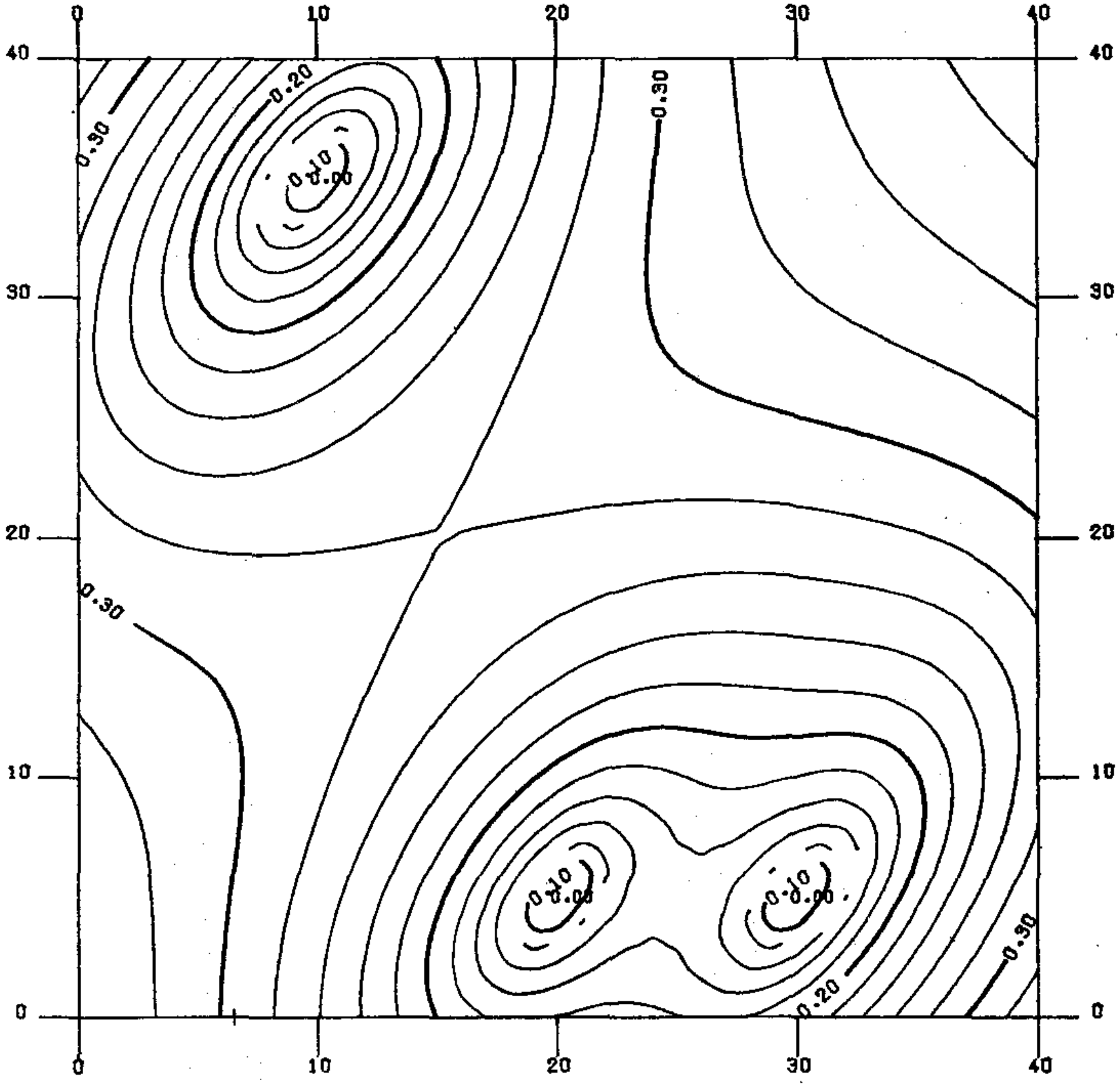


Figure 5.6 : Interpolation error for an anisotropic correlation function: optimal interpolation with means equal but unknown. (orientation 3)

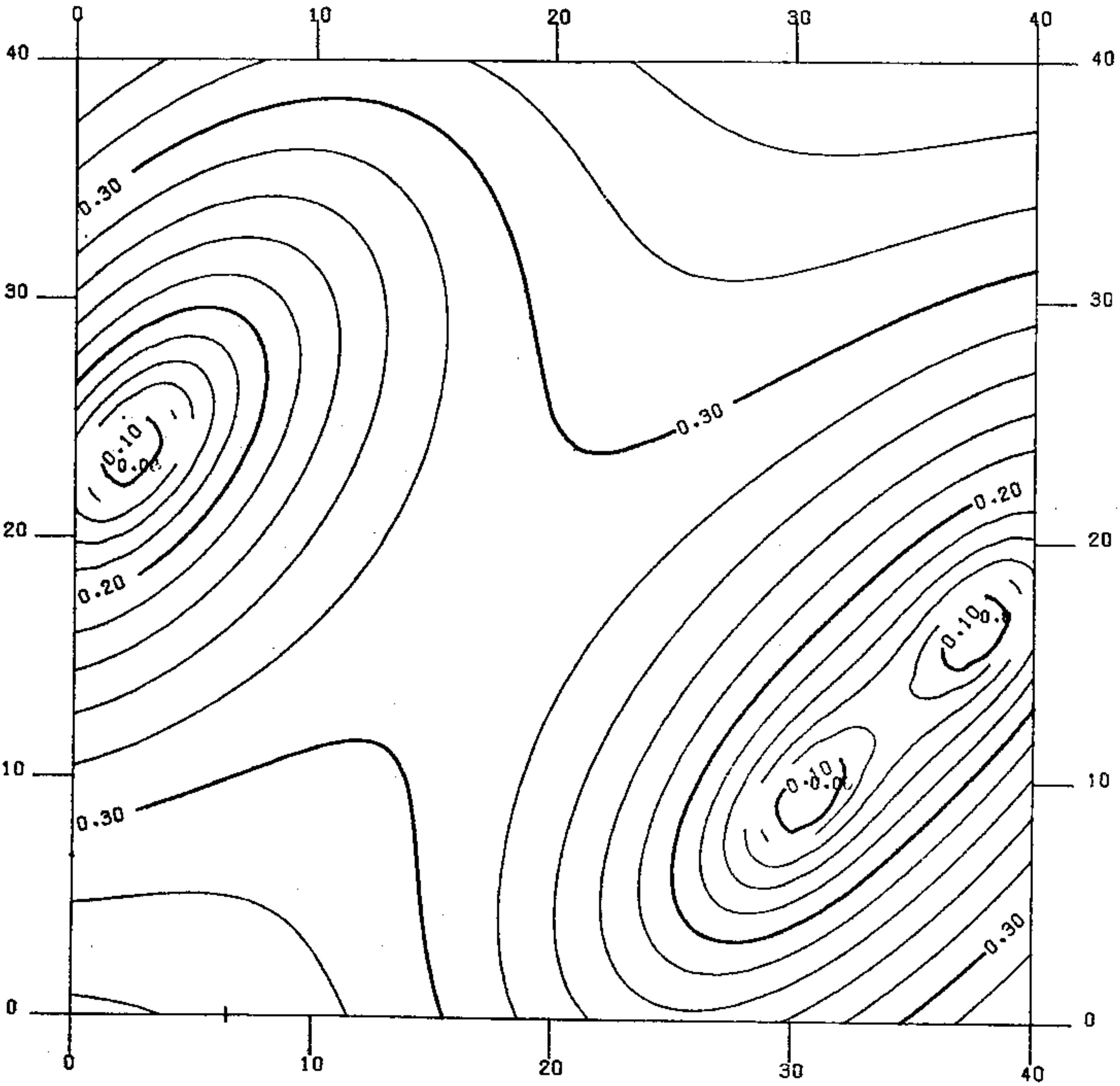


Figure 5.7 : Interpolation error for an anisotropic correlation function: optimal interpolation with means equal but unknown. (orientation 4)

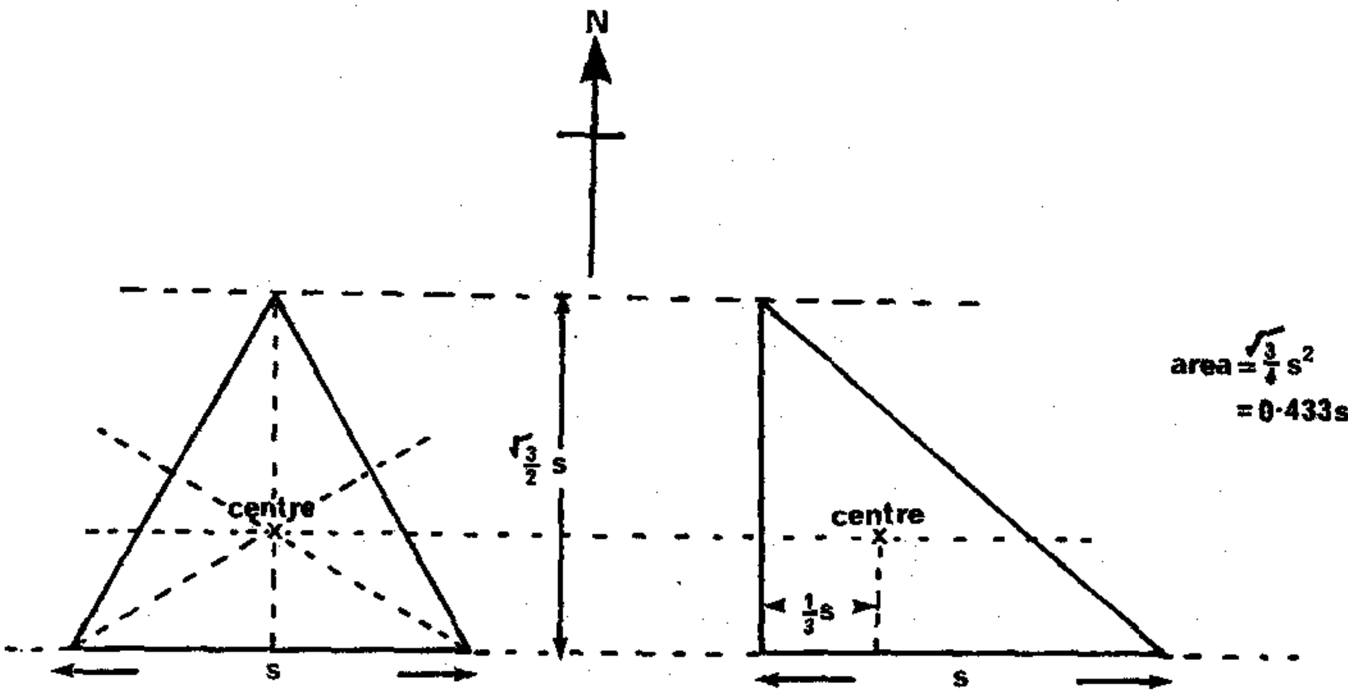


Figure 5·8 : *Hypothetical arrangement of gauges in equilateral and right-angled triangle.*

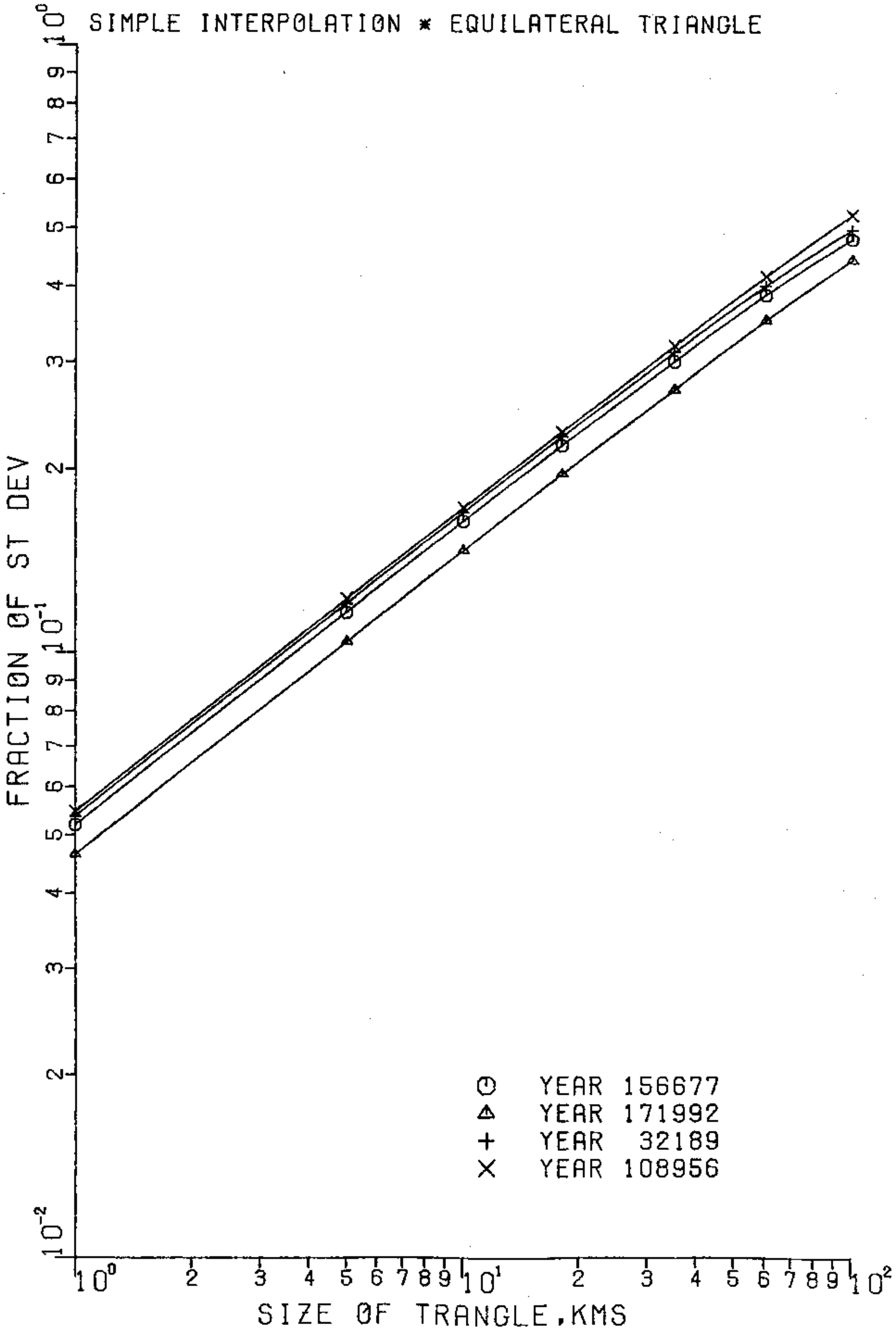


Figure 5.9 : Interpolation error to centre of equilateral triangle: yearly totals, Eastern and Northern England.

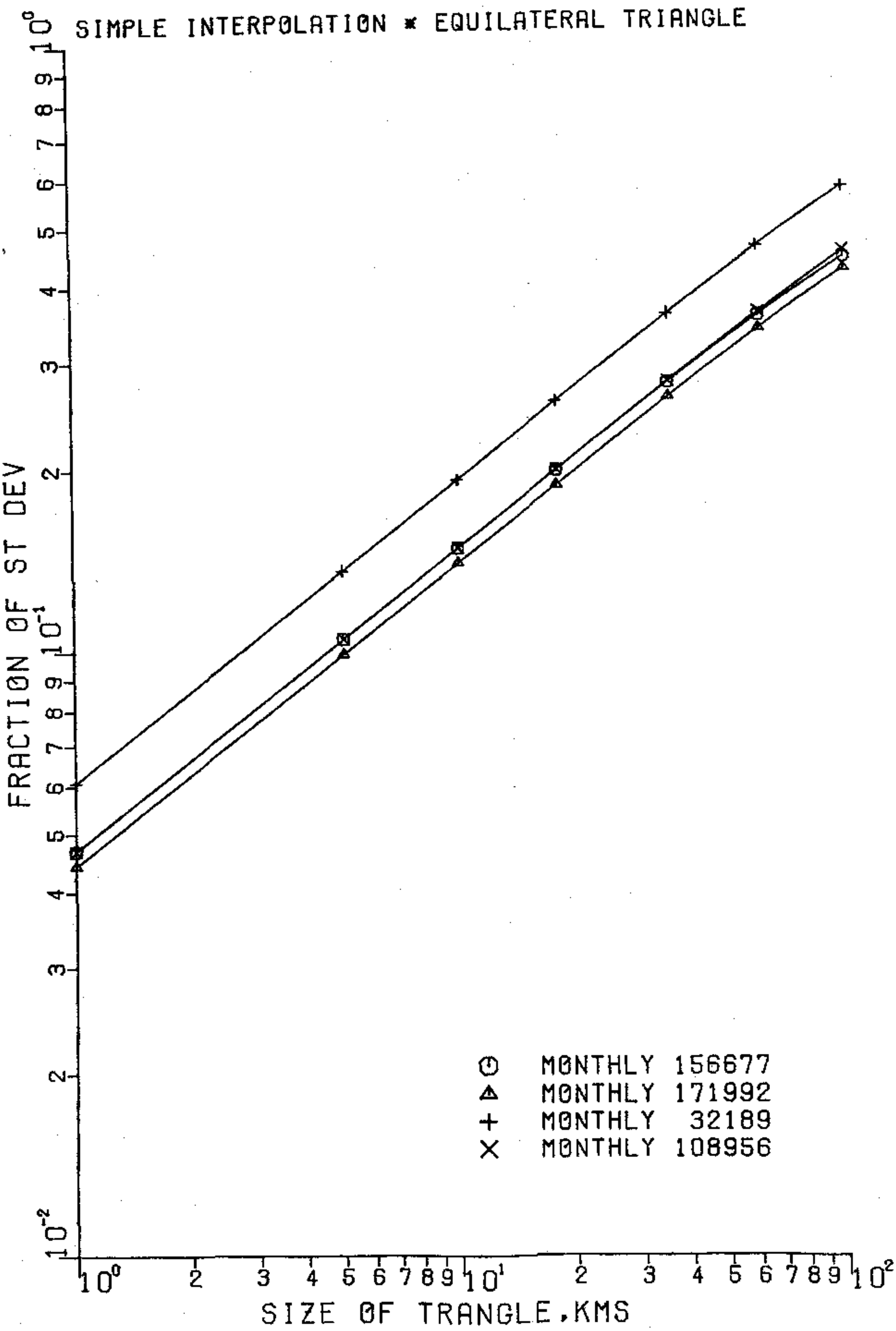


Figure 5.10 : Interpolation error to centre of equilateral triangle: monthly totals, Eastern and Northern England.

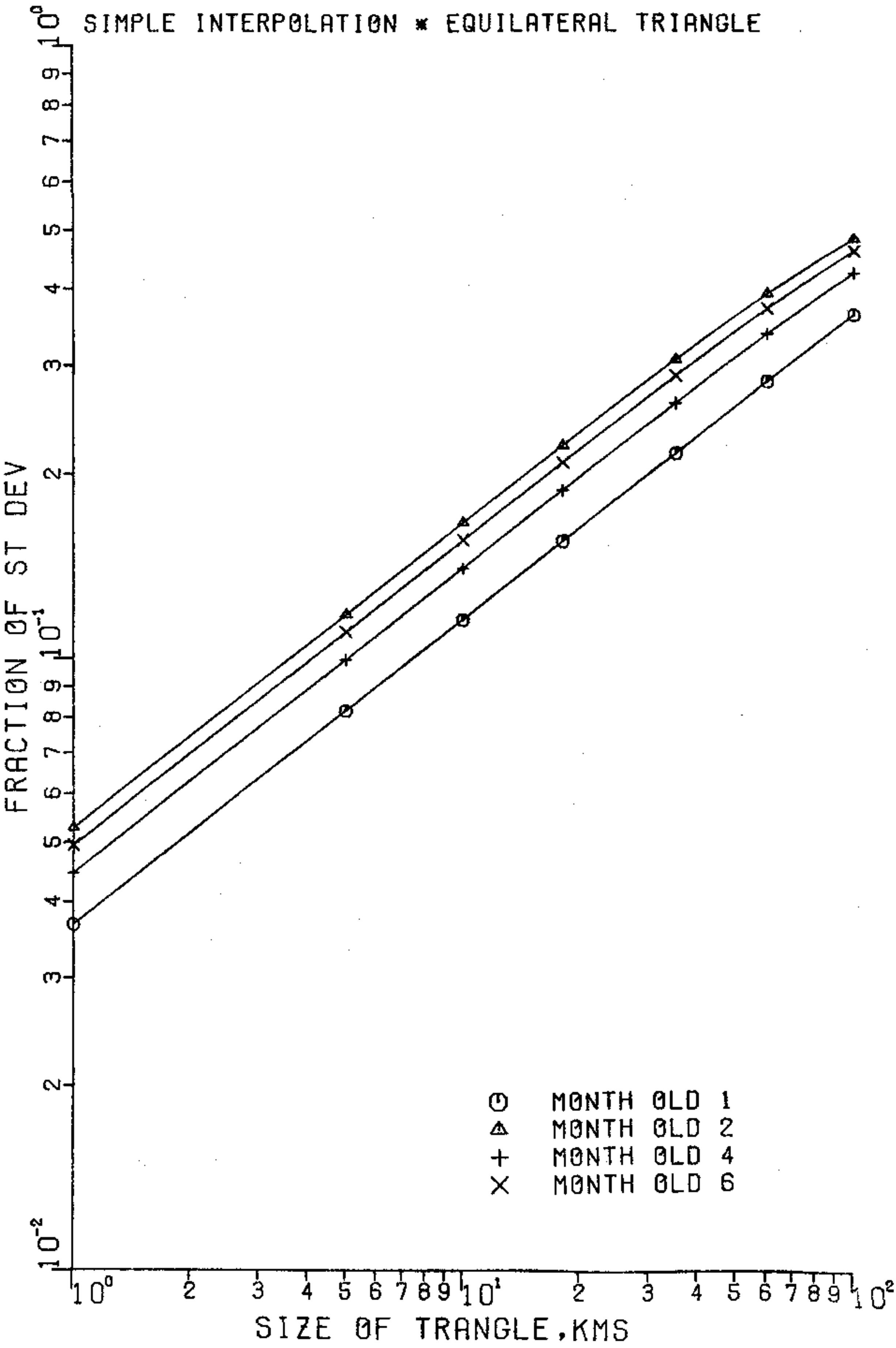


Figure 5.11 : Interpolation error to centre of equilateral triangle: monthly totals, Eastern England (1875-1890)

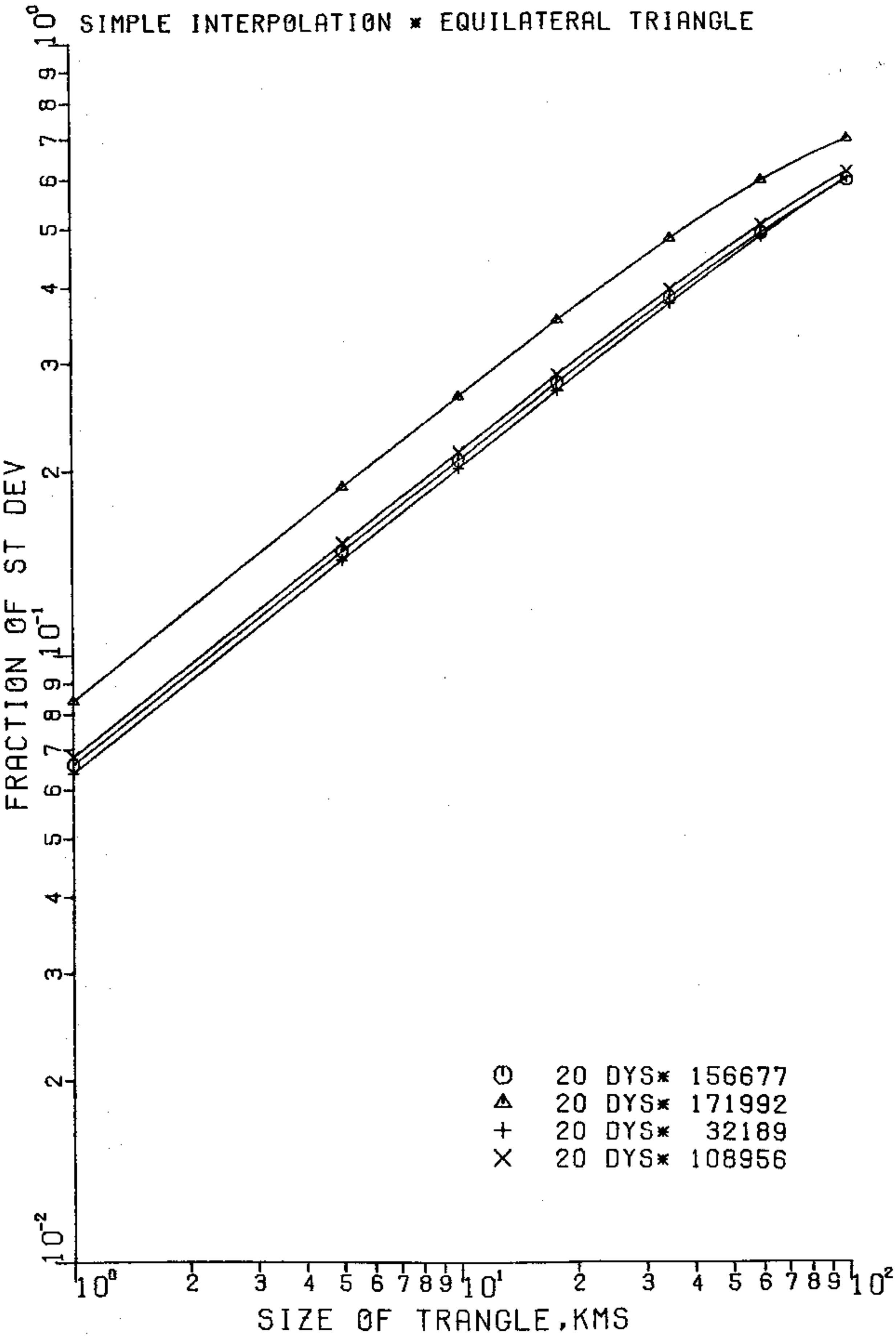


Figure 5.12 : Interpolation error to centre of equilateral triangle: daily totals, Eastern and Northern England.

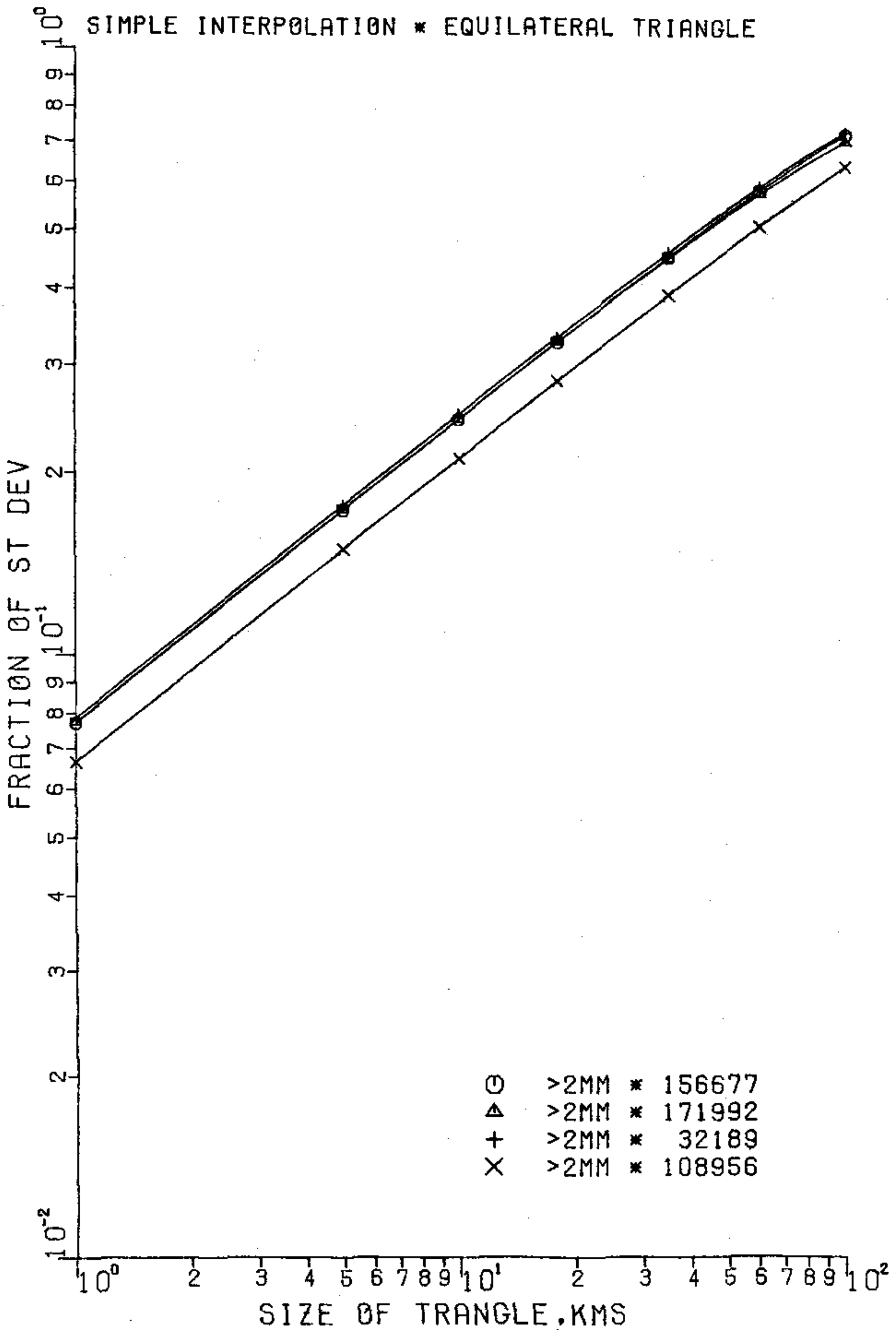


Figure 5.13 : Interpolation error to centre of equilateral triangle: days with rainfall over 2 mm, Eastern and Northern England.

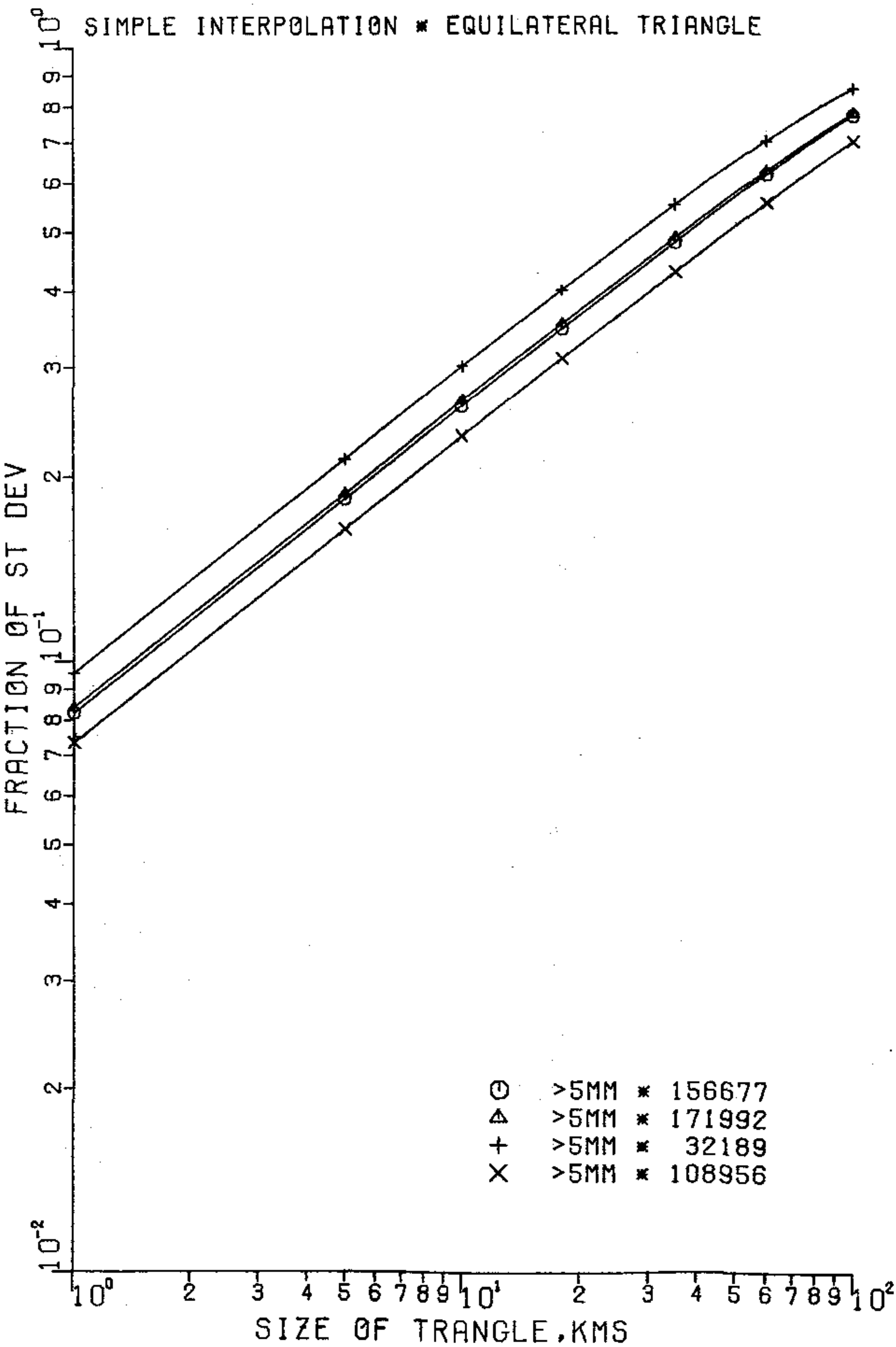


Figure 5.14 : Interpolation error to centre of equilateral triangle: days with rainfall over 5 mm, Eastern and Northern England.

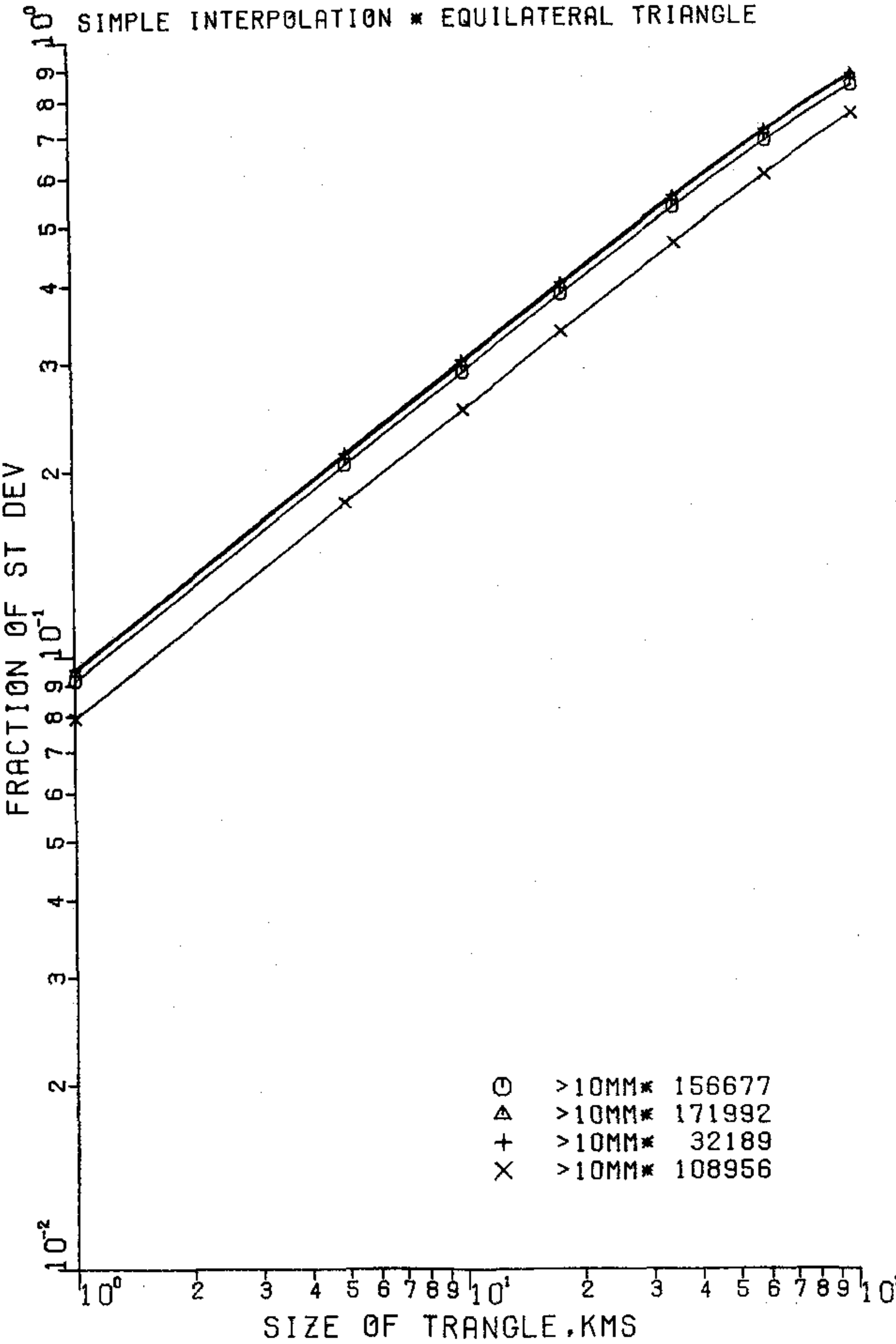


Figure 5.15 : Interpolation error to centre of equilateral triangle : days with rainfall over 10 mm, Eastern and Northern England.

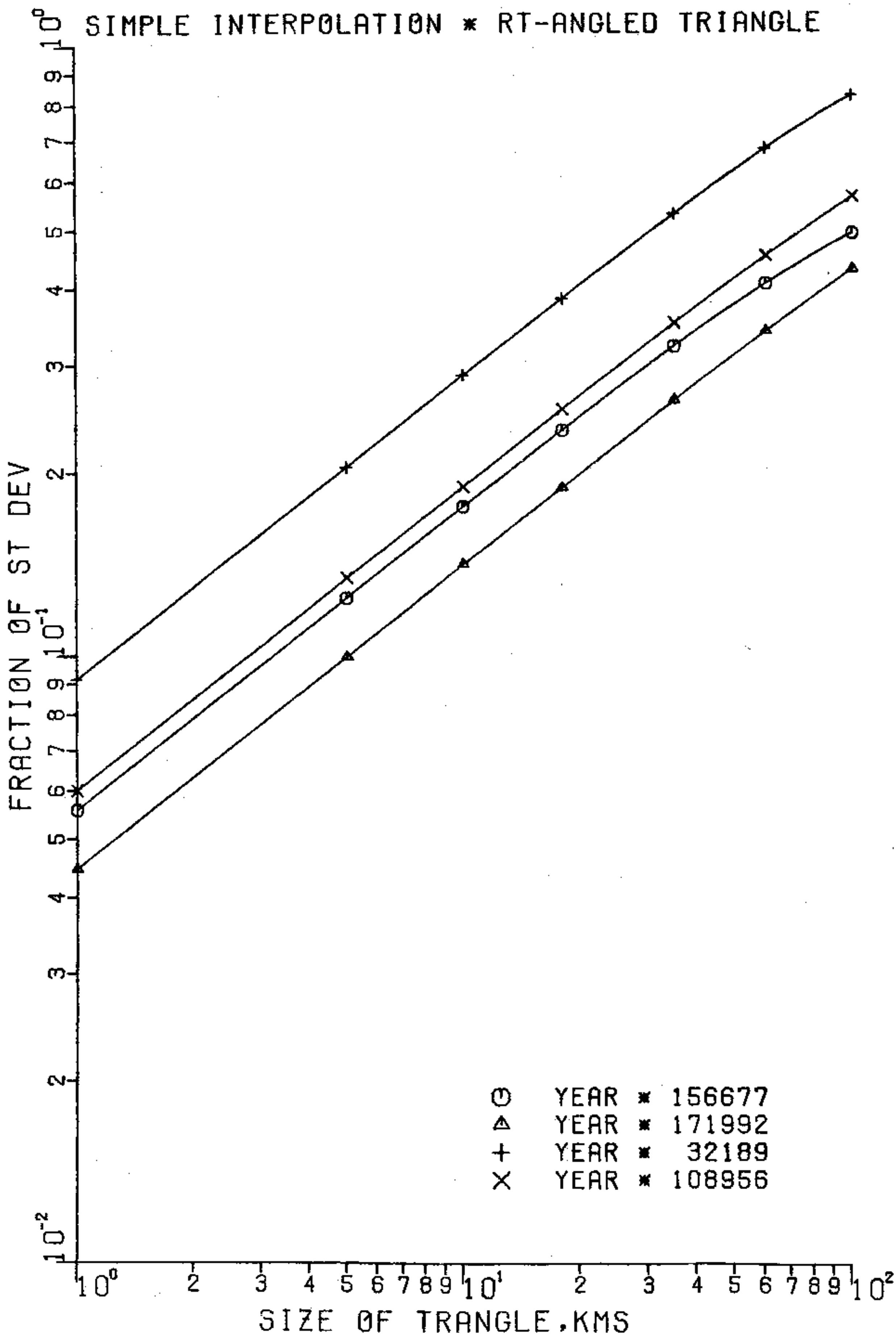


Figure 5.16 : Interpolation error to centre of right-angled triangle: yearly totals; Eastern and Northern England.

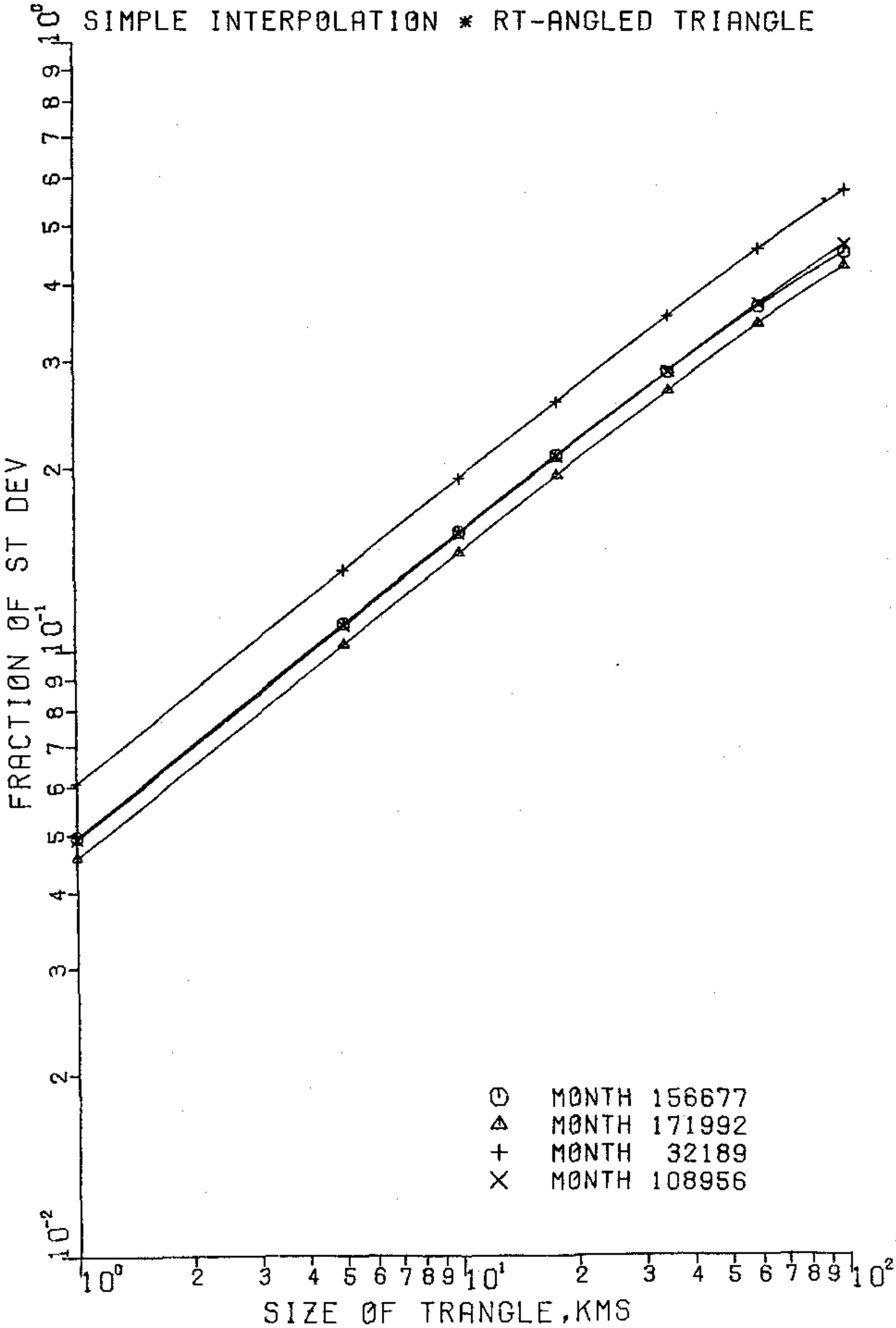


Figure 5.17 : Interpolation error to centre of right-angled triangle: monthly totals, Eastern and Northern England.

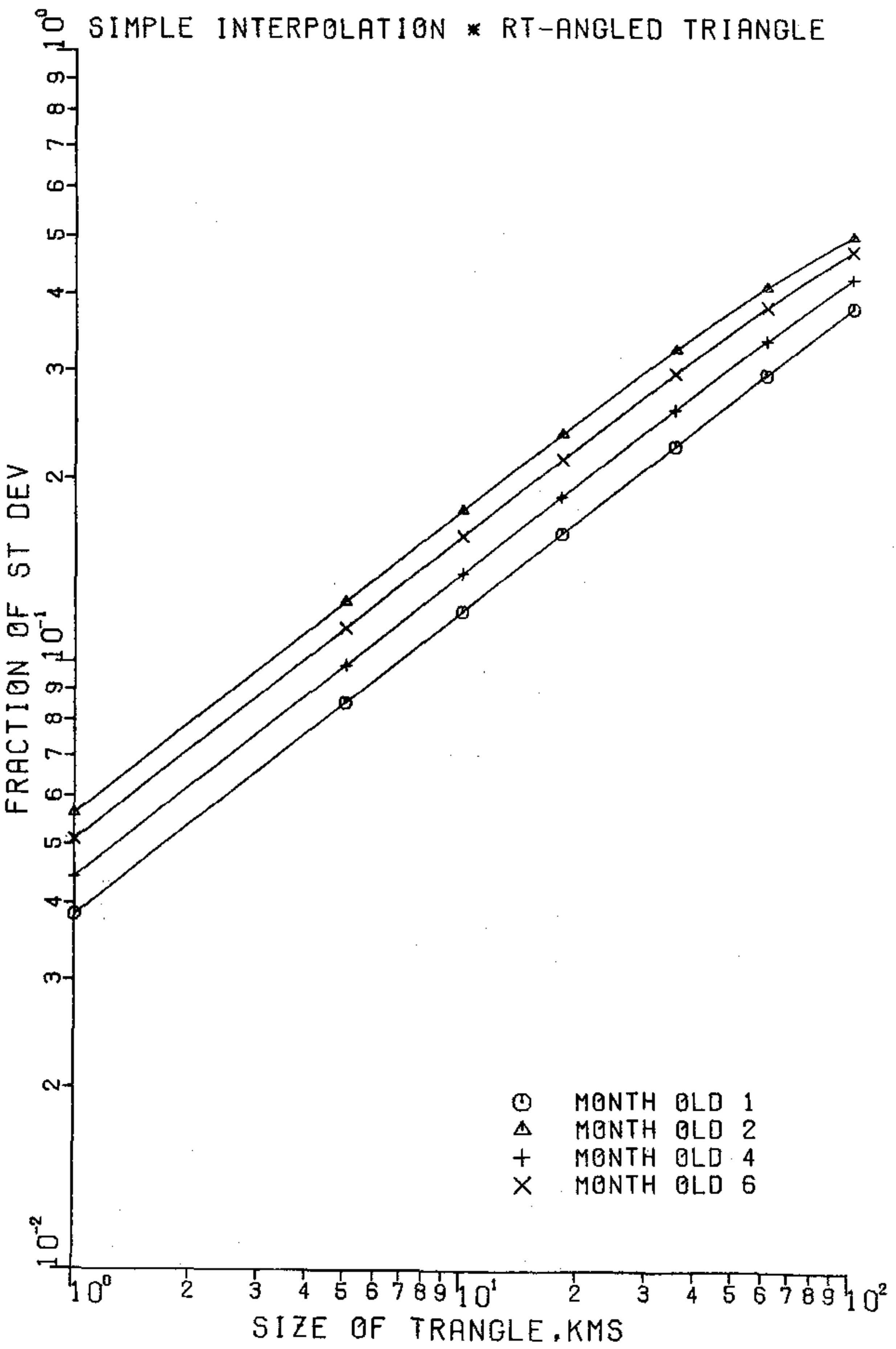


Figure 5.18 : Interpolation error to centre of right-angled triangle: monthly totals, Eastern England (1875-1890).

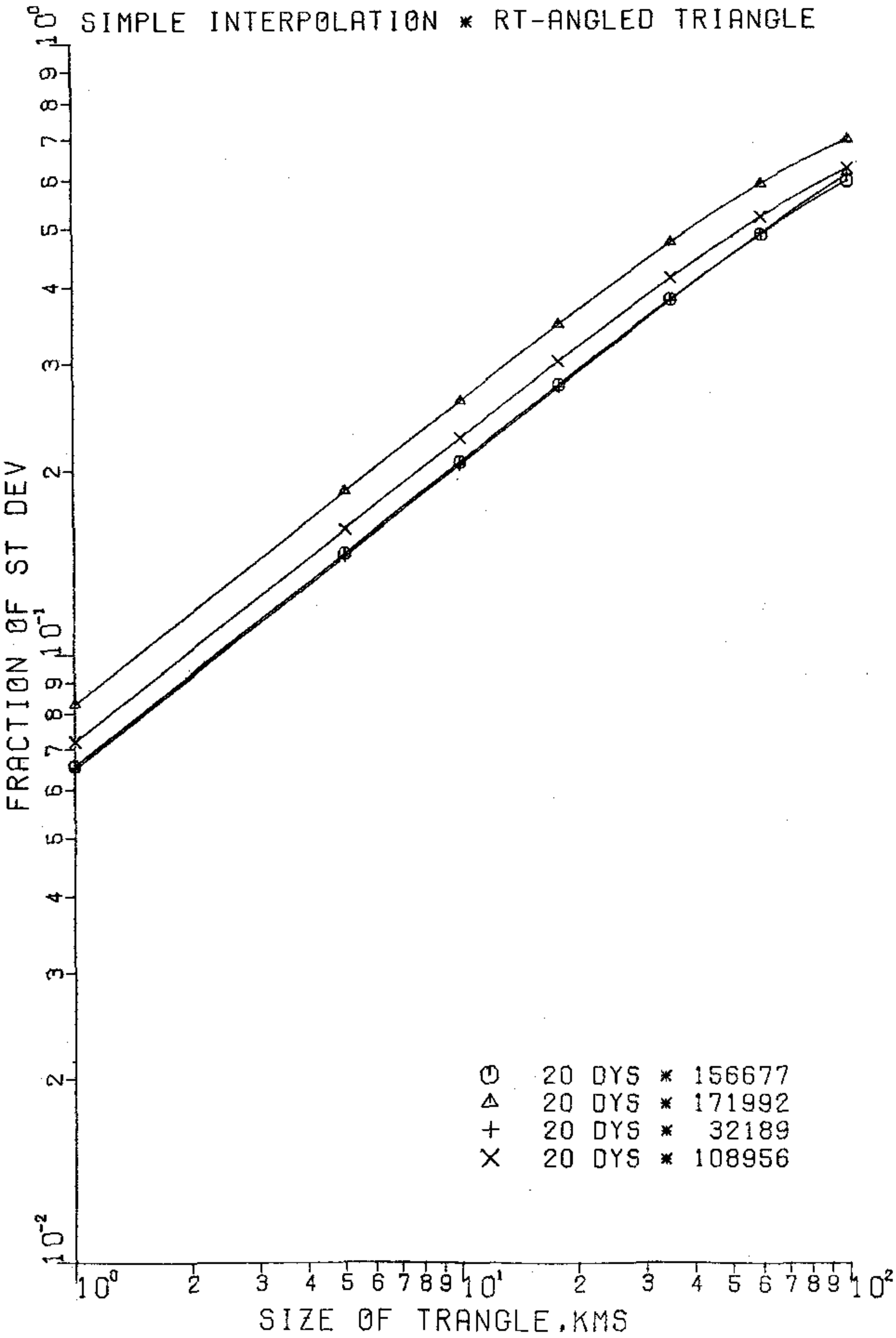


Figure 5.19 : Interpolation error to centre of right-angled triangle: daily totals, Eastern and Northern England.

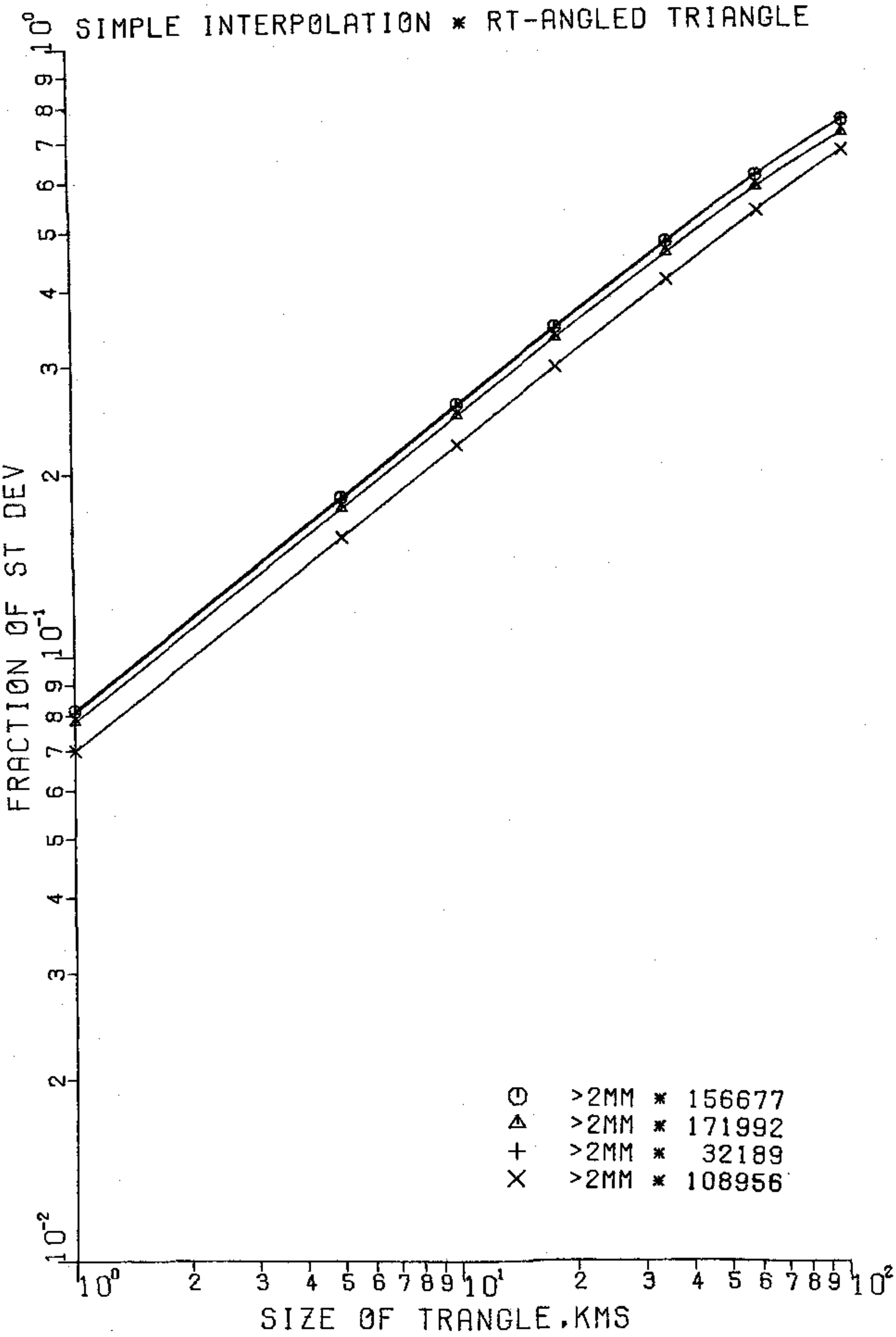


Figure 5.20 : Interpolation error to centre of right-angled triangle: days with rainfall over 2 mm, Eastern and Northern England.

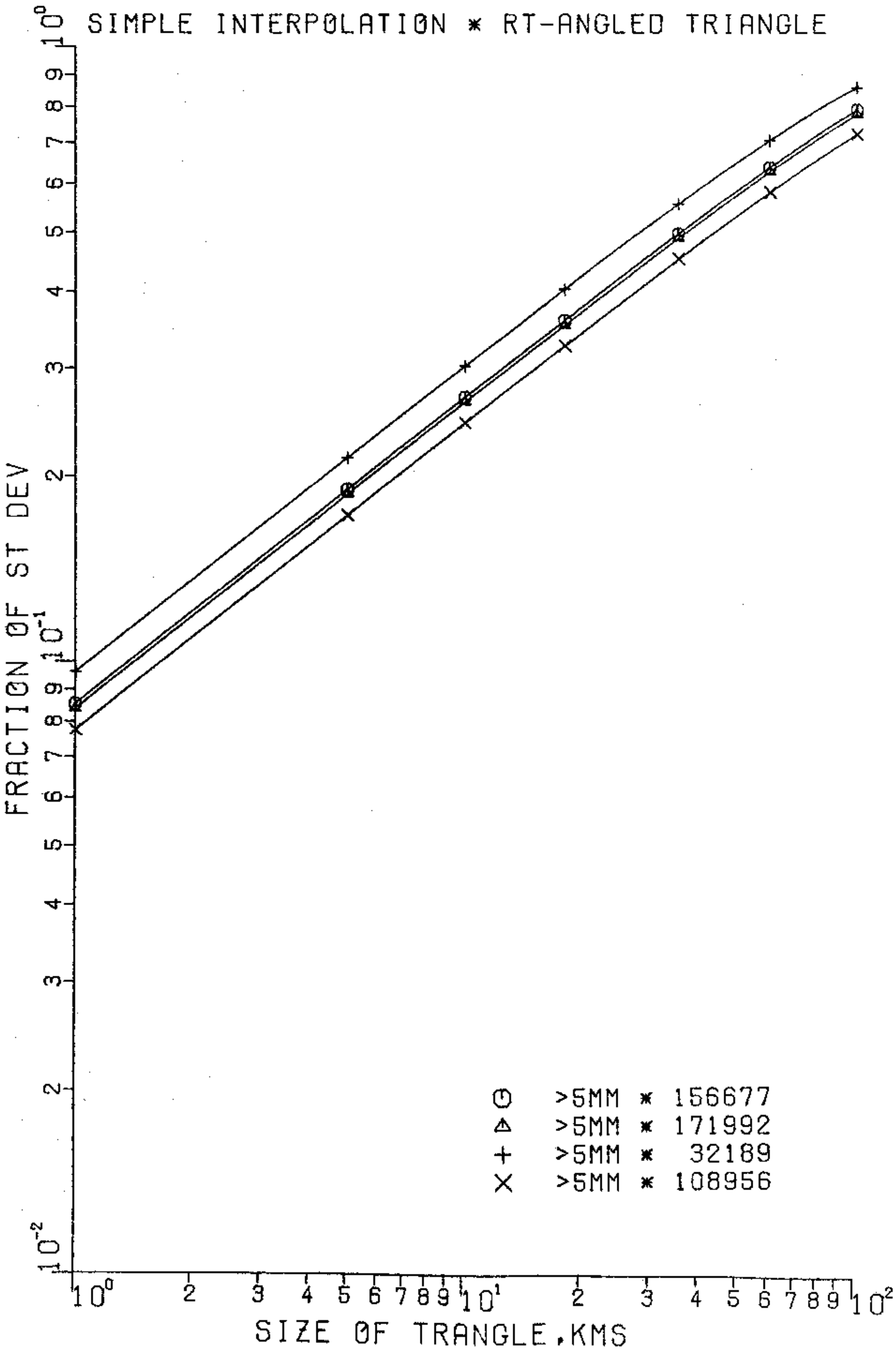


Figure 5.21 : Interpolation error to centre of right-angled triangle: days with rainfall over 5 mm, Eastern and Northern England.

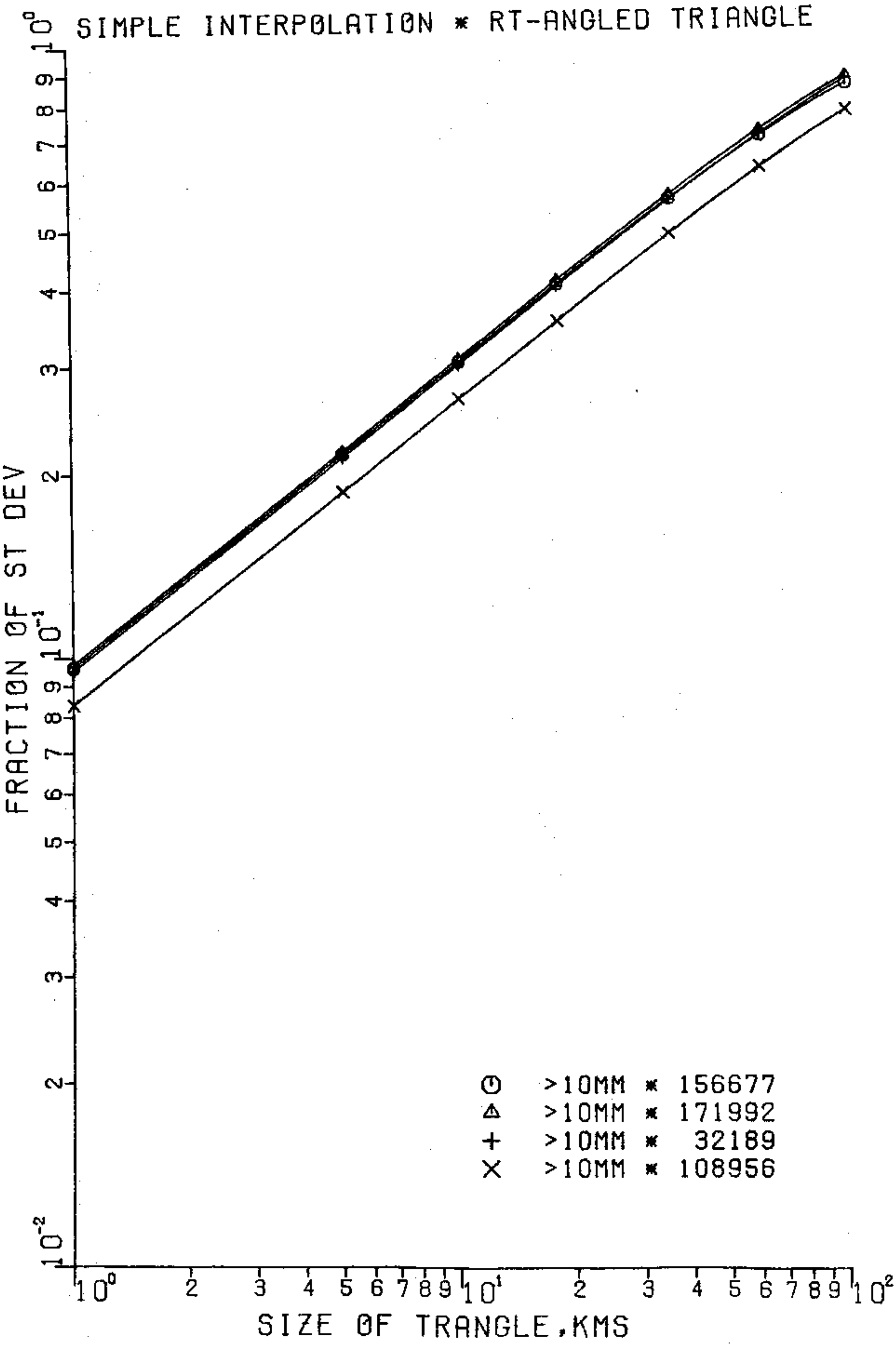


Figure 5.22 : Interpolation error to centre of right-angled triangle: days with rainfall over 10 mm, Eastern and Northern England.

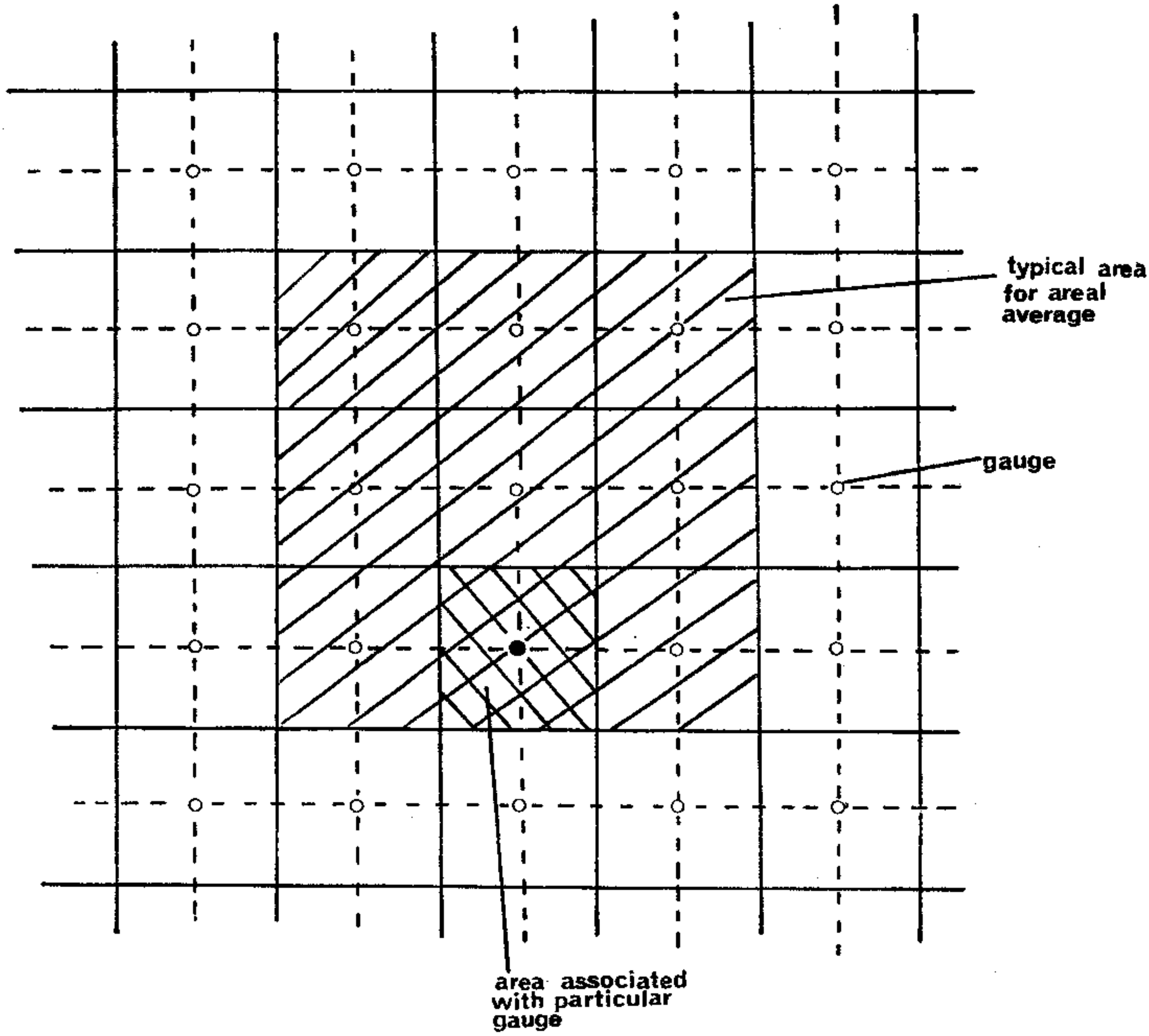


Figure 5.23 : Arrangement of gauges on square grid with area over which the average rainfall is to be found.

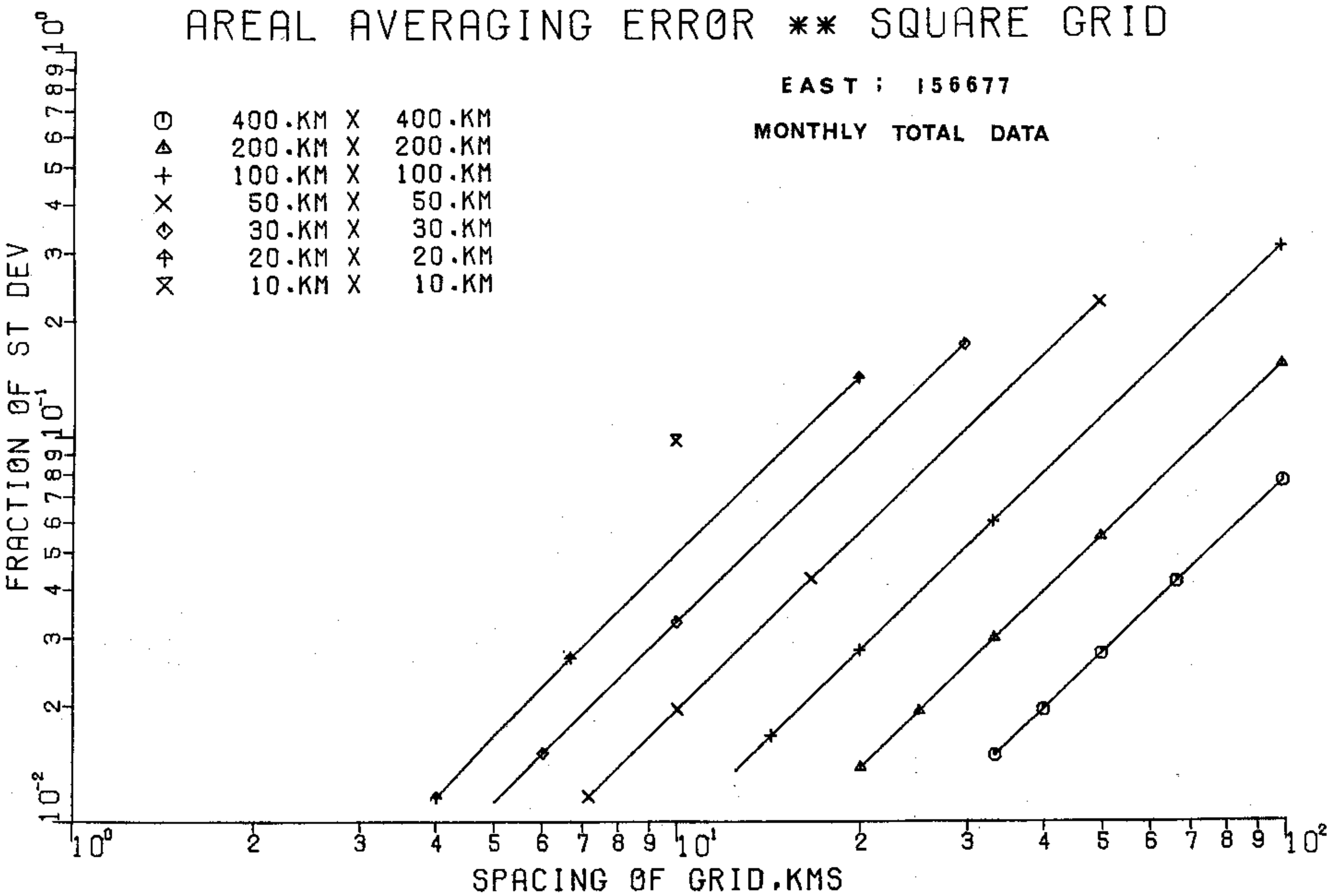


Figure 5.24 : Error for estimate of average rainfall over areas of different sizes: monthly total rainfalls.

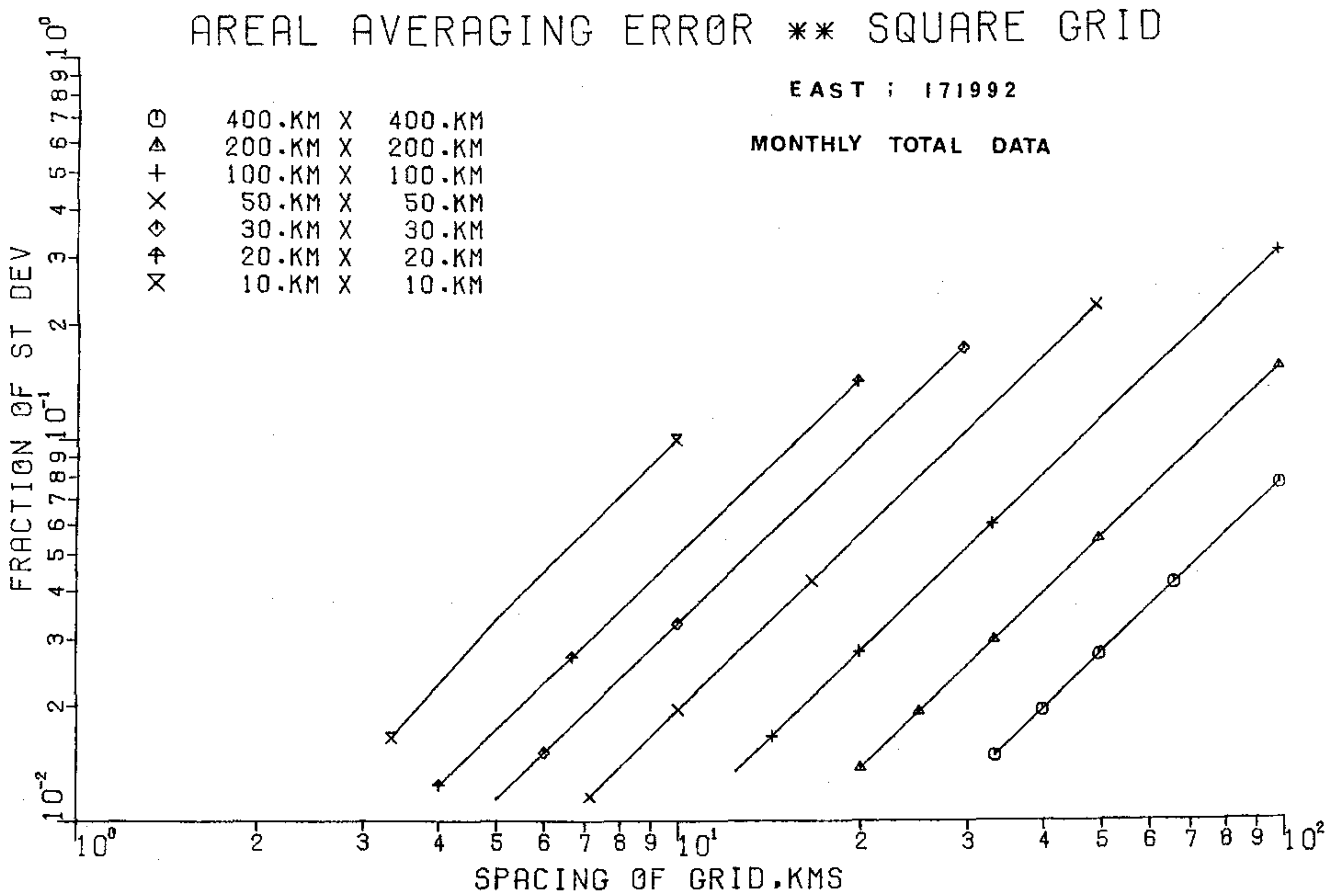


Figure 5.25: Error for estimate of average rainfall over areas of different sizes: monthly total rainfalls.

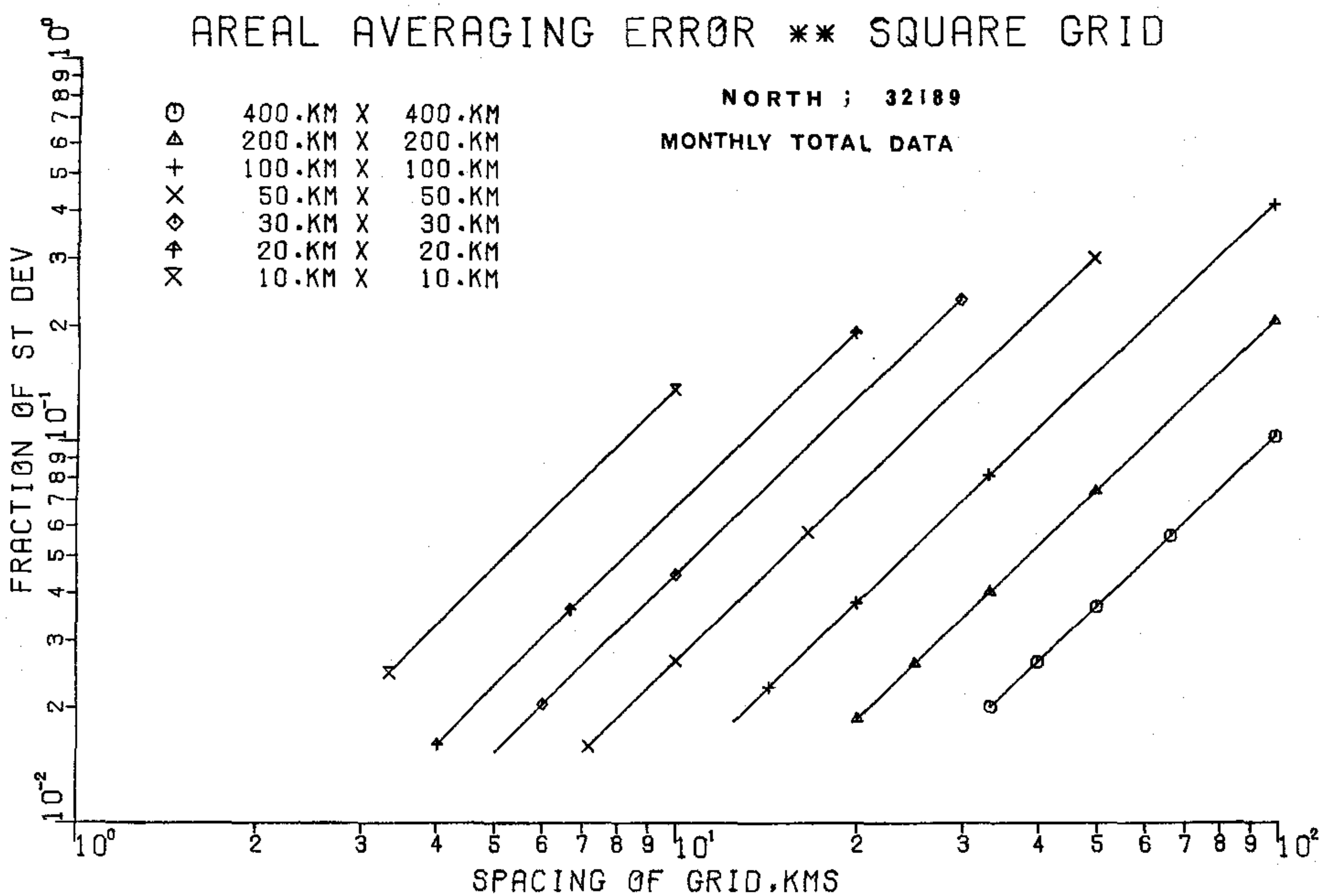


Figure 5.26: Error for estimate of average rainfall over areas of different sizes: monthly total rainfalls.

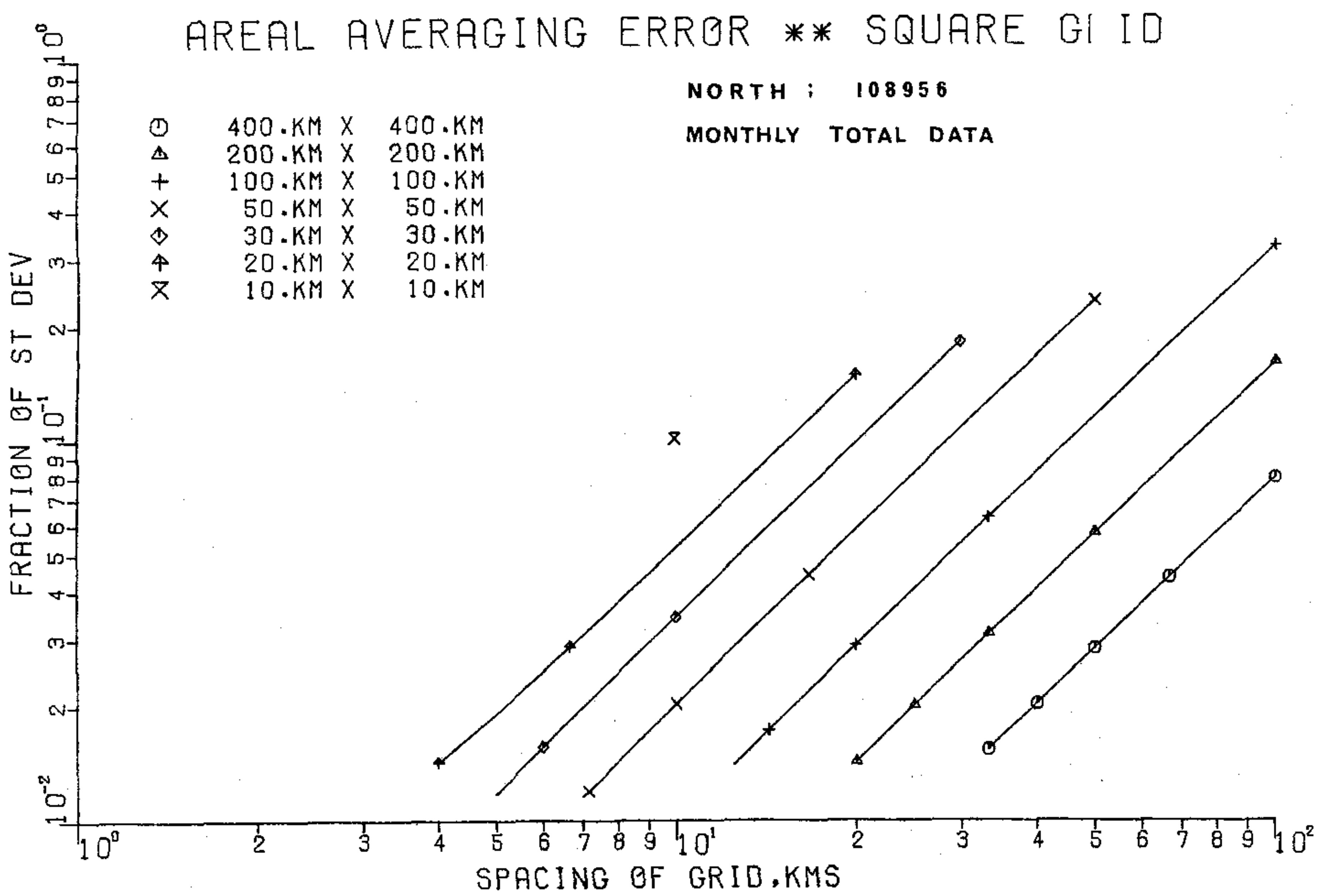


Figure 5.27 : Error for estimate of average rainfall over areas of different sizes: monthly total rainfalls.

6. METHODOLOGY FOR RAINGAUGE NETWORK EVALUATION: INDIRECT METHODS

6.1 Introduction

It is often difficult to specify what the accuracy requirements of rainfall data are when rainfall acts as an input to a complex decision-making process, and is not itself the basis of the decision. Under such circumstances, two approaches are possible. One approach analyses the nature of the decision and an attempt is then made empirically to infer what the accuracy requirement of the rainfall input should be to increase the likelihood of a 'good' decision. This is essentially the approach that has been adopted in specifying some of the accuracy requirements in Table 3.2. An alternative approach is to attempt to model the decision-making process in toto; the model can then be used to examine the sensitivity of the decision to various levels of accuracy for the rainfall input. This is invariably too formidable a task; however, where water resources studies are in question, the decision variable will quite often be some function of streamflow, or some closely related variable, and so a viable approach should be to model the rainfall-runoff process on the time scale required. The model can then be used to determine what accuracy of rainfall input is required to give the required accuracy of estimation for whatever function of streamflow is in question and thereby allow a 'good' decision to be made. Such an approach will be referred to here as an indirect method of network evaluation.

Where short duration rainfall data are being considered, there is a further justification for adopting a modelling approach. As noted in Sections 1.4 and 3.3.1, the extreme non-normality of short duration rainfall data poses problems for a conventional spatial analysis approach. Where such data are to be used in rainfall-runoff modelling, the most likely applications are liable to be within the areas of urban hydrology (e.g. for storm sewer network design) or streamflow forecasting (e.g. for real-time flood control). Within the time available, no attempt has been made to study rainfall network requirements for urban hydrology problems; however, a brief study has been made of the latter requirements for real-time flood forecasting.

In the following sections, the influence of raingauge network density on the accuracy with which streamflow may be modelled has been investigated for two distinct problems. The first problem is concerned with the behaviour of streamflow in the longer term and, in hydrology, is generally classified as a prediction problem; the second is primarily concerned with streamflow behaviour in the shorter term, and is classified as a forecasting problem. A daily rainfall-runoff modelling approach is adopted to address the former problem (Section 6.2) while, for the latter, a rainfall-runoff model is employed which is applied to the real-time forecasting of streamflow on a half-hourly time scale (Section 6.3). An important attribute of the real-time forecasting model is that, in real-time, its forecast is updated on the basis of the latest observed value of streamflow, a facility which the majority of conceptual models do not have.

6.2 Raingauge network requirements for daily rainfall-runoff modelling

6.2.1 General approach

Rainfall-runoff modelling studies are now being undertaken more frequently than hitherto by Water Authorities, and in many cases it is important to model the spatial variability of rainfall and runoff. A typical case occurs when it is found that rainfall records within the study area are considerably longer than streamflow records, and a *model is then used to* extend the streamflow records. A typical example of such a study is given by Manley and Sexton (1976).

A number of studies have been carried out to determine what raingauge network requirements are appropriate for various rainfall-runoff modelling problems; a review of some of these studies has been presented in Section 2.4, where it was noted that most tend to be of a specialised nature. Accordingly, it is difficult to generalise the results. For some of these specialised problems, an analytical treatment is feasible; however, for more general problems, a semi-empirical approach involving simulation is usually required. This latter approach will be adopted here.

The effect of the spatial variability in rainfall on rainfall-runoff modelling would be expected to be most important where distributed catchment models are being used; here it is important to have both a configuration and *density* of gauges over a catchment to ensure that any spatial variation in rainfall is directly measured on a time scale which is sufficiently short to accurately model the catchment response.

Where a catchment model with a lumped input is to be used, the problem of accurately defining the rainfall input is invariably taken to be equivalent to having a sufficient density of gauges to define the mean rainfall over the catchment as accurately as possible for the time unit in question. Provided all the gauges are representative, then the more gauges that are sited within a catchment, the more stable will the estimate of mean areal rainfall become. Accordingly, errors in modelling streamflow arising from instability in mean areal rainfall should be minimised as a result.

In defining a lumped input to a catchment model, it is usually assumed that the best strategy is to use as many gauges as possible within and adjacent to the catchment to define mean areal rainfall. This may not always, in practice, prove to be justified. Due to the presence of unrepresentative gauges, there may be a sub-set of gauges which provide an input which is more highly correlated with observed streamflow than the input provided by the maximum number of gauges. However, this possibility will not be considered further in relation to the rainfall-runoff modelling carried out here, but will be discussed in Section 6.3 in relation to the short-term forecasting of streamflow.

Assuming that decreasing sub-sets of raingauges have been identified so that a lumped rainfall input to a model can be provided for different gauge densities (discussion of how such sub-sets might be identified is deferred to Section 6.2.2), a procedure is required whereby the effect of gauge density on model output can be quantified. A general measure of agreement between observed streamflow, Y_t, and predicted streamflow, Y_t^*, is provided by the coefficient of determination, R^2, which is defined as

$$R^2 = \frac{F_o^2 - F^2}{F_o^2} \qquad \ldots. \qquad (6.1)$$

where

$$F_o^2 = \sum_{t=1}^{N} (Y_t - \overline{Y})^2 \quad .... \quad (6.2)$$

$$F^2 = \sum_{t=1}^{N} (Y_t^* - Y_t)^2 \quad .... \quad (6.3)$$

and $\overline{Y}$ is the sample mean of Y_t over the period of record. For perfect agreement between Y_t and Y_t^*, clearly $F^2 = 0$ and $R^2 = 1$. The object of model calibration by least squares is to minimise F^2 and thus maximise R^2.

In calculating the effect of raingauge density on the coefficient of determination R^2, two approaches are possible. One is to recalibrate the model for each input corresponding to a particular sub-set of gauges, as carried out by Richards (1975). However, the difficulty with this approach is that the inaccurate definition of mean areal rainfall input tends to be compensated for by re-adjustment of the model parameters on calibration, thus masking the true effect of inadequate definition of rainfall input on R^2. The resulting conclusion may then be that network density has little influence on R^2 (Richards, 1975).

The fact that the effects of model fitting and network density on R^2 are inseparable using the above approach does not mean that if the true model were available and its parameters known, the effect of network density on R^2 could not be quantified. An approach whereby such effects can be separated is the following. Using the total number of gauges to define the rainfall input, a catchment model is calibrated to achieve a maximum coefficient of determination R^2. At this stage the model is assumed to be the true model, and its parameters known without error. Thus, a time series of 'synthetic' runoff values may be generated by the model; for this particular model, its associated set of parameter values and 'synthetic' runoff record, $R^2 = 1$. This 'synthetic' runoff record is henceforth treated as Y_t. By successively decreasing the number of gauges used in computing the rainfall input and keeping the model parameters unchanged, a set of model outputs, Y_t^*, with associated values of R^2 can be computed. The resulting relationship between R^2 and i, the number of gauges, $i = 1, 2, ...P$, will obviously be specific to the model; nevertheless, provided a reasonably good fit is obtained in fitting the model in the first place, the relationship between R^2 and i should be a reasonable approximation to reality. A computational advantage is that re-calibration of the model is avoided for each sub-set of gauges. This approach is applied in the following sections.

In Section 6.2.2, methods for selecting sub-sets of gauges are considered, while in Section 6.2.3, the catchment model used is described. Section 6.2.4 gives a brief description of the catchments modelled, and the results are presented in Section 6.2.5.

6.2.2 Selection of sub-sets of gauges

In order to assess the effect of raingauge density on rainfall-runoff modelling, a procedure is required for selecting decreasing sub-sets of gauges from the total set. If all possible sub-sets are considered, then for P gauges there will be ${}^{P}C_r$ possible ways of carrying out a sub-set selection, $r = 1, 2, ..., (P-1)$ which for even moderate values of P will clearly lead to an unmanageable number of inputs. This number can be reduced significantly by eliminating a gauge from consideration as the sub-set size is decreased by one, so that ${}^{P-r}C_{P-r-1}$ selections are then possible, r = 0, 1, ...(P-2); however, the number of inputs will still create computational problems. Thus, some means of reducing the number of possible combinations further is desirable. This will largely depend on the criterion used for eliminating gauges.

One possible basis for an elimination criterion is cross-correlation analysis. By computing the cross-correlation function for various lags between each possible rainfall input and streamflow over the concurrent period of record, the sub-sets of gauges yielding the highest overall correlation could be selected. A difficulty with using the cross-correlation function is that there is no obvious objective measure of overall correlation which can be computed from it. Nevertheless, even if such a measure could be defined, for a catchment with 15 gauges and using the more restricted selection procedure mentioned above, 120 cross-correlation functions would have to be computed. As it was felt that more could be gained from carrying out a less sophisticated selection procedure on more than one catchment, the following approach was adopted.

Firstly, the maximum number of gauges to be used in calculating mean areal rainfall on a daily time scale is selected: this is assumed to give the best possible definition of rainfall input to the model. If $\overline{U}_t$ denotes the mean areal rainfall at time t for P gauges, and $U_{t,i}$ denotes the daily rainfall at gauge i, then an error criterion for gauge i, i = 1, 2,....,P is defined as

$$e_i = \frac{1}{N} \sum_{t=1}^{N} (U_{t,i} - \overline{U}_t)^2 \qquad \ldots. \qquad (6.4)$$

where N is the number of time points. The quantity e_i thus provides a measure of the extent to which an individual gauge deviates from the mean areal rainfall over the period of record. If the e_i values are ranked from largest to smallest, then the order of the e_i values can be used to decide the order in which the gauges should be eliminated as the sub-set size is decreased. Accordingly, as each gauge is eliminated, the mean areal rainfall is re-computed; thus for P gauges, only P inputs are generated by this procedure which is manageable within the limits of the present study.

In the case of one gauge, the procedure provides a unique selection; however, in identifying larger sub-sets, the sub-set which, when used to compute mean areal rainfall, gives the smallest value of e_i should ideally be selected; however, this as previously noted creates an unmanageable amount of computation. It is assumed that the elimination of a single gauge based on its e_i value leaves a sub-set providing good definition of the mean areal rainfall for that sub-set size.

6.2.3 The CLS Model

Over the past 15 years or so, a vast number of rainfall-runoff models have been developed; such models are usually based on quasi-physical concepts and are characterised by a set of parameters which govern a set of operations to be performed on model input to generate model output. Such models may be simple in structure and involve a limited number of parameters (e.g. O'Connell et al. 1970) or invoke multiple parameters in an attempt to achieve a more complete description of catchment behaviour, (e.g. Linsley and Crawford, 1968). The majority of these models embody lumped soil moisture stores of one form or another, which are recognised to be at variance with physical reality; nevertheless, by calibrating such models on observed streamflow records, it is often possible to achieve good measures of agreement between observed and predicted streamflows. More recently, there has been evidence of a move towards building models which are physically based and which attempt to formulate and solve the partial differential equations of flow over and through the soil and in channels (e.g. Freeze, 1972; Beven, 1977). Such models when developed to the level at which they can be applied in practice offer the promise of being able to predict accurately the effects of land use changes and to model ungauged catchments. For the present, however, the basic choice for a catchment model lies between that of a conceptual model of the type referred to above or a 'black-box' model of the type advocated by Todini and Wallis (1977) in which more emphasis is placed on statistically efficient parameter estimation methods.

A recent intercomparison study of conceptual models for operational hydrological forecasting (W.M.O. 1975) sought to establish the relative merits of a set of models which were either of the conceptual or 'black-box' type. The results of this study indicated that, provided that sufficient high quality data were available, then somewhat better results could be derived through the use of more complex models of the Stanford Watershed type as opposed to simpler alternatives. However, in the absence of high quality data, the 'black-box' CLS model (Todini and Wallis, 1977) which utilises statistically efficient parameter estimation procedures, and has minimal computational requirements, was found to perform as well if not better than some of its more complicated competitors.

The CLS model was adopted for the present study for a number of reasons. In view of the time constraint on the study, a model was required with low computational demands which could be calibrated quickly. The procurement of data other than daily rainfall and runoff data, which were readily available, and which represent the total data requirements of the CLS model, would have entailed further delays. In addition, if a model of the conceptual type were employed, then its application to a range of catchments might well have involved model re-structuring which could prove time consuming. Finally, the computer coding for the model in basic form was available.

The basis of the CLS model is a multiple-input single-output linear system which has been developed for hydrological application by Natale and Todini (1976a, b) and has been applied in non-linear form to daily rainfall-runoff modelling by Todini and Wallis (1977). In linear form, the model is defined as follows. Associated with each input is an impulse response $\underset{\sim}{\Sigma}_j = (v_{j1}, v_{j2}, \ldots\ldots v_{jk})^T$ where k is the length of each impulse response; the model can thus be written as

$$\underset{\sim}{Y} = \underset{\sim}{U}\ \underset{\sim}{V} + \underset{\sim}{\varepsilon} \qquad \ldots. \qquad (6.5)$$

where Y is an (N x 1) vector of discrete outputs (streamflow) sampled at a time interval Δt, $\underset{\sim}{U}$ is an (N x nk) partitioned matrix of discrete time input vectors, $\underset{\sim}{V}$ is an (nk x 1) vector of impulse responses, and n and N are respectively the number of inputs and the number of concurrent observations on each input and the output. Usually the estimate of $\underset{\sim}{V}$ is obtained through a straightforward application of least squares involving the inversion of the matrix $(\underset{\sim}{U}^T \underset{\sim}{\Sigma}_{\varepsilon}^{-1} \underset{\sim}{U})$ where $\underset{\sim}{\Sigma}_{\varepsilon}$ is the variance-covariance matrix of the errors.

However, this approach has a number of disadvantages, among which are

(i) the matrix $(U^T \Sigma_\varepsilon^{-1} U)$ is frequently ill-conditioned (Abadie, 1970), and errors introduced through matrix inversions may introduce errors comparable to the values of the parameters to be estimated

(ii) the estimated impulse responses may be oscillatory with a large proportion of negative values, which is in conflict with physical principles

(iii) continuity is not necessarily maintained.

Natale and Todini (1976a, b) have developed estimation procedures for the impulse responses which do not have shortcomings (i), (ii) and (iii). Their formulation of the problem is to minimise the functional

$$J(\underset{\sim}{\varepsilon}^T\underset{\sim}{\varepsilon}) = \tfrac{1}{2}\, \underset{\sim}{V}^T \underset{\sim}{U}^T \underset{\sim}{\Sigma}_\varepsilon^{-1} \underset{\sim}{U}\, \underset{\sim}{V} - \underset{\sim}{V}^T \underset{\sim}{U}^T \underset{\sim}{\Sigma}_\varepsilon^{-1} \underset{\sim}{Y} \qquad \ldots\ldots \quad (6.6)$$

subject to the constraints that

$$v_{ji} \geq 0 \quad i = 1, 2, \ldots k,\ j = 1, 2, \ldots, n \qquad \ldots\ldots \quad (6.7)$$

$$\sum_{i=1}^{k} v_{ji} = \phi_j \qquad \ldots\ldots \quad (6.8)$$

The coefficient ϕ_j is defined as

$$\phi_j = \frac{\sum_{t=1}^{N} Y_t}{\sum_{t=1}^{N} U_{t,j}} \qquad \ldots\ldots \quad (6.9)$$

where Y_t is streamflow at time t, and $U_{t,j}$ is the j^{th} rainfall input at time t, and thus corresponds to a coefficient of runoff for the j^{th} input. The minimisation of $J(\varepsilon^T\varepsilon)$ subject to the above constraints is achieved through quadratic programming. In the above description, the lengths of the impulse response vectors have been assumed equal for ease of presentation; no essential difficulty is encountered with non-equal values of k.

As formulated, the model may be used as a spatially distributed multiple input linear system; however, for daily and longer period rainfall-runoff modelling, the assumption of linearity is restrictive. For such modelling, Todini and Wallis (1977) introduce non-linearity into the model as follows using a threshold mechanism applied to the rainfall input. This they do for a single lumped rainfall input, although there is no reason why the procedure could not be applied to multiple inputs. If U_t denotes the lumped rainfall input, then the threshold mechanism is applied at time t as follows. An antecedent rainfall sum is computed as

$$AR_t = \sum_{w=1}^{L} U_{t-w}\ ; \qquad \ldots\ldots \quad (6.10)$$

if this sum does not exceed a preselected threshold T, then the rainfall input vector remains unchanged. If the threshold is exceeded, then the rainfall values $(U_t, U_{t-1}, \ldots U_{t-M})$ where $M<L$ are set to zero in the basic

rainfall input vector, and these values are transferred to the corresponding locations in a new input vector of the same length. At time (t+1), the same procedure is applied to the basic rainfall input vector as if it had remained unchanged, and so on. Thus, for one threshold, two inputs are generated from a single basic input; the multiple input capability of the basic CLS model is then utilized to derive the impulse responses for these inputs, and ultimately to derive a model output which has a non-linear relationship with the original input. The basic notion underlying the model is that different response regimes operate in a catchment in response to different states of catchment wetness, with the switch from one response to another achieved through the threshold, which introduces non-linearity into the model.

Todini and Wallis (1977) applied the CLS model with either one or two thresholds to the catchments used in the W.M.O. Intercomparison Study (1975) with success. They found that reasonable values of the threshold(s), and associated values of the parameters M and L could be found using a trial and error procedure involving a small number of model runs. More recently, Todini (personal communication) has replaced the three parameter threshold (T, M, L) with a two parameter version (T, K) where K is the parameter of an antecedent precipitation index computed at time t as

$$API_t = K\,API_{t-1} + U_{t-1} \; ; \qquad \ldots\ldots \quad (6.11)$$

if $API_t > T$, then the current rainfall value U_t is set to zero in the basic rainfall input vector, and U_t is transferred to the corresponding location in a second input vector. Further thresholds can be incorporated in a similar manner.

Further modifications to the CLS model have been incorporated during the present study. Seasonal variation has been introduced into the parameter K as follows

$$K_t = \bar{K} + \alpha \ \text{Sin}\left(\frac{2\pi t}{365}\right) + \beta \ \text{Cos}\left(\frac{2\pi t}{365}\right) \qquad \ldots\ldots \quad (6.12)$$

This allows greater flexibility in describing antecedent catchment conditions. An optimisation routine has been used to derive good estimates of the parameters, $\bar{K}$, α and β, and also of the threshold value(s) through minimising the quantity F^2 defined by equation (6.3). Thus, two optimisation procedures work in tandem; one to estimate the threshold and API parameters, and the quadratic programming routing to estimate the ordinates of the impulse response functions.

6.2.4 Catchments used

Some initial trials with the CLS model indicated that difficulty could be encountered in fitting the model to catchments where evaporation has a dominant effect on the runoff regime. It was decided accordingly to concentrate on catchments with moderate to large coefficients of runoff. The catchments used were selected from those listed in Table 6.13 of the Flood Studies Report (N.E.R.C. 1975) to satisfy the following criteria

(i) the gauging station should have an A1 rating in the Flood Studies Report (N.E.R.C. 1975)

(ii) a sufficient number of gauges should be located in and adjacent to the catchment; this was arbitrarily selected as 10.

Six catchments satisfying the above criteria were finally selected. Three of these are located in Devon, (the Exe at Stoodleigh, the Culm at Woodmill and the Dart at Austin's Bridge) and one each in Lancashire, (the Wyre at St Michael's), Yorkshire, (the Ure at Kilgram Bridge), and Northumberland (the North Tyne at Tarset). They are thus mostly located in the wet western areas of the country and all are moderately steep. Their geographical location can be seen in Figure 6.3 of the Flood Studies Report (N.E.R.C. 1975). Various characteristics of the catchments are listed in Table 6.1. A brief description of each catchment is given in Appendix (C).

The four year period, January 1966 - December 1969 was chosen over which to calibrate the CLS model on each catchment, with the exception of the Ure at Kilgram Bridge, for which the period January 1968 - December 1971 was chosen. Daily rainfall data for a set of gauges within and adjacent to each catchment were retrieved subject to the criterion that each gauge had operated for over 75% of the total period. This resulted in a considerable reduction in the number of gauges identified initially, with the result that data for only 7 gauges were available for the North Tyne catchment. The number of gauges available for each catchment is given in Table 6.1 and their location on the catchment maps is shown in Figures 6.1-6.6. Mean daily flows for the selected catchments had previously been obtained from the Water Data Unit.

6.2.5 Results

6.2.5.1 Model calibration

For calibration of the CLS model, the lumped daily rainfall input $\overline{U}_t$, $t = 1, 2, \ldots, N$ was calculated as the arithmetic mean daily rainfall over the total number of gauges on each catchment. Other procedures (e.g. calculation of Thiessen polygons) might have been used to assign weights other than unity to the individual gauges but these were not considered here.

For a CLS model with one threshold and a seasonal API parameter K_t, the parameter set consists of $\overline{K}$, α, β, T_1 and k_1 and k_2, the lengths of the two impulse responses, the ordinates of which are also required to be estimated; for two thresholds there will also be the second threshold T_2 and the length of the third impulse response k_3 and its ordinates. In calibrating the model, the parameters $\overline{K}$, α, β and T_1 and where relevant, T_2 were estimated through a standard non-linear optimization or 'hill climbing' routine; a brief description of the latter is given in O'Connell et al. (1970). The lengths of the impulse responses can be set to correspond to the points at which the ordinates become effectively zero; a few trial runs is usually sufficient to establish these, and they did not accordingly have to be optimized. The ordinates of the impulse responses were estimated through a quadratic programming routine; both of the constraints represented by equations (6.7) and (6.8) were applied.

In using the two optimization routines conjunctively, it was found that, as the ordinates of the impulse responses are subject to random variation, this created a locally irregular response surface for the 'hill climbing' routine, and quite large step sizes were required to ensure that the search did not terminate near the starting point. A number of different sets of starting values were used to try to ensure that a global optimum was found; however, an exhaustive exploration of the response surface was not made. In general, further searches were not initiated once no significant improvement in values of R^2 could be obtained.

Table 6.1 Details of six catchments

Station Number	Station Name	N.G. Ref.	Area (km^2)	Annual Average Rainfall (mm)	2 day 5 yr. Return Period Rainfall (mm)	Channel Slope (m/km)	Flood Studies Soil Index (Range 0.15-0.50)	No. of raingauges
23/5	North Tyne at Tarset	NY776861	285	1255	69.7	4.94	0.500	7
27/34	Ure at Kilgram Bridge	SE190860	510	1390	84.1	4.10	0.472	14
45/2	Exe at Stoodleigh	SS943178	422	1402	84.3	5.70	0.326	18
45/3	Culm at Woodmill	ST021058	226	965	67.6	5.93	0.276	12
46/3	Dart at Austins Bridge	SX751659	248	1821	110.0	6.50	0.361	21
72/2	Wyre at St Michaels	SD465411	275	1257	69.9	3.69	0.458	19

The results of the CLS model calibrations for the six catchments are given in Table 6.2, where runoff coefficients for the total period of record, defined as

$$\phi = \frac{\sum_{t=1}^{N} Y_t}{\sum_{t=1}^{N} \bar{U}_t} \qquad \ldots\ldots \quad (6.13)$$

are also presented; these are obviously independent of calibration. The North Tyne at Tarset and Ure at Kilgram Bridge both have almost the same high runoff coefficient even though the periods of record are not wholly concurrent; the Dart at Austin's Bridge has a runoff coefficient of a similar order. The Exe at Stoodleigh ($\phi = 0.746$) contrasts with its neighbour, the Culm at Woodmill ($\phi = 0.535$); the Wyre at St Michael's ($\phi = 0.662$) has a value intermediate to these. While the six catchments all lie in the wetter areas of the country, a reasonable range in percentage runoff is still observed.

For five of the six catchments, two thresholds T_1 and T_2 were employed in the CLS model; for the remaining catchment, the Dart, no significant improvement could be achieved with a second threshold. The threshold(s) were applied to the antecedent precipitation index API_t defined by equation (6.11) with seasonally varying parameter K_t described by equation (6.12). The optimized values of the harmonic coefficients α and β all imply larger values of K_t during winter than summer. In that the index API_t seeks to track the changes in soil moisture deficit (SMD) over the year smaller values of K_t in summer would be consistent with larger day to day changes in SMD. As might be expected higher values of the thresholds are associated with higher values of the parameter $\bar{K}$ and average annual rainfall. If at any point during the optimization of the parameters $\bar{K}$, α and β, the value of K_t was found to be $\leqslant 0$ or $\geqslant 1$, then K_t was set to 0.001 or 0.999, respectively.

The impulse responses may be thought of as corresponding to different runoff regimes within a catchment; the first, second and third impulse responses will *invariably have increasing associated coefficients of runoff.* Examples of the impulse responses obtained are given in Figures 6.7–6.8 for two catchments, the Exe and the Ure. For the Exe, the impulse responses are smooth on the whole; the magnitude of the ordinates for the third impulse response suggests that very high runoff rates do not occur on this catchment. For the Ure, the impulse responses are much more erratic; this could be attributable to a basically erratic catchment response to rainfall, or errors in the input and/or output data. The impulse responses obtained for the Tyne were somewhat more erratic, and for the Wyre, somewhat less so. The impulse responses for the Culm and Dart were essentially smooth.

The coefficients of determination for daily flows R^2_d, range from 0.659 for the Ure to 0.789 for the Dart; while the monthly values R^2_m range *from 0.661 for the Culm to 0.889 for the Dart.* The values of R^2_m were computed by aggregating observed and computed daily flows, rather than fitting the CLS model at the monthly level. The values of R^2_m are greater than R^2_d, except in the case of the Exe and Culm; however, in comparing values of R^2_m and R^2_d, it should be remembered that the former has much larger sampling variability than the latter. The lowest values of R^2_d were obtained on the Ure and the Tyne, which were those with the most erratic impulse responses.

Table 6.2 **Runoff coefficients, parameter values (excluding impulse response ordinates), and monthly and daily coefficients of determination for fitted CLS model**

Catch. No.	Period of Record	$\bar{K}$			T_1(mm)	T_2(mm)	k_1	k_2	k_3	R_m^2	R_d^2	
23/5	1966-1969	0.849	0.042	0.061	32.80	80.0	20	10	5	0.760	0.666	0.838
27/34	1968-1971	0.800	0.120	0.200	34.00	180.0	12	20	5	0.777	0.659	0.839
45/2	1966-1969	0.599	0.106	0.176	31.30	70.0	20	25	10	0.780	0.786	0.746
45/3	1966-1969	0.700	0.057	0.207	36.70	80.0	20	20	5	0.661	0.733	0.535
46/3	1966-1969	0.904	0.004	0.094	76.00	-	25	20	-	0.889	0.789	0.853
72/2	1966-1969	0.843	0.011	0.120	28.90	83.0	20	10	10	0.836	0.737	0.662

6.2.5.2 Raingauge network requirements

The effect of raingauge density on R^2_m and R^2_d was assessed using the procedure described in Section 6.2.1. Values of e_i were computed for each of the P gauges on a catchment, and these were used to determine the order of deletion of the gauges; for the six catchments this is given in Table 6.3. Reference to Figures 6.1-6.6 shows that the deletion orders obtained were reasonable from a practical viewpoint in that gauges outside and adjacent to the catchment boundaries were generally deleted first. It may not always happen that the gauges most representative of mean rainfall lie within the boundaries.

By successively deleting the gauges for each catchment in the order given in Table 6.3, recomputing the mean rainfall for the remaining (P-r) gauges, r = 1, 2, ...(P-1), and calculating the CLS model output, Y_t*, values of R^2_m and R^2_d were calculated for each value of (P-r); these values were calculated on the basis that for the output from P gauges, Y_t, $R^2_m = R^2_d = 1$. Plots of R^2_m and R^2_d against the number of gauges for the six catchments are given in Figures 6.9-6.14. These plots all indicate a decrease in R^2_m and R^2_d with a decreasing number of gauges; however, there is variation in the rates of decrease and in the smoothness of the plots.

Table 6.3 Order of deletion of raingauges for six catchments

Order of Deletion	Catchment Number 23/5	27/34	45/2	45/3	46/3	72/2
1	603649	53904	393315	357633	362452	573796
2	916382	47474	356262	358549	363474	573426
3	915054	61249	397532	357983	365962	573338
4	12213	53253	355639	358100	363697	573345
5	10774	57403	393432	402190	365999	578833
6	9519	48001	357633	354352	366134	582925
7	-	53511	392894	358942	363624	579720
8	-	51065	401620	353964	361123	575649
9	-	52286	356984	358232	362187	576634
10	-	48417	397445	358511	369919	578682
11	-	57427	355650	358300	368633	583359
12	-	49132	357435	-	363871	575752
13	-	48217	397713	-	365974	579611
14	-	-	357187	-	365974	576577
15	-	-	401668	-	366105	577793
16	-	-	393059	-	361957	577881
17	-	-	356466	-	368714	579384
18	-	-	-	-	364177	578008
19	-	-	-	-	368715	-
20	-	-	-	-	362061	-
Remaining Gauge	8887	49210	356616	358670	363294	577975

For the Exe, the Culm and the Dart, the plots are essentially smooth and R^2_d reduces to very similar values for one gauge (0.936, 0.929, 0.939), with similar behaviour observed for values of R^2_m which are generally higher.

These results indicate that, if the CLS model is taken as providing a good representation of catchment behaviour, then the predictive ability of the model is not greatly affected by the number of gauges used to compute the lumped input. The rate of decrease of R^2_d and R^2_m increases progressively as the number of gauges approaches one; this starts to occur for approximately 8 gauges in the case of the Dart, and 5 gauges in the cases of the Culm and Exe, corresponding to approximate densities of 1 gauge per 31 km^2, 45 km^2 and 80 km^2 respectively. For larger numbers of gauges, the values of R^2_d and R^2_m observed suggest that stability of the input has been achieved as far as output prediction is concerned. Thus, for these three catchments, there is no obvious suggestion that the density of gauges used in calibrating the model is not adequate, provided that the fit of the CLS model is acceptable in the first place. If not, then to increase the values of R^2_m and R^2_d, further improvements to the model would be required, or a new more complex model adopted, which, to get a much better fit, might have to resort to spatially distributed inputs. If the ultimate decision to be made on the basis of model prediction is not likely to be sensitive to this improved accuracy, then the added complexity would not be justified. If, however, the extra accuracy was required, and a model was capable of providing it, then the input requirements of such a model could be more stringent than the CLS model used here.

The above considerations might also be applied to the Wyre catchment; however, the value of R^2_d for one gauge (0.905) is somewhat less than that for the Exe, Culm and Dart, and the value of R^2_d drops off more quickly initially. An increase in the rate of R^2_d occurs around 5 gauges; the corresponding approximity density (1 per 55 km^2) would appear to be adequate for fitting the CLS model.

The results for the North Tyne (Figure 6.9) suggest that seven gauges (approx. density 1 per 41 km^2) do not provide a stable estimate of the input to the CLS model. There is a relatively rapid decrease in R^2_d and R^2_m with decreasing gauge number, with a value of $R^2_d = 0.780$ observed for one gauge. For the case of the Ure, a similar result is observed. From $R^2_d = 1$ for a total of 14 gauges (approx. density 1 per 36 km^2), the value of R^2_d drops to 0.822 for one gauge. (Figure 6.10). The relationship between R^2_d and the number of gauges is somewhat erratic; for example, there is the suggestion that the selected sub-set of four gauges provides better input definition than the sub-sets of 5, 6, 7 and 8 gauges. As previously noted, the impulse responses observed for the North Tyne and Ure were also noisy, which may reflect the variability in AAR noted in Appendix (C) for both catchments, although their contrast with the other catchments in this respect was not remarkable.

6.2.6 Discussion and Conclusions

The extent of the experiments reported on in Section 6.2.5 is of necessity limited; had time permitted, the study could have included an exploration of the following questions, among others:

(i) the effect of different gauge sub-set selection procedures on the relationship between R^2 and the number of gauges;

(ii) the effect of different methods of estimating mean areal rainfall on the relationship between R^2 and the number of gauges;

(iii) the relationship between R^2 and number of gauges that would be obtained for a period of record independent of that on which the model was calibrated.

Having calibrated a rainfall-runoff model, a split-sample test of the model's predictive ability is frequently carried out (e.g. W.M.O. 1975). It would have been of interest to apply such a test to the CLS model in the case of each catchment studied; however, because of the statistically efficient parameter estimation procedures used in estimating the ordinates of the impulse responses, it is to be expected that the difference observed between values of R^2 obtained from a calibration period and a test period would be minimized (Todini and Wallis, 1977). Nonetheless, it would be of interest to observe if the relationship between R^2 and number of gauges for a catchment varied from one period of record to another.

On the basis of the experiments which have been carried out, some tentative conclusions may be drawn about the effect of raingauge density on the accuracy of streamflow prediction. A much more extensive study involving a much larger number of catchments, a number of different models and different methods of defining sub-sets of gauges would be required before general network density requirements could be established for streamflow prediction. The results obtained here suggest that such requirements would be likely to vary depending on the rainfall-runoff regime to be modelled, and the variability of the rainfall over the catchment in question. No obvious classification can be attempted on the basis of the present results; those obtained for the Exe, Culm, Dart and Wyre suggest that the present network density would be adequate for modelling studies of the type carried out here. However, the results for the Ure and Tyne suggest that higher densities than are currently available would be required to provide adequate definition of lumped inputs for the CLS model.

The study carried out could be expanded on considerably, and a number of alternative approaches could be tried. Detailed consideration of suggestions for further work is deferred to Section 8.

6.3 Raingauge network requirements for real-time flow forecasting

6.3.1 Introduction

The aim here is to examine in detail the influence of an existing gauge network on the calibration of a simple rainfall-runoff model, and its consequent implication for the design of raingauge networks for real-time flow forecasting. The questions considered are:

(i) How many gauges are required to attain a given accuracy of forecast for a given lead time?

(ii) Can the number required be reduced by considering spatially distributed inputs, rather than a lumped input?

(iii) Is it better to use only those gauges transmitting data in real-time to calibrate the model, or all gauge records which more probably will estimate the 'true' catchment average rainfall?

(iv) What is the optimum gauge location or configuration?

The approach adopted differs markedly from that of Bras and Rodriguez-Iturbe (1976b)(see Section 2.4.2) who assume a perfect rainfall-runoff model and uncertainty in the rainfall input. Here it is shown by direct statistical inference that a causal relation exists between rainfall measured at each gauge and the river flow. The strength of this relationship will indicate the relative importance of each gauge in determining the flow response. Linear models relating one or more of these gauges (treated as lumped or distributed inputs) to river flow are identified (Section 6.3.4) and estimated (Section 6.3.5), allowing for the fact that the model is not a perfect representation of reality. Implications for raingauge network design and rainfall-runoff modelling are obtained in Sections 6.3.5.3 and 6.3.5.4, and qualified conclusions made in Section 6.3.6.

6.3.2 Data

The analysis is based on rainfall and runoff data for the 33.9 km^2 Hirnant subcatchment, situated south-east of Bala Lake in the River Dee catchment in North Wales. Impervious rocks, providing very little storage for precipitation, combine with steep slopes to give a fast streamflow response to rainfall. The river is gauged at Plas Rhiwaedog at a natural river section, and flow data at half-hour intervals are held at IH for the period January 1970 to August 1974.

Rainfall data from a dense network of recording Plessey tipping bucket raingauges set up over the Dee catchment as part of the Dee Weather Radar Project were used. Six gauges in and around the Hirnant catchment were selected from this network, and their locations are mapped in Figure 6.15. Concurrent flow and rainfall data are available for the two year period beginning 1 July 1972. The rainfall data however were found to be highly discontinuous with many periods where data were missing at one or more gauges (Figure 6.16). The longest period when all gauges were recording and which included a suitable storm event for rainfall-runoff modelling was the 39 days from 12.30 hrs on the 20 July to 15.00 hrs on the 28 August 1973, comprising 1878 half-hourly values. Time series plots of flow, and measured rainfall at each gauge are given in Figures 6.17 and 6.18, and Table 6.4 provides summary statistics of the series.

Table 6.4 Summary statistics for the Hirnant catchment

Series	Units	Mean	Variance	Minimum	Maximum
Flow	mm/hr	.1416	.1036	.0264	3.593
Gauge 18	mm/$\frac{1}{2}$hr	.0897	.1973	0	9.6
Gauge 20	"	.0613	.0988	0	4.8
Gauge 21	"	.1048	.3698	0	17.0
Gauge 23	"	.1023	.2738	0	9.2
Gauge 24	"	.0830	.1895	0	8.8
Gauge 25	"	.0880	.4069	0	16.8

The time series plots show that the period encompasses two main storm events, the first being of high intensity and short duration compared to the second. The spatial variability of the rainfall is very different for the two storms, the first varying in maximum intensity from 1.5 mm/$\frac{1}{2}$hr at gauge 20 to 17 mm/$\frac{1}{2}$hr at gauges 21 and 25. The peak intensity for the second storm shows remarkable spatial uniformity, all gauges measuring quantities within the range 5 to 6 mm/$\frac{1}{2}$hr. This change in spatial variability with storm type may complicate the problem of choosing a network configuration that is optimal for flow forecasting purposes.

6.3.3 Model specification

A model is required to relate measurements of river flow, y, made at discrete equidistant intervals of time, to measurements of rainfall, u, falling over this same time interval. Measured river flow at time t, y_t, is assumed to be the sum of a deterministic component $\tilde{y}_t$, and a stochastic component, η_t, so that

$$y_t = \tilde{y}_t + \eta_t \quad \ldots . \quad (6.14)$$

The $\tilde{y}_t$ component is that part of y_t which is causally related to present and past measured rainfall by the linear model

$$\tilde{y}_t + \delta_1 \tilde{y}_{t-1} + \ldots + \delta_r \tilde{y}_{t-r} = \omega_o u_{t-b} + \omega_1 u_{t-b-1} + \ldots + \omega_s u_{t-b-s} \quad \ldots . \quad (6.15)$$

where b denotes the delay in river flow response to rainfall, and is here called the *pure time delay*. Introducing the backward difference operator, B, defined as

$$B^b u_t \quad u_{t-b} \quad \ldots . \quad (6.16)$$

the linear model may be expressed in notationally efficient difference equation form

$$\delta(B)\tilde{y}_t = \omega(B) u_{t-b} \quad \ldots . \quad (6.17)$$

where $\delta(B)$, $\omega(B)$ are polynomials in B of degree r and s respectively, i.e.

$$\delta(B) \equiv 1 + \delta_1 B + \delta_2 B^2 + \ldots + \delta_r B^r$$

$$w(B) \equiv \omega_0 + \omega_1 B + \omega_2 B^2 + \ldots + \omega_s B^s$$

Equation (6.17) is called the *process model*. Note that the measured rainfall is assumed to be a deterministic input variable which is causally related to the deterministic part of the measured river flow.

The additive noise component, η_t, is attributed to the disturbance effects of model and measurement errors and is referred to as the *process noise*. Model errors include the omission of other system inputs that influence the output, as well as any inadequacy in the form of the deterministic model structure assumed above. Note that the variance of the noise component, η_t indicates the relative strength of the causal relation between rainfall and runoff. If the rainfall input, u_t, is defined as the average of several gauges, the effect of the number of gauges on the noise variance can be observed. This may aid a decision as to the 'best' number of gauges required for flow forecasting.

In general the process noise, n_t, will not form a white noise sequence (an independent sequence of random variables) but a reasonable assumption is that it may be related to white noise by the difference equation

$$\phi(B)\eta_t = \theta(B)a_t \qquad \ldots. \qquad (6.18)$$

where $\phi(B) \equiv 1 + \phi_1 B + \phi_2 B^2 + \ldots.. + \phi_p B^p$

$$\theta(B) \equiv 1 + \theta_1 B + \theta_2 B^2 + \ldots.. \quad \theta_q B^q$$

The white noise sequence, a_t, is assumed to have zero mean and variance σ_a^2, and to be independent of the input, u_t, i.e.

$$E\{a_t\} = 0, \; E\{a_t a_{t+k}\} = \begin{cases} \sigma_a^2 & k=0 \\ 0 & \text{otherwise} \end{cases}, \quad E\{a_t u_{t+k}\} = 0 \text{ for all } k$$

The component relating process noise, η_t, to white noise, a_t is called the *noise model*, and is defined by (6.18).

The *composite model* obtained by combining the process and noise models using (6.14) and depicted in Figure 6.19, is completely defined by

$$y_t = \delta^{-1}(B)\omega(B)u_t + \phi^{-1}(B)\theta(B)a_t \qquad \ldots. \qquad (6.19)$$

This is used in later sections for single-gauge and lumped rainfall-runoff modelling.

The process model of (6.17) is easily extended to the case where each of P gauges are considered as separate inputs, rather than as an averaged (lumped) input. This multi-input single-output process model may be represented succinctly by

$$\delta(B)\tilde{y}_t = \sum_{i=1}^{P} \omega_i(B)u_{i,t-b_i} \qquad \ldots. \qquad (6.20)$$

where $u_{i,t-b_i}$ denotes the measured rainfall at gauge i at time t less b_i time units, and $\omega_i(B)$ is a polynomial as defined in (6.17) but of order r_i. The *spatially-distributed model* is then represented by

$$y_t = \delta^{-1}(B) \sum_{i=1}^{P} \omega_i(B)u_{i,t-b_i} + \phi^{-1}(B)\theta(B)a_t \qquad \ldots. \qquad (6.21)$$

For convenience the mean values of the output and input series have been ignored for clarity of exposition, and it is assumed throughout that y_t and u_t are the deviations from their respective mean values.

The process model may be improved by considering transformation of the rainfall input series to account for soil moisture deficit variations, and seasonal fluctuations in evaporation, which affect the flow response (Moore, 1977). Due to the short data sequence employed, and the objective of assessing raingauge requirement, such refinements were not included.

The following two sections describe how process models for both single-gauge and multi-gauge inputs are identified, and their parameters estimated. Attention is focussed on the number of raingauges required to explain variations in flow, and consequently little attention is given to identification and estimation of the noise model. The assumed structure of the noise series (6.18) however is an important consideration in obtaining consistent and unbiased parameter estimates of the process model (Section 6.3.5).

6.3.4 Process model identification

6.3.4.1 Problem statement

The spatially distributed model of (6.21) and its single gauge equivalent (6.17) represents too broad a category of models to be fitted directly, and techniques are required to isolate a more likely subset of models worthy of further attention. Identification of the process model involves making a choice as to which past flow and rainfall measurements are to be included in the model; specifically this involves identifying the order, r and s, of the polynomials $\delta(B)$, $\omega(B)$, together with the pure time delay, b. Process model identification consequently involves examining the causal relation between rainfall and runoff.

The interrelationship between the two time series u_t and y_t may be investigated using the cross-correlation function (XCF), $\rho_{uy}(.)$. But because both the output, y_t, and input, u_t, are autocorrelated (Figures 6.20(a) and 6.21(a)), the ordinates of the estimated XCF, $r_{uy}(.)$, between u_t and y_t have large variance, and correlation exists between estimates at different lags. Consequently, even if u_t and y_{t+k}, $k = 0, \pm 1, \pm 2, \ldots\ldots$, are not cross-correlated, the autocorrelation of each could give spuriously significant cross-correlation estimates. The effect of autocorrelation in both u_t and y_t in obscuring the underlying causal relation between rainfall and runoff is illustrated by the sample XCF between gauge 20 and flow in Figure 6.22. Laugh and Box (1977) suggest that the interpretive power of the XCF may be enhanced by considering the XCF of the 'whitened' y and u sequence, denoted a_y, a_u where whitening is performed by the *stochastic models* (Figure 6.19)

$$\phi_y(B)y_t = \theta_y(B)a_{y,t}$$
$$\phi_u(B)u_t = \theta_u(B)a_{u,t} \qquad \ldots . \quad (6.22)$$

Note that these stochastic models are of the same form as the noise model, the polynomials being similarly defined, and $a_{y,t}$, $a_{u,t}$ are white noise. An equivalent representation of (6.22) is the infinite moving average form

$$y_t = \psi_y(B)a_{y,t} + a_{y,t}$$
$$u_t = \psi_u(B)a_{u,t} + a_{u,t} \qquad \ldots . \quad (6.23)$$

where $\psi_y(B) = \sum_{i=1}^{\infty} \psi_{yi}B^i$, $\psi_u(B) = \sum_{i=1}^{\infty} \psi_{ui}B^i$

The white noise sequences $a_{y,t}$, $a_{u,t}$ may be related linearly by

$$a_{y,t} = v(B)a_{u,t} + \eta'_t \qquad \ldots . \quad (6.24)$$

where $v(B) = \sum_{k=-\infty}^{\infty} v_k B^k$, and η'_t is a noise sequence independent of $a_{u,t}$.

The ordinates of the cross-correlation function between $a_{y,t}$, $a_{u,t}$ are then proportional to $v(B)$, since

$$\rho_{a_u a_y}(k) = E(a_{u,t}a_{y,t+k})/\sigma_{a_u}\sigma_{a_y}$$

$$= E(v(B)a^2{}_{u,t} + a_{u,t}\,\sigma'_t)/\sigma_{a_u}\sigma_{a_y}$$

$$= v(B)\sigma_{a_u}/\sigma_{a_y} \qquad \ldots. \qquad (6.25)$$

This XCF may be interpreted as the correlation between y_t and u_t after allowance has been made for the correlation of y_t and u_t with their previous values; it is therefore closely analagous to the partial cross-correlation function (Jenkins and Watts 1968). The ordinates of $\rho_{a_u a_y}(\cdot)$ are used to select which past rainfalls significantly affect the present output, y_t, thereby identifying the lumped and distributed process models.

6.3.4.2 <u>Identification procedure</u>

The above provides a four-stage procedure for process model identification as follows:

(a) Identification and estimation of stochastic models (6.22) for the streamflow series, and each rainfall series, to produce estimates of the whitened flow series, $a_{y,t}$, and each whitened rainfall series, $a_{u,t}$, required at stage (ii).

(ii) Estimation of the XCF, $\rho_{a_u a_y}(\cdot)$, between each whitened rainfall series, and the whitened flow series estimated at stage (i). This is used to identify which past rainfall measurements significantly influence the streamflow, y_t. Each single-gauge process model is identified by its order (r, s, b) indicating that flow, y_t, is related to r previous values, y_{t-1}, y_{t-2},..... y_{t-r}, and to s previous values of rainfall u_{t-b}, u_{t-b-1},...... $u_{t-b-s-1}$, where b is the pure time delay.

(iii) Estimation of single-gauge process models identified at stage (ii) and their associated process noise series, η_t, (see Section 6.3.5.2).

(iv) The variance of the process noise for each single-gauge process model obtained at stage (iii) is used to rank the gauges in order of importance with regard to explaining the variation in flow. Spatially distributed models of increasing complexity and their lumped model equivalents are identified by including new gauges in accordance with this ranking. Which past rainfalls are to be included for a particular gauge is decided using the results obtained at stage (ii) (see Section 6.3.5.3).

Note that stages (i) and (ii) complete the identification of the single-gauge process models. Stages (iii) and (iv) more properly belong to the parameter estimation stage of model building, and discussion of them is deferred to Section 6.3.5. They are included here because they facilitate identification of the order in which gauges are to be included in the lumped and spatially distributed models.

The need to model each component of a multivariate time series model separately has been suggested by Parzen (1969). Also Jenkins (1974) and Granger and Newbold (1976) adopt a similar approach to multivariate model identification. In the present context the approach has the additional advantage of providing an objective procedure for deciding which gauge has the strongest causal relationship to streamflow. It allows gauges to be included in the spatially-distributed model in order of importance, using the variance of the process noise from each single-gauge process model as the criterion of selection. The procedure however ignores the problem of multicollinearity when, for example, the optimal three gauge set does not necessarily include one or both of the optimal two gauge set. A procedure to circumvent this problem using techniques already described here is possible, but would prove too lengthy a task.

6.3.4.3 Application of the identification procedure

Stage 1: *Identification of the stochastic models of the flow sequences*, and the rainfall sequences at each gauge was performed using the autocorrelation function (ACF), and partial autocorrelation function (PACF), following the approach of Box and Jenkins (1970). Examples of both the ACF and PACF are given in Figures 6.20 and 6.21, and the properties of *the final fitted* models are summarised in Table 6.5. Preliminary identification of the stochastic flow model using the ACF and PACF suggested a (1,2) model, using the notation (p.q) where p and q are the orders of the polynomials $\phi_y(B)$, $\theta_y(B)$ respectively, in (6.22). This proved inadequate in that $\hat{a}_{y,t}$ did not approximate white noise. Surprisingly a (3,3) model was also inadequate and recourse to a (7,0), suggested as feasible by the PACF cutting off after lag 7, was finally adopted. This structure is unusual in hydrological experience, but it was felt statistically justified to adopt it for identification purposes. The stochastic models for rainfall were identified as (1,2) models, except for gauge 21 where a (3,1) model was appropriate. Parameter estimation for the seven identified models was performed using a recursive Approximate Maximum Likelihood algorithm (Young et al., 1971; Moore, 1977) and an estimate of the residual or innovation sequence was obtained for each model, $\hat{a}_{y,t}, \hat{a}_{u_1,t}, \ldots \hat{a}_{u_6,t}$.

Stage 2: The sample XCF between the flow residual, and each gauge residual is given in Table 6.6 up to lag 10; no values for negative lags are given, all being insignificantly different from zero, in agreement with the obviously one-way causal relation between rainfall and runoff. The significant sample cross-correlations in Table 6.6, indicate which previous rainfalls are to be included in the single-gauge, lumped, and spatially distributed models. On this basis, the number of preceding rainfall values to be included was 2 in all cases except for gauge 18 where s = 1. The pure time delay, b, equalled 2 indicating a delay between $\frac{1}{2}$ and 1 hour between rain and the flow response, except for gauges 21 and 24 where b = 3.

One previous flow measurement was *included in all models* (i.e. $r = 1$, $\delta(B) = 1 + \delta_1 B$), and thus the importance of each gauge in determining the flow is uninfluenced by different orders of $\delta(B)$ in different models. A lag 1 autoregressive component concurs with the observation that the hydrograph recession is approximately exponential. Table 6.7 summarises *the selected single-gauge process models* using the notation (r, s, b) to indicate the model structure.

6.3.5 Parameter estimation for the process model

6.3.5.1 Introduction

Having identified the structure of the process models relating each gauge to river flow, attention may now be turned to process model parameter

Table 6.5

Stochastic models for flow, and all gauges

(i) **The (7,0) stochastic flow model**

$$(1-1.45B + .30B^2 + .37B^3 - .34B^4 + .07B^5 - .02B^6 + .09B^7)y_t = a_{y,t}$$

$$\bar{a}_y = -1.7 \times 10^{-6}, \quad \hat{\sigma}^2_{a_y} = 7.36 \times 10^{-4}$$

(ii) **The (1,2) rainfall models**

Gauge No.	Parameters ϕ_1	θ_1	θ_2	Innovation statistics $\bar{a}_u$	$\hat{\sigma}^2_{a_u}$
18	-.90	-.38	-.12	-.0037	.1056
20	-.92	-.32	-.01	-.0017	.0264
23	-.91	-.15	-.37	-.0060	.1167
24	-.89	-.19	-.25	-.0009	.0870
25	-.80	-.16	-.43	-.0054	.2309

(iii) **The (3,1) rainfall model for gauge 21**

$$(1-.21B - .11B^2 - .34B^3)u_t = (1 + .20B)a_{u,t}$$

$$\bar{a}_u = .0021, \quad \hat{\sigma}^2_{a_u} = .2389$$

Table 6.6

Sample cross-correlation functions, $r_{a_u a_y}(.)$, between the flow residual and each gauge residual (approximate two standard errors = $2/\sqrt{N}$ = 0.46)

Lag / Gauge	0	1	2	3	4	5	6	7	8	9	10
18	.05	.04	.28	.09	-.03	.11	.03	.06	.01	.02	.11
20	-.03	-.04	.15	.19	-.02	.17	.16	-.09	.08	.03	.20
21	-.04	.18	.01	.44	.29	-.18	-.02	.04	-.12	-.02	.00
23	.06	-.04	.35	.42	.06	-.17	.12	.03	-.09	.09	-.07
24	.05	.02	.07	.37	.32	-.06	-.08	.23	-.06	-.09	-.03
25	-.00	-.01	.16	.58	.03	-.20	.03	.03	-.12	.04	-.01

Table 6.7

Identified Structure of Process models relating flow to each gauge

	Gauge Number					
Model	18	20	21	23	24	25
(r, s, b)	(1, 1, 2)	(1, 2, 2)	(1, 2, 3)	(1, 2, 2)	(1, 2, 3)	(1, 2, 2)

estimation. The lumped model may be considered a special case of the more general spatially-distributed model (6.21), which may be written in vector form as

$$y_t = \underset{\sim}{x}_t^T \underset{\sim}{\theta} + n_t \qquad \dots \qquad (6.26)$$

where

$$\underset{\sim}{x}_t^T = \left[y_{t-1}, y_{t-2}, \dots\dots, y_{t-r}, u_{1,t-b_1}, \dots, u_{1,t-b_1-s_1}, \dots, u_{n,t-b_n}, \dots u_{n,t-b_n-s_n}\right]$$

and

$$\underset{\sim}{\theta}^T = \left[\delta_1 \; \delta_2 \; \dots\dots \; \delta_r \; \omega_{10} \; \omega_{11} \cdots \omega_{1s_1} \cdots \omega_{n0} \cdots \omega_{ns_n}\right]$$

The least squares solution of (6.26) is inconsistent and biassed because in general the process noise, n_t, is both autocorrelated, and not independent of $\underset{\sim}{x}_t$ due to the autoregressive terms $y_{t-1}, y_{t-2}, \dots, y_{t-r}$. An unbiased estimate of the parameter vector, θ, may however be obtained by the *instrumental variable* procedure (Young et al., 1971; Moore, 1977). The estimation problem posed by the spatially distributed model is overcome by a simple extension of the instrumental variable algorithm to handle multiple lagged deterministic inputs. The reader is referred to the references previously cited for further details of the technique.

Once the model parameters have been estimated an estimate of the process noise is obtained from

$$\hat{n}_t = y_t - \hat{\delta}^{-1}(B)\hat{\omega}(B)u_t \qquad \dots \qquad (6.27)$$

It was noted in Section 6.3.3 that the variance of the noise component is indicative of the causal relation between flow and each gauge. This is conveniently expressed using the R^2 statistic defined by $(\sigma^2_y - \sigma^2_n)/\sigma^2_y$ where σ^2_y, σ^2_n are the variances of the flow series ($\sigma^2_y = .1036$), and process noise series respectively. This is the same R^2 used in daily modelling, and defined in equation (6.2). Note that 100 x R^2 indicates the percentage of the variation in flow explained by the process model. The R^2 values, and later the percentage error in the peak, obtained from the process model are used to indicate at what level increasing the number of gauges fails to provide appreciable additional information to explain the observed variation in flow. They should not be used to assess the performance of the model for real-time flow forecasting, as the process model taken in isolation is a sub-optimal forecasting algorithm. Only when combined with the noise model can one assess how well the model will perform in real-time; and a different performance criterion, such as the variance of the one-step ahead prediction error, is used. Note also that the noise sequence, n_t, is generated given the rainfall sequence, u_t, and the first flow value, y_1, only. In real-time, estimates of flow are replaced by their observed values as they become available; this is not done in calculating n_t.

6.3.5.2 The single-gauge process models

The parameter estimates, $\hat{\delta}(B)$, $\hat{\omega}(B)$, together with the summary statistics of the process noise are given in Table 6.8 for each single-gauge process model.

Table 6.8 Single-gauge process models

Gauge No.	Model (r, s, b)	Parameters δ_1	ω_0	ω_1	Process Noise Mean	Variance	R^2	% error in peak
18	(1, 1, 2)	-.968	.037	-	-.00016	.02580	.75	+54
20	(1, 2, 2)	-.957	.032	.042	-.00017	.01108	.89	+27
21	(1, 2, 3)	-.944	.031	.020	-.00020	.02003	.81	+44
23	(1, 2, 2)	-.941	.030	.031	-.00012	.01177	.89	+31
24	(1, 2, 3)	-.945	.041	.025	-.00020	.01617	.84	+38
25	(1, 2, 2)	-.956	.013	.032	-.00009	.02318	.78	+48

R^2 ranges from .75 (i.e. 75% explained variance) for gauge 18 to .89 for gauge 20, with the order of the gauges indicated as 20, 23, 24, 21, 25, and 18. The same order is obtained ranking the gauges according to the error in the peak, which ranges from 54% for gauge 18, to 27% for gauge 20. Discussion of the results is deferred to the next section, where the single-gauge and spatially distributed results are related.

6.3.5.3 The spatially distributed process models

Spatially distributed models were fitted starting with the two gauge models using gauges 20 and 23, and then including new gauges in the order of importance found in the previous section. This ignored the possible problem of multicollinearity mentioned in Section 6.3.4.2. The fitted process models are summarised in Table 6.9 together with the statistics of their associated noise sequences. The two-gauge model gave the highest R^2 value of .91, only .2 higher than the single-gauge model using gauge 20. The addition of further gauges had the effect of slightly increasing R^2, although this may possibly be attributed to parameter estimation inaccuracy. Comparing the R^2 values obtained for single-gauge and distributed models (Tables 6.8, 6.9) it appears that only one or two gauges are sufficient to forecast the flow sequence in question. Even if the gauges were poorly sited an R^2 of .75 for gauge 18 suggests forecasts may still be adequate. This conclusion may however be misleading, because of the criterion employed, which sums the squares of the errors over the whole sequence. For large parts of the series no rainfall occurs and the model is wholly autoregressive on the river flow. Rainfall affects primarily the rising limb and peak of the hydrograph, and consequently attention must be concentrated here when the efficacy of the raingauge network is of prime concern. The error in peak forecasting by the spatially distributed model is remarkably consistent at 24%, irrespective of the number of gauges used. For the single-gauge process models, however, Table 6.8 indicates the error ranges from 27% to 54%. Selecting the correct gauge for forecasting is consequently of greater importance than the R^2 criterion at first suggests. Although the 27% error in peak may appear high, remember that this should not be used to assess the potential performance of the model for real-time flow forecasting (see Section 6.3.5).

Table 6.9 <u>The Spatially Distributed Process Models</u>

No. of Gauges	δ_1	ω_o 18	ω_1 18	ω_o 20	ω_1 20	ω_o 21	ω_1 21	ω_o 23	ω_1 23	ω_o 24	ω_1 24	ω_o 25	ω_1 25	Process Noise Mean	Process Noise Variance	R^2	% Peak Error
1	-.957			.032	.042									-.00017	.01108	.893	+27
2	-.938			.012	.011			.024	.027					-.00020	.00935	.910	+24
3	-.936			.011	.011			.025	.020	-	.012			-.00024	.00970	.906	+24
4	-.935			.011	.012	.006	.003	.023	.015	-.003	.011			-.00026	.01002	.903	+24
5	-.942			.014	.012	-	-.003	.037	-.009	-.005	.010	-.012	.030	-.00024	.00981	.905	+24
6	-.940	.012		.005	.012	.001	-.003	.034	-.010	-.005	.010	-.011	.031	-.00027	.01037	.900	+24

6.3.5.4 The lumped process models

To investigate whether treating the raingauge measures as spatially distributed inputs, rather than as a single lumped input has any advantage, the arithmetic average of the rainfall at P gauges was computed. A (1,2,2) process model was fitted for P = 2, 3, 4, 5, 6, selecting the gauge to be included next according to the importance ranking obtained in Section 6.3.5.1. In terms of R^2, little advantage appeared to accrue using the distributed model (compare Tables 6.10, 6.11). In fact the two-gauge lumped model performed marginally better than its distributed equivalent; this was the 'best' process model obtained. But in terms of forecasting the peak, the distributed model performed consistently better for models using more than 3 gauges, only the two-gauge model predicting less well. These poor peak forecasts from the lumped model when several gauges are used favour adoption of distributed models, but the additional effort required to fit such models must be borne in mind.

What the single-gauge, lumped, and distributed model results reveal is that it may be possible to forecast flow more effectively using one or two carefully selected gauges than using a larger number of gauges to define the input.

Table 6.10 The Lumped Process Models

No. of Gauges	Included Gauge	Parameters δ_1	ω_0	ω_1	Process Noise Mean	Variance	R^2	% peak error
1	20	-.957	.032	.042	-.00017	.01108	.893	+27
2	23	-.938	.038	.044	-.00015	.00876	.938	+21
3	24	-.940	.024	.055	-.00016	.00945	.909	+24
4	21	-.943	.012	.058	-.00016	.01116	.892	+29
5	25	-.944	.013	.055	-.00015	.01252	.879	+31
6	18	-.941	.017	.055	-.00016	.01154	.889	+28

Note that the most important gauge 20 is located near Bala on the flood-plain of the River Dee (Figure 6.15). The part of the catchment contributing to the rising limb and peak of the hydrograph for a small catchment will be confined to those areas near the channel and the catchment outlet. Location of gauges in valley bottom areas near to the catchment outlet should therefore be preferred to more distant sites. Gauge 20 whilst not in the Hirnant catchment satisfies this requirement, whereas the worst gauge (18) although situated in the catchment, measures rain that probably will only contribute to the falling limb of the hydrograph. The results also suggest that where a strong causal link exists between flow and one or two gauges, the aggregation of rainfall at all gauges to form an areal average may diffuse this relation, resulting in process noise amplification.

6.3.5.5 The composite models

Because interest is focussed on gauge requirement and thus the information contained in each rainfall series to explain variations in flows, identification and parameter estimation of the noise sequences to obtain the complete composite model was not done for all models. To demonstrate how well one may forecast in real-time using a single gauge, however, a noise model was fitted to the process noise sequence for gauge 20, and the two-step ahead forecast (i.e. 1 hour ahead) from the composite model plotted in Figure 6.23. Note that for lead times greater than one hour (b = 2), forecasts will depend on the accuracy of rainfall forecasts. The inclusion of a soil moisture accounting term in the model (Moore, 1977) would help rectify the observed overprediction of the first peaks when the catchment is dry, and slight underprediction of the major peak.

6.3.6 Conclusions

Any conclusions to be drawn from this investigation must be tempered by the fact that only a short record for a single small catchment was examined. The following tentative conclusions appear to have some substance to aid the design of raingauge networks for real-time flow forecasting:

(i) Given an existing network of gauges, selection of those gauges having the strongest causal relation with river flow may provide the basis of a better forecasting model than the areal average computed using all gauges.

(ii) One or two carefully sited gauges may be adequate for real-time flow forecasting in small catchments. The number of gauges and their siting should pay regard to the catchment area contributing to the rising limb and peak of the storm hydrograph.

(iii) The problem of instrument failure (Figure 6.16) may prove to be the most important criterion affecting gauge density requirement. The reliability of telemetry installations to transmit gauge measurements must also be considered in network design.

Missing observations in parts of a gauge network appear to have important implications for rainfall-runoff modelling. A model calibrated using a particular subset of gauges may perform poorly if only a different subset is available in real-time. Calibration of a model using the dense network provided as part of the Dee Weather Radar Project, when only a single telemetering raingauge may be available in real-time, will probably result in more serious forecast errors than had the single gauge been used for calibration. In essence the real-time flow forecaster is not interested in estimating the 'true' catchment average rainfall, but rather in measurements of point rainfall that have the strongest causal link with the river flow response.

The storage effect of a catchment is often held to diminish the importance of raingauge network design for flow forecasting, damping out the extremes of rainfall intensity so that river flow forms a delayed and smoothed function of the highly variable original rainfall sequence. Bras and Rodriguez-Iturbe (1976b) present evidence contrary to this belief, finding that a dense network will significantly reduce errors in the hydrograph rising limb and peak where most uncertainty exists. In the small Hirnant catchment the area contributing to this portion of the hydrograph appears to be confined closely to the channel and towards the catchment outlet. A single gauge sampling this area appears adequate. The catchment used by Bras and Rodriguez-Iturbe was six times larger (212 km^2), and the hydrograph peak was influenced by rainfall on upstream locations. Both investigations consequently accord with conclusion (ii) but indicate that the number of gauges required will depend on (a) the size of the catchment contributing to the hydrograph rising limb and peak, and (b) the heterogenity of rainfall over this area, and consequently the storm type. The storage effect of the catchment does alleviate the network design problem in that rain falling on more remote parts need not be sampled, the falling limb being adequately modelled by an autoregressive flow component. This only pertains to real-time flow forecasting, where the facility exists to continually reset forecast flow values to their measured value in real-time.

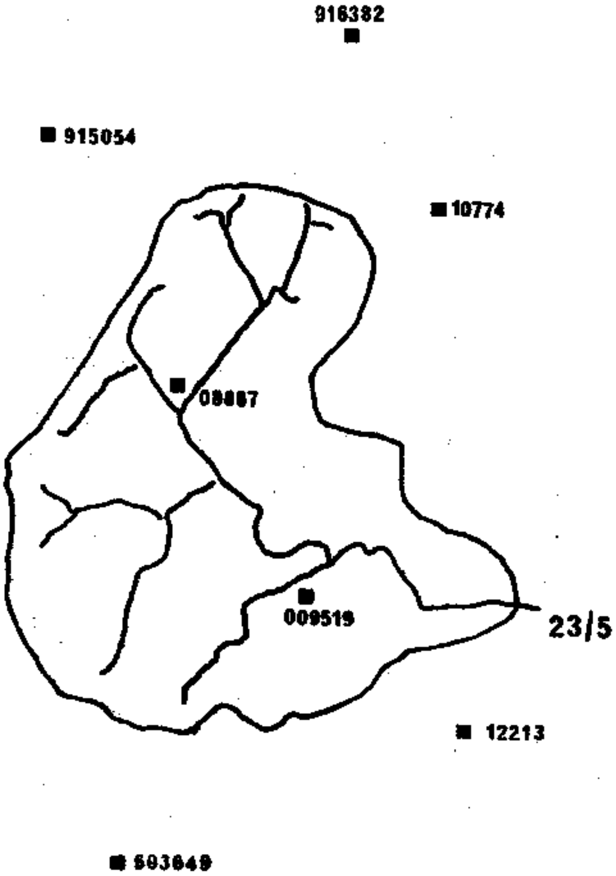


Figure 6.1 Catchment map for the North Tyne at Tarset (23/5), with locations of raingauges used

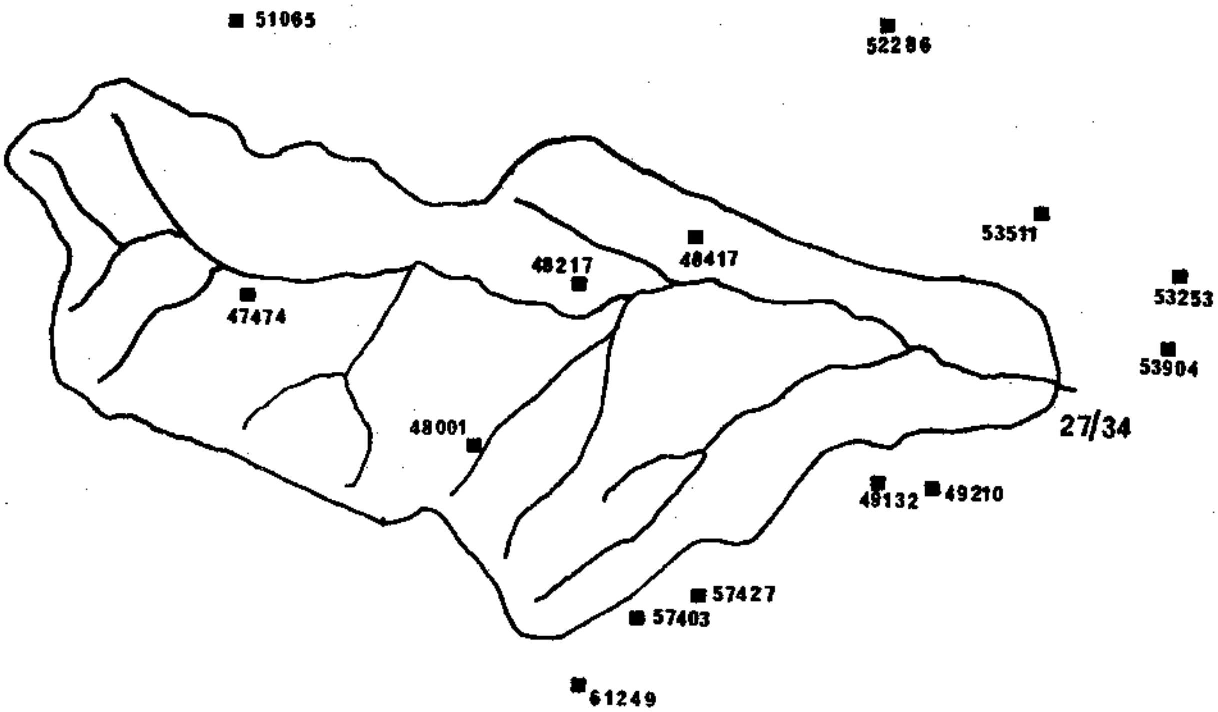


Figure 6.2 Catchment map for the Ure at Kilgram Bridge (27/34), with locations of raingauges used

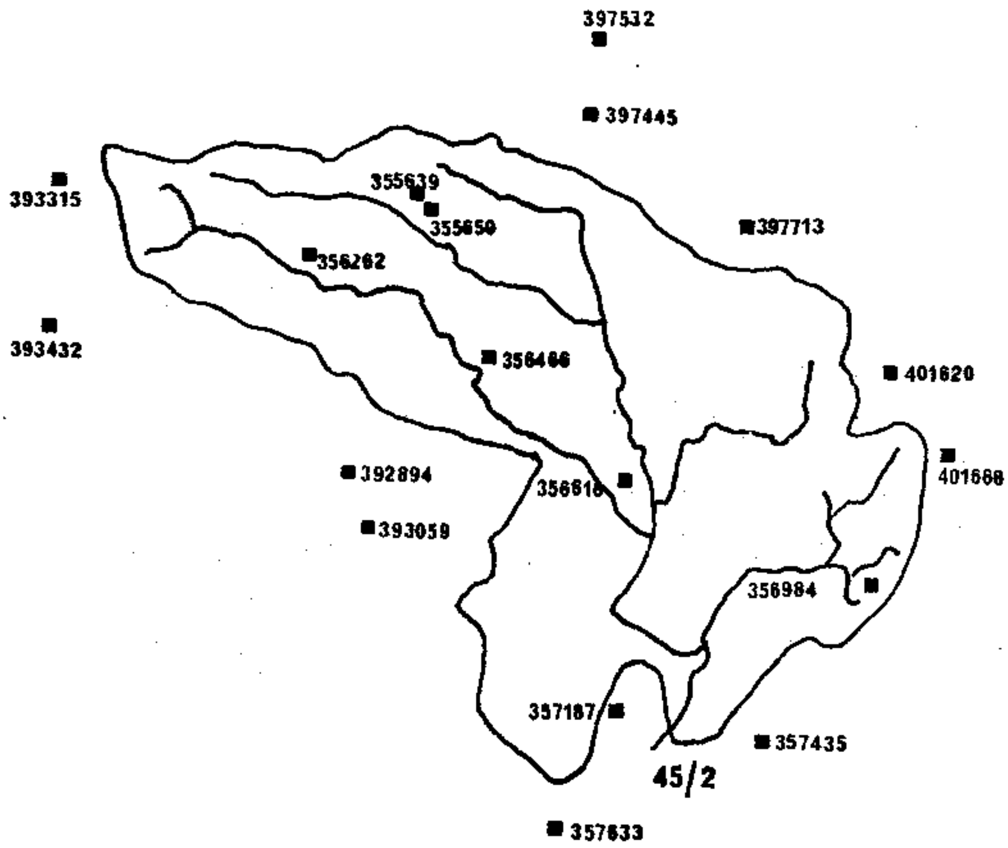


Figure 6.3 Catchment map for the Exe at Stoodleigh (45/2), with locations of raingauges used

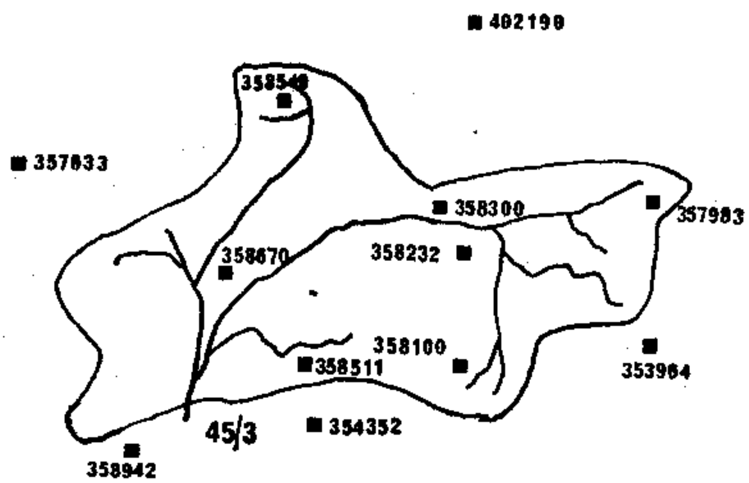


Figure 6.4 Catchment map for the Culm at Woodmill (45/3), with locations of raingauges used

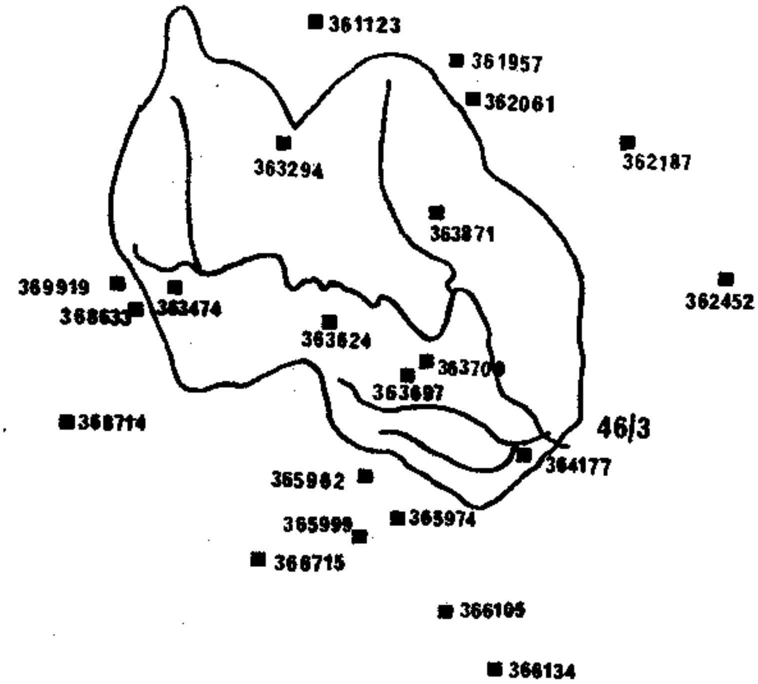


Figure 6.5 Catchment map for the Dart at Austin's Bridge (46/3), with locations of raingauges used

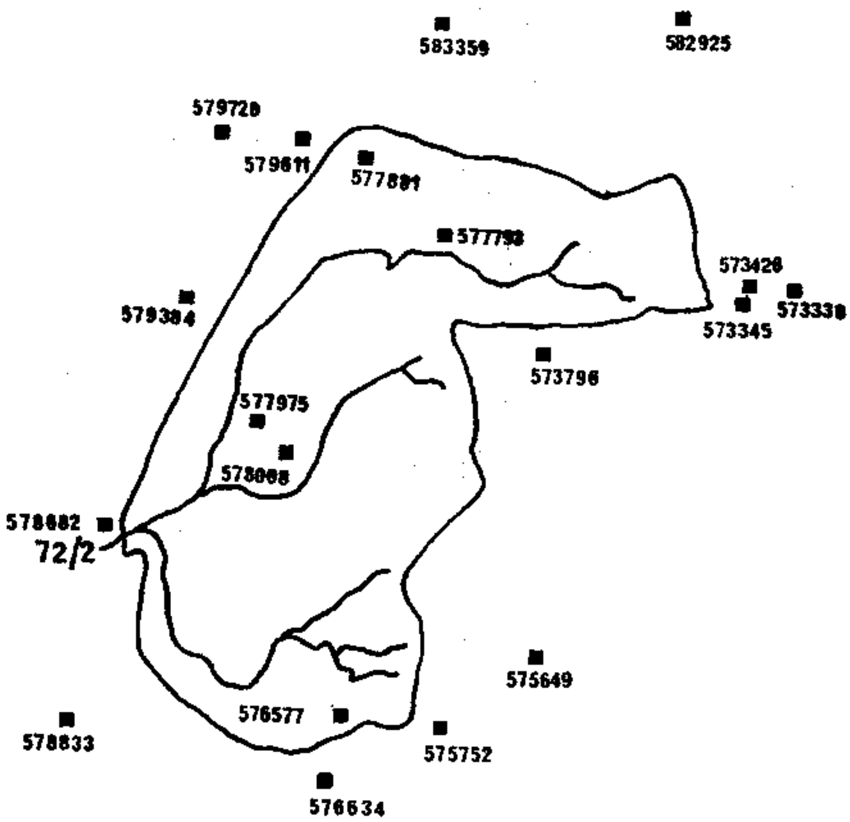


Figure 6.6 Catchment map for the Wyre at St Michael's (72/2), with locations of raingauges used

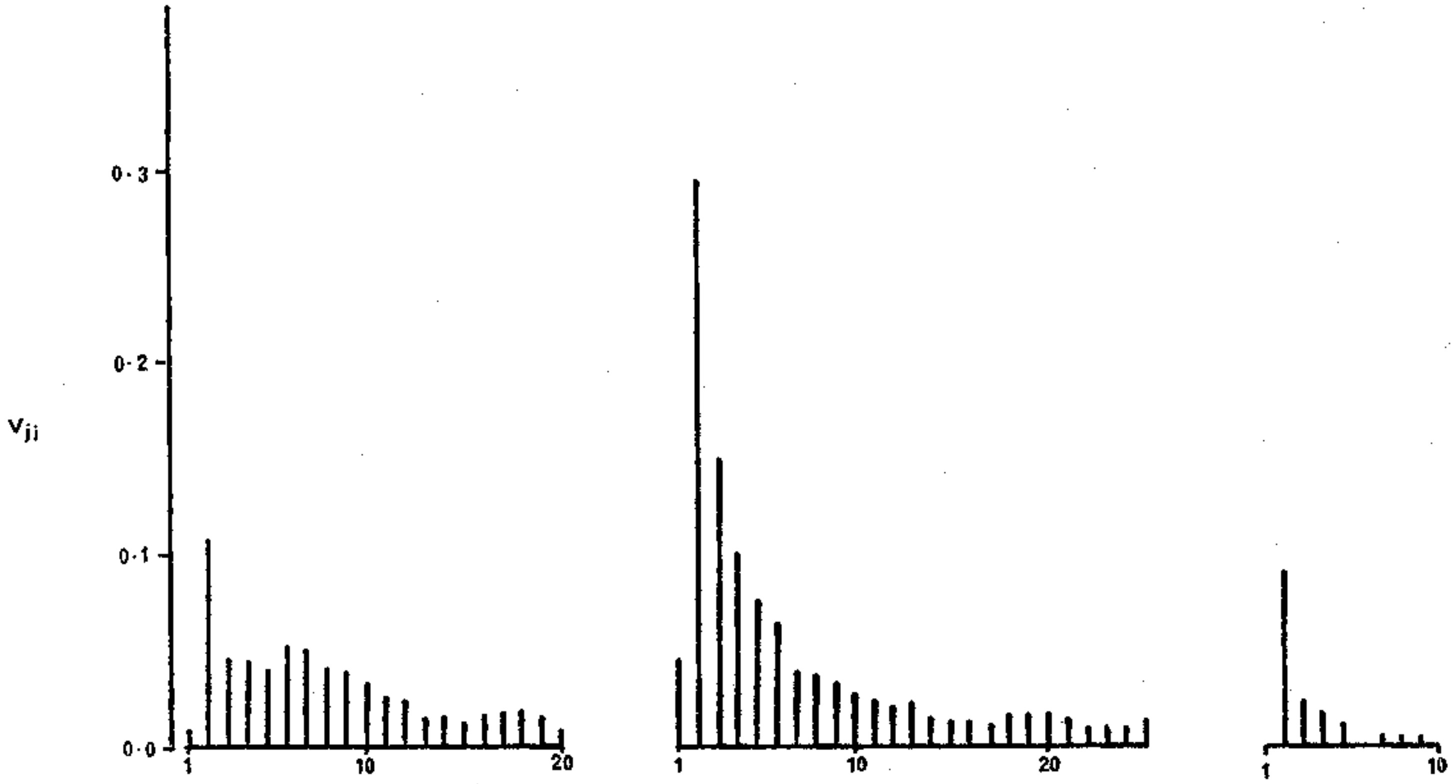


Figure 6.7. The ordinates v_{ji} of the jth impulse response, j = 1, 2, 3, i = 1, 2, ..., k_j for the Exe at Stoodleigh (45/2). The values of k_j were k_1 = 20, k_2 = 25 and k_3 = 10.

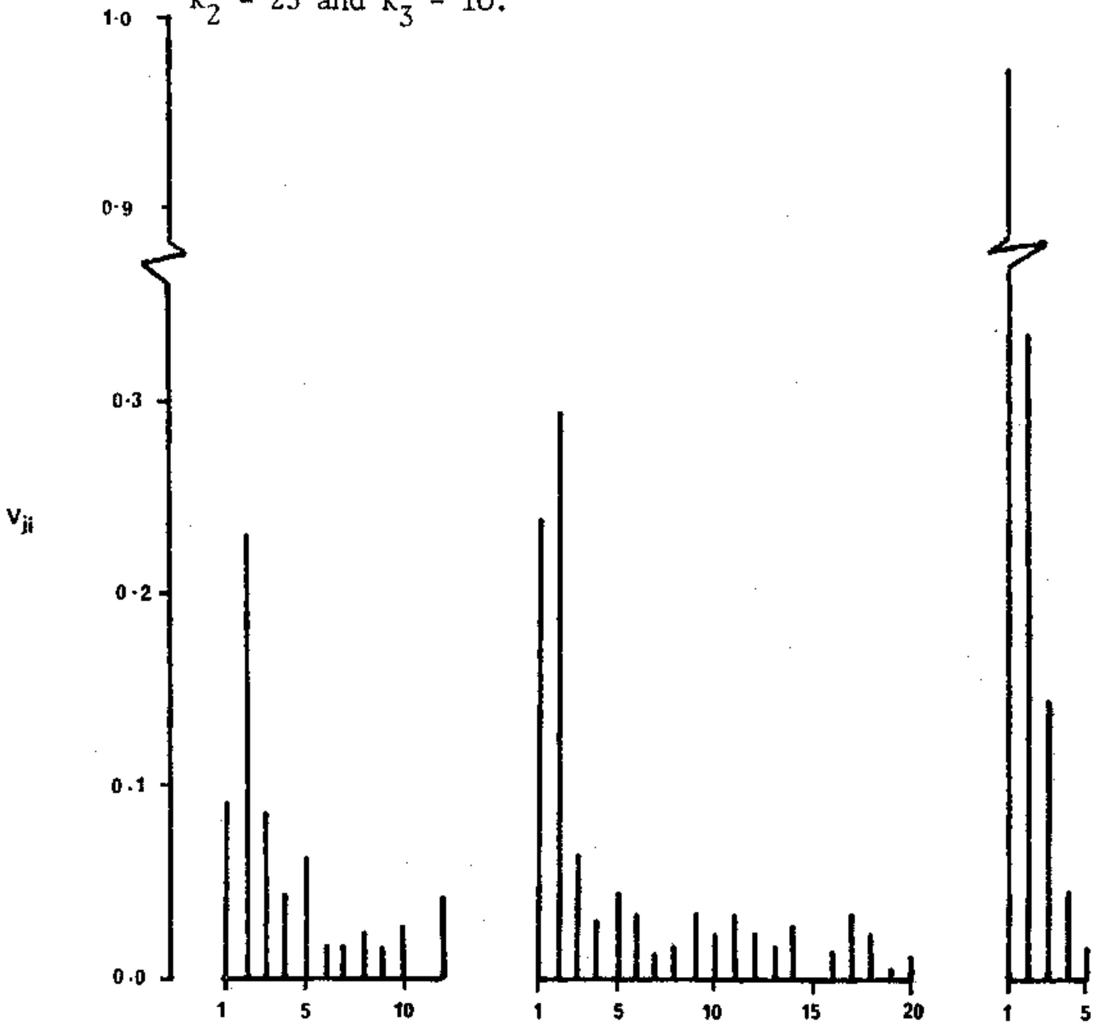


Figure 6.8 The ordinates v_{ji} of the jth impulse response, j = 1, 2, 3, i = 1, 2, ..., k_j, for the Ure at Kilgram Bridge (27/34). The values of k_j are k_1 = 12, k_2 = 20 and k_3 = 5.

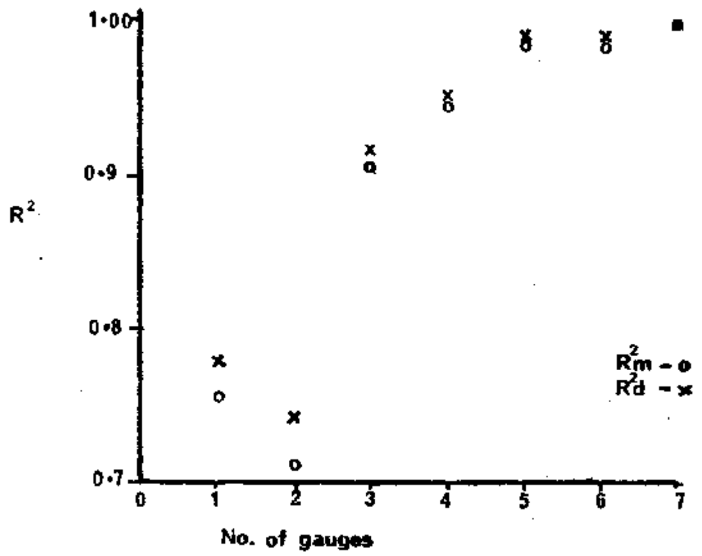


Figure 6.9 Relationship between R_d^2, R_m^2 and number of gauges for the North Tyne at Tarset (23/5)

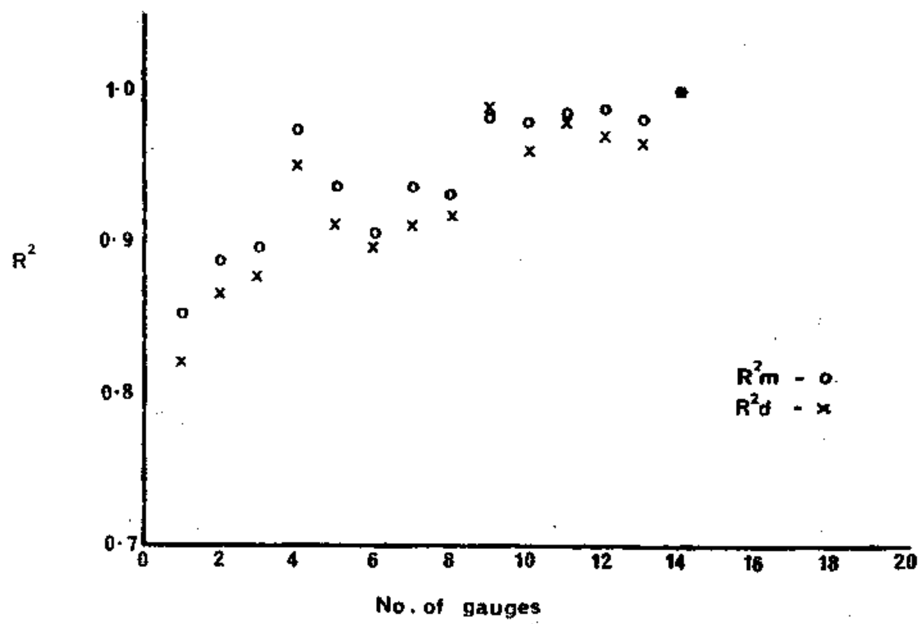


Figure 6.10. Relationship between R_d^2, R_m^2 and number of gauges for the Ure at Kilgram Bridge (27/34)

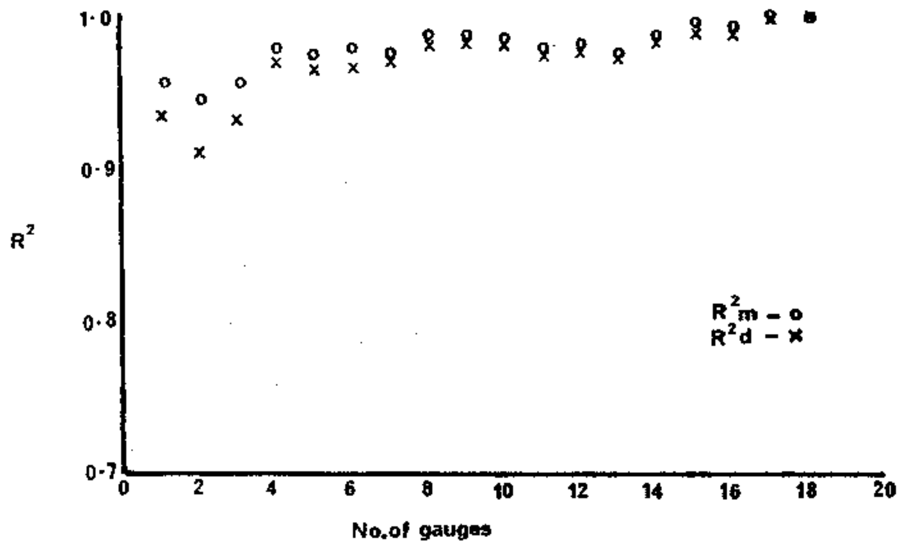


Figure 6.11 Relationship between R_d^2, R_m^2 and number of gauges for the Exe at Stoodleigh (45/2)

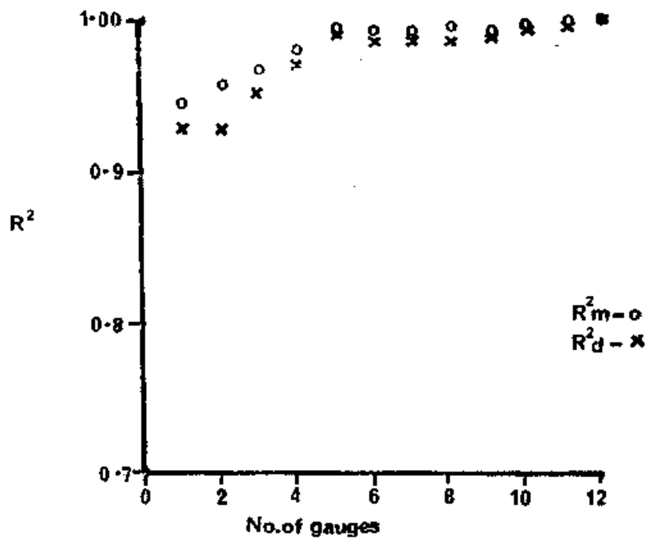


Figure 6.12 Relationship between R_d^2, R_m^2 and number of gauges for the Culm at Woodmill (45/3)

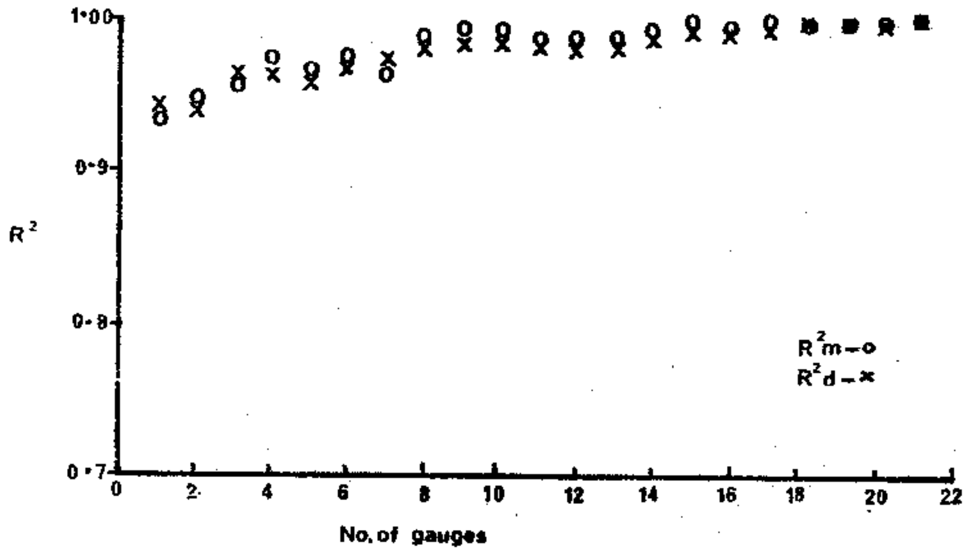


Figure 6.13 Relationship between R_d^2, R_m^2 and number of gauges for the Dart at Austin's Bridge (46/3)

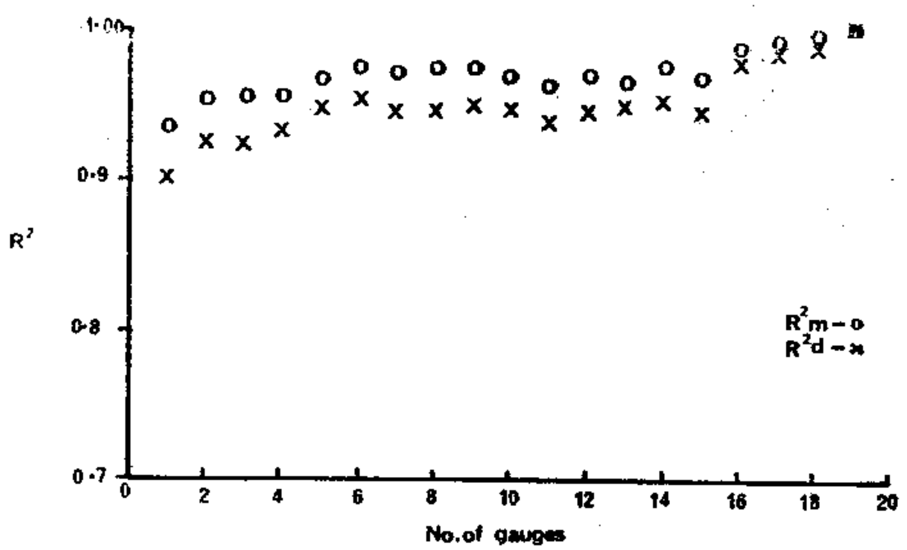


Figure 6.14 Relationship between R_d^2, R_m^2 and number of gauges for the Wyre at St Michael's (72/2)

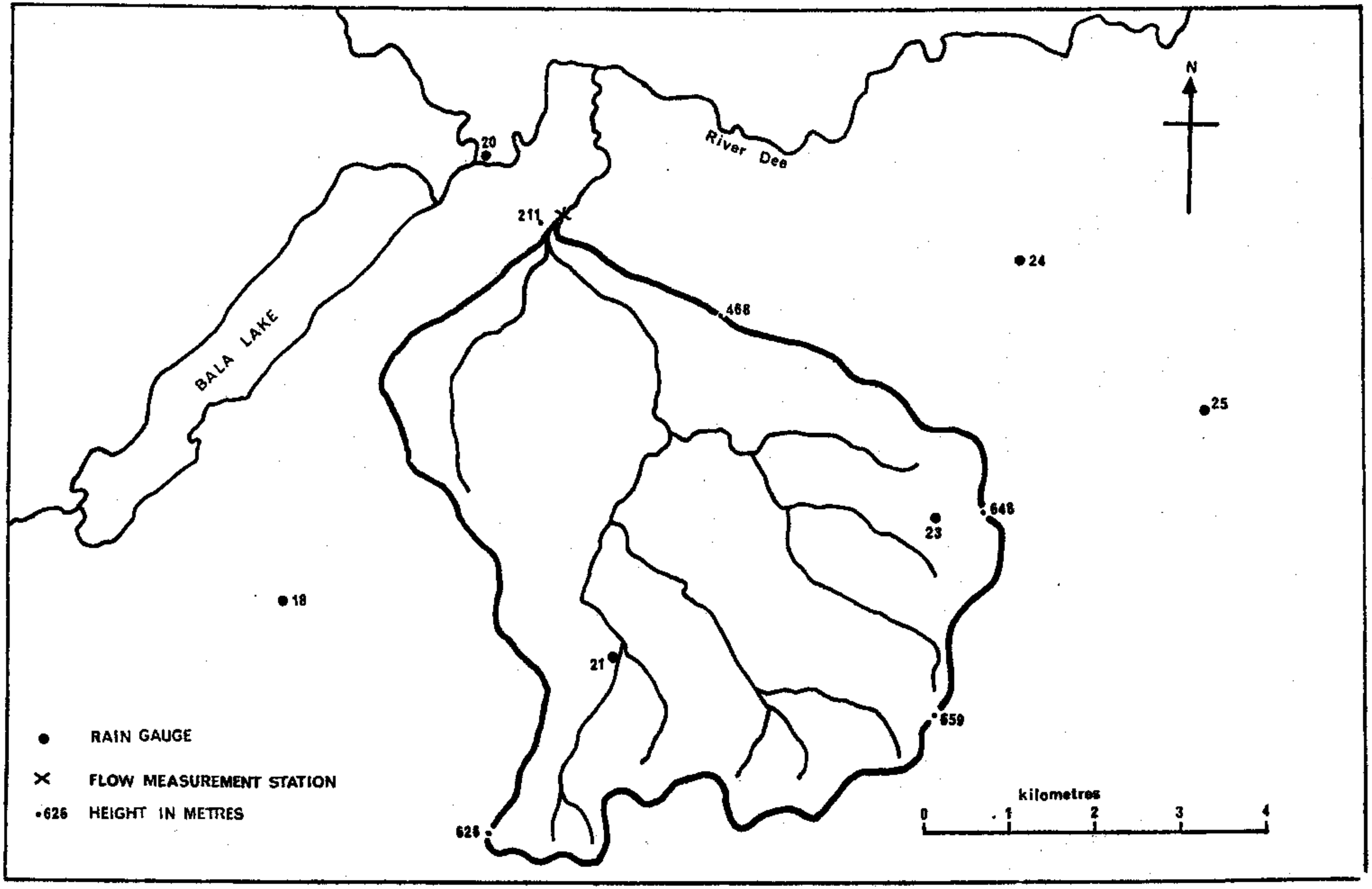


Figure 6.15 The Hirnant Catchment

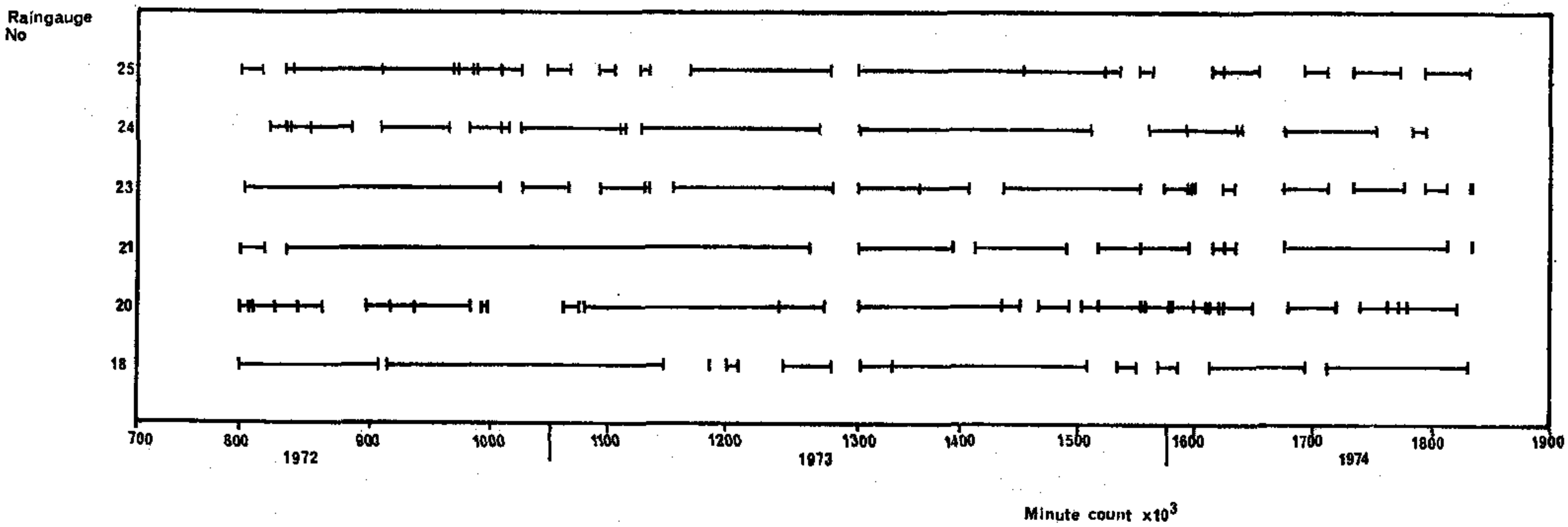


Figure 6.16 Available raingauge data for the Hirnant, 1/7/72 to 30/6/74 (ı ı indicates missing data)

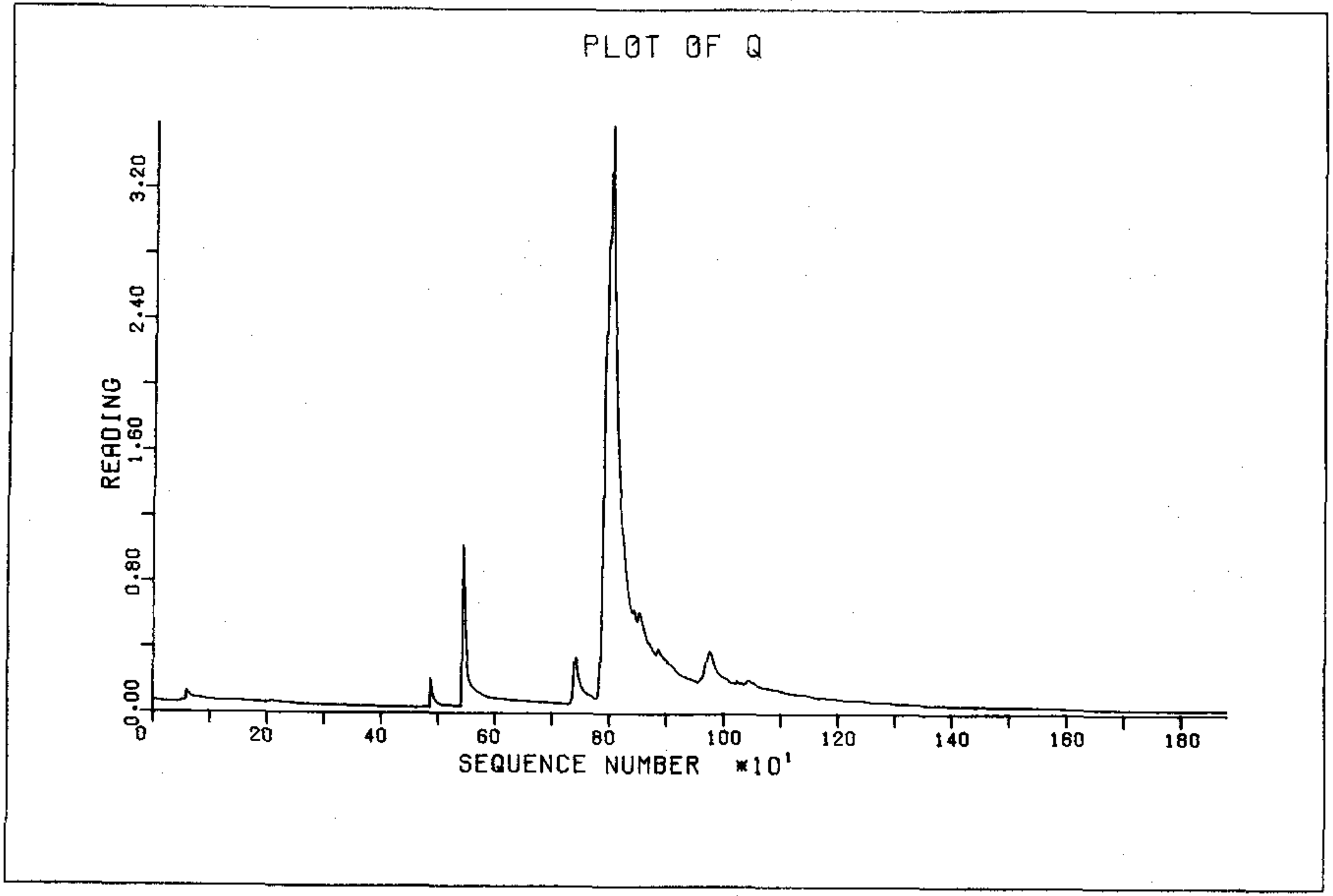


Figure 6.17 Flow of the Hirnant at Plas Rhiwaedog, 20/7/73 to 28/8/73, in mm/hr

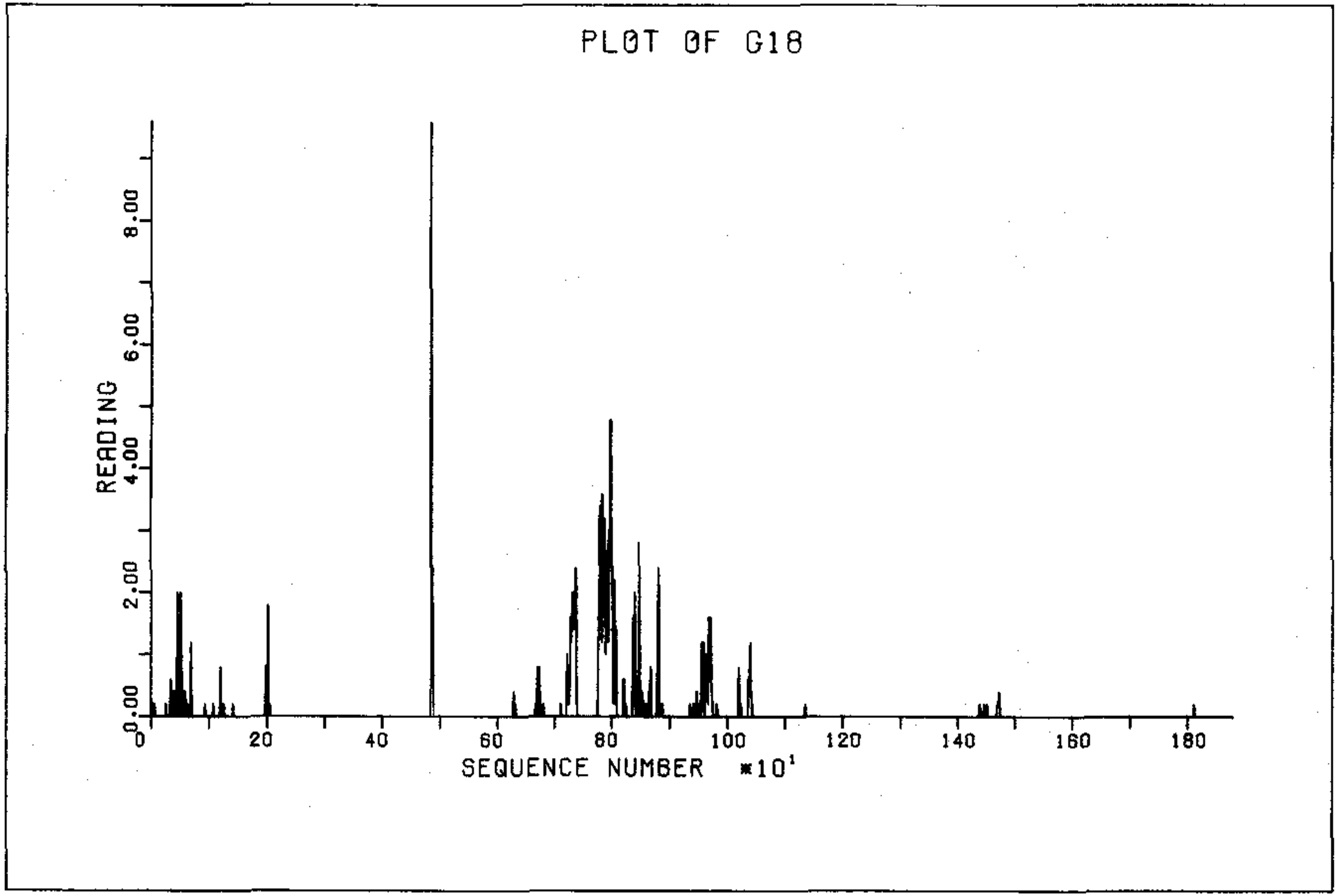


Figure 6.18 (a) Half-hour totals of rainfall in millimetres for gauge 18, 20/7/73 to 28/8/73

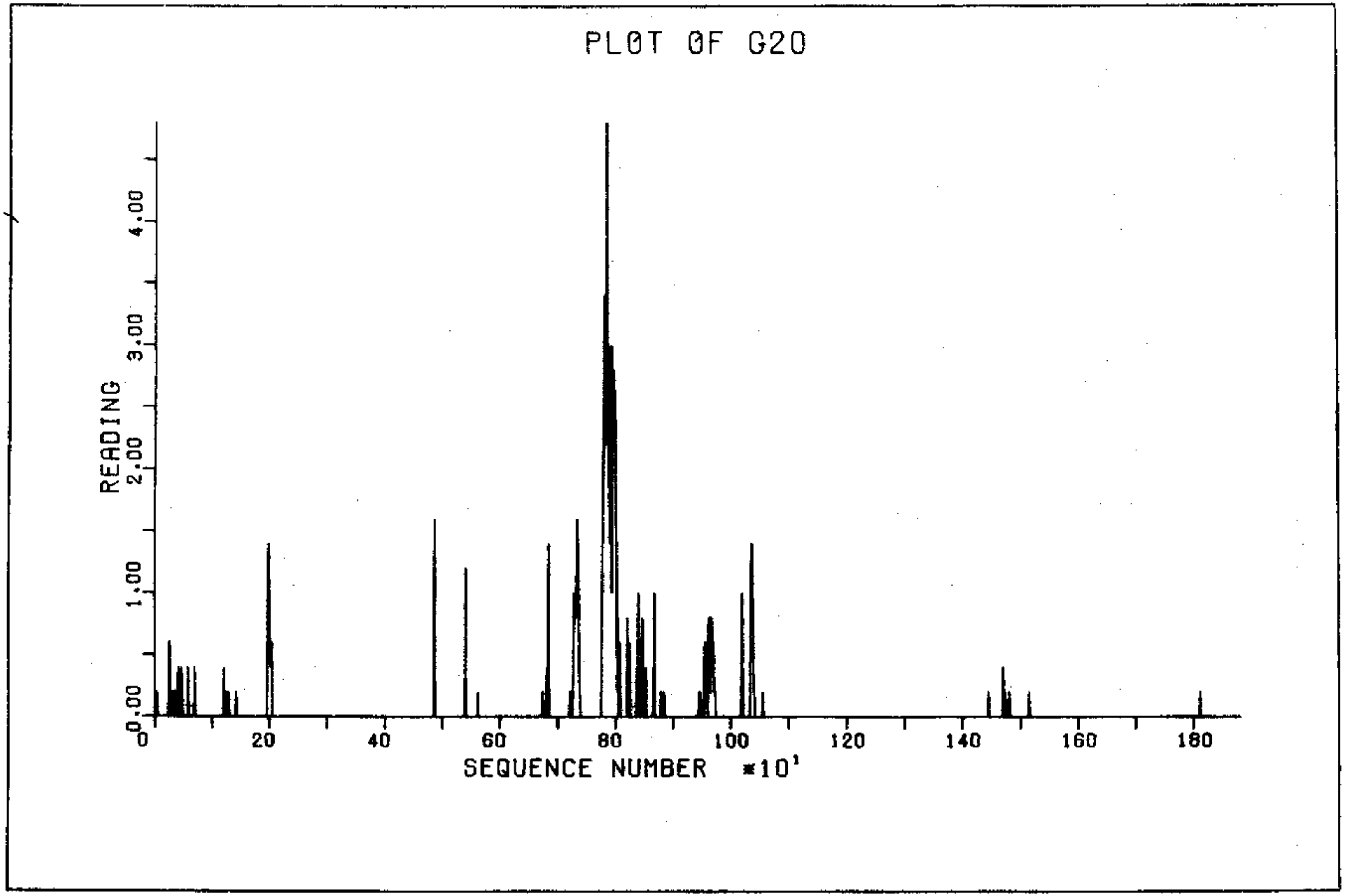


Figure 6.18 (b) Half-hour totals of rainfall in millimetres for gauge 20, 20/7/73 to 28/8/73

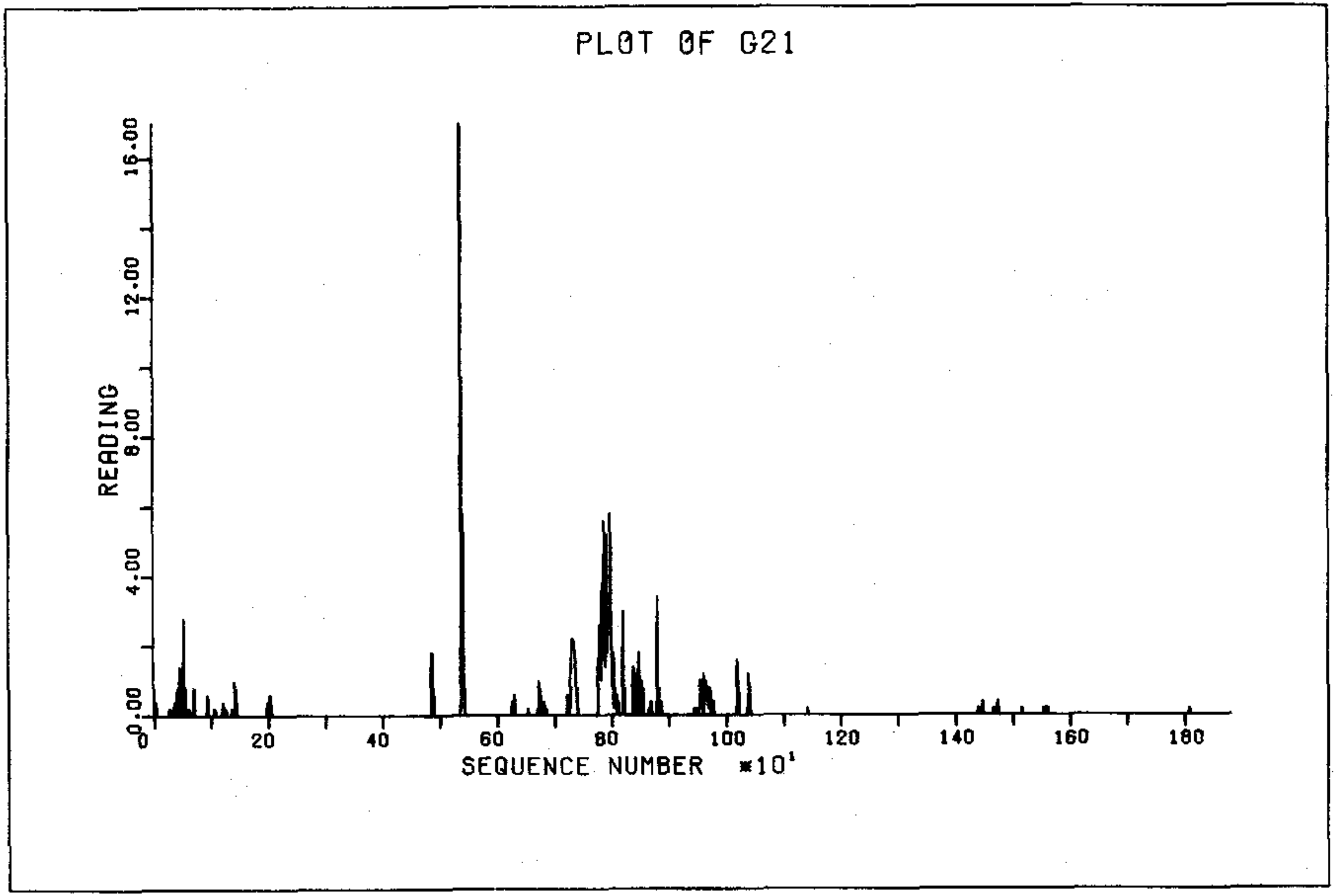


Figure 6.18 (c) Half hour totals of rainfall in millimetres for gauge 21, 20/7/73 to 28/8/73

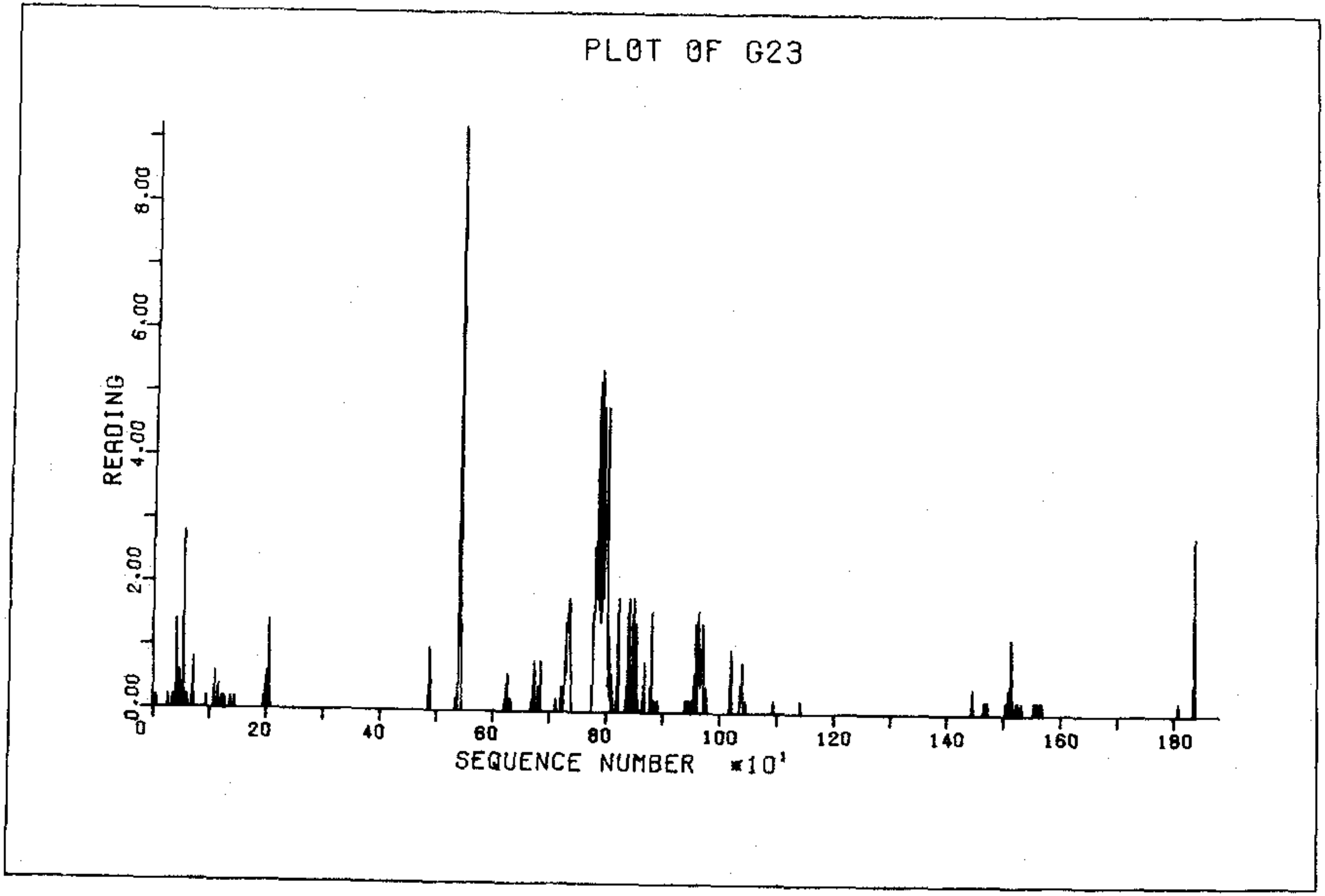


Figure 6.18 (d) Half hour totals of rainfall in millimetres for gauge 23, 20/7/73 to 28/8/73

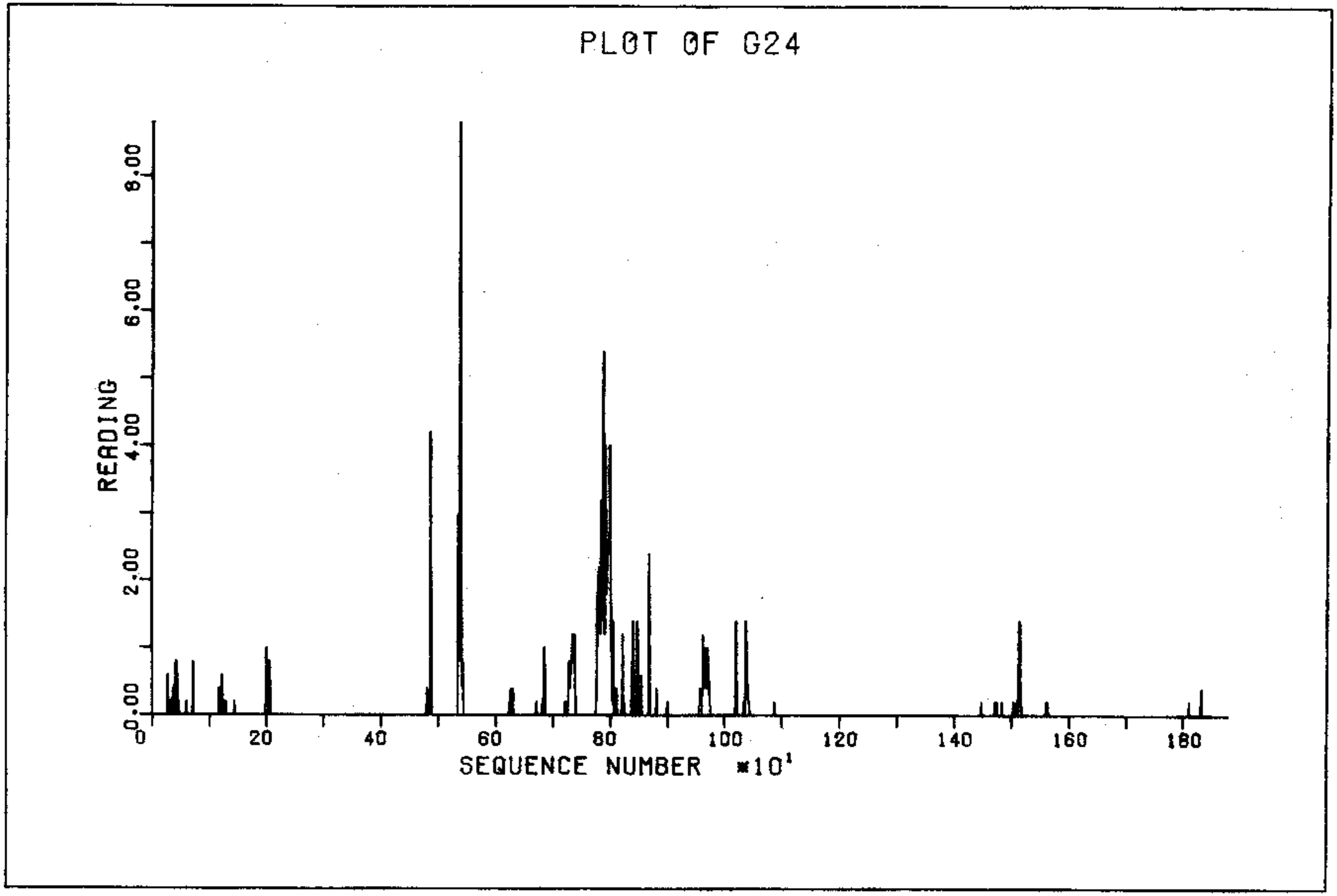


Figure 6.18 (e) Half hour totals of rainfall in millimetres for gauge 24, 20/7/73 to 28/8/73

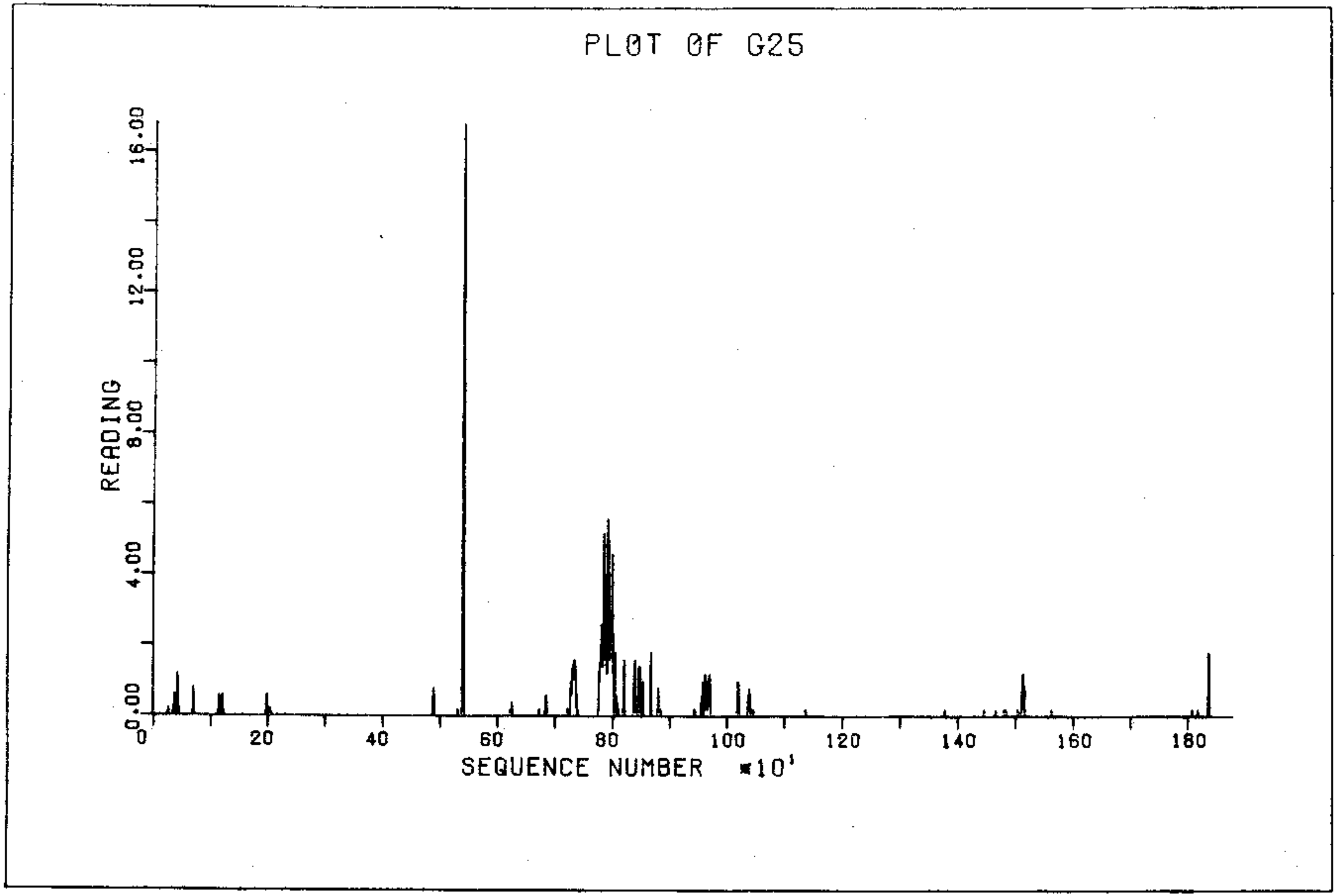
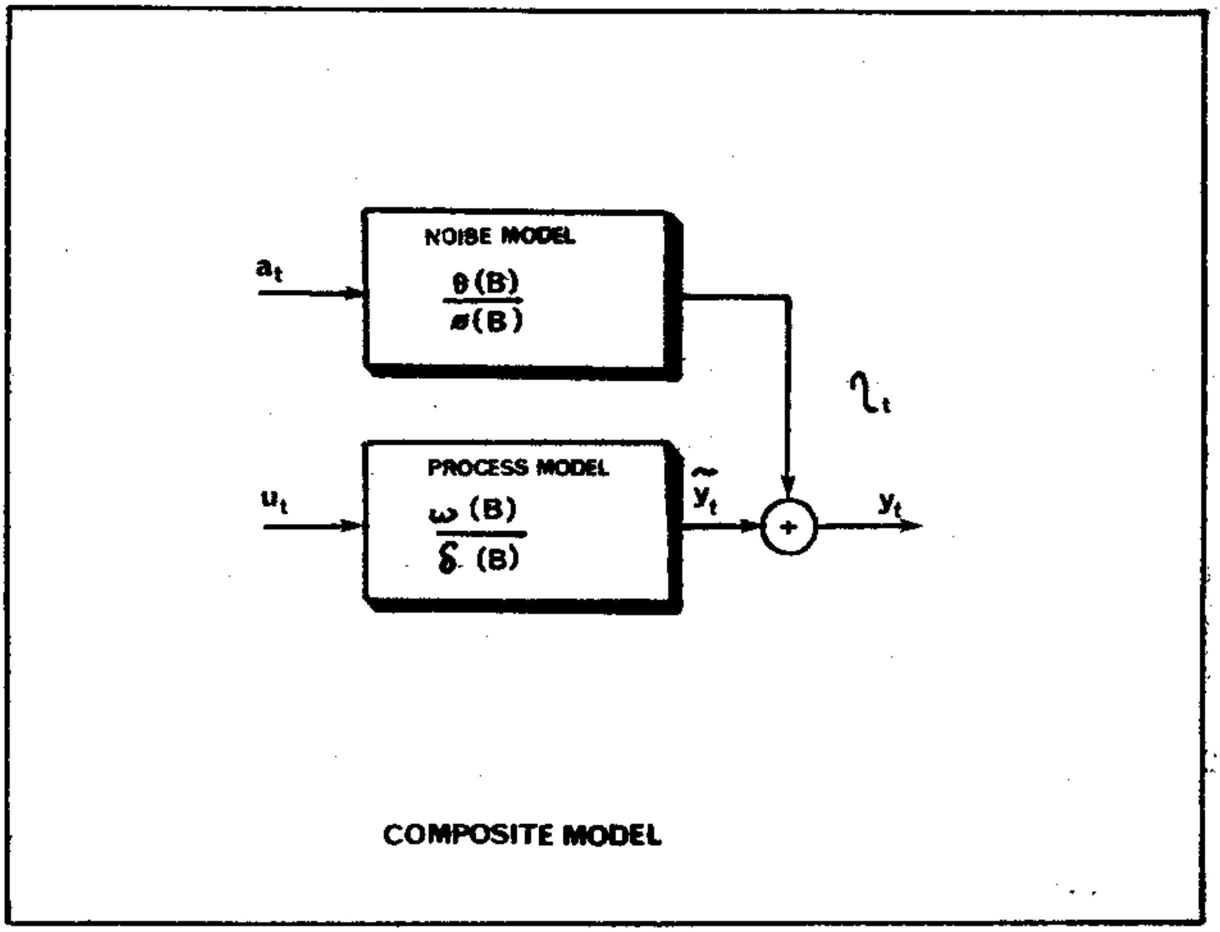


Figure 6.18 (f) Half hour totals of rainfall in millimetres for gauge 25, 20/7/73 to 28/8/73



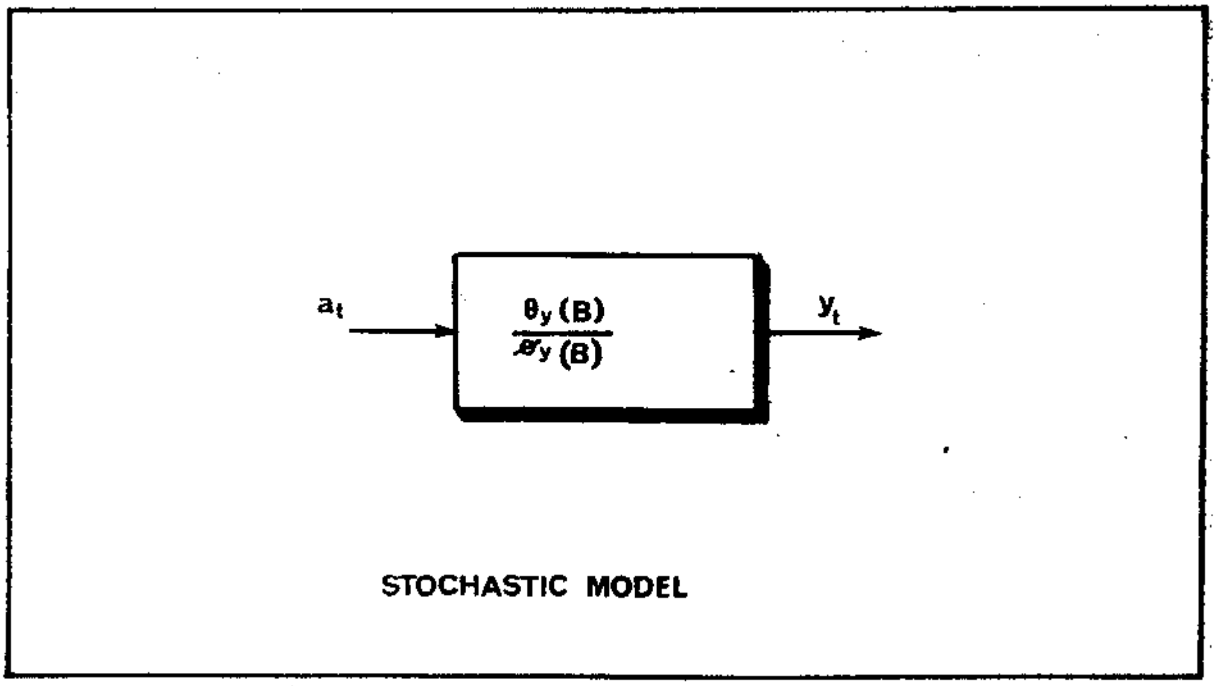


Figure 6.19 Schematic representation of composite and stochastic models

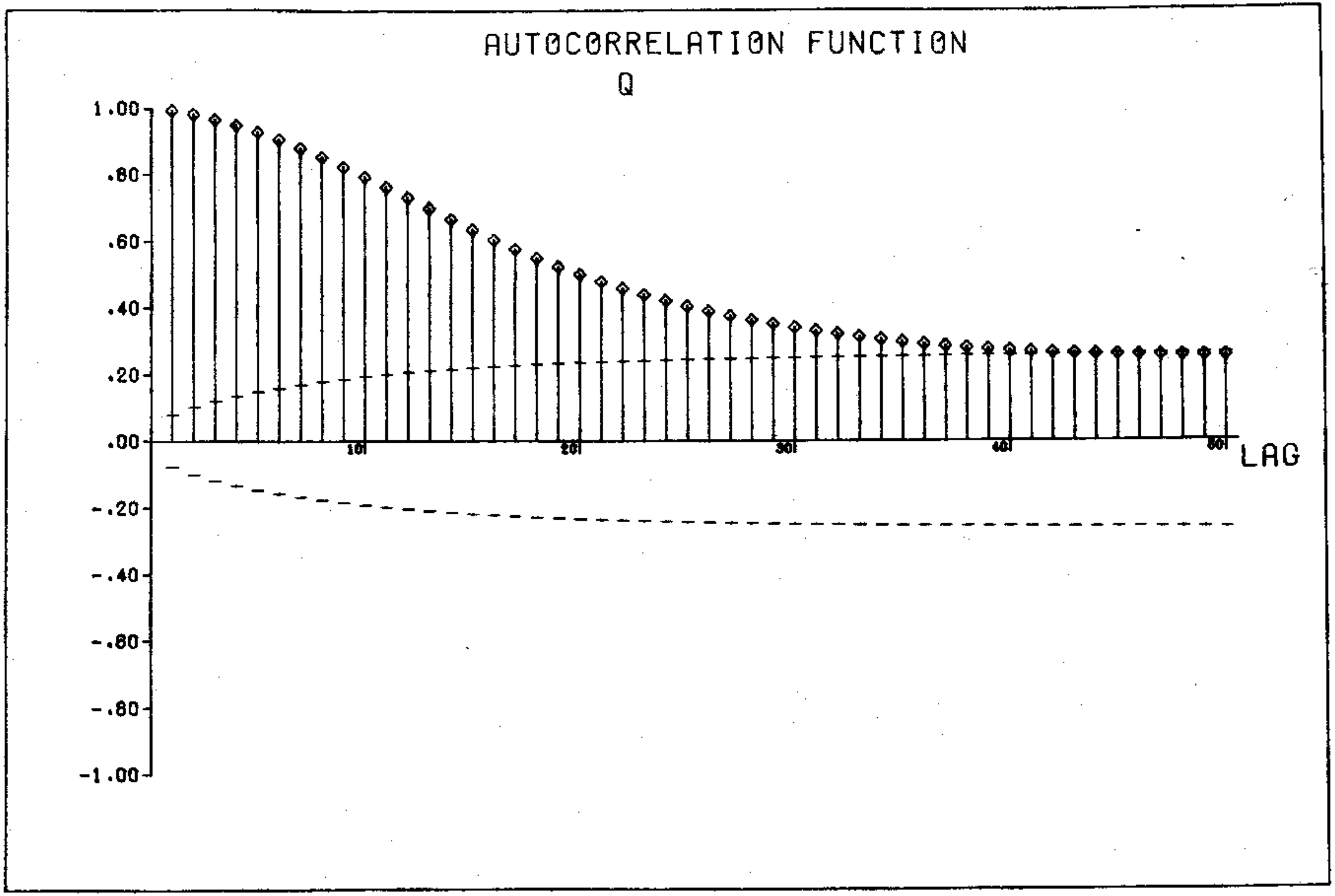


Figure 6.20 (a) Autocorrelation function of Hirnant flow series

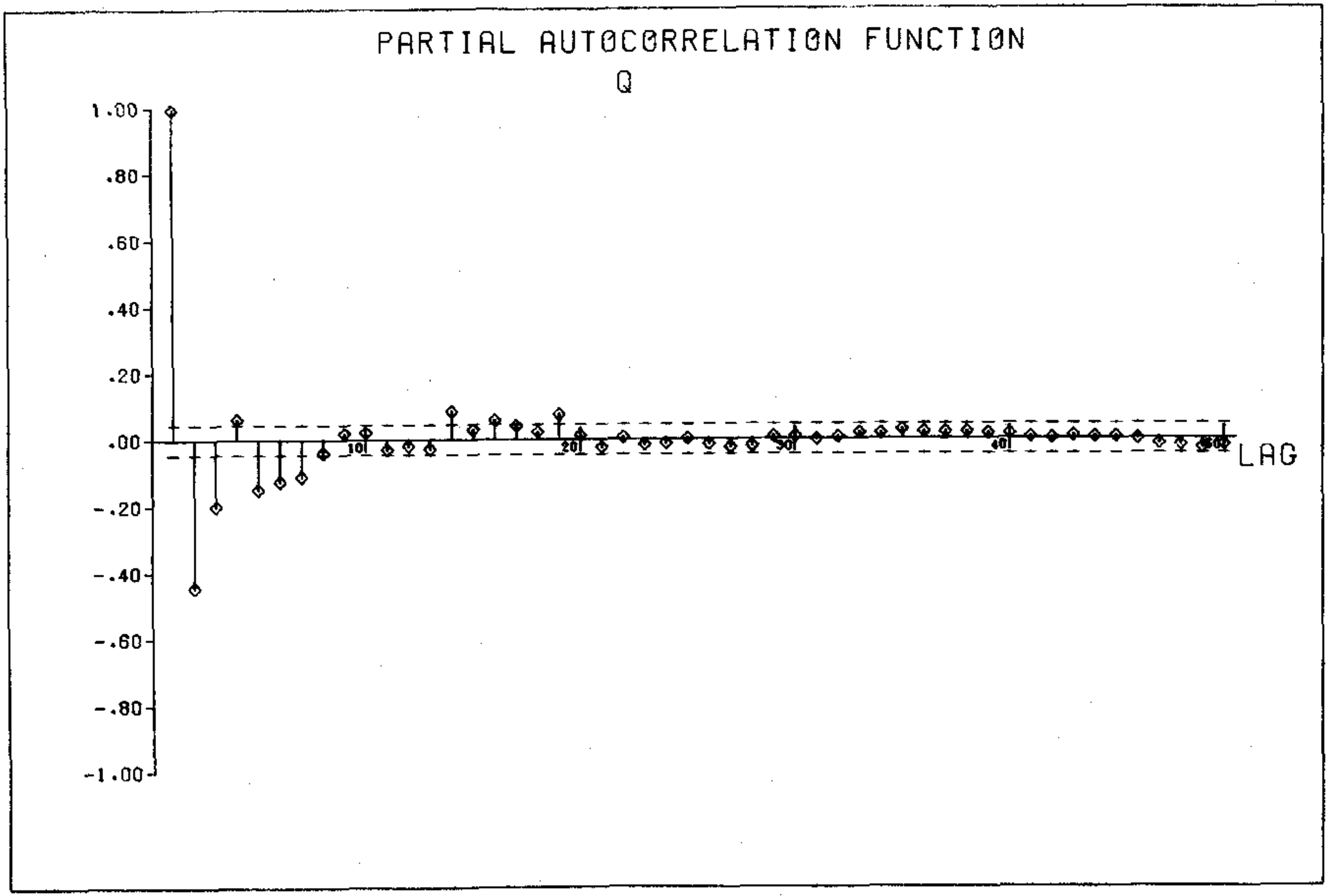


Figure 6.20 (b) Partial autocorrelation function of Hirnant flow series

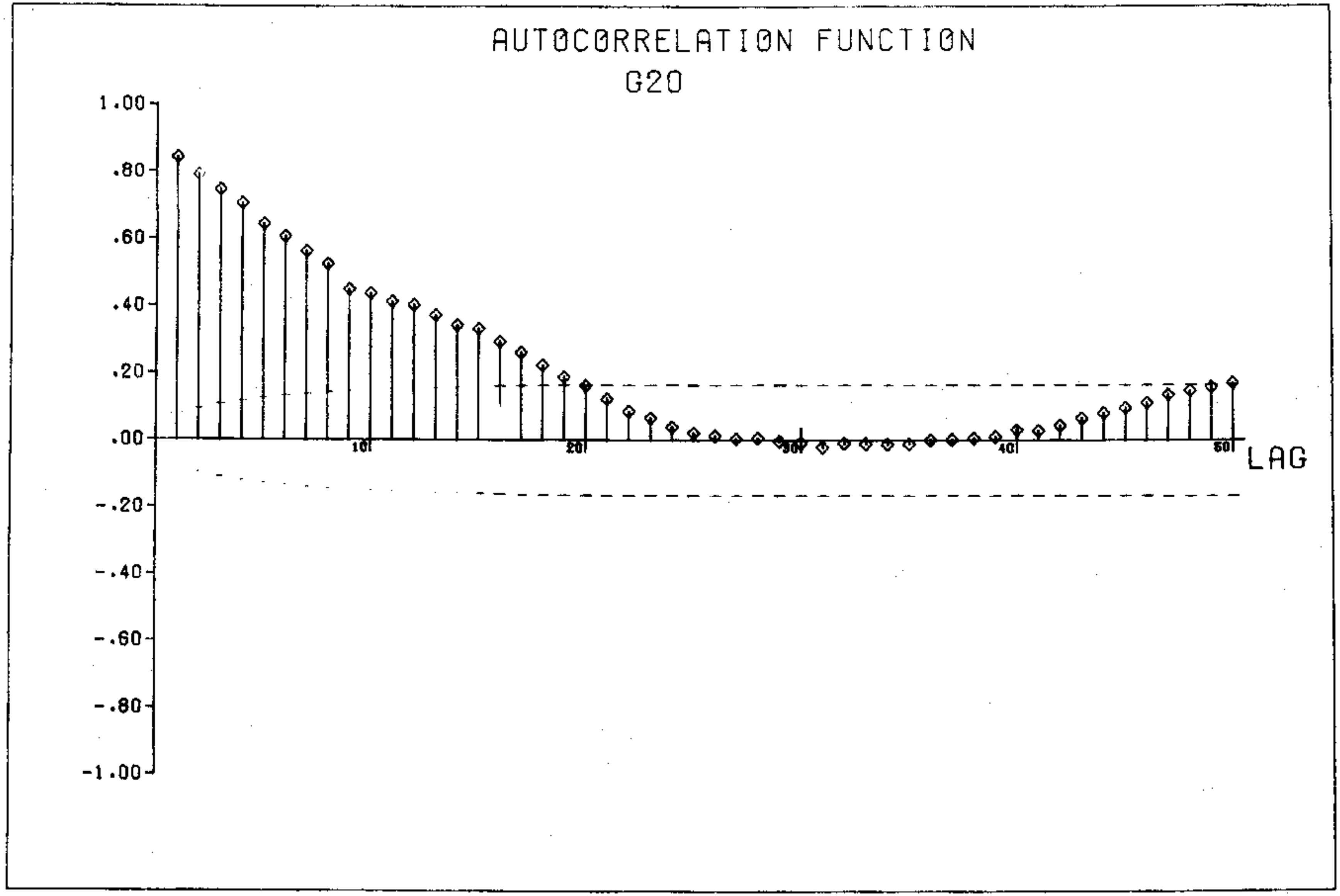


Figure 6.21 (a) Autocorrelation function of gauge 20 rainfall

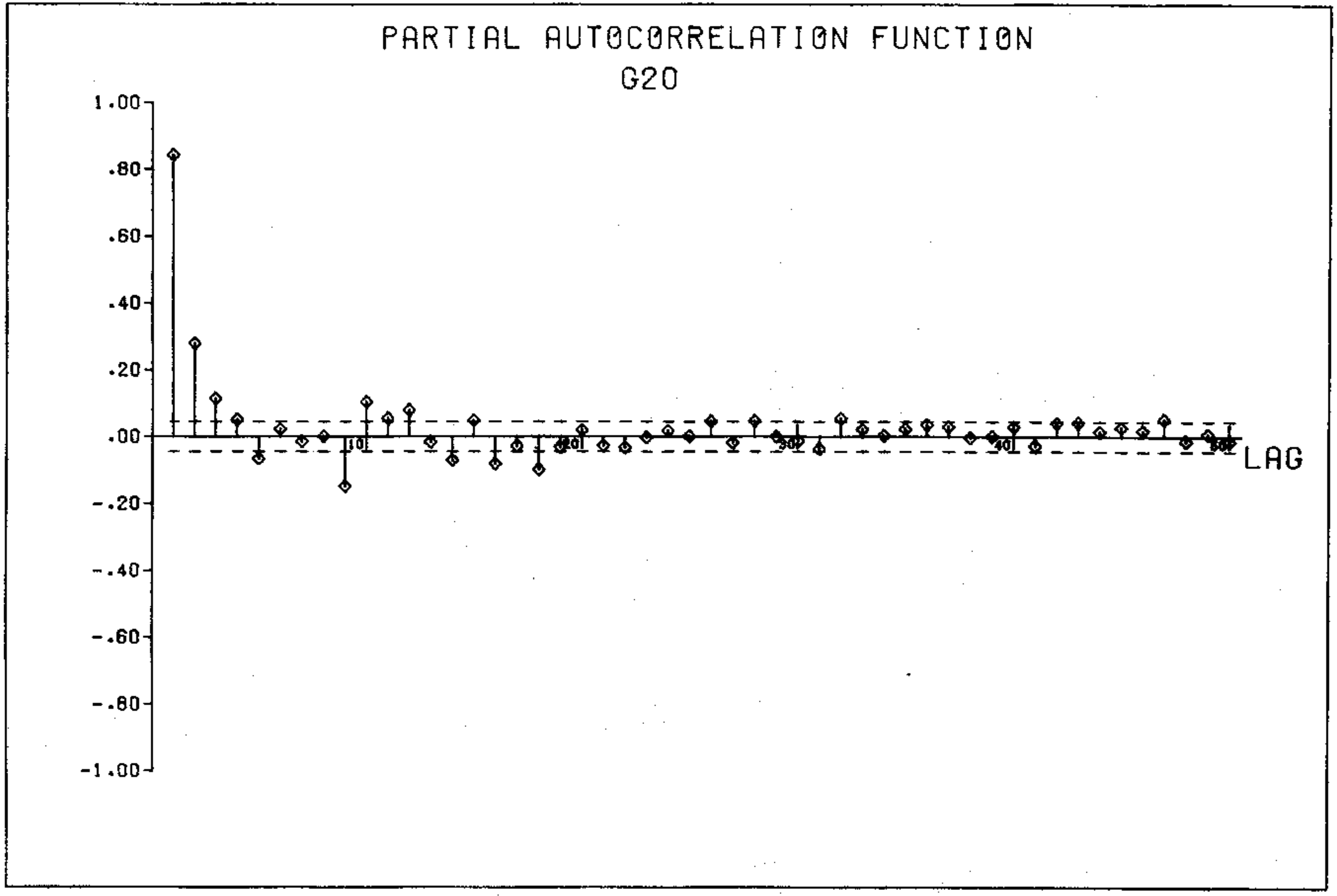


Figure 6.21 (b) Partial autocorrelation function of gauge 20 rainfall

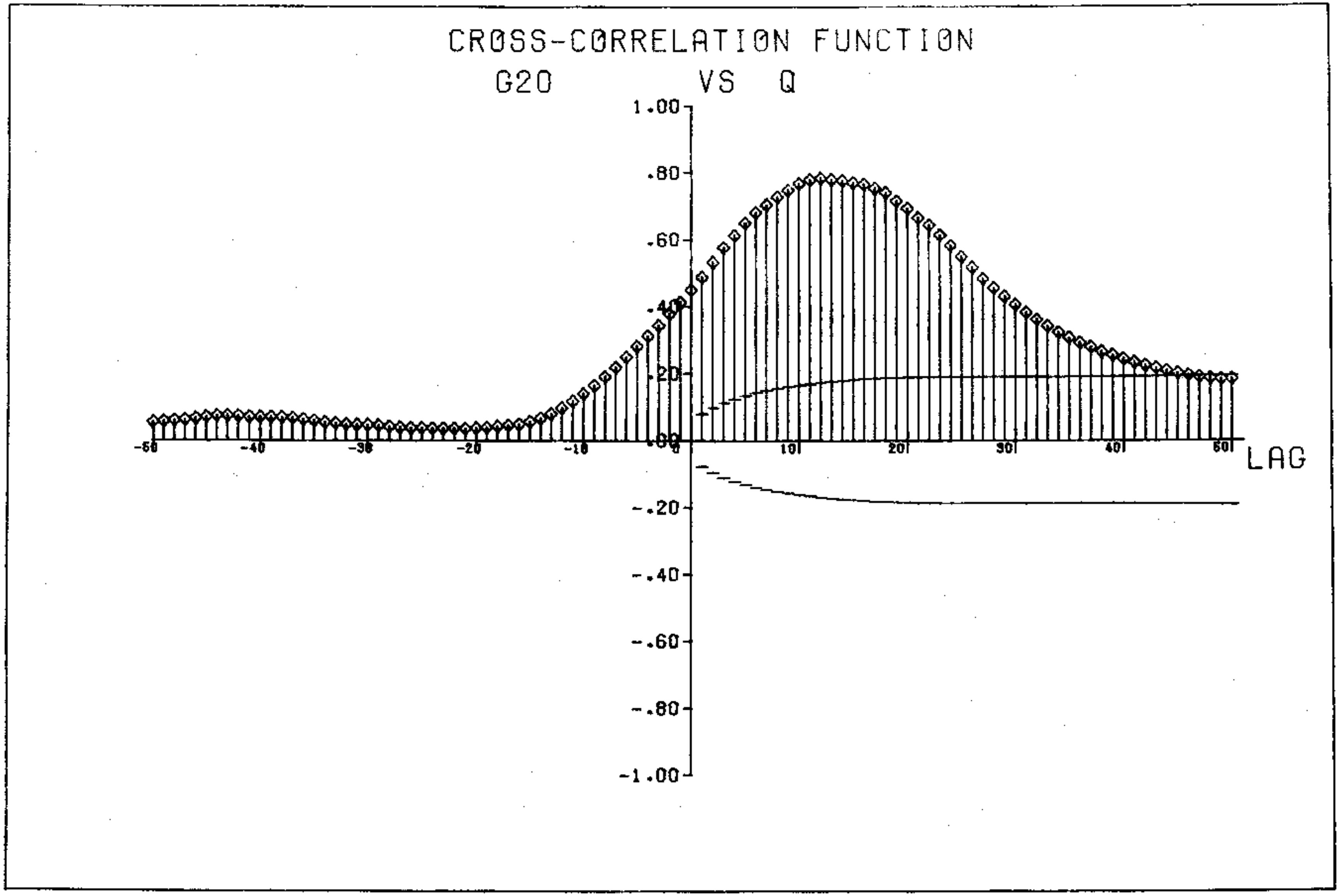


Figure 6.22 Cross-correlation function of gauge 20 rainfall against Hirnant flow

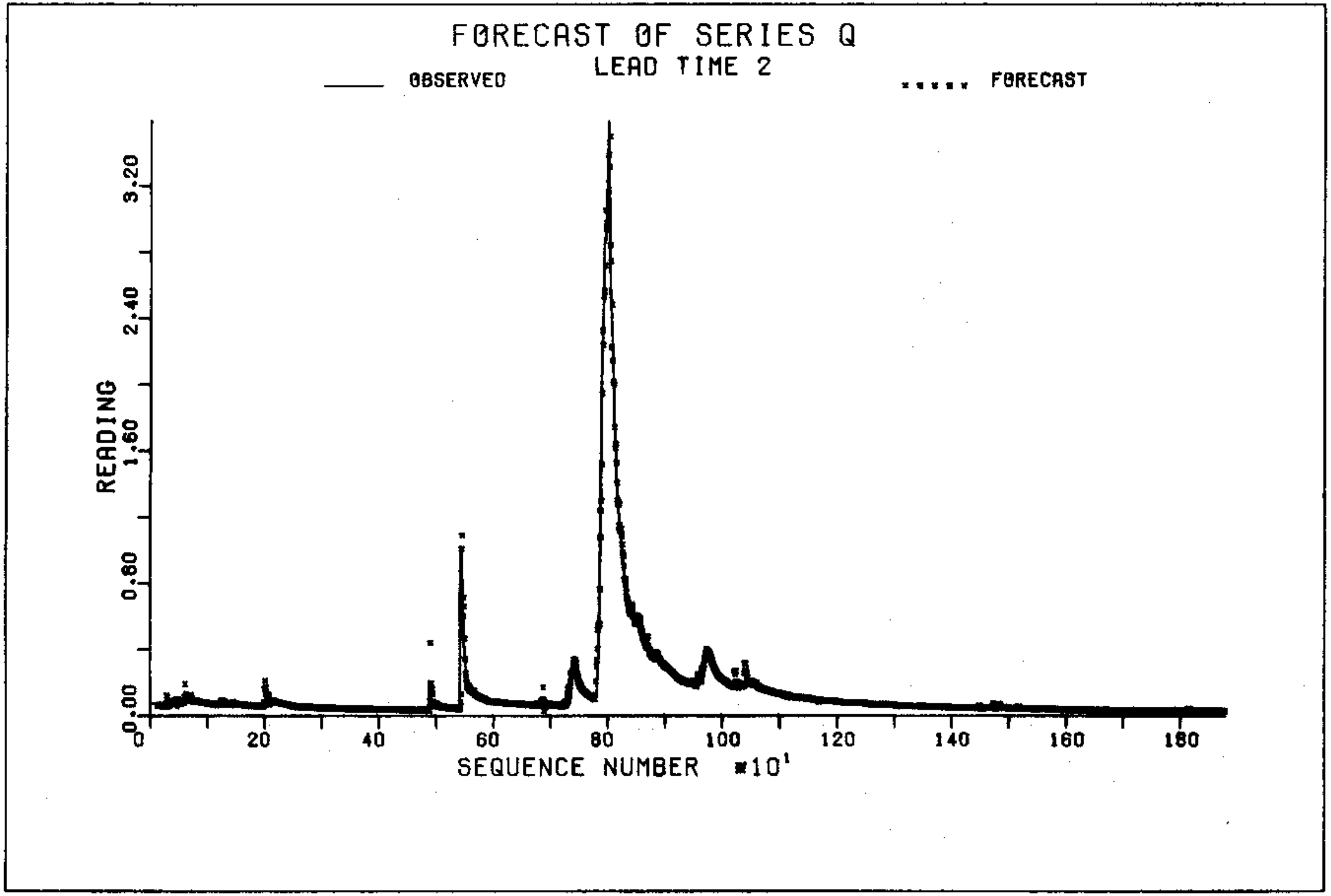


Figure 6.23 Two-step ahead forecast for the Hirnant catchment

7. EVALUATION OF EXISTING NETWORK

7.1 Introduction

Direct and indirect approaches to network design have been described in Sections 5 and 6. The implications of these analyses will be discussed in this section, together with some considerations of objective ways of modifying a raingauge network. Section 7.2 discusses the adequacy of the existing network of raingauges for meeting the requirements of users who can specify directly error criteria for rainfall data. Network requirements for streamflow forecasting and prediction are discussed in Section 7.3. In Section 7.4 the problem of modifying an existing network is considered within a statistical framework - the aim is to identify those areas in which the rainfall process is not well represented by the network and thus to suggest localities for new gauges. The question of choosing specific sites is not under consideration here.

7.2 Implications for network design using direct methods

7.2.1 Interpolation of point values

The adequacy of a raingauge network may be judged in several ways, one of which is to determine the accuracy with which rainfall may be interpolated and then to adjust the configuration of the network until a specified accuracy is attained everywhere. The accuracy of point interpolation was considered in Section 5.2. Using the methods described there, it is possible to calculate the accuracy of any particular estimator, based on measurements at specified stations. It is difficult to evaluate fully the adequacy of a network, as any evaluation is dependent on the number and configuration of stations used, and on the type of estimator employed. The general method is illustrated in this section using a few examples.

In Section 5.2 fixed shapes of triangle, equilateral and right-angled, of various sizes were considered. The mean square error of interpolation at the centroid of each triangle was calculated and plotted in Figures 5.9 to 5.22. The actual raingauge network is not arranged on a regular lattice but this has often been assumed in order to obtain an indication of the raingauge density required to attain a specified accuracy. Here it is assumed for illustrative purposes that a network of gauges is to be sited on a regular lattice consisting of equilateral triangles, and that the major purpose of the network is to provide point rainfall information over the whole area. The graphs in Figures 5.9-5.15 may then be used to decide the maximum size of triangle that can be permitted to give a specified accuracy or to decide the necessary spacing of gauges in a triangular lattice.

As an example, interpolation to an ungauged point will be performed using the simple interpolation method of Section 5.2.3, using only information from the gauges at the vertices of the triangle in which the point lies. The accuracy quoted in Figures 5.9-5.15 is the accuracy of interpolation to the centroid of the triangle. The accuracy of interpolation to points near the edges (in some undefined sense) of a triangle can be larger than at the centre but it may be assumed that they would be treated using an interpolation based on points outside the triangle under consideration.

The accuracy of interpolation, which in Figures 5.9-5.22 is in terms of fractions of the standard deviation of rainfall in the relevant periods, may be expressed in the measured units of rainfall using information about sample standard deviations. Some sample standard deviations are given in Table 1.1. In practice these estimates are calculated at the same time as the correlations.

The values of accuracy as plotted in Figures 5.9–5.22 were derived on the assumption that the correlation functions fitted, as described in Section 4, are the true functions. However these are simply estimates of the true functions and are themselves subject to sampling variations. It would be incorrect to read these graphs as giving exact values for the accuracy of interpolation. Some idea of the sampling variation is given by the spread between the lines corresponding to different stations, although this may also reflect a true spatial variation.

As an example of the implications that may be drawn concerning network densities, consider three stations at the vertices of an equilateral triangle of side 10 kms. The fraction of the standard deviation that is the accuracy of simple interpolation to the mid-point of the triangle for monthly data is between about 0.15 and 0.18 (from Figure 5.10).

Assuming a standard deviation of 24 mm, which is typical for monthly total rainfalls in Eastern England, the standard error of interpolation is between 3.6 mm and 4.3 mm, the monthly mean rainfall in this area being about 50 mm. Similar calculations may be performed for daily and annual values, and for other spacings and configurations of stations.

This method may be used to assess the network needed to meet the requirements given in Table 3.2. For example, if the flood design requirement is interpreted as being that interpolation should be to within 10% of the mean value on 95% of the occasions when more than 10 mm of rainfall occurs over a large area on that day, then the spacing of the gauges required is 1 km or less on the assumption that the mean rainfall on such occasions is 15 mm, with standard deviation 9 mm. Similarly, the requirement for general enquiries about annual rainfall (accuracy 25 mm, 98% of the time) can be met with an inter-gauge spacing of 6 km to 9 km in Eastern England (mean rainfall 575 mm, standard deviation 85 mm).

More results of this nature will be described in Section 7.4.2, but first similar calculations for areal rather than point interpolated values are described.

7.2.2 Areal averaging

The network required to enable estimates to be made of areal average rainfall of given accuracies can be considered using the method of Section 7.2.1. Again, a regular arrangement of gauges has usually been treated, one such arrangement being the square lattice discussed in Section 5.3. The accuracy of an estimate of the areal average rainfall over a particular area based on measurements at particular stations can be calculated using the formulae of Section 5.3. However, it should be noted that in this section a square lattice was assumed, and also that the gauges were configured in the area considered for interpolation in a particular way.

If the requirement for an estimator of areal average monthly rainfall over a 10 km x 10 km square is that it should be within 10% of the true value 98% of the time, then Figure 5.2.4 indicates that this can be met with a single raingauge at the centre of the square, assuming that the mean rainfall is 50 km with a standard deviation of 24 mm. An estimator with the same accuracy of the average rainfall over a 100 km x 100 km square can be obtained with a spacing of just over 40 km between gauges: nine gauges with a regular spacing of 33 km would suffice.

7.2.3 Comparison of existing network with user requirements

Table 3.2 gave a list of required accuracies as specified by users of rainfall information. As already shown in the preceding sections, this information can be used in conjunction with Figures 5.9–5.27 to give some indication of the required maximum spacing of raingauges. Table 7.1 was compiled using this method. It relates the requirements of users to the spacing necessary to meet those requirements, assuming that the raingauges are arranged on a triangular lattice for point interpolation, and a square lattice for areal interpolation. Alternatively, for point interpolation, the distance given as the required spacing can be interpreted as giving the size of the largest triangle of gauges that can be allowed in the area if the requirement for interpolation accuracy is to be met.

It may be seen that network densities based on the requirements of areal interpolation are slightly less than those based on point interpolation. It may also be seen that the required network density in Northern areas is generally slightly greater than in Eastern England. The network densities required differ widely between uses.

Table 7.1 may be compared with Figure 1.1, which gives the mean spacing of the existing raingauge network in Great Britain. The actual network is not arranged on a triangular or square lattice, and so comparisons are necessarily approximate, quite apart from the approximate nature of the figures for required spacings. Nevertheless, it can be seen that the most stringent requirements are met nowhere in the existing network, apart from in local clusters. The only requirements to be met everywhere are those for long range weather forecasting and for general enquiries to the Meteorological Office for point rainfall information on rain days. The requirement for daily areal data for the proposed soil moisture deficit calculations and for point information for monthly totals are met in only a few locations, while the requirements for areal monthly information for water balances and point information for general enquiries about annual data are not met in the Chilterns, Cornwall and parts of Devon, west Wales, the Vale of York and North York Moors, and the Southern Uplands and Highlands of Scotland.

These tentative conclusions are subject to many assumptions, and should not be taken to be at all definitive. The assumptions of stationarity both of the mean rainfall and of the parameters of the correlation function have not been examined in many of the areas listed. In calculating the figures of Table 7.1, the percentages in the last column of Table 3.2 have been taken as equivalent to 2 standard deviations (95%) and 2.33 standard deviations (98%). Strictly this interpretation holds only for the Normal distribution but, in view of the subjective nature of the requirements of the table, the results obtained should not be misleading. As has already been stated the network densities required for rainfall data for different purposes vary greatly. It may be that some users have overstated their requirements for accuracy, although there are real differences in the accuracies required for different users.

Table 7.1 Required spacing of gauges needed to meet different uses

Use	Data	Space	rmse req'd mm	mean for interval mm	st dev for interval mm	fract st dev req'd	required spacing (kms)	Region
Flood design	day 10 mm	point	0.76	15.0	9.0	0.085	0.9 - 1.2	E
			0.87	17.0	10.0	0.087	0.9 - 1.2	N
Seed germination	1 month	point	2.15	50.0	24.0	0.090	2 - 4	E
			2.15	60.0	30.0	0.072	1.4 - 3	N
General enquiries	day 2 mm	point	2.15	5.5	6.0	0.36	20 - 29	E
			2.15	6.0	6.5	0.33	18 - 25	N
General enquiries	year	point	10.7	600.0	83.0	0.13	5 - 7.5	E
			10.7	720.0	104.0	0.103	3 - 5	N
Water balance	month	10 km x 10 km	2.14	50.0	24.0	0.089	7.5 - 9	E
			2.57	60.0	30.0	0.086	7.5 - 9	N
SMD map	1 day arbitrary	20 km x 20 km	0.04	1.6	3.0	0.013	3.5 - 4	E
			0.05	2.0	3.7	0.013	3.5 - 4	N
Long range forecast	½ month	100 km x 100 km	0.64	25.0	17.0	0.038	20 - 25	E
			0.76	30.0	21.0	0.036	20 - 25	N

The overall mean spacing of the existing network is of the same order as required by some users but is greater than that required to meet other users' requirements. In particular local areas of the network, users' requirements for accuracy may or may not be met depending on whether the network is more or less dense than the average for the country. The network densities suggested in this section have been calculated on the basis that the rainfall will be recorded for all periods at all gauges in the network. This means that the densities required in practice would be still greater than those stated here - some have suggested that the required densities should be doubled but this is possibly rather extreme. These considerations suggest that the raingauge network needs augmentation rather than reduction.

These conclusions are corroborated by the work of the Meteorological Office, which has applied the localised analysis described in Section 4.5 to a region in Northern England centred near Huddersfield and covering the part of the South Lancashire Plain and Southwest Yorkshire. As described earlier a separate analysis is performed for the group of gauges situated within 25 km of each chosen point. For the above region a 15 km square grid plan was used, each vertex of the grid being used as the centre point for a 25 km circle and analyses for each circle being separate. In the contour diagrams that follow, values for each group of stations were plotted at the centroid of the group rather than at the centre of the original circle. Figure 7.1 shows a map of the distance between raingauges in the area for the period 1969-1974: a similar map, for mainland Britain, of the current network is presented in Figure 1.1. These maps were obtained by calculating for each circle the mean intergauge spacing d

$$\bar{d} = \left(\frac{\pi 25^2}{N}\right)^{\frac{1}{2}} \text{ kms},$$

where N denotes the number of gauges within the 25 km circle. The map of Figure 7.1 shows a maximum density near Shipley, falling away steadily east into Yorkshire. However, note that this is the density of the gauges recording for more than 365 days over the period 1969-74 and this is not necessarily the same as the current network represented by Figure 1.1.

The analysis described in Section 4.5 was performed using daily rainfall for each circle, providing local estimates of the correlation function and noise index. From these, it is possible to calculate the interpolation error as in Section 5.2. Results are presented for the relative mean square error of optimal interpolation (assuming known means), ξ_{opt}, to the mid-points of regularly arranged equilateral triangles allowing for the existence of measurement errors (Sections 1.4.3 and 5.2.2; Gandin, 1970, equation 50); this quantity is the square of the measure of accuracy used in Section 5 (equation 5.20)) and in Figures 5.1-5.27. Figure 7.2 shows a map of the spacing required to give a maximum relative mean square error (i.e. ξ_{opt}) of 0.05; the variations in spacing are due to variations in the estimated correlation parameters. Figures 7.3 and 7.4 show maps of the interpolation error if the raingauge densities are those presented in Figure 7.1; this assumes that the existing gauges would be rearranged onto a locally regular triangular grid of the same average density. Figures 7.3 and 7.4 are equivalent: the first is in terms of fractions of the standard deviation of rainfall, while Figure 7.4 is in terms of millimetres of rainfall. The interpolation error increases in areas of higher relief. Figure 7.5 shows the variance of the rainfall over the area, and the estimated measurement error index,

$\bar{\lambda}^2$, (averaged locally) is plotted in Figure 7.6. A graph of the interpolation mean square error index (ξ_{opt}) plotted against gauge spacing is shown in Figure 7.7, for a particular estimated correlation function. It may be seen from this latter figure that the required spacings for a given accuracy criterion are of the same order as for the analyses described earlier. However, there are moderately large geographical variations of measurement error: in particular, a higher network density is required to the east of the Pennines, in Yorkshire, than in Lancashire to the west. This probably reflects the type of rainfall experienced and indicates the order of variations in required density that may occur over short distances.

7.3 Implications for network design using indirect methods

7.3.1 Daily rainfall-runoff modelling

In Section 6.2 a lumped input rainfall-runoff model was applied to six catchments in the U.K. with a view to determining raingauge density requirements for streamflow prediction. Here, prediction is associated with the behaviour of streamflow in the longer term, as opposed to short-term forecasting, which was the subject of Section 6.3. However, some of the results obtained in Section 6.3 have implications for the modelling carried out in Section 6.2, and so will be considered here.

Raingauge network requirements for streamflow prediction would be expected to be primarily influenced by the spatial variability of rainfall over a catchment, and by the nature of the catchment response to rainfall. If the response is heavily damped, reflecting a large groundwater discharge component in runoff (e.g. chalk catchments), then, intuitively, network density requirements would be expected to be less than for catchments where the response to rainfall is governed by surface runoff. Spatial variability in rainfall would be expected to be most pronounced where the topography changes rapidly; this may result in significant spatial variability in runoff, particularly if catchment vegetation and soil cover also change significantly over space.

In Section 6.2, it has been assumed that catchment response can be adequately represented by the non-linear CLS model with a spatially lumped input. The effect of spatial variability in rainfall on runoff prediction enters only through the lumped rainfall input. In assessing the effect of network density on streamflow prediction, it has been assumed that the maximum number of gauges (in and adjacent to a catchment) provides the 'best' estimate of model input; this assumes that the lumped input should estimate as accurately as possible mean areal rainfall over the catchment. However, it is possible that a lumped input defined from a sub-set of gauges may be more highly correlated with streamflow than that provided by the total number of gauges, as observed in Section 6.3. If this is the case, then the network density problem could be posed as follows: how dense should a set of gauges be, and where should they be positioned, before a sub-set is obtained which provides a desired level of accuracy in streamflow prediction? While the location of gauges may appear relatively unimportant for defining a lumped input, some of the results obtained in Section 6.3 suggest that this might not be so, at least for short-term forecasting. Thus, future work on the effect of network density on streamflow prediction should explore this question more fully.

From the experiments conducted in Section 6.2.5, it has not been possible to establish what the relative influences of catchment response and the spatial variability of rainfall are on network density requirements for streamflow prediction. The results for the catchments in Devon (the Exe, Culm and Dart) suggest that the raingauge network in that area is adequate for streamflow prediction. This may be because the spatial variability of rainfall in these catchments is not significant; a further factor worth considering is that the network in this area has been formally designed (Shaw, 1965). The results for the Wyre also indicate that the present density is satisfactory. For the Ure and North Tyne, however, there is distinct evidence that if a higher network density were available, then higher values of R^2 could be obtained; this is borne out by the fact that the lowest values of R^2 were obtained on both these catchments. It is not possible to assess if this is because of the basic nature of the catchment responses, or because the spatial variability of rainfall over these catchments is relatively high.

Accordingly, the results obtained in Section 6.2 suggest that, in some areas, the present network is adequate for streamflow prediction, but is inadequate in others. This conclusion is similar to those resulting from the use of direct methods of network evaluation, in that the network has been found to satisfy most requirements in some areas, but not in others. An extensive application of the methods of Section 6.2 and of direct methods on a localised scale would allow an interesting comparison of the areas found to be deficient by each approach.

7.3.2 Real-time flow forecasting

The number and siting of gauges for real-time flow forecasting should pay regard to the area of the catchment contributing to the rising limb and peak of the hydrograph. This area will depend on the physical characteristics peculiar to individual catchments, and no firm general guidance can be given. Important factors affecting the network density required will include catchment area, overland and channel slope, the nature of the valley bottom, and the storage characteristics of soil and bedrock. More rigid guidelines for network design may be obtained by investigating a range of catchments which differ in the physical attributes itemised above. If the network is to be used specifically for real-time flow forecasting, the results obtained in Section 6.3 suggest that the design should not follow procedures which attempt to estimate the catchment areal average. This result merits further exploration using data from a range of catchments.

Given a network of raingauges, these may be best used for real-time flow forecasting if the following points are borne in mind:

(i) The indiscriminate use of all gauges in and around the catchment as input to a lumped model may result in sub-optimal forecasts.

(ii) Only those gauges which have a strong causal link with variations in flow should be used. A suitable selection procedure is given in this report (Section 6.3.4).

(iii) Model calibration should use only those gauges transmitting rainfall data in real time.

(iv) Benefits may accrue if a spatially distributed model is used. An extension of the work described herein to larger catchments is required to assess more clearly the potential benefits of spatially distributed models.

7.4 Methods of network modification

7.4.1 Mapping of interpolation error

The methods of obtaining a required network density discussed in the last section are based on the assumption of a regular arrangement of gauges and are thus difficult to apply in practice to the existing irregular network. More direct methods of modifying an existing network are available. Figure 7.8 shows the raingauges operating in a 20 km x 20 km square near Northampton (in 1965), these being 20 in number. It is assumed that measurements at those twenty gauges are to be used to estimate point rainfall within the area and that the optimal linear estimator (assuming a constant but unknown mean) is employed as the interpolator. The contours on this map show the accuracy (as a fraction of the standard deviation of point rainfall) of interpolation to all points in the area. Here the correlation function was taken as being that fitted to a station just outside the area by the methods described in Section 4. By considering such a contour map it is possible to identify the areas for which interpolation performs least well and to consider whether the accuracy of interpolation is sufficient at any point. Points outside the immediate area covered by the gauges would usually be disregarded but the points of locally highest interpolation error are clearly candidates for the positioning of new gauges. Network reduction could be effected by assessing the corresponding maps of interpolation accuracy when various gauges, which are candidates for removal, are deleted.

Figure 7.8 is based on a particular correlation function but it is clear that the points of highest interpolation error would be almost the same for a fairly wide range of such functions - specifically points relatively distant from any existing gauge. Ideally, the calculations made to produce Figure 7.8 would be repeated including all the gauges in the national network; however, this is not computationally feasible. A practical alternative would be to produce maps for a large number of overlapping areas each containing about 20 or 30 existing raingauges.

It is clear from Figure 7.8 that because the raingauge network is not arranged on a regular lattice, the lowest accuracies of point and areal interpolation will be worse than those for an identical number of gauges arranged on a regular lattice. As a result, even greater network densities would be required than those given in Table 7.1.

7.4.2 Information Transfer

This subsection gives a brief outline of a method of network assessment which is directed towards the reduction of an existing network of gauges. The method aims to choose which gauges should be deleted from an existing network. This is done by balancing two opposing requirements: firstly, the amount of information gathered by the resulting network should be as large as possible, and secondly, the cost of running the network should be kept in check. Thus, given both the costs of keeping the individual gauges in operation and a set of past records from the gauges, the procedure evaluates the extent to which the information recorded by different gauges is duplicated. On this basis, gauges which should be discontinued are identified so that the total running cost of the reduced network is less than some specified budgetary constraint, and so that the amount of *information gathered by the reduced network is maximised.*

The method is designed to cover a situation in which some long-term properties of the statistical distributions at different points are to be estimated: e.g. the long term mean rainfalls at various places. It may be possible to modify the procedure to deal directly with estimating the rainfalls actually occurring. A brief technical description of the concepts of information transfer analysis is given in Appendix D: the method is more fully described in Maddock (1974) and the results of analyses of some streamflow gauging networks are given by Carrigan and Golden (1975). A programme for applying the technique has been written within the Surface Water Branch of the U.S. Geological Survey: unfortunately, this program did not become available until April 1977 and consequently an application was not possible in the time available.

7.5 Summary

The requirements for accuracy by users of rainfall data, reported in Section 3, have been compared with the accuracies obtainable from the existing network of raingauges. This has been done by assuming that interpolated values of point rainfall and estimates of areal average rainfalls are calculated in certain simple and easily applied ways. It has been shown that, on this basis, some users requirements for accuracy are more than met by the existing average densities of raingauges, while, for certain users requiring data of greater statistical accuracy, the stated accuracy requirements would be met in only very few places where the existing network has large numbers of gauges in small areas.

By carrying out some rainfall-runoff modelling studies, it has been possible to carry out a very limited evaluation of the present network using an indirect approach. For daily streamflow prediction the results suggest that, for some catchments, the present density of gauges is adequate but not so for others. A study of raingauge network requirements for real-time flow forecasting suggests that, if a model with an efficient updating procedure is employed, one or two gauges sited strategically may be sufficient for real-time forecasting in small catchments.

Taking into consideration the very irregular placement of existing gauges and the possibility that data will not be recorded by every gauge for every day, the evaluation conducted here suggests that the present network is in need of a more thorough local reassessment. This would consider moving currently operating gauges to new sites as well as introducing new gauges. Two methods of modifying a network have been described: one considers maps of interpolation error and can be used to add or delete gauges; the second could be used to choose which of a locally dense set of gauges to discontinue giving due consideration to their operating costs.

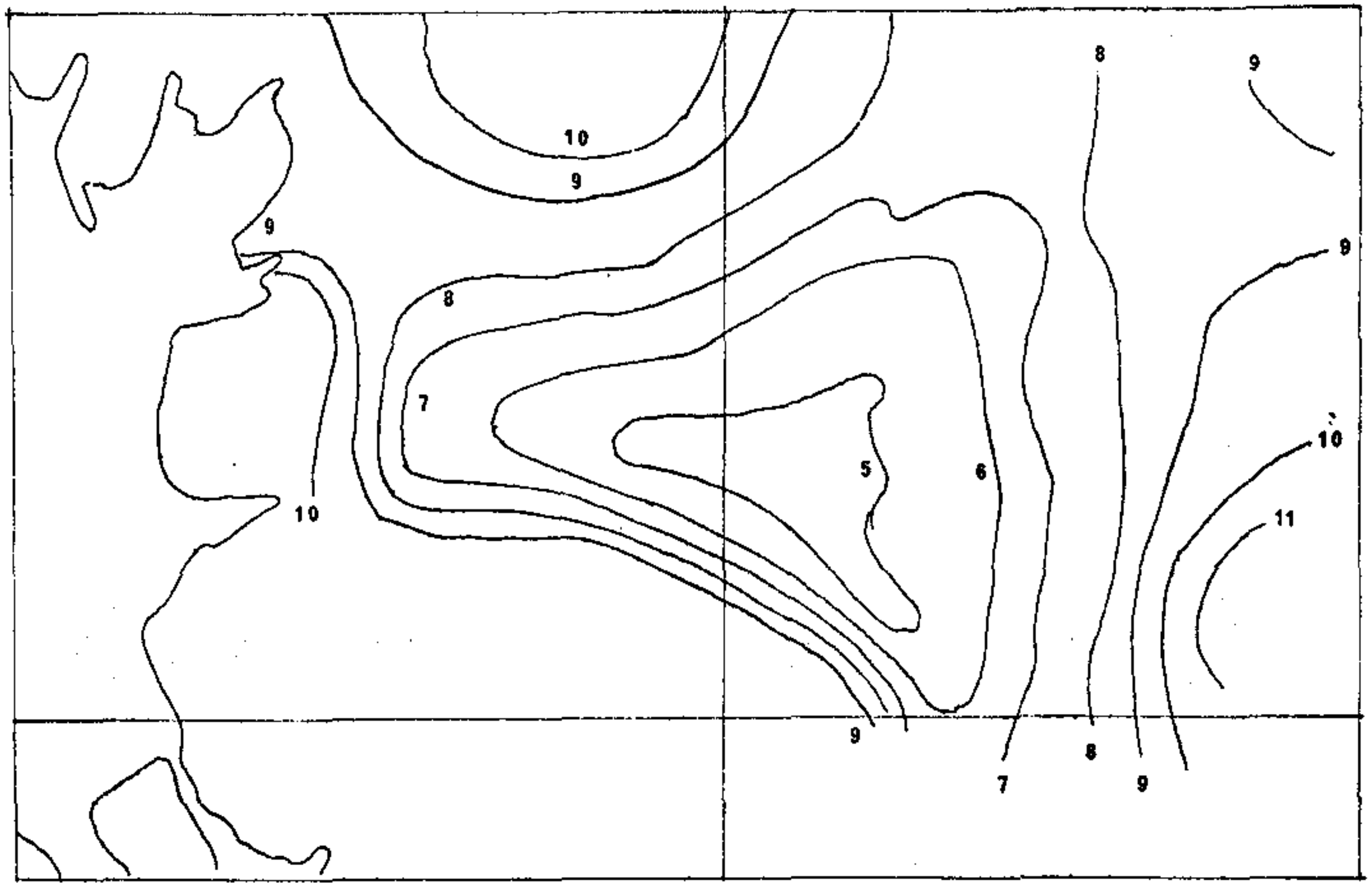


Figure 7.1: Approximate spacing for daily read raingauges used in the analyses after reduction for missing data. Spacing in kilometres.

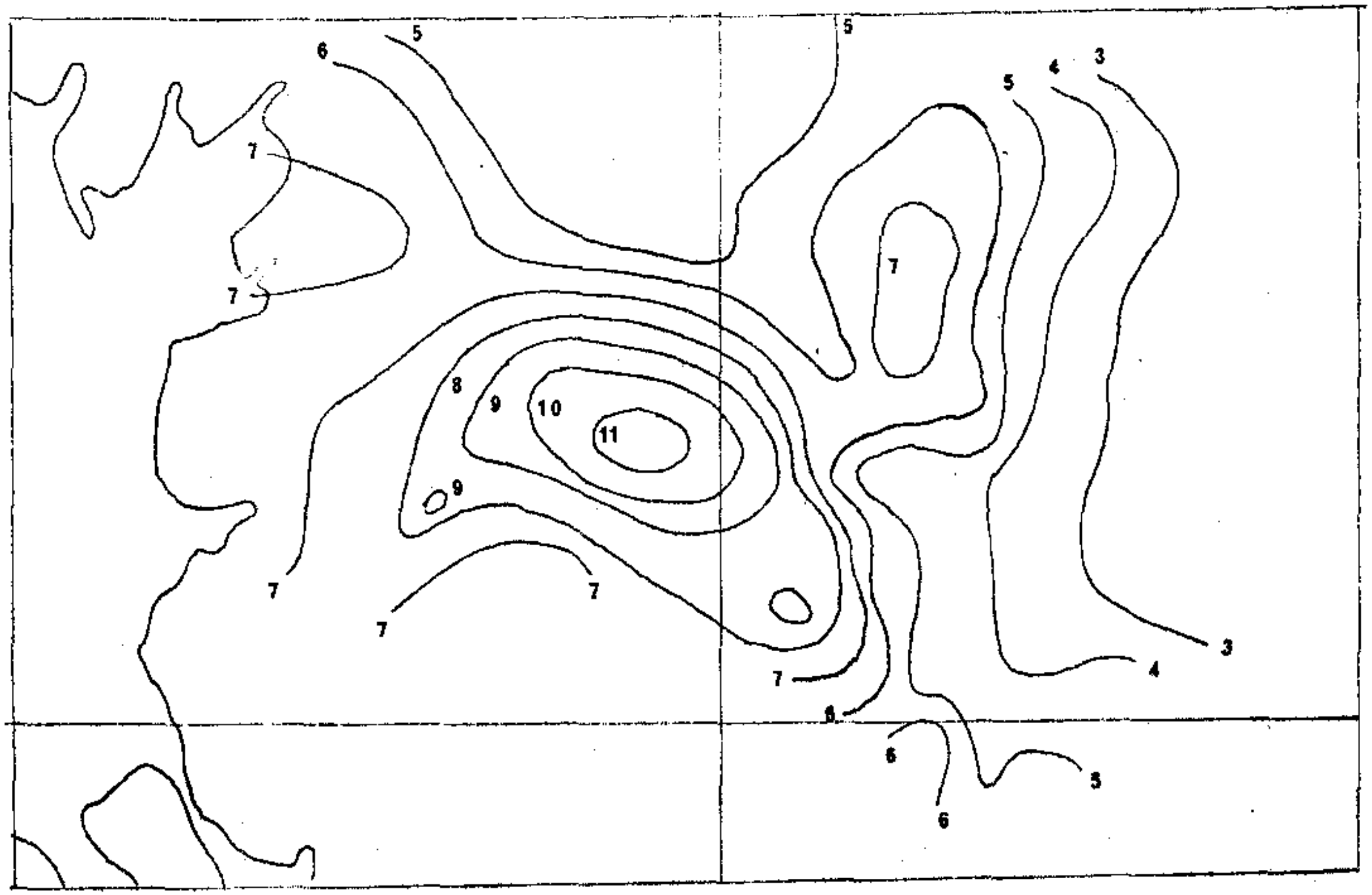


Figure 7.2: The required raingauge spacings (in kms) to attain an interpolation error index of $\xi_{opt} = 0.05$ ie an interpolation error of 0.224 of a standard deviation.

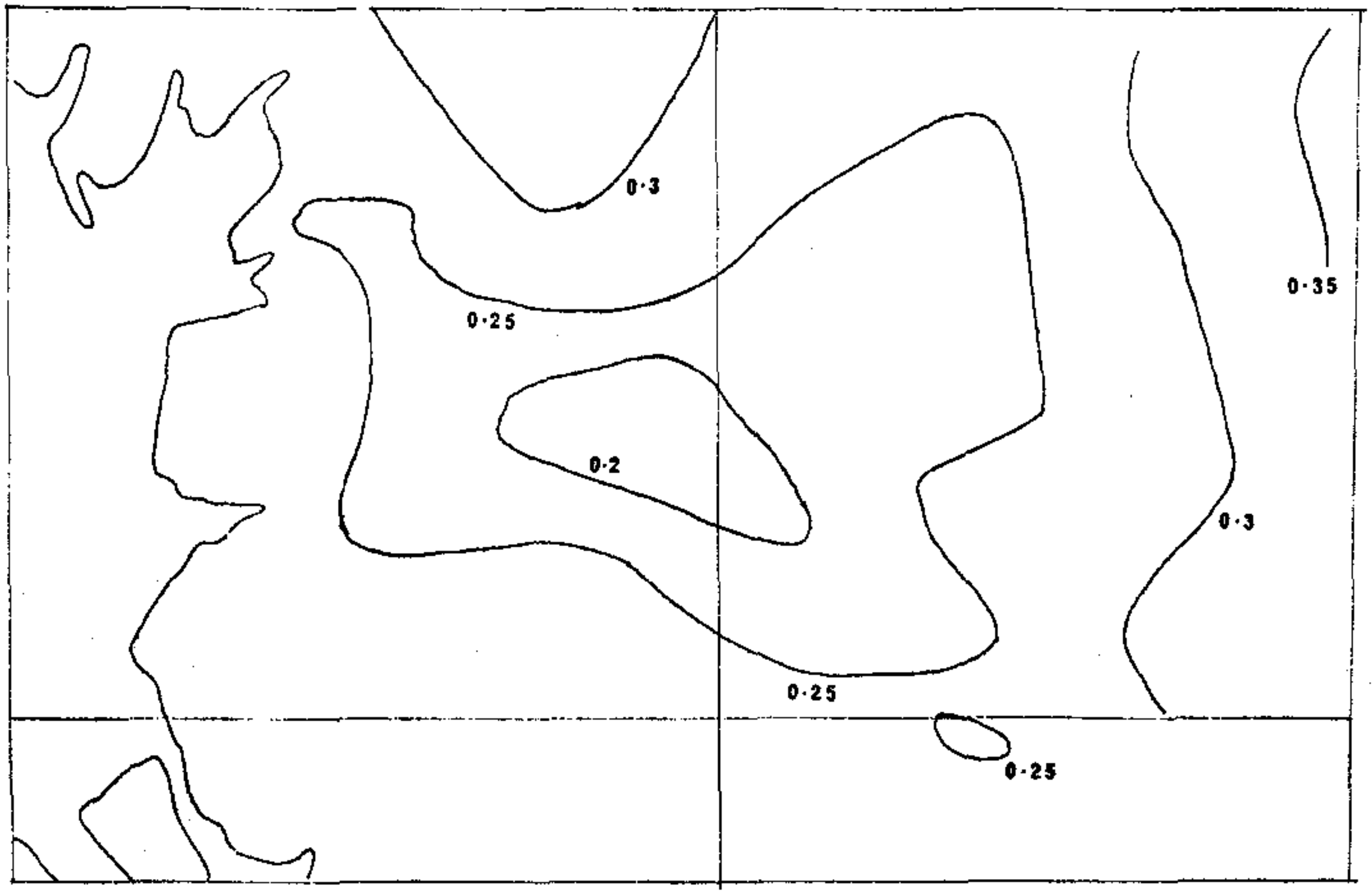


Figure 7.3: The interpolation error (as fraction of standard deviation) of a network with average raingauge densities as shown in fig'7.1.

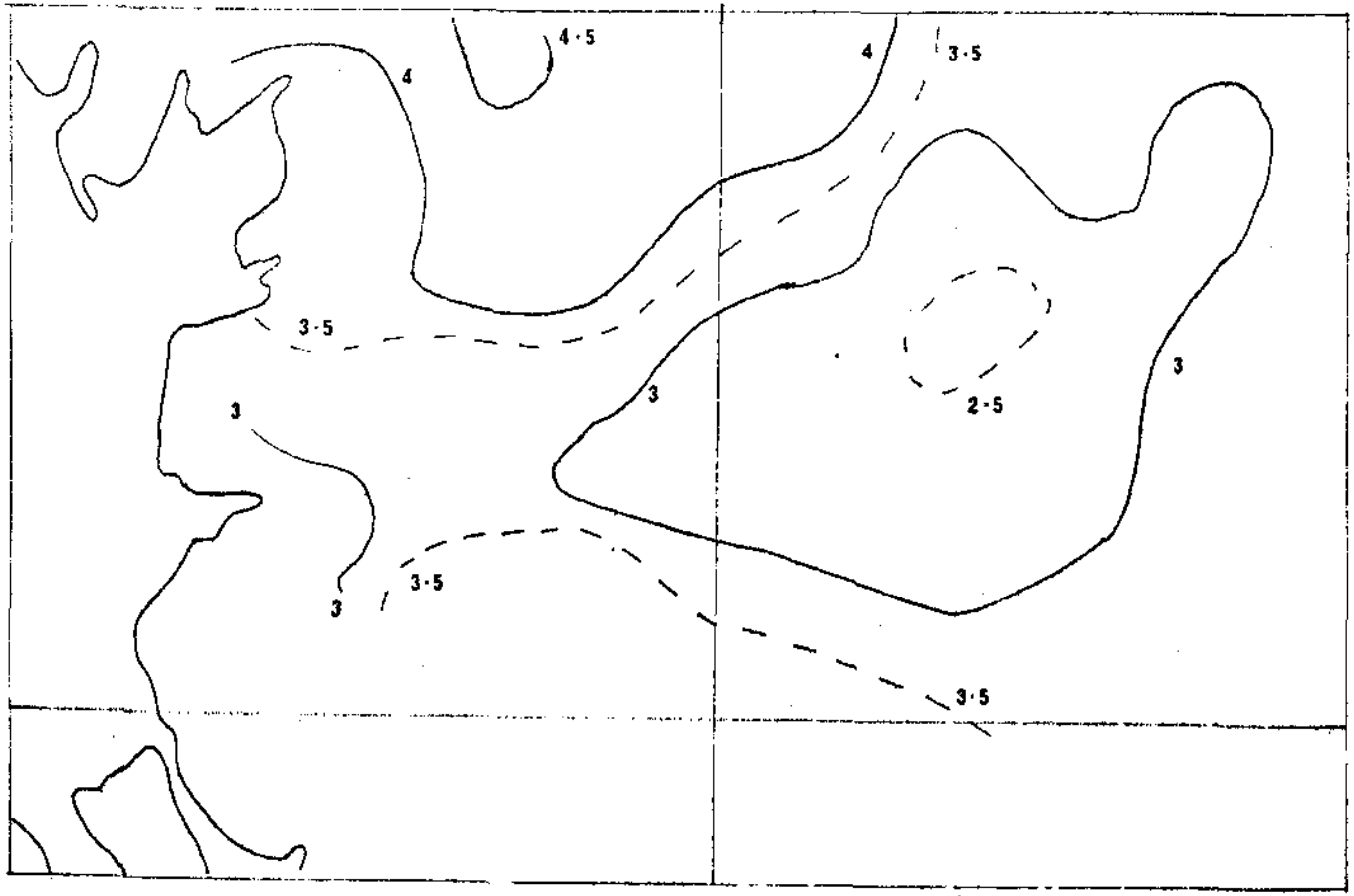


Figure 7.4: The interpolation error (in millimetres of rainfall) of a network with average raingauge densities as shown in fig 7.1.

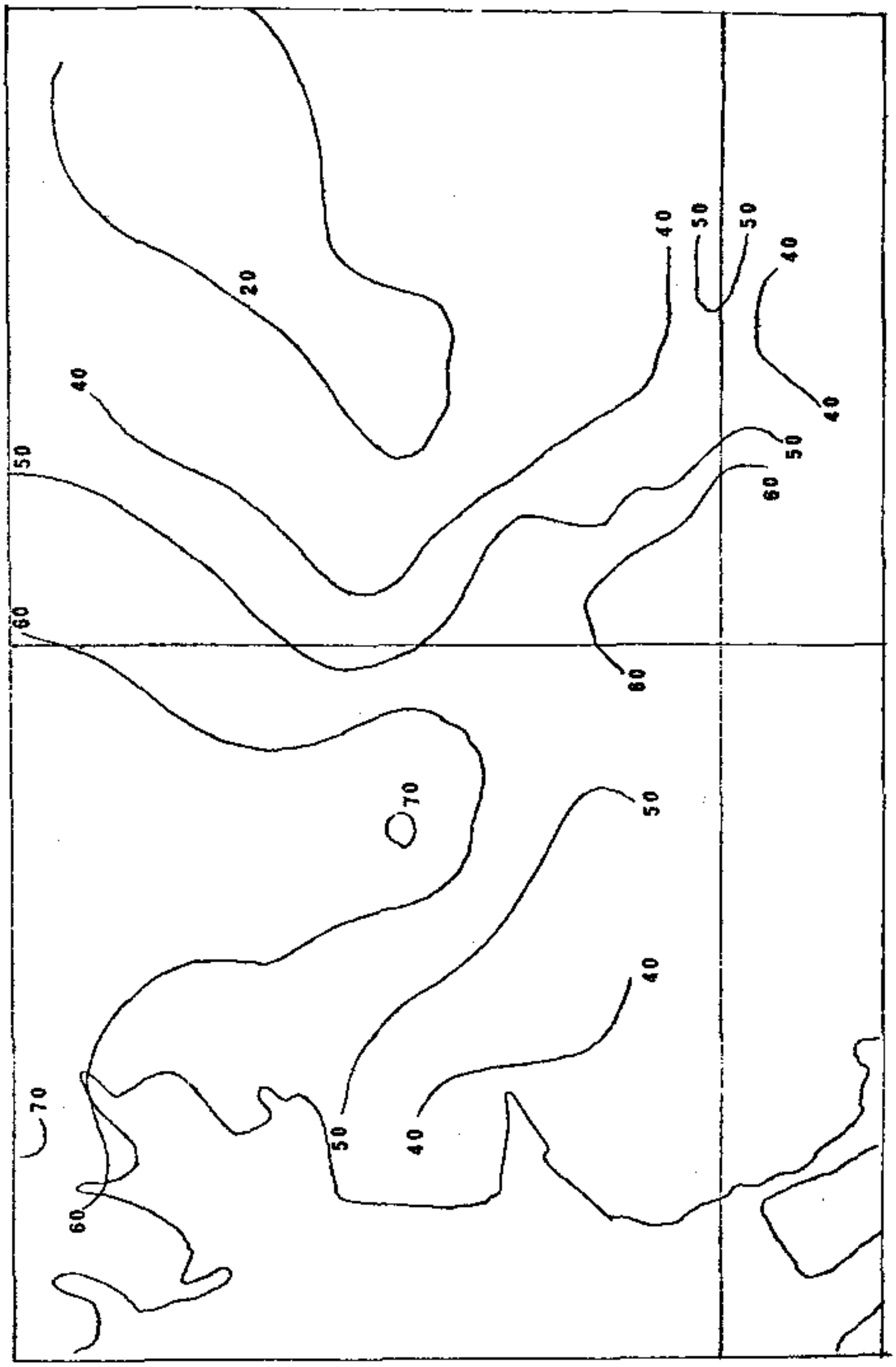


Figure 7.5: The variances of the measured rainfalls over the region.

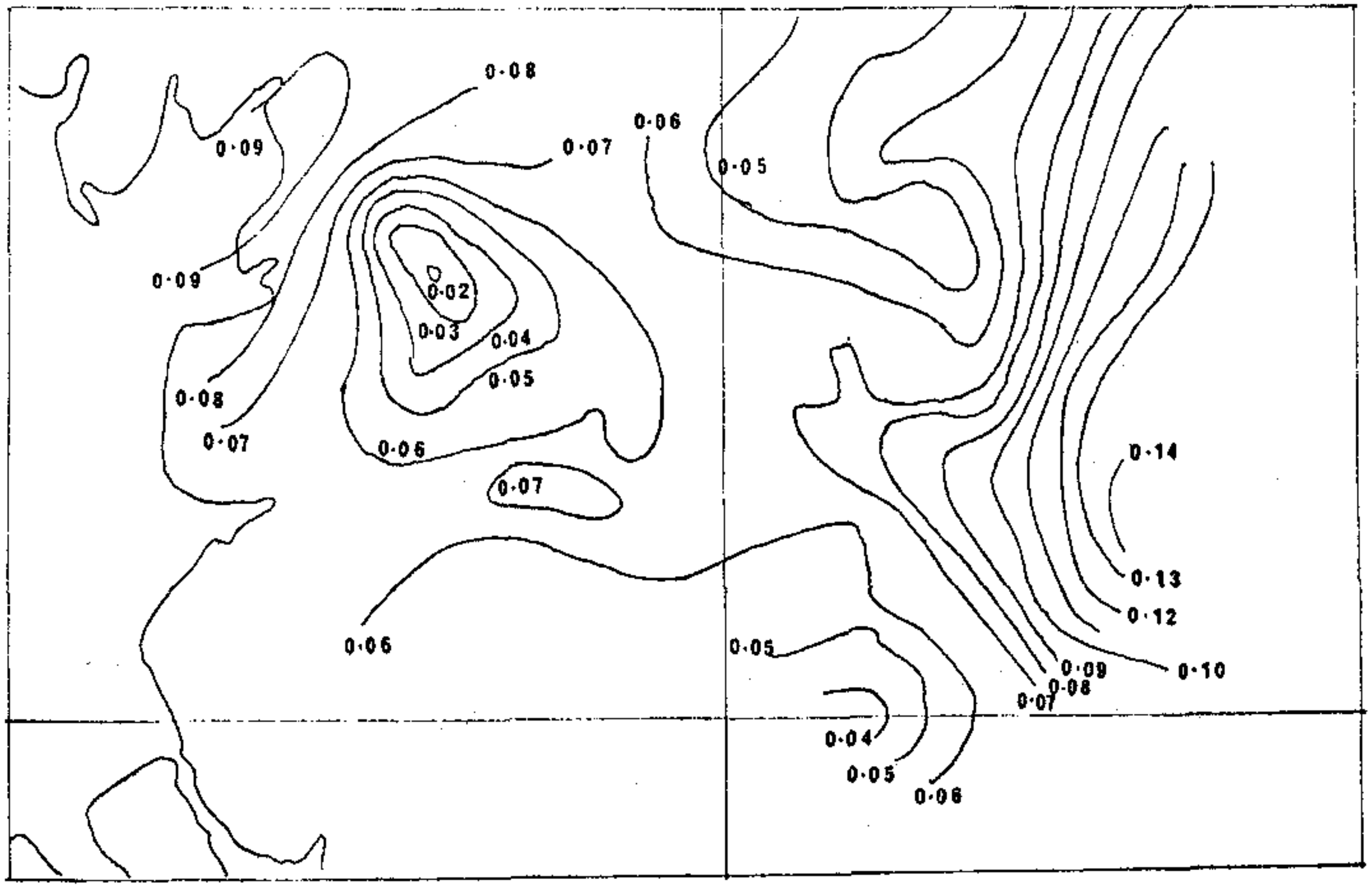
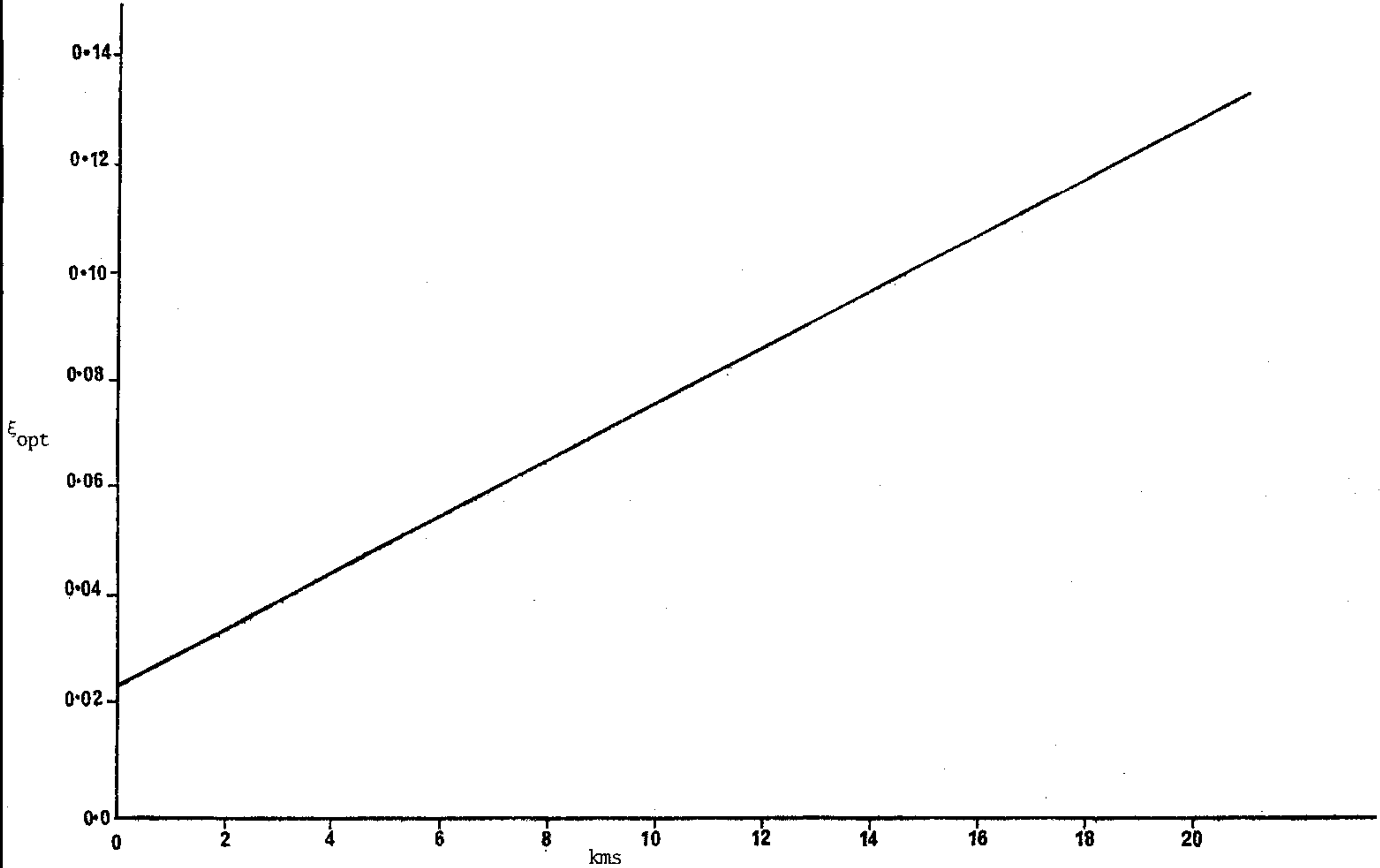


Figure 7.6: The average noise index $\bar{\lambda}^2$ over local areas of the region.

Figure 7.7: Interpolation error index ξ_{opt} as a function of gauge spacing showing effect of measurement error. Gauges sited at apexes of an equilateral triangle, correlation function of observed rainfalls: $\rho(d) = 0.93 - 0.00919d$.



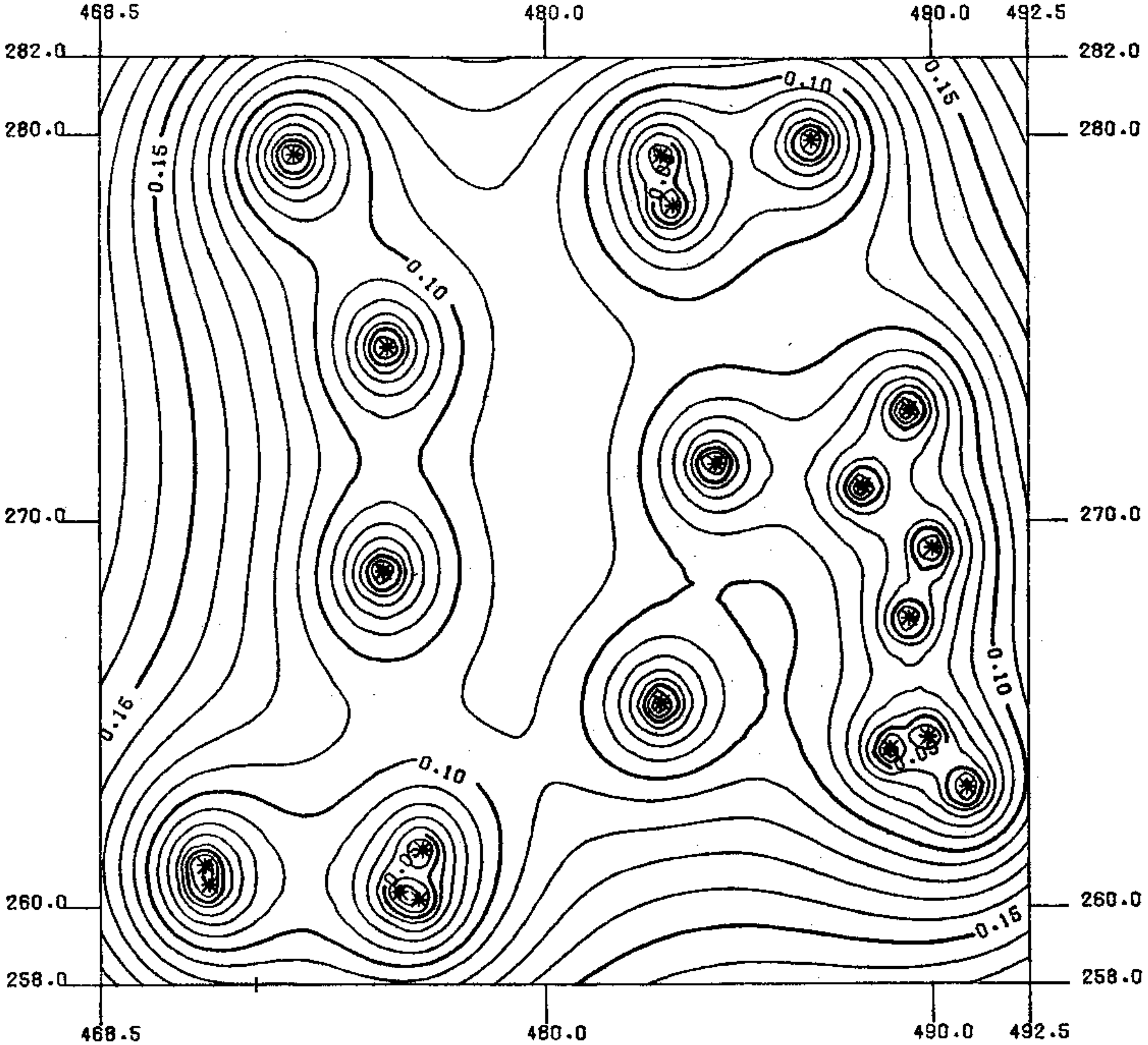


Figure 7.8: Contours of interpolation error: Optimal interpolation from twenty stations assuming constant but unknown mean rainfalls.

8. SUMMARY, CONCLUSIONS, AND SUGGESTIONS FOR FURTHER WORK

8.1 Summary

This report set out to investigate methods for determining the adequacy of the United Kingdom raingauge network in providing data on rainfall required by users. A survey of user requirements identified many organisations that make use either of their own locally gathered information or of the national archive of rainfall data. For most of these organisations, the amount of error that they were prepared to allow in estimated rainfall quantities could only be determined subjectively, and may have been overstated. The benefits of being able to obtain more accurate estimates had been analysed neither in monetary terms nor in terms of the sensitivity of some eventual decision on the rainfall information used in that decision. The possibility of interpolating rainfall values between raingauges was not fully appreciated by many users; they preferred to collect data at a particular site or to use, without adjustment, the data from a nearby gauge.

In the assessment of a raingauge network the issue of most immediate practical importance is whether a sufficiently close spacing of gauges is available to permit a user to interpolate rainfall at a point of interest with less than a tolerable amount of error. In the course of this report, many subsidiary problems have been identified and, while not all of these have been resolved, it has been possible to conclude that there is no clear case for a reduction in density of the UK raingauge network: indeed, an increased number of raingauges may be called for.

As well as considering records of rainfall for the current period, monthly totals of rainfall were obtained for a period in the late nineteenth century when "blocked high" pressure systems were a more prominent feature of the British climate than at present, in order to study whether medium to long term climatic fluctuations would markedly alter network requirements. The outcome of this part of the investigation was that the network requirements for the two situations are extremely similar. The data used for the remainder of the work was for the period 1961-1974, and consisted of daily, monthly and annual totals of rainfall; in addition, daily rainfalls were categorised according to the amount of rain falling over a large area. The accuracy of interpolation of each of these types of data was considered, together with the accuracy of estimates of the average total rainfall over an area for a particular month. These accuracies are related to the correlation between the rainfalls recorded at different sites and therefore spatial correlation analyses of the different types of data were performed. These lead to the fitting of spatial correlation functions which in turn allows the accuracy of interpolation from any particular set of gauges to be calculated.

Based on the assumption that, when interpolating to a point, only data from three nearby gauges would be used, a measure of the accuracy of a network has been devised; this can be presented in a convenient graphical form in which the interpolation error at the centre of a triangle is related to the size of the triangle. Thus, given a desired level of accuracy for the network, the maximum size of interpolation triangle that can be allowed in the network is determined. A similar relation between accuracy and spacing of gauges was also found for estimates of average rainfall over an area.

On comparing the stated requirements of accuracy of different types of user of rainfall information with the accuracies calculated by the above procedures corresponding to the spacing of the current network of raingauges, it was found that for some requirements the present network is adequate while for others it is not. A consideration of the gauges in a small area has suggested that the current network of gauges is so irregularly spaced that it is more relevant to consider the maximum inter-neighbour distance as representing the density (and thus the accuracy) of the network in a particular area rather than simply the number of gauges in an area.

Some suggestions have been made as to how an existing section of the network might be objectively analysed with a view to adding or deleting gauges. In practice, this would have the effect of eliminating the clustering effect noticeable in maps of the network and of placing gauges in the empty spaces. Thus a more regularly distributed network would be produced. However, the method is adaptable so that higher accuracy of interpolation can be specified for regions of particular importance.

The potentialities of radar as a measurement instrument for rainfall have been noted but the view was taken that a great deal of further investigation would be needed before radar could be used in place of a conventional raingauge network.

A second method of assessment is relevant when the main use of rainfall information is to provide data for indirect purposes; for example, for streamflow forecasting or prediction. Here the aim is to arrive at accurate estimates of quantities other than rainfall. The assessment of the adequacy of the UK raingauge network for indirect usage of rainfall information depends on the user being able to define his objectives, and thus on being able to state a tolerable error for the quantity (e.g. streamflow, soil moisture deficit) to be estimated. Provided a relationship between rainfall and the quantity of interest can be established either analytically or through simulation, then in principle this should allow an accuracy requirement for rainfall to be stipulated. This is rarely feasible using an analytical approach, but semi-empirical procedures can be adopted involving the use of rainfall-runoff models in experiments with actual records of data from different selections of existing raingauges; the accuracies of the results of these experiments are then compared. Clearly this approach usually involves a great deal of computational effort even when considering a small number of raingauges. Further there is no immediate way of evaluating the benefits of establishing new raingauges at specific but previously ungauged points. Certain of the indirect uses of rainfall information (e.g. for flood forecasting) require measurements made at short intervals ($\frac{1}{2}$ hr–$\frac{1}{2}$ day) and in these circumstances the network of raingauges recording daily totals is not of direct relevance. However, the design of networks of autographic or telemetering gauges is a problem which, for reasons stated in Sections 1.4.1 and 3.3.1, presents difficulties when the conventional methods of Sections 4 and 5 are applied. Accordingly, a design procedure is required which is based specifically on the use which is to be made of the network. A typical use of short duration rainfall data is for flow forecasting in real time, and this has been taken as an example to illustrate how the design of the required network should be approached. By employing a daily rainfall-runoff model, an evaluation of the daily raingauge network can be made on the basis of the accuracy with which streamflow can be predicted.

The daily rainfall-runoff model employed in this study was of a non-linear 'black-box' type with low computational requirements and statistically efficient parameter estimation procedures. The model was applied to six catchments in the North and South-West of England; reasonably good agreement between observed and computed streamflow was obtained in each case. Experiments carried out with sub-sets of gauges indicated that the present network density was adequate for this purpose in four of the six cases, but inadequate for the remaining two. This may reflect the particular rainfall-runoff regimes on these catchments, or greater spatial variability of rainfall in these areas.

The approach adopted to assess the rainfall requirements for real-time flow forecasting was to investigate, by direct statistical inference, the strength of the linear causal relation between rainfall measured at each gauge and the river flow. The strength and nature of this causal relationship was used to indicate the importance of each gauge in determining the flow response, and to identify spatially distributed models of increasing complexity until the inclusion of further gauges provided little additional information to explain variations in river flow. The procedure was applied to one small catchment only, the Hirnant in North Wales, where a dense raingauge network had been set up as part of the Dee Weather Radar Project. Although no firm conclusions can be drawn from this limited study, the results did suggest the following: the number of gauges required and their optimal location will depend on the area contributing to the rising limb and peak of the storm hydrograph where most of the uncertainty lies. Because this area is typically confined to a restricted area near to the channel, and towards the catchment outlet for small catchments, it may adequately be sampled by only one or two strategically placed gauges. The indiscriminate aggregation of rainfall from all gauges in and around a catchment to define a lumped input to a flow forecasting model may result in poorer forecasts than that obtained if a careful gauge selection procedure had been adopted initially. Operational considerations appear to loom large in real-time when instrument and/or telemetry failure may necessitate a denser network, or gauge duplication, to ensure adequate information is transmitted at a given level of reliability. The results obtained for both daily and flow-forecasting models depend on the model used; a larger number of catchments would need to be studied, and perhaps more than one model applied before general inferences could be made about network densities and configurations on a country wide basis.

8.2 Conclusions

The investigations described in this report allow conclusions to be drawn about the adequacy of the existing UK raingauge network as well as about technical aspects of the analyses. The conclusions relating to the former will be presented first.

1. On the basis of the errors which users are prepared to tolerate in point or areal estimates of rainfall measured over various durations, the present UK raingauge network

(i) provides the required accuracy everywhere for the following uses:

(a) Long-range meteorological forecasts

(b) General enquiries for daily rainfall

(ii) provides the required accuracy in some areas (these are listed in Section 7.2.3) for the following uses:

(a) Water balances

(b) General enquiries for annual data

(iii) provides the required accuracy in only a few local areas for the following uses

(a) Soil moisture deficit map

(b) Seed germination

(iv) does not provide the required accuracy anywhere for the following use:

(a) Flood design.

The above conclusions are based on analyses which assume that the network is arranged on a regular triangular grid; the requirement is said to be met if the average spacing of gauges in an area is less than the maximum spacing allowable in the triangular grid. Thus, while a user requirement may or may not be met on average in an area, local variations in density may sometimes preclude this.

2. On the basis of the foregoing conclusion, there would appear to be no grounds for any significant reduction in density of the UK raingauge network. If any modification to the network is to be considered, then user requirements suggest that it should be augmented by adding new gauges or, if this is not feasible, transferring gauges from areas of locally high density to areas of low density.

3. The conclusion that the present network does not meet some users' requirements could reflect the fact that these have been overstated. If their present operations are not adversely affected by the lack of accuracy in rainfall data then this would be the case. However, some users' accuracy requirements may reflect anticipated future demands to be placed on the network.

4. If an evaluation of the network is required for a specific area, then the 'average' accuracies obtained from the analyses carried out in Sections 4 and 5 should not be used as a basis for design; rather, a more detailed localized analysis should be carried out which would consist of:

(i) identifying and quantifying as far as possible the user requirements for the specific area in question.

(ii) mapping the interpolation error for the area as described in Section 7.4.1. This should take into account specifically the local correlation structure of the rainfall process and the configuration of gauges, and would allow the identification of areas of maximum interpolation error. These would then be candidates for the siting of new gauges.

(iii) if network reduction must be considered, and the budget available for running the reduced network can be specified, then network reduction should be undertaken such that the amount of statistical information on the rainfall process provided by the network is maximized subject to the budgetary constraint.

5. The effective density of the network at any time is less than the nominal figure because of inoperative gauges. Any modifications to be applied to the network should take account of this.

6. The present network probably satisfies the input requirements of a daily rainfall-runoff model on some catchments, but not on others.

7. In the design of special networks of autographic or telemetering gauges, the reliability of the gauges is likely to be a major consideration in determining network requirements. For small catchments, one or two functioning gauges strategically sited should be sufficient for flow forecasting in real time.

The above conclusions have been corroborated where relevant by the work of the Meteorological Office.

The following conclusions pertain to technical aspects of the analyses carried out in this report:

1. The spatial correlation functions fitted to data for the East and North of England all display anisotropy in varying degrees. It has not been possible to investigate whether spatial variation of the estimated parameters (which was not large) was due to a sampling effect or to non-stationarity.

2. The rate of decay of correlation with distance is faster for daily data than for monthly and annual data; on days with large amounts of rainfall, the decay is much faster than on days with low to moderate rainfall.

3. The spatial correlograms for the North of England suggest that quite large changes in correlation may occur over short distances across the Pennines.

4. For most spatial correlograms, the effort of measurement error on the estimated correlations was not clearly discernible. Accordingly, it is to be expected that the overall effect of measurement error would not be significant, although local exceptions to this may occur.

5. Spatial correlation functions fitted to data for a period with a markedly different climatic regime did not show any significant difference from functions fitted to more recent data.

6. Under a number of assumptions for mean rainfall at each gauge, optimal point interpolation to the centre of an equilateral triangle formed by three gauges was not found to be significantly better (in terms of mean square error) than that obtained with the use of a simple linear estimator. It was also observed that the maximum error may not occur at the centre of the triangle.

7. Where the spatial correlation function is anisotropic, the shape and orientation of a triangle will influence the error of interpolation to the centre of the triangle.

8. The variation in the parameters of the spatial correlation functions within and between the selected regions in the East and North of England did not lead to a significant range in the errors associated with point interpolations and areal estimates.

9. Interpolation procedures which take account of the existence of serial correlation in daily rainfall data would not be expected to be significantly better than procedures which do not.

10. The gauge spacing required to meet a user requirement was found to be smallest in the case of daily rainfalls in excess of 10mm, reflecting the rapid rate of decay of correlation with distance in this case.

11. 'Black-box' models, which have as their data requirements rainfall and runoff data only, can achieve moderately good results in predicting and forecasting streamflow on British catchments with moderate to high runoff coefficients.

12. The manner in which the input to rainfall-runoff models is defined merits careful consideration; using all available gauges to define the model input may not be optimal.

8.3 Suggestions for further work

If further work on the evaluation of the UK raingauge network is to be undertaken, then this should be along the lines suggested in the conclusion of Section 8.2. This would allow the whole of the UK raingauge network to be assessed in a consistent way, taking into account local variations in user requirements and density. The necessary information would then be available to define how the network should be modified; operational problems would also be considered at this stage.

While the above suggestion is immediately practicable using the techniques described earlier, there are several points which need further investigation and which could radically affect the efficacy with which the project could be undertaken; some mention has already been made of these points as they have arisen in earlier sections. A more thorough statistical analysis of the rainfall data is needed covering the following points:

(a) The existence of local differences in long-term average rainfall and of differences in variation of rainfall and their modelling and treatment in interpolation situations.

(b) The construction of improved estimators for correlation appropriate for the statistical distribution of the data under consideration.

(c) The establishment of a better representation of the correlation structure of the data, possibly using a non-homogeneous correlation function and incorporating an allowance for measurement errors.

The foregoing relates only to network evaluation using direct methods. However, an overall evaluation of the network should also include the use of indirect methods, depending on what the local requirements for streamflow forecasting and prediction are, and on other possible indirect uses of rainfall. So far, network requirements have been explored for only a limited number of catchments with moderate to large coefficients of runoff. A much larger number of catchments typifying the various rainfall-runoff regimes in the UK should be selected, and experiments conducted along the lines of those in Section 6 to enable the interaction between spatial rainfall variability, model input definition and the nature of the rainfall-runoff regime to be more clearly understood. Modelling approaches of the type used in Section 6 offer most potential for such experiments because of their computational efficiency and because of their use of statistically efficient parameter estimation procedures. Multiple-input single-output rainfall-runoff models could be applied to catchments whose responses are governed by (a) significant soil moisture deficits, such as occur in the East and South of the UK in summer, and (b) groundwater discharge.

For example, the CLS model could be modified to incorporate daily measurements of soil moisture deficit through applying the threshold(s) to the catchment wetness index (CWI) defined in the Floods Studies Report (1975). For catchments dominated by groundwater discharge, spatially distributed water table level measurements could be used as further inputs to either of the models used in Section 6.2 and 6.3; the following specific questions could then be addressed:

(a) Will the storage characteristics of baseflow dominated catchments smooth out both spatial and temporal variations in rainfall, resulting in a less demanding raingauge network requirement?

(b) As model errors are likely to be higher for rainfall-runoff regimes dominated by baseflow and large soil moisture deficits, can much be gained from a more accurate specification of the spatial and temporal nature of the rainfall input?

(c) Can a network of groundwater level gauges reduce the necessity for a dense network of raingauges for rainfall-runoff modelling?

(d) Do the most stringent requirements for raingauge network density occur on catchments with shallow soils overlying impermeable bed-rock?

(e) How are lumped and spatial model inputs best defined so that errors in streamflow forecasting and/or prediction are minimised?

The foregoing suggestions relate to work which might be undertaken as a further phase of the evaluation of the UK raingauge network. The proposed analyses would mainly make use of conventional techniques, suitably adapted where necessary. However, there is a need to improve on some of the methodology currently available; specific problems which should be investigated are:

(a) The construction of non-linear estimators of rainfall at ungauged sites, after consideration of the statistical distribution of the data.

(b) Assessment of the potentialities and applicability of an eigenvalue (principal component) analysis of the data.

(c) The use of radar estimates of rainfall for real-time flow forecasting.

(d) The use of radar rainfall patterns to study instantaneous spatial variations of rainfall which may then assist in planning networks of autographic raingauges.

(e) The application of a Bayesian decision theory approach to rainfall network design.

(f) Investigation of more general methods of information transfer between gauges.

REFERENCES

Abadie, S. 1970 Integer and non-linear programming, North Holland, Amsterdam, pp 544.

Betts, A.K. 1973 Non-precipitating cumulus convection and its parameterization, Quart. J. Roy. Met. Soc., 99, pp 178-196.

Beven, K.J. 1977 Hillslope hydrographs by the finite element method, Earth Surf. Processes, 2, pp 13-28.

Bleasdale, A. 1965 Raingauge networks development and design with special reference to the United Kingdom, WMO/IASH Symposium, Design of hydrological networks, Quebec, Publ. 67, pp 146-154.

Box, G.E.P., Jenkins, G.M. 1970 Time series analysis forecasting and control, Holden-Day, San Francisco, pp xix + 553.

Brandes, E.A. 1975 Optimizing rainfall estimates with the aid of radar, J. Appl. Met., 14, pp 1339 - 1345.

Bras, R.F., Rodriguez-Iturbe, I. 1976a Evaluation of mean square error involved in approximating the areal average of a rainfall event by a discrete summation, Water Resour. Res., 12, pp 181 - 184.

Bras, R.F., Rodriguez-Iturbe, I. 1976b Rainfall network design for runoff prediction, Water Resour. Res., 12, pp 1197-1208.

Bussell, R.B. 1976 Rainfall v. radar : evaluation of benefit, Internal report of Central Water Planning Unit (private communication).

Carrigan, P.H., Golden, H.G. 1975 Optimizing information transfer in a stream gauging network, U.S. Geol. Surv. Water Resour. Investigations 30 - 75, Reston.

Cislerova, M., Hutchinson, P. 1974 The redesign of the raingauge network of Zambia, Hydrol. Sciences Bull., 19, pp 423-434.

Clarke, R.T., Leese, M.N., Newson, A.J. 1973 Analysis of data from Plynlimon raingauge networks, April 1971 - March 1973, Institute of Hydrology, Wallingford, Report No. 27, pp viii + 76.

Clemesha-Smith, C., Barnes, A.A., Bilham, E.G., Glasspoole, J., Mill, H.R., McClean, W.N. 1937 Report of the Joint Committee to consider methods of determining the general rainfall over any area, Trans. Inst. Water Engin., 42, pp 231-276.

Collier, C.G. 1975 Rainfall measurement by radar, Paper 1 and conference supplement, Weather Radar and Water Management, Water Research Centre and Royal Radar Establishment, Chester.

Delfiner, P., Delhomme, J.P. 1975 Optimum interpolation by Kriging, in Display and Analysis of Spatial Data, Proceedings of a NATO Advanced Study Institute, ed by J.C. Davis and M.J. McCullagh, Wiley and Sons, London, pp 96-114.

Delhomme, J.P. and Delfiner, P. 1974 Application of Kriging to the optimization of a raingauge network in an arid region, UNESCO/WMO/IASH Proceedings of symposium, Design of water resources projects with inadequate data, Madrid, 1, pp 171-190.

Dickinson, W.T., Douglas, J.R. 1972 A conceptual runoff model for the Cam catchment, Institute of Hydrology, Wallingford, Report No. 17, pp iv + 44.

Eagleson, P.S. 1967 Optimum density of rainfall networks, Water Resour. Res., 3, pp 1021 - 1033.

Freeze, R.A. 1972 Role of subsurface flow in generating surface runoff 1. Base flow contributions to channel flow, Water Resour. Res., 8, pp 601-623.

Gandin, L.S. 1963 The objective analysis of meteorological fields, Gidrometeorzdat, Leningrad, pp 282.

Gandin, L.S. 1970 The planning of meteorological station networks, W.M.O. Tech. Note No. 111, No 265, T.P. 149, Geneva, pp 35.

Granger, C.W.J., Newbold, P. 1976 Identification of the two-way causal systems, in Frontiers of Quantitative Economics, III, ed. by M.D. Intriligator, North-Holland, Amsterdam.

Grayman, W.M., Eagleson, P.S. 1971 Evaluation of radar and raingauge systems for flood forecasting, M.I.T. Dept of Civil Engineering, Report 138, pp 427.

Hardison, C.H. 1969 Accuracy of streamflow characteristics in Geological Survey research, U.S. Geol. Surv. Prof. Pap. 650-D, D210-214.

Harrold, T.W., English, E.J., Nicholass, C.A. 1973 The Dee weather radar project: the measurement of area precipitation using radar, Weather, 28, pp 332-338.

Harrold, T.W., English, E.J., Nicholass, C.A. 1974 The accuracy of radar-derived rainfall measurements in hilly terrain, Quart. J. Roy. Met. Soc., 100, pp 331-350.

Herbst, P.H., Shaw, E.M. 1969 Determining raingauge densities in England from limited data to give a required precision for monthly areal rainfall estimates J. Inst. Water Engin. 23, pp 218-230.

Hershfield, D.M. 1965 On the spacing of raingauges - WMO/IASH symposium: Design of hydrological networks, Quebec, Publication 67, pp 72-79.

Hill, F.F., Whyte, K.W., Browning, K.A. 1977 The contribution of a weather radar network to forecasting frontal precipitation: a case study, Met. Mag, 106, pp 69-88.

Huff, F.A. 1967 The adjustment of radar estimates of storm mean rainfall with raingauge data, J. Appl. Met., 6, pp 52-56.

Jacobi, S., Dawdy, D.R. 1973 The relation of rainfall network density to accuracy of runoff prediction in a mountainous basin, WMO Symposium, Distribution of precipitation in mountainous areas, Geilo, Publ. 326, 1, pp 214-218.

Jenkins, G.M. 1974 The interaction between the muskrat and mink eagles in North Canada, in The International Biometric Conference Proceedings, Constanta, Romania, pp 55-71.

Jenkins, G.M., Watts, D.G. 1968 Spectral analysis and its applications, Holden-Day, San Francisco, pp. xviii + 525.

Kendall, M.G., Stuart, A. 1968 The advanced theory of statistics, vol 3 Griffin, London, 2nd ed. pp x + 557.

Kendall, M.G., Stuart, A. 1969 The advanced theory of statistics, vol 1. Griffin, London, 3rd ed. pp xii + 439.

Kohler, M.A. 1958 Design of hydrological networks. WMO Tech. Note No. 25, pp 16.

Laugh, L.D., Box, G.E.P. 1977 Identification of dynamic regression (distributed lag) models connecting two time series, J. Amer. Stat. Assn, 72, pp 121-130.

Linsley, R.K., Crawford, N.H. 1966 Digital simulation in hydrology: Standford watershed model IV, Technical Report 39, Dept. Civ. Eng., Stanford Univ., pp 210.

Maddock, T. 1974 An optimum reduction of gauges to meet data program constraints, Bull. IASH, 19, pp 337-345.

Manley, R.E., Sexton, J.R. 1976 The extension of records using a deterministic hydrologic simulation model, Paper presented at IMA Conference on "The Mathematics of Hydrology and Water Resources" University of Lancaster.

Mason, B.J. 1971 The physics of clouds, Clarendon Press, Oxford, 2nd ed., pp xvi + 672.

Mason, B.J., Jonas, P.R. 1974 The evolution of droplet spectra and large droplets by condensation in cumulus clouds, Quart. J. Roy. Met. Soc., 100, pp 23-38.

Matheron, G. 1971 The theory of regionalised variables and its applications, Centre de Morphologie Mathematique de Fontainebleau, pp 211.

Matheron, G. 1975 A simple substitute for conditional expectation: the disjunctive Kriging, Proceedings of a NATO Advanced Study Institute, Geostatistics, Nice, pp 15.

McCullagh, M.J. 1975 Estimation by Kriging of the reliability of the proposed Trent telemetry network, Computer Applications, 2, pp 357-374.

Mejia, J.M., Rodriguez-Iturbe, I. 1973 Multidimensional characterization of the rainfall process. I. Synthetic generation of hydrologic spatial processes. II. On the transformation of point rainfall to areal rainfall, MIT Dept. of Civil Eng., Report No. 177, pp iv + 78.

Moore, R.J. 1977 A recursive time series analysis approach to hydrological modelling. Institute of Hydrology Report (to appear).

Moran, M.A. 1974 On estimators obtained from a sample augmented by multiple regression. Water Resour. Res., 10, pp 81-85.

Morrison, D.F. 1971 Expectations and variances of maximum likelihood estimates of the multivariate normal distribution parameters with missing data. J. Amer. Stat. Assn., 66, pp 602-604.

Moss, M.E. 1976 Design of surface-water data networks for regional information. Hydrol. Sciences Bull., 21, pp 113-128.

Moss, M.E., Karlinger, M.R. 1974 Surface water network design by regression analysis simulation. Water Resour. Res., 10, pp 427-433.

N.E.R.C. 1975 Flood Studies Report, 5 vols.

Natale, L., Todini, E. 1976a A stable estimator for linear models. 1. Theoretical development and Monte Carlo experiments. Water Resour. Res., 12, pp 667-671.

Natale, L., Todini, E. 1976b A stable estimator for linear models. 2. Real world hydrologic applications. Water Resour. Res., 12, pp 672-676.

Nicks, A.D. 1965 Field evaluation of raingauge network design principles. WMO/IASH symposium, Design of hydrological networks, Quebec, Publ. 67, pp 82-93.

O'Connell, P.E., Nash, J.E., Farrell, J.P. 1970 River flow forecasting through conceptual models. Part II - The Brosna catchment at Ferbane. J. Hydrol. 10, pp 317-329.

Ord, J.K., Rees, M. 1976 Spatial processes; a review of recent developments. Paper presented at IMA Conference on "The Mathematics of Hydrology and Water Resources." Univ. of Lancaster, pp 22.

Parzen, E. 1969 Multiple time series modelling, in Multivariate Analysis II, ed P.R. Krishnaiah, Academic Press, New York, pp 389-409.

Richards, D.M. 1975 The effects of reducing raingauge network density on goodness of conceptual model fit and prediction, Institute of Hydrology Report 28, Wallingford, pp 25.

Rodda, J.C. 1969 Hydrological network design - needs, problems and approaches, WMO Report No. 12, pp x + 58.

Rodda, J.C. 1971 The precipitation measurement paradox : the instrument accuracy problem, WMO Report No. 16, pp xii + 42.

Rodda, J.C., Downing, R.A., Law, F.M. 1976 Systematic Hydrology, Newnes-Butterworth, London, page 38.

Rodriguez-Iturbe, I., Mejia, J.M. 1973 The design of rainfall networks in time and space, MIT Dept. of Civil Eng., Report No. 176, pp iv + 78.

Schlatter, T.W. 1975 Some experiments with a multivariate statistical objective analysis scheme, Mon. Weather Rev., 103, pp 246-257.

Sharon, D. 1972 Spatial analysis of rainfall data from dense networks, Hydrol. Sciences Bull., 17, pp 291-300.

Shaw, E.M. 1965 A raingauge network prepared for the Devon River Board area of the United Kingdom, WMO/IASH symposium, Design of hydrological networks, Quebec, publ. 67, pp 63-71.

Shearman, R.J. 1977 The speed and direction of movement of storm rainfall patterns, Met. Office, Bracknell.

Smagorinsky, J. 1967 The role of numerical modelling, Bull. Amer. Met. Soc., 46, pp 388-391.

Stephenson, P.M. 1968 Objective assessment of adequate numbers of raingauges for estimating areal rainfall depths, IUGG/IASH General Assembly, Berne, Publ. 78, pp 252-264.

Sutcliffe, J.V. 1966 The assessment of random error in areal rainfall estimation, Bull. IASH, 11, pp 35-42.

Thiebaux, H.J. 1976 Anisotropic correlation functions for objective analysis, Mon. Weather Rev., 104, pp 994-1002.

Todini, E., Wallis, J.R. 1977 Using CLS for daily or longer period rainfall-runoff modelling in "Mathematical Models in Surface Water Hydrology" edited by T.A. Ciriani, U. Maione, and J.R. Wallis, J. Wiley and Sons, London, pp 149-168.

WMO 1975 Intercomparison of conceptual models used in operational hydrological forecasting, Oper. Hydrol. Report No. 7, WMO No. 429, Geneva.

Water Data Unit 1975 Water data 1975, HMSO, London, pp iv + 72.

Water Resources Board, Meteorological Office, Plessey Radar Limited (-) Dee Weather Radar Project : the use of a radar network for the measurement and quantitative forecasting of precipitation - Undated summary report by the Operations Systems Group, Water Resources Board, Reading, pp 7.

Whittle, P. 1954 On stationary processes in the plane, Biometrika, 41, pp 434-449.

Wilson, J.W. 1970 Integration of radar and raingauge data for improved rainfall measurement, J. Appl. Met., 9, pp 489-497.

Wilson, J.W. 1975 Radar-gauge precipitation measurements during the IFYGL, Centre for the Environment and Man Inc.

Young, P.C., Shellswell, S.H., Neethling, C.G. 1971 A recursive approach to time series analysis, CUED/B - Control/TR16. Dept. of Engin., Univ. of Cambridge, pp iv + 69.

APPENDIX A: CLASSIFIED BIBLIOGRAPHY ON RAINGAUGE NETWORK DESIGN

A.1 Spatial variations of the rainfall process

ATKINSON, B.W. and SMITHSON, P.A. 1974 Mesoscale circulations and rainfall patterns in an occluding depression. Quart. J.R. Met. Soc., 100, pp 3-22

BERGERON, T. 1960 Preliminary report of synoptic-quantitative rainfall studies with dense networks. Jordbrukets forsk Under Tio

DAWDY, D.R. and LANGBEIN, W.M. 1960 Mapping mean areal precipitation. IASH Bull., 19, pp 16-23

DYER, T.G.J. 1976 On the components of time series: the removal of spatial dependence. Quart. J.R. Met. Soc., 102, pp 157-165

ENGLISH, E.J. 1973 An objective method for calculating areal rainfall. Met. Mag., 102, pp 292-298

EPSTEIN, E.S. 1966 Point and area precipitation probabilities. Mon. Wea. Rev., 94, pp 595-598

HENDRICK, R.L. and COMER, G.H. 1970 Space variations of precipitation and implications for raingauge network design. J. Hydrol., 10, pp 151-163

HOUNAM, C.E. 1960 Temporal and spatial distribution of rainfall: a review. Comm. Bur. Met., 5, pp 1-20

HUFF, F.A. 1956 Relation between point and areal rainfall frequencies. Amer. Met. Soc. Bull., 37, 243, pp 618-20, 1958, p 363

HUFF, F.A. 1970 Spatial distribution of rainfall rates. Water Resour. Res., 6, pp 254-260

HUFF, F.A. 1970 Sampling errors in measurement of mean precipitation. J. Appl. Met., 9, pp 35-44

HUFF, F.A. and NEILL, J.C. 1956 Frequency of point and areal mean rainfall rates. Trans. Amer. Geophys. Un., 37, pp 679-681

HUFF, F.A. and NEILL, J.C. 1957 Areal representativeness of point rainfall. Trans. Amer. Geophys. Un., 38, pp 344-345

HUFF, F.A. and SHIPP, W.L. 1969 Spatial correlations of storm, monthly, and seasonal precipitation. J. Appl. Met., 8, pp 542-550

KOTZ, S. and NEUMANN, J. 1963 On the distributions of precipitation amounts for periods of increasing length. J. Geophys. Res., 68, pp 3635-3640

LENTON, R.L. and RODRIGUEZ-ITURBE, I. 1974 On the collection, the analysis, and the synthesis of spatial rainfall data. MIT Dept. Civ. Eng. Report 194

MCCULLOCH, J.S.G. 1961 Statistical assessment of rainfall. Nairobi: Inter-African Conf. Hydr.

NEWTON-SMITH, W.H. and WILLIAMS, G.D.V. 1964 The relationship between raingauge network density and areal rainfall estimates on the prairies. Torronto, Met. Branch Circ., 4137, pp i + 17

RIGG, J.B. 1960 A statistical study of the distribution of rainfall over small and medium sized areas. Weather, 15, pp 377-379

SAKUMA, K., ARAI, T. and ARAKI, K. 1951 A note on the distribution of rainfall amounts. Tokyo: J. Met. Res., 3, pp 455-457

SCHICKEDANZ, P.T. 1973 A statistical approach to rainfall patterns. Boston. Amer. Met. Soc. Conf. on Prob. Stat. Atmos. Sci., pp 104-9

SEMPLAK, R.A. and KELLER, H.A. 1969 A dense network for rapid measurement of rainfall rate. Bell Syst. Tech. J., 48, pp 1745-1756

SHARON, D. 1972 Spatial analysis of rainfall data from dense networks. IASH Bull., 17, pp 291-300

SMITH, R.E. 1974 Point processes of seasonal thunderstorm rainfall 3. Relation of point rainfall to storm areal properties. Water Resour. Res., 10, pp 424-426

A.2 <u>Raingauge network accuracy considerations</u>

AKIN, J.E. 1971 Calculation of mean areal depth of precipitation. J. Hydrol., 12, pp 363-376

ANDERSSON, T. 1963 On the accuracy of rain measurements and statistical results from rain studies with dense networks (Project Pluvius). Uppsala K. Univ. Met. Inst. Meddelande 68, pp 307-332

ANDERSSON, T. 1968 On the representativity of rain measurements. Uppsala K. Univ. Met. Inst., Report 7

BETHLAMY, N. 1976 The two-axis method: a new method to calculate average precipitation over a basin. IASH Bull., 21, pp 379-385

CLARKE, R.T. and EDWARDS, K.A. 1972 The application of the analysis of variance to mean areal rainfall estimation. J. Hydrol., 15, pp 97-112

COLLIER, C.G., HARROLD, T.W. and NICHOLASS, C.A. 1975 A comparison of areal rainfall as measured by a raingauge-calibrated radar system and raingauge networks of various densities. Proc. Amer. Met. Soc., 16th Met. Radar Conf., pp 467-472

DAHLSTROM, B. 1973 Investigations of errors in rainfall observations. Uppsala K. Univ. Met. Inst., Report 34

EDWARDS, K.A. 1972 Estimating areal rainfall by fitting surfaces to irregularly spaced data. WMO No. 326, 2, pp 565-587

GLASSPOOLE, J. 1962 Assessment of areal rainfall amounts. Weather, 17, pp 312-313

HARROLD, T.W., ENGLISH, E.J. and NICHOLASS, C.A. 1973 The Dee Weather Radar Project: the measurement of area precipitation using radar. Weather, 28, pp 332-338

HARROLD, T.W., ENGLISH, E.J. and NICHOLASS, C.A. 1974 The accuracy of radar-derived rainfall measurements in hilly terrain. Quart. J.R. Met. Soc., 100, pp 331-350

HUTCHINSON, P. 1969 Estimation of rainfall in sparsely gauged areas. IASH Bull., 14, pp 101-119

HUTCHINSON, P. 1971 Accuracy of estimates of daily areal mean rainfall. Weather, 26, pp 366-371

KELWAY, P.S. 1974 A scheme for assessing the reliability of interpolated rainfall estimates. J. Hydrol., 21, pp 247-267

LEE, P.S., LYNN, P.P. and SHAW, E.M. 1974 Comparison of multiquadric surfaces, for the estimation of areal rainfall. IASH Bull., 19, pp 303-317

MCDONALD, J.E. 1957 A note on the provision of estimation of missing precipitation data. Trans. Amer. Geophys. Un., 38, pp 657-661

MOSS, J.M. 1960 Exposure factor affecting the accuracy of precipitation measurements. Melbourne, Comm. Bur. Met. Seminar Rain, 4, pp 1-20

PAULHUS, J.L.H. and KOHLER, M.A. 1952 Interpolation of missing precipitation records. Mon. Wea. Rev., 80, pp 129-133

RAINBIRD, A.F. 1967 Methods of estimating areal average precipitation. WMO/IHD Report 3

RAMAMURTI, K.S. 1948 On estimating the average depth of rainfall over an area and the distribution of raingauges. Bangalore: Curr. Sci., 17, pp 317-318

RAMAN, P.K. and VIRJI, H. 1971 On estimation of areal average precipitation by a weighted mean method. WMO No. 301, 2, pp 280-287

RODDA, J.C. 1962 An objective method for the assessment of areal rainfall records. Weather, 17, pp 54-59

RODDA, J.C. 1971 The precipitation measurement problem: the instrument accuracy paradox. WMO Report 16, pp xii + 42

RUSIN, N.P. 1973 Estimation of areal precipitation. WMO No. 326, 1, pp 115-128

SALTER, P.M. 1973 Areal rainfall analysis by computer. WMO No. 326, 2, pp 497-509

SAMUELSSON, B. 1973 Filling in gaps in rainfall records by simulated data. Nordic Hydrol., 4, pp 17-27

SCHLATTER, T.W. 1975 Some experiments with a multivariate statistical objective analysis scheme. Mon. Wea. Rev., 103, pp 246-257

SEAMAN, R.S. 1973 Analysis accuracy achievable with specified observational networks. Austr. Met. Mag., 21, pp 93-111

SEVRUK, B. 1974 Comparison of mean rain catch of various gauge networks. Nordic Hydrol., 5, pp 56-63

SHEARMAN, R.J. and SALTER, P.M. 1975 An objective rainfall interpolation and mapping technique. IASH Bull., 20, pp 353-363

SINGH, V.P. and BIRSOY, Y.K. 1975 Comparison of the methods of estimating mean areal rainfall. Nordic Hydrol., 5, pp 222-241

STIDD, C.K. 1953 Cube-root-normal precipitation distributions. Trans. Amer. Geophys. Un., 34, pp 31-5

UNWIN, D.J. 1969 The areal extension of rainfall records: an alternative model. J. Hydrol., 7, pp 404-14

VAHL, H. 1972 Computerised calculation of areal precipitation and its accuracy. WMO No. 326, 2, pp 510-524

ZAWADZKI, I.I. 1973 Errors and fluctuations of raingauge estimates of areal rainfall. J. Hydrol., 18, pp 243-255

A.3 <u>Principles of network design</u>

ALVAREZ, F. and HENRY, W.K. 1970 Raingauge spacing and reported rainfall. IASH Bull., 15, pp 97-107

AMOROCHO, J., BRANDSTETTER, A. and MORGAN, D. 1967 The effects of density of recording raingauge networks on the description of precipitation amounts. IASH Gen. Assy., 25, pp 189-202

BLEASDALE, A. 1965 Raingauge networks development and design with special reference to the United Kingdom. WMO/IASH Symposium, 67, pp 146-154

CLARKE, R.T., LEESE, M.N. and NEWSON, A.J. 1973 Analysis of data from Plynlimon raingauge networks, April 1971-March 1973. Wallingford, IH Report 27, pp viii + 76

CLEMESHA-SMITH, C., BARNES, A.A., BILHAM, E.G., GLASSPOOLE, J., MILL, H.R. and MCCLEAN, W.N. 1937 Report of the Joint Committee to consider methods of determining the general rainfall over any area. Trans. Instn. Wat. Engrs., 42, pp 231-276

FERGUSON, H.L. 1973 Precipitation network design for large mountainous areas. WMO No. 326, 1, pp 85-105

HATCH, D.J. 1976 The estimation of local rainfall by sample gauging. Weather, 31, pp 79-83

KOHLER, M.A. 1958 Design of hydrological networks. WMO Tech. Note 25, pp 16

NICKS, A.D. 1965 Field evaluation of raingauge network design principles. WMO/IASH Symposium, 67, pp 82-93

NOBLE, J.R.H. 1970 The degree of uniformity required in network operations. Boston, Amer. Met. Soc. Met. Mon. 11, No. 33, pp 15-19

RODDA, J.C. 1969 Hydrological network design: needs, problems, and approaches. Geneva: WMO/IHD Report 12, pp ix + 57

SHAW, E.M. 1965 A raingauge network prepared for the Devon River Board in the United Kingdom. Quebec: WMO/IASH Symposium, 67, pp 63-71

THOMPSON, P.D. 1963 An essay on the technical and economic aspects of the network problem. WMO Bull. 12, pp 2-6

WILKINSON, J.H. 1955 Study of rainfall station locations in South Chickamanga Creek watershed. Trans. Amer. Geophys. Mon., 36, pp 1021-1028

A.4 Direct methods of network design: 1: Local

BRADY, P.J. 1976 Multivariate experimental design in meteorology. Unpublished Ph.D thesis, Univ. Oklahoma, pp xviii + 121

CISLEROVA, M. and HUTCHINSON, P. 1974 The redesign of the raingauge network of Zambia. IASH Bull., 19, pp 423-434

DELFINER, P. and DELHOMME, J.P. 1975 Optimum interpolation by Kriging. Wiley: London: Display and analysis of spatial data, pp 96-114

DELHOMME, J.P. and DELFINER, P. 1974 Application of Kriging to the optimisation of a raingauge network in an arid region. UNESCO/WMO/IASH Symposium, Madrid, 1, pp 171-190

GANDIN, L.S. 1963 The objective analysis of meteorological fields. Lenningrad: Gidrometeorzdat: pp 282

GANDIN, L.S. 1970 The planning of meteorological station networks. WMO Tech. Note 111, No. 265, T.P. 149, pp 35

GANDIN, L.S. and KAGAN, R.L. 1967 On an economical approach in the planning of a network of meteorological stations. Soviet Hydrol., pp 597-606

MARSHALL, R.J. 1975 A stochastic model to simulate moving storms. Amer. Geophys. Un. Symposium: Precipitation Analysis for Hydrologic Modelling, pp 289-300

MATHERON, G. 1971 The theory of regionalised variables and its applications. Fontainebleau: Centre de Morphologie Mathematique, pp 211

MATHERON, G. 1975 A simple substitute for conditional expectation: the disjunctive Kriging. Nice: NATO Symposium, pp 15

MCCULLAGH, M.J. 1975 Estimation by Kriging of the reliability of the proposed Trent telemetry network. Computer Applications, 2, pp 357-374

MEJIA, J.M. and RODRIGUEZ-ITURBE, I. 1973 Multidimensional characterisation of the rainfall process (1) Synthetic generation of hydrologic spatial processes (2) On the transformation of point rainfall to areal rainfall. MIT Dept. Civ. Eng. Report 177, pp iv + 78

RODRIGUEZ-ITURBE, I. and MEJIA, J.M. 1973 The design of rainfall networks in time and space. MIT Dept. Civ. Eng. Report 176, pp iv + 78

SHARON, D. 1974 On the modelling of correlation functions for rainfall studies. J. Hydrol., 22, pp 219-224

THIEBAUX, H.J. 1976 Anisotropic correlation function for objective analysis. Mon. Wea. Rev., 104, pp 994-1002

A.5 Direct methods of network design: 2: Global

ALAKA, M.A. 1970 Theoretical and practical considerations for network design. Boston, Mass., Amer. Met. Soc. Met. Mon. No. 33, pp 20-27

EAGLESON, P.S. 1967 Optimum density of rainfall networks. Water Resour. Res., 3, pp 1021-1033

HERBST, P.H. and SHAW, E.M. 1969 Determining raingauge densities in England from limited data to give a required precision for monthly areal rainfall estimates. J. Inst. Water Engrs., 23, pp 218-230

HERSHFIELD, D.M. 1965 On the spacing of raingauges. WMO/IASH Symposium: Design of Hydrological Networks Publ., 67, pp 72-79

HOWARD, C.D.D. and BRADLEY, B. 1975 Data needs for modelling flood frequency. A.G.U. Symposium: Precipitation Analysis for Hydrologic Modelling, pp 279-288

STEPHENSON, P.M. 1968 Objective assessment of adequate numbers of raingauges for estimating areal rainfall depths. IUGG/IASH General *Assembly Berne*, 78, pp 252-264

SUTCLIFFE, J.V. 1966 The assessment of random error in areal rainfall estimation. IASH Bull., 11, pp 35-42

A.6 Indirect methods of network design

BRAS, R.F. and RODRIGUEZ-ITURBE, I. 1976 Rainfall network design for runoff prediction. Water Resour. Res., 13, pp 1197-1208

DAVIS, D.R., KISIEL, C.C. and DUCKSTEIN, L. 1972 Optimum design of mountainous raingauge networks using Bayesian decision theory. WMO No. 326, 2, pp 416-420

GRAYMAN, W.M. and EAGLESON, P.S. 1971 Evaluation of radar and raingauge systems for flood forecasting. MIT Dept. Civ. Eng. Report 138, pp 427

JACOBI, S. and DAWDY, D.R. 1973 The relation of rainfall network density to accuracy of runoff prediction in a mountainous basin. WMO No. 326, 1, pp 214-218

RICHARDS, D. 1975 The effects of reducing network density on goodness of conceptual model fit and prediction. Wallingford, IH Report 28, pp 25

APPENDIX B: INVENTORY OF RAINFALL DATA USERS AND THEIR REQUIREMENTS

Appendix B.1 Inventory of rainfall data users

The following schedule lists the organisations approached in connection with the review of data requirements in Section 3. Enquiries within any organisation were not exhaustive although in every case as much information as possible was obtained through representative contacts. They are entered here under their primary field of activity even though discussions may have focussed on some specialist aspect of their expertise. Organisations visited by the Institute of Hydrology are marked (H) and those who were circulated by the Meteorological Office are marked (M).

Water industry

The ten Regional Water Authorities (H), Central Water Planning Unit (H), Water Data Unit (H), Scottish Development Department (H), Northern Ireland Departments of the Environment and Finance (H), Strathclyde Regional Council (H), Clyde River Purification Board (H), Greater London Council (H), Departments of Environment and Transport (M), Welsh Office (M), Institute of Hydrology (H,M), Water Research Centre (H), Central Lancashire Development Corporation (M).

Civil engineering, consultancy, building and construction

Construction Industry Research and Information Association (H), Building Research Establishment (H), Sir William Halcrow and Partners (M), R. Cuthbertson and Partners (H), Binnie and Partners (H), George Wimpey (H).

Agriculture

Soil Survey of England and Wales (H), Agricultural Development Advisory Service, - Headquarters (H,M) - Cambridge (H) and Bristol (M) Regional Offices, Letcombe Laboratory (H), Plant Pathology Laboratory (H,M), Alice Holt Research Station (Forestry Commission) (H), National Institute of Agricultural Botany (H,M), Grassland Research Institute (M), Central Veterinary Laboratory (M), National Institute of Agricultural Engineering (M), West of Scotland Agricultural College (M), Field Drainage Experimental Unit (H,M), Wright Rain (M), University of Nottingham (Sutton Bonington) (M), Hannah Research Institute (M), Hill Farming Research Organisation (M), North of Scotland College of Agriculture (M), Jealott's Hill Research Station (ICI) (M), West of Scotland Agricultural College (M), Rothamsted Research Station (M), Chief Agronomist MAFF (H), National Farmers' Union (M).

Meteorology

Meteorological Office branches concerned with agrometeorology, instrumentation and observational practices, forecasting research (M); Enquiry branches at Bracknell (M) and Belfast (H,M); Meteorological Office Weather Centres at London, Newcastle, Southampton, Manchester and Glasgow (M); branch concerned with soil moisture deficit calculations (H,M). Climatic Research Unit of University of East Anglia (H).

Public Utilities

North of Scotland Hydro-Electric Board (H), Appleton Laboratory (representing Post Office telecommunications) (H), Central Electricity Generating Board (H).

Law, Insurance, pollution and health

Eagle Star Insurance (H), Thames Valley Police (H), Warren Springs Laboratory (H), Atomic Energy Research Establishment (H), Institute of Terrestrial Ecology (H).

Appendix B.2 Inventory of user requirements

USE	DESCRIPTION OF PROCEDURE WATER INDUSTRY	ACCURACY AND SPECIAL REQUIREMENTS
a. General background for management decisions:		
i. time series data for operational purposes	Weekly summary of rainfall depths at key stations with comparisons with long term averages at the same location	Long term stations, usually at major population centres. 6 per Water Authority area probably sufficient
ii. statistical data for planning purposes	Annual average rainfall map in order to make ready comparison of rainfall regime between various sites of interest	Annual values are required with accuracy at a point of 25 mm, 2 s.e. to produce adequate maps of average rainfall
b. Water balances for such purposes as providing the input to aquifer recharge studies, yield studies, and gauging station checks	Monthly catchment average rainfall is used in conjunction with the other elements of balance. Area of integration is normally over 50 km^2 but occasionally can be less	1 month, 50 km^2, 10%, 2.33 s.e. is a typical accuracy. Occasionally a more stringent accuracy is necessary but basic network could be augmented for the duration of the study. A skeleton network of ground level gauges for assessment of 'true' rainfall may also be required
c. Rainfall-runoff models for water resource studies	Monthly catchment average rainfall is used as input to both stochastic and hydrological models	1 month, 50 km^2, 10%, 2.33 s.e. is a typical accuracy. Long term stations should be retained in order to provide input for data extension exercises
d. Rainfall-runoff models for flood analysis:		
i. Analysis of past flood events to determine cause and frequency of occurrence on natural catchments	Catchment average rainfall resolved to 1 hour or less. At least one recording raingauge is used to distribute daily catchment average rainfall. This is thought to provide sufficiently accurate input data for current spatially lumped models used in engineering	1 day, 10 km^2, 20%, 2 s.e. is a typical accuracy for the daily raingauges. No comparable figures can be quoted for the recording raingauges; storm cell size could provide a basis for gauge spacing. An additional practical requirement is for the early availability of the rainfall data, e.g. within one week of the flood event

USE	DESCRIPTION OF PROCEDURE	ACCURACY AND SPECIAL REQUIREMENTS
	WATER INDUSTRY (Cont'd)	
	hydrology. Catchment average daily rainfall is also used to index the antecedent wetness prior to the flood event	
ii. As d(i) but for urban catchments	As d(i) except that the rainfall data should be resolved to 5 minutes	1 day, 1 km^2, 20%, 2 s.e. is a typical accuracy, otherwise use (d(i))
e. Rainfall-runoff models for flood design	Design techniques are developed from analysis of past flood events (d). Statistics of extreme rainfall over specified durations are also required. Miscellaneous ways of describing the rainfall process such as areal reduction factor, spatial variability, intensity time profile, and evaluation effect are also necessary	Long term stations are needed for statistical purposes. The network requirement may be equated with that for annual average rainfall evaluation. Clusters of stations are useful for the miscellaneous descriptors; 1 per 30 km^2 was found adequate for a recent study of areal reduction factors
f. Rainfall-runoff models for forecasting future runoff from current rainfall	Similar to (d(i)) for natural catchments and (d(ii)) for urban catchments	As (d(i)) and (d(ii)) except that telemetering gauges are required to produce the data in real time
g. Testing leaks into and out of foul sewers	Daily rainfall is compared with inflows to the sewage works. A visual inspection of the two quantities enables the presence of leaks to be detected	Daily raingauges in remote parts of sewerage system as well as at sewage works.
	CIVIL ENGINEERING, CONSULTANCY, BUILDING AND CONSTRUCTION	
a. Time lost on construction sites due to	Average time lost has been related to rainfall statistics; in	Network requirement is for long recording raingauge records and mean annual rainfall for interpolation

USE	DESCRIPTION OF PROCEDURE	ACCURACY AND SPECIAL REQUIREMENTS
	AGRICULTURE (Cont'd)	
ii. Day to day advice to farmers wishing to irrigate	Advice by the Ministry of Agriculture is based upon a limited network of agro-meteorological stations. Ground conditions and temperature are important factors as well as knowledge of the rain depth over the previous two days	The stated requirement is 2 days, point, 5%, 2.33 s.e. More 'advanced' farmers would provide their own raingauge. It is sometimes necessary to discriminate between rain falling at night and by day
iii. Planning total irrigation need	Irrigation planning for such purposes as the purchase of spray equipment and farm dam design is based on soil moisture deficit statistics	Network requirement as for soil moisture deficit calculation (Meteorology)
b. Crop planting needs, including advice on and testing of seed suitability, and applicability of direct drilling technique	Advice is based upon the same network as (a(ii)) and has similar accuracy requirements. Testing of seed suitability is carried out at a very large number of farms for which rain data have to be interpolated	The stated requirement for seed testing is 2 weeks, point, 5 mm, 2.33 s.e.
c. Crop health and issuing warning of disease prone conditions such as Barley mildew and Wheat Septoria. Advice on timing of spraying and respraying if rain removes initial application	Bulletins warning of the onset of disease conditions are issued on a regional basis and use the same network as (a(ii)). The forecasts make use of formulae involving rainfall and other climatic variables. Typically the rainfall input to the formulae appears in the form of a threshold quantity over a period of days. The regional network is not regarded as adequate to provide warnings at any particular site	The accuracy requirement varies with the nature of the disease; some develop in rainy conditions and others in dry conditions. While the major requirement is in the arable areas, bracken control spraying in the Scottish highlands also calls on the raingauge network

USE	DESCRIPTION OF PROCEDURE	ACCURACY AND SPECIAL REQUIREMENTS
	CIVIL ENGINEERING, CONSULTANCY, BUILDING AND CONSTRUCTION	
adverse weather conditions	particular, percentage of time rainfall intensity exceeds 0.1 mm/hr. Effect of rain is of greater importance than the rainfall itself; however large construction sites do maintain a raingauge for diary purposes and for negotiating wage rates. Rain falling on construction sites is a common source of enquiry to the Meteorological Office	
b. Setting building standards, and design criteria for roofs, downpipes, soakaways and wall cladding materials	Statistics of rainfall intensity (and also snow loading) are used in structural design. An index of driving rain compounded of depth, wind velocity and direction is used by architects to determine suitable wall cladding and window sealant materials	Similar to (a) for rainfall statistics. A sparse network of gauges measuring driving rain would be useful to provide updated information to calibrate the index
	AGRICULTURE	
a. Irrigation:		
i. Experimental work to establish crop water use and to improve application practices	Most experimentation is carried out at research establishments or experimental farms which are equipped with the necessary recording equipment. A common requirement is for point data integrated over 12 hour periods especially in the growing season	This requirement seems to be self supporting in the main

USE	DESCRIPTION OF PROCEDURE	ACCURACY AND SPECIAL REQUIREMENTS
	AGRICULTURE (Cont'd)	
d. Advice on dates of harvesting and statistical work on past years yield. Statistical studies of yield reduction due to disease. Development of guidelines for maximising yield (blue-prints).	Harvesting advice is based upon the same network as (a(ii)). The statistical work is carried out on a national and regional basis and uses a sparse network only. 'Blueprint' development is at present at a research stage and is carried out at experimental farms. When these yield maximising practises become more widespread it will be necessary to have site rainfall data available	
e. Design of field drainage work, i.e. pipe size and spacing, drain works and suitability of alternative practices	Statistics of high rainfall are used such as 10 year return period one and five day falls, and total rainfall after the date at which field capacity in the soil is reached	100 stations in England and Wales have been used to develop existing rules
f. Surveying and classifying soil types to produce maps of soil types and soil conditions	The map of annual average rainfall and its seasonal distribution. This is combined with Soil Moisture Deficit data to produce maps related to ease of cultivation and risk of damage through grazing (poaching)	The requirement for rainfall data is similar to that for Water Industry (a(ii))
g. Forecasting the onset of animal disease condition	A liver fluke forecast is issued and makes use of monthly rainfall totals. Future forecasts still under development for other diseases envisage the use of 12 hour totals at times of peak infection risk	Liver fluke forecast is of importance in highland regions as well as in the main agricultural areas. Like crop disease forecasts reporting raingauges are essential for an operational system

USE	DESCRIPTION OF PROCEDURE	ACCURACY AND SPECIAL REQUIREMENTS
	AGRICULTURE (Cont'd)	
h. Design of animal housing and silage systems	To meet Water Authority requirements it is becoming necessary to provide storage for animal waste based on similar principles to storm storage above sewage works	Design recommendations are under preparation for both storage systems and for animal housing standards
i. Forestry:		
i. Work conditions	'Down time' and work conditions are judged using autographic raingauge data. Wage rates are negotiated	A gauge at the work site is needed. Requirement is similar to Civil Engineering (a)
ii. Assessing fire risk	A formula is used which includes rainfall data	A raingauge is usually provided close to the forest
iii. Fire risk statistics and of days on which harvesting is possible	Useful planning information is obtained from past rainfall data near forest sites	Similar to (i(ii)) using retrospective rainfall data. Daily data are adequate
iv. Yield studies including species suitability and requirement for forest and nursery irrigation	The selection of suitable sites for growing different tree species uses monthly rainfall totals as well as other climatic and soil parameters. The occurrence of dry springs is especially important	Network stations are used but the density of gauges in upland areas is thought to be inadequate. Information on elevation dependency of rainfall would be welcomed
v. Wildlife studies	Wildlife studies require monthly rainfall amounts within particular valleys	As (i(iv))

USE	DESCRIPTION OF PROCEDURE	ACCURACY AND SPECIAL REQUIREMENTS
	AGRICULTURE (Cont'd)	
j. Uptake of nitrate fertilizer and naturally occurring nitrogen by crops and leaching of nitrate to groundwater	This problem is currently being studied at 60 experimental farms across UK. Leaching is related to winter rainfall following field capacity date	Experimental work is self sufficient but mass balances will require data similar to that for Water Industry (b)
	METEOROLOGY	
a. The validation of weather forecasts	Forecasts are scored for accuracy and there are plans for putting these on a firmer footing using a systematic procedure based upon a mesoscale network	Reporting raingauges read twice daily and at a 50 km grid with accuracy .5 mm, 1 s.e. when rainfall forecast exceeds 10 mm has been stated as the present requirement
b. The calibration of short and long-term forecasting models	Physically based process models are commonly used for short-term forecasts while statistical ones are used for long term forecasts	The stated requirement is $\frac{1}{2}$ month, 10^4 km^2, 5%, 2 s.e. Long-term records are an additional requirement to provide the statistical data for long-term forecasts
c. General enquiries from members of the public	A large variety of enquiries are directed at the Meteorological Office who maintain a staffed enquiry bureau. The information given can be quantitative in terms of event rainfall, or statistical, or qualitative in terms of whether rain was likely to have occurred at a particular place and time	Satisfactory responses to most enquiries could be given with 1 day, point, 5 mm, 2.33 s.e. accuracy. Statistical enquiries have a similar requirement to Water Industry (e)

USE	DESCRIPTION OF PROCEDURE	ACCURACY AND SPECIAL REQUIREMENTS
	METEOROLOGY (Cont'd)	
d. The production of a fortnightly map of soil moisture deficit throughout the growing period	Soil moisture deficit is calculated as a daily water balance between rainfall and transpiration. The current method will be updated using the MORECS system which divides the country into 40 km squares. Local refinements even to this system are envisaged to provide a finer space and time resolution	Current mark of SMD maps use 180 raingauges reporting within a few days of recording. The MORECS system requires 1 day, 40 km x 40 km, 5%, 2 s.e. accuracy. A rainfall breakdown to 12 hour resolution will be needed for some areas. A future requirement for 20 km squares has been in the network assessment
e. Calibration of radar rainfall measurements	The network replacement possibilities of radar are discussed in Section 3.4. The present mark used primarily for flood warning requires on line reporting raingauges to update hourly the reflectivity/rainfall conversion factor	For present purposes 15 raingauges per radar installation are thought necessary. This might increase for other uses to which the data might be put
f. Search for climatic trends, cycles and other repeated patterns	As much historic data as available is assembled and subjected to time series and other analyses	Long and continuous records are essential for this research
g. Instrumentation: measurement standards and observational practices are set, the recorded data is subjected to quality control prior to archiving and publication	Quality control of daily rainfall data consists of checks on internal consistency and comparisons with nearest neighbour gauges	While this Meteorological Office function does not itself call upon the network it is essential for the smooth running and servicing of the network

USE	DESCRIPTION OF PROCEDURE	ACCURACY AND SPECIAL REQUIREMENTS
	PUBLIC UTILITIES	
a. Electricy supply:		
i. *Forecasting demand*	A small network of reporting rain-gauges is used in a demand forecast formula	
ii. Hydro-power efficiency monitoring	Monthly catchment average rainfall on Scottish hydro-power catchments is compared with power output to assist with performance monitoring	As Water Industry (b). Elevation dependancy information is particularly important
iii. Operating rule design for hydro-power installation	Analysis of past events and flood warning techniques are used to design and implement operating rules	As Water Industry (d(i)) and (e)
iv. Precipitation of pollutants from power station smoke plumes	Pollution due to 'rain-out' of pollution is of national and international interest. Chemical quality and rain depth data are used in mass balances	A small purpose-augmented network of stations has been employed well downwind of power production areas
b. Telecommunications:		
i. Production of rainfall intensity statistics	High frequency waves are sensitive to rainfalls in excess of 10 mm/hour. Two experimental networks are used to derive duration curves for rainfall intensity	Experiments use specially designed equipment
ii. Design of communication links both between towers and from ground to satellite	Results from (b(i)) are generalised to line statistics and used in other parts of the country by reference to the spatial pattern of rainfall statistics in the Flood Studies Report (N.E.R.C., 1975)	Similar to Water Industry (e)

USE	DESCRIPTION OF PROCEDURE	ACCURACY AND SPECIAL REQUIREMENTS
	LAW, INSURANCE AND HEALTH	
a. Police work:		
i. Forensic and detective work	The use normally revolves on the presence or absence of rain at a particular location and time, e.g. in assessing blame in a road accident	Use is made of the Meteorological Office's enquiry service, Meteorology (e)
ii. Emergency relief	Flood relief, traffic diversion and direct effects of rain are normally dealt with in conjunction with Water Authority and Meteorological Office's forecasting arrangements	
b. Insurance against rain:		
i. Calculation of premiums	Because most rainfall insurances (Pluvius policies) are to aid open-air event organisers, the statistics which are most use are for summer weekend afternoons. The United Kingdom business is largely concentrated in the south-east	As Water Industry (e)
ii. Rainfall observers during the insured event	The insurance takes the form of a sliding scale depending on the depth of rain recorded within the insured time interval at an agreed site	Stated preference for an experienced observer using their standard raingauge within 2 miles of the event. If one is unavailable then ad hoc arrangements are made

USE	DESCRIPTION OF PROCEDURE	ACCURACY AND SPECIAL REQUIREMENTS
	LAW, INSURANCE AND HEALTH (Cont'd)	
c. Health:		
i. Epidemiological studies	A weak secondary correlation has been observed between rainfall and the prevelance of certain diseases	Few raingauges only are needed for this usage
ii. Monitoring air pollution	Rainfall is a major agency in depositing air pollution but so far only special purpose instruments have been installed. Important gaps in knowledge remain such as variation with intensity, coastal and geographic effects on deposition, effect of natural low pH and self cleansing abilities of rainfall	Main current requirement is for improved techniques for chemical and biological monitoring of 'rained-out' pollutants. Use of quantitative network is to produce mass balances, and, except for some experimental sites, is recorded as adequate presently

APPENDIX C: DESCRIPTION OF CATCHMENTS USED FOR DAILY RAINFALL-RUNOFF MODELLING

The six catchments used for daily rainfall-runoff modelling all lie in the wet western areas of the country and have many similarities. They all have steep upland headwater tributaries with high annual average rainfalls and gently sloping lower sections. Thus the values of main channel slope are all similar and are typical of moderately steep catchments with areas of over 100 km^2. In each of the six catchments the isohyets follow the topographic contours fairly closely. Thus those catchments with the greatest topographic variation such as the Exe and North Tyne also have the greatest variation in rainfall. This is true for both annual average rainfall (AAR) and a shorter duration rainfall index developed for the Flood Studies Report, a 2 day, 5 year return period rainfall (M52D). A brief description of each of the six study catchments follows. Various catchment characteristics are given in Table 6.1.

The North Tyne at Tarset is a natural rated section and drains south-eastwards from the Cheviot Hills over an area of mainly Carboniferous Limestone with poorly drained peaty soils. Slopes are moderately steep with a channel slope of 4.94 m/km. The annual average rainfall (AAR) is 1255 mm and varies from greater than 1800 mm to 900 mm over the catchment. It is probably useful to consider the shorter duration rainfall index also, the 2 day 5 year return period rainfall (M52D); for the North Tyne catchment this is 69.7 mm and varies with topography as does annual average rainfall.

The Ure at Kilgram Bridge is a natural rated section controlled by the bridge and drains from Carboniferous Limestone onto Millstone Grit. Soils are peaty on the hill summits with poorly drained gleyed soils in the valleys. Headwater streams are steep but the main valley is relatively flat with a slope of 4.1 m/km. The annual average rainfall is 1390 mm and varies from over 2000 mm on the hills to 750 mm at Kilgram Bridge. M52D is 84.1 and the variation in both AAR and M52D follows the topography but is rather less variable than on the North Tyne.

The Exe at Stoodleigh is also a rated section and drains southwards from the old Devonian sandstones of Exmoor onto Carboniferous sandstones and grits. Soils are mainly well drained brown podzolic soils but there are some small areas of peat on Exmoor. Headwater streams are steep but the valley as a whole is again only moderately steep although it has a rather higher channel slope than the catchments in the northern part of the country. The AAR is 1402 mm, varying from over 2000 mm to 1100 mm and M52D is 84.3 mm. The variation is similar to the Ure at Kilgram Bridge and follows the topography.

The Culm at Woodmill is a natural rated section and drains from an area of mixed Permian and Triassic sandstones and marls. The soils are moderately well drained clays on the Keuper Marl and well drained brown earths on the sandstones. Slopes are moderate and, although the channel slope is greater than either the North Tyne or Ure, the Culm is basically a more gently rolling catchment than either of these. The AAR is 965 mm and only varies from just over 1000 mm to about 850 mm. M52D is 67.6 mm and so although the AAR is lower than for either the North Tyne or Wyre the shorter duration index is similar for all three.

The Dart at Austins Bridge is again a rated section and the river rises on the granites of Dartmoor with poorly drained peat soils. However, the lower half of the catchment is a mixture of Carboniferous sandstones and grits and soils are well drained brown podzols. Slopes are steep in the headwaters but only moderately so on the lower sandstone areas. This is however, the steepest of the six catchments and has the highest AAR of 1821 mm. AAR varies from over 2400 mm on Dartmoor to 1350 mm and is very variable although the variation follows the topography closely. The variation in AAR is mirrored by variations in M52D which is 110 mm.

The Wyre at St Michaels is a natural rated section which drains from the Forest of Bowland. This area is a mixture of Carboniferous sandstones and limestones and is covered by poorly drained peaty soils. The lower part of the catchment is Bunter Sandstone with poorly drained gleyed brown earths. Slopes are fairly steep in the headwaters but relatively flat on the Bunter Sandstones and overall this is the flattest catchment of the six. The AAR is 1257 mm, varying from over 2000 mm to 980 mm and M52D is 69.9 mm. Both AAR and M52D vary with topography but since the topography only varies gradually so does rainfall and the AAR gradient is slight over the catchment. The catchment has a similar AAR and M52D to the North Tyne and a similar M52D to the Culm catchment which has a lower AAR. However, the North Tyne has a greater variability in both rainfall indices.

APPENDIX D: NETWORK REDUCTION USING INFORMATION TRANSFER

A brief summary of a method of network reduction using the concept of information transfer is given in this appendix. This was mentioned in Section 7.4.2 as a possible method of modifying an existing network of raingauges.

Consider a sample $X_1, \dots X_{n_1}$ of statistically independent records of rainfall. Then one measure of the information in the sample about the long term mean μ_X is given by the reciprocal of the variance with which this may be estimated

$$I(n_1) = \{var(\hat{\mu}_X)\}^{-1} = \frac{n_1}{\sigma_X^2} \quad \dots \quad (D.1)$$

where σ_X^2 is the variance of the rainfall. If a further n_2 independent measurements $X_{n_1+1}, \dots X_{n_1+n_2}$ are made then the information in the combined sample is

$$I(n_1+n_2) = \frac{n_1 + n_2}{\sigma_X^2} \quad . \quad \dots \quad (D.2)$$

Now suppose that there is station for which the corresponding measurements $Y_1, \dots Y_{n_1+n_2}$ available. Then the information in the combined sample $(X_1, \dots X_{n_1+n_2}, Y_1, \dots Y_{n_1+n_2})$ about the mean μ_X remains as in (D.2). If only the measurements $(X_1, \dots X_{n_1}, Y_1, \dots Y_{n_1+n_2})$ are available then, using a particular estimator, $\hat{\mu}_X^*$ say, based on all the available data leads to an estimate of the long term mean for which

$$var(\hat{\mu}_X^*) = \frac{\sigma_X^2}{n_1} \{1 - \frac{n_2}{n_1+n_2} \frac{1 - (n_1-2)\rho^2}{1 - (n_2-2)}\} \quad \dots \quad (D.3)$$

this variance is less than the variance of the usual estimator based only on $(X_1, \dots X_{n_1})$ if

$$\rho^2 > \frac{1}{n_1-2} \quad . \quad \dots \quad (D.4)$$

Here ρ is the correlation between concurrent measurements (X_t, Y_t). When (D.4) holds it is said that information can be transferred from the station Y to station X and the measure of information content about the mean is given by the reciprocal of (D.3). If (D.4) does not hold then it is said that information cannot be transferred from station Y to station X. The above considerations are relevant when a pair of raingauges have been operating for n_1 time intervals and it is required to know whether it is possible to cease measurements at one station and still be in a position to obtain increasingly accurate estimates of the mean rainfalls at both stations using the measurements at the remaining station. The above formulae all assume that the measurements are normally distributed.

Consider now a situation where a network of gauges (numbered 1, 2, ... P) has been in operation for n_1 years and it is proposed to continue operating a possibly reduced network for a further n_2 years. Let the number $I^*(i,j)$ be a measure of the amount of information that can be transferred from station j to station i. These quantities would be based on the above considerations and be rescaled versions of I, defined to have the properties that

$$0 \leqslant I^*(i,j) \leqslant 1, \quad I^*(i,i) = 1.$$

It is assumed that, in the modified network, if a gauge is deleted it is replaced by information transferred from only one other gauge. Let $\varepsilon(i,j)$ be a decision variable such that, for $i = j$, $\varepsilon(i,j) = 1$ indicates that a station is to be kept in the network and $\varepsilon(i,j) = 0$ indicates it is to be deleted. For $i \neq j$, $\varepsilon(i,j) = 1$ if station j transfers information to station i, otherwise $\varepsilon(i,j) = 0$. The method of designing the reduced network is then to choose the quantities $\varepsilon(i,j)$ such that

$$Z = \sum_{i=1}^{P} \sum_{j=1}^{P} I^*(i,j) \; \varepsilon(i,j) \qquad \ldots\ldots \qquad (D.5)$$

is maximised. This maximisation is performed subject to three sets of constraints. Firstly a budget constraint is imposed:

$$\sum_{i=1}^{P} C(i) \; \varepsilon(i,i) \leqslant B \qquad \ldots\ldots \qquad (D.6)$$

where C(i) is the cost of operating station i and B is the amount of funds available for the total network. The second set of constraints, namely

$$\sum_{j=1}^{P} \varepsilon(i,j) = 1 \qquad (i=1, \ldots P), \qquad \ldots\ldots \qquad (D.7)$$

ensures that a station i receives information from only one other station. The further constraints

$$A \; \varepsilon(j,j) \geqslant \sum_{\substack{i=1 \\ i \neq j}}^{P} \varepsilon(i,j) \qquad (j=1, \ldots P), \qquad \ldots\ldots \qquad (D.8)$$

where A is a large number (A>P), ensures that a gauge is included in the network if it is to be used to transfer information to another gauge. The problem of maximising the expression (D.5) subject to the constraints (D.6) - (D.8) is one of the type known as integer programming problems: there are several computer programme packages available to deal with such problems once the required arrays ($I^*(i,j)$, C(i) etc) have been set up.