123

Illinois State Water Survey
at the
University of Illinois
Urbana, Illinois

APPLICATION OF DENSE RAINGAGE DATA TO REGIONS OF SPARSE DATA COVERAGE

by

Paul T. Schickedanz

Final Report on
Theoretical Frequency Distributions
For Rainfall Data

Principal Investigator: Paul T. Schickedanz

National Science Foundation
Atmospheric Sciences Section
NSF GA-16168

October 1, 1971

ACKNOWLEDGMENTS

This research was supported largely by funds from the Atmospheric Sciences Section of the National Sciences Foundation, NSF GA-16168. Additional support came from the State of Illinois.

The research was performed under the general supervision of Stanley A. Changnon, Jr., Head of the Atmospheric Sciences Section of the Illinois State Water Survey. Useful suggestions were provided by J. Loreena Ivens, Survey Editor, and the drafting of all figures was supervised by John W. Brothers, Jr.

Appreciation is expressed to Marion Busch who supervised and/or performed a large portion of the statistical and data analyses. Student employees who assisted were Barbara Binch, Patricia Coughlin, Susan Lewandowski, and Nancy Thun. The typing of the report was done by Gail Fairfield of the Survey Staff.

CONTENTS

ILLUSTRATIONS

TABLES

INTRODUCTION

Recent research involving dense raingage networks has added materially to our basic knowledge concerning the duration, intensity, and areal extent of storm rainfall. Such studies have often been concerned with interrelationships between storm rainfall parameters and with relationships of these parameters to hailfall, synoptic type, weather type, etc. (Changnon, 1967; Huff and Neill, 1957; Huff, Shipp, and Schickedanz, 1969; Huff, 1970; Huff and Schickedanz, 1970; and Hershfield, 1971). Sometimes the raingage data have been used to determine how well rainfall can be measured by radar (Jones, 1966; Huff, 1966; Huff, et al., 1956; and Stout and Mueller, 1968) or in conjunction with scientific field projects (Wilk, 1961; and Changnon, 1969). It is not the intent of this report to repeat these previous analyses, but rather to concentrate our attention on the applicability of dense raingage data to estimate rainfall parameters in areas of sparse data coverage.

Unfortunately, the economic cost of installing and maintaining dense raingage networks of appreciable size over the continental United States is prohibitive. Often, in weather modification experiments, the size of area in which it is economically feasible to install and operate a dense raingage network is too small to properly study many of the desired features, in particular extra-area effects due to seeding. Therefore, it is of utmost importance to determine which results from dense raingage networks are applicable in larger areas, in areas of limited data coverage, and in different climatic regimes. Often, the only data available in the region of interest will be the daily rainfall (24 hr) amounts from the existing climatological network. These networks have roughly a density of 250 mi^2/gage in the midwest. If it is desired to use hourly data, one is faced with the dismal prospects of 600 mi^2/gage. All analyses in this report will be based on either daily or hourly amounts. Data terms used in this report are defined in the appendix.

The specific purposes of the research described in this report are twofold. The first purpose is to present a method of estimating the required density of a raingage network in order to insure that various rainfall parameters will be measured with a desired level of precision. This method will be based on the available climatological network in the area of interest. The second purpose is to determine the reliability of measurements made in an area of limited data coverage as compared with the reliability that would be available from a more dense network. Knowledge of this reliability enables the user to determine which measurements would or would not be improved if a more dense network were available. The set of techniques presented in this report can be used to fulfill both purposes. These techniques are based on the use of various statistical methods; namely, sampling, regression analysis, and theoretical frequency distributions. In particular, the relations between frequency distributions of point and areal rainfall and the relation of these distributions to distance, the correlation between points, and the time scale on which the event is measured will be investigated.

The techniques are very useful in the design, planning, and evaluation of weather modification experiments. The methods presented are especially useful in evaluating the economic cost of installing additional raingages as opposed to increasing the duration of the experiment.

ESTIMATION OF RAINGAGE DENSITY NECESSARY TO INSURE A DESIRED LEVEL OF PRECISION FROM EXISTING CLIMATOLOGICAL DATA

Hydrological and meteorological field investigations of precipitation, such as rain enhancement experiments, are often required in geographical regions where an inadequate number of raingages exist. A goal of such investigations is to obtain measurements of various rainfall parameters that are measurable within a specified error range or precision. In many applications, it is the areal mean rainfall, not the areal pattern that is desired. For example, in many weather modification experiments, the rainfall is averaged over the area of interest for seeded and non-seeded days, and a test of significance is eventually made without regard to the areal pattern. Thus, in this section a method will be presented for estimating raingage density to insure specified precision, without regard to areal pattern, from the existing climatological network.

In other applications, the areal rainfall pattern is of interest. If the effect of a weather modification experiment, an urban-industrial complex, or some orographic-marine feature on an isolated high and low in the rainfall pattern is being determined, then the density of gages becomes a crucial factor. In order to estimate the required raingage density, it will be necessary to have prior knowledge of the relationship of correlation decay with distance. Therefore this section also will present a method for estimating the raingage density with regard to areal pattern using correlation decay information obtained from the climatological network.

Analysis of Variance Technique for Estimating Raingage Density without Regard to Areal Pattern

Such a method requires prior estimates of the raingage density, the number of experimental units sampled per year, and the number of years of investigation to insure the desired precision in the parameter of interest. These prior estimates of density must be made on the basis of the existing climatological raingage network, usually 250 mi^2/gage. The prior estimates thus obtained are tested by empirically determing the precision of the areal mean rainfall by taking progressively smaller samples of data from a dense raingage network.

Although the method should be generally applicable to various areas throughout the country, the results were tested by using dense raingage data from a 400 mi^2 area. The effect of areal size on the procedure will be discussed in a later section.

Expected Mean Squares and Components of Variance

The variability inherent in the mean rainfall for any period (ignoring areal size, measurement errors, and areal pattern) is composed of variations due to raingages, days, and years. The proposed technique is to estimate the sampling density of the various components from the expected mean squares of the appropriate analysis of variance model. Such a model provides suitable relationships between sampling sizes and density so that it will be necessary only to estimate the components of variability from existing raingages in the area under consideration. The sampling size will then be computed through algebraic relationships.

The analysis of variance model pertaining to an extended completely random design with subsampling is proposed (Ostle, 1963; Steele and Torrie, 1960). This design utilizes the nested or hierarchical classification. It is assumed that the rainfall amounts from year to year are random quantities. It is further assumed that the arrangement of gages in the network constitutes a random sampling of the storms which move through the network.

A gage observation of rainfall from the above design is composed of a sum of components, namely, a mean and several random elements. The appropriate statistical model is

$$X_{ijk} = \overline{X}... + Y_i + D_{ij} + G_{ijk} \qquad (1)$$

$$i = 1, r$$
$$j = 1, s$$
$$k = 1, t$$

where $\overline{X}...$ is the overall mean rainfall, $Y_i.$ is the effect of year i, D_{ij} is the effect of day j within year i, and G_{ijk} is the effect of gage k on the jth day. It is assumed that $\overline{X}...$ is a constant, and that Y_i, D_{ij}, and G_{ijk} are normally and independently distributed with a mean of zero and a common variance.

The above considerations lead to the analysis of variance table presented in Table 1. The table illustrates that the expected mean square for gages within days (sampling error) contains only one component of variance, in this case, the variance due to gages within days. This term contains only one component of variance because the only factor which affects the variation among samples (gages) within days is the G_{ijk} factor (equation 1). However, the expected mean square for days within years (experimental error) contains two components of variance since this source of variation reflects the variation among the means of the gage amounts on each day, and these means will vary not only because of the variation from day to day but also because of the variation among the gages on each day. The expected mean square for years reflects the variation among the means of all the gage observations recorded for each year. These will vary because of three contributing factors: 1) variation among years, 2) variation among days within years, and 3) variation among gages within days.

Table 1. Analysis of variance model for the random design with subsampling.

Source of variation	Sum of squares	Degrees of freedom	Mean square	Expected mean square
Between years	$SS = st\sum_{i=1}^{r}(\overline{X}_{i..} - \overline{X}_{...})^2$	r-1	SS/r-1	$\sigma^2_g + t\sigma^2_d + ts\sigma^2_y$
Among days within years	$SS = t\sum_{i=1}^{r}\sum_{j=1}^{s}(\overline{X}_{ij.} - \overline{X}_{i..})^2$	r(s-1)	SS/r(s-1)	$\sigma^2_g + t\sigma^2_d$
Among gages within days	$SS = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(X_{ijv} - \overline{X}_{ij.})^2$	rs(t-1)	SS/rs(t-1)	σ^2_g
Total	$SS = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(X_{ijv} - \overline{X}_{...})^2$	rst-1		

When the experimental error is designated as σ^2_e, manipulation of the expected mean square for days within years yields the following relationship for σ^2_d (the variance due to the differences between days)

$$\sigma^2_d = \frac{\sigma^2_e - \sigma^2_g}{t} \tag{2}$$

When the variance for years is designated as σ^2_x, manipulation of the expected mean square for years yields the following relationship for σ^2_y (the variance due to differences between years)

$$\sigma^2_y = \frac{\sigma^2_x - (\sigma^2_g + t\sigma^2_d)}{ts} \tag{3}$$

The variance of daily areal mean, $V(\bar{X}_{ij.})$, can be obtained by dividing the estimated variance σ^2_g of the individual gages contribution to $\bar{X}_{ij.}$ (the daily areal mean) by the number of gages per day. Thus the equation for the variance of the daily areal mean becomes

$$V(\bar{X}_{ij.}) = \sigma^2_{\bar{g}} = \sigma^2_g / \sqrt{t} \tag{4}$$

The variance of a yearly areal mean, $V(\bar{X}_{i..})$, can be obtained by dividing the estimated variance σ^2_e of the individual items contributing to $\bar{X}_{i..}$ by the number of gage amounts averaged to obtain the mean. The division of the experimental error by ts yields the following relationship

$$V(\bar{X}_{i..}) = \frac{\sigma^2_g + t\sigma^2_d}{ts} = \frac{\sigma^2_g}{ts} + \frac{\sigma^2_d}{s} \tag{5}$$

An examination of equation 5 leads to certain conclusions: 1) If the estimates of the components of variance σ^2_g and σ^2_d remain relatively constant, an increase in t or s will result in a smaller estimated variance of a yearly mean. 2) An increase in s (the number of days per year) will have more of an effect than an increase in t (the number of gages per day) in reducing $V(\bar{X}_{i..})$. This suggests that in a weather modification experiment, differences in weather conditions from day to day contribute more to the variance than differences in gage amounts within an area. 3) If either σ^2_g or σ^2_d, or both, can be made smaller, the variance of the yearly mean could be made smaller also. This can be done by choosing more homogeneous days (stratification) or improving the experimental technique in a weather modification experiment.

In order to estimate the variance of the overall mean, the expected mean square for years is divided by rst to yield the following relationship

$$V(\bar{X}...) = \frac{\sigma^2_g}{rst} + \frac{\sigma^2_d}{rs} + \frac{\sigma^2_y}{r} \tag{6}$$

Examination of equation 6 leads to the conclusion that an increase in r (the number of years) has more of an effect on $V(\bar{X}...)$ than s (the number of days per year), and s has more of an effect on $V(\bar{X}...)$ than t (the number of gages per day).

The precision of a mean is a measure of the repeatability of the mean. Precision may be expressed in terms of the variance of the mean, with a large variance indicating lack of precision and a small variance indicating high precision. Absolute precision would be represented by zero variance. Thus, precision is an expression of the variance of the mean, whereas the coefficient of variation is an expression of the variability of individual values about the mean.

Examples of the Technique

Error of the daily areal mean. The computation of required sample sizes to insure that the daily areal mean (areal mean on an individual day) will be measured within a given precision level requires an estimate of σ^2_g . Certainly, σ^2_g could be estimated from a dense raingage network, but it would be strictly applicable for that region only. Thus, estimates of σ^2_g will be made by using all gages in the ECI network (49) and from using only the four corner gages (see Figure 4). Since the four corner gages are approximately 16.8 miles apart, they represent a climatological network of 282 mi^2/gage. When estimates of a^2_g or σa^2_g are based on the 4-gage network, they will be designated as "predicted". The computations are made by using the mean square relationship from Table 1 and equation 4. The predicted estimates will then represent the degree to which σ^2_g or $\sigma^2_{\bar{g}}$ can be approximated from the climatological network.

Figure 1 shows estimates of the relative standard error, RSE, ($\sigma_{\bar{g}}$ expressed as a percentage of the daily mean*) for differing distances between gages and for different raingage amounts for the period 1964-1967. (Raingage amounts DRHA, DHA, CRHA, and CHA are defined in the appendix.) Both the predicted and the actual curves show a considerable increase in the error of the daily mean as the distance between gages is increased. The actual RSE increases from 4% at

* Whenever a standard error is expressed as a percentage of the mean, it will be defined as the relative standard error (RSE).

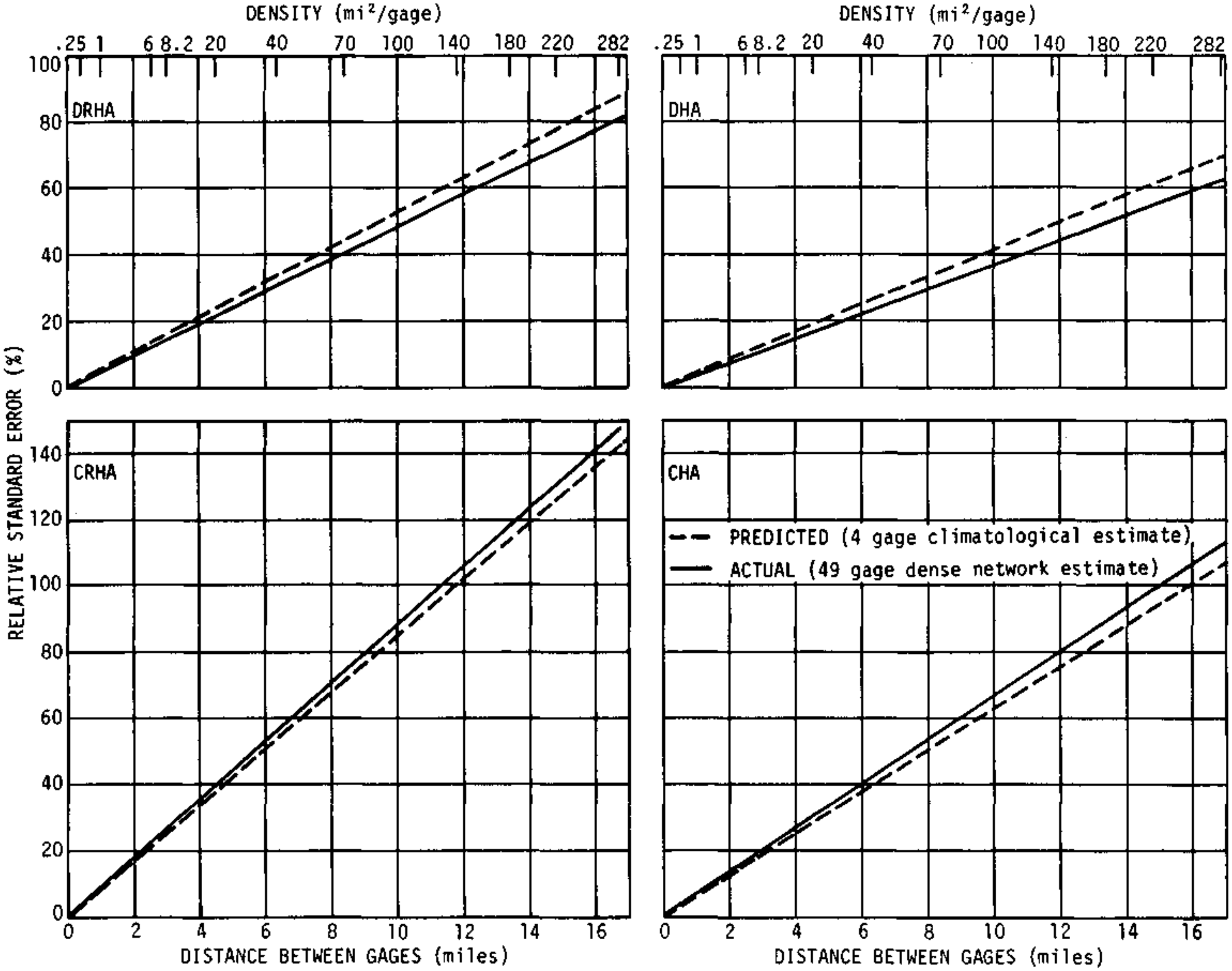

Figure 1. Effect of gage density on the relative standard error (RSE) of daily areal means.

a distance of 1 mi to 60% at the distance of 16 mi for DHA. There is excellent agreement between predicted and actual values of RSE. This agreement indicates that an adequate approximation of σ^2_g and σ^2_g can be made from a climatological network.

At a density of 282 mi^2/gage or a distance separation between gages of 16.8 mi the actual error in measuring the *areal mean rainfall on a particular day* will be 80% of the mean for DRHA and 62% for DHA. Convective rainfall has more variability than other types, and this is evidenced by the curves for convective hour rainfall which show that the actual error is 142% of the daily areal mean for CRHA and 114% of the mean for CHA when the climatological density is 282 mi^2/gage. There is a huge reduction in the error as the density of gages increases. At a density of 8.2 mi^2/gage (the density of the Illinois dense networks) the error is 12, 11, 25, and 19% respectively for DRHA, DHA, CRHA, and CHA. For a very dense network of 1 mi^2/gage the error is 4, 4, 9, and 6% respectively for DRHA, DHA, CRHA, and CHA.

Error of the yearly areal mean. The computation of required sample size to insure that the yearly areal mean (average of all daily areal means within an individual year) will be measured within a given precision level requires estimates of σ^2_d and σ^2_e in addition to σ^2_g . Again, the 4-gage climatological network was used to obtain predicted values. Both predicted and actual estimates were made of σ^2_e and σ^2_d using the mean square relationship of Table 1 and equation 5. Estimates of RSE for the yearly areal mean are shown in Figure 2. The predicted and actual curves are relatively flat, indicating a very small effect of gage density on the yearly areal mean. The actual RSE increases from 21% at a distance of 1 mi to 22% at a distance of 16 mi for DHA, 50 days. This is in direct contrast to the large effect of gage density on the daily areal mean.

There are differences according to the duration of the network operation within the year (Figure 2). If the network operation involves 25 days with rain,

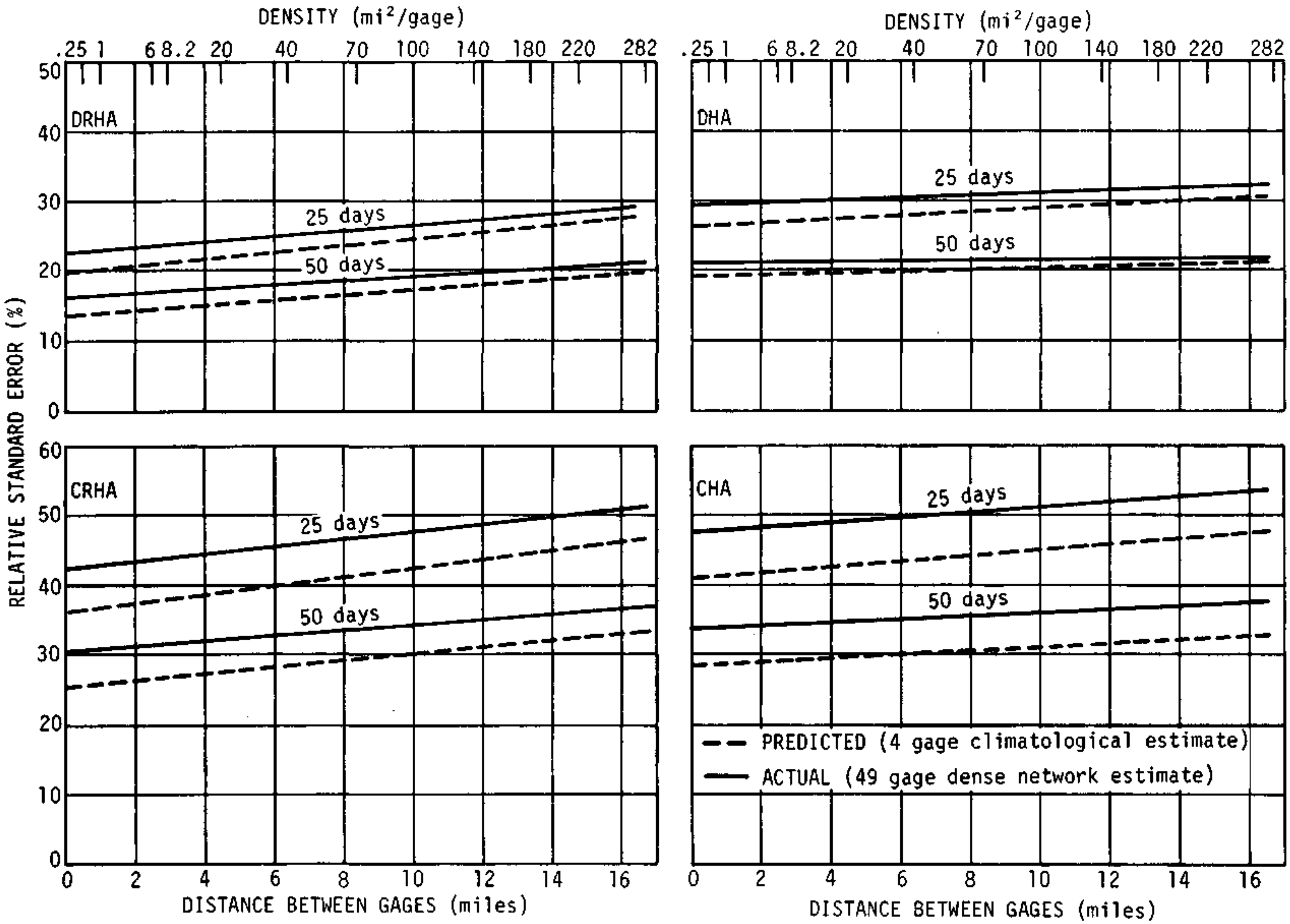

Figure 2. Effect of gage density on the relative standard error (RSE) of yearly areal means.

the actual RSE is 30%, as opposed to 21% for 50 days at a distance of 1 mi for DHA. For a distance of 16 mi, the RSE is 32% and 23% respectively for 25 and 50 days with rain which indicates a very small change for variation in gage density. There is excellent agreement between predicted and actual values of RSE. This agreement indicates that a good approximation can be made of the standard error of the yearly areal mean from the climatological network.

At a density of 282 mi^2/gage, the actual error in measuring the *yearly areal mean rainfall in a partieular year* will be 20% of the yearly mean for DRHA and 23 % of the mean for DHA (50 day duration). The error is larger for convective rainfall with RSE being 51% for CRHA and 53% for CHA (50 day duration). The RSE differences between rainy and all-hour amounts, and the RSE differences between convective and total-day amounts are much less for the yearly areal mean than similar differences for the daily areal mean (Figures 1 and 2).

Error of the period areal mean. The computation of required sample sizes to insure that the period areal mean will be measured within a given precision level requires estimates of a^2_y in addition to a^2_g and a^2_d. Again, the 4-gage climatological network was used to obtain predicted values. Both predicted and actual estimates were made of $\sigma^2_{\bar{x}}$ and σ^2_y using the mean square relationship of Table 1 and equation 6, and the results are shown on Figure 3. Again the predicted and actual values are relatively flat, indicating a very small effect of gage density on the period areal mean. For a 5-year period mean, there is no change in the actual RSE with density for DHA.

There are differences in RSE as the duration is increased. For a measurement period of one year, the actual RSE is approximately 30% for all densities, whereas it is only 13% for 5 years and 9% for 20 years for DHA.

The agreement of predicted with actual period means is good for DHA. For DRHA the agreement is adequate for the longer periods but not for a period of 1 year. The agreement between predicted and actual values is certainly not as good for the period means as it was for the daily and yearly means. The greater

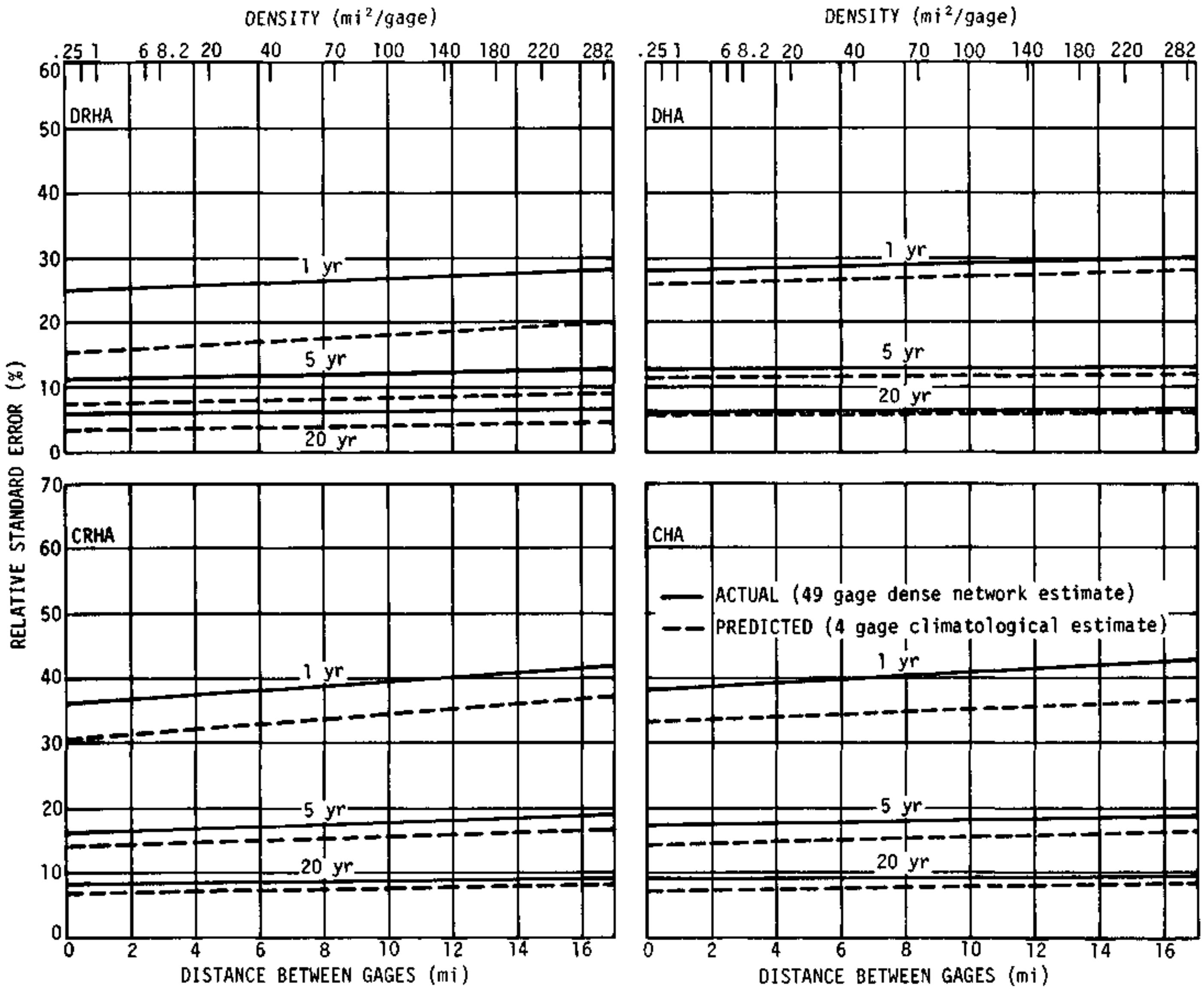

Figure 3. Effect of gage density on the relative standard error (RSE) of period areal means.

discrepancy is because the equation for σ^2_x has more terms than the equations for σ^2_e and σ^2_g and therefore the chance for error is greater.

The net results of these examples indicated that there is considerable trend of error with density for daily areal means (Figure 1). However, the variability between days is a much more important factor for yearly and period means. In fact, for the yearly and period means, there is considerable trend of error with duration of the measurement period but very little trend with density.

Regression Technique for Estimating Raingage Density with Regard to Areal Pattern

For the estimation of raingage density with regard to areal pattern, a knowledge of the relationship of correlation decay with distance will be necessary. The assumption will be made that the correlation-distance relationship can be obtained from existing climatological data for the purpose of estimating sample sizes. Dense raingage data will be used to evaluate the validity of this assumption.

Method for Estimating the Correlation Pattern from the Climatological Network

Two estimates will be made of the correlation pattern from the climatological network. As stated previously, the density of stations recording daily rainfall amounts is approximately 250 mi^2/gage, while the density of stations recording hourly amounts is approximately 600 mi^2/gage.

The ECI network depicted on Figure 4 is a 49-gage network representing a density of 8.2 mi^2/gage. For purposes of simulating a network of daily recording stations from a climatological network, the four corner gages 1, 7, 43, and 49 are used. The average distance $\bar{D}_1$ between these gages (along the boundaries) is

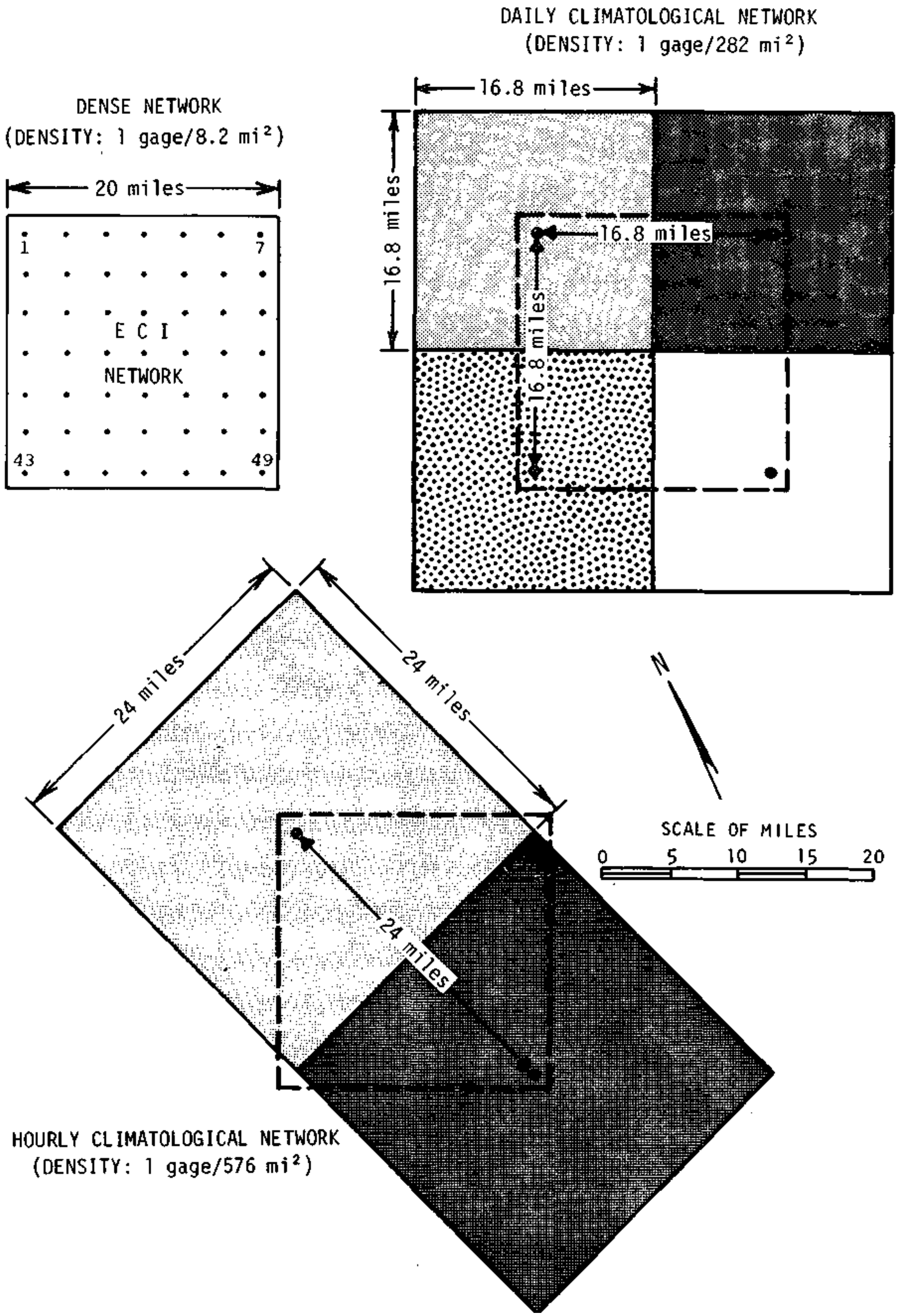

Figure 4. The network configurations used for the climatological estimates of the C-D relationships.

16.8 mi, corresponding to a climatological network of 282 mi^2/gage. In order to approximate the correlation decay with distance, the correlations between gages 1 and 43, 43 and 49, 7 and 49, and 1 and 7 were used to obtain the average correlation ($\bar{R}_{282}$). If we then assume that the correlation approaches 1.0 as the distance between gages approaches 0.0, the slope B_{282} of the regression line of the C-D (correlation-distance) relationship is given by

$$B_{282} = \frac{1.0 - \bar{R}_{282}}{\bar{D}_1} \tag{7}$$

The estimate of correlation R_1 for varying distances between gages is then given by the relationship

$$R_1 = 1.0 - B_{282} D \tag{8}$$

For the distance between gages 1 and 49 and the distance between gages 7 and 43, the average distance $\bar{D}_2$ is 24 mi. This distance corresponds to a climatological network of 576 mi /gage. The slope B_{576} of the regression lines of the C-D relationship is given by

$$B_{576} = \frac{1.0 - \bar{R}_{576}}{\bar{D}_2} \tag{9}$$

The estimated correlation R_2 at varying distances between gages is then given by the relationship

$$R_2 = 1.0 - B_{576} D \tag{10}$$

In order to test the representativeness of these estimates of the C-D relation, an estimate of the C-D relation based on the 49-gage network is needed. The 49-gage estimate R_{act} is obtained by fitting a regression to the data in the form

$$R_{act} = \exp(-BD) \tag{11}$$

Equation 11 was used as the actual estimate of the C-D relationship.

Method for Estimating Raingage Density Given the C-D Relationship

Once an estimate is obtained for the C-D relationship, it is possible to obtain the required raingage density when pattern information is desired. The criteria used in this study to determine the error involved in pattern recognition are based on the predictability of the gage rainfall value from the gage nearest to it. Such an estimate will tend to be "conservative," that is, it will tend to indicate a higher raingage density than is necessary to insure a specified precision in the prediction of the gage value. The higher estimate of density occurs because, in drawing a pattern, the analyst has other supporting data such as other stations and the overall areal trend. However, the predictability of an individual gage from the gage nearest to it serves as a first approximation of the error at various points from the predictor gage in the rainfall pattern.

The error involved in predicting an individual value from the regression line is given by Steele and Torrie (1960) as:

$$SIP = S_{y.x} \left[1 + \frac{1}{n} + \frac{x - \bar{x}}{\sum(x-\bar{x})^2} \right]^{\frac{1}{2}} \qquad (12)$$

where $S_{y.x}$ is the standard error of estimate, n is sample size, and the quantity $x-\bar{x}$ represents a deviation from the mean. We have applied this equation to determine the error in predicting the rainfall at a point from a predictor gage on a particular day. This application involves the simple regression and correlation relationship between gages over time.

In order to apply equation 12 to the determination of error from the climatological network, estimates of $S_{y.x}$ and $\sum(x-\bar{x})^2$ are needed. Our analysis of the ECI network implies that $\sum(x-\bar{x})^2$ has a. small variation over the 49-gage network when it is computed over a period of several years. Therefore, an estimate of $\sum(x-\bar{x})^2$ can be obtained from any gage in the network.

The estimate of $S_{y.x}$ will vary according to the distance between gages, and can be obtained from

$$S_{y.x} = S_y \sqrt{1 - r^2} \qquad (13)$$

where the value of r is obtained from the C-D relationships described previously. The standard deviation S_y has a small variation over the network and therefore any gage can be used to provide an estimate of S_y .

The error involved in predicting the mean rainfall (over a particular time period at a point) from a predictor gage can be determined from

$$SM = S_{y.x} \left[\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x-\bar{x})^2} \right]^{\frac{1}{2}} \tag{14}$$

where n, $S_{y.x}$, and $(x-\bar{x})$ are defined and estimated as in equation 12. The use of equations 12 and 13 provides conservative estimates of the error involved in determining the areal pattern on a given day (equation 12) and in determining the areal pattern for a given time period (equation 13).

Examples of the Technique

As stated previously, the purpose of these analyses is first to estimate the C-D relationship from the climatological network, and then to estimate the required density to insure a given level of precision in the areal pattern. The following examples indicate the extent to which such a scheme is feasible.

Prediction of the C-D relationship from the climatological network. The data from gages corresponding to densities of 282 mi^2/gage and 576 mi^2/gage and equations 7, 8, 9, and 10 were used to estimate the C-D relationship for distances of separation of gages up to 15 miles. The actual estimate was obtained by using the data from all four gages and equation 11. Figure 5 illustrates these estimates for each year of the period 1956-1967 for DRHA. The estimates of the C-D relationship are better for a gage density of 282 mi^2/gage than for one of 576 mi^2/gage. In some years the agreement between predicted and actual is quite good (1956, 1957, 1958, 1961, 1963) and for some years the agreement is poor (19 59, 1962, and 1966). In the years 1960, 1964, and 1967 the estimate is good for the 282 mi^2/gage density but not for a density of 576 mi^2/gage.

Figure 6 illustrates estimates of the C-D relationship for DHA (24 hour average) which is probably the most useful from a climatological point of view. In general,

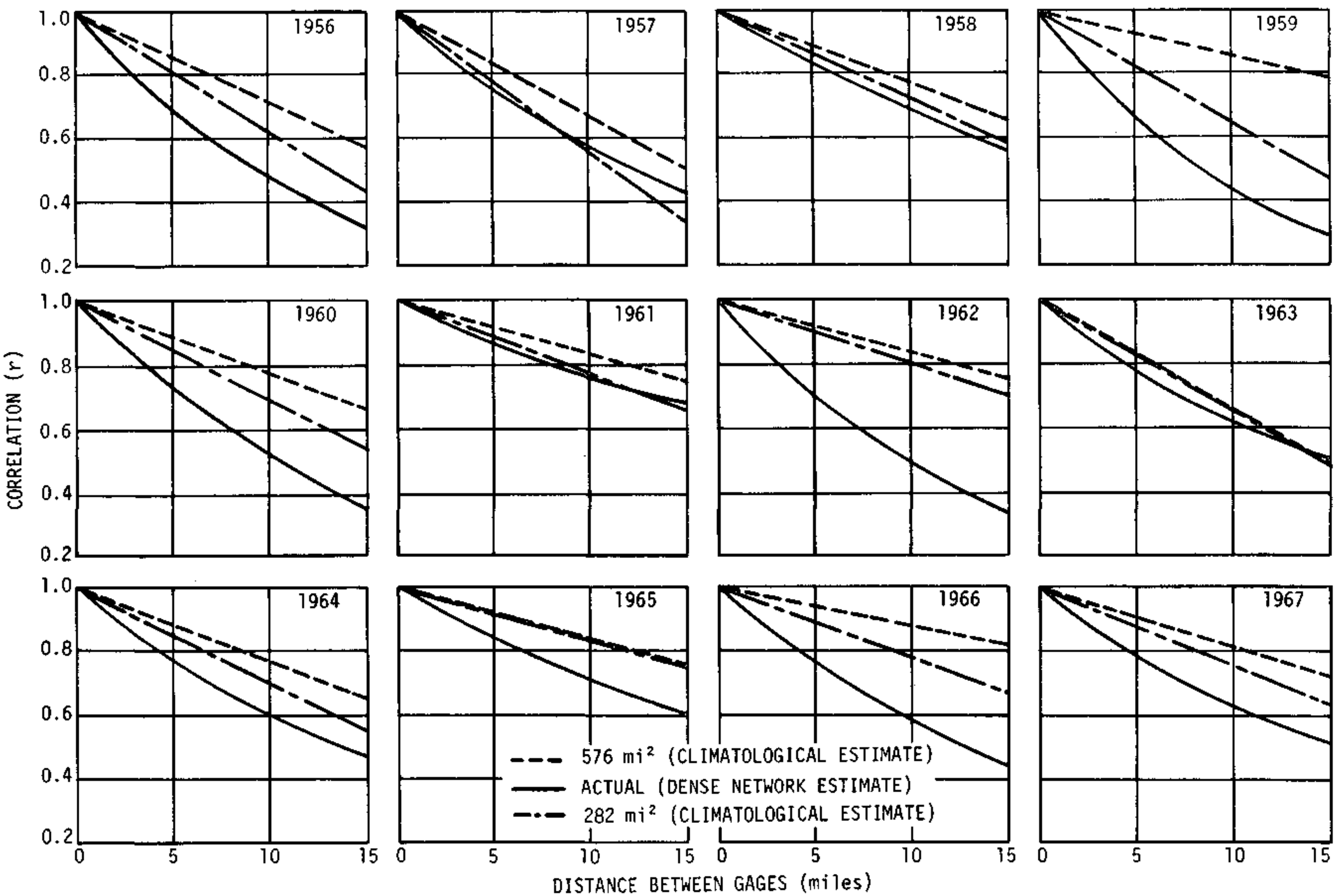

Figure 5. The correlation of gage rainfall amounts and its relationship to the distance between gages (DRHA) by years.

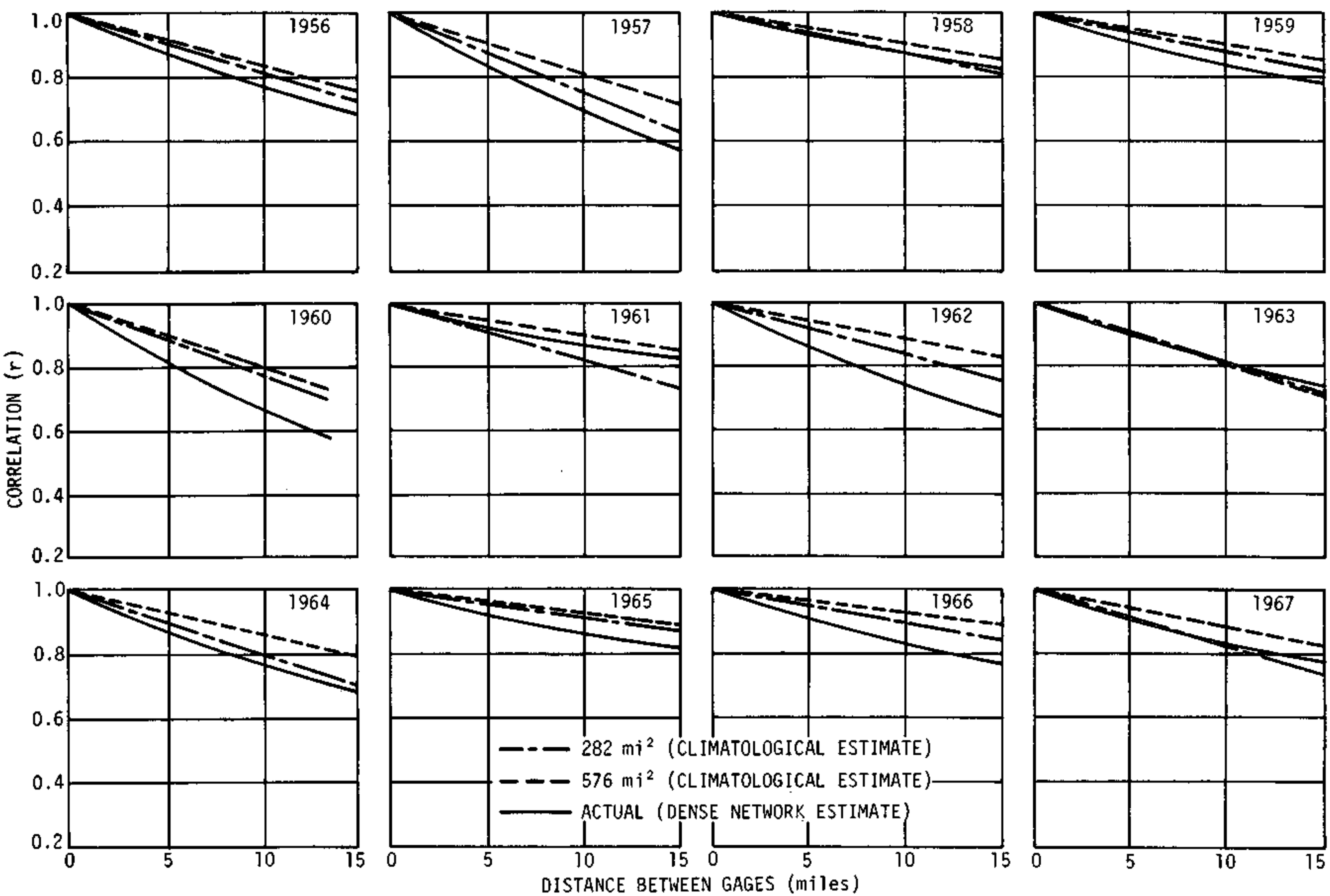

Figure 6. The correlation of gage rainfall amounts and its relationship to the distance between gages (DHA) by years.

the estimates are very good and are much better than those for DRHA. Also, the correlation decay with distance is not as large as with DRHA.

Since it is impractical to vary the raingage density of a network each year, an estimate of the C-D relationship over a period of years is more useful. Figure 7 illustrates estimates of the C-D relationship based on 12 years of data. Actual estimates of the data from a 12-yr and a 4– yr period are included for comparison. The figure indicates that there is less difference between the 4– and 12-yr actual values than between the actual and predicted. There is also a greater difference between the estimated and actual values for CHRA than for DRHA, and for CHA than for DHA. This simply illustrates the larger variability of convective rainfall over that of rainfall in general, which makes it more difficult to determine the required density for convective rainfall. It is also apparent that DRHA has a greater difference between estimated and actual than DHA and that CRHA has a greater difference than CHA. This is a direct reflection of the fact that all-hour rainfall tends to cancel out the extremes typical of the more localized condition in time and space of rainy hour rainfall.

The estimates of the C-D relationship in these examples tend to overestimate the correlation between gages. Thus when the C-D relationship is used to estimate error in the areal pattern, the effect will be to understate the error. These opposite effects tend to make the estimate of error more realistic.

Error in estimating an individual point in a daily pattern. Equation 12 was used to estimate the error SIP involved in predicting an individual value in the rainfall pattern. Values of S_y and $\sum(x-\bar{x})^2$ from gage 25 were used in equations 12 and 13 to obtain an estimate of SIP. Estimates were obtained for three different magnitudes of gage rainfall amounts, x. These were values equal to the mean, the mean plus 1 standard deviation, and the mean plus 2 standard deviations. The SIP values were then expressed as a percentage of the mean* (RSEp). The RSE for the daily areal mean was also included on FLgure 8 for comparison.

* Whenever the error for an individual point is expressed as a percentage of the mean it will be defined as the relative standard error for a point (RSEp).

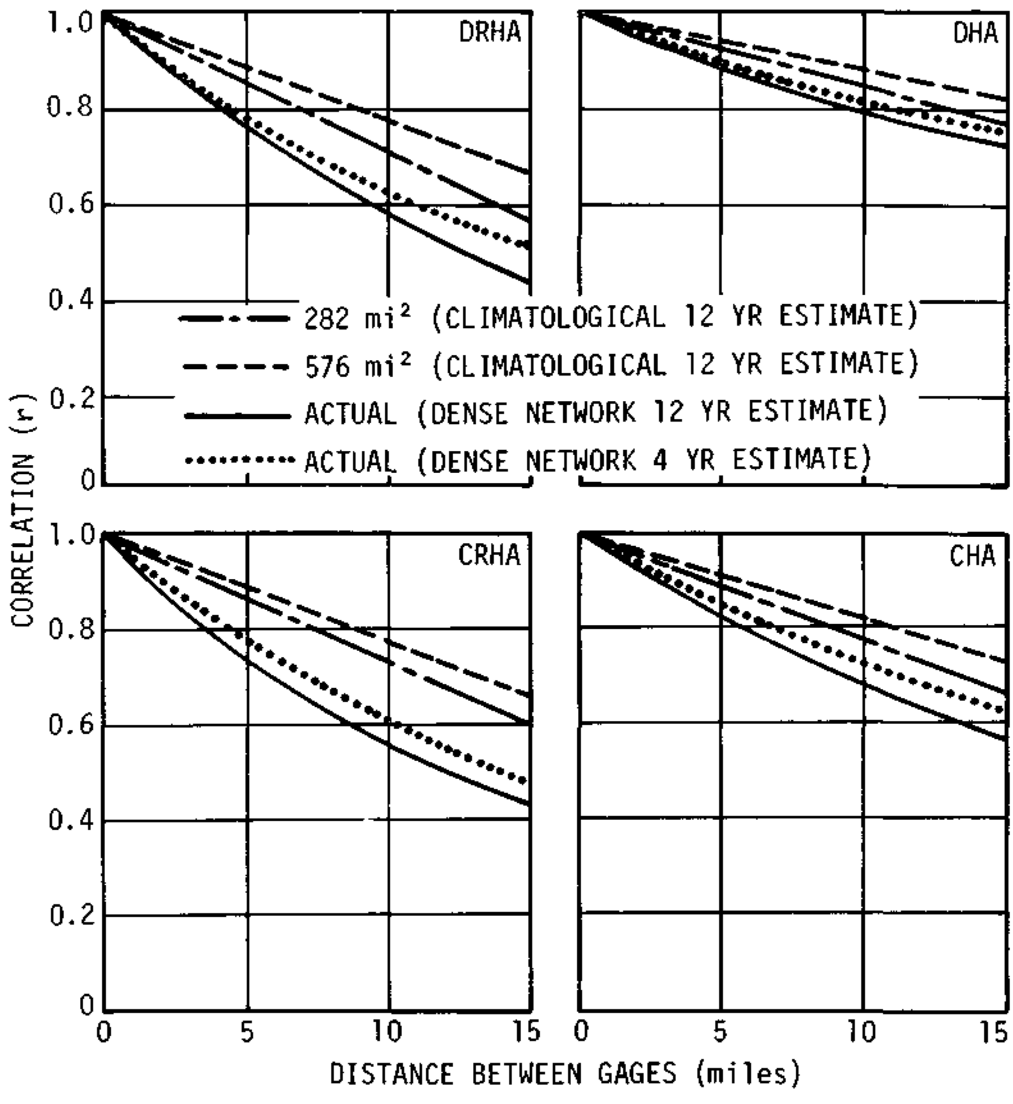

Figure 7. The correlation of gage rainfall amounts and its relationship to the distance between gages for the period 1956-1967.

The error is certainly larger for an individual point on a particular day than it is for the areal mean. For separation distances of 4 mi and larger the difference in error is approximately 60% for DHA. The trend of RSEp is much steeper at the higher raingage densities than it is for RSE. For example, RSEp ranges from 15 to 80% while RSE ranges from 20 to 10% when the density is varied from 0.25 mi^2/gages to 8.2 mi^2gage for DHA.

The errors for the convective hour amounts are much higher than those for the total-day amounts. At a density of 8.2 mi^2/gage RSEp is 80% for DRHA compared with 135% for CRHA, and this again illustrates the greater variability of convective rainfall.

Instead of predicting the gage rainfall at a point from the gage nearest to it, it would also be possible to use all surrounding gages to predict the gage rainfall amount. Such a procedure would entail the use of multiple regression instead of simple regression. For DRHA and DHA, the error involved in predicting an individual point (equal to the mean) from a multiple regression relationship was determined. These results are also shown on Figure 8. The reduction in RSEp is on the order of 20% when a multiple regression relationship is used instead of a simple linear relationship. When we consider the order of magnitude of RSE and RSEp, the simple relationship certainly yields a sufficient estimate of point error in the rainfall pattern if it is treated as a conservative estimate.

The differences between the three magnitudes of RSEp are of little practical importance in light of the wide range in RSEp over the raingage densities, so the predictions of RSEp were made only for individual points equal to the mean. Predictions were made from the 282 mi^2 and the 586 mi^2 climatological network and are depicted in Figure 9.

The differences between the RSEp curves for the two types of predictions are small. Although there are differences between the predicted and actual, they are sufficiently small to adequately depict the trend of error with density. Also, it was indicated in Figure 8 that a multiple regression using the 8 surrounding stations yields smaller RSEp values, which are very comparable to the predicted values in Figure 9.

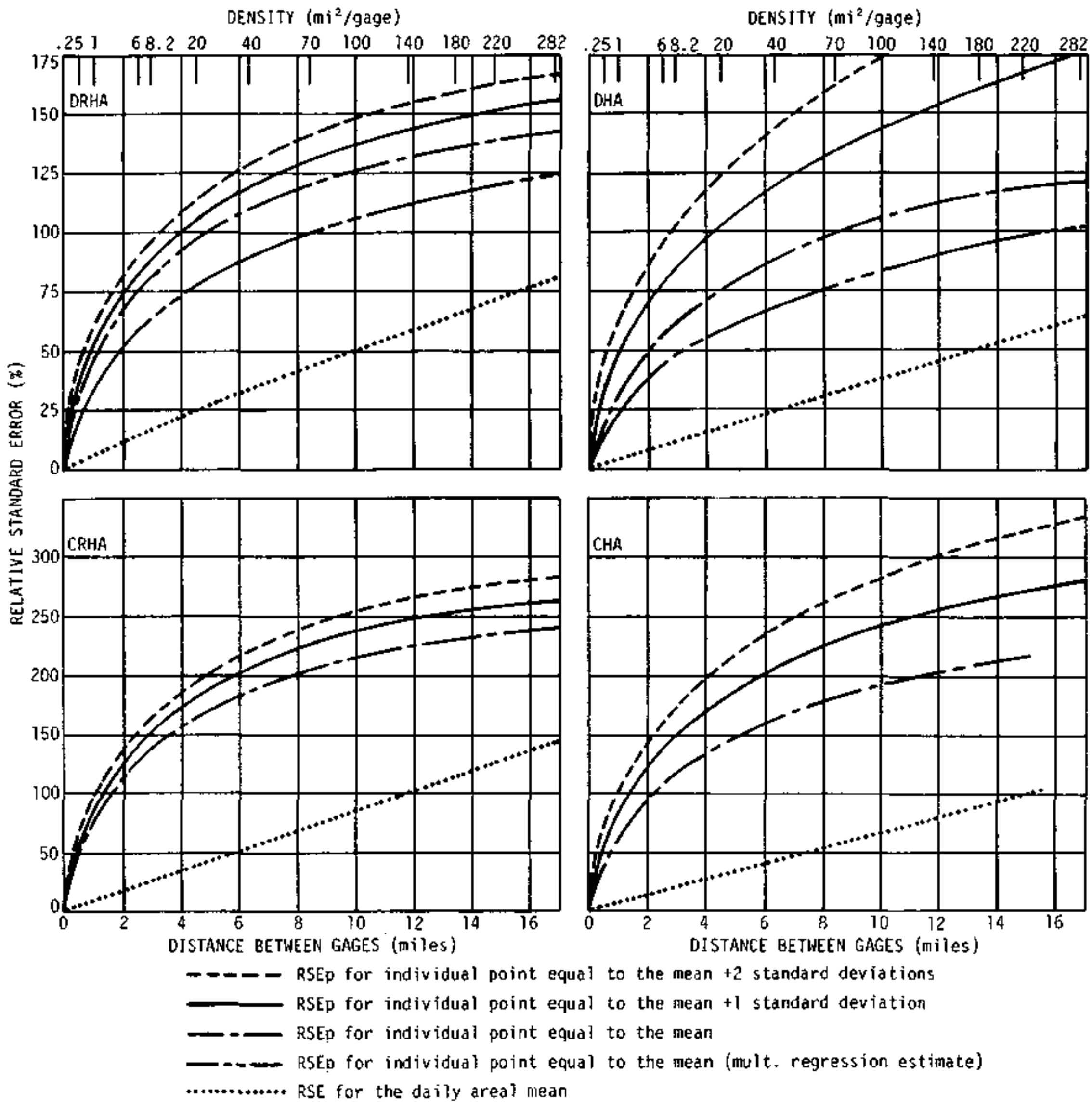

Figure 8. Effect of gage density on the relative standard error (RSEp) of point rainfall estimation in a daily pattern (actual).

Error in estimating an individual point in yearly and period patterns. Equation 14 was used to estimate the error SM involved in predicting the rainfall at a point in the yearly rainfall pattern. Values of S_y and $\sum(x-\bar{x})^2$ from gage 25 were used in equations 12 and 13 to obtain an estimate of SM. The average number of warm season rain days per year was used for n. The results of these computations for DRHA and DHA are shown in Figure 10 (1-yr curves). For values of rainfall equal to the mean, the actual RSE is 20% at a density of 282 mi^2/gage and 13% at a density of 8.2 mi^2/gage for DRHA. For similar values of DHA, the RSE is 17 and 9% respectively. These values of RSE represent a considerable reduction from the error for the daily areal pattern (see Figures 8 and 9).

Equation 14 and an n equal to the appropriate multiple of the average number of rain days per warm season were used to determine the error (SM) for the period patterns. For DRHA, as the period of measurement is increased from 1 year to 20 years, RSE is reduced (Figure 10). For values of rainfall equal to the mean, the actual RSE is reduced from 20 to 9% as the number of years is increased from 1 to 5 years, and the actual RSE is reduced further to 5% as the number of years is increased to 20 for a climatological density of 282 mi^2/gage. Comparable figures for DHA are 17, 8, and 4% for 1, 5, and 20 years, respectively.

The reduction of RSE for periods of 5 and 20 years was so small for values of rainfall equal to the mean plus 1 standard deviation and the mean plus 2 standard deviations that curves for these are not shown. However, the curves for a period of 1 year show that the error is much larger for values away from the mean than for values at the mean.

This agreement between predicted and actual values appears to be close enough to enable one to make the estimates from the climatological network.

Inclusion of the Areal Factor into Estimation Procedures for Raingage Density

The emphasis of this section will be on estimating the relative standard error for a specified raingage density in areas larger than the areal size of most

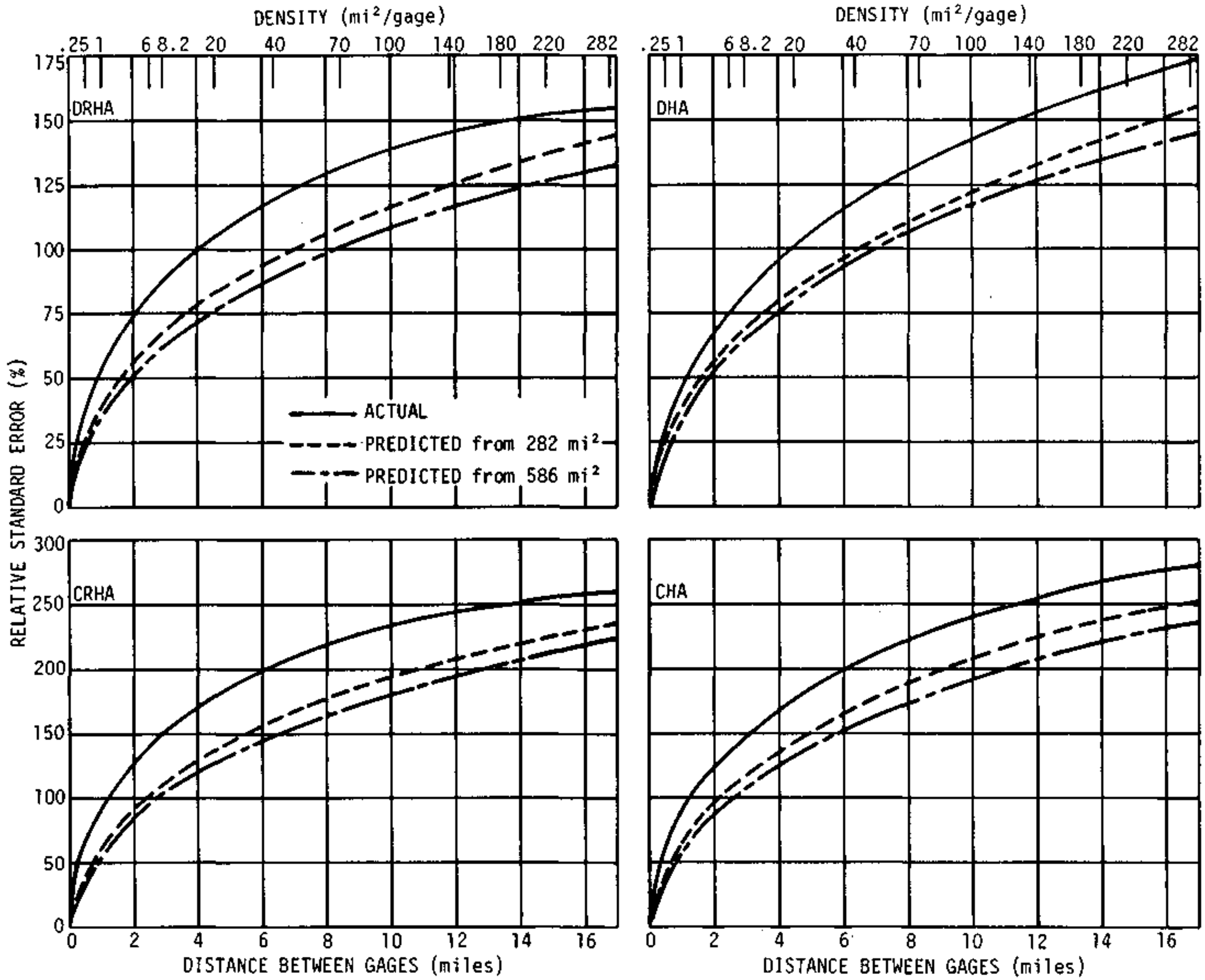

Figure 9. Effect of gage density on the relative standard error (RSEp) of point rainfall estimation in a daily rainfall pattern (predicted versus actual).

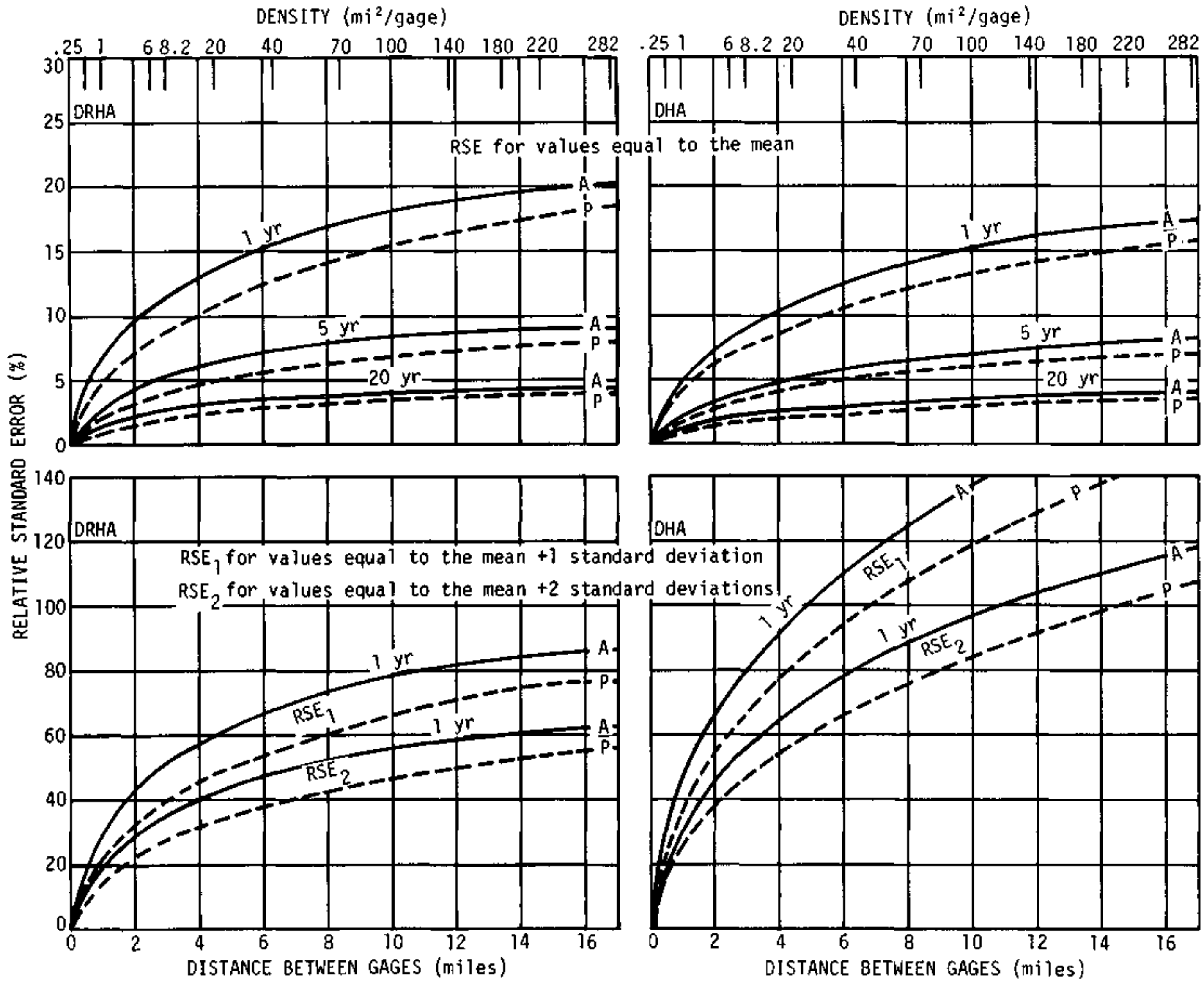

Figure 10. Effect of gage density on the relative standard error (RSE) of point rainfall estimation in a yearly and period pattern (predicted versus actual).

dense raingage networks. The estimation of error could be made in either of two ways: 1) from the dense network in a small subarea or 2) from a climatological network in the larger area. The examples in this section are based on the former but the results will also be applicable for the latter.

Technique without Regard to Areal Pattern

Certainly one of the handicaps of a dense network is that they are often small in areal size. The data used in this report were obtained from a 400 mi^2 network. In weather modification experiments the area of interest is often larger. In Project Whitetop, a major weather modification experiment in the midwest (Decker and Schickedanz, 1967), the experimental area was 11,310 mi^2.

The analysis of variance technique is generally applicable to larger areas, provided estimates of σ^2_g and σ^2_d are available for the larger area. In actual practice, the estimation of σ^2_g would of necessity come from the climatological network. The results of the previous section indicate that an estimate of σ^2_g from a climatological network is sufficient for estimating the error from various densities of gages. However, for the purpose of demonstrating the effect of the areal size for larger densities, the error will be estimated from the 400 mi^2 network.

We will consider areas which are multiples of the 400 mi^2 network. The estimation of σ^2_g for a larger network can then be expressed as the pooled estimate of the subareas as follows:

$$\sigma^2_g = [\sigma^2_{g_1}(t_1-1) + \sigma^2_{g_2}(t_2-1) + ---\sigma^2_{g_k}(t_k-1)]/[(t_1 + t_2 + t_3 + t_4)-K] \qquad (15)$$

where t_1, is the number of gages in the subarea 1, t_2 is the number of gages in the subarea 2, etc., and k represents the number of subareas under consideration. Under the special conditions that $\sigma^2_{g_1} = \sigma^2_{g_2} = ---\sigma^2_{g_k}$ and $t_1 = t_2 = ---t_k$, σ^2_g for the large area is equal to σ^2_g of any subarea.

Therefore, for the areal extension of the estimate of error it is sufficient to use a pooled estimate of σ^2_g values from the subareas; or, when the values for each subarea are equal, the value of any subarea is equal to the value of the total area. Therefore, the estimate of σ^2_g for the 400 mi^2 area can be used in equation 4 to obtain the variance of a daily areal mean.

This procedure will be used in an example to illustrate the effect of areal size on the RSE of daily areal means utilizing network data. However, in actual practice one would use an estimate of σ^2_g based on the climatological stations in the area of interest.

In order to estimate the RSE of a yearly area mean, of an area larger than the dense network, an estimate of σ^2_d must be available for the larger area. In actual practice, σ^2_d like σ^2_g could be estimated from the climatological stations available. However, for our purposes, σ^2_d for the larger area can be obtained from the pooled estimate of σ^2_d for the k subareas. Again, if $\sigma^2_{d_1}$ = $\sigma^2_{d_2}$ = $--\sigma^2_{d_k}$ and $t_1 = t_2 = ---t_k$, the value of σ^2_d from a subarea will equal that of the total area. Therefore, the estimates of σ^2_g and σ^2_d for the 400 mi^2 area can be used in equation 5 to obtain an estimate of the variance of a yearly areal mean for the larger area. In a similar fashion, a pooled estimate of σ^2_y can be used in conjunction with σ^2_g and σ^2_d in equation 6 to get an estimate of the variance of a period areal mean.

Figure 11 illustrates estimates of the relative standard error (RSE) for differing distances between gages for the period 1964-1967. The main feature of the graph is the tendency for the estimate of RSE to be less as the areal size increases. The other feature is the tendency for the response to density to be much less as the size of the area increases. For example, the change in RSE from a density of 8.2 mi^2/gage to a density of 282 mi^2/gage is 72% for a 400 mi^2 area, but is only 12% for a 12,800 mi^2 area.

Estimates of the relative standard error for the yearly areal means were made for the areal sizes depicted in Figure 11. The results of these computations indicated that the difference between curves for 400, 1600, and 12,800 mi^2 areas were so small that the curves of Figure 2 (400 mi^2) are sufficient for the larger areal sizes. The change in relative error in response to the number of days was also similar to that indicated in Figure 2.

Technique with Regard to Areal Pattern

The estimation of error with regard to areal pattern was based on the C-D relationships discussed previously. As a smaller area is expanded into a larger area, the C-D relationship is never based on a gage separation more than that of a climatological network. Thus, even though there is a decrease in correlation as the density of raingages is increased, an increase in areal size does not affect the estimate of relative standard error. Since the climatological estimates of the C-D relationships for the 282 mi^2 network were adequate for the purpose of estimating error, the curves of Figures 8 and 9 can also be used as a good approximation of larger areas.

In actual practice the estimate of the C-D relationship would be made from the large climatological network, and the differences between various subareas would be accounted for.

FREQUENCY INFORMATION FOR A CLIMATOLOGICAL NETWORK COMPARED WITH THAT FROM A DENSE NETWORK

The previous sections of this report have been directed toward the estimation of raingage density in order to insure a given level of precision in the daily areal means (all gage average). We now turn our attention to the amount of error involved in deriving frequency information from a climatological network in relation to a more dense network. This phase of the analysis is based on theoretical frequency distributions. A determination will be made of the error involved in parameter estimation in both areal and temporal frequency distributions. In addition, trends of the parameters with density and area and the goodness of fit of the distributions will be investigated.

Estimation of Areal Distributions from Climatological Data

The approach to the determination of error will again involve the analysis of variance technique which was described previously. In addition, relationships of the distributional parameters with density and area, interrelationships between parameters, and goodness of fit information will be presented.

Determining Relative Standard Error for the Distributional Parameters

The treatment of frequency distributions will be limited to the 2-parameter log-normal and gamma distributions. Initially the extreme value and Weibull distributions were considered. However, the data samples used included the complete range of rainfall values and were not samples of extremes. Furthermore, a technique involving the 3rd and 4th moments (Hahn and Shapiro, 1967) which indicates distributions most likely to fit was used. (Graphs for this purpose can be found in Hahn and Shapiro, 1967.) The employment of this technique led to the conclusion that the data were more likely to be gamma or log-normal rather than the other distributions considered.

The 2-parameter log-normal and gamma distributions include only values greater than zero. Hence, all areal distributional parameters are based only on gages with rain. The analyses of the previous sections of the report included all gages, irrespective of whether they had rain.

Description of the technique. The equations for the estimation of distributional parameters as well as the density functions are presented elsewhere (Schickedanz et al., 1969; Thorn, 1958; Hahn and Shapiro, 1967) and will not be repeated here.

In addition to the gamma parameters (scale β and shape γ) and the log-normal distributional parameters (scale $\overline{\ln x}$ and shape $\sigma_{\ln x}$), the non-transformed means X and standard deviations σ for the gages with rain were also considered. The non-transformed means and standard deviations represent the normal distributional parameters and are useful for comparison purposes.

The technique used for the determination of RSE for the mean $\bar{x}$ and log mean $\overline{\ln x}$ is the same as was described and used previously in determining error

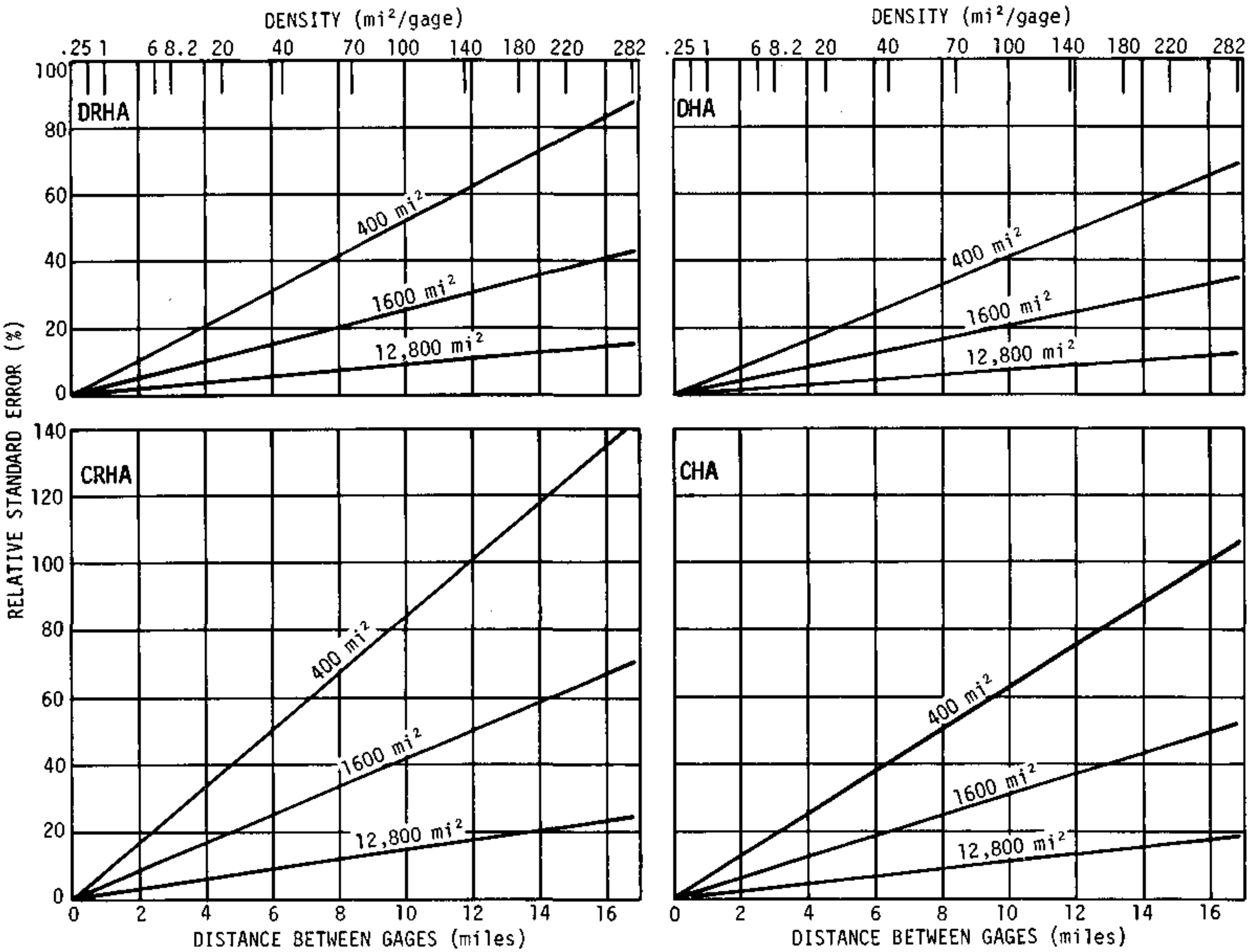

Figure 11. Effect of areal size and gage density on the relative standard error (RSE) of daily areal means.

for the daily areal means (all gage average). However, a subset of 6 gages from the ECI network was used for the climatological estimate instead of the set of 4 gages used previously. This was necessary because a subset of 6 gages had an average of only 4 gages with rain, and a distributional parameter based on less than 4 gages would have little meaning. This means that in actual practice, the areal distributional parameters can not be estimated unless the area is larger than 400 mi^2.

The variance of the gamma shape parameter γ was estimated from the following equation (Thorn, 1958)

$$V(\gamma) = \frac{\gamma}{N\,[\gamma\;\Psi(\gamma) - 1]} \tag{16}$$

where $\Psi(\gamma)$ is the tri-gamma function and can be evaluated from tables of the tri-gamma function (Davis, 19 33). The value of gamma used for daily areal values in this research was a pooled estimate of all areal daily values at a particular density and areal size for the 1964-1967 data. The value of γ obtained from the distribution of daily areal means was used for yearly areal values. The variance of the gamma parameter ß was estimated from the following equation (Thorn, 1958)

$$V(\beta) = \frac{\beta^2\;\Psi(\gamma)}{N\;(\gamma\;\Psi(\gamma) - 1)} \tag{17}$$

where $\Psi(\gamma)$ is again the tri-gamma function and the value of ß used for daily areal values is the pooled estimate of all areal daily ß values at a particular density and areal size. The value of ß obtained from the distribution of daily areal means was used for the yearly areal values.

Examples of the technique. The appropriate equations from Table 2 was used to estimate the a^2 for the non-transformed rainy gage data and for the log-transformed data. For computational purposes, the value to t was replaced by the average number of gages with rain on a given day. Equation 4 was used to compute the variance of $\overline{X}$ and $\overline{\ln x}$ for the rainy gage data. The results are presented on Figure 12. The actual values are based on the 49-gage network for which the average number of gages with rain is 33. The predicted (climatological) values are based on a subset of 6 gages for which the average number of gages with rain is 4. As mentioned previously, this would indicate that in actual practice (density of 282 mi^2/gage), the network would have to be more than 400 mi^2 before the areal distributions could be calculated. However, because of the similarity of the actual and predicted values of the log-normal mean, an estimate of error based on 600 mi^2 or more with the usual climatological density would be adequate. However, there are differences in the curves for convective amounts (Figure 12), and it is doubtful if the error can be approximated adequately from the climatological network for areal means (rainy gage average).

The striking feature of Figure 12 is the small error involved in measuring the daily areal log-normal scale parameter with a coarse network of gages. For example, the actual curves for DRHA indicated only a 18% error in the log-normal mean for a density of 282 mi^2/gage compared with a 4% error at a density of 8.2 mi^2/gage. Corresponding numbers for the non-transformed mean are much larger, 60 and 10% respectively.

The variance of the daily areal values of γ and β were determined by using equations 16 and 17 and the pooled values of γ and β. The relative standard error was also computed and the results are depicted on Figures 13 and 14.

The agreement between predicted and actual values is very good for the total-day amounts DRHA and DHA, whereas the agreement for the convective averages is not as good. The striking feature of Figures 13 and 14 is the sharp response of γ and β to the density of raingages. The response is much greater for the gamma distribution parameters than for the log-normal scale parameter $\overline{\ln x}$ or for the nontransformed mean $\bar{X}$. In fact the results indicated that the estimation of gamma areal parameters on a particular day will produce errors of such magnitude that the parameter would be virtually useless for the typical climatological network. In general, to keep the error less than 30% a gage density of 8.2 mi^2/gage is required.

The appropriate equation from Table 1 and equation 2 were used to estimate $\sigma^2_{\bar{e}}$ for the yearly areal mean (rainy gage average) for non-transformed and transformed data. The relative standard error for various densities was then computed, and the results are shown on Figure 15.

For both the non-transformed mean and the log-normal scale parameter $\overline{\ln x}$, the response of relative error to density is nil. The striking feature of the graph is the

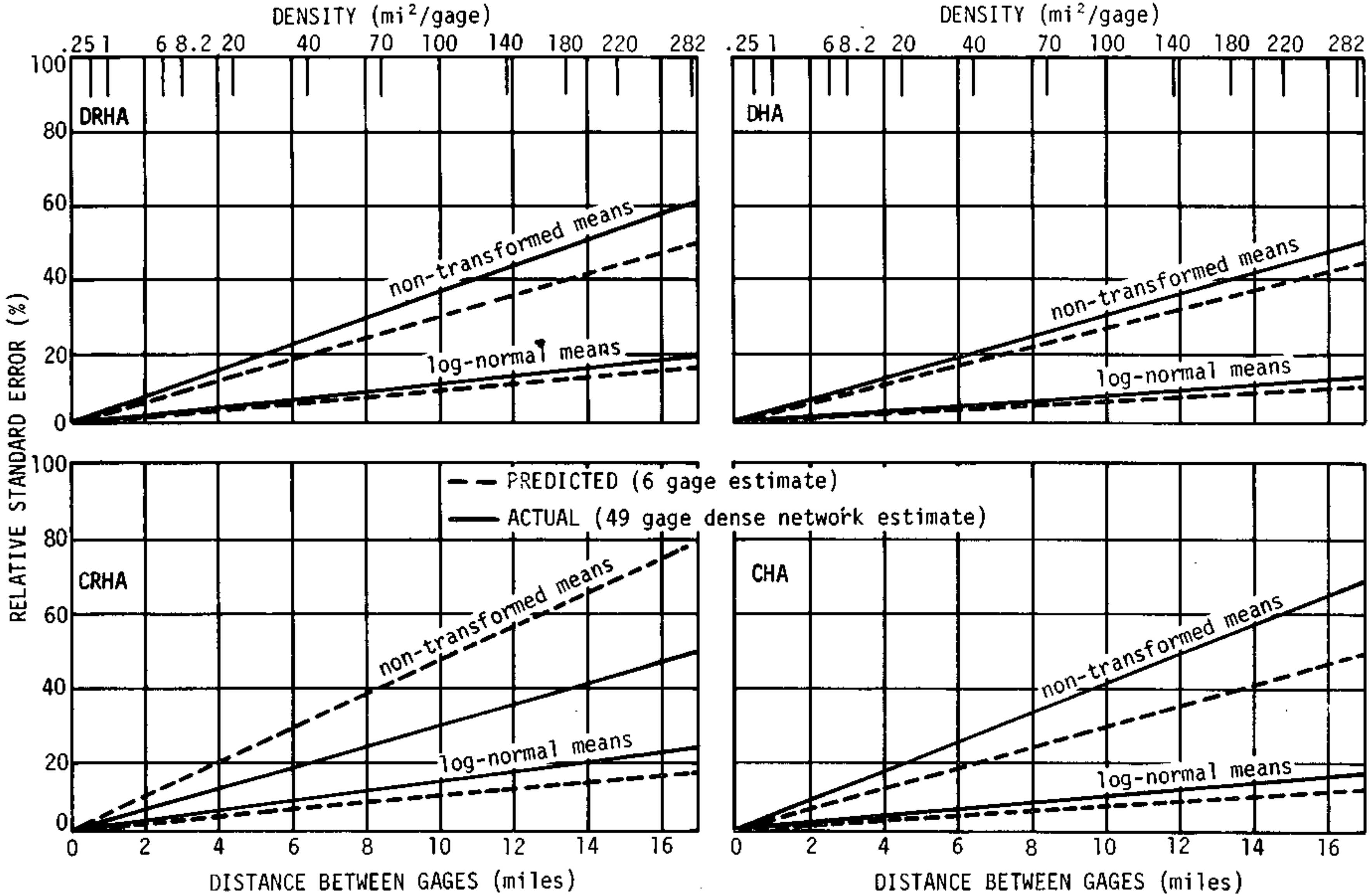

Figure 12. Effect of gage density on the relative standard error (RSE) of daily non-transformed and log-normal areal means.

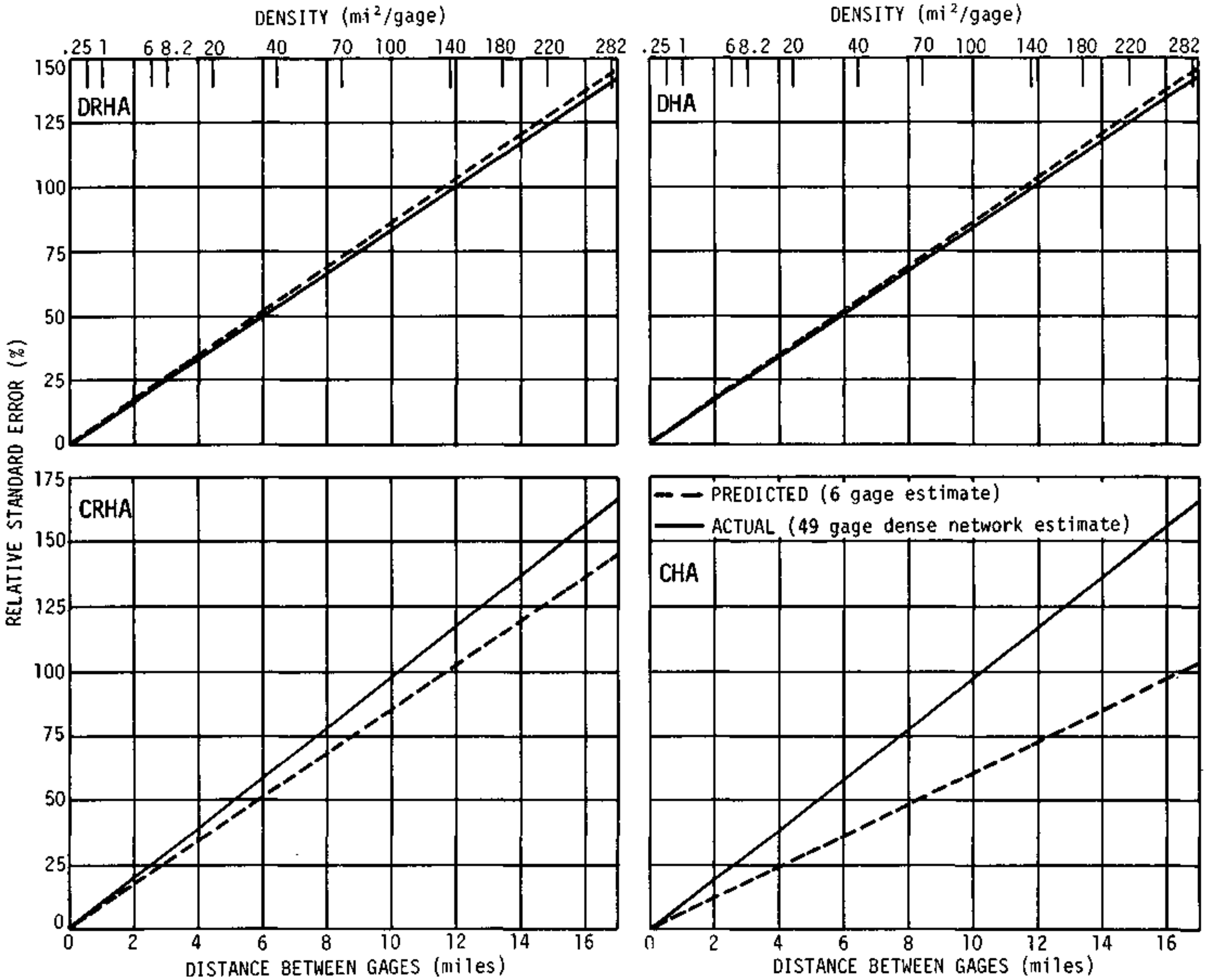

Figure 13. Effect of gage density on the relative standard error (RSE) of daily γ parameters.

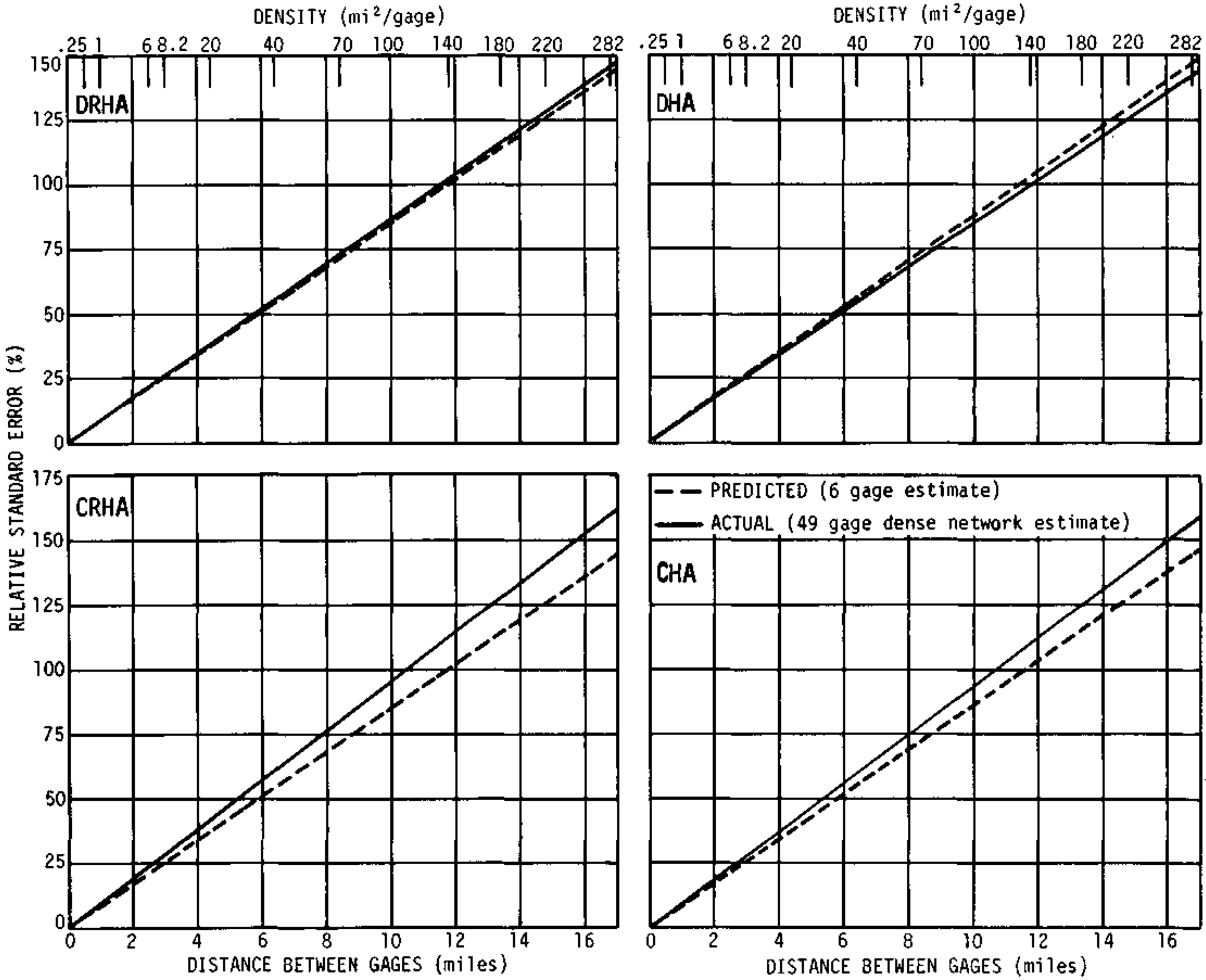

Figure 14. Effect of gage density on the relative standard error (RSE) of daily ß parameters.

the small change in error using a period of 50 days instead of 25 days.

Since the response of the relative standard error for $\bar{x}$ and $\overline{\ln x}$ is small, the relative standard error for the gamma parameters was computed only by equations 16 and 17 for the density of 8.2 mi /gage. The results are tabulated in Table 2. The results indicate that the error in the yearly areal values are nearly the same for all types of gage, amounts and that the error for ß is larger than that for γ. Also, the difference between the estimate of error for a period of 25 days and 50 days is larger than it was for $\overline{\ln x}$ and approximately the same as that for $\bar{x}$ (see Figure 15).

Table 2. Effect of sample size on the relative standard error (RSE) of yearly γ and ß values (density of 8.2 mi^2/gage).

		Relative standard error (%)		
Days	DRHA	DHA	CRHA	CHA
		γ parameter		
25	26.0	24.5	25.9	25.1
50	18.4	17.4	18.3	17.8
		3 parameter		
25	30.1	32.2	30.4	31.7
50	21.2	22.8	21.4	22.4

Inclusion of the areal factor. An estimate of error for areas larger than the dense network can be obtained from the climatological network. This is possible because the trend of the pooled estimates of σ, $\sigma_{\ln x}$, γ and ß with density is small, as will be shown in a later section. Conceivably, the estimate could be made by extending the estimates from small areas into longer areas (page 27), This might be possible in the case of $\bar{x}$ and $\overline{\ln x}$ because the trend of σ and $\sigma_{\ln x}$ with areal size is relatively small. However, in case of γ and ß the estimate must be made from the climatological network, because there is a strong trend of these parameters with area. In addition, for practicable use the estimate would of necessity come from the climatological network.

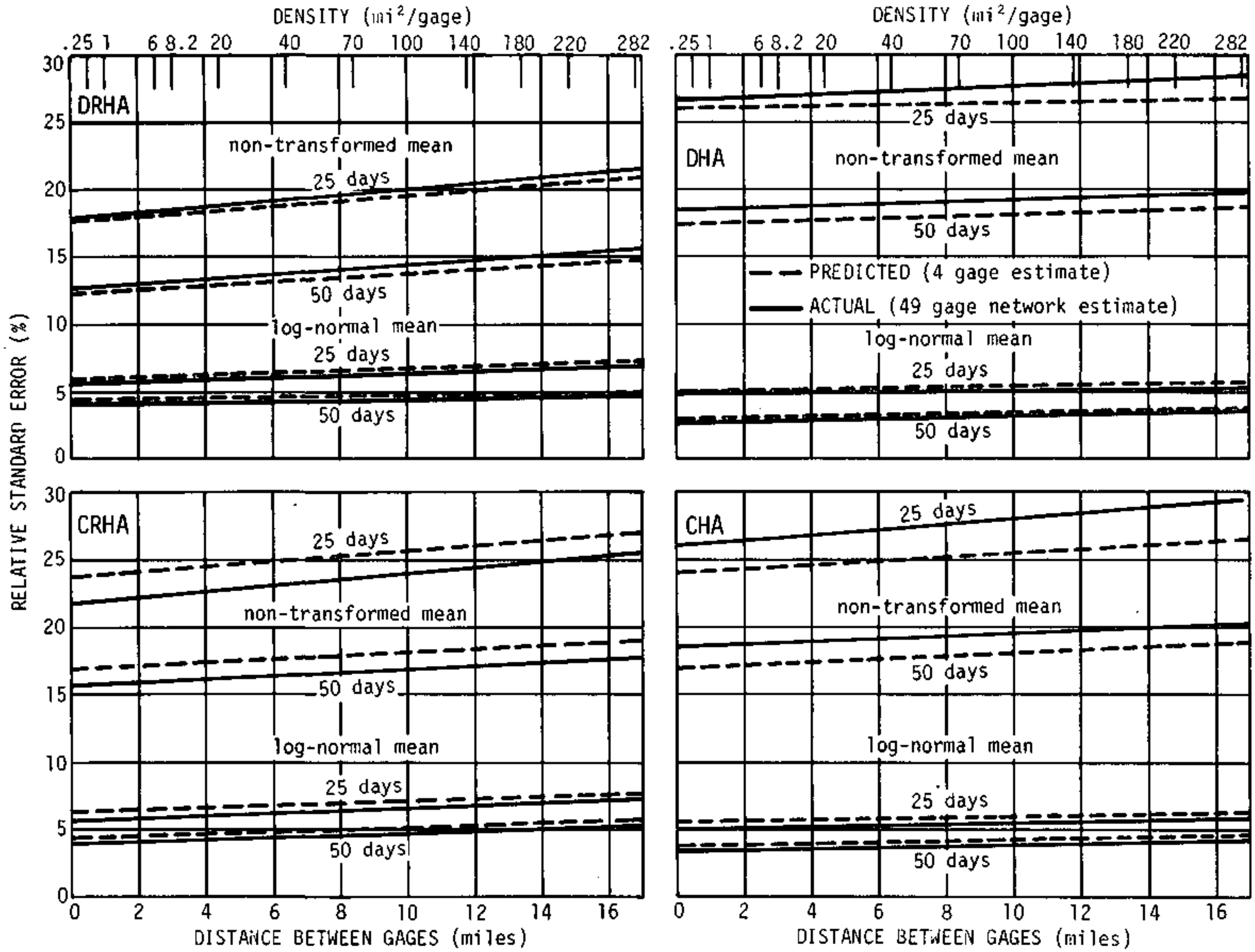

Figure 15. Effect of gage density on the relative standard error (RSE) of yearly non-transformed and log-normal means.

Relationships of Areal Distribution Parameters with Raingage Density and Area! Size

In the previous sections information was presented on the effect of density and areal size on the relative standard error of various distributional parameters. We now turn our attention to the expected value of the distributional parameter for a given density and areal size.

Relationships with raingage density. The average values of areal distributional parameters for a given density and areal size were computed for the period 1964-1967 and the results are shown in Table 3, for DRHA, DHA, CRHA, and CHA.

Table 3. Relationship of density with areal distributional parameters based on rainy gage amounts.

	Value of the parameter in percent*						
Density (mi^2/gage)	Mean $\bar{X}$	Log mean $\overline{\ln x}$	Standard deviation σ	Log standard deviation $\sigma_{\ln x}$	Gamma γ	Beta β	% of gages with rain
				DRHA			
8.2	93	97	99	103	73	96	67
16.	99	99	106	108	88	104	68
30.8	102	100	104	102	94	109	69
66.7	106	104	91	87	109	92	67
				DHA			
8.2	92	97	92	103	79	91	67
16.	96	99	101	107	83	102	68
30.8	101	101	101	100	115	100	69
66.7	111	103	106	90	124	107	67
				CRHA			
8.2	91	97	100	105	63	75	51
16.	101	97	102	115	75	84	52
30.8	94	103	110	99	125	131	54
66.7	114	106	88	81	137	111	67
				CHA			
8.2	87	96	96	110	61	91	51
16.	92	99	100	107	71	96	52
30.8	102	101	104	101	87	106	54
66.7	118	104	101	82	180	107	67

* Percent is computed as the percentage of the average of all densities for a given parameter.

In all cases the areal mean and areal log-normal mean increases in magnitude as the density of gages decreases. There is a greater change in the non-transformed mean than in the log-normal scale parameter. Also there is a greater change for all-hour amounts, CHA and DHA, than for the rainy hour amounts, DRHA and CRHA, in the case of the non-transformed mean. For the log means, the amount of change is nearly the same for DRHA, DHA, CRHA, and CHA. Note that in the case of the non-transformed means, the change in density is greater for the convective amounts than for the total-day amounts. Why does the daily areal mean (rainy gage average) increase as the density of gages decrease? The explanation lies in the fact that the areal distributions have a positive skew with more values below the mean than above the mean. When the density of gages decreases, more of the small values are missed than large values. The net effect of this is to decrease n in the computation of the mean $(\sum x/n)$ at a greater rate than $\sum x$. Thus, the mean increases as the density decreases. This effect is apparent only when the areal means are based on rainy gage amounts.

Although σ and $\sigma_{\ln x}$ vary as the density decreases, there is a lack of trend with density. This is an expected result since the variance of a subsample should be the same as for the total sample.

The estimate of ß has a weak tendency to increase as the density decreases. This increase is explained by the same reasoning employed with the mean.

The parameter γ has a strong increase with decreasing sampling size in every case. Given that the mean increases and ß is nearly constant, the parameter γ must increase because of the relationship $\beta = \bar{x}/\gamma$.

Relationships with areal size. The average values of the distributional parameters for areal sizes of 400, 200, 100, and 50 mi^2 were computed for the 49-gage network (8.2 mi^2/gage) and the results are shown in Table 4 for DRHA and DHA.

There is a tendency for $\bar{x}$ and $\overline{\ln x}$ to increase with decreasing areal size. The trend is much weaker than it is for the relationship with density. The explanation is the same as it was for density; i.e., there are more values below the mean than above the mean. The trend is weakened, however, because local highs

Table 4. Relationship of areal size with areal distributional parameters based on rainy gage amounts.

Area (mi^2)	Mean $\overline{X}$	Log mean $\overline{\ln x}$	Standard deviation σ	Log standard deviation $\sigma_{\ln x}$	Gamma γ	Beta β	% of gages with rain
	Value of the parameter in percent*						
				DRHA			
400	95	99	111	111	62	116	67
200	101	99	106	105	82	110	72
100	103	100	99	100	95	101	75
50	101	103	83	83	161	73	78
				DHA			
400	94	94	109	114	58	125	67
200	100	92	106	105	83	109	72
100	103	106	102	99	99	101	75
50	103	108	83	82	161	65	78

* Percent is computed as the percentage of the average of all densities for a given parameter.

and lows in the rainfall pattern are encountered as the areal size is decreased and this tends to destroy the trend.

The parameters σ and $\sigma_{\ln x}$ decrease as the areal size decreases. This is an expected result because raingage amounts are more homogeneous in a small area than in a large area due to physiographic features and the correlation between gages.

Beta (ß) decreases with decreasing areal size while γ increases with decreasing areal size. The explanation of the trend for γ lies in the fact that the raingage amounts are more homogeneous in a small area than in a large area. Thus the shape of the distribution tends toward a normal distribution. When the shape of a gamma distribution approaches a normal distribution, the value of γ increases (Hahn and Shapiro, 1967). Given that $\bar{x}$ increases weakly and γ increases strongly, ß must decrease because of the relationship $\beta \approx \bar{x}/\gamma$.

Relationships between Distributional Parameters

It is quite useful to be able to specify the distribution of rainfall over an area from only the areal mean rainfall. Such a specification resolves into the problem of establishing the relationships of the daily areal mean rainfall $\bar{x}$ with the log-normal scale parameter $\overline{\ln x}$, the log-normal shape parameter ($\sigma^2_{\ln x}$), the gamma shape parameter γ and the gamma scale parameter ß. Such a relationship could be established by regressing $\overline{\ln x}$, σ^2, $\sigma^2_{\ln x}$, γ, and ß on the $\bar{x}$ values. However, such a method would involve empirical constants and the results in one area may not be applicable in another. Thus, it would be highly desirable to establish such relationships without the aid of empirical constants.

With use of the relationships which follow, it was found that the distributional parameters could be obtained from a knowledge of $\bar{x}$ and $\sigma^2_{\ln x}$. However, the relationships shown below apply only when the data are stratified into more homogeneous subsamples (i.e., storm type or synoptic type). One of the chief applications of these relationships is in the "Monte Carlo" generation of data such as was used by Schickedanz and Decker (1969). It should be noted that the relationships presented below hold only if the data are log-normal or gamma distributed.

We will assume that from a particular area, an estimate of the daily areal mean rainfall $\bar{x}$ has been obtained for each day as well as an estimate of $\sigma^2_{\ln x}$. If the data are log-normally distributed, $\sigma^2_{\ln x}$ will be independent of $\overline{\ln x}$. Thus, there will be no trend in $\sigma^2_{\ln x}$ with $\overline{\ln x}$, and the best estimate of $\sigma^2_{\ln x}$ for all values of $\overline{\ln x}$ is the average value. It follows that the estimate of $\sigma^2_{\ln x}$ will also be a constant for all values of $\bar{x}$. Figure 16b shows the relation between $\bar{x}$ (daily rainy hour average) and the corresponding $\sigma^2_{\ln x}$ for days classified as cold frontal days for the year 1964. The horizontal line represents the value of $\sigma^2_{\ln x}$ as estimated by the average value. Although there is considerable scatter to the points, there is certainly no trend in the sample.

If the variate X is log-normal distributed, the parameters $\bar{x}$ and σ^2 are given by

$$\bar{x} = \exp\left[\overline{\ln x} + \frac{\sigma^2 \ln x}{2}\right] \tag{18}$$

and

$$\sigma = \bar{x} \, [\exp (\sigma^2_{\ln x}) - 1]^{\frac{1}{2}} \tag{19}$$

Manipulation of equation 18 yields the following relationship:

$$A = \ln \bar{x} - \overline{\ln x} = \frac{\sigma^2 \ln x}{2} \tag{20}$$

The shape parameter γ of the gamma distribution is given by

$$\gamma = \frac{1 + \sqrt{1 + 4/3A}}{4A} \tag{21}$$

and the scale parameter of the gamma distribution is given by

$$\beta = \bar{x}/\gamma \tag{22}$$

Thus, from a knowledge of $\sigma^2_{\ln x}$ and $\bar{x}$ it is possible to obtain an estimate for the shape parameter γ and the location parameter ß of the gamma distribution. From equation 20 it follows that γ will be a constant over the range of $\bar{x}$ as was $\sigma^2_{\ln x}$. Figure 16a shows the relation between $\bar{x}$ (daily rainy hour average) and the corresponding values of γ for the cold frontal case. The horizontal line represents the estimate of γ obtained from equation 20 using the constant estimate of $\sigma^2_{\ln x}$, and the points are the actual sample estimates. It is apparent that there is scatter about the line, but there is no trend of γ with $\bar{x}$. Figure 17b shows the relation between $\bar{x}$ and ß for the cold frontal case. The curved line is the estimate of ß obtained from equations 20, 21, and 22, and the points are the sample estimates of ß. Although there is considerable scatter about the line, the line appears to be a reasonably good estimate of the trend in ß with $\bar{x}$. Again, it should be noted that regression techniques could have been used to obtain the relationship between $\bar{x}$ and ß but such techniques would have involved empirical constants.

Manipulation of equation 20 yields

$$\overline{\ln x} = \ln \bar{x} - A \tag{23}$$

and from this equation an estimate of the location parameter $\overline{\ln x}$ of the log-normal distribution is obtained. Figure 16c shows the relation of $\overline{\ln x}$ with $\bar{x}$ as estimated from equation 23 (curved line) and the actual sample points for the cold frontal case, 1964. Estimates of σ were obtained from equation 19, and the relationship of σ with $\bar{x}$ is shown in Figure 17a. This figure is an excellent illustration of the dependence of σ on $\bar{x}$ when the distribution of rainfall values are positively skewed.

Goodness of Fit of the Areal Frequency Distributions

In addition to determining the precision of distributional parameters, the log-normal and gamma distributions were also tested for goodness of fit. For sample sizes <40 the Kolmogorov-Smirnov goodness of fit test was applied. The chi-square test was used for sample sizes ≥40. The chi-square test was based on the method described by Hahn and Shapiro (1967) with one exception: the number of class intervals was chosen on the basis of the relation 5 $\log_{10}$ N, where N is the number in the sample. This method insures that the choice of class interval boundaries will depend on the theoretical values and not on the sample values. It also insures that, except for modification of class intervals due to rounding and measurement errors, equal numbers of expected values will result in each interval. The above rule also insures that there will be at least five expected values in each interval as long as the sample is 40 or more. This chi-square procedure makes comparisons between different distributional fits more objective.

The areal log-normal and gamma distributions for each day were tested for goodness of fit and the results are shown in Table 5. The tabled values represent the percentage of distributions which has goodness of fit probabilities ≥0.05 and ≥0.20. A goodness of fit probability of less than 0.05 implies that observed differences between the data sample and the given distribution could have occurred by random chance in less than 5% of the time. This implies that the distribution does not fit the data sample. A goodness of fit probability of ≥0.20 indicates a better fit than a probability of ≥0.05.

The log-normal distribution provides a better fit to the data than the gamma distribution does. For the all-hour averages DHA and CHA, there are almost as many gamma distributions which fit the data (65 and 58%) over the 12-year period as there are log-normal (68 and 62%). However, for the rainy hour averages DRHA and CHA, the log-normal is clearly superior.

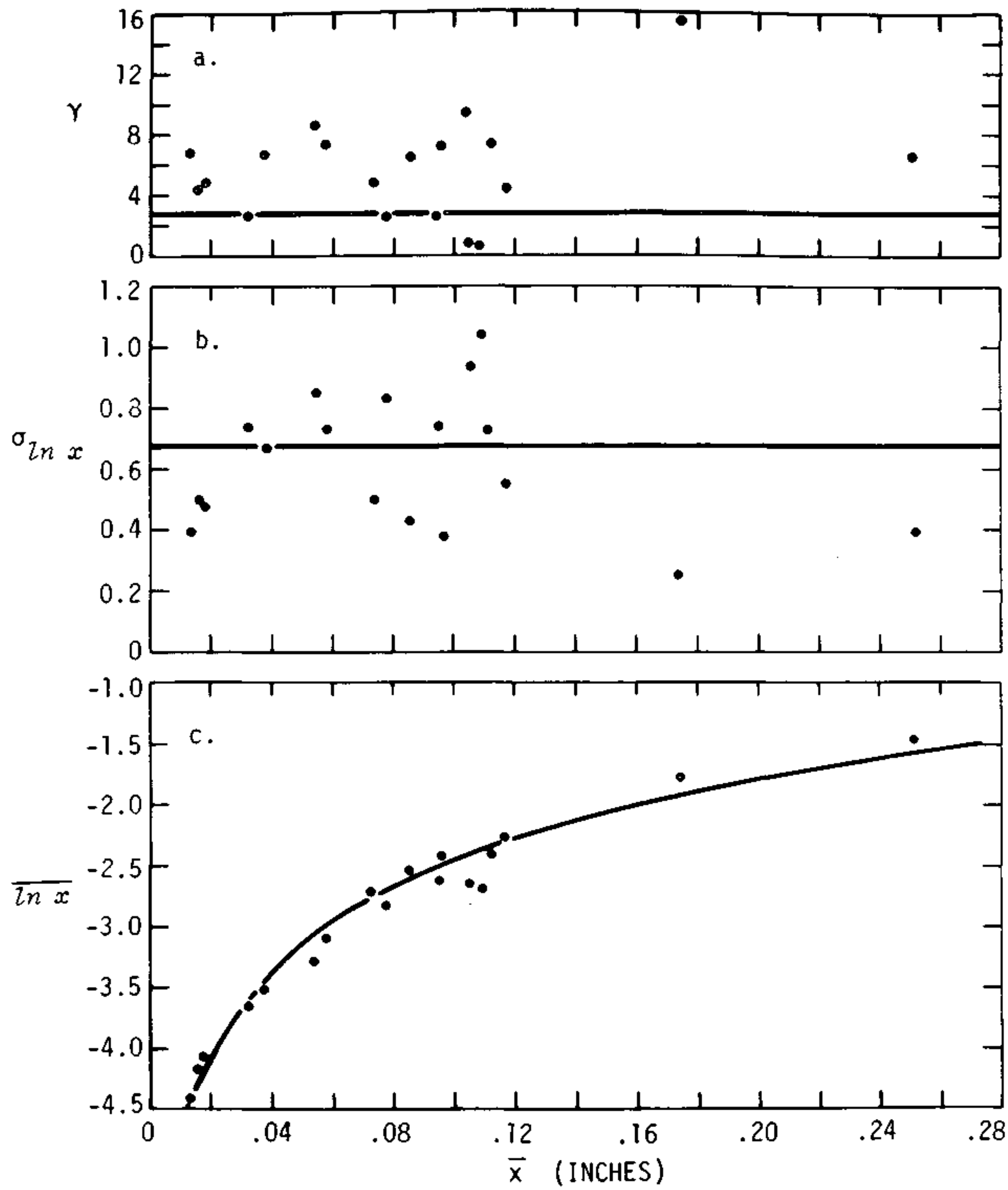

Figure 16. The relationships of the average rainfall on rainy hours ($\bar{x}$) with the gamma shape parameter (γ), the log-normal shape parameter ($\sigma_{\ln x}$), and the log-normal scale parameter ($\overline{\ln x}$).

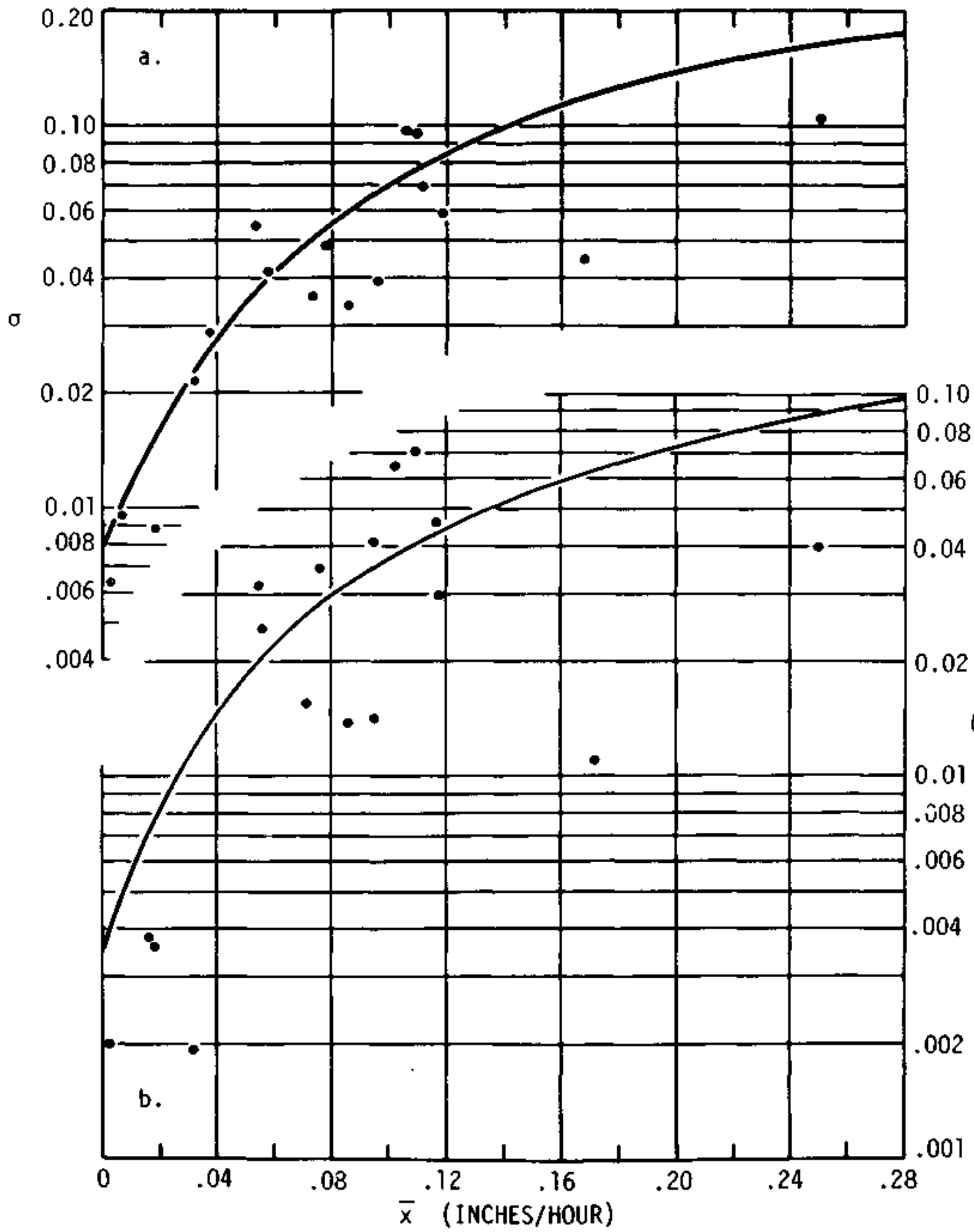

Figure 17. The relationships of the average rainfall on rainy hours ($\bar{x}$) with the standard deviation (σ), and the gamma scale parameter (β).

Table 5. Comparison of the goodness of fit probabilities* for areal distributions during the period 1956-1967 (400 mi).

Time period	Distribution	Percentage of distribution with specific probability							
		DRHA		DHA		CRHA		CHA	
		≥.20	≥.05	≥.20	≥.05	≥.20	≥.05	≥.20	≥.05
1956	log-normal	57	76	49	63	53	67	47	63
	gamma	46	66	46	66	43	57	47	63
1957	log-normal	51	70	51	68	40	48	51	63
	gamma	44	65	44	67	26	49	40	48
1958	log-normal	56	75	47	76	56	71	44	71
	gamma	43	68	46	62	37	63	37	56
1959	log-normal	42	58	36	51	37	46	41	56
	gamma	36	49	25	51	27	41	24	44
1960	log-normal	41	62	39	59	41	54	29	44
	gamma	37	60	34	54	42	50	39	46
1961	log-normal	47	65	42	62	35	45	26	39
	gamma	43	65	44	65	40	50	29	45
1962	log-normal	57	88	65	90	77	92	62	85
	gamma	51	71	61	78	69	81	62	81
1963	log-normal	59	83	51	71	41	63	41	59
	gamma	41	74	36	64	22	52	33	52
1964	log-normal	66	84	52	75	46	75	58	79
	gamma	59	70	49	79	50	71	50	88
1965	log-normal	69	82	47	67	61	89	56	78
	gamma	72	87	40	66	64	78	47	78
1966	log-normal	58	81	54	85	59	82	32	64
	gamma	41	74	43	76	41	73	41	64
1967	log-normal	39	58	42	58	36	44	42	50
	gamma	37	56	42	58	31	46	36	47
1956-67	log-normal	53	73	47	68	48	64	44	62
	gamma	45	66	42	65	40	58	39	58

* Probability that the observed difference between the data sample and the given distribution could have occurred by random chance.

Both the gamma and log-normal distributions fit the total-day amounts better than the convective amounts. For example, 73 and 68% of the log-normal distributions fit the data for the total-day amounts compared with 64 and 62% for the convective averages.

There is some variability in the goodness of fit probabilities between years. The percentage of log-normal distributions that fit ranges from 58 to 88 for DRHA, 51 to 90 for DHA, 44 to 92 for CRHA, and 39 to 85 for CHA.

Table 6 shows a comparison of the goodness of fit probabilities for areal distribution according to variation in density. As the density decreases, the percentage of distributions that fit the data increases. It is felt that this is a fictitious trend for there is no apparent reason for the fit to be

Table 6. Comparison of the goodness of fit probabilities" for areal distributions according to variation in density (400 mi^2, 1964-1967).

Density	Percentage of distribution with specific probability							
	DRHA		DHA		CRHA		CHA	
(mi^2/gage)	≥.20	≥.05	≥.20	≥.05	.20	≥.05	≥.20	≥.05
	Log-normal							
8.2	57	75	49	71	50	71	47	67
16.0	79	92	74	85	75	86	73	86
30.8	88	97	88	97	90	96	89	95
66.7	97	98	98	100	98	100	96	98
	Gamma							
8.2	52	71	44	69	47	66	43	68
16.0	75	88	71	83	74	87	72	82
30.8	85	94	83	96	89	95	82	94
66.7	94	98	98	100	98	100	98	98

* Probability that the observed difference between the data sample and the given distribution could have occurred by random chance.

better in the smaller areas. In fact, as indicated earlier, the error in the parameters tends to increase as the density is reduced. On the basis of these considerations, it is believed that the trend is strictly due to the lack of power of the goodness of fit test to reject an inappropriate model. This illustrates a problem in using the goodness of fit test on small samples. When the samples are small, one should compute the variance of the parameters in addition to choosing a higher probability level to make the decision as to whether distribution does or does not fit.

Table 7 gives a comparison of the goodness of fit probabilities for areal distributions according to variation in areal size. As the areal size decreases, the sample size also decreases, thus confounding the "apparent trend." Thus the conclusion is drawn that it is impossible to discern trends in the goodness of fit according to areal variations in areas less than 400 mi^2. It is felt that in these cases, the variance of the parameters as described earlier yields a better estimate of the trend.

Table 7. Comparison of the goodness of fit probabilities* for areal distributions according to variation in areal size (1964-1967).

	Percentage of distribution with specific probability							
	DRHA		DHA		CRHA		CHA	
Areal size (mi^2)	≥.20	≥.05	≥.20	≥.05	≥.20	≥.05	≥.20	≥.05
				Log-normal				
400	57	75	49	71	50	71	47	67
200	75	91	73	86	76	90	73	87
100	86	93	81	91	86	96	82	91
50	95	99	93	97	94	99	93	97
				Gamma				
400	52	71	44	69	47	66	43	68
200	66	82	69	83	69	82	70	82
100	78	93	73	88	80	99	77	88
50	93	97	91	97	93	97	90	96

* Probability that the observed difference between the data sample and the given distribution could have occurred by random chance.

Estimation of Temporal Frequency Distributions from Climatological Data

A treatment of the relative standard error of temporal frequency distributions of daily areal means is included on Figure 15, page 38, and in Table 2, page 37. It was found that for a density of 282 mi^2/gage the relative error was less than 20% for the yearly mean, less than 8% for yearly log-normal mean, less than 20% for γ, and less than 23% for β. This illustrates that the distributional parameters of temporal distributions can be estimated with a relatively small degree of error. The fact that the error is nearly the same for a climatological network as for a more dense network (Figure 15) indicates that the parameter for a temporal distribution at a point can also be estimated with a small degree of error. We now turn our attention to two more aspects of the problem. First, we will consider the relationship of distributional parameters with distance and, second, the goodness of fit of the distributions. To study these two aspects, the log-normal and gamma distributions were fitted to all daily amounts within a year at each gage, and for each year of the period 1956-1967.

Relationships between Temporal Distributional Parameters and Distance

For the purpose of determining the relationship of the temporal distributional parameters to distance, a line of gages was selected from northwest to northeast. The temporal distributional parameter for each gage along the line was correlated with the distance from the edge of the network. This was done to find whether there was any predictability of the distributional parameter at one gage from the distributional parameter at another gage. If a relationship exists, the magnitude and sign of the correlation should remain approximately the same from year to year. If the magnitude and sign of the correlation fluctuate from year to year, then the relation of these parameters with distance is a random quantity from year to year.

The results of the correlation study are presented in Table 8 for the period

1964—1967. In a given year the correlation is occasionally high for a given distributional parameter (-0.90 for σ in 1964, DRHA; -0.97 for ß in 1964, DRHA; etc.). However, the magnitude and sign for the correlation for a given parameter varies considerably from year to year. This implies that even though there is a relationship between distance and the parameter for a given year, the relation is entirely unpredictable from year to year indicating a random arrangement of highs or lows in the areal pattern from one year to the next.

Table 8. Relationship of the distributional parameters with distance.

	Correlation coefficient					
Year	Mean $\bar{X}$	Standard deviation σ	Log mean $\overline{\ln x}$	Log standard deviation $\sigma_{\ln x}$	Gamma γ	Beta β
			DRHA			
1964	-.64	-.90	.46	-.53	.49	-.97
1965	.11	.14	.03	.42	-.52	.28
1966	.02	-.31	-.10	-.05	.36	-.19
1967	-.05	.13	.23	.13	-.29	.08
			DHA			
1964	-.81	-.82	.53	-.52	.34	-.73
1965	.32	.34	-.14	.38	-.20	.27
1966	.31	.78	-.11	.12	.22	.10
1967	-.38	.06	.79	.64	-.68	.42
			CRHA			
1964	-.43	-.69	.28	-.35	.37	-.71
1965	-.29	-.31	.23	.26	-.12	-.03
1966	-.33	-.26	.38	.33	-.30	-.09
1967	.40	.33	-.44	.29	-.36	.42
			CHA			
1964	-.10	-.04	.14	-.16	.12	-.25
1965	-.44	-.10	.48	.31	-.21	-.06
1966	.06	.36	.44	.86	-.79	.78
1967	.01	-.21	.05	.02	.02	-.004

Goodness of Fit of the Temporal Distributions

In this section, two sets of distributions are considered. First the temporal distributions at a point for each year of the period 1956-1967 are investigated for goodness of fit. Then the temporal distribution of dailyareal means (rainy gage average) as well as daily areal means (all gage average) in the area are investigated for goodness of fit.

Goodness of fit of the distributions at a point. The results for the point distributions are listed in Table 9. The overall results for the 12-yr period reveal clearly that the temporal distributions fit the data much better than the areal distributions (Table 5). Again, it is clear that the gamma distribution provides a better fit than the log-normal distribution. Another striking feature is the large difference between log-normal and gamma distributions for the percentage of distributions having probabilities ≥0.20. For example, in case of DRHA, there are more than twice as many log-normal as gamma distributions which fit the data.

Overall, the log-normal distribution fits the temporal distribution quite well. In case of DHA, only 87% of the log-normal distributions fit the data, but for the other average, more than 92% fit the data.

There is some variability from year to year in the goodness of fit. The range of the percentage of log-normal distributions which fit the data is from 84 to 100 for DRHA, from 61 to 100 for DHA, from 73 to 100 for CRHA, and from 86 to 100 for CHA. The range in the percentage of gamma distributions which fit the data is from 37 to 86 for DRHA, 63 to 100 for DHA, from 8 to 90 for CRHA, and from 55 to 92 for CHA. Clearly, the gamma distribution has a greater variation in the number of distributions fitting from year to year.

Goodness of fit of the distributions of the daily areal means. Log-normal and gamma distributionsof daily areal means (rainy gage average) and daily areal means (all gage average) were tested for goodness of fit and the results are shown in Table 10.

In this case, the tabled values are the actual goodness of fit probabilities rather than the percentage of distributions that fit. For the daily areal means

Table 9. Comparison of the goodness of fit probabilities* for temporal distributions at a point during the period 1964-1967.

		Percentage of distribution with specific probability							
Time		DRHA		DHA		CRHA		CHA	
period	Distribution	≥.20	≥.05	≥.20	≥.05	≥.20	≥.05	≥.20	≥.05
1956	log-normal	92	98	73	90	86	98	86	96
	gamma	43	76	78	94	63	90	65	86
1957	log-normal	76	90	63	78	84	98	88	100
	gamma	49	86	73	92	45	76	63	82
1958	log-normal	73	92	57	94	78	98	92	98
	gamma	45	71	78	98	22	61	47	88
1959	log-normal	49	84	57	94	73	96	88	100
	gamma	14	37	27	67	16	53	53	82
1960	log-normal	57	86	37	61	47	73	61	90
	gamma	16	53	49	86	0	8	22	59
1961	log-normal	65	96	39	73	94	100	78	96
	gamma	47	76	80	100	51	88	78	90
1962	log-normal	88	100	49	67	84	96	98	100
	gamma	29	63	59	88	33	59	67	90
1963	log-normal	92	100	82	98	92	98	96	100
	gamma	29	61	41	63	35	61	67	90
1964	log-normal	90	96	90	100	92	98	94	100
	gamma	45	67	47	69	73	88	53	90
1965	log-normal	80	90	80	96	88	100	88	100
	gamma	59	82	41	71	51	73	51	82
1966	log-normal	86	100	82	100	96	100	98	100
	gamma	22	43	53	82	51	84	78	92
1967	log-normal	63	90	67	94	71	92	55	86
	gamma	31	69	47	71	10	57	27	55
1956-67	log-normal	80	93	65	87	82	96	85	97
	gamma	36	65	56	82	38	67	56	82

* Probability that the observed difference between the data sample and the given distribution could have occurred by random chance.

Table 10. A comparison of the goodness of fit probabilities* for temporal distributions of the daily areal means during the period 1964-1967.

Time period	Distribution	DRHA	DHA	CRHA	CHA
		Daily areal means (rainy gage average)			
1964	log-normal	.17	.07	>.20	>.20
	gamma	.27	.04	>.20	.18
1965	log-normal	.43	.50	.11	.49
	gamma	.36	.16	.29	.54
1966	log-normal	.27	.22	>.20	>.20
	gamma	.09	.02	<.01	.03
1967	log-normal	.29	<.01	.15	.03
	gamma	.04	<.01	<.01	<.01
		Daily areal means (all gage average)			
1964	log-normal	<.01	<.01	>.20	>.20
	gamma	<.01	.01	.06	.05
1965	log-normal	<.01	.75	.02	.49
	gamma	.54	.75	.64	.19
1966	log-normal	.15	.74	>.20	>.20
	gamma	.86	.37	>.20	.14
1967	log-normal	.09	.14	.40	.17
	gamma	.93	.34	.02	<.01

* Probability that the observed difference between the data samples and the given distribution could have occurred by random chance.

(rainy gage average) there are 16 out of 18 log-normal distributions which fit the data. For the gamma distributions, 8 out of 16 fit the data. Again, the log-normal distribution fits the data better than the gamma does. For daily areal means, there are 12 out of 16 log-normal distributions which fit the data. For the gamma distribution there are also 12 out of 16 distributions which fit the data. Thus, the log-normal distributions fit the data better than the gamma for daily areal means (rainy gage average), while the two distributions fit the data equally well for daily areal means (all gage averages).

RESUME

This research had two specific purposes. The first was to present a method of estimating the required density of a proposed raingage network in order to insure that various rainfall parameters will be measured with a desired level of precision. The second purpose was to determine the reliability of measurements made in an area of limited data coverage compared with the reliability that would be available from a denser network. The determination of this reliability enables the user to decide which measurements would or would not be improved if a more dense network were available. The techniques and methods presented in the report can be used to fulfill both purposes. In other words, the techniques will aid in the planning of the density of a network to insure that desired accuracy will be attained in the measurements; or, faced with a sparse network, the techniques yield information on the reliability of the measurements and the quality of the information obtained. In the paragraphs that follow, a summary is made of the techniques, the analyses, and the implication of the methods.

Basic Data

This report deals with the following forms of data estimates: 1) daily areal means (all gage average), 2) daily areal means (rainy gage average), 3) yearly areal means, 4) the period areal means, 5) point estimation in a daily areal pattern, 6) point estimation in a yearly areal pattern, 7) point estimation in a period areal pattern, 8) log-normal and gamma distributional parameters for areal distributions, 9) log-normal and gamma distributions of the daily areal means (rainy gage average), 10) log-normal and gamma distributions of the daily areal means (all gage average), and 11) the log-normal and gamma temporal distributions at a point. Terms are defined in the Appendix.

The basic data from which these estimates were derived were based on convective and total-day time periods, and on hours with rain as well as all hours of the day (DRHA, DHA, CRHA, and CHA).

Approach

The relative standard error for each of the data estimates was determined for data from a dense network in East Central Illinois (see Figure 4). The relative standard error for each of these parameters was then determined from a subset of gages which had a density corresponding to that of a climatological network. The error was then determined for various raingage densities and for various areal sizes. The errors as determined from the dense network and the climatological network were then compared to determine the accuracy of the data measurements from a climatological network.

In addition to these error determinations, the relationship of the distributional parameters with distance and density, relationships between distributional parameters, and the goodness of fit of areal and temporal distributions were investigated.

The equations and techniques necessary to perform the above investigation are described throughout the report.

Selected Results

Equations and methods for the determination of error in the daily, yearly, and period areal means (all gage average) are given on pages 4 - 6. The climatological estimates of the relative standard error of daily areal means were found to be nearly the same as dense network estimates. For example, the RSE from the climatological network was 14% for a density of 8.2 mi^2/gage and 86% for a density of 282 mi^2/gage for DRHA. Comparable estimates for a dense network were 12 and 80% respectively.

For the yearly and period areal means the error was less, but the differences between estimates from the climatological and dense networks were greater. The greater differences were caused by the additional sources of variance contributing to the error. However, the climatological network was found to be adequate for the estimation of error in the daily, yearly, and period areal means (all gage average).

Equations and techniques for the determination of error in the daily, yearly, and period areal patterns are described on pages 13-14. The error of point estimation in the daily rainfall pattern was found to be very large unless the network is very dense. For example, with dense network estimates, the error was 155% at a density of 282 mi^2/gage, 85% at a density of 8.2 mi^2/gage, and 55% for a density of 1 mi^2/gage for DRHA. With estimates from the climatological network, the trend is comparable with values of 140, 70, and 35% respectively. When the pattern is measured over a period of a year or longer, the error is reduced considerably.

Techniques for estimating the error in larger areas are discussed on pages 27-29. The results indicated that the best method is to estimate the error in the larger climatological network at the available density and then estimate the error for the smaller densities. This approach appears to be superior to estimating the error in the dense network (small area) at the largest density and then extrapolating to a large area. This conclusion is based on the fact that the prediction of various densities from the climatological network was found to be adequate. The results also indicated that the standard error decreases as the areal size increases.

The estimation of the error at a point in the areal pattern for larger areas is relatively simple. The technique is never based on a gage separation of more than that of a climatological network. Thus, once the error is computed for the subareas of the larger areas, the estimate can be combined to obtain the estimate for the total area.

In general, it was found that the climatological network could be used to estimate the error in areal distributional parameters. However, there were large differences between the estimates of the climatological and dense networks in the convective rainfall for the gamma shape parameter.

Although the error of the areal gamma distributional parameters could be estimated from a climatological network in most cases, the magnitude of the error for the climatological density is so large that the estimates are virtually useless.

The areal log-normal scale parameter could be estimated from the climatological network, and the error was small enough that the parameter was of value. For example, the error was 20% or less for the various raingage amounts DRHA, DHA, CRHA, and CHA.

For the distributions of daily areal means, the relative error of the distributional parameters is practically the same for all densities. This is in direct contrast to the areal distributional parameters, which have a sharp trend with density. However, similar results were found for the relative error of the yearly areal and period areal means (see page 10). The distributional parameters for the daily areal means had an error of less than 25% if the sampling period during the year had 50 rain days.

An investigation of the relationships of the distributional parameters with distance revealed that the log-normal scale parameter increases and the shape parameter remains relatively constant as the density of raingages decreases. The log-normal scale parameter was found to increase more for convective amounts than for total-day amounts. The gamma shape and scale parameters increase as the density of raingages decreases.

In regard to areal size, the log-normal scale and shape parameter decrease as the areal size decreases. The gamma scale parameter decreases with decreasing areal size, while the gamma shape parameter increases with decreasing areal size.

In regard to relationships between parameters, it was found that the other distributional parameters could be obtained from a knowledge of $\overline{\ln x}$ and $\sigma^2_{\ln x}$. However, the relationships between distributional parameters were adequate only when the data were stratified into more homogeneous subsamples.

The log-normal and gamma distributions were also tested for goodness of fit. The areal distributions, temporal distributions on individual gages, distributions of daily areal means (rainy gage averages), and distributions of daily areal means (all gage averages) were tested for goodness of fit. The goodness of fit tests indicated that the log-normal distributions fit the data samples used in this study better than the gamma distribution. The temporal distributions were found to fit the data better than the areal distributions. For the log-normal distributions, the percentage of temporal distributions fitting the data were

93, 87, 96, and 97; this compares with 73, 68, 64, and 62 for the areal distributions for the various raingage amounts DRHA, DHA, CRHA, and CHA. For the gamma distributions, the percentage of temporal distributions fitting the data was 65, 82, 67, and 82; this compares with 66, 65, 58, and 5 8 for the areal distributions for the various raingage amounts DRHA, CHA, CRHA, and CHA.

For the temporal distributions of daily areal means (rainy gage average) there were 16 out of 18 log-normal distributions and 8 out of 16 gamma distributions which fitted the data for the period 1964-1967. For the temporal distributions of daily areal means (all gage averages) there were 12 out of 16 log-normal distributions and 12 out of 16 gamma distributions which fitted the data.

Thus, it is illustrated clearly that the temporal distributions fit the data better than the areal distributions which agrees with the results found for the relative error of the parameters.

Implications

Implications to weather modification. The results of this research have implications in the area of weather modification. Although much research effort has been directed to the problem of planning and verification of weather modification experiments according to experimental design, storm types, weather types, and duration of the experiment (Neyman and Scott, 1967; Schickedanz and Changnon, 1970; Schickedanz and Huff, 1971), little attention has been directed to the problem of density of the raingage network in relation to duration of the experiment. An investigation of the sampling error according to density in areal mean storm rainfall was made by Huff and Schickedanz (1970). However, the important contribution of the present research is to show that the estimate can be made for several rainfall units from the existing climatological network of raingages.

As an example, the researcher must often make a decision as to whether the network should be made more dense in order to measure the individual points in the daily rainfall pattern with a higher degree of precision. If higher precision is desired, the network would need to be very dense. Because of the cost of installing and maintaining a very dense network, the dense network may of necessity be limited to a small area. In the small area, other expensive instrumentation would likely be installed and would eventually be correlated with radar data.

However, the rainfall over a larger area would also be of interest. In the larger area, the interest might be limited to differences between areal means on seeded and non-seeded days. In this case the additional gages would contribute very little to the end result and a greater reduction in experimental error might be obtained by increasing the duration. With the techniques presented in this report the network could be designed to serve both purposes. Also, through the use of these techniques information could be gleaned from the climatological network which would help estimate the cost of the experiment.

Implications for climatological applications. For many climatological studies, the only data which are available will be that obtained from the Weather Bureau networks which have a density in the midwest of approximately 250 mi^2/gage. Often the smallest unit of measurement will be 24 hour amounts. In this case, one needs to know the reliability of the various estimates that can be made. Had there been more sampling points in the area, would the results have been different? The techniques and figures presented in this report allow one to make the decision in an orderly and systematic way. By the use of these techniques, figures and tables similar to those presented here can be constructed for any area.

STUDENT PERSONNEL AND SCIENTIFIC PAPERS

Several undergraduate students from the University of Illinois were employed during the course of this investigation. The students were Susan Lewandowski, Cecelia Semonin, Patricia Coughlin, Barbara Binch, and Nancy Thun.

A paper summarizing initial results from the frequency distributions was presented at the International Symposium on Probability and Statistics in the

Atmospheric Sciences at Honolulu, Hawaii, in June 1971. In addition, a paper which summarizes the pertinent results from this project will be submitted to the Journal of Applied Meteorology.

REFERENCES

Changnon, S. A., 1967: Hail evaluation techniques project. Annual Report, NSF Contract GA-482, Illinois State Water Survey, Urbana, 11 pp.

Changnon, S. A., 1969: Hail evaluation techniques. Part I, Final Report, NSF GA-482, Illinois State Water Survey, Urbana, 97 pp.

Davis, H. T., 1933: Tables of higher mathematical functions. Vol. I and II, Bloomington, Indiana, Principal Press.

Decker, W. L., and P. T. Schickedanz, 1967: "The evaluation of rainfall records from a five year cloud seeding experiment in Missouri." Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, California, December 27-31, 1965, pp. 55-63.

Hahn, G. J., and S. S. Shapiro, 1967: Statistical models in engineering. New York, John Wiley and Sons, Inc., 355 pp.

Hershfield, D. M., 1971: Agricultural research service precipitation facilities and related studies. ARS 41-176, Agricultural Research Service, United States Department of Agricultural, 117 pp.

Huff, F. A., 1966: "The adjustment of radar estimates of storm mean rainfall with raingage data." Proceedings of the Twelfth Conference on Radar Meteorology, Norman, Oklahoma, October 17-20, 1966, pp. 19 8-203.

Huff, F. A., 1970: Rainfall evaluation studies. Final Report, Part I, NSF Grant GA-1360, Illinois State Water Survey, Urbana, 53 pp.

Huff, F. A., and J. C. Neill, 1957: Rainfall relations on small areas in Illinois. Bulletin 44, Illinois State Water Survey, Urbana, 224 pp.

Huff, F. A., and P. T. Schickedanz, 1970: Rainfall evaluation studies. Final Report. Part II, NSF Grant GA-1360, Illinois State Water Survey, Urbana, 224 pp.

Huff, F. A., W. L. Shipp, and P. T. Schickedanz, 1969: Evaluation of precipitation modification experiments from precipitation rate measurements. Final Report, INT 14-06-D-6575, U.S. Dept. of Interior, Bureau of Reclamation, Office of Atmospheric Resources, 122 pp.

Huff, F. A., J. C. Neill, and M. Spock, Jr., 1956: Evaluation of a low power 3-cm radar for quantitative rainfall measurement. Report of Investigation 29, Illinois State Water Survey, Urbana.

Jones, D. M. A., 1966: "The correlation of raingage-network and radar detected rainfall." Proceedings of the Twelfth Conference on Radar Meteorology, Norman, Oklahoma, October 17-20, 1966, pp. 204-207.

Neyman, J., and E. L. Scott, 1967: "Some outstanding problems relating to rain modification." Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, California, Dec. 27-31, 1965, pp. 293-350.

Ostle, Bernard, 1963: Statistics in research. Ames, Iowa, The Iowa State University Press, 585 pp.

Schickedanz, P. T., and S. A. Changnon, 1970: "The design and evaluation of hail suppression experiments." Monthly Weather Review, Vol 98, No. 3, March 1970, pp. 242-252.

Schickedanz, P. T., S. A. Changnon, and C. G. Lonnquist, 1969: A statistical methodology for the planning and evaluation of hail suppression experiments in Illinois. Part II, Final Report, NSF GA-482, Illinois State Water Survey, Urbana, 140 pp.

Schickedanz, P. T., and W. L. Decker, 1969: "A Monte Carlo technique for designing cloud seeding experiments." Journal of Applied Meteorology, Vol. 8, No. 2, April 1969, pp. 220-228.

Schickedanz, P. T., and F. A. Huff, 1971: "The design and evaluation rainfall modification experiments." Journal of Applied Meteorology, Vol. 10, No. 3, June 1971, pp. 502-514.

Stout, G. E., and E. A. Mueller, 1968: "Survey of relationships between rainfall rate and radar reflectivity in the measurement of precipitation." Journal of Applied Meteorology, Vol. 7, No. 3, June 1968, pp. 465-474.

Steele, R. G. D., and J. H. Torrie, 1960: Principles and procedures of statistics.

First Edition, New York, McGraw-Hill Book Company, Inc., 481 pp.

Thorn, H. C. S., 1958: "A note on the gamma distribution." Monthly Weather Review. Vol. 86, No. 4, April 1958, 117-122.

Wilk, K. E., 1961: Radar investigations of Illinois hailstorms. Scientific Report 1, AF 19(604)-4940, Illinois State Water Survey, Urbana, 42 pp.

A P P E N D I X

APPENDIX

The purpose of this appendix is to describe in one place the basic data and some basic data definitions that are used throughout the report.

Source of Data

The East-Central Network (ECI) of the Illinois State Water Survey supplied the basic data used in this study (see Figure 4). It consisted of 49 raingages arranged in a nearly uniform grid pattern in a 400-mi^2 rural area of relatively flat terrain in which elevations ranged from 650 to 910 ft MSL. The network was operated from 1956 to 1967 with no significant changes in gage locations.

Basic Data

In some cases, data from the entire 12-yr period (1956-1967) were used while in other cases only data from the 4-yr period (1964-1967) were used. A 4-yr data sample was used whenever it was deemed that the analysis of a 12-yr data sample would not contribute enough additional information to warrant the effort, or when a particular phase of the analysis involved lengthy and costly analyses which could not be justified for a longer period from the amount of information gained.

Early in the analyses, it was decided to work with four basic units of data. These were chosen to approximate measurements often used in weather modification experiments. These units were chosen to distinguish between measurements based only on hours with rain and between measurements made for all hours. The units were also chosen to distinguish between rainfall measured during the convective period and rainfall measured for the entire day. For purposes of analysis, it was found convenient to divide the total gage amounts by the number of hours so that the values would be more readily comparable. Thus, the four basic data units used were: 1) Daily Rainy Hour Average (DRHA) which is the average of the gage rainfall amounts on hours with rain during the 24–hr period, from midnight to midnight; 2) Daily Hourly Average (DHA) which is the average of gage rainfall amounts over all hours during the period from midnight to midnight; 3) Convective Rainy Hour Average (CRHA) which is the average of gage rainfall amounts on hours with rain during the period 1100-1900; and 4) Convective Hourly Average (CHA) which is the average of gage rainfall amounts over all hours during the period 1100-1900.

For purposes of clarity in the writing and reading of the report, the symbols DRHA, DHA, CRHA, and CHA are often referred to as raingage amounts rather than averages, for they in fact represent normalized amounts.

Definitions

The data were often summarized according to various averages and data units which are repeated many times throughout the text. The following is a list of definitions frequently used.

Convective amounts. CRHA and CHA, or the amounts based on convective hours only.

Total-day amounts. DRHA and DHA, or the amounts based on the total 24- hour period of the day.

Rainy-hour amounts. DRHA and CRHA, or the amounts based on only the hours with rain.

All-hour amounts. DHA and CHA, or the amounts based on all hours.

Warm season. The months of May through September. Only data from this period were used in the analyses.

Daily areal mean (all gage average). A mean computed by dividing the summation of the averages DRHA, CHA, CRHA, and CHA over the area of interest by the total number of gages in the area.

Daily areal mean (rainy gage average). A mean computed by dividing the summation of the averages DRHA, DHA, CRHA, and CHA over the area of interest by the number of gages with rain.

<u>Yearly areal mean</u>. A mean computed as the average of all the non-zero areal means over the year (warm season). Sometimes it is referred to as the average of the daily areal means over a period of a year.

<u>Period areal mean</u>. A mean computed as the average of all the yearly areal means over several years. It is sometimes referred to as the average of the daily areal means over a period of several years.

<u>Areal Distribution</u>. The distribution of non-zero gage amounts (DRHA, DHA, CRHA, and CHA) over the area on a particular day. Thus, there is an areal distribution for every day with rain.

<u>Temporal distribution of individual gages</u>. The distribution of the gage amounts (DRHA, DHA, CRHA, and CHA) over a period of time at a particular gage. Thus, there is temporal distribution for every gage.

<u>Temporal distribution of the daily areal means (rainy gage average)</u>. The distribution of the daily areal means (rainy gage averages) over a period of time.

<u>Temporal distribution of the daily areal means (all gage average)</u>. The distribution of the daily areal means (all gage average) over a period of time.