Effects of Inhomogeneity and Record Length on Estimates of Correlation and Variability of Precipitation Data

JAMES E. McDONALD AND CHRISTINE R. GREEN

Institute of Atmospheric Physics
The University of Arizona
Tucson, Arizona

Abstract. Empirical estimates of the extent to which internal inhomogeneity and variable length of record can influence the coefficient of correlation r and the coefficient of variation C are summarized. The basic data used comprise 50-year records of winter and summer half-year precipitation totals for 50 stations well distributed throughout the United States. On the basis of an examination of double-mass plots for 37 stations, it is concluded that a 20 per cent increase imposed on the latter half of each test station's record would constitute a comparatively large degree of synthetic inhomogeneity for test purposes. The resulting effects on r and C are regarded as not being serious for most purposes. About 85 per cent of all r values computed from such synthetically inhomogeneous records do not depart by more than one standard error from the corresponding unadjusted r values, and 72 per cent of the synthetic C values lie within one standard error of the corresponding unadjusted C values. By computing r and C for record lengths varying from 10 to 50 years, the way in which r and C converge to their 50-year values are examined. It is concluded that at least 30-year samples are desirable in correlation and variability studies, a result in agreement with previous estimates of record lengths required to obtain stable estimates of mean precipitation amounts. Record lengths of 40 years are found to be essentially as good as 50-year series in specifying r or C.

Introduction. In many hydrometeorological and climatological studies, questions concerning internal homogeneity and record length arise in the selection and treatment of precipitation data. Since the investigator must almost invariably make some kind of compromise in his choice and handling of such data, he needs reference data on likely magnitudes of error in making his compromises. Theory alone cannot provide more than a few not very useful hints. Hence, *empirical* studies are needed to establish the magnitude of errors that may be expected to result from the ever-present shortcomings of actual data. In this paper empirical investigations carried out with reference to problems of estimating degree of *correlation* and *variability* of precipitation data are summarized.

Effects of inhomogeneity. The typical source of inhomogeneity within a given published precipitation record is a gage-location shift great enough to alter the effective catch of the gage but not great enough to lead to an official change of station name. When a gage is near buildings, trees, or other local obstructions to airflow, as is frequently the case, a shift of a few tens of feet can alter the effective gage exposure in such a manner as to systematically raise or lower the average catch by a discernible amount. In other instances, a gage may be shifted from, say, one side of a city or town to another in a region of irregular terrain, thereby moving the gage across a real isohyetal gradient and rendering the record for that station inhomogeneous. The U. S. Weather Bureau has recently published an excellent series of compilations of individual station histories that are very helpful in anticipating the more drastic instances of the effects of such changes, but all too often research requirements plus paucity of long-term records limit the investigator's freedom to choose records wholly free from such obvious internal discontinuities. It is important to know, therefore, how serious an effect on various statistics can be produced by these breaks.

The standard method for discerning inhomogeneity in precipitation records is that of double-mass analysis, originally developed by Merriam and subsequently discussed by many writers [cf. *Linsley, Kohler, and Paulhus,* 1959]. To explain briefly the nature of double-mass analysis we may suppose we wish to test station X's homogeneity with respect to annual precipitation amounts for the period 1906–1955, the period used here. We first compute the cumulative sum of X's annual totals, proceeding, say, forward in time from 1906 toward 1955. Then, selecting about 10 nearby stations (distance preferably under 100 miles from X), we compute a similar cumulative sum of the *combined* annual totals for all 10 of these comparison stations. Then, plotting X's cumulative sums against corresponding 10-station-total sums, year by year, we obtain a set of 50 points constituting the 'double-mass plot' for X. By laying a straightedge along this set of points, we see whether they do or do not form an essentially straight line. If they exhibit distinct breaks between subsets which themselves constitute straight arrays of points, we have evidence of inhomogeneity in X's record, the seriousness of which is proportional to the angularity of the breaks. The double-mass technique enables one not only to detect such breaks but also to apply objectively an adjustment to render the record homogeneous. However, the practical question of just what constitutes a serious break is not definitely answered in the literature.

In the course of an extensive study of seasonal precipitation correlation patterns for the entire United States, the present writers used double-mass methods to select base stations for correlation with large numbers of secondary stations. A qualitative finding of general hydrologic interest was the following: The fraction of records subjected to double-mass analysis for which there appeared breaks as pronounced as those usually exhibited in textbook and other references on the double-mass technique was *surprisingly small*. Out of 37 stations, well distributed over the United States, whose records were double-mass tested, the one showing the most pronounced break was the published record for Big Springs, Texas; yet this station's break is not large compared with the kind usually exhibited in the literature. In Figure 1 its double-mass curve is displayed as a specimen of what might be regarded as a comparatively large break, on the basis of this sample of 37 cases. The break occurred in 1940, in which year the responsibility for the rain gage at Big Springs changed from the Civil Aeronautics Administration to the U. S. Weather Bureau; presumably, the raingage was subjected to some change of exposure that produced a systematic decrease of catch. From the double-mass plot, the magnitude of the decrease can be shown to be about 25 per cent, so, again extrapolating from our sample of only 37 cases, we might say that most records will contain discontinuities that are not much greater than this percentage. A more pertinent observation is that a crude estimate of the modal value of double-mass breaks in this sample would be about 10 per cent. That the preceding conclusions drawn from a sample of 37 cases will scarcely cover all possibilities, especially in mountainous areas of strong isohyetal gradients, will be obvious; but even the above rough statements are believed to be of some use to the investigator who encounters warnings about the indispensability of homogeneity tests but who finds no indication in the standard published data of the order of magnitude of the breaks he must expect to find.

A second unexpected finding of the double-mass analyses concerned the degree of internal homogeneity in precipitation records from large cities in which the records are taken from gages at Weather Bureau city offices where complex series of exposure changes have often occurred during the growth of the city. In the search for homogeneous base stations, it was initially felt that observations taken in rural areas or in very small towns should be uniformly better than records for large cities; but experience did not confirm this expectation. City records as a whole seemed to contain no greater degree of inhomogeneity than did records for small towns and rural areas.

Presumably the latter finding reflects the fact that if any structure, large or small, is built near a rain gage some nonhomogeneity is introduced into the record. The former finding suggests that, by and large, all such effects may be of somewhat smaller magnitude than has been implied in past warnings concerning the homogeneity problem. However, it remains true that, when in doubt, one should apply the double-mass test, particularly when the final inferences must rest upon only a small number of individual records.

More useful to the investigator than the preceding merely qualitative remarks are quantitative measures of how large an effect inhomogeneity of commonly occurring magnitudes can produce in those statistics frequently used in climatological and hydrologic studies. Two such statistics of basic importance are the correlation coefficient r and the coefficient of variation C. In the remainder of this section results of an empirical test of effect on r and C of synthetically produced inhomogeneity will be summarized.

Clearly, maximum distortion of any statistic will result from an inhomogeneity that splits the full record into two equal parts, since if any break occurs either very early or very late in a given record, we need not speak of that record as being seriously inhomogeneous. (The question whether a given gage provides a *representative* sample of natural precipitation is also closely related to exposure changes, but this question is, of course, distinct from that concerning homogeneity.) Hence, in the present study, a series of actual records were subjected to a synthetic break at their midpoints, the magnitude of the break being chosen, on the basis of the experience cited earlier, as a 20 per cent increase in every seasonal total during the *latter half* of the period of record. A sample of original and synthetic double-mass plots for the case of Albany, New York, is shown in Figure 1. All records were for the 50-year period 1906–1955; therefore every synthetic record had a break at 1930 of the type illustrated for Albany.

It should be emphasized that the reason one cannot precisely predict on any theoretical basis the quantitative effects of a specified synthetic break on statistics such as r of C (or even on the record mean itself) is that climatic fluctuations shuffle high and low rainfall years in nonrandom but still unpredictable fashion, so a discontinuity imposing, say, a uniform 20 per cent increase after the record midpoint may be increasing either the predominantly larger or the predominantly smaller readings of the full record, or it may be influencing a nearly equal number of high and low readings, depending on the exact nature of secular trends and fluctuations in the region concerned. For this reason there appears to be real value in the type of *empirical* test used here. The test affords a measure of the consequence of breaks operating under *actually occurring* conditions of precipitation fluctuation.

A total of 50 station records were analyzed; but, because each record was separated into a 'winter' half (defined as the half-year from November through the following April) and a 'summer' half, and because precipitation characteristics for summer and winter half-years are governed by quite dissimilar circulation features, the study may properly be regarded as applying to 100 series of 50-year records of half-year precipitation totals. The results exhibited no systematic differences between the synthetic inhomogeneity effects on the winter versus the summer data, as a whole, and support the viewpoint that these two seasonal series may be treated as being essentially independent.

All calculations were done on an IBM 650 computer, the program involving parallel computations on the original data and on the synthetically adjusted data in order to permit evaluation of the effects of the artificial 20 per cent break at the midpoints. The 50 stations were chosen as follows. First, 10 *base* stations were selected from among the 37 stations whose double-mass plots had been previously prepared. Only stations free from breaks were used and, from these, 10 were chosen to secure a good geographic spread over the country, though more lay in western than in eastern

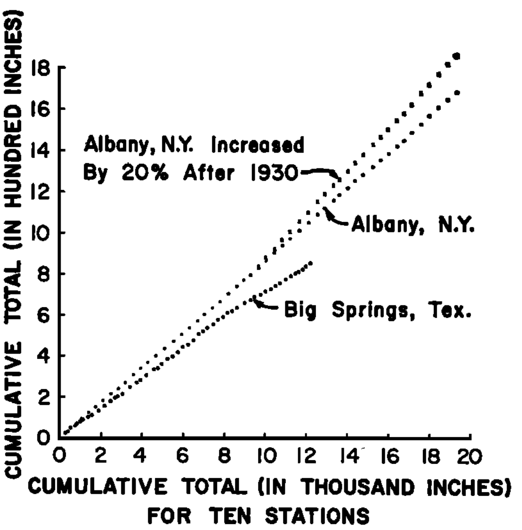

Fig. 1. Sample double-mass plots (annual data).

United States. Then, for each base station, 5 nearby stations were chosen so that distances between stations were generally less than 300 miles. The station locations are shown in Figure 2. For each base station, the 5 winter and 5 summer correlation coefficients were computed, yielding 100 correlations. The base station data were not subjected to the synthetic break, so there was no opportunity for similarity of displacements in the base station's and the secondary stations' data to suppress the distorting effect produced by the synthetic break. The procedure used is, in this respect, comparable to the problem of correlating actual data when one can never assume that compensatory breaks occur simultaneously at two or more independent stations within a region.

A scatter diagram is shown (Fig. 3) of the values of the false or synthetic correlation coefficients r_f plotted against the true value r_t. Use of the simple standard error of r as a yardstick for assessing the seriousness of inhomogeneity effects seems to be entirely adequate for present purposes, so in Figure 3 dashed lines have been entered for reference, one standard error s_r above and below the $r_f = r_t$ diagonal. Most of the points lie below the diagonal, revealing a tendency for inhomogeneity to *lower* the degree of correlation, as should be expected. However, the clustering of most of the points close to the diagonal and generally well inside the dashed lines shows how slight is the effect of even a 20 per cent break at the midpoint if a precipitation record is used in correlation analysis.

The corresponding plot for values of the coefficient of variation is shown in Figure 4, C_f being the false or synthetic value and C_t being the true value. The quantity C_f was, of course, computed only for the secondary station records, not for the base stations, since no synthetic break was imposed on any of the base station records that were used in calculating correlations. The fact that most of the points in

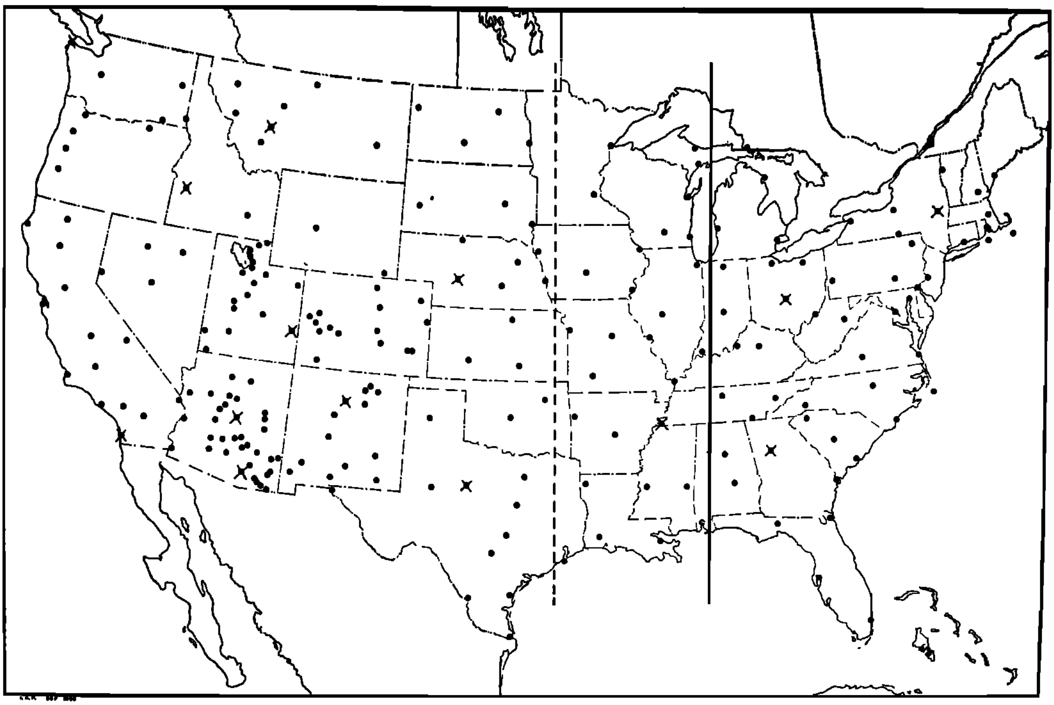

Fig. 2. Map of precipitation stations. Crosses mark stations used as correlation base stations in parent study. Ten used here are Tucson, San Diego, Santa Fe, Helena, North Platte, Abilene, Columbus, Atlanta, Washington, and Albany. Secondary stations for western base stations are all those shown west of the solid vertical line; those for eastern base stations are east of the dashed vertical line.

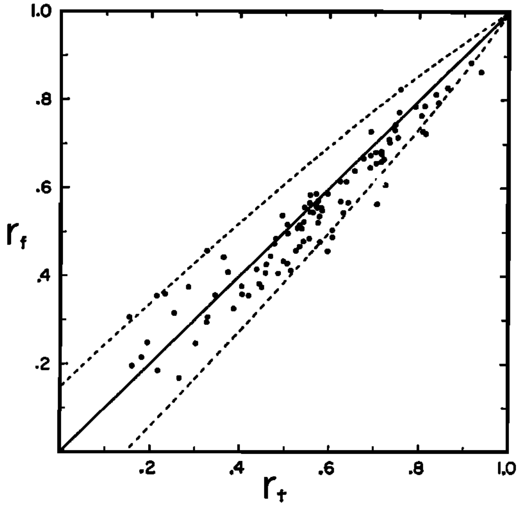

Fig. 3. Scatter diagram of r_f against r_t. The dashed lines lie one standard error of r_t above and below the solid-line locus $r_f = r_t$.

Figure 4 lie slightly above the diagonal indicates that inhomogeneity tends in general to *increase* apparent variability, an understandable effect. The dashed curves lie one standard error s_c above and below the diagonal $C_f = C_t$. One notes that the large values of C_t are associated with C_f values exhibiting smaller increases due to inhomogeneity than do the smaller C_t values. It is believed that this reflects the climatic fluctuations of the Southwest, which contributes most of the high C_t values. Generally declining precipitation in the latter half of the 50-year period is characteristic of the Southwest, so application of the synthetic 20 per cent increases probably tended to cancel some of the total variance for these records.

The computer was programed to calculate quantities $Q_r = (r_f - r_t)s_r$ and $Q_c = (C_f - C_t)/s_c$ as very simple measures of the significance of the inhomogeneity effects. Statistical purism might lead to objections about such simple use of s_r and s_c as yardsticks for assessing the seriousness of inhomogeneity effects, but the writers feel that greater refinements are not in order here. The present objectives of displaying empirical indicators of effects of inhomogeneity seem to be quite sufficiently served by use of Q_r and Q_c as defined. Frequency-distribution data for Q_r and Q_c are given in Table 1. Eighty-five per cent of the values of r_f lie within one standard error of their corresponding r_t values, and 72 per cent of the values of C_f lie within one standard error of C_t. (The data from which Table 1 is drawn showed only 2 and 5 per cent of all cases for which the absolute magnitudes of Q_r and Q_c, respectively, exceed 2, roughly the 0.05 level of significance for the departure from a zero difference.) These Q values confirm what is graphically evident in Figures 2 and 3: a 20 per cent midpoint break in a precipitation record does not impose extremely serious distortion on either the coefficient of correlation or the coefficient of variation, particularly for those applications in which r and C are employed only as descriptive statistics. If either of these statistics is used in some *predictive* capacity, the investigator must decide in each individual context whether inhomogeneity is likely to introduce serious bias. It is hoped that, in making such decisions, the above data will prove useful as empirical indicators of the magnitude of errors that might actually be encountered.

Effects of record length. The problem of selecting record lengths adequate for specifying climatological 'normals' has received attention by a number of workers. *Landsberg and Jacobs* [1951] presented a table of record lengths needed to obtain a stable frequency distribution for five different climatic elements and four

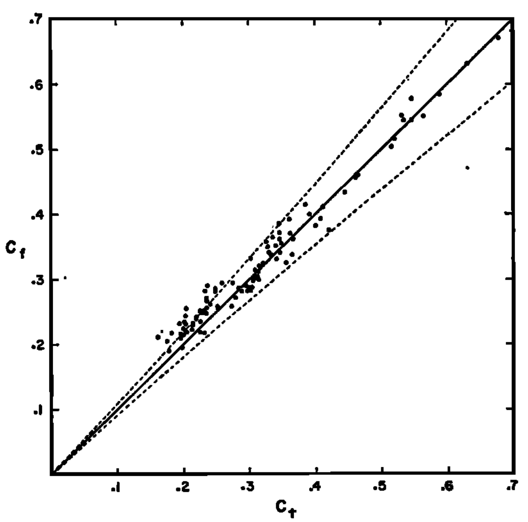

Fig. 4. Scatter diagram of C_f against C_t. The dashed lines lie one standard error of C_t above and below the solid-line locus $C_f = C_t$.

TABLE 1. Percentage Frequencies of Q_r and Q_c

	Q_r or Q_c					
	<−1.1	−0.6 to −1.0	−0.1 to −0.5	0.0 to 0.4	0.5 to 0.9	>0.1
Per cent Q_r	12	23	40	17	5	3
Per cent Q_c	0	5	28	23	16	28

types of climatic regions. For precipitation amounts, record lengths of up to 50 years are suggested (for the least favorable case of mountainous regions), though these writers did not precisely define their meaning of 'stable distribution.' *Lenhard and Baum* [1954] discussed the same problem with respect to temperature normals, and *Beaumont* [1957] and *Enger* [1959] recently treated the related question of optimum record length for predicting precipitation, runoff, or temperature values for 1 year in the future, and found that records extending back only 15 to 20 years prove generally most effective in this 1-year-hence prediction problem.

Here we will examine the rate at which sample estimates of r and C for precipitation data converge to their 50-year values as we increase the number of years n of record length used in computing the estimates. Lengths of $n = 10$, 20, 30, and 40 years were used, the record length being extended backward in time in jumps of 1 decade beginning with the 1955 date that terminated all records of precipitation used in the present study. The computing program was arranged to yield ratios, Q_r and Q_c, of differences between the n-year minus the 50-year value of r and C of the standard error of $\hat{r}$ and C, respectively, very much as was done above in connection with the inhomogeneity data. The same set of data used in the inhomogeneity analysis was employed here, again yielding 100 essentially independent time series for analysis.

Table 2 presents the results in terms of the percentage frequency distributions of values of Q_r and Q_c for the four n-values examined. It will be seen that 10-year records are entirely unsatisfactory in estimating the 50-year values of r and C, 71 per cent of all the 100 estimates being more than a standard error away from the 50-year value in the case of both r and C for $n = 10$. Also 20-year records are unacceptable because 39 per cent and 48 per cent of the values of Q_r and Q_c, respectively, lie more than a standard error away for $n = 20$. One might say that 30-year samples are marginally acceptable, since 39 per cent of all r estimates and 26 per cent of all C estimates are then beyond a standard error away from their corresponding 50-year values. There are no Q_c values exceeding 2 in absolute magnitude for $n = 30$, and only 5 per cent of all Q_r values for $n = 30$ exceed 2 in absolute magnitude. The 40-year values prove to be entirely adequate estimators of 50-year r and C values, only 8 per cent of all cases missing by as much as one standard error for each of the two statistics.

It may be concluded that, whereas record lengths of less than 30 years are rather unsatisfactory, values of r and C derived from record lengths of 40 years or more are in such good agreement with those based on 50-year values that further increases in record length will probably not materially improve the estimate of the degree of long-term correlation or of variability. In this, the present results for r and C are in good accord with the record lengths suggested by Landsberg and Jacobs for specifying stable averages.

To use, as in the above, a 50-year sample as a fiducial standard is, of course, a rather arbitrary step. If an investigator's criterion for optimal record length is in any sense similar to the predictive criterion employed by Beaumont or Enger, then some test of the least-squares type that they used becomes more appropriate than the simple convergence criterion employed here. The criterion used here is chiefly relevant to studies in which one is interested in historically descriptive statistics pertaining to degree of correlation or to relative variability of precipitation data.

Summary. A double-mass study of 37 precipitation records for stations well distributed over the United States revealed an unanticipated tendency for precipitation records taken in large cities to be about as homogeneous as those for rural areas and small towns. The study further indicated that internal breaks as large as those

TABLE 2. Percentage Frequencies of Q_r and Q_c

	Years of Record	Q_r or Q_c <−1.1	−0.6 to −1.0	−0.1 to −0.5	0.0 to 0.4	0.5 to 0.9	>1.0
Per cent Q_r	10	39	8	7	4	10	32
	20	29	7	18	12	15	19
	30	15	17	17	25	15	11
	40	3	8	37	32	15	5
Per cent Q_c	10	52	7	8	5	9	19
	20	26	15	16	15	15	13
	30	16	12	20	18	21	13
	40	8	5	33	25	29	0

offered as type cases in the literature on double-mass analysis are probably rather rare, at least in data pertaining to nonmountainous regions. The greatest observed break amounted to a 25 per cent change in rain gage catch. A very rough estimate puts the modal break at about 10 per cent for the sample of 37 cases examined.

Using 50 stations' 50-year records of winter- and summer-half-year precipitation totals, an empirical test of effects of inhomogeneity was carried out. Each of the 100 seasonal records was used in originally published form to compute measures of both correlation and relative variability, and the computations were then repeated for the data in an adjusted form in which a synthetic, positive 20 per cent break was imposed on the latter half of each record. Correlation coefficients tended to be lowered and coefficients of variation to be raised by the inhomogeneity, but by amounts that would, for most climatological or hydrometeorological purposes, be regarded as rather unimportant as measured against the associated standard errors of r and C.

From examination of the rate of improvement in the agreement between n-year and 50-year estimates of coefficients of correlation and variation, as n was increased in 10-year steps from 10 to 40, it was concluded that acceptable estimates of the 50-year values are not obtained until n is greater than 30 years, with 40-year samples giving very good estimates, a result that is similar to previous estimates of record lengths required to obtain stable arithmetic means for precipitation amounts.

Acknowledgments. We wish to express our appreciation to Robert W. Mitchell of the University of Arizona Numerical Analysis Laboratory for his assistance in computational work. Support of the Office of Naval Research, under Contract NR 082-164, is acknowledged.

References

Beaumont, R. T., A criterion for selection of length of record for a moving arithmetic mean for hydrologic data, *Trans. Am. Geophys. Union, 38,* 198–200, 1957.

Enger, I., Optimum length of record for climatological estimates of temperature, *J. Geophys. Research, 64,* 779–787, 1959.

Landsberg, H. E., and W. C. Jacobs, Applied climatology, in *Compendium of Meteorology,* edited by T. F. Malone, Am. Meteorol. Soc., Boston, 976–992, 1951.

Lenhard, R. W., and W. A. Baum, Some considerations on normal monthly temperatures, *J. Meteorol., 11,* 392–398, 1954.

Linsley, R. K., M. A. Kohler, and J. L. H. Paulhus, *Applied Hydrology,* McGraw-Hill Book Co., New York, 689 pp., 1959.

(Manuscript received April 14, 1960.)