THE USE OF CORRELATION FIELDS IN RELATING PRECIPITATION TO CIRCULATION

By C. K. Stidd

U. S. Weather Bureau [1]

(Manuscript received 14 December 1953)

ABSTRACT

The physical meaning of a correlation-field pattern is explored, and such patterns are shown to be analogous to composite patterns.

Correlation fields are used to investigate the relationship of precipitation to mean monthly, 5-day and synoptic pressure-contour patterns over the United States and adjacent areas.

1. Introduction

The long-range forecasts of temperature and precipitation made by the U. S. Weather Bureau depend upon prognostic mean pressure or height[2] charts for the forecast period (Namias, 1947). The preparation of these charts is the major problem of the Extended Forecast Section, but the subsidiary problem of interpreting the prognostic charts in terms of the weather also demands a large share of attention and is somewhat complex (Martin and Hawkins, 1950).

This study is directed toward one aspect of this subsidiary problem, the interpretation of mean monthly 700-mb charts in terms of precipitation. In order to relate a given precipitation record to the pressure field, a large number of coefficients have been computed of the correlation between the given precipitation and various point-pressures. Maps are used to display these coefficients, each one being plotted at the point where its pressures were measured. This field of correlation coefficients may be analyzed by the drawing of isopleths, a process which usually yields a smooth and well-defined pattern. Most of the figures in this paper consist of correlation-field patterns for various rainfall parameters.

Correlation fields have been used by a number of investigators, an early example being found in Pearson and Lee's (1897) study of barometric heights. For the most part, however, it appears to the writer that the science of statistics has concerned itself with simple point variables, distributed in either time or space alone. In meteorology, we are frequently concerned with variables that are distributed in space as well as in time; precipitation, for instance, has two-dimensional distributions in space, and pressure and temperature are distributed in all three space dimensions. Many of the well-known statistical tests and methods are designed for use with random samples or with simple point variables, and in dealing with variables distributed in two or more dimensions one must be wary of forming conclusions on the basis of experience with the statistics of ordinary point variables. For example, Pearson and Lee came to this conclusion: It would thus seem . . . within our power to predict almost exactly the height (barometer) at any selected station from a knowledge of the heights at two other selected stations at selected intervals of time. Meteorological experience since that time indicates that this statement is over-optimistic.

Since there is a continuous distribution of correlation coefficients available in a correlation field, Pearson and Lee assumed that a systematic search of such fields would reveal two variables which would have the precise value of intercorrelation necessary to bring the multiple correlation to unity. They did not seem to realize that the intercorrelations, themselves, would vary in a systematic manner precisely calculated to defeat their purpose.

For example, in fig. 1c, the isopleth running through the Tennessee Valley is for a correlation coefficient of 0.50. If any two points along this line were used in multiple correlation to forecast Tennessee Valley rainfall, the multiple correlation coefficient,

$$R = \sqrt{\frac{r_{1y}^2 + r_{2y}^2 - 2r_{1y}r_{2y}r_{12}}{1 - r_{12}^2}},$$

would depend also on the intercorrelation, r_{12}, of the 700-mb heights at the two points. If in the above equation R is set equal to unity (expressing a perfect relationship), the necessary value of r_{12} can be determined, since both r_{1y} and r_{2y} are known to be equal to $+0.5$. Thus r_{12}, for a perfect relationship, is found to have a value of -0.5.

Since the value of the intercorrelation may be expected to diminish from unity when the two points

[1] Published, with permission of the Director, as Technical Paper No. 223 of the Pineapple Research Institute.

[2] To avoid repetition, the word "pressure" alone frequently will be used in reference to heights of pressure surfaces as well as sea-level or station pressures.

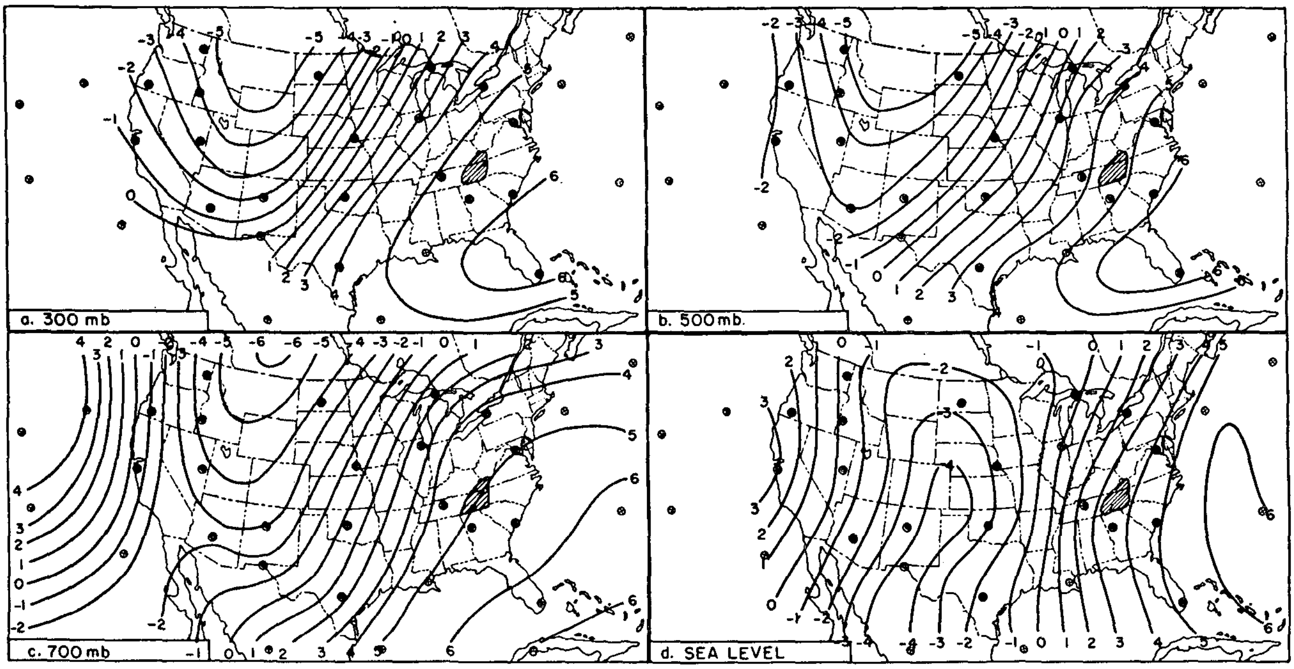

FIG. 1. Correlation-field patterns of monthly precipitation averaged over Tennessee Valley *vs.* mean monthly heights of (a) 300-, (b) 500-, (c) 700-mb pressure surfaces and (d) sea-level pressure. Winter months, 39 cases. Decimal points have been omitted in labelling isopleths.

are contiguous to some low value when they are far apart, it might be supposed that two points a sufficient distance apart on the +0.5-correlation isopleth might have the desired intercorrelation of −0.5. In practice, however, it is invariably found that the correlation isopleth will come full circle and close upon itself without the desired low value of intercorrelation ever having been attained.

The earlier parts of this study will be concerned with the problem of interpreting a correlation field pattern, and with analogies between these patterns and those developed by other techniques. In the rest

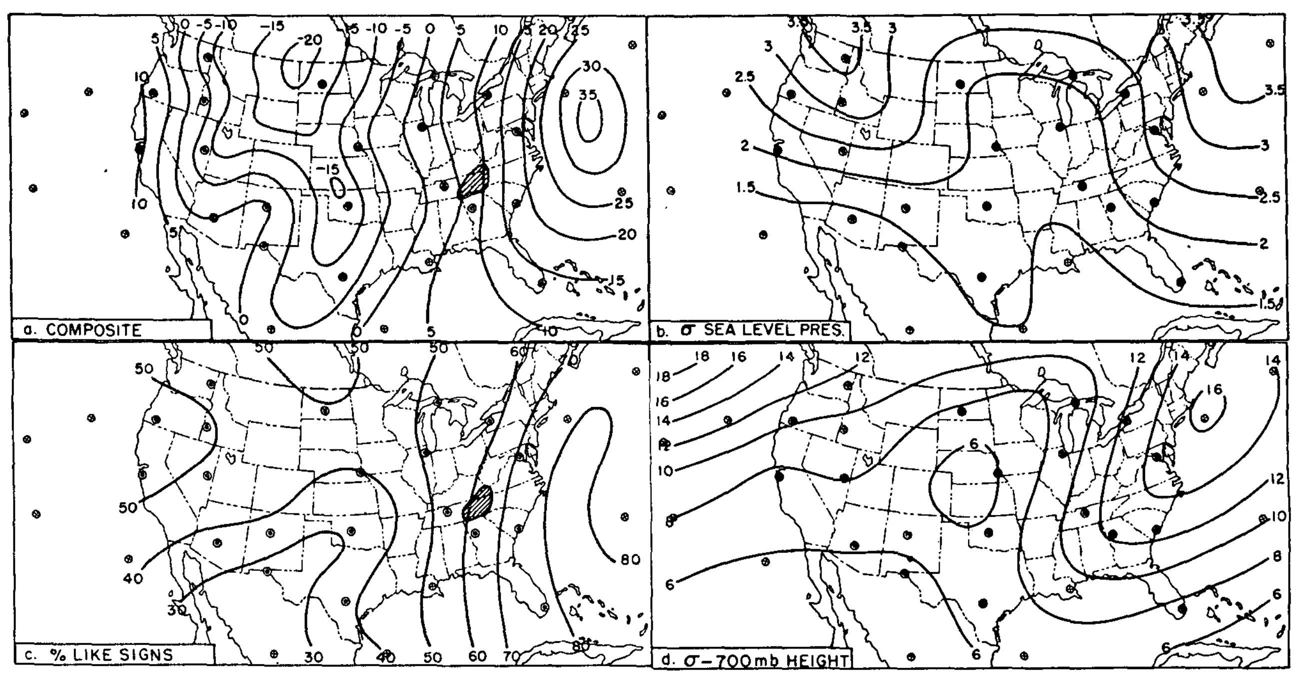

FIG. 2. December, January and February data, 1939–1952 (39 cases). (a): composite-difference chart of sea-level pressures; sum of pressures (mb) for five wettest months of Tennessee Valley data minus sum for five driest months. (b): standard deviation of mean monthly sea-level pressures (mb). (c): per cent of months for which mean monthly sea-level pressure departure from normal had same sign as departure from normal of Tennessee Valley precipitation. (d): standard deviation of mean monthly 700-mb heights (tens of feet).

of the study, the correlation-field technique will be used to investigate the relationship of the precipitation of various areas in the United States to the contours of the 700-mb pressure surface.

2. A three-dimensional correlation field

An example of a correlation field is found in fig. 1, which relates the wintertime precipitation of the Tennessee Valley to the mean monthly values of sea-level pressure and of 700-, 500- and 300-mb heights at various points throughout the United States and adjoining areas. Details of the data treatment, normalization of the precipitation, removal of trends, and methods of averaging will be found in a later section. Time periods, number of cases, and data sources may be found in table 1.

This initial example employs four different levels, since it was deemed desirable to obtain a three-dimensional view of the correlation field. The four levels are closely related and, while the axes of the "troughs" and "ridges" in the correlation field seem to slope with height, there is little change with height in the magnitudes of the maximum correlations, and it was found that one level would serve about as well as another for the purpose of making linear-regression estimates of rainfall. In considering one of the conditions which

TABLE 1. Data sources.

Fig.	Type	Period	Total cases	Source
1	Monthly precipitation, Tennessee Valley	Dec. 1939 through Feb. 1952 (winter months)	39	Obtained from Tennessee Valley Authority.
	Mean monthly 300-, 500- and 700-mb heights for each of 21 stations			Mean monthly soundings published in *Mon. Wea. Rev.* and (since 1950) in *Climat. Data, Natl. summary.*
	Mean monthly 700-mb heights at grid-points and all sea-level pressures			Mean monthly northern-hemisphere 700-mb maps and sea-level maps prepared by Extended Forecast Section of U. S. Weather Bureau.
2	Sea-level pressure data	Same as fig. 1	39	Same as fig. 1.
3	Precipitation and mean monthly Oklahoma City surface pressure	Same as fig. 1	39	Same as fig. 1.
4	Monthly precipitation averaged over Tennessee Valley above Chattanooga (based on 33 gages)	Jan. 1890 through Feb. 1952 (winter months)	188	Obtained from Tennessee Valley Authority.
5	Monthly precipitation averaged by states	Same as fig. 1	39	*Mon. Wea. Rev.* and *Climat. Data, Natl. summary.*
	Mean monthly 700-mb heights			Same as fig. 1.
6	Same as fig. 5	Same as fig. 5	39	Same as fig. 5.
7	Monthly precipitation for United States as a whole	Same as fig. 1	39	Tannehill (1947); brought up-to-date by figures furnished by Earl Thom, U. S. Weather Bureau.
	Mean monthly 700-mb heights, station and grid-point			Same as fig. 1.
8	July precipitation averaged by states	1891–1940	50	Same as fig. 3.
	Mean July surface pressures (24 stations)			*World Weather Records.*
9	5-day average precipitation amounts, Tennessee Valley	One 5-day period each week, Dec. 1940 through Feb. 1945 (winter months)	60	Obtained from Glenn Brier, U. S. Weather Bureau.
	5-day mean 10,000-ft pressures			Interpolated from 5-day mean 10,000-ft maps of Extended Forecast Section, U. S. Weather Bureau.
	24-hr average precipitation amounts, Tennessee Valley	Every fifth day, Jan. 1941 through Feb. 1945 (winter months)	79	Obtained from Glenn Brier, U. S. Weather Bureau.
	Synoptic 10,000-ft pressures			Interpolated from microfilmed 10,000-ft map series in Library of U. S. Weather Bureau.

produces heavy rainfall, attention is directed to the sea-level chart, fig. 1d, where the centers are closest together and the resulting tight gradient of isopleths, if these isopleths could be considered analogous to those of pressure patterns, would suggest a strong flow of air from the Gulf of Mexico to the Tennessee Valley. This would support the conclusion of Klein (1949) in his study of 5-day rainfall amounts in the Tennessee Valley. Such an interpretation is valid, of course, only to the extent that the correlation-field patterns may be considered analogous to pressure-difference charts. This analogy is established in the next section.

3. Analogy of correlation patterns to composite patterns

A popular method of revealing relationships of the type discussed in this paper is based on the use of composites (*e.g.*, Solot, 1948). Using this method, the investigator selects from his record those months which had the heaviest rainfall and makes a composite, or average, of the pressure charts for that group of months. He might then take an equal number of the driest months, make a composite of their pressure charts, and subtract the dry composite from the wet composite to obtain a composite-difference chart. Such a chart, based on the sea-level pressure data of fig. 1d, is shown in fig. 2a. It reveals the differences in flow between wet months and dry months, and has physical meaning to the meteorologist. It is the purpose of this section to show that the correlation-field chart (fig. 1d) is analogous to, and more reliable for some purposes than, the composite-difference chart (fig. 2a).

Fig. 3 is a scatter diagram of the mean monthly sea-level pressures at Oklahoma City *versus* Tennessee Valley rainfall. The values are taken as differences from their means, and the cube roots will be explained in the section on data treatment. The correlation coefficient of this relationship is -0.37, the figure

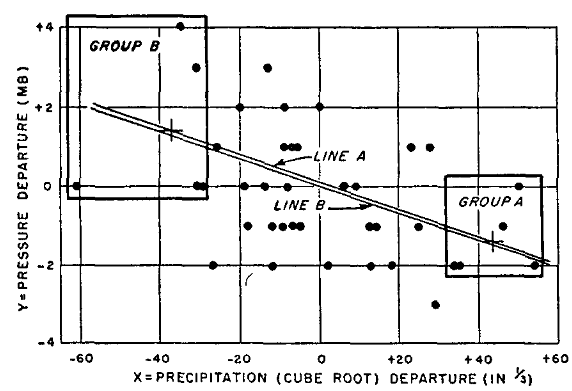

FIG. 3. Scatter diagram of Oklahoma City mean monthly surface pressures *vs.* Tennessee Valley monthly precipitation, winter months (December, January and February) from December 1935–February 1952, with seasonal trends removed. Cube roots taken of precipitation amounts prior to removal of trend. Groups A and B designate five highest and five lowest precipitation cases. Means of Groups A and B, designated by crosses (+), determine line A. Line B is least-squares regression line.

which was plotted at Oklahoma City in fig. 1d. If the composite-difference technique were being used, the pressures of an arbitrary number of points at the extreme right-hand end of the pattern would be averaged, and from this average would be subtracted the average of the same number of points at the extreme left-hand end of the pattern. In this example, the sum of the pressure differences of the five points in group A is -7 mb, and the sum of the five points in group B is $+7$ mb, the difference being -14 mb, the figure which was plotted at Oklahoma City on the composite-difference chart (fig. 2a).

If a scatter diagram similar to fig. 3 were plotted for some station other than Oklahoma City, the plotted points would shift to different vertical positions; but they would not move horizontally, since the rainfall data would remain the same. This means, in effect, that the composite differences are simply rough estimates of the slopes of the regression lines of pressures at various stations on precipitation in the Tennessee Valley. Line A in fig. 3 is such an estimate. (The regression line of precipitation on pressure is actually the relationship of direct interest. The composite technique cannot reveal this, since it would involve the selection of the five highest and five lowest pressures.)

A more accurate estimate of the slope of the regression line is given by the equation $m = \sum xy/\sum x^2$, where m is the slope, and y and x are the individual departures from the means. This gives the regression line which minimizes the sums of the squares of the errors of y (Hoel, 1947, p. 79). Line B in fig. 3 is this regression line. It thus is evident that a composite-difference chart is proportional to, but less accurate than, a chart of regression coefficients.

One other weighting factor is desirable, however, since the slopes of the lines depend in part on the overall spread, or degree of dispersion, of the y values, a factor which may vary from place to place, and which has no bearing on the relationship between pressure and precipitation.

This correction can be made by dividing m by the standard deviation of the y values. This results in an amount proportional to the correlation coefficient, r, since $r = m\sigma_x/\sigma_y$, and σ_x is constant throughout a given chart.

Thus, if the composite differences were divided by the standard deviations of pressure, the resulting chart would be proportional to a correlation-field pattern. Since the standard deviation of pressure varies slowly from point to point on a map, the correlation field might be regarded as analogous to the composite difference chart, even as regards direction and gradient of isopleths, although the strict analogy is between a chart of regression coefficients and a composite-difference chart.

It is desirable to develop the formula which expresses precisely the relationship between the correlation of the geostrophic wind and the gradient of isopleths in the correlation field. This can be done as follows.

The geostrophic wind equation is

$$\lambda V = dh/dn, \tag{1}$$

where $\lambda = 2(\omega/g)\sin\phi$, V = speed of geostrophic wind, h = height of pressure surface, n = distance on pressure surface, normal to contours, ω = earth's angular speed, g = acceleration of gravity, and ϕ = latitude. It may be expressed as

$$\lambda V = \lim_{\Delta n \to 0} \frac{\Delta h}{\Delta n} = \frac{h_1 - h_2}{\Delta n}. \tag{2}$$

The correlation of velocity with precipitation is

$$r_{vw} = (\textstyle\sum vw)/N\sigma_v\sigma_w, \tag{3}$$

where r = correlation coefficient, σ = standard deviation, w = precipitation, expressed as a departure from the mean, v = difference in speed between the vector mean V and that component of the individual V which is parallel to the vector mean V, and N = number of pairs.

Since (2) is linear, and Δn may be held constant at a given point, the value for V in (2) may be substituted for v in (3), giving

$$r_{vw} = \frac{\sum (h_1 - h_2)w}{\lambda\, \Delta n\, N \sigma_v \sigma_w}, \tag{4}$$

$(h_1 - h_2)$ being expressed as a departure from the mean. Thus,

$$\lambda\sigma_v r_{vw} = (\textstyle\sum h_1 w - \sum h_2 w)/N\sigma_w\, \Delta n, \tag{5}$$

and, since

$$(\textstyle\sum hw)/N\sigma_w\sigma_h = r_{hw}, \tag{6}$$

equation (5) may be written as

$$\lambda\sigma_v r_{vw} = (\sigma_{h_1} r_{h_1 w} - \sigma_{h_2} r_{h_2 w})/\Delta n, \tag{7}$$

which, in the limit, is equal to

$$\lambda\sigma_v r_{vw} = d(\sigma_h r_{hw})/dn. \tag{8}$$

Since $\sigma_v r_{vw}$ and $\sigma_h r_{hw}$ are, throughout a given chart, proportional to the regressions of precipitation on geostrophic wind and height of pressure surfaces, respectively, (8) shows that these regressions may be substituted into the geostrophic wind equation for the simple elements, v and h, themselves. A somewhat similar development, with use of the hydrostatic relationship, shows that the vertical change in the correlation field is related to the correlation of temperature with precipitation.

Equation (8) may be found useful when correlation-field patterns are being used in the search for predictors to be used in objective methods. Equation (8), in combination with a correlation field of pressures and a chart of the standard deviations of pressures, will allow the investigator to select those areas where the geostrophic wind will be expected to correlate well with the predictand (provided that some inference can be made as to the variance of the geostrophic wind). With reference to fig. 1d, for instance, the initial assumption might be that the point where the geostrophic wind has the strongest correlation with Tennessee Valley precipitation is near the western border of Tennessee, where the isopleths are most crowded. Reference to fig. 2b indicates that, while the standard deviation of sea-level pressures is increasing to the east, the relative rate of increase does not seem to be as great as the relative rate of decrease of the gradient of the isopleths in fig. 1d. Disregarding possible changes in the standard deviation of the geostrophic wind, we may assume that the geostrophic wind near the western border of Tennessee has a higher correlation with Tennessee Valley precipitation than does the geostrophic wind of the Tennessee Valley itself.

As a rough check, the correlations with Tennessee Valley precipitation were computed for the gradients from Oklahoma City to Nashville and from Nashville to Charleston, giving $r = -0.68$ in the first case and $r = -0.40$ in the second.

Instead of composite-difference charts as described above, some investigators work with the composite for heavy precipitation alone, expressing this in terms of difference from "normal". As long as only straight-line relationships are being considered, this again is equivalent to an estimate of the slope, the difference in this case being from one extreme to the mean rather than from one extreme to the other.

The composite "difference from normal" chart for heavy precipitation will often be found to be not the precise opposite of the composite for light precipitation. In most cases this lack of similarity between the two patterns is due to the fact that, for some areas, the "normal" value is quite different from the mean value of the sample from which the heavy and light cases were selected. In other words, the departures of groups A and B in fig. 3 are taken not from the zero axis but from some independently determined "normal" value, which may, in some cases, lie much closer to one group than to the other. If both groups lie close to the zero axis, it might happen that both will be above or below "normal."

Dissimilarities of this sort will not be found if all differences are taken from the sample means. Any remaining dissimilarities are likely to be accidental in source, since it seems unlikely that this technique is sensitive enough to reveal any curvilinear relationships in patterns such as that of fig. 3.

Comparison of fig. 2a with fig. 1d indicates that the centers of maximum value on the composite-difference chart are not in the same places as the centers of maximum "relationship" as revealed by the correlation field. This is due to the fact that the composite method makes no allowance for the differing degrees of variance for different points on the chart. Thus a point at some high latitude, where the pressure fluctuates over an exceedingly wide range, may be expected (if a relationship exists) to have a high value on the composite difference chart, while a point in the tropics, where the pressure is comparatively steady, will have only a moderate value. Note that both the maximum and minimum in fig. 2a are displaced well to the north of the centers of maximum correlation on fig. 1d. The negative center, in fact, appears to lie at a point where the correlation is only about -0.20, whereas the maximum negative correlation on fig. 1d is probably greater than -0.45.

The fact that these centers have been displaced in the direction of greater variance may be verified by reference to fig. 2b, which maps the standard deviation of mean monthly sea-level pressure.

A second simplified approach to the problem avoids the above mentioned disadvantage, but is somewhat less sensitive than the correlation technique. This approach, sometimes called the "frequency-chart technique," simply characterizes all values as above or below normal, and the figures plotted on the chart are the percentages (or frequency) of times that like signs were observed between rainfall at a given point and pressure at the plotted point.

In its essence, this method is equivalent to the correlation technique, the values, however, being classified into only two categories. The technique was used quite extensively by Bundgaard and Martin (1954) in exploring point-to-point pressure relationships, and its use is recommended in extensive, exploratory investigations of this sort.

Fig. 2c shows the analysis of the data by this technique. It is evident that the pattern is quite similar to that of fig. 1d, and that the method gives an unbiased estimate of the correlation field. The frequency patterns, however, are often not as smooth and definitive as the correlation-field patterns.

Most of the time-saving in both the "composite" and "frequency" methods stems from the fact that incomplete use is made of the data.

4. Data sources and treatment

If two variables are correlated, it frequently is possible to increase the magnitude of their correlation coefficient by using mean values of either or both the variables. Space means or time means may be used for this purpose. While high correlation coefficients are not necessarily more meaningful than low ones, they should be as high as possible in correlation-field studies to throw the patterns into sharp focus. Averaging eliminates some of the accidental or extraneous elements which reduce correlation coefficients, and the limit to which this averaging process may be carried depends on the homogeneity of the elements that go into the mean, homogeneity, that is, with respect to the relationship in question. (The correlation of random samples drawn from correlated populations is the same as the correlation of individuals. Time or space means of meteorological data are not random, however, and, in practice, they are found almost invariably to correlate more strongly than do the individuals.) A further limitation on the use of time means is the fact that combining individual cases into mean values greatly reduces the number of cases available for testing. These considerations, together with that of the availability of data, account for the predominant use of mean monthly and state-wide average values.

The data sources, with the periods and total number of cases, are listed in table 1. The data received special treatment as follows.

Heights of pressure surfaces from mean monthly soundings at given stations.—In the earlier years, these mean soundings were expressed as series of pressures for given heights; linear interpolation was used to convert these figures to heights of given pressure surfaces, spot checks indicating that the small systematic errors introduced by this process would not materially affect the correlations.

Precipitation data.—One difficulty commonly encountered in working with precipitation stems from the fact that precipitation time-distributions are skewed, the bulk of the observations falling in the lower half of the total range. In correlation computations, the extremely high precipitation values thus may be weighted heavily in comparison to the extremely low values.

The averaging process does not completely eliminate this skewness. To obtain precipitation indices which have better balance between the high and low values, the cube roots of the precipitation amounts, rather than the amounts themselves, have been used.

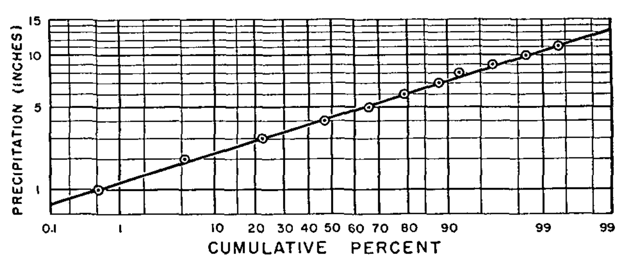

FIG. 4. Cumulative frequency diagram of January, February and December precipitation amounts for Tennessee Valley above Chattanooga, January 1890 through February 1952. Amounts are averages for valley, based on approximately thirty-three gages. Scales are cube-root scale (ordinate) and probability scale (abscissa).

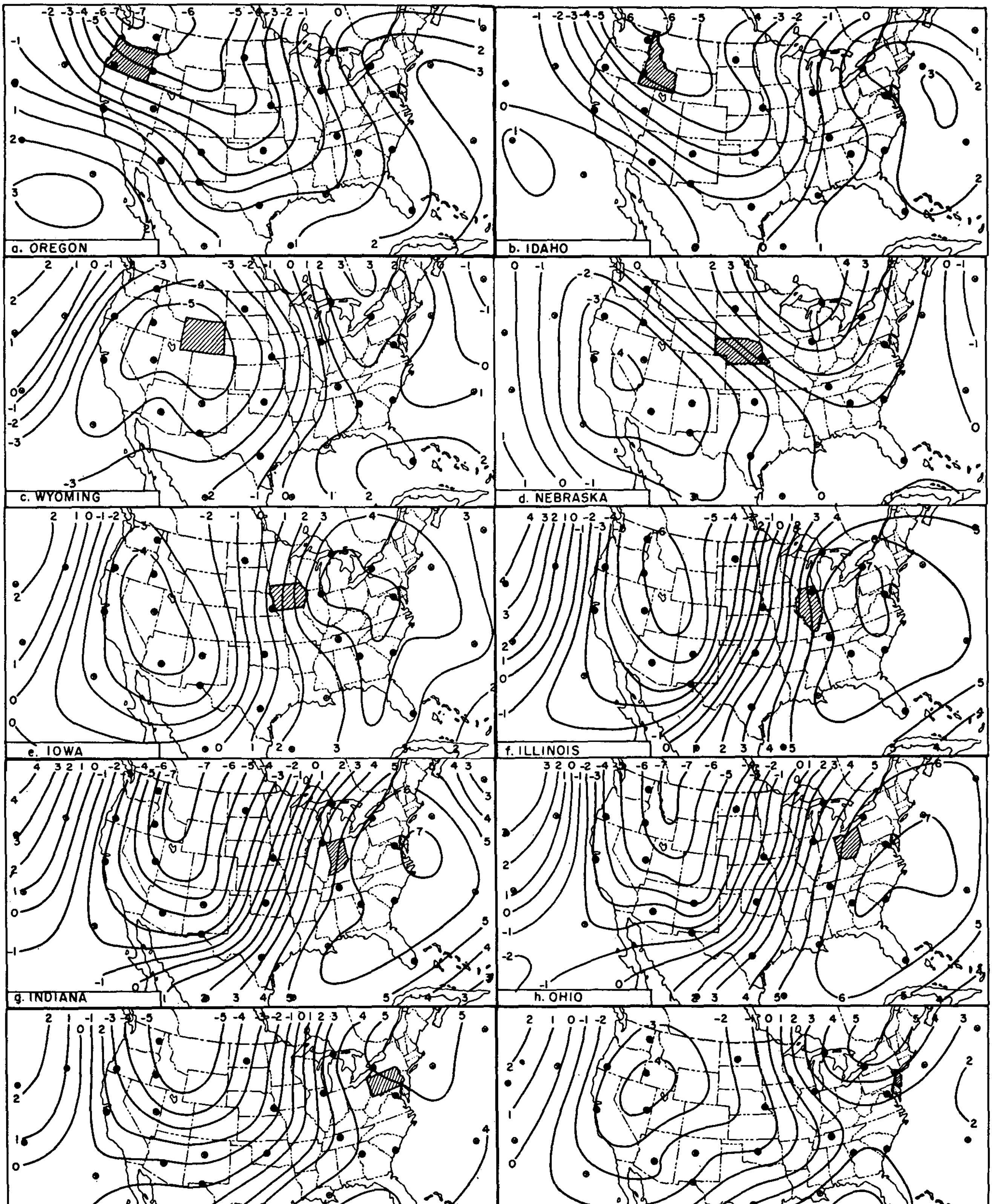

FIG. 5. Correlation-field patterns of mean monthly 700-mb height *vs.* monthly precipitation averaged over given state. Winter months, 39 cases. For ten states, side-by-side, approximately along 40th parallel.

An earlier study (Stidd, 1953) indicates that the cube roots of single-station precipitation amounts may be considered to have normal distributions. Areal-average precipitation amounts often do not fit this scheme, since, unless the area is completely homogeneous with respect to precipitation, the time distribution of the areal-average precipitation may be quite complicated. While the cube-root transformation may not "normalize" such distributions, it can effectively remove much of the skewness, so that the influence of the extremely heavy rainfalls will not mask the smaller variations. In this study, the cube roots of precipitation amounts have been used in every case, except that for the precipitation of the United States as a whole. In this particular case, the average has been taken over such a large and inhomogeneous area that the total variability is small compared to the mean, and the degree of skewness is very slight.

An example of the cube-root distribution is given in fig. 4 which shows a cumulative frequency diagram, or ogive, for Tennessee Valley precipitation, plotted on a graph with a cube-root ordinate scale and a probability abscissa. The fact that the points tend to plot as a straight line confirms the assumption of the normal distribution of the cube roots of the precipitation. The months December, January and February in this case are treated as a single group, since their separate distributions are very much alike.

In this respect, the Tennessee Valley precipitation amounts are like those from a single station, and the area appears to be very homogeneous. State-wide precipitation averages should not, in general, fall into line as precisely as this, but it has been assumed that the cube-root transformation would remove much of the skewness without introducing other serious errors.

Trends in time series.—In all cases, seasonal trends have been removed by expressing each value as a difference from the average for the month in which it appears.

5. Variation in correlation pattern with position of precipitation area

The three-dimensional aspects of a typical correlation-field pattern, as shown in fig. 1, suggest that a fairly close relationship exists between the patterns for adjacent levels. For the rest of this study, then, the 700-mb level, alone, will be considered.

Fig. 5 shows the 700-mb correlation patterns for a series of state-wide monthly rainfall averages. Each map relates the 700-mb pattern to the rainfall of a given state, from Oregon eastward approximately along the fortieth parallel to New Jersey.

The following paragraphs supply for these patterns an interpretation based on the analogy which has been shown to exist between correlation fields and anomalous flow patterns. The purpose of the discussion is to draw attention to what the writer believes to be the more interesting and consistent features of the patterns; the discussion is by no means complete; it is not necessarily correct; and other possible explanations should not be ruled out.

Considering this series in its entirety, we note that the strongest relationships (greatest magnitudes of maximum or minimum correlation coefficient) are found for those states which are within reach of a major moisture source, and the patterns are invariably arranged with a tight gradient of anomalous flow in a direct path from the moisture source to the state in question. Oregon and Idaho, on the west slopes of the continental divide, obtain moisture from the Pacific. The westerlies perhaps have lost most of their moisture by the time they reach Wyoming; here the pattern taps no moisture source, and the relationship is weaker. For Nebraska, the relationship is weakest of all; the moisture source, as indicated by the pattern, is the Gulf of Mexico, but this moisture could reach Nebraska only during temporary breakdowns in the normal flow of the westerlies. The weakness of this correlation field may be due to the fact that, in the absence of a moisture source, Nebraska precipitation stems from a variety of causes. From Iowa through Ohio, strong southerly anomalous flows from the Gulf prevail. For Pennsylvania, the pattern is so arranged as to provide anomalous flow from the Atlantic; but a slight eastward shift of the pattern would place Pennsylvania in the anomalous flow from the Gulf. This fact suggests that both sources of moisture are important to Pennsylvania. For New Jersey, the Atlantic source is strongly emphasized.

It is an interesting fact that all the states east of Nebraska lie in regions of positive correlation (pressure *above* normal favorable to precipitation). It is suggested that this is due, in general, to the desirability, from the standpoint of precipitation, of providing a southerly anomalous flow from the Gulf of Mexico up into the central part of the country. The moisture thus provided may eventually be distributed over the entire eastern half of the United States. Some such consideration as this must be a strong factor, to outweigh the commonly accepted association of precipitation with low pressure or cyclonic curvature.

Low pressure and cyclonic curvature on the mean map may still be important rainfall producers, however, since they seem to play a prominent part in the precipitation of the far western states. The negative center over Wyoming on the chart for that state is probably directly related to precipitation, and in Oregon and Idaho there are deep negative centers in close proximity to the states, although the relationship between rainfall and cyclonic activity may be less direct in these two latter cases, since the direction of anomalous flow with relation to orography is probably a large factor in the effectiveness of the westerlies over Oregon and Idaho. Anomalous flow with relation to orography also may be a principal factor in Nebraska, where the nearly easterly gradient over that state on the correlation pattern signifies a reversal in the normal, down-slope, westerly flow in that area.

Fig. 6 shows the correlation patterns associated with precipitation in a number of other states or areas throughout the country. Discussion of these patterns is, for the most part, unnecessary, since the principles already mentioned are seen to apply. It might be remarked, however, that Florida presents the only case where a negative center is found in the eastern half of the United States. Since Florida is almost surrounded by water, the direction of flow with respect to moisture source is of little importance, and, as in the case of Wyoming, cyclonic activity appears to be the important factor.

6. Precipitation averaged over the entire United States

The fact that all but one of these patterns displays a negative center in the west and a positive center in the east suggests that a strong relationship may exist between mean 700-mb heights and rainfall for the United States as a whole. Such is shown to be the case in fig. 7. In these data, the state-wide rainfall averages are weighted in accordance with the state areas, to obtain an average for the whole country. The strength of the relationship is indicated by the maximum correlation values of −0.72 at Boise and +0.60 at Buffalo, which, having an intercorrelation of −0.58, yield a multiple correlation coefficient of 0.76. This relationship has been checked by reference to surface-pressure data (Clayton, 1927, *etc.*) for January values from 1886 through 1938, a total of 53 winter months. These data give correlation coefficients as follows:

1. between United States rainfall and Salt Lake City, Utah, surface pressure, −0.71;
2. between United States rainfall and Eastport, Maine, surface pressure, +0.48;
3. between Salt Lake City surface pressure and Eastport surface pressure, −0.17,

with a multiple correlation of 0.80 between United States rainfall as the dependent variable and surface pressures at Salt Lake City and Eastport as the independent variables. The Salt Lake City correlation coefficient by itself reveals the significance of the relationship. Salt Lake City was the only point selected for test in this area of apparent negative correlation, and the coefficient is based on data independent of that on which the selection was based; therefore, if the effects of serial correlation in a time series of consecutive Januarys can be neglected, it is evident that the relationship is real.

It should be pointed out here that part of this relationship, that between pressures in the Great Basin and rainfall throughout the United States, has been recognized by Tannehill (1947). He discusses a seasonal variation in the relationship and suggests some physical mechanisms to account for it. Klein (1952), in a

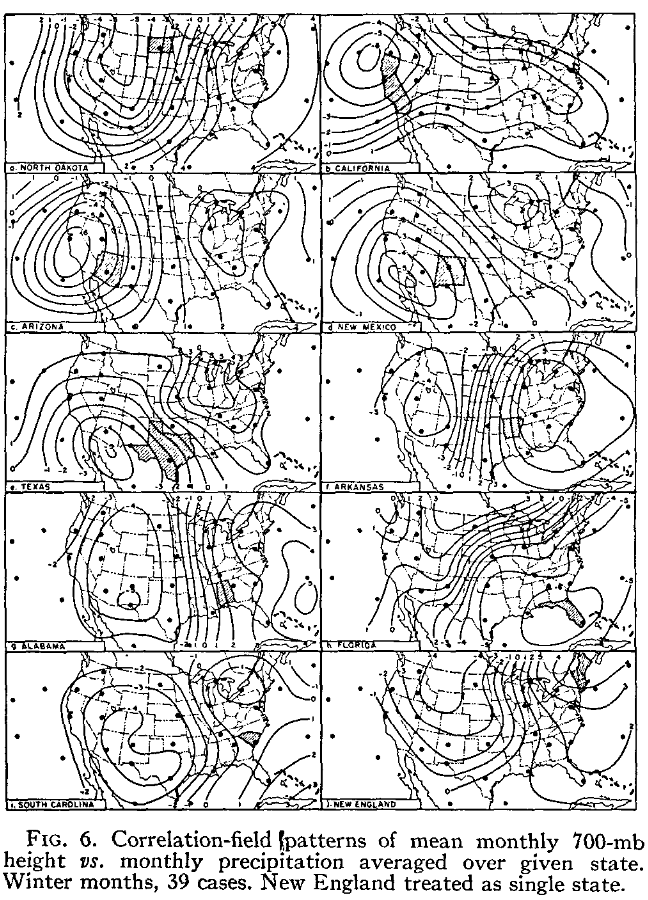

FIG. 6. Correlation-field patterns of mean monthly 700-mb height *vs.* monthly precipitation averaged over given state. Winter months, 39 cases. New England treated as single state.

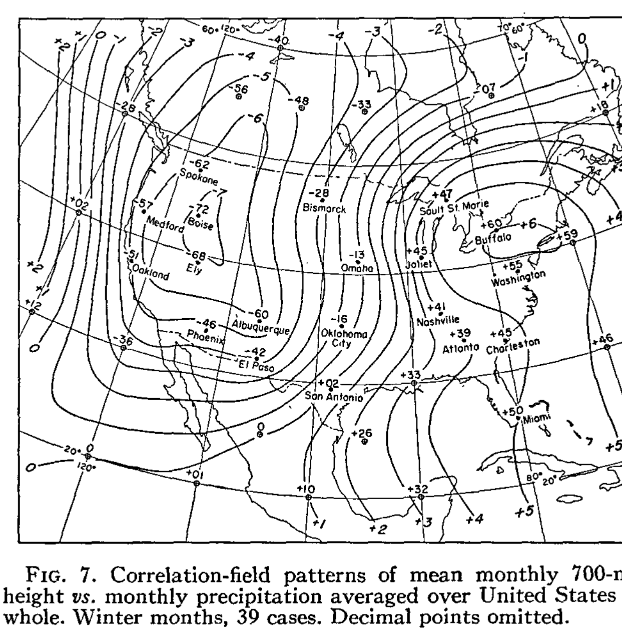

FIG. 7. Correlation-field patterns of mean monthly 700-mb height *vs.* monthly precipitation averaged over United States as whole. Winter months, 39 cases. Decimal points omitted.

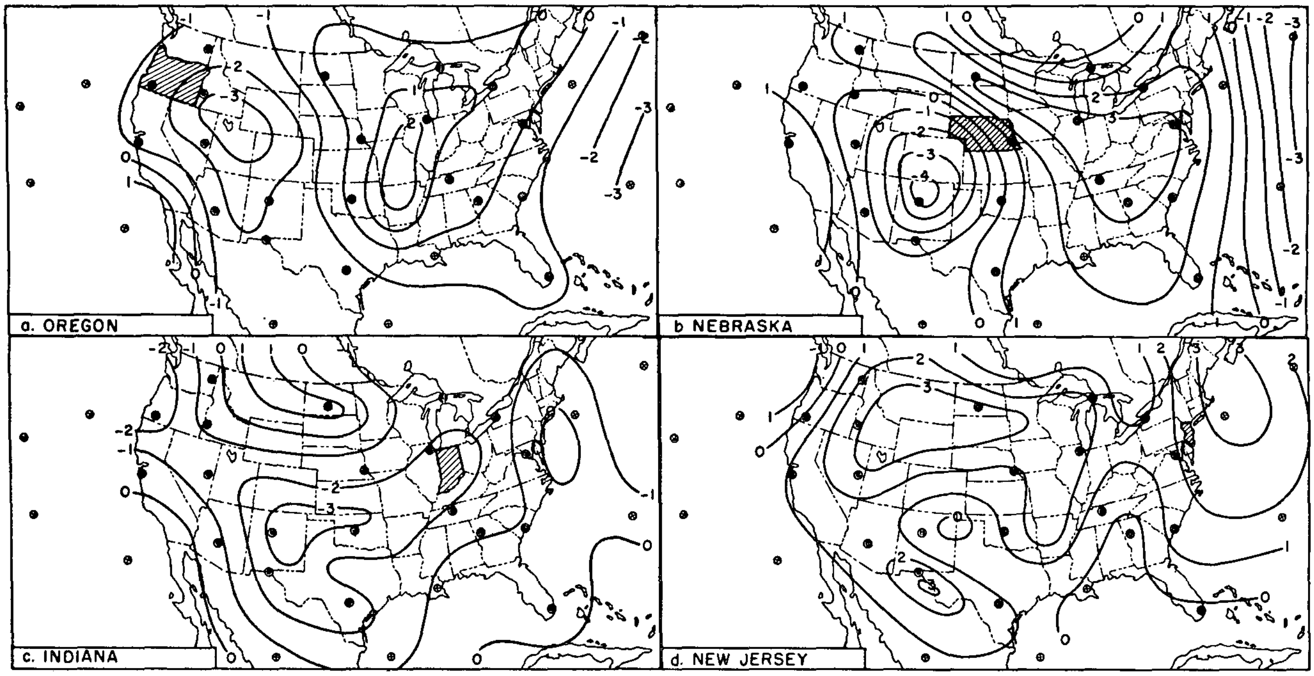

FIG. 8. Correlation-field patterns of mean July surface pressures *vs.* July precipitation averaged over given state. 50 cases.

discussion of long waves on mean charts, demonstrates that a ridge in the eastern half of the United States is normally associated with a trough in the West, and *vice versa*.

Although the data for fig. 7 are from the months of December, January and February only, the most striking illustration of the relationship is found during the month of October 1952. Winston (1952) points out that this was the driest month yet recorded for the country as a whole, the average rainfall being only 0.56 in, 27 per cent of normal. The mean 700-mb anomaly pattern for October 1952 is almost precisely the opposite of fig. 7. Winston also points out that Florida was the only state that had excessive rainfall, a fact which ties in nicely with the state-by-state map series, figs. 5 and 6, which shows Florida to be the only state having precipitation associated with a low anomaly in the eastern half of the country.

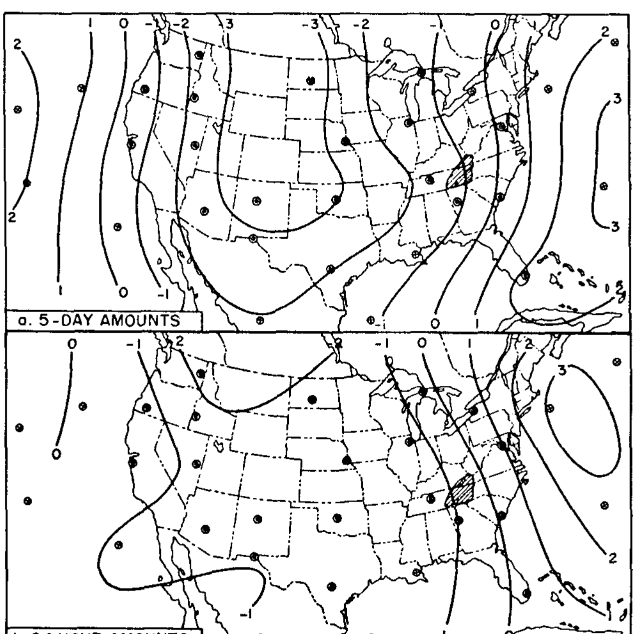

FIG. 9. Correlation-field patterns. (a): 5-day mean 10,000-ft pressures *vs.* 5-day precipitation amounts averaged over Tennessee Valley; winter-time data, 60 cases, data taken once each week. (b): 10,000-ft pressures *vs.* 24-hr precipitation amounts averaged over Tennessee Valley; winter-time data, 79 cases, data taken once every five days.

The large correlation coefficients obtained for country-wide precipitation must be due largely to the reduction of accidental or extraneous values in data through use of the averaging process. It is probably this reduction that allows the correlations to be generally stronger between related sets of monthly data than between related sets of data for shorter periods. An average taken over the entire United States must contain many elements that are unrelated or even negatively related to the main body of precipitation, but, in spite of this lack of homogeneity, high correlations are found.

7. Summer-time relationship

In the summer, this strong relationship breaks down, the July correlation between Salt Lake City surface pressure and U. S. rainfall being only −0.12.

Since the most abundant data are those from *World Weather Records*, surface pressures serve as the basis for a brief summer-time investigation. Correlation fields of July surface pressures with rainfall for Oregon, Nebraska, Indiana and New Jersey are shown in fig. 8. These patterns indicate that the relationship between summer-time monthly rainfall amounts and pressure patterns must be extremely tenuous, since chance alone could easily provide correlation coefficients as high as the ones on these charts.

8. Shorter-period relationships

Fig. 9 shows correlation-field patterns which are based on periods shorter than one month. These relate 10,000-ft pressure to Tennessee Valley precipitation, and may be compared with the 700-mb height *versus* Tennessee Valley monthly precipitation pattern of fig. 1c. Fig. 9a is based on 5-day mean 10,000-ft pressure values and 5-day total precipitation amounts. A major difference between this pattern and that of fig. 1c is found in the magnitudes of the maximum and minimum correlation coefficients. On the monthly chart these are greater than ± 0.6, whereas on the 5-day chart they may be less than ± 0.4. A similar reduction in magnitude is noted in fig. 9b, which compares mid-period synoptic 10,000-ft pressures with 24-hr precipitation amounts. This reduction in correlation is due in part to the higher percentage of "noise," or accidentally high or low values, found in data averaged over shorter periods. It is also likely, however, that much of this reduction stems from the possibility that short-period precipitation amounts are not as simply related to the large-scale flow pattern of a single level as are monthly amounts. On the monthly chart, fig. 1c, the correlation in the neighborhood of the Tennessee Valley, itself, is greater than $+0.4$, while on the 5-day chart the zero line has moved well to the east and the Tennessee Valley lies in an area of negative correlation. This eastward shift of the zero line has been made without any appreciable zonal displacement of the positive and negative centers. A negative correlation in the vicinity of the Tennessee Valley suggests an increased need for cyclonic activity for the production of shorter-period precipitation, and it may be that this need is in conflict with the ideal large-scale pattern which provides a flow of moisture from the Gulf of Mexico.

The latitudinal shift of the correlation centers between the long-period and short-period charts is hard to explain. A comparison of the 5-day 10,000-ft correlation field with that for sea level, if one were available, would give a rough idea of the mean virtual temperature correlation-field, which, in turn, might suggest an explanation.

9. Formulae for objective estimates

With regard to multiple correlation, using variables at points selected from the correlation field, the very limited experience of the writer indicates: (1) that no great improvement will be found beyond the best simple correlation in the field, and (2) that the best pair of points is the center of maximum and the center of minimum simple correlation.

The data used in obtaining fig. 1 are used to illustrate the possibilities of multiple correlation. In fig. 1, the best simple correlation, $+0.66$, is found at Miami, Florida, at the 700-mb level. The center of negative correlation (not shown in fig. 1c) reaches its maximum magnitude of -0.64 at 55°N and 105°W. Combination of these two in multiple correlation gives a value of 0.71. Experience indicates that this value will not be exceeded by the use of any other two points on this chart.

If another level is considered, it might be logical to try the sea-level pressure at El Paso, Texas, since this is near the negative center on the sea-level pattern. The correlation between Tennessee Valley precipitation and sea-level pressure at El Paso is -0.43. Combination of sea-level at El Paso and 700 mb at Miami gives a multiple correlation of 0.76, slightly higher than was obtained by combining the best two 700-mb points. If, finally, the 700-mb height at 55°N and 105°W is added, to give three independent predictors, the multiple correlation coefficient is boosted to 0.77. There seems not much likelihood of boosting this figure appreciably through the use of other combinations of variables, and it appears that even in a three-dimensional field, two key points, if properly selected, will be all that are needed.

The appearance in many of these charts of more than one center of correlation is to be expected from our knowledge of the inter-relationships in the pressure field itself. These inter-relationships have been discussed at length by Bundgaard and Martin (1954). Their charts show that a strong relationship, as revealed by a very large correlation coefficient in the Great Basin, would also reveal large correlation coefficients (but of opposite sign) both east and west of the Great Basin at distances of from 35 to 40 deg long. There need be no direct relationship between these secondary centers and the rainfall in question. In the event that the correlation in one of the centers of a correlation-field pattern is very much stronger than that of any other center, it may be suspected that the secondary centers are merely reflections due to internal relationships in the pressure field. If, however, two centers have nearly equal magnitudes, it may be supposed that both are directly important to the rainfall, their relative positions, however, being greatly influenced by the internal relationships in the pressure field itself.

10. Orthogonal methods

A major objection to the use of one or two "key points," as revealed by the correlation field, in objective-forecast methods is that the key points alone

TABLE 2. Multiple correlation coefficients of ten independent variables (map description) versus precipitation.

Area	Corresponding to figure	R
Tennessee Valley	1c	0.73
Nebraska	5d	0.81
Iowa	5e	0.79
Illinois	5f	0.79

are considered and the rest of the chart ignored. A variety of pattern configurations might be associated with a given set of key-point values. As a possible alternative to the correlation-field method, some exploratory work has been done on the use of orthogonal polynomials to describe individual mean monthly 700-mb charts by a method adapted from that described by Wadsworth (1948). Each chart was described by a formula of ten terms which, in effect, gave the height of the 700-mb surface in terms of latitude and longitude over an area extending from 20 to 60°N and from 60 to 130°W. The ten terms are so designed that, for a given map, each is independent of all the others (orthogonal). Use of these ten terms in multiple correlation involved the inverse of the ten-by-ten matrix of intercorrelations, a task which was done on punched-card machines by the U. S. Weather Bureau's Machine Tabulation Unit. The ten map-description parameters were then multiply correlated with precipitation, the resulting coefficients being given in table 2. The data indicate that, for the stronger correlation-field patterns, little improvement is to be gained through the use of the orthogonal-polynomial, multiple-correlation method. The multiple correlation for Nebraska of 0.81 is quite surprising, however, inasmuch as Nebraska has one of the weakest correlation-field patterns of any state tested. The high multiple-correlation figure for this state would seem to indicate that the orthogonal-polynomial method might be quite useful in dealing with relationships which are too complex to be revealed by the correlation field. On the other hand, the magnitude of this multiple correlation figure may be largely accidental.

The results of this preliminary test are encouraging, however, and this line of investigation is being continued.

11. Conclusions

Correlations between a space variable (such as pressure) and a point variable (as precipitation has been considered to be for the purpose of this study) may have little meaning unless the two- or three-dimensional correlation-field pattern is considered. In selection of data for correlation studies, the elimination of randomness through the averaging of many values will usually lead to larger correlation coefficients despite the loss in homogeneity. This leads to stronger, more sharply defined patterns.

Short-cut techniques, involving composite charts or frequency charts, may yield useful indications but should be supplanted by the correlation-field technique where detailed studies leading to objective methods are desired.

Correlation-field charts may be interpreted, in the first approximation, as though they were anomalous flow charts. The configurations of the patterns may thus lead to a better understanding of the physical processes.

Correlation-field charts provide a systematic and reliable tool for the selection of parameters for use in objective forecasting.

Acknowledgment.—The writer gratefully acknowledges the aid of Mrs. Adelene Dingman, who alone computed the several thousand correlation coefficients on which this article is based.

REFERENCES

Bundgaard, R. C., and D. E. Martin, 1954: Some inter-relationships in the broad-scale features of the general circulation. *Quart. J. r. meteor. Soc.*, to be published.

Clayton, H. H., 1927, 1944, 1947: World weather records. *Smithson. misc. Coll.*, **79**, **90** and **105**.

Hoel, R. G., 1947: *Introduction to mathematical statistics.* New York, John Wiley and Sons, 258 pp.

Klein, W. H., 1949: An objective method of forecasting five-day precipitation for the Tennessee Valley. *U. S. Wea. Bur., Res. Paps.*, No. 29, 60 pp.

——, 1952: Some empirical characteristics of long waves on monthly mean charts. *Mon. Wea. Rev.*, **80**, 203–219.

Martin, D. E., and H. F. Hawkins, 1950: Forecasting the Weather: The relationship of temperature and circulation aloft. (In six parts.) *Weatherwise*, **3**, Nos. 1–6.

Namias, J., 1947: *Extended forecasting by mean circulation methods.* Washington, U. S. Wea. Bur., 89 pp.

Pearson, K., and A. Lee, 1897: On the distribution of frequency of the barometric heights at divers stations. *Phil. Trans. roy. Soc. London*, A, **190**, 423–467.

Solot, S. B., 1948: Possibility of long range precipitation forecasting for the Hawaiian Islands. *U. S. Wea. Bur., Res. Paps.*, No. 28, 52 pp.

Stidd, C. K., 1953: The cube-root-normal precipitation distribution. *Trans. Amer. geophys. Union*, **34**, 31–35.

Tannehill, I. R., 1947: *Drought, its causes and effects.* Princeton, Princeton Univ. Press, 264 pp.

Wadsworth, G. P., 1948: Short range and extended forecasting by statistical methods. *Air Wea. Serv., Tech. Reps.*, No. 105–37, 202 pp.

Winston, J. S., 1952: The weather and circulation of October 1952—the driest month on record. *Mon. Wea. Rev.*, **80**, 190–194.