THE USE OF BIPLOTS TO EVALUATE SPATIAL EFFECTS OF WEATHER MODIFICATION

K. Ruben Gabriel
Department of Statistics
and
Division of Biostatistics
The University of Rochester, New York, 14627

Abstract. Weather modification effects are likely to be spread unevenly over an area rather than have equal effects everywhere. This paper describes graphical techniques for exploring such variable effects. They display the slopes of the effects' variations over the area. Such slopes may be obtained by regression and displayed on a map, or by principal component analysis and displayed with respect to the principal coordinates of the stations at which the effects were observed. Both these displays are biplots and the differences between the two may reveal special features of the effect variability over the area.

WEATHER MODIFICATION AND SPATIAL VARIATION. This paper proposes a statistical technique which may help in revealing some simple spatial patterns of weather modification. Whether intended or inadvertent, weather modification activities are likely to affect an entire geographic region rather than a single location, though their influence may be to be uneven over that region. Thus, the effects of cloud seeding or of an industrial concentration may be spread unevenly over a region of unknown boundaries. This spatial variability of effects makes it difficult to evaluate them statistically.

We propose a two stage technique for evaluating the effects of weather modification in a given region. The technique requires weather data for a number of stations in the region, both during the modification activity -- "seeded" periods -- and during periods without that activity -- "unseeded" periods. In the first stage, regressions of weather data on longitudes and latitudes are displayed in a "map-biplot" (Kempton, 1984; Gabriel and Mather, 1986) to see whether the spatial patterns of the weather were different during seeded periods from what they were in unseeded periods. In the second stage, principal component analysis is used to produce a "statistical map" in which stations with similar weather patterns are plotted near each other and stations with different patterns are plotted farther apart. The geographical location of the stations is not used at all in constructing that statistical "map" which can then be compared with the geographical map to see if and how the statistical configuration of the stations differs from their geographical scatter. If such differences are found they might be ascribed to the effects of the weather modification activity.

The proposed statistical technique is exploratory and may lead only to suggestions and formulation of hypotheses; one may hope for the future development of related confirmatory techniques that would also permit significance testing of apparent modification effects. The present discussion stresses the application of the techniques rather than their mathematical basis, of which a brief summary is given in the Appendix.

DATA. The typical data set considered here has precipitation y[i,j] (or other weather variables) recorded during time periods j=1,...,m at stations i=1,...,n, whose coordinates are given by vectors c[1],...,c[n]. The modification activity is presumed to have occurred only, or mainly, during certain of the m periods -- the "seeded" periods -- and the investigator is likely to have some hunches as to roughly at which of the n stations it should have had the largest effects -- the "target" -- and where the effects were likely to have been least and possibly null -- the "control".

THE TRANSVAAL EXAMPLE. Cloud seeding took place around Nelspruit, Transvaal, in the summers of 1972 to 1981 (October of the previous year to March of the year itself) and was thought to possibly augment precipitation downwind. Gabriel and Mather (1986) studied this by using rainfall data for the 1951 to 1982 summers, i.e., the 10 seeded summers as well as 22 unseeded summers which provided a baseline for comparison. Their data was from 36 stations in that area, of which 16 were in the potential target region -- northeast of Nelspruit. The present paper follows their analysis in stressing unusual patterns by displaying residuals from the average for each summer and from the average for each station, so that y[i,j] reflects how the j-th summer's precipitation at station i differed from the additive model of (overall average)+(summer effect)+(station effect).

THE MAP-BIPLOT. A map is used as the basis for this type of display, and individual stations are plotted at their locations c[1],...,c[n]. Superimposed on this station map are markers for the different time periods, the marker for period j plotted at coordinates r[j] which are calculated so as to reflect the direction and slope of the planar trend of precipitation in that period. Thus, if marker r[j] is directly the right of the origin, this reflects a West-East trend of increasing precipitation during the j-th period, and the farther out r[j] is from the origin, the steeper the trend's slope. Similarly, if r[j] were in any other direction from the origin, that would be the direction of the maximum precipitation trend in the j-th period, and the distance of marker r[j] from the origin would show how strong that trend was. (This type of biplot was originally proposed by Kempton, 1984, for an agricultural application.)

*Fig. 1. Transvaal Summer Rainfall: Map-biplot.
Station markers: C - control, T - target.
Summer markers: * - unseeded, SD - seeded.*

THE TRANSVAAL MAP-BIPLOT. The map-biplot of Figure 1 displays both the 36 stations around Nelspruit and the 32 summers' precipitation residuals. The stations are plotted at their map locations; target stations (in the northeastern quadrant) are labeled T, whereas control stations (mostly in the western quadrants) are labeled C. The summers are plotted at the locations of the vector r[j] of the spatial trend of their precipitation residuals. Seeded summers are labeled SD, whereas unseeded summers are indicated by asterisks.

APPARENT SEEDING EFFECTS IN THE TRANSVAAL. The goodness of fit of this map-biplot is only 25.5%, but it shows appreciable regularity of precipitation residuals. The markers for the 22 unseeded summers are spread out in a roughly elliptical scatter which is centered close to the origin. The orientation of the ellipse suggests that when summer precipitation patterns deviate from the average -- recall that this is a plot of residuals -- these deviations are mostly in the form of an unusual SW-to-NE precipitation trend -- increasing in that direction in some summers, decreasing in others.

The distribution of the markers for the 10 seeded summers is quite distinct: Five of them are at the top and right of the plot, showing unusually strong SW-to-NE trends, one shows an unusually strong Northward trend, and the remaining four show weak trends to the SW or South. What is striking is that the first six markers are extreme in their directions; only two or three of the 22 unseeded summers are near them. Since these unusual slopes are in the direction of the target relative to the control -- to the NorthEast -- this strongly suggests that a positive seeding effect has occurred. It also suggests that this effect may have occurred only on some of the summers, without any apparent effects during the other summers. (It would be intriguing to discover possible explanations for this apparent hiatus, but that requires more than the present data base.)

THE PRINCIPAL COMPONENT BIPLOT. The pc-biplot also displays markers for the stations and markers for the summers, but these are constructed differently from those of the map-biplot. Its station markers a[1],...,a[n] are at their "principal coordinates", which are their locations on a statistical map that reflects their weather patterns. Its summer markers b[1],...,b[m] indicate the precipitation residuals' trend with respect to that statistical map rather than with respect to the geographical map. Clearly, if weather patterns are simply linear functions of latitudes and longitudes, then the principal coordinates will correspond to map coordinates and the pc-biplot will be essentially the same as the map-biplot. But if weather patterns are related to geography in a more complex manner, the pc-biplot will describe them better, i.e., will fit the weather data more closely. (This form of biplot was originally proposed by Gabriel, 1971 -- see also Gabriel 1981a,b -- and its application to meteorology discussed in Gabriel, 1972, 1980 and 1985a,b. See the Appendix for a mathematical discussion.)

Fitting a pc-biplot requires, essentially, that individual stations' markers be moved around on the map so as to improve the regressions of precipitations on station coordinates. If the pc-biplot fits the data much better than the map-biplot, it becomes of interest to trace the displacement of stations from their map coordinates c[i] to their principal coordinates a[i]; this expresses the way weather patterns deviate from simple linear dependence on geography. In particular, the position of target stations relative to control stations may indicate the way seeding has modified the weather.

THE TRANSVAAL PC-BIPLOT. Figure 2 is the pc-biplot of these data. Its goodness of fit is 52.1%, appreciably more than that of the map-biplot. Control station a[] markers are found to cluster much more tightly than r[] coordinates on the map-biplot, but their separation from target a[] markers is no less evident; indeed, there is hardly any overlap between the two subsets. Apparently, precipitation residuals had much the same patterns at all control stations, but these were quite different from the patterns at target stations. The trends for seeded summers differed very clearly from those for unseeeded summers; eight of the latter ten were clearly on the periphery of the biplot, mostly in the same direction as the target station markers.

MORE ON THE APPARENT EFFECTS OF SEEDING IN THE TRANSVAAL. The evidence of the pc-biplot of Figure 2 strengthens the impression gained from the map-biplot of Figure 1, that several seeded summers had appreciably higher target precipitation than unseeded summers. In addition to the six summers noted for this effect in Figure 1. two more of the seeded summers are seen to have had unusual precipitation patterns of a different kind: could it be that seeding occasionally had local effects on part of the target area?

*Fig. 2. Transvaal Summer Rainfall: pc biplot. Station markers: C control, T target. Summer markers: * - unseeded, SD - seeded.*

More detail might be gleaned from a comparison of the geographical map of longitudes and latitudes with the statistical map of principal coordinates. The displacement of the stations from the former's r[i]'s to the latter's a[i]'s is indicated on Figure 3. Again, what is most evident is the crowding together of the control stations. The displacements of the target stations is less uniform, though the tendency to move away from the control stations is pretty much in evidence.

CENTRAL ILLINOIS SUMMER PRECIPITATION AND THE POSSIBLE INFLUENCE OF THE ST. LOUIS URBAN AREA. There has been some evidence of the effect of urban areas and industrial concentrations on downwind weather. Studies of weather events in and around St. Louis have suggested that the urban area might have influenced certain precipitation related patterns. (Changnon, 1976; Changnon, Huff, Schickendanz, and Vogel, 1977; Changnon, Semonin, Auer, Braham, and Hales, 1981; Detwiller and Changnon, 1976; Vogel and Huff, 1978). We therefore thought it worth while to use our proposed techniques to check whether St. Louis might have affected precipitation downwind of the city. We used 1901-80 summer precipitation data for 29 stations forming a West-East swath of 29 stations across central Illinois, from St. Louis to the Indiana border -- Figure 4a. We thought that if there were any effect, it would be likely to have produced increasing rainfall near St. Louis, at least until the 40ies or 50ies when pollution controls might have dampened such an urban effect. Thus, there should have been a time trend of increasing East-to-West (or, in accordance with the prevailing wind directions, NE-to-SW) rainfall gradients during the first four or five decades of the 20th century.

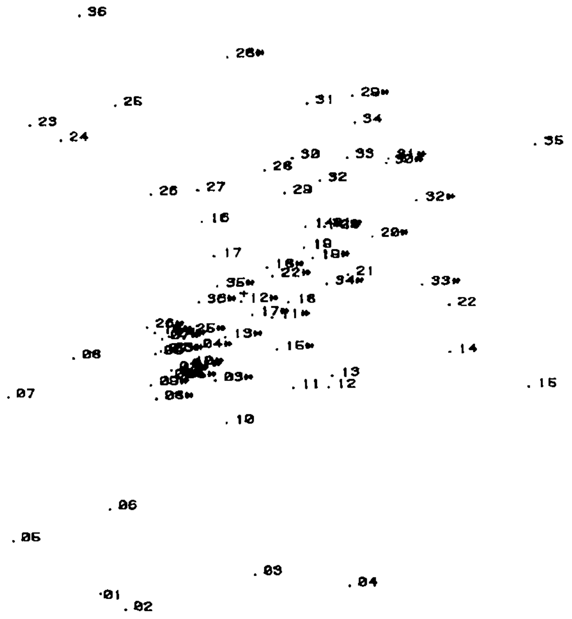

Fig. 3. Transvaal Summer Rainfall: Map biplot station markers (labeled by numbers) and pc-biplot station markers (labeled by numbers and asterisk).

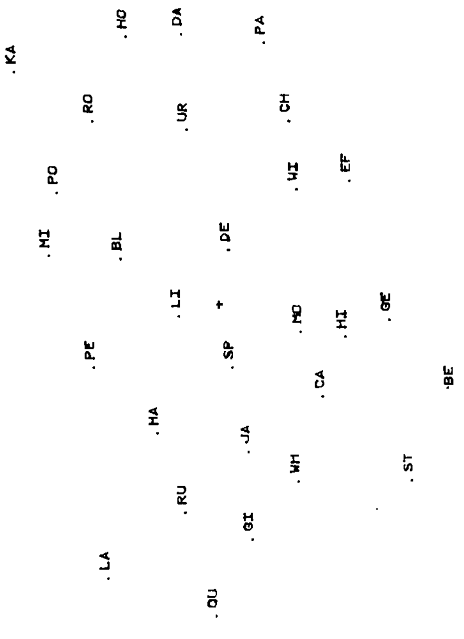

Fig. 4a. Central Illinois Summer Rainfall: Map-biplot station markers (labeled by letters).

Biplot display of these data for central Illinois required markers for each of 29 stations and for each of 80 summers. Joint display on one page was much too crowded, so the c[i] markers for the stations are displayed in Figure 4a -- which is simply a map of the stations labeled by the first two letters of their names -- and the r[j] markers for the summers are shown separately in Figure 4b. These two displays together constitute the map-biplot of the precipitation residual data for central Illinois, whose goodness of fit was 34.8%. Similarly, Figures 5a and 5b show, respectively, the a[i] station markers and the b[j] summer markers of the pc-biplot of the same data. That biplot had a goodness of fit of 38.2%.

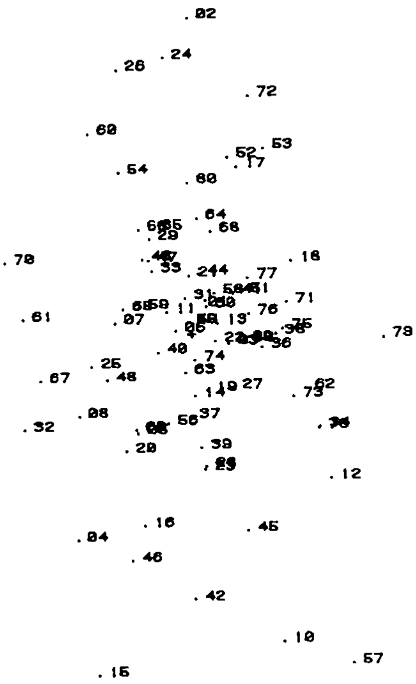

Fig. 4b. Central Illinois Summer Rainfall: pc-biplot summer markers (labeled by years).

Figure 4b of the slopes of spatial precipitation patterns during the 80 summers is still rather cluttered. Most of the r[j] markers are clustered at the center and indicate the prevalence of "average" patterns. The unusual summer patterns are easier to discern at the edges of this cloud of markers. There are SW-to-NE slopes during the summers of 1912, 1952 and 1953 as well as 1972; opposite slopes during 1904, 1911, 1920, 1932 and 1967; strong SE-to-NW slopes during 1960 and 1970 and an unusually strong NW-to-SE slope in 1957. No time trend is evident, nor do standard time series methods reveal such trends.

Turning next to the pc-biplot, one notices first that its goodness of fit is very little better than that of the map-biplot. Evidently, there is little difference between the two biplots. Indeed, the statistical map of stations by their principal coordinates -- Figure 5a -- corresponds very closely to the geographical map of the stations by longitudes and latitudes -- Figure 4a. No distortions of any magnitude are evident. The plots of spatial slopes -- Figures 4b and 5b -- are also very similar.

Judging from these displays, there is no evidence of an urban effect of St. Louis on summer precipitation in central Illinois. There are neither clear clear time trends in the spatial patterns of precipitation, nor is there any sizable distortion of the geographic pattern, which might have indicated local effects. Of course, this negative finding is not proof of the lack of any effects at all; indeed such effects have been traced within a much shorter distance of the city than that covered by the sparse station network whose records have been used here (Changnon et al, 1981).

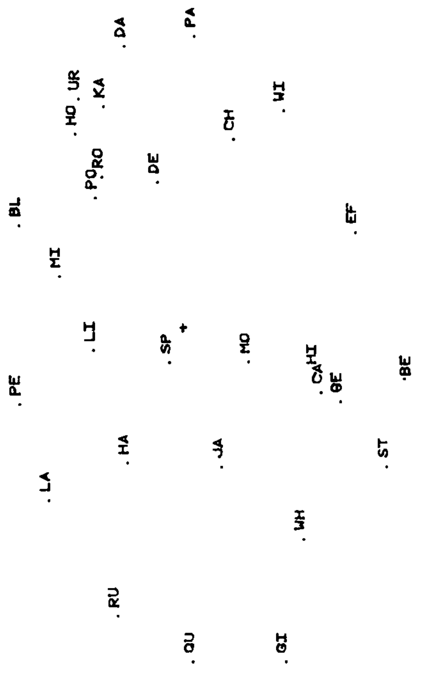

Fig. 5a. Central Illinois Summer Rainfall: pc-biplot (pc-biplot station markers labeled by letters).

Fig. 5b. Central Illinois Summer Rainfall: pc-biplot (pc-biplot summer markers labeled by years).

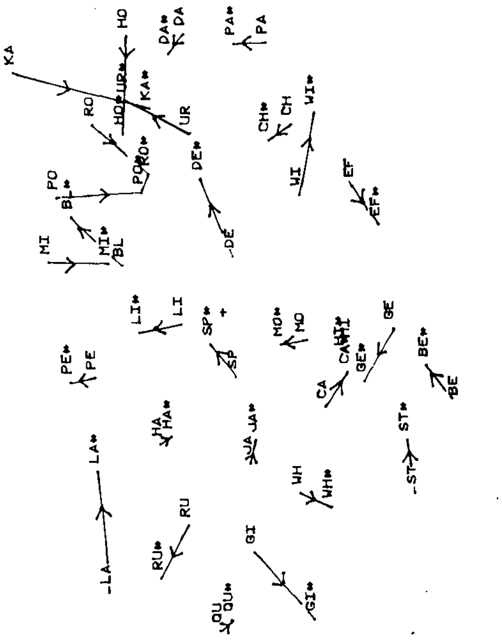

Fig. 6. Central Illinois Summer Rainfall: Map-biplot station markers (labeled by letters) and pc-biplot station markers (labeled by letters and asterisk).

POSSIBLE FURTHER ANALYSES. The two data sets used in this paper have served to illustrate the proposed technique, rather than to present definitive findings about intended and inadvertent weather modification. Indeed, more complete analyses of both these instances have been published elsewhere. However, previous analyses of such data have been incomplete in that they did not take into account effects on changing distribution over the target area. The technique we propose can take these into account to some extent and may thus provide additional information. Of course, in a real application one would further supplement these analyses by examining the residuals from the two biplots for possible special effects.

ACKNOWLEDGMENT. This research was supported by National Science Foundation Grant ATM-8314063 awarded to the University of Rochester (K.R.Gabriel, Principal Investigator).

REFERENCES

Changnon, S.A.,Jr., 1976. Inadvertent weather modification. Water Resources Bulletin 12:695-718.

Changnon, S.A.,Jr., Huff, F.A., Schickendanz, P.T., and Vogel, J.L., 1977. SUMMARY OF METROMEX, Volume 1: Weather Anomalies And Impacts. Champaign: Illinois State Water Survey Bulletin, 62, 260 pp.

Changnon, S.A.,Jr., Semonin, R.G., Auer, A.H., Braham, R.R.,Jr., and Hales, J.M., 1981. METROMEX: A Review And Summary. Meterorological Monographs, Vol.18, No. 40. Boston: Amer. Meteor. Soc., 181 pp.

Detwiller, J., and Changnon, S.A.Jr., 1976. Possible urban effects on maximum daily rainfall at Paris, St.Louis, and Chicago. J. Appl. Meteor. 15:517-519

Eckart, C. ad Young, G., 1936. The approximation of one matrix by another of lower rank. Psychometrika 1:211-218.

Gabriel, K.R., 1971. The biplot - graphic display of matrices with application to principal component analysis. Biometrika 58:453-467.

Gabriel, K.R., 1972. Analysis of meteorological data by means of canonical decomposition ad biplots. J. Appl. Meteor. 11:1071-1077.

Gabriel, K.R., 1980. Comments on the re-analyses of the Santa Barbara II cloud seeding experiments. In Statistical Analysis Of Weather Modification Experiments (E.J.Wegman and D.J.DePriest, eds.). New York, Dekker, 113-129.

Gabriel, K.R., 1981a. Biplot. In Encyclopedia Of Statistical Sciences, Vol. I (S.Kotz, N.L.Johnson and C.Read, eds.). New York, Wiley, 262-265.

Gabriel, K.R., 1981b. Biplot display of multivariate matrices for inspection of data and diagnosis. In Interpreting Multivariate Data (V.Barnett, ed.). London, Wiley, 147-173.

Gabriel, K.R., 1985a. Exploratory multivariate analysis of a single batch of data. In Probability, Statistics And Decision Making In The Atmospheric Sciences (A.H.Murphy and R.W.Katz, eds.). Boulder, Westview, 101-152.

Gabriel, K.R., 1985b. Multivariate comparisons of data from several batches. In Probability, Statistics And Decision Making In The Atmospheric Sciences (A.H.Murphy and R.W.Katz, eds.). Boulder, Westview, 153-187.

Gabriel, K.R. and Mather, G.K., 1986. Exploratory data analysis of 1951-1982 summer rainfall around Nelspruit, Transvaal and of possible effects of 1972-1981 cloud seeding. Submitted for publication.

Householder, A.S. and Young, G., 1938. Matrix approximation and latent roots. Am. Math. Mon. 45:165-171.

Kempton, R.A., 1984. The use of biplots in interpreting variety by environment interactions. J. Agric. Sc., Cambridge 103:123-135.

Vogel, J.L., and Huff, F.A., 1978. Relation between the St.Louis urban precipitation anomaly and synoptic weather factors. J. Appl. Meteor. 17:1141-1152.

"REVIEWED"

APPENDIX. MATHEMATICAL PROPERTIES OF THE BIPLOTS

For each of n stations, we are given an m-variate data vector y[i] and a 2-element map coordinate vector c[i], i=1,...,n. Let these vectors be arrayed in, respectively, an n-by-m matrix Y of data and an n-by-2 matrix C of geographical coordinates, and let both of these matrices be centered so that each column sums to zero.

Consider y[j], the j-th column of Y, and compute its regression onto the map coordinates. The resulting coefficients are well known to be

$$r[j] = (\mathrm{inv}(C'C))C'y[j] .$$

Calculating these for all columns and arraying them as rows of m-by-2 matrix R, we can write

$$R' = (\mathrm{inv}(C'C))C'Y .$$

The regression fit of data matrix Y then is

$$Y^{\wedge} = CR',$$

and its goodness of fit is calculated as the ratio of the sum of squares of the elements of Y^ to the sum of squares of the elements of Y.

Rewriting the regression fit for each element y^[i,j] of Y^, we obtain

$$y^{\wedge}[i,j] = <c[i], r[j]> ,$$

where the right hand side denotes the inner product (sum of products of corresponding elements) of vectors c[i] and r[j]. Now, the map-biplot of Y displays both the c[i]'s and the r[j]'s and it can therefore be interpreted in terms of inner products. That actually is the defining property of all kinds of biplots (Gabriel, 1981a).

A principal component (pc) biplot of a matrix Y displays row and column markers a[i], i=1,...,n, and b[j], j=1,..,m, whose inner products

$$y^{\sim}[i,j] = <a[i], b[j]> ,$$

represent the elements y~[i,j] of the matrix Y~ of rank 2 which approximates Y better than any other rank 2 matrix. This approximation is obtained by principal component analysis, a technique based on the singular value decomposition (SVD)

$$Y = P'LQ,$$

which is computed so that the above three factors satisfy

$$PP' = QQ' = I[r],$$

and L is a positive diagonal r-by-r matrix whose diagonal elements are in descending order. The SVD is used to define n-by-2 matrix G which consists of the first two columns of the P' factor, and m-by-2 matrix H that consists of the first two columns of the product Q'L. The rank 2 approximation can then be obtained as the product Y~=GH' (Eckart and Young, 1936; Householder and Young, 1938), and the goodness of fit of Y~ to Y measured by the ratio of the sums of squares of their Elements.

Choice Of The Matrix T Allows Some Non-Uniqueness Of The Pc-Biplot, Though It Always Satisfies The Above Inner Product Representation Of The Rank 2 Approximation Of Y. In Its Original Presentation (Gabriel, 1971) T Was Chosen To Be The Identity Matrix, Since That Also Allowed Standard Deviations And Correlations Of Column Variables To be represented by lengths and angles of the b[j]=h[j] vector markers. In the present context, the choice of

$$T = C'G$$

is made since it minimizes the sum of squared differences between A=GT' and C. In other words, that is the form of the pc-biplot for which the scatter of the rows a[i], i=1,...,n, of AT' is as similar as possible to the map scatter of the stations at locations c[1],...,c[n].