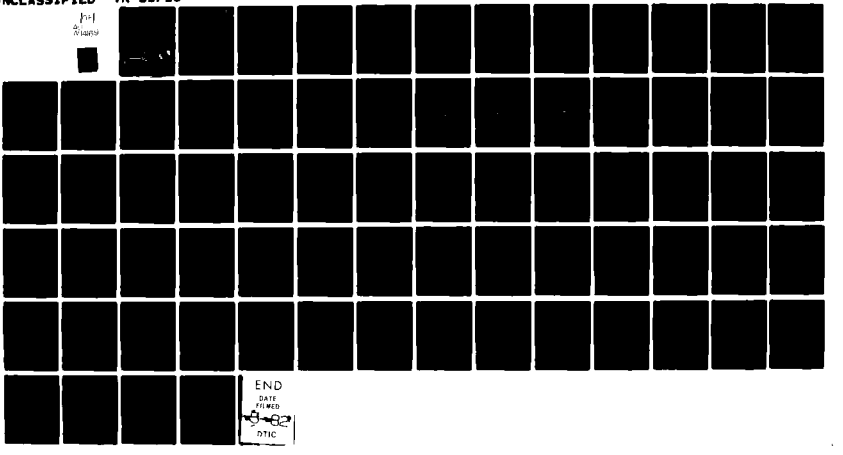
AD-A114 169 ROCHESTER UNIV NY F/G 4/2
MODELING TEMPERATURE DATA: AN ILLUSTRATION OF THE USE OF BIPLOT--ETC(U)
SEP 81 M C TSIANCO, K R GABRIEL N00014-80-C-0387
UNCLASSIFIED TR-81/15 NL



END
DATE FILMED
8-82
DTIC

MODELING TEMPERATURE DATA:
AN ILLUSTRATION OF THE USE OF BIPLOTS AND BIMODELS IN NON-LINEAR MODELING

by

MICHAEL C. TSIANCO*

and

K. RUBEN GABRIEL

September 1981

TECHNICAL REPORT 81/15

Department of Statistics and Division of Biostatistics
University of Rochester, Rochester, NY 14627

*This work is part of Michael C. Tsianco's Ph.D. Dissertation in Statistics at the University of Rochester under the supervision of K. Ruben Gabriel. It was supported in part through ONR Contract 042-442 on "Biplot Multivariate Graphics" of which K. Ruben Gabriel is Principal Investigator

1. Introduction

This paper applies a statistical strategy to the exploration of multivariate meteorological data, the diagnosis of suitable models and their fitting and the examination of fits and residuals. It shows that the inspection of biplots and bimodel graphical displays - in 2 and 3 dimensions, respectively - can lead to formulation of harmonic models that approximate actual observations closely. Both the parameters of the models and the magnitudes of certain residuals are found to allow physically meaningful interpretations. Though some of the findings may be of meteorological interest, the thrust of this paper is methodological. A straightforward example has been chosen in order to illustrate how the investigators' interaction with the data, and their hunches about possible patterns, can play a role within this statistical strategy of exploratory data analysis.

The principal tool used for data exploration is the biplot display of a data matrix (or its three-dimensional analogue - the bimodel). This is a descriptive tool which can be used to guide the investigator in examining complex data and in suggesting mathematical models to fit it. The proposed strategy is avowedly exploratory; since it does not presuppose any patterns or hypotheses but infers these from the data themselves, the strategy cannot include tests of significance; such tests are valid only for hypotheses formulated in advance.

The reader who is not acquainted with the biplot may wish to refer to Gabriel (1972) or to Gabriel, Rave and Weber (1976)

for a brief practical introduction and to Bradu and Gabriel (1978) for a more mathematical discussion (other references are Corsten and Gabriel, 1976; Gabriel 1978a,b, 1981a,b,c,d, Gabriel and Zamir, 1979; Haber, 1975; Kester, 1979; Tsianco, 1980). Computer programs for calculating biplot coordinates and providing a printer plot are available from the Division of Biostatistics, University of Rochester Medical Center, Rochester, NY 14642. Program BGRAPH for graphical display of three and higher dimensional bimodels (which produced the displays in this text) is also available from the authors Tsianco et al, 1981).

2. The Temperature Data

The data consist of monthly mean temperatures at each of 50 weather stations in North, Central and South America - See the map in Figure 1 and the list in Appendix Table A - for all 12 months of 1951 and 1952 - Appendix Table B. For a preliminary view of the data, we present a three-way analysis of variance - Table 1 - and a sequence of possible linear models with their numbers of parameters, residual sums of squares and residual mean squares (RMS) - Figure 2. It is evident that there are strong effects of stations and months as well as interactions between these factors. A model with stations, months and station x month effects indeed fits quite well with RMS=91. The "effect" of the years' difference is more equivocal, though clearly not negligible. This is evidenced by the slightly better fit - RMS=76 - of the model which further incorporates year x month interactions.

- Figures 1, 2 and Table 1 -

Figure 1: Map of the weather stations

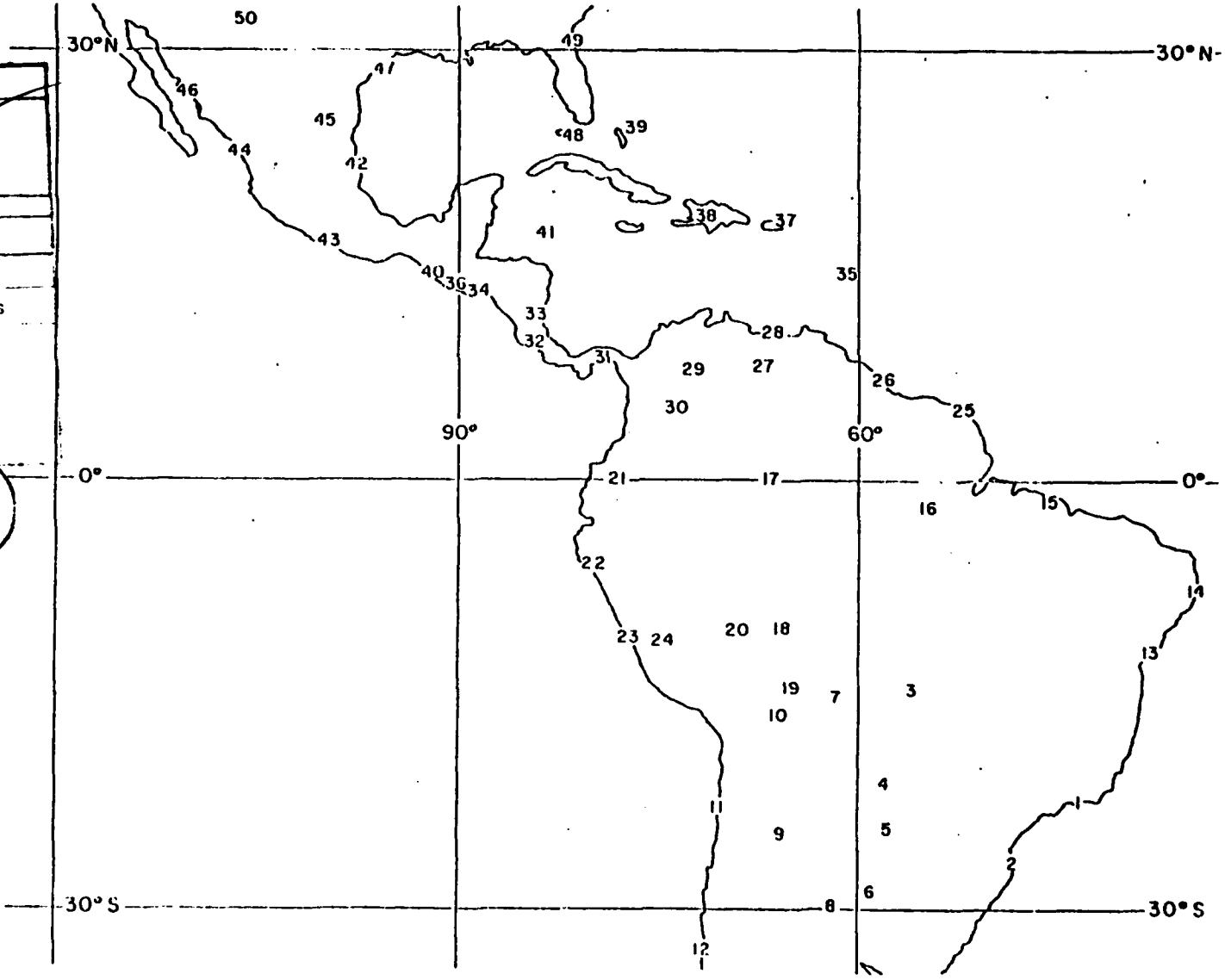


Figure 2: Diagram of linear models (model, number of parameters, RSS/df = RMS)

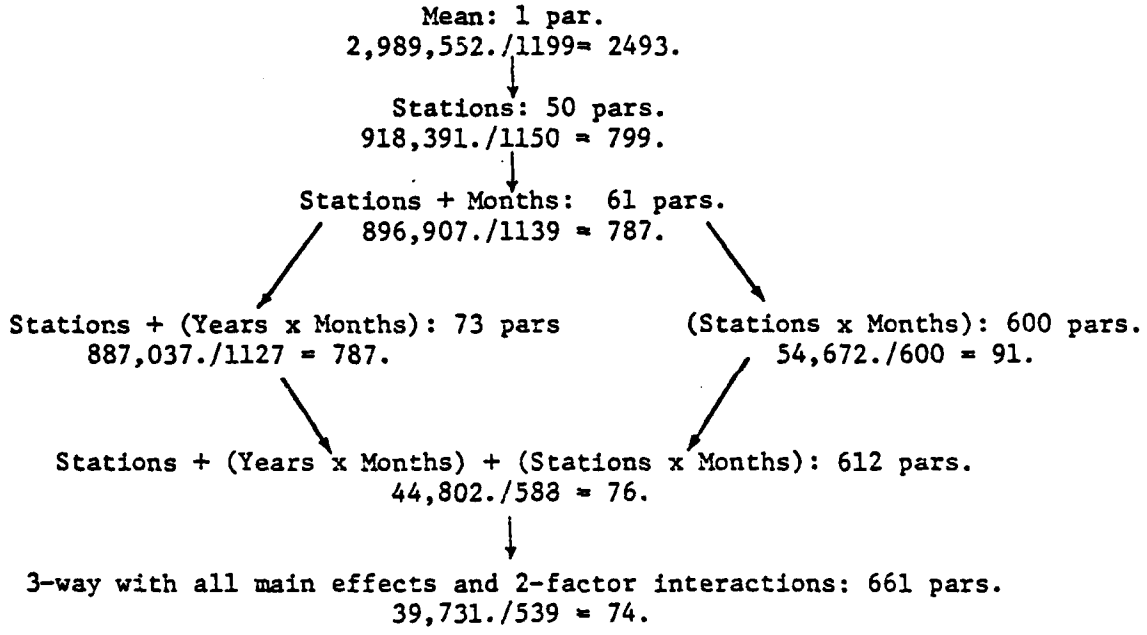


Table 1: Three factor ANOVA of monthly average temperatures (C°x10)

Source of variation	df	SS	MS	F
Stations	49	2,071,161.	42,268.6	573.5
Years	1	3.	3.0	0.0
Months	11	21,484.	1,953.1	26.5
Stations x Years	49	5,071.	103.5	1.4
Stations x Months	539	842,235.	1,562.6	21.2
Years x Months	11	9,867.	897.0	12.2
Error	539	39,731.	73.7	
Total (c.f.m.)	1199	2,989,552.		

The above models provide baselines against which the fits of other models will be judged. An RMS of, say, between 70 and 130, will be considered to show quite a good fit. Of course, we would prefer a model with rather fewer parameters than the 600-odd used above - even if at the expense of a little higher RMS - especially if the residuals would not reveal any left-over structure.

3. The Biplot - Inspection of the Data

We start graphical exploration of these data with the matrix of deviations

$$Y_{(50x24)} = (\text{Temperatures}) - \underline{1}_{50}\ (\underline{\text{Means}})'. \qquad (1)$$

This was obtained from the 50 stations-by-24 months matrix of temperatures -Appendix Table B - by subtracting out the 24 column means, i.e., the months' temperatures averaged over the 50 stations. To allow graphical display, the matrix Y was then approximated by rank 2 matrix

$$Y_{[2]} = G_{[2]}\ H'_{[2]}, \qquad (2)$$

where the right hand side of (2) is a factorization into a (50 x 2) matrix $G_{[2]}$ subject to $G'_{[2]}\ G_{[2]} = I_2$ and a (24 x 2) matrix $H_{[2]}$. (Reduced-rank approximation is due to Householder and Young (1938) and its use for graphical display was described by Gabriel (1971). The goodness of fit of this rank 2 approximation is

$$1 - ||Y - Y_{[2]}||^2 / ||Y||^2 = 0.961.$$

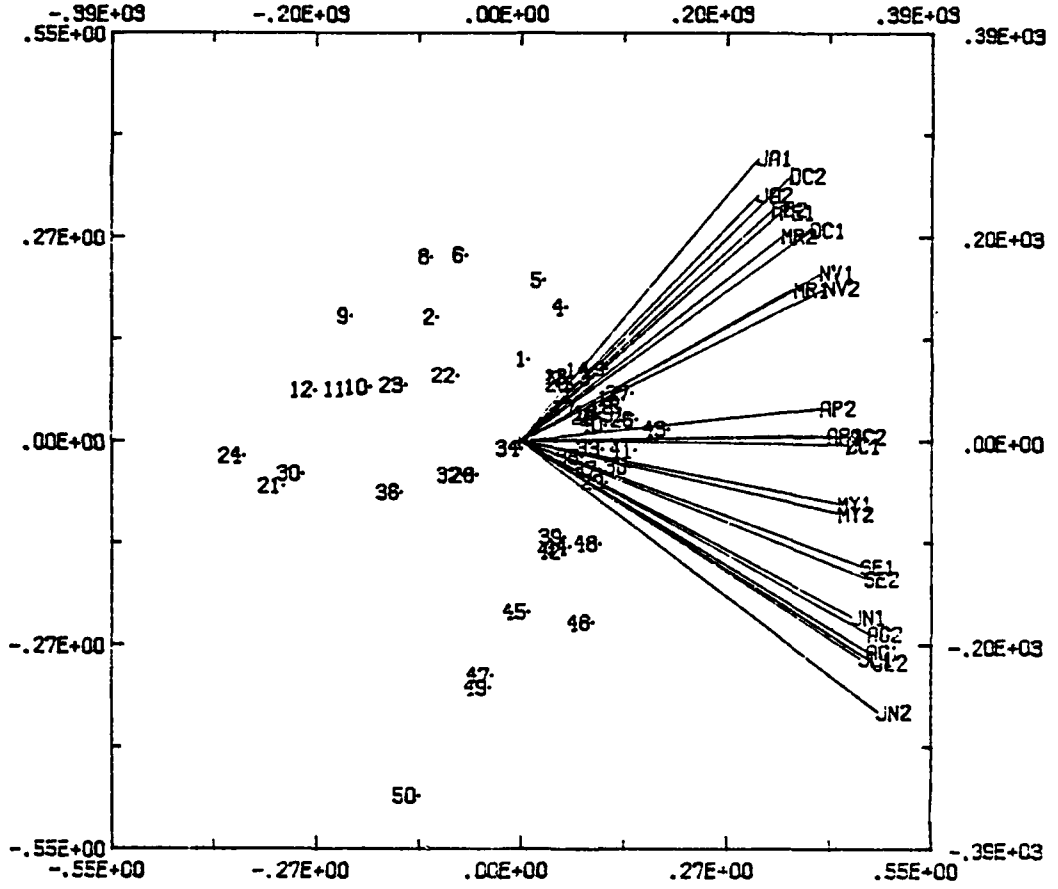
Approximating Y by $G_{[2]} H'_{[2]}$ is equivalent to fitting the temperature matrix by $\underline{1}_{50}\underline{t}' + G_{[2]} H'_{[2]}$, where $\underline{t}$ is the (24 x 1) vector of monthly means. Thus, 24 means $\bar{t}_{(j)}$, 100 $g_{i,k}$'s and 48 $h_{j,k}$'s are used, subject to 6 constraints ($\Sigma_i g_{i,k} = 0$, k=1,2; $\Sigma_i g^2_{i,k} = 1$, k=1,2; $\Sigma_i g_{i,1} g_{i,2} = \Sigma_j h_{j,1} h_{j,2} = 0$). This leaves 166 independent parameters and the residuals reduce to RMS = 115,574/(1200 - 166) = 112, where $||Y - Y_{[2]}||^2 = 115,574$ and 1200 is the number of entries in the matrix of temperatures. This compares quite well with the RMS of the linear models discussed above. The coordinates for the first five components of fit are shown in Appendix Table C (The biplot uses only the first two of these).

- Figure 3 -

To allow visual inspection of the data, the rank 2 approximation $G_{[2]} H'_{[2]}$ is displayed in the biplot of Figure 3. The month markers -- vectors $\underline{h}_{[2]j}$, j=1,...,24 -- are displayed as arrows; the station markers -- vectors $g_{[2]i}$, i=1,...,50 -- are shown as numbered points. This method of display and its inspection have been discussed by Gabriel (1971, 1972).

The first thing that strikes us in this biplot is the almost perfectly linear configuration of the months' $\underline{h}$-arrows. Furthermore, we note turn-of-the-year months to have $\underline{h}$-arrows pointing right and up, whereas mid-year months have $\underline{h}$-arrows pointing right and down. Other months have $\underline{h}$-arrows in between, in a regular annual sequence. Since small angles between $\underline{h}$-arrows represent high correlations, this pattern indicates that correlations are highest between nearby months and lowest

Figure 3: GH' - biplot of deviations from monthly means



between January and July (angles close to 90° represent zero correlation).

We next turn to the scatter of the fifty stations. No particular pattern is apparent, but we note the g-points to cluster pretty much according to geographical location. (This may be seen by relating the biplot of Figure 3 to the map of Figure 1). North American stations' g-points are in the lower half of the biplot; South American stations are generally in the top half; stations near the equator cluster in the middle of the display. Since biplot proximity expresses statistical similarity, this geographical clustering expresses the fact that stations in the same region generally have similar temperature profiles.

The relation between the h-arrows and the g-points further shows which stations have high or low temperatures in what months. Thus, we note that the bottom of the biplot has the g-points for North American stations and the h-arrows for mid-year months: This concurrence represents the fact that mid-year temperatures in North America are high. Similarly, the top of the biplot has South American stations' g-points and turn-of-the-year months' h-arrows: That represents the occurrence of the Southern hemisphere's summer at the turn of the year. Furthermore, North American g-points are generally opposite turn-of-the-year h-arrows and South American g-points opposite mid-year h-arrows: That represents the low temperatures in the two hemispheres' winters. By the same graphical criterion, the equatorial stations' g-points which are in the middle of the plot are not particularly far out, either in the direction of

any of the h-arrows or in the direction opposite such arrows. This represents the relatively steady annual temperature profile at equatorial stations -- none of the months are much hotter or colder than the others.

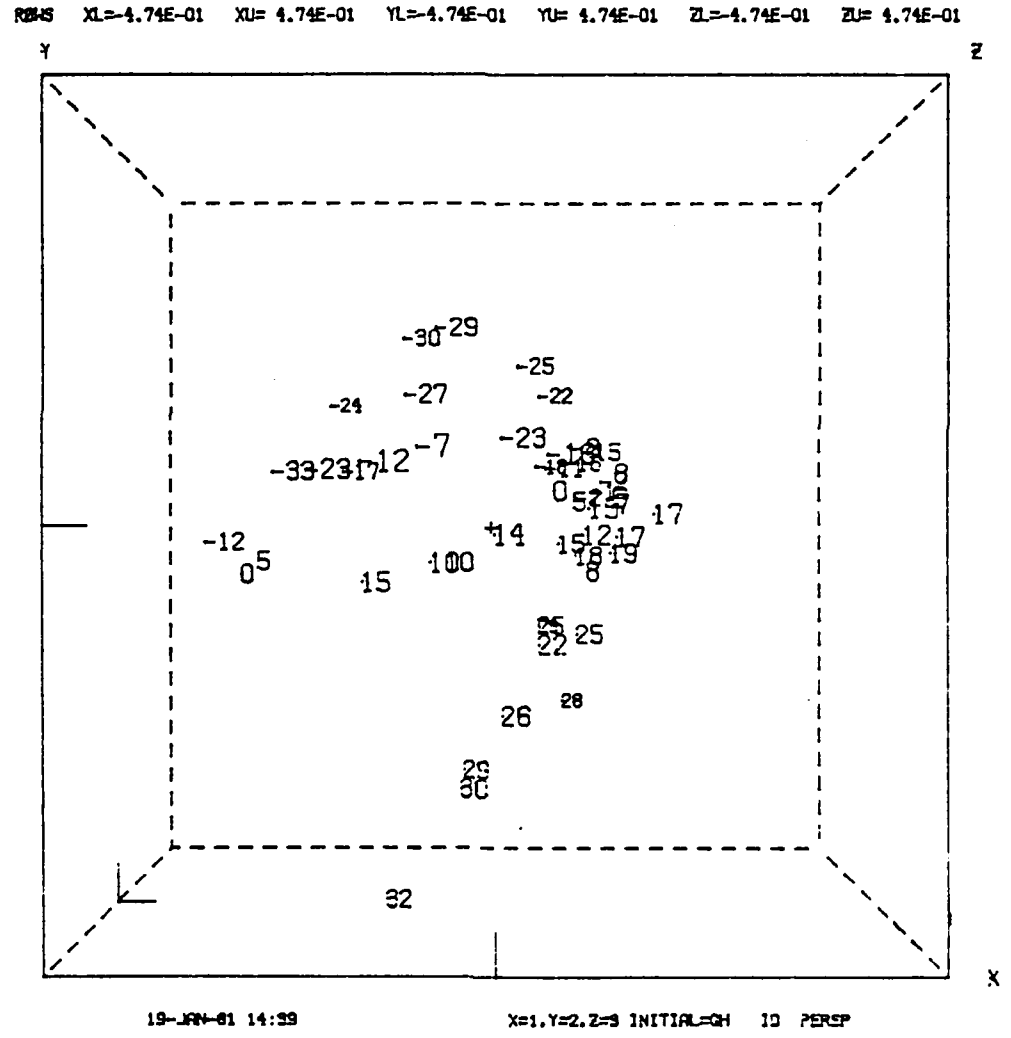
4. Further Exploration with the Biplot: Relation of Temperature to Latitude, Longitude and Altitude

After the initial general inspection of the temperature data biplot, we turn to correlating these patterns with concomitant variables which were not used in constructing the biplot. We consider data on latitudes, longitudes and altitudes of the 50 stations as shown in Appendix Table A. (The following displays are perspective views of the 3D bimodel. In these the large labels indicate proximity to the viewer in the third dimension, whereas small labels show markers distant from the viewer. The present discussion - Section 4 - is concerned only with the biplot, so the reader may ignore the third dimension in these displays until it is discussed in Section 5, below.)

- Figure 4 -

To display the relation of temperatures to latitudes we reproduce the biplot with the stations labelled by their latitudes - Figure 4 - Northern latitudes being indicated by positive numbers, Southern latitudes by negative numbers. It is immediately evident that the top of the biplot displays Southern stations at high latitudes and the bottom high latitude Northern stations. The equatorial stations -- low latitudes -- generally form a middle belt. The trend is very obvious, even though it does not account for all the variation on the biplot,

Figure 4: Perspective view of GH' - bimodel of deviations: Station markers labelled by latitudes (positive for Northern Hemisphere, negative fcr Southern)



especially that in the horizontal direction which evidently cannot be explained by latitude differences.

Another way of looking at this trend is by breaking the stations up into 4 latitude groups: (1) South of 20°S; (2) Between 20°S and 0°; (3) Between 0° and 20°N, and (4) North of 20°N. Concentration ellipsoids for each of these groups are shown in Figure 5. The downward South-North gradient is very apparent, with successive groups' ellipsoids being farther down on the biplot.

- Figures 5,6 and 7 -

The horizontal elongation of most of these ellipsoids again indicates the existence of additional variability that is not accounted for by latitude. Figure 5 also displays the individual g-points which are outside the concentration ellipsoids. Several g-points are well to the left of their respective groups' ellipsoids. They represent stations whose temperatures are generally well below those of most stations at those latitudes (The left on the biplot indicates low average temperatures since all h arrows point generally to the right).

Identification of the stations whose g-points are left-outliers (i.e., stations 24, 21, 30, 50, 36, 12, 11 and perhaps 9, 10, 23, 32 and 28) shows most of these to be at relatively high altitudes. Displaying the altitudes of the stations in the biplot of Figure 6 indeed shows most of these stations -- (with the exception of 12, 11, 23) to be at elevations of 1000 meters and more, whereas all 38 other stations are below 1000 meters. The right-to-left gradient of increasing altitude is very clear from Figure 6 and suggests that the two dimensional

Figure 5: Perspective view of GH' - bimodel with concentration ellipsoids (1.5 SD) for four groups of stations (Individual station markers shown if outside ellipsoids)

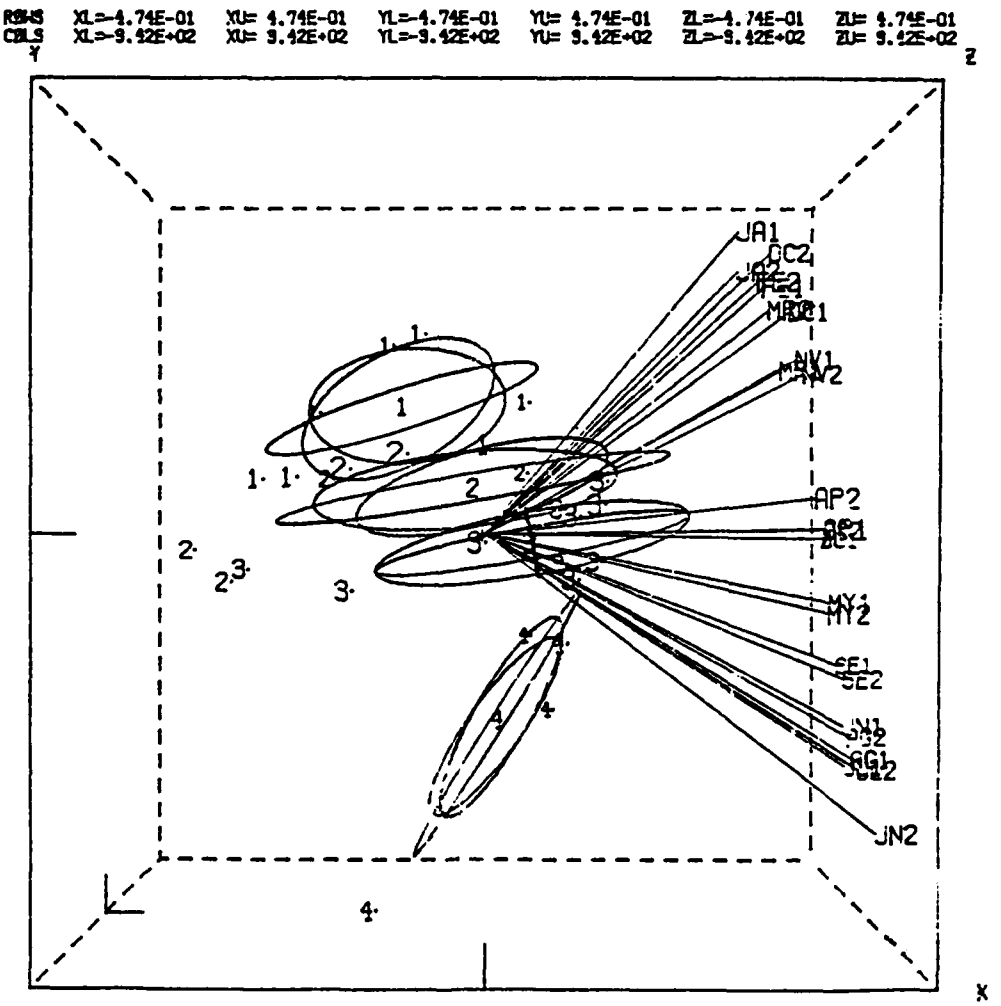


Figure 6: Perspective view of GH' - bimodel: Station markers labelled by altitudes (100 meters).

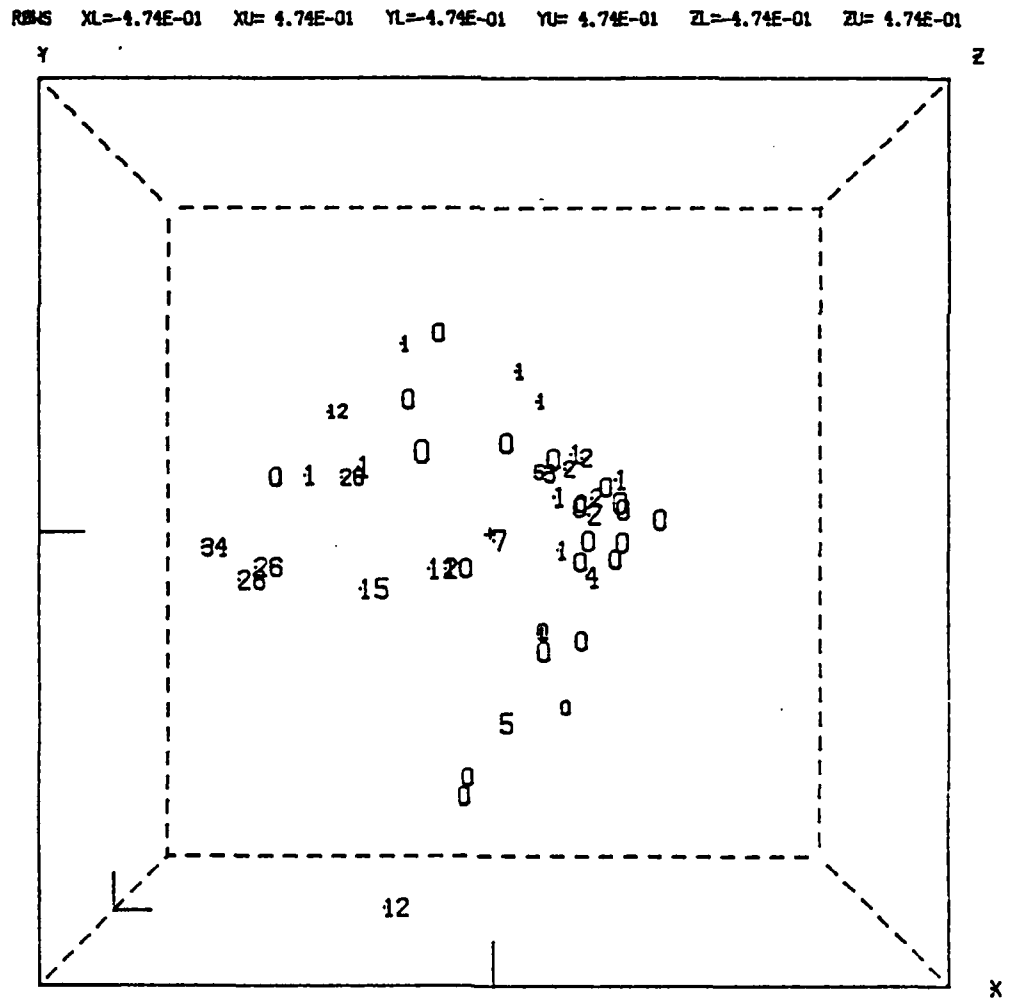
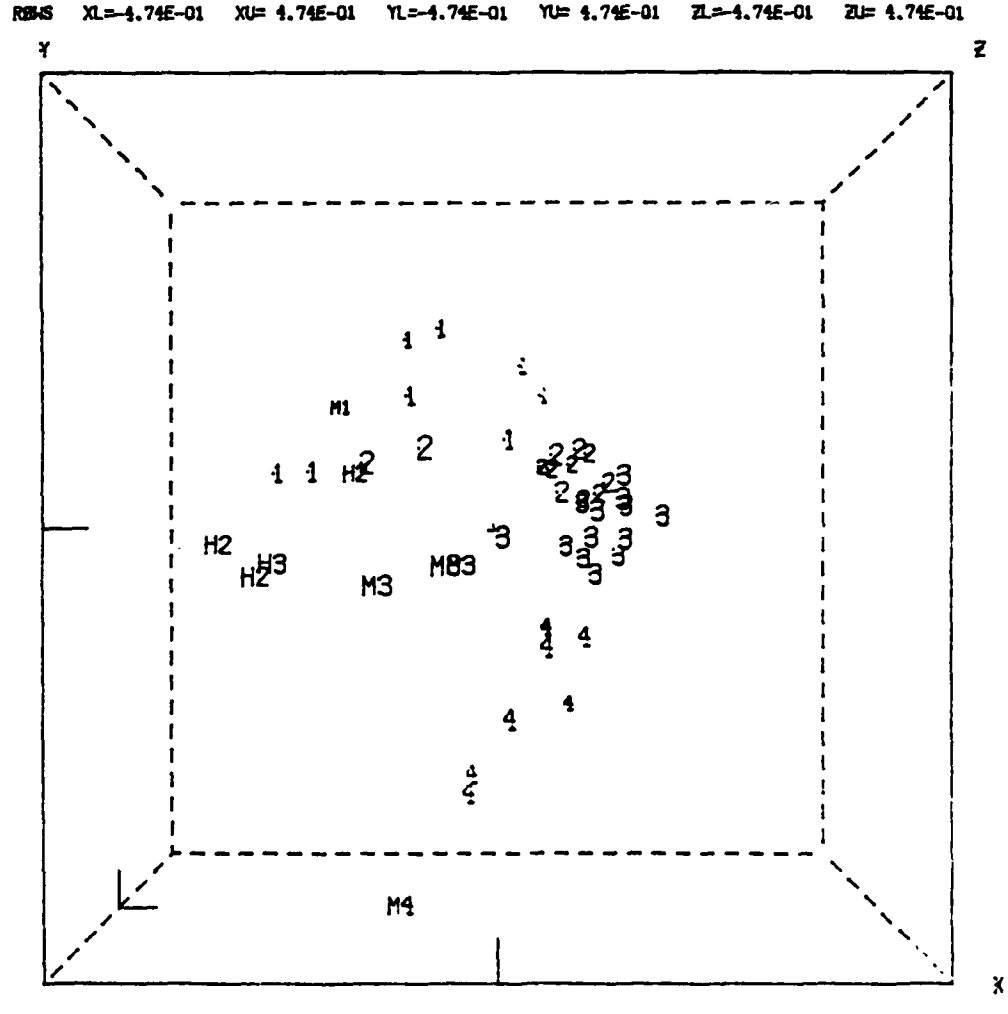


Figure 7: Perspective view of GH' - bimodel: Station markers labelled by altitude and latitude categories



g-point scatter is well accounted for by latitudes and altitudes. Evidently the stations' temperature profiles are strongly correlated with latitude and altitude.

To further check the dependence of temperature profiles on both altitude and latitude, we have provided double indices for the g-points in Figure 7. The letter indicates the altitude (H - > 2000 meters; M - 1000-2000 meters; no letter - < 1000 meters) and the digit the latitude (1,2,3,4 from South to North as in Figure 5). This display clearly shows the consistent latitude/altitude pattern of the g-scatter.

Figure 7 also shows that a small number of g-points of low altitude stations are much farther to the left of the biplot than expected from the general pattern. These stations - 12, 11, 23 and possibly 22, (which had been noticed before) -- were colder than expected from their location and elevation. Reference to the map - Figure 1 - shows these to be the stations on the Pacific Coast of South America. A likely explanation for the relative coolness of that coast is the cold Peru Current which flows northward along these shores.

Another group of stations that might be considered to be colder than warranted by their altitudes and latitudes are 47, 49 and 50, the three US stations.

5. The Three-Dimensional Bimodel: Identifying a Pattern for the Months

We now return to a detailed consideration of the h-arrows for the 24 months of 1951 and 1952. The striking near collinearity of the arrowheads would generally have suggested diagnosis

of a rows regression model (Bradu and Gabriel, 1978). This would have expressed each station's (i=1,...,50) temperatures $t_{i,1},\ldots,t_{i,24}$ as a linear function $t_{i,j} = \eta_i + \psi_i x_j$ of some constants $x_1,\ldots,x_{24}$ for the twenty-four months. It would not, however, have explicitly taken into account the cyclical nature of annual temperatures. And yet, on the biplot both years' $\underline{h}$-arrows followed a perfect order from January on top, through February, March, April, May and June to July at the bottom, and up again to December and January. This indicated that the above linear model should be replaced by one incorporating an annual cycle. It suggested to us that another dimension was necessary.

To inspect the temperature data in a higher dimension than the plane of the biplot, we fitted a rank three approximation

$$Y_{[3]} = G_{[3]} H'_{[3]} \qquad (3)$$

by least squares, using the Householder-Young method again. Our fit produced a (50x3) matrix $G_{[3]}$ whose first two columns equal those of $G_{[2]}$ and a (3x24) matrix $H'_{[3]}$ whose first two rows equal those of $H'_{[2]}$. Moreover, the method of fitting ensured that $G'_{[3]} G_{[3]} = I_3$. Goodness of fit was improved from 0.961 to 0.975 and the RMS reduced from 112 to 73 (now using 234 independent parameters).

The rank three approximation $Y_{[3]} = G_{[3]} H'_{[3]}$ provides a three-dimensional bimodel with rows $\underline{g}'_{[3]i}$ of $G_{[3]}$ serving as station markers and columns $\underline{h}_{[3]j}$ of $H'_{[3]}$ as month markers. A perspective view of this approximation is given in Figure 8 and

a stereo pair of perspective views in Figure 9. Since these may be difficult for the reader to inspect, we provide three orthogonal projections in Figures 3, 10a and 10b - Figure 3 is the $G_{[2]}H'_{[2]}$ biplot which displays the projections on the first two principal axes, whereas Figures 10a and 10b display projections on the first and third principal axes and the second and third principal axes, respectively.

- Figures 8, 9, 10a and 10b -

A clearer picture of the h-configuration now emerges. On Figure 10b the heads of the h-arrows are seen to be pretty close to an ellipse, which, in view of Figure 3, is noted to be on a plane almost perpendicular to the biplot. The h's for successive months are found to be arranged in a counterclockwise order around the center of the ellipse. Since the major axis of the ellipse is considerably longer than the minor axis, (which lies close to the third principal axis) the arrows for January and June are much farther apart than those for April and October. Furthermore, nearby months such as April and May, September and October, are close together. This pattern reflects the correlations between the months of both years, as described above.

Exploration of the third dimension of the bimodel is seen to add a great deal to our ability to model the seasonal temperature cycle over the months. The third axis seems to be one of Fall temperatures versus Spring temperatures. We would hope that inspection of the scatter of g-points in this direction might also add to our understanding of variation in temperature profiles between stations. However, all we are able

Figure 8: Perspective view of GH' - bimodel: Months and stations

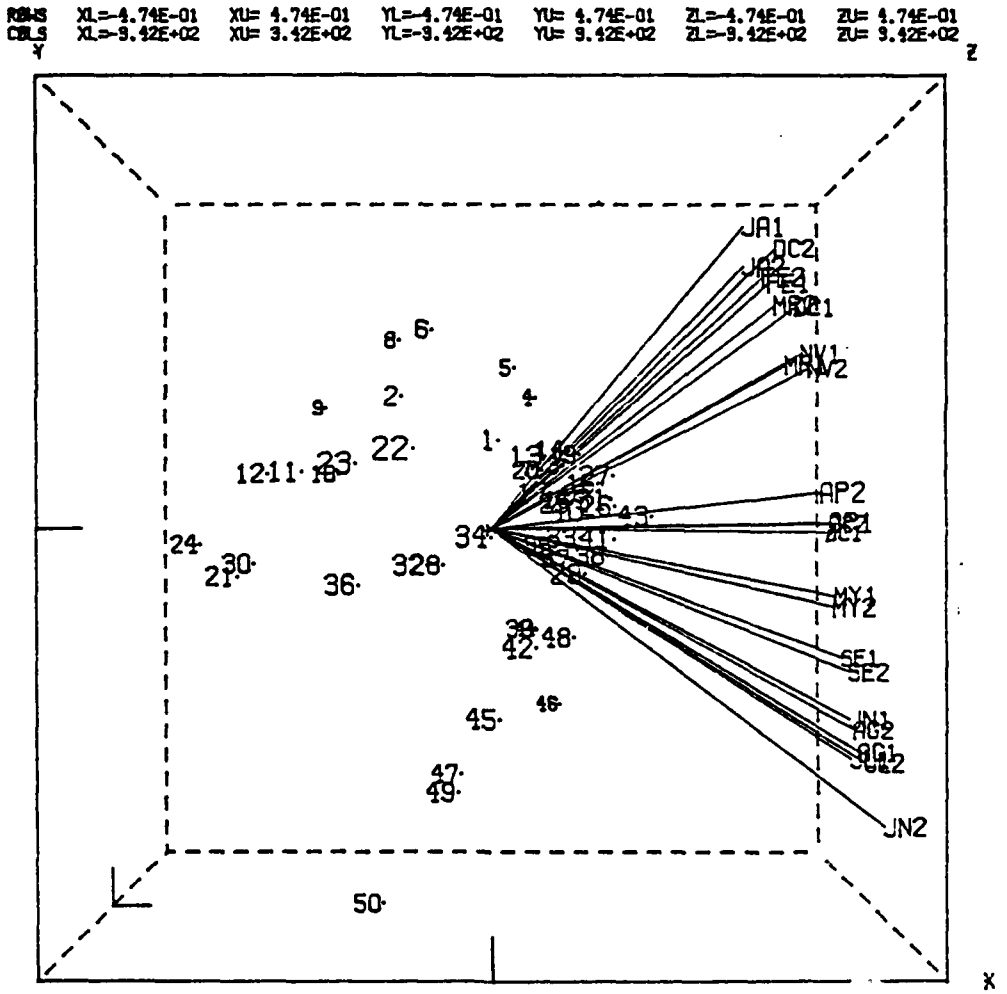


Figure 9: Stereoscopic display of GH' - bimodel

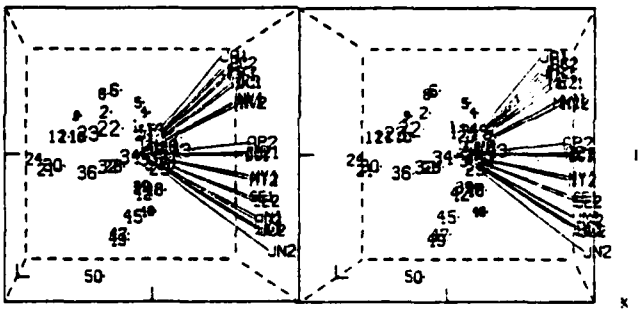


Figure 10a: GH' - bimodel: Plane of axes 1 and 3

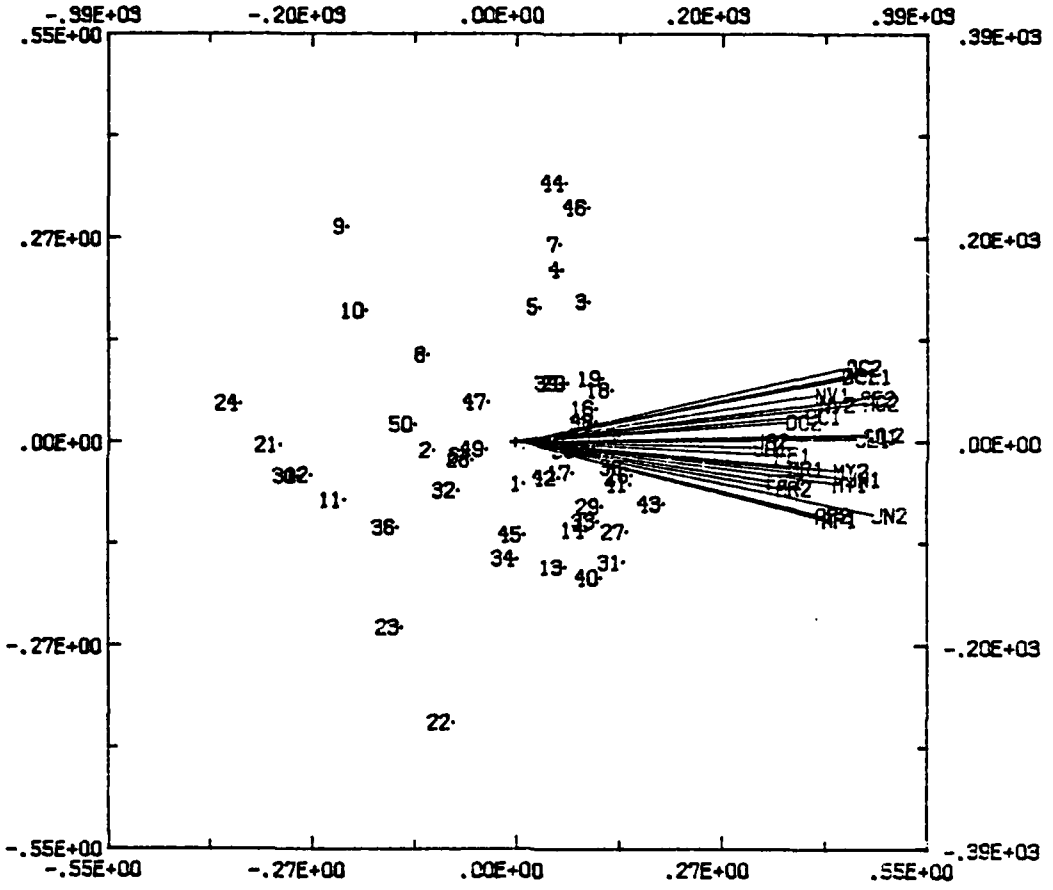
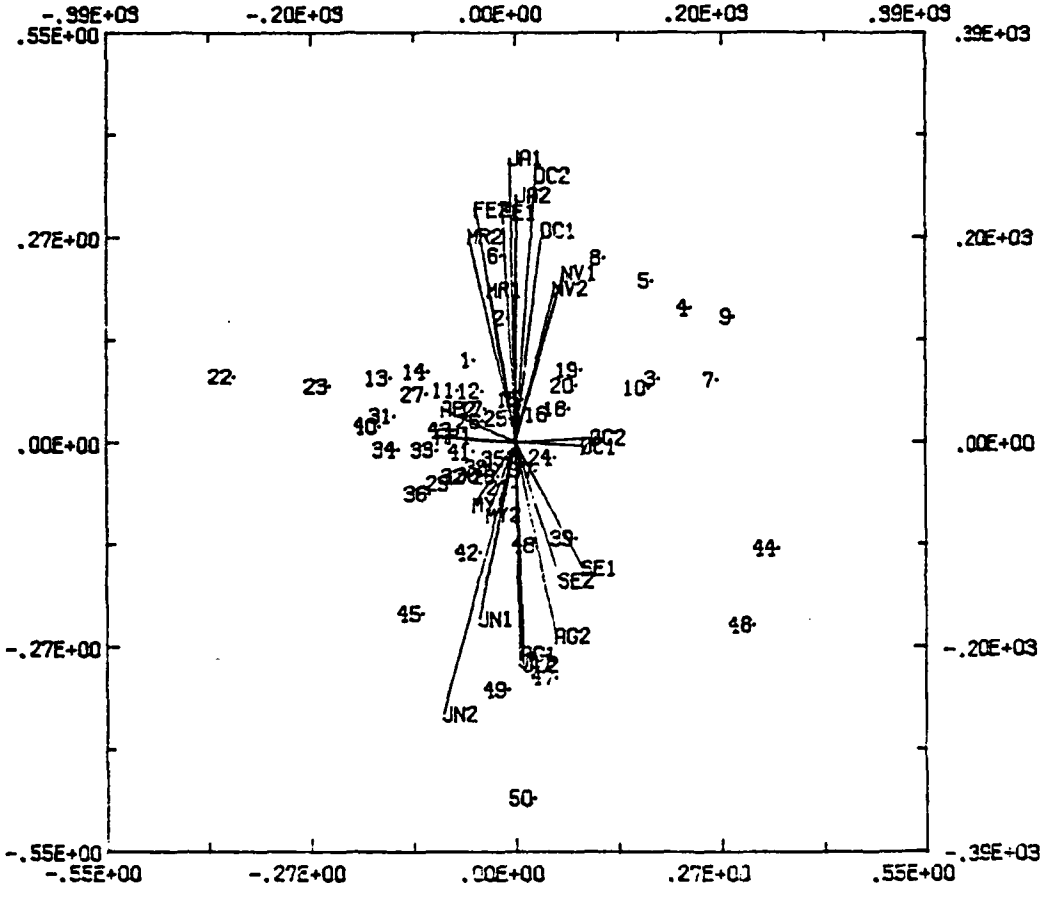


Figure 10b: GH' - bimodel: Plane of axes 3 and 2



to see from Figures 8, 9, 10a and 10b is that Fall temperatures are relatively high, and Spring temperatures low, at stations as diverse as 44 and 46 on the West Coast of Mexico and a cluster of stations in the interior of South America. On the other hand, Spring temperatures are relatively higher at stations 22 and 23 on the Peruvian coast. We do not at present see a physical explanation for that. Also we do note that inland South American stations 4, 5, 6, 8 and 9 seem to be rather extreme in their seasonal pattern, with apparently unusually hot turn-of-the-year temperatures.

6. Diagnosis and Fit of Models: First Stage

6.1 The decision to model the months first

We now use the insights obtained from inspecting the biplot and bimodel of the temperature data to formulate simplified models which we then fit to the data. We prefer to begin by modeling only the months' configuration since its elliptic shape was the single most striking pattern observed. We could, at this stage, have jointly modeled the stations' (rows') scatter and the months' (columns') configuration. Section 3 and 4, above, certainly give enough leads to begin with - but we prefer to proceed stage-wise, first fitting a partial model, examining its residuals and then proceeding to the next stage of fitting in which we also model the scatter of the stations -- Section 7, below.

6.2 Models for the months

The near elliptic configuration of the h-markers for the

months suggests that we try to fit the model

$$\underline{h} \doteq \underline{\lambda} + \underline{\gamma} \sin\theta + \underline{\delta} \cos\theta, \quad \theta\epsilon[0,2\pi], \tag{4}$$

where $\underline{\lambda}$, $\underline{\gamma}$ and $\underline{\delta}$ are three-dimensional vectors of, respectively, the location of the ellipse's center, its major axis and its minor axis, and $\underline{h}$ is a point on the ellipse's circumference corresponding to angle θ. For a particular month j, then

$$\underline{h}_j \doteq \underline{\lambda} + \underline{\gamma} \sin\theta_j + \underline{\delta} \cos\theta_j \ , \tag{5}$$

for some suitable angle θ_j, j = 1, ..., 24.

The above is a model for the $\underline{h}$-markers. It can be used to derive a model for the temperature deviations Y in the following way. To begin with, we have

$$\begin{aligned} Y &\doteq G_{[3]} \, H'_{[3]} \\ &\doteq G_{[3]} \, (\underline{\lambda},\underline{\gamma},\underline{\delta}) \begin{pmatrix} 1 & ,\ldots, & 1 \\ \sin\theta_1 & ,\ldots, & \sin\theta_{24} \\ \cos\theta_1 & ,\ldots, & \cos\theta_{24} \end{pmatrix} \\ &\doteq Y_{[M_\theta]} \ , \end{aligned}$$

where $[M_\theta]$ denotes this particular model. The model can be rewritten

$$Y_{[M_\theta]} = (G_{[3]}\underline{\lambda}, \, G_{[3]}\underline{\gamma}, \, G_{[3]}\underline{\delta}) \begin{pmatrix} 1 & ,\ldots, & 1 \\ \sin\theta_1 & ,\ldots, & \sin\theta_{24} \\ \cos\theta_1 & ,\ldots, & \cos\theta_{24} \end{pmatrix}. \tag{6}$$

If we relabel

$$G_{[3]}\underline{\lambda} = (\eta_1 \ ,\ldots, \ \eta_{50})' \tag{7a}$$

$$G_{[3]}\underline{\gamma} = (A_1 \ ,\ldots, \ A_{50})' \tag{7b}$$

and

$$G_{[3]}\underline{\delta} = (B_1 \ ,\ldots, \ B_{50})' \tag{7c}$$

we obtain

$$y_{[M_\theta]i,j} = \eta_i + A_i \sin\theta_j + B_i \cos\theta_j \ , \quad \begin{matrix} i=1,\ldots,50, \\ j=1,\ldots,24 \end{matrix} . \tag{8}$$

If we further reparametrize

$$\psi_i = \sqrt{(A_i^2 + B_i^2)} \tag{9a}$$

and

$$\phi_i = \sin^{-1}(A_i/\psi_i) = \cos^{-1}(B_i/\psi_i) \ , \tag{9b}$$

then

$$\begin{aligned} y_{[M_\theta]i,j} &= \eta_i + \psi_i(\cos\phi_i\cos\theta_j + \sin\phi_i\sin\theta_j) \\ &= \eta_i + \psi_i \cos(\phi_i + \theta_j) \ . \end{aligned} \tag{10}$$

For the temperatures themselves this model thus becomes

$$t_{[M_\theta]i,j} = \eta_i + \xi_j + \psi_i\cos(\phi_i + \theta_j) \ . \tag{11}$$

This is a harmonic model with additive station η_i and month ξ_j effects and a cosine term in monthly argument θ_j and station-specific phase ϕ_i and amplitude ψ_i .

Model $[M_\theta]$ of (10) and (11) makes obvious physical sense in having a harmonic annual cycle of temperatures whose phase and amplitude vary from station to station, as well as additive terms for station and monthly average temperature levels. It is, in fact, a specialization of the model mentioned above,

$$y_{[RR]i,j} = \eta_i + \psi_i x_j \ . \tag{12}$$

In the present specialization the interaction is not a simple $\psi_i x_j$ product of a station by a month term, but the latter term

becomes $\cos(\phi_i + \theta_j)$ which depends both upon stations and months.

Model $[M_\theta]$ involves 195 independent parameters: 24 ξ_j's, 49 η_i's, 49 ϕ_i's, 49 ψ_i's and 24 θ_j's. Tsianco (1980, Appendix II) has described an iterative procedure for fitting this model by least squares. It converges to the parameter estimates $\hat{\eta}_i$, $\hat{A}_i$, $\hat{B}_i$, $\hat{\theta}_j$ given in Appendix Table D (the $\hat{\xi}_j$ estimates are the monthly means in Appendix Table B) and has a residual sum of squares of 79,405. The RMS $[M_\theta]$ = 79,405/(1200 - 195) = 79. Evidently, this model fits the data very satisfactorily.

Inspection of the $\hat{\theta}_j$'s fitted by Tsianco (1980, Section 2.5) showed that they were very close to evenly spaced in each of the two years. We therefore simplified model $[M_\theta]$ by substituting $\pi[j(\text{mod } 12) - \frac{1}{2}]/6$ for θ_j. The resulting model, denoted [M], becomes

$$t_{[M]i,j} = \eta_i + \xi_j + \psi_i \cos\left([j(\text{mod } 12) - \tfrac{1}{2}]\pi/6 + \phi_i\right). \tag{13}$$

Upon rewriting this as

$$y_{[M]i,j} = \eta_i + A_i \sin\left([j(\text{mod } 12) - \tfrac{1}{2}]\pi/6 + \phi_i\right) + B_i \cos\left([j(\text{mod } 12) - \tfrac{1}{2}]\pi/6 + \phi_i\right), \tag{14}$$

it is readily fitted by linear least squares methods -- estimates are given in Appendix Table D and goodness-of-fit is compared in Table 2 with that of the bimodel and $[M_\theta]$.

- Table 2 -

Table 2: Goodness of fit of bimodel and harmonic models

Model	Rows	Columns	No. of Parameters	Residual S of Sqs.	Residual Mean Sq.
Bimodel	3D	3D	234	70,743	73
$[M_\theta]$	3D	Any harmonic	195	79,405	79
[M]	3D	Equi-spaced harmonic	171	92,888	90
[M2]	3D	Equi-spaced double harmonic	269	73,129	78

It is evident that the harmonic models fit almost as well as the bimodel itself and they have the advantage of mathematical elegance and somewhat fewer parameters. The loss of fit due to making the θ angles equidistant is not large enough to cause concern, especially in comparison to the original sum of squares of deviations 2,958,198 from monthly means. In the following, we will therefore confine the discussion to models with equidistant angles. (This also has the practical advantage of avoiding the need for non-linear fitting procedures).

6.3 Residuals from the bimodel and models $[M_\theta]$ and $[M]$

The adequacy of any of these models will be judged not only by the smallness of the residuals but also by whether these residuals seem to behave like random "noise". If the residuals, however small, were to show clear systematic patterns, these would have to be incorporated into a model before the fitting could be considered satisfactory.

-Figures 11, 12 and 13 -

Figure 11 is the GH' biplot of the residuals from the rank 3 bimodel of Y. In other words, it is a plot of the fourth and fifth components of the rank 5 GH' bimodel of Y, which account for 0.9 and 0.3 percent of $||Y||^2$. Regarding the month markers as vectors radiating from the origin, we note that the cosines of angles between these vectors approximate correlations between residuals for the corresponding months. Interpreting angles between month markers accordingly, we see that November

Figure 11: The 4th and 5th dimensions of the rank 5 GH' - bimodel of Y

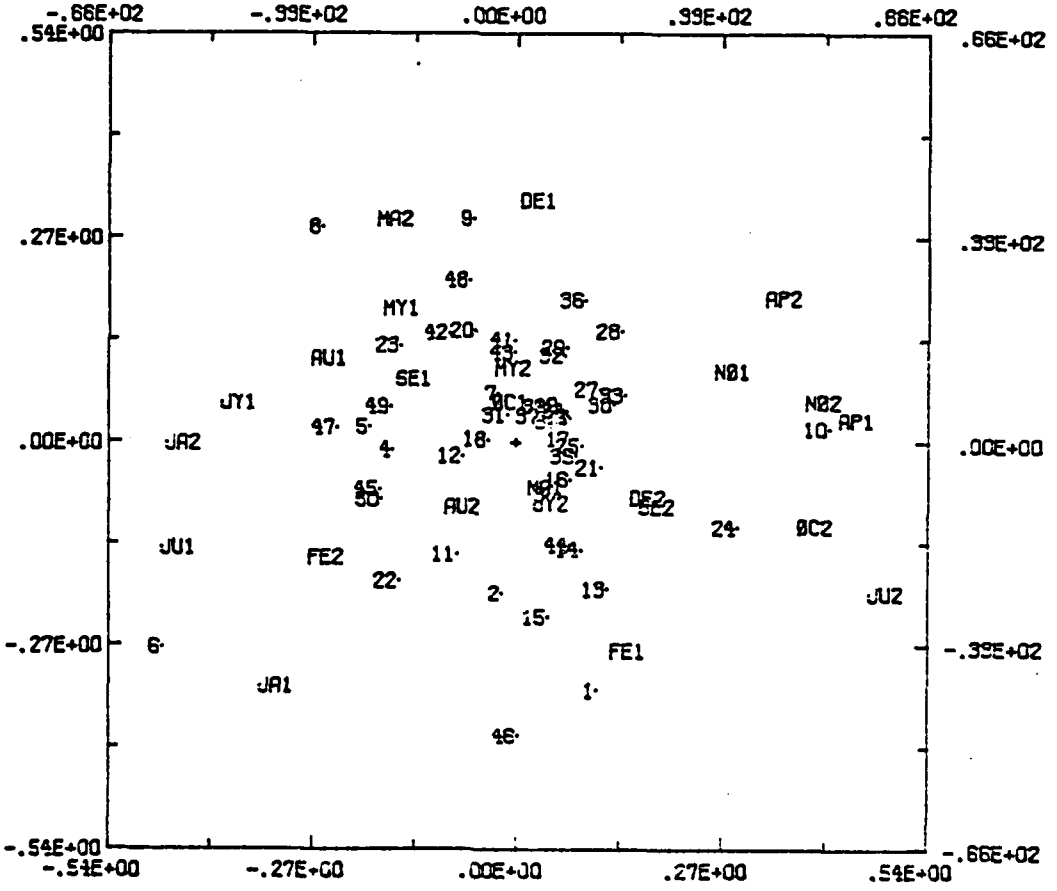


Figure 12: GH' - biplot of the residuals from model [M_c]

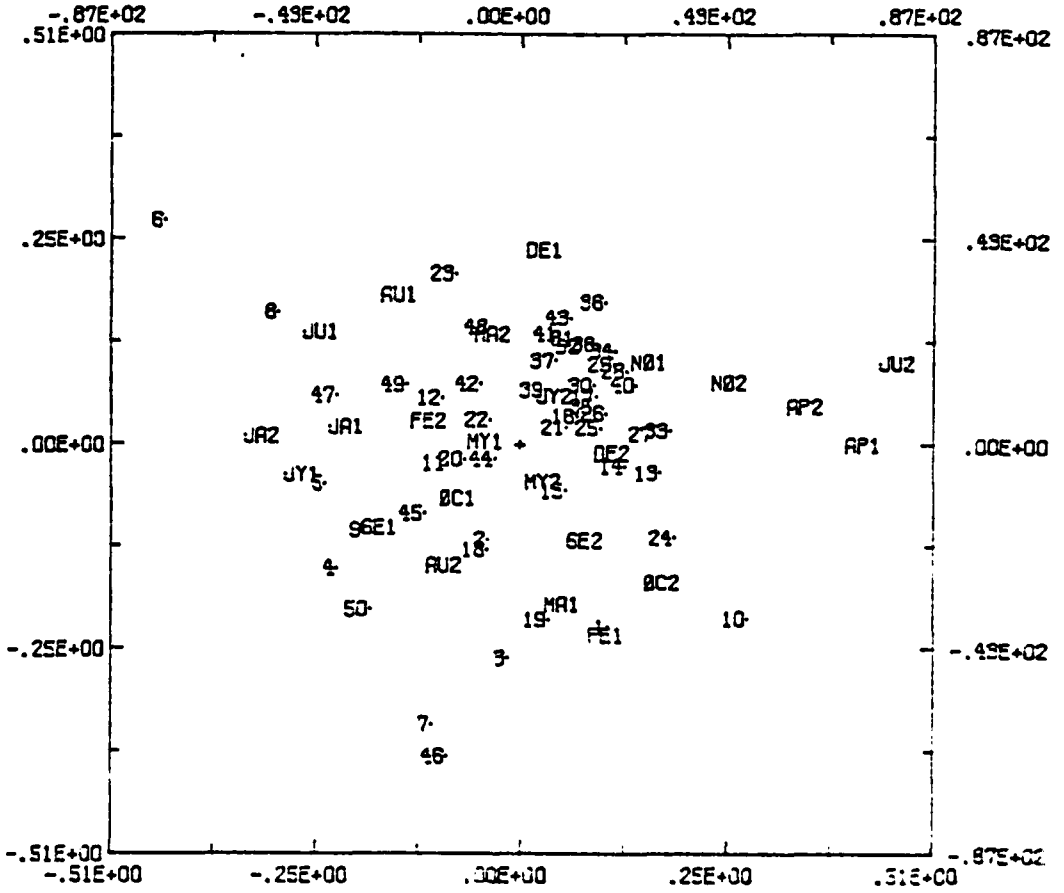
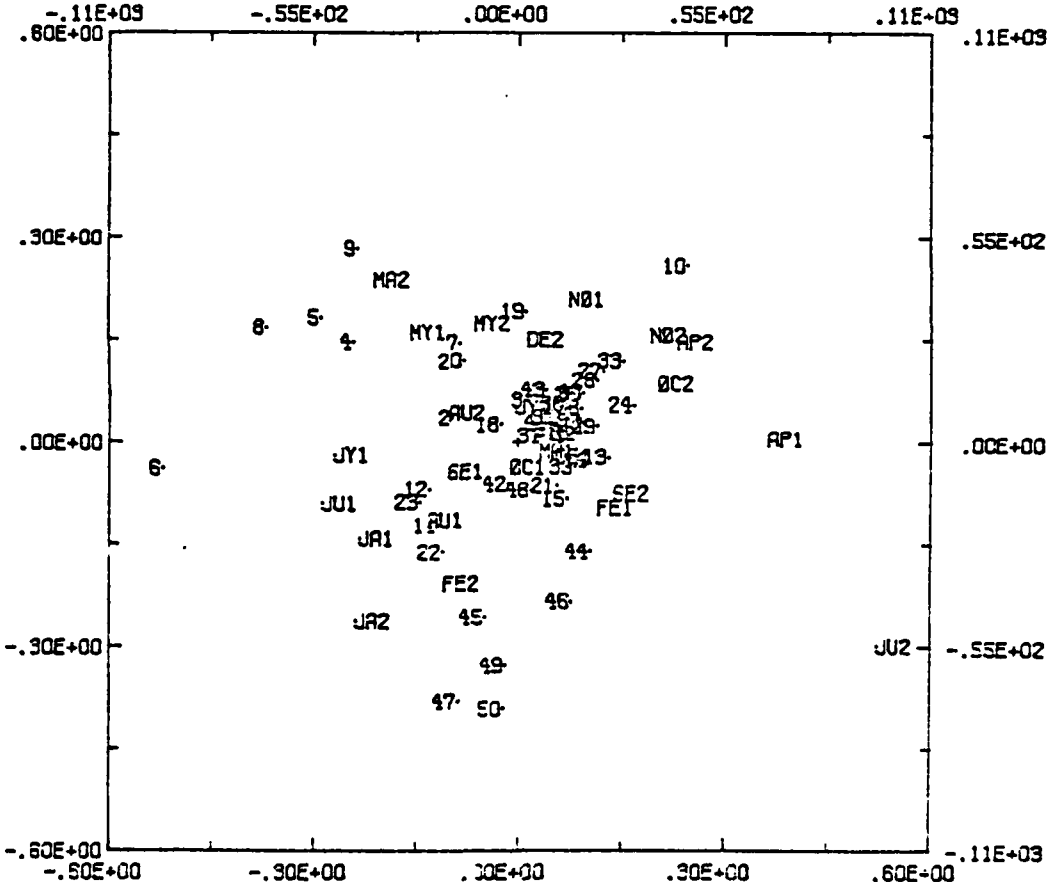


Figure 13: GH' - biplot of the residuals from model [M]



and April residuals from both years are positively correlated amongst themselves and negatively correlated with 1951 January, June and July, residuals as well as with January, 1952 residuals. We also note that the high numbered (northern) row markers are not all in the lower portion of the plot as they were in Figure 3.

Figure 12 is the GH' biplot of the residuals from model $[M_\theta]$. Its components account for 38 and 13 percent of the residual sum of squares, or in other words, essentially the same proportions of $||Y||^2$ as the components in Figure 11. The row and column marker configurations in Figure 12 are very similar to those in Figure 11. This is not at all surprising since $[M_\theta]$ fits the bimodel very closely.

Figure 13 is the GH' biplot of the residuals from model [M]. Its components account for 35 and 18 percent of the residual sum of squares from [M]. The row marker scatter in this biplot looks much more like that of the biplot of Y (Figure 3) than like the scatter in the other biplots of residuals. Here, the high numbered stations are again at the bottom of the plot. The column marker configuration exhibits the same April-November versus January-June-July axis seen in the two preceeding biplots.

The assumption that θ_j equals $[j(\text{mod } 12) - \frac{1}{2}]\pi/6$ does not appreciably increase the residual mean square, but it does reintroduce a systematic element into the residuals, as evidenced by the difference between row marker configurations in Figures 12 and 13. Apparently, this assumption affects the fit for

northern stations differently than that for southern stations.

The April-November versus January-June-July axis is a systematic feature of the residuals of all three models. The correlations between January, April, July and November residuals under [M] are:

	JA1	AP1	JY1	NO1	JA2	AP2	JY2	NO2
JA1	1.0							
AP1	-.5	1.0						
JY1	.4	-.7	1.0					
NO1	-.3	.2	-.1	1.0				
JA2	.4	-.6	.5	-.6	1.0			
AP2	-.7	.6	-.5	.5	-.6	1.0		
JY2	.1	-.2	-.0	.2	-.1	-.1	1.0	
NO2	-.4	.5	-.6	.6	-.7	.6	.2	1.0 .

With the exception of correlations involving July 1952, these alternate in sign. Also, with the exception of November, these months are at three month intervals, suggesting the presence of a double annual oscillation in the residuals.

The residuals from the bimodel and from model $[M_\theta]$ also show evidence of a double annual oscillation. In the next Section we will introduce terms into [M] to account for this variation.

6.4 A model with annual and double annual terms, and its residuals

To account for the double annual oscillation noted in the preceeding Section, we fit the double harmonic model [M2],

$$y_{[M2]i,j} = \eta_i + \psi_i \cos\left([j(\text{mod } 12) - \tfrac{1}{2}]\pi/6 + \phi_i\right) + \psi_{2,i} \cos\left([j(\text{mod } 6) - \tfrac{1}{2}]\pi/3 + \phi_{2,i}\right) \quad . \tag{15}$$

As with model [M], to fit [M2] we transform to a linear model with 269 independent parameters. The residual sum of squares is 73,129 so the residual mean square is 78, a slight improvement on that of the single harmonic model [M].

Craddock (1956) found that, after averaging over many years, mean monthly temperatures for many stations in northern and central Europe could be represented "very accurately" by a model of the same form as [M2].

This double harmonic model [M2] is highly parameterized -- later fits will attempt to simplify the representation and use fewer parameters. It is therefore appropriate to check the residuals from [M2] in some detail. A box-plot (Tukey, 1977, Chapter 2) - Figure 14 - of the residuals shows the great majority to be between ± 15. The more salient deviations are shown in Table 3. Evidently, June 1952, is the single month with the poorest fit, temperatures in the central South American stations being well below the model values, US stations being slightly above model values.

- Figures 14, 15 and Table 3 -

Figure 14: Box-plot of residuals from [M2]

Table 3: Salient residuals from model [M2].

Station	1951 JA	FE	MR	AP	MY	JN	JL	AU	SE	OC	NO	DE	1952 JA	FE	MR	AP	MY	JN	JL	AU	SE	OC	NO	DE
1										+17		-24						+16						
2												-17								+23				
4				-17													+15	-34						
5					+17												+17	-40						
6	+42	-27		-18		+48						-17			+28			-49						
7									+25									-19						
8		-17		-20	+16		+16						+17		+29		+16	-44						
9				-27	+17		+20								+16		+17	-39						
10					-17			-19										+16				+15		
11												-17												
20					+16													-20						
22	+15					+18						-17							-21					
23		-17																						
33										-15														
44															-19									-19
46		+16													-29					-17		+22		
47										+19		+24												
49										+16	-19							+24						-16
50										+15		+21			-24			+20	-19				-17	

Figure 15: GH' - biplot of the residuals from model [M2]

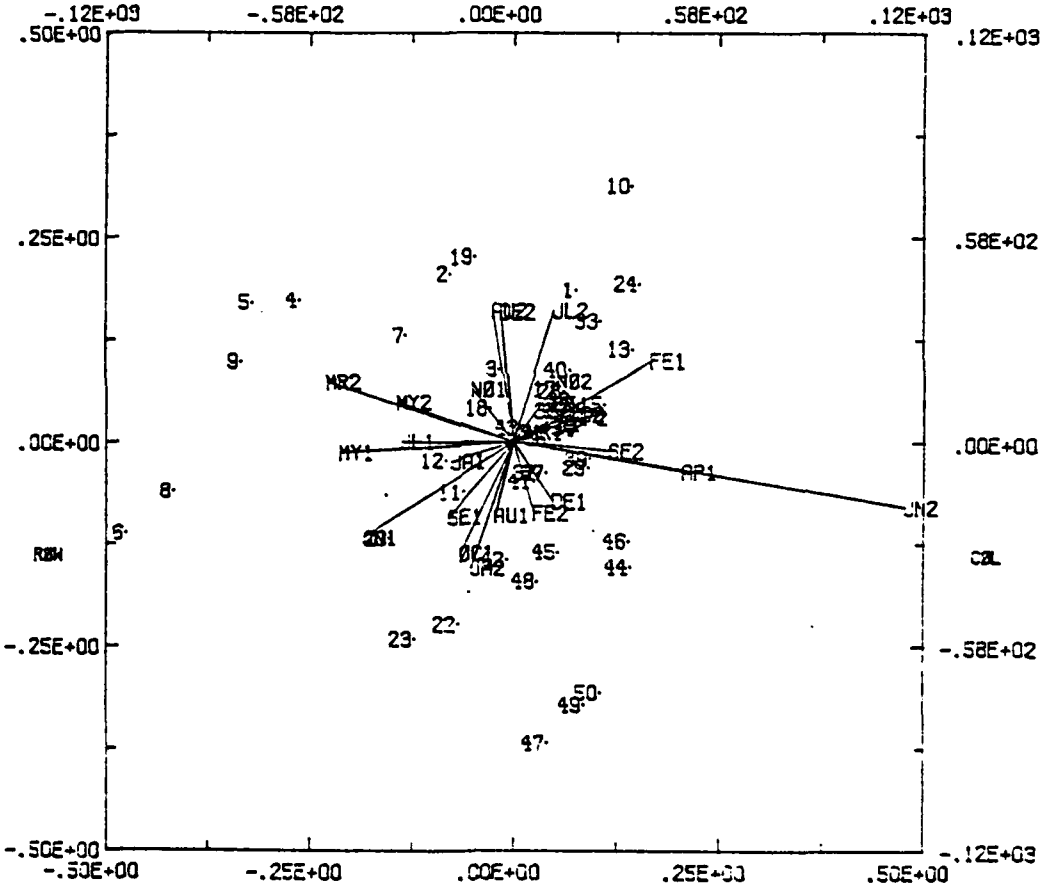


Figure 15 is the GH' biplot of the residuals from model [M2]. Its components account for 39 and 15 percent of the residual sum of squares. The row marker scatter once again shows separation between low and high numbered stations. The April-November versus January-June-July axis is no longer present in the column marker configuration. June 1952 clearly is an unusual month. Station 6 is seen to have a particularly large negative value on that month, with other central South American stations 8, 5, 9, 4 and perhaps 7 not far behind. The biplot also shows the three US stations to be outliers and perhaps, to a lesser extent, Peruvian Current stations 22 and 23. No single month accounts for the latter deviations, but October 1951, January and June 1952, seem to share in the relatively high USA temperatures.

Since these residuals occur for several geographically contiguous stations, they are unlikely to result from errors but presumably show some special local temperature features, for which the model does not account.

6.5 Station parameters of the harmonic model [M]

As noted in the foregoing analyses, model [M] with the single annual harmonic terms fits almost as well as the double harmonic model [M2]. Since [M] is simpler, and has fewer parameters, we continue the analysis with the single annual harmonic.

Model [M] is

$$Y_{[M]i,j} = \eta_i + \psi_i \cos\left([j(\text{mod } 12) - \tfrac{1}{2}]\pi/6 + \phi_i\right) \quad , \qquad (16)$$

and has an "effect" η_i, an amplitude ψ_i and a phase ϕ_i for each station i. It will be of interest to see the geographical variation of the estimates of these three parameters, each of which is readily interpreted in physical terms.

- Figure 16 -

Figure 16 shows a map labeled with the values of η_i under model [M]. Negative values indicate relatively cold stations. Stations 50, 36, 32, 30, 28, 24, 21, 10 and 9 are all at elevations over 1000 meters above sea level, and, consistent with our discussion in Section 4, above, all are cold spots. Also consistent with our earlier discussion is the relationship between latitude and η which can be seen in this map. After excluding the stations listed above, we can best see the nature of this relationship by focusing on three separate groups of stations, those north of the equator, those on the west coast of South America and the remaining stations south of the equator. In each group, as latitude (North or South) increases, η tends to decrease. The map also confirms that the stations on the west coast of South America tend to be colder than their counterparts on the east coast, a phenomenon linked in Section 4 to the cold Peru Current.

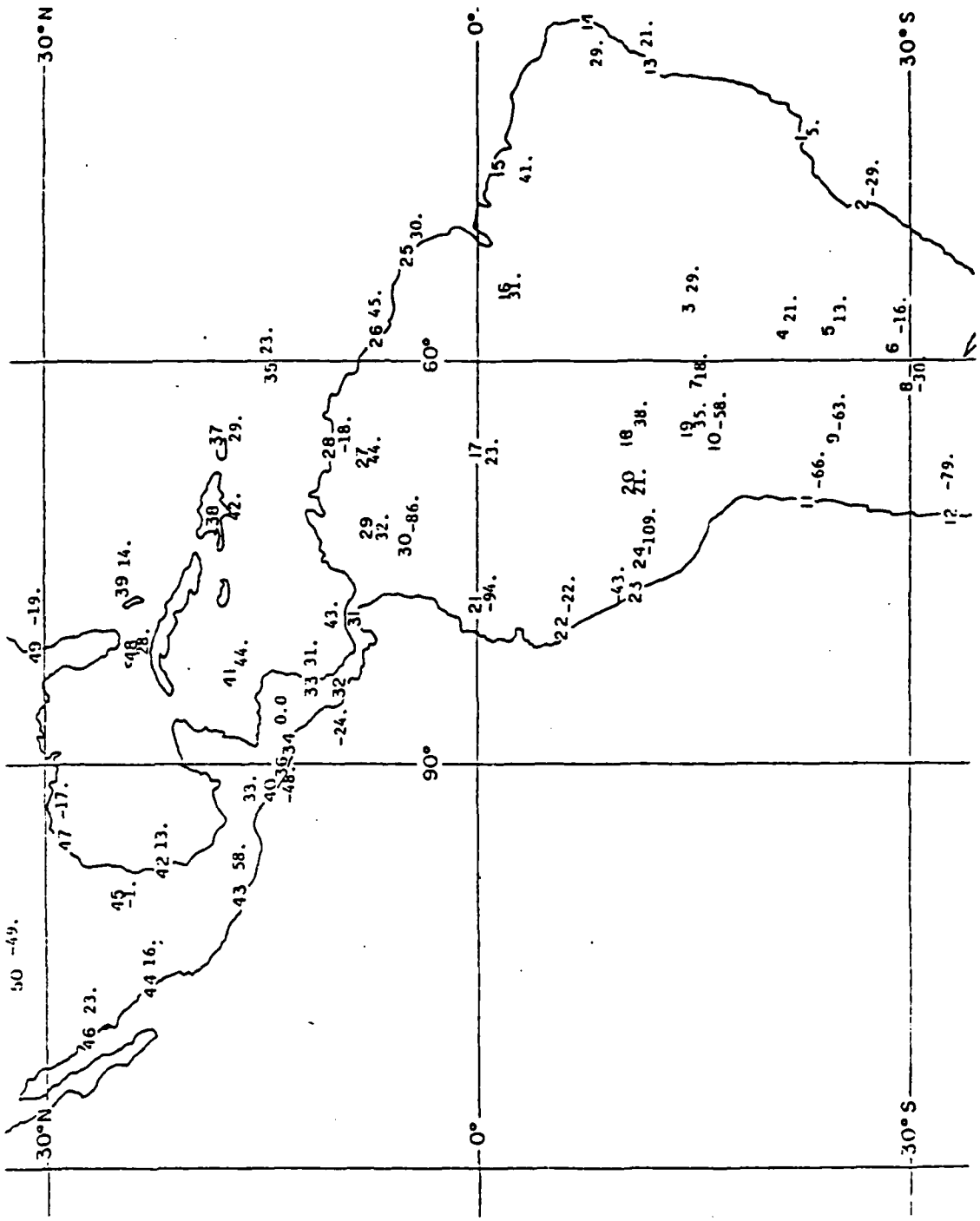


Figure 16: A map indicating station "effects" η_i under model [M]

- Figure 17 -

Figure 17 shows the amplitudes, ψ_i, which reflect the amount of annual cyclic within station variation. In both hemispheres there is an obvious positive relationship between ψ_i and latitude. A close look at the stations south of the equator reveals that among the west coast stations, there is no relationship, and among the inland stations there is a steeper relationship than among the east coast stations.

- Figure 18 -

Figure 18 shows the phases, ϕ_i, under model [M]. This map indicates that north of the equator most phases are close to π. South of the equator, coastal stations tend to have phases somewhat below 2π, while inland stations have phases very close to either 0 or 2π. There is obviously a strong relationship between ϕ_i and latitude. It is interesting to consider also the difference between inland and coastal phases which presumably reflects the fact that water loses and gains heat more slowly than land.

7. Modeling the Geographical Dispersion of the Stations

7.1 Observed geographical temperature patterns

The geographical dispersion of the monthly temperatures has been examined twice. In Section 4, the station's g-markers on the biplot/bimodel were inspected and in Section 6.5 the

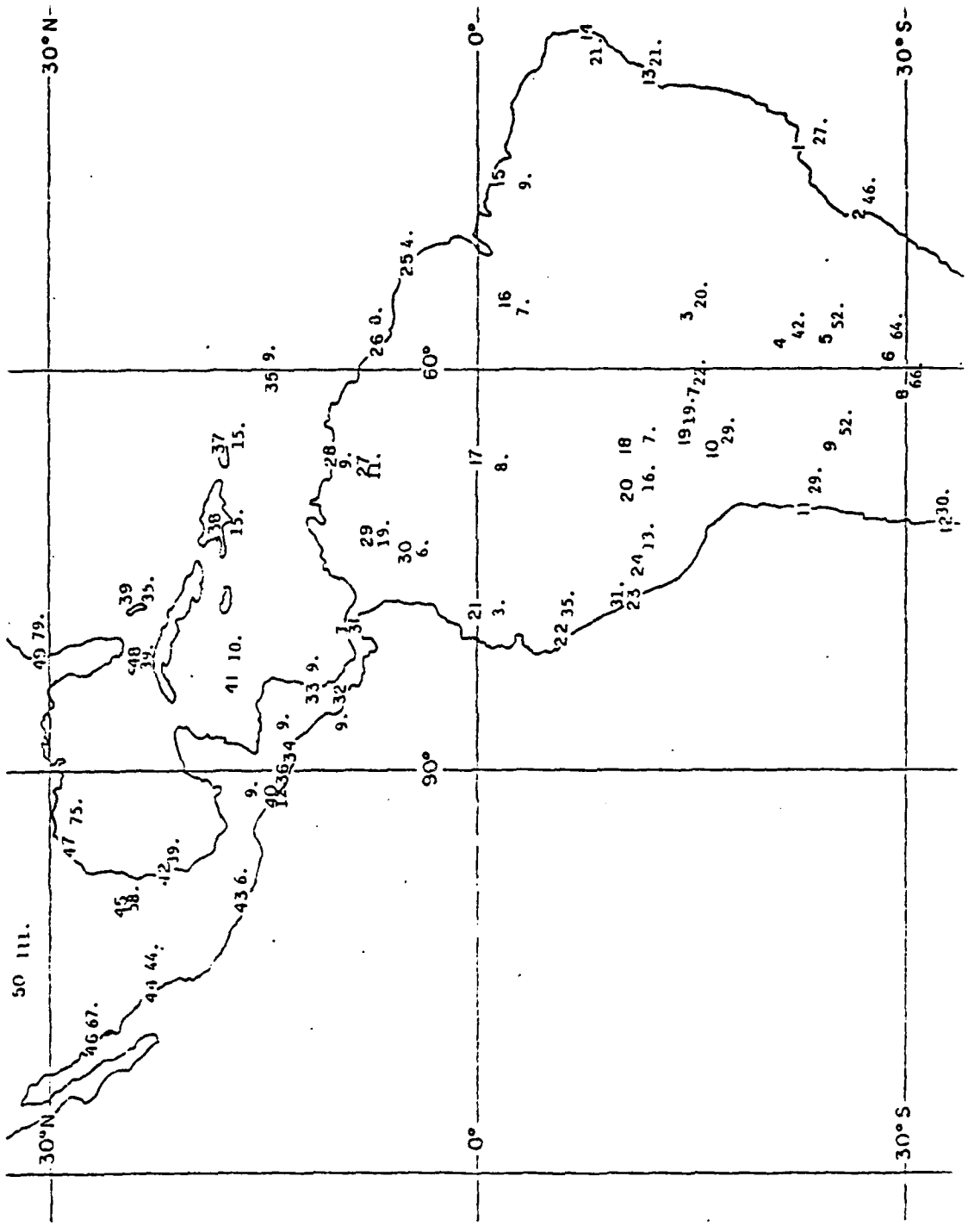
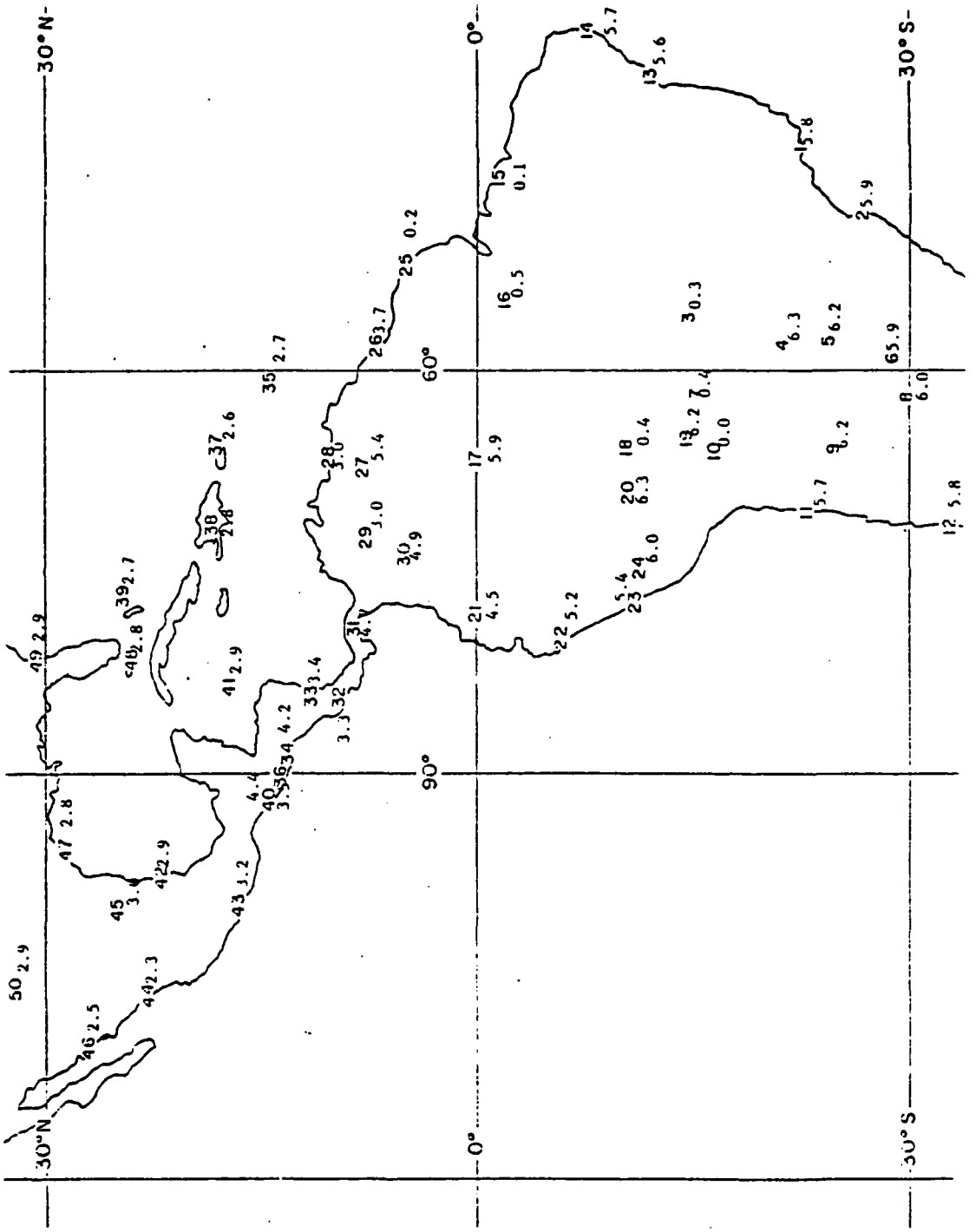


Figure 17: A map indicating station amplitudes, ψ_i, under model [M]

Figure 18: A map indicating station phases ϕ_i under model [M]



stations' temperature parameters (average, amplitude and phase) were surveyed. Both inspections essentially agreed that temperatures were related to altitude, to hemisphere (North versus South) and to latitude. Longitude did not seem to play a noticeable role. It was also noted that some clusters of stations did not quite seem to fit the geographical/topographical pattern: primarily stations 11, 12, 22, 23 close to the Peru Current, less notably US stations 47, 49, 50 and also, possibly, the inland stations 4, 5, 6, 8, and 9 of South America which seemed to differ from coastal stations.

7.2 Polynomial models

The most obvious and mathematically most tractable models are polynomials in latitudes and altitudes with indicator variables for special groups. Thus, for example, one model for the $\underline{g}$-markers would be

$$\underline{g}_{\{P,\bar{P},A,A^2,AL,L^2\},i} = t_{i,P}\underline{\mu}_P + (1 - t_{i,P})\underline{\mu}_{\bar{P}} + a_i\underline{\alpha}_1 + a_i^2\underline{\alpha}_2 + a_i \ell_{i-1,1}\underline{\rho} + \ell_{i-2}^2\underline{\beta} \quad , \tag{17}$$

where the $\underline{g}$, $\underline{\mu}$, $\underline{\alpha}$, $\underline{\rho}$, $\underline{\beta}$,'s are three dimensional vectors of parameters ($\underline{\mu}_P$ is a mean for stations in P, $\underline{\mu}_{\bar{P}}$ for other stations), a_i and ℓ_i are station i's altitude and latitude (Southern latitudes being prefixed by a negative sign), $t_{i,P} = 1$ if i = 11, 12, 22 or 23 and zero otherwise. The notation $\{P,\bar{P},A,A^2,AL,L^2\}$, for this model readily extends to other models of this type.

After some trial and error it was decided to try out models for the stations involving latitudes as ℓ_i, ℓ_i^2, ℓ_i^3, altitudes as a_i and a_i^2 and interactions $a_i\ell_i^2$ and $a_i^2\ell_i^2$. Station groups that were identified by indicator variables were as follows

Southern	S	i = 1, 2, 3, 7, 10, 13, 14, 18, 19, 20, 24
Inland South	I	i = 4, 5, 6, 8, 9
Peru Current	P	i = 11, 12, 22, 23
Equatorial	E	i = 15, 16, 17, 21
Northern	N	i = 25-46, 48
US	U	i = 47, 49, 50

Reference to the map -- Figure 1 -- to biplot patterns -- Figures 8, 10 -- and parameters -- Figures 16, 17, 18 -- show these groupings to be suitable both geographically and in terms of temperature patterns.

7.3 Least squares fitting

The fit of polynomial models to the g-vectors will not be further discussed since it was preliminary to joint modeling of g's and h's for models to the data Y (or the temperatures T). Thus, for example, a joint model with (17) for the rows and (5) for the columns would become

$$Y_{\{P,\bar{P},A,A^2,AL,L^2\}[M_\theta]i,j} = \tag{18}$$

$$(t_{i,P}\ ,1 - t_{i,P}\ ,a_i\ ,a_i^2\ ,a_i\ell_i\ ,\ell_i^2)\begin{pmatrix} \underline{\mu}'_P \\ \underline{\mu}'_{\bar{P}} \\ \underline{\alpha}'_1 \\ \underline{\alpha}'_2 \\ \underline{\rho}'_{1,1} \\ \underline{\beta}'_2 \end{pmatrix} (\underline{\lambda},\underline{\gamma},\underline{\xi}) \begin{pmatrix} 1 \\ \sin\theta_j \\ \cos\theta_j \end{pmatrix}.$$

To describe the computations, it will be convenient to write the (50 x 6) matrix

$$X_{\{P,\bar{P},A,A^2,AL,L^2\}} = \left(\underline{1}_P\ ,\ \underline{1}_{\bar{P}}\ ,\ \underline{a}\ ,\ (\underline{a^2})\ ,\ (\underline{a\ell})\ ,\ (\underline{\ell^2})\right)\ , \tag{19}$$

where $\underline{1}_P$ is a vector with ones in rows $i \in P$, zeroes elsewhere, $\underline{a}$ is the vector $(a_1, a_2, \ldots, a_{50})'$, $(\underline{a}^2) = (a_1^2, a_2^2, \ldots, a_{50}^2)'$, etc.

Similarly, one may write the (24 x 3) matrix

$$Z_{[M_\theta]} = \left(\underline{1},\ (\underline{\sin\theta}),\ (\underline{\cos\theta})\right) \tag{20}$$

and the (6 x 3) parameter matrix

$$\begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \tag{21}$$

The model then becomes

$$Y_{\{P,\bar{P},A,A^2,AL,L^2\}[M_\theta]} = X_{\{P,\bar{P},A,A^2,AL,L^2\}} \Omega\, Z'_{[M_\theta]} \quad , \tag{22}$$

which can be fitted by standard least squares methods for given θ's, i.e., when $\theta_j = [j(\text{mod } 12) - \frac{1}{2}]\pi/6$. Thus, the sum of squares $\|Y - X\Omega Z'\|^2$ is minimized for

$$\hat{\Omega} = (X'X)^{-1} X'YZ\ (Z'Z)^{-1} \quad , \tag{23}$$

a straightforward computation.

A few comments about this parametrization may be helpful. The fitted parameters $\hat{\omega}_{u,v}$ of $\hat{\Omega}$ must be interpreted in terms of the model for the rows and columns. Thus, in model $\{P,\bar{P},A,A^2,AL,L^2\}[M]$, $\hat{\omega}_{3,2}$ would be the coefficient of the sine of θ for altitudes, $\hat{\omega}_{1,3}$ the coefficient of the cosine of θ for the Peru Current. Each station group, and altitude, latitude combination or power would have a coefficient for the mean and one each for the sine and cosine of θ.

The following reparametrization -- analagous to that used in (8), (9), above -- may lead to easier interpretation. Set

$$\hat{\eta}_u = \hat{\omega}_{u,1} \quad , \tag{24a}$$

$$\hat{\psi}_u = \sqrt{(\hat{\omega}_{u,2} + \hat{\omega}_{u,3})} \tag{24b}$$

and

$$\hat{\phi}_u = \tan^{-1} (\hat{\omega}_{u,2}/\hat{\omega}_{u,3}) \quad , \tag{24c}$$

and regard these as coefficients of the u-th criterion (group of stations, altitude, latitude, etc.) for a constant and the amplitude and phase of the months' factor θ. Thus

$$y_{i,j} = \Sigma_u \, x_{i,u} \left(\hat{\eta}_u + \hat{\psi}_u \cos(\hat{\phi}_u + \theta_j)\right) \quad . \tag{25}$$

The fit is of course

$$\hat{Y} = X\hat{\Omega}Z' \quad , \tag{26}$$

and the residual sum of squares is $\|Y - X\hat{\Omega}Z'\|^2$.

7.4 Fit of various models

Table 4 shows the sums of squared residuals from a variety of models which were tried out. A number of things are evident. (i) None of the models for stations fit anywhere as well as the general model with a separate harmonic for each station. (ii) Clustering the stations into 6 groups reduces the number of parameters at the expense of leaving huge residuals. (iii) Adding in altitudes makes for a reasonable, but not good, fit. (iv) Better fits are obtained by adding in latitude factors and altitude x latitude interactions. (v) With sufficient

such factors one may as well pool some of the station groupings and single out only the Peruvian, USA and Southern Inland clusters. (vi) The second harmonic improves the fit little. Proportionally, this improvement is negligible unless the best fitting model with 50 station factors is used. (In other words, the second harmonic does not noticeably improve the fit of any of the low parameter models).

- Tables 4, 5 and 6 -

The regression coefficients $\hat{\omega}$ of (24) are shown for two of the better fitting models in Tables 5 and 6. These make quite good sense. Thus, in Table 5 the mean temperatures are low on the Peru Current, and the amplitude of the annual harmonic generally increases with distance from the equator but is particularly large in Southern Inland areas. The phases are near zero for Southern Hemisphere stations, near 180° for Northern Hemisphere stations - again a not unexpected result. The mean temperatures fall off with increased altitude and latitude, the amplitude increases with latitudes but does not depend much on altitude - all very much as one would expect from elementary geographical considerations. The harmonics for altitude and altitude x latitude are negligible so their phase angle is of no interest. However, it would be interesting to understand the 221° phase of the $(\text{latitude}/10)^2$ harmonic.

The coefficients in the model of Table 6 are generally similar. Here three groups of stations have been merged and

Table 4: Residual sum of squares for fit of various models. (Number of parameters in parentheses.)

Station Model	Months Model	
	[M]	[M2]
All Stations	92,888 (171)	73,129 (269)
S,I,P,E,N,U	1,813,447 (39)	
S,I,P,E,N,U;A	468,167 (42)	
S,I,P,E,N,U;A, L^2	350,285 (45)	
S,I,P,E,N,U;A, L^2, AL^2	328,995 (48)	
S,I,P,E,N,U;A, L^2 A^2L^2	322,185 (48)	
S,I,P,E,N,U;A, L^2,L^3, A^2L^2	252,439 (51)	
P,$\bar{P}$;A, L,L^2	395,494 (36)	
P,$\bar{P}$; A^2,L,L^2	559,915 (36)	
P,$\bar{P}$;A,A^2,L,L^2	392,813 (39)	385,164 (49)
P,$\bar{P}\cup\bar{U}$, U;A,A^2,L,L^2	348,923 (42)	340,321 (54)
P,$\bar{P}$;A,A^2,L,L^2,L^3	347,102 (42)	337,164 (54)
P,$\bar{P}\cup\bar{U}$, U;A,A^2,L,L^2,L^3	311,764 (45)	301,107 (59)
P,$\bar{P}$;A,A^2,L,L^2,L^3,AL^2	326,623 (45)	
P,$\bar{P}$;A,A^2,L,L^2,L^3, A^2L^2	318,191 (45)	306,718 (59)
P,$\bar{P}$, U;A,A^2,L,L^2,L^3, A^2L^2	283,139 (48)	270,957 (64)
I,P,$\bar{U}$,U;A, L,L^2,L^3, A^2L^2	261,096 (48)	

Table 5: Least squares coefficients for model $\{S,I,P,E,N,U;A,L^2,AL^2\}$ [M]

Station Parameters	Mean	Amplitude	Phase Angle
South	44.4	25.3	16°
South Inland	22.9	78.2	21°
Peru Current	-28.9	47.2	42°
Equatorial	35.0	3.9	-7°
North	46.1	8.4	164°
USA	23.8	54.0	182°
Altitude/100	-4.87	0.27	34°
$(\text{Latitude})^2/100$	-4.49	3.58	221°
$\text{Alt}(\text{Lat})^2/1000$	2.26	1.69	203°

Table 6: Least squares coefficients for model $\{I,P,U;A,L,L^2,L^3,A^2L^2\}$ [M]

Station Parameters	Mean	Amplitude	Phase Angle
South Inland	29.9	38.4	23°
Peru Current	-24.3	25.6	68°
USA	9.7	18.1	127°
Other Stations	42.7	11.5	25°
Altitude/100	-5.10	0.31	179°
Latitude/10	0.07	7.88	170°
$(Lat/10)^2$	-4.10	3.63	220°
$(Lat/10)^3$	0.42	1.19	222°
$(Alt)^2(Lat)^2/10^6$	0.017	0.003	5°

two more latitude factors included. The regional patterns and the altitude and latitude effects are much the same as in Table 5 except for the phases for the altitude factors which, however, seem to be of little relevance in view of the small amplitudes.

The residuals from the last two models are biplotted in Figures 19 and 20, respectively. An annual elliptic pattern remains in the former biplot - showing that the 6 station group did not completely account for their different annual harmonics. In the latter biplot this elliptic pattern is much distorted and quite irregular -- evidently the extra latitude factors account for much of the annual harmonic.

- Figures 19 and 20 -

7.5 Summary of modeling

Modeling the 50 stations' differences by means of a few groupings and/or altitude and latitude factors was only partially successful. The resulting RMS was about 3 times that obtained by fitting the stations separately and systematic elements remain in the residuals. However, these models have a greatly reduced number of parameters and are much easier to describe and interpret. Thus, there is some trade-off between the not-quite-so-well-fitting simple systematic models as described in Tables 5 and 6 and the better fitting but more complex 50 station models of the preceding Section.

8. Time Extrapolated Eigenvector Prediction (TEEP)

Brier and Meltesen (1976) used the same data we have been

Figure 19: GH' - biplot of residuals from $\{S,I,P,E,N,U;A,L^2,AL^2\}$ [M]

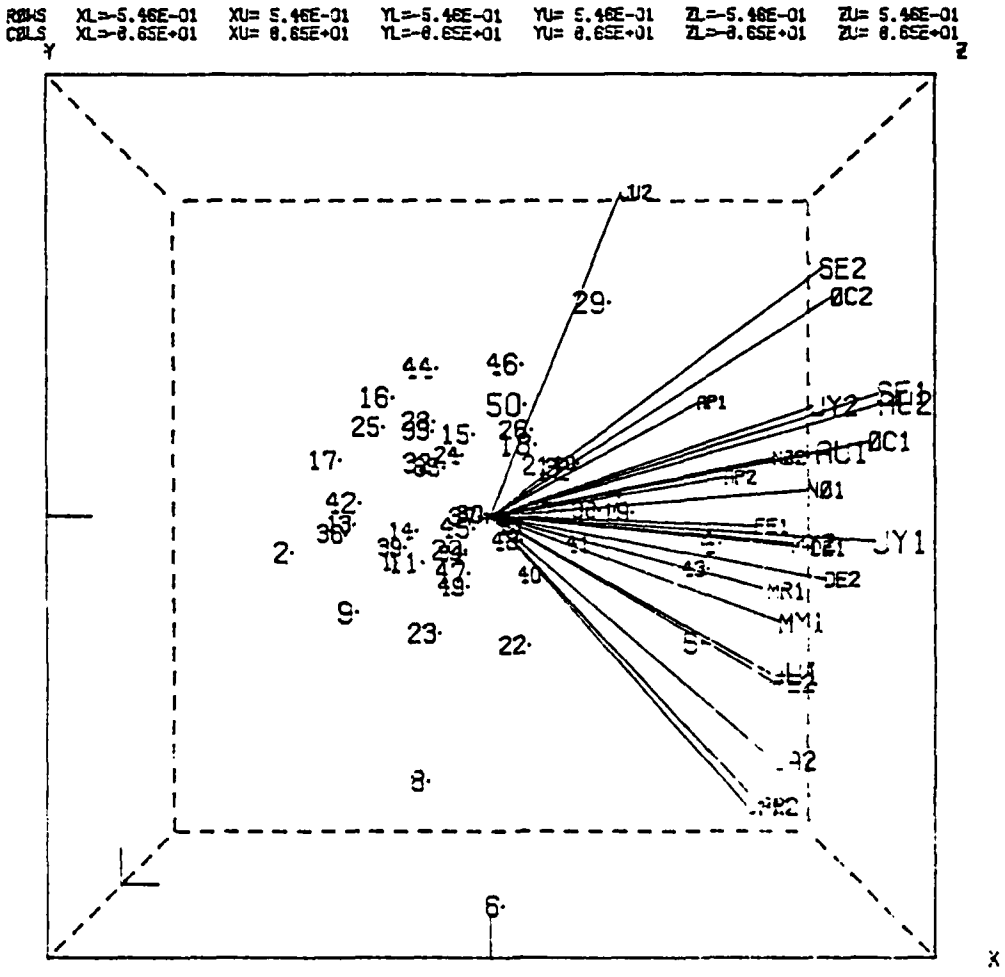
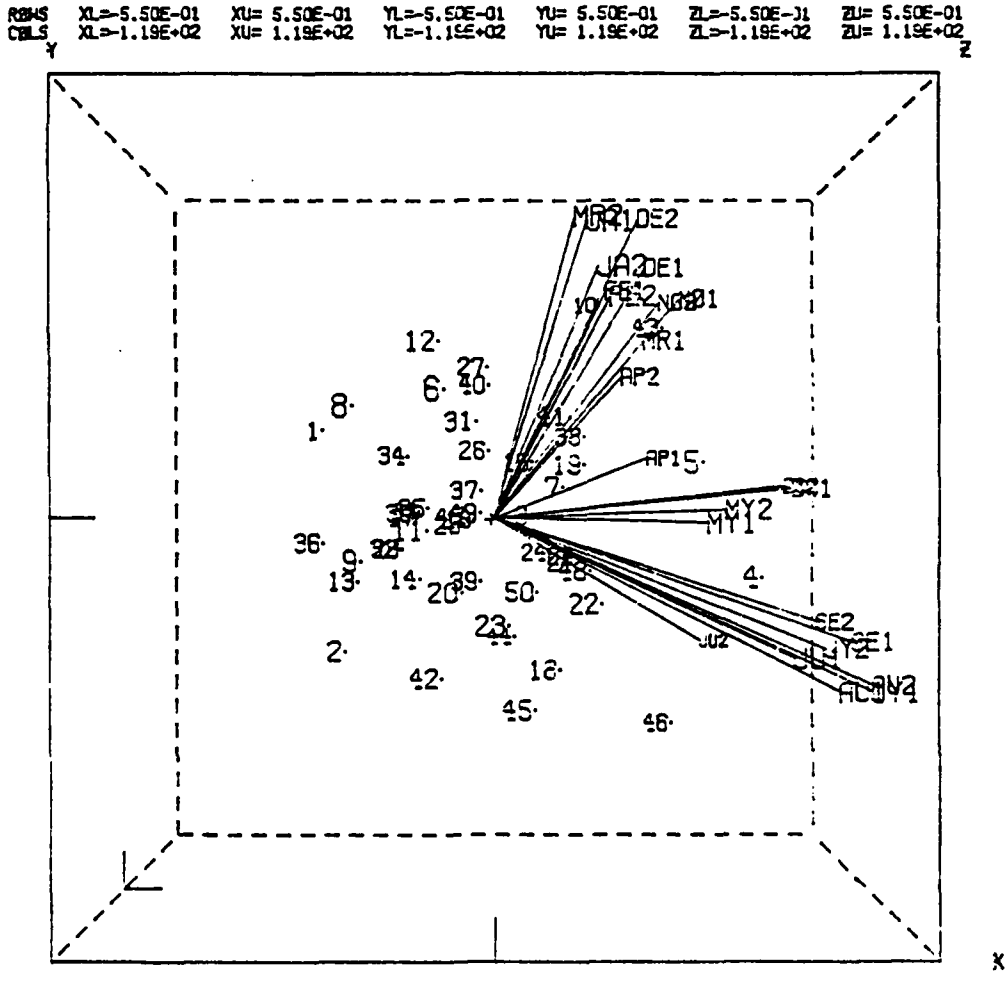


Figure 20: Perspective view of GH'-bimodel of residuals from $\{I,P,U;A,L,L^2,L^3,A^2L^2\}$ [M]



using to illustrate TEEP. They did not actually produce a TEEP estimate, but they indicated how to obtain one from these data. It is best to show the difference between TEEP estimation and what has been done here, by producing a TEEP estimate.

Instead of deriving a model from the elliptical configuration of $\underline{h}_j$'s, an ellipse was fitted directly to the $\underline{h}_j$'s. This entailed a special non-linear fitting routine (described in Appendix II.5 of Tsianco, 1980) and resulted in the fitted ellipse represented by

$$\{\underline{P}\epsilon R^3: \ \underline{P} = \hat{\underline{\mu}} + \hat{\underline{\alpha}} \sin\theta + \hat{\underline{\beta}} \cos\theta, \ \theta\epsilon[0,2\pi]\} \ ,$$

where

$$\hat{\underline{\mu}} = \begin{pmatrix} 296.3 \\ 5.0 \\ -0.1 \end{pmatrix} , \quad \hat{\underline{\alpha}} = \begin{pmatrix} -49.2 \\ 269.2 \\ 8.4 \end{pmatrix} \quad \text{and} \quad \hat{\underline{\beta}} = \begin{pmatrix} -14.2 \\ 9.4 \\ -51.8 \end{pmatrix} .$$

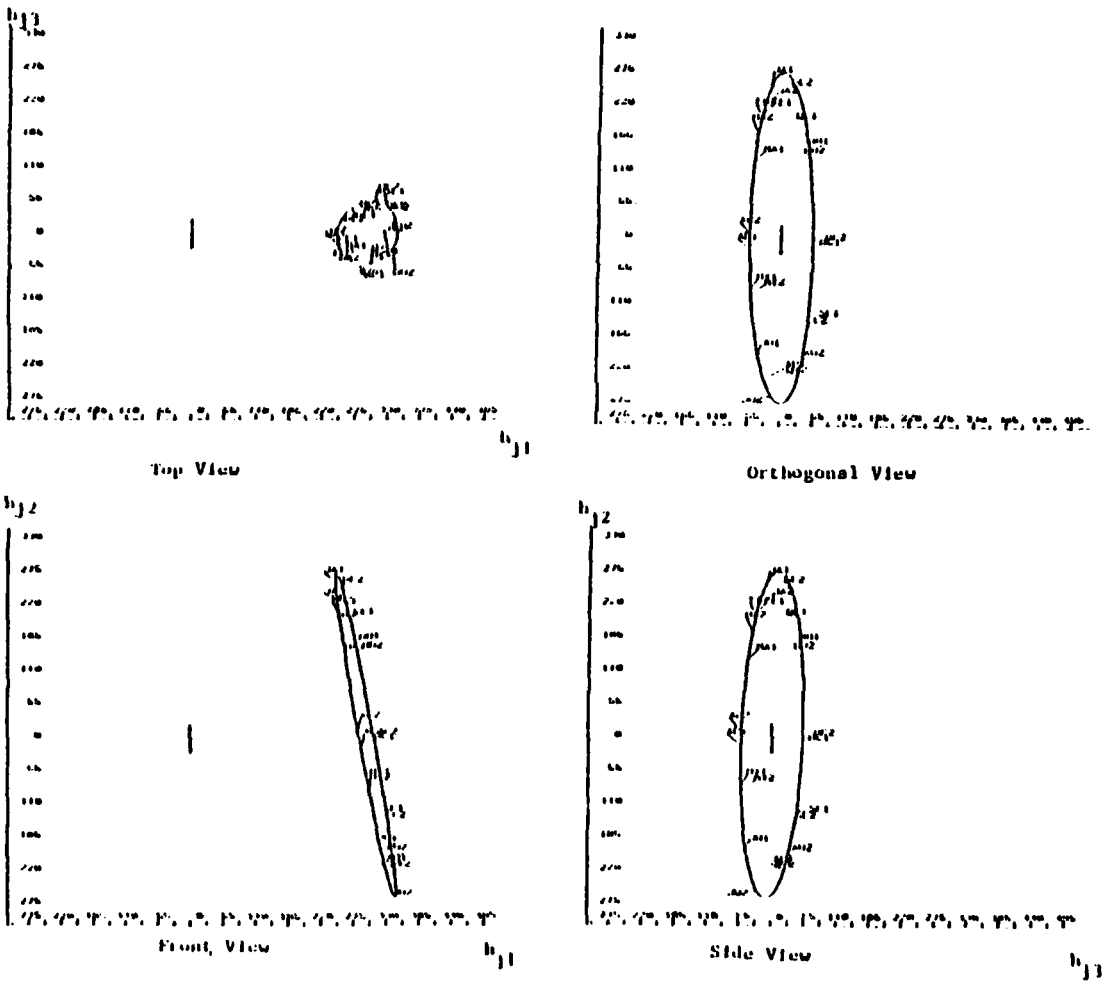
The major and minor axes of this ellipse are

$$\hat{\underline{\alpha}}_{\text{Major}} = \begin{pmatrix} -49.8 \\ 269.3 \\ 6.4 \end{pmatrix} \quad \text{and} \quad \hat{\underline{\beta}}_{\text{Minor}} = \begin{pmatrix} -12.3 \\ -1.0 \\ -52.0 \end{pmatrix} .$$

Figure 21 shows the $\underline{h}_j$'s and the fitted ellipse.

- Figure 21 -

Figure 21: Projections of the $\underline{h}_j$'s and the fitted ellipse onto the three coordinate planes and onto the plane of the fitted ellipse.



For months in 1951-52, the $\underline{h}_j$'s are arranged in order around this ellipse. To produce TEEP estimates for some future time τ, one needs to predict where along the ellipse a column marker for time τ would fall if data at time τ were available. In the process of fitting an ellipse to the $\underline{h}_j$'s, 24 values $\hat{\theta}_j$ were obtained; these generate points on the fitted ellipse that are closest to the $\underline{h}_j$'s. To extrapolate to a future position on the ellipse, one must extrapolate the relationship between $\hat{\theta}_j$ and j to time $j = \tau$.

Regression of the $\hat{\theta}_j$'s on j(mod 12) reveals a strong linear relationship ($R^2 = .981$). The fitted line is

$$-.44 + .52 \times j(\text{mod } 12).$$

The slope is essentially $\pi/6$.

To obtain a TEEP estimate of the i-th station's deviation from the average temperature for all stations at time τ in the future, one computes

$$\hat{y}_{i\tau} = \underline{c}_i' \; (\hat{\underline{\mu}} : \hat{\underline{\alpha}} : \hat{\underline{\beta}}) \begin{pmatrix} 1 \\ \sin[-.44 + \pi/6 \times \tau(\text{mod } 12)] \\ \cos[-.44 + \pi/6 \times \tau(\text{mod } 12)] \end{pmatrix}.$$

To allow comparisons with other approximations, this was used to approximate Y, resulting in a residual sum of squares of 131,926. Of all the other approximations in this paper, model [M] is most like this. Its residual sum of squares was 92,888.

The distinction between [M] and TEEP is in the manner in which they use the structure found in the eigenvectors of Y. In [M] we used the elliptical structure we had found to derive

a model for the data and we fitted that model directly to the data. In TEEP estimation we fitted a model to eigenvectors (by fitting an ellipse to the $\underline{h}_j$'s, we have implicitly modeled the rows of $H_{[3]}$, which are scaled eigenvectors) and constructed an approximation to the data from the fitted values obtained. Although the name TEEP reflects a context in which elements of the modeled eigenvector are a function of time, the idea of modeling eigenvectors is applicable in other contexts as well.

The choice between the derived model approach and TEEP is a choice between two different methods of fitting equivalent functional forms. In the former, fitting is done in one stage, while in the other it is done in two stages. Obviously, a model fitted directly to the data will fit at least as well as the equivalent TEEP approximation.

In our development of a TEEP estimate, we have simultaneously extrapolated three eigenvectors by extrapolating to a future position among an ellipse. Brier and Meltesen considered each eigenvector separately. In some instances, modeling each eigenvector separately and then constructing a TEEP estimate may be computationally easier than the derived model approach.

9. Summary and Comments

We have used a meteorological example to illustrate in detail a strategy of inspecting data and modeling. Our strategy has relied heavily on biplot/bimodel displays of approximations to the data in 2 or 3 dimensions. The example has shown this approach to be successful, as have a number of other examples

reported elsewhere (Bradu and Grine, 1979; Bradu and Gabriel, 1978; Gabriel, 1972, 1980, 1981a, 1981b, Gabriel, Hill and Law-Yone, Strauss et al., 1979).

Biplot/bimodel display is only one of many available techniques for display of data (Greenacre, 1980; Guttman, 1968; Kruskal, 1978; Lingoes and Guttman, 1967) but we have argued (Gabriel, 1981a; Cox and Gabriel, 1981) that it is particularly suitable for diagnosis of models. Briefly, that is because it is the only one of these techniques whose display is directly related to an approximation of the data: Thus, markers $\underline{g}_{[2],i}$ and $\underline{h}_{[2],j}$ (or $\underline{g}_{[3],i}$ and $\underline{h}_{[3],j}$) are related to datum $y_{i,j}$ by $\underline{g}'_{[\ ],i}\,\underline{h}_{[\ ],j} \doteq \underline{g}'_i\,\underline{h}_j = y_{ij}$. Other techniques display approximations to summary functions (i.e., statistics) of the data such as distances, correlations, etc., and might be helpful in modeling those. In the biplot, on the other hand, an observed pattern of $\underline{g}_{[\ ]i}$'s or $\underline{h}_{[\ ]j}$'s translates algebraically into a structure for the y_{ij}'s and thus provides a model for the Y data.

ACKNOWLEDGEMENT

The authors are grateful to G. W. Brier and G. T. Meltesen for providing them with the data and discussing the results.

BIBLIOGRAPHY

Bradu, D. and Gabriel, K.R. "The Biplot as a Diagnostic Tool for Models of Two-Way Tables," Technometrics, 20, (1978) 47-68.

Bradu, D. and Grine, F.E. "Multivariate Analysis of Diademontive Crania from South Africa and Zambia," South African Journal of Science, 75, (1979) 441-448.

Brier, G.W. and Meltesen, G.T. "Eigen Vector Analysis for Prediction of Time Series," Journal of Applied Meteorology, 15, (1976) 1307-1312.

Corsten, L.C.A. and Gabriel, K.R. "Graphical Exploration in Comparing Variance Matrices," Biometrics, 32, (1976) 851-863.

Cox, C. and Gabriel, K.R. "Some Comparisons of Biplot Display with Pencil-and-Paper E.D.A. Methods," in Advances in Data Analysis (A. Siegel and R.L. Launer, eds.), New York: Academic Press (1981).

Craddock, J.M. "The Representation of the Annual Temperature Variation over Northern and Central Europe by a Two-Term Harmonic Form," Quarterly Journal of the Royal Meteorological Society, 82 (1956) 275-288.

Gabriel, K.R. "The Biplot -- Graphic Display of Matrices with Application to Principal Component Analysis," Biometrika, 58, (1971) 453-467.

Gabriel, K.R. "Analysis of Meteorological Data by Means of

Canonical Decomposition and Biplots," Journal of Applied Meteorology, 11, (1972) 1071-1077.

Gabriel, K.R. "The Complex Correlational Biplot," in Theory Construction and Data Analysis in the Behavioral Sciences, (S. Shye, ed.), San Francisco: Jossey-Bass, 350-370 (1978a).

Gabriel, K.R. "Least Squares Approximation of Matrices by Additive and Multiplicative Models," Journal of the Royal Statistical Society, B, 40, (1978b) 186-196.

Gabriel, K.R. "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, (1980) 113-129.

Gabriel, K.R. "Biplot Display of Multivariate Matrices for Inspection of Data and Diagnosis," Chapter 8 of Interpreting Multivariate Data, (V. Barnett, ed.), London: John Wiley (1981a).

Gabriel, K.R. "Biplot," Encyclopedia of Statistical Sciences (S. Kotz and N.L. Johnson, eds.), New York: John Wiley (1981b).

Gabriel, K.R. "Exploratory Multivariate Analysis of a Single Batch of Data," in Probability, Statistics and Decision Making in Meteorology, (A.H. Murphy and R.W. Katz, eds.), Boulder, Colorado: Westview Press (1981c).

Gabriel, K.R. "Multivariate Comparisons of Data from Several Samples," in Probability, Statistics, and Decision

Making in Meteorology, (A.H. Murphy and R.W. Katz, eds.), Boulder, Colorado: Westview Press (1981d).

Gabriel, K.R., Hill, M. and Law-Yone, H. "A Multivariate Statistical Method for Regionalization," Journal of Regional Science, 14, (1974) 89-106.

Gabriel, K.R., Rave, G. and Weber, E. "Graphische Darstellung von Matrizen durch das Biplot," EDV in Medizen und Biologie, 7, No. 1 (1976) 1-15.

Gabriel, K.R. and Zamir, S. "Lower Rank Approximation of Matrices by Least Squares with any Choice of Weights," Technometrics, 21, (1979) 489-498.

Greenacre, M. "Basic Structure Displays of a Data Matrix," University of South Africa, 80/2, Department of Statistics Research Report (1980).

Guttman, L. "A General Nonmetric Technique for Finding the Smallest Coordinate Space for a Configuration of Points," Psychometrika, 33, (4), (1968) 469-506.

Haber, M. "The Singular Value Decomposition of Random Matrices," Jerusalem, Hebrew University. (1975) Ph.D. Thesis (in Hebrew with English summary).

Householder, A.S. and Young, G. "Matrix Approximation and Latent Roots," American Mathematics Monthly, 45, (1938) 165-171.

Kester, N. "Diagnosing and Fitting Concurrent and Related Models for Two-Way and Higher-Way Layouts," Ph.D. Thesis

at University of Rochester, New York, (1979).

Kruskal, J.B. "Factor Analysis and Principal Components I Bilinear Methods," International Encyclopedia of of Statistics (W.H. Kruskal and J.M. Tanur, eds.), New York: Free Press, (1978) 307-330.

Lingoes, J.C. and Guttman, L. "Nonmetric Factor Analysis: A Rank Reducing Alternative to Linear Factor Analysis," Multivariate Behavioural Research, 2, (1967) 485-505.

Strauss, J.S.; Gabriel, K.R.; Kokes, R.F.; Ritzler, B.A.; VanOrd, A. and Tarana, E. "Do Psychiatric Patients Fit Their Diagnosis? Patterns of Symptomatology as Described with the Biplot," Journal of Nervous and Mental Disease, 167, (1979) 105-113.

Tsianco, M.C. "Use of Biplots and 3-D Bimodels in Diagnosing Models for Two-Way Tables," Ph.D. Thesis, (1980) University of Rochester.

Tsianco, M.C.; Odoroff, C.L.; Plumb, S. and Gabriel, K.R. "BGRAPH -- Program for Biplot Multivariate Graphics: User's Guide," (1981) University of Rochester, Statistics Technical Report 81/20.

Tukey, J.W. Exploratory Data Analysis, New York: Addison-Wesley (1977).

APPENDIX

Table A: Weather stations with latitudes, longitudes and altitudes

	Station	Latitude	Longitude	Altitude
1.	Rio de Janeiro	-22.90	43.17	27.
2.	Brusque	-27.10	48.93	24.
3.	Cuiaba	-15.58	56.10	165.
4.	Puerto Casado	-22.28	57.87	87.
5.	Asunción	-25.27	57.63	64.
6.	Goya	-29.13	59.27	36.
7.	Concepcion	-16.25	62.05	490.
8.	Ceres	-29.88	61.95	88.
9.	Salta	-24.85	65.48	1226.
10.	Cochabamba	-17.38	66.17	2570.
11.	Antofagasta	-23.47	70.43	122.
12.	Valparaiso	-33.02	71.63	41.
13.	Salvador	-12.95	38.48	9.
14.	Olinda	-8.02	34.85	62.
15.	Turiacu	-1.72	45.40	6.
16.	Santarém/Taperinha	-2.42	54.70	20.
17.	Uaupés	-0.13	67.08	85.
18.	Riberalta	-1.00	66.08	172.
19.	Trinidad	-14.75	64.80	236.
20.	Cobija	-11.07	68.73	260.
21.	Quito	-0.22	78.50	2818.
22.	Lambayeque	-6.70	79.90	18.
23.	Lima	-12.07	77.03	137.
24.	Huancayo	-12.03	75.33	3350.
25.	Cayenne	4.83	52.37	9.
26.	Georgetown	6.82	58.18	2.
27.	San Fernando	7.90	67.42	73.
28.	Caracas	10.50	65.92	1042.
29.	San Antonio	7.85	72.45	404.
30.	Bogota	4.63	74.08	2556.
31.	Balboa Heights	8.95	79.55	36.
32.	San José	9.93	84.08	1172.
33.	Bluefields	12.00	83.72	12.
34.	San Salvador	13.72	89.20	695.
35.	Fort-de-France	14.62	61.07	146.
36.	Guatemala	14.58	90.53	1502.
37.	San Juan	18.47	66.12	14.
38.	Port-au-Prince	18.55	72.33	41.
39.	Nassau	25.05	77.47	10.
40.	Tapachula	14.90	92.27	168.
41.	Swan Island	17.40	83.93	3.
42.	Tampico	22.20	97.85	20.
43.	Acapulco	16.83	99.93	3.
44.	Mazatlan	23.18	106.42	78.
45.	Monterrey	25.67	100.30	534.
46.	Guaymas	27.92	110.88	4.
47.	Galveston	29.27	94.85	8.
48.	Key West	24.55	81.75	7.
49.	Jacksonville	30.42	81.65	12.
50.	El Paso	31.80	106.40	1200.

Table B: 1951-52 monthly average temperatures (C° x 10) at each station

	1	2	3	4	5	6	7	8	9	10	11	12
1	260.	260.	251.	227.	224.	207.	196.	196.	215.	255.	240.	228.
	255.	253.	255.	234.	227.	219.	211.	224.	217.	238.	239.	260.
2	243.	236.	225.	183.	182.	164.	142.	161.	183.	187.	216.	217.
	252.	240.	241.	188.	190.	149.	149.	190.	173.	193.	221.	236.
3	259.	267.	268.	242.	249.	231.	226.	245.	275.	285.	268.	271.
	268.	268.	263.	241.	245.	217.	239.	266.	269.	273.	273.	265.
4	282.	273.	268.	213.	231.	214.	218.	221.	246.	270.	278.	286.
	293.	278.	277.	224.	232.	164.	214.	241.	237.	255.	267.	287.
5	289.	263.	256.	212.	225.	197.	207.	205.	229.	262.	275.	279.
	299.	286.	277.	219.	226.	147.	198.	220.	212.	240.	262.	299.
6	325.	238.	226.	179.	193.	218.	166.	162.	184.	203.	234.	255.
	288.	277.	271.	191.	192.	112.	153.	164.	165.	200.	224.	260.
7	248.	255.	255.	228.	237.	215.	228.	233.	293.	264.	255.	262.
	257.	252.	257.	232.	229.	186.	221.	257.	255.	263.	259.	264.
8	256.	229.	225.	169.	180.	155.	152.	156.	182.	212.	239.	257.
	286.	248.	262.	183.	181.	90.	134.	145.	154.	187.	222.	244.
9	204.	184.	188.	127.	149.	113.	129.	128.	157.	192.	205.	211.
	223.	192.	206.	159.	150.	64.	109.	133.	148.	181.	188.	204.
10	182.	173.	185.	170.	141.	122.	128.	127.	164.	186.	196.	186.
	178.	175.	195.	177.	159.	144.	141.	161.	176.	209.	205.	190.
11	192.	183.	172.	160.	156.	141.	147.	144.	145.	159.	172.	162.
	207.	209.	189.	159.	159.	127.	135.	141.	152.	152.	163.	165.
12	173.	174.	163.	136.	142.	129.	139.	134.	137.	142.	157.	167.
	186.	189.	180.	146.	141.	112.	119.	120.	135.	136.	151.	166.
13	257.	264.	265.	256.	244.	235.	225.	223.	229.	243.	249.	254.
	263.	270.	262.	262.	243.	242.	236.	233.	245.	248.	256.	258.
14	269.	267.	275.	265.	254.	238.	233.	238.	247.	259.	263.	267.
	272.	278.	267.	262.	252.	245.	240.	238.	253.	256.	263.	268.
15	273.	279.	278.	261.	263.	261.	260.	272.	273.	273.	277.	273.
	272.	267.	263.	257.	255.	265.	259.	264.	272.	272.	276.	277.
16	255.	255.	256.	252.	253.	254.	250.	263.	267.	268.	271.	271.
	257.	255.	256.	256.	253.	252.	253.	261.	263.	270.	265.	262.
17	248.	247.	251.	250.	247.	241.	242.	248.	254.	255.	260.	256.
	251.	254.	264.	253.	249.	245.	241.	245.	248.	256.	254.	255.
18	260.	262.	265.	259.	266.	252.	263.	267.	283.	275.	266.	276.
	269.	270.	269.	259.	260.	239.	254.	279.	273.	279.	260.	266.
19	265.	270.	277.	256.	261.	231.	250.	240.	262.	283.	273.	279.
	273.	265.	282.	259.	255.	223.	245.	266.	265.	278.	270.	282.
20	259.	248.	256.	249.	259.	242.	236.	252.	268.	268.	261.	264.
	256.	250.	260.	241.	238.	202.	216.	232.	238.	264.	257.	252.
21	124.	127.	136.	137.	132.	137.	130.	141.	136.	131.	127.	132.
	138.	135.	133.	134.	135.	134.	134.	138.	137.	136.	130.	135.
22	241.	225.	240.	229.	216.	214.	191.	188.	186.	195.	191.	195.
	235.	248.	239.	222.	204.	184.	160.	173.	176.	177.	189.	211.
23	203.	192.	202.	196.	194.	180.	169.	177.	172.	179.	182.	195.
	219.	233.	230.	199.	174.	153.	148.	153.	160.	164.	167.	187.
24	131.	133.	120.	137.	112.	99.	93.	107.	118.	122.	132.	120.
	120.	117.	118.	124.	112.	106.	111.	118.	118.	128.	126.	129.
25	247.	250.	256.	255.	253.	251.	252.	259.	266.	268.	265.	261.
	255.	260.	262.	262.	256.	255.	250.	255.	263.	268.	262.	258.

Table B (continued)

	1	2	3	4	5	6	7	8	9	10	11	12
26	258.	261.	271.	273.	274.	269.	268.	279.	279.	277.	280.	272.
	266.	270.	282.	278.	274.	271.	269.	275.	280.	284.	272.	269.
27	261.	268.	281.	287.	277.	260.	250.	263.	276.	279.	278.	273.
	270.	280.	295.	278.	273.	265.	259.	261.	271.	277.	273.	272.
28	186.	188.	193.	215.	217.	211.	206.	219.	220.	219.	212.	205.
	195.	200.	213.	218.	225.	217.	212.	220.	218.	217.	206.	196.
29	223.	225.	240.	267.	266.	269.	270.	276.	276.	270.	250.	248.
	240.	251.	258.	271.	283.	279.	275.	276.	278.	264.	246.	238.
30	134.	136.	143.	145.	146.	143.	137.	139.	142.	141.	143.	141.
	144.	142.	148.	149.	146.	142.	138.	138.	136.	140.	140.	142.
31	263.	263.	269.	277.	271.	274.	271.	272.	269.	267.	268.	273.
	272.	274.	283.	287.	273.	266.	268.	270.	267.	264.	264.	262.
32	186.	186.	192.	212.	216.	214.	208.	213.	216.	212.	204.	200.
	192.	196.	206.	210.	215.	210.	206.	207.	204.	203.	193.	194.
33	240.	243.	252.	271.	270.	257.	257.	262.	262.	253.	253.	252.
	247.	239.	262.	270.	278.	267.	264.	268.	275.	264.	263.	248.
34	216.	217.	232.	238.	236.	234.	227.	233.	221.	227.	223.	226.
	225.	222.	233.	241.	238.	234.	230.	228.	226.	222.	223.	217.
35	233.	234.	236.	248.	255.	256.	258.	260.	259.	259.	254.	245.
	237.	239.	245.	251.	261.	261.	256.	261.	259.	262.	250.	240.
36	160.	155.	178.	195.	187.	191.	185.	194.	186.	180.	173.	173.
	164.	170.	190.	193.	191.	186.	186.	188.	185.	167.	176.	162.
37	238.	232.	233.	253.	271.	267.	268.	278.	273.	270.	263.	249.
	242.	239.	247.	253.	263.	268.	263.	272.	268.	269.	256.	239.
38	242.	242.	249.	270.	273.	282.	284.	284.	280.	276.	270.	260.
	247.	257.	268.	267.	281.	285.	283.	284.	278.	281.	269.	253.
39	198.	204.	211.	237.	250.	267.	275.	276.	278.	252.	233.	224.
	207.	203.	229.	228.	248.	273.	276.	275.	269.	259.	232.	202.
40	248.	255.	266.	275.	269.	262.	258.	264.	254.	259.	260.	255.
	254.	261.	272.	278.	268.	260.	262.	261.	258.	257.	258.	252.
41	251.	246.	260.	274.	281.	277.	280.	286.	284.	283.	271.	271.
	263.	258.	274.	275.	281.	277.	278.	282.	282.	269.	267.	253.
42	192.	186.	218.	239.	266.	274.	278.	277.	273.	258.	215.	209.
	213.	212.	232.	240.	261.	271.	274.	282.	268.	238.	216.	195.
43	261.	267.	279.	280.	291.	293.	289.	299.	282.	297.	293.	285.
	282.	280.	282.	300.	298.	282.	291.	293.	282.	284.	286.	272.
44	198.	193.	204.	223.	237.	269.	278.	287.	284.	281.	254.	228.
	213.	207.	186.	212.	236.	268.	283.	283.	286.	281.	248.	202.
45	164.	176.	204.	235.	255.	276.	286.	280.	257.	228.	173.	180.
	186.	199.	199.	219.	255.	268.	283.	299.	264.	219.	176.	153.
46	188.	194.	210.	234.	243.	301.	312.	313.	317.	287.	231.	186.
	189.	195.	180.	216.	274.	292.	312.	310.	317.	299.	219.	192.
47	130.	129.	171.	191.	240.	275.	285.	297.	267.	236.	158.	153.
	169.	156.	164.	191.	233.	278.	281.	289.	261.	203.	161.	131.
48	202.	203.	234.	248.	268.	287.	292.	302.	296.	268.	234.	242.
	225.	214.	247.	244.	272.	288.	285.	292.	284.	261.	239.	207.
49	132.	129.	174.	194.	238.	276.	283.	289.	272.	230.	146.	154.
	149.	142.	172.	191.	248.	292.	286.	281.	257.	202.	163.	118.
50	68.	87.	125.	163.	228.	271.	298.	281.	255.	199.	106.	80.
	96.	86.	104.	172.	222.	282.	274.	287.	246.	186.	97.	67.
Means	220.	217.	226.	224.	230.	228.	228.	232.	237.	237.	228.	226.
	229.	227.	233.	226.	230.	219.	225.	234.	232.	232.	225.	221.

Table C: GH' - biplot/bimodel coordinates of temperatures
(Deviations from means of each month)

G [5]					H [5]				
0.010	0.110	-0.056	0.104	-0.327					
-0.112	0.166	-0.012	-0.020	-0.199					
0.095	0.085	0.188	0.039	-0.069					
0.060	0.179	0.231	-0.164	-0.011					
0.031	0.216	0.180	-0.194	0.020					
-0.073	0.249	-0.019	-0.466	-0.271					
0.058	0.083	0.265	-0.025	0.065					
-0.119	0.246	0.116	-0.256	0.282					
-0.227	0.167	0.288	-0.058	0.293					
-0.202	0.073	0.175	0.406	0.017					
-0.229	0.069	-0.078	-0.079	-0.148					
-0.274	0.068	-0.045	-0.072	-0.019					
0.064	0.086	-0.168	0.118	-0.194					
0.093	0.095	-0.117	0.083	-0.141					
0.135	0.057	0.008	0.040	-0.230	226.59	273.06	-6.13	-41.71	-39.76
0.105	0.037	0.045	0.069	-0.049	248.57	220.99	-12.91	15.16	-34.06
0.075	0.044	-0.041	0.070	0.004	262.02	146.40	-26.04	1.84	-7.37
0.126	0.045	0.070	-0.038	0.003	294.12	5.42	-75.09	51.94	3.56
0.114	0.097	0.085	0.056	0.049	304.24	-60.19	-40.69	-21.59	21.68
0.066	0.076	0.079	-0.056	0.146	316.96	-169.97	-35.97	-57.59	-17.30
-0.317	-0.060	-0.005	0.109	-0.034	324.92	-209.38	2.65	-47.77	6.28
-0.085	0.088	-0.376	-0.155	-0.183	332.44	-203.32	5.59	-33.15	13.49
-0.154	0.075	-0.249	-0.152	0.126	326.22	-120.94	63.79	-19.43	10.21
-0.369	-0.020	0.051	0.288	-0.113	312.19	-3.83	63.59	-3.82	6.52
0.101	0.032	-0.008	0.084	-0.005	286.31	162.60	45.41	31.99	11.44
0.152	0.028	-0.044	0.069	0.032	277.49	203.67	24.92	0.67	39.24
0.146	0.064	-0.120	0.108	0.069	227.17	237.45	0.90	-56.85	-0.30
-0.059	-0.045	-0.025	0.135	0.145	240.76	223.85	-38.51	-33.63	-18.81
0.114	-0.055	-0.086	0.065	0.124	250.01	198.21	-45.20	-22.58	36.13
-0.291	-0.043	-0.047	0.126	0.049	287.46	31.75	-70.90	40.15	23.38
0.142	0.035	-0.162	-0.013	0.036	304.96	-70.29	-27.92	-3.36	11.99
-0.078	-0.045	-0.066	0.062	0.114	342.12	-261.28	-70.60	56.68	-24.67
0.107	-0.010	-0.106	0.141	0.061	338.22	-213.46	7.03	2.65	-9.99
0.001	-0.009	-0.157	0.065	0.041	333.01	-185.99	37.83	-11.58	-10.54
0.080	-0.021	-0.011	0.077	-0.019	329.83	-132.63	41.98	19.78	-10.57
-0.159	-0.068	-0.115	0.088	0.187	316.67	4.54	72.42	45.31	-13.72
0.103	-0.037	0.026	0.031	0.034	290.94	147.62	36.56	46.65	6.53
0.144	-0.033	-0.034	0.057	0.027	258.69	256.51	19.32	18.34	-9.13
0.058	-0.128	0.078	0.040	0.047					
0.112	0.022	-0.182	0.066	0.034					
0.151	-0.012	-0.056	-0.003	0.134					
0.055	-0.147	-0.048	-0.090	0.143					
0.195	0.016	-0.083	-0.003	0.117					
0.065	-0.141	0.347	0.069	-0.136					
0.010	-0.228	-0.125	-0.181	-0.062					
0.095	-0.243	0.315	0.001	-0.387					
-0.038	-0.313	0.053	-0.236	0.018					
0.104	-0.137	0.027	-0.063	0.213					
-0.042	-0.330	-0.010	-0.166	0.046					
-0.137	-0.474	0.023	-0.179	-0.076					

Table D: Station Parameter Estimates for Models [M_θ] and [M].

	[M_θ]				[M]		
	η	ψ	ϕ	θ	η	ψ	ϕ
1	4.	30.	6.2	0.4	5.	27.	5.8
2	-31.	51.	6.2	0.7	-29.	46.	5.9
3	29.	23.	0.5	1.1	29.	20.	0.3
4	20.	49.	0.2	1.6	21.	42.	6.3
5	12.	59.	0.1	1.9	13.	52.	6.2
6	-18.	71.	6.2	2.3	-16.	64.	5.9
7	18.	25.	0.5	2.6	18.	22.	0.4
8	-31.	74.	0.0	3.8	-30.	66.	6.0
9	-64.	60.	0.1	4.2	-63.	52.	6.2
10	-58.	34.	0.3	4.7	-58.	29.	0.0
11	-67.	31.	6.0	5.3	-66.	29.	5.7
12	-80.	33.	6.1	5.5	-79.	30.	5.8
13	20.	21.	5.9	0.6	21.	21.	5.6
14	29.	21.	6.1	0.7	29.	20.	5.7
15	41.	10.	0.3	0.9	41.	9.	0.1
16	31.	8.	0.7	1.5	31.	7.	0.5
17	23.	8.	0.0	1.9	23.	8.	5.9
18	38.	8.	0.6	2.8	38.	7.	0.4
19	35.	22.	0.2	3.7	35.	19.	6.2
20	21.	18.	0.2	3.9	21.	16.	6.3
21	-94.	2.	3.0	4.2	-94.	3.	4.5
22	-24.	34.	5.5	4.7	-22.	35.	5.2
23	-44.	31.	5.7	5.2	-43.	31.	5.4
24	-109.	14.	0.1	5.8	-109.	13.	6.0
25	30.	4.	0.5		30.	4.	0.2
26	45.	1.	3.3		45.	0.	3.7
27	44.	10.	5.9		44.	10.	5.4
28	-18.	9.	4.7		-18.	8.	3.0
29	33.	21.	1.6		32.	19.	3.0
30	-87.	5.	5.2		-86.	6.	4.9
31	42.	6.	4.9		43.	8.	4.7
32	-24.	9.	1.9		-24.	9.	3.3
33	31.	9.	1.9		31.	9.	3.4
34	0.	8.	2.6		0.	9.	4.2
35	23.	10.	4.5		23.	9.	2.7
36	-48.	13.	2.0		-48.	12.	3.5
37	30.	16.	4.5		29.	15.	2.6
38	42.	17.	4.6		42.	15.	2.8
39	15.	38.	4.5		14.	35.	2.7
40	33.	8.	3.0		33.	9.	4.4
41	44.	11.	1.6		44.	10.	2.9
42	14.	44.	1.6		13.	39.	2.9
43	58.	6.	1.8		58.	6.	3.2
44	17.	44.	4.2		16.	44.	2.3
45	0.	65.	1.7		-1.	58.	3.0
46	25.	71.	4.5		23.	67.	2.5
47	-16.	84.	4.7		-17.	76.	2.8
48	29.	43.	4.6		28.	39.	2.8
49	-17.	89.	4.7		-19.	79.	2.9
50	-47.	124.	1.6		-49.	111.	2.9

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A114 169	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Modeling Temperature Data: An Illustration of the Use of Biplots and Bimodels in Non-linear Modeling		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Michael C. Tsianco K. Ruben Gabriel		8. CONTRACT OR GRANT NUMBER(s) N00014-80-C-0387
9. PERFORMING ORGANIZATION NAME AND ADDRESS Division of Biostatistics University of Rochester Medical School Rochester, New York 14642		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of the Naval Research Arlington, Virginia, 22217		12. REPORT DATE September, 1981
		13. NUMBER OF PAGES 63
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The paper describes a strategy for exploration and modeling of multivariate data. The data illustrated are temperatures at 50 North, Central and South American stations; the techniques rely heavily on lower rank approximation, biplot and binomial display, least squares fitting, and examination of residuals. The biplot/bimodel display is shown to reveal the relation of temperatures to various geographical features, (over)

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20 continued:

(No new meterorological findings are claimed -- the example is used mainly to illustrate the technique.) More strikingly, the bimodel shows an elliptical configuration of the markers for the different months. The paper shows how this ellipse implies a harmonic model for the temperatures themselves. This model makes good physical sense and its parameters are readily understood in meteorological terms. Some summary comments are made about the usefulness of biplots and bimodels for diagnosing models that may fit the data.


END
DATE
FILMED
5-82
DTIC