

A Note on the Spatial Structure of the Covariability of Observed Northern Hemisphere Surface Air Temperatures

J. B. ELSNER¹ and A. A. TSONIS²

Abstract—The spatial structure of covariability in the surface air temperature records of the last 109 years is examined using factor analysis. From the results it is indicated that the Northern Hemisphere can be objectively divided into 15 regions with each region having a similar covariance structure of surface air temperatures. From such a division it is noted that warming is evident over 11 out of the 15 regions. The strongest warming is not, however, confined to specific latitudinal belts, for example, the tropics or the arctic, nor is the strongest warming confined to a specific continent. As a consequence of the coarse spatial distribution of available data, results should be considered preliminary.

Key words: Hemispheric temperature records, factor analysis, long-term trends.

1. Introduction

There have been indications recently that there exists a distinct warming trend in near-surface global air temperatures over the past century (FOLLAND *et al.*, 1984; JONES *et al.*, 1986; HANSEN and LEBEDEFF, 1987). The global mean temperature trend which amounts to nearly 0.5°C/100 yrs is based on station observations on land as well as over the oceans. The warming appears most significantly in the recent past with five or six of the warmest years on record (depending on which data set is considered) all occurring since 1980 (TSONIS and ELSNER, 1989). Warming trends in surface temperature are evident in both the Northern and Southern Hemispheres. The warming trend over the Northern Hemisphere is most pronounced since about 1970 (Figure 1).

The observed global and hemispheric warming trends are in the right direction and magnitude hypothesized for a warming resulting from increases in the concentration of greenhouse gases in the atmosphere. Increases in the concentration of the primary greenhouse gas carbon dioxide (CO₂) have been observed to be greater than 8% since 1960 and approximately as much as 25% since the beginning of the

¹ Department of Meteorology, Florida State University, Tallahassee, FL 32306, U.S.A.

² Department of Geosciences, University of Wisconsin-Milwaukee, Milwaukee, WI 53201, U.S.A.

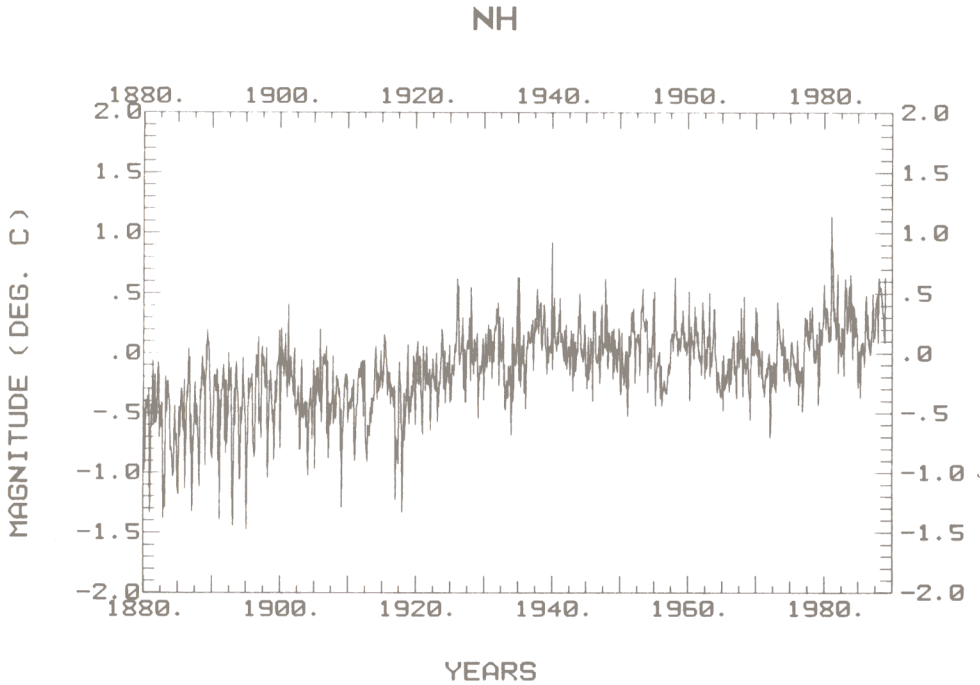


Figure 1

Monthly surface air temperature departures in °C degrees over the Northern Hemisphere for the period 1880–1988 from the data of HANSEN and LEBEDEFF (1987, 1988). Note that the most pronounced warming occurs from 1920–1940 and again from 1970–present.

industrial age, largely as a result of burning fossil fuels (oil, gas and coal) and global deforestation (MAXWELL, 1987). Due to increasing concentrations of other greenhouse gases such as methane and chlorofluorocarbons, the combined radiative effect of the gases is equivalent to an increase of 40% in the CO_2 concentration, at least half of which has occurred in the last 30 years (PARKER, 1989).

Numerical studies using general circulation models suggest that globally a warmer lower troposphere can be expected as a result of increases in CO_2 concentrations in the model atmosphere (WASHINGTON and MEEHL, 1984; SCHLESINGER and MITCHELL, 1987; HANSEN *et al.*, 1988). Although there is no consensus among the various numerical model studies regarding regional temperature trends, there is an agreement on the interpretation that the atmospheric response to increased greenhouse gases will not be uniform over the entire globe. That is, some regions may experience a marked surface warming while other regions will experience no warming or perhaps even a cooling. However, regional details are still quite uncertain in the models.

Observational studies of regional long-term temperature variations also suggest the possibility that all areas of the globe may not experience a significant warming

trend. JONES (1988), for example, shows, using linear trends, that the recent warming (past 20 years) has been strongest in the middle latitudes and weaker over the polar latitudes. In this study we are interested in objectively determining regions of the Northern Hemisphere based on the covariability structure of the temperature data and then assessing the long-term trends in each of the regions.

2. Data and Method of Analysis

The data set used in this study was developed by HANSEN and LEBEDEFF (1987, 1988). The data are long-term (1880–1988) series of area-averaged surface temperature departures based on monthly mean surface values and a reference period of 1951–1980. A time series of temperature departures is available at all 80 nearly-equal area boxes (40 in the Northern and 40 in the Southern Hemisphere) covering the entire globe. Figure 2 shows the arrangement of the boxes in the Northern Hemisphere. Data used in the area-averaging procedure are primarily from monthly station reports published in *World Weather Records* and continued in *Monthly Climate Data for the World*. A complete description of the procedure employed for defining the temperature departures used in this study is given in HANSEN and LEBEDEFF (1987). Table 1 is a list of the temporal means and standard deviations of the records from each of the 40 Northern Hemisphere boxes along with the number of observations comprising each record.

It should be noted that there exist many difficulties in constructing representative and useful temperature records. For example, the oceans are not adequately presented, the quality of the measurements and the density of the reporting stations have not remained consistent throughout the period of record and urbanization effects are not included. Nevertheless, a careful construction with what is available has been performed. In fact, using different data archives and largely independent construction techniques, research groups in England (JONES *et al.*, 1986) and in the USSR (GRUZA *et al.*, 1988) have developed quite similar hemispheric monthly temperature records, although long-term trends in these data have been ascertained as significantly different (ELSNER and TSONIS, 1991).

The primary limitation of using these data in large-scale analyses like this one is the incomplete spatial coverage of reporting stations. Spatial coverage is particularly limited in the Southern Hemisphere where, not until the middle 1960s, do we have observations in all of the 40 Southern Hemisphere boxes. Since the statistical technique we employ in this study requires that the number of variables (in our case boxes) be much less than the number of observations, we are forced to consider only the Northern Hemisphere. The size or dimension of the data set used is thus determined by the number of boxes which total 40 (Northern Hemisphere only) and by the number of observations in each box which total 1308 (monthly temperature departures for the period 1880 through 1988).

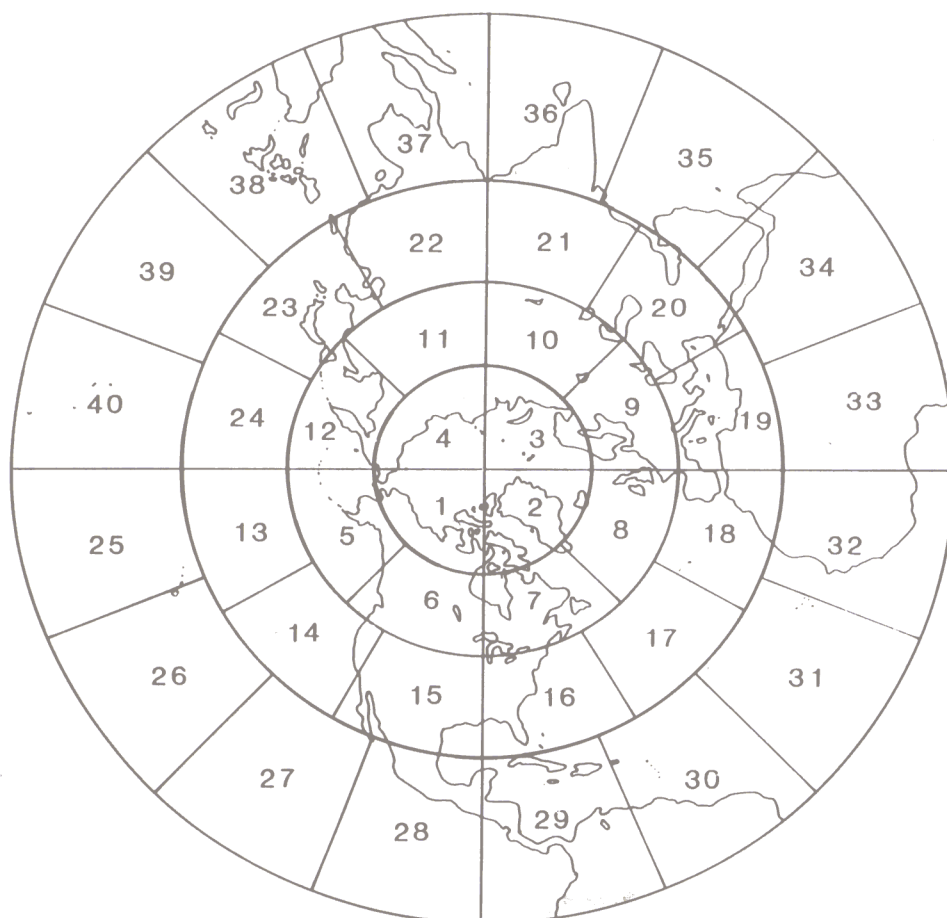


Figure 2

Locations of the 40 Northern Hemispheric nearly equal-area boxes used in the study. A single temperature departure record is available for each box. Table 1 is a list of the temporal mean and standard deviation of the temperature record along with the number of stations used to estimate the area departure for each box. Factor analysis is performed using the 40 temperature records.

The objective is to divide the Northern Hemisphere into distinct regions and then determine which of these regions supports a trend in mean temperature departures in the past 109 years. Therefore, as a first step we seek to reduce the dimensions of the data by finding a smaller number of regions that have the same covariability of temperature departures as the original temperature departures in the 40 Northern Hemisphere boxes. One of the most efficient methods of reducing the dimensionality of the data in this way is the use of factor analysis. Factor analysis is a commonly used procedure that reveals how many statistical "factors" are necessary to describe the correlations that exist in a data set (RICHMAN, 1986).

Table 1

Temporal mean and standard deviation of the temperature records along with the number of observations comprising the record in each of the 40 Northern Hemisphere boxes

Box No.	Temporal Mean	Standard Deviation	No. of Observations
1	-0.35	2.460	27
2	-0.38	1.842	23
3	-0.10	1.942	45
4	0.01	2.157	18
5	0.10	1.626	21
6	-0.02	1.851	49
7	-0.30	1.324	53
8	-0.08	0.618	52
9	-0.05	1.380	224
10	-0.15	1.979	37
11	0.00	1.765	40
12	-0.40	1.364	24
13	-0.38	0.703	2
14	-0.31	1.031	11
15	-0.17	1.052	61
16	-0.09	1.042	52
17	-0.06	0.627	4
18	0.00	0.647	63
19	0.00	0.803	138
20	-0.04	0.959	119
21	-0.10	0.900	64
22	0.03	0.754	83
23	-0.39	0.895	84
24	-0.34	0.879	2
25	-0.42	0.626	7
26	-0.35	0.646	7
27	-0.27	1.142	1
28	-0.24	0.599	28
29	-0.26	0.469	142
30	-0.32	0.577	54
31	0.17	0.613	3
32	0.18	0.512	73
33	0.18	0.700	110
34	-0.07	0.609	64
35	-0.18	0.458	9
36	-0.07	0.384	41
37	0.15	0.405	69
38	-0.14	0.466	30
39	-0.10	0.588	7
40	0.11	0.461	7

It has been successfully applied to a variety of fields, including archaeology, climatology, economics and geography (ARCHER *et al.*, 1988).

The factor analysis procedure begins with computing linear correlation coefficients between all temperature time series (one in each of the 40 boxes) to form the spatial correlation matrix (see Figure 3). The correlation matrix is then

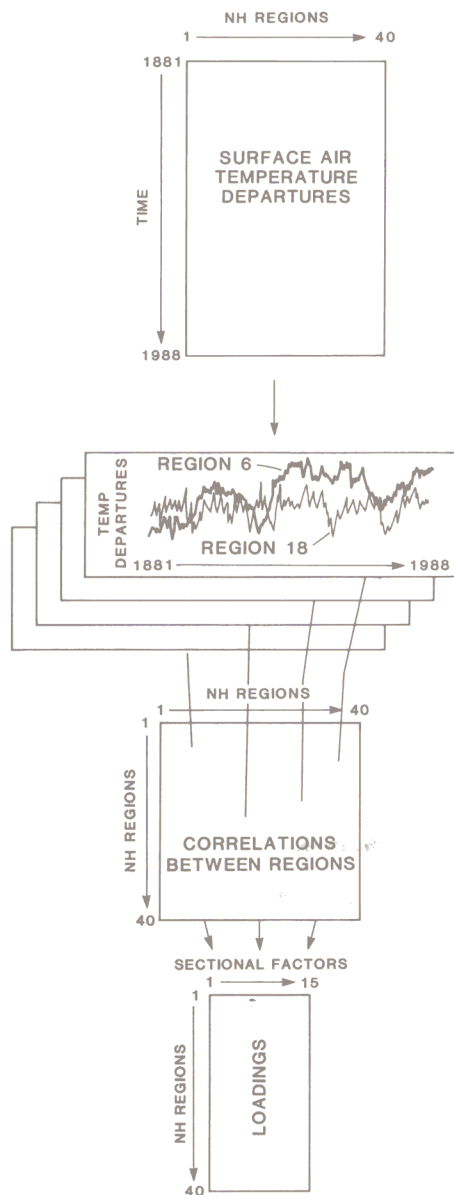


Figure 3

Schematic diagram showing the procedure of factor analysis. Factor analysis begins with a data matrix, in this case consisting of the monthly temperature records over the past 109 years. In search of factors that underlie similarities in the temperature records of the different boxes (regions), the correlation matrix is formed consisting of correlations between every pair of temperature records. Factor analysis revealed that 15 factors could account for most of the correlations when the temperature records were given "loadings" for each factor.

decomposed using a principal component (eigenvector) common factor model (JOHNSON and WICHERN, 1982) whereby each of the 40 boxes receives a specific "loading" for each factor, representing the degree to which the factor reflects the boxes' temperature trends. After the loadings have been extracted from the correlation matrix they are linearly transformed (rotated) as an aid in their interpretation. In this study we apply the commonly used varimax rotation method (RICHMAN, 1986) to linearly transform the loadings.

3. Results

The key decision to be made in any factor analysis is how much factors are important above the noise level in the data to adequately describe the covariance relations. Our decision was based on the so-called Rule-U procedure (PREISENDORFER *et al.*, 1981) where a matrix of random normal deviates, of the same size as the original data matrix (1308 observations by 40 variables) is formed into a correlation matrix and then decomposed into a series of eigenvalues. A number of trials (e.g., 100) are completed, confidence intervals drawn and then compared to the original data eigenvalues. Only the original data eigenvalues exceeding the upper 5% significance level are retained. In this way we find that 15 (out of a possible 40) factors are needed to adequately explain the spatial covariance in the Northern Hemisphere monthly temperature departure records, as is evident in Figure 4. The first twenty eigenvalues from the original data and their corresponding cumulative percentage of explained variance are displayed in Table 2.

To aid in their interpretation, the 15 extracted factors are subjected to a varimax rotation resulting in patterns which are constrained to be orthogonal. Evidence of the effectiveness of this linear least-squares transformation is seen in Figure 5, where the rotated factors are plotted against one another with the points tending to cluster near the origin and along the orthogonal axes. The tendency for the rotated loadings to cluster along the axes is indicative of simple structure in the original temperature fields and provides confidence in a correct application of the factor analysis procedure.

The relative loadings on each of the 40 boxes for the first factor are shown in Figure 6a. The deepest yellow region indicates the boxes that have largest loadings (≥ 0.80) with lighter shades indicating boxes having smaller loadings or loadings having an opposite sign. Largest loadings for the first factor are found over the central North Pacific Ocean including Hawaii. This result is consistent with the study done by BARNETT (1978), who showed that the components of the first eigenvector of surface temperature anomalies over the Northern Hemisphere for the period 1950–1977 are most significant over the central and eastern North Pacific Ocean.

Table 2

Magnitude and corresponding percentage of variance explained for the first 20 eigenvalues of the correlation matrix

	Eigenvalue	Cumulative % of Explained Variance
1	5.80	21.1
2	3.19	32.6
3	2.44	41.5
4	2.29	49.8
5	1.79	56.3
6	1.69	62.4
7	1.36	67.3
8	1.34	72.2
9	1.25	76.7
10	0.98	80.3
11	0.86	83.4
12	0.71	86.0
13	0.67	88.4
14	0.48	90.1
15	0.46	91.8
16	0.36	93.1
17	0.32	94.3
18	0.31	95.4
19	0.29	96.4
20	0.24	97.3

Table 3

Regions based on the factor analysis and their corresponding color as shown in Figure 6b

Factor	Region	Color
1	Central Pacific	Yellow
2	Europe	Beige
3	Southeast Asia	Light Orange
4	Central Asia	Light Yellow
5	Northern North Atlantic	Purple
6	Western North America	Orange
7	India	Dark Blue
8	Western Arctic	Light Blue
9	Central Africa	Light Green
10	Eastern North Atlantic	Light Brown
11	Eastern Arctic	Red
12	Eastern Canada	Green
13	Central America	Blue
14	Marshall Islands	Gray
15	Japan	Dark Brown

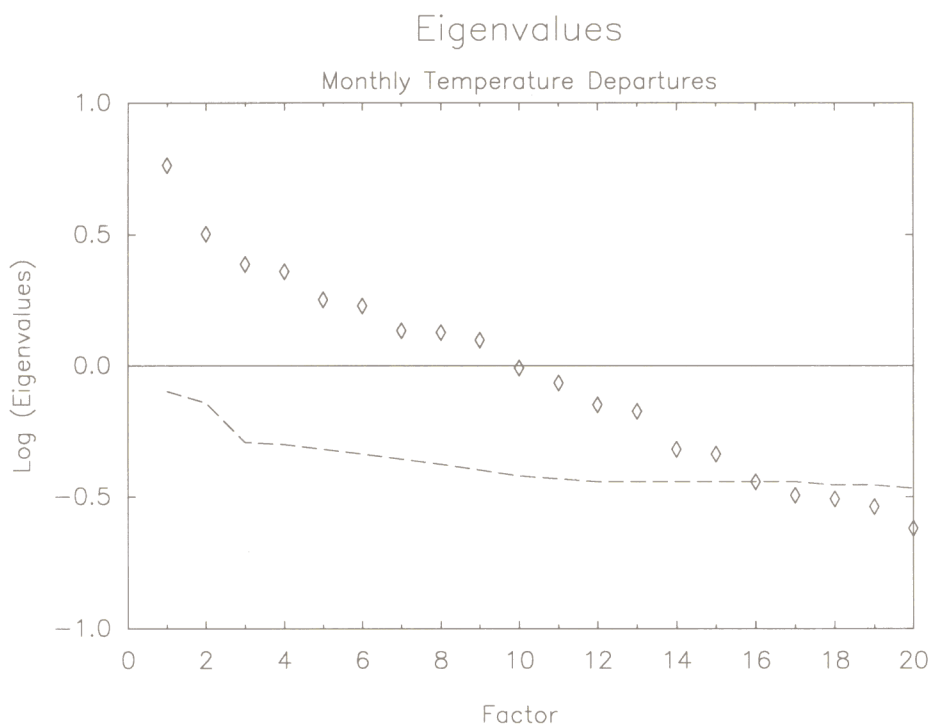


Figure 4

A plot of the logarithm of the eigenvalues extracted from the temperature departure correlation matrix. The solid line indicates the 95% significance level determined by 100 simulations using correlation matrices constructed from random normal deviates. Based on this graph we conclude that 15 factors are significant.

Rotation of the factor loadings makes it easy to decide to which region a particular box belongs since, for each factor, loadings are generally large on a few boxes and small on the remaining boxes as is evident in Figure 5. We thus divide the hemisphere by first finding the box having the largest magnitude loading for that factor and then including all other boxes of the same sign having a magnitude of 0.40 or greater into that factor region (group). There are a few exceptions to this grouping procedure. The exceptions occur with boxes (variables) which are called "complex" (RICHMAN, 1986). Complex variables are characterized by moderate loadings over several factors with no large loading in any single factor. In this analysis complex variables (boxes) account for 5 out of the 40, or 12% of the boxes. In a sense, the number of complex boxes represents the degree of overlapping between the regions. For these cases we choose an appropriate region based on the value of the moderate loading and its location with respect to nearby factors so as not to geographically divide any factors. A composite map is drawn (Figure 6b) displaying

the regions of largest loadings for each of the 15 factors. Each region defined by a single factor groups boxes with similar temperature records.

As is evident in Figure 6, factor analysis is useful in objectively dividing the Northern Hemisphere into 15 distinct regions. Each region represents boxes of monthly temperature departures that have a similar covariance structure. The next step in our analysis is to average the records from each box within each region and evaluate the 110 year trends separately. We find that 11 of the 15 regions show positive trends. The trend and therefore the warming, however, is not the same in all these regions. The most pronounced positive trends are found over the central Pacific (factor 1), western North America (factor 6), and portions of the Arctic (factor 8). Three regions, including much of Europe (factor 2), eastern North Atlantic (factor 10), and the Marshall Islands (factor 14) show no significant trends and the region including the United Kingdom and portions of the northern Atlantic (factor 5) shows a slight negative (cooling) trend. Note that trend estimations should be viewed with caution due to differences in the number of data points in the various regions (i.e., regions have varying number of boxes and boxes have varying

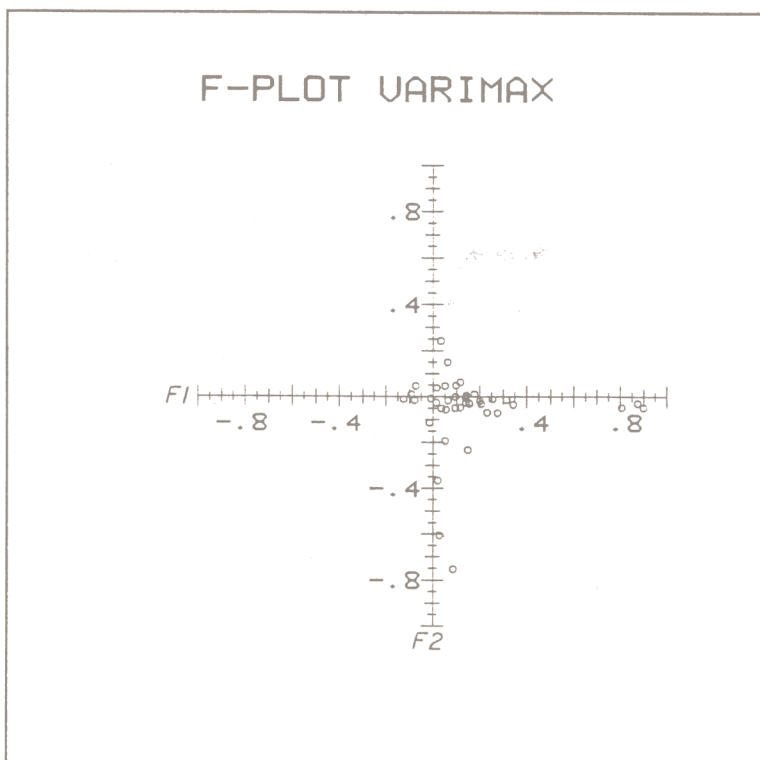


Figure 5(a)

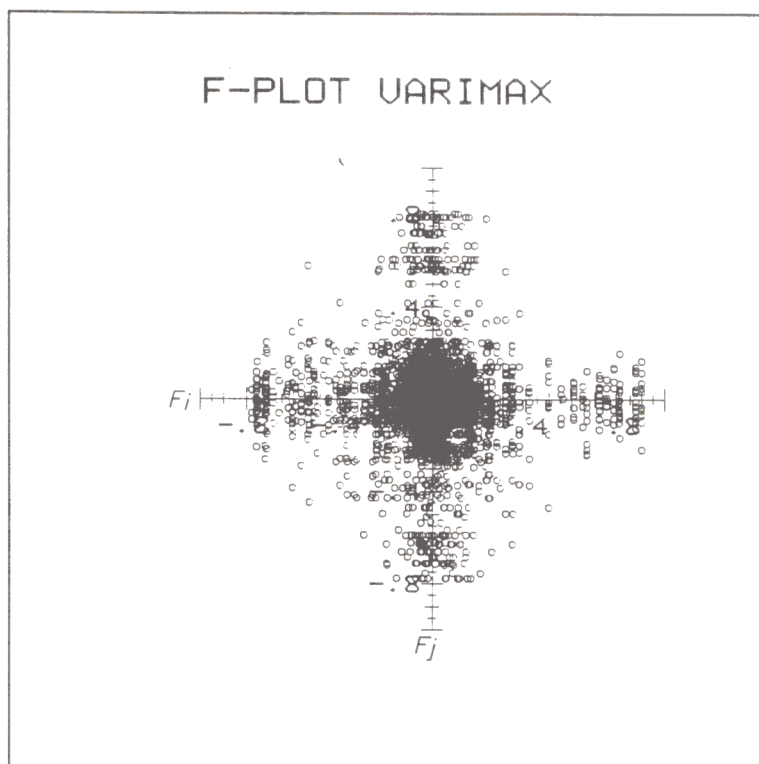


Figure 5(b)

Varimax rotated factor plots constructed by plotting one factor loadings against another factor loadings: (a) factor one versus factor two, and (b) combination of all factor plots. Note that the rotated loadings tend to cluster near the origin and along the axes, indicative of simple structure in the hemispheric temperature field.

number of observations). In fact some of the oceanic boxes (e.g., boxes 13, 24, and 27) contain only a few observations over the 110-year period. It should also be noted that since the factor analysis procedure groups the boxes based on maximizing correlations within a particular group while minimizing intergroup correlations, temperatures over a particular region will in general not be correlated with temperatures over another region even if both regions show warming trends.

4. Conclusions

The results presented here are new, indicating that it is possible to objectively divide the Northern Hemisphere into 15 coherent regions based on the covariability of the long-term temperature records using factor analysis. Based on the coherent regions we substantiate a warming trend over most regions of the hemisphere. The

data reveal that the most pronounced warming is not confined to a specific latitudinal belt, such as the tropics. The significance of these results is that they provide a challenge for the models that are trying to predict the regional variations in temperature. Is such an observed spatial structure, for instance, a result of natural variability of the system? Can the models reproduce such a temperature covariability structure as the concentration of CO_2 is increased by some amount? A worthwhile model verification technique would be to compare a factor decomposition of model generated hemispheric monthly temperatures with the results of this study.

It should be noted that covariability, being a higher order statistic than the mean, is more sensitive to nonuniformities in time and space. Therefore, the results presented here should be regarded as preliminary.

Finally, we note the emergence of methodologies for analyzing and/or making predictions on time-series data that exploit the inherent nonlinearities involved in the underlying dynamical processes responsible for the observed signal (see e.g., TSONIS and ELSNER, 1988; SUGIHARA and MAY, 1990). In order for these techniques to be of practical use a low-dimensional state space (embedding space) is necessary. Present results indicate that a high-dimensional embedding space for large-scale surface air temperatures can be reduced to a considerably lower dimensional subspace, allowing for more effective analysis of the data. We are currently exploring these ideas with the hemispheric temperature data.

Acknowledgement

We thank Dr. Hansen for providing us with the data. Part of this research was supported by NOAA grant # NA-16RC 0454-01.

REFERENCES

- ARCHER, J. C., SHELLEY, F. M., TAYLOR, P. J., and WHITE, E. R. (1988), *The Geography of U.S. Presidential Elections*, Sci. Amer. 259, 44–51.

Figure 6

(a) Relative loadings for each box for the first factor revealed by factor analysis of the boxes monthly temperature departure record. A deeper color shows a higher loading. Boxes 13, 25, and 26, over the central Pacific Ocean have factor loadings exceeding 0.4 and are thus shaded a deep yellow, while the lighter yellow region indicates boxes having the first factor loading between 0.2 and 0.4. The uncolored regions in this figure indicate boxes having factor loadings less than 0.2. (b) This figure is produced by continuing the procedure of isolating boxes having factor loadings exceeding 0.4 as in Figure 6a and assigning a different color for each of the 15 factors. Each region, defined by a single factor, groups boxes having a similar temperature departure record. Some slight overlapping of the factor loadings did occur, where a few boxes had loadings exceeding 0.4 for two different factors. How these few cases were handled is explained in the text. Table 3 is the color scheme used in the figure.

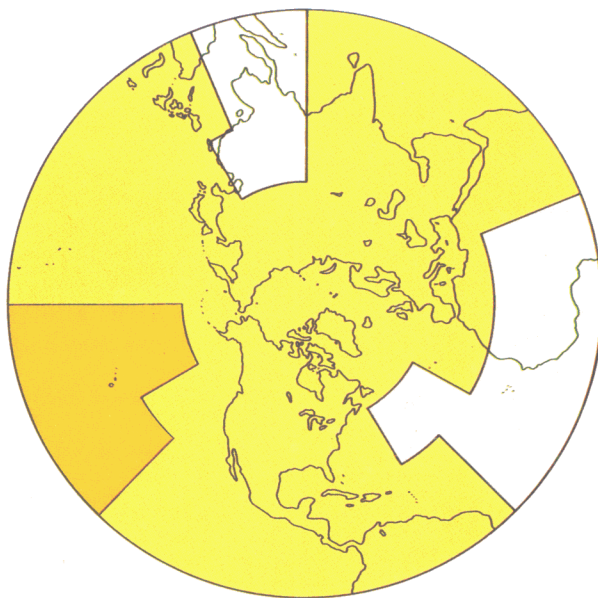


Figure 6(a)

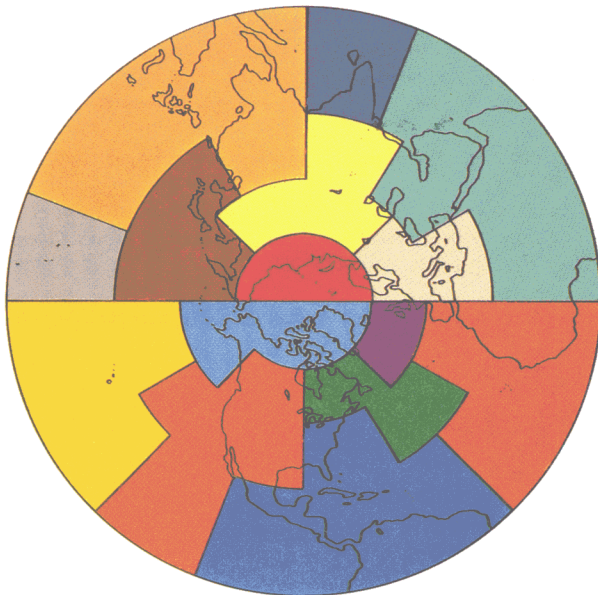


Figure 6(b)

- BARNETT, T. P. (1978), *Estimating Variability of Surface Air Temperature in the Northern Hemisphere*, Mon. Wea. Rev. 106, 1353–1367.
- ELSNER, J. B., and TSONIS, A. A. (1991), *Comparisons of Observed Northern Hemisphere Surface Air Temperature Records*, Geophys. Res. Lett. 18, 1229–1232.
- FOLLAND, C. K., PARKER, D. E., and KATES, F. E. (1984), *Worldwide Marine Temperature Fluctuations 1856–1981*, Nature 310, 67–673.
- GRUZA, G. V., RAN'KOVA, E. YA., and ROCHEVA, E. V. (1988), *Analysis of Global Data Variations of Surface Air Temperature during Instrument Observation Period*, Meteor. Gridr. 16–24.
- HANSEN, J., and LEBEDEFF, S. (1987), *Global Trends of Measured Surface Air Temperature*, J. Geophys. Res. 92, 13,345–13,372.
- HANSEN, J., and LEBEDEFF, S. (1988), *Global Surface Air Temperatures: Update through 1987*, Geophys. Res. Lett. 15, 323–326.
- HANSEN, J., FUNG, I., LACIS, A., RIND, D., LEBEDEFF, S., RUEDY, R., RUSSELL, G., and STONE, P. (1988), *Global Climate Changes as Forecast by Goddard Institute for Space Studies Three-dimensional Model*, J. Geophys. Res. 93, 9341–9364.
- JOHNSON, R. A., and WICHERN, D. W., *Applied Multivariate Statistical Analysis* (Prentice-Hall, Inc. 1982).
- JONES, P. D. (1988), *Hemispheric Surface Air Temperature Variations: Recent Trends and an Update to 1987*, J. Climate 1, 654–660.
- JONES, P. D., WIGLEY, T. M. L., and WRIGHT, P. B. (1986), *Global Temperatures Variations between 1861 and 1984*, Nature 322, 430–434.
- MAXWELL, B. (1987), *Atmospheric and Climate Change in the Canadian Arctic: Cause, Effects, and Impacts*, North. Persp. 15, 2–6.
- PARKER, D. E. (1989), *Observed Climatic Change, and the Greenhouse Effect*, Met. Mag. 118, 128–131.
- PREISENDORFER, R. W., ZWIERS, F. W., and BARNETT, T. P. (1981), *Foundations of Principal Component Selection Rules*, SIO Reference Series 81–4, Scripps Institute of Oceanography, La Jolla, CA.
- RICHMAN, M. B. (1986), *Review Article: Rotation of Principal Components*, J. Climatol 6, 293–335.
- SCHLESINGER, M. E., and MITCHELL, J. F. B. (1987), *Climate Model Simulations of the Equilibrium Climatic Response to Increased Carbon Dioxide*, Rev. Geophys. 25, 760–798.
- SUGIHARA, G., and MAY, R. M. (1990), *Nonlinear Forecasting as a Way of Distinguishing Chaos from Measurement Error in Time Series*, Nature 344, 734–741.
- TSONIS, A. A., and ELSNER, J. B. (1988), *The Weather Attractor over Very Short Time Scales*, Nature 333, 545–547.
- TSONIS, A. A., and ELSNER, J. B. (1989), *Testing the Global Warming Hypothesis*, Geophys. Res. Lett. 16, 795–797.
- WASHINGTON, W. M., and MEEHL, G. A. (1984), *Seasonal Cycle Experiment of the Climate Sensitivity Due to a Doubling of CO₂ with an Atmospheric General Circulation Model Coupled to a Simple Mixed Layer Ocean Model*, J. Geophys. Res. 89, 9475–9503.

(Received February 4, 1991, revised August 12, 1991, accepted August 31, 1991)