

NOTES AND CORRESPONDENCE

Removal of Systematic Biases in S-Mode Principal Components Arising from Unequal Grid Spacing

DIEGO C. ARANEO AND ROSA H. COMPAGNUCCI

Departamento de Ciencias de la Atmósfera y los Océanos, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

10 October 2002 and 19 June 2003

ABSTRACT

Frequently, physical variables are analyzed using gridded fields, on regular latitude–longitude frameworks. Such networks often concentrate a disproportionate number of observations over polar regions. If these types of grids are used for an S-mode principal component analysis, they produce a bias of the component patterns toward the temporal patterns observed at higher latitudes. A method to potentially eliminate this effect, while employing the covariance similarity matrix, is to weight the variables by the square root of the cosine of the latitude of the point at which the datum was observed. However, this processing is not useful when using the correlation similarity matrix. In this case, a spatially uniform or equal density grid can be designed by means of the criteria of a constant density of points in each circle of latitude. Then, the variable values are linearly interpolated into this new equal-density grid. This technique is easy to program and to adapt to any regular latitude–longitude network.

As an example, an application of the technique is presented for monthly anomalies of 70-hPa temperature data collected by the Microwave Sounding Unit (MSU) flown on board the NOAA Television Infrared Observation Satellites (TIROS-N). A regular latitude–longitude network generates an overestimation of the polar area significance with respect to those obtained by the equal-density criterion. By comparing two sets of point distributions, it is shown that the representative patterns of the temporal evolution of the variables at low or midlatitudes are modified little, if any, by creating a network of equal density points. Nevertheless, changes may be observed in the explained variances because the eigenvalues are impacted directly to the change in the number of points in the networks caused by processing a non-equal-density grid to the equal-density one.

1. Introduction

Eigentechniques, such as the principal component analysis (PCA) and the empirical orthogonal function (EOF) analysis, are statistical tools that have been widely used in climatological research, particularly in the study of spatial–temporal variability of physical variables. These techniques can be applied to data in various ways according to the aim of the analysis (Cattell 1952). One such mode, S-mode analysis, that treats stations or grid points as variables, is mostly used to discriminate spatial groupings of temporal patterns from a set of physical variables observed over a given space. The popularity of this mode arises from its ability to determine homogeneous spatial areas. Each eigentechnique

is derived directly from a parent similarity matrix that typically consists of either a correlation or covariance matrix. Furthermore, EOF uses unit-length eigenvectors, whereas in PCA each eigenvector is weighted by the square root of its corresponding eigenvalue (Ehrendorfer 1987). The weighted eigenvectors (or component loadings) represent the correlations or covariances between each variable and each principal component, depending upon which parent similarity matrix is employed (Jolliffe 1995). The eigenvalues and eigenvectors of the correlation matrix have no simple relationship with those of the corresponding covariance matrix. Therefore, the principal components (PCs) for correlation and covariance matrices do not give equivalent information and cannot be derived directly from each other (Jolliffe 1986).

Historically, the various similarity matrices have been discussed and compared in meteorological literature, with correlation or covariance being the most commonly applied (Glahn 1965; Craddock and Flood 1969), though cross products (Molteni et al. 1983) and Eu-

Corresponding author address: Dr. Diego C. Araneo, Departamento de Ciencias de la Atmósfera y los Océanos, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria—Pabellón 2—2do piso, Ciudad de Buenos Aires 1428, Argentina.
E-mail: araneo@at.fcen.uba.ar

clidean metrics (Elmore and Richman 2001) have also been used. An argument for using correlation matrices to define PCs, rather than covariance matrices, is that the results of analyses for different sets of random variables are more directly comparable than for analyses based on covariance matrices. Sizes of variances of PCs have the same implications for different correlation matrices, but not for different covariance matrices. Also, patterns of coefficients in PCs can be readily compared for different correlation matrices to see if the two correlation matrices are giving similar PCs, whereas informal comparisons are often much more involved for PCs from covariance matrices. Furthermore, if there are differences between the magnitudes of the variances of the variables, then those variables whose variances are largest will tend to dominate the first few PCs (Jolliffe 1986). Consequently, if the analyzed variables show temporal variances similar to one another throughout the spatial domain, the researcher can legitimately perform the study from the covariance matrix of the variables. If not, the variables should be standardized, that is to say, one should perform the analysis on the correlation matrix.

In both cases, the resulting patterns produced by the PC or EOF analysis are often physically interpreted and/or used in subsequent analyses. When irregularly spaced data with systematic density biases in the location of the data points are used, the relation between the PC/EOF patterns and the actual physical patterns is likely to be distorted to some extent. A PCA, run in S mode, is sensitive to changes in the differences in the density of stations or grid points on a given area. This problem can be overcome by interpolating the data to an equal-area grid, as in Karl et al. (1982), or by using a fairly uniform distribution of stations, as in Richman and Lamb (1985).

Frequently, physical variables are analyzed using fields gridded on a regular latitude–longitude network. Such a network concentrates a large number of observational points over polar regions, producing a bias of the patterns derived from the PCA toward the temporal patterns observed at higher latitudes. When covariance similarity matrices are used it is possible to eliminate this effect through the use of weighted variables. Weights are chosen to reflect some a priori idea of the relative importance of the variable. Thompson and Wallace (2000) multiplied the original variables by the square root of the cosine of the latitude of the data point. Thus, they reduced the variances proportionally to their proximity to the Poles. However, this processing is not useful when using the correlation similarity matrix because any weighting is removed by the standardization employed while calculating the correlation coefficients. Consequently, analyses of variables that exhibit differences in temporal variances over the spatial field are good candidates for use of the correlation matrix. In such cases, the solution would be to apply PCA over an equal spatial density data grid in order to avoid the

overemphasis of PCs in the polar regions that might be produced by the use of regular grids. Examples of this general idea can be seen in the literature. The general circulation model (GCM) data analyzed in Richman and Gong (1999) was initially on a Gaussian grid that was not useful for statistical processing. In order to lessen this problem of convergence of the longitude lines toward the Poles, the authors interpolated the model grid onto the National Meteorological Center's octagonal grid, which more closely approximate equal area than the original output grid; however, it was not a *perfect* equal-area grid. In a similar approach, Barnston and Livezey (1987) devised a more equal-area grid by reducing gradually the number of grid points between 60° and 85°N. Nonetheless, although this grid is more uniform than the regular original one, the areal density of points increases between 20°–55° and 60°–65°N.

A spatially uniform density grid can be designed by means of the criteria of a constant density of points in each circle of latitude. Then, the variable values can be interpolated into this new equal-density grid by any interpolation procedure (e.g., a linear model) to minimize negative effects of a non-equal-density grid when the PCA is used with a correlation input matrix. The technique developed herein is easy to program and adapt to any regular latitude–longitude network. As an example, we show its application to monthly anomalies of 70-hPa temperature data collected by the Microwave Sounding Unit (MSU) flown on board the National Oceanic and Atmospheric Administration (NOAA) Television Infrared Observation Satellites (TIROS-N). Formerly, Compagnucci et al. (2001) studied the behavior of these temperature anomalies by using a correlation-based PCA based on the original latitude–longitude grid output by MSU. The authors commented that the bias produced by the systematic poleward increase in data points could be responsible for the Antarctic temporal pattern, which explained the second highest variance. Therefore, this dataset is a perfect example to show the differences in results obtained between both a regular latitude–longitude grid and an equal-density grid, given that the studied variable was shown to discriminate the equatorial zone from the polar zone in different PCs. The comparison between these results and those obtained by the equal-density criterion will be developed below.

2. Data and methodology

In order to devise an equal-density grid that shows a more uniformly spaced distribution of points, the criterion of constant density in points per latitude is adopted. If $n(\varphi)$ is the number of data points on a latitude φ , the density of points on a latitudinal circle is given by the ratio

$$\lambda(\varphi) = \frac{n(\varphi)}{L(\varphi)},$$

where $L(\varphi) = 2\pi R \cos(\varphi)$ is the approximated length of the latitudinal circle and R is the mean radius of the earth. Assuming that this density is constant it is equal to the density of a reference latitude φ_0 on which a number $N = n(\varphi_0)$ of data points is spaced [i.e., $\lambda(\varphi) = \lambda(\varphi_0)$]. The number of data points at the latitude φ in the new equal-density grid is given by

$$n(\varphi) = \text{int} \left[\frac{N \cos(\varphi)}{\cos(\varphi_0)} \right],$$

where the function $\text{int}[\]$ represents the integer part of the expression. Usage of the integer-part function will not yield a “perfect” equal-density grid. Nevertheless, we can consider it to be so because the effect obtained by the application of the integer-part function is typically negligible, as the considered grids are discrete. The point-to-point distance (in degrees of longitude) in the equal-density grid is expressed by

$$\Delta(\varphi) = \frac{360}{n(\varphi)},$$

with $n(\varphi)$ given by the previous equation. The maximum latitude of estimation for the equal-density grid depends on the density of points at the reference latitude φ_0 in accordance with the condition $\Delta(\varphi) < 360$. Therefore, the maximum latitude is $|\varphi_{\max}| < \cos^{-1}[\cos(\varphi_0)/N]$. These equations define a grid that is easily adaptable to any regular latitude–longitude-gridded network.

The monthly temperature anomaly data of the lower stratosphere (~ 70 hPa) collected by the MSU (channel 4) in the sun-synchronous orbit of NOAA’s TIROS-N family of operational satellites are used to illustrate the differences in results from the two approaches. The time period of analysis is from January 1979 to December 1997. The fields of anomalies are estimated by removing the monthly mean (1982–91), as in Compagnucci et al. (2001). The MSU information is output over a global regular 2.5° latitude–longitude grid, from -178.75°W to $+178.75^\circ\text{E}$ longitude (144 values) and -88.75°S to $+88.75^\circ\text{N}$ latitude (72 values). In this study, only the fields of the variable in the Southern Hemisphere are considered.

The equal-density grid is estimated from the above-mentioned equations taking the equator line as the reference latitude (i.e., $\varphi_0 = 0$) and $N = 144$ points of longitude. The latitudes coincide exactly with the original ones (from -1.25° to -88.75°S of latitude increasing every 2.5°) and the longitudes for each latitude φ are calculated from -178.75°W of longitude with an increment of $\Delta(\varphi) = 360/\text{int}[144 \cos(\varphi)]$. The upper panels of Fig. 1 show the regular latitude–longitude grid in a polar projection (Fig. 1A1) and in a 3D perspective (Fig. 1A2) for the Southern Hemisphere. Analogously, the lower panels of Fig. 1 show the equal-density gridded distribution in a polar projection (Fig. 1B1) and in a 3D perspective (Fig. 1B2).

Once the equal-density grid has been established, the

values of the variable at each new point are estimated by linear interpolation. For that, the original values before and after each new point on latitude circle are used.

To test the efficiency of the interpolation method, the data has been reinterpolated back to the original locations from the new equal-density dataset. Differences between the original data and the reinterpolated data have been assessed across all grid points as a mean standard error (mse). In addition, they have been plotted back to the domain and isoplethed. Figure 2a shows the spatial distribution of mse values in $^\circ\text{C } 100^{-1}$. The maximum error is 0.0785°C . The higher values are observed between 60° and 70°S . The mse values go to near-zero close the equator where both grids (latitude–longitude and equal density) almost match each other. The mse values also vanish in the polar area since the estimations on the equal density grid are good because the original locations are close to one another. On the other hand, it is of note that these mse values are probably larger than the real error as two interpolations are used in their estimation: the first, from the latitude–longitude grid to the equal-density grid (first error source) and the second, when reinterpolating data back onto the latitude–longitude grid (second error source). Therefore, it is possible that the mse values of interest (corresponding to the first interpolation) are lower than 0.08°C for all points. Anyway, whether or not a linear interpolation is sufficient might depend on the (spatial–temporal) resolution of the data in question, and perhaps on the characteristics of the variable under investigation as well. However, the selection of different reference latitudes would modify the spatial distribution of the errors according to degree of precision required by the investigator over a region of interest. Additionally, the application of more sophisticated interpolation techniques might further reduce the MSEs. In the current case, the interpolation processing is efficient as the visual assessment of the reproduction of the original field is practically undistorted. Figure 2 shows the original anomaly field of January 1979 for both the regular (Fig. 2b) and equal-density (Fig. 2c) grids.

A correlation-based S-mode PCA is performed using these data from both the regular and the equal-density grids. Results of those analyses will indicate practical differences between the two sets of grid points.

3. Results

Figure 3 displays the first five PC patterns for the regular grid (case A) and for the equal-density grid (case B). These patterns explain roughly 76% of the variance for both cases. For each PC, the PC loading (PCL) fields and the PC score (PCS) time series are shown together with their respective explained variances and eigenvalues.

The *first* pattern shows roughly the same PCL field for both cases, with maximum values over the subtropical region. The congruence coefficient (Richman 1986)

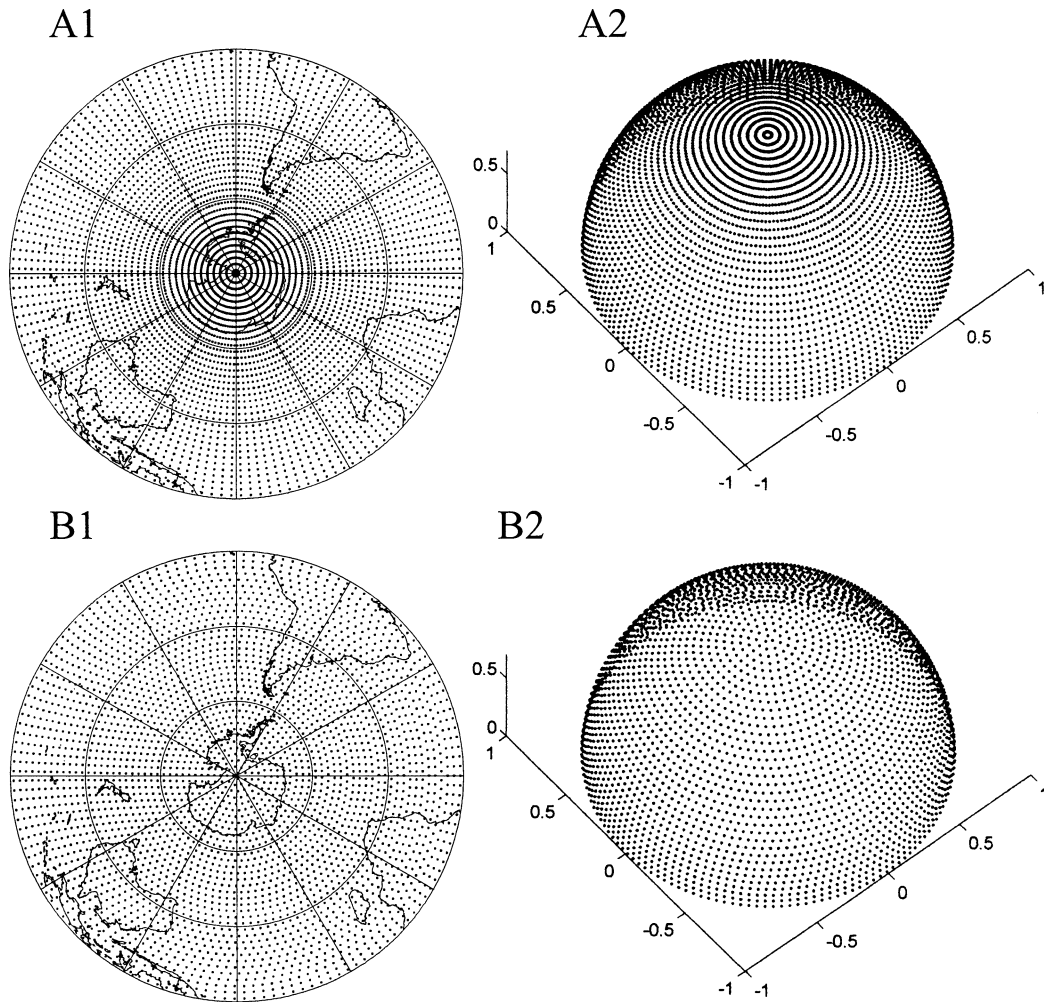


FIG. 1. The regular lat-lon grid (A1) in polar projection and (A2) the 3D perspective for the Southern Hemisphere; the equal-density grid (B1) in polar projection and (B2) the 3D perspective.

between the respective PCS is 0.99 (very high) indicating a marked or excellent concordance between both results. However, the explained variance is 38.2% in case B and 25.9% in case A. The increased variance in case B is caused by the corresponding eigenvalues (variances of the PCs) diminishing in smaller proportion than the reduction in the network points (5184 points for case A and 3286 for case B).

The relative variation between eigenvalues is 6.64%, whereas that of the number of points is 36.61%. In turn, this is due to the sum of the square of the loadings being equal to the corresponding eigenvalues. In the polar regions, these values are lower and the difference between the number of gridded points is large.

The PCs of higher order show major differences. The explained variances differ in such a way that the order of the components is altered, especially those representative of the temporal behaviors at high latitudes. In the case of a regular grid, the *second* PC corresponds to the temporal behavior of the temperature over Antarctica,

with maximum values close to the South Pole (specifically, at 86.25°S and 61.25°E, point P onward). The correlation coefficient between this PCS and the real temporal series at P is 0.957. In the case of the equal-density grid, the points are maximally removed over the polar zone producing a notorious bias, not only in the PCL field but also in the PCS time series. The third PCL from case B is most similar to the second PCL from case A. However, one can observe that the maximum PCL values (that indicate the major correlation values between the PCS temporal pattern and real time series) are displaced from the Pole toward the Pacific, at about 70°S. The correlation values between both PCS series is 0.65. This magnitude of correlation indicates that the temporal behavior depicted by the third PC in case B no longer corresponds exactly to that observed over the South Pole, as explained by the second PC in case A. This result is anticipated because the area in case B that controls the temporal behavior is different and it corresponds to the southern Pacific sector with a

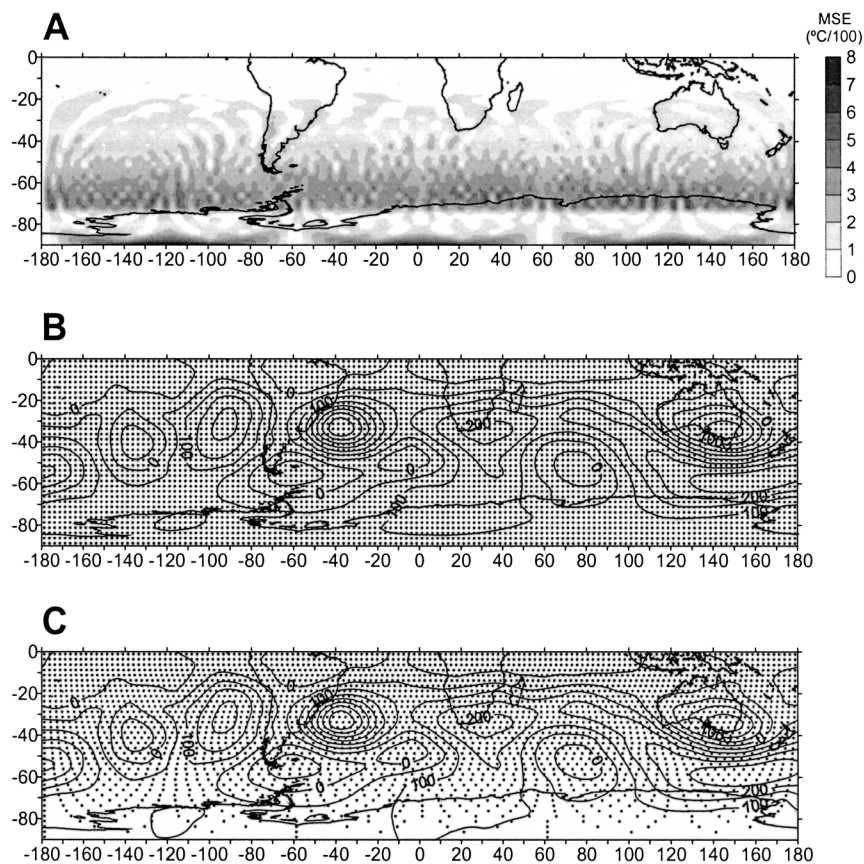


FIG. 2. (a) Differences between the original data and those reinterpolated back to the original locations from the previous equal-density interpolation, as a mse. The spatial distribution of mse values are plotted in $^{\circ}\text{C } 100^{-1}$. The original anomaly field of Jan 1979 for both the (b) regular and (c) equal-density grids.

maximum goodness of fit over the northern portion of the Ross Sea. Furthermore, the fifth PCS for case B shows a significantly high correlation with the second PCS for case A, maximizing the PCL values over the Antarctic and the western southern Atlantic. Therefore, the behavior over higher latitudes shown by the second PC in case A can be seen to be separated into two components for case B. That is to say, the real time series corresponding to the Pole can be well fitted by a linear combination of the PCSs 3 and 5 in case B. For example at point P, this linear combination is given by $T = 0.5748 \text{ PCS}_3 + 0.5684 \text{ PCS}_5$, where the weighted factors are the corresponding factor loadings at point P. The correlation between the linear combination time series T and the second PCS for case A is 0.85. This correlation value is considerably higher than that obtained by considering the third PCS only (i.e., 0.65).

The *third* component of case A reveals a pattern that is clearly similar to that of the second PC in case B, with a correlation between the PCS time series of 0.88. The explained variances are also similar, indicating that the associated eigenvalues and the amount of gridded points changed in equal proportion.

The *fourth* component shows similar spatial configurations for both cases with a dipole of maximum weighted values (absolute value) crossing over the South American continent from the southern Atlantic toward the Pacific, south of Australia and New Zealand. The correlation between the associated PCS time series is of 0.79. This coefficient value indicates that the time series experience similar temporal developments. The relative explained variances are also alike between both of them.

4. Conclusions

The equal-density technique is easily programmable and adaptable to any regular network in latitude–longitude points. For the case currently studied, the linear interpolation to the new network is shown to be efficient and to reproduce the real values with little distortion. Both fields and original temporal series show no perceptible modification either, which would be shown by reduced magnitude of mse values (Fig. 2a). Consequently, for the MSU analysis, more sophisticated interpolation techniques were deemed unnecessary. The

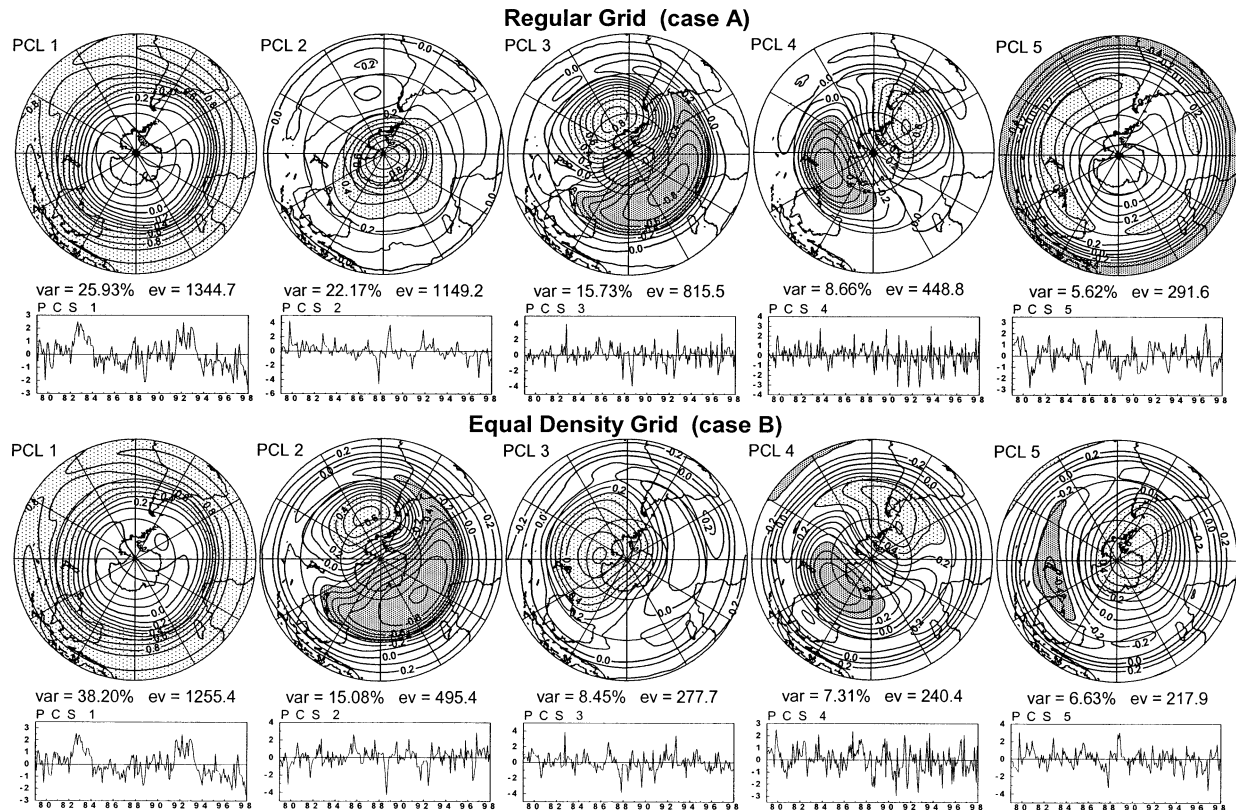


FIG. 3. The first five PC patterns for the regular grid (case A) and for the equal-density grid (case B). For each PC, the PCL field and the PCS time series are shown together with their respective explained variance (var) and eigenvalue (ev).

procedure outlined herein essentially reproduces fields identical to the original ones but on a new uniform areal density network, where all the regions are weighted according to their areal extent. Anyway, there might be a potential problem with variables with smaller scales of variability.

The PCA application to the MSU dataset for both networks (latitude–longitude grid and equal density) shows the bias introduced by the latitude–longitude grid in the estimation of the homogeneous areas and, even more so, in the relative relevance of each area. A regular latitude–longitude grid network generates an overestimation of the significance of polar areas that can be corrected by the application of the equal-density technique. The comparison of the two sets of principal component patterns shows that the temporal evolution of the variable is modified little, if at all, at low and midlatitudes, by the conversion to the equal-density network. Nevertheless, the explained variances can be altered due to the diminution of the eigenvalues in differing proportions relative to the numbers of grid points, when passing from the latitude–longitude grid to the equal-density grid. Consequently, the explained variances obtained from both cases are not directly comparable and should be only compared in a relative sense between patterns resulting from the same gridding method.

Acknowledgments. We would like to thank Dr. Michael Richman for his discussions, suggestions, and comments. We would also like to thank Dr. Peter Waylen for his suggestions. The authors also wish to thank the Marshall Space Flight Center for the access to the MSU data. This work was supported by Consejo Nacional de Ciencia y Técnica (CONICET-Argentina, The Argentine Research and Technology Council) PIP N° 0428/98 project, by the Universidad de Buenos Aires UBACYT 01X002 project and by the Agencia Nacional Para la Promoción de la Ciencia y la Tecnología PICT-99 06588 project.

REFERENCES

- Barnston, A. G., and R. E. Livezey, 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.*, **115**, 1083–1126.
- Cattell, R. B., 1952: *Factor Analysis*. Harper and Row, 462 pp.
- Compagnucci, R. H., M. A. Salles, and P. O. Canziani, 2001: The spatial and temporal behavior of the lower stratospheric temperature over the Southern Hemisphere: The MSU view. Part I: Data, methodology and temporal behavior. *Int. J. Climatol.*, **21**, 419–437.
- Craddock, J. M., and C. R. Flood, 1969: Eigenvectors for representing the 500 mb geopotential surface over the Northern Hemisphere. *Quart. J. Roy. Meteor. Soc.*, **95**, 576–593.
- Eherendorfer, M., 1987: A regionalization of Austria's precipitation

- climate using principal component analysis. *J. Climatol.*, **7**, 71–89.
- Elmore, K. L., and M. B. Richman, 2001: Euclidean distance as a similarity metric for principal component analysis. *Mon. Wea. Rev.*, **129**, 540–549.
- Glahn, H. R., 1965: Objective weather forecasting by statistical methods. *Statistician*, **15**, 111–142.
- Jolliffe, I. T., 1986: *Principal Component Analysis*. Springer-Verlag, 271 pp.
- , 1995: Rotation of principal components: Choice of normalization constraints. *J. Appl. Stat.*, **22**, 29–35.
- Karl, T. R., A. J. Koscielny, and H. F. Diaz, 1982: Potential errors in the application of principal component (Eigenvector) analysis to geophysical data. *J. Appl. Meteor.*, **21**, 1183–1186.
- Molteni, F., P. Bonelli, and P. Bacci, 1983: Precipitation over northern Italy: A description by means of principal components analysis. *J. Climate Appl. Meteor.*, **22**, 1738–1752.
- Richman, M. B., 1986: Rotation of principal components. *Int. J. Climatol.*, **6**, 293–335.
- , and P. J. Lamb, 1985: Climatic pattern analysis of three- and seven-day summer rainfall in the Central United States: Some methodological considerations and a regionalization. *J. Climate Appl. Meteor.*, **24**, 1325–1343.
- , and X. Gong, 1999: Relationships between the definition of the hyperplane width to the fidelity of principal component loading patterns. *J. Climate*, **12**, 1557–1676.
- Thompson, D. W. J., and J. M. Wallace, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate*, **13**, 1000–1016.