

A ROTATED PRINCIPAL COMPONENT ANALYSIS OF
TOTAL COLUMN OZONE OBTAINED FROM TOMS FOR 1984 - 1989

Brian K. Eder* and Sharon K. LeDuc*

Atmospheric Sciences Modeling Division
Air Resources Laboratory
National Oceanic and Atmospheric Administration
Research Triangle Park, NC USA

1. INTRODUCTION

The global distribution of total column ozone, arguably the most important trace gas in the atmosphere, is attracting great international attention as concerns over reduced ozone abundances escalate (WMO, 1990). In addition to its biological importance, ozone also plays a critical role in the chemical and meteorological dynamics of the atmosphere. Ozone is a precursor of the hydroxyl radical, the major cleansing agent in the troposphere and its absorption of ultraviolet radiation is the major heat source in the stratosphere, ultimately driving the global circulation of the stratosphere.

Despite its great importance, the spatial and temporal distribution of ozone is poorly understood. To assess anthropogenic changes to date and to better understand how ozone abundance may respond to future perturbations requires a better understanding of its natural intra- and interannual variability and the processes that contribute to this variability. Unfortunately, many current studies (Herman et. al., 1991; Stolarski et. al., 1991) use either globally, hemispherically or zonally averaged (i.e. data from 30°N to 40°N) data to try and detect cycles in the TOMS data. Although these studies provide a simple, conceptually appealing picture of ozone abundance, much of its variability is attributable to more complex phenomena that are neither globally or zonally symmetrical. As a result, such approaches often mask the true strength of temporal cycles and do not allow identification of the exact spatial extent of such cycles.

Accordingly, the purpose of this analysis is to develop a better understanding of these natural variations across various spatial and temporal scales. This will be achieved through the application of a multivariate statistical technique called rotated Principal Component Analysis (PCA) to the total column ozone data derived from Version 6.0 TOMS (Total Ozone Mapping Spectrometer) for the period 1984 through 1989. The main objective of principal component analysis (PCA) is

to identify, through a reduction in data, the characteristic, recurring and independent modes of variation across all potential spatial and temporal scales (Eder et al., 1993). The analysis sorts initially correlated data into a hierarchy of statistically independent modes of variation which explain successively less and less of the total variation; thereby summarizing the essential information of that set so that meaningful and descriptive conclusions can be achieved. This technique is ideal for application to the TOMS data set where the total number of observations exceeds 3 million. Utilization of Kaiser's (1958) varimax orthogonal rotation will allow delineation of homogeneous subregions - that is, areas of the globe that experience unique total ozone characteristics. Examination of the time series associated with each unique subregion will be based on spectral density analysis. This will allow further elucidation (across many wavelengths) of the physical phenomena (i.e. annual and semi-annual cycles, Quasi-Biennial Oscillation (QBO), El-Nino-Southern Oscillation (ENSO)) responsible for the natural variability of total column ozone.

2. TOMS DATA

The daily TOMS (Version 6.0) data were obtained from the archived data sets available on CD-ROM at the U.S. National Space Science Data Center (NSSDC) located at NASA's Goddard Space Flight Center in Greenbelt, MD. TOMS is aboard the Nimbus 7 satellite, which is in a sun-synchronous nearly polar orbit, launched in October, 1979. The instrument consists of a single Ebert-Fastie monochromator that measures the ultraviolet sunlight backscattered from the Earth's atmosphere and surface at six wavelengths. Four of these wavelengths (312.5, 317.5, 331.2 and 339.8 nm) are used in pairs to infer, from the differential absorption of scattered sunlight, the total column ozone. For instance, ozone is calculated from the ratio of two wavelengths, 312.5 nm and 331.2, where one wavelength (312.5 nm) is strongly absorbed by ozone, while the other is only weakly absorbed. The two remaining wavelengths (360 and 380 nm) are used to measure surface reflectivity.

*On assignment to the National Exposure Research Laboratory, U.S. Environmental Protection Agency.

TOMS collects 35 measurements every 8 s as it rapidly scans from right to left in a plane perpendicular to the orbital plane, producing roughly 200,000 daily ozone measurements with a resolution of between 50 and 150 km. The total ozone column is then expressed as the depth that the ozone alone would occupy at standard temperature and pressure. This depth is measured in thousands of centimeters or Dobson Units [1 DU = 2.69×10^{16} molecules cm^{-2}], such that typical column abundances of between 250 and 400 DU would correspond to pure ozone column depths of between 0.25 and 0.40 cm. The TOMS data used in this analysis were regridded from the original resolution of 1° lat. by 1.25° long. using an averaging technique which yielded a total of 1440, 5° by 5° daily fields extending from 50°S to 50°N. Six years of data were examined, from January 1, 1984, through December 31, 1989. All missing data (< 5 %) were eliminated through a temporal interpolation scheme resulting in 3,156,480 observations.

3. METHODOLOGY

3.1 Spatial Analysis

Mathematically, this analysis began with the calculation of a square, symmetric correlation ${}_{1440}\mathbf{R}_{1440}$ from the original data matrix, which had dimensions of 1440 (grid cells) x 2192 (days). Selection of a correlation matrix (as opposed to a covariance matrix) has two advantages in PCA. First, use of a correlation is much more suitable for resolving spatial patterns; second, use of a correlation matrix allows maps of component loadings (the correlation coefficient between the grid cell and component) to be drawn (Overland and Preisendorfer, 1982). By using \mathbf{R} and the identity matrix \mathbf{I} of the same dimensions, 1440 characteristic roots or *eigenvalues* (λ) were derived, which satisfied the following polynomial equation:

$$\det [{}_{1440}\mathbf{R}_{1440} - \lambda_{1440}\mathbf{I}_{1440}] = 0 .$$

For each root (λ), a non-zero vector \mathbf{e} can be derived:

$${}_{1440}\mathbf{R}_{1440}\mathbf{e}_1 = \lambda_{1440}\mathbf{e}_1 ;$$

where the vector \mathbf{e} is called the characteristic vector, or eigenvector of the correlation matrix \mathbf{R} , associated with its corresponding eigenvalue (λ). The eigenvectors derived from the correlation matrix represent the mutually orthogonal linear combinations (or modes of variation) of the matrix. Their associated eigenvalues, which are scalars, represent the amounts of total variance that are explained by each of the eigenvectors. By retaining only

the first few eigenvector-eigenvalue pairs, or principal components, a substantial amount of the total variance can be explained while ignoring higher order principal components which explain minimal amounts of the total variance and can be viewed as noise. The exact number of components that should be retained was determined by the Scree Test (Eder et al., 1993), and indicated that eleven components should be retained. Therefore, the original data set, which contained 1440 intercorrelated and noisy variables (the 5° by 5° grid cells), has been reduced to one containing only eleven orthogonal and thereby independent variables (the principal components), yet still explains nearly three-fourths of the total variance.

Since one of the major goals of this research was to define areas of homogeneous total column ozone, a rotation was performed on the principal components in order to better segregate the areas that have similar ozone characteristics. Of the many types of rotation available, an orthogonal method developed by Kaiser (1958) was selected because it rigidly rotates the predetermined principal components while retaining the constraint that the individual components remain orthogonal or uncorrelated. This method increases the segregation between loadings, allowing better definition of distinct groupings of intercorrelated data, thereby making spatial interpretation easier (Horel, 1981).

The elements of each eigenvector were then multiplied by the square root of the associated eigenvalue to obtain the component loadings (L_{ij}), for grid cell j on principal component i . These loadings represent the correlations between the component and the grid cell. The square of the component loading indicates the proportion of variance in the individual grid cell that can be attributed to that component. Maps of the loadings associated with the first seven retained components can be found at the tops of Figures 1 through 4. Space constraints prohibit inclusion of all eleven components.

3.2 Temporal Analysis

Another useful parameter that can be derived from PCA is the component score (PCS). Initially, the rotated PCA replaced the 1440 grid cells, which were measured over 2192 days, with 11 rotated principal components having no temporal measure. By introducing the PCS, a derivation of similar temporal measurements for the rotated principal components over the same 2192 days can be achieved. The rotated principal components are identified in terms of the original grid cells, the larger the loading the more important the grid cell is in the interpretation of the component. Therefore, if a day has high values for grid cells with large loadings, it should have a large value on the component. The PCS for day

i on component k is designed to meet this requirement are defined as follows:

$$(PCS)_{ik} = \sum_j O_{ij} L_{jk} ;$$

where O_{ij} is the observation for day i on grid cell j , L_{jk} is the loading of grid cell j on component k , and the PCS is the component score for day i on component k and is summed over all 1440 grid cells. As seen from the equation, the PCSs are simply weighted summed values for the days over the grid cells, the weights being the component loadings. The larger the value that a day has on a grid cell with large loadings, the larger the PCS. When plotted as a time series (bottoms of Figures 1 through 4), the PCS, which are standardized, provide excellent insight into the spectrum of temporal variance experienced by each of the subregions.

Spectral density analysis (SDA) using the finite Fourier transformation was then employed to examine each of the 11 time series associated with the subregions.

Such analysis yields a measure of the distribution of variance of the time series across all possible wavelengths, each arbitrarily close to the next. It is used to look for non-random, physically generated cyclical patterns or periodicities in the time series data, which would be represented by a peak in the spectrum at a particular frequency. SDA decomposes the time series into a sum of sine and cosine waves of varying amplitudes and wavelengths as defined by:

$$(PCS)_i = \overline{(PCS)} + \sum_k [a_k \cos \omega_k(i-1) + b_k \sin \omega_k(i-1)] ;$$

where the a_k 's are the cosine coefficients, the b_k 's are the sine coefficients and $\omega_k = 2\pi k/2191$. From this decomposition, the periodogram can be obtained and is defined as follows:

$$I_k = \frac{2191(A_k^2 + B_k^2)}{2} ;$$

where A_k and B_k are estimates of a_k and b_k and are given by:

$$A_k = \left[\frac{2}{N} \right] \sum_i [((PCS)_i - \overline{(PCS)}) \cos \omega_k(i-1)] ;$$

$$B_k = \left[\frac{2}{N} \right] \sum_i [((PCS)_i - \overline{(PCS)}) \sin \omega_k(i-1)] .$$

The spectral density is then estimated by smoothing the periodogram using a triangular weight (often referred to as the kernel or spectral window) as follows:

$$\text{Spectral density estimates} = \sum_{j=-p}^p W_j I_{k+j} ;$$

where the W_j 's are smoothing weights normalized to $\pi/4$. At the risk of decreasing the statistical stability of the spectral estimates, a relatively narrow window (1 2 3 4 5 4 3 2 1) was used in order to prevent any bias that might occur through the over smoothing of "real" peaks.

4. RESULTS

The first Rotated Principal Component (RPC) ($\lambda = 409.48$), which explains 28.44% [(409.48/1440) x 100] of the total variance, defines an area encompassing much of the *Southern Hemisphere* (SH). A large area of this subregion contains grid cells with loadings in excess of 0.70 (meaning more than half of the variance of these cells can be attributed to this component). The time series and spectra analysis of this component reveal a strong annual cycle (maximum spectral power at $f = 0.01720$, corresponding to a periodicity of $(2\pi/f) = 365.33$ days) that peaks during the austral Spring (September and October) of each year and falls to a broader minimum during the period of February through May. This cycle is likely responding to the changing solar cycle which alters the chemical and dynamical process in the atmosphere. In terms of magnitude, the annual cycle of this SH subregion exhibits similar strength during the years: 1984, 1986, 1988 and 1989; however during 1987 and especially 1985 its magnitude and range is strikingly reduced.

The second RPC ($\lambda = 287.49$), which explains 19.96% of the total variance defines a comparable area comprising much of the *Northern Hemisphere* (NH); however, the poleward extent of this NH subregion is limited to roughly 35°N, resulting in a more narrow subregion. This difference is likely attributable to the extensive landmasses found in the NH, which play a large role in perturbing the Arctic circumpolar vortex and hence the hemispheric scale circulation patterns. These perturbations, called planetary waves, occur much more frequently in the NH than in the SH. As a result, the Antarctic circumpolar vortex is much less disturbed, resulting in less variability in the ozone abundance of the SH. The NH land masses, which coincide fairly well with loadings on RPC2 of less than 0.30 induce additional variability into the total column ozone which are not represented well by this RPC. The time series and spectral of this subregion likewise indicate a strong annual

cycle, however its peak is much more broad and occurs during the period of March through July. The minimum tends to occur during December and January. It is interesting that like the SH, the annual cycle associated with the NH is weakest during 1985 and 1987. Generally, the relative annual maximum is greater in the SH than the NH. However, during 1985 and 1987, the maxima are nearly equal with the NH slightly stronger.

The third RPC ($\lambda = 207.85$, 14.43%) defines an area that is fairly symmetrical about the equator, dominating between 10°N and 10°S. This subregion coincides well with a region associated with the *Quasi-Biennial Oscillation* (QBO) of the tropical winds in the lower stratosphere, where dominance of the NH and SH annual variability is replaced with QBO dominated variance (Ziemke and Stanford, 1994). This association is confirmed by the SDA which reaches a maximum power at $f = 0.00860$, corresponding to a periodicity of 731 days (roughly two-years). This component is slightly more significant in the southeastern Pacific Ocean near Ecuador and Peru. Its time series peaks in 1985 and 1987, the two years corresponding with the weak annual cycles observed in both the NH and the SH.

The fourth RPC ($\lambda = 33.50$, 2.33%) defines an area in the SH from New Guinea in the Western Pacific to central South America. This area coincides well with that impacted by the *El Nino - Southern Oscillation* (ENSO), which is the likely driving force. Note that during the El Nino of 1986-1987, the PCs are lower, while in the non-El Nino years, especially, 1988 and 1989, the scores are high.

The fifth and sixth RPCs are both associated with subregions where a strong semi-annual variance dominates. The fifth subregion ($\lambda = 29.43$, 2.04%) defines a region from the western Indian Ocean into the western Pacific Ocean. It is strongest north of the Australian Continent over Indonesia and appears to be associated with the *Equatorial Semi-Annual Oscillation*, which is physically driven by the negative temperature effect of ozone photochemistry associated with the semi-annual modulation of the temperature field (Varotsos et al., 1992). This component has a very strong semi-annual cycle, with maximum spectral power at $f = 0.03440$, corresponding to a periodicity of 182.67 days. It peaks during the transitional seasons, (Spring and Autumn). Similarly, the sixth subregion ($\lambda = 25.87$, 1.80%), which defines an area in central Asia may be related to the *Polar Semi-Annual Oscillation*, which is driven by two processes: the photochemical production of ozone and dynamical transport from the equatorial regions of maximum ozone production. This subregion also has a strong periodicity of roughly 182 days. Reasons for the areas in South America that load highly on this

component and any potential teleconnection implications are not understood at this time.

The seventh RPC ($\lambda = 18.74$, 1.30%) is somewhat different from the first six RPCs in that it is not contiguous, but rather a compilation of three separate subregions, that are none-the-less driven by the same physical process - namely *Baroclinic Waves* associated with the polar jet stream. Each of these three areas are associated with favored areas of tropospheric jet stream cores, principally, the Aleutian coast, the Canadian North Atlantic Coastal region and eastern Siberia/Asia. The time series and spectra support this contention in that this component generally reaches a maximum twice a year, during transitional seasons which are periods of maximum baroclinicity.

The remaining RPCs (8 through 11), like RPC7, define various, small areas in the higher latitudes of both hemispheres, and are likewise thought to be associated with baroclinic waves.

5. REFERENCES

- Eder, B.K., J.M. Davis and P. Bloomfield (1993) "A characterization of the spatiotemporal variability of non-urban ozone concentrations over the Eastern United States", *Atmos. Environ.*, **27A**, 2645-2668.
- Herman, J.R., R. McPeters and R. Stolarski (1991) "Global average ozone change from November 1978 to May 1990", *J. Geophys. Res.*, **96**, D9, 17297-17305.
- Horel, J.D. (1981) "A rotated principal component analysis of the interannual variability of the Northern Hemisphere 500 mb height field", *Mon. Wea. Rev.*, **109**, 2080-2092.
- Kaiser, H.F. (1958) "The varimax criterion for analytical rotation in factor analysis", *Psychometrika*, **23**.
- Overland, J.E. and R.W. Preisendorfer (1982) "A significance test for principal components applied to a cyclone climatology", *Mon. Wea. Rev.*, **110**, 1-4.
- Stolarski, R.S., P. Bloomfield and R.D. McPeters (1991) "Total ozone trends deduced from Nimbus 7 TOMS Data", *Geophys. Res. Lett.*, **18**, No. 6, 1015-1018.

Varotsos, C., C. Helmis and C. Cartalis (1992) "Annual and semi-annual waves in ozone as derived from SBUV vertical global ozone profiles", *Geophys. Res. Lett.*, **19**, No. 9, 925-928.

World Meteorological Organization (1990) *Scientific Assessment of Stratospheric Ozone: 1989*, Global Ozone Research and Monitoring Project, Report No. 20, Geneva, Switzerland.

Ziemke, J.R and J. L. Stanford (1994) "Quasi-biennial oscillation and tropical waves in total ozone", *J. Geophys. Res.*, **99**, **D11**, 23041-23056.

The information in this document has been funded by the United States Environmental Protection Agency. It has been subjected to Agency review and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Varotsos, C., C. Helmis and C. Cartalis (1992) "Annual and semi-annual waves in ozone as derived from SBUV vertical global ozone profiles", *Geophys. Res. Lett.*, 19, No. 9, 925-928.

Ziemke, J.R and J. L. Stanford (1994) "Quasi-biennial oscillation and tropical waves in total ozone", *J. Geophys. Res.*, 99, D11, 23041-23056.

World Meteorological Organization (1990) *Scientific Assessment of Stratospheric Ozone: 1989*, Global Ozone Research and Monitoring Project, Report No. 20, Geneva, Switzerland.

The information in this document has been funded by the United States Environmental Protection Agency. It has been subjected to Agency review and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

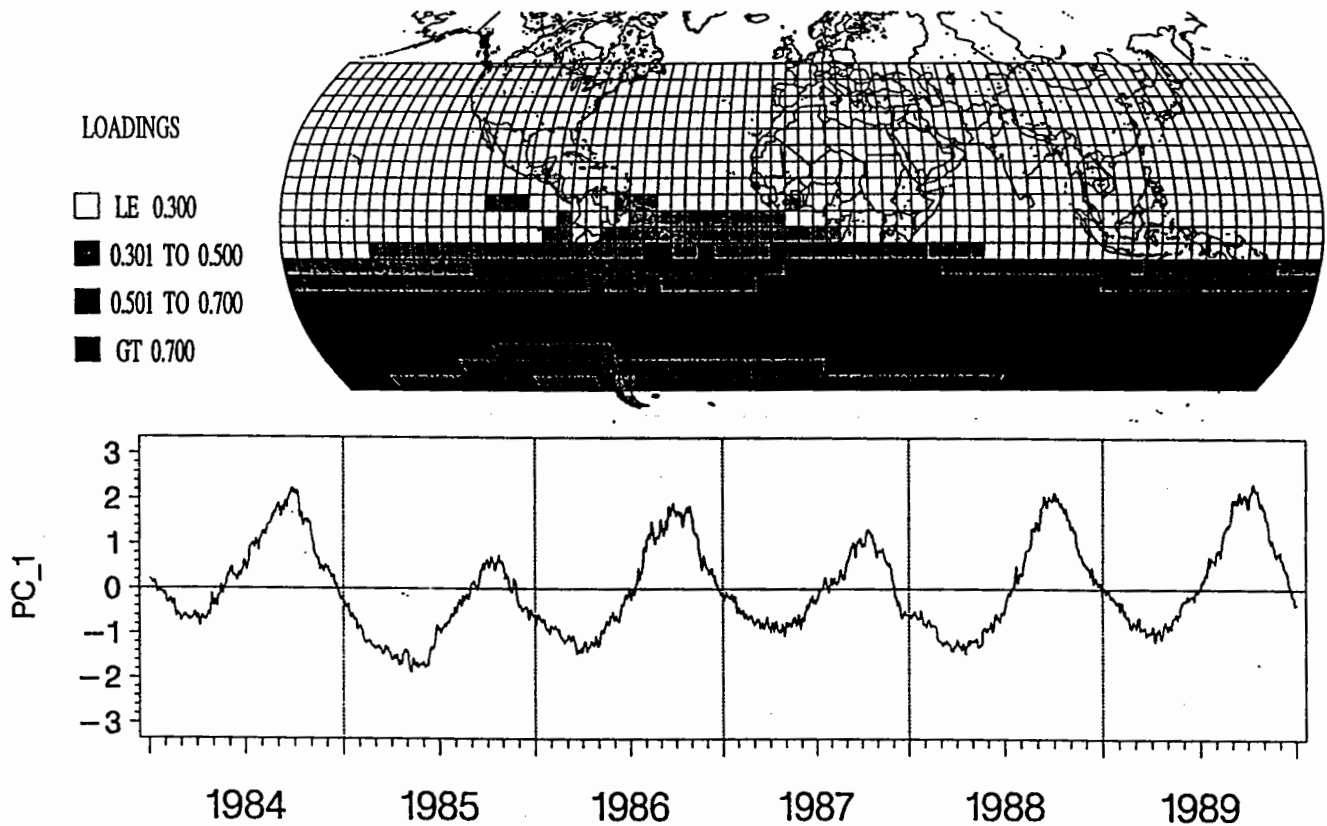


Figure 1. Rotated principal component loadings and scores associated with the *Southern Hemisphere* Subregion.

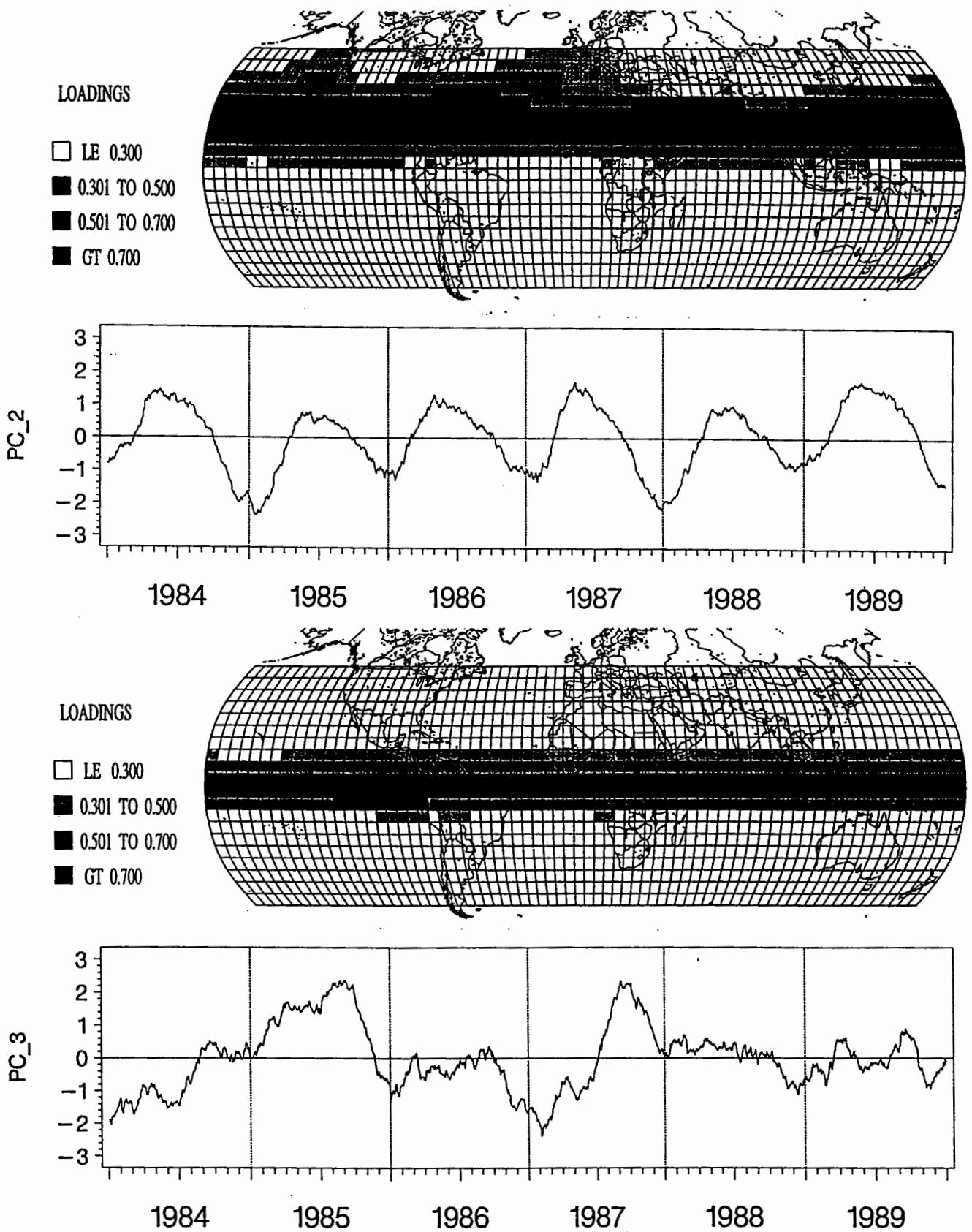


Figure 2. Rotated principal component loadings and scores associated with the *Northern Hemisphere* Subregion (top figures) and the *Quasi-Biennial Oscillation* Subregion (bottom figures.)

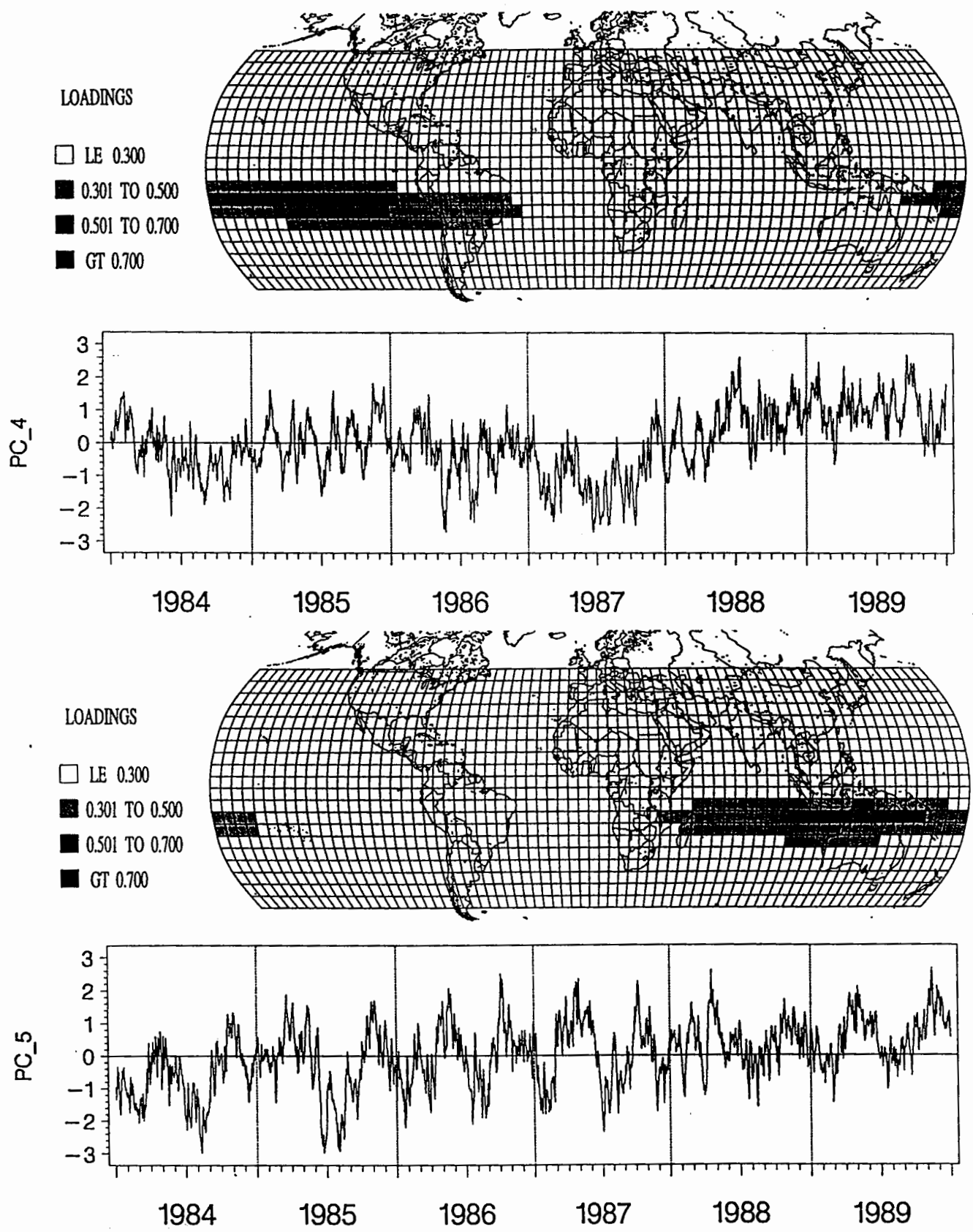


Figure 3. Rotated principal component loadings and scores associated with the *El Niño - Southern Oscillation* Subregion (top figures) and the *Equatorial Semi-Annual Oscillation* Subregion (bottom figures).

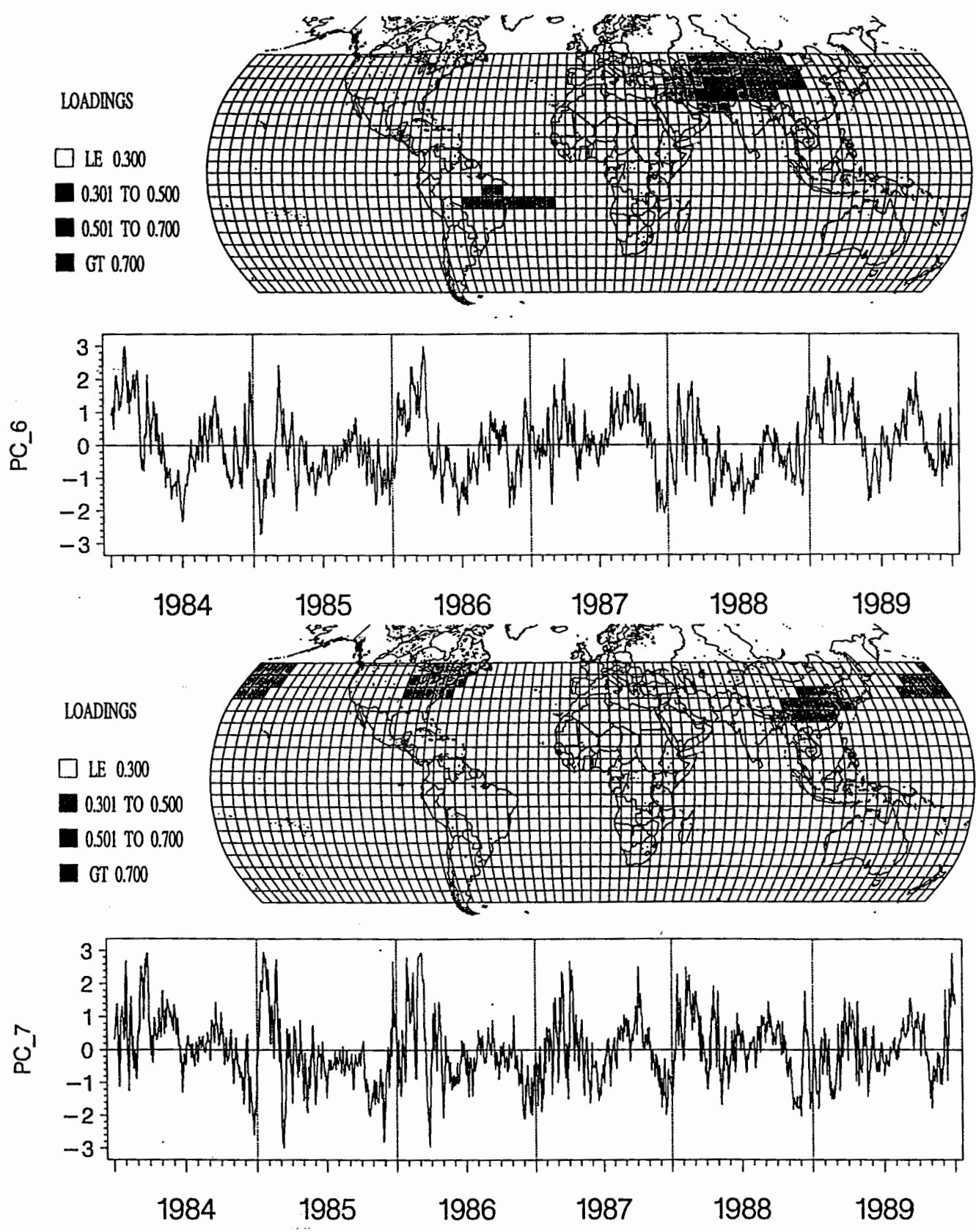


Figure 4. Rotated principal component loadings and scores associated with the *Polar Semi-Annual Oscillation* Subregion (top figures) and the *Baroclinic Waves* Subregion (bottom figures).

TECHNICAL REPORT DATA

1. REPORT NO. EPA/600/A-96/073	2.	3. REC)	
4. TITLE AND SUBTITLE A Rotated Principal Component Analysis of Total Column Ozone Obtained from TOMS for 1984-1989		5. REPORT DATE	
		6. PERFORMING ORGANIZATION CODE	
7. AUTHOR(S) Brian K. Eder and Sharon K. LeDuc		8. PERFORMING ORGANIZATION REPORT NO.	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Same as Block 12		10. PROGRAM ELEMENT NO.	
		11. CONTRACT/GRANT NO.	
12. SPONSORING AGENCY NAME AND ADDRESS National Exposure Research Laboratory Office of Research and Development U.S. Environmental Protection Agency Research Triangle Park, NC 27711		13. TYPE OF REPORT AND PERIOD COVERED	
		14. SPONSORING AGENCY CODE	
15. SUPPLEMENTARY NOTES			
<p>16. ABSTRACT</p> <p>The purpose of this analysis is to develop a better understanding of these natural variations across various spatial and temporal scales. This will be achieved through the application of a multivariate statistical technique called rotated Principal Component Analysis (PCA) to the total column ozone data derived from Version 6.0 TOMS (Total Ozone Mapping Spectrometer) for the period 1984 through 1989. The main objective of principal component analysis (PCA) is to identify, through a reduction in data, the characteristic, recurring and independent modes of variation across all potential spatial and temporal scales (Eder et al., 1993). The analysis sorts initially correlated data into a hierarchy of statistically independent modes of variation which explain successively less and less of the total variation; thereby summarizing the essential information of that set so that meaningful and descriptive conclusions can be achieved. This technique is ideal for application to the TOMS data set where the total number of observations exceeds 3 million. Utilization of Kaiser's (1958) varimax orthogonal rotation will allow delineation of homogeneous subregions - that is, areas of the globe that experience unique total ozone characteristics. Examination of the time series associated with each unique subregion will be based on spectral density analysis. This will allow further elucidation (across many wavelengths) of the physical phenomena (i.e. annual and semi-annual cycles, Quasi-Biennial Oscillation (QBO), El-Nino-Southern Oscillation (ENSO)) responsible for the natural variability of total column ozone.</p>			
17. KEY WORDS AND DOCUMENT ANALYSIS			
a. DESCRIPTORS	b. IDENTIFIERS/ OPEN ENDED TERMS	c. COSATI	
18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS (This Report) UNCLASSIFIED	21. NO. OF PAGES	
	20. SECURITY CLASS (This Page) UNCLASSIFIED	22. PRICE	