EPA/600/A-92/124

THE SPATIAL AND TEMPORAL ANALYSIS OF NON-URBAN OZONE CONCENTRATIONS OVER THE EASTERN UNITED STATES USING ROTATED PRINCIPAL COMPONENT ANALYSIS

Brian K. Eder*

Atmospheric Sciences Modeling Division
Air Resources Laboratory
National Oceanic and Atmospheric Administration
Research Triangle Park, NC, USA 27711

1. INTRODUCTION

Traditionally, ozone (O_3) has been considered an urban-scale pollutant. More recently, however, it has been recognized by scientists as a regional (Logan, 1989) and even global-scale phenomenon (Liu et al., 1987) as high concentrations are routinely observed over large non-urban areas of most industrialized countries where forest retardation and crop injury are becoming growing environmental concerns (Lefohn and Lucier, 1991). Daily maximum O_3 concentrations in these areas are often comparable to those found in urban areas (Meagher et al., 1987) and daily average levels can even exceed urban levels due to a lack of nitric oxide (NO) scavenging.

Clearly, a better understanding of the spatial and temporal variability of non-urban O_3 concentrations is critical to the development and implementation of standards designed to mitigate any adverse effects on forest and crop productivity as well as ecosystem well-being. In an attempt to enhance this understanding, the eastern United States will be segregated, using a Rotated Principal Component Analysis (RPCA), into areas whose O_3 concentrations exhibit unique, homogeneous characteristics. Examination of these characteristics will then be achieved using a variety of interpretive analyses.

The RPCA approach has been used successfully in the examination of other aerometric data including $SO_4^=$ concentrations in precipitation (Eder, 1989) and SO_2 ambient air concentrations (Ashbaugh et al., 1984). The advantages of using such an approach are numerous (Cox and Clark, 1981; Crutcher et al., 1986; Eder, 1989). First, because of the copious amount of data often resulting from such large-scale studies (nearly 100,000 observations for this study), and because individual data tend to be erratic or noisy, it is advantageous to employ an analysis technique which identifies, through a reduction of data, the recurring and independent modes of variation within the larger data set. Secondly, RPCA will allow comparison of O_3 characteristics (i.e. trends, distributions, etc.) between regions whose segregation is statistically and physically based as opposed to those based on arbitrary or geo-political boundaries. And finally, the analysis of O_3 characteristics and trends within these regions would be based upon an aggregation of data from many stations over a long time period, as opposed to individual stations, minimizing the effects of anomalous or even erroneous data often associated with a particular station.

2. DATA

The data employed in this analysis were obtained from the United States Environmental Protection Agency's Aerometric Information Retrieval System (AIRS). This database contains a multitude of hourly aerometric data, including O_3 concentrations, collected by local agencies at nearly 800 stations nationwide. The data are monitored using criteria established by the EPA (1985), which includes multipoint calibrations, independent audits and data validation based upon frequent zero, span and precision checks and a rigorous quality assurance program. The O_3 measurements were made during the O_3 "season" (April through October for this study) using either chemiluminescence analyzers, which are sensitive to light emitted by the reaction between O_3 and ethylene or ultraviolet photometers, which measure the absorption of light by O_3.

A major consideration of this study was the establishment and utilization of a complete, regionally-representative O_3 data base, unencumbered by either missing data or local-scale variability. Attainment of this goal was achieved by employing numerous strict selection criteria. First, to avoid NO scavenging affects found in close proximity to urban areas, only those stations classified as either rural or suburban and reporting a land use of either forest, agriculture or residential were selected. Rural stations received highest priority. However, to meet the second criteria, spatial completeness, several suburban stations were also included. Third, only those stations reporting a capture rate of 90.0% or better for the study period were considered. Finally, to keep the amount of data manageable, only the daily maximum O_3 concentrations were used. Use of this statistic is justified by Vukovich and Fishman (1986) who have shown that at the time of maximum surface concentration (typically between 1 and 3 pm LST), the boundary layer is uniformly mixed and the surface concentration is very nearly equal to the mean boundary layer concentration.

---

* On assignment to the Atmospheric Research and Exposure Assessment Laboratory, U.S. Environmental Protection Agency.

These selection criteria resulted in the inclusion of 77 stations across the eastern half of the United States, the majority of which (55) are classified as rural. Several combinations of station-"seasons" were examined before the optimum period of 1985-1990 was selected. The total capture rate for the period was 95.7% (94,617 out of a possible 98,868 observations), which translates into an average of 205 out of 214 days per station-"season". Missing data were estimated using a linear interpolation scheme.

3. METHODOLOGY

Mathematically, this analysis began with the extraction of a square, symmetrical correlation matrix **R** (having dimensions of 77 x 77) from the original data matrix (having dimensions of 77 stations x 1284 days and containing 98,868 observations). Selection of a correlation matrix (as opposed to a covariance matrix) has two advantages. First, use of a correlation matrix is much more suitable for resolving spatial oscillations (Overland and Preisendorfer, 1982), which is a major goal of the analysis. Second, use of a correlation matrix allows isopleths of component loadings (which can be regarded as the correlation coefficient between the component and the individual station) to be drawn. By using **R** and the identity matrix (I), of the same dimensions, 77 eigenvector-eigenvalue pairs were derived. The eigenvectors represent the mutually orthogonal linear combinations or modes of variation of the matrix, while their respective eigenvalues represent the amount of variance explained by each of the eigenvectors.

By retaining only the first few eigenvector-eigenvalue pairs and their corresponding principal components, a substantial amount of the total variance can be explained. The higher order principal components, which explain minimal amounts of the variance, are viewed as noise. The exact number of components that should be retained remains a subject of some controversy and uncertainty. Retaining too few components results in the blending of spatial patterns which are actually more discrete; whereas retaining too many components will delineate ares which represent nothing more than randomly generated variations (Richman, 1981). The method selected for this analysis was developed by Overland and Preisendorfer (1982) and is based upon a significance test at a 95% Confidence Level (C.L.) using a random Monte Carlo simulation. When applied to this data set, the significance test indicated that the first six principal components should be retained and that the higher order principal components are considered noise. The Scree Test, which entails the plotting of the percentage variance explained by each component in the order extracted and then looking for a discontinuity in the curve, also supported a six principal component solution. The fraction of total variance explained by these six components are presented in Table 1.

The original data set, which contained 77 intercorrelated and noisy variables (stations) has therefore been reduced to six orthogonal and thereby independent variables (principal components) and yet still explains almost two-thirds of the total variance of the original data set.

Table 1. Statistics for the first six unrotated principal components.

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
Eigenvalue	27.61	7.44	5.51	3.52	2.86	2.36
% Variance	35.86	9.66	7.15	4.58	3.71	3.06
Cumul. %	35.86	45.52	52.67	57.25	60.96	64.02

These results are very promising, especially given the highly variable nature of O_3 and the fact that daily, non-smoothed data were used.

Since one of the major goals of this paper is to define areas of homogeneous O_3 concentrations, a rotation was performed on the components in order to better segregate the areas that have similar concentration characteristics. Of the many types of rotation possible, an orthogonal method developed by Kaiser (1958) was selected because it rigidly rotates the predetermined principal components while retaining the constraints that the individual components remain orthogonal or uncorrelated to each other. This method increases the segregation between component loadings, which in turn better defines a distinct grouping or clustering of intercorrelated data, thereby making spatial interpretation easier (Horel, 1981).

Comparison between the statistics associated with the six rotated (Table 2) and unrotated (Table 1) components reveals that the total explained variance remains the same, 64.02%, but the distribution of the variance explained over the individual components is more uniform.

Table 2. Statistics for the first six orthogonally rotated principal components.

	PC1	PC2	PC3	PC4	PC5	PC6
Eigenvalue	13.55	10.82	8.58	7.58	5.46	3.30
% Variance	17.60	14.05	11.14	9.84	7.09	4.30
Cumul. %	17.60	31.65	42.79	52.63	59.72	64.02

4. RESULTS AND DISCUSSION

When the elements of each eigenvector are multiplied by the square root of the associated eigenvalue one obtains the *component loading*, which represents the correlation between the component and the station. (The square of a component loading indicates the proportion of variance at the individual station that can be attributed to that component). When these principal component loadings are spatially mapped onto their respective stations for each component, isopleths of component loadings can be drawn. Because of space limitation, individual maps of each of the rotated components are not presented. However, all six maps can be incorporated into one by plotting the maximum loadings for each station and their respective rotated principal component. The resulting map (Figure 1) depicts six separate, contiguous subregions, each exhibiting statistically unique O_3 concentration characteristics. Their uniqueness is likely attributable to commonality of forcing factors such as meteorological and emissions patterns.

Loadings associated with the first rotated component define an area encompassing the *Great Lakes* region north and west of the Ohio River. The second component encompasses the *Northeast* States from Pennsylvania and northern New Jersey, northward. The highest loadings associated with the third rotated component are found over the *South*, while the fourth component defines the *Mid-Atlantic* region from North Carolina to Maryland. The fifth component defines the *Southwest* part of the study area from Texas and Louisiana northward to Kansas, and loadings associated with the sixth component highlight stations in *Florida*.

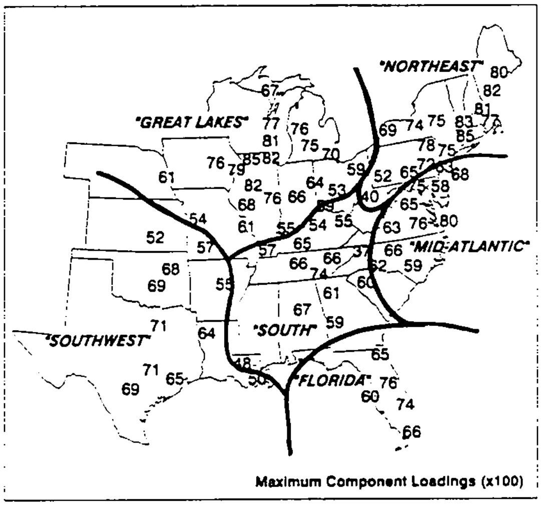

Figure 1. Six homogeneous O_3 concentration subregions as defined by maximum component loadings.

In order to determine the stability of this rotated principal component solution (i.e. stability of the six homogeneous subregions), and to determine the amount, if any, of correlation between subregions, the orthogonality constraint was relaxed. This "relaxation" was achieved through the application of an Harris-Kaiser case II Oblique rotation, which has the same general objective as a Varimax rotation (maximization of the variance of the component loadings between components for a given station) but it entirely relaxes the orthogonality constraint. If there is little correlation between subregions, the obliquely rotated solution will closely resemble the orthogonally rotated solution. For this analysis, the oblique solution was very similar to the orthogonal solution with only 4 of the 77 stations transferring subregions. With only a few exceptions, which were found between several adjacent subregions, cross-correlations were quite weak, ranging mostly from -0.20 to 0.20. This suggests that the physical factors determining the uniqueness of O_3 concentrations for each of these subregions are indeed different.

Examination of the temporal characteristics of these six homogeneous subregions can be accomplished with *principal component scores*, which are weighted summed values for the days over the stations, the weights being the component loadings. The component scores are standardized, so they have a mean of zero and standard deviation of one. When plotted as a time series, the scores provide excellent insight into the spectrum of temporal variance experienced by each subregion - insight that would prove very useful for the modeler trying to determine the optimum time for model simulations.

For the sake of brevity, only three of the six time series are presented (Figure 2). They reveal several interesting features, most notably the presence and timing of strong seasonalities. The Northeast Subregion (Fig. 2a) exhibits a strong seasonality in which the highest principal component scores (corresponding to high O_3 concentrations) occur much more frequently during the months of June through August. Low scores occur predominantly during the months of April, September and October. This strong seasonality is not nearly as evident in the Southwest time series (Fig. 2b) as the timing of high and low scores is more sporadic. Florida's time series (Fig. 2c) reveals another strong seasonality, however, one which is out-of-phase with that occurring in the Northeast. For this subregion, the highest concentrations tend to occur most frequently during April and May, with low concentrations dominating the remainder of the season.

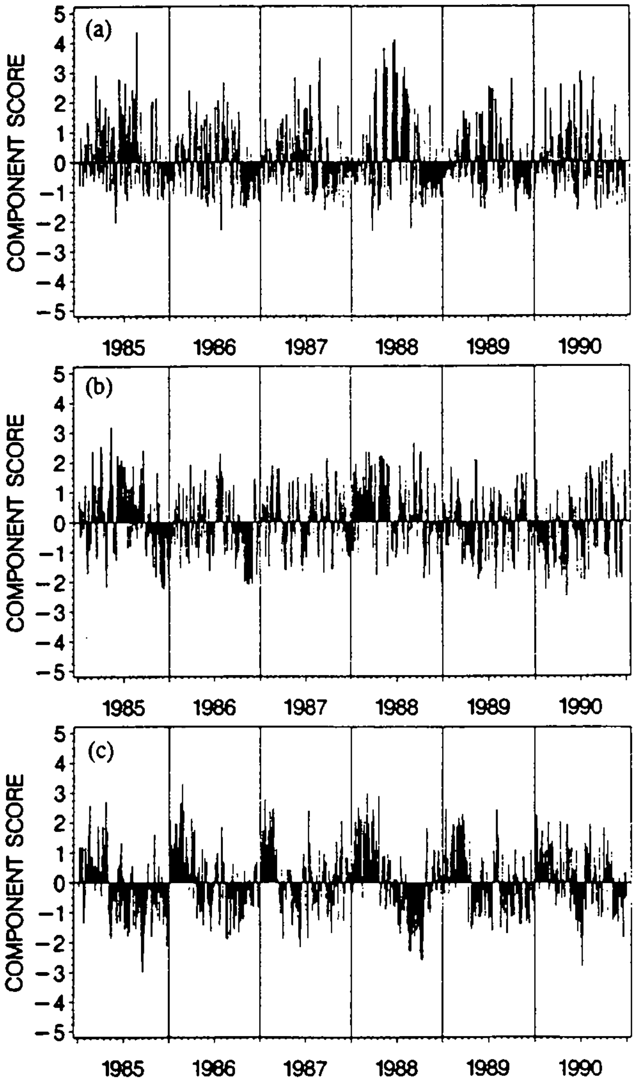

Figure 2. Time series of the daily, standardized principal component scores associated with three of the subregions: (a) Northeast, (b) Southwest and (c) Florida.

Insight into the day-to-day variability occurring within each subregion can also be gained from the time series. For illustrative purposes lets examine 1988, when the Northeast subregion recorded 10 days with principal component scores greater than 3.0 standard deviations above the six year normal. No other year for this subregion recorded more than one score greater than 3.0. Additionally, eight of these days occurred within a 31-day period, from June 16 through July 16; signifying the most intense episode experienced by the subregion. Similar insight can be made upon quick visual examination of the time series associated with the other subregions. Modelers can easily select, for each subregion, the worst period (3-day, 7-day, month, season, etc.) on which to perform their simulations.

Having delineated the eastern United States into areas that exhibit homogeneous O_3 characteristics, further elucidation and comparison of these characteristics can now be accomplished through a variety of simple analyses involving the raw ozone data found at each station within subregions. Table 3 reveals that in comparison to the entire study domain (mean of 56.06 ppb), the Florida Subregion experiences considerably lower concentrations (48.74 ppb), and that concentrations exceeding 80 ppb (120 ppb) occur only one-half (one-fifth) as often as the domain concentrations. The mean concentrations for the Great Lakes (53.83), Northeast (54.61) and Southwest (55.15) are all slightly lower than the domain average, while those for the South (58.64) are slightly higher. Mean concentrations for the Mid-Atlantic Subregion (63.23) are considerably higher than the domain, as are the percentage of days exceeding 80 ppb (20.3%) and 120 ppb (1.9%).

Examination of the Coefficient of Variation (%) reveals the compared to the entire domain (39.02), less variability is found in the Mid-Atlantic (35.97), South (37.04), and Great Lakes (37.48) Subregions, while more is found in the Florida (40.17), Southwest (40.60) and Northeast (40.93) Subregions.

Table 3. Summary statistics of the O_3 (ppb) concentrations associated with each subregion and the entire domain.

SUBREGION	MEAN	C. V.	% > 80	% > 120
Great Lakes (n=26,964)	53.83	37.48	9.7	0.6
Northeast (19,260)	54.61	40.93	13.1	1.3
South (17,976)	58.64	37.04	15.0	0.9
Mid-Atlantic (14,124)	63.23	35.97	20.3	1.9
Southwest (14,124)	55.15	40.60	14.0	1.5
Florida (6,420)	48.74	40.17	7.1	0.2
Domain (98,868)	56.06	39.02	13.9	1.1

Examination of the monthly distributions associated with each subregion is also revealing (Figure 3). Statistics are presented in the form of a boxplot and include: 1st and 99th percentiles (open circles), 5th and 95th percentiles (bars), 25th and 75th percentiles (rectangle ends) and median (center line) and mean (filled circle). Figure 3(a) supports the earlier analysis in that a strong seasonality, centered around July, is observed in the Northeast Subregion. (This seasonality is also observed in the Mid-Atlantic, South and Great Lakes subregions; however, there are several interesting variations in its timing, strength and symmetry). Statistics for the Northeast Subregion reveal that concentrations in July are substantially higher than the next two closest months June and August (which are very similar). The next highest concentrations are observed in May, with April and September following closely, although September exhibits considerably more variability than April. Concentrations observed during October are much lower than any other month.

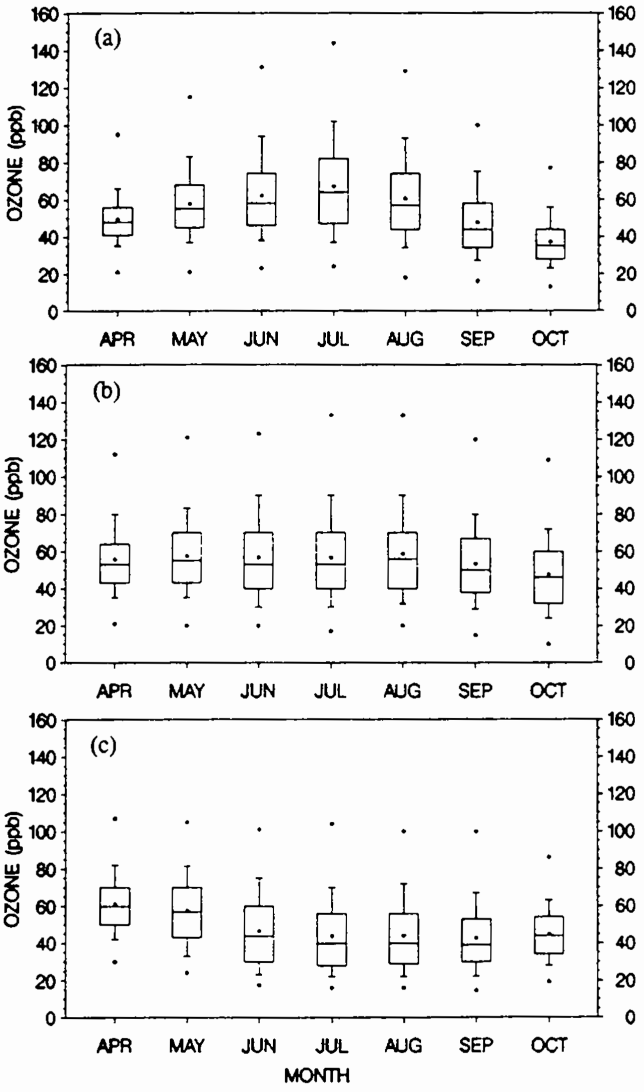

Figure 3. Boxplots of the monthly O_3 concentrations associated with three subregions: (a) Northeast, (b) Southwest and (c) Florida.

Also supporting the earlier analysis is the lack of a strong seasonality in the Southwest Subregion (Fig. 3b) and an out-of-phase seasonality observed in the Florida Subregion (Fig. 3c). The monthly statistics remain virtually the same through the Southwest Subregion's season, with perhaps a slight decrease in September and October. Monthly O_3 concentrations observed in the Florida Subregion are completely different than the other subregions. Highest concentrations are found during the months of April and May. Concentrations for the remaining months are considerably lower and exhibit little variation.

Having examined the monthly variability within subregions, we can now focus on the annual (April through October) variability as seen in Figure 4. The annual boxplots for the Northeast Subregion (Fig. 4a) reveal that the highest concentrations (means and especially the higher percentiles) occurred during 1988. 1985 provided the next highest concentrations, with the remaining years, 1986, 1987, 1989 and 1990 all experiencing more similar concentrations.

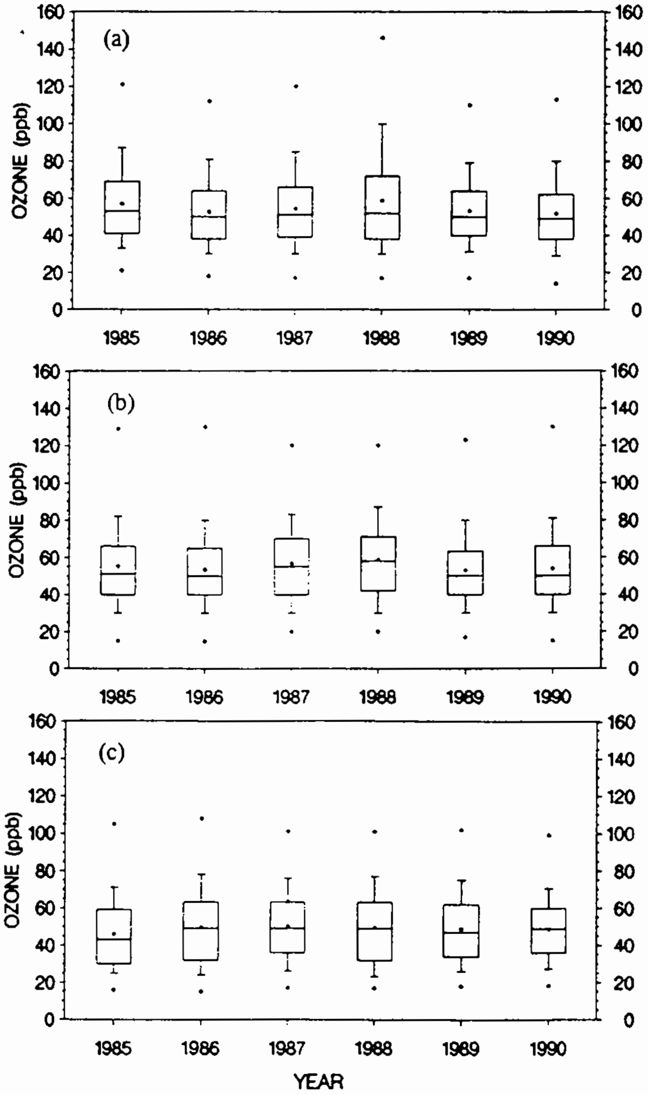

Figure 4. Boxplots of the annual O_3 concentrations associated with three subregions: (a) Northeast, (b) Southwest and (c) Florida.

Mean concentrations for the Southwest Subregions (Fig 4b.) were similarly high during 1987 and 1988. Concentrations for the years 1985 and 1990 were also similar and the next highest, followed by 1986 and 1989. Florida (Fig. 4c) exhibited the smallest annual variability. With the exception of 1985, which recorded slightly lower concentrations, annual statistics remained virtually constant.

In order to determine if these annual summary statistics exhibited trends within this limited six-year period, a simple linear regression was performed for all six subregions and the estimated slopes (and their standard errors) of the models calculated (Table 4). Although only two of the trends were statistically significant (90% C.L.), they did generally exhibit within-subregion consistency. For instance, all O_3 statistics associated with the Southwestern, Mid-Atlantic and Northeast Subregions show a slight decreasing tend. (Only the median O_3 concentrations associated with the Mid-Atlantic Subregion was significant). Conversely, the South Subregion has experienced a slight increase in all statistics over the six year period. Both the Great Lakes and Florida Subregions exhibit mixed results, with the means and medians increasing over the period and the higher percentiles generally decreasing. (The 99th% in Florida was significant). Statistics associated with the entire domain (all stations) indicated a general decreasing though not statistically significant trend.

Table 4. Estimated trend line slopes (ppb year $^{-1}$) for selected statistics for each subregion and the entire domain. (Standard error of estimate provided in parenthesis.)

SUBREGION	MEAN	MEDIAN	75th %	95th %	99th %
Great Lakes	0.16 (1.00)	0.23 (0.93)	-0.09 (1.36)	0.26 (2.19)	-0.23 (2.92)
Northeast	-0.57 (0.67)	-0.54 (0.29)	-0.83 (0.90)	-0.51 (2.55)	-0.57 (3.55)
South	0.19 (0.92)	0.00 (0.73)	0.23 (1.23)	0.77 (1.88)	1.46 (2.46)
Mid-Atlantic	-0.87 (0.64)	-0.94* (0.32)	-1.20 (0.86)	-0.69 (2.49)	-0.23 (3.54)
Southwest	-0.18 (0.60)	-0.06 (0.90)	-0.14 (0.81)	-1.31 (0.73)	-0.46 (1.28)
Florida	0.34 (0.35)	0.69 (0.55)	0.06 (0.47)	-0.31 (0.49)	-1.37¯ (0.54)
Domain	-0.16 (0.66)	-0.17 (0.62)	-0.26 (0.94)	-0.20 (1.83)	0.06 (2.51)

* Significant at 95%

¯ Significant at 90%

5. SUMMARY

The spatial and temporal variability of O_3 concentrations over the eastern United States during the period 1985 through 1990 was examined through the use of a multivariate statistical technique called Principal Component Analysis. The original data set, which contained 77 correlated variables (monitors) was reduced to six uncorrelated principal components, while still explaining almost two-thirds (64.02) of the total variance. Application of Kaiser's Varimax rotation led to the identification of six separate, contiguous subregions which each exhibit statistically unique O_3 concentration characteristics.

When compared to the entire domain, the Great Lakes, Northeast, Southwest and Florida Subregions observed lower mean O_3 concentrations. Conversely, the South and Mid-Atlantic Subregions recorded higher than domain means. Variability, as defined by the Coefficient of Variation, was highest in the Northeast, Southwest and Florida Subregions, and lowest in the Great Lakes, South and Mid-Atlantic Subregions. The percentage of observations exceeding concentration thresholds of 80 and 120 ppb, were higher in the Mid-Atlantic and Southwest Subregions, lower in the Great Lakes and Florida Subregion and near the domain average in the Northeast and South Subregions.

A strong seasonal cycle was observed in the Great Lakes, Northeast, Mid-Atlantic and South Subregions. The timing, strength and symmetry varied across these subregions. The Southwest Subregion exhibited little seasonality, while the Florida Subregion contained a strong, out-of-phase seasonality with the maximum occurring during the months of April and May.

Annually, the highest O_3 concentrations generally occurred during 1988; however, several subregions (Florida and the Southwest) recorded equally high concentrations in other years. No one other year stood out statistically across all subregions. Trend analyses indicated a slight, though not statistically significant, decrease in O_3 statistics for the Northeast, Mid-Atlantic and Southwest Subregions and a slight increase over the South Subregion. Ozone concentration trends for the Great Lakes and Florida Subregions were mixed indicating slight increases in mean and median concentrations and slight decreases in the higher percentiles.

These results have provided a statistically and physically based rationale for choosing distinctive geographical areas for interpreting O_3 air quality distributions and trends. Since data from stations within subregions exhibit homogeneous variability, we have been able to develop regionwide O_3 indicators which have provided meaningful insight into the seasonal and annual concentration trends of the six subregions. The analysis has also suggested that trends analyses for determining general progress in improving O_3 air quality could be based on aggregate statistics from clusters of monitors rather than from individual stations.

Disclaimer

This paper has been reviewed in accordance with the U.S. Environmental Protection Agency's policies and approved for presentation and publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

6. REFERENCES

Ashbaugh, L. L., L. O. Myrup, and R. G. Flocchini (1984) A principal component analysis of sulfur concentrations in the western United States. *Atmos. Environ.*, **18**, 783-791.

Cox, W. M. and J. Clark (1981) Ambient ozone concentration patterns among eastern United States urban areas using factor analysis. *J. Air Pollut. Control Assoc.*, **31**, 762-766.

Crutcher, H. L., R. C. Rhodes, M. E. Graves, B. Fairbairn, and A. C. Nelson (1986) Application of cluster analysis to aerometric data. *J. Air Pollut. Control Assoc.*, **36**, 1116-1122.

Eder, B. K. (1989) A principal component analysis of $SO_4^=$ precipitation concentrations over the eastern United States. *Atmos. Environ.*, **23**, 2739-2750.

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**, 187-201.

Lefohn A. S. and A. A. Lucier (1991) Spatial and temporal variability of ozone exposure in forested areas of the United States and Canada: 1978-1988. *J. Air Waste Manage. Assoc.*, **41**, 694-701.

Liu, S. C., M. Trainer, F. C. Fehsenfeld, D. D. Parrish, E. J. Williams, D. W. Fahey, G. Hubler, and P. C. Murphy (1987) Ozone production in the rural troposphere and the implications for regional and global ozone distribution. *J. Geophys. Res.*, **92**, 4191-4207.

Logan, J. A. (1989) Ozone in rural areas of the United States. *J. Geophys. Res.* **94**, 8611-8532.

Meagher, J. F., N. T. Lee, R. J. Valente and W. J. Parkhurst (1987) Rural ozone in the southeastern United States. *Atmos. Environ.*, **21**, 605-615.

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.

Richman, M.B. (1981) Obliquely rotated principal components: An improved meteorological mapping technique? *J. Appl. Meteor.*, **20**, 1145-1159.

United States Environmental Protection Agency (1985) Quality Assurance requirements for state and local air monitoring stations (SLAMS). 40 CFR, Part 58, Appendix A.

Vukovich, F. M. and J. Fishmann (1986) The climatology of summertime O_3 and SO_2. *Atmos. Environ.*, **20**, 2423-2433.

TECHNICAL REPORT DATA

(Please read Instructions on the reverse before comple

1. REPORT NO.
EPA/600/A-92/124

2.

3.

4. TITLE AND SUBTITLE
THE SPATIAL AND TEMPORAL ANALYSIS OF NON-URBAN OZONE CONCENTRATIONS OVER THE EASTERN UNITED STATES USING ROTATED PRINCIPAL COMPONENT ANALYSIS

5. REPORT DATE

6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)
Brian K. Eder

8. PERFORMING ORGANIZATION REPORT NO.

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Atmospheric Research & Exposure Assessment Laboratory
Office of Research and Development
U. S. Environmental Protection Agency
Research Triangle Park, North Carolina 27711

10. PROGRAM ELEMENT NO.
A101T/G/51

11. CONTRACT/GRANT NO.

12. SPONSORING AGENCY NAME AND ADDRESS
Same as in 9

13. TYPE OF REPORT AND PERIOD COVERED

14. SPONSORING AGENCY CODE
EPA/600/09

15. SUPPLEMENTARY NOTES
To be presented at the 12th Conference On Probability and Statistics In The Atmospheric Sciences, June 22-26, 1992, Toronto, Ontario, Canada

16. ABSTRACT

The spatial and temporal variability of O_3 concentrations over the eastern United States during the period 1985 through 1990 was examined through the use of a multivariate statistical technique called Principal Component Analysis. The original data set, which contained 77 correlated variables (monitors) was reduced to six uncorrelated principal components, while still explaining almost two-thirds (64.02) of the total variance. Application of Kaiser's Varimax rotation led to the identification of six separate, contiguous subregions which each exhibit statistically unique O_3 concentration characteristics.

When compared to the entire domain, the Great Lakes, Northeast, Southwest and Florida Subregions all observed lower mean O_3 concentrations. Conversely, the South and Mid-Atlantic Subregions recorded higher than domain means. Variability, as defined by the Coefficient of Variation, was highest in the Northeast, Southwest and Florida Subregions, and lowest in the Great Lakes, South and Mid-Atlantic Subregions. The percentage of observations exceeding concentration thresholds of 80 and 120 ppb, were higher in the Mid-Atlantic and Southwest Subregions, lower in the Great Lakes and Florida Subregion and near the domain average in the Northeast and South Subregions.

A strong seasonal cycle was observed in the Great Lakes, Northeast, Mid-Atlantic and South Subregions. The timing, strength and symmetry varied across these subregions. The Southwest Subregion exhibited little seasonality, while the Florida Subregion contained a strong, out-of-phase seasonality with the maximum occurring during the months of April and May.

Annually, the highest O_3 concentrations generally occurred during 1988; however, several subregions (Florida and the Southwest) recorded equally high concentrations in other years. No one other year stood out statistically across all subregions. Trend analyses indicated a slight, though not statistically significant, decrease in O_3 statistics for the Northeast, Mid-Atlantic and Southwest Subregions and a slight increase over the South Subregion. Ozone concentration trends for the Great Lakes and Florida Subregions were mixed indicating slight increases in mean and median concentrations and slight decreases in the higher percentiles.

These results have provided a statistically and physically based rationale for choosing distinctive geographical areas for interpreting O_3 air quality distributions and trends. Since data from stations within subregions exhibit homogeneous variability, we have been able to develop regionwide O_3 indicators which have provided meaningful insight into the seasonal and annual concentration trends of the six subregions. The analysis has also suggested that trends analyses for determining general progress in improving O_3 air quality could be based on aggregate statistics from clusters of monitors rather than from individual stations.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT
RELEASE TO PUBLIC

19. SECURITY CLASS *(This Report)*
UNCLASSIFIED

20. SECURITY CLASS *(This page)*
UNCLASSIFIED

21. NO. OF PAGES
7

22. PRICE

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE