Addendum 3: Presented at the Air and Waste Management Association International Specialty Conference 'Visibility and Fine Particles', October 1989, Estes Park, CO.

EMPIRICAL ORTHOGONAL FUNCTION ANALYSIS OF THE PARTICULATE SULFATE CONCENTRATIONS MEASURED DURING WHITEX

Kristi A. Gebhart
National Park Service
CIRA, Colorado State University
Fort Collins, CO 80523

Douglas A. Latimer
Latimer & Associates
P.O. Box 4127
Boulder, CO 80306-4127

James F. Sisler
Cooperative Institute for Research in the Atmosphere
Colorado State University
Fort Collins, CO 80523

Abstract

Empirical orthogonal function (EOF) analysis is used to systematically examine the spatial and temporal trends which existed in the fine ammonium sulfate concentrations collected during the Winter Haze Intensive Tracer Experiment (WHITEX) during January and February 1987 on the Colorado Plateau. EOF analysis reduces the many observed spatial and temporal patterns in the data to only a few which are then examined to see if they are consistent with the meteorology and SO_2 sources which existed during that time. It is found that they are. Evidence of the impact of Navajo Generating Station as well as transport from sources outside the study region to the WHITEX monitoring sites are seen in the EOF maps. The time factors which indicate which EOF dominates during each time period are consistent with what would be expected based on the observed meteorology.

EMPIRICAL ORTHOGONAL FUNCTION ANALYSIS OF THE PARTICULATE SULFATE CONCENTRATIONS MEASURED DURING WHITEX

Introduction

WHITEX

The Winter Haze Intensive Tracer Experiment (WHITEX)[1] was carried out from January 7 through February 18, 1987 in the Colorado River area of the Colorado Plateau. It was designed to study the ability of various receptor modeling techniques to attribute visibility impariment to a single point source, the Navajo Generating Station (NGS), from which a unique tracer was released during the experiment.

During WHITEX, 6 and/or 12-hour averages of particulate ammonium sulfate were measured concurrently at 13 different sites in the study region. Examination of the spatial patterns and temporal trends in these data can yield insight into the sources and meteorological conditions which influenced concentrations in the study area during WHITEX. For example, relatively uniform concentrations thoughout the area would indicate long range transport of sulfur into the area from distant sources. Conversely, local mesoscale influences from sources within the region would be expected to cause highly nonuniform concentrations.

SO_2 Sources

There are several sulfur dioxide sources in the region which could potentially influence sulfur concentrations in the study area. Nearby coal-fired power plants and those in the Colorado River drainage, and their annual average SO_2 emission rates in tons/day in 1987 include: NGS (163), Bridger (146), San Juan (116), Four Corners (106), Mohave (52), Cholla (45), Naughton (41), Hayden (41), Huntington Canyon (33), Craig (24), Coronado (18), Hunter (16), and Carbon (16). Most of these power plants, however, are more than 200 Km from NGS. Other large sources of SO_2 which are relatively close are the copper smelters in southern Arizona and Mexico, including Magma (480), Nacozari (380), Cananea (240), Asarco-Hayden (92), and Inspiration (54). A smelter in Douglas, AZ ceased operations sometime in January, 1987. The locations of these sources in relationship to the WHITEX study area are shown in Figure 1.

Relationship of EOFs to Spatial Patterns

Empirical orthogonal function (EOF) analysis is a method which can be used to systematically examine the spatial and temporal patterns in the measured sulfur concentrations (or other data) across several monitoring sites in a region. While these patterns can be determined simply by creating isopleth maps of the concentrations for each of the measurement time periods, EOF analysis has the advantage of simplifying these many maps into a few patterns (EOFs) which can be linearly recombined to reproduce the observations for all time periods. It has been used by meteorologists to study patterns in precipitation, temperature, and other data for several decades[2] and has more recently been applied to air pollution data[3-8].

EOF analysis is the decomposition of a single data matrix (time by site) of pollutant concentrations into two matrices, one independent of time (EOF by site) and the other independent of location (time by EOF). Multiplication of the resulting two matrices regenerates the original matrix. The EOF by site matrix can be thought of as a series of isopleth maps (one for each EOF), of common concentration patterns in space, while the time by EOF matrix, plotted as a series of time plots, also one for each EOF, reveals the relative importance of each EOF for each sampling period. The advantage of the analysis is that while there are as many EOFs as there are sites, in general only a few are needed to explain most of the variance in the data. For example, in the WHITEX study, there are 85 12-hour time periods which we would like to examine but, only 2 EOFs are needed to explain more than 80% of the variance in the data.

Data

Discussion of Sulfur Concentrations

Table I gives the latitude, longitude, and elevation coordinates and the monitor types and laboratory techniques used to obtain the sulfur data at each of the 13 WHITEX sites.

Figure 2 shows time plots of the sulfur concentrations at these sites. The most dramatic feature is the mid-February episode which occurred during Julian Days (JD) 40-44. The highest concentrations measured during WHITEX occurred at Page and Bullfrog during this time period and concentrations at all sites were high. The episode ended with a frontal passage on JD 45. Three other episodes of interest are those occuring on JD 25-28 when concentrations were elevated at many sites along the Colorado River, including Cisco, Green River, Hite, Canyonlands, and Bullfrog; JD 14-18 when the highest concentrations at Hopi Point were measured; and JD 34-36 when concentrations were relatively high at Hopi Point. The Hopi Point site is of special interest because it is located in Grand Canyon National Park where visual air quality is a valuable resource.

Missing data

Data must exist for every site for every time period which is included in the analysis. Therefore, three steps were taken to deal with non-missing data. First, sites and times with more than 25% missing data were not used. This eliminated Meadview because samples there were collected as 24-hour averages once every three days. Cisco was also eliminated because there were too many missing data (48%) there. In addition, 6 of the 85 12-hour time periods from 1/7/87 through 2/18/87 were not used because data were missing at several sites. These time periods are those which began at Julian day (JD) 7 hour 8 and JD 7 hour 20, when data were missing at 4 sites, JD 10 hour 8, JD 11 hour 20, and JD 48 hour 20 when data were missing at 3 sites, and JD 48 hour 20 when data were missing at 6 sites. These deletions left 79 time periods and 11 sites for use in the EOF analysis.

Missing data preceded and followed by good data at the same site were filled by linear interpolation across time. 0.9% of the data were filled in this way. Other missing data were replaced by a distance weighted (inverse of the squared distances) average of concentrations at sites with non-missing data for the same time period. Approximately 3% of the data were replaced by this method.

Methodology

The matrix of centered sulfur concentrations (mean for the site subtracted from each concentration) is decomposed into two matrices as follows:

$$Z = A \quad P \qquad (1)$$

where:

- Z = Centered concentration matrix with the rows corresponding to times of observation and the columns to locations.
- A = Orthonormal matrix of dimensionless time weighting factors, where the rows correspond to the time periods and the columns correspond to the EOFs.
- P = Orthogonal matrix of spatial empirical orthogonal functions. One row for each EOF, one column for each site. Values are weighted deviations from the mean concentration at each site.

P is the eigenvectors of Z^TZ (Z transpose times Z) multiplied by the matrix which contains the square roots of the eigenvalues of Z^TZ on the diagonal and zeros elsewhere. The eigenvalues have units of concentration squared so entries in the P matrix are units of concentration. A is found by postmultiplying the centered data, Z, by the inverse of P. More details of the decomposition can be found in texts on multivariate analysis[9,10].

EOFs are orthogonal, statistically simple, and each explains a decreasing amount of the total variance, however they are not a unique decomposition of the data and frequently the maps associated with each are more physically interpretable and better represent the patterns in the input data after rotation. The advantages and disadvantages of orthogonally and obliquely rotated and unrotated EOFs are discussed in detail by Richman[2]. For the WHITEX study, the EOFs were rotated using the Varimax[11] criteria which forces each site to load as stongly as possible onto one EOF while loading weakly onto all others. This can be useful for determining the source areas and meteorolgical conditions which most influence each site.

Results

Eigenvalues of Correlation Matrix

The eigenvalues of the correlation matrix (correlation between sites) are approximately equal to the signal to noise ratio for each factor. These were used to determine how many EOFs to retain for Varimax rotation. The results are shown in Table II. Only the first 2 eigenvalues show a signal to noise ratio greater than 1 and therefore only these EOFs were retained for the Varimax rotation.

Eigenvalues and variance explained by each EOF

The EOF analysis was then carried out using the centered data cross product matrix. EOFs of this matrix are desirable because the values on the maps are weighted deviations from the mean concentration at each site and, thus they have physical meaning. The weights are the time factors.

The eigenvalues of each EOF as well as the variance explained by each before rotation and after Varimax rotation of 2 EOFs are shown in Table III.

Discussion of unrotated EOFs

The time weighting factors for the first two unrotated EOFs are shown in Figure 3. The corresponding maps for these EOFs are shown in Figures 4 and 5. The signs of the values indicate positive or negative deviation from the mean, however, the signs for any given time period depend also on the sign of the time weighting factor. For example, for EOF 2, on JD 25-28, the time factor is negative, so all signs on the map would be reversed for that time period.

When the EOFs are not rotated, the first EOF should resemble the spatial pattern which is most like a contour plot of the mean sulfur concentrations at each site. By comparing the map of unrotated EOF 1 (Figure 4) which explains 70% of the variance, to a contour plot of the actual mean sulfur concentrations shown in Figure 6, it can be seen that this is true. The two maps have many similar features. The highest mean concentration, 342 ng/m^3 is at

Green River, but the mean at Page is only slightly lower at 318 ng/m^3. On EOF map 1, the highest loading is at Page with values decreasing as the distance from Page increases. The map of mean sulfur concentrations is similar, though it shows concentrations increasing again north of Hite. Because Page is the monitoring site nearest to NGS, the EOF map as well as the map of mean concentrations both suggest that NGS has a strong influence on mean sulfur concentrations in the study area.

The succeeding EOFs, though not all shown, are also interesting because they represent other common patterns in the measured sulfur concentrations. Based on the gradients, NGS appears to be a dominant source of sulfur in EOFs 1, 3, and 4; while in 5 and 6, it does not appear to be the dominant source, but its influence is seen in perturbations in the gradients near Page. Transport of sulfur into the study region from the south and southwest, perhaps indicating the copper smelters or Mohave Power Plant are sources is seen in EOF 2 when weighted positively. When EOF 2 is weighted negatively transport of sulfur into the region from sources north of the study area such as the Hunter and Huntington power plants in central Utah is suggested. Transport from the east is indicated by EOF 5. It should be remembered, however, that only the first two EOFs represent more signal than noise. Also, as seen in Table III, the variance explained by each of the EOFs ranked 3 through 11 is 5% or less.

The time weighting factors (Figure 3) show that most sulfur episodes, are not uniquely singled out by any one of the unrotated EOFs. Sulfur concentrations are seen to be a linear combination of several patterns.

Discussion of the rotated EOFs

The rotated EOFs are potentially more useful for examining the conditions which may influence each receptor because each site should load strongly to only one factor.

The time factors which result after Varimax rotation of 2 EOFs are shown in Figure 7 and the maps in Figures 8 and 9. EOF 1, associated with 58% of the variance, indicates dispersion of sulfur outward from NGS. A preferred transport direction appears to be to the north and northeast of Page. However, sulfur concentrations are greater than the means at all monitoring sites in the study region.

EOF 2 (32% of the variance) also suggests that NGS is the dominant source of sulfur for time periods when this EOF is important, although the preferred transport direction appears to be southwestward from Page.

The JD 13-17 episode is attributed solely to EOF 2. This is consistent with the synoptic winds during this time period which were northeasterly, leading to transport of NGS emissions to Hopi Point in the Grand Canyon.

The JD 25-28 episode is a combination of a positive contribution from EOF 1 and a negative contribution from EOF 2. Synoptic winds were moderate from the southeast during this time. This would transport NGS emissions towards Bullfrog, Hite, and Green River just as the EOF maps indicate.

JD 34-36 is positively weighted for EOF 2 which indicates transport of NGS emissions southwestward towards Hopi Point. High pressure and light winds from the northeast existed during these two days.

The major mid-February episode is the most complicated. The magnitudes of time factors are large for both patterns during this episode (JD 40-45). EOF 1 is weighted strongly positive for JD 42-45 and EOF 2 is positive from 40-44. EOF 2 is negative during JD 44-45. This is consistent with the meteorology which existed during this time period. For several days ending JD 39 synoptic winds were moderately strong from the southeast. Then a high pressure with weak synoptic winds (mesoscale influences dominating) persisted during JD 40-44. On JD 44 the synoptic pressure gradient strengthened leading to southeasterly winds again. A frontal passage with strong winds cleared pollutants from the area on JD 45.

Conclusions

EOF analysis has been used to systematically examine the spatial and temporal trends which existed in the fine sulfur data collected at 11 sites during the 6-week WHITEX experiment of January and February 1987. Only 2 of the EOFs have signal/noise ratios greater than 1, however, these factors, explain approximately 80% of the variance. Examination of the EOF maps and time factors shows that they are consistent with the meteorology and sources which existed during this time. The results qualitatively indicate that emissions from NGS had a more dominant influence on spatial sulfur patterns in the study area during WHITEX than any other sources. This is consistent with the quantitative results of the differential mass balance[12] and tracer mass balance[13] models which both indicate that NGS was the largest contributor of sulfur to both Hopi Point and Page during WHITEX.

References

1. W. C. Malm, M. Pitchford, and H. K. Iyer, "Design and Implementation of the Winter Haze Intensive Tracer Experiment-WHITEX", Presented at the 81st Annual APCA Meeting, Dallas, TX, June 19-24, Paper No. 88-52.1, 1988.

2. M. B. Richman, "Rotation of Principal Components", Journal of Climatology. Vol. 6, pp. 293-335, 1986.

3. J. T. Peterson, "Distribution of Sulfur Dioxide over Metropolitan St. Louis as Described by Empirical Orthogonal Functions, and its Relation to Meteorological Parameters", Atmospheric Environment, Vol. 4, pp. 501-518, 1970.

4. R. C. Henry and K.A. Gebhart, "Empirical Orthogonal Functions of Airborne Sulfur in the Western United States Related to Sources and Upper Level Winds", Presented at the 81st Annual APCA Meeting, Dallas, TX, June 19-24, Paper No. 88-52.7, 1988.

5. R. C. Henry, "Location of Airborne Particulate Sources Empirical Orthogonal Function Analysis", Preprint Volume: Third Joint Conference on Applications of Air Pollution Meteorology, San Antonio, TX. Published by the American Meteorological Society, 45 Beacon Street, Boston, MA 02108, 1982.

6. L. L. Ashbaugh, L. O. Myrup, and R. G. Flocchini, "A Principal Component Analysis of Sulfur Concentrations in the Western United States", Atmospheric Environment, Vol. 18, pp. 783-791, 1984.

7. W. C. Malm, K. A. Gebhart, and R. C. Henry, "Source Areas of Fine Sulfur in the Western United States as Investigated by Principal Component Analysis and Residence Time Analysis", Presented at the 81st Annual APCA Meeting, Dallas, TX, June 19-24, Paper No. 88-54.5, 1988.

8. P. K. Mueller and G. M. Hidy, "The Sulfate Regional Experiment: Report of Findings", Electric Power Research Institute (EPRI), 3412 Hillview Avenue, Palo Alto, CA 94304, EPRI Document Number EA-1901, Vol. 3, 1983.

9. C. L. Lawson and R. J. Hanson, Solving Least Squares Problems, Princeton-Hall, Englewood Cliffs, NJ, 1974.

10. P. E. Green, Mathematical Tools for Applied Multivariate Analysis, Academic Press, New York, 1976.

11. H. F. Kaiser, "The Varimax Criterion for Analytic Rotation in Factor Analysis", Psychometrika, Vol. 23, pp. 187-200, 1958.

12. D. A. Latimer, "Application of a Differential Mass Balance Model to Attribute Haze in the Southwest", AWMA/EPA International Specialty Conference--'Visibility and Fine Particles', Estes Park, CO, Oct. 1989.

13. W. C. Malm, H. K. Iyer, and K. A. Gebhart, "Application of Tracer Mass Balance Regression to WHITEX data", AWMA/EPA International Specialty Conference--'Visibility and Fine Particles', Estes Park, CO, Oct. 1989.

14. R. A. Eldred, T. C. Cahill, M. Pitchford, and W. C. Malm, "IMPROVE--A New Remote Area Particulate Monitoring System for Visibility Studies", Presented at the 81st Annual APCA Meeting, Dallas, TX, June 19-24, Paper No. 88-54.3, 1988.

15. R. A. Eldred, T.A. Cahill, and P. J. Feeney, "Particulate Monitoring at US National Parks Using PIXE", 4th International Conference on PIXE, Tallahassee, FL, Nuclear Instruments and Methods in Physics Research, B22 III Atmospheric Applications, pp. 289-295, 1987.

16. C. F. Rogers, J. G. Watson, and J. Zwicker, "Fine Particle Measurements with a Size Classifying Isokinetic Sequential Aerosol Filter Sampler", Presented at the 81st Annual APCA Meeting, Dallas, TX, June 19-24, Paper No. 88-53.5, 1988.

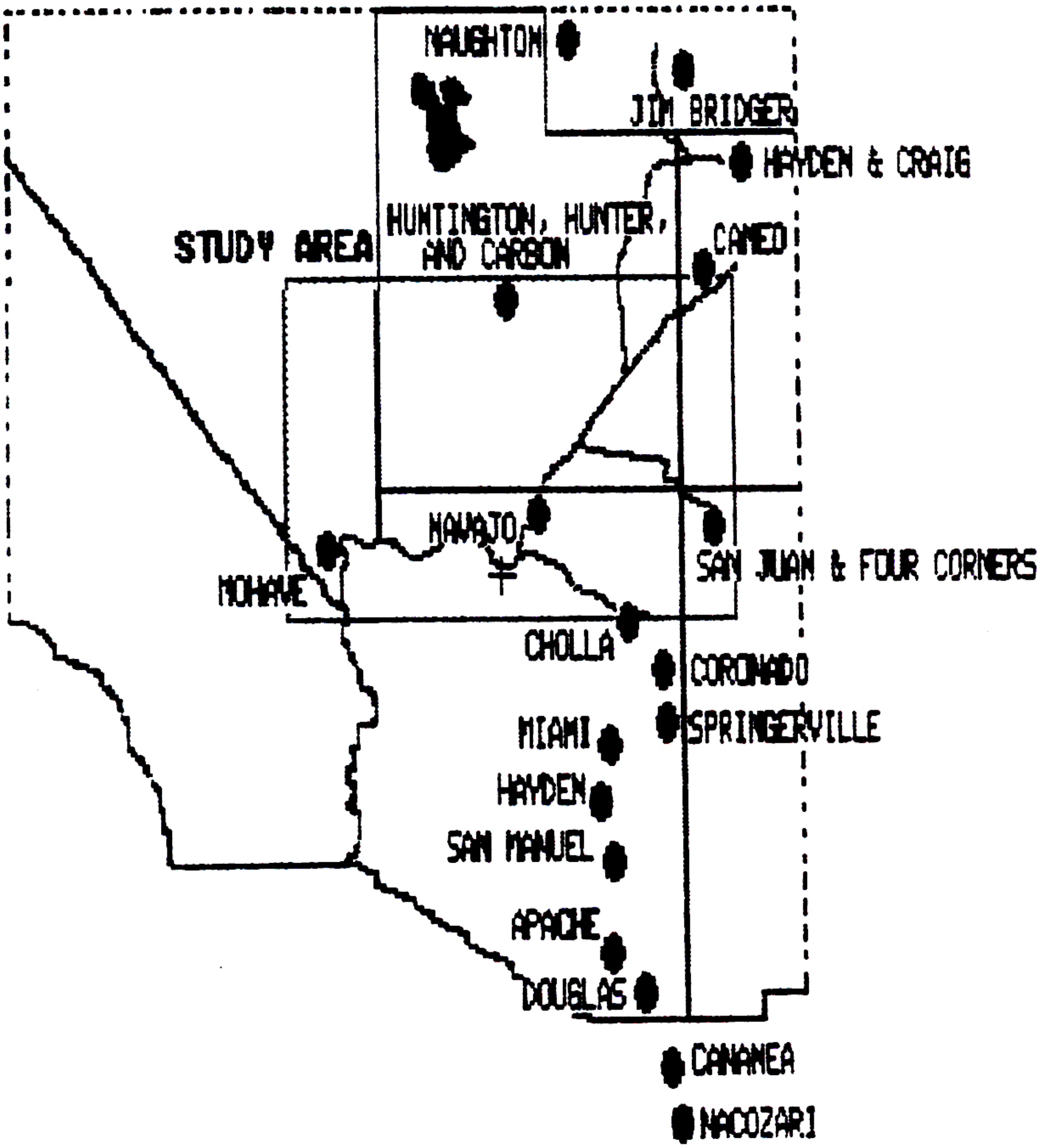

Figure 1. SO_2 sources in the southwestern United States.

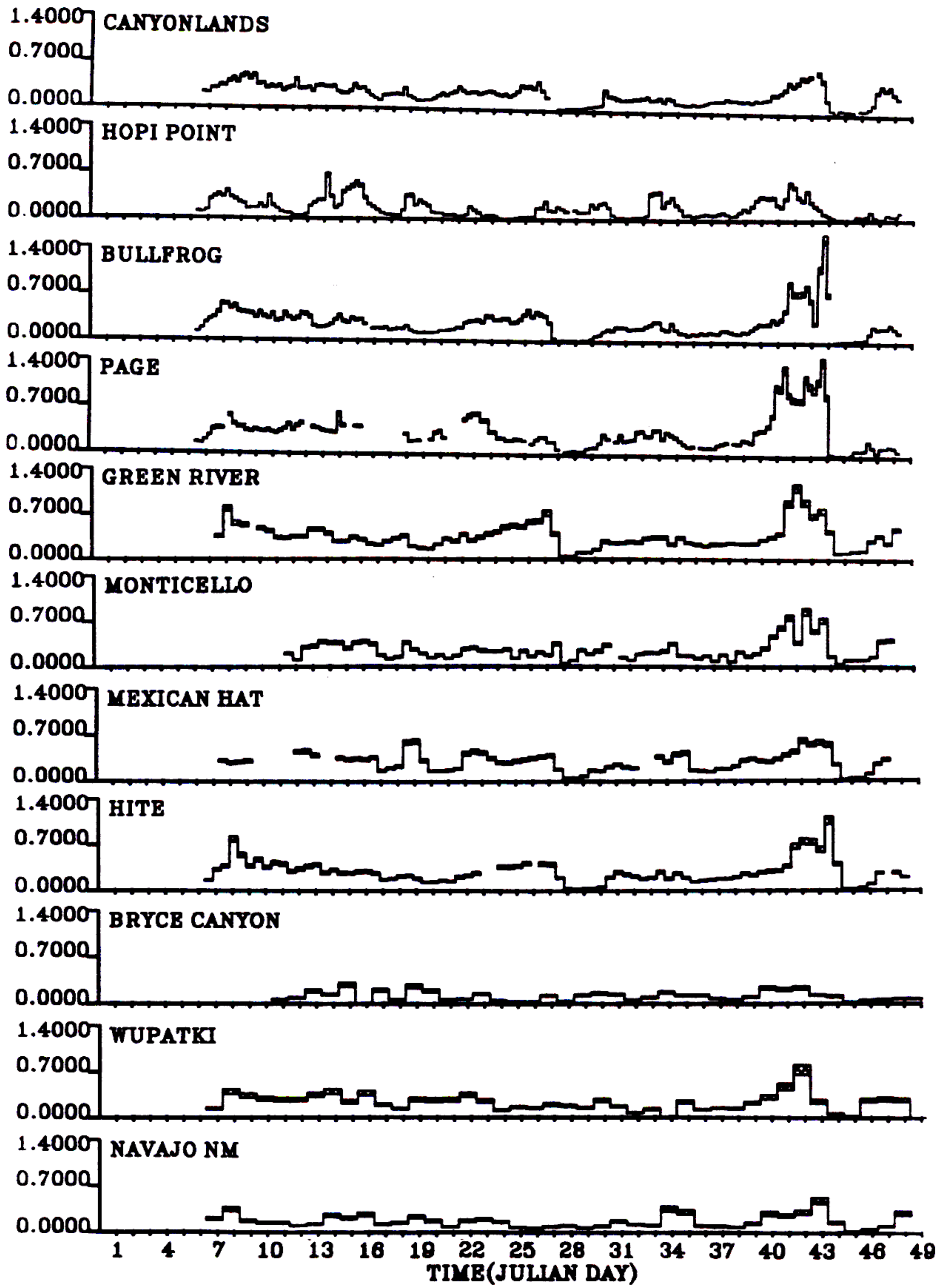

Figure 2. Time history of the fine sulfur concentrations measured during WHITEX.

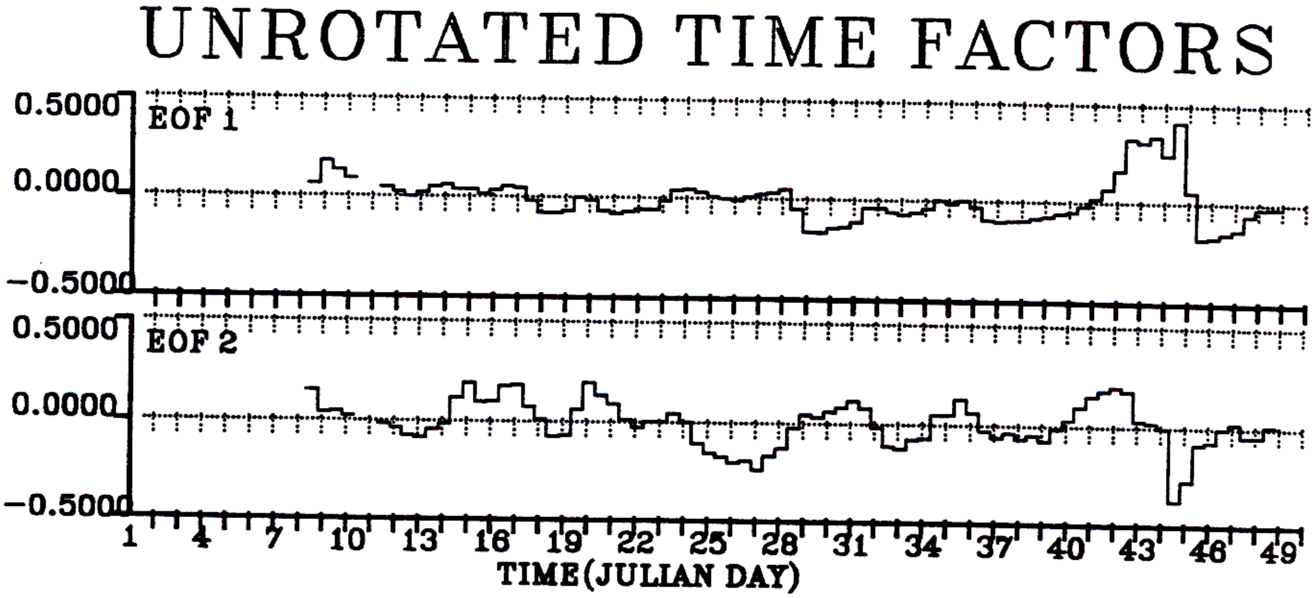

Figure 3. Time weighting factors for the first 2 unrotated EOFs.

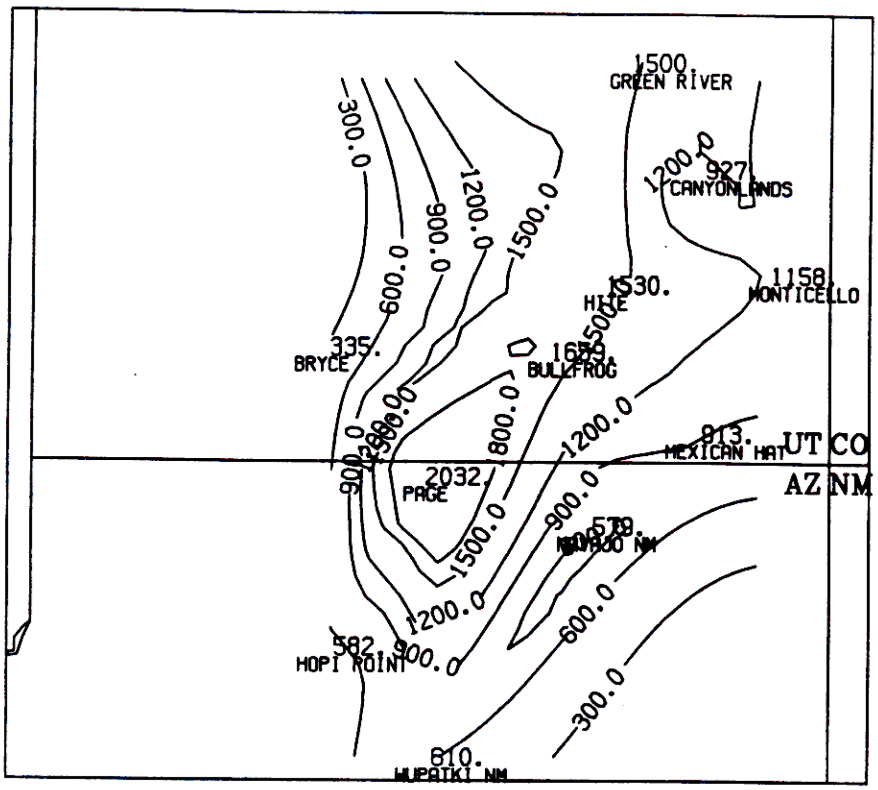

Figure 4. Map of unrotated EOF 1 explaining 70% of the variance. Values on the map are weighted deviations from the mean sulfur concentrations at each site (ng/m^3).

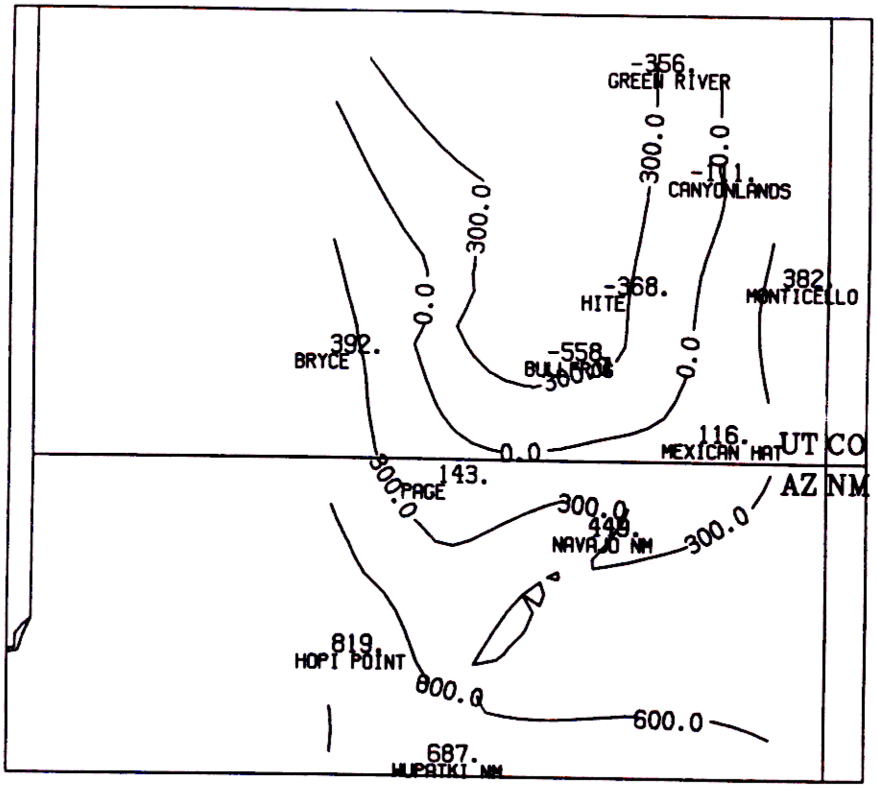

Figure 5. Map of unrotated EOF 2 explaining 10% of the variance. Values on the map are weighted deviations from the mean sulfur concentrations at each site (ng/m^3).

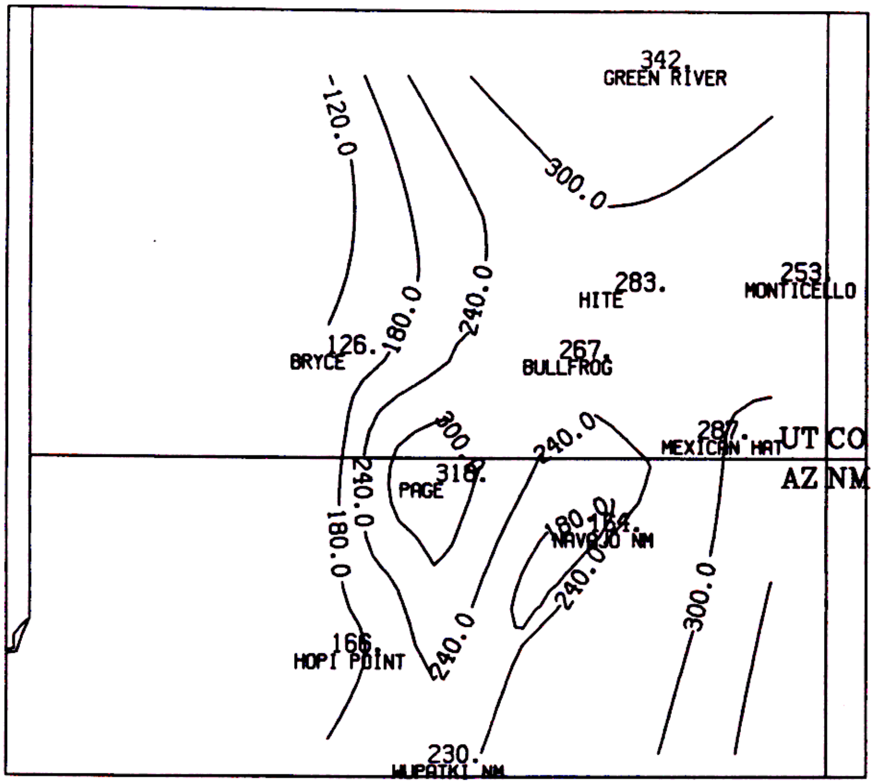

Figure 6. Map of mean fine sulfur concentrations (ng/m^3).

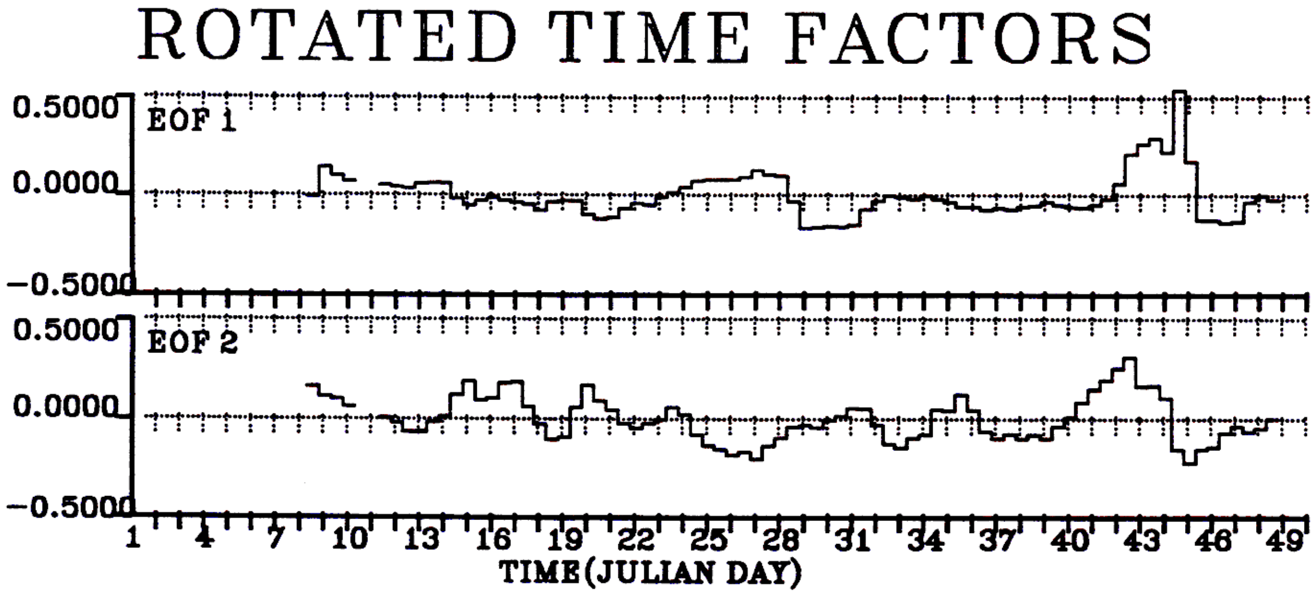

Figure 7. Time weighting factors for the 2 rotated EOFs.

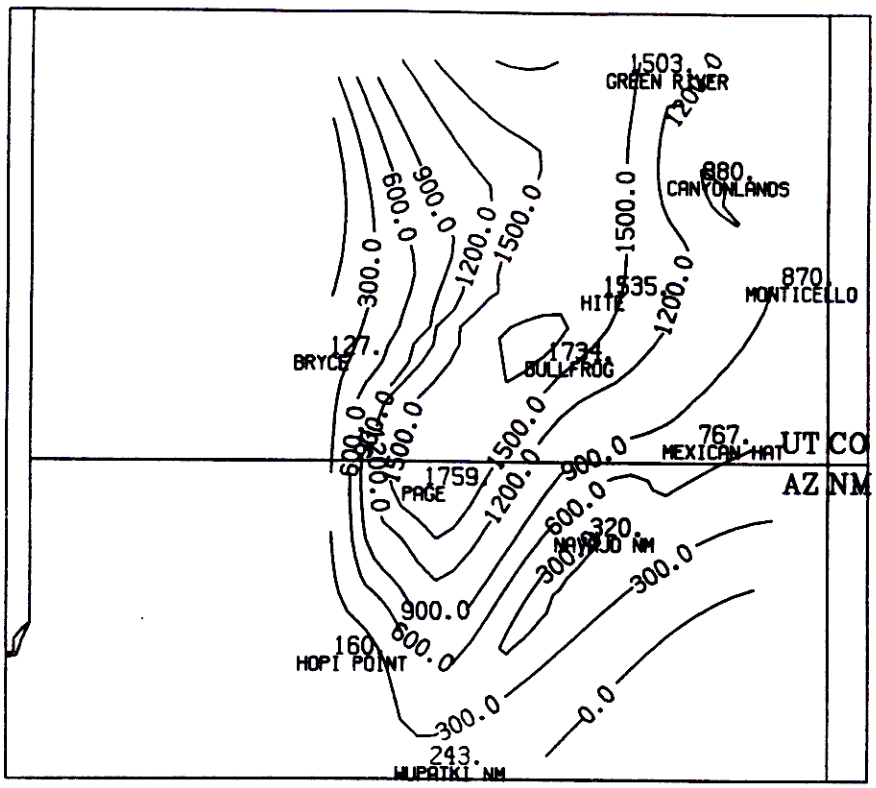

Figure 8. Map of rotated EOF 1 explaining 58% of the variance when 2 EOFs are rotated. Values on the map are weighted deviations from the mean sulfur concentrations at each site (ng/m^3).

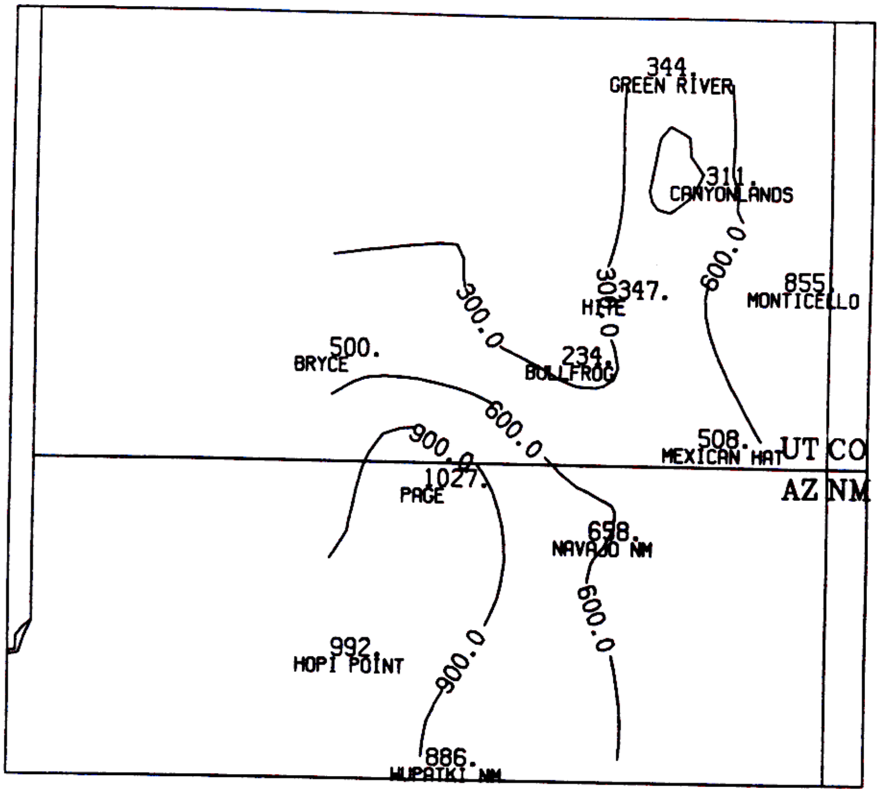

Figure 9. Map of rotated EOF 2 explaining 22% of the variance when 2 EOFs are rotated. Values on the map are weighted deviations from the mean sulfur concentrations at each site (ng/m^3).

Table I. Locations, elevations (feet), monitors, and analysis techniques used at each site. PIXE is proton induced x-ray emission, XRF is x-ray fluoresence. SFU is stacked filter unit.

Site	Latitude	Longitude	Elevation	Monitor	Analysis Technique
Canyonlands	38.45	109.83	5925	IMPROVE[14]	Ion Chrom.
Hopi Point	36.07	112.15	7100	IMPROVE	Ion Chrom.
Bullfrog	37.55	110.72	4167	IMPROVE	Ion Chrom.
Page	36.93	111.49	4300	IMPROVE	Ion Chrom.
Green River	38.97	110.22	4100	IMPROVE	PIXE
Cisco	39.17	109.17	4450	IMPROVE	PIXE
Monticello	37.93	109.33	6998	IMPROVE	PIXE
Mexican Hat	37.15	109.85	4700	IMPROVE	PIXE
Hite	37.88	110.37	3840	IMPROVE	PIXE
Bryce Canyon	37.57	112.18	7950	SFU[15]	PIXE
Wupatki NM	35.52	111.53	5250	SFU	PIXE
Navajo NM	36.68	110.53	7100	SFU	PIXE
Meadview	36.01	114.04	9709	SCISAS[16]	XRF

Table II. Eigenvalues of the sulfur correlation matrix. These are approximately equal to signal/noise for each factor.

Rank	Eigenvalue
1	6.70
2	1.57
3	0.65
4	0.59
5	0.46
6	0.33
7	0.29
8	0.17
9	0.13
10	0.08
11	0.03

Table III. Eigenvalues and variance explained by each EOF.

Rank	Eigenvalue ($ng^2\ m^{-6}$)	Variance Explained (%)	Cumulative Variance Explained (%)
a) Before Rotation			
1	1.6×10^7	70.1	70.1
2	2.3×10^6	10.1	80.2
3	1.2×10^6	5.4	85.6
4	8.0×10^5	3.6	89.2
5	7.0×10^5	3.1	92.4
6	5.2×10^5	2.3	94.7
7	4.7×10^5	2.1	96.8
8	3.0×10^5	1.4	98.1
9	1.9×10^5	0.9	99.0
10	1.2×10^5	0.6	99.6
11	9.7×10^4	0.4	100.0
b) After Varimax Rotation of 2 EOFs			
1	1.3×10^7	58.4	58.4
2	4.9×10^6	21.9	80.2