



Project Summary

Modified Factor Analysis of Selected RAPS Aerosol Data

Stephanie C. Davis

Target transformation factor analysis (TTFA) has been applied to a subset of the aerosol-composition data acquired during the Regional Air Pollution Study (RAPS) for St. Louis, Missouri. The RAPS program collected a large number of samples with 10 continuously operated dichotomous samplers from March 1975 to March 1977. The purpose of the present study was to evaluate the capability of TTFA to resolve sources of airborne particulate matter in a large set of ambient-aerosol samples. Only the samples from July and August 1976, both fine and coarse fractions, were examined in this study. To determine the most appropriate way to apply TTFA, two separate sets of data were analyzed—all the samples collected during the two months at a single station and all the samples collected during a single week at all 10 RAPS stations. A more detailed source resolution was obtained when the data were further divided into subsets corresponding to the fine- and coarse-particle fractions. Because of the large number of different sources in the St. Louis area, superior results were obtained from the examination of the variation in aerosol composition with time at a single location rather than the spatial variation over multiple sites during a shorter time period.

This Project Summary was developed by EPA's Environmental Sciences Research Laboratory, Research Triangle Park, NC, to announce key

findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).

Introduction

In 1971, the first National Ambient Air Quality Standards were established, setting limits on permissible levels of total suspended particulate matter. Because these regulations concentrated only on the absolute amounts of suspended particulate matter, control efforts were directed primarily toward the removal of the larger, more massive particles that comprise most of the total mass of particle emissions. There is growing evidence, however, that the smaller particles, although representing a smaller fraction of the total mass, present the greater threat to public health and welfare. Because of their light-scattering properties, the smaller particles also contribute to a larger extent to the reduction in visibility. As a result of the growing concern over the possible health effects of smaller diameter particulate matter in the atmosphere, there has been an increase in the development and use of aerosol sampling equipment that measures the concentrations of particles in the inhalable and fine particle-size fractions. New regulations, now under review by the Environmental Protection Agency, may include standards not only on the mass of total suspended particles but also on the levels of inhalable and fine particles.



The effectiveness of these regulations would be enhanced greatly if the sources and relative emission rates of these particles were known. Determining the sources of airborne particulate matter is a very difficult problem because of the complexity of urban ecosystems. The problem is aggravated by the presence of fugitive sources and the lack of understanding of the modes of formation and transport of airborne particles. The advent of receptor methodology and recent developments in the areas of trace-element analysis and multivariate statistical techniques now permit a much more detailed analysis of ambient aerosol samples. By providing detailed information on the sources of fine and inhalable particles, such techniques could become a major part of any strategy for controlling airborne particulate matter.

The Study

One approach to the problem of determining the contribution of each source to the total level of suspended particulate matter is the method of source-emissions inventories and atmospheric dispersion models. Using estimated emission rates for all known sources along with a model for their behavior following emission, this method attempts to determine the contribution of each source to the total aerosol concentration. The results of source-emissions inventories are too often unsatisfactory because of a lack of accurate source-emission profiles and the inability to include fugitive sources. The development of improved aerosol sampling equipment permitted the development of receptor methods for the determination of source contributions. These methods employ statistical analysis techniques to calculate the ambient aerosol samples. Receptor methods have primarily taken the form of the chemical element balance (CEB) method. In this method, it is assumed that the number and composition of the sources are known. The observed elemental concentrations in a set of ambient aerosol samples are fit by means of a regression analysis. Early applications of the chemical element balance were to ambient aerosol samples from Pasadena, California; Ghent, Belgium; Heidelberg, Germany; and Chicago, Illinois.

These early applications of the chemical element balance method suffered from a lack of reliable data for the source

emission profiles needed for the analysis. Recent improvements in the quality of source profile determinations have produced good agreement between the calculated and measured elemental concentrations for airborne particulate matter in Washington, DC, and for dichotomous sampler data in St. Louis, Missouri, and the Great Smoky Mountains. The scope of the chemical element balance method was further expanded in an extensive study of the Portland, Oregon, aerosol as part of the PACS program. This study included components such as vegetative and grass burning and chemical species such as volatile and nonvolatile carbon, nitrate, and sulfate that had not previously been used. By including the larger number of source components, very good agreement with the measured concentrations was achieved for the 23 elements and chemical species included in the analysis.

Chemical element balance methods have a major drawback in that they require an *a priori* knowledge of both the number and composition of the suspected source emissions. The only quantities calculated are the contribution of each of the presumed sources to each of the measured samples. Though extensive efforts by the groups in Maryland and Oregon have produced much better determinations of the elemental profiles of a number of sources, these determinations are time consuming, expensive, and, in many cases, inapplicable to the resolution of aerosol sources in different locations. In addition, source determinations of in-stack material cannot account for chemical and physical transformations of volatile species following emission.

Multivariate statistical techniques, including factor and cluster analyses, represent a different approach to aerosol source resolution. A major advantage of these methods is that they require no prior assumption about the nature of the system under study. In an early application of multivariate statistical techniques to elemental source resolution, factor and cluster analyses were used to identify sources of airborne particulate matter in the Boston urban aerosol. In this study, the data were first transformed into a standardized form that removed the effects of using different metrics in describing the data. The identification of sources was made by associating the largest calculated factor loading with the marker element for

each source. In a similar fashion, factor and cluster analyses were used to characterize the aerosol of Tucson, Arizona. Factor analysis was used to identify sources of particulate matter in St. Louis, Missouri. Factor analysis of a matrix of standardized variables resolves only the variance in the system, not the actual data values. In addition, the calculated factor scores provide only a measure of the relative importance of each source rather than the actual amount of contributed material.

Recently, a new form of factor analysis called target transformation factor analysis (TTFA) has been employed. This technique permits an analysis similar in nature to the chemical element balance method, but without the presumptions as to the number or nature of the sources in the system. Unlike the prior applications of factor analysis, TTFA allows the investigator to use his understanding of the system to determine the best representation of the actual source concentration profiles. TTFA also permits the direct calculation of the contribution of each source to each sample.

Target transformation factor analysis has been applied to a subset of the aerosol composition data acquired during the Regional Air Pollution Study (RAPS) for St. Louis, Missouri. In the RAPS program, automated dichotomous samplers were operated over a two-year period at 10 sites in the St. Louis metropolitan area. Ambient aerosol samples were collected in fine, 2.4 μm , and coarse, 2.4 to 20 μm , fractions. Samples were analyzed at the Lawrence Berkeley Laboratory for total mass by beta-gauge measurements and for 27 elements by x-ray fluorescence.

To determine the most appropriate way to apply target transformation factor analysis to a large set of aerosol samples, two subsets of this data were analyzed—all the samples collected during July and August at one station and the samples collected during a single week at all 10 RAPS stations. The data from station 112 were selected for the first subset and the data from the week beginning July 31, 1976, for the second. The relative completeness of the data in these two subsets was the primary criterion for their selection.

Station 112 was located near Washington University, west of downtown St. Louis. During the 62 days of July and August, filters were changed at 12-hour intervals, producing a total of 124

samples in each of the fine and coarse fractions. Data were missing for 48 of the samples, leaving a total of 200. Of the 27 elements determined for each sample, a majority of the determinations of 10 elements were below the detection limits. Since a complete and accurate data set is required to perform a factor analysis, these 10 elements were eliminated from the analysis. Table 1 lists the full 27 elements and the number of values below detection limits. Note that the element arsenic was excluded because of the large number of determinations below detection limits. Arsenic determinations by x-ray fluorescence are often unreliable because of an interference between the arsenic K x-ray and the lead L x-ray. A neutron activation analysis of these samples should produce better arsenic determinations. Reliable data for arsenic may be important to the differentiation of coal fly ash and crustal material, two materials with very similar source profiles. The remaining 200 samples were analyzed in two separate parts. Two different divisions in the data were made to determine which division per-

Table 1. RAPS Station 112, July and August 1976.

| Element | Number of Values Below Detection Limits |
|---------|---|
| Al | 15 |
| Si | 2 |
| P* | 179 |
| S | 0 |
| Cl | 35 |
| K | 0 |
| Ca | 0 |
| Ti | 39 |
| V* | 162 |
| Cr* | 151 |
| Mn | 29 |
| Fe | 0 |
| Ni | 95 |
| Cu | 39 |
| Zn | 0 |
| Ga* | 194 |
| As* | 196 |
| Se | 113 |
| Br | 1 |
| Rb* | 130 |
| Sr | 73 |
| Cd* | 161 |
| Sn* | 160 |
| Sb* | 180 |
| Ba | 131 |
| Hg* | 196 |
| Pb | 0 |

*Elements excluded from the analyses.

mitted the extraction of the most information. Separate analyses were then performed on each group. First, the data were divided into groups corresponding approximately to the months of July and August. In the second analysis, the data were divided into groups corresponding to the fine and coarse fractions of both months.

During the week beginning July 31, 1976, filters were changed at 12-hour intervals at five of the stations and at 6-hour intervals at four of the stations. A total of 182 samples were collected in each of the fine and coarse fractions. Data were missing for 43 samples, and another 50 samples had a majority of determinations below the detection limits; these 93 samples were excluded from the analysis. Of the 27 elements, a majority of the determinations of 12 had values below detection limits. These 12 elements were excluded. Table 2 lists the number of values below detection limits in the data for one week. Again, two separate divisions in the data were made and each individually analyzed. The data were first divided into two groups composed of days Saturday,

Table 2. One Week Beginning July 31, 1976.

| Element | Number of Values Below Detection Limits |
|---------|---|
| Al | 135 |
| Si | 27 |
| P* | 265 |
| S | 16 |
| Cl | 115 |
| K | 0 |
| Ti | 170 |
| Ca | 1 |
| V* | 318 |
| Cr* | 288 |
| Mn | 149 |
| Fe | 1 |
| Ni* | 267 |
| Cu | 120 |
| Zn | 18 |
| Ga* | 321 |
| As* | 316 |
| Se | 201 |
| Br | 0 |
| Rb* | 294 |
| Sr* | 224 |
| Cd* | 273 |
| Sn* | 258 |
| Sb* | 309 |
| Ba | 291 |
| Hg* | 320 |
| Pb | 3 |

*Elements excluded from the analyses.

Sunday, and Monday (SSM), and of days Tuesday, Wednesday, Thursday, and Friday (TWTF). The first group, with 115 samples, corresponds to days 213, 214, and 215, and the second group, with 156 samples, to days 216 through 219. The analysis was then repeated with the data divided into groups corresponding to the fine and coarse fractions with 139 and 132 samples, respectively.

The 27 elements determined for each sample account for only a small fraction of the total sample mass. The remaining mass consists primarily of oxygen, nitrogen, and carbon. Determining the source of particulate carbon is of growing concern because of the potential mutagenicity associated with carbonaceous particles. Measurements have shown that both soot and organic carbon can account for a sizeable fraction of the total mass of airborne particulate matter. Though no measurements of carbon are included in the RAPS data, that portion of the sample mass must be accounted for by the other sources. This can often lead to distortions in the scaling factors produced by the multiple regression analysis. In order to produce the best possible source resolutions, it is vital to have both accurate TSP measurements and determinations for as many elements as possible.

For use in the target transformation portion of the analysis as test vectors, a set of 17 potential source profiles was assembled. These include nine different crustal factors, two determinations of coal fly ash, three automobile factors, and three industrial sources. The 17 test vectors are listed in Table 3.

Conclusions

Using these data, target transformation analysis was applied, and several conclusions were made; they are as follows:

- Target transformation factor analysis (TTFA) can be used to resolve sources of airborne particulate matter from large sets of ambient aerosol samples. With this method, the contribution of each source to each sample can be calculated. In addition, TTFA requires no *a priori* assumptions as to the number or nature of the sources in the system. This method could be an important tool in the achievement of mandated limits on suspended particulate matter in the urban areas of this country by identifying those sources where imposition of control

Table 3. Source Profiles Used as Test Vectors, mg/g

| Element | Shale* | Sandstone | Carbonate | Soils | | | | | Rock | Soil 4 |
|---------|---------|-----------|----------------|--------|--------|------------|---------|---------|--------|--------|
| | | | | IAEA | Soil 1 | Soil 2 | Soil 3 | | | |
| Al | 80 | 25 | 4.2 | 82 | 82 | 69 | 71 | 82 | 60 | |
| Si | 273 | 368 | 24 | 332 | 200 | 260 | 330 | 282 | 350 | |
| P | 0.7 | 0.17 | 0.4 | 1 | | | 0.8 | 1.1 | | |
| S | 2.4 | 0.24 | 1.2 | | | | 0.85 | 0.26 | | |
| Cl | 0.18 | 0.01 | 0.15 | | | | 0.1 | 0.13 | | |
| K | 26.6 | 10.7 | 2.7 | 19 | 15 | 25 | 13.6 | 20.9 | | |
| Ca | 22.1 | 39.1 | 302 | 22 | 15 | 10 | 13.7 | 41.5 | | |
| Ti | 4.6 | 1.5 | 0.4 | 5 | 4 | 3 | 4.6 | 5.7 | | |
| V | 0.13 | 0.02 | 0.02 | 0.15 | 0.06 | 0.04 | 0.1 | 0.14 | 0.14 | |
| Cr | 0.09 | 0.035 | 0.011 | .03 | | | 0.2 | 0.1 | 0.06 | |
| Mn | 0.85 | 10 | 1.1 | 15 | 1.1 | 0.6 | 0.85 | 0.95 | 0.99 | |
| Fe | 47.2 | 9.8 | 3.8 | 45 | 32 | 20 | 38 | 56.3 | 37.8 | |
| Ni | 0.068 | 0.002 | 0.02 | 0.013 | 0.04 | 0.02 | 0.04 | 0.075 | | |
| Cu | 0.045 | 0.001 | 0.004 | 0.077 | 0.08 | 0.02 | 0.02 | 0.055 | | |
| Zn | 0.095 | 0.016 | 0.02 | 0.4 | | | 0.05 | 0.07 | 0.08 | |
| Ga | 0.02 | 0.02 | 0.004 | 0.02 | | | 0.03 | .015 | | |
| As | 0.013 | 0.001 | 0.001 | 0.09 | | | 0.005 | 0.002 | 0.005 | |
| Se | 0.0006 | 0.00005 | 0.00008 | 0.001 | | | 0.00001 | 0.00005 | 0.0006 | |
| Br | 0.004 | 0.001 | 0.0062 | 0.005 | | | 0.005 | 0.003 | | |
| Rb | 0.14 | 0.06 | 0.003 | 0.14 | | | 0.1 | 0.09 | | |
| Sr | 0.3 | 0.002 | 0.61 | 0.33 | | | 0.3 | 0.4 | | |
| Cd | 0.0003 | | | 0.0015 | | | | 0.002 | | |
| Sn | 0.006 | | | | | | | 0.002 | | |
| Sb | 0.0015 | | 0.0002 | 0.56 | | | | 0.0002 | 0.0006 | |
| Ba | 0.58 | 0.01 | 0.01 | 0.6 | 0.6 | 1 | 0.5 | 0.43 | | |
| Hg | 0.0004 | | | 0.001 | | | | 0.0001 | | |
| Pb | 0.02 | 0.007 | 0.009 | 0.13 | 0.2 | 0.06 | 0.01 | 0.013 | | |
| | Fly Ash | | Motor Vehicles | | | Industrial | | | | |
| Element | NBS | Gladney | Auto 1* | Auto 2 | Auto 3 | Cement | Paint | Steel | | |
| Al | 127 | 133 | | | | 24 | | 11 | | |
| Si | 210 | 208 | | | | 107 | | 7 | | |
| P | | | | 0.9 | | | | 1 | | |
| S | | | | 6 | | | | | | |
| Cl | 0.042 | | 0.1 | 19 | 68 | | | | | |
| K | 16.1 | 13.6 | | | | 5.3 | | | | |
| Ca | 47 | 10.6 | 0.054 | | | 460 | | 12 | | |
| Ti | 7.4 | 8.1 | | | | 1.4 | 422 | | | |
| V | 0.24 | 0.26 | | | | | | | | |
| Cr | 0.13 | 0.19 | | | | | | | | |
| Mn | 0.5 | 0.39 | | 0.6 | | | 4 | 31 | | |
| Fe | 62 | 123 | 0.05 | 1 | 4 | 11 | 204 | 570 | | |
| Ni | 0.16 | 0.0006 | | | | | | | | |
| Cu | | | 0.005 | | | | | | | |
| Zn | 0.2 | 0.23 | 0.01 | 3 | 1.4 | | | | | |
| Ga | | 0.078 | | | | | | | | |
| As | 0.058 | 0.16 | | | | | | | | |
| Se | 0.01 | 0.025 | 0.0005 | | | | | | | |
| Br | 0.012 | 0.006 | 0.38 | 37 | 79 | | | | | |
| Rb | 0.13 | 0.11 | | | | | | | | |
| Sr | 1.7 | 0.81 | | | | | | | | |
| Cd | | | 0.002 | | | | | | | |
| Sn | | | | | | | | | | |
| Sb | 0.007 | 0.006 | 0.0008 | | | | | | | |
| Ba | 2.7 | 1.14 | 0.013 | 0.9 | | | | | | |
| Hg | | | | | | | | | | |
| Pb | 0.075 | 0.056 | 1 | 148 | 400 | | | | | |

*Normalized to Pb = 1.0.

technology would be the most cost-effective approach.

- TTFA can resolve strong sources that are present in only a few of the samples, and a source-profile vector for the source can be predicted from the data. Other methods for performing quantitative source resolutions cannot account for such sources unless their presence is initially known and a measurement of the source composition is available.
- The presence of several samples with one very strong source was found to conceal the presence of other, weaker, factors in the full data set. Removal of the affected samples permitted the resolution of additional sources.
- Source profiles of coal-burning power plant emissions may be identified by TTFA. Distinguishing crustal material from fly ash would be more certain if sources were available.
- TTFA can resolve sources with dependence on the same element such as lead in particles from motor vehicle exhaust and lead and/or zinc smelters.
- The analysis of the July and August data from one RAPS station produced a reasonable source resolution. The data from one week at all RAPS stations seem to be composed from a number of sources that are significantly different at each station. The large number of sources contributing to the 10 RAPS stations precluded the resolution of any but the most important sources.
- By dividing both data sets into fine- and coarse-fraction subsets, additional sources were resolved, indicating that the sources contribute primarily to either the fine or coarse fractions, but not both.

Recommendations

With some additional development, target transformation factor analysis could be a valuable tool in the efforts of local authorities to devise cost-effective strategies for the control of airborne particulate matter. To ensure the credibility of any sort of statistical analysis, it is important that there be some indication of the reliability of the calculated results. In the present study, the actual nature of the sources was not known, so the accuracy of the quantitative results could not be determined. Testing, under

more controlled conditions, is essential to ensure that the calculated results are both accurate and reliable. One way to evaluate the accuracy of this method would be to analyze computer-generated data sets containing randomly generated errors that approximate typical analytical uncertainties and source-composition variability. Comparison of the starting data with the calculated results would give an indication of the reliability of the analysis. Another way to verify the quantitative results would be to analyze samples prepared in the lab from known amounts of well-characterized materials. Comparison of the calculated results with the known concentration of each sample would determine the quality of the source resolution. A third, and more complex, way to verify the quantitative results would be to compare the results of the factor analysis of real samples with results obtained by x-ray diffraction and/or microscopic examinations of the same samples.

Further development of target transformation factor analysis would include error models to give a better indication of the reliability of the calculated source profiles and to provide limits on the uncertainties of the calculated contributions of each factor. In addition, more objective and explicit criteria for determining the number of correct factors to be retained are needed. The analyses of computer-generated data sets could shed additional light on the still nebulous process of determining the number of factors.

Performing a source resolution of ambient aerosol samples requires a complete and accurate data set. If all the sources that contribute to a system are to be resolved, it is important that the data set include as many elements as possible. In the present study, the measured elements accounted for less than half of the total measured mass of each sample. The inclusion of other elements such as carbon and other gaseous species such as SO₂ would have greatly facilitated the analyses. In addition, the exclusion of the element arsenic because of the large number of determinations below the detection limits inhibited the resolution of a factor for coal fly ash. Future air quality studies should attempt to produce accurate data for both the total level of suspended particulate matter and as many elements and chemical species as practical.

This Project Summary was authored by Stephanie C. Davis of WAPORA, Inc., Cincinnati, OH 45233.

T. G. Dzubay and C. W. Lewis are the EPA Project Officers (see below).

The complete report, entitled "Modified Factor Analysis of Selected RAPS Aerosol Data," was authored by Daniel J. Alpert and Philip K. Hopke of the Institute for Environmental Studies and Nuclear Engineering Program, University of Illinois, Urbana, IL 61801.

The above report (Order No. PB 81-120 784; Cost: \$9.50, subject to change) will be available only from:

National Technical Information Service

5285 Port Royal Road

Springfield, VA 22161

Telephone: 703-487-4650

The EPA Project Officers can be contacted at:

Environmental Sciences Research Laboratory

U.S. Environmental Protection Agency

Research Triangle Park, NC 27711

United States
Environmental Protection
Agency

Center for Environmental Research
Information
Cincinnati OH 45268

Postage and
Fees Paid
Environmental
Protection
Agency
EPA 335



Official Business
Penalty for Private Use \$300

PS 0000329
U S ENVIR PROTECTION AGENCY
REGION 5 LIBRARY
230 S DEARBORN STREET
CHICAGO IL 60604