

SOURCE IDENTIFICATION AND APPORTIONMENT FOR ATMOSPHERIC AEROSOL BY FACTOR ANALYSIS

Wang Mingxing (王明星)

Institute of Atmospheric Physics, Academia Sinica, Beijing

Received July 11, 1984

ABSTRACT

The statistical basis for factor analysis is briefly discussed. A method combining common factor analysis with orthogonal transformation has been proposed and discussed in detail. The method has been tested with a data set consisting of 16 elemental concentrations in 40 samples collected at the Beijing meteorological tower, March 1980. Seven sources—soil+coal fly ash, sulfate, cement and lime dust, oil, plant burning, motor, and marine aerosol—corresponding to the first seven factors have been identified. The source profiles and their contributions have been obtained.

I. INTRODUCTION

Aerosol is the major pollutant in many cities, especially in the north part of this country where the climate is generally dry. In Beijing the concentration of total suspended particulate often exceeds the air pollution standard, and the control of particulate emissions is the most important measure for improving the environment. In order to devise economical and effective control strategies we need a better knowledge of the sources of aerosols, in addition to the knowledge of concentration, composition and their spacial and temporal variation of ambient aerosols. So far the distribution of air pollutants has been investigated, largely based on source-emission inventories, by dispersion modeling which is quite adequate for the calculation of the distribution of air pollutants emitted from a point-source. However, most of the sources, such as wind-blown dust, road dust and motor exhaust, can not be considered as point source. Furthermore, our present knowledge on atmospheric chemistry can not describe the transformation and transport of particles between source and receptor site precisely in a dispersion model. In addition, the source-emission inventories are barely accurate enough. Thus dispersion modeling is quite inadequate for the more sophisticated evaluation of aerosol problems. Over the past few years much effort has been directed toward developing new techniques, namely receptor-oriented models that start with the receptor and reconstruct the source contributions. Over the past ten years or so, there have been enormous improvements in our ability of collecting and analysing aerosol for characterizing the chemical composition of ambient aerosol. With the enormous increase in our knowledge of both particle size distribution and chemical composition of aerosol it is possible to develop a more useful model, the receptor model, for devising more effective control strategies. The receptor model is a technique that starts with measured properties of aerosol at the receptor site, including total mass and chemical composition, and constructs a statistical model to calculate contributions of various sources. This technique has the

advantages of being able to avoid the usage of source emission inventories and knowledge on physical and chemical processes between the sources and receptor site. Several forms of receptor models have been applied to the problem of aerosol source resolution, which may be grouped into two basic approaches: chemical element balance and factor analysis. The chemical element balance method is simple in mathematics and has obvious physical meaning but needs specific data on various sources concerned, which can only be obtained by extensive measurements for each source under specific conditions. The factor analysis does not need a priori knowledge of the sources while using a large data base of measured properties of aerosol. In this paper the statistical background for factor analysis is briefly discussed, while a method combining factor analysis with orthogonal transformation is discussed in more detail. The method is applied to the data set of elemental concentrations of aerosol collected in Beijing, March 1980. Source profiles and relative contributions to the elemental concentrations of each source are obtained.

II. STATISTICAL BACKGROUND

The basic principle in factor analysis is the usage of interrelationship that exists within the data. The goal of factor analysis is to simplify the quantitative description of a data set, thus reducing the original data matrix from one having M variables necessary to describe N samples to a matrix with p factors ($p < M$) for each of the N samples.

In the study of atmospheric aerosol one collects a large amount of samples (N), which may be analysed for tens of chemical components (M). Then we have a data set that can be properly described by N vectors in a M -dimensional space. However, the chemical components in a specific sample are interrelated with those from one particular source having higher degrees of correlation. Thus, the dimensions of the space used to characterize the data set can be reduced and p ($p < M$) factors can be used to describe the original data set.

The basic assumption is that the measured concentrations of elements in a particular sample are the sum of contributions from several sources,

$$C_x = C_{1x} + C_{2x} + C_{3x} + \dots + C_{px} \quad (1)$$

where C_x is the measured concentration of element x at the receptor site, C_{kx} ($k=1, 2, \dots, p$) is the contribution of source k .

The contribution of each source can be rewritten as a product of two factors

$$C_{kx} = a_{kx} f_k, \quad (2)$$

where a_{kx} is the relative concentration of element x in source k , and f_k is the contribution of source k to the measured mass of aerosol. For all the M elements and N samples we have

$$C_{ij} = a_{i1} f_{1j} + a_{i2} f_{2j} + \dots + a_{ip} f_{pj}, \quad (3)$$

$$i=1, 2, \dots, M, \quad j=1, 2, \dots, N,$$

where C_{ij} is the concentration of i th element in j th sample, a_{ik} the chemical character of source k , f_{kj} the contribution of source k to j th sample, and p the number of sources that have a significant contribution to the aerosol mass at the receptor site. In matrix notation we have

$$\mathbf{C} = \mathbf{A} \cdot \mathbf{F} \quad (4)$$

where \mathbf{C} is the $M \times N$ data matrix, \mathbf{A} the $M \times p$ factor loading matrix, and \mathbf{F} the $p \times N$ source contribution matrix. Factor analysis must permit determination of matrices \mathbf{A} and \mathbf{F} from \mathbf{C} with suitable statistical method. Empirical orthogonal expansion may be proved to be a good approach for this purpose. Suppose \mathbf{A}_1 is a set of p normalized orthogonal vectors,

the components of the projections of sample vectors on A_1 can be written as

$$F_i = A_1^t \cdot C. \quad (5)$$

Projecting the projected vector onto the original M -dimensional space, we obtain the expression in the original coordinates, which is denoted as C' for distinction. Then we have

$$C' = A_1 \cdot A_1^t \cdot C. \quad (6)$$

Now the problem is reduced to finding a suitable A_1 so as to replace C by C' with a minimum error. Then a linear transformation that produced a transformed matrix of A_1 , each row of which corresponds to a source profile, is performed. The problem is thus reduced to the calculation of eigenvectors of the covariance matrix. If A_1 consists of the p eigenvectors of the covariance matrix, which correspond to the first p eigenvalues, the mean error produced by replacing C with C' can be measured by δ_p^2 :

$$\delta_p^2 = \sum_{m=p+1}^M \lambda_m, \quad (7)$$

where λ_m stands for the eigenvalues of the covariance matrix, which are arranged in the monotonically decreasing order. Detailed discussion can be found in Ref. [4].

Given A_1 , source profile matrix can be found by a linear transformation as discussed further below.

III. APPLICATION OF FACTOR ANALYSIS TO AEROSOL DATA COLLECTED IN THE NORTH SUBURB OF BEIJING, MARCH 1980

A comprehensive aerosol study program was carried out at Xinglong Astronomical Observation Station, 6–21 March 1980. As a comparative study, aerosol samples were also collected at the Beijing meteorological tower using a time sequence filter sampler^[5]. The samples were analysed by particle induced X-ray emission for up to 16 elements. The meteorological tower is located at the conjunction of the suburb and city proper, where there are no major pollution sources. The aerosol samples collected may reflect the characteristics of aerosol in the area, among which those collected under northern air flow conditions may reflect the feature of the suburban air and those under southern air flow may represent the urban plume. The 84 samples, each representing 4 hr average, were examined based on the wind data provided by the meteorological tower. Two groups of samples were selected for the factor analysis: one with 17 samples (denoted as N) was collected when the wind direction was NE–NW and the wind velocity was in the range of 1–5 m/s, and the other with 23 samples (denoted as S) was obtained when the wind direction was generally southern with the velocity in the range of 1–5 m/s. The method presented in the last section was applied to these two groups of samples. Table 1 lists the first seven eigenvectors and the corresponding eigenvalues for data set N . As can be seen from the table, the first factor has high loadings for elements Al, Si, Ca, Fe and S and has significant loadings for most of the trace elements, which may be identified as soil plus coal smoke. The second factor has high loading only for S and can be identified as sulfate. The third has high loadings for Mg, Cl and Ca with highest value for Ca and may be attributed to cement and lime dust from construction sites. The fourth has highest loading for Cl and may represent the contribution of marine air. The fifth has significant loadings for Cl, Zn, Si, and Mg with highest values for Cl and Zn and may be attributed to the plant burning and incineration. The sixth factor has a high loading for Mg and a moderate loading for Zn. It is not clear what type of source this factor represents, although it may indicate an aged aerosol from oil burning, which

should have a high loading for V, in addition to high loadings for Mg and Zn. The explanation is that the concentration of V in Beijing aerosol is generally low and the measured concentrations for many of the selected samples were essentially zero. Moreover, particulate V may be easily lost during transport due to high volatility. Therefore the sixth factor may be identified as aged aerosol from oil burning that has transported a long way. The seventh factor has higher loading only for Pb and may represent the motor exhaust which should have a high loading also for Br. The lower loading for Br in this factor may also be explained by the higher volatility of particulate Br and the very low concentration of Br in the aerosol samples.

Table 1 Factor Loadings

Number	1	2	3	4	5	6	7
Eigenvalues	45.36	0.686	0.146	0.048	0.017	0.011	0.003
Mg	0.232	-0.118	0.676	-0.508	1.076	5.706	0.020
Al	2.761	-1.667	-2.278	2.894	-0.162	0.694	-0.107
Si	5.546	-1.439	0.606	-4.00	2.398	-0.106	-0.008
S	2.038	6.100	-1.490	0.067	-0.395	0.301	-0.567
Cl	0.644	0.617	2.182	5.470	6.692	-0.103	-0.157
K	0.665	0.405	0.846	0.650	1.216	-0.027	5.160
Ca	1.920	0.567	6.124	1.344	-3.466	0.073	-0.503
Ti	0.156	-0.024	-0.117	0.125	-0.040	0.105	0.059
V	0.007	0.014	-0.006	-0.099	-0.079	-0.007	0.261
Mn	0.042	-0.007	0.131	0.180	0.482	0.028	-0.117
Fe	1.115	-0.023	0.730	0.562	-0.524	-0.449	0.335
Cu	0.005	0.017	-0.037	-0.104	-0.053	-0.020	0.145
Zn	0.093	0.106	0.291	0.548	1.317	1.234	0.322
Br	0.009	0.007	-0.008	-0.048	0.022	0.062	-0.012
Sr	0.028	0.019	-0.051	-0.005	-0.037	-0.156	-0.074
Pb	0.069	0.203	0.027	0.270	-0.009	0.018	2.019

The eigenvectors were rotated to find the corresponding source profiles. This is achieved by an iterative procedure using a set of possible source profiles as test vectors. The rotated eigenvectors, the refined source profiles, are denoted as A and listed in Table 2. Given A , the transformed source contributions can be obtained and denoted as F . Denoting the original data set expressed in terms of the rotated eigenvectors as C_1 , we have

$$C_1 = AF = A_1 R R^{-1} F_1 = C', \quad (8)$$

where R is the rotation matrix. Thus the relative error produced when using C_1 to represent the original data set is given by^[4]

$$\varepsilon_p^2 = \sum_{m=p+1}^M \lambda_m / \sum_{m=1}^M \lambda_m, \quad (9)$$

which is the measure of the accuracy of the factor analysis. In the case of data set N of Beijing aerosol, $\varepsilon_p = 0.061$. In order to see clearly the meaning of the computation error in the factor analysis, we may compare C_1 calculated directly through matrices A and F with the original data. This is shown in Table 3 for the first sample in data set N of

Beijing aerosol. As can be seen, they are in good agreement for most of the elements, with a relative error below 10%. The mean relative error for the whole data set is 0.089, which is close to ϵ_p calculated by Eq. (9). Similar results are obtained for all other aerosol samples. This indicates that the elemental concentrations of aerosol at the Beijing meteorological tower may be well described by 7 factors. This means that the aerosol has 7 significant sources.

In order to quantitatively describe the contributions of individual sources to the measured elemental concentrations, we have calculated the apportionments of elemental concentrations among the sources for each of the samples. Listed in Table 4 are the mean results for data set *N*. The last column in Table 4 gives the ratios of the calculated concentration to the observed. It can be seen that the ratios are very close to 1 for most of the elements except those with very low concentrations and large measurement errors. This indicates that the apportionment is successful.

The negative number in the factor loading matrix indicates that the corresponding contribution of that source to the elemental concentration is generally lower than the background value. These negative values produce negative values in the apportionments, which imply that the corresponding contributions are essentially zero. These were marked with bar in the table.

Similar results are obtained for data set *S*, but are not presented here. In order to clearly show the contributions of each source for urban and suburban air in Beijing, the percentage contributions of the sources for both data sets *N* and *S* are calculated, with the mean results listed in Tables 5 and 6. It can be seen that the most abundant elements

Table 2 Transformed Factor Loadings (Source Profiles)

Factor No.	1	2	3	4	5	6	7
Source	Soil+Coal Ash	Sulfate	Cement Lime Dust	Marine	Incine- rator	Oil	Motor
Mg	61.53	-2.16	273.68	33.80	7.42	6.96	-0.36
Al	2698.58	11.62	1034.34	1.11	4.14	1.74	0.21
Si	6938.42	21.70	697.60	-2.90	2.21	-0.147	0.04
S	396.98	2669.69	-0.16	0.20	0.09	0.02	-
Cl	11.48	12.78	-14.25	589.96	134.89	29.80	0.18
K	485.11	18.82	263.90	10.36	16.19	59.01	0.11
Ca	250.40	0.85	2122.19	13.83	92.06	1.47	0.19
Ti	152.20	16.39	50.20	1.57	2.46	-0.36	0.39
V	16.01	0.67	6.18	-11.71	-1.72	1.69	0.08
Mn	55.18	-19.55	-21.60	34.47	7.97	0.71	0.18
Fe	1096.36	7.34	298.84	13.96	16.72	5.09	-0.25
Cu	19.29	7.81	-7.67	-9.58	-2.41	0.91	-0.02
Zn	19.48	5.28	2.27	81.81	18.86	9.45	0.36
Br	15.80	5.46	-5.12	-0.79	-0.58	-0.11	-0.02
Sr	36.52	13.79	-7.47	-0.77	-0.04	-1.26	0.08
Pb	8.80	4.10	89.70	-0.17	10.19	14.72	1.89

Table 3 Comparison between Calculated and Observed Elemental Concentrations

Element	Concentrations ng/m ³		Relative Error
	Calculated	Observed	
Mg	158.13	160.55	0.015
Al	1668.29	1667.85	0.0003
Si	3540.44	3539.92	0.0001
S	2861.48	2861.43	0
Cl	637.73	639.02	0.002
K	667.17	666.79	0.0006
Ca	1451.85	1451.08	0.0005
Ti	103.95	125.51	0.1718
V	10.06	10.79	0.0676
Mn	24.43	26.07	0.0629
Fe	782.63	813.66	0.0381
Cu	9.05	11.84	0.2356
Zn	102.88	58.11	0.4351
Br	7.87	8.90	0.1157
Sr	21.08	15.23	0.2775
Pb	102.37	102.26	0.0001

Table 4 Source Apportionment for Data Set N

Element	Contributions from Components (ng/m ³)							Total Concentrations (Cal.)	Cal./Obs.
	Coal + Soil	Sulfate	Cement + Lime	Marine	Incinerator	Oil	Motor		
Mg	61.9	—	262.6	22.9	20.0	15.0	—	382.4	1.17
Al	2715.5	12.2	992.6	0.8	11.2	3.8	4.0	3740.1	1.00
Si	6981.9	22.8	669.4	0	6.0	0	0.7	7680.8	1.00
S	399.5	2807.3	0	0	0	0	0	3206.8	1.00
Cl	11.5	13.4	—	399.6	364.2	64.4	3.4	856.5	1.09
K	488.1	19.8	253.2	7.0	43.7	127.5	2.2	941.5	1.01
Ca	252.0	0.9	2036.5	9.4	248.6	3.2	3.6	2554.2	1.00
Ti	153.1	17.2	48.2	1.1	6.6	—	7.3	233.5	1.08
V	16.1	0.7	5.9	—	—	3.6	1.5	27.8	2.11
Mn	55.5	—	—	23.3	21.5	1.5	3.5	105.3	2.04
Fe	1103.2	7.7	286.8	9.5	45.2	11.0	0	1463.4	0.97
Cu	19.4	8.2	—	—	—	2.0	—	29.5	2.76
Zn	19.6	5.5	2.2	55.4	50.9	20.4	6.8	160.8	1.31
Br	15.9	5.7	—	—	—	—	—	21.6	1.52
Sr	36.8	14.4	—	—	—	—	1.6	52.8	1.30
Pb	8.6	4.3	86.1	—	27.5	31.8	36.0	194.5	1.86

in Beijing aerosol, i. e. Al, Si, Fe and Ti, mainly come from soil dust and coal fly ash, while Ca mainly comes from cement and lime dust, which is the explanation that Ca in Beijing aerosol is largely enriched with respect to average earth crust. The sources of Pb in Beijing aerosol are somewhat striking. The largest source is cement and lime dust (accounts for more than 40% of the total), while motor contributes only a small portion. It should be pointed out here that the concentrations of Pb in Beijing aerosol are generally low, and were close to PIXE detection limit for many samples collected in March 1980. Therefore the elemental concentration of Pb has larger error, and the source apportionment is not very reliable.

IV. TOTAL MASS CONCENTRATION OF AEROSOL

If the total mass concentrations were known for all the samples used in factor analysis, the factor analysis should have been able to produce the mass contributions of each source to each sample. Then, Eq. (3) can be rewritten as

$$C_{ij} = \sum_{k=1}^p a_{ik} f_{kj} = \sum_{k=1}^p \frac{a_{ik}}{b_k} (b_k f_{kj}), \quad (10)$$

where a_{ik}/b_k represents the relative concentration of element i in k th source, and $b_k f_{kj}$ stands for the mass contribution of that source to j th sample. Thus the total mass of j th sample is given by

$$M_j = \sum_{k=1}^p b_k f_{kj}. \quad (11)$$

Therefore, by use of the matrix F obtained in section III and the measured masses of all samples used, b_k can be calculated with a least-square procedure. The mass concentrations

Table 5 Percentage Contributions of Sources for Data Set N

Element	Percentage Contributions (%)						
	Soil + Coal	Sulfate	Cement + Lime	Marine	Incinerator	Oil	Motor
Mg	16.2	0	68.7	6.0	5.2	3.9	0
Al	72.6	0.3	26.5	0	0.3	0.1	0.1
Si	90.9	0.3	8.7	0	0.1	0	0
S	12.5	87.5	0	0	0	0	0
Cl	1.3	1.6	46.7	42.5	7.5	0.4	0
K	50.8	2.1	26.9	0.7	4.6	13.5	0.2
Ca	9.9	0	79.7	0.4	9.7	0.1	0.1
Ti	65.6	7.4	20.6	0.5	2.8	—	3.1
V	57.9	2.5	21.2	—	—	12.9	5.4
Mn	52.7	—	—	22.1	20.4	1.4	3.3
Fe	75.4	0.5	19.6	0.6	3.1	0.8	0
Cu	65.8	27.8	—	—	—	6.8	—
Zn	12.2	3.4	1.4	34.4	31.7	12.7	4.2
Br	73.6	26.4	—	—	—	—	—
Sr	69.7	27.2	—	—	—	—	3.0
Pb	4.5	2.2	44.3	—	14.1	16.3	18.5

of each source to all samples can be calculated by the use of Eq. (11). Unfortunately, the total mass of aerosol sample used in this study was not measured, and thus the mass

Table 6 Percentage Contributions of Sources for Data Set S

Element	Percentage Contributions (%)						
	Soil + Coal	Sulfate	Cement + Lime	Marine	Incinerator	Oil	Motor
Mg	27.7	0.5	40.7	21.1	3.7	2.3	3.9
Al	59.5	0.2	39.7	0	0.1	0.5	0
Si	83.9	0.3	15.5	0	0	0	0.2
S	10.6	89.3	0	0	0	0	0.1
Cl	0.5	0.5	—	83.6	6.3	9.0	0
K	30.2	1.5	38.4	2.6	0.3	26.9	0
Ca	5.5	0	90.9	0.9	2.2	0.4	0
Ti	35.5	20.6	40.6	3.2	0	0	0
V	2.2	42.1	20.2	25.7	2.7	—	7.1
Mn	43.3	—	—	16.5	—	40.2	—
Fe	82.6	1.6	1.4	0.4	1.9	0	12.1
Cu	53.9	0.7	—	—	—	29.4	16.0
Zn	—	15.7	43.3	37.4	3.2	—	—
Br	—	55.9	17.7	24.7	1.6	—	—
Sr	41.6	18.8	—	13.2	2.6	—	23.8
Pb	—	24.2	49.8	—	10.4	15.5	—

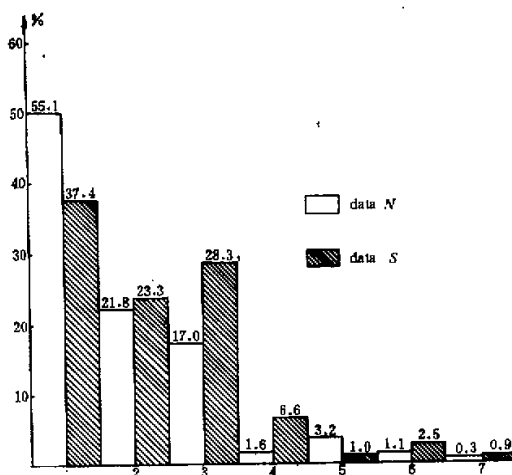


Fig. 1. Mass contribution of various sources to total aerosol mass. The numbers on the abscissa correspond to the sources indicated in Table 2.

concentrations can not be obtained directly. Alternatively, the mass contributions may be estimated from Table 4 according to the relative abundances of some specific elements in each source^[3]. The results for both data sets are shown in Fig. 1.

It is clearly shown in the figure that the major sources of Beijing aerosol are soil + coal fly ash, sulfate, and cement and lime dusts. They account for more than 90% of the total mass. As can be seen from the figure, the main difference between the two data sets is the mass contribution of soil + coal fly ash, which is significantly larger for data set N. This is logical, as the suburban air carries more soil dust while the urban air has more pollutants. In addition, the mass contribution of marine air is larger for data set S due to the southeast wind bringing such air to the sampling site.

V. DISCUSSIONS

The factor analysis for the aerosol samples collected at the north suburb of Beijing, March 1980 shows that this is a useful method for the identification and apportionment of sources of urban aerosol.

For the elements used in this study, the soil dust and coal fly ash have so similar compositions that the factor analysis has not been able to resolve these two sources. To resolve soil dust and coal fly ash is of great importance in practice. To do this, it is necessary to have data on those components, the relative abundances of which differ greatly in soil dust and coal fly ash. These components can be element As and polycyclic aromatic hydrocarbons. However, these components have very low concentrations in aerosol samples, and should be measured with great care and probably need to be weighted in the factor analysis. This is one of the difficulties in factor analysis. In addition, it is necessary to measure the total mass of samples in order to determine the mass contribution of each source.

In Beijing, the major sources of aerosol are soil dust + coal fly ash, sulfate, and cement and lime dust. They account for more than 90% of the total aerosol mass. It should be noted that the absolute concentration of sulfate in Beijing is much higher than that in European and U. S. cities, although its relative contribution to total mass is much lower. The total mass concentration of aerosol in Beijing is 5 times as high as that in New York City, and may be 10 times as high as in winter.

The valuable support and encouragement provided by Prof. Zeng Qingcun is gratefully acknowledged.

REFERENCES

- [1] Gordon G. E., *Ann. of N.Y.A.S.*, **338**(1980), 93—105.
- [2] 王明星, 科学探索, **3** (1983), 2: 13—20.
- [3] Hopke P. K. *ACS Symposium Series*, **167**(1981), 21—49.
- [4] 曾庆存, 大气红外遥测原理, 科学出版社, 1974, 158—166.
- [5] 王明星等, 环境科学丛刊, **2**(1981), 1: 1—10.
- [6] 任丽新等, 大气科学, **6**(1982), 1: 11—17.
- [7] Winchester J. W. et. al., *Nuclear Instr. and Meth.*, **181**(1981), 3: 391—398.
- [8] 王明星等, 大气科学, **5**(1981), 2: 136—144.