数理解析研究所講究録 1253 巻 2002 年 87-94

Recent development of estimating source apportionment in air pollution

広島県立保健福祉大学・総合教育センター　佐藤学 (SATO, Manabu)
Department of Human Sciences
Hiroshima Prefectural College of Health Sciences

1 Concept of estimating source apportionment

The concept of *source apportionment* is identifying the sources of air pollution and estimating the apportionments of the concentrations of the pollutants observed at the sites in the environment. The results obtained can be used to help manage effectively the quality of the environment.

2 Receptor-oriented model

For estimating source apportionment, several receptor-oriented models of chemical mass balance (CMB) and multivariate statistical methods (factor analysis, principal component analysis and so on) have been widely used. See for example, Hopke(1985, 1991), Heinsohn and Kabel(1998), Dey and Schnelle(1999), Park et al.(2002). These models are substantially based on the assumption of mass conservation and a mass balance analysis.

Suppose that there are n observations $(x_{i1}, \ldots, x_{ip})$, $i = 1, \ldots, n$, of p elements in material where x_{ij} is the concentration (μg m^{-3}) of the j-th component in the i-th sample. If there are m ($< p$) possible sources, then *the general receptor-oriented model* can be expressed as

$$x_{ij} = \sum_{k=1}^{m} a_{jk} g_{ki}, \quad i = 1, \ldots, n, \; j = 1, \ldots, p, \tag{1}$$

where a_{jk} is the mass concentration (μg g^{-1}) of the j-th element in material from the k-th source, and g_{ki} is the volume concentration (g m^{-3}) of suspended particulate matter (SPM) from the k-th source in the i-th receptor. In studies of air pollution, the number m of important sources that affect the air quality is usually small and profiles a_{jk} of

the potential sources are available. The following assumptions (i) to (iii) are required to construct the model:

(i) Emission components do not convert chemically.

(ii) Emission components are not removed selectively.

(iii) Source emissions have a constant composition.

3 Necessity of constructing a modified model

Although the receptor models are potentially useful tools for demonstrating the determinants of exposure to air pollutants, they have several disadvantages (Henry et al.(1984), Henry(1987), Sexton and Hayward(1987)). Few explicit ways of estimation based on the classical model were offered for treating random fluctuations in observations (Wiens et al.(2001)). Further, the validity of receptor modeling is mostly dependent on the sufficiency of the source profiles used in the analysis. Therefore many efforts to update and revise the database of source emission profiles have been made by some organizations. For example, the United States Environmental Protection Agency has been developed "Receptor Model Source Composition Library" (Shareef and Bravo(1988)). In the real world, however, source profiles should depend on many varying factors such as wind, temperature and distance from sources.

Indoor pollutant concentrations have a large contribution to total human exposures to airborne particles because urban residents spend a great portion of their life indoors. Thus, it is of interest to estimate the relative contributions of outdoor sources to indoor air quality (Koutrakis et al.(1992), Moffat(1997)). In analysis of indoor data, we should pay careful attention to a treatment of tobacco effects. The profiles of tobacco should have sharp fluctuations because they depend on various causes such as ventilation, puffing, and kinds of cigarettes.

4 A modified model based on source profiles with fluctuations

Taking into acount fluctuations in concentration and source profiles, Ohtaki et al.(1997) modified a model and developed a method which can estimate the apportionments. Model (1) is extended as follows:

$$x_{ij} = \sum_{k=1}^{m}(a_{jk} + r_{jk}^{(i)})(\lambda_k + \tau_{ki})y_i, \quad i = 1, \ldots, n, \quad j = 1, \ldots, p, \tag{2}$$

where y_i denotes the mass concentration (μg m^{-3}), $r_{jk}^{(i)}$ and τ_{ki} denote fluctuations of the source profile and of the source apportionment at the i-th receptor, respectively. We assume here that $r_{jk}^{(i)}$ is a random variable with mean 0, which is independent of $\boldsymbol{\tau}_i = (\tau_{1i}, \ldots, \tau_{mi})'$, and that $\boldsymbol{\tau}_i$ $(i = 1, \ldots, n)$ are mutually independent and identically distributed random vectors with mean $\mathbf{0}$. The expression $'$ (prime) denotes the transposition of a vector or a matrix in this text. Our aim is to estimate the mean vector of source apportionment $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)'$ under the natural two constraints such that

$$C_1 : \sum_{k=1}^{m} \lambda_k = 1 \quad \text{and} \quad C_2 : \lambda_k \geq 0, \ k = 1, \ldots, m.$$

5 Estimation of source apportionment for the modified model

Model (2) can be rewritten to a usual regression model. Setting

$$z_{ij} = x_{ij}/y_i, \quad i = 1, \ldots, n, \ j = 1, \ldots, p,$$

and $\mathbf{z}_i = (z_{i1}, \ldots, z_{ip})'$, $i = 1, \ldots, n$, we can express Model (2) as

$$\mathbf{z}_i = (A + R_i)(\boldsymbol{\lambda} + \boldsymbol{\tau}_i), \qquad i = 1, \ldots, n, \tag{3}$$

where

$$A = (a_{jk}), \quad j = 1, \ldots, p, \ k = 1, \ldots, m,$$

and

$$R_i = (\mathbf{r}_1^{(i)}, \ldots, \mathbf{r}_m^{(i)}), \quad \mathbf{r}_k^{(i)} = (r_{1k}^{(i)}, \ldots, r_{pk}^{(i)})', \quad i = 1, \ldots, n, \ k = 1, \ldots, m.$$

Further, if we set $\bar{z}_{\cdot j} = \sum_{i=1}^{n} z_{ij}/n$, $j = 1, \ldots, p$,

$$V = \mathrm{diag}(v_1, \ldots, v_p), \quad v_j = \sum_{i=1}^{n} (z_{ij} - \bar{z}_{\cdot j})^2/(n-1), \quad j = 1, \ldots, p,$$

and

$$\mathbf{u}_i = (A'V^{-1}A)^{-1}A'V^{-1}\mathbf{z}_i, \qquad i = 1, \ldots, n,$$

then, Formula (3) is transformed into

$$\mathbf{u}_i = \boldsymbol{\lambda} + \boldsymbol{\varepsilon}_i, \qquad i = 1, \ldots, n, \tag{4}$$

where

$$\boldsymbol{\varepsilon}_i = (A'V^{-1}A)^{-1}A'V^{-1}R_i\boldsymbol{\lambda} + [I + (A'V^{-1}A)^{-1}A'V^{-1}R_i]\boldsymbol{\tau}_i, \quad i = 1, \ldots, n,$$

are mutually independent with $E\{\boldsymbol{\varepsilon}_i\} = \mathbf{0}$, and $\mathrm{var}\{\boldsymbol{\varepsilon}_i\} = \Omega$, say, $i = 1, \ldots, n$. Letting $\mathbf{u} = (\mathbf{u}_1', \ldots, \mathbf{u}_n')'$ and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', \ldots, \boldsymbol{\varepsilon}_n')'$, we obtain from (4) a usual regression model

$$\mathbf{u} = (\mathbf{1}_n \otimes I_m)\boldsymbol{\lambda} + \boldsymbol{\varepsilon},$$

where $\mathbf{1}_n = (\underbrace{1, \ldots, 1}_{n})'$ and I_m means the m-order unit matrix and the symbol $\otimes$ represents Kronecker product of matrices.

We consider an estimator of $\boldsymbol{\lambda}$ under Constraints C_1 and C_2. First, we assume that $\Omega = (\omega_{k\ell})$ is known. Using the general theory of statistical estimation, we obtain *the generalized least squares estimator* $\hat{\boldsymbol{\lambda}}(\Omega)$ of $\boldsymbol{\lambda}$ under Constraint C_1 as follows:

$$\hat{\boldsymbol{\lambda}}(\Omega) = \bar{\mathbf{u}} + \left(\sum_{k=1}^{m} \bar{u}_k - 1\right) \omega_{\cdot\cdot}^{-1}(\omega_{1\cdot}, \ldots, \omega_{m\cdot})', \tag{5}$$

where $\bar{\mathbf{u}} = (\mathbf{u}_1 + \cdots + \mathbf{u}_n)/n = (\bar{u}_1, \ldots, \bar{u}_m)'$, $\omega_{\cdot\cdot} = \sum_{k,\ell=1}^{m} \omega_{k\ell}$, and $\omega_{k\cdot} = \sum_{\ell=1}^{m} \omega_{k\ell}$, $k = 1, \ldots, m$. Since Ω is usually unknown in practice, we substitute the best linear unbiased estimator

$$\hat{\Omega} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{u}_i - \bar{\mathbf{u}})(\mathbf{u}_i - \bar{\mathbf{u}})' = (\hat{\omega}_{k\ell})$$

for Ω in (5). Thus we have a provisional estimator $\hat{\boldsymbol{\lambda}}(\hat{\Omega})$ of $\boldsymbol{\lambda}$ as

$$\hat{\boldsymbol{\lambda}}(\hat{\Omega}) = \bar{\mathbf{u}} + \left(\sum_{k=1}^{m} \bar{u}_k - 1\right) \hat{\omega}_{\cdot\cdot}^{-1}(\hat{\omega}_{1\cdot}, \ldots, \hat{\omega}_{m\cdot})',$$

where $\hat{\omega}_{\cdot\cdot} = \sum_{k,\ell=1}^{m} \hat{\omega}_{k\ell}$ and $\hat{\omega}_{k\cdot} = \sum_{\ell=1}^{m} \hat{\omega}_{k\ell}$, $k = 1, \ldots, m$. If $\hat{\boldsymbol{\lambda}}(\hat{\Omega}) = (\hat{\lambda}_1(\hat{\Omega}), \ldots, \hat{\lambda}_m(\hat{\Omega}))'$ satisfies Constraint C_2, that is, $\hat{\lambda}_k(\hat{\Omega}) \geq 0$ for all k's, then we adopt it as a proper solution. Otherwise, reduce the model by excluding the sources such that $\hat{\lambda}_k(\hat{\Omega}) < 0$, and retry to calculate $\hat{\boldsymbol{\lambda}}(\hat{\Omega})$ with using the reduced model. Iterate this procedure if necessary.

As for property of the estimator, $\hat{\boldsymbol{\lambda}}(\Omega)$ is an unbiased estimator. Its variance-covariance matrix $\mathrm{var}\{\hat{\boldsymbol{\lambda}}(\Omega)\}$ can be expressed as follows:

$$\mathrm{var}\{\hat{\boldsymbol{\lambda}}(\Omega)\} = \left(\Omega - \omega_{\cdot\cdot}^{-1}(\omega_{1\cdot}, \ldots, \omega_{m\cdot})'(\omega_{1\cdot}, \ldots, \omega_{m\cdot})\right)/n.$$

Thus, we may evaluate the variation of the estimates.

Figure I shows that a flow chart of the procedure for estimating source based on source profiles with fluctuations.

6 An example with a Real Data Set

An example with a real data set demonstrates this procedure owing to Ohtaki et al.(1997). Fine fraction samples were collected in living rooms of houses within 150 meters of the roadside in the Metropolitan Tokyo, Japan. Each house was monitored for

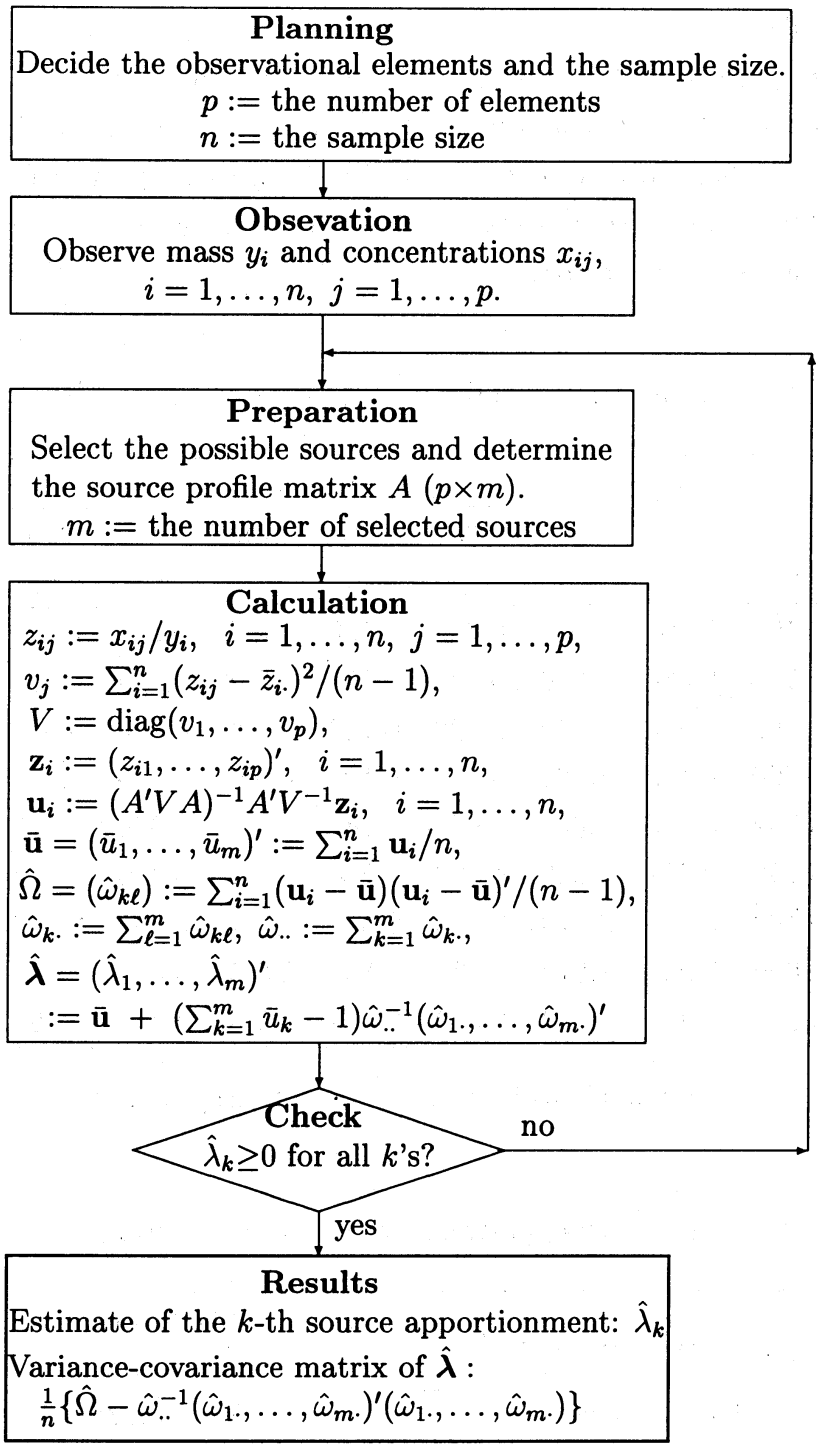

Figure I. Flow chart of the procedure for estimating source apportionments based on source profiles with fluctuations

consecutive four weekdays during July. In this study, we used the data of the concentrations of ten elements. The technical details of handling the original data are described in Nitta et al.(1994).

Table I represents the observed mass y_i and element concentrations x_{ij} ($i = 1, \ldots, 9$, $j = 1, \ldots, 10$). Referring to the previous studies (see Nitta et al.(1994)), we selected the seven possible sources and the source profile matrix $A = (a_{jk})$ was compiled from several references; See Ohtaki et al.(1997) in detail. Table II represents the estimates $\hat{\boldsymbol{\lambda}}$ of source apportionments and standard errors.

Table I. Observed mass y_i and elemental concentration x_{ij} [μg m^{-3}] ($p = 10$, $n = 9$)

ID	mass	element									
		Al	Si	S	K	Ca	Mn	Fe	Cu	Zn	Br
i	y_i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}	x_{i6}	x_{i7}	x_{i8}	x_{i9}	x_{i10}
1	31.870	0.101	0.405	0.991	0.398	0.180	0.029	0.265	0.026	0.251	0.019
2	18.870	0.143	0.388	2.040	0.288	0.253	0.044	0.484	0.025	0.130	0.000
3	11.740	0.095	0.236	0.779	0.396	0.159	0.000	0.170	0.026	0.124	0.000
4	92.980	0.000	0.152	1.329	0.508	0.120	0.020	0.310	0.036	0.291	0.137
5	19.340	0.232	0.286	2.032	0.647	0.144	0.015	0.359	0.013	0.139	0.000
6	63.140	0.156	0.352	0.829	0.588	0.206	0.009	0.319	0.017	0.242	0.023
7	53.530	0.182	0.544	2.770	0.365	0.302	0.023	0.657	0.000	0.265	0.026
8	26.590	0.000	0.247	1.870	0.339	0.187	0.051	0.522	0.035	0.487	0.128
9	19.100	0.137	0.362	1.280	0.389	0.294	0.021	0.394	0.023	0.210	0.032

ID: The identificaiton number of receptors.

Table II. Estimates $\hat{\boldsymbol{\lambda}}$ for apportionment and their standard errors (s.e.)

sources ($m = 7$)	estimates	
	$\hat{\boldsymbol{\lambda}}$	s.e.
tobacco smoke	0.317	0.124
soil	0.016	0.007
fuel combustion (heavy oil)	0.468	0.098
steel works	0.060	0.012
incinerators (waste)	0.056	0.011
gasoline-powered automobile	0.023	0.006
diesel-powered automobile	0.060	0.024

Acknowledgements

The author would like to thank Professors Umegaki and Kawabe for encouragement. ıis research was partially supported by Japan Society for the Promotion of Science, :ant-in-Aid for Scientific Research under contract numbers 11640121 and 12308011.

References

Dey, P. R. and Schnelle, K. B. Jr. (1999). *Atmospheric Dispersion Modeling Compliance Guide,* McGraw Hill.

Heinsohn, R. J. and Kabel, R. L. (1998). *Sources and Control of Air Pollution,* Prentice Hall.

Henry, R. C. (1987). Current factor analysis receptor models are ill-posed. *Atmospheric Environment*, **21**, 1203–1217.

Henry, R. C., Lewis, C. W., Hopke, P. K. and Williamson, H. J. (1984). Review of receptor models fundamentals. *Atmospheric Environment*, **18**, 1507–1515.

Hopke, P. K. (1985). *Receptor Modeling in Environmental Chemistry,* Wiley.

Hopke, P. K. (1991). *Receptor Modeling for Air Quality Management (Data Handling in Science and Technology, Vol. 7)*, Elsevier Science.

Koutrakis, P., Briggs S. L. K. and Leaderer B. P. (1992). Source apportionment of indoor aerosols in Suffolk and Onondaga counties, New York. *Environmental Science and Technology*, **26**, 521–527.

Moffat, D. (1997). *Handbook of Indoor Air Quality Management,* Simon & Schuster.

Nitta, H., Ichikawa, M., Sato, M., Konishi, S. and Ono, M. (1994). A new approach based on a covariance structure model to source apportionment of indoor fine particles in Tokyo. *Atmospheric Environment*, **28**, 631–636.

Ohtaki, M., Sato, M. and Nitta H. (1997). Estimating source apportionment of particulate matters based on source profiles with fluctuations. *Environmetrics*, **8**, 341–350.

Park, E. S., Spiegelman, C. H. and Henry, R. C. (2002). Bilinear estimation of pollution source profiles and amounts by using multivariate receptor models. *Environmetrics*, in press.

Sato, M. (2002). Importance of identification of the model: some comments on the paper "Bilinear estimation of pollution source profiles and amounts by using multivariate receptor models." *Environmetrics*, in press.

Sato, M. and Ohtaki, M. (2001). Source apportionment. An entry of *The Encyclopedia of Environmetrics*, Wiley, in press.

Sexton, K. and Harward, S. B. (1987). Source apportionment of indoor air pollution. *Atmospheric Environment*, **21**, 407–418.

Shareef, G. S. and Bravo, L. A. (1988). *Air Emissions Species Manual. Vol. 2, Particulate Matter Species Profiles.* EPA-450/2-88-003b.

Wiens, D., Florence, L. Z. and Hiltz, M. (2001). Robust estimation of chemical profiles of air-borne particulate matter. *Environmetrics*, **12**, 25–40.