Supporting Information for:

Mode and place of origin of carbonaceous aerosols transported from East Asia to Cape Hedo, Okinawa Japan

Kojiro Shimada[1,2*], Masamichi Shimada[1], Akinori Takami[3], Shuichi Hasegawa[4], Akihiro Fushimi[3], Takemitsu Arakaki[5], Izumi Wanatabe[1] Shiro Hatakeyama[1]

[1] Graduate School of Agriculture, Tokyo University of Agriculture and Technology, 3-5-8 Saiwaicho, Fuchu, Tokyo 183-8509, Japan

[2] Tokyo Metropolitan Research Institute for Environmental Protection, 1-7-5 Shinsuna, Kouto, Tokyo 136-0075, Japan

[3] National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki 305-8506, Japan

[4] Center for Environmental Science in Saitama, 914 Kamitanadare, Kazo, Saitama 347-0115, Japan

[5] University of the Ryukyus, 1 Senbaru, Nishihara-cho, Okinawa 903-0213, Japan

The Supporting Information contains 12 Figures and 1 Table.

Supplementary Figures and Table: Choosing the optimal number of factors for the model is challenging. To determine the optimal number of factors, it is necessary first to determine the minimum Q values for different numbers of factors. Different Q functions can be defined that $Q_{\text{(true)}}$ is the goodness of fit parameter calculated including all points. $Q_{\text{(robust)}}$ is the goodness of fit parameter calculated excluding points not fit by the model, defined as samples for which the uncertainty scaled residual is greater than 4. We basically followed the method presented by Norris *et al.* (2008), although several different methods have been proposed (Bhanuprasad *et al.*, 2008; Kitayama *et al.*, 2010; Cherian *et al.*, 2010; Norris *et al.*, 2008). In this study, we first investigated minimum Q values (Fig. S1(a)).

Fig. S1(a) shows that the slope of the relationship between the minimum Q value and the number of factors began to level off at five factors. The reduction in Q with the increase in the number of factors and the agreement of estimated Q with its theoretical value, Q_{theo}, were used to identify possible optimal solutions. However, with this method, the results for 4, 5, and 6 factors are similar.

To identify the minimum number of factors needed for a well-constrained solution, we examined the maximum individual column mean (IM) and the maximum individual column standard deviation (IS) (Fig. S1(b)). These scaled residual matrix values dropped sharply for solutions with 5 or more factors. Therefore, we adopted 5 as the optimal number of factors. In, We then evaluated the resultant PMF calculations (Figs. S2 and S3 and Table S1). We removed several chemical components with poor correlation factors (R^2) because the model results for these chemical components seemed unreasonable. In these cases, the number of data were insufficient, or the uncertainties were high. Table S1 also shows the number of missing values and of values below the detection limit. The PMF model does not allow values below the detection limit or missing values to be implemented. Instead, they are replaced by "virtual" values having larger uncertainties in order to lower their influence on the PMF modeling result (Polissar *et al.*, 1998; Reff *et al.*, 2007). Some metallic components had values that were below the detection limit in this study; these values were replaced with half the detection limit value, and their estimated errors were evaluated as 5/6 of the detection limit (Norris *et al.*, 2008). In the case of missing data, they are applied geometric or arithmetic values (Bhanuprasad *et al.*, 2008; Cherian *et al.*, 2010). In this study, missing values were replaced with the median value for that species, and the uncertainty was set at 4 times the species-specific median, as suggested by Norris *et al.* (2008). Some OC and EC measurement data were defective in both summer (18 and 20

July) and winter (10, 11, and 13 January). These data, because they were only "virtual," were excluded from the analysis results presented in Fig. 6. The measurement uncertainties (i.e., the method detection limits) were determined following the procedure of Norris *et al.* (2008). In this study, these values were low. Bhanuprasad *et al.* (2008) used one standard deviation of the blank measurement values as the MDL data set.

To evaluate the uncertainties, we examined the interquartile ranges of the contributions of the 5 factor contributions and the concentrations of key species (Fig. S2 and S3). These ranges were around 20% or less. Bhanuprasad *et al.* (2008), Mehta *et al.* (2009), and Cherian *et al.* (2010) reported that the estimated uncertainty for the contribution of each factor was < ~15%, and Norris et al. (2008) evaluated the estimated uncertainty to be around 30%.

REFERENCES

Bhanuprasad, S. G., Venkataraman, C. and Bhushan, M. (2008). Positive matrix factorization and trajectory modelling for source identification: A new look at Indian Ocean Experiment ship observations. *Atmos. Environ*. 42: 4836–4852, doi:10.1016/j.atmosenv.2008.02.041.

Cherian, R., Venkataraman, C., Kumar, A., Sarin, M. M., Sudheer, K. A. and Ramachandran, S. (2010). Source identification of aerosols influencing atmospheric extinction: Integrating PMF and PSCF with emission inventories and satellite observations. *J. Geophys. Res.* 115: D22212, doi:10.1029/2009/JD012975.

Mehta, B., Venkataraman, C., Bhushan, M. and Tripathi, N. S. (2009). Identification of sources affecting fog formation using receptor modeling approaches and inventory estimates of sectoral emissions. *Atmos. Environ.* 43: 1288–1295, doi:10.1016/j.atmosenv.2008.11.041.

Norris, G., Vedantham, R., Wade, K., Brown, S., Prouty, J., Foley, C. and Martin, L. (2008). EPA Positive Matrix Factorization (PMF) 3.0 Fundamentals & User Guide. E*nvironmental Protection Agency (EPA)*. 600/R-08/108.

Kitayama, K., Murao, N. and Hara, H. (2010). PMF analysis of impacts of SO_2 from Miyakejima and Asian Continent on precipitation sulfate in Japan. *Atmos.*

Environ. 44: 95-105.

Polissar, A., Hopke, P., Paatero, P. (1998). Atmospheric aerosol over Alaska - 2. Elemental composition and sources, *J. Geophys. Res.*, 103(D15), 19045–19057, doi:10.1029/98JD01212.

Reff, A., Eberly, S. I. and Bhave, P. V. (2007). Receptor modeling of ambient particulate matter data using positive matrix factorization: review of existing methods, *J. Air. Waste. Manage. Assoc.,* 57(2), 146, 2007.

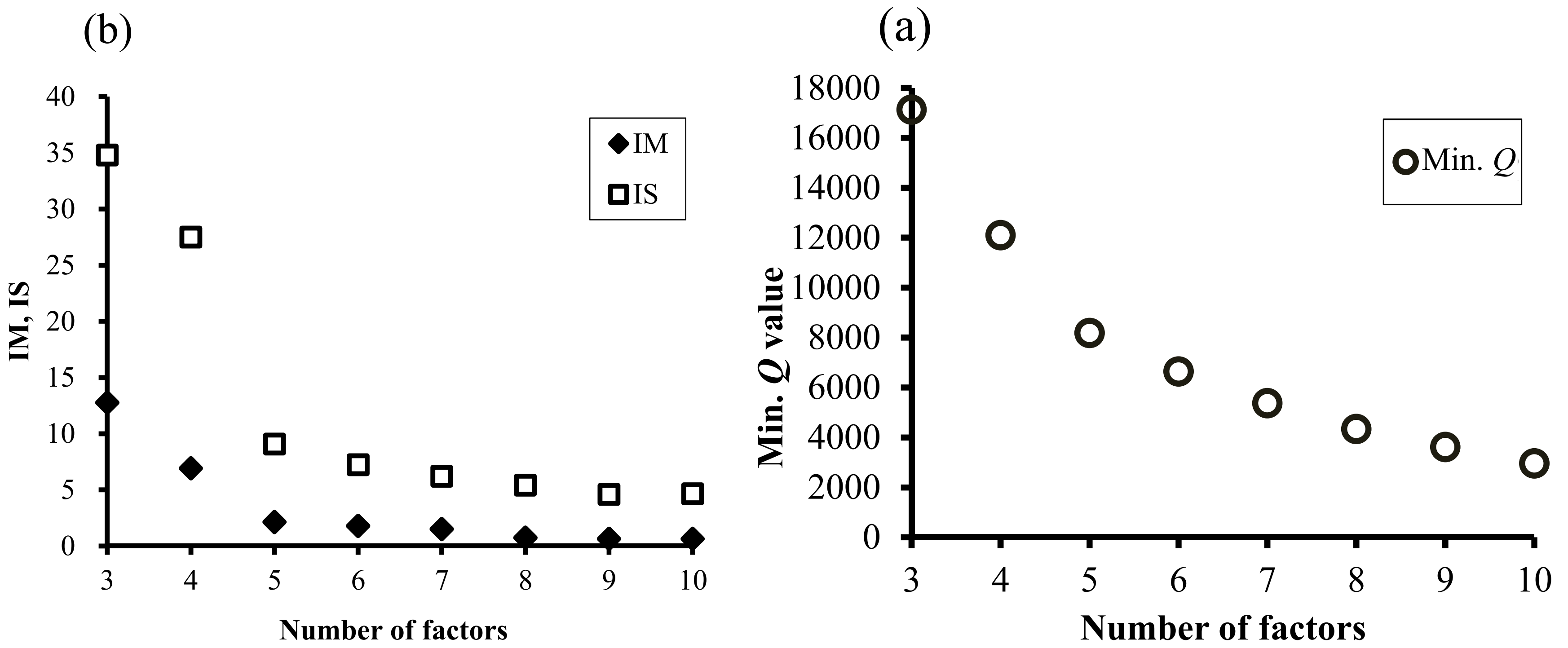

Fig. S1. Determination of the minimum number of factors needed for a well-constrained solution based on (a) minimum Q and (b) IM and IS values.

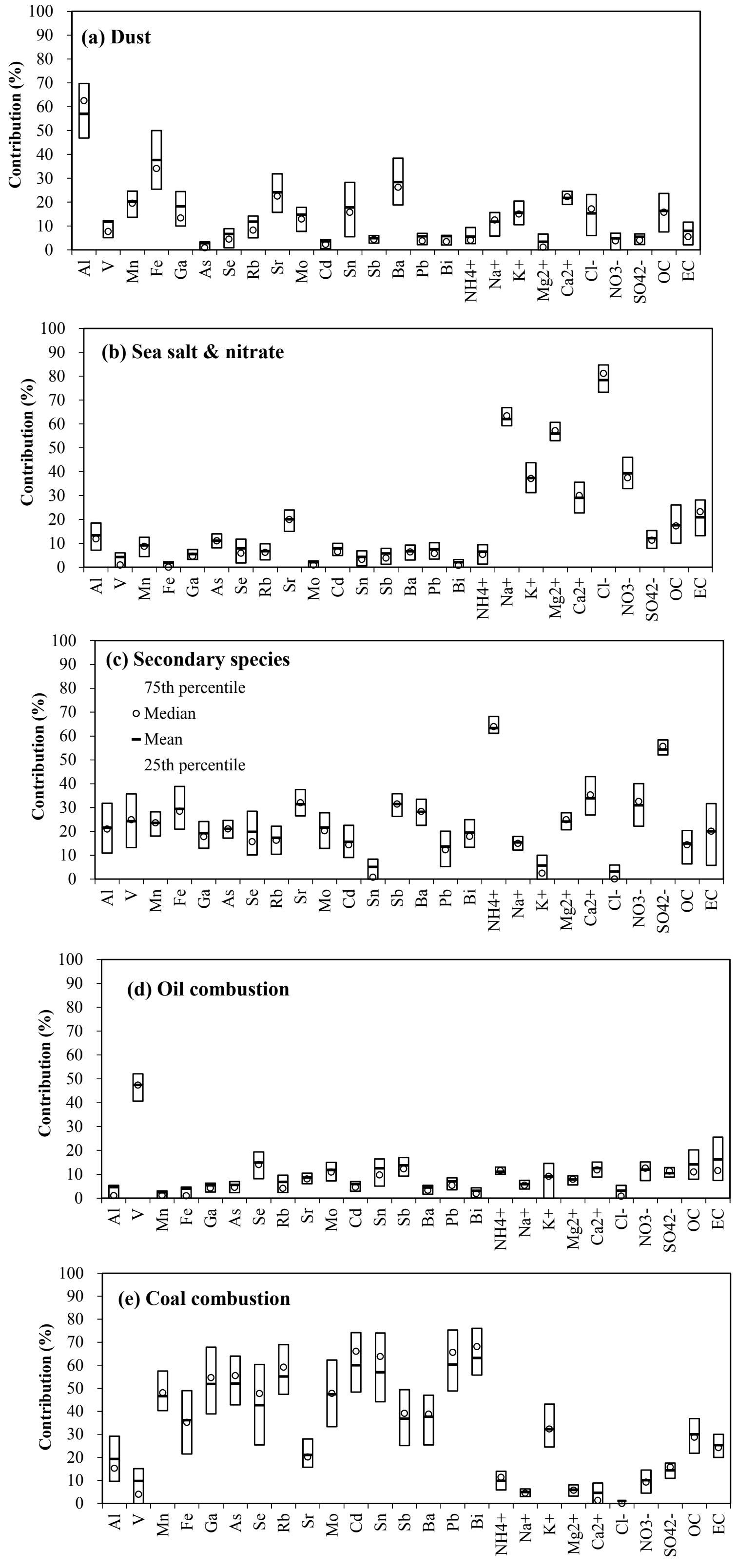

Fig. S2. The profiles obtained by bootstrapping of the relative contributions of (a) dust, (b) sea salt and nitrate, (c) secondary species, (d) oil combustion, and (e) coal combustion.

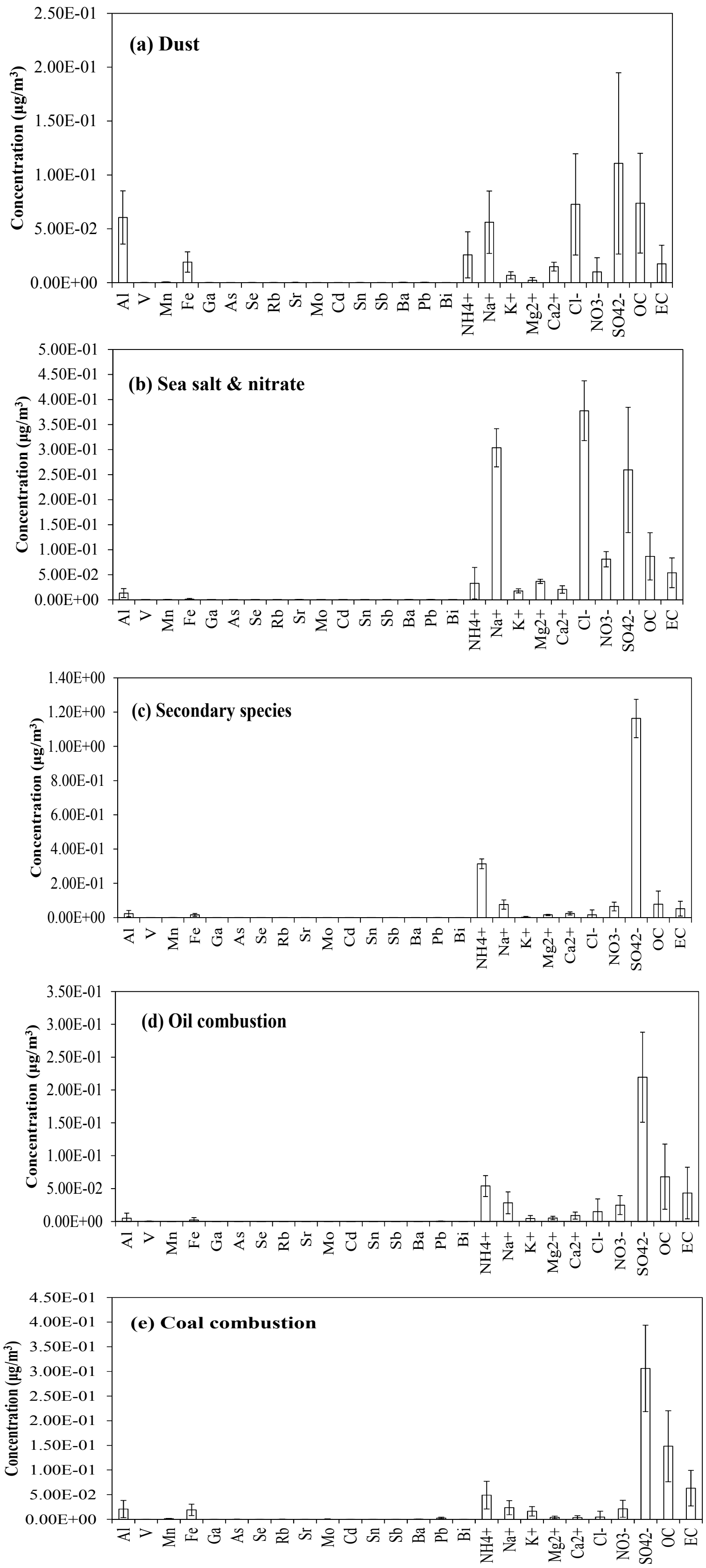

Fig. S3. The profiles obtained by bootstrapping of the absolute contributions (μg/m^3) of (a) dust, (b) sea salt and nitrate, (c) secondary species, (d) oil combustion, and (e) coal combustion. Error bars mean one standard deviation

Table S1. Summary of aerosol chemical measurements made throughout 2010 at CHAAMS

Spceies	Slope and R^2 (predicted versus observed)	No. of values below the detection limit	No. of missing values	Number of samples
Al	0.99 (R^2 = 0.99)	0	0	45
V	0.99 (R^2 = 0.99)	0	0	45
Mn	0.87 (R^2 = 0.93)	0	0	45
Fe	0.78 (R^2 = 0.83)	11.1	0	45
Ga	0.96 (R^2 = 0.94)	0	0	45
As	0.92 (R^2 = 0.94)	0	0	45
Se	0.76 (R^2 = 0.7)	11.1	0	45
Rb	0.88 (R^2 = 0.86)	0	0	45
Sr	0.78 (R^2 = 0.78)	0	0	45
Mo	0.57 (R^2 = 0.78)	0	0	45
Cd	0.97 (R^2 = 0.96)	11.1	0	45
Sn	0.78 (R^2 = 0.78)	0	0	45
Sb	0.89 (R^2 = 0.85)	0	0	45
Ba	0.87 (R^2 = 0.92)	0	0	45
Pb	0.96 (R^2 = 0.95)	0	0	45
Bi	0.96 (R^2 = 0.96)	0	0	45
NH_4^+	0.99 (R^2 = 0.99)	0	0	45
Na+	0.99 (R^2 = 0.99)	0	0	45
K^+	0.51 (R^2 = 0.78)	0	0	45
Mg^{2+}	0.89 (R^2 = 0.93)	0	0	45
Ca^{2+}	0.89 (R^2 = 0.88)	8.8	0	45
Cl^-	1.0 (R^2 = 0.96)	0	0	45
NO_3^-	0.60 (R^2 = 078)	0	0	45
SO_4^{2-}	1.0 (R^2 = 0.97)	0	0	45
OC	0.87 (R^2 = 0.87)	0	11.1	40
EC	0.90 (R^2 = 0.90)	0	11.1	40