

**Non-negative matrix
factorization of
particulate air
pollution in Beijing**

A. Thiem et al.

Using non-negative matrix factorization for the identification of daily patterns of particulate air pollution in Beijing during 2004–2008

A. Thiem¹, U. Schlink¹, X.-C. Pan², M. Hu³, A. Peters⁴, A. Wiedensohler⁵,
S. Breitner⁴, J. Cyrys^{4,6}, B. Wehner⁵, C. Rösch¹, and U. Franck¹

¹Core Facility Studies, Helmholtz Centre for Environmental Research – UFZ, 04318 Leipzig, Germany

²Peking University, School of Public Health, Department of Occupational and Environmental Health, Beijing, China

³Peking University, State Key Joint Laboratory of Environmental Simulation and Pollution Control, College of Environmental Sciences and Engineering, Beijing, China

⁴Institute of Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany

⁵Physics Department, Leibniz Institute for Tropospheric Research, Leipzig, Germany

⁶Environment Science Center, University of Augsburg, Augsburg, Germany

13015

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Received: 14 May 2012 – Accepted: 14 May 2012 – Published: 25 May 2012

Correspondence to: A. Thiem (alexander.thiem@ufz.de)

Published by Copernicus Publications on behalf of the European Geosciences Union.

ACPD

12, 13015–13052, 2012

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Abstract

Increasing traffic density and a changing car fleet on the one hand as well as various reduction measures on the other hand may influence the composition of the particle population and, hence, the health risks for residents of megacities like Beijing. A suitable tool for identification and quantification of source group-related particle exposure compositions is desirable in order to derive optimal adaptation and reduction strategies and therefore, is presented in this paper.

Particle number concentrations have been measured in high time- and space-resolution at an urban background monitoring site in Beijing, China, during 2004–2008. In this study a new pattern recognition procedure based on non-negative matrix factorization (NMF) was introduced to extract characteristic diurnal air pollution patterns of particle number and volume size distributions for the study period. Initialization and weighting strategies for NMF applications were carefully considered and a scaling procedure for ranking of obtained patterns was implemented. In order to account for varying particle sizes in the full diameter range [3 nm; 10 μm] two separate NMF applications (a) for diurnal particle number concentration data (NMF-N) and (b) volume concentration data (NMF-V) have been performed.

Five particle number concentration-related NMF-N factors were assigned to patterns mainly describing the development of ultrafine (particle diameter $D_p < 100\text{ nm}$ instead of D_p) as well as fine particles ($D_p < 2.5\ \mu\text{m}$), since absolute number concentrations are highest in these diameter ranges. The factors are classified into primary and secondary sources. Primary sources mostly involved anthropogenic emission sources such as traffic emissions or emissions of nearby industrial plants, whereas secondary sources involved new particle formation and accumulation (particle growth) processes. For the NMF-V application the five extracted factors mainly described coarse particle ($2.5\ \mu\text{m} < D_p < 10\ \mu\text{m}$) variations, generated by processes like dust storm events. Because particle volume depends on particle diameter in a cubic manner, larger particles are emphasized in the latter application.

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



In order to gain insight in the day-by-day varying source-associated composition of the particle burden non-negative linear combinations of individual source-associated pollution patterns were used to approximate the original particle data. Consequently, this NMF-based procedure provides a reasonable numerical-statistical tool for the description of daily patterns of particle pollution, source identification and reconstruction of daily patterns by summarizing weighted factors.

1 Introduction

In the last decades traffic density and energy consumption increased rapidly in Beijing and other Chinese megacities, with enormous impacts on air quality. European limit values for particle mass (EU, 1999) are exceeded tremendously day by day (Liu et al., 2008; Wehner et al., 2008; Wu et al., 2007, 2008), resulting in rising health risks for the residents (Breitner et al., 2011; Guo et al., 2009; Pan et al., 2007; Zhang et al., 2007). In order to derive source-dependent measures for reduction of particle concentrations it is necessary to gain insight into the exact composition of the particle burden. Hence, it is critical to understand the processes dominating particle concentrations within different particle size fractions in highly polluted areas.

Particulate matter (PM) is commonly defined as a mixture of particles of different size ranges and different physical and chemical properties. That means airborne particles do not constitute a uniform population. Regarding particle size, PM is most often distinguished between coarse (2.5 to 10 μm , CP), fine (< 2.5 μm , FP), and ultrafine particles (< 0.1 μm , UFP). Since fractions of different particle size originate from different sources and are associated to various chemical compositions hazardousness for human health differs. Numerous studies have faced PM composition and adverse health effects of elevated PM concentrations (Andersen et al., 2007, 2008; Arhami et al., 2010; Atkinson et al., 2010; Bell et al., 2009; Branis et al., 2011; Cyrus et al., 2003; Duvall et al., 2008; Franck et al., 2011a; Leitte et al., 2009; Pöschl, 2002; Suwanwaiphatthana et al., 2010). Previous research has shown that fine and ultrafine particles might be

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

more harmful than coarse particles (Chalupa et al., 2004; Franck et al., 2011b; Heinrich et al., 2002; Leitte et al., 2011b,c; Peters et al., 1997, 2002; Wichmann et al., 2000), because of their ability to penetrate more deeply into the lungs. Especially UFP may attain the pulmonary alveoli and can even translocate into the bloodstream (Penttinen et al., 2001; Rundell and Caviston, 2008). Because of high particle number concentrations and large active surface area per unit mass in the UFP range, these particles may contribute to the observed health effects largely (Brook, 2008; Delfino et al., 2005). Reviews are given in (Brook et al., 2010; Pope and Dockery, 2006; R uckerl et al., 2011).

In addition, mode of action and hazardousness of airborne particles are predetermined essentially by their origin. The majority of particles in Beijing's air originate from primary sources comprising emissions from traffic, industry, power plants, private households (e.g. domestic heating) or construction sites, which are all anthropogenically caused. Latest research also focused on soot particles (Cheng et al., 2011). Secondary sources are e.g. exhaust gases, which form the basis for particle nucleation processes. Weather conditions like wind speed, wind direction, relative humidity or temperature considerably affect urban air particle concentrations (Wehner and Wiedensohler, 2003). For example, lower mixing layer heights or temperature inversions in the lower troposphere – occurring frequently in the winter months – elevate the level of particulate air pollution heavily. The basin-like topography of Beijing's urban area and local weather systems are further important factors influencing air pollutant concentrations (Ren et al., 2004). Additionally, not only local emission sources but also long-range transported air pollutants and especially particles carried in by dust storms contribute substantially to particulate pollution in the urban air (Sun et al., 2004; Yu et al., 2011).

The issue of temporal and spatial variation of the particle burden is still an unresolved question in atmospheric and exposure research. Cyrus et al. (2008) focused the problem in a case study for Augsburg, Germany, by using 4 different monitoring sites. The authors concluded that a single monitoring site is not sufficiently capable of approximating the particle number concentration (NC) level for a whole urban area, but the temporal variation of particle concentrations can be represented rather satisfactorily as

meteorological conditions and industrial emissions are similar for the whole area. Other studies analyzed the spatio-temporal correlation between different types of monitoring sites like traffic, urban background and regional background (Buzorius et al., 1999; Krudysz et al., 2009; Puustinen et al., 2007; Tuch et al., 2006; Wang et al., 2009).

Positive matrix factorization (PMF) is a widely used receptor model, developed by Paatero (1997). It is commonly used for characterizing aerosol sources by calculating dominant positive factors on the basis of observation without detailed prior knowledge of the sources and source profiles. It decomposes the measured PM composition data or particle size distribution (PSD) data into sub factor profiles and factor contributions. Compared to the previously used methods like principal components analysis (PCA), it has the advantage of more realistic non-negative constraints on factor profiles and contributions, and better scaling of the data by individually assigned uncertainties (Paatero and Tapper, 1994; Paatero, 1997). The application of PMF to PSD data has been successfully demonstrated in several studies (Gu et al., 2011; Ogulei et al., 2007; Yue et al., 2008; Zhou et al., 2004).

Following this idea we decided to use non-negative matrix factorization (NMF), which was introduced by Lee and Seung (1999). Actually, NMF is a basic principle realized by a group of algorithms and differing extensions comprising the PMF method as a special case. NMF aims on decomposing a given non-negative data matrix into non-negative characteristic features and non-negative linear coefficients. Contrary to traditional approaches for PSD data, our data matrix was arranged by diurnal data sets comprising quarter-hourly particle size distributions in the diameter range 3 nm–10 μ m. Because of the high dimensions of the data matrix and the need for better physical reliability of the resulting factors we used a modified NMF algorithm based on the fast projected gradients method of Lin (2007) in combination with a special initialization technique called non-negative double singular value decomposition (NNDSVD, Boutsidis and Gallopoulos, 2008).

The aim of this study was, on the one hand, to enhance NMF by several adjustments as a pattern recognition tool to the field of air pollution. On the other hand, it was

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



our intention to use this method to categorize diurnal particle burden patterns for the urban area of Beijing in the period 2004–2008 into few categories, which could then be associated to specific sources and weather conditions and classified into primary (mostly anthropogenic) and secondary (mostly based on natural processes) source groups.

2 Material and methods

2.1 Particle measurements

The station for particle measurements was located on the campus of the Peking University in an urban background area in the Haidian district in the northwestern part of Beijing. It was placed 20 m above ground on top of a campus building and more than 500 m away from major roads to reduce local source effects on particle measurements. Because the surrounding area is primarily residential and commercial, local emission sources mainly are vehicular traffic, construction sites, and fuel combustion for domestic cooking and heating. Industrial emissions are to be expected as regional variable. Yue et al. (2009) confirmed that the measuring station can be considered being an urban background station.

Particle measurements were performed by the State Key Joint Laboratory of Environmental Simulation and Pollution Control, College of Environmental Sciences and Engineering, Peking University (PKU), Beijing, and the Leibniz Institute of Tropospheric Research (IfT), Leipzig, using a Twin Differential Mobility Particle Sizer (TDMPMS) together with two condensation particle counters (CPC) for the size range 3 nm to 800 nm as well as an Aerodynamic Particle Sizer (APS) for particles between 800 nm and 10 μm . APS data were transformed from aerodynamic to Stokes diameter assuming spherical particles and a particle density of 1.7 g cm^{-3} for particles > 800 nm for Beijing (Yao et al., 2002). Afterwards, both data sets were concatenated truncating the APS data to obtain particle number size distributions in the size range from 3 nm to 10 μm . A more

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



detailed description of the measurement setup can be found elsewhere (Wehner et al., 2004, 2008).

The period of investigation was 1 July 2004 to 14 December 2008 with a measurement interval of 10 min. Measurements of both DMPS and APS were available for 1213 days out of 1750. To obtain a most comprehensive data base for pattern recognition, measurement failures of less than half an hour were linearly interpolated with existing adjacent values, whereas records containing larger failure intervals had to be excluded from the pattern recognition process. The records for 6 January 2008 to 10 January 2008 and 1 December 2008 to 6 December 2008 reported obviously false measurements with extremely elevated UFP levels, which could not be explained by weather fluctuations or known particle emission processes. Consequently, these records were removed precautionally prior to the analysis. After all these restrictions in total 864 out of 1750 days with nearly complete and usable data records were determined during the sample period.

Size distributions for particle volume V are obtained from particle number N by

$$\frac{dV}{d(\log D_p)} = \frac{\pi \cdot D_p^3}{6} \cdot \frac{dN}{d(\log D_p)} \quad (1)$$

assuming spherical particles and depending on particle diameters D_p . Since the particle diameter affects particle volume in a cubic manner, coarse particles contribute more efficiently to total particle volume and mass concentrations. Hence, high particle number concentrations, e.g. in the ultrafine particle range, do not necessarily correlate with high volume or mass concentrations.

2.2 Weather and climate of Beijing

Beijing is characterized by a temperate monsoon climate. Summers are warm and wet with the warmest period from June to August, whereas winters are cold and dry. During this dry season (November to March) the air pollution level rises significantly,

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



e.g. through increased domestic heating. Additionally, occurring temperature inversions and lower mixing layer heights favor high particle number concentrations (Janhäll et al., 2006; Silva et al., 2007). These conditions are characterized by cold air close to the ground and warm air above. Emitted particles are trapped beneath the boundary layer and may accumulate over time.

Due to its location in the North China Plain – surrounded by the Mongolian Plateau in the north and west – besides the typical westerlies Beijing's weather conditions are marked by local weather phenomena like mountain and valley breezes, sea land breezes and the urban heat island effect (Liu et al., 2009). Dust storms are also distinctive processes, originating in the Gobi Desert and exerting influence on Beijing's urban area by transporting dust particles. These wind systems clearly affect particle concentrations in the Chinese capital.

Figure 1a presents the diurnal particle number size distribution pattern of a typical winter day without dust storm influence, but characterized by particle nucleation in the UFP range and growing processes during the course of the day. To the contrary, Fig. 1b illustrates elevated particle number concentrations in the coarse particle range between 8 a.m. and 2 p.m. for a typical dust storm day.

For the purpose of interpretation of pattern recognition results and corresponding time series meteorological data were obtained from the data archive on www.weatherunderground.com. The corresponding measurement site was located at Beijing international airport in about 25 km distance to the particle measurement site.

2.3 NMF

Non-negative matrix factorization is a novel statistical approach for pattern recognition introduced by Lee and Seung (1999), which since then has been applied to several research areas, e.g. to meteorological fields of atmospheric pressure (Schlink and Thiem, 2010). The main idea of the method is to decompose a data matrix into a limited set of characteristic patterns, called NMF factors, and associated weighting coefficients. Combining the NMF factors again by their weights approximates the original data in

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



a non-negative additive manner. That way, a reduced-dimensional representation of large data sets can be obtained.

Restricted by the non-negativity of the raw data matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$ as well as the arising NMF factors $\mathbf{W} \in \mathbb{R}^{n \times r}$ and their corresponding weights $\mathbf{H} \in \mathbb{R}^{r \times m}$, the NMF factorization

$$\mathbf{X} \approx \mathbf{W}\mathbf{H}$$

$$w_{ia} \geq 0, \quad h_{bj} \geq 0 \quad (2)$$

$$i = 1, \dots, n; \quad j = 1, \dots, m; \quad a, b = 1, \dots, r.$$

is approximate, since for r holds $r \ll \min(n, m)$. Factors and weights are calculated iteratively by minimizing a suitable measure f for the distance between \mathbf{X} and $\mathbf{W}\mathbf{H}$. Starting with a reasonable initialization ($\mathbf{W}^0, \mathbf{H}^0$) the pair of matrices is improved step by step until convergence. The most commonly used distance function for this purpose is the Euclidean distance. Hence the NMF problem can be rewritten as the following constrained optimization problem (Chu et al., 2004):

$$f(\mathbf{W}, \mathbf{H}) = \frac{1}{2} \|\mathbf{V} - \mathbf{W}\mathbf{H}\|_F^2 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^m (x_{ij} - (\mathbf{W}\mathbf{H})_{ij})^2 \rightarrow \min_{\mathbf{W}, \mathbf{H}}, \quad \mathbf{W}\mathbf{H} \geq 0 \quad (3)$$

where $\|\mathbf{A}\|_F$ denotes the Frobenius norm for a matrix \mathbf{A} . Other distance measures, all of them special cases of the so called β -divergence like the generalized Kullback-Leibler divergence (Dhillon and Sra, 2005), are possible, but were not considered in this study.

Several NMF algorithms based on gradient descent or other non-linear optimization methods were developed. Lee and Seung (2001) presented simple but slow multiplicative update formulas. In most applications alternating non-negative least squares methods are commonly used (Berry et al., 2007; Cichocki et al., 2006; Gonzalez and Zhang, 2005), among others the projected gradient algorithm suggested by Lin (2007).

We adjusted this algorithm for use in our NMF calculations. After careful consideration of stopping criteria, we stopped NMF runs at a maximum iteration number of 1000.

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



2.3.1 Initialization for NMF

Unfortunately the NMF problem is non-convex and non-linear in the variables w_{ia} and h_{bj} . Hence all optimization methods are solely able to find a local despite of a global minimum of the objective function f (Chu et al., 2004). In other words, NMF does not assure uniqueness of its results, if starting values for the algorithm are chosen randomly. Every run of the NMF algorithm may find another local minimum of the objective function. However, uniqueness and consistence of the NMF factors are desirable properties for this kind of analysis, because physical interpretability of the obtained patterns is intended. Therefore a reasonable preprocessing step for initialization of matrix factors \mathbf{W} and \mathbf{H} is necessary and helps to find an appropriate solution.

We tested three different kinds of initialization techniques for calculating NMF starting values. In the first method a PCA was applied to the covariance matrix of the original data, followed by a shifting of these PCA factors into the non-negative orthant and a diagonal scaling by the mean of the original data. Secondly, a random selection based on a Gaussian distribution with original data sample mean and standard deviation was used. Eventually, an initialization algorithm called non-negative double singular value decomposition NNDSVD was applied (Boutsidis and Gallopoulos, 2008). Because this algorithm is based on singular value decomposition (SVD), which is an equivalent procedure to PCA, it provides starting matrices that are calculated by maximizing the explained data variance, but incorporating the non-negativity constraint.

Testing procedures showed that, as expected, in this special case of application each of the three initialization strategies leads to different starting values and calculation times for NMF. NNDSVD initialization resulted in fastest NMF algorithm convergence. Unexpectedly, for each initialization method NMF generated extremely resembling results. Nevertheless NNDSVD was regarded best appropriate, because it is built on PCA and hence accounts for reasonable and physically interpretable starting points. We therefore used it for our final calculations.

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



2.3.2 Application to particle number and volume data

Given size distributions for particle number N and calculated data for particle volume V were considered on a daily basis ranging from 0 a.m. to 12 p.m. as particle pollution patterns analogous to Figs. 1–3 in Wehner et al. (2004). For each day, 15-min interval data values for diameters from 3 nm to 10 μm were stored together in one vector. In other words, each data “pixel” of the charts in Fig. 1 corresponds to one vector component. The whole set of data vectors yields a matrix $\mathbf{X} = \mathbf{X}(D_p, t, d)$ of daily size distribution patterns for particle number N or volume V , respectively. Each matrix element depends on time of day t , particle diameter D_p (rows of \mathbf{X}) on the one hand and the date of measurement d (columns of \mathbf{X}) on the other hand.

Applying non-negative matrix factorization to the matrix \mathbf{X} decomposes the set of diurnal size distribution measurements into a matrix \mathbf{W} of characteristic NMF factors and a matrix \mathbf{H} of time-dependent indices for each of these factors. By specifying the number of NMF factors r , the r most important factors are extracted, which can then be interpreted as physically caused air pollution patterns. A linear combination of the obtained NMF factors by coefficients in \mathbf{H} enables to reconstruct the original measured data for each single day. Fig. 2 illustrates this favorable ability of the NMF results.

The specification of the number of NMF factors had to be considered carefully. Values in the range $r = 3 \dots 10$ were tested on the basis of 864 daily data sets. In this special case best physical interpretation and dimension reduction at the same time were obtained for $r = 5$. The choice $r < 5$ produces combined factors consisting of more than one physical pattern. Hence interpretation becomes more difficult. By contrast, in the case of $r > 5$ in addition to maintained patterns (results for $r = 5$) further small-scale patterns occur, that only marginally contribute to the true particle concentrations within the original measured patterns. Residual analysis showed that the relative error of reconstructed data based on NMF-superposition with regard to the original measured data did not reduce significantly when increasing the number of patterns (r). On the other hand, no systematic variations are missed for $r = 5$.

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Since particle number is dominated by small particles in the ultrafine and fine range, (whereas the number of coarse particles is usually considerably less) obtained NMF patterns for particle number emphasize these ranges. In order to make statements about features of coarse mode particles a certain weighting for the original measured data has to be accomplished before NMF application. One way to achieve this implicitly is to use particle volume instead of particle number data. Since the particle diameter affects the volume in a cubic relationship, the behavior of larger particles then sticks out more clearly. Alternatively, we considered direct weighting of particle number size data in \mathbf{X} by the reciprocal of the long-term data mean, depending on particle diameter D_p , as well as a weighting by the decadic logarithm of the corresponding particle diameter (since particle size varies by several powers of ten).

Hence, separate applications of NMF were calculated for particle number and volume size distribution data, respectively. Resulting patterns were compared according to their representativeness and physical interpretability. Both explicit weighting procedures were also tested; however, since results did not provide any further information, they are omitted here.

2.3.3 Relevance of NMF factors

In contrast to other pattern recognition procedures like PCA, where patterns are automatically sorted according to decreasing explained variance of the original data, the NMF method does not incorporate a certain measure of importance for NMF factors. Thus, the definition of an additional sorting strategy is necessary. An appropriate importance measure could be the order of magnitude of the corresponding weights in \mathbf{H} . If the average weight of one factor is higher than the average weight of another one, its influence on the linear combination and hence, on the original data is regarded more important. But this is only valid, if the patterns themselves have all equal means. This claim has to be incorporated as a constraint in the NMF procedure. For simplification, variance was assumed to be constant for all factors. Therefore, after NMF calculation

we applied a diagonal scaling step

$$\hat{\mathbf{W}} = \mathbf{W}\mathbf{D}, \quad \hat{\mathbf{H}} = \mathbf{D}^{-1}\mathbf{H} \quad (4)$$

on the pattern and weight matrices, aiming to achieve simultaneous scattering around a predefined mean value for each of the NMF factors, where

$$\mathbf{D} = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_r \end{pmatrix} \in \mathbb{R}^{r \times r} \quad (5)$$

Following Berry et al. (2007) the minimum property of the NMF solution is maintained during this procedure, because

$$\hat{\mathbf{W}} \cdot \hat{\mathbf{H}} = \mathbf{W}\mathbf{D} \cdot \mathbf{D}^{-1}\mathbf{H} = \mathbf{W} \cdot \mathbf{D}\mathbf{D}^{-1} \cdot \mathbf{H} = \mathbf{W} \cdot \mathbf{I} \cdot \mathbf{H} = \mathbf{W} \cdot \mathbf{H} = \mathbf{W}\mathbf{H} \quad (6)$$

holds for each non-negative regular matrix $\mathbf{D} \in \mathbb{R}^{r \times r}$ with (r, r) -identity matrix \mathbf{I} . Diagonal elements of \mathbf{D} were formed as

$$d_k = \frac{C}{\bar{w}_k} = \frac{C \cdot n}{\sum_{i=1}^n w_{ik}}, \quad k = 1, \dots, r \quad (7)$$

with an arbitrary constant C , where in this case C was chosen as the total original data mean. Consequently, all NMF factors were normalized regarding their mean, scattering around C and hence time series in \mathbf{H} could be analyzed and compared to each other regarding their mean or median amplitude. It has to be pointed out, that in doing so time-dependent coefficients in \mathbf{H} , called weights, do not necessarily sum up to 1.

In our application measured data means for number and volume data were $C = 7208 \text{ cm}^{-3}$ and $C = 2.414 \times 10^{10} \text{ nm}^3 \text{ cm}^{-3}$, respectively.

3 Results and discussion

3.1 Particle number NMF (NMF-N)

In the following, the five obtained NMF factors (called NMF-N) for particle number N are analyzed and related to different physical conditions, including meteorological processes and various sources. Table 1 lists main characteristics of the five factors and their associated sources. The corresponding patterns of these factors are shown in Fig. 4. NMF factors contribute to the total particle burden by time-varying coefficients. In other words, these coefficients represent the diurnal impact of each of the factors at specific days. Physical interpretation of the factors as characteristic pollution patterns therefore also depends on time-dependent coefficients in H . The factors are sorted according to their average contribution to the particle burden, measured in terms of the median value of the corresponding time series. Most impact is exerted by the pattern NMF-N1, followed by NMF-N2, NMF-N3 and so on. Figure 3 presents median contributions, divided into seasons.

NMF-N1: The first factor NMF-N1 describes contributions to the particle number concentration in the size range (10 nm, 1000 nm). Permanent emissions of private households, traffic, and industry are combined to form this underlying pattern. It provides a background pattern with a typical diurnal particle number size distribution and, thus, a kind of basic load. The impact of more dynamical external influences is described by other NMF factors. Consequently, the important process of particle nucleation, for example, is completely excluded from NMF-N1 resulting in an artificial “zero particles” area around noon as seen in Fig. 4a. NMF-N1 shows highest median time-dependent weights (0.318) and thus, most impact on the particle burden among all five factors and a rather small temporal variability. It exerts most influence during the winter months (December, January, February, compare Fig. 4b and Fig. 3), which is on average 1.4, 1.7 and 1.2 times higher compared to summer (June, July, August), spring (March, April, May) and autumn (September, October, November), respectively. This supports the assumption that NMF-N1 describes the basic urban background pollution, which is

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



generally higher during winter, enhanced by typical weather situations like temperature inversions.

NMF-N2: In NMF-N2 the maximum value of the particle number size distribution shifts towards larger particle diameters with increasing time of day indicating particle growth. The growing process starts in the morning and ends up at midnight Fig. 4c. The pattern is also associated to traffic emissions. Although Beijing traffic is characterized by high volume during the whole day, there still exist rush hour peaks in the morning and the late afternoon. At these times the second factor NMF-N2 reaches maximum concentrations for particle diameters around 10 to 100 nm, which can be related to exhaust emissions (Costabile et al., 2009; Shi et al., 2007; Uhrner et al., 2011). Highest impact of the traffic pattern is achieved in autumn and winter, being on average 1.6 and 1.9 times higher as in summer.

NMF-N3: The third factor NMF-N3 accommodates for primary particle emissions from industrial and power plants, which are transported to the city of Beijing by local wind systems. The location of Beijing in the North China Plain surrounded by mountains to the north and to the west provides the basis for a mountain and valley breeze circulation. At night, mountains cool down much faster than the lowlands. This effect is even amplified by the urban heat island effect in the area of Beijing. As a consequence, cool air sags in the mountains and flows southwards (so-called mountain breeze). At early noon this process turns upside down. The mountains heat up earlier after daybreak than Beijing urban area so that warmer air masses rise and mixing layer height increases. Reduced air pressure in the mountains causes an upstream flow of air masses from the lowlands resulting in the valley breeze from southern directions. By transporting particle emissions of southwards located industry to Beijing this wind system contributes significantly to the whole particle burden. The correlation coefficient of NMF-N3 weights and wind velocity at Beijing airport is 0.47. However, the correlation with wind direction cannot be computed, because of hourly wind direction data and daily NMF coefficients. But for randomly picked data sets, where the mountain and valley breeze wind system was prevalent, we observed significantly higher weights for

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



NMF-N3. Figure 4d illustrates the influence of this system, which divides the diurnal particle number pattern into two parts. At night and during the morning hours a typical particle number size distribution occurs, but at approximately 10 a.m. the valley breeze starts and transports especially ultrafine particles to Beijing, whereas the concentration of larger particles (100 nm to 1000 nm) decreases. The industrial pattern is most important in spring (2.5 and 2.1 times higher compared to summer and winter, respectively), when the local mountain and valley breeze is prevalent more often.

NMF-N4: The factor NMF-N4 displays the process of new particle formation and subsequent particle growth in the course of the day. This pattern (Fig. 4e) is well known in the literature and often observed in diurnal particle number concentration data sets. Because of its shape it is often called the “banana-like” pattern. A rising mixing layer height in the morning hours firstly causes a slight decrease of particle concentrations at about 8 a.m. Afterwards, due to increasing solar radiation, new particle formation begins and increases the number of ultrafine particles (Metzger et al., 2010; Wu et al., 2007; Yue et al., 2009). In the sequel these particles subsequently grow up until 6 p.m. in the evening due to coagulation (Meier et al., 2009). The “banana-like” pattern shows most impact during spring and summer, probably caused by increased solar radiation, but less during winter (0.62 or 0.56 times lower median weight). One reason could be, that the effect of hygroscopic particle growth is less important in the winter months (Meier et al., 2009), because of drier weather and thus less condensation of water droplets on the particle’s surface.

NMF-N5: A red area in NMF-N5 indicates elevated UFP levels showing a clear particle nucleation pattern. It comprises extremely high concentrations for the so-called nucleation mode particles ($D_p < 30$ nm). The particle formation event starts at 8 a.m. in the morning and ends at around 2 p.m. (Fig. 4f). Compared to factor NMF-N4, formed particles are even smaller and do not grow up afterwards. However, a combination of NMF-N5 and NMF-N4 is also possible. Especially in winter and spring, when temperature inversion appears frequently and mixing layer height is very low in the morning (Janhäll et al., 2006; Schaefer et al., 2006; Seibert et al., 2000; Silva et al., 2007),

NMF-N5 exerts much influence, whereas during summer it is almost negligible (11.8 or 19.4 times higher median in winter and spring compared to summer).

The variation in time of all five NMF-N factors follows mainly variations of particles in the ultrafine and fine size ranges, since the total particle number concentration is dominated by these ranges and absolute number concentrations of coarse particles vary much less than those of finer. Consequently, Beijing-typical coarse mode variations like dust storm events are not captured by particle number NMF-N factors. To account for variations of coarse particles, a weighting of the measured data is required. The most obvious method is to use particle volume concentration data.

3.2 Particle volume NMF (NMF-V)

A second NMF application was performed for particle volume concentration data (called NMF-V). Again, five factors were extracted and considered to be most suitable for pattern description. Factors have been sorted according their importance (median weights) and compared to each other. Contrary to the NMF-N results, the first factor of NMF-V does not describe a basic particle burden pattern. It is just one of five factors describing several distinct variations in the coarse mode. Fine particles' fluctuations are not covered. A direct interpretation of the factors in terms of physical conditions, however, is not as clearly possible as for particle number patterns. Therefore, a detailed description of these patterns is omitted (see Supplement).

For the Beijing urban area dust storms are the main reason for elevated particulate air pollution levels for coarse particles (Wu et al., 2008). Especially concentrations of particles $> 1 \mu\text{m}$ are significantly higher on dust storm days. As these dust storm events occur at different times in the course of the day, not just one but several NMF-V factors obviously present patterns associated with this phenomenon at different times of the day. Figure 1 illustrates diurnal particle patterns for a typical non-dust storm day (Fig. 1a) and a typical dust storm day (Fig. 1b), respectively. In the dust storm example on 18 March 2008, the time-dependent coefficient for NMF-V4 is maximal (Fig. 5), i.e. this pattern exerted enormous influence on the daily particle volume values. Further

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



dust events are indicated by elevated weights in Fig. 5b. Whereas NMF-V4 mainly accounts for events in the morning hours (Fig. 5a), the other factors describe evening and night variation patterns of coarse particles.

3.3 Data reconstruction

By the help of time-dependent weighting coefficients in the matrix \mathbf{H} each day an additive and non-negative superposition of the NMF-factors in \mathbf{W} can be obtained to reconstruct the measured data patterns. This superposition enables us to trace the behavior of particle concentrations for certain particle diameters D_p over time divided into sources. Each day, each factor contributes in a different magnitude to the whole diurnal PSD.

Since the representation of the measured data \mathbf{X} by its factors is approximate, differences between non-negative NMF factor superpositions and the real measured data arise. Additional factors controlling the rest of the particle variations (in other words the absolute error pattern) may exist. However, they are rather negligible in terms of their average contribution, i.e. very infrequent or permanently less important.

Figure 6b shows the NMF-N superposition of five obtained particle number factors for 10 consecutive days in March 2008 (13 March 2008 to 22 March 2008). For three different particle diameters (3000 nm, 100 nm, 10 nm) the resulting curve is compared to a time profile for the measured particle number data (black line). In the nucleation mode, represented by the 10 nm area plot, the number of particles is largely determined by the two NMF-N factors NMF-N3 and NMF-N5. They mainly describe new particle formation events and influence of industrial emissions in the first days of the studied period (Fig. 6a).

For increasing diameter values the composition of the curve changes. The 100 nm-particles are mainly affected by NMF-N1, NMF-N2 and NMF-N3, which are now essential for the reconstruction of the original data, while factor NMF-N5 shows almost no importance. The reason is that the associated pattern for NMF-N5 (Fig. 4f) only accounts for small shares to particle concentrations in this diameter range. In other

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



words, the influence of an NMF-factor on the superposition curve is not only determined by daily weighting coefficients in \mathbf{H} , but depends also on the diameter and thus on the pattern itself. Likewise, the composition of the curve changes once again when proceeding to the coarse particle mode, represented by $D_p = 3000\text{nm}$. NMF-N1 – the background pattern – is then contributing most.

The area plots for three different diameters in Fig. 6b illustrate clearly that the five factor NMF-N model performs best for concentrations of UFP and FP particles, worse for CP. To put it another way, the larger the particle diameter, the worse the ability of approximating the initial data curve by a particle number based model. For example, the dust storm on 18 March 2008, which transported enormous amounts of coarse particles into downtown Beijing, is not detected by the NMF-N superposition curve. None of the factors has the ability to reproduce increases in the particle number for large particle diameters around $D_p = 3000\text{nm}$.

For the detection and simulation of variations in the coarse mode, NMF-V factors for particle volume are used. They emphasize larger particles and thus obtained patterns also reconstruct behavior during dust storm events. Figure 6d presents a superposition of NMF-V factors and $D_p = 3000\text{nm}$, 100nm and 10nm , respectively. In contrast to the approximation curve for NMF-N factors, the enormous increase in the concentration of coarse particles on 18 March 2008 is now clearly modeled by NMF-V4, which is also dominating for $D_p = 3000\text{nm}$ during the whole study period (Fig. 6c). Conversely, fluctuations of smaller particles are not captured as well by NMF-V factors.

3.4 Categorization of NMF patterns

In Beijing urban area, primary aerosol particles are mostly emitted by anthropogenic caused particle sources, especially in the fine and ultrafine range. The first three NMF-N factors are dominated by these primary PM sources. NMF-N factors 4 and 5, however, account for particle growing processes and particle nucleation, which can be considered secondary particle processes. Although in urban air the nucleation of particles from gaseous precursors is mostly based on anthropogenically emitted air pollutants

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



like SO₂ (Metzger et al., 2010) and thus is also influenced by primary processes, we decided to categorize factors 1 to 3 into primary and factors 4 to 5 into secondary source categories. Both secondary factors are mainly favored by the prevalent meteorological conditions, whereby there exist important differences for summer and winter months. During summer particle nucleation is fostered by higher solar radiation (Birmili and Wiedensohler, 2000) and higher relative humidity supports hygroscopic particle growth. Otherwise, during winter occurring nucleation events are often amplified by lower mixing layer height in the morning hours and temperature inversion (Janhäll et al., 2006). Despite their anthropogenic origin, factors 1 to 3 are also influenced by atmospheric conditions. Temperature inversions increase the concentration of traffic emitted particles in the lower atmosphere. Hence, the strength of NMF-N2 depends to some extent on weather. Besides temperature and relative humidity, wind plays an important role. The impact of industrial emissions from southwards plants in NMF-N3, for example, is obviously increased by the local mountain and valley breeze wind system.

NMF results for particle number concentrations suggest that the primary (mostly anthropogenic) particle emission sources show on average more influence on UFP and FP air pollution levels in the Beijing urban area than secondary factors, since the corresponding weights are on average higher. The bar diagram in Fig. 3, presenting median values for time-dependent weights of NMF-N factors, underlines that the major part of the particle burden is explained by the first three factors (NMF-N1, NMF-N2, NMF-N3).

For Beijing high concentrations in the coarse mode are mainly fostered by dust particles (primary source). The particle volume NMF application captures variations of larger particles and therefore comprises factors for dust storm events. Since other natural primary and secondary emission sources like volcanic ash or sea salt are less important for the period of investigation and the area of Beijing, respectively, they were not found as NMF-V patterns.

3.5 Advantages of NMF results

One of its major advantages compared to other matrix factorization or pattern recognition techniques is otherwise the strongest restriction of NMF – the non-negativity constraint. It is essential for obtaining a physical interpretable decomposition of the measured data by assuring that no cancellations appear when superposing the factors. As putting one brick over another, several factors associated to primary and secondary particle sources are combined additively to form the whole building, i.e. the particle burden. The height of each brick (represented by the weight of each factor) varies day by day indicating changing contributions from various sources. PCA and similar techniques do not restrict the factors to be non-negative (Jolliffe, 1986). A meaningful superposition is not possible, because single effects in the patterns amplify, whereas others neutralize. PCA uses another constraint to reduce data dimensions: factors need to be orthogonal. True conditions however are not necessarily formed by orthogonal processes and are in fact somehow related to each other. NMF omits this restriction. Further advantages of the NMF method with respect to PCA were discussed more in detail in Schlink and Thiem (2010).

4 Conclusions

The aim of this study was to extract typical diurnal features of particulate air pollution in Beijing for the period 2004–2008. By the help of non-negative matrix factorization characteristic PSD patterns on a daily basis were identified as vectors in the factor matrix and assigned to specific sources. To account for variations of fine and ultrafine particles on the one hand, as well as coarse particles on the other hand, two separate NMF decompositions for particle number and particle volume were calculated. By this means, an itemized view on particulate matter is maintained throughout the pattern recognition process. As a consequence of our results, we recommend using particle number (NMF-N) and particle volume applications (NMF-V) simultaneously to receive

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



characteristic patterns for fine and coarse particle variations separately and to gain a comprehensive view on the composition of particulate air pollution in its full size range.

Whereas processes involving fine and ultrafine particles like new particle formation events or emissions from traffic and industry were captured best in the NMF-N application to particle number, coarse mode fluctuations like dust storm events were detected by an NMF-V on particle volume, which implies a kind of implicit weighting on the original data. Explicit weighting of the measured data was also investigated and considered reasonable. It is envisaged to further address the pros and cons of different weighting strategies in future work.

Regarding the NMF initialization, it was found that for the data set used in this study in most runs of the algorithm random initialization led to similarly good NMF results as found by the NNDSVD initialized method. Nevertheless, in general random starting values do not guarantee unique and interpretable results. We consider NNDSVD or similar initialization strategies very reasonable and therefore recommend them for this type of NMF applications.

The daily varying contribution of each of the extracted components to the whole particle burden was analyzed by the corresponding time-dependent coefficients. In this way, a source-associated development of particle exposure during the investigation period was retrieved and the ability of NMF factors to reconstruct measured data in a non-negative additive manner was proven. Further studies could start at this point, using NMF factors as a technique for prediction purposes as well as data imputation for particle number and mass concentrations taking up the ideas in e.g. Leitte et al. (2011a).

Consequently, our analyses point out that based on the introduced modifications our NMF approach is a suitable tool for source identification based on daily PSD data.

Future studies could take up the idea of decomposing the particle burden at several measurement sites at a time. Including different nearby sites in a single application has the advantage of adding another dimension for analyzing spatial variability. In this way better localization of sources could be achieved.

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



For the Olympic Games 2008 in Beijing it is intended to use the obtained NMF factors for the investigation of source-associated reduction effects caused by source control measures, which were introduced for improvement of air quality during the Beijing Olympics, by means of their corresponding factor weights, extending the outcomes of already conducted studies (Gao et al., 2011).

Supplementary material related to this article is available online at:
[http://www.atmos-chem-phys-discuss.net/12/13015/2012/
acpd-12-13015-2012-supplement.zip](http://www.atmos-chem-phys-discuss.net/12/13015/2012/acpd-12-13015-2012-supplement.zip).

Acknowledgements. This study was supported by DFG (Deutsche Forschungsgemeinschaft) grant FR 1417/4-1. The authors thank all co-workers, especially from the IfT Leipzig and the PKU Beijing for performing and providing particle measurements. The supply of weather archive data by www.weatherunderground.com is gratefully acknowledged.

References

- Andersen, Z. J., Wahlin, P., Raaschou-Nielsen, O., Scheike, T., and Loft, S.: Ambient particle source apportionment and daily hospital admissions among children and elderly in Copenhagen, *J. Expo. Sci. Env. Epid.*, 17, 625–636, 2007.
- Andersen, Z. J., Wahlin, P., Raaschou-Nielsen, O., Ketzel, M., Scheike, T., and Loft, S.: Size distribution and total number concentration of ultrafine and accumulation mode particles and hospital admissions in children and the elderly in Copenhagen, Denmark, *Occup. Environ. Med.*, 65, 458–66, 2008.
- Arhami, M., Minguillon, M. C., Polidori, A., Schauer, J. J., Delfino, R. J., and Sioutas, C.: Organic compound characterization and source apportionment of indoor and outdoor quasi-ultrafine particulate matter in retirement homes of the Los Angeles Basin, *Indoor Air*, 20, 17–30, 2010.
- Atkinson, R. W., Fuller, G. W., Anderson, H. R., Harrison, R. M., and Armstrong, B.: Urban ambient particle metrics and health: a time-series analysis, *Epidemiology*, 21, 501–511, 2010.

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)




[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)


- Bell, M. L., Ebisu, K., Peng, R. D., Samet, J. M., and Dominici, F.: Hospital admissions and chemical composition of fine particle air pollution, *Am. J. Resp. Crit. Care.*, 179, 1115–1120, 2009.
- Berry, M. W., Browne, M., Langville, A. N., Pauca, V. P., and Plemmons, R. J.: Algorithms and applications for approximate nonnegative matrix factorization, *Comput. Stat. Data. An.*, 52, 155–173, 2007.
- Birmili, W. and Wiedensohler, A.: New particle formation in the continental boundary layer: meteorological and gas phase parameter influence, *Geophys. Res. Lett.*, 27, 3325–3328, 2000.
- Boutsidis, C. and Gallopoulos, E.: SVD based initialization: a head start for nonnegative matrix factorization, *Pattern Recogn.*, 41, 1350–1362, 2008.
- Branis, M., Safranek, J., and Hytychova, A.: Indoor and outdoor sources of size-resolved mass concentration of particulate matter in a school gym-implications for exposure of exercising children, *Environ. Sci. Pollut. R.*, 18, 598–609, 2011.
- Breitner, S., Liu, L., Cyrus, J., Bruske, I., Franck, U., Schlink, U., Leitte, A. M., Herbarth, O., Wiedensohler, A., Wehner, B., Hu, M., Pan, X. C., Wichmann, H. E., and Peters, A.: Sub-micrometer particulate air pollution and cardiovascular mortality in Beijing, China, *Sci. Total Environ.*, 409, 5196–5204, doi:10.1016/j.scitotenv.2011.08.023, 2011.
- Brook, R. D.: Cardiovascular effects of air pollution, *Clin. Sci.*, 115, 175–187, 2008.
- Brook, R. D., Rajagopalan, S., Pope, C. A., Brook, J. R., Bhatnagar, A., Diez-Roux, A. V., Holguin, F., Hong, Y. L., Luepker, R. V., Mittleman, M. A., Peters, A., Siscovick, D., Smith, S. C., Whitsel, L., and Kaufman, J. D.: Particulate matter air pollution and cardiovascular disease – an update to the scientific statement from the American Heart Association, *Circulation*, 121, 2331–2378, 2010.
- Buzorius, G., Hameri, K., Pekkanen, J., and Kulmala, M.: Spatial variation of aerosol number concentration in Helsinki city, *Atmos. Environ.*, 33, 553–565, 1999.
- Chalupa, D. C., Morrow, P. E., Oberdorster, G., Utell, M. J., and Frampton, M. W.: Ultrafine particle deposition in subjects with asthma, *Environ. Health Persp.*, 112, 879–882, 2004.
- Cheng, Y. F., Su, H., Rose, D., Gunthe, S. S., Berghof, M., Wehner, B., Achtert, P., Nowak, A., Takegawa, N., Kondo, Y., Shiraiwa, M., Gong, Y. G., Shao, M., Hu, M., Zhu, T., Zhang, Y. H., Wiedensohler, A., Andreae, M. O., and Pöschl, U.: Size-resolved measurement of the mixing state of soot in the megacity Beijing, China: diurnal cycle, aging and parameterization, *Atmos. Chem. Phys. Discuss.*, 11, 32161–32204, doi:10.5194/acpd-11-32161-2011, 2011.

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Costabile, F., Birmili, W., Klose, S., Tuch, T., Wehner, B., Wiedensohler, A., Franck, U., König, K., and Sonntag, A.: Spatio-temporal variability and principal components of the particle number size distribution in an urban atmosphere, *Atmos. Chem. Phys.*, 9, 3163–3195, doi:10.5194/acp-9-3163-2009, 2009.

5 Cyrus, J., Stolzel, M., Heinrich, J., Kreyling, W. G., Menzel, N., Wittmaack, K., Tuch, T., and Wichmann, H. E.: Elemental composition and sources of fine and ultrafine ambient particles in Erfurt, Germany, *Sci. Total Environ.*, 305, 143–156, 2003.

Cyrus, J., Pitz, M., Heinrich, J., Wichmann, H. E., and Peters, A.: Spatial and temporal variation of particle number concentration in Augsburg, Germany, *Sci. Total Environ.*, 401, 168–175, 2008.

10 Delfino, R. J., Sioutas, C., and Malik, S.: Potential role of ultrafine particles in associations between airborne particle mass and cardiovascular health, *Environ. Health Persp.*, 113, 934–946, 2005.

Duvall, R. M., Norris, G. A., Dailey, L. A., Burke, J. M., McGee, J. K., Gilmour, M. I., Gordon, T., and Devlin, R. B.: Source apportionment of particulate matter in the US and associations with lung inflammatory markers, *Inhal. Toxicol.*, 20, 671–683, 2008.

EU: Council Directive 1999/30/E C., Official Journal of the European Communities, L163, 41–60, 1999.

20 Franck, U., Herbarth, O., Roder, S., Schlink, U., Borte, M., Diez, U., Kramer, U., and Lehmann, I.: Respiratory effects of indoor particles in young children are size dependent, *Sci. Total Environ.*, 409, 1621–1631, 2011a.

Franck, U., Odeh, S., Wiedensohler, A., Wehner, B., and Herbarth, O.: The effect of particle size on cardiovascular disorders – the smaller the worse, *Sci. Total Environ.*, 409, 4217–4221, 2011b.

25 Gao, Y., Liu, X., Zhao, C., and Zhang, M.: Emission controls versus meteorological conditions in determining aerosol concentrations in Beijing during the 2008 Olympic Games, *Atmos. Chem. Phys.*, 11, 12437–12451, doi:10.5194/acp-11-12437-2011, 2011.

30 Gu, J., Pitz, M., Schnelle-Kreis, J., Diemer, J., Reller, A., Zimmermann, R., Soentgen, J., Stoelzel, M., Wichmann, H. E., Peters, A., and Cyrus, J.: Source apportionment of ambient particles: comparison of positive matrix factorization applied to particle size distribution and chemical composition data., *Atmos. Environ.*, 45, 1849–1857, 2011.

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)




[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)


Guo, Y., Jia, Y., Pan, X., Liu, L., and Wichmann, H. E.: The association between fine particulate air pollution and hospital emergency room visits for cardiovascular diseases in Beijing, China, *Sci. Total Environ.*, 407, 4826–4830, 2009.

Heinrich, J., Grote, V., Peters, A., and Wichmann, H. E.: Gesundheitliche Wirkungen von Feinstaub: Epidemiologie der Langzeiteffekte., *Umweltmed. Forsch. Prax.*, 7, 91–99, 2002.

Janhäll, S., Olofson, K. F. G., Andersson, P. U., Pettersson, J. B. C., and Hallquist, M.: Evolution of the urban aerosol during winter temperature inversion episodes, *Atmos. Environ.*, 40, 5355–5366, 2006.

Jolliffe, I. T.: *Principal Component Analysis*, Springer, New York, NY, 1986.

Krudysz, M., Moore, K., Geller, M., Sioutas, C., and Froines, J.: Intra-community spatial variability of particulate matter size distributions in Southern California/Los Angeles, *Atmos. Chem. Phys.*, 9, 1061–1075, doi:10.5194/acp-9-1061-2009, 2009.

Lee, D. D. and Seung, H. S.: Learning the parts of objects by non-negative matrix factorization, *Nature*, 401, 788–791, 1999.

Lee, D. D. and Seung, H. S.: Algorithms for Non-negative Matrix Factorization, *Adv. Neu. In.*, 13, 556–562, 2001.

Leitte, A. M., Petrescu, C., Franck, U., Richter, M., Suci, O., Ionovici, R., Herbarth, O., and Schlink, U.: Respiratory health, effects of ambient air pollution and its modification by air humidity in Drobeta-Turnu Severin, Romania, *Sci. Total Environ.*, 407, 4004–4011, 2009.

Leitte, A. M., Herbarth, O., Wichmann, H. E., Pan, X. C., Hu, M., Wiedensohler, A., Schlink, U., Wehner, B., Tuch, T., Wu, Z., Liu, L., Breitner, S., Cyrus, J., Peters, A., and Franck, U.: Modelling daily size-segregated particle number concentrations for a possible later use in epidemiological time series analyses – a case study for Beijing, submitted, 2011a.

Leitte, A. M., Schlink, U., Herbarth, O., Wiedensohler, A., Pan, X. C., Hu, M., Richter, M., Wehner, B., Tuch, T., Wu, Z., Yang, M., Liu, L., Breitner, S., Cyrus, J., Peters, A., Wichmann, H. E., and Franck, U.: Size-segregated particle number concentrations and respiratory emergency room visits in Beijing, China, *Environ. Health Persp.*, 119, 508–513, 2011b.

Leitte, A. M., Schlink, U., Herbarth, O., Wiedensohler, A., Pan, X. C., Hu, M., Wehner, B., Breitner, S., Peters, A., Wichmann, H. E., and Franck, U.: Associations between size-segregated particle number concentrations and respiratory mortality in Beijing, China, *Int. J. Environ. Heal. R.*, 22, 119–133, doi:10.1080/09603123.2011.605878, 2011c.

Lin, C. J.: Projected gradient methods for nonnegative matrix factorization, *Neural Comput.*, 19, 2756–2779, 2007.

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Liu, S., Hu, M., Wu, Z. J., Wehner, B., Wiedensohler, A., and Cheng, Y. F.: Aerosol number size distribution and new particle formation at a rural/coastal site in Pearl River Delta (PRD) of China, *Atmos. Environ.*, 42, 6275–6283, 2008.

Liu, S. H., Liu, Z. X., Li, J., Wang, Y. C., Ma, Y. J., Sheng, L., Liu, H. P., Liang, F. M., Xin, G. J., and Wang, J. H.: Numerical simulation for the coupling effect of local atmospheric circulations over the area of Beijing, Tianjin and Hebei Province, *Sci. China Ser. D*, 52, 382–292, 2009.

Meier, J., Wehner, B., Massling, A., Birmili, W., Nowak, A., Gnauk, T., Brüggemann, E., Herrmann, H., Min, H., and Wiedensohler, A.: Hygroscopic growth of urban aerosol particles in Beijing (China) during wintertime: a comparison of three experimental methods, *Atmos. Chem. Phys.*, 9, 6865–6880, doi:10.5194/acp-9-6865-2009, 2009.

Metzger, A., Verheggen, B., Dommen, J., Duplissy, J., Prevot, A. S. H., Weingartner, E., Riipinen, I., Kulmala, M., Spracklen, D. V., Carslaw, K. S., and Baltensperger, U.: Evidence for the role of organics in aerosol particle formation under atmospheric conditions, *Proc. Natl. Acad. Sci.*, 107, 6646–6651, 2010.

Ogulei, D., Hopke, P. K., Ferro, A. R., and Jaques, P. A.: Factor analysis of submicron particle size distributions near a major United States-Canada trade bridge, *J. Air Waste Manage.*, 57, 190–203, 2007.

Paatero, P.: Least squares formulation of robust non-negative factor analysis, *Chemometr. Intell. Lab.*, 37, 23–35, 1997.

Paatero, P. and Tapper, U.: Positive matrix factorization – a non-negative factor model with optimal utilization of error-estimates of data values, *Environmetrics*, 5, 111–126, 1994.

Pan, X. C., Yue, W., He, K. B., and Tong, S. L.: Health benefit evaluation of the energy use scenarios in Beijing, China, *Sci. Total Environ.*, 374, 242–251, 2007.

Penttinen, P., Timonen, K. L., Tiittanen, P., Mirme, A., Ruuskanen, J., and Pekkanen, J.: Number concentration and size of particles in urban air: effects on spirometric lung function in adult asthmatic subjects, *Environ. Health Persp.*, 109, 319–323, 2001.

Peters, A., Doring, A., Wichmann, H. E., and Koening, W.: Increased plasma viscosity during an air pollution episode: a link to mortality?, *Lancet*, 349, 1582–1587, 1997.

Peters, A., Heinrich, J., and Wichmann, H. E.: Gesundheitliche Wirkungen von Feinstaub: Epidemiologie der Kurzzeiteffekte, *Umweltmed. Forsch. Prax.*, 7, 101–115, 2002.

Pope, C. A. and Dockery, D. W.: Health effects of fine particulate air pollution: lines that connect, *J. Air Waste Manage.*, 56, 709–742, 2006.

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Pöschl, U.: Formation and decomposition of hazardous chemical components contained in atmospheric aerosol particles, *J. Aerosol Med.*, 15, 203–212, 2002.
- Puustinen, A., Hömeri, K., Pekkanen, J., Kulmala, M., de Hartog, J., Meliefste, K., ten Brink, H., Kos, G., Katsouyanni, K., Karakatsani, A., Kotronarou, A., Kavouras, I., Meddings, C., Thomas, S., Harrison, R., Ayres, J. G., van der Zee, S., and Hoek, G.: Spatial variation of particle number and mass over four European cities, *Atmos. Environ.*, 41, 6622–6636, 2007.
- Ren, Z. H., Wan, B. T., and Yu, T.: Influence of weather system of different scales on pollution boundary layer and the transport in horizontal current field, *Res. Environ. Sci.*, 17, 7–13, 2004.
- Rückerl, R., Schneider, A., Breitner, S., Cyrys, J., and Peters, A.: Health effects of particulate air pollution: a review of epidemiological evidence, *Inhal. Toxicol.*, 23, 555–592, 2011.
- Rundell, K. W. and Caviston, R.: Ultrafine and fine particulate matter inhalation decreases exercise performance in healthy subjects, *J. Strength Cond. Res.*, 22, 2–5, 2008.
- Schaefer, K., Emeis, S., Hoffmann, H., and Jahn, C.: Influence of mixing layer height upon air pollution in urban and sub-urban areas, *Meteorol. Z.*, 15, 647–658, 2006.
- Schlink, U. and Thiem, A.: Non-negative matrix factorization for the identification of patterns of atmospheric pressure and geopotential for the Northern Hemisphere, *Int. J. Climatol.*, 30, 909–925, 2010.
- Seibert, P., Beyrich, F., Gryning, S.-E., Joffre, S., Rasmussen, A., and Tercier, P.: Review and intercomparison of operational methods for the determination of the mixing height, *Atmos. Environ.*, 34, 1001–1027, 2000.
- Shi, Z. B., He, K. B., Yu, X. C., Yao, Z. L., Yang, F. M., Ma, Y. L., Ma, R., Jia, Y. T., and Zhang, J.: Diurnal variation of number concentration and size distribution of ultrafine particles in the urban atmosphere of Beijing in winter, *J. Environ. Sci.-China*, 19, 933–938, 2007.
- Silva, P. J., Vawdrey, E. L., Corbett, M., and Erupe, M.: Fine particle concentrations and composition during wintertime inversions in Logan, Utah, USA, *Atmos. Environ.*, 41, 5410–5422, 2007.
- Sun, Y. L., Zhuang, G. S., Wang, Y., and Han, L. H.: The air-borne particulate pollution in Beijing – concentration, composition, distribution and sources, *Atmos. Environ.*, 38, 5991–6004, 2004.
- Suwanwaphatthana, W., Ruangdej, K., and Turner-Henson, A.: Outdoor air pollution and children's health, *Pediatr. Nurs.*, 36, 25–32, 2010.

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Tuch, T. M., Herbarth, O., Franck, U., Peters, A., Wehner, B., Wiedensohler, A., and Heintzenberg, J.: Weak correlation of ultrafine aerosol particle concentrations < 800 nm between two sites within one city, *J. Expo. Sci. Env. Epid.*, 16, 486–490, 2006.
- Uhrner, U., Zallinger, M., von Lowis, S., Vehkamäki, H., Wehner, B., Stratmann, F., and Wiedensohler, A.: Volatile nanoparticle formation and growth within a diluting diesel car exhaust, *J. Air Waste Manage.*, 61, 399–408, 2011.
- Wang, X. Y., Hu, W., and Tong, S.: Long-term exposure to gaseous air pollutants and cardio-respiratory mortality in Brisbane, Australia, *Geospatial Health*, 3, 257–263, 2009.
- Wehner, B. and Wiedensohler, A.: Long term measurements of submicrometer urban aerosols: statistical analysis for correlations with meteorological conditions and trace gases, *Atmos. Chem. Phys.*, 3, 867–879, doi:10.5194/acp-3-867-2003, 2003.
- Wehner, B., Wiedensohler, A., Tuch, T. M., Wu, Z. J., Hu, M., and Slanina, J.: Variability of the aerosol number size distribution in Beijing, China: new particle formation, dust storms, and high continental background, *Geophys. Res. Lett.*, 31, L22108, doi:10.1029/2004GL021596, 2004.
- Wehner, B., Birmili, W., Ditas, F., Wu, Z., Hu, M., Liu, X., Mao, J., Sugimoto, N., and Wiedensohler, A.: Relationships between submicrometer particulate air pollution and air mass history in Beijing, China, 2004–2006, *Atmos. Chem. Phys.*, 8, 6155–6168, doi:10.5194/acp-8-6155-2008, 2008.
- Wichmann, H. E., Spix, C., Tuch, T., Wolke, G., Peters, A., Heinrich, J., Kreyling, W. G., and Heyder, J.: Daily mortality and fine and ultrafine particles in Erfurt, Germany – Part I: role of particle number and particle mass, *Res. Rep. Health Eff. Inst.*, 98, 5–86, 2000.
- Wu, Z. J., Hu, M., Liu, S., Wehner, B., Bauer, S., Maßling, A., Wiedensohler, A., Petaja, T., Dal Maso, M., and Kulmala, M.: New particle formation in Beijing, China: statistical analysis of a 1-year data set, *J. Geophys. Res.-Atmos.*, 112, D09209 doi:10.1029/2006JD007406, 2007.
- Wu, Z. J., Hu, M., Lin, P., Liu, S., Wehner, B., and Wiedensohler, A.: Particle number size distribution in the urban atmosphere of Beijing, China, *Atmos. Environ.*, 42, 7967–7980, 2008.
- Yao, X. H., Chan, C. K., Fang, M., Cadle, S., Chan, T., Mulawa, P., He, K. B., and Ye, B. M.: The water-soluble ionic composition of PM_{2.5} in Shanghai and Beijing, China, *Atmos. Environ.*, 36, 4223–4234, 2002.

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Yu, Y., Schleicher, N., Norra, S., Fricker, M., Dietze, V., Kaminski, U., Cen, K. A., and Stuben, D.: Dynamics and origin of PM_{2.5} during a three-year sampling period in Beijing, China, *J. Environ. Monitor.*, 13, 334–346, 2011.

5 Yue, D., Hu, M., Wu, Z., Wang, Z., Guo, S., Wehner, B., Nowak, A., Achtert, P., Wiedensohler, A., Jung, J., Kim, Y., and Liu, S.: Characteristics of aerosol size distributions and new particle formation in the summer in Beijing, *J. Geophys. Res.*, 114, 1–3, 2009.

10 Yue, W., Stolzel, M., Cyrys, J., Pitz, M., Heinrich, J., Kreyling, W. G., Wichmann, H. E., Peters, A., Wang, S., and Hopke, P. K.: Source apportionment of ambient fine particle size distribution using positive matrix factorization in Erfurt, Germany, *Sci. Total Environ.*, 398, 133–144, 2008.

Zhang, M., Song, Y., and Cai, X.: A health-based assessment of particulate air pollution in urban areas of Beijing in 2000–2004, *Sci. Total Environ.*, 376, 100–108, 2007.

Zhou, L. M., Kim, E., Hopke, P. K., Stanier, C. O., and Pandis, S.: Advanced factor analysis on Pittsburgh particle size-distribution data, *Aerosol Sci. Tech.*, 38, 118–132, 2004.

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Table 1. 5 NMF-N factors for daily particle number size distributions and assigned sources.

NMF factor	diameter range	physical pattern	assigned source	median weight
NMF-N1	10 ... 10 000 nm, full diameter range	basic air pollution level	traffic background and industrial background, fuel combustion	0.32
NMF-N2	10 ... 300 nm (UFP and FP)	increasing peak of number size distribution during daytime, elevated concentrations in the evening	traffic emissions and particle growing processes	0.22
NMF-N3	morning: 10 ... 1000 nm, afternoon: 3 ... 100 nm	ordinary distribution in the morning, replaced by increasing UFP levels in the afternoon	industrial emissions from southwards directions, mountain and valley breeze system	0.13
NMF-N4	10 ... 100 nm (UFP)	elevated UFP level, increasing peak of number size distribution in the course of the day	particle nucleation and particle growth during summer months due to specific meteorological conditions (hygroscopic growth)	0.09
NMF-N5	3 ... 70 nm (UFP)	extremely elevated UFP level in the morning until noon	primary aerosol from domestic heating and particle nucleation from precursor gases during winter months fostered by specific meteorological conditions (temperature inversion)	0.01

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

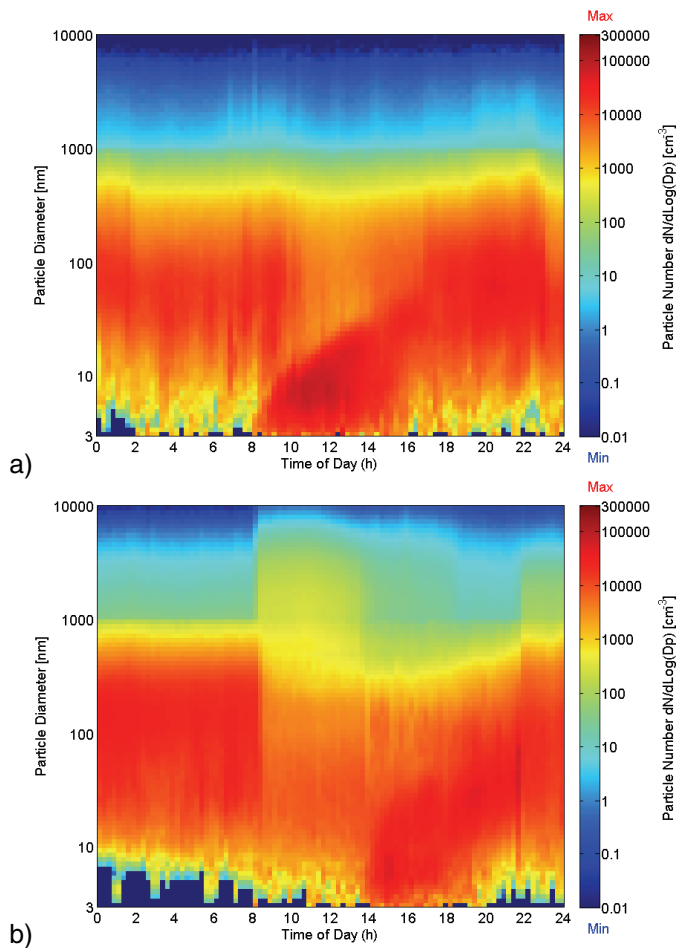


Fig. 1. Diurnal particle number size distribution patterns for a typical non-dust storm day on 13 February 2008 **(a)** and a typical dust storm day on 18 March 2008 **(b)**.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

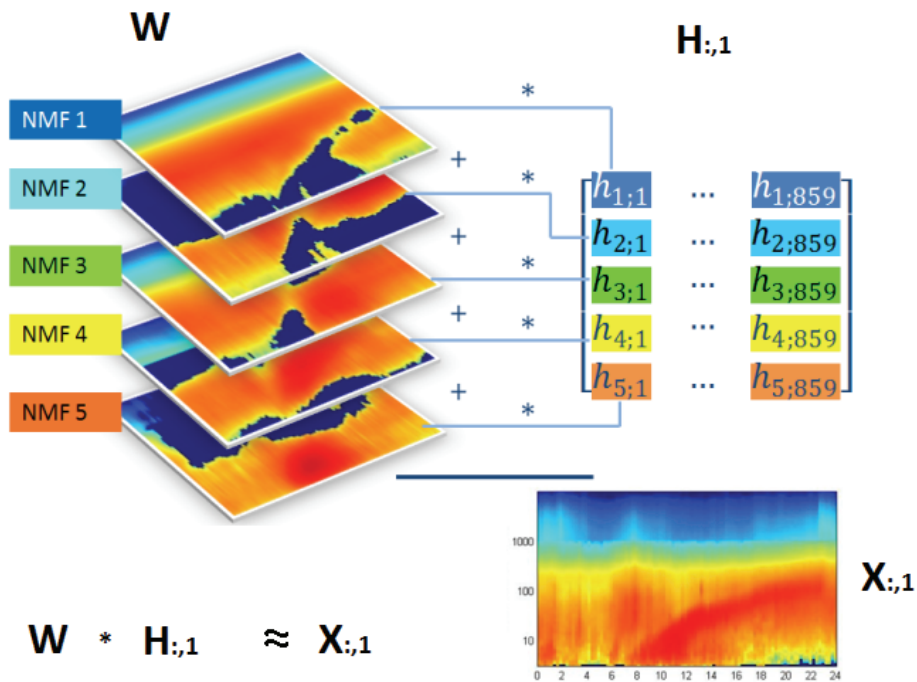


Fig. 2. Linear combination of NMF factors for approximation of a diurnal particle burden pattern of particle number N .

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

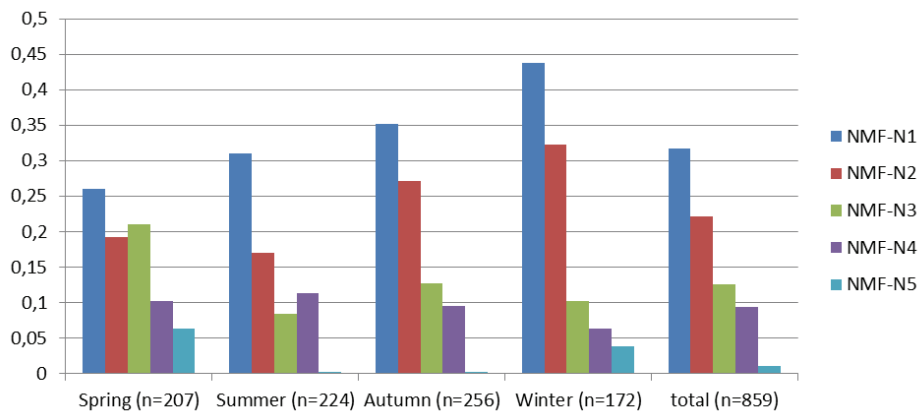


Fig. 3. Medians of the time-dependent coefficients for NMF-N factors 1 to 5 for spring, summer, autumn, winter and all days in the period 2004–2008.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

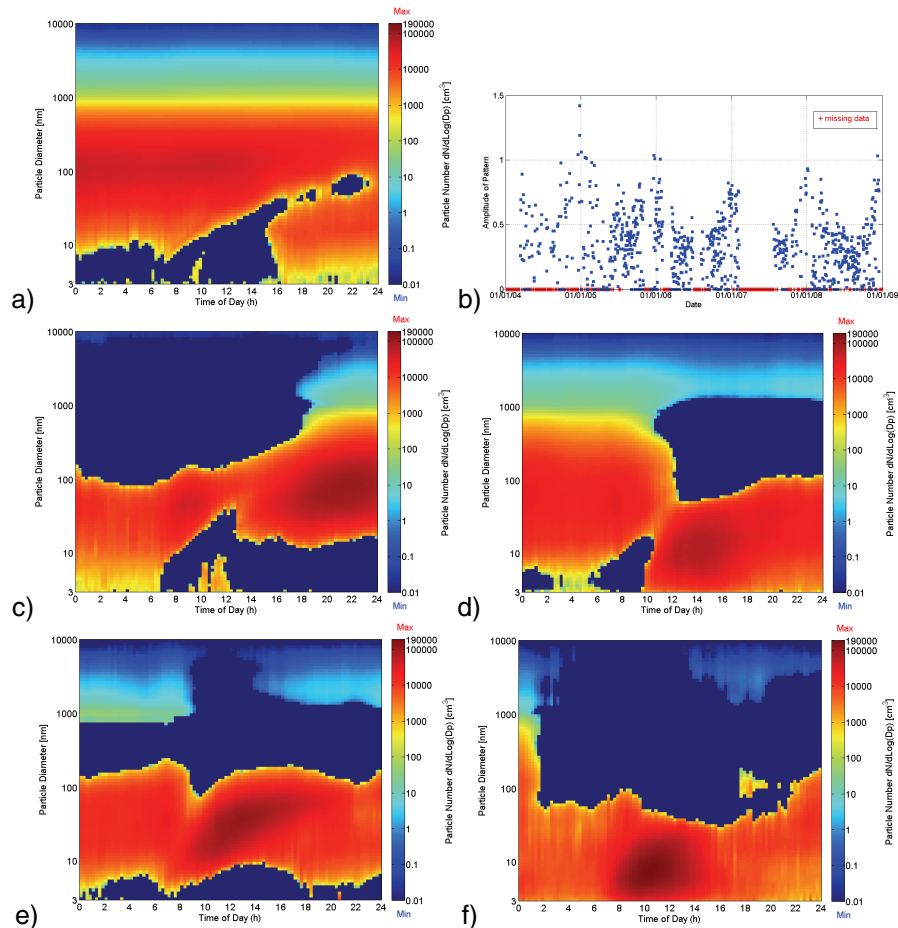


Fig. 4. Factor NMF-N1 (a), its corresponding time-varying coefficients (b), factors NMF-N2 (c), NMF-N3 (d), NMF-N4 (e) and NMF-N5 (f) for particle number size data for the investigation period 2004–2008 in Beijing.

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

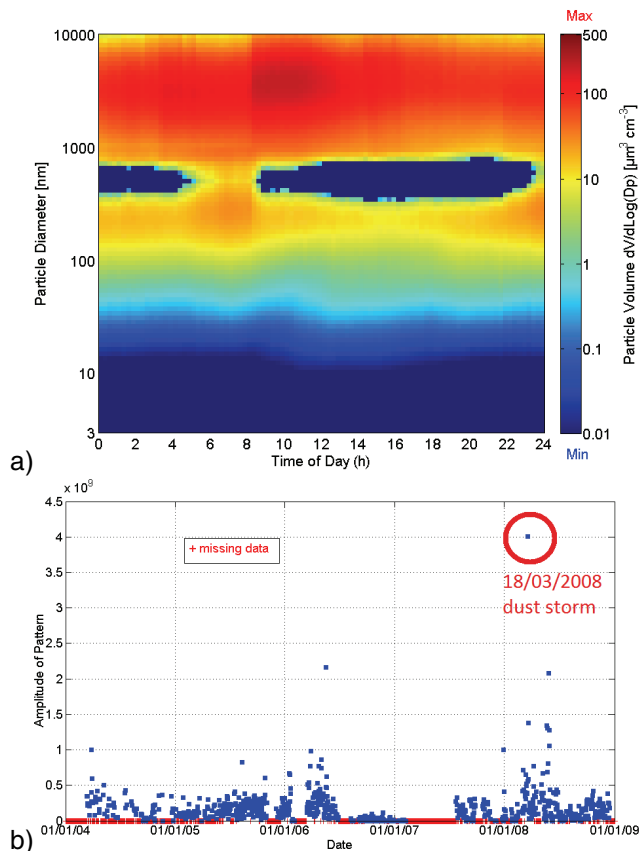


Fig. 5. NMF-V4 (a) and its time-dependent coefficients (b) for the investigation period 2004–2008 in Beijing.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Non-negative matrix factorization of particulate air pollution in Beijing

A. Thiem et al.

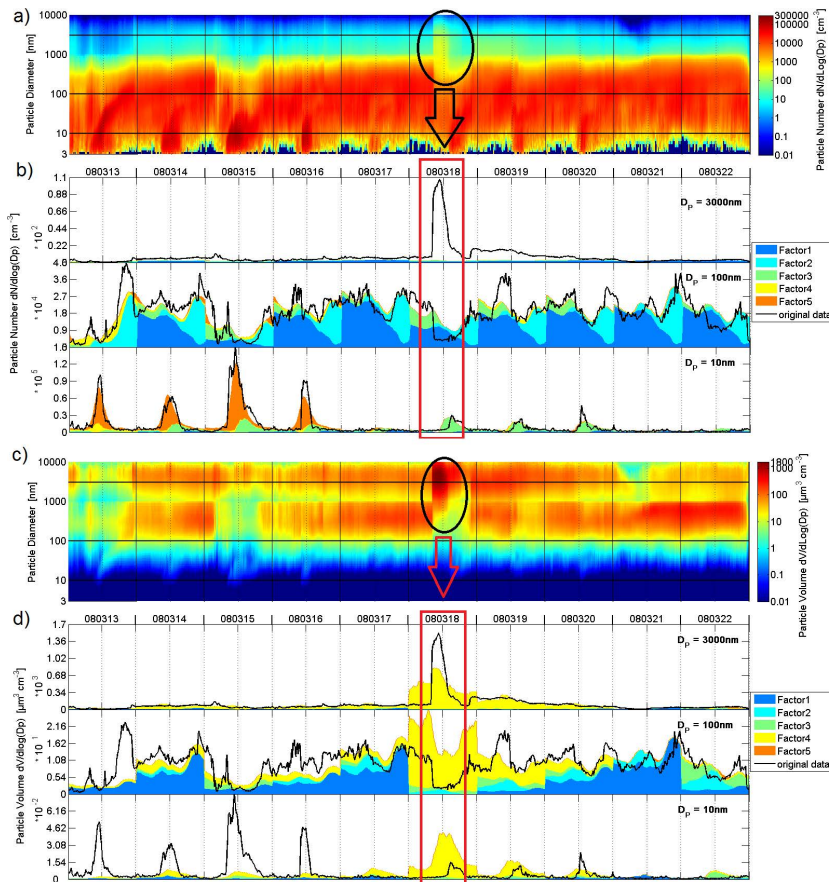


Fig. 6. Particle number **(a)** and volume size distribution **(c)** for a 10-day period (13 March 2008 to 22 March 2008) and reconstruction of particle number and volume data by NMF-N **(b)** and NMF-V superposition **(d)** for different diameters $D_p = 10$ nm, 100 nm, 3000 nm.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion