



HAL
open science

Non-negative matrix factorization for the analysis of particle number concentrations: Characterization of the temporal variability of sources in indoor workplace

Rachid Ouaret, Anda Ionescu, Olivier Ramalho

► To cite this version:

Rachid Ouaret, Anda Ionescu, Olivier Ramalho. Non-negative matrix factorization for the analysis of particle number concentrations: Characterization of the temporal variability of sources in indoor workplace. *Building and Environment*, 2021, 203, pp.108055. 10.1016/j.buildenv.2021.108055 . hal-04320146

HAL Id: hal-04320146

<https://hal.u-pec.fr/hal-04320146>

Submitted on 4 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Non-negative Matrix Factorization for the analysis of particle number
2 concentrations: characterization of the temporal variability of sources
3 in indoor workplace.

4 R. Ouaret^{a,*}, A. Ionescu^b, O. Ramalho^c

5 ^aLaboratoire de Génie Chimique, Université de Toulouse, CNRS, INPT, UPS, 4 Allée Emile Monso, 31432
6 Toulouse, Cedex 4, France

7 ^bUniversity Paris-Est, Center for Study and Research on Thermics, Environment and Systems (CERTES,
8 EA3481), University Paris-Est, 61, avenue du Général de Gaulle, Créteil Cedex, 94010, France

9 ^cFrench Indoor Air Quality Observatory, Scientific and Technical Center for Building (CSTB), University
10 Paris-Est, 84 Avenue Jean Jaurès, Champs-sur-Marne, 77420, France

11 **Abstract**

12 The temporal variability of indoor Particle Number (PN) concentrations, their determinants and their relative con-
13 tributions in an occupied workspace were investigated. The presented study is based on the receptor modeling
14 approach, focusing on Non-negative Matrix Factorization (NMF) to provide new insights on the source time vari-
15 ability. Continuous size distribution from 0.3 μm to 20 μm were collected with a short time step sampling (1 min)
16 over six months in 2015. The measurements were made inside and outside an open-plan office occupied by 6-8
17 persons. NMF distinguished five major patterns obtained from PN concentrations time series. The apportionment
18 results were expressed as source diurnal profiles and strengths by relating the obtained source contributions to the
19 source information provided by the office occupancy and natural ventilation (the effect of opening windows). Factor
20 2 contributes to 75% of the total contributions for finer size fraction ($< 0.5 \mu\text{m}$). Combining NMF results with
21 indoor occupancy and windows states successfully demarcated the main sources of fluctuation. The diurnal profiles
22 of the third factor (F3) and $\text{PN}_{0.9-1.8}$ concentrations time series are very similar ($r = 0.95$). The diurnal variation of
23 factor 1 is very similar to that observed for CO_2 variations and $\text{PN}_{6.25-12.5}$ time series. Coarse particles ($> 17.5 \mu\text{m}$)
24 are associated with the 4th factor. The latter does not contribute to any of the other particle ranges. The NMF
25 factors interpretation was supported by correlation analysis and statistical tests, as well as by temporal variation
26 comparison.

27 *Keywords:* Non-negative Matrix Factorization (NMF), Temporal source apportionment, Particle
28 Number (PN) concentration, Open-plan Office.

29 **1. Introduction**

30 Nowadays, it is becoming increasingly evident that indoor environments play a critical role to
31 understand and assess total human exposure to air pollution. Indoor air quality became a matter
32 of particular interest for the following three main reasons: (i) people spend about 85% of their
33 time indoors, (ii) indoor pollutant concentrations can be significantly higher from those outdoors

*Corresponding author
Email address: rachid.ouaret@toulouse-inp.fr (R. Ouaret)
Preprint submitted to Elsevier

34 and (*iii*) many potentially health-hazardous pollutants are emitted from indoor sources [40]. The
35 assessment of the indoor source variability is now necessary for the design and implementation
36 of effective control strategies. Airborne Particulate Matter (PM) is one of the major types of
37 contaminants in indoor air due to their ubiquitous occurrence and toxicity [42, 38]. Particulate
38 matter is a mixture of many different chemical species and the assessment of human exposure to
39 PM requires some knowledge about the sources, and the determination of the time variability of
40 the PM time series.

41 PM represent a complex mixture of organic and inorganic species which vary in size, composition,
42 and origin. Regarding particle size, PM is usually classified according their aerodynamic diameter
43 such as coarse (2.5 μm -10 μm), fine (<2.5 μm), and ultra-fine particles (< 0.1 μm) [12]. Since differ-
44 ent particles originate from different sources and their concentration in a given place is influenced
45 by different factors, scientific knowledge about the occurrence, strength, and temporal variability
46 of the sources is required to reduce the indoor exposure impact. Several studies focused the extent
47 to which human exposure to outdoor PM occurs indoors [64, 41]. From outdoors to indoors, PM
48 enters by infiltrating through cracks and gaps in the building envelope, via natural or mechanical
49 ventilation [11, 26]. Consequently, the temporal patterns of indoor PM concentrations are resulting
50 from both indoor and outdoor sources variability which constantly varies over time. In addition, the
51 PM temporal patterns are influenced by many factors: (*i*) penetration and ventilation efficiency, (*ii*)
52 indoor occupancy and occupants' activity, and (*iii*) building's volume and interaction with surfaces.

53 Important indoor sources of PM include, but not limited to, household chores (combustion, candles,
54 and cooking, ...) [1, 26] and different human activities [18]. In indoor offices, particles also can
55 be generated from some equipment, such as copier and printers machines, computers, and other
56 electronic devices [39, 9, 56]. Furthermore, many studies claim that particle re-suspension from
57 walking is an important indoor source of PM [53, 57]. The source emissions and sinks are generally
58 variable and depend on many environmental factors, such as temperature, humidity, and air ex-
59 change rates. Particle deposition and resuspension within the microenvironments are mechanisms
60 that can extremely change the time variability of the indoor PM concentration. Overall, as detailed

61 in [40], three main factors determine indoor pollutant variations in occupied spaces: (i) properties
62 of pollutants, (ii) occupants' behavior and (iii) building characteristics. In the case of the testing
63 chambers or experimental automated houses, these parameters are usually well defined [68, 61].

64 To the best of our knowledge there are few studies in the literature focusing on the impact of the
65 occupancy and windows state in offices for a long period with a short time step.

66 The originality of the study presented in this paper is due to the following features: it is based
67 on a long-term monitoring campaign in real conditions; parameters are monitored with a short
68 time step; it contributes to indoor source identification taking into account the occupancy and the
69 window opening state.

70 The objective of this work is to get a better understanding of the temporal variability of PM's and
71 their determinants, *i.e.* the influencing factors, as well as their relative contribution to PM indoor
72 concentrations. The measurement campaign was carried out in an occupied open-plan office.

73 By retracing pollutants to their origins, emission sources variability can be characterized. The
74 source identification and the assessment of their relative contribution to total indoor exposures can
75 provide valuable information for reducing their emission with the aim to protect human health.

76 The practice of deriving information about pollution sources and the amount they emit is called
77 source apportionment [5].

78 In principle, there are two basic approaches used in the environmental field to perform a source
79 apportionment analysis: source-oriented models (dispersion model) and receptor-oriented models.
80 Dispersion models require the knowledge of emission rates and dispersion factors together with local
81 topography and meteorology for the estimation of source impacts [16, 22]. It is a direct modeling
82 (from sources to receptors). By contrast, receptor modeling, which is an inverse one (from receptor
83 to sources), is based on the mass conservation and it can be used to identify and apportion sources
84 of contaminants in the air [23]. The main idea in the latter is to solve a mass balance equation
85 using multivariate factor analysis. Thus, the receptor models are used to estimate the contribution
86 of different sources to ambient PM concentrations based on PM measurements and subsequent
87 chemical analysis [49]. Exogenous variables can be used for interpretation.

88 Indoor environments are modeled sometimes by means of experimental chambers, where climatic
89 parameters are controlled. In this case, physical models can be developed easier than for the real
90 environments. Indeed, in the case of the real buildings the development of such models is more
91 complicated for the following reasons: difficulty in measuring source emissions [58]; unknown or
92 changing airflow patterns [33, 10]; difficulty to model room configuration and lack of information
93 about occupancy and its impact on the indoor environment.

94 The advantage of receptor models in occupied spaces is that this type of models doesn't require all
95 the information mentioned previously, as input. They are based on the decomposition of measured
96 concentration signals without predetermining the pollutant transformation, transportation, and
97 sink processes [63].

98 In recent years, there have been tremendous advances in the development of source apportionment
99 techniques based on statistical analysis, particularly the use of Factor Analysis (FA) for temporally
100 varying constituent concentrations [5, 22, 23, 67]. All these techniques are based on the factor-
101 ization of the initial matrix (database of concentrations in our case) in two matrices: matrix of
102 source profiles and matrix of source contributions, under the constraint of the RMSE minimiza-
103 tion and by imposing different constraints, such as: decorrelation (Principal Component Analysis),
104 statistical independence (Independent Component Analysis), non-negativity (Non-negative Matrix
105 Factorization and Positive Matrix Factorization).

106 Positive Matrix Factorization (PMF), developed by Paatero et al. 1994 [47] is widely used for source
107 apportionment in environmental applications such as atmospheric pollution. This method imposes
108 positivity constraints in the factor computational process by calculating dominant positive factors
109 based on measurements. The factorization is achieved without detailed prior knowledge of source
110 profiles or chemical fingerprints. In contrast to many other linear factorizations such as (PCA)
111 and Independent Component Analysis (ICA), Non-negative Matrix Factorization (NMF) [31, 32],
112 as well as PMF, makes positive latent structure explicit.

113 The Principal Component Analysis (PCA) method is one of the most used tools in factorial methods.
114 In indoor environment, we can find hybridization with other linear methods (regressions, ...) to

115 quantify the factors [3].

116 In the same vein, we have used NMF and simulated several factorizations, which are obtained by
117 a group of initialization-computation algorithms. For this reason, the different extensions on NMF
118 comprising the PMF method which is considered as a special case [59].

119 The NMF factors were analyzed and connected with their influencing parameters, such as opening
120 windows and occupancy rate of the open-plan office. In other words, we identify patterns that
121 shared by several sources. To date, few research studies have been conducted to identify temporal
122 patterns of different sources. Most of them focus on the chemical fingerprint. The contribution of
123 this study lies in the characterization of the time variability of the most frequent source-patterns
124 in an occupied open office. The interpretation of NMF factors and time variability patterns were
125 supported by correlation analysis as well as by temporal variation visualization.

126 **2. Materials and Methods**

127 *2.1. Materials*

128 *2.1.1. PN concentration monitoring*

129 Particle Number concentrations (PN) were acquired continuously using an optical particle counter,
130 a GRIMM 1.108 Dust Monitor (Grimm Technologies, Inc., Douglasville, GA, USA). This monitor
131 counts the number of particles with a diameter within the range of 0.3 to 20 μm , and more in
132 detail in 15 different size channels limited by: 0.3, 0.4, 0.5, 0.65, 0.8, 1.0, 1.6, 2.0, 3.0, 4.0, 5.0,
133 7.5, 10, 15 and 20 μm diameters (Figure 1, sensor's location). The flow rate of the instrument is
134 1.2 L/min⁻¹. It can count particles up to 2000 particles.cm⁻³ without coincidence effects with a
135 sensitivity of 0.001 particles.cm³ and a reproducibility of 2%. The optical counter uses two laser
136 powers to perform its measurements. Between 0.3 and 2 μm , the high laser power is used. Between
137 2 and 20 μm , the lowest laser power is used. The measurement at 2 μm is carried out twice at high
138 and low laser power, the final measurement is the average of the two values obtained. During the

139 calibration of each device, the maximum error tolerated is 10% between 0.3 and 2 μm and 20%
140 between 2 and 20 μm .

141 The period of investigation covers 6 months with a measurement interval of 1 minute. The indoor
142 PN concentrations were monitored in an open-plan office occupied by 6-8 persons. Outdoor PN con-
143 centrations were collected with the same type of instrument, which housed in a dedicated enclosure
144 outdoors on the roof of the building. Particle number concentrations are recorded every minute in
145 a memory card with an autonomy of more than 45 days and the data is recovered (approximately)
146 every 15 days.

147 The monitoring period runs from January, 1st, 2015, to June, 30th, 2015 representing 4 344 hours.
148 In our experiment, the database called later raw data matrix \mathbf{X} , consists of the number of particles
149 for the 15 fractions (size ranges) (15 variables $\#.\text{cm}^{-3}$) and 260580 rows (minutes). In the case of
150 this study, there are 485 missing values scattered (randomly). So, about 0.0018% of missing values
151 have been detected. The missing data are imputed using interpolation of the available data or by
152 simple calculation of the medians if there are less than 5 consecutive missing values. For particle
153 measurements, we rarely find more than 30 consecutive minutes of missing values.

154 2.1.2. Climatic data, occupancy and window opening state

155 Meteorological, windows opening states and occupancy parameters were collected simultaneously
156 in the same open-plan office.

157 The presence of occupants is estimated using motion detectors, while the state (opened/closed)
158 of the windows and doors are recorded by contactors. The measuring device has three motion
159 detectors, two contactors on two doors, and five contactors on windows facing exterior environment.

160 Motion detectors are infrared passive sensors able to detect motion in a half-spherical field (cf.
161 section 2.1.3). The door and windows contactors are soft-leaf switches that record a binary signal
162 at each state change: *opened* \rightarrow *closed* or *closed* \rightarrow *opened*. We retrieve the amount of movement
163 which tells us about occupancy, but we do not measure the number of occupants per minute. More
164 specifically, when no movement of occupants is detected, no data is transmitted to the *CSTBox*

165 (Telemonitoring Box manufactured by CSTB). As soon as motion (of occupants) is detected, the
166 information quantity is recorded during 10 seconds and a frame containing an array of 10 samples
167 is sent to the *CSTBox*. From the recorded raw data of the motion quantity variations, data is
168 transformed and re-coded to binary time series with 1-minute time step.

169 It should be noted that motion detection greatly underestimates the actual occupancy of the office
170 space. The situation where the occupant is static is considered to be a vacancy state. This would
171 reflect a high vacancy rate over the measurement period ($\approx 94\%$).

172 *2.1.3. The open-plan environment*

173 The concentration of indoor particles is highly variable and indoor-specific. The measurements
174 were performed in a building (Scientific and Technical Center for Building ([CSTB](#)), cf. Figure A.1
175 in supplementary material) located in a suburban area, at 30 km East of Paris. The measurement
176 campaign was conducted in an open-plan office space with a total area of 132 m^2 and a volume of
177 364 m^3 . Figure 1 shows the plan of the open space located at the 2^{nd} floor. A virtual tour of the
178 open space office is given in the supplementary material of this article.



Figure 1: Plan of the office space (132 m^2 , 364 m^3). The configuration of the tables varies with the number of occupants. An example of motion detection is materialized by two pink half spheres.

179

180 This office has a permanent mechanical exhaust ventilation. A single flow ventilation system provides a constant air extraction rate of $228 \text{ m}^3 \cdot \text{h}^{-1}$ (measured in 2014 at $\pm 6\%$). The six extract units are located by a black cross on the Figure 1. On the opposite side, 10 air inlets are attached to the joinery of the 5 sliding windows. as described in [43]. In addition, natural ventilation is possible due to the windows. Consequently, the effective ventilation rate greatly depends on the opening state of the windows. The actions made by the occupants on the windows (opening / closing) depend on the indoor comfort as well as on the individual sensitivity of the occupants [62, 21]. These actions on windows together with the other environmental factors provide different time-scale variations that are difficult to model by pure physical laws.

189 The indoor materials were carpets on the floor, painted walls, and ceiling tiles. The furniture comprises typical L-shaped desks melamine-faced particleboard and aluminium closets. A laser

191 multi-function copier was in use in the office plan. No specific major sources of particles, such as
192 combustion, were identified in the office.

193 *2.2. Methods*

194 *2.2.1. Related work*

195 The number of existing receptor models applied to the environmental field is relatively large and
196 it includes methods such as Principal Component Analysis [60, 20], Positive Matrix Factorization
197 (PMF) [48, 47], Independent Component Analysis (ICA) [4] and recently Non-negative Matrix
198 Factorization (NMF) in outdoor [17, 59, 27, 28, 37] and indoor [55, 46, 44, 45] environment.

199 Compared to the PCA and ICA, the PMF and NMF methods have the advantage of more realistic
200 non-negative constraint on factor profiles and contributions. Following this criterion, this study
201 used non-negativity for indoor source apportionment and particularly employed NMF that offers a
202 wide range of algorithms and extensions compared to PMF.

203 NMF and its generalizations have been used for different purposes such as dimensionality reduction,
204 feature extraction, clustering, blind source separation (BSS), and classification. In this paper, NMF
205 was used only for BSS purposes.

206 In the environmental field, NMF is a new method of a wide range of receptor modeling. It is used
207 to analyze the series of chemical concentration measurements and to find underlying explanatory
208 sources [59, 28]. Some research papers proposed an extension of the standard NMF form to incor-
209 porate some physical proprieties. Limem et al. (2013) proposed an informed NMF with a specific
210 parametrization which involves constraints about some known components of the factorization [35].
211 Plouvin et al. (2014) extended the latter work by adding some information provided by a phys-
212 ical dispersion model [50]. Limem et al. (2014) introduced the use of basic equality constraints
213 and have derived theoretical expressions of constrained Weighted NMF (WNMF) to characterize
214 industrial source apportionment of PM₁₀ [34]. To take into consideration both constraints simul-
215 taneously, in the previous works [35, 34] a new parametrization was proposed by incorporating a
216 new unconstrained matrix [13]. The update rules in [13] are based on the framework of the Split

217 Gradient Method of Lantéri *et al.* [29]. To identify distinctive PM₁₀ patterns across Europe on the
218 airborne data obtained from 1097 monitoring stations for 3 years, Žibert *et.al* [70] used NMF and
219 the autocorrelation function (ACF) to enhance the interpretation of factors features. In the indoor
220 air quality (IAQ) field, the positive (non-negative) matrix factorization is not such popular as for
221 outdoor air quality. We find only one study assessing the use of NMF in IAQ, which is conducted by
222 Rösch *et al.* in 2014 [55]. Also in the aircraft cabins context, PMF coupled with information related
223 to VOC sources have been applied to identify the major VOC sources [65]. The study pointed out
224 the importance of service and humans as major source (29%) of the total VOC emissions. Overall,
225 there is not sufficient research to provide meaningful information on indoor source time variability.
226 As the IAQ continues to face the realities of climate change [40], our understanding of temporal
227 patterns of exposure is crucial and this research gap requires further investigation. Therefore, this
228 study aims to address the elements of the research gap in indoor source identification. This paper
229 is an extended version of the works presented in [44], [46] as well as in [45].

230 2.2.2. Non-negative Matrix Factorization

231 Non-negative matrix factorization (NMF) is a multivariate data analysis technique which is aimed
232 to estimate physically meaningful latent components from non-negative data. Mathematically, the
233 factorization is carried out by a linear superposition of non-negative basis components and non-
234 negative weights.

235 NMF was initially introduced by Paatero’s works [48, 47], which refer to the problem as Positive
236 Matrix Factorization (PMF) that corresponds to the mass-balance model. The NMF has gained
237 popularity by the works of Lee and Seung who presented multiplicative update algorithms for com-
238 puting the NMF to optimize a cost function based on either a Euclidean distance measure or a
239 generalized Kullback–Leibler divergence [30, 31, 32]. A multitude of NMF variants and generaliza-
240 tions is summarised in Cichocki’s concise lecture note [15].

241 In this paper, a matrix is denoted with an uppercase bold letter, *e.g.*, \mathbf{X} , its elements with the
242 corresponding lowercase letter, *e.g.*, x_{it} , and a column vector in lowercase boldface, *e.g.*, \mathbf{x}_i .

243 Given an input non-negative raw data matrix $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_T] \in \mathbb{R}_+^{I \times T}$ and a positive integer
 244 reduced rank J , ($J \leq \min(I, T)$), the non-negative matrix factorization problem consists in finding
 245 two non-negative matrices $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_J] \in \mathbb{R}_+^{I \times J}$ and an encoding matrix $\mathbf{H} = h_{jt} \in \mathbb{R}_+^{J \times T}$
 246 that approximate \mathbf{X} , i.e.

$$\mathbf{X} \approx \mathbf{W}\mathbf{H} = \mathbf{x}_t \approx \sum_{j=1}^J \mathbf{w}_j h_{jt}. \quad (1)$$

247 Depending on the application field, these factors \mathbf{W} and \mathbf{H} are interpreted differently. In environ-
 248 mental source apportionment, \mathbf{H} plays the role of mixing matrix (weights) which represents the
 249 source contributions, while \mathbf{W} expresses temporal factor profiles (time series, see Appendix C).
 250 Thus, if a weight h_{jt} in a column of \mathbf{H} is high, then the corresponding basis vector \mathbf{w}_j is very
 251 important in approximating \mathbf{x}_t . Geometrically the basis vectors generate a simplicial cone and the
 252 columns of the matrix \mathbf{W} are basis vectors spanning a subspace in $J \leq I$. Once estimated, the \mathbf{H}
 253 and \mathbf{W} matrices will be presented in sections 3.2.1 and 3.2.2, respectively. Indeed, the \mathbf{H} matrix
 254 will allow us to analyze the contribution of the factors and the \mathbf{W} matrix to visualize the diurnal
 255 profiles of the factors (time series).

256 The time series of the sources and their weights are calculated iteratively by minimizing a suitable
 257 measure f for the divergence between \mathbf{W} and \mathbf{H} :

$$\arg \min_{\mathbf{W}, \mathbf{H} \geq 0} f(\mathbf{W}, \mathbf{H}) = \arg \min_{\mathbf{W}, \mathbf{H} \geq 0} [\mathcal{D}(\mathbf{X} \parallel \mathbf{W}\mathbf{H}) + \mathcal{R}(\mathbf{W}, \mathbf{H})], \quad (2)$$

258 where $\mathcal{D} : \mathbb{R}_+^{I \times J} \times \mathbb{R}_+^{J \times T} \rightarrow \mathbb{R}_+$ is a loss function and \mathcal{R} is an optional regularization function that
 259 enforce desirable properties (e.g. smoothness, sparsity, ...) on matrices \mathbf{W} and \mathbf{H} [15]. In this
 260 study, only $\arg \min_{\mathbf{W}, \mathbf{H} \geq 0} [\mathcal{D}(\mathbf{X} \parallel \mathbf{W}\mathbf{H})]$ is optimized. The simplest loss function measure is based on
 261 the Frobenius norm:

$$\mathcal{D}_F(\mathbf{X} \parallel \mathbf{W}\mathbf{H}) = \frac{1}{2} \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2 = \frac{1}{2} \sum_i^I \sum_t^T (\mathbf{X}_{it} - (\mathbf{W}\mathbf{H})_{it})^2. \quad (3)$$

262 The Frobenius similarity measure is a special case of the so-called β -divergence [15]. In this study
 263 we will consider only Kullback-Leibler divergence in case of Brunet algorithm [8] as follows:

$$\mathcal{D}_{KL}(\mathbf{X} \parallel \mathbf{WH}) = \sum_i^I \sum_t^T \mathbf{X}_{it} \ln \left(\frac{\mathbf{X}_{it}}{(\mathbf{WH})_{it}} \right) - \mathbf{X}_{it} + (\mathbf{WH})_{it}. \quad (4)$$

264 *2.2.3. Algorithms for solving the NMF problem*

265 Many numerical algorithms have been developed to solve the NMF problem expressed in the equa-
 266 tion (2). They can be divided into three general classes: (i) Alternating Least Squares (ALS)
 267 algorithms, (ii) multiplicative update algorithms, and (iii) gradient descent algorithms [6, 15]. The
 268 ALS algorithm computes the optimal solution of the unconstrained least squares problem, then it
 269 optimizes alternatively over one of the two factors \mathbf{W} or \mathbf{H} while keeping the other fixed. This
 270 subproblem is then reduced to one-factor convex.

271 By minimizing two criteria the squared Euclidean distance (or equivalently the squared Frobenius
 272 norm \mathcal{D}_F) and the generalized Kullback-Leibler divergence \mathcal{D}_{KL} , Lee and Seung [32] proposed the
 273 multiplicative update algorithm to solve the equation (1). Simple multiplicative update formulas
 274 based on \mathcal{D}_F are given by

$$w_{ij} \leftarrow w_{ij} \frac{[\mathbf{XH}^\top]_{ij}}{[\mathbf{XHH}^\top]_{ij} + \varepsilon}, \quad (5)$$

$$h_{jt} \leftarrow h_{jt} \frac{[\mathbf{W}^\top \mathbf{X}]_{jt}}{[\mathbf{W}^\top \mathbf{WH}]_{jt} + \varepsilon}. \quad (6)$$

275 For the implementation purpose, a small positive constant ε is added to the denominator in each
 276 update rule to avoid division by zero. Lee and Seung claimed that the above algorithm converges
 277 to a local minimum [32], which was later shown to be incorrect (see for example [14, 6, 36]): the
 278 above algorithm 5, 6 can only keep the non-increasing property of the objective.

279 Based on the Kullback-Leibler divergence, Brunet et al. [8] used modified versions of Lee and
 280 Seung’s (2001) [32] simple multiplicative updates to avoid numerical underflow. At each step, \mathbf{W}
 281 and \mathbf{H} are updated by using the coupled divergence equations:

$$\mathbf{W}_{ia} \leftarrow \mathbf{W}_{ia} \frac{\sum_{\mu} \left[\frac{\mathbf{H}_{a\mu} \mathbf{X}_{i\mu}}{(\mathbf{W}\mathbf{H})_{i\mu}} \right]}{\sum_{\nu} \mathbf{H}_{a\nu}}, \quad (7)$$

$$\mathbf{H}_{a\mu} \leftarrow \mathbf{H}_{a\mu} \frac{\sum_i \left[\frac{\mathbf{W}_{ia} \mathbf{X}_{i\mu}}{(\mathbf{W}\mathbf{H})_{i\mu}} \right]}{\sum_j \mathbf{W}_{ja}}. \quad (8)$$

282 The stopping criterion for Lee and for Burnet optimization is the variance of the connectivity
 283 matrix.

284 2.2.4. Initialization of NMF

285 Most NMF objectives are not convex and they are sensitive to the initialization of matrices \mathbf{W} and
 286 \mathbf{H} . A good initialization can then sidestep some of the convergence problems, especially for a large
 287 input dataset. If the starting values for the algorithm are chosen randomly, every run of the NMF
 288 algorithm may find a different local minimum of the objective function. Therefore a reasonable
 289 initialization of matrix \mathbf{W} and \mathbf{H} is necessary and helps the physical interpretation of the obtained
 290 patterns [59, 50]. Particular emphasis has to be placed on the initialization of NMF because of its
 291 local convergence. Several approaches have then been proposed to choose the appropriate NMF
 292 initialization (see [2] for a review). Thiem et al. [59] suggested the use of Non-Negative Double
 293 Singular Value Decomposition (NNDSVD, developed in [7]) for PM source apportionment and did
 294 not recommend random initialization. Meanwhile, Hutchins et al. [24] claim that performing 30-50
 295 runs for random initialization is considered sufficient to get a robust estimate of the factorization.
 296 We tested three different kinds of initialization techniques for the two different NMF algorithms
 297 described above. In the first seeding method, a random starting point has been used, where the
 298 entries of \mathbf{W} and \mathbf{H} are drawn from a uniform distribution, within the same range as the \mathbf{X} matrix’s
 299 entries. That is the entries of each factor are drawn from a uniform distribution over $[0, \max\{\mathbf{x}\}]$,

300 where \mathbf{x} is the column vector of \mathbf{X} . We used a maximum of 100 runs for each algorithm to
 301 achieve stability. The second initialization method that we tested consists in using the results of
 302 the Independent Component Analysis (ICA) (FastICA algorithm [25]) where only the positive part
 303 is used to initialize the factors. The last tested initialization algorithm is the Non-Negative Double
 304 Singular Value Decomposition (NNSVD) [7].

305 2.2.5. Determining the number of components

306 One of the critical parameters in NMF is the number of components J to select for the factorization
 307 in equation 1. An appropriate decision on the value of J is critical in practice, but it is usually
 308 chosen such that $J \ll \min(I, T)$ in which case \mathbf{WH} represents a compressed form of the initial
 309 matrix \mathbf{X} . The factorization rank parameter can be estimated by computing the Residual Sum of
 310 Squares (RSS) or by the explained variance (EVAR) between a target matrix \mathbf{X} and its estimate
 311 $\widehat{\mathbf{X}}$:

$$\text{RSS} = \sum_i \sum_t (\mathbf{x}_{it} - \widehat{\mathbf{x}}_{it})^2 \quad (9)$$

$$\text{EVAR} = 1 - \frac{\text{RSS}}{\sum_{i,t} \widehat{\mathbf{x}}_{it}} \quad (10)$$

312 The “optimal” rank is chosen using the graph of the EVAR (or RSS); it corresponds to the first
 313 point where the graph shows an inflection point, as Hutchins et al. [24] did with the algorithm of
 314 Lee et al. [31].

315 3. Results and discussion

316 In the following, a preliminary analysis of the original data is conducted (section 3.1). Then, the
 317 choice of the number of factors and their profiles are discussed in the section 3.2. Finally, for
 318 interpretation purposes, these factors are analyzed and related to different other parameters such
 319 as occupancy and windows state.

320 In this paper, the terms factor, component and patterns are equivalent. The results and analyses
321 were conducted in R [54] using the NMF R package [19] and figures were produced using the ggplot2
322 package [66].

323 Since particles in the fine range ($<2.5\ \mu\text{m}$) dominate the total particle number, the obtained NMF
324 patterns emphasize this range. As the range variation of the number of particles can be very
325 different from one bin to another, we considered very carefully the standardization need of the raw
326 data matrix as input for NMF. This standardization procedure produces negative values that should
327 be avoided due to the non-negativity constraint. For this reason we have made a translation of the
328 standardized values by adding a constant equal to $2\ \text{particles}\cdot\text{cm}^{-3}$ to each one. This constant shift
329 does not impact the factorization results.

330 3.1. Data description and preliminary analysis

331 During the sampling campaign, the indoor temperature values varied from 15°C to 40°C and relative
332 humidity ranged between 18 % and 65%. The highest humidity values were recorded during May
333 and June.

334 Fifteen size bins were collected, but only five representative fractions are described in this paper:
335 $\text{PN}_{0.35} = \text{PN}_{[0.30-0.40]}$, $\text{PN}_{0.9} = \text{PN}_{[8-10]}$, $\text{PN}_{2.5} = \text{PN}_{[2-3]}$, $\text{PN}_{8.75} = \text{PN}_{[6.25-12.25]}$ and $\text{PN}_{17.5} =$
336 $\text{PN}_{[15.5-19.5]}$. The indoor $\text{PN}_{0.35}$ varied from 1 to $263\ \#\cdot\text{cm}^{-3}$ ($\#\cdot\text{cm}^{-3}$ for number per cm^3) with
337 a median of $16\ \#\cdot\text{cm}^{-3}$ and 90 % (percentile P_{90}) of the values are less than $18\ \#\cdot\text{cm}^{-3}$. There is
338 no significant difference in the monthly levels excepting for March. The range variations for $\text{PN}_{0.9}$,
339 $\text{PN}_{8.75}$ and $\text{PN}_{17.5}$ are 2611, 119 and $26\ \#\cdot\text{L}^{-1}$, respectively.

340 The variability expressed by means of the coefficient of variation ($CV = \text{Standard Deviation}/\text{Mean}$)
341 depends on the size of the particle. Indeed, the CV values are: 113% for $\text{PN}_{0.35}$, 80% for $\text{PN}_{0.9}$,
342 92% for $\text{PN}_{2.5}$, 250% for $\text{PN}_{8.75}$ and 750% for $\text{PN}_{17.5}$.

343 Descriptive statistics for the five size bins concentrations according to occupancy and windows
344 opening state are shown in Table 1. The mean values show very large variations; mean values are

Table 1: Descriptive statistics of the selected indoor PN size bins ($\#.cm^{-3}$ for number per cm^3) according to occupancy (Occup. or Non Occup.) and windows opening state (Open or Close). The percentile P corresponds to percentage of the total values are the same as or below the measurement concentrations: the 25th percentile (P_{50}), the 50th percentile corresponds to the median and the 75th percentile (P_{75}). n represents the number of samples (minutes).

PN	Group	n	Mean(SD)	Min	Max	P_{10}	P_{25}	P_{50}	P_{75}	P_{90}
PN _{0.35}	Close	230266	28(32.1)	0.68	262.8	5.73	8.76	16.1	34.83	63.6
	Open	29765	20.4(16.4)	3.16	125	7	9.46	14.6	26	41.8
PN _{0.9}	Close	230266	0.33(0.27)	0	2.6	0.1	0.15	0.2	0.42	0.64
	Open	29764	0.32(0.19)	0.025	1.4	0.12	0.18	0.2	0.41	0.57
PN _{2.5}	Close	230266	0.031(0.02)	0	0.6	0.006	0.013	0.02	0.042	0.06
	Open	29764	0.081(0.05)	0	0.6	0.029	0.04	0.06	0.11	0.15
PN _{8.75}	Close	230266	3×10^{-4} (0.001)	0	0.11	0	0	0	0	0.001
	Open	29764	0.001(0.002)	0	0.03	0	0	0	0.002	0.004
PN _{17.5}	Close	230266	3×10^{-5} (0.0002)	0	0.02	0	0	0	0	0
	Open	29764	1×10^{-4} (0.0004)	0	0.006	0	0	0	0	0
PN _{0.35}	Non Occup	245326	27.25(31)	0.68	262.8	5.9	8.8	16	34	60.8
	Occup	14705	25.54(28.5)	2.74	188.1	6	8.7	14.8	30.15	57.93
PN _{0.9}	Non Occup	245326	0.32(0.26)	0	2.6	0.1	0.16	0.26	0.41	0.63
	Occup	14704	0.36(0.29)	0.02	2.4	0.12	0.17	0.28	0.44	0.72
PN _{2.5}	Non Occup	245326	0.036(0.03)	0	0.67	0.006	0.01	0.027	0.05	0.073
	Occup	14704	0.061(0.04)	0	0.6	0.02	0.03	0.049	0.08	0.121
PN _{8.75}	Non Occup	245326	0.0004(0.001)	0	0.07	0	0	0	0	0.001
	Occup	14704	0.002(0.002)	0	0.11	0	0	0.001	0.002	0.004
PN _{17.5}	Non Occup	245326	0.00003(0.0002)	0	0.02	0	0	0	0	0
	Occup	14704	0.0002(0.001)	0	0.02	0	0	0	0	0.001

345 higher than the median ones. As the data distribution was positively skewed, the median (P_{50}) was
346 preferred for interpretation purposes instead of the arithmetic mean.

347 For most of the particle sizes, the median values of PN are very similar regardless the windows state,
348 while the 10th and 90th percentile values are different. For instance, 90 % of PN_{0.35} concentrations
349 are less than 64 $\#.cm^{-3}$ in the case of closed windows state and 42 $\#.cm^{-3}$ when at least one
350 window is opened. These observations are related to the total number of minutes in each opening
351 state. On average, windows remain closed around 85% of the total time throughout the study
352 period (24/7), week-end and holidays included.

353

354 Indoor particles can originate from outdoor sources and also from various indoor sources. Thus, it is

355 possible that the levels of indoor particles exceed outdoor ones. Table 2 shows the Indoor/Outdoor
 356 (I/O) ratio for the selected five size bins levels in different configurations. Diurnal time variability
 357 of the ratio (I/O) is provided in the supplementary material of this paper (cf. Figure B.1).

358 Median I/O ratios are less than 1 for smaller range ($\leq 2.5 \mu\text{m}$) regardless the windows state or
 359 indoor occupancy.

360 It is worth noting that median I/O ratios of PN during the non-occupancy are very similar to those
 361 observed for closed windows. In the absence of known indoor sources, the reported median I/O
 362 ratios have been ranged from 0.37 to 0.53 in the case of opening windows.

Table 2: Descriptive statistics of I/O PN ratios for the selected size ranges according to occupancy (Occup. and Non Occup.) and opening windows state (Open and Close). n represents the number of samples (minutes) and P the percentile values.

I/O ratio	Group	n	Mean(sd)	min.	max.	P_{25}	P_{50}	P_{75}	P_{90}
PN _{0.35}	Close	104393	0.49(0.2)	0.008	4.318	0.363	0.47	0.59	0.71
	Open	496	0.73(0.21)	0.347	1.476	0.601	0.7	0.88	1
PN _{0.9}	Close	104393	0.57(0.25)	0.002	7.436	0.425	0.53	0.65	0.81
	Open	496	0.76(0.18)	0.36	1.378	0.643	0.76	0.87	1
PN _{2.5}	Close	104393	0.59(4.57)	0.0003	611	0.243	0.37	0.57	0.91
	Open	496	0.62(0.33)	0.1	2.212	0.376	0.57	0.83	1.07
PN _{8.75}	Close	104393	2.29(6.22)	0.0004	311	1	1	1	1.91
	Open	496	6(11.33)	0.012	71	0.524	1	3.72	21
PN _{17.5}	Close	104393	1.17(1.9)	0.002	201	1	1	1	1
	Open	496	2.43(5.2)	0.032	51	1	1	1	11
PN _{0.35}	Non Occup.	99902	0.49(0.2)	0.043	4.318	0.365	0.47	0.56	0.71
	Occup.	4987	0.51(0.26)	0.008	4.143	0.342	0.47	0.63	0.8
PN _{0.9}	Non Occup.	99902	0.57(0.25)	0.002	7.436	0.424	0.53	0.65	0.8
	Occup.	4987	0.62(0.27)	0.075	3.849	0.448	0.58	0.73	0.9
PN _{2.5}	Non Occup.	99902	0.6(4.6)	0.0003	611	0.239	0.36	0.56	0.8
	Occup.	4987	0.9(4.25)	0.008	291	0.398	0.59	0.2	1.4
PN _{8.75}	Non Occup.	99902	2.1(5.67)	0.0004	301	1	1	1	1
	Occup.	4987	6.6(12.7)	0.001	311	0.524	1	11	21
PN _{17.5}	Non Occup.	99902	1.15(1.7)	0.002	201	1	1	1	1
	Occup.	4987	2(3.8)	0.009	71	1	1	1	1

363

364 The Figure 2 shows the diurnal variability of the median value of indoor PN concentrations according
 365 to occupancy and windows states for the following bins: PN_{0.35}, PN_{0.9}, PN_{2.5} and PN_{17.5}. The

366 diurnal pattern for $PN_{0.35}$ shows a convex shape when all the windows are closed and no occupancy
367 detected. The I/O ratio of the corresponding bin is always less than 0.6 for closed windows and it
368 ranges between 0.6 and 1 in the case of opened windows. The variation profile can be explained
369 by the fact that the outdoor particles penetrate inside by mechanical ventilation and infiltration
370 through the gaps in the building, given that the windows are closed. These fine particles come
371 mostly from outdoor sources, such as traffic which is higher during the rush hours. The diurnal
372 profile of the PN of size $0.9\mu\text{m}$ when the windows are opened is characterized by a significant peak
373 between 8:00 and 9:30 a.m. (when usually work starts) and then it decreases until 7:00 p.m. On
374 the other hand, the same variation is observed when a movement of occupants is detected, even
375 when the windows are closed. For $PN_{2.5}$, there is a significant difference between the case when
376 the windows are opened or closed, regardless the occupancy. The values are higher for $PN_{2.5}$ when
377 windows are opened showing the importance of outdoor sources for these size bins. By contrast,
378 for the coarse particles ($PN_{17.5}$), the occupancy variable discriminates the profiles revealing that
379 these particles are generated indoors (occupants' activities, such as walking).

380 To summarize, opening the windows results in similar trends in PN for both occupation and non-
381 occupation conditions (except for $PN_{17.5}$ due to less contribution of coarse particles in the air).
382 Mostly, PN concentrations with occupancy are higher than PN concentrations without occupancy.
383 However, real contribution of occupation is marked when the windows are closed.

384 The particulate number concentration decrease in the indoor environment occurs mainly by two
385 mechanisms: ventilation and deposition. In general, ventilation could play a positive role in the
386 loss of particles from indoor air, but sometimes it may cause entering the outdoor pollutants via
387 the supplied air coming indoors.

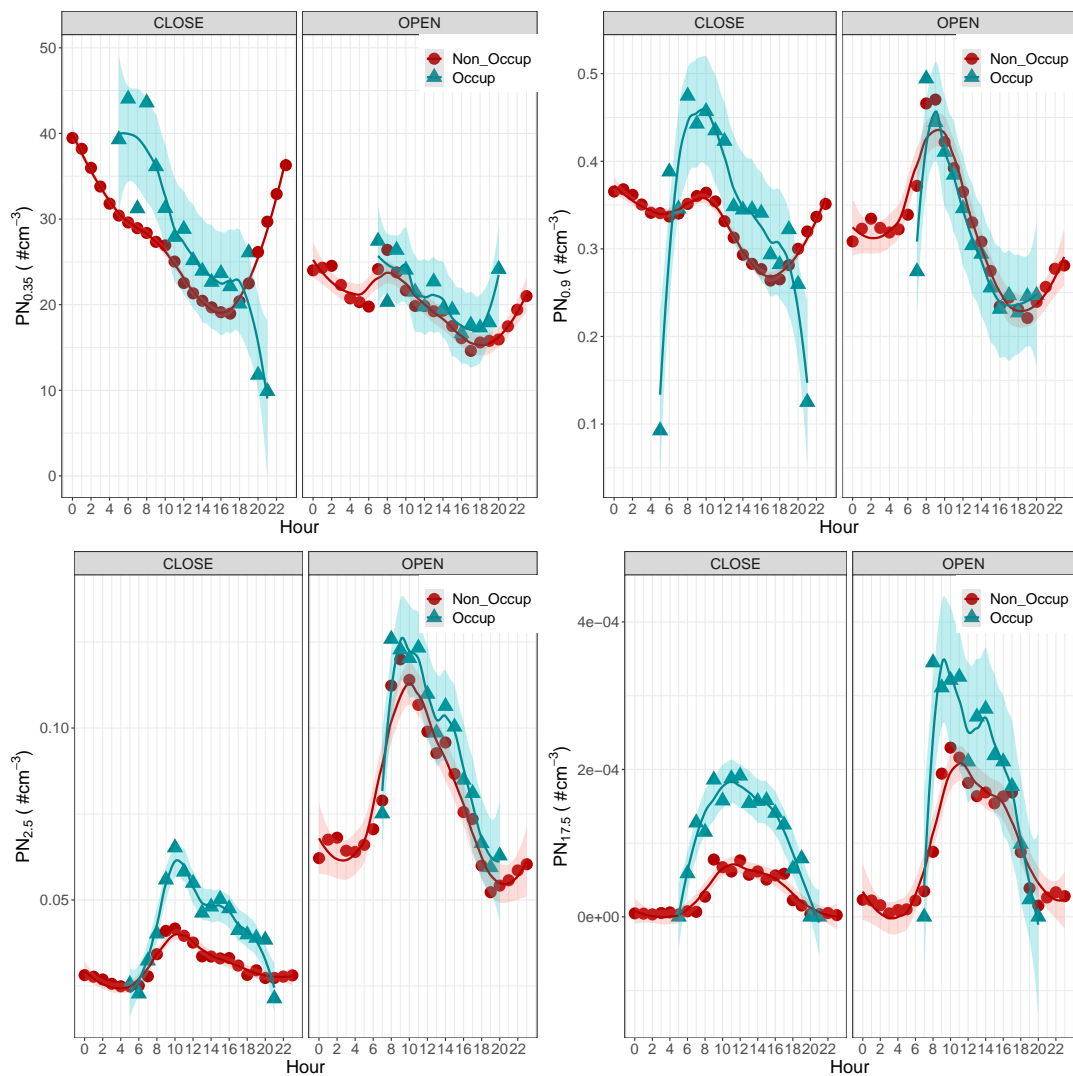


Figure 2: Diurnal variability of the median value of indoor PN concentrations for the following size ranges: $0.35\ \mu\text{m}$, $0.9\ \mu\text{m}$, $2.5\ \mu\text{m}$ and $17.5\ \mu\text{m}$. $\text{PN}_{0.35} = \text{PN}_{[0.30-0.40]}$, $\text{PN}_{0.9} = \text{PN}_{[8-10]}$, $\text{PN}_{2.5} = \text{PN}_{[2-3]}$ and $\text{PN}_{17.5} = \text{PN}_{[15.5-19.5]}$.

389 3.2. NMF results

390 As presented in the subsection 2.2.4 and 2.2.5, the method should be initialized and an optimal
391 number of components should be determined.

392 Values in the range $J = 2, \dots, 10$ were tested based on 20 000 randomly sampled 1-minute dataset
393 using three different initialization algorithms: ICA, NNSVD and Random for the two optimization
394 methods proposed by Brunet [8] and Lee [31, 32].

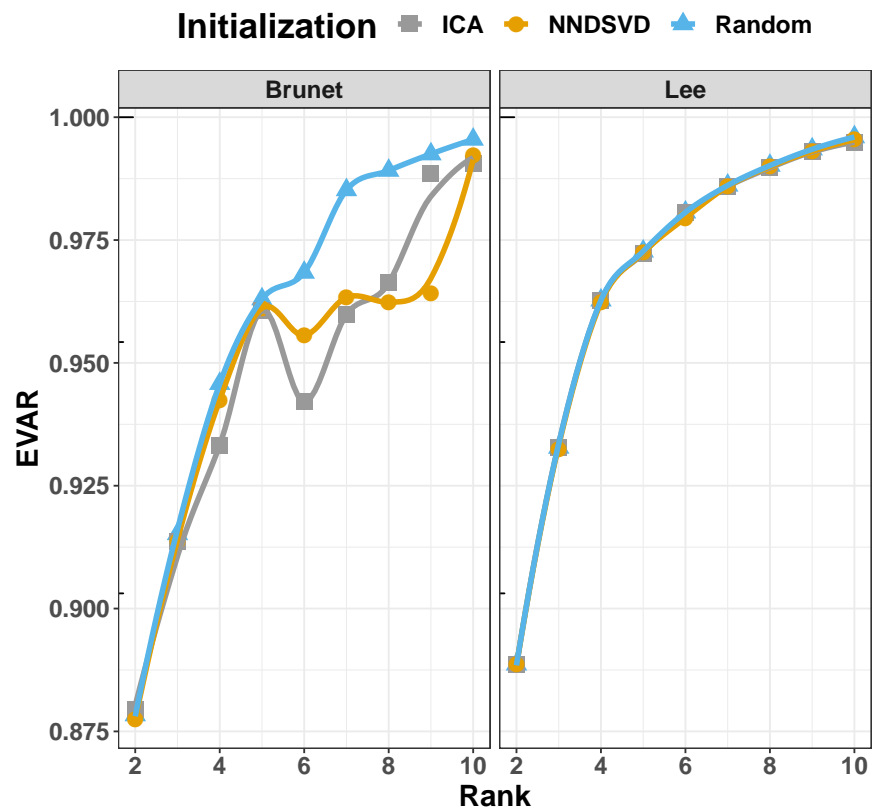


Figure 3: Explained Variance (EVAR) variations of Brunet [8] or Lee [31, 32] algorithms according to different initializations.

395

396 Figure 3 shows the evolution of the explained variance according to initialization algorithms (ICA,
397 NNSVD, and Random) according to the rank number. Unexpectedly, for each initialization method,
398 NMF generated extremely resembling results for Lee’s algorithm. A similar observation is observed
399 for the first 5 rank factors in Brunet’s algorithm. Note that random initialization may make the
400 experiments unrepeatable because of their local minima property. Several studies have revealed
401 that methods with non-random seeding demonstrate their superiority either in the fast convergence
402 in prophase or the structure preservation [69, 7]. This is yet another reason for not choosing random
403 initialization in our experiments.

404 The interpretability of the factors can be a selection criterion for the rank choice. The use of expert
405 insight in the case of indoor air quality was taken into account. It can be observed that using
406 the interpretability criterion of the value of J by increasing or decreasing by 1, either the new
407 NMF-factor shows a mixture of parts of existing factors or contains completely new environmental
408 patterns. To represent as many different patterns as possible and taking into account the explained
409 variance variation, we found for the considered dataset that J should be about 4-6. To draw
410 comparisons between different results, the best physical interpretation and factorization error at
411 the same time were obtained for $J = 5$. With this rank corresponding thus to 5 factors, the
412 explained variance corresponds to 96% and 97.5% for Brunet’s and Lee’s algorithms, respectively.

413 3.2.1. Factors contributions

414 The time series of the factors obtained by NMF are provided in the supplementary material of this
415 article (cf. Figure C.1). The relative contribution of factors to each particle size is shown in Figure
416 4. The relative contributions of the five components are distributive *i.e.* all the fractions can be
417 found in at least one factor.

418 For the fine size fractions ($0.3\ \mu\text{m}$, $0.45\ \mu\text{m}$, $0.57\ \mu\text{m}$), the contribution of the component 2 obtained
419 by NMF is nearly 75 % of the total contributions. By contrast, this component contributes the
420 least for larger sizes of particles, so it is specific to the fine size.

421 The fourth component is related to coarse fraction, which represents 80 % of all the other contri-

422 butions. For fractions between $1.3\ \mu\text{m}$ and $4.5\ \mu\text{m}$, the components 3 and 5 share over 50 % of all
 423 contributions. The first component is involved in about 20% for particles smaller than $2.5\ \mu\text{m}$ and
 424 can reach 80% for particles with a size of $12.5\ \mu\text{m}$.

425 To facilitate the interpretation of each component, we compute the correlations between the origi-
 426 nal data (the monitored time series corresponding to each fraction) and each NMF component.
 427 The interpretation of the NMF components is based on finding which PN time series are the most
 428 strongly correlated with each component. Figure 5 shows the correlation values between the com-
 429 ponents time series and fraction variables time series. The color variations express the strength and
 430 direction of the correlation relationship.

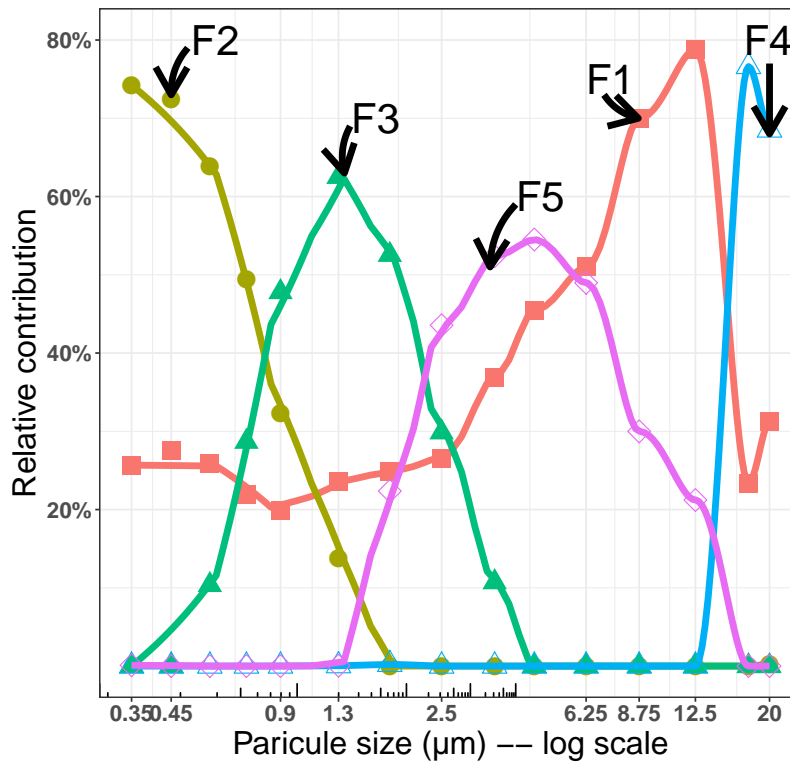


Figure 4: Relative contributions of the factors to temporal variability of the 15 fractions.

431

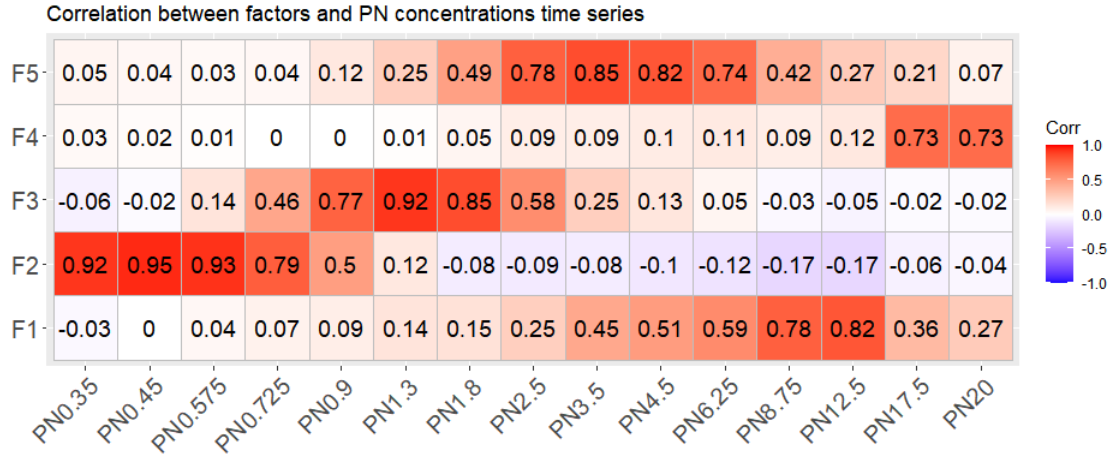


Figure 5: Pearson correlation coefficient between NMF factors (\mathbf{H} matrix) and the PN concentrations time series of each fraction.

432 Note that all the significant correlations, i.e. greater than 0.6, are positive. The first NMF compo-
 433 nent is strongly correlated with particles ranging from 6.25 μm to 12.5 μm , the correlation coefficients
 434 vary from +0.6 and +0.82. The contribution of this component to the variability of the of particles
 435 of sizes between 3.5 μm and 4.5 μm is not negligible ($\sim 40\%$) although their correlations are not
 436 very strong (+0.45, +0.51).

437 The 2nd factor is associated with particle size ranges below 0.8 μm with a strong positive correlation
 438 coefficient (> 0.8). Its contribution is higher than 63% for particles of sizes 0.575 μm , 0.45 μm and
 439 0.35 μm .

440 Factor 3 contributes (up to 50%) to sizes between 0.9 μm and 2.5 μm , with a correlation varying
 441 between +0.58 and +0.95. We notice that the diurnal profiles of the F3 and PN_{0.9–1.8} are very
 442 similar (cf. Figure 2); both of them are characterized by a significant peak between 8:00 a.m. and
 443 9:30 a.m. followed by a decreases until 7:00 p.m.

444

445 Note that for particles of size 2.5 μm , two factors contribute simultaneously to its variation: F3
 446 (correlation = 0.58) and F5 (correlation = 0.78). The same observation can be made for the particles
 447 of size 6.25 μm : two factors contribute to its variation (F1 and F5). Coarse particles ($> 17.5 \mu\text{m}$)
 448 are associated with the 4th factor with a correlation of +0.73. The latter does not contribute to any

449 of the other particle ranges. In Figure 3, there is a peak of contribution of about 80% for particle
450 sizes $17.5\ \mu\text{m}$ and $> 20\ \mu\text{m}$. Particles in the intermediate size range (between 2.5 and $6.25\ \mu\text{m}$) are
451 associated to F5 with correlations ranging from $+0.74$ to $+0.85$. The contribution of F5 to the
452 variability of particles of $3.5\ \mu\text{m}$ size, is close to 50%.

453 3.2.2. Time variation of the factors

454 Having examined the overall contributions of NMF components, we now take a closer look to
455 interpretate them. This section sheds light on the nature of the hidden components obtained using
456 NMF. To do so, we combine three sources of information: (i) diurnal variation of different time
457 series (NMF factors and CO_2), (ii) windows states and (iii) indoor occupancy. According to the
458 states of the windows (i.e. open/closed) and to the occupancy (occupied/ unoccupied), we further
459 subset the factor's time series and plotted the diurnal variation in Figure 6. More specifically, two
460 figures are presented for each component corresponding to the window states (closed or open) and
461 in each figure, two curves (red dot and green triangle) are associated with the occupancy status.
462 To facilitate the interpretation, diurnal variations in CO_2 concentrations have been added in the
463 same Figure. The CO_2 variations are used here as a fingerprint of occupant presence in order to
464 allow identification of similar factors with this type of variation.

465

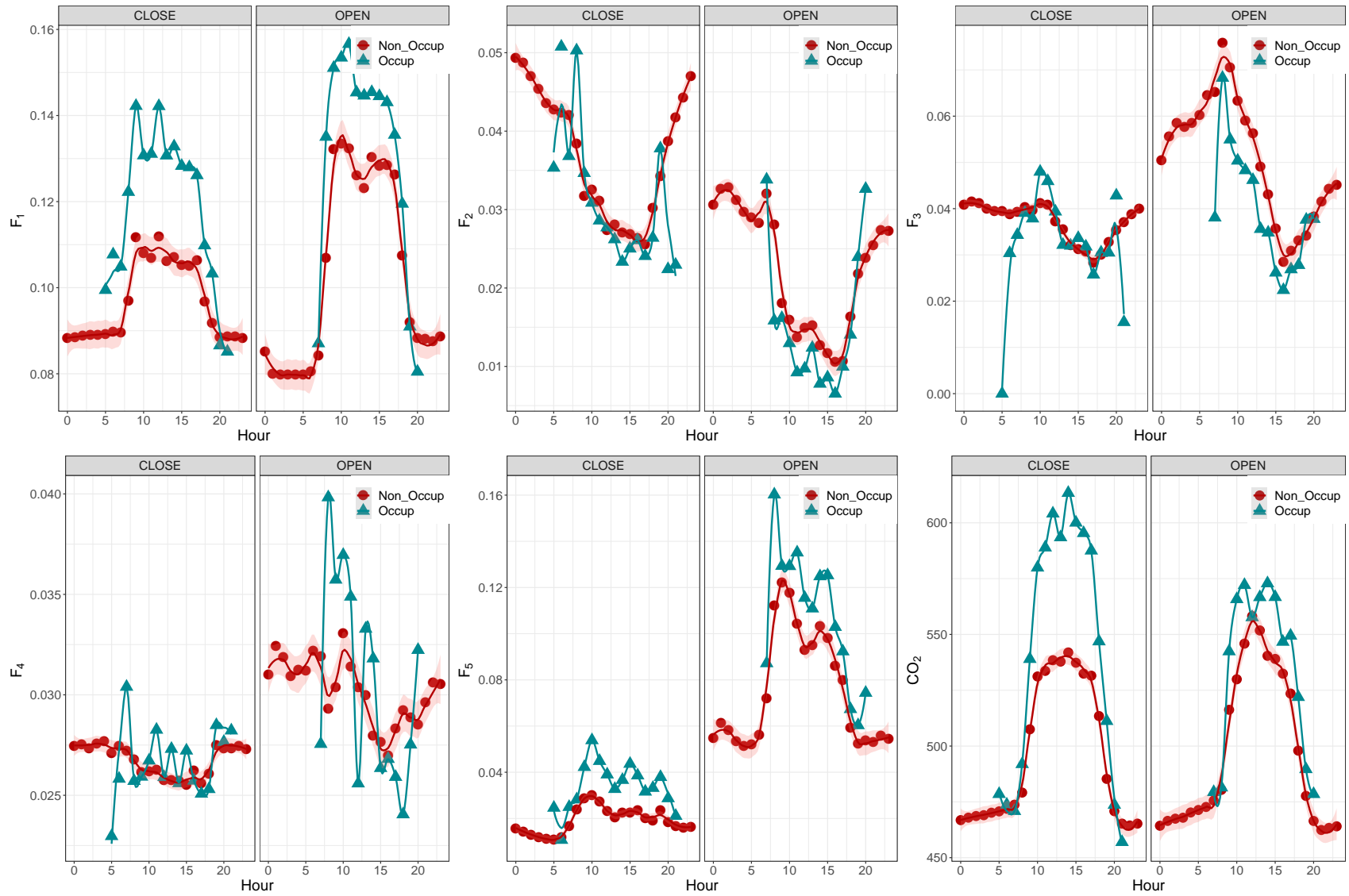


Figure 6: Diurnal variation of the five NMF factors and CO_2 concentrations according to windows state and occupancy.

466 Table 3 shows statistical tests performed to determine whether the differences between the average
 467 values of all the NMF factors are significant in two situations: (i) when all windows are closed
 468 *vs.* when at least one window is open, and (ii) in the case of occupancy *vs.* non-occupancy. For
 469 a *p-value* <0.05, the null hypothesis " H_0 : No difference in mean" is rejected. For example, for
 470 the first factor, we can say that there is no significant difference when all the windows are closed
 471 or at least one window is open (*p-value* >0.05, gray cells in the table 3). Whereas occupancy is
 472 an important parameter for component 1 (F1): the variability of the two profiles (occupancy *vs.*
 473 non-occupancy) are different (*p-value* <0.05).

Table 3: Significant differences between diurnal profiles to discriminate the role of windows and occupancy parameters using statistical tests.

<i>p - value</i>	Windows		Occupancy	
	<i>t - test</i>	Wilcoxon	<i>t - test</i>	Wilcoxon
F1	0.36	0.98	4.69E-06	2.54E-06
F2	0.0003301	0.0006599	0.0005568	0.0009668
F3	0.0001015	3.69E-05	0.00245	0.004618
F4	6.77E-08	1.97E-08	0.1147	0.02706242
F5	0E-10	0E-10	0.02706242	0.002432

474

475 The diurnal profile of the first factor is very similar to the diurnal variations of indoor CO₂ concen-
 476 trations (cf. Figure 6), suggesting that this component is related to the presence of the occupants,
 477 who are the main source of CO₂. When all the windows are closed, the occupancy curve (red dots)
 478 is separated (*p-value* < 0.05) from the non-occupancy one.

479 When at least one window is open, the values of component 1 are higher. It is as if another source
 480 coming from outside is added or the ventilation changes the transport mechanisms of large particles
 481 (resuspension). Nevertheless, several studies report that the mechanism of resuspension is especially
 482 related to particles of sizes less than 10 μm [52, 51]. As previously mentioned (Figure 5), the peak
 483 contribution of the first component is associated with particles of sizes between 8.75 and 12.5 μm.

484 The diurnal profile of the second component illustrates a decreasing effect during the period from
 485 6 a.m. to 7 p.m.. For F2, there is no significant difference for the "occupancy" variable, the two
 486 curves overlap. As shown in previous sections, this component is related to fine particles (less than
 487 0.75 μm).

488 It turns out that the parameter "opening windows" perfectly discriminates the component 5 (*p-value*
 489 <0.05), but the importance of occupancy is not conclusive. When at least one window is open, the

490 component 5 levels are high. The diurnal profiles of the two components 1 and 5 have comparable
491 patterns, for the case of open windows. As a matter of fact, the relative contribution of F1 and F5
492 are important when the open-plan is occupied (7 a.m.-7 p.m.).

493 For component 3, a peak at 9 a.m. is observed in the case of open windows and the diurnal profile
494 gives a sinusoidal appearance. On the other hand, NO_x emissions vary according to the time of
495 day: they are very high in the morning and late afternoon (the rush hours). A hypothesis can be
496 put forward suggesting that the peak of component 3 is of external origin, related to some sources
497 such as road traffic.

498 Components F3 and F5 could be attributed to the influence of outdoor sources on the indoor
499 environment. The seasonal appearance (of component 3) is mainly due to diurnal variations of
500 outdoor sources. This finding provides additional insight into the sources of particles ranging in
501 size from $1.8\ \mu\text{m}$ to $6.25\ \mu\text{m}$. Thus, by combining them with the information provided in Table 2, we
502 notice the important influence of outdoor conditions on indoor concentration levels. At this stage,
503 one of the major points of interest has been the outdoor source identification i.e. when outdoor
504 concentrations and characteristics are the main contributing factors. Overall, outdoor sources have
505 been mainly associated with fine particles in accumulation mode ($0.1\text{--}1\ \mu\text{m}$), probably because
506 these particles can persist in the air since they are too small for inertial deposition and too large for
507 diffusion removal processes. These particles are capable of entering in the buildings and remaining
508 airborne for longer periods.

509 The forth component (F4) is characterized mainly by the coarse particles ($> 17.5\ \mu\text{m}$). Figure 6
510 shows a strong variability in the diurnal profile during the occupation period. The profile is very
511 random because it is mainly associated with the activities of the occupants. These activities, and in
512 particular walking, are responsible of resuspension of the coarse particles. We remind that the office
513 is equipped with a carpet covering the entire floor. The statistical tests cannot be used because the
514 probability density is bimodal for this profile. We notice that when at least one window is open,
515 the profile is further modified. Statistical tests confirm this observation ($p\text{-value} < 5\%$).

516 3.3. Limitations and future directions

517 Firstly, a generic NMF, which is a method to solve a linear system, was employed and the underlying
518 model was used as a set of linear Chemical Mass Balance (CMB) to estimate individual source
519 components from their mixtures. However, the indoor mixing phenomena are better represented
520 by non-linear relationships, such as infiltration, sink, inertial deposition, and diffusive removal
521 processes. A research question that remains is how these latter processes can be considered in a
522 more general model. In the same vein, future work would include the hybridization of physical
523 models with factorial source separation methods.

524 Secondly, although NMF has a realistic non-negative constraint on factor profiles and contributions,
525 it still does not cover all the indoor environment characteristics. Thus, many parameters as occupant
526 behavior and windows opening, which have “*random impacts*” over the primary determinants of the
527 decay rate, have to be incorporated in the NMF optimization problems. Similarly, the including
528 time-activity patterns could improve the interpretability of the NMF results.

529 One critical question that remains to be answered about the NMF method for indoor air quality
530 is: how to integrate inherent indoor specificity and constraints in the NMF formulation?

531 Recently the NMF field expanded to multidimensional data arrays, called Non-negative Tensor
532 Factorization (NTF) [15], which could offer valuable new insights on IAQ modeling issues. Future
533 work also involves validating the performance of the NMF model using data over a much longer
534 period and above all including more information about the different activities such as walking,
535 cleaning, printing,

536 From the instrumentation point of view, it is clear that the motion detection has been largely un-
537 derestimated by the measurement method. It might be more appropriate to introduce the “*number*
538 *of occupants*” as a parameter in the data processing as well as in the modeling step.

539 3.4. Conclusions

540 Evidence continues to mount that indoor particles concentrations are one of the major determinants
541 for individual exposure. That is why it is necessary to characterize the source time variability and

542 to assesses the different source impacts on the total personal exposure. Attempts to estimate
543 the source of indoor particles concentrations are complicated because indoor air quality is both
544 building-specific and occupant-specific. Many researchers are using experimental chambers; however
545 quantifying continuous ventilation rates and penetration factors for an occupied building can be
546 tedious, time-consuming and expensive. While most receptor-oriented methods use the pollutant
547 compositions (chemical fingerprint) to apportion the contribution of individual sources, our work
548 focuses on the time variability source characterization, namely “*temporal fingerprint*”. This study
549 outlines the basic temporal characterization and source apportionment using Non-negative Matrix
550 Factorization (NMF).

551 This study has shown that continuous measurements of indoor particle number concentrations with
552 additional information of occupancy and windows opening are useful for the source apportionment
553 models. The factorization has been successfully employed to study particle number time series
554 fluctuations in an occupied open-plan office. Five distribution profiles were resolved using different
555 initialization and optimization algorithms. Thanks to NMF and correlation analysis, the impact of
556 occupancy and windows states associated with outdoor traffic were identified. The outdoor sources
557 are captured by two components (very likely the traffic impact) with major number modes at $1.3\ \mu\text{m}$.
558 Besides, a potential association with primary outdoor pollutants (NO_x) has been captured and a
559 diurnal pattern similar to traffic can be associated. A common pattern between F1 and $\text{PN}_{0.9-1.8}$
560 has been identified, such as diurnal profile which is characterized by a significant peak between
561 8:00 a.m. and 9:30 a.m, decreasing until 7:00 p.m.

562 Taking into consideration the results presented above, throughout the six-month measurements of
563 indoor and outdoor PN concentrations, it can be concluded the followings:

- 564 • This study demonstrates the importance of recording real-time concentrations over a longer
565 duration (i.e., several months with a short time step). The exploitation of such data has the
566 potential to extract and capture the different patterns of temporal variability and their major
567 determinants.
- 568 • These results contribute significantly to the very small data set available in the literature on

569 the time variability source characterization.

- 570 • The NMF technique -and its variants- shows considerable promise for further application to
571 the indoor environment and the possibility to identify other sources and their contributions.
- 572 • Continuous monitoring of climatic parameters as well as active instrumentation of the build-
573 ing (windows, occupancy) are necessary for the evaluation of the total exposure to indoor
574 pollutants.

575 To sum up, the results of this study entail that the degree of human's exposure to different sources
576 varies with many parameters (building and occupants specific). These sources could be captured
577 using NMF and its variants.

578 **Acknowledgments**

579 This work received financial support from the French research program on air quality (PRIME-
580 QUAL) through the Grant No 12-MRES-PRIMEQUAL-4-CVS-09.

581 **References**

- 582 [1] Eileen Abt, Helen H Suh, Paul Catalano, and Petros Koutrakis. Relative contribution of outdoor and
583 indoor particle sources to indoor concentrations. *Environmental science & technology*, 34(17):3579–
584 3587, 2000. [1](#)
- 585 [2] Russell Albright, James Cox, David Duling, Amy N Langville, and C Meyer. Algorithms, initializations,
586 and convergence for the nonnegative matrix factorization. Technical report, Tech. rep. 919. NCSU
587 Technical Report Math 81706. <http://meyer.math.ncsu.edu/Meyer/Abstracts/Publications.html>,
588 2006. url: <http://citeseerx.ist.psu.edu/viewdoc/download>, 2006. [2.2.4](#)
- 589 [3] Mohd Yasreen Ali, Marlia M Hanafiah, Md Firoz Khan, and Mohd Talib Latif. Quantitative source
590 apportionment and human toxicity of indoor trace metals at university buildings. 121:238–246. [1](#)
- 591 [4] JP Barnard and C Aldrich. Modelling of air pollution in an environmental system by use of non-linear
592 independent component analysis. *Computer Aided Chemical Engineering*, 9:81–86, 2001. [2.2.1](#)
- 593 [5] Claudio A Belis, Bo R Larsen, Fulvio Amato, Imad El Haddad, Olivier Favez, Roy M Harrison,
594 Philip K Hopke, Silvia Nava, Pentti Paatero, and André Prevot. European guide on air pollution
595 source apportionment with receptor models, 2014. [1](#)
- 596 [6] Michael W Berry, Murray Browne, Amy N Langville, V Paul Pauca, and Robert J Plemmons. Algo-
597 rithms and applications for approximate nonnegative matrix factorization. *Computational statistics &*
598 *data analysis*, 52(1):155–173, 2007. [2.2.3](#), [2.2.3](#)
- 599 [7] Christos Boutsidis and Efstratios Gallopoulos. Svd based initialization: A head start for nonnegative
600 matrix factorization. *Pattern Recognition*, 41(4):1350–1362, 2008. [2.2.4](#), [3.2](#)
- 601 [8] Jean-Philippe Brunet, Pablo Tamayo, Todd R Golub, and Jill P Mesirov. Metagenes and molec-
602 ular pattern discovery using matrix factorization. *Proceedings of the national academy of sciences*,
603 101(12):4164–4169, 2004. [2.2.2](#), [2.2.3](#), [3.2](#), [3](#)
- 604 [9] Carmen Cacho, G Ventura Silva, Anabela O Martins, Eduardo O Fernandes, Dikaia E Saraga, Chrysan-
605 thi Dimitroulopoulou, JB Bartzis, Diana Rembges, Josefa Barrero-Moreno, and Dimitrios Kotzias. Air
606 pollutants in office environments and emissions from electronic equipment: A review. *Fresen. Environ.*
607 *Bull*, 22(9), 2013. [1](#)

- 608 [10] Christopher Y Chao and Eddie C Cheng. Source apportionment of indoor pm2. 5 and pm10 in homes.
609 *Indoor and Built Environment*, 11(1):27–37, 2002. [1](#)
- 610 [11] Chun Chen and Bin Zhao. Review of relationship between indoor and outdoor particles: I/o ratio,
611 infiltration factor and penetration factor. *Atmospheric Environment*, 45(2):275–288, 2011. [1](#)
- 612 [12] Judith C Chow, John G Watson, et al. Guideline on speciated particulate monitoring. Technical
613 report, US Environmental Protection Agency, 1998. [1](#)
- 614 [13] Robert Chreiky, Gilles Delmaire, Matthieu Puigt, Gilles Roussel, Dominique Courcot, and Antoine
615 Abche. Split gradient method for informed non-negative matrix factorization. In *International Con-*
616 *ference on Latent Variable Analysis and Signal Separation*, pages 376–383. Springer, 2015. [2.2.1](#)
- 617 [14] Moody Chu, Fasma Diele, Robert Plemmons, and Stefania Ragni. Optimality, computation, and
618 interpretation of nonnegative matrix factorizations. In *SIAM Journal on Matrix Analysis*. Citeseer,
619 2004. [2.2.3](#)
- 620 [15] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and Shun-ichi Amari. *Nonnegative matrix and tensor*
621 *factorizations: applications to exploratory multi-way data analysis and blind source separation*. John
622 Wiley & Sons, 2009. [2.2.2](#), [2.2.2](#), [2.2.2](#), [2.2.3](#), [3.3](#)
- 623 [16] John A Cooper and John G Watson Jr. Receptor oriented methods of air particulate source appor-
624 tionment. *Journal of the Air Pollution Control Association*, 30(10):1116–1125, 1980. [1](#)
- 625 [17] Gilles Delmaire, Gilles Roussel, Dany Hleis, and Frédéric Ledoux. Une version pondérée de la factorisa-
626 tion matricielle non negative pour l’identification de sources de particules atmospheriques. application
627 au littoral de la mer du nord. *Appl Jesa Journal Europeen Des Systemes Automates*, 44(4):547, 2010.
628 [2.2.1](#)
- 629 [18] Andrea R Ferro, Royal J Kopperud, and Lynn M Hildemann. Source strengths for indoor human
630 activities that resuspend particulate matter. *Environmental science & technology*, 38(6):1759–1764,
631 2004. [1](#)
- 632 [19] Renaud Gaujoux and Cathal Seoighe. A flexible r package for nonnegative matrix factorization. *BMC*
633 *bioinformatics*, 11(1):367, 2010. [3](#)

- 634 [20] Hai Guo. Source apportionment of volatile organic compounds in hong kong homes. 46(11):2280–2286.
635 [2.2.1](#)
- 636 [21] Per Heiselberg, Erik Bjørn, and Peter V Nielsen. Impact of open windows on room air flow and thermal
637 comfort. 1(2):91–100. [2.1.3](#)
- 638 [22] Ronald C Henry, Charles W Lewis, Philip K Hopke, and Hugh J Williamson. Review of receptor model
639 fundamentals. *Atmospheric Environment (1967)*, 18(8):1507–1515, 1984. [1](#)
- 640 [23] Philip K Hopke. Review of receptor modeling methods for source apportionment. *Journal of the Air
641 & Waste Management Association*, 66(3):237–259, 2016. [1](#)
- 642 [24] Lucie N Hutchins, Sean M Murphy, Priyam Singh, and Joel H Graber. Position-dependent motif char-
643 acterization using non-negative matrix factorization. *Bioinformatics*, 24(23):2684–2690, 2008. [2.2.4](#),
644 [2.2.5](#)
- 645 [25] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural
646 networks*, 13(4):411–430, 2000. [2.2.4](#)
- 647 [26] Christina Isaxon, Anders Gudmundsson, EZ Nordin, Leif Lönnblad, Andreas Dahl, Gunilla Wieslan-
648 der, Mats Bohgard, and Aneta Wierzbicka. Contribution of indoor-generated particles to residential
649 exposure. *Atmospheric Environment*, 106:458–466, 2015. [1](#)
- 650 [27] Adib Kfoury, Frédéric Ledoux, Abdelhakim Limem, Gilles Delmaire, Gilles Roussel, and Dominique
651 Courcot. The use of a non negative matrix factorization method combined to pm2. 5 chemical data for
652 a source apportionment study in different environments. In *Air Pollution Modeling and its Application
653 XXIII*, pages 79–84. Springer, 2014. [2.2.1](#)
- 654 [28] Adib Kfoury, Frédéric Ledoux, Cloé Roche, Gilles Delmaire, Gilles Roussel, and Dominique Courcot.
655 Pm 2.5 source apportionment in a french urban coastal site under steelworks emission influences using
656 constrained non-negative matrix factorization receptor model. *Journal of Environmental Sciences*,
657 40:114–128, 2016. [2.2.1](#)
- 658 [29] Henri Lantéri, Céline Theys, Cédric Richard, and Cédric Févotte. Split gradient method for nonneg-
659 ative matrix factorization. In *Signal Processing Conference, 2010 18th European*, pages 1199–1203.
660 IEEE, August 23–27, 2010 2010. [2.2.1](#)

- 661 [30] Daniel D Lee and H Sebastian Seung. Unsupervised learning by convex and conic coding. *Advances*
662 *in neural information processing systems*, pages 515–521, 1997. [2.2.2](#)
- 663 [31] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factoriza-
664 tion. *Nature*, 401(6755):788–791, 1999. [1](#), [2.2.2](#), [2.2.5](#), [3.2](#), [3](#)
- 665 [32] Daniel D Lee and H Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances*
666 *in neural information processing systems*, pages 556–562, 2001. [1](#), [2.2.2](#), [2.2.3](#), [2.2.3](#), [3.2](#), [3](#)
- 667 [33] Weihui Liang and Xudong Yang. Indoor formaldehyde in real buildings: Emission source identification,
668 overall emission rate estimation, concentration increase and decay patterns. *Building and Environment*,
669 69:114–120, 2013. [1](#)
- 670 [34] A Limem, G Delmaire, Matthieu Puigt, G Roussel, and D Courcot. Non-negative matrix factorization
671 under equality constraints—a study of industrial source identification. *Applied Numerical Mathematics*,
672 85:1–15, 2014. [2.2.1](#)
- 673 [35] Abdelhakim Limem, Gilles Delmaire, Matthieu Puigt, Gilles Roussel, and Dominique Courcot. Non-
674 negative matrix factorization using weighted beta divergence and equality constraints for industrial
675 source apportionment. In *Machine Learning for Signal Processing (MLSP), 2013 IEEE International*
676 *Workshop on*, pages 1–6. IEEE, 2013. [2.2.1](#)
- 677 [36] Chih-Jen Lin. On the convergence of multiplicative update algorithms for nonnegative matrix factor-
678 ization. *IEEE Transactions on Neural Networks*, 18(6):1589–1596, 2007. [2.2.3](#)
- 679 [37] Haitao Liu, Chongguo Tian, Zheng Zong, Xiaoping Wang, Jun Li, and Gan Zhang. Development and
680 assessment of a receptor source apportionment model based on four nonnegative matrix factorization
681 algorithms. 197:159–165. [2.2.1](#)
- 682 [38] Christopher M Long, Helen H Suh, Lester Kobzik, Paul J Catalano, Yao Yu Ning, and Petros Koutrakis.
683 A pilot investigation of the relative toxicity of indoor and outdoor fine particles: in vitro effects of
684 endotoxin and other particulate properties. *Environmental Health Perspectives*, 109(10):1019, 2001. [1](#)
- 685 [39] Marianna Luoma and Stuart A Batterman. Characterization of particulate emissions from occupant
686 activities in offices. *Indoor air*, 11(1):35–48, 2001. [1](#)

- 687 [40] William W Nazaroff. Exploring the consequences of climate change for indoor air quality. *Environmental Research Letters*, 8(1):015022, 2013. [1](#), [2.2.1](#)
688
- 689 [41] Ioannis Nezis, George Biskos, Konstantinos Eleftheriadis, and Olga-Ioanna Kalantzi. Particulate matter and health effects in offices-a review. 156:62–73. [1](#)
690
- 691 [42] S Oeder, S Dietrich, I Weichenmeier, W Schober, G Pusch, RA Jörres, R Schierl, D Nowak, H Fromme, H Behrendt, et al. Toxicity and elemental composition of particulate matter from outdoor and indoor air of elementary schools in munich, germany. *Indoor air*, 22(2):148–158, 2012. [1](#)
692
693
- 694 [43] Rachid Ouaret, Anda Ionescu, Viorel Petrehus, Yves Candau, and Olivier Ramalho. Spectral band decomposition combined with nonlinear models: application to indoor formaldehyde concentration forecasting. 32(4):985–997. [2.1.3](#)
695
696
- 697 [44] Rachid Ouaret, Anda Ionescu, Viorel Petrehus, Yves Candau, and Olivier Ramalho. Particulate matter variability sources in an open-plan office: comparison of two monitoring campaigns. In *European Aerosol Conference*, page 1, Tours, France, September 2016. [2.2.1](#)
698
699
- 700 [45] Rachid Ouaret, Anda Ionescu, Olivier Ramalho, and Yves Candau. Indoor air pollutant sources using blind source separation methods. In *25th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning.*, Bruges, Belgium, April 2017. [2.2.1](#)
701
702
- 703 [46] Rachid Ouaret, Anda Ionescu, Olivier Ramalho, Yves Candau, Evelyne Gehin, and Viorel Petrehus. Analysis of the temporal variability of indoor particulate matter concentrations using Blind Source Separation methods: a comparative study. In *International Aerosol Conference*, page 1, Busan, South Korea, August 2014. [2.2.1](#)
704
705
706
- 707 [47] Pentti Paatero and Unto Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2):111–126, 1994. [1](#), [2.2.1](#),
708
709 [2.2.2](#)
- 710 [48] Pentti Paatero, Unto Tapper, Pasi Aalto, and Markku Kulmala. Matrix factorization methods for analysing diffusion battery data. 22:S273–S276. [2.2.1](#), [2.2.2](#)
711
- 712 [49] Pallavi Pant. *Receptor modelling studies of airborne particulate matter in the United Kingdom and India*. PhD thesis, University of Birmingham, 2014. [1](#)
713

- 714 [50] Marc Plouvin, Abdelhakim Limem, Matthieu Puigt, Gilles Delmaire, Gilles Roussel, and Dominique
715 Courcot. Enhanced nmf initialization using a physical model for pollution source apportionment. In
716 *22nd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine*
717 *Learning (ESANN 2014)*, pages 261–266, 2014. [2.2.1](#), [2.2.4](#)
- 718 [51] Jing Qian and Andrea R Ferro. Resuspension of dust particles in a chamber and associated environ-
719 mental factors. 42(7):566–578. [3.2.2](#)
- 720 [52] Jing Qian, Andrea R Ferro, and Kathleen R Fowler. Estimating the resuspension rate and residence
721 time of indoor particles. 58(4):502–516. [3.2.2](#)
- 722 [53] Jing Qian, Jordan Peccia, and Andrea R Ferro. Walking-induced particle resuspension in indoor
723 environments. *Atmospheric Environment*, 89:464–481, 2014. [1](#)
- 724 [54] R Core Team. *R: A language and environment for statistical computing*, 2019. [3](#)
- 725 [55] Carolin Rösch, Tibor Kohajda, Stefan Röder, Martin von Bergen, and Uwe Schlink. Relationship
726 between sources and patterns of vocs in indoor air. *Atmospheric Pollution Research*, 5(1):129–137,
727 2014. [2.2.1](#)
- 728 [56] Heidi Salonen, Tunga Salthammer, and Lidia Morawska. Human exposure to ozone in school and office
729 indoor environments. *Environment international*, 119:503–514, 2018. [1](#)
- 730 [57] Norbert Serfozo, Sofia Eirini Chatoutsidou, and Mihalis Lazaridis. The effect of particle resuspension
731 during walking activity to pm 10 mass and number concentrations in an indoor microenvironment.
732 *Building and Environment*, 82:180–189, 2014. [1](#)
- 733 [58] Ken Sexton and Steven B Hayward. Source apportionment of indoor air pollution. 21(2):407–418. [1](#)
- 734 [59] A Thiem, U Schlink, X-C Pan, M Hu, A Peters, A Wiedensohler, S Breitner, J Cyrus, B Wehner,
735 C Rösch, et al. Using non-negative matrix factorization for the identification of daily patterns of
736 particulate air pollution in beijing during 2004–2008. *Atmospheric Chemistry and Physics Discussions*,
737 12(5):13015–13052, 2012. [1](#), [2.2.1](#), [2.2.4](#)
- 738 [60] George D Thurston and John D Spengler. A quantitative assessment of source contributions to inhal-
739 able particulate matter pollution in metropolitan boston. *Atmospheric Environment (1967)*, 19(1):9–
740 25, 1985. [2.2.1](#)

- 741 [61] Bruce A Tichenor, Leslie A Sparks, James B White, and Merrill D Jackson. Evaluating sources of
742 indoor air pollution. 40(4):487–492. [1](#)
- 743 [62] Eus JW Van Someren, Kim Dekker, Bart HW Te Lindert, Jeroen S Benjamins, Sarah Moens, Filippo
744 Migliorati, Emmeke Aarts, and Sophie van der Sluis. The experienced temperature sensitivity and
745 regulation survey. 3(1):59–76. [2.1.3](#)
- 746 [63] M Viana, TAJ Kuhlbusch, X Querol, A Alastuey, RM Harrison, PK Hopke, W Winiwarter, M Vallius,
747 S Szidat, ASH Prévôt, et al. Source apportionment of particulate matter in europe: a review of
748 methods and results. *Journal of Aerosol Science*, 39(10):827–849, 2008. [1](#)
- 749 [64] Lance Wallace and Wayne Ott. Personal exposure to ultrafine particles. *Journal of Exposure Science
750 and Environmental Epidemiology*, 21(1):20–30, 2011. [1](#)
- 751 [65] Chao Wang, Xudong Yang, Jun Guan, Zheng Li, and Kai Gao. Source apportionment of volatile
752 organic compounds (vocs) in aircraft cabins. 81:1–6. [2.2.1](#)
- 753 [66] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. [3](#)
- 754 [67] Elen Yakovleva, Philip K Hopke, and Lance Wallace. Receptor modeling assessment of particle total
755 exposure assessment methodology data. *Environmental Science & Technology*, 33(20):3645–3652, 1999.
756 [1](#)
- 757 [68] Xudong Yang. *Study of building material emissions and indoor air quality*. PhD thesis. [1](#)
- 758 [69] Zhonglong Zheng, Jie Yang, and Yitan Zhu. Initialization enhancer for non-negative matrix factoriza-
759 tion. *Engineering Applications of Artificial Intelligence*, 20(1):101–110, 2007. [3.2](#)
- 760 [70] Janez Žibert, Jure Cedilnik, and Jure Pražnikar. Particulate matter (pm10) patterns in europe: An
761 exploratory data analysis using non-negative matrix factorization. *Atmospheric Environment*, 132:217–
762 228, 2016. [2.2.1](#)