Atmospheric Pollution Research 5 (2014) 129-137

Atmospheric Pollution Research

www.atmospolres.com

Relationship between sources and patterns of VOCs in indoor air

Carolin Rösch [1], Tibor Kohajda [2,3], Stefan Röder [4], Martin von Bergen [2,5], Uwe Schlink [1]

[1] Department of Urban and Environmental Sociology, Helmholtz–Centre for Environmental Research–UFZ, Permoserstrasse 15, 04318 Leipzig, Germany
[2] Department Metabolomics, Helmholtz–Centre for Environmental Research–UFZ, Permoserstrasse 15, 04318 Leipzig, Germany
[3] Department Monitoring & Exploration Technologies, Helmholtz–Centre for Environmental Research–UFZ, Permoserstrasse 15, 04318 Leipzig, Germany
[4] Department Core Facility Studies, Helmholtz–Centre for Environmental Research–UFZ, Permoserstrasse 15, 04318 Leipzig, Germany
[5] Department of Biotechnology, Chemistry and Environmental Engineering, Aalborg University, Fredrik Bajers Vej 5, 9220 Aalborg Ø,Denmark

ABSTRACT

People spend most of their daytime in indoor environments. Their activities influence the composition of the indoor air by emitting volatile organic compounds (VOCs). The increasing number of different VOCs became the focus of attention in recent years as the question arises from the relationship between exposure to air pollutants and diseases. The present study of flats in Leipzig (Germany) is based on measurements of 60 different VOCs and is unique in the field of indoor air quality due to its enormous size of samples (n=2 242) and questionnaire data. The main purpose of our analysis was to identify the sources and patterns that characterize airborne VOCs in occupied flats. We combined two methods, principal components analysis (PCA) and non–negative matrix factorization (NMF), to assign compounds to their origin and to understand the coinstantaneous existence of several VOCs. PCA clustering provided a source apportionment and yielded 10 principal components (PCs) with an explained variance of 72%. However, real indoor air quality is often affected by combined sources. NMF reveals characteristic compositions of VOCs in indoor environments and emphasizes that constantly recurring structures are not single sources, but rather fusions of them, so called patterns. Interpreting these sources, we realized that homes were strongly influenced by ventilation, human activities, furnishings, natural processes (such as solar radiation) or their combinations. The very large set of samples and the combination with questionnaires applied on this comprehensive assessment of VOCs allows generalizing the results to homes in middle–scale cities with minor industrial pollution. As a conclusion, single VOC–dose–response relationships are inopportune for situations when indoor sources occur in combination. Further studies are necessary to assess associated health risks.

Keywords: *Volatile organic compounds, non–negative matrix factorization (NMF), PCA, pattern*

Corresponding Author:
Carolin Rösch
☎ : *+49-341-235-1548*
🖷 : *+49-341-235-4518*
✉ : *carolin.roesch@ufz.de*
uwe.schlink@ufz.de

Article History:
Received: 05 August 2013
Revised: 11 November 2013
Accepted: 12 November 2013

doi: 10.5094/APR.2014.016

1. Introduction

Due to the increased time periods spent in different interior spaces in recent years, indoor air quality became a matter of particular interest (Ayoko, 2004) and affects behavior, e.g. ventilation (Qian et al., 2010), and the health of people (Rumchev et al., 2004; Arif and Shah, 2007; Billionnet et al., 2011).

Indoor air quality is extremely variable and depends on activities of the people (Morawska et al., 2003; Edwards et al., 2006; Eklund et al., 2008; Buonanno et al., 2009; Buonanno et al., 2012), home furnishings (Yrieix et al., 2010), building materials (Missia et al., 2010) and season (Schlink, 2004). Current research is involved with the constantly rising amount of sources, the complexity of mixtures, and the role of outdoor air (Carslaw et al., 2009). The diversity of compounds, their variable toxicity and the addressed peer group complicate the determination of guidelines for concentrations of volatile organic compounds (VOCs) in indoor environments.

The formation of VOCs in indoor environments is difficult to understand and to reconstruct (e.g. in experiments). On the one hand, compounds originate exclusively from indoor sources (a point of origin of gases or other materials, which appears constantly in a similar way) and, on the other hand, they are formed by mixtures of indoor and outdoor pollutants. In most cases, indoor VOC concentrations are significantly higher than outdoor levels (Batterman et al., 2007). This is influenced by type and age of building materials (Missia et al., 2010) and personal activities, e.g. renovation processes, that cause elevated levels. Increased levels occur directly after renovations and then normalize to lower concentrations (Jia et al., 2008a; Herbarth and Matysik, 2010). Seasonal variations cause higher indoor levels to accumulate due to abated ventilation in winter (Dodson et al., 2008; Matysik et al., 2010). Furthermore, local conditions, such as industry or busy roads, create emission sources that differentiate the pollution amount of homes in industrial, urban, and nonurban regions (Jia et al., 2008b). This high variety of possible sources in indoor and ambient air poses a big challenge for scientists to assign different compounds to their point of origin.

By retracing compounds to their origins, emission sources can be recognized and eliminated in order to protect the human health. Several methods for indoor and outdoor source apportionment are possible, e.g. chemical mass balance (Badol et al., 2008; Gokhale et al., 2008) and positive matrix factorization–PMF (Cai et al., 2010; Pindado and Perez, 2011), although PCA and related procedures were mainly used for indoor air (Jia et al., 2008b; Ohura et al., 2009; Jo and Kim, 2010; Guo, 2011). The majority of studies concern source apportionment of outdoor air. The high variability of indoor VOC combinations, caused by various activities, differing indoor equipment and its age, complicates source apportionment and pattern recognition. Lately, PMF is the new means of choice through its positive character, producing stable results for small data sets and the independence on source strength in contrast to PCA (Chan et al., 2011; Demir et al., 2012). The elaborate correction of zero values, that changes the original data, and the creation of error estimates brings disadvantages (Pekey et al., 2013) and makes other methods more suitable, e.g. non–negative matrix factorization.

As a consequence of the studies by Han et al. (2011) and (2012), it seems more appropriate to consider mixtures of VOCs from multiple sources, because various source combinations of numerous air pollutants, including interactions as well as superposition of compounds, possess different effects on human health.

We suppose in the case of VOCs that items, activities, and buildings emit a fixed compound spectra. Depending on the VOC lifetime and the emission strength, the source intensity is variable but usually the composition is relatively constant. These so called "patterns" describe recurring structures or regular sequences, characterized by the way in which something occurs and may contain different sources or sub–patterns (Oxford Dictionaries, 2013).

For this VOC analysis both terms, sources and patterns were used, analyzed and connected owing to the fact that the link between "single VOC/indoor activities" (Shin and Jo, 2013) as well as "indoor activities/disease outcomes" (Herbarth et al., 2006) produced promising results, but a direct link of "single VOC/disease outcome" is debated and limited to a restricted number of VOCs (Diez et al., 2000; Rumchev et al., 2004; Fuentes–Leonarte et al., 2009). Main reasons might be that indoor air is influenced by its sources. Just a minority of sources emit single VOCs while the majority give rise to various compounds, and that source emissions interfere to pollution patterns. The innovation of this study is the identification of the most frequent patterns in homes that might be more harmful to human health than single VOCs.

Many studies on indoor air suffer from small sample sizes. This study analyses 2 242 measurements of more than 60 VOCs of 622 homes in Leipzig. The high number of included VOCs improves the matching of predicted model values to real measured concentrations. Additional questionnaires helped to differentiate between several living spaces and their effects on air quality.

At first, we identified the sources of VOCs in indoor environments by means of principal component analysis (PCA), which relates increased compound concentrations with normal average levels of homes and aims to discover harmful VOC sources caused by occupant activities and natural processes. Secondly, non–negative matrix factorization (NMF) was applied to the compound data set to find recurrent structures of air pollutants in homes. The interpretation of sources and patterns was supported by regression analysis.

Differences between sources and patterns might be caused by the analyzing techniques, PCA and NMF. Several sources can be combined to patterns but not vice versa. We found astonishing results on recovering PCA sources in NMF patterns. The easement of pattern description argues for the combination of both methods, PCA and NMF.

2. Material and Methods

2.1. Sample collection and measurement site

This study was conducted in Leipzig, central Germany. Approximately 523 000 people live in this midscale city with a small industrial impact. Passive sampling with Organic Vapor Monitors (3M, OVM 3500) was used to gain information about indoor compound levels (Begerow et al., 1995). The VOCs are collected by the principle of diffusion on a single charcoal sorbent wafer that was placed at a height of 1.5 to 2 m in the middle of the rooms. VOC data was collected in the mother–child study LINA (Lifestyle and environmental factors and their Influence on new–born's Allergy risk), which is an on–going birth cohort study in Leipzig. The study included 622 homes of pregnant mothers, who were recruited from May 2006 to December 2008. To get a wide variety of indoor activities, with emissions covering a large range of VOCs (e.g. smoking or recent renovation), homes were chosen randomly. Samples were collected over a period of approximately 4 weeks in a room where the child spent most of its time (preferential living or child's room). Dwellers were asked to keep their usual behaviors in order to reproduce typical indoor environments. 3M–samplers were then returned to laboratory for GC/MS analysis using previously reported extraction methods (Matysik et al., 2010). Sampling was performed through the whole year (summer: 1 140, winter: 1 102).

The VOC dataset contained 2 246 measurements of 61 VOCs (Table 1). Missing values were replaced by VOC–specific half detection limits (HDL) and VOCs were included into analyses as long as the detection frequency exceeded 70%. The total dimension of the additive questionnaire involved 2 242 cases.

2.2. Questionnaires

The housings were characterized by recent renovations (53.6%), painting of walls (62.6%) and arrangements of new furniture (68.3%). Further features with an expected long–term influence on VOC concentration, such as new flooring (19.6%) or smoking (3.9%), were rare. Lower importance proved questions for the use of solvents, mothballs or cleansing agents (see Figure 1). Small response to questions was a reason for attaching little importance to their results in multivariate regression (all items <10%).

The majority of homes were close to roads with residential traffic (78.2%); almost one third was located next to streets with transit traffic (31.4%). Traffic exhaust may be the most effecting outdoor source of Leipzig due to the fact that heavy industry was not located near by the housings.

2.3. Principle component analysis (PCA)

PCA is a full spectral dimensionality reduction method and uses Euclidean distances to classify the measured data (van der Maaten, 2009). The analysis identified special sources in environments which may be detectable in numerous housings in medium–sized cities like Leipzig. All presented results of PCA followed from Varimax–rotation of logarithmically transformed VOC concentrations. The number of factors was invigorated by eigenvalues >1 with at least one variable covering a factor loading >0.5. Scree plots were further checked. The algorithms of PCA were executed with STATISTICA 10 (StatSoft, Inc. 1984–2011) and Matlab Version 7.12.0.635 (Mathworks, 1984–2011).

2.4. Non–negative matrix factorization (NMF)

NMF is a method that was first implemented in computer science to characterize pictures by dividing images in rows and columns. Through iterative minimizing algorithms, the factorization of the input matrix V (dimension: m×n) ends in two matrices W (m×r) and H (r×n) with smaller dimensions. In the case of VOC patterns, W is the pattern of VOC in indoor environments and H reflects the weight of each pattern. Every vector of V can be represented by linear combinations of W and H, which is very specific for NMF. The reduction of dimension and the final number of factors (r) must be chosen intuitively. In this case, the number of separated factors from PCA was tested.

Table 1. *Standard deviation, mean, median, minimum and maximum concentrations (μg/m³) of each compound used in NMF and PCA*

Species	Abbreviation	Mean	Median	S.D.	Min.	Max.
1-Butanol	C1	8.08	7.45	7.02	0.0200	115.26
2-Ethyl-1-Hexanol	C2	6.83	5.53	6.82	0.3055	153.59
2-Methyl-1-Propanol	C3	1.32	0.68	2.87	0.0221	57.49
Ethyl acetate	C4	10.86	4.74	27.03	0.0250	597.94
n-Butyl acetate	C5	6.23	2.76	16.42	0.0098	449.92
Hexane	C6	2.40	1.12	6.59	0.0231	145.82
Heptane	C7	4.20	1.20	16.84	0.0154	407.52
Octane	C8	1.25	0.56	3.77	0.0373	73.40
Nonane	C9	1.60	0.56	6.60	0.0221	198.93
Decane	C10	4.33	1.40	18.82	0.0119	527.55
Undecane	C11	3.81	1.26	13.54	0.0138	342.31
Dodecane	C12	3.35	1.83	7.06	0.0151	132.23
Tridecane	C13	1.78	0.85	5.20	0.0191	165.41
Tetradecane	C14	6.16	2.18	20.51	0.1545	467.28
Pentadecane	C15	3.91	0.93	17.29	0.0469	356.11
Hexadecane	C16	3.71	1.46	12.42	0.0175	194.77
Methylcyclopentane	C17	1.31	0.42	5.18	0.0326	129.84
Cyclohexane	C18	1.60	0.41	11.29	0.0113	342.77
Tetrahydrofuran	C19	0.45	0.14	1.02	0.0072	18.66
Texanol isobutyrate	C20	1.63	0.60	5.82	0.0084	133.59
Tetrachloromethane	C21	0.40	0.33	1.13	0.0208	45.42
Trichlorethylene	C22	0.17	0.11	0.47	0.0171	16.22
Chlorobenzene [a]	C23	2.93	1.98	2.18	0.0000	22.70
Benzene	C24	1.51	1.09	1.76	0.0077	31.57
Toluene	C25	13.18	8.06	16.91	0.0590	249.55
Ethylbenzene	C26	1.51	0.90	2.59	0.0724	47.76
m-p-Xylene	C27	3.27	1.84	7.25	0.0091	174.33
Styrene	C28	0.83	0.37	1.90	0.0137	39.71
o-Xylene	C29	0.97	0.61	1.83	0.0079	47.46
Isopropylbenzene	C30	0.23	0.13	0.51	0.0063	18.67
Propylbenzene	C31	0.46	0.30	0.83	0.0105	20.26
4-Ethyltoluene	C32	0.86	0.45	2.03	0.0084	52.21
1,3,5-Trimethylbenzene	C33	0.41	0.21	1.13	0.0081	35.57
1,2,4-Trimethylbenzene	C34	1.35	0.71	3.78	0.0079	126.07
1,2,3-Trimethylbenzene	C35	0.40	0.20	1.07	0.0075	31.74
Pentanal	C36	1.71	1.08	2.16	0.0184	33.92
Hexanal	C37	4.22	2.47	5.68	0.0370	110.09
Benzaldehyde	C38	1.40	1.05	1.38	0.0197	28.47
Octanal	C39	1.19	0.82	1.72	0.0213	50.69
Nonanal	C40	2.43	1.06	15.04	0.0392	598.09
Methylisobutylketone	C41	0.63	0.28	1.91	0.0087	49.47
3-Heptanone	C42	0.70	0.51	1.07	0.0119	31.87
Cyclohexanone	C43	1.13	0.52	1.83	0.0458	20.20
Acetophenone	C44	0.26	0.19	0.30	0.0066	5.57
2-Heptanone	C45	0.55	0.44	0.45	0.0095	7.90
α-Pinene	C46	31.69	15.53	51.93	0.0117	854.27
β-Pinene	C47	3.69	1.84	13.16	0.0119	575.59
δ-3-Carene	C48	15.54	7.06	25.34	0.0189	303.54
Limonene	C49	28.31	13.03	43.76	0.0730	641.97
Longifolene	C50	0.71	0.48	0.84	0.0566	11.45

[a] *Total number of measurements is smaller than 2 242, and VOCs were excluded from analysis*

To receive stable results, initialization matrices for *W* and *H* were produced by singular value decomposition (SVD) (Boutsidis and Gallopoulos, 2008). Otherwise, inconsistent and irreproducible results would be created due to permanently new calculated starting matrices. With SVD, initialized matrices converge to the given and always identical local minimum. Initialization is followed by iterative calculation of reconstructed matrices *W* and *H* with alternating least squares and without the requirement of orthogonality.

The non–negativity of the data matrix is the main advantage of this application in comparison to positive matrix factorization because it is often not feasible to gain measurement results without zero entries. Furthermore, physical processes and source contribution are usually not orthogonal what argues for an application of NMF in addition to PCA. The calculation of VOC patterns was obtained by using MathWorks Matlab, version 7.10.0.499.

3 Results and Discussion

3.1. Source identification via PCA

PCA was used to define frequently occurring VOCs in indoor environments and to apportion them to household products and indoor activities (arisen from questionnaires). Due to normal distribution of input data for PCA, the VOC concentrations were logarithmically transformed. For detailed results of elemental analysis and PCA, we refer to the Supporting Material (SM).

Reproducibility was 72% and contained the following sources: ventilation and season (18.7%), which can include intrusions of vehicle emissions and environmental tobacco smoke. Further, wooden furniture/parquet (8.2%); flooring, wallpapers, and gluing emissions (8.2%), as well as recent renovations (7.6%) accounted for a high percentage of the variability of compound levels. Natural processes, e.g. solar radiation, aging of materials, and indoor climate added nearly 16.3%; the background concentration contributed to 4.2% (Table 2).

3.2. Results of NMF

NMF was used to quantify the existence of special patterns in typical indoor environments after rearrangement procedures and to research the difference to PCA. Concentrations of compounds were logarithmically transformed for comparability with PCA results.

The results of PCA and regression analysis as well as comparison with literature helped to identify the different patterns. A table including the percentage share of each compound in all patterns is given in the SM.

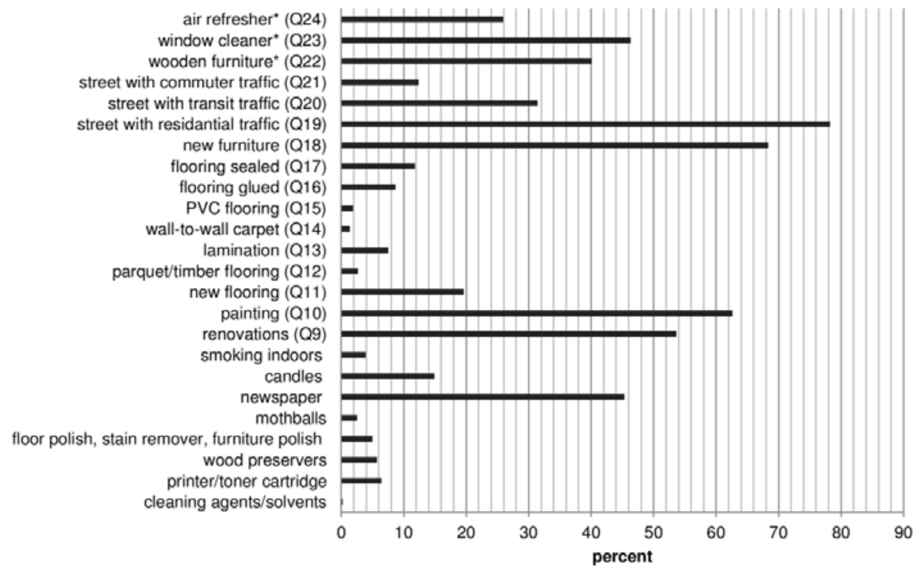

Figure 1. *Characterization criterions of housings that were inquired in LINA-study, given in percent. Abbreviations in brackets are used in Table S1 of Supporting Material. Total number of questionnaire items amounts to N(max)=2 242, items with * were collected in 1 632 cases.*

Table 2. *Sources of VOCs in indoor environments developed by PCA*

Factors	% Variance PCA	Possible source	Included compounds
PCA1	18.75	Ventilation/season	Aromatic hydrocarbons, nonane
PCA2	8.72	Wood products (furniture, parquet), personal care and cleansing products	Terpenes, 2-heptanone
PCA3	8.16	Solvents, adhesive emissions, PVC	Hexane, heptane, methylcyclopentane, cyclohexane
PCA4	7.6	Renovation, paints/traffic emissions	Tri-, tetra-, penta-, hexadecane
PCA5	7.17	Solar radiation/secondary emissions	Texanol isobutyrate, benzaldehyde, cyclohexanone
PCA6	6.48	Renovation, new furniture/traffic emissions	Decan, un-, do-, tridecan
PCA7	5.27	Aging of materials	Pentanal, hexanal, octanal, nonanal
PCA8	4.23	Background	Tetrachloromethane
PCA9	3.84	Indoor climate	1-Butanol, 2-ethylhexanol
PCA10	2.81	PVC flooring	Tetrahydrofuran

Pattern 1 is strongly influenced by 1–butanol, pentanal, hexanal, octanal and nonanal. These aldehydes can be emitted by linoleum floor coverings, floor lacquers, and aging of materials (Wolkoff, 1995). The combined occurrence of these sources is obvious, because flooring is a long–lasting source, which emits specific compounds in low levels over a long period of time. Emission of VOCs from aging materials was observed after thoroughly drying, also.

The almost balanced contribution of alkanes in combination with aromatic hydrocarbons in pattern 2 may follow from lowered ventilation during and after renovation and painting of walls. The portion of aromatic hydrocarbons is high and correlates with ventilation behavior and renovation in regression analysis. The classification of this pattern to smoker households is inadequate due to a low number of smoker households.

NMF found a pattern of indoor air that is influenced, in most instances, by alkanes due to renovation activities (pattern 3).

Pattern 4 shows a combination of terpenes and alkanes, a pattern that is conceivable for households with renovation activities and arrangement of new furniture. In this study, the majority of the expecting parents had combined activities [renovation and arrangement of new furniture: 963 instead of having each separately (renovation without furniture: 239; furniture without renovation: 569)]. The usage of cleansing agents and air fresheners after renovation is increased, so the level of terpenes might be lifted, additionally.

Pattern 5 is dominated by terpenes but equal contributions of alkanes and aromatics are observable. These compounds are emitted in indoor environments by renovations/painting of walls (alkanes), arrangement of furniture/laying of parquet (terpenes) and ventilation/traffic emission (aromatics). Elevated concentrations of alkanes, aromatics and terpenes might occur when cleaning was carried out, but in most cases, renovation events are followed by ventilation that diminishes the terpene level due to chemical reactions with ozone and air mass transport (Morawska et al., 2009; Salthammer and Bahadir, 2009). As we found, a high number of study participants renovated their homes in winter months when ventilation is reduced, which explains the shape of pattern 5.

2–ethyl–1–hexanol, 2–methyl–1–propanol and ethyl acetate are crucial for pattern 6 and appear in elevated levels after renovation activities and in conjunction with laying of PVC flooring and its sealing. Emission from PVC flooring is arguable due to a small number of participants who chose this type of flooring. The number of sealed floorings is five times higher but not reliable either.

Pattern 7 is affected by high contributions of hexane, heptane, methylcyclopentane and cyclohexane, which accumulate through gluing emissions and renovation in homes.

Factor 8 includes a small proportion of VOCs, specific for renovations (alkanes, C_{10}–C_{16}), and a higher amount of compounds related to aging of materials (pentanal, hexanal, octanal, nonanal).

Styrene, pentanal, hexanal and methylisobutylketone define the shape of pattern 9. These VOCs are affected by solar radiation, arise from aging of materials, and ventilation. The combination of solar radiation and aging of materials is not astonishing because solar radiation and increased temperatures may be correlated and influence the aging of materials.

The last extracted pattern is clearly influenced by tetrahydrofuran, mainly emitted by PVC, and a small contribution of terpenes. This mixture might occur due to cleaning activities after the laying of PVC–flooring or the usage of air fresheners to minimize the olfactory irritation of the new flooring.

3.3. Combination of PCA and NMF results

PCA showed more or less special compositions of compounds, which are thought to be emitted by a single or by inseparable sources, which emit the same compounds in similar concentrations (Geng et al., 2009). It is not possible to differentiate several sources with equal VOC spectra. For example, accumulations of aromatics were found to be characteristic for smoking indoors, solvents of paints, or traffic emissions. Hence, this fact allows only an assumption about sources and patterns. However, it was noticeable, that almost all factors of PCA were contained in the patterns of NMF, so that patterns seem to be modifications and mixtures of sources. Due to this reason, we tested the correlation of NMF patterns and PCA factors and linked NMF and PCA via correlation coefficients (CCs). This resulted in some NMF patterns, having high CCs for several PCA sources (NMF1, NMF2, NMF3 and NMF8). In contrast, some NMF patterns only correlated with single PCA components, e.g. NMF9 and NMF10 (see Table 3).

The CCs showed that the linear relationship between single patterns and sources is barely reproducible with two influencing items. Therefore, two regression analyses were conducted, one with PCA sources as dependent variable and the second one with NMF patterns as dependent variable.

The linear regression with PCA factors as dependent variables failed as it did not reveal statistically significant findings suggesting that it is unsuitable for describing PCA components as a combination of NMF factors.

Table 3. *Correlation coefficients (CCs) for NMF patterns and PCA factors*

	NMF1	NMF2	NMF3	NMF4	NMF5	NMF6	NMF7	NMF8	NMF9	NMF10
PCA1	-0.36 [a]	0.84 [b]	-0.35 [a]	-0.09	0.44 [a]	-0.10	0.19	-0.48 [a]	-0.01	-0.04
PCA2	-0.32 [a]	-0.23	-0.33 [a]	0.71 [b]	0.70 [b]	0.06	-0.15	-0.35 [a]	0.01	0.14
PCA3	-0.41 [a]	0.46 [a]	-0.40 [a]	-0.11	0.17	0.08	0.93 [b]	-0.47 [a]	-0.21	-0.01
PCA4	0022	-0.30 [a]	0.92 [b]	-0.15	0.00	-0.30 [a]	-0.30 [a]	0.14	0.01	-0.13
PCA5	-0.34 [a]	-0.28	-0.08	-0.07	-0.17	0.37 [a]	-0.09	-0.08	0.72 [b]	-0.06
PCA6	-0.19	0.27	0.21	0.48 [a]	-0.16	-0.09	-0.18	-0.11	-0.22	-0.14
PCA7	0.63 [b]	-0.36	-0.15	0.01	-0.42 [a]	-0.16	-0.31 [a]	0.86 [b]	0.23	-0.28
PCA8	-0.17	-0.02	-0.13	-0.13	0.33 [a]	0.45 [a]	0.01	-0.12	-0.12	0.02
PCA9	0.41 [a]	-0.13	0.01	-0.07	-0.11	0.16	0.07	-0.19	-0.05	-0.10
PCA10	-0.04	0.00	-0.15	-0.17	-0.08	0.09	-0.17	-0.21	-0.12	0.78 [b]

[a] *Significance levels of p≤0.05*

[b] *Significant CCs higher than 0.5*

The contrary analysis with PCA sources as independent variables showed astonishing results (Table 4). Each pattern of NMF was calculated by a linear combination of all PCA factors with positive and negative coefficients, as output of the linear regression analysis.

$$Y(\mathrm{NMF})=\beta_0+\beta_1 x_{(\mathrm{PCA1})}+\beta_2 x_{(\mathrm{PCA2})}+\ldots+\beta_{10} x_{(\mathrm{PCA10})} \quad (1)$$

The analysis demonstrated that nearly all single sources, found by PCA, were observable in the patterns of NMF. In addition, it highlights that the best recovery of each NMF pattern is reached when multiple sources were combined. Table 4 shows the different influences of one source on various patterns (from top to bottom) and the mixture of sources, which create the shape of the pattern (from left to right hand side) by their individual strength.

Figure 2 shows the representation of NMF factors with varying PCA coefficient factors. For this figure, the calculated NMF patterns out of all PCA factors were used and the R^2 correspond to the values in Table 4.

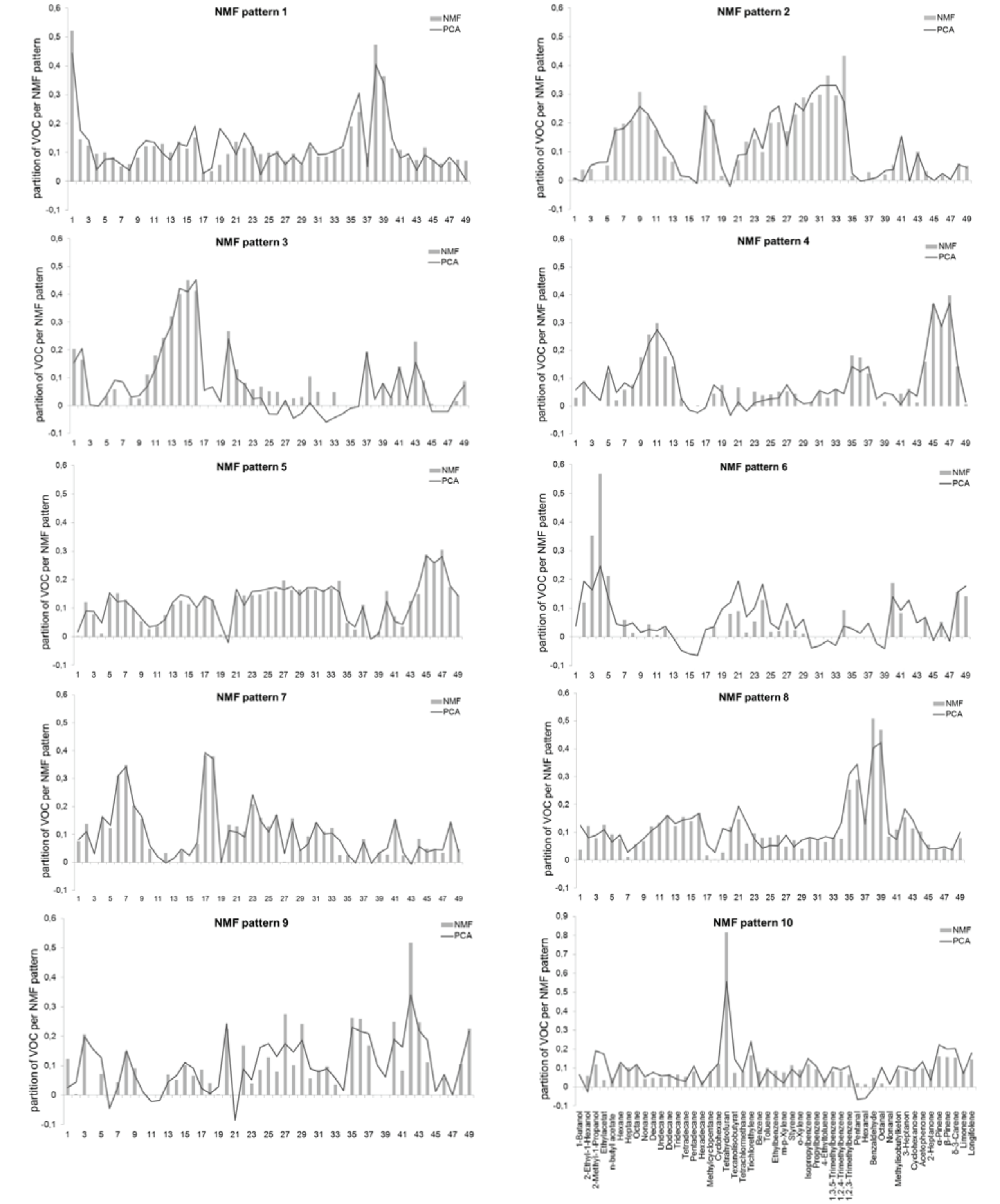

Figure 2. *NMF patterns represented by the share of each VOC (grey bars) and by their reconstruction as a linear-combination of 10 principal components (PCA, black line). The numbers at the abscissa represent the VOCs. Full names are given at the bottom right panel.*

Table 4. *The R^2 of linear regression analysis by taking all PCA factors into account in an additive way*

NMF factor	R^2 (%)	Contribution of PCA factors (%)									
		1	2	3	4	5	6	7	8	9	10
1	84	4.57^{d}	9.61^{c}	15.69^{c}	6.47^{a}	13.91^{c}	3.05^{d}	14.58^{c}	5.78^{d}	20.02^{c}	6.33^{d}
2	88	34.15^{c}	13.53^{c}	16.73^{c}	1.81^{d}	10.70^{b}	7.79^{a}	2.40^{d}	1.88^{d}	9.91^{a}	1.10^{d}
3	95	9.14^{c}	11.84^{c}	10.47^{c}	28.11^{c}	2.28^{d}	5.86^{b}	15.17^{c}	1.28^{d}	4.36^{a}	11.49^{c}
4	90	8.82^{c}	27.61^{c}	4.03^{d}	8.24^{c}	10.82^{c}	19.92^{d}	0.17^{d}	10.55^{c}	3.41^{d}	6.43^{a}
5	91	11.82^{c}	26.23^{c}	6.28^{b}	10.05^{c}	9.33^{c}	11.05^{d}	4.32^{d}	8.83^{c}	7.86^{a}	4.23^{d}
6	40	10.92^{a}	7.42^{d}	6.98^{d}	16.09^{a}	15.17^{b}	2.77^{d}	15.71^{a}	18.68^{b}	3.09^{d}	3.18^{d}
7	96	5.54^{c}	7.27^{c}	30.48^{c}	5.62^{c}	5.65^{c}	10.87^{d}	13.05^{c}	3.33^{d}	3.82^{b}	14.37^{c}
8	84	5.31^{d}	13.52^{c}	15.15^{b}	1.38^{d}	6.88^{d}	3.38^{d}	35.13^{c}	4.82^{d}	5.66^{d}	8.77^{d}
9	62	12.55^{b}	0.07^{d}	0.62^{d}	9.91^{a}	36.37^{c}	4.80^{d}	17.58^{b}	0.25^{d}	8.41^{d}	9.44^{d}
10	65	6.40^{d}	7.49^{d}	1.95^{d}	5.71^{d}	0.38^{d}	4.88^{d}	12.15^{d}	6.81^{d}	10.82^{a}	43.41^{c}

[a] *Significance level $0.05 \geq p > 0.01$*
[b] *$0.01 \geq p > 0.001$*
[c] *$p \leq 0.001$*
[d] *$p > 0.05$*

Hence, we found a mathematical way to identify NMF factors as patterns, resulting from various combinations of PCA factors, which can be described as underlying sub–patterns or sources. The best model had a recovery efficiency of 96%, whereas the least effective one had just about 40%. This might be caused by the methodical differences and, in particular, by the exclusion of negative values from NMF. The link between sources and patterns is the individual impact of activities and environmental factors to the whole indoor environment.

3.4. Discussion

PCA is a standard method for dimensionality reduction. However, it tends to find more PCs than necessary and therewith over–interprets the data structures. Kaiser criterion is a good choice for the number of factors that should be included but fails if explained variance is spread, which argues for underlying (non–linear) structures. In this case, PCA calculated 73% of explained variance containing 10 PCs. This is not much and highlights that indoor air is extremely variable. The interpretation of all 10 factors was complicated and is not suitable to extract more PCs for gaining higher explained variances. To increase the cumulative variance the involvement of more PCs is necessary, but this counteracts with an improvement of variable number and makes source identification more difficult.

The factors of PCA are always orthogonal, meaning that the occurrence of two factors, and so two sources, is not very probable at all. For example, source 1 of PCA is supposed to be the influence of ventilation on indoor air quality but is not directly related to indoor activities. We suggest that ventilation is intensified through VOC–emitting activities, but the factor does not account for mixtures. In reality, there are numerous possibilities for the combinations of indoor activities, accommodation characters and outdoor influences that have an impact on indoor air. Having this in mind, PCA can be used to get an impression of the single factors influencing air quality because VOCs with similar concentration characteristics are pooled.

The combination of different factors, for example ventilation, renovation and new furniture, is only combined in one principal component when the overall variability of the whole VOC dataset is strongly influenced by the most important compounds which must not differ in their variances. So, PCA is more feasible when profiles of specific sources are of concern.

NMF results are a combination of PCA–clustered VOCs (Figure 2). Hence, it is possible that the method of NMF shows more or less characteristic situations of VOCs in indoor environments; constantly recurring structures are not single PCs but rather a combination of them. This study gives evidence for simultaneously occurring sources in indoor environments, describing the various patterns of indoor compounds. Different numbers for *r* were chosen but smaller quantities showed patterns difficult to explain, and larger *r* split the patterns.

One would guess that seasonality and smoking indoors create own PCs and patterns, but this was not observable in the analyses. Although VOC measurements were conducted throughout the whole year, only multiple regression analysis showed a relationship. Seasonal variations, which are supposed to account for a high variability in the VOC dataset, were not detected as a special pattern of seasonality in PCA or NMF. PC1 describes the influence of ventilation on indoor VOC concentrations and, indirectly, ventilation is driven by season. Ventilation is more pronounced in summer compared to winter months, and total VOC concentrations do not differ much in smoker and nonsmoker households. In winter months, VOCs accumulate indoors due to inappropriate ventilation behaviors in nonsmoker houses, raising the total VOC level as it is observable in factor 1 of PCA (Schlink et al., 2010). In contrast, smoker housings tend to ventilate their rooms independently of season. Hence, concentrations of smoking–related VOCs are significantly elevated, but unrelated VOC concentrations are free of influences from indoor smoking and there was no significant difference between both housing types. This may be one reason why smoking and seasonality did not result in any single factor or PC.

Aromatics, especially the toluene/benzene (T/B) ratio are a good marker for the influence of outdoor air and traffic. Gelencser et al. (1997) showed that the strength of the ratio depends on the proximity to roads. This ratio is inappropriate in our analysis because benzene and toluene are emitted by indoor sources (e.g. newspaper) as well and references for T/B ratios in indoor air are still missing. The ratio between toluene and benzene is 0.18 (mean) and there are seven cases with a ratio >1 and 54 cases with ratio >0.5. Therefore, there is no clear evidence that traffic affected the indoor concentrations as a result of ventilation.

The main aim of the application of NMF to this data set was to identify characteristic situations of VOC occurrences via combinations of origins, as specific and independent sources without forcing requirements of orthogonality. PCA and NMF are both used for data acquisition and processing. The comparison of PCA and NMF showed that the characterizations of housings should be done with NMF as combinations of VOCs out of different sources from PCA so that results reflect distinctive VOC "pictures" from different homes. PCA is a good choice for revealing sources of VOC groups; NMF joins the pieces of a puzzle (sources) to form a big picture (pattern).

4. Conclusions

VOC concentrations were measured over a period of 16 months in 2 246 homes in Leipzig, Germany. To determine the influence of sources and patterns of VOCs on indoor air, different analyzing methods were applied to reduce the dataset to its main structures (PCA) and to find recurring characteristics (NMF). Factors were found that described the following sources: influence of ventilation (PC1), wooden furniture, cleaning agents and plants (PC2), gluing emissions (PC3), renovation activities (PC4 and PC6), solar radiation (PC5), aging of materials (PC7), background concentrations (PC8), indoor climate (PC9) and PVC flooring (PC10). PCA is appropriate when the highest possible explained variance is of interest. The aim of this study was, ideally, to find characteristic sources of VOCs in urban housings in areas without strong outdoor influences. The method was able to extract 10 factors, explaining 73% of the total variance.

Sources and patterns differ in their occurrence. We defined sources to be a point of origin for related compounds in contrast to patterns, which contain various, often unrelated compounds. PCA is a statistical analyzing method for source–oriented analyses whereas the NMF method is used for pattern–oriented purposes. The most affecting difference between both methods is the combination of sources to form patterns without any reverse option. So, the catenation of typical sources for indoor environments was established by NMF, which gave information about potential combinations of single sources. The recovery rate of PCA sources in NMF patterns averaged at 79.5%. So, indoor air of participant housings is an interaction of renovation, flooring, cleaning activities, and natural aging processes. Patterns of homes showed various combinations of anthropogenic and natural sources. These results affirm the latest studies by Guo (2011) and Lau et al. (2010).

This study showed that most complex data relations can be reproduced with NMF while data that do not contain several local behaviors and has a limited dispersion can be represented using PCA.

We suggest that adverse impacts of a combination of sources should be researched in further studies as well as a verification of existing results to the relationship between single/group VOC sources and health effects.

Acknowledgments

The authors acknowledge the sponsoring by the scholarship program of the German Environmental Foundation (Deutsche Bundesstiftung Umwelt – DBU).

Supporting Material Available

Additional information on: Data analysis (S1), Results of regression analysis (S2), Results of Principal Component Analysis (S3), Results of NMF (S4), Contributions of VOC concentrations for measurements of the whole year and divided by season (Figure S1), Results of regression analysis of all indoor VOC for different conditions (Table S1), Factor loadings from PCA of 49 VOCs (Table S2), Results of NMF analysis for patterns of homes in Leipzig (Table S3). This information is available free of charge via the internet at http://www.atmospolres.com.

References

Arif, A.A., Shah, S.M., 2007. Association between personal exposure to volatile organic compounds and asthma among US adult population. *International Archives of Occupational and Environmental Health* 80, 711–719.

Ayoko, G.A., 2004. Volatile organic compounds in indoor environments, in *The Handbook of Environmental Chemistry: Indoor Air Pollution*, Volume 4 – Air Pollution Part F, edited by Pluschke, P., Springer–Verlag Berlin Heidelberg and New York, pp. 1–35.

Badol, C., Locoge, N., Galloo, J.C., 2008. Using a source–receptor approach to characterise VOC behaviour in a French urban area influenced by industrial emissions. Part II: Source contribution assessment using the Chemical Mass Balance (CMB) model. *Science of the Total Environment* 389, 429–440.

Batterman, S., Jia, C.R., Hatzivasilis, G., 2007. Migration of volatile organic compounds from attached garages to residences: A major exposure source. *Environmental Research* 104, 224–240.

Begerow, J., Jermann, E., Keles, T., Ranft, U., Dunemann, L., 1995. Passive sampling for volatile organic compounds (VOCs) in air at environmentally relevant concentration levels. *Fresenius Journal of Analytical Chemistry* 351, 549–554.

Billionnet, C., Gay, E., Kirchner, S., Leynaert, B., Annesi–Maesano, I., 2011. Quantitative assessments of indoor air pollution and respiratory health in a population–based sample of French dwellings. *Environmental Research* 111, 425–434.

Boutsidis, C., Gallopoulos, E., 2008. SVD based initialization: A head start for nonnegative matrix factorization. *Pattern Recognition* 41, 1350–1362.

Buonanno, G., Morawska, L., Stabile, L., Wang, L., Giovinco, G., 2012. A comparison of submicrometer particle dose between Australian and Italian people. *Environmental Pollution* 169, 183–189.

Buonanno, G., Morawska, L., Stabile, L., 2009. Particle emission factors during cooking activities. *Atmospheric Environment* 43, 3235–3242.

Cai, C.J., Geng, F.H., Tie, X.X., Yu, Q.O., An, J.L., 2010. Characteristics and source apportionment of VOCs measured in Shanghai, China. *Atmospheric Environment* 44, 5005–5014.

Carslaw, N., Langer, S., Wolkoff, P., 2009. New directions: Where is the link between reactive indoor air chemistry and health effects? *Atmospheric Environment* 43, 3808–3809.

Chan, Y.C., Hawas, O., Hawker, D., Vowles, P., Cohen, D.D., Stelcer, E., Simpson, R., Golding, G., Christensen, E., 2011. Using multiple type composition data and wind data in PMF analysis to apportion and locate sources of air pollutants. *Atmospheric Environment* 45, 439–449.

Demir, S., Saral, A., Erturk, F., Kuzu, S.L., Goncaloglu, B.I., Demir, G., 2012. Effect of diurnal changes in VOC source strengths on performances of receptor models. *Environmental Science and Pollution Research* 19, 1503–1514.

Diez, U., Kroessner, T., Rehwagen, M., Richter, M., Wetzig, H., Schulz, R., Borte, M., Metzner, G., Krumbiegel, P., Herbarth, O., 2000. Effects of indoor painting and smoking on airway symptoms in atopy risk children in the first year of life results of the LARS–study. *International Journal of Hygiene and Environmental Health* 203, 23–28.

Dodson, R.E., Levy, J.I., Spengler, J.D., Shine, J.P., Bennett, D.H., 2008. Influence of basements, garages, and common hallways on indoor residential volatile organic compound concentrations. *Atmospheric Environment* 42, 1569–1581.

Edwards, R.D., Schweizer, C., Llacqu, V., Lai, H.K., Jantunen, M., Bayer–Oglesby, L., Kunzli, N., 2006. Time–activity relationships to VOC personal exposure factors. *Atmospheric Environment* 40, 5685–5700.

Eklund, B.M., Burkes, S., Morris, P., Mosconi, L., 2008. Spatial and temporal variability in VOC levels within a commercial retail building. *Indoor Air* 18, 365–374.

Fuentes–Leonarte, V., Tenias, J.M., Ballester, F., 2009. Levels of pollutants in indoor air and respiratory health in preschool children: A systematic review. *Pediatric Pulmonology* 44, 231–243.

Gelencser, A., Siszler, K., Hlavay, J., 1997. Toluene–benzene concentration ratio as a tool for characterizing the distance from vehicular emission sources. *Environmental Science & Technology* 31, 2869–2872.

Geng, F.H., Cai, C.J., Tie, X.X., Yu, Q., An, J.L., Peng, L., Zhou, G.Q., Xu, J.M., 2009. Analysis of VOC emissions using PCA/APCS receptor model at city of Shanghai, China. *Journal of Atmospheric Chemistry* 62, 229–247.

Gokhale, S., Kohajda, T., Schlink, U., 2008. Source apportionment of human personal exposure to volatile organic compounds in homes, offices and outdoors by chemical mass balance and genetic algorithm receptor models. *Science of the Total Environment* 407, 122–138.

Guo, H., 2011. Source apportionment of volatile organic compounds in Hong Kong homes. *Building and Environment* 46, 2280–2286.

Han, K., Zhang, J.S., Wargocki, P., Knudsen, H.N., Varshney, P.K., Guo, B., 2012. Model-based approach to account for the variation of primary VOC emissions over time in the identification of indoor VOC sources. *Building and Environment* 57, 403-416.

Han, K.H., Zhang, J.S., Knudsen, H.N., Wargocki, P., Chen, H., Varshney, P.K., Guo, B., 2011. Development of a novel methodology for indoor emission source identification. *Atmospheric Environment* 45, 3034-3045.

Herbarth, O., Matysik, S., 2010. Decreasing concentrations of volatile organic compounds (VOC) emitted following home renovations. *Indoor Air* 20, 141–146.

Herbarth, O., Fritz, G.J., Rehwagen, M., Richter, M., Roder, S., Schlink, U., 2006. Association between indoor renovation activities and eczema in early childhood. *International Journal of Hygiene and Environmental Health* 209, 241–247.

Jia, C., Batterman, S., Godwin, C., 2008a. VOCs in industrial, urban and suburban neighborhoods – Part 2: Factors affecting indoor and outdoor concentrations. *Atmospheric Environment* 42, 2101–2116.

Jia, C., Batterman, S., Godwin, C., 2008b. VOCs in industrial, urban and suburban neighborhoods, Part 1: Indoor and outdoor concentrations, variation, and risk drivers. *Atmospheric Environment* 42, 2083–2100.

Jo, W.K., Kim, J.D., 2010. Personal exposure of graduate students attending the college of natural sciences or social sciences to volatile organic compounds on campus. *Chemosphere* 81, 1272–1279.

Lau, A.K.H., Yuan, Z.B., Yu, J.Z., Louie, P.K.K., 2010. Source apportionment of ambient volatile organic compounds in Hong Kong. *Science of the Total Environment* 408, 4138–4149.

Matysik, S., Ramadan, A.B., Schlink, U., 2010. Spatial and temporal variation of outdoor and indoor exposure of volatile organic compounds in Greater Cairo. *Atmospheric Pollution Research* 1, 94–101.

Missia, D.A., Demetriou, E., Michael, N., Tolis, E.I., Bartzis, J.G., 2010. Indoor exposure from building materials: A field study. *Atmospheric Environment* 44, 4388–4395.

Morawska, L., He, C.R., Johnson, G., Guo, H., Uhde, E., Ayoko, G., 2009. Ultrafine particles in indoor air of a school: Possible role of secondary organic aerosols. *Environmental Science & Technology* 43, 9103–9109.

Morawska, L., He, C.R., Hitchins, J., Mengersen, K., Gilbert, D., 2003. Characteristics of particle number and mass concentrations in residential houses in Brisbane, Australia. *Atmospheric Environment* 37, 4195–4203.

Ohura, T., Amagai, T., Shen, X.Y., Li, S.A., Zhang, P., Zhu, L.Z., 2009. Comparative study on indoor air quality in Japan and China: Characteristics of residential indoor and outdoor VOCs. *Atmospheric Environment* 43, 6352–6359.

Oxford Dictionaries, 2013. http://www.oxforddictionaries.com/, accessed in July 2013.

Pekey, H., Pekey, B., Arslanbas, D., Bozkurt, Z.B., Dogan, G., Tuncel, G., 2013. Source apportionment of personal exposure to fine particulate matter and volatile organic compounds using positive matrix factorization. *Water Air and Soil Pollution* 224, art. no. 1403.

Pindado, O., Perez, R.M., 2011. Source apportionment of particulate organic compounds in a rural area of Spain by positive matrix factorization. *Atmospheric Pollution Research* 2, 492–505.

Qian, H., Li, Y.G., Seto, W.H., Ching, P., Ching, W.H., Sun, H.Q., 2010. Natural ventilation for reducing airborne infection in hospitals. *Building and Environment* 45, 559–565.

Rumchev, K., Spickett, J., Bulsara, M., Phillips, M., Stick, S., 2004. Association of domestic exposure to volatile organic compounds with asthma in young children. *Thorax* 59, 746–751.

Salthammer, T., Bahadir, M., 2009. Occurrence, dynamics and reactions of organic pollutants in the indoor environment. *Clean–Soil Air Water* 37, 417–435.

Schlink, U., Thiem, A., Kohajda, T., Richter, M., Strebel, K., 2010. Quantile regression of indoor air concentrations of volatile organic compounds (VOC). *Science of the Total Environment* 408, 3840–3851.

Schlink, U., Rehwagen, M., Damm, M., Richter, M., Borte, M., Herbarth, O., 2004. Seasonal cycle of indoor–VOCs: Comparison of apartments and cities. *Atmospheric Environment* 38, 1181–1190.

Shin, S.H., Jo, W.K., 2013. Longitudinal variations in indoor VOC concentrations after moving into new apartments and indoor source characterization. *Environmental Science and Pollution Research* 20, 3696–3707.

van der Maaten, L., 2009. Learning a parametric embedding by preserving local structure. *Proccedings of 12^th^ International Conference on Artificial Intelligence and Statistics (AISTATS),* April 16–19, 2009, Florida, USA, pp. 384–391.

Wolkoff, P., 1995. Volatile organic compounds – sources, measurements, emissions, and the impact on indoor air quality. *Indoor Air* 5, 5–73.

Yrieix, C., Dulaurent, A., Laffargue, C., Maupetit, F., Pacary, T., Uhde, E., 2010. Characterization of VOC and formaldehyde emissions from a wood based panel: Results from an inter–laboratory comparison. *Chemosphere* 79, 414–419.