PARALLEL FACTOR ANALYSIS (PARAFAC) POUR L'INTERPRETATION DE LA VARIATION DES POLLUANTS DE L'AIR INTERIEUR DANS UN ENVIRONNEMENT DE BUREAUX

T.H Nguyen*[1], A. Ionescu[1], O. Ramalho[2], M. Mathis[3], E. Gehin[1]

[1]CERTES, Université de Paris Est Créteil, 94000, France
[2]Centre scientifique et technique du bâtiment, 77420, France
[3]Université de Paris Est Créteil, 94000, France
*Courriel de l'orateur : thi-hao.nguyen@univ-paris-est.fr

TITLE

Parallel factor analysis (PARAFAC) for interpretation of variation of indoor air pollutants in an open-plan office

RESUME

Cette recherche se concentre sur l'étude de la qualité de l'air intérieur (QAI) en développant une approche optimale du traitement des données. En observant les composantes de sortie de la décomposition de PARAFAC, nous sommes en mesure d'expliquer les fluctuations des processus sous-jacents et leurs causes dans un environnement intérieur (un bureau paysager) et ainsi mieux anticiper les conséquences. Par ailleurs, PARAFAC est une méthode très prometteuse pour la QAI lorsqu'il est possible de traiter des tableaux de grandes dimensions et de fournir une sortie unique. Sur cette base, nous pouvons appliquer différentes situations de QAI et utiliser les résultats de chargement pour le modèle de régression ultérieurement.

ABSTRACT

This research focuses on indoor air quality (IAQ) study by developing an optimal approach of data processing. By observing the output components of PARAFAC decomposition, we are able to explain the fluctuations of the underlying processes and their causes in an indoor environment (an open-plan office) and thus, better anticipate their consequences. Besides, PARAFAC is a very promising method for IAQ when it is possible to deal with high dimensional array and gives the unique output. Based on that, we can apply for different situations of IAQ for different purposes and also use the loading outputs for the regression model afterward.

MOTS-CLES : qualité de l'air intérieur, décomposition tensorielle, particule / **KEYWORDS:** indoor air quality, tensor decomposition, particle matter

1. INTRODUCTION

Human being spend almost more than 90 percent of their time in the indoor environment, but they are not yet or never aware of how damaging it is the pollutant of indoor air quality (IAQ). A number of researches about this topic have been increasing overtime. However, the lack of data and its complexity make it become more difficult to deal with.

In this report, our work concentrates on analyzing a tensor data of an open-plan office to explain their variation and causes, by applying PARAFAC – a recent method of decomposition.

1.1. Study case

The Scientific and Technical Center for Building (CSTB - France) is one of the most famous center working with IAQ in France and it has been archived many successes in this topic.

The studied open-plan office is located in the building of CSTB, Champs-sur-Marne. This is an urban area, and it is surrounded by many departmental and highway roads. The office receives a lot of sunlight and wind every day, as it has many slide windows (6), and is not covered by the other buildings nearby.

The surface and volume of the office are 132 m^2 and 364 m^3 respectively, and it is reserved for about 15~20 people, from 8:30 to 18:00 from Monday to Friday. In this office and outdoors, many measurement devices are equipped for monitoring, including:

- Concentration of indoor and outdoor particles [0.3 - 20 *μm*] (optical measurement Grimm Dust Monitor 1.108);
- Concentration of indoor formaldehyde (using AL1021 Aerolaser);
- Concentration of indoor NO_x, O_3, CO (by Microstation Environnement SA);
- Opening factor of door and windows, occupation (by sensors); and

- Climate information (temperature, relative humidity, irradiance) and outdoor (temperature, relative humidity, irradiance, rainfall, pressure, wind).

The monitoring is performed every minute, therefore we have a very high frequency database (Ramalho, 2016).

1.2. Literature review

Principal Component Analysis (PCA), Positive Matrix Factorization (PMF) or Non-Negativity Matrix Factorization (NNMF), Independent Component Analysis (ICA) are well-known methods for decomposition of matrix data (2D array). They can create the model which fits well with the data but they have to deal with the problem of uniqueness of the solution and complexity of data if the dimension is increased to more than two. It is necessary to have a method that can deal with data array of a greater complexity than that. Therefore, tensor decomposition is invented for this purpose. By definition, tensors are generalizations of matrices to higher dimensions and can consequently be treated as multidimensional fields (Hitchcock, 1927).

PARAFAC (Bro, 1997) is one of the tensor decomposition methods and it can be considered as PCA or PMF for higher order array. Even if its model fitting degree is not as good as other matrix decompositions, it gives us a unique output and is very easy to increase the complexity of the data dealt with. For these reasons also we want to introduce other kind of information: inside/outside, climate, etc, we decide to use PARAFAC in this work.

The simple PARAFAC model for a 3-dimension array is given by three loading matrices **A**, **B** and **C.** With the elements a_{if}, b_{jf}, and c_{kf} , the model tries to minimize the sum of squares of the residuals e_{ijk}. This model can be descried by the equation below, and its example for the data of IAQ is represented in Figure 1.

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk}$$

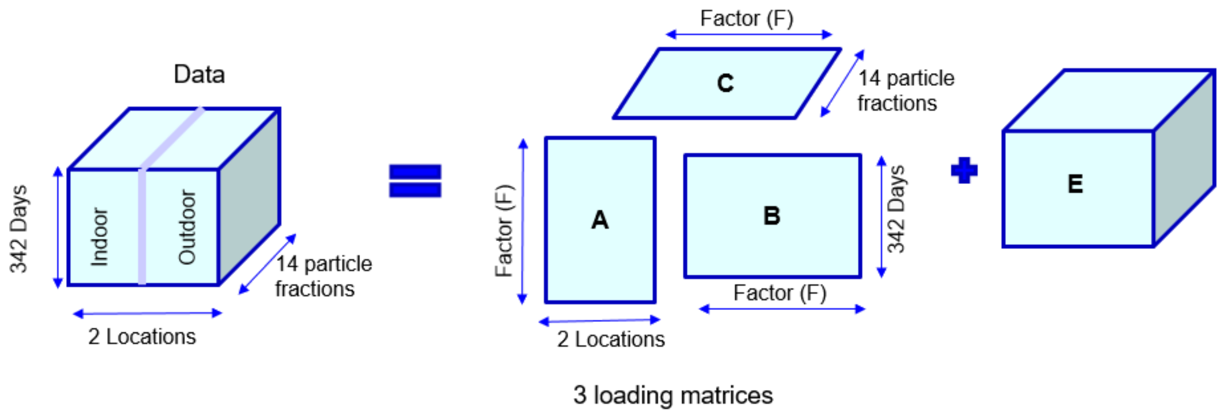

Figure 1: An example about PARAFAC model for three dimensional data of IAQ

2. METHOD

We analyzes the variation of them according to different times and locations of measurement (Indoor/Outdoor).

2.1. Data presentation

Our data in this work consists of the particle concentration for different fractions and information about specific humidity which is measured during one year (2014). The former is presented as a number of particle per liter of air (so called PN - Particle Number). Originally, there are 14 fractions in total, named as: PN0.35, PN0.45, PN0.575, PN0.725, PN0.9, PN1.3, PN1.8, PN2.5, PN3.5, PN4.5, PN6.25, PN8.75, PN12.5, PN17.5, according to their sizes. However, in order to avoid numerical instability due to the correlation high among the different fraction, we decide to categorize them into 3 groups based on their sizes and correlations between each other as: Fine (PN0.35 to PN0.725), Medium (PN0.9 to PN1.8) and Coarse (PN3.5 to PN17.5).

Regarding the specific humidity, based on the relative humidity and temperature measured by Q-Track Probe (TSI Inc), it is calculated using the Rankine formula below:

$$H_{abs}\left(\frac{g}{kg} humide\ air\right) = \frac{HR}{100} \times \frac{M_{water}}{M_{air}} \times e^{\left(13.7 - \frac{5120}{T+273.5}\right)} \times 1000$$

$$H_{spec}\left(\frac{g}{kg} dry\ air\right) = \frac{H_{abs}}{1000 - H_{abs}} \times 1000$$

2.2. Implementation

The 3-dimensional array of $342_{days} \times 4_{3\ PN\ groups\ and\ Hspec} \times 2_{locations}$ is obtained as input data of PARAFAC after we perform the following steps below:

- **Pre-processing data**: Every sample with missing data is removed (23 days). Then, the maximum normalization is applied for each type of variable. This is to make sure that all the data is in the range of 0 – 1 but still keeps its variation.
- **Constructing input data:** Depending on the aim of this experiment, we calculate the average value of each variable during one day under different locations (Indoor/Outdoor).
- **Determining number of components:** Core consistency diagnostic method is used in this work (Bro, 2003). The most suitable number of components is chosen as the highest number that has the valid value of core consistency (80% - 100%)**.**
- **Applying constrains for the algorithm**: Given that the concentration of pollutants should be positive, we apply the Non-negativity constrains for the loading output of PARAFAC.

3. RESULTS AND DISCUSSIONS

In our case, the most suitable number of components is determined as 3 with the value of core consistency being up to 99.63%. The outputted loadings of PARAFAC are displayed in Figure 2.

We see that the 1st component depends mainly on the concentration of Hspec (loading > 0.99) and is not affected by the location (same loadings for indoor and outdoor). In addition, using the loading value of this component according to time series of 342 days, we detect that it has the periodicity of the weekly data. Therefore, the algorithm accurately decomposes the **background component** of particle concentration.

Furthermore, the 2nd component with high loading of fine and medium particles (loading > 0.68) is affected by outdoor environment. In contrast, the 3rd component (coarse particle, loading > 0.98) is mainly affected by indoor environment. These results fit for real-life conditions, where the concentration of fine and medium particles is normally manipulated by the outdoor environment when people let doors and windows open. Meanwhile, the coarse sized particles are strongly impacted by human presence (indoor environment).

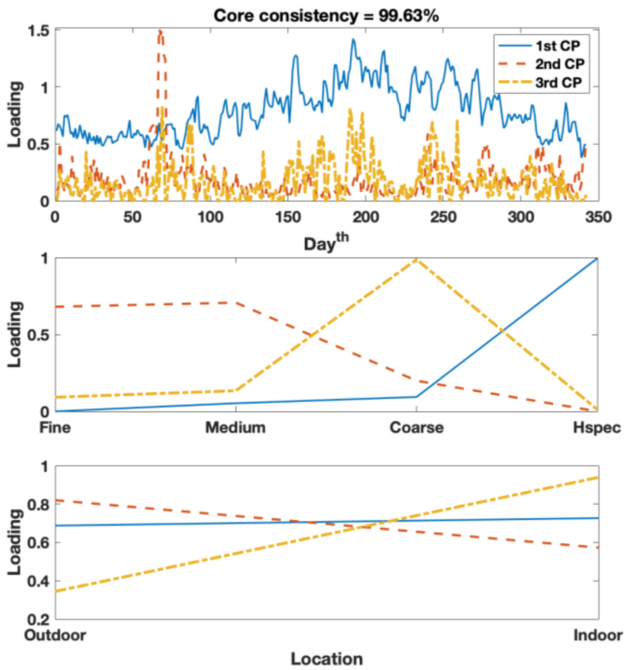

Figure 2. 3-Dimensional PN data at 2 locations (indoor/outdoor)

DOI : 10.25576/ASFERA-CFA2020-19781

4. CONCLUSION

In this report, we succeed to decompose the 3-dimentional data consisting of particles in indoor and outdoor air by using PARAFAC as a tensor decomposition algorithm. From the matrices of loadings, we are able to analyze the variation of data, identify the sources of pollutants and assess their relative contributions. In addition, these outputs can be used as inputs of a regression model. From this, an optimal model, which can be applied for real-time monitoring of IAQ and gives the prediction several hours after, could be created in next step.

O. Ramalho et al, 2016, TRIBU rapport, PRIMEQUAL APR EIAI / projet TRIBU.

F. L. Hitchcock, 1927. The expression of a tensor or a polyadic as a sum of products. J. Math. Phys 6, 1 (1927), 164–189.

R. Bro, 1997, Parafac tutorial and applications, Chemometrics and Intelligent Laboratory Systems.

R. Bro and H. A. L. Kiers, 2003, A new efficient method for determining the number of components in PARAFAC models, Journal of chemometrics, pp. 274–286

DOI : 10.25576/ASFERA-CFA2020-19781