# J4.3 APPLICATION OF POSITIVE MATRIX FACTORIAZATION FOR ATMOSPFERIC AEROSOL SOURCES IDENTIFICATION IN SÃO PAULO CITY

Beatriz S. Oyama*, Maria de Fátima Andrade
University of Sao Paulo, Sao Paulo, Brazil

## 1. INTRODUCTION

This work is part of a comprehensive project conducted by Atmospheric Sciences Department and Medical School, both from the University of Sao Paulo. The project aims to evaluate the health and economic impact of fine particulate matter ($PM_{2.5}$) in six Brazilian urban cities (Sao Paulo, Rio de Janeiro, Belo Horizonte, Recife, Curitiba and Porto Alegre). The focus of the present study was Sao Paulo Mega-city fine particles source identification.

Sao Paulo is situated in the Metropolitan Region of Sao Paulo (MRSP). House of 11 million inhabitants, it holds a lot of polluter industries and a 7 million car's fleet. These features are responsible for strong air quality degradation and complex mixture of aerosols and gases in the atmosphere (Cetesb – Environmental Protection Agency of Sao Paulo, 2008).


Figure 1: Localization of Metropolitan Region of São Paulo in southeast Brazil.

The MRSP is one of the largest urban motor vehicles fleet of the word. This fleet is divided into 61% based gasohol, that is a mix of 78% of gasoline and 22% of ethanol, 11.4% ethanol, 5.5% diesel, 10% flex fuel and 12% are motorcycles (running with gasohol).

* *Corresponding author address:* Beatriz Oyama, University of Sao Paulo, Sao Paulo, Brazil. e-mail: beatrizoyama@gmail.com

Recently, new technology was introduced that allows cars to be fueled with any mixture of gasoline and ethanol (called Flex Fuel technology), with means that the engine adapts itself to any ratio of fuel being used at each particular moment. Brazil was pioneer in the implementation and use of an alternative fuel as ethanol in its fleet. The program of sugar-cane ethanol started in the 1970's (PROALCOOL: http://projects.wri.org/sd-pams-database/brazil/national-alcohol-program-proalcool).

### *1.1 AEROSOL*

The aerosol consists in solid or liquid particles suspended in the atmosphere. The chemical composition (origin and source) and size distribution (nanometers to micrometers) may vary according to emission and secondary process in the atmosphere.

The aerosol can be split in two different size modes: course particulates (diameter between 2.5 and 10 μm), and fine ones (diameter below 2.5 μm). The last one, although harmful to health, is not regulated in Brazil by the national governmental environmental agency (Seinfield and Pandis, 1998, CESTESB, 2008). The aerosols can be directly emitted, like particulates (primary aerosol), or formed in the atmosphere, by gas-particulate process (secondary aerosol). The aerosols can vary their diameter from some nanometers (nm) to tents of micrometers (mm) (Raes et. al., 2000, Seinfield e Pandis, 1998).

The characterization of particles in the atmosphere of São Paulo has started in late 70s with research at the Physics Institute in the University of Sao Paulo In 80´s started the application of receptor models to identify the sources of the particles. Principal Component analysis and Absolute Principal Component Analysis were applied with the identification of mobile emission as the principal source for fine particles and soil ressuspension for the coarse mode for São Paulo Metropolitan Area (Andrade et al., 1994). Castanho & Artaxo , 2001, showed that a significant fraction of fine particulate matter is constituted by organic carbon (OC) representing approximately 40% of the concentration.

The BC mean concentration represented 21% of PM2.5, in total 60% of the fine particles mass concentration is explained by carbon compounds, mainly due to vehicular emissions, being the heavy duty diesel the main source of fine particles (Sanchez-Ccoyllo et al., 2009). The crustal elements were estimated by the sum of the oxides representing 12 of $PM_{2.5}$.

The fine and course particulates differ not only by the size distribution but by the chemical composition and formation process too. The fine particulates are essentially formed by combustion, high-temperature process and atmospheric reactions.

Nucleation, condensation, coagulation are the mainly process for the $PM_{2.5}$ formation. In their composition it is very common to find sulfate ($SO_4^{2-}$), nitrate ($NO_3^-$), ammonium ($NH_4^+$), Black Carbon (BC) and metals (such as Pb, V, Ni, Cu, Zn, Mn, Fe).

## 2. OBJECTIVE

The goal of this work is the identification of the $PM_{2.5}$ sources for Sao Paulo city considering the elemental concentration data using the Positive Matrix Factorization technique as the methodology for receptor modeling.

## 3. METHODOLOGY

The samples were collected at the Medical School building; close of a heavy traffic avenue in downtown city (23°32'51'' S, 46°38'10'' W) using a particulate sampler, during 24 hours, between August 2007 and August 2008, totalizing one year dataset. The total of 160 samples were analyzed for the determination of mass and BC concentration by X-Ray Fluorescence analysis for the quantification of trace-elements concentration

### *3.1 EXPERIMENTAL METHODOLOGY*

The samples were submitted to analysis for identification of: mass concentration ($PM_{2.5}$), Black Carbon, elementary constitution, and soluble species. The method for identification of each concentration is described below.

The first proceed was the gravimetric analysis (weighing samples), before and after the measurements, in order to determine the $PM_{2.5}$ concentration. This procedure consists on weighing the polycarbonate filters; using a balance which the nominal precision is 1 μg, after their static electricity has been removed.

Before the quantification of $PM_{2.5}$, the samples were submitted to analysis for identification of Black Carbon. This technique utilizes the light reflectance induced by the particulate. The result is the light absorption by the particulate deposited on the sample. The fraction of absorbing radiation is classified as BC, which is essentially composed by Elemental Carbon.

The third analysis was the elemental concentration by the X-Ray Fluorescence Analysis. This technique is used to evaluate quality quantitative chemical composition of various samples types (Nascimento Filho, 1999). Were identified in this analysis the follow elements: Al, Si, P, S, Cl, P, K, Ca, Ti, V, Cr, Mn, Ni, Cu, Zn, Br, and Pb.

### *3.2 POSITIVE MATRIX FACTORIZATION*

A bilinear model was used in order to expresses the observations of PM species as the sum of contributions from a number of time-invariant source profiles.

The main objective of multivariate receptor models is to identify the number of factors, the profile of the species in each source, the quantity of mass distributed in each factor for each sample. To accomplish this objective the Paatero and Tapper, 1994 model was used.

$$X = G\,F + E \qquad (1)$$

Using index notation:

$$x_{ij} = \sum_{k=1}^{p} g_{ik}\, f_{kj} + e_{ij} \qquad (2)$$

Where:
$x_{ij}$ = Species concentration $j$ in the sample $i$;
$p$ = Factors number that contribute to the samples;
$f_{kj}$ = Concentration of species $j$ in the profile factor $k$;
$g_{ik}$ =relative contribution of the factor $k$ in the sample $i$;
$e_{ij}$ = PMF model error for species $j$ measured in the sample $i$ (Reef et al, 1999).

The PMF model allows that each point be individually weighted. This characteristic permits that the user adjusts the significance of each point, depending of the confidence in the measure (Norris et al, 2008).

The values $g_{ik}$ and $f_{kj}$ are adjusted until a minimum Q-value is founded or a predicted iteration is exceeded (not-convergent case), with the condition that all the matrix elements be positive. Q-value is called object-funtion, defined following (Paatero, 1997, Hopke, 1998, Reff et al, 2007, Norris et al, 2008). The non-negativity restriction is accomplished minimizing the Q-value based on G and F with reservation of:

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \left[\frac{e_{ij}}{\sigma_{ij}}\right]^2 \qquad (3)$$

Where:

$\sigma_{ij}$ = uncertainty of concentration $j$ in the specie $i$.
$e_{ij}$ = residual value; equal to $x_{ij} - \sum_{k=1}^{p} g_{ik}\, f_{kj}$

### *3.2.1. ROBUST MODE*

Higher values of environmental data can affect disproportionally the results. In order to avoid this pattern, it can be used a filter function ($h_{ij}$) in the Q calculation by least squares method.

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{e_{ij}}{h_{ij} s_{ij}} \qquad (4)$$

and:

$$h_{ij}{}^{2} = \begin{cases} 1 & if \ |e_{ij}/s_{ij}| \leq \alpha \\ \dfrac{|e_{ij}/s_{ij}|}{\alpha} & \end{cases} \qquad (5)$$

Where α is a parameter called: outlier threshold distance. A common value is: α = 4 (Paatero, 1997 Lee et al, 1999, Hopke, 1998).

### *3.2.2 DETERMINING THE NUMBER OF FACTORS*

One of the most important details of PMF model is the determination of the number of factors utilized in the models adjust. There is no simple way to do it, i.e. it is difficult to determine how many factors should be extracted from a dataset. The better way to obtain this value is by analyzing the Q and P values. The Q values (also called $Q_{true}$ – that considers all the dataset including outliers – or $Q_{robust}$ – that excludes the outliers) need to converge to $Q_{theoretical}$, i.e. the $Q_{true}$ or $Q_{robust}$ need to be close to $Q_{theoretical}$. Therefore, if the best solution does not presents physical validity, other values of P should be tested until a reasonable result should be founded (Reef et al, 2007, Norris et al, 2008, Hedberg et al, 2005).

## 4. RESULTS

The following results refer to the concentrations found on experimental methodology. From these results, it was possible to use the PMF to obtain the sources contributing to the formation of fine particulate matter. These results are showed after the concentration results.

### 4.1 CONCENTRATION RESULTS

The figure 3 shows the dataset obtained by gravimetric and reflectance analysis. The BC concentration follows the $PM_{2.5}$ variability. The higher concentrations (75 µg $m^{-3}$ $PM_{2.5}$ and 25 µg $m^{-3}$ BC) occur during the winter. This season, in Sao Paulo city, is characterized by low precipitation and stable atmosphere days, these patterns difficult the $PM_{2.5}$ removal and dispersion. The lower concentrations occur (in both cases) during the summer, a rainy and airy period.

From the data obtained by X-Ray Fluorescence Analysis, the figure 3 presents the elementary concentrations (note the logarithmic scale). The lower concentrations were identified in V, Cr, and Ni, below than 1.6 ng $m^{-3}$ and the higher concentrations were observed in S, K, and Fe (important urban center markers). The higher Al, Si, and Fe concentrations indicate soil dust re-suspension. The sulfur concentrations are associated to gas-particle conversion (mainly $SO_2$ gas), so the presence of sulfur in the sample indicates that is secondary particle.

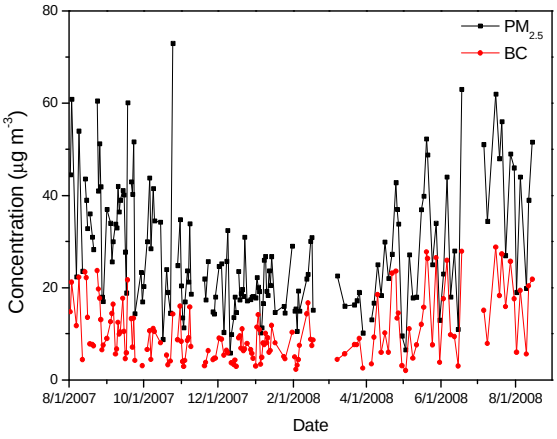


Figure 3: $PM_{2.5}$ and BC concentration variability.

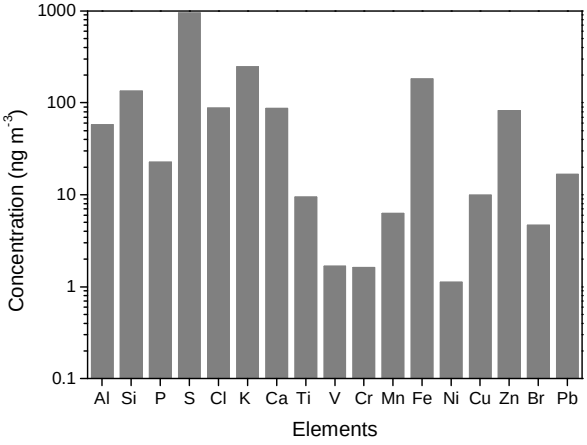


Figure 4: Mean trace-elements concentrations.

### 4.2 SOURCES IDENTIFICATION RESULTS

#### *4.2.1 Dataset preparation*

The source identification was accomplished using the data treated in the methodology. It was necessary to prepare the dataset before use the PMF technique. The missing data were substituted by the median and their respective uncertainties. These uncertainties were equaled to 3 times the median

value; therefore these values assume a less important role in the analysis process.

The objective of this procedure is to reduce the contribution of these values with high uncertainties to the concentrations that were measured by the instrument. This procedure includes a selection of species, because it is important to avoid a double concentration, e.g. the sulphur (elemental) and sulfate (ion). It should be a repetition of information that could change the final result of source division. Considering that PMF uses the least square method, it is necessary to exclude the outliers. It was accomplished by using a visual technique inspection of the temporal series. The concentrations that present outliers values were excluded.

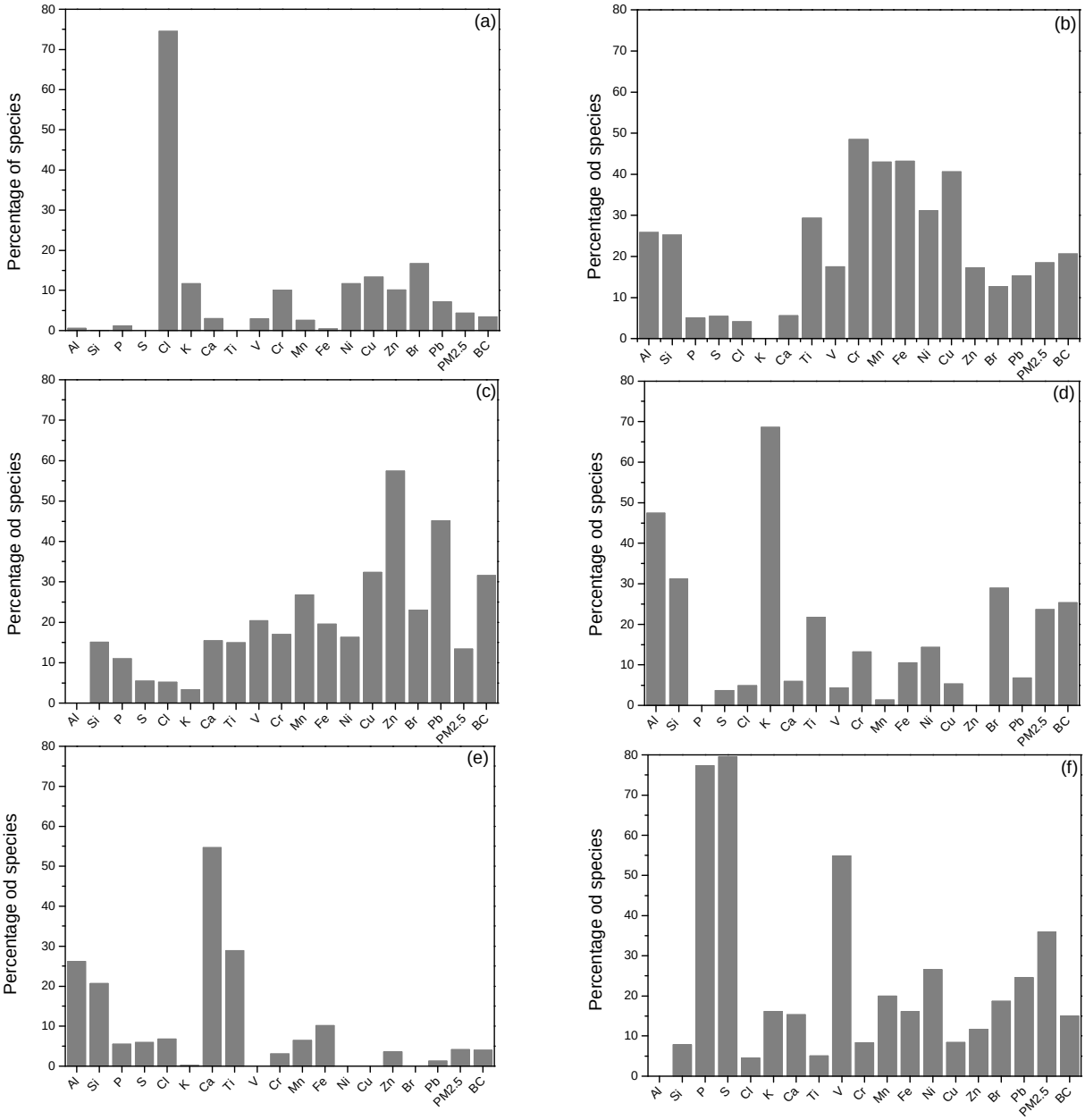


Figure 5: Distribution of identified species participation: (a) Vehicular emission (light fleet), (b) and (c) Vehicular emission (heavy fleet), (d) Biomass burning, (e) Soil, (f) Oil Fuel Burning/ Secondary Aerosol

### *4.2.2 Number of factors*

Several tests were accomplished focusing to found the best number of factors that generates physically real results. It was executed tests using 5, 6, 7 and 8 factors. In order to analyze these results the Q values were used. Using 5, 7 and 8 factors the results were unrealistic considering that Q values were far away from each other and far away from the theoretical prediction. Therefore, the run with 6 factors was selected, because it presents the better results. The 3 Q values were closer, $Q_{theoretical}$ = 1966.00, $Q_{robust}$ = 1930.85 and $Q_{true}$ = 2175.21.

### *4.2.3 Source apportionment*

The PMF model was executed 7 times, with the objective of finding the Q local minimum. The results presented hereafter are related to this minimum Q value. The figure 5, (attached in the end of article), presents the results obtained in the sources profile apportionment, more specifically, the participation of species in the analysis process. It can be inferred from the Fig.5a that higher percentages are related to the following elements: Cl, Cu e Br. These elements indicate the emission by lighy-duty vehicles. The heavy-duty vehicles emission was identified in the factors 2 and 3, as showed in the Figs. 5b and 5c. The higher percentages of Cr, Mn and Fe were founded in the factor 2; the factor 3 presents Zn, Pb and BC as higher percentages. Due to the Al, Si and K, presented in the factor 4 (Fig.5d), biomass burning could be characterized. The Fig.5e shows higher percentages for Al, Si, Ca and Ti, therefore it can be concluded that the soil is the most important source. The higher participations of Si, P, V and $PM_{2.5}$ in factor 6, Fig.5f, indicate a probable source of fossil fuel burn or secondary aerosol.

**Table 1: Contribution (%) of identified sources**

| Source Identified | % |
|---|---|
| Vehicular emission (light fleet) | 4.33 |
| Vehicular emission (heavy fleet) 1 | 18.5 |
| Vehicular emission (heavy fleet) 2 | 13.4 |
| Biomass burning, | 23.66 |
| Soil | 4.17 |
| Oil Fuel Burning/ Secondary Aerosol | 35.94 |

Following the PMF results, the source identification can be quantified. The results are displayed in the table 1, (attached in the end of article). It can be concluded from table 1 that vehicles have a large contribution in the fine particulate formation, with a total of 36%. The oil fuel burning/secondary aerosol presents almost the same contribution. Moreover, with approximately 24%, the biomass burning presented a significant participation. With a lower percentage, the soil do not has a large influence in the fine particulate formation, mainly because its source is a course particulate predominant compound ($PM_{10}$).

For a better evaluation of the results obtained by the PMF model, a comparison was accomplished between the predicted data and the concentration observed, Fig.6. In this figure it is possible to observe a good agreement between the two variables. This result can be reinforced by the $R^2$ = 0.8.

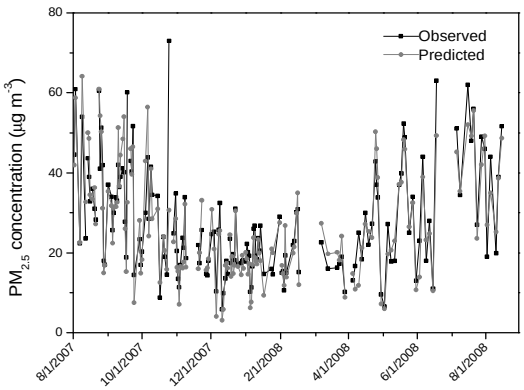


Figure 7: Comparison of the $PM_{2.5}$ concentration predicted and observed.

## 5. CONCLUSIONS

The most significant $PM_{2.5}$ and BC concentration occur during the winter, season characterized by low precipitation and more stable atmosphere. The lower concentrations were observed during the summer.

It can be concluded that 6 factors were adjusted by the PMF receptor model and 5 sources were identified: vehicular emission, lighter and heavier ones, biomass burning, soil and oil fuel/ secondary aerosol. The vehicular emission showed the more significantly participation in the fine particulate formation, with about 36%. Moreover, the heavier vehicles show the larger contribution, due to the diesel burn. The oil burning presents a comparable participation with the vehicular emission, about 36%. This fact can be justified by a local source of boilers, due to the fact that the measurement point is located near a large hospital. In addition, other sources could be identified, like soil and biomass burning, but the last ones with a minor importance.

The predicted data fits very well the concentrations observed, $R^2$=0.8. The $R^2$ result indicates a good agreement between the model and the observed data. Therefore it can be inferred that the PMF technique is an important tool in the fine particulates source identification.

## 6. REFERENCES


Andrade, M.F., Orsini, C., Maenhaut, W., *Relation between aerosol sources and meteorological parameters for inhalable atmospheric particles in São Paulo city, Brazi*l. Atmospheric Environment, Vol. 28, No. 14, p.2307-2315, **1994**.

Castanho, A.D.A, Artaxo, P. *São Paulo Aerosol Source Apportionment for Wintertime and Summertime*. Atmospheric Environment, vol 35, p. 4889 – 4902, **2001**.

Cetesb – *Relatório de Qualidade do Ar no Estado de São Paulo – 2007*, Secretaria do Meio Ambiente, São Paulo ,(INSS - 0103 - 4103), **2008**.

Hedberg, E., Gidhagen, L., Johansson, C., *Source contributions to PM10 and arsenic concentrations in Central Chile using positive matrix factorization*, Atmospheric Environment, 39, 549 – 561, **2005**.

Hopke, P. K., *A guide to Positive Matrix Factorization*, Prepared for Positive Matrix Factorization Program, Postdam, NY, by Department of Chemistry, Clarkson University, **1998**.

Lee, E., Chan, C. K., Paatero, P., *Application of Positive Matrix Factorization in source apportionment of particulate pollutants in Hong Kong*, Atmospheric Environment, 33, 3201 – 3212, **1999**.

Norris, G., Wade, K., Foley, C., *EPA Positive Matrix Factorization (PMF) 3.0 Fundamentals & User Guide*, U. S. Environmental Protection Agency, Office of Research and Development, Washington, **2008**.

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

Paatero, P., *Least square formulation of robust non-negative factor analysis, Chemometrics and intelligent laboratory system*, vol. 37, 23 – 35, **1997**.

Raes, F., Dingenen, R. V., Vignati, E., Wilson, J.,Putaud, J.P., Seinfeld, J. H., Peter Adams, *Formation and cycling of aerosol in the global troposphere*, Atmospheric Environment, 34, 4215 – 4240, **2000**.

Reff, A., Eberly, S., Bhave, P*., Receptor Modeling of Ambient Particulate Matter Data Using Positive Matrix Factorization: Review of Existing Methods*, Air & Waste Manage. Assoc., 57, 146 – 154, **2007**.

Sánchez-Ccyllo, O. R., Martins, L. D., Ynoue, R. Y., Astolfo, R., Miranda, R. M., Freitas, E. D., Borges, A. S., Fornaro, A., Moreira, A., Andrade, M. F., 2009. Vehicular particulate Matter Emission in Road Tunnels in São Paulo, Brazil. Environmental Monitoring and Assessment, v. 149, 241-249.

Seinfeld, J. H., Pandis, S. N., *Atmospheric Chemistry and Physics: from Air Pollution to Climate Change*, Willey, New York, **1998**.