Oponentní posudek na doktorskou disertační práci

(Ph.D. Thesis opponent review)

Devraj THIMMAIAH, M.Sc.

Source Apportionment of Sub-Micron Aerosols from Combined Particle Number Size Distribution and Gaseous Composition Data by Bilinear Positive Matrix Factorization

Vypracoval (opponent) : RNDr. Vladimír Havránek, CSc., Ústav jaderné fyziky AVČR, v.v.i.

The Thesis deals with the searching for the possible sources of aerosol particles and gaseous pollution in the Prague city region. The high time resolution gaseous and particle number size distribution data from two short range acquisition campaigns (February to March 2007 and January 2008) were used as input data for the advanced multivariate statistical method of source appointment, the Positive Matrix Factorization(PFM). There are nowadays several approaches and algorithms which solve the Positive Matrix Factorization problem (special case of the Factor Analysis with constraints on values of factor profiles and factor contribution matrices, allowing only solution with positive values both in the factor and loading matrices). The author chose the PMF2 computer code implementation of PMF from P.Paatero to decompose data obtained during described campaigns.

The Thesis is divided into five parts. The first part deals mainly with the brief theoretical introduction into the PFM, data preparation, PFM2 code settings and additional approaches of source identification.
The second part reviews available literature about the PMF decomposition. It is presented mainly as a short historical introduction, justification of PMF approach and only chronologically sorted list of papers dealing with the PMF method. In this part I would expected more detailed analysis of listed papers. Evaluation of the papers content, reporting the advantages and disadvantages of PMF approach. Trying to asses the area of feasible application (used input data - particle data, mass data, chance to identify local sources, long range transport atc., what factors can be explained and what not, comparison with other approaches). Some critical overview of present experience with the method also included the negative experience or critical attitudes if they are present in literature.
The third part describes the experimental procedures and used analysers. This part is clear. It also contains the measured data a deals with the data preparation and technical details of the PMF2 code settings. Concerning this part I will have some minor technical question later in the text.
The fourth part shows the obtained results of the PMF decomposition and subsequent CPF analysis. There is also a detailed description of obtained factors from both sampling campaigns and descriptive attribution of obtained factors with possible pollution sources. Concerning this part, I will have same principal comments about the consistency and interpretation of obtained sources later in the text.
The last part is a conclusion with a brief recapitulation of obtained results.

What is missing to me, is a clear description of the Thesis aims, the author contribution to the presented research and the declaration of the main scientific output of this work. I would like the author to clear it up during the Thesis defence.

In the next part a list of questions and remarks concerning the text clarity, input data and obtained results is enumerate. It can be divided into the technical questions and principal questions.
The technical questions are as follows:

- On page 2. the Prague area should be 496 km^2 (wikipedia source). What is the origin of the presented data?
- There are also much more receptor studies dealing with the air quality and environmental pollution in the Czech Republic, then the author presented on page 2. For instance following works published in reviewed journals: (E.Swietlicky,R.Krejčí, NIMB 109/110(1996) 519-525, P.Bohm et.al, Environmental pollution 102(1998) 243-250, partly also in J.Kučera et.al.,Biol.Trace el. Res. Vols.71-72(1999)233-245). And there are probably more.
- Next, I would appreciate if the author can comment on the technical details of the PMF algorithm, uniqueness of the solution, definition of FPEAK parameter, result dependence on the input matrix scaling (dependence on units used for the species concentrations, description of Fkey and Gkey matrices, uncertainty matrix preparation (special the justification of the equation 7 on page 40).

- What number of bins (particle size intervals) is optimal for PMF (more or less?), What can caused the collinearity of neighbouring bins values and its errors.
- The NOx concentrations on graphs on p.23 and p.26 seems to be higher then the sum of NO + NO2. Is it a printing error or it has a real base?
- I have some doubt about the measurement of the relative humidity RH (why it is saturated at aprox. 85% - plateau in graphs- and never reach 100%, even during the precipitation periods).
- There is a strong clustering of CO data on the graphs on p.73 and p.75, around the 200,300,400, and 500 ug/m3, this effect probably also corresponds with the some straight horizontal parts in CO concentrations on graf.IV.13 p.61. Is it a sampling or data reading artefact? There are still some fine distributed data between these values, lake that there are two data sets of input data.
- What is the meaning of error bars in the graphs of source profiles, weekday/weekend diurnal graphs and weekday/weekend averages graphs. How they are defined and how they are calculated?
- The S1 and S4 factor in the graph IV.17. are probably swapped.
- The CH4, NMHC and THC concentrations are reported in „ppm". Is it volume or mass ppm? What is the meaning of the parameter k in equations on p.21? How the k reflects the variability of individual hydrocarbons in total NMHC?

There are also a principal questions about the stability, feasibility and utilization of derived factors. The best is, if the obtained factors can be interpreted as general source profiles with a general validity. Is this the case of presented study?
Despite the fact, that both the sampling campaigns were performed at similar seasonal period (in winter or early spring) the derived factors are not comparable. Simply if we compare the graph IV.9 (p.55) and IV.17(p.71) and the traffic source(s) contribution to the total particle concentration (S2-first champaign and S2,S3-second champaign), they contribute only 2.5% (150 particle/cm3) in winter 2007 and 72% (6100particle/cm3) in winter 2008. This seems to be unrealistic result as the traffic is relatively stable and well defined source of the pollution over the Prague region. Also the corresponding source profiles has a different shape and they are not representative for longer period. There are only a slight similarity for the heating sources.
The main problem, in my opinion, is that the particle number concentration distributions is not a good „tracer" for the source discrimination over larger areas or longer times. The main problem (mentioned also on p.28) is the variability of the particle size distribution during the transport from source to receptor size, caused by particles coagulation, nucleation episodes, hygroscopic effects and ambient atmospheric chemistry processes. The necessary requirement for successful source factor decomposition is the conservation of the original source fingerprint during the pollution transport. This is probably rarely fulfilled for the particle size distribution. The reason, why we can still observed some meaningful diurnal patterns in the derived factors, can be the presence of some very local sources or small portion of original particles which are not internally mixed during the transport to the receptor site. It would be nice to try to reconstruct the mass contribution of individual sources and compare it with some dispersion studies for the city and expected source contribution (as heating, traffic, industry and long range transport).
There is also a question about ozone data utilisation for this kind of receptor studies. The tropospheric ozone is mainly produced continually under the solar radiation, with enhance production in polluted atmosphere, it sinks rapidly to surface, so during the calm periods without solar irradiation its near surface concentrations decrease rapidly- night minima (in fact this behaviour resembles a negative source), but it also can be transported over large distances on windy periods. In this particular case we can asked if ozone data can bring a new information into the receptor model or they represent so called trivial factor which has no correlation with rest of the data. In this case, some authors of factor analysis studies recommend to leave out the trivial factor and perform the factorization using the remaining data only. This may result in better source decomposition special in the case of second champaign.
Despite of my critical remarks, both the formal and the scientific level of the presented work is satisfactory for successful defence of this Thesis. The author did considerable peace of work. In fact it was first time that the PMF algorithm was used for this kind of environmental studies at Prague region, and it shows its advantages in fine diurnal pattern recognition of possible sources of pollution around the receptor site. It also shows some drawbacks if particle number concentration with only limited number of other pollutants is used for the source identification and quantification. I hope this finding will stimulate the discussion about the further improvement of the source appointment studies in the environmental pollution studies at the Czech republic and elsewhere.

I recommend the Thesis to be accepted for successful defence.

V Praze 26.3.2009

Vladimír Havránek