

Interactive comment on “Measurements and receptor modeling of volatile organic compounds in south-eastern Mexico City, 2000–2007” by H. Wöhrnschimmel et al.

Anonymous Referee #1

Received and published: 5 April 2010

Measurements and receptor modeling of volatile organic compounds in south-eastern Mexico City, 2000–2007 by Wöhrnschimmel et al.

This paper describes the analysis of time series of 13 volatile organic compounds determined during 2000–2007 period at South-east of Mexico City using linear regression. A brief description of season and time of the day of the 13 species trend is provided claiming a concentration decrease during the studying period. One of the objectives of the manuscript is to identify and evaluate the contribution of major sources of VOC using a receptor model (Chemical Mass Balance model). Important sources identified included domestic use of liquefied petroleum gas and vehicular exhaust with 70 and

C1240

80% of total VOC. However, the manuscript has severe shortcomings and the uncertainty is not estimated so there is not way to judge the reasonableness of the input data, and hence no way to evaluate the final results. There are numerous editorial errors that need to be corrected. General comments: Certainly the above mentioned constitutes a unique long term database of VOC for a specific area within Mexico City. Nevertheless, there should be more discussion about the overall composition of the VOC, and also it is important to include the uncertainty of the measurements for both ambient and emission profiles. The manuscript does not show unambiguously how errors were used in different parts of the work, nor what their signal-to-noise ratios were in the experimental work. This is important when using receptor models as the uncertainty plays an important role in calculations. Authors did not mention on how important the inclusion of uncertainty is to the whole process, it is claimed that they represent "the standard deviation of the average value". For most variables, this "std-dev of average" is higher than the average itself. I wonder if the std-dev showed in table 1 represents the variability of the compounds for the whole period, which is different from what is used in the model. Overall, using so uncertain data base in CMB appears questionable or perhaps outright wrong. Authors need to convince the reader that using less than 50% of the total mass, which is approximately the percentage of the 13 species in the VOC total mass, is valid to apply CMB. To have a better apportionment of sources it is necessary to have other key compounds that are important to trace specific sources. Authors analyzed 13 chemical species (ethane, propane, propylene, butane, acetylene, pentane, hexane, heptane, benzene, octane, toluene, nonane, o-xylene) that accounts for 50% of total VOC mass. This must be some form of artifact. When used in CMB, such an artifact invalidates the results entirely. Together, propane and butane account for more than 60% of the 13, it is highly likely that this fact bias the source apportionment results, giving an overestimation of the LPG contribution. Authors claim that this source has a median contribution of 68% at night, 60% during the day and 42% in the evening. How the authors can explain this? If they mentioned that this is due to leakages and why this is more important at night. They also mentioned a higher increase in November to

C1241

February with 70%, claiming that this is due to usage of LPG for heating, but in Mexico which percentage of the population have such a heating systems? It is more likely that people use electricity rather than LPG heating systems. According to previous reported results in the area, the proportions of LPG-related species are in disagreement, how can the authors explain these results? Is this reflected in the CMB results, i.e. lower contribution of LPG during the afternoon period? How is explained the weekend lower levels if most of the contribution come from the LPG? Authors need to include the validation of their database for the CMB results, as the results do not sound reasonable. They need to mention how they decide the source profiles and number of fitting species. How it can be explained that all sources reported (LPG, EXHAUST, HOT-SOAK, SOLVENT and FOOD) are higher during the evening, if the important activities in the urban area take place in the morning, especially vehicular emissions and LPG handling, which as been reported by several authors, On the other hand, seems that the model has not sensitivity to differentiate HOTSOAK, SOLVENT AND FOOD as the results are quite similar at different time of the day. Authors use erroneously the term source activity throughout the document, do they mean contribution of the source?

The main assumptions on which CMB models rely are: (a) all the sources, contributing significantly to a receptor site, have been identified and have had their emissions chemically characterized (b) chemical species do not react with each other, i.e. they add linearly. In this sense authors should not use xylene as a fitting specie, (c) compositions of source emissions are constant over the period of ambient and source sampling (d) the number of sources is less than the number of chemical species (e) the source compositions are linearly independent of each other (f) measurement uncertainties are random, uncorrelated and normally distributed The two fit indices, R^2 and χ^2 , are considered by the US EPA as primary performance measures of an EFWLS solution, along with the percent mass accounted for %mass, although the latter one can be misleading when the total mass concentration that has been measured for the VOC of the ambient sample is small, as is the case in this study. The fraction FracEst of those source contributions that have acceptably large projection lengths in the eligible

C1242

space is another primary performance measure that has been established by the US EPA so as to validate the well-conditioning of the least squares system being solved. Each T-statistic ratio T_{statj} , is an additional performance measure, established by the US EPA, as an indicator of whether j source contribution is below detection limit or not. Low T-statistic values for several source contributions may be caused by collinearity among their profiles; the presence of collinearity could be uncovered in this case by the source contribution's projections in a properly adjusted eligible space. Finally, each ratio $(Res/Unc)_i$ is also an additional performance measure, established by the US EPA, which specifies the number of uncertainty intervals by which the calculated and measured concentrations of i species differ. In general, authors mentioned only the primary performance measures and do not mentioned the other at all.

I question whether the single paragraph is sufficient background for their methodological application. Neither the instruments used for ambient samples nor the analytical techniques or the methodology to run the receptor model are mentioned in the manuscript. References are given where this information can be found, especially for emission profiles. It would be opportune to give at least a brief summary, though. I do not find adequate to put a personal communication as a reference for a new profile for vehicle exhaust determined in Guanajuato, as this should be describe, how do they account for changes in vehicle technology and fuel characteristics, this will be interesting. How were gasoline catalytic vs. noncatalytic profiles discerned from each other? Why diesel was not take into account?. Give more detail about the CMB results (uncertainty, apportionment for individual species if any). It was confusing, as authors state in the abstract that VOC species were decreasing during this period and in the introduction the mentioned that air pollution increase, and that in Mexico ozone formation is VOC-limited, so authors should clarify this.

In the introduction the toxicity of some VOC as well a debate on the LPG contribution. Would it not be possible and advisable to discuss the findings on this regard in the Conclusion?

C1243

C1244