

Interactive
Comment

Interactive comment on “Factor analytical modeling of C₂–C₇ hydrocarbon sources at an urban background site in Zurich (Switzerland): changes between 1993–1994 and 2005–2006” by V. A. Lanz et al.

V. A. Lanz et al.

Received and published: 18 February 2008

We would like to thank referee 2 for carefully evaluating our manuscript and raising interesting questions.

Comment 1: How did the ethane factor for 1993-94 get converted to ppb (Fig 10) since the ethane concentrations were only measured quantitatively in 2005-06 (pg 912)?

The ethane series was converted to ppb before PMF analysis by a calibration factor accounting for its trapping efficiency below 100%. However, we should add that this holds true for both campaigns and for ethene and ethyne as well. Thus, we will replace

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



the present sentence on p. 912, lines 6-9 by “Trapping efficiency was below 100% for the C₂ hydrocarbons, which however was corrected by a specific calibration factor.”

Comment 2: Using a modeling uncertainty of 10% is not unusual for PMF analysis (pg 914). However the suggestion that this accounts for the variability in the atmospheric lifetime of the compounds is surprising. Presumably the concentration of OH varies by much more than 10% between days. In addition the lifetime of the species vary by a factor of 10 or more, as pointed out by the authors. If this uncertainty is accounting for the variability in the lifetime, why is a value 10% appropriate?

The OH variability is probably not too important for the modeling uncertainty here, because – as pointed out in the paper – we measured close to the sources and ageing of the profiles did not drive the variability in the data (p. 914, lines 14-24). But we generally agree with the referee that by using 10% modeling uncertainty, we can not necessarily account for the different reactivities. The argument needs to be put the other way round: as the substances are subject to reactivity and react differently, a modelling uncertainty >0% needs to be defined (in real data sets other assumptions about receptor modelling, see p. 909, lines 15-16, can be violated, too.). 10% is an arbitrary choice (we have to state this more clearly in the updated manuscript). As far as we know, there is no theory from which the correct modeling uncertainty can be derived. Instead, we can show the sensibility of the results to modeling uncertainties = [5%, 10%, 15%, 20%]. For a detailed reply concerning the reactivity issue we refer to our answer to Bo Larsen.

Comment 3: The use of k-nearest neighbor to impute missing values in PMF (pg 915) is an interesting and, to my knowledge, novel strategy worthy of further discussion. How much difference did this make as compared to using the more common approach of assigning values based upon a fraction of the detection limit?

In the earlier data (years 1993-94), ethane and ethyne measurements were not labelled

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



“missing” because they were below detection limit (BDL) values, but missing due to data acquisition problems. Polissar et al. (1998) proposed to impute such values by their geometric mean and assumed an uncertainty of four times the geometric mean. The ratio x_{ij}/s_{ij} for missing values is constant at 0.25 following this approach. Using our approach (*knn* method; for an evaluation of the *knn* in atmospheric science see Junninen et al., 2004), x_{ij}/s_{ij} is variable, accounting for different concentration levels. When the missing concentration values are non-randomly distributed, the geometric mean can be biased, e.g. if the missing values coincide with persistent temperature inversions and VOC accumulation phases.

Comment 4: Were the less than 1% of the values that exceeded the model outlier threshold for the scaled residuals down-weighted (pg 918) or excluded (pg 916)?

We differentiated between two distinct types of outliers. One type stands for contaminated samples in the data (p. 916) (e.g. gas grill emissions). These samples were excluded. On the other hand, “model outliers” were defined in terms of large deviations from the uncertainty weighted error (eq. 2), $|e_{ij}/s_{ij}| > 4$. Errors of the latter type were down-weighted. In the revised article, this difference will be pointed out.

Comment 5: What does the “n=13” on page 919 represent? Is this the number of components in the profile or the number of reference profiles compared? Assuming it's the former, what type of correlation coefficient is reported? Was this correlation biased by the dominant components in the source profile?

In this context, n represents the number of components and R is the Pearson's correlation coefficient, retrieved from linear curve fitting. In addition to high R -values, the regression slope follows approximately the 1:1-line (modeled vs. computed profile) and, thus, the distribution of the values within the literature source profile can be derived from the modeled profiles shown in Figs. 2 and 3. In the first example, on p. 919, line 9 removing the highest values (isopentane) from the fit still yields $R=0.88$

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

($n=13-1$), after removal of the second highest values (butane), R remains high (0.61, $n=13-2=11$), the third highest value is the sum of isohexanes ($R=0.73$, $n=13-3=10$), fourth isobutene ($R=0.68$, $n=13-4=9$) etc. A similar behavior can be observed for the other correlations reported here. Therefore, we can say that these high correlations are not due to leverage effects.

Comment 6: The conversion of ratios of emission factors to concentration ratios on page 920 is not clear. How does a change from 2:1 to 5:2 get equated to a doubling from 1:1 to 2:1?

The ratio toluene:benzene (on a mass concentration basis) was converted to toluene:benzene in ppb/pbb as follows:

$$\text{toluene:benzene [mg/mg]} * ((1/M_{\text{toluene}})/(1/M_{\text{benzene}})) = \text{toluene:benzene [ppb/pbb]},$$

where M are the molecular weights. Thus, the ratio 16:10 [$\text{mg km}^{-1} / \text{mg km}^{-1}$] as found in Staehelin et al. (1998) will be converted to 1.4:1.0, which is approximately 1:1, and 5:2 [$\text{mg km}^{-1} / \text{mg km}^{-1}$] will be converted to 1.3:0.6, which equals approximately 2:1. The numbers in the latter ratio have no absolute meaning each, but a relative one. In the revised article we replace the (rounded) ratio 2:1 (p. 920, line 11) by 16:10, which we actually used for conversion.

Comment 7: Is the correlation of the gasoline source with temperature ($R = 0.45$ and $R = 0.3$) statistically significant? Should the temperatures on figures 7 and 10 correspond with each other? For example the average temperature 10-15°C for January 2006 (Fig. 7) does not match with the value of about 4°C on Fig 10. Are the labels for the values in Fig. 7 for January 2006 and October 2005 perhaps reversed?

The (Pearson) correlation coefficient (R -value) is just a measure for the linear dependence of two variables. Statistical tests for the significance of R are rather unusual.

This is different for the regression coefficients. The regression slope of “factor 1” vs. temperature ($n=12$) as shown in Fig. 7 is not significant on a 95%-confidence level for both campaigns. However, using hourly data ($n=10^4$ - 10^5 , which can not be as clearly arranged as monthly values) yields less correlation (15% and 29% for 2005-06 and 1993-94, respectively), but the (positive) regression slope estimates are significantly different from 0 on the 95%-confidence level. We will summarize the details of these additional linear regression fits in the caption of Fig. 7 in the revised version. Technical correction in Fig. 7 (top): “January 2006” should read “October 2005”; “October 2005” should be replaced by “November 2005”.

Comment 8: In Table 5, how were the mean concentrations for the factors (last column) combined to give the mean values for the source groups? The method should be described, as it does not appear to be a simple summation.

Table 5 was mixed up during the production process and contains some typos as well. We apologize for this. In the upper part of the table (“PMF factors 2005-2006”), the numbers from factor 5 and factor 6 were reversed. The lower part represents the summation involving the factors named in the first column. “Gas leakage” equals the sum of factor 3 and factor 4 (rather than the sum of factor 1 and factor 2). Table 5 will be compiled anew; Figure 13 represents an accurate assignment of labels and VOC source contributions.

Comment 9: The low concentrations of both the ethane and wood burning factors in Dec 1993-94 is curious and perhaps worthy of discussion on page 926. Was there actually much less wood burning in December than November?

The wood burning and ethane contribution were comparatively high in November 1993 and February 1994, while they were relatively low in December 1993 and January 1994. (Fig. 10, top). This can be (partly) explained by low wind speed and temperature in November 1993 and February 1994, indicating thermal inversions and VOC accu-

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

mulation. Please note that the legend in Fig. 10 (bottom) needs a technical correction: the dashed lines represent the earlier meteorological data (1993-94), whereas the solid lines represent the data of 2005-06.

Comment 10: It is recommended that the wording in a few places be revised. Title: "Factor Analytical modelling" P908, line 23 "complex regarding." P909, line 4: "got special.." P911, line 5: "located at a .." P911, line 23: "extracted from hydrocarbons.." P918 line 10: "first-guess attribution.."

We will rephrase the passages mentioned above.

Comment 11: The word factorial in the figure and table captions (e.g. 6-factorial PMF solution) may result in misunderstanding due to its mathematical meaning. "6-factor PMF solution" might be better.

We will change the manuscript accordingly.

Comment 12: Table 2 - 4 captions: many of the coefficients are less than 0.7. Should the caption read $0 < R < 0.8$?

*R-values between $0.7 < R < 0.8$ are shown in italics. *R*-values $R < 0.7$ are written as standard numbers.*

Reference:

Polissar, A. V., P. Paatero, P. K. Hopke, W. C. Malm, and J. F. Sisler (1998), Atmospheric aerosol over Alaska 2. Elemental composition and sources, *J. Geophys. Res.*, 103(D15), 19,045–19,057.

Interactive comment on *Atmos. Chem. Phys. Discuss.*, 8, 907, 2008.

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)