**Atmospheric
Chemistry
and Physics
Discussions**

Interactive comment on "**Statistical analysis of non-methane hydrocarbon variability at a European background location (Jungfraujoch, Switzerland)**" *by* V. A. Lanz et al.

V. A. Lanz et al.

Received and published: 8 April 2009

We are grateful to referee 1 for his/her comments. Detailed answers to his/her critical comments can be found below. Excerpts from these comments appear in *italics*, directly followed by our answer. The involved changes to the manuscript are written in **bold**.

Comment 1a: *'[. . .] my main concern is that 4 factors do not adequately describe their observations; [. . .]'*

We were also surprised that only 4 factors can explain the main part of the variability of such a complex system: from a statistical point of view, all necessary diagnostics

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

were provided and show that 4 factors sufficiently describe the data (see report of the calculated versus expected scaled residuals, P. 19535, Lines 12–17; variability explained by the 4-factorial model, Section 3.3 on PP. 19537–19538; projection of the PMF-profiles along with real samples in the principal components' space, Figure 1 on P. 19554, and the corresponding discussion on P. 19535, Line 18 – 19536, Line 8). In order to stress this important result, **the variability of each compound that is explained by the PMF model (described on PP. 19537–19538 and deducible from Figure 2) was added to Table 1 (last column).**

In an ideal case of source apportionment based on factor analysis, each factor can be related to one individual source. However in real data cases, several sources have either a similar VOC fingerprint and/or sources with different VOC fingerprints coincide temporally (i.e. they exhibit similar time series at the receptor site and hence the resulting PMF profiles can represent a variety of sources each) and/or the variability in the data is influenced by aging. These latter circumstances will result in factors that include several source-types, leading to a reduction of the actual number of factors needed in factor analysis, and factors that must not necessarily be equated with emission sources (Factor 1, as an example, is also representing oxidized air and therefore can not be equated with an individual emission source). We did not say, nor mean to imply, that there are generally only four relevant VOC sources (with an influence on the VOC levels on Jungfraujoch) that are represented by fours factors retrieved from PMF modeling of the selected NMVOCs. Please note that factor analysis is used here to reduce the dimensionality of the dataset (i.e. to 'explain' the variability of the samples) and therefore the individual factors cannot by viewed to represent individual sources (which is indeed different when studying measurements at a high mountain site compared to a 'classical' receptor modelling study for an urban site where the investigation of the individual sources' contribution to ambient air concentrations is the primary goal (P. 19530, Lines 2–7; P. 19533, Lines 10–11).

Comment 1b:*'[. . .] they should consider incorporating other gases [. . .]'*

As both referees suggest the incorporation of more trace gases into the PMF calculations we understand that we have not been explicit enough about the selection of the compounds' subset as presented in this paper. We differentiate between the consideration of a) additional organics and b) additional inorganics for PMF:

a) On the choice of organic compounds

We agree that many receptor studies included a larger number of measured species in the factor analysis. The currently chosen dataset for PMF includes all organic compounds measured at Jungfraujoch with atmospheric lifetimes between 2 days and 6 months (P. 19535, Lines 2–5; P. 19542, Line 26; Table 1 on P. 19551) which have been measured during the last 8 years. The concentrations of trace gases with longer lifetimes have different properties which are not studied in this paper. Shorter lived organic species (e.g. MVK) have only been measured during four seasonal campaigns in one year and therefore they were not included in this study. (The results of these campaigns were published by Legreid et al., 2008). Very short-lived compounds (e.g., isoprene) have too low signal-to-noise ratios at this site to interpret their variability, including more persistent substances (such as HFCs) yields factors that do not represent sources but rather long-term trends or hemispheric exchange, because emissions of the latter species were drastically reduced during the past years. Furthermore, compounds with shorter (< 2 days) or longer atmospheric lifetimes (> 6 months) are expected to define their own factors, when included into a data matrix that consists of compounds with intermediate lifetimes (in these cases factor analysis is probably not the best tool for the analysis).

b) Why not to include inorganic species and methane in the PMF matrix

We agree that the inclusion of inorganic species and methane would increase the number of species in the factor analysis. The latter compounds were included in the presented analysis already as 'external PMF markers' (i.e. the time series of the PMF factors were compared with the time series of the inorganics; P. 19540 – P. 19541,

Table 2 on P. 19552, and Figure 5 on 19558). However, we neither included inorganic species (such as NO_x and CO) nor did we include CH_4 in the factor analysis because of the following reasons:

• Based on the concentration ratios of the species we first attributed the factors to different source regions/sources and different atmospheric processes/lifetimes. However, we found it very useful to have independent information in order to critically discuss the attribution of the factors to source regions and atmospheric processes. For this discussion the concentrations of the inorganic species and methane was very useful.

• Entity of chosen organics. The presented data set for PMF (Table 1, P. 19551) consists of primary and predominantly anthropogenic substances. In this respect, the chosen organic gases represent an entity. However, other species have other sources such as the oxidation from other hydrocarbons (e.g. CO in the free troposphere) or other emission sources (such as methane which includes strong biogenic sources). The inclusion of largely biogenic and partially secondary species breaks down the entity of the chosen organics and furthermore bears some conceptual problems: either the secondary portion of carbon monoxide, say, is attributed to a factor representing primary emissions or, conversely, secondary portions of some inorganic species may form their own factor, but the exclusively primary organics might nonetheless be forced by PMF to positive values within the same 'secondary factor'. Such artifacts and more, similar ones might be introduced to the PMF analysis when these inorganic species are added to the organics *a priori*. This was avoided by the current selection of PMF species and making use of the inorganics *a posteriori*. It would be rather unfortunate if primary compounds were mathematically used to reconstruct secondary components and vice versa. By using exclusively the organics for PMF, a straightforward interpretation of the factors is possible.

• Data weighting. A critical trait of the algorithm used in PMF is that each data point is weighted by its uncertainty. PMF is based on an uncertainty weighted least-square approach (Equation 3, P. 19532). This implicitly also means that ideally only compounds

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

that have comparable instrumental errors should be input together into a data matrix for PMF. Different instruments need different calibrations etc. and their errors are often not directly comparable and, therefore, they should not be put into the same PMF data matrix (or a rather highly-subjective weighting of the reliability of the different instrumental errors needs to be introduced). This is different for approaches that do not take data uncertainty into account, e.g. for the simpler setting of a standard PCA (principal component analysis). The concept of PCA however assumes all data points have equals weights. This is not realistic and Paatero and Tapper (1993, 1994) have shown that data-weighting is an advantage of PMF over other techniques of factor analysis.

• Data availability. About 10000 samples of organic trace gases are available for the presented analysis. Since methane (CH_4) was not measured at Jungfraujoch before the year 2005 less than 3000 samples match the other organic samples (see Table 2, P. 19552). Numeric values are needed for each element in the PMF data matrix.

We realize that the issues above were not discussed in the current manuscript and may have caused some confusion. For clarification of the article we added the following sentences to Section 2.2: **'The organic compounds selected for PMF (Table 1) represent an entity with respect to their origin (primary and predominantly anthropogenic sources) as well as to error calculations (all compounds were measured by the same GC-MS technique). OVOC measurements (including compounds with strong biogenic sources) were only measured during four seasonal campaigns in 2005 (Legreid et al., 2008) and were therefore not included in the study. In this analysis correlations between the inorganic compounds and the PMF factors were used to assist factor interpretation (e.g., see Sect. 3.3).'**

Comment 1c: *'[. . .] that may elucidate the i) confounding problem with overlapping gases in multiple factors and ii) no single grouping of gases presented can succinctly delineate each factor/source type. For example, similar distributions of the same chlorinated compounds are present in factors 2 and 4 while factors 1 and 3 have similar*

hydrocarbon distributions.'

First we would like to address the comment under *i)*: as a matter of fact many species, e.g. hexane, can be present in combustion-related sources, as well as in freshly-emitted solvents (see e.g. Passant, 2002). It therefore seems natural that this overlap also is represented here by the factor profiles derived from ambient data. The presence of one single compound in several factors does not pose a problem, as long as the factors are discernable by their characteristic VOC ratios (e.g. different benzene/toluene ratios) or by certain key species (as presented in Table 2 on P. 19553), which is the case here. Concerning comment *ii)* we realize that we have not been explicit enough about the succinct discrimination of factors 1–4. **As a consequence we inserted the following sentences to the Conclusions section (P. 19543, Line 5): 'Factors 1 and 3 mainly explain the C_xH_y variabilities. While factor 1 explains aged combustion- and natural gas distribution-related NMHCs, factor 3 covers NMHCs associated with fresh emissions and solvent-use. Correspondingly, factor 1 is correlated with CO and CH_4 (probably anthropogenic CH_4 as its source regions match the anthropogenic CH_4 emissions), whereas factor 3 is rather correlated with NO_x, a tracer for fresh emissions. The variability in chlorinated compounds ($C_xH_yCl_z$) is covered by factors 2 and 4. While both factors can be associated with industrial VOC use in Northern Italy and Eastern Europe, the contributions of factor 4 (key species C_2HCl_3) are relatively high when air arrives from the West at Jungfraujoch. No such dependency can be observed for factor 2 (key species CH_2Cl_2, C_2Cl_4).'**

Comment 1d: *'Is it possible to attribute factor 1 to combustion alone by including gases such as ethyne and CO? Also, while factor 4 contains a larger fraction of trichloroethene, are other short-lived halocarbon data available to further substantiate this difference?'*

Factor 1 describes oxidized air and is ultimately certainly linked to combustion sources.

Unfortunately (continuous) ethyne concentrations are not available for the receptor site in order to confirm the expected similarity to benzene concentrations: We also agree that further short-lived chlorinated species could shed more light on industrial emissions, but they are not available either. We decided to use CO in this analysis as an external PMF marker, enabling us to confirm the interpretation of factor 1 as aged combustion (please consider our answer to Comment 1b). The correlation of factor 1 (time series) with methane in winter also points to a combustive source.

Comment 2: *'Another important point that must be considered is that this paper appears to focus solely on anthropogenic emissions; no biogenic compounds/emissions were considered. In studying the background levels of tracers, wouldn't biogenic emissions be important at this particular site? Thus, the title is misleading.'*

We agree that the paper focuses on anthropogenic VOCs and we adapted the title of the paper accordingly. The VOC compounds that are suitable for factor analysis AND at the same time available at the Jungfraujoch site are commonly believed to be predominantly of anthropogenic origin (isobutane, butane, isopentane, hexane, benzene etc.). In the study of Legreid et al (2008) additional OVOCs measurements of the Jungfraujoch are included showing indeed that a large proportion of all VOCs are of biogenic origin. However, such measurements are only available for four seasonal campaigns in one single year, whereas in our paper we wanted to explore much longer data series. Therefore (see also our answer to Comment 1b), we do not intend to cover biogenic VOCs in the study. However, we did not intend to claim that biogenic emissions are not important for t-NMVOCs. In contrast, it is possible and even plausible that fresh anthropogenic VOC emissions coincide with local biogenic emissions when mountain venting and turbulence are the driving agents of VOC transport to Jungfraujoch. This process could also be of relevance for toluene (a compound included in this analysis) and we hence modified the manuscript as described in our response to Comment 5 below. We furthermore indicated the possible influence of marine/terrestrial sources (P. 19542, Lines 20–23), but clearly stated that the e.g. impact of biogenic sources on

t-NMVOC can not be inferred from the present study (P. 19543, Lines 5–9). Taking into account your comment we changed the title of the paper into:

'Statistical analysis of anthropogenic non-methane VOC variability at a European background location (Jungfraujoch, Switzerland)'

Comment 3: *'Moreover, it is surprising that no factor exhibits biogenic or terrestrial contributions 8211; one would expect this to be persistent and likely contribute (especially during the summer months), if not make up an additional factor describing the sources and source regions. Also, were any marine tracers measured that may provide additional insight on air mass transport and processing?'*

We did not rule out, nor mean to implicitly rule out that certain factors (derived from predominantly anthropogenic NMVOCs) could be influenced by biogenic or terrestrial VOC contributions. We mentioned this possibility already in the ACPD paper for certain chlorinated VOCs (P. 19542, Lines 20–23) and we tried to clarify this point in the revised article for toluene (see our answer to Comment 5). With the current model solved by PMF the shape of factors is not prescribed. A chemical mass balance (CMB) approach is not judged to be an appropriate method for the Jungfraujoch measurements because (*i*) natural emission profiles are often poorly characterized or not known at all (example: terrestrial emissions of chlorinated VOCs) and (*ii*) it is hardly feasible to include oxidation process in such an approach.

Comment 4: *'Furthermore, the C2 and C3 NMHCs would be useful to include because they may help to better delineate each factor/source type. Isoprene and its oxidation products, methyl vinyl ketone and methacrolein, could be used to constrain air mass processing and transport times during summer months. Also, other pairs of gases such as acetone and propane could provide insight on source regions, air mass processing and transport times. In my opinion, to accurately assess each of the factors, the authors should seriously consider including additional gases – most manuscripts that I have*

read carrying out similar types of analyses use a larger, more comprehensive suite of compounds to constrain each factor/source type.'

We agree that measurements of such compounds would be valuable either to strengthen our interpretation (C2 NMHCs, C3 NMHCs) or to elucidate the role of biogenic emissions (isoprene, methyl vinyl ketone (MVK), methacrolein, acetone etc.). Unfortunately such measurements do not exist or are only available for single intensive campaigns, e.g. (biogenic) OVOCs which were not the focus of the paper. For the latter compounds we refer to the publication of Legreid et al. (2008). (We justified the selection of compounds for factor analysis in our answer to Comment 1b).

Comment 5: *'Because the authors use the benzene/toluene ratio to determine the relative photo chemical age of air masses, it would prove useful for the authors to examine a recent manuscript by M. L. White et al. currently in ACPD regarding a biogenic source of toluene in the rural northeastern US. If a biogenic source of toluene is persistent, this would affect the photochemical age determinations used in the paper - even though they are only referred to as "relative". This air mass age indicator should be used with caution, especially at a site like theirs. Here is the citation for the manuscript: M. L. White et al. (2008), Are biogenic emissions a significant source of summertime atmospheric toluene in rural Northeastern United States?, Atmos. Chem. Phys. Discuss., 8, 12283 -12311. Following suite, the authors bring about an interesting point that should be addressed more comprehensively. For L20 on P19536 the authors state that 'Toluene, on the other hand, does not show a distinct seasonal variation'. Benzene is in Factor 1 and Factor 1 is most dominant in winter, which means there could be a large amount of toluene from anthropogenic emissions. The fact that there is little seasonal variation in toluene year-round (Figure 4) indicates that there is likely to be an equally important contribution from vegetation to toluene in the warm season. This is an interesting point that the authors should expand upon.'*

We agree with the referee that relevant local sources of toluene would clearly im-

pede the calculated photochemical ages as determined via the benzene/toluene ratio. However, earlier investigations of JFJ data showed that (for the studied periods) benzene/toluene ratios can indeed be used for JFJ as a qualitative indication of the photochemical age of the arriving air masses (P. 19536, Lines 18–20). Additionally, in our ACPD article we already pointed out the potential existence of such a local toluene source and its influence on the toluene mixing ratios determined on Jungfraujoch (P. 19536, Lines 4–6) and also discussed the weaknesses of the benzene/toluene ratio as a surrogate for the photochemical age (P. 19536, Line 9 *sqq.*). Thanks to the referee's comment we realized that we should have brought these two latter statements together and therefore added on P. 19536, Line 6 in the revised paper:

'(the presence of such local toluene sources can affect the determination of the absolute and relative photochemical age for the arriving air masses)'

We further thank the referee for hinting at interesting findings by White et al. (2008; ACPD), who successfully related ambient toluene in part to vegetative sources. In the light of the results by White et al., a modification of the manuscript seems appropriate (though there is no clear indication of the nature of this potentially local toluene source mentioned above). Page 19539, Line 24 now reads:

'The seasonal variation in toluene is relatively small compared to other hydrocarbons with similar atmospheric lifetimes (iso-hexane and iso-butane; cf. Table 2 and Fig. 4). A possible explanation for this seasonal behavior is increased evaporative loss of toluene from solvents in summer (as indicated by the seasonal patterns of factors 2 and 3 in Fig. 4). In addition, contributions from vegetation to toluene in the warm season (as reported for a rural site in the Northeastern US; White et al., 2009) can not be ruled out.'

The ACP-version of White et al. is now available and the following citation was added: **'White, M. L., Russo, R. S., Zhou, Y., Ambrose, J. L., Haase, K., Frinak, E. K., Varner, R. K., Wingenter, O. W., Mao, H., Talbot, R., and Sive, B. C.: Are biogenic**

emissions a significant source of summertime atmospheric toluene in the rural Northeastern United States?, Atmos. Chem. Phys., 9, 81–92, 2009.'

Comment 6: *'Additionally, while not necessary, it would be useful to have a figure displaying the time series of the data in addition to the time series of the factor contributions.'*

The original VOC data is shown as monthly boxplots in Figure 4. **In addition to Figure 4, the original VOC data is also shown as time series in a Supporting Section (Figure S1) in the revised article.**

Specific comments

P19528, L6: Reviese 'Variabilities in the NMHC series were modeled by factor analysis.' to 'Variabilities in the NMHC TIME series DATA SET were modeled by factor analysis.'

The manuscript was updated as suggested by the reviewer.

P19529, L2: Begin sentence with 'therefore' to read 'Therefore, Jungfraujoch is a...'

We modified the article correspondingly.

P19529, L16: Replace 'due to' with 'because of'

We changed the sentence accordingly.

P19529, L24: Should read either 'positively constrained' or 'positive constraint'

'Constraint' was replaced by 'constrained' at this instance.

P19530, L2-5: Revise this sentence, very awkward as written.

This section was modified as follows: 'The aim of the present study is to inves-

tigate the variability of predominantly anthropogenic NMVOCs at Jungfraujoch. NMVOC time series for the past eight years (2000–2007) are described by means of factor analytical modeling. This study further evaluates the capability and caveats of factor analysis when applied to reactive substances retrieved from remote measurement sites.'

P19530, L19: Do not start sentence with 'Due to', should say something more like 'As a result of' or 'Because of'

This sentence was started with 'Because of' in the revised manuscript.

P19530, L23: 'measurement' should be 'measurements'

Done.

P19531, L19: 'species' should be replaced with 'gases' or 'compounds'

We replaced 'species' by 'compounds'

P19531, L28: 'is determined' should be replaced with 'has been measured'

The manuscript was updated accordingly.

P19533, L3: Doesn't 'non-negative' just mean 'positive' values? If so, then why not just say so?

'Non-negative' also includes zeros, 'positive' does not necessarily.

P19533, L10: As written, the following does not make sense, please revise: 'As we measured distant to anthropogenic NMHC sources, the factor profiles F can not be directly related to emission profiles.' Specifically, what are 'distant to anthropogenic NMHC sources'?

For the revised manuscript, this sentence was written anew: 'As the measurement site Jungfraujoch is located away from considerable anthropogenic source regions, the factor profiles F retrieved from anthropogenic NMVOC concentrations not be expected to equal fresh emission profiles.'

P19536, L4-6: The 'overwhelming toluene peak' from 'local sources' could be of biogenic origin - please consider looking in to this.

Done (see our answer to Comment 5).

P19536, L26: As written, this says the 'trichloromethane:benzene ratio' - should be 'trichloroethene'

'Trichloromethane' was replaced by 'trichloroethene'

P19541, L22: Please quantify or specifically define what 'nice agreement' is in this case.

The corresponding passage was re-written in the revised manuscript: 'The sources of factor 1-species are potentially located in the Netherlands, Eastern Europe, Southern Italy, and England. These regions are known for their anthropogenic emissions as also corroborated by the EDGAR inventory for anthropogenic CH_4-emissions in Europe (`http://www.mnp.nl/edgar/model/v32ft2000edgar/25edgv32ftghg/edgv32ft-ch4.jsp`) that identifies the same areas as strong methane source regions. Thus, the identification of these source regions points to gas/oil distribution, combustion etc. or wood burning as sources related to factor 1. Furthermore, the same regions show relatively strong CO-emissions as can be derived from the EMEP emission in-

ventory; `http://www.ceip.at.`'

Additional references

Passant: Speciation of UK emissions of non-methane volatile organic compounds, AEAT/ENV/R/0545, Issue 1, 2002.

Interactive comment on Atmos. Chem. Phys. Discuss., 8, 19527, 2008.