We would like to thank the reviewer for his/her helpful comments which improved the manuscript considerably. Answers to comments are written in blue. Changes in the manuscript are marked with red.

Interactive comment on "Exploring sources of biogenic secondary organic aerosol compounds using chemical analysis and the FLEXPART model" by Johan Martinsson et al.

Anonymous Referee #3
Received and published: 27 March 2017

Review of Martinsson et al., "Exploring sources of biogenic secondary organic aerosol compounds using chemical analysis and the FLEXPART model"

Synopsis
In this work, the abundance and time variability of molecular compounds identified in filter samples of ambient air measured at the Vavihill site in Sweden is investigated. The authors then use a numerical Lagrangian particle dispersion model (FLEXPART) and statistical analysis (PCA) to identify source regions for the different compounds they have identified. They find that coniferous forests contribute strongly to monoterpene tracers found in organic aerosol samples.

General comments
The authors attempt to use backtrajectory calculations to qualitatively connect exposure of an air mass to land surface types to molecular markers found in organic aerosol samples, thereby investigating their formation formation processes. This is a nice and straightforward idea. The authors struggle, however, to convince the reviewer that their analysis and modeling has been conducted in a knowledgeable way. Their use of beta release software without justification, plus a number of other vaguenesses in the methods description are a warning sign that considerable caution should be taken before this manuscript can be published. Apart from questions regarding the methods used, the final results of the analysis don't seem to provide much new information. The fact that coniferous forest emits monoterpenes which then forms SOA has been shown numerous times. If there are other important findings, they are not apparent to the reviewer.

In summary I am tempted to recommend rejecting the manuscript due to the deficiencies in the methods section and the lack of scientifically new findings. It should be noted that in my review I cannot comment on aspects of the chemical analysis, as this is not my field of expertise. Hence I will recommend 'major revisions' here, as the chemical analysis might contain information that is novel for other readers.

Specific comments
* Use of beta software in analysis
The current stable release version of FLEXPART is 9.02, while the authors (claim to?) use version 10.0. Software in beta versions is considered unstable and for testing purposes only and can surely not be used in a scientific publication.

The justification for using FLEXPART 10.0 over FLEXPART 9.02 is purely practical: we already had an installation of FLEXPART 10.0 (or based on FLEXPART 10.0: some of the

output subroutines were modified by us) running for another project, and it was simple to re-use it for computing these aerosol footprints.
We haven't encountered any instability, neither have we discovered any weirdness in the results that could point to a specific problem of aerosol simulations in FLEXPART 10.0. Therefore, unless there is such a known problem, we don't see the justification for re-computing the footprints with FLEXPART 9.02.

* HYSPLIT and FLEXPART together
It is unclear why simulations using the HYSPLIT model are 1) done at all and 2) presented as auxiliary analysis which is different from the FLEXPART analysis. Both HYSPLIT and FLEXPART solve the transport equations backwards in time. HYSPLIT as used here calculates single, deterministic trajectories, while FLEXPART calculates a large number (100000 in the present case) of trajectories, applying processes like turbulence and convection stochastically. FLEXPART by default delivers mass-weighted center trajectories and clusters (see documentation), which provide information equivalent to HYSPLIT. There is no additional information gained from the use of the HYSPLIT model, unless the authors start and compare the model results in detail. I recommend removing this completely, at most leaving a sentence stating that they evaluated HYSPLIT and it gave similar results.

We compute these HYSPLIT trajectories routinely, as auxiliary data of the measurements. They are, as the reviewer notices, very simplistic, and we didn't use them in the interpretation of the data. The FLEXPART-based analysis was done in a later stage, and is meant to be more thorough. We however agree that the interest of showing the HYSPLIT data is limited; therefore we removed them from Figure 2.

* Uncertainties due to neglected sources and sinks during transport
On several occasions the authors caution that what they are doing is neither a full source inversion, nor a modeling effort considering (non-linear) effects of chemistry and other sources and sinks in the atmosphere. Statements like: "a formal source apportionment would require precise accounting of these factors, which is extremely complicated and is clearly out of scope of this study" (p 5, l 14-15) leave the reader wondering what this study is about, then, as more then half of the manuscript deals with exactly this kind of analysis on a simple level. This leaves the reader with the uneasy feeling that he/she cannot attribute significance to the findings. How large are those uncertainties? Where do they come from?

Aerosols (some of them at least) contain reactive chemical species. The aerosol mass concentration of an air mass can change with factors such as the amount of solar radiation, the presence of reactive gas species in the air mass (OH, NOx, O_3, etc.), temperature and humidity, etc. Furthermore the aerosol mass concentration can itself influence the aforementioned parameters (i.e. reaction with chemical species will deplete these chemical species, it can change the albedo of the Earth in some wavelength ranges of the solar spectrum, which can in turn affect the temperature, etc.).

It is possible to attempt to reproduce these processes in a numerical model, and to use observations to evaluate the model results and possibly to provide an estimation of aerosols sources/sinks, in a top-down approach. This requires however 1) a more specialized model than FLEXPART, in particular one that can handle non-linear reaction between different transported species, 2) at least some prior knowledge on the aerosol production/destruction associated to each land surface type, and 3) a lot more observations. Two month of measurements of aerosol chemical composition at just one site is not nearly enough to provide constraints on aerosols sources and sinks, even locally.

The main objective of this paper is therefore not to produce an estimate of aerosol production, but to publish our measurements, in the hope that they will be useful to future studies. We attempted to interpret the data, within the limits of what is permitted by the size of the dataset: we chose a relatively simplistic modeling approach, and the scientific conclusions are limited, but it is unlikely that a more complex modeling approach would have led to more robust results. The bottleneck is the amount of data.

* "Surface vertical level" method

The "surface vertical level" definition as half the PBL height for each particle is not a standard FLEXPART output product that I am aware of (at least not in v 9.02). There is no documentation of this feature judging by a cursory look over the available publication (Stohl et al., 2005, ACP) and a quick source code survey. While surely useful, I don't see how the authors have achieved this without coding it themselves. This would have to be described accordingly, if this is what they did. Furthermore, the choice of 1/2 the PBLH is arbitrary, and the reasoning ("non-linear processes" again) is insufficient.

The feature is indeed not standard FLEXPART, we implemented it ourselves. It involves mainly two changes to the code:

- The PBL height at each particle position that is calculated in advance.f90 (in standard FLEXPART) is saved in an array.
- A new output module has been written, which accumulates in a 3D (lat, lon, time) array the residence time of particles between the surface and a user-defined threshold altitude, which can be either a fixed altitude or a fraction of the PBL. In the latter case, the height of these "virtual" surface grid-boxes varies from one place to another, and from one time-step to another, therefore we do not accumulate the residence time directly, but the residence times divided by the "virtual grid box" height and density, so that the resulting response function has a unit of $s.m^2/kg$.

These changes to FLEXPART were done for a different project, for which a manuscript is in preparation. The objective was to improve the representation of the diurnal variability in CO_2 and CH_4 simulations.

For these aerosol simulations, the impact is in fact negligible (the samples are taken over a 24 hours period, so the diurnal variability is smoothed), and that feature was just used as a default settings (there is no strong argument against or in favor of it). We recognize however that it should be better described and evaluated, and since this paper is not the good place for this, all the simulations in the revised manuscript use a more standard fixed surface level thickness of 400 m.

* AERO-Tracer

Justify the use of the particle diameter used, as this has considerable effect on the lifetime of the particles and hence the exposure calculation. I suggest recalculating for large and small particles.

FLEXPART distributes the particle diameters according to the mean diameter setting ("dquer", which we set to 250 nm), and to the "dsig" parameter (which controls the spread of the size distribution. We use a value of 12.5 for dsig, which means that the particles in a 250/12.5=20 nm to 12.5*250=3125 nm make 68% of the total particles mass.

The results are indeed dependent on the particle diameter. Smaller particles travel longer, and will therefore show higher sensitivity to remote land areas. We do not know accurately the size distribution of the particles we measured, but previous size distribution measurements at the measurement station during summer have shown a mean size distribution around a central value of 100 nm (Kristensson et al., 2008).

We experimented different mean particle diameters, i.e. we performed new simulations of 50 nm and 1 µm particles. However, these new simulations did not result in very different PCA results, at least not for the components which explained most of the variance in the data. PCA tables for particles of 50 nm and 1 µm are found in the supplement.

* Source apportionment
Given the description on page 6 (top paragraph), you are simply multiplying the response function output fields (units of s m3 kg-1) by the fractional land cover - did you correct for grid area and level thickness?
Yes (see previous point for the level thickness correction).

* Principal component analysis (PCA) method
The method description of the PCA is insufficient. Citing a commercial, non-free software package is not an appropriate source of information for the reader. Also: for a PCA to be meaningful, a number of preconditions have to be met, out of which I wonder if two are met: 1) sample size: 38 data points (filter samples) is quite small, can you show that the results are still reasonable? 2) outliers: did you remove them?

We have added more information about the PCA setup and performance in the methods-section.

1. We agree with the reviewer that it would have been desirable to have more samples than the 38 samples we have in this study. More samples would result in a more robust PCA and perhaps results that were easier to interpret. However, we believe that the current interpretation of the PCA output is sound, logical and satisfactory.

   We have not seen any strict recommendations regarding the number of samples vs. number of variables in PCA. By searching the literature we found that this ratio varies significantly between studies. In our study we had 38 samples and 32 variables (22 chemical species and 10 surface categories), that gives a sample-to-variable ratio of 1.19. van Pinxteren et al. (2010) used 29 samples and 60 variables (sample-to-variable ratio of 0.48). Viana et al. (2006) used 41 samples and 28 variables (sample-to-variable ratio of 1.46).

   Further, we have stressed the importance of larger datasets in future atmospheric PCA appliances. This is found in the discussion section.

2. Initially, we did not remove any potential outliers prior to the PCA. This is because we have no reason to distrust the data, even though it contains outliers. Outlier selection and removal is not trivial. Outliers can contain valuable information, in this particular case they can provide information regarding the sources of aerosols. The most obvious example of suspicious outliers in this dataset is the concentration of adipic acid which peaked during the 27th of June and 6th of July. Being aware of the potential disruption in the PCA caused by outliers, we removed the concentration peaks of adipic acid and re-analyzed the data. Hence, the manuscript now contains analysis with and without the adipic acid concentration peaks. A discussion on this matter is given in section 3.3 in the manuscript.

* PCA results
The kind of PCA performed should be described in the methods section, see above.

We have added more information about the PCA setup and performance in the methods-section.

References

Kristensson, A., Dal Maso, M., Swietlicki, E., Hussein, T., Zhou, J., Kerminen, V.-M., Kulmala, M. Characterization of new particle formation events at a background site in Southern Sweden: relation to air mass history. Tellus, 60B, 330-344, 2008.

van Pinxteren, D., Bruggemann, E., Gnauk, T., Muller, K., Thiel, C., and Herrmann, H.: A GIS based approach to back trajectory analysis for the source apportionment of aerosol constituents and its first application, J Atmos Chem, 67, 1-28, 10.1007/s10874-011-9199-9, 2010.

Viana, M., Querol, X., Alastuey, A., Gil, J. I., and Menendez, M.: Identification of PM sources by principal component analysis (PCA) coupled with wind direction data, Chemosphere, 65, 2411-2418, 10.1016/j.chemosphere.2006.04.060, 2006.