Referee comments regarding: Modeling organic aerosol composition at the puy de Dôme mountain (France) for two contrasted air masses with the WRF-Chem model, by C. Barbet et al.

General comments

This (quite lengthy) paper compares model calculated aerosol particle concentrations (especially the organic aerosol) to AMS-measurements at the mountain site Puy de Dôme. The comparisons were made for two very short time periods (ca 4 and 5 days) and the standard model/emission setup underestimated the observations severely; a number of adjustments of model parameters and emissions were tested in order to reproduce the observed OA-concentrations.

In my opinion these model adjustments are not well motivated and the tests are not extensive enough to be useful as a more general sensitivity test of the model. To just tune the model/emissions to get total OA-concentrations in relatively good agreement with a few days of measurements at a single site does not prove anything regarding which model adjustments are useful in general for a regional scale model. There are so many different things that can be tuned in the model type used in this study that it is relatively easy to get "good" results for total OA concentrations (at a single site over short time periods) for completely wrong reasons.

The authors draw the conclusion that "This work shows that the model underestimation seems to be mainly due to a lack in SOA formation processes". I do not agree at all! The work does not show this, it only shows that tuning the parameterisation of these processes (in a rather poorly motivated way) can give very large OA-concentrations. I am convinced that the four-fold increase in SOA-aging rates and doubling of SOA yields (to unrealistic levels of up to several hundred percent SOA-yield) suggested by the authors would lead to massive overestimation of OA-concentrations at many sites and time-periods!

To really test such extreme changes in the model set-up the study must be extended to cover many more sites and longer time-periods. And it is important to not only compare the results to the total OA concentration but also to more detailed data (source-apportionment data, tracers, detailed AMS-data). But even if this was done it would still not be easy to *prove* that the changes in the SOA-formation processes were the "correct" changes.

In summary, I think that no new concepts, ideas, methods or data are presented in the manuscript and the evaluation of the model (including the sensitivity tests) is far too limited to be of significant use for readers of ACP. Because of this no substantial conclusions can be drawn from the work presented in this manuscript.

Specific comments

Page 13401, lines 16-18
What is the thickness of the model layers of interest for the PUY site (the surface layer and the level for which model results are compared to the measurements)?

Page 13403, lines 16-19
There is a discussion about LVOCs, SVOCs and IVOCs but do you really include any IVOCs in your model? Do you treat POA-emissions using a VBS-approach? Or do you consider them as non-volatile (fixed) emissions?

Page 13403, lines 27-29
"Moreover, the organic condensable vapours (OCVs), i.e. first generation VOCs oxidation products that condense on particles, may undergo a chemical aging by oxidizing with the hydroxyl radicals."
Does this mean that you include aging in the particulate phase?

Page 13403-13404 (About the SOA-aging and yields)
If you use the 4-bin VBS yields from Ahmadov et al. (2012) and add (possibly multi-generational) aging by OH-radical reactions (including mass increases by oxygen addition), and no fragmentation reactions, you will produce unrealistic SOA-yields for several of the VOC-precursors (very far from what has been observed in experiments).

For example, for SESQ the initial (mass) yields in the model are:

$$\text{SESQ} + \text{oxidant} \rightarrow 0.075\ C^*_1 + 0.15\ C^*_{10} + 0.75\ C^*_{100} + 0.9\ C^*_{1000}$$
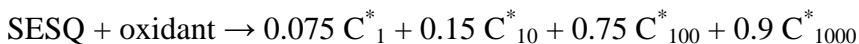
After aging in the atmosphere (which will continue as long as the semi-volatile species reside in the gas phase) you could thus end up with a total particulate mass-yield of more than 180%!

Similarly for the monoterpenes and aromatics the total mass yields in the model after aging can be about 120% (for low-NOx conditions; slightly lower, ca 80-90% for high-NOx-conditions).

These model yields are extremely high and there is no support for such high yields from chamber studies under atmospheric concentration levels.

The VBS-yields used in the WRF-Chem model are not constructed for the type of aging that is applied in the current study but to simulate "first-generation" SOA-products in smog-chamber studies.

If the scheme with these extremely high SOA-yields gives good results for total organic aerosol concentrations it is almost certainly due to compensating for some missing or severely underestimated OA-sources. Biogenic VOC emissions are extremely uncertain and there may, e.g., be large stress-induced emissions from vegetation (of a number of different SOA-forming VOCs) that probably is not included in the BVOC-emission module used. Primary organic aerosol emissions may also be underestimated and especially emissions of semi- and intermediate volatility organic compounds (S/IVOCs) may be missing in the inventory used. After atmospheric aging these can potentially form large amounts of SOA.

Page 13404, lines 10-11
"However, the 1-D-VBS approach still shows a better performance than a functionalization and fragmentation scheme (Murphy et al., 2012)"

This is not correct. Murphy et al. (2012) got better results for OA mass with the detailed 2D-VBS, including both functionalization and fragmentation, than with their base-case simulation (lower RMSE for five of six cases, lower fractional error for four of six cases).

However, it is true that the detailed 2D-VBS was not much better than the simple base case (for OA-mass) and it was definitely much worse for O:C ratios. But it is important to notice that the base-case simulation by Murphy et al. was different than the 1D-VBS aging used in the present study! Murphy et al. did not change the volatility of the biogenic SOA upon aging (they only increased the oxygen content and assumed functionalization and fragmentation in a rough balance).

A number of studies have shown that simple 1D-VBS-schemes that allow multi-generational aging of biogenic SOA (without fragmentation) may lead to substantial overestimation of OA-concentrations (e.g., Lane et al., 2008; Dzepina et al., 2011; Fountoukis et al., 2011) and this is not surprising considering the very high yields that can be achieved if all semi-volatile VBS-components are allowed to age until they are so low-volatile that they end up in the particulate phase.

Page 13404
There is a short description of the dry deposition but nothing about wet deposition. How is the wet deposition handled in the model?

Section 2.3 Emissions
There is no information about how the vertical distribution of the emissions is handled. What emission heights do you assume for different source categories (both anthropogenic sources and wildfires)? Do you include any plume rise calculation or do you use some simplified effective emission heights?

Page 13405, lines 13-14
Regarding NOx-emissions: Do you consider that 100% of the NOx-emissions are in the form of NO? And 0% NO_2, even for diesel emissions?

Page 13407, lines 2-3
A PMF analysis of the organic aerosol mass spectra is mentioned here. Why are the model results not compared to this more detailed description of the OA?

Page 13407 line28 - 13408 line 6
"In this paper, two particular episodes have been selected out of those observed periods of time to illustrate typical extremes that are encountered at the top of the PUY: continental vs. Mediterranean air masses (i.e. polluted vs. remote air masses), for two seasons. Also, the case study period was chosen over days without any changes in the air mass origins so to be able to study the evolution of aerosol chemical composition within the same air mass. This was the case in autumn 2008 (from the 14 to the 18 September 2008) and in summer 2010 (from the 23 to the 26 June 2010)."

Why do you choose <u>only</u> "typical extremes" at the top of the PUY? This makes no sense if you really want to explore possible tuning of the 1D-VBS scheme that could be of somewhat general applicability.

As mentioned in the "General comments" above, the model-measurement comparison needs to be extended substantially for this study to be possibly useful. For PUY you obviously need to include all of the available AMS-data (not only the extreme episodes); the autumn 2008 campaign includes data for 12/9 – 30/9 + 9/10-18/10 and the summer 2010 includes 22/6 – 29/6 2010. But it is even more important to include several other sites than PUY in the evaluation.

Another question regarding the summer episode (2010):
The AMS-measured OM-concentrations are extremely high (for Puy de Dome) for this period - 12.5 µg m^{-3} as average for the whole several-day long episode. This can hardly be considered a "typical" extreme concentration but an extremely high concentration for this mountain top site. Organic carbon (in $PM_{2.5}$) has been measured for a number of years (at least 2008 – 2010) at PUY as part of the EMEP programme (published in various EMEP reports, e.g., http://www.nilu.no/projects/ccc/reports/emep4-2012.pdf and probably also in a number of peer-reviewed papers; and the data have been available on ebas.nilu.no). The yearly average $OC_{2.5}$ concentration has been around 0.9 – 1.0 µg m^{-3} for the years 2008 – 2010 and the highest measured $OC_{2.5}$ concentrations (2-day averages) have been 2.4 – 2.8 µg m^{-3}. Unless there were some serious measurement errors (either with the AMS campaign in 2010 or in all the $OC_{2.5}$-measurements performed at PUY) this means that the summer episode 2010 is the most polluted episode at PUY in many years (from an OA-perspective)! This needs to be discussed in the paper. It seems like a poor idea to tune a model to work only for the worst episode ever recorded at a single site.

Since there have been more or less continuous measurements of $OC_{2.5}$ for several years at PUY and the present manuscript only evaluates the total OA-concentration it would be natural to compare the model results to the measured $OC_{2.5}$-timeseries, in addition to the extremely short AMS-campaign.

Page 13409, lines 25-27
"Consequently, in the rest of the paper, results will be discussed at the real PUY height, where the WRF-Chem model correctly simulates the location of the summit with respect to the boundary layer."
This is difficult to understand (at least for me). What do you mean? Do you use model results from a model level at 1465 m a s l (instead of the model surface layer)? And if so:
- what is the thickness of the model level used?
- how would the effect of dry deposition on the model concentrations (at the measurement station) be influenced by using a model-level far above the model surface?

Page 13410, Section 3.3 (and Supplement Table S2 & S3)
The evaluation of the meteorological parameters is strange. The statistical performance seems to be based on recommendations for air quality models (by Chang and Hanna, 2004) but I would, e.g., not consider a meteorological model that predicts temperatures within a factor of two from the observed 51% of the time as having "good performance". Such poor performance may be acceptable for some air quality parameters (that have uncertain emissions etc.) but for "simple" meteorological variables (T, P, RH) I think the quality recommendations from Chang and Hanna are not useful. I also wonder how you can have 20-30% fractional bias for temperature and consider it good model performance?

The model performance for the met-parameters should be illustrated (in the supplement) by time series plots of T, P and RH and include the actual RMSE, bias, correlation, etc. (with proper accuracy for these parameters).

Also for CO and O_3 the "criteria of acceptance" used in this study are not good enough. Both these compounds are usually better modelled than the acceptance criteria given by Chang and Hanna and, from the results presented in Table S3, I do not think you can "conclude that the WRF-Chem model is able to correctly simulate both the chemical reactivity and the transport". To check this you need to show time series performance for CO and O_3 (and preferably also NOx and SO_2 if such measurements are available, to check if transport and mixing are well modelled); the actual statistics (NMSE, bias, correlation etc.) should also be given.

How did you choose criteria of acceptance for the five different statistical measures? Only two of them seem to be from Chang and Hanna (2004) [fraction of predictions within 50% >0.5 and relative mean bias within 30%; in addition Chang and Hanna suggests that random scatter should be about a factor of two and three].

Page 13411, lines 19-21 (regarding the comparison of modelled and measured BVOC)
"their concentration levels are underestimated by the model in average: by factors 2 and 10 respectively for pinenes and limonene."

This seems inconsistent with the results shown in Figure 2.

For the Pinenes it looks like 5 measurements are underestimated, 8 are overestimated and 9 are within the measurement uncertainties. Only the first 12 hours (3 measurements) seem to be severely underestimated in the model – for the rest the model results look fairly good (but with some zero-measurements substantially overestimated).

For Limonene 17 of 22 measurements gave zero concentration (and at least 6 of these were overestimated by the model) and the factor of 10 underestimation seem to be for the max concentration rather than the average for the whole time period? The description of the comparison of modelled and measured concentrations should show the correct statistics.

Page 13411, lines 22-24 (regarding the comparison of modelled and measured AVOC)
"The time evolutions of the AVOCs (alkanes, xylene and toluene) simulated and observed (Fig. 3) highlight a strongest underestimation of the concentration levels by the model compared to biogenic gaseous precursors of SOA."
This sentence is difficult to understand. What do you mean by "a strongest underestimation of the concentration levels by the model compared to biogenic gaseous precursors of SOA"?

As far as I can see from Figure 3 the correlation between modelled AVOCs and the measured concentrations is very poor. This needs to be discussed, since it may indicate problems with the temporal and/or spatial variation of the model emissions and/or the transport and/or chemistry in the WRF-Chem model. However, the average model bias for the three "model-AVOCs" does not seem to be very large?

Page 13412-13413, Section 3.5 (first part, about inorganic particles) + Figure 4 and Table S4
This section is not very well written. A table showing the concentrations and statistics would help making it easier to read (instead of just giving a large number of concentration values in the text).

Some questions/comments:

From Figure 4 it seems like the correlations between the measured and modelled concentrations are very poor! This clearly indicates problems with the temporal and/or spatial variation of the model emissions and/or transport, chemistry etc. in the WRF-Chem model. It is difficult to expect that the model should be able to model the organic aerosol well unless the problems with the relatively "simple" (compared to OA) inorganic model species are reasonably well understood.

To make it possible to evaluate the model performance the mean measured and observed concentrations, mean absolute error (or RMSE), correlation coefficients (for hourly mean concentrations), and some measure of the variance should be given in a table.

I find it hard to understand that the results for sulfate and ammonium are classified as "good model performance" for the Autumn 2008 period (in Table S4) – from Figure 4 it looks like the model performance is very poor (unless the only thing you are interested in is long-term mean concentrations).

Page 13413,
lines 12-14:
"For the autumn case, the model correctly reproduces the sulphate mass concentration observed at the PUY station"
This seems to be in error – the modelled sulfate concentration seems very poorly correlated with the measured concentration.

lines 14-15:
"ammonium, nitrate and black carbon mass concentrations are quite well captured by the model."
This also seems to be erroneous – The modelled BC concentrations are completely wrong! And nitrate is also extremely poorly modelled. And the modelled ammonium seems to be poorly correlated with the measurements.

lines 19-22:
"The underestimation of the black carbon mass concentration is probably due to the anthropogenic emission inventory used in this study which is too coarse to correctly represent the local emissions of black carbon which is a primary particle, directly emitted into the atmosphere from fossil fuel and biomass combustion."

Do you really have substantial problems with local emissions of black carbon near the PUY measurement site? If you have that problem for BC you will likely have the same problem for OA since most BC-sources also emit large amounts of organic particles and SOA-precursors. To be able to properly evaluate the model performance you should choose model sites that are representative for the resolution of your model (and emission data).

Page 13414, lines 1-4
"POA is the main OA component during both seasons since it contributes respectively 34 and 62% to the total OA."

What does these two percentage figures refer to? The situation at PUY or averages for the whole modelling domain?

If it is for PUY, it is not consistent with the results from the PMF analysis by Freney et al. (2011) that shows only oxidised OA (OOA) and no hydrocarbon like OA (HOA). In any case, the large POA-concentrations indicates that the model needs to take into account the aging of primary OA to form SOA from the S/IVOC emissions (possibly using a scheme similar to the one employed by Shrivastava et al., 2008). Aging of primary S/IVOC emissions may be a dominant source of SOA and the fact that the POA dominates in the model indicates that you have to take this processing of the semi- and

intermediate volatility OC emissions into account to be able to model particulate OA realistically.

Page 13419, lines 14-17
"The effect of dry deposition of gas-phase OCVs on SOA concentrations is largely unconstrained since in many cases the deposition velocity of OCVs is proportional to the one of nitric acid (HNO_3), which is very soluble."
What do you mean by "in many cases" (references are missing for these cases)? Why would the deposition velocity of OCVs be proportional to the one of nitric acid?

Page 13420-13423, Sections 4.2, 4.2.1, 4.2.2
General comment: A lot of effort is spent on trying to reproduce the total OA concentrations for the very short (just a few days) and extreme OA-episode at PUY. Unfortunately the sensitivity tests done here do not really give any new information about the June 2010 episode or any new insights into how to better model OA (compared to several earlier more extensive VBS-model sensitivity studies).

The AVOC and BVOC-emissions are definitely very uncertain and can easily be much more than a factor of two wrong, especially for very short time periods and limited regions. As mentioned above I think that SOA-formation from neglected primary S/IVOC emissions could also be an important part of the missing OA for this episode.

The doubling of the SOA-yields that were tested in this study are, however, not "realistic". Already the original SOA-yields are probably too high (when aging without fragmentation is included) as discussed in the comment about "Page 13403-13404". Doubling the yield could (after aging reactions) lead to >360% SOA-yield for sesquiterpenes and ca 160-240% yield for monoterpenes and aromatics! If this leads to improved total OA results it must be for the wrong reason!

The oxidation (aging) rate for the VOC oxidation products also seem to be used mostly as a tuning parameter in this study (just as in several earlier studies by other groups).

Lane et al. (2008) tested two different aging rates for AVOC- and BVOC-oxidation products: 4×10^{-11} cm^3 molecule^{-1} s^{-1} (based on the rate for aging of S/IVOC species suggested by Robinson et al., 2007), and an order of magnitude slower, 4×10^{-12} cm^3 molecule^{-1} s^{-1}; even with the lower reaction rate the model overestimated observed OA concentrations at rural locations severely.
Murphy and Pandis (2009) used the aging rate 4×10^{-11} cm^3 molecule^{-1} s^{-1} for "POA" (i.e., S/IVOCs) and 1×10^{-11} cm^3 molecule^{-1} s^{-1} for AVOC-oxidation products (based on "OH oxidation of the products of aromatic VOC oxidation") – note that they assumed <u>no</u> aging of BVOC-oxidation products based on "the available laboratory studies and the results of Lane et al.".
Other studies have tested similar setups regarding aging rates, e.g., Fountoukis et al. (2011) used the same aging rates as Murphy and Pandis (2009) as base case; as a sensitivity test aging of BVOC-oxidation products with the same rate (1×10^{-11} cm^3

molecule^{-1} s^{-1}) was also included; including BSOA-aging led to a significant overprediction of the OA levels at all four stations investigated.

Bergström et al. (2012) tested several different aging assumptions 4×10^{-11} cm^3 molecule^{-1} s^{-1} for S/IVOCs in combination with either 4×10^{-12} cm^3 molecule^{-1} s^{-1}, for BSOA and ASOA, or 1×10^{-11} cm^3 molecule^{-1} s^{-1}, for only ASOA, or no aging at all of ASOA and BSOA.

Shrivastava et al. (2011) used WRF-Chem with a VBS-treatment of OA and included only aging of S/IVOCs and no aging of V-SOA since Dzepina et al. (2011) found that including both SOA from S/IVOCs and aging of V-SOA led to large overestimations of OA-concentrations in the MILAGRO campaign.

The aging of VOC-oxidation products are not yet well constrained by laboratory measurements so it is difficult to know what assumptions to use in modelling these processes. The present study uses comparatively high aging rates for ASOA and BSOA in the base case and even higher in the sensitivity test(s). Several other studies have shown that even lower aging rates (especially for BSOA) may lead to overestimation of OA concentrations. This could indicate that the "improvements" seen in the present study for the extreme summer episode at PUY during 2010 when using high aging rates (in combination with extremely large SOA-yields) are due to compensation of other model errors (or large underestimation of some emissions).

Page 13426, lines 9-14
"If the same tests as those for the 2010 situation are applied, one notices that modifying the SOA yields and the rate constant for aging OCVs is enough to retrieve the levels observed in the SOA mass concentration as shown in Fig. 6b. This shows that increases in the oxidation rates are an important parameter to be considered when modeling OA concentrations, regardless of the meteorological conditions."

Yes, the oxidation rates are important (at least during the summer months June and September); the modelled OA concentrations are clearly sensitive to large increases in SOA yields and aging rates for OCVs. But this does not imply that these very high rates and yields are realistic – only that they can compensate for other errors that lead to underestimated OA-concentrations.

The fact that the model sensitivity test that gave "good" results for the "summer" episode does not work well for the "autumn" period shows that it is not possible to draw general conclusions regarding what parameters in the model are "best", based on the present study.

Page 13426, lines 17-19
"In 2010, it is likely that POA could have been emitted in the north eastern part of Europe (as seen in Fig. 5) and that the WRF-Chem model does not simulate correctly their aging during their transport from these source areas."

This is a strange sentence. As far as I understood the manuscript the model does not consider aging of POA (as primary S/IVOC) emissions at all? I think it would be a good idea to include S/IVOC emissions (and SOA-formation from these emissions).

Page 13426 line 25-page 13427 line 1
"The results of the constrained ME-2 analysis provided by Crippa et al. (2014) for OA data from the PUY station show that both HOA and BBOA generally contribute to less than 10% of the total OA. So, emission sources of POA used with the WRF-Chem model are probably underestimated."

Another strange statement; if the results of Crippa et al. show that HOA and BBOA contribute less than 10% these cannot be the main reason for the very large underestimation by the WRF-Chem model? I guess the other 80-90% of the observed OA must be more important?

Page 13427 lines 1-2
"This could also explain the great discrepancy found between AVOCs simulated and measured at the PUY station."

I do not understand how the results of Crippa et al. (2014) can explain the "great discrepancy" for the AVOCs. This needs to be explained better (or removed).

Page 13427 lines 6-7
"A particular emphasis is made on decreasing the observed differences between modeled and measured OA concentrations."

This sentence summarizes the main problem with this work. It seems like the work has focused only on "tuning" the model to minimize the difference between the modelled and observed OA-concentrations without thinking about if the tuning has anything to do with the real atmospheric processes. If such a tuning had been done for long time periods and many sites it could perhaps have given a model that could be "useful" for predicting OA-concentrations but when it is done for only a very short time period at a single site I do not see the usefulness.

Page 13427 lines 23-25
"During the 2010 summer period, AVOCs were much higher than biogenic ones and rather support the importance of an anthropogenic component of OA over biogenic one."

But the atmospheric residence time is much longer for the AVOCs than for the BVOCs! So even if the measured AVOC-concentrations are much higher than the BVOC-concentrations BVOC-emissions could still be responsible for larger amounts of SOA.

Page 13428 lines 1-2
"This work shows that model underestimation seems to be mainly due to a lack in SOA formation processes."
No, it does not "show" this! Just because it is possible to tune the parameters to get reasonably good agreement between modelled and measured total OA-concentrations at one site for a few days does not *show* much at all about the true reasons for the model underestimation – just that the model results are sensitive to a number of parameters.

Technical corrections

Page 13402, lines 20-22
As far as I understand Papiez et al. (2009) did only add gas phase sesquiterpene chemistry (to study influences on ozone etc.) and not SOA production from sesquiterpenes.

Page 13404, lines 21-22
to constrain it. (instead of "to contstraint it.")

Page 13405, line 3
ammonia (instead of ammoniac)

Page 13407, line 5
To assess the black carbon (instead of "To access the black carbon")

Page 13407, line 15
"carbosievesIII" – should this be carbosieves SIII?

13408 line 6 and Page 13454 Figure 4
The time period for the summer period is given as: "from the 23 to the 26 June 2010" but the Figure 4 only shows the period 23/6 – 25/6 (less than 72 hours of data!).
Which exact time period did you really include in the model simulations?
Figure 4 should show the same time period as used for the other results presented in the paper.

Page 13409, line 20
"is not practised" – do you mean: is not practical?

Page 13420, line 24 and page 13421, line 2 and line 13
"the random errors (NMSE, VG)" and "non-systematic errors (NMSE, VG)"

The NMSE and VG are not purely random/non-systematic errors. They measure both systematic bias and random (unsystematic) scatter (see Chang and Hanna, 2004).

Page 13428, line 8
"to constraint model" – should be "to constrain model"

Page 13454, Figure 5
The font size is too small for the concentration scales.
Also, for ease of comparison, use the same scale for POA, ASOA and BSOA.
I also suggest that the episode dates should be specified in the Figure caption.

Supplement, Tables S2 and S3
The tables need better headers. Abbreviations need to be explained and, e.g., what do you mean by a value of 0.0 in the column NMSE<1.5?

References

Dzepina, K., Cappa, C. D., Volkamer, R. M., Madronich, S., De-Carlo, P. F., Zaveri, R. A., and Jimenez, J. L.: Modeling the multiday evolution and aging of secondary organic aerosol during MILAGRO 2006, Environ. Sci. Technol., 45, 3496–3503, 2011.

Fountoukis, C., Racherla, P. N., Denier van der Gon, H. A. C., Polymeneas, P., Charalampidis, P. E., Pilinis, C., Wiedensohler, A., Dall'Osto, M., O'Dowd, C., and Pandis, S. N.: Evaluation of a three-dimensional chemical transport model (PMCAMx) in the European domain during the EUCAARI May 2008 campaign, Atmos. Chem. Phys., 11, 10331–10347, 2011.

Shrivastava, M. K., Lane, T. E., Donahue, N. M., Pandis, S. N., and Robinson, A. L.: Effects of gas particle partitioning and aging of primary emissions on urban and regional organic aerosol concentrations, J. Geophys. Res., 113, D18301, 2008.