We thank Anonymous Referee #2 for helpful suggestions and comments, which helped us improve the manuscript. Our point by point answers to the comments are presented below. Referee comments are in bold and our replies in body text.

Referee #2 comments:

General comments:

The TM5-MP model description lacks the spatio-temporal model resolution used for this study. In the evaluation and conclusions, the time resolution becomes clear. In section 2.4 the simulation period, and later in the text also the horizontal resolution, are mentioned. For the reader it would be easier to get the full information on spatial and temporal resolution, such as the time period for the simulations either right at the beginning in Section 2.1 or in Section 2.4 in a more coherent way.

We will collate the simulation details in Section 2.4 for more clarity.

In Section 2.3 the earlier version of TM5 is described and it is stated that it uses biogenic monoterpene emissions from MEGAN v1 and does not include isoprene emissions for SOA formation. In contrast, the new version calculates SOA from isoprene and monoterpenes, based on MEGAN v2 inventories. In section 2.4 the authors describe the simulations to be compared, one simulation using the older version "OLDSOA" and a second simulation for the same time period using the new formation schemes "NEWSOA". The difference between these two simulation configurations are not only the novel SOA and new particle formation schemes, but also the changed biogenic emissions plus, in the old scheme only monoterpene-derived SOA was included. For disentangling the contributions from the updated biogenic emissions from the contributions by the extended SOA treatment, I would suggest to add sensitivity simulations. I wonder how does the NEWSOA scheme perform compared to OLDSOA while using the same MEGAN v1 emissions? Also, how do NEWSOA and OLDSOA compare when both using MEGAN v2 emissions, even if OLDSOA does not form isoprene-derived SOA? I would not expect the difference in emissions shown in Figure 2 apply in a linear way to SOA formation. In Section 3.1 the last paragraph points into the direction. Additional sensitivity simulations with both emission inventory versions would help to understand what difference is caused by the changed emissions and what difference is related to the novel SOA formation pathways.

This is a good suggestion and we have performed an additional sensitivity simulation similar to the OLDSOA simulation but using surrogate SOA emissions with the MEGAN v2 monoterpene emissions. The main results will be discussed in the paper.

In section 2.3.2 the production of ELVOC and SVOC is described. In the introduction, the authors do mention low volatility compounds (LVOC), but do not explicitly state here why they assume two products of the isoprene and monoterpene reactions (ELVOC and SVOC) and leave out the possible LVOC products. It would be helpful for the reader to know the reason for this decision.

We apply a hybrid method that can consider kinetic condensation to surface area but also consider products that reach equilibrium with gas and particle phases (e.g. Riipinen et al., 2011). We lump the products into two different volatility classes. ELVOC represents the fraction that has such a low volatility it can participate in the growth of the nanometer scale particles (e.g.Tröstl el al. 2016). The other SOA precursors in our mechanisms are assumed to represent those that reach equilibrium with the gas phase within one time step. These are partitioned to the aerosol phase according to mode OA mass. Even though SOA cannot re-evaporate this represents the semi-volatile fraction of the condensable VOC. In essence, the LVOC is lumped into the same fraction as SVOC in our model.

We have extended the text in the introduction:

“...are often separated into lumped species. In the present model we separate low-volatility products into semi-volatile VOC (SVOC) and extremely low-volatility VOC (ELVOC). ” .

Furthermore, we have added following to Sect. 2.3.2:

“In this work we assume that SVOC does not re-evaporate.”

And in Section 2.3.3 describing the partitioning to particles:

“The equilibrium model is assumed to be irreversible, since the yields are determined in the equilibrium state.”

Riipinen, I., Pierce, J. R., Yli-Juuti, T., Nieminen, T., Häkkinen, S., Ehn, M., Junninen, H., Lehtipalo, K., Petäjä, T., Slowik, J., Chang, R., Shantz, N. C., Abbatt, J., Leaitch, W. R., Kerminen, V.-M., Worsnop, D. R., Pandis, S. N., Donahue, N. M., and Kulmala, M.: Organic condensation: a vital link connecting aerosol formation to cloud condensation nuclei (CCN) concentrations, Atmos. Chem. Phys., 11, 3865–3878, https://doi.org/10.5194/acp-11-3865-2011, 2011.

Tröstl, J., Chuang, W., Gordon, H. *et al.* The role of low-volatility organic compounds in initial particle growth in the atmosphere. *Nature* 533, 527–531 (2016). https://doi.org/10.1038/nature18271

Table 2 nicely shows the SOA budget together with some literature values. In this study, the authors differentiate between monoterpene-derived SOA and isoprene-derived SOA, a distinction which cannot be found in the literature cited in Table 2. A study focusing on isoprene derived SOA could add some more information on how to interpret the values for isoprene-derived SOA. Please consider comparing to the global isoprene-derived SOA values in Table 4 given in Stadtler et al. 2018 (https://doi.org/10.5194/gmd-11-3235-2018). I did not find a comparable study for monoterpene- derived SOA on a global scale, but it might also exist.

Just as the reviewer, most of the studies we have found only report the total SOA production without information on the production pathway. However, we will add the data from Stadtler et al. (2018) and a discussion in the paper.

In Section 3.3 the authors start to compare the OLDSOA and NEWSOA simulations to observed particle number concentrations (Figure 8). Although this kind of scatter plot is a common evaluation method, the comparison shown here does not include any error or uncertainty indication, either for the observations, nor for the statistical error caused by comparing grid cells to point measurements. Swall et al. 2009 (https://doi.org/10.1016/j.atmosenv.2008.10.057) describe and discuss the incommensurability problem arising from comparison of point measurements to grid cell values. According to Swall et al. even if a perfectly performing model and measurements without observational error, the scatterplot will not show modeled and observed values on a one-to-one line. The same holds true for Figure 14 and comparison to AERONET stations, Figure 10 in comparison to IMPROVE stations and Figure 11 in comparison to EMEP stations. The authors should consider including information on uncertainty derived from the comparison approach and, if available, the observational errors. A model error would be nice-to-have, but I dare not to ask for that.

We agree that there are several sources of errors as also stated in the studies by Schutgens et al. (2016a, 2016b and 2017). To limit the errors from temporal sampling we have done colocation of the model data with the observations at the hourly level. In Schutgens et al. (2016a) it is estimated that constructing yearly averages from daily sampling will produce errors of 7–17%. However, we use hourly sampling (when possible, e.g. EMEP can have daily or monthly data) which should lead to smaller sampling errors. In Schutgens et al. (2016b) they show errors of at least 30% and up to

80% in simulated black carbon concentrations compared to in-situ observations. Furthermore, in monthly data they note errors to be typically between 10%–40%. They also point out that one should do spatio-temporal collocation which we have done, but still some error will remain. We have done this to limit the errors as much as possible.

We will add error bars into the figures in question using standard errors for the means. Furthermore, we will add a discussion on the error in Sect. 2.5 with a description of the data used in evaluation where we also note that we are doing colocation of observations and model data.

Schutgens, N. A. J., Partridge, D. G., and Stier, P.: The importance of temporal collocation for the evaluation of aerosol models with observations, Atmos. Chem. Phys., 16, 1065–1079, https://doi.org/10.5194/acp-16-1065-2016, 2016 ab.

Schutgens, N. A. J., Gryspeerdt, E., Weigum, N., Tsyro, S., Goto, D., Schulz, M., and Stier, P.: Will a perfect model agree with perfect observations? The impact of spatial sampling, Atmospheric Chemistry and Physics, 16, 6335–6353, https://doi.org/10.5194/acp-16-6335- 2016, https://www.atmos-chem-phys.net/16/6335/2016/, 2016 b.

Schutgens, N., Tsyro, S., Gryspeerdt, E., Goto, D., Weigum, N., Schulz, M., and Stier, P.: On the spatio-temporal representativeness of observations, Atmos. Chem. Phys., 17, 9761–9780, https://doi.org/10.5194/acp-17-9761-2017, 2017.

In whole Section 3 the description of the results and comparison is blended with the discussion of the model errors and limitations. The authors should consider editing the subsections in Section 3 such that the description of the results is separated from the discussion and interpretation of the model strengths and weaknesses. A clear strength of this study is the evaluation with a variety of observations, but it makes the manuscript difficult to read if the interpretation and discussion is mixed into the comparisons to the old model and observations. The authors should consider to collect all the discussion points of NEWSOA schemes' limitations and advantages and to write them into a single "Discussion" Section in a concise way.

Thank you for the suggestion. We will do so.

Out of curiosity, why was it not possible to include isoprene and monoterpene oxidation by NO3 in this study? As you mention, it limits the current model performance, but should be considered in a future study. I wonder what reason was prohibiting to implement this additional oxidation reaction in the NEWSOA scheme.

We started out to reproduce the method used by Jokinen et al. (2015), which only includes yields for reactions with OH and O_3. It is true that it lacks the NO_3 oxidation. We had a discussion to include oxidation by NO_3 but due to time constraints of CMIP6 we were unable to implement this reaction at this time. As noted in the text, the chemistry does include NO_3 which will allow this reaction to be added in the future as we have noted in the text.

Technical corrections:

In the abstract line 11 – 14 it is written "Compared to the old scheme, the new schemeis increasing the number concentrations across the observation stations while still underestimating the observations. The total aerosol mass concentrations in the US show a much better seasonal cycle and removal of a clear overestimation of concentrations.", which is confusing the reader. First the authors speak of "still underestimating" but the next sentence reads "removal of overestimation". Please make it clearer here.

These two things are not exactly the same. The first sentence refers to number concentrations and the latter to mass concentrations. Nevertheless, it is a bit unclear so we have changed the latter sentence to:

"The organic aerosol mass concentrations in the US show a much better seasonal cycle and no clear overestimation of mass concentrations anymore."

In Line 34 there is some number in the word "vegetation" "vegetat589527ion".
Line 431 and 432: Two sentences in a row starting with "However".

We will correct these.

*I**n Figure 13 the plot labels indicate [NEWSOA-OLDSOA], but according to the caption of Figure 13 the annual mean difference of NEWSOA and MODIS (a), NEWSOA and AATSR (b) are shown. Please change the label of the plots.***

This will be corrected. It was an error introduced during the production of the preprint version, where we mixed another figure with the article figure.

Line 504 it says "However, compared to OLDSOA the absolute bias across all stations increases in NEWSOA by 0.007 to a low bias of 0.003", should the verb be "decreases"?

We will correct this as suggested.