Response to reviewer #2

We thank reviewer #2 for reviewing our manuscript and providing valuable feedback on our work. Below we address first the general and then the specific comments. The reviewer's comments are in **blue**, our responses are in **black** and any additions to the manuscript are in **red**

General comment 1

It is noteworthy that the authors do not use numerical values of their measure of model result-data correlation (fitness) for their argumentation. The performance of an optimization algorithm should also be evaluated in terms of how quickly (or how reliably) it reaches a certain threshold of model result-data correlation that justifies consideration of the optimized parameter set (e.g. model result-data residue is comparable to the noise in the experimental data). The authors denote the spread in fitted parameters, but it would also be interesting to see spreads in their measure for correlation, especially in section 3 where two methods are compared.

Response

We considered comparing the model result-data correlation, but decided not to include those because of the methodological differences between the MCGA and Bayesian inference. The main problem we encountered in comparing the two methods was that the Bayesian analysis fundamentally assumes that the estimated parameters have distributions whereas the MCGA algorithm looks for an optimal value which fits the data. This Bayesian vs. frequentist conflict presents challenges when the two methods are compared.

Considering only a situation where one parameter would be optimized, the MCMC algorithm would sample from the posterior distribution of this parameter such that sometimes during the execution also values that fits less well to the experimental data are visited but less often than those parameter values that produce a better fit to the data. For this reason we feel that it would not be fair to compare how the model result-data correlation evolves during the MCMC calculations. This comparison would show that in the MCMC algorithm the model result-data residue sometimes spikes whereas MCGA algorithm consistently moves toward lower residues.

We believe that the best way to compare the algorithms is to think that over multiple optimization rounds the MCGA algorithm also produces a distribution of parameter values which can be compared to the posterior distribution calculated with the MCMC algorithm.

We have clarified our reasoning that led us to only show the parameter estimate distribution in the manuscript by adding the following to Sect. 3 line 276 in the revised version.

The Bayesian inference and the MCMC algorithm assume estimated parameters are random variables with probability distributions, whereas the MCGA algorithm tries to find a single set of input parameters that best fit to the observations. The MCMC algorithm explores the posterior distribution by randomly drawing samples from it. The samples are drawn such that the parameter values that better fit to the data and thus are more probable are drawn more often than the lower probability parameter values that produce worse fits. Because of this fundamental difference between these two approaches, it would not be appropriate to compare how the model output-data residuals evolve during the execution of both algorithms and only the distributions of parameter

estimates from MCMC and MCGA over multiple optimization rounds are used to compare the two methods.

General comment 2

What is the reason to choose MCGA over Bayesian inference for further analysis in this manuscript? Did one method outperform the other in any regard or was it just easier to operate?

Response

As shown in Sect. 3 both methods produce similar distributions for the estimated parameters and there is no obvious evidence that one method is better than the other. We feel that once this comparison is made the choice is up to the person performing the optimization, as rigorous model optimization needs hands-on expertise with the selected method. MCGA was selected for further comparison because we had more experience with Genetic Algorithms as our group has used GA before in Yli-Juuti et al., (2017) and the MCGA method in Buchholz et al., (2019).

General comment 3

I wonder how big the error is that is introduced by combination of the topmost bulk layers in the KM-GAP model during evaporation. This practice in conjunction with discrete layers can introduce step-profiles in evaporation, where evaporation slows down as volatile constituents are depleted from the topmost bulk layer and picks up pace again as soon as layers are merged (and hence volatile component is mixed into the topmost bulk layer). Kinetic models operating with fixed layer sizes and merging schemes have to have mechanisms in place that prevent these artifacts from happening, especially when input parameter optimization is automatized and hence numerical convergence not always manually checked for each combination of input parameters. Which mechanisms are in place in this study to prevent this?

Response

Reviewer #2 raises an important point here. After each KM-GAP simulation we did not check numerically whether the model output contained step-profiles. An efficient method to remove these step-profiles caused by high viscosity in the topmost layer is to increase the number of layers in the model, which decreases the thickness of the topmost layer and leads to layers being merged more frequently, i.e. in practice before all of the volatile component (glycerol in mixtures 3 and 4) is gone from the topmost layer. Based on a random search of the parameter space, large step-profiles are rare in our system but do exist with some combination of the allowed input parameters. However, increasing the number of layers from 30 to e.g. 100 led more often to situations where the diffusion fluxes are too high compared to the amount of molecules in a layer, which in turn resulted in the ODE solver to stop integration as the required time step to solve the molecular transport became too short.

In figure 1 we compare KM-GAP output using the parameter estimates calculated by optimizing to low RH evaporation data of mixtures 3 and 4 with original number of layers (30) and with 100 layers. With both mixtures we see small deviation between the outputs. However, this variation is small compared to the variation in the experimental data.

To check that this numerical problem does not cause a systematic deviation to optimized parameter values, we tested how the best parameter estimates for mixture 3 would change if the candidates in the last generation of an optimization round were rerun with 300 layers (instead of the original 30 layers). The best mole fraction estimates would change by 0.001, C_{sat} estimates would stay the same, $b_{sucrose}$ would change from $10^{12.0}$ to $10^{11.4}$ and $b_{glycerol}$ from $10^{-3.88}$ to $10^{-3.33}$. Based on this analysis, not checking the convergence might have caused the $b_{sucrose}$ estimates to be slightly lower and $b_{glycerol}$ estimates slightly higher. However, the deviations in the estimated b_i parameters are small and do not affect our conclusions.

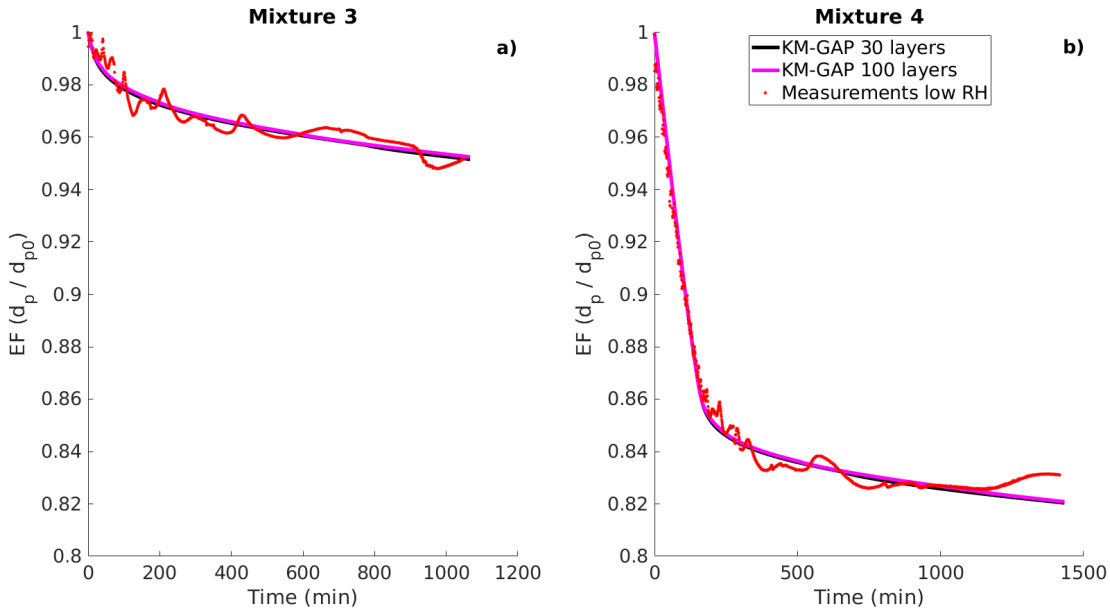

Figure 1: Low RH measurements and KM-GAP output calculated with a set of estimated parameters (C_{sat},x_{mole} and b_i) and with 30 layers (black solid line) or 100 layers (magenta solid line). a) Low RH measurement data and parameter estimates of mixture 3 in the manuscript. b) Low RH measurement data and parameter estimates of mixture 4 in the manuscript

General comment 4

A stylistic suggestion: A concept that is usually used when talking about model parameter optimization is that of a parameter's (local) sensitivity. Talking about conditions under which model output is sensitive to the numerical value of an input parameter could simplify the (sometimes a little slow-moving) discussion in this manuscript considerably. However, usage of this concept is left to judgment of the authors.

Response

We have elected not to change the discussion as we deal with data sets that can be said to have "correct" answers in terms of literature values or input values used in calculating the artificial data sets. We feel that talking about the distribution of estimates we get from independent optimization runs fits better to the context of our work instead of parameter sensitivity.

Specific comment 1

Sect. 1, l. 45 – What do you mean by "increasing attention has also been given to modeling the particle dynamics to better understand the measurements"? Particle dynamics could refer to processes like deposition and coagulation, but probably means evaporation dynamics here, please clarify.

Response

We have modified the text on line 45 to the following

In addition to experimental methods, increasing attention has also been given to modeling the evaporation process to better understand the measurements

Specific comment 2

Sect. 1, l. 50 – The literature review on kinetic parameter determination through inverse modelling seems a little sparse here. Even in aerosol research, there have been more studies detailing such procedures. Examples include Berkemeier et al. (2016), who determined both diffusion coefficients and reaction rates by inverse modelling or Lowe et al. (2016), who used a Monte Carlo Markov Chain (MCMC) algorithm on artificial data as a tool for sensitivity analysis.

Response

We thank Reviewer #2 for pointing out these excellent publications that we hadn't referred to in our manuscript. We have added references to both of these articles starting from line 51 reading:

Berkemeier et al., (2016) derived kinetic parameters influencing ozone uptake of shikimic acid by fitting to multiple measurements of ozone uptake by the acid at different relative humidities. Lowe et al., (2016) studied the sensitivity of various parameters of Köhler theory by studying the goodness-of-fit to artificial cloud condensation nucleus spectra. ~~The same approach was also used~~ Yli-Juuti et al. (2017) ~~to~~ inferred ...

Specific comment 3

Sect. 2.2, l. 132 – It is counterintuitive to talk about a good candidate having low fitness here since in the biological sense a high fitness would be considered better. "Fitness" is thus usually defined as the inverse of least-squares deviation.

Response

We have changed the manuscript so that when we talk about high fitness we mean an inverse of the mean squared error and thus a good fit candidate now has a higher fitness than a candidate with higher mean squared error. We have also modified the explanation starting from the line 133 in the revised version to the following:

Below, the fitness of a candidate is determined as the inverse of its goodness of fit statistic. The goodness of fit in this work is calculated as a mean squared error (MSE) between the evaporation simulation produced with the candidate's parameter set and the measured evaporation. A lower value for goodness of fit statistic, i.e. lower MSE, means higher value for fitness and a better candidate.

Specific comment 4

Sect. 2.2, l. 153 – Can you expand on why the Metropolis criterion ensures variability in the population? While this might be intuitive for experts, it might not be for clear to the general reader of ACP. A couple of follow-up questions on the Metropolis algorithm: Did the MCGA algorithm perform significantly better with this criterion? Do you make sure that a newly created parameter set with improved fitness over all previous ones is not discarded?

Response

The idea behind the Metropolis criterion (together with the mutation scheme) is that sometimes candidates with worse goodness-of-fit are accepted to the next generation. The idea is to keep the population more variable by removing members from the previous generation that would otherwise be included in the next generation and thereby to allow the MCGA algorithm to explore the parameter space more thoroughly.

The Metropolis criterion does not significantly increase the performance of the algorithm when the versions with and without metropolis criterion are tested to artificial data set 1. Figure 2 below shows the distribution of the median mean squared error for generations 2-10 over the 500 optimization rounds performed with the Metropolis criterion (blue lines) and MSE over 50 optimization rounds performed without the Metropolis criterion (red lines). Figure 1 shows that at first the blue distribution is wider but not by much and in the end (generation 10) the version without the Metropolis criterion has more probability mass over smaller and higher median MSE values.

The Metropolis criterion is applied only for candidates that have higher MSE than the maximum MSE in the previous generation, which ensures that a newly created parameter sets with improved fitness are not discarded.

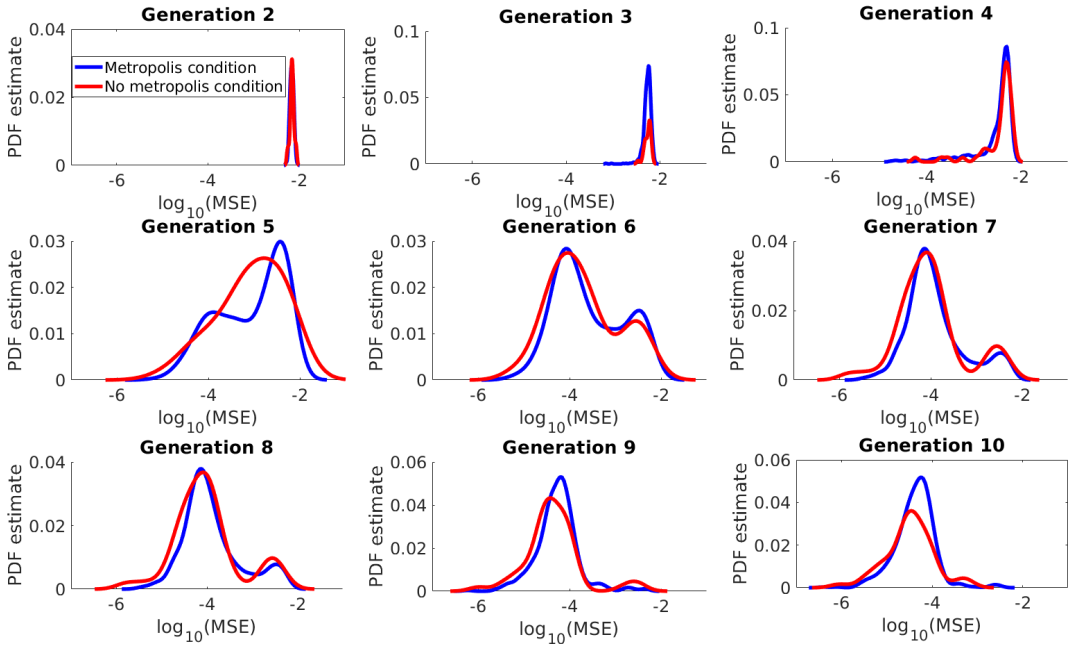

Figure 2: Distribution of median mean squared error (MSE) in a generation for 500 optimization rounds with the Metropolis condition and 50 optimization rounds without the Metorpolis for artificial data set 1 in the manuscript. The distribution is calculated with a kernel density estimation method using Gaussian kernel.

We have added a following explanation to the manuscript Sect 2.2 line 157:

In genetic algorithms the goal is to keep the population as variable as possible while at the same time the goal is that the fitness of the candidates improves when new generations are calculated. If only the candidates that best produce the observations were chosen the algorithm might get stuck in a local minimum and, on the other hand, if new random parameters are drawn too often the genetic algorithm does not converge.

Specific comment 5

Sect. 2.3.1, l. 201 – The comment on subconscious bias made me wonder what choices the person performing the optimization has. In my understanding, there should be none in a fully automated optimization.

Response

Even though the optimization is fully automated there exists a couple of things that the person performing the optimization can do to guide the optimization towards correct results, even though their effect might not be substantial to the optimization. The person performing the optimization can narrow the parameter space so that it is more likely for the algorithm to find correct values. The person might tune lower the mutation rate or increase the number of elite members in a generation if the algorithm is found to converge fast towards the correct values. One could also try different goodness-of-fit functions so that the correct values were found more likely. When analyzing real SOA evaporation experiments the correct volatility distribution is not known, therefore, the above mentioned factors cannot be used to guide the results towards correct values. In practice, the above mentioned factors include decision making by the operator but the operator has to be able to make these decisions based on technical aspects of the method and the optimization problem at hand and not by knowing the correct answers.

Specific comment 6

Sect. 4, l. 294 – Please define the term “optimization rounds”. From context, it seems to be completed MCGA runs. In contrast, how many process model evaluations were performed in each MCGA run? How has this number been chosen?

Response

We have defined the optimization rounds to mean MCGA runs in Sect 2.2.1 line 161.

The number of process model evaluations was not chosen when performing the optimization, instead we chose the number of generations and population size such that further increase in either number did not decrease the mean squared error between model output and data significantly, while at the same time the choice kept the computation time as small as possible.

The number of process model evaluations can be calculated from these two numbers considering that we choose the number of parameters to sample in the MC phase to be the same as in the GA phase and that the process model evaluations in the GA part is $0.95N_{population}\cdot(N_{generation}-1)$, where

$N_{population}$ is the population size and $N_{generation}$ the number of generations. The coefficient 0.95 comes from the fact that 5% of the next generation is always populated with the elite members of the previous generation. This results in 6840 process model evaluations with $N_{population} = 400$ (corresponding to mixtures 3 and 4 at low RH in our study) or to .10260 evaluations with $N_{population} = 600$ (corresponding to all other data sets in our study).

We have added Table S1 to the supplement which shows the population size, number of elite members and the number of parameter sets in the MC part for each data set. This table is referred to in the manuscript in Sect 2.2.1 lines 167-168. The old table S1 showing the composition of the particle at the start of the evaporation for artificial data set 4 is now labelled as Table S2 in the revised version of the manuscript

All the MCGA parameters for each data set in this work are listed in the supplementary material Table S1.

Specific comment 7

Sect. 5.3 – You give two potential reasons for the process model failing at describing the experimental data using only literature values for viscosity and volatility. Can you add a brief discussion on what you think contributes more to the observed discrepancy, the depressed viscosity or hindered evaporation due to Raoult's law?

Response

In Sect 5.3. we speculate with the possibility that since the AIOMFAC model gives lower than unity activity coefficients for the initial particle composition, there might be more water in the particle, which would hinder the rate of evaporation of other compounds and lower viscosity. It is noteworthy that these changes would affect the rate of evaporation opposite ways. The decreased viscosity would increase the rate of evaporation, whereas solution effect would decrease the rate of evaporation of an organic compound.

We have performed the AIOMFAC calculations only with one composition (the initial particle composition) and only to consider with the possibility of more water being present at low RH measurements of mixtures 3 and 4 and at high RH measurement of mixture 3. Discussion about the magnitudes of these effects would need detailed calculations of the activity coefficients as a function of composition which is out of the scope of this study.

Specific comment 8

Sect. 6, l. 600 – "(. . .) the few shortcomings of the method could be largely attributed to the fact that the method can only characterize properties that influence the quantity that is measured". As stated, this is too trivial, please rephrase. You may want to refer to it as "model parameters that have sufficient sensitivity in the probed time and concentration range". For reference, some general concepts of model parameter optimization (at least for application in aerosol research) have been discussed in section 3 in Berkemeier et al. (2017) or can be found in the discussion of sensitivity analysis and kinetic regimes in Berkemeier et al. (2013), section 6.

Response

We have modified the text starting at line 622 in the revised version of the manuscript to read:

For some of the tested data sets, the few shortcomings of the method could be largely attributed to the fact that the model output was not sensitive to the changes in the estimated parameters with respect to the experimental timescale and parameter range.

Specific comment 9

Fig. S9 – Please add literature values from Table 5 into the plot since you are referring to this comparison in Sect. 5.3.

Response

We have added the literature values to Fig. S9.

References

Berkemeier, T., S. Steimer, S., K. Krieger, U., Peter, T., Pöschl, U., Ammann, M. and Shiraiwa, M.: Ozone uptake on glassy, semi-solid and liquid organic matter and the role of reactive oxygen intermediates in atmospheric aerosol chemistry, Phys. Chem. Chem. Phys., 18(18), 12662–12674, doi:10.1039/C6CP00634E, 2016.

Buchholz, A., Lambe, A. T., Ylisirniö, A., Li, Z., Tikkanen, O.-P., Faiola, C., Kari, E., Hao, L., Luoma, O., Huang, W., Mohr, C., Worsnop, D. R., Nizkorodov, S. A., Yli-Juuti, T., Schobesberger, S. and Virtanen, A.: Insights into the O;C-dependent mechanisms controlling the evaporation of α-pinene secondary organic aerosol particles, Atmospheric Chem. Phys., 19(6), 4061–4073, doi:10.5194/acp-19-4061-2019, 2019.

Lowe, S., Partridge, D. G., Topping, D. and Stier, P.: Inverse modelling of Köhler theory – Part 1: A response surface analysis of CCN spectra with respect to surface-active organic species, Atmospheric Chem. Phys., 16(17), 10941–10963, doi:https://doi.org/10.5194/acp-16-10941-2016, 2016.

Yli-Juuti, T., Pajunoja, A., Tikkanen, O.-P., Buchholz, A., Faiola, C., Väisänen, O., Hao, L., Kari, E., Peräkylä, O., Garmash, O., Shiraiwa, M., Ehn, M., Lehtinen, K. and Virtanen, A.: Factors controlling the evaporation of secondary organic aerosol from α-pinene ozonolysis, Geophys. Res. Lett., 44(5), 2016GL072364, doi:10.1002/2016GL072364, 2017.