Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2019-1036-AC1, 2020

Atmospheric Chemistry and Physics Discussions
Open Access

Interactive comment on "Identification of molecular cluster evaporation rates, cluster formation enthalpies and entropies by Monte Carlo method" *by* Anna Shcherbacheva et al.

Anna Shcherbacheva et al.

anna.vlad.shcher@gmail.com

Received and published: 27 June 2020

[12pt,oneside,a4paper]article

authblk algpseudocode hyperref

amssymb,amsmath,amsthm,mathrsfs

color [normalem]ulem

Response to discussion-stage referee comments for the paper "Identification of molecular cluster evaporation rates, enthalpies and entropies by Monte Carlo method"

June 27, 2020

1 Overview

In this document we respond to the referee comments for the paper “Identification of molecular cluster evaporation rates, enthalpies and entropies by Monte Carlo method”. These comments were provided at the public discussion stage of the review process for publication in Atmospheric Chemistry and Physics.

In Section 2 we list each of Referee’s comments. We also include our comment-by-comment responses. Each of the referee’s comments are denoted with “**C**” and our responses to the referee’s comments are denoted with “**R**”.

We thank the referee for his/her time, thoughtfulness, and feedback. All the remarks and suggestions for our paper have been very helpful.

2 Referee 2 comments and our responses

Referee 2's summary: The author proposes to use the Markov chain Monte Carlo (MCMC) algorithm to solve the problem of cluster evaporation rate based on cluster distribution, and this is a novel idea for us to evaluate the thermal stability of clusters. But I have a question about the cluster distribution. The author uses ACDC to simulate the cluster distribution (from 1SA.1NH3 to 5SA.5NH3 box) instead of experimental data. Is this simulation result good enough to replace the experimental data? Simulation results are affected by accurate structure, calculation method and basis set. So I suggest that first the author expand the SA.NH3 system to a larger size (1.7 nm). Before using MCMC, simulate the SA.NH3 formation rate and compare it with the experiment data (Nature 502, 359-363, 2013) to illustrate the reliability of the simulation cluster distribution.

R: The answer to reviewer's summary:

1. "The objective of the present study is to investigate if we can extract evaporation rates from the type of data generated by experiments. Here we search to identify the combination of estimated parameters and experimental data which enables to obtain the estimates for evaporation rates with fair accuracy (i.e., the estimates with the variances comprising less then one order of magnitude).

   In Besel et all, 2020 (J. Phys. Chem. A.) is was shown that the 5x5 simulation box (which is used for generation of the synthetic data is the present study) produces results in a good agreement with the measurements obtained from the CLOUD chamber experiment. However, the quality of data is not a major issue for our parameter estimation procedure, since the main point is not here to reproduce CLOUD data with the quantum chemical calculations, but to find the settings which will give fair estimates of the evaporation rates in case if the data are available.

The MCMC results are not specific for the simulation box considered in the present study, but rather general. This is supported by the fact that although the size of the system (the number of clusters included into simulations) has impact on the particle formation rates at high temperatures ($>$ 278 K), the particle formation rates and cluster concentrations produced using different simulation boxes are qualitatively similar. Thus the changes of the ACDC outputs due to the difference in the simulation box does not change for MCMC parameter estimation results.

The experimental data can differ from the synthetic data in the sense that they contain noise which originate from measurement instruments and uncertainies associated with experimental conditions (e.g., in CLOUD chamber experiments). Treating the noise inherent for experimental data will be the topic of our future studies. "

2. **C:** "time-independent steady-state" in abstract could be revised to be "steady-state"

   R: We have made this change of wording.

3. **C:** The motivation and test results about the case of single temperature steady-state cluster distributions should be mentioned in the abstract;

   R: At the end of line 12, we have added:

   “We also estimated the evaporation rates using synthetic steady-state cluster concentration data at one temperature (which has appeared in previous literature) and compared our two study cases to this setting. Both the transient concentration data and two-temperature steady-state concentration data estimated the evaporation rates with less variance than the steady-state one temperature case. ”

4. **C:** The best result in this study is the case for steady-state concentration with two temperatures. Is this conclusion general or very specific? How sensitive towards the number of ammonia concentrations and the box size (referring to the cluster types here) is this conclusion?

   R: The MCMC results are not specific for the simulation box considered in the present study, but rather general. This is supported by the fact that although the size of the system (the number of clusters included into simulations) has impact on the particle formation rates at high temperatures (> 278 K), the particle formation rates and cluster concentrations produced using different simulation boxes are qualitatively similar. Thus the changes of the ACDC outputs due to the difference in the simulation box does not change for MCMC parameter estimation results. In Besel et all, 2020 (J. Phys. Chem. A.) is was shown that the 5x5 simulation box (which is used for generation of the synthetic data) produces reasonable results with a good agreement with the measurements obtained from the CLOUD chamber experiment. Additionally, the boundary conditions for the outgrowing clusters (the choice of the clusters that are considered as formed particles) has only minor influence on the simulation results, given that the simulated system of clusters is defined in a reasonable way (see Besel at al., 2020, J. Phys. Chem. A).

   In general, the accuracy of the MCMC results increases when we include additional data. In particular, including more concentration data measured at different ammonia concentrations will yield better estimates for the evaporation rates. The sensitivity of the estimates to the number of ammonia concentrations will be considered in the future work. In the present study we rather focus on the question which combination of estimated parameters and concentration data will produce an accurate estimates for the evaporation rate.

The data of steady-state concentration with two temperatures allowed us to apply two general principles of inverse problems/Bayesian estimation to the problem of estimating evaporation rates. First, the two temperature data set enabled us to reformulate the problem in a numerically effective way (in terms of enthalpy and entropy) that reduced the number of unknown parameters we sought to estimate. Second, the reformulated differential equation describing the time evolution of the concentrations was more numerically stable than the original expression (the stiffness of the equation was reduced in the reformulated form). This made our estimates for the rates less sensitive to small perturbations/errors.

However, the reformulation we used was to parametrize the evaporation rates in terms of enthalpy and entropy. The fact that the entropies and enthalpies were strongly correlated made them an effective parametrization. The strong inverse correlations have a physical explanation. Firstly, both enthalpy and entropy follow from the partition function of the molecular complex, and their functional forms are partly similar. Practically, if a cluster has really strong bonds between the molecules, then that means the formation enthalpy is very negative, and also the intermolecular vibrational frequencies corresponding in a broad sense to vibrations involving those bonds (note that these frequencies dominate the "variable part" of the formation entropy, as the entropy effect from the loss of translational and rotational degrees of freedom is almost a constant factor) are fairly high, meaning that the entropy loss in forming the cluster is large. So if the formation enthalpy is very negative so is also the formation entropy. Conversely, if the cluster is only quite weakly bound, the formation enthalpy is only slightly negative, and the intermolecular frequencies can be very low, leading to a less negative (though still negative of course) formation entropy.

In line 343 we add the Section 3.5."Discussion and future work", where we place

Printer-friendly version

Discussion paper

the above-written answer to the reviewer's question.

5. **C:** VODE mentioned in L107 may be different from the solver used in McGrath et al. (2012) (ode15s). If so, "A detailed description of this program was published in McGrath et al. (2012)." should be deleted and a simple benchmark should be made to compare different solvers.

   R: We compared the ode15s with those for the vode when creating synthetic data, and they were producing practically identical results.

6. **C:** For table 3, why the minimal values of H and S are set to be -400?

   R:

   (a) A narrower range could have been used for the formation enthalpies, since the upper limit correspond to evaporation which in practice almost always happens before growth. The lower limit formally corresponds to zero evaporation. Physically, an upper limit of 0 can be justified by the fact that > 0 formation enthalpies would mean no attractive interactions at all, which is obviously physically wrong for polar, H-bonding molecules such as H_2SO_4 and NH_3. For the lower limit (-400) we mean that on average each H_2SO_4 cluster is bound more strongly than in the (extremely strongly bound) $HSO_4^- * H_2SO_4$ cluster, for which the best available computational studies indicate a binding enthalpy roughly around -40 kcal/mol. So it seems unlikely that the average binding per H2SO4 could be tens of kcal/mol stronger than that in the larger clusters where the effect of charge should be much smaller. In any case, a formation enthalpy below -400 kcal/mol means practically zero evaporation so it makes no difference if this is set to a lower value. On the other hand, the largest cluster included into the system has 5 H_2SO_4 and 5 NH_3, so 10 molecules, and -400 kcal/mol would mean -40 kcal/mol per molecule, which 1) corresponds to the strongest known cluster in the system and 2) means evaporation of practically zero.

Printer-friendly version

Discussion paper

(b) For the formation entropies, the 0 cal/Kmol upper limit can be justified as follows: clustering has to have a negative ΔH, as we are reducing the number of gas molecules (and converting translational and rotational degrees of freedom into much more constrained vibrational degrees of freedom). Probably a much lower upper limit could have been used, but certainly the ΔS values can never be $>$ 0. For the lower limit, we state that the typical per-molecule ΔS for clustering is around -30 cal/Kmol, with a typical variation of up to +-10 cal/mol K, see Kürten, 2019. So for the largest clusters the upper limit corresponds to a per-molecule ΔS of -40 cal/Kmol. In this case, all the new vibrational degrees of freedom formed in the product clusters are quite rigid, i.e. have very low entropy.

(c) After the line 153 we edit an explanation on the sampling limits selected for the thermodynamic parameters: "Next we justify the limits selected for data setting 2, where we sample thermodynamic parameters. For the formation enthalpies an upper limit of 0 kcal/mol is chosen by the fact that a positive ΔH would mean an absence of attractive interactions in the molecular cluster, which is physically incorrect for polar, H-bonding molecules such as H_2SO_4 and NH_3. For the lower limit (-400 kcal/mol) we mean that on average each H_2SO_4 is bound substantially stronger than in the $HSO_4^- * H_2SO_4$ cluster, for which the most recent computational studies indicate a binding enthalpy roughly around -40 kcal/mol. Another motivation for the prior distribution selected for the cluster formation enthalpies comes from the fact that the largest cluster included into the system has 5 H_2SO_4 and 5 NH_3, so 10 molecules, and -400 kcal/mol would give an enthalpy of -40 kcal/mol per molecule, which 1) corresponds to the strongest known cluster in the system and 2) which implies that the evaporation rate is zero for all purposes of measurement.

Next, we set the upper limit for the formation entropies to 0 cal/K/mol, since molecule clustering must have a negative ΔH,as the number of gas

Printer-friendly version

Discussion paper

molecules is reduced (and translational and rotational degrees of freedom are converted into much more constrained vibrational degrees of freedom). For the lower limit of -400 cal/K/mol, we state that the typical per-molecule ΔS for clustering is around -30 cal/K/mol, with a typical variation of up to +-10 cal/mol K, see Kürten, 2019. So for the largest clusters the upper limit corresponds to a per-molecule ΔS of -40 cal/Kmol. In this situation,all the new vibrational degrees of freedom formed in the product clusters are quite rigid, i.e. have very low entropy."

7. **C:** L156, “ACDC plus VODE” should be revised to be “ACDC based on VODE”

   R: We have rewritten this paragraph for clarity, and this emphasis for ACDC has been redirected to Section 2.1. The new paragraph which includes the old line 156 is as follows:

   "We make our initial guess $\boldsymbol{\theta} = \boldsymbol{\theta}_{old}$, where $\boldsymbol{\theta}_{old}$ is the flat distribution which obeys the estimates in Table 4. We also assume that the conditional probability distributions for the parameters given the concentration data are of Gaussian type. Once initialized, the following iterative steps take place. From the likelihood probability distribution for $\boldsymbol{\theta}_{old}$, a new candidate for the unknown parameter values, $\boldsymbol{\theta}_{new}$, is sampled using the proposed Gaussian likelihood distribution. We then use the algorithm in Section 2.1 to obtain concentration outputs from the evaporation rates $\boldsymbol{\theta}_{new}$. In the first stage of DRAM, we chose to accept the new proposed values $\boldsymbol{\theta}_{new}$ with probability ... "

8. **C:** L233, “upper limit” needs to be explained further.

   R: We have edited the sentence to read “... all the parameter chains for the evaporation rates have values bounded above by an upper limit which differs for different evaporation rates.’

9. **C:** L244, “well-defined” need to be defined.

Printer-friendly version

Discussion paper

R: We have rewritten the sentence to state:

“All the evaporation rates larger than $10^{-3}s^{-1}$ are well-identified (see subfigures labelled 1, 2, 4, 5, 7, 10, 12, 16, 18, 22, 27, 31 and 35 in Figures 3- 4), in the sense that the variances for these cluster types are well within our accepted error range of less then one order of magnitude.”

Printer-friendly version

Discussion paper