

I would like to thank Dr. Nadykto for his comments, which helped me improve the manuscript. Below are my point-by-point answers (normal font) to the comments (bold font) as well as the additions (yellow highlight) made to the manuscript (italic font). The line numbers refer to the revised manuscript.

The manuscript presents a series of MCMC simulations aimed at determining cluster evaporation rates from concentration measurements. The topic of the paper is interesting and important . The paper’s well-written and easy to follow. After a thorough validation, the proposed approach could possibly be developed into a useful theoretical tool linking cluster concentrations and evaporation rates. However, I have to recommend major revisions because the number of issues to be addressed before the paper can be further considered for publication is quite large and some of them are serious..

Comments

I. Introduction is a way too self-referential, dedicated almost exclusively to own work and fails to acknowledge important contributions made by others. It also contains some misleading statements that need correction.

There were 15 references in the Introduction, and only two of them were papers where I am a co-author. I have trouble seeing this as “way too self-referential”. The new count after the revisions I have made is 19 references including 4 where I am a co-author, which still seems quite reasonable.

1.1 The clusters considered in the paper are relevant directly to the Ion -Mediated Nucleation (IMN), which is an important source of new particles in the Earth’s atmosphere (see e.g. Geophys. Res. Lett., 27, 883-886, 2000; J. Geophys. Res., 106, 4797-4814, 2001; Atmos. Chem. Phys., 8, 2537-2554, 2008; Atmos. Chem. Phys., 12, 11451-11463, 2012). A brief discussion on these matters accompanied by the corresponding references should be included in the Introduction to the revised manuscript.

A mention of ions and ionic clusters was indeed missing. However, as Yu and Turco (2000) were neither the first to suggest ion-induced cluster formation nor the first to demonstrate it experimentally, I decided to cite the CLOUD experiments instead.

Page 1, lines 15–18:

“The experiments of Kirkby et al. (2011); Almeida et al. (2013) have also shown that the first steps of cluster formation can proceed along an ionic pathway, and that this process can dominate over the electrically neutral pathway when there are not enough base molecules or other impurities available to stabilize the small neutral sulfuric acid clusters.”

1.2. The discussion on quantum-chemical studies on charged sulfuric acid-ammonia and sulfuric acid-ammonia-water clusters is limited to Almeida et al., 2013; Olenius et al., 2013b and fails to acknowledge a number of relevant contributions made by others (e.g. JPC A 116(24) 5886-5899, 2011; J. Phys. Chem., A, 117, 133-152, 2013; Atmos. Chem. Phys., 9, 4031-4038, 2009; PCCP, 10, 7073 - 7078, 2008). References to the aforementioned and other relevant studies should be included in the revised manuscript.

The discussion about which is the best quantum chemistry method for atmospheric clusters has been going on more than long enough (Nadykto et al., Entropy 2011, 13, 554–569; Kurtén, Entropy 2011, 13, 915–923; Nadykto et al., Nadykto et al., Chem. Phys. Lett. 2014, 609, 42–49; Kupiainen-Määttä et al., Chem. Phys. Lett. 2015, 624, 107–110), and I see no reason to continue it. As the cluster energies cannot be measured directly, there is no way to find out which method gives the

best predictions, or whether there even is any method that could be trusted. The whole point of this paper is to find a new way to obtain information on cluster properties, so that we no longer need to rely on quantum chemistry calculations at all. I have now tried to explain this more clearly in the Introduction.

Page 2, lines 11–15:

“As evaporation rates depend exponentially on the cluster formation energies, theoretical evaporation rates may easily be wrong by several orders of magnitude. Different quantum chemistry methods can give qualitatively very different predictions for cluster concentrations (Kupiainen-Määttä et al., 2013; Kupiainen-Määttä et al., 2015), and it is not clear whether any of the methods can be trusted. Also the treatment of the collision rates is highly simplified, but errors of more than a factor of two or perhaps ten are unlikely.”

The quantum chemistry data is used only to provide a test case for the MCMC data analysis method. I could just as well have used some other quantum chemistry data set or simply random numbers, but it seemed more sensible to use cluster energies that reproduce the measured cluster distributions qualitatively, if not quantitatively. However, the test data is not claimed to mimic perfectly the true cluster concentrations.

1.3. MC has been widely used in nucleation and cluster formation research since 2000s. In particular, a well-known MC-based DNT (Dynamic Nucleation Theory) has been developed by Kathmann and Garrett with co-workers at the PNNL (e. g. PRL82(17):3484-3487, 1999. JPC B 105(47):11719-11728, 2001, J.Chem. Phys. 120(19):9133-914, 2004; . It would be useful to include a brief discussion on earlier applications of MC to nucleation and cluster formation in the revised manuscript.

In DNT, Monte Carlo has been used for computing an integral. This is different from using Monte Carlo for parameter estimation as is done in the present paper.

1.4. The statement that "At the same time, modeling of particle formation has also advanced greatly in the past few years. For the first time theoretical predictions of cluster distributions (Olenius et al., 2013b) and particle formation rates (Almeida et al., 2013) agree qualitatively with experimental findings." is partly misleading because predictions of particle formation rates in Almeida et al., 2013 clearly disagree with the experimental data (Chem. Phys. Lett., 624, 111-118, 2015). The statement should be corrected.

The results are stated to agree qualitatively. This does not imply that they would match perfectly.

1.5. The author states that "This approach has been shown to give qualitative agreement with experiments (Almeida et al., 2013; Olenius et al., 2013b), but several very drastic assumptions are involved. First-principles molecular dynamics simulations (Loukonen et al., 2014a, b) have shown that one harmonically oscillating cluster structure is far from a realistic description of the thermal motion of molecules in a cluster, implying that the traditional way of computing cluster formation free energies may be a rough approximation". However, this statement is obviously misleading because conclusions obtained using lower level theory such as ab initio MD Loukonen et al., 2014a, b are not applicable to results obtained using higher level theory such as ab initio or DFT. Unharmonic corrections for DFT level with typical scaling factors of 0.95-0.99 are very low and cannot significantly impact cluster formation rates. Also , the impacts of local minima on resulting thermochemical properties can be easily calculated using the Gibbs-Boltzmann distribution. This statement should be either modified or deleted.

The simulations of Loukonen et al. used DFT to compute the energies and forces, so it is unclear

why the conclusions “are not applicable to results obtained using higher level theory such as [...] DFT”. In any case, the problem is not so much the anharmonicity of vibrations within or even between the molecules, but the observation that the molecules rotate inside the cluster breaking bonds and forming new ones. This has been clarified in the revised paper.

Page 2, lines 8–11:

“First-principles molecular dynamics simulations (Loukonen et al., 2014a, b) have shown that one harmonically oscillating cluster structure is far from a realistic description of the thermal motion of molecules in a cluster, as molecules may rotate inside the cluster continuously breaking intermolecular bonds and forming new ones. This implies that the traditional way of computing cluster formation free energies may be a rough approximation.”

II. The source of thermochemical data used for computing evaporation rates in Table 1 is unclear. The MC fitting data were compared to Ortega et al. (2014) only. The author states that “Also evaporation rates estimated from quantum chemical Gibbs free energies Ortega et al. (2014) are presented in Table 1 for comparison”. However, I wasn’t able to find any data on Gibbs free energies in Ortega et al. (2014). Neither delta H nor delta S values were found in there. Delta G values seem to be missing in Ortega et al. (2014), too. Please, clarify the source of the data and include computations of evaporation rates based on quantum data obtained by others in Table 1 of your paper.

The evaporation rates were discussed by Ortega et al., while the cluster energies were published already by Almeida et al. (2013). This has been corrected in the revised manuscript.

Page 15, lines 15–16:

“Also evaporation rates estimated from quantum chemical Gibbs free energies (Ortega et al., 2014; Almeida et al., 2013) are presented in Table 1 for comparison.”

Unfortunately there don't seem to exist any other data sets with quantum chemical Gibbs free energies for all the relevant clusters, which could be included in the Table. Herb et al. (2013) have done calculations on some of the smaller clusters, and these results were used as a comparison already in the submitted version of the manuscript.

Page 18, lines 6–9:

“The stability of the dimer and trimer and the instability of the tetramer are consistent with cluster formation energies calculated with different quantum chemical methods (Ortega et al., 2014; Herb et al., 2013) and with semi-empirical estimates combining measurements and quantum chemistry (Lovejoy and Curtius, 2001; Curtius et al., 2001).”

III. Temperature dependency of evaporation rates is very important; however, the analysis of the temperature-dependent evaporation rates is missing. I would suggest the author to perform a study of evaporation rates for a few clusters at the room temperature and T=273.15 K and compare MC fitted evaporation rates with those obtained using quantum methods in Ortega et al. (2014) and other related studies.

Studying the temperature dependence of the evaporation rates would certainly be interesting, but it cannot be done using MCMC analysis without experimental input data. Olenius et al. (2013) do present some data also at 292 K and 248 K, but the data sets are unfortunately too small.

IV. It is well-known that uncertainties in measured cluster concentrations may be pretty big due to impurities, charging and other issues. The influence of the experimental uncertainties on MC fitted evaporation rates and fragmentation in the mass spectrometer should be

discussed in some detail.

It is unclear how impurities and charging would cause uncertainties in these measurements: the clusters are detected with a high-resolution mass spectrometer, so any impurities in them could not go unnoticed, and the clusters are ionic to begin with, so they do not need to be charged before detection.

I would like to thank Referee #2 for his/her thorough reading of the paper and very useful and constructive comments. Below are my point-by-point answers (normal font) to the comments (bold font) as well as the additions (yellow highlight) made to the manuscript (italic font). The line numbers refer to the revised manuscript.

Kupiainen-Määttä presents a Markov chain Monte Carlo (MCMC) study to derive sets of evaporation rates from observed cluster distributions of negatively charged sulfuric acid ammonia clusters. The simulations are expanded by also treating the fragmentation rates of the clusters in the mass spectrometer as unknown parameters that are varied with MCMC as well.

The paper is generally well written. It presents a useful modelling exercise to gain insight into cluster evaporation rates that are difficult to access. The MCMC is especially useful to realize that several different sets of fitting parameters are well suited to describe a set of experimental cluster measurements, and finding one well-fitting solution does not necessarily mean that this is the correct set of parameters. Exploring MCMC for this type of data is valuable. For larger data sets, covering larger ranges of conditions, hopefully in the future more and more firm conclusions can be drawn from this type of analysis. The paper is publishable in ACP after addressing the following comments:

1) p1 l12: The Sipilä et al. 2010 paper is not a good reference for this statement because it claimed that the H₂SO₄/H₂O system alone would be sufficient to explain the nucleation rates as observed in the BL.

I would expect that undetected impurities in the laboratory air affected the particle formation rate, but perhaps it is indeed better to leave this reference out. The sentence was modified accordingly.

Page 1, lines 11–15:

*“Recent laboratory experiments (Berndt et al., 2010; **Sipilä et al., 2010;** Benson et al., 2011; Almeida et al., 2013) have confirmed that particle formation rates of the magnitude observed in the atmosphere can be produced with ambient sulfuric acid concentrations and **either impurities present in laboratory air or intentionally added** low concentrations of base molecules, giving new support for sulfuric acid being at least one of the compounds driving atmospheric particle formation.”*

2) p1 l17: The high res ToF mass spectrometers certainly allowed a lot of advances for characterizing the clusters during nucleation, but also earlier MS studies such as described by Hanson and Eisele, JGR, 2000 and 2002, already allowed to study the first steps of cluster formation for the sulfuric acid/water and sulfuric acid/ammonia systems.

The sentence has been modified and new references have been added.

Page 1, lines 19–21:

*“**The development of highly sensitive** mass spectrometers has enabled the detection and characterization of individual ionic clusters consisting of only a few molecules (**Eisele and Hanson, 2000; Zhao et al., 2010;** Junninen et al., 2010), opening a new window into the first steps of cluster formation.”*

3) p2 l26 and line 30/31: Besides Olenius et al., 2013b, also other references for the CLOUD data should be included: At least Kirkby et al., Nature, 2011, Schobesberger et al., ACP, 2015, and Duplissy et al., JGR, 2016, should be cited here as well. These papers are from the experimental groups and describe the experimental set-up and the experimental data in much

more detail. Referencing only Olenius et al. does not give credit to the many other groups that contributed in order to set up and perform the CLOUD experiments and to obtain the experimental data that are used here (note, for example, that only authors from U Helsinki are part of Olenius et al. but many more groups were involved running the experiments and obtaining the H₂SO₄ and NH₃ concentrations that are used here).

A reference to Kirkby et al. (2011) was added, as that paper describes the experiments from which the data used of Olenius et al. was obtained. The later CLOUD publications are, to my knowledge, related to other later campaigns.

Page 2, lines 32–34:

“The method is applied to measurement data from the CLOUD experiment (Olenius et al., 2013b; see Kirkby et al., 2011 for more details on the CLOUD experiment).”

4) p5 l110: The assumption of a size-independent wall loss coefficient is problematic. The diffusion coefficient is strongly size dependent, and a cluster consisting of 5 sulfuric acid molecules will diffuse much slower to the walls than the monomer or dimer. This needs to be mentioned, and it should be discussed in how far it may influence the results.

The size dependence was already mentioned, but I have now explained in more detail that this seemed to be the least bad option.

Page 5, lines 16–22:

“A wall loss rate of $1.7 \times 10^{-3} \text{ s}^{-1}$ was determined for the electrically neutral H₂SO₄ monomer in the CLOUD chamber (Almeida et al., 2013). This rate decreases with increasing cluster size, but ions may have a higher loss rate. The probability of an individual cluster being lost on a wall also varies with location inside the chamber, or in practice with time as the air is continuously circulated around the chamber by large fans. As the size, charge and composition dependence of the wall losses is not known, all clusters were, for simplicity, assumed to have the same wall loss rate, and its value was sampled from the range 0 and 10^{-2} s^{-1} . The size-independence of the wall loss rate may cause some uncertainty to the results, but introducing even more free parameters in order to vary the value separately for each cluster would also be problematic.”

5) Section 3.2.1. and Section 4.3: Besides fragmentation also the transmission efficiency of the mass spectrometer should be discussed (see, e.g. Heinritzi et al., AMT, 2016). The mass dependent transmission efficiency also influences the observed cluster distributions. While fragmentation can only lead to an overestimation of the measured small clusters and underestimation of the large clusters, changes in the transmission efficiency can also have the opposite effect. Transmission efficiency is very dependent on the tuning of the individual mass spectrometer. Influences on the observed distributions due to uncertainties of the transmission efficiency or mass discrimination should be discussed.

This is a good point and should certainly be taken into account in future studies. In this paper, however, adding new fitting parameters is really not possible due to the small number of experimental data points. The transmission efficiency is now mentioned in the revised manuscript.

Page 18, line 32 – page 19, line 2:

“However, if it is not possible to suppress fragmentation completely for some instrument type or experimental setup, it is important at least to gain some understanding of the fragmentation processes, and MCMC analysis appears to be a suitable tool for this. In this study, the mass spectrometer was assumed to have been calibrated so that there was no mass discrimination, but in the future, also the mass dependent transmission efficiency of mass spectrometers could be studied

using MCMC analysis.”

6) Section 3.3. At some point the limits of the MCMC should be discussed in more detail. Currently this discussion is distributed over the paper and limitations become evident from the results but it would be helpful to state the limitations already in the beginning of Section 3.3. When just 22 experimental distributions can be used to derive a large set of parameters, and additionally the input parameters are correlated, then the solutions will not be unambiguous. More discussion of this is needed.

This is not really a limitation of MCMC, but a problem caused by not having enough experimental data available. Therefore the problem is mentioned in Sect. 4 instead of Sect. 3.3 as suggested.

Page 9, lines 28–32:

“Even when the MCMC simulation finds a good fit to the observed distributions, the interpretation of the output parameter distributions is not always clear. The number of input data points from the CLOUD experiment is so small that unambiguous values were not reached for most of the evaporation rates. To get better insight into what conclusions can safely be drawn, Sect. S2 of the Supplementary Material presents test simulations for synthetic input cluster distributions with known evaporation rates and fragmentation probabilities.”

7) Table 1 and section 4.2: The “alternative solutions” and cases (A)-(E) are listed but not explained at all. The differences need to be briefly described so that the reader has some idea about what is different in these cases without reading the Supplementary Material (see also comment #11).

An explanation has been added.

Page 13, line 1 – page 14, line 3

“These alternative solutions correspond to different cluster types being stable and unstable, but they all still give an equally good fit to the measured cluster distributions. For instance, when assuming a background ammonia concentration of 5 ppt, the posterior distribution of parameter number 5 shows three separate peaks (purple line in Fig. 4). Looking only at the sets of parameter values in the right hand side peak, it can be noted that the value of parameter number 3 corresponds always to the left hand side peak of this distribution (see Fig. S14 of the Supplementary Material), and parameter number 8 always has a low value due to correlations between the different parameters. This set of ranges for the parameter values is denoted as solution (E), and similarly the two other peaks in the distribution of parameter number 5 correspond to solutions (C) and (D). The observation that the distributions can be divided into separate solutions in this way implies that, for instance, either an evaporation rate of 100 s^{-1} for ammonia from the $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_3 \cdot \text{NH}_3$ cluster and an evaporation rate of 3 s^{-1} of the pure sulfuric acid tetramer or an evaporation rate of 0.2 s^{-1} for ammonia from the $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_3 \cdot \text{NH}_3$ cluster and an evaporation rate of 60 s^{-1} of the pure sulfuric acid tetramer could produce a good fit to the experimental cluster distributions, but an evaporation rate of 100 s^{-1} for ammonia from the $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_3 \cdot \text{NH}_3$ cluster and an evaporation rate of 60 s^{-1} of the pure sulfuric acid tetramer would not reproduce the data. For the simulations with an ammonia concentration of 1 ppt, the separate solutions (A) and (B) correspond to the two peaks in the distribution of coefficient number 6.”

8) Figure 6 shows the total fragmentation probabilities, e.g. the upper left panel, displaying $\text{A}_2\text{A}^- \rightarrow \text{A}_{0-1}\text{A}^-$, should be formed from #18 and #19 from Fig 5. Why does the peak at about 0.2, where #19 has its maximum, not show up in the upper left panel of Fig 6? Adding a scale to the y-axis could be helpful.

It should be kept in mind that the distribution in Fig. 6 is not the sum of the *distributions* in Fig. 5 but a distribution of the *sum of the parameters*. This has now been written out more clearly. The scale of the y-axis depends only on the number of iterations in the MCMC simulation, and is completely irrelevant.

Page 16, lines 3–4:

“The *posterior distributions of the overall fragmentation probabilities (that is the sums of the probabilities of all fragmentation processes in which a given cluster can be lost)* of these clusters are shown in Fig. 6.”

9) P17 l11-21:

a) An unexpected result is the high stability of the pentamer while the tetramer is less stable. It is mentioned that the stability could be due to hydration of the pentamer but hydration should also stabilize the tetramer. Please discuss.

What I meant is that perhaps the pentamer is stabilized more strongly by hydration than the other clusters. This has now been clarified in the text.

Page 18, lines 11–13:

“On the other hand, these previously determined cluster energies correspond to dry clusters, and hydration is likely to stabilize clusters at least to some extent. *It is, in principle, possible that the pentamer could have a very stable hydrated structure, while the tetramer would only be moderately stabilized by hydration.*”

b) Could it be that the pentamer forms in a “closed shell” cluster configuration that is more stable than the tetramer?

This is possible, although it would be somewhat surprising as it has not been observed in quantum chemistry studies. However, this would still not explain the discrepancy with the semi-empirical results of Lovejoy and Curtius (2001) and Curtius et al. (2001).

c) The stabilities can also compared with the lifetimes of clusters discussed in Hanson and Eisele, JGR, 2002, Section 2.3.2 and 3.1.

The reason why I had not discussed these results is that Hanson and Eisele only obtained estimates for the lifetimes of the clusters $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_3 \cdot (\text{NH}_3)_3$ and $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_4 \cdot (\text{NH}_3)_4$, while no estimates were obtained for the evaporation rates of these clusters in the present study. Thus there is nothing to compare.

10) Acknowledgment: p18 l14-16. The CLOUD team and CERN resources should be acknowledged for provision of the experimental data.

This is quite a strange suggestion, as I have used previously published data and cited the article from which I took that data, and I have had no contact whatsoever with the CLOUD community about this study.

11) Supplementary Material. I am lost in section S2.6. It is not clear how the separation was made and why it was made in the way it was made. At the end of p7 the separation of several cases is briefly explained. I do not understand why parameters 3 and 5 are selected for the separation of the synthetic data and why is parameter 6 selected for the posterior distributions with 1 ppt ammonia and parameter 5 for the 5 ppt ammonia simulations, respectively. It is stated that “First, it can be seen in Fig. S4 that the posterior distribution of

coefficient number 3 has two peaks.” I think, Figure 4 is meant here. But even then, only the purple line (5ppt) has two peaks (are we supposed to look only at the purple line? Why not blue and green?). The selection process seems to be arbitrary. Furthermore, the five lines of description of S3.3 on page 9 are much too short. It is still unclear what makes the difference for cases (A) and (B), and (C)-(E).

Some clarifications were added.

Page 7, 2nd paragraph:

“The correlations can be found by grouping the sets of parameter values according to the value of one specific parameter, and comparing the posterior distributions of the other parameters for these groups. If the groups of parameters form completely separate peaks also in the distributions of one or more of the other parameters, a correlation has been found.”

Page 7, last paragraph:

“First, it can be seen in Fig. S4 that the posterior distribution of coefficient number 3 has two peaks (light green line in Fig. S4).”

Page 9, last paragraph – page 10:

“Figures S13-S18 show the posterior distributions of Figs. 4, 5 and 6 of the main article corresponding to MCMC simulations with a fixed background ammonia concentration separated into the alternative solutions (A)–(E) of Table 1. The posterior distributions were split into the different solutions similarly as described in Sect. S2.6. For the simulation with 1 ppt, the solutions were separated based on the two peaks in the distribution of coefficient number 6 (blue line in Fig. 4 of the main text), as this produced a neat separation also for coefficients number 7 and 8 (Fig. S13). For the case with 5 ppt ammonia the solutions were separated based on the three peaks in the distribution of coefficient number 5 (purple line in Fig. 4 of the main text), as this produced neat separation also for coefficients number 3, 6, 7 and 8 (Fig. S14).”

12) Section S2.6. Second line: “consider a case were” → “consider a case where”

The typo was corrected.

A Monte Carlo approach for determining cluster evaporation rates from concentration measurements

Oona Kupiainen-Määttä¹

¹University of Helsinki, Department of Physics, P.O. Box 64, FI-00014 University of Helsinki, Finland

Correspondence to: Oona Kupiainen-Määttä (oona.kupiainen@alumni.helsinki.fi)

Abstract. Evaporation rates of small negatively charged sulfuric acid–ammonia clusters are determined by combining detailed cluster formation simulations with cluster distributions measured at CLOUD. The analysis is performed by varying the evaporation rates with Markov chain Monte Carlo (MCMC), running cluster formation simulations with each new set of evaporation rates and comparing the obtained cluster distributions to the measurements. In a second set of simulations, the fragmentation of clusters in the mass spectrometer due to energetic collisions is studied by treating also the fragmentation probabilities as unknown parameters and varying them with MCMC. This second set of simulations results in a better fit to the experimental data, suggesting that a large fraction of the observed HSO_4^- and $\text{HSO}_4^- \cdot \text{H}_2\text{SO}_4$ signals may result from fragmentation of larger clusters, most importantly the $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_2$ trimer.

1 Introduction

Gas-phase sulfuric acid has long been believed to be an important precursor for particle formation in the atmosphere (Doyle, 1961; Kiang et al., 1973; Cox, 1973; Mirabel and Katz, 1974). The details of the process have, however, remained poorly understood until lately. Recent laboratory experiments (Berndt et al., 2010; Benson et al., 2011; Almeida et al., 2013) have confirmed that particle formation rates of the magnitude observed in the atmosphere can be produced with ambient sulfuric acid concentrations and low concentrations of base molecules, giving new support for sulfuric acid being at least one of the compounds driving atmospheric particle formation. The experiments of Kirkby et al. (2011); Almeida et al. (2013) have also shown that the first steps of cluster formation can proceed along an ionic pathway, and that this process can dominate over the electrically neutral pathway when there are not enough base molecules or other impurities available to stabilize the small neutral sulfuric acid clusters.

The development of highly sensitive mass spectrometers has enabled the detection and characterization of individual ionic clusters consisting of only a few molecules (Eisele and Hanson, 2000; Zhao et al., 2010; Junninen et al., 2010), opening a new window into the first steps of cluster formation. However, measurements alone cannot fully uncover the dynamics of the process, as they only provide information on the concentrations, not the collision and evaporation fluxes from one cluster type to another.

At the same time, modeling of particle formation has also advanced greatly in the past few years. For the first time theoretical predictions of cluster distributions (Olenius et al., 2013b) and particle formation rates (Almeida et al., 2013) agree qualitatively with experimental findings.

Cluster formation simulations require as input the collision and evaporation rates of clusters. The collision frequencies are usually computed simply using classical physics, and an estimate of the evaporation rates can be obtained by relying on equilibrium considerations and using the formation free energies of the clusters computed by quantum chemistry. This approach has been shown to give qualitative agreement with experiments (Almeida et al., 2013; Olenius et al., 2013b), but several very drastic assumptions are involved. First-principles molecular dynamics simulations (Loukonen et al., 2014a, b) have shown that one harmonically oscillating cluster structure is far from a realistic description of the thermal motion of molecules in a cluster, as molecules may rotate inside the cluster continuously breaking intermolecular bonds and forming new ones. This implies that the traditional way of computing cluster formation free energies may be a rough approximation. As evaporation rates depend exponentially on the cluster formation energies, theoretical evaporation rates may easily be wrong by several orders of magnitude. Different quantum chemistry methods can give qualitatively very different predictions for cluster concentrations (Kupiainen-Määttä et al., 2013; Kupiainen-Määttä et al., 2015), and it is not clear whether any of the methods can be trusted. Also the treatment of the collision rates is highly simplified, but errors of more than a factor of two or perhaps ten are unlikely.

An alternative approach for estimating the rate constants is to start from experimental cluster concentrations and find rate constants that reproduce these results. This has been done previously by Bzdek et al. (2010) who measured time series of cluster concentrations in order to study base exchange in positively charged clusters containing a fixed number of sulfuric acid molecules, and by Jen et al. (2014) who measured concentrations of neutral clusters containing two sulfuric acid molecules in the presence of different base compounds. However, in both cases the studied system consisted of only a few cluster types, and the theoretical description was highly simplified. Bzdek et al. (2010) assumed sequential pseudo-first order substitution reactions, and used the analytic solution of the time evolution of the concentrations to fit the pseudo-first order rate constants. Jen et al. (2014), on the other hand, used a heuristic cluster formation model with only two free parameters to optimize. In both cases, the optimization problem was simple enough that traditional fitting tools could be used. More recently, Chen et al. (2015) used a more complicated model with tens of unknown parameters to describe measured particle concentrations in an experiment involving methanesulfonic acid, trimethylamine and water, but they used effective reaction rates instead of separate collision and evaporation rates, and only presented one reasonably good fit instead of attempting to find either the best fit or all sets of parameter values giving a good fit.

In this study, measured cluster distributions are combined with detailed cluster formation simulations describing explicitly all possible collision and evaporation processes. Theoretical estimates are used for the collision rates, while all evaporation rate coefficients as well as some parameters related to experimental details are optimized to reproduce the experimental data. Due to the large number of unknown parameters, the fitting is done by Monte Carlo simulation. The method is applied to measurement data from the CLOUD experiment (Olenius et al., 2013b; see Kirkby et al., 2011 for more details on the CLOUD experiment). This study focuses solely on ion clusters, but a similar approach could also be used for determining evaporation rates of neutral clusters based on cluster distributions measured with a chemical ionization mass spectrometer.

2 Experimental ion cluster distributions

The experimental cluster distributions used in this study are from an earlier publication from the CLOUD experiment at CERN (Olenius et al., 2013b). Concentrations of negatively charged sulfuric acid–ammonia clusters were measured in steady-state conditions with sulfuric acid vapor concentrations between 10^7 and 10^9 cm^{-3} and ammonia mixing ratios from below 35 ppt up to 250 ppt. The clusters were detected using a high resolution APi-TOF (Atmospheric Pressure interface Time-Of-Flight) mass spectrometer. The largest clusters considered in the study contained one HSO_4^- ion, four H_2SO_4 molecules and four ammonia molecules. However, it is likely that most of the clusters initially also contained some water molecules, although none were detected, and water was concluded to evaporate from the clusters inside the APi-TOF. The clusters were also assumed to lose some or all of the ammonia molecules inside the instrument prior to detection. Therefore, the concentrations were reported separately for ammonia containing and ammonia-free clusters, but the ammonia containing ones were not sorted further by number of ammonia molecules. The bisulfate ion HSO_4^- and the two smallest clusters, $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_{1-2}$, were only observed with no ammonia molecules attached. Olenius et al. (2013b) presented a total of 25 cluster distributions measured with ion production from natural ionization, a temperature of 278 K and different sulfuric acid and ammonia vapor concentrations, but three of these distributions had very low concentrations for some of the cluster types and were thus omitted from the present study.

3 Simulation methods

Cluster dynamics simulations were performed with ACDC (Atmospheric Cluster Dynamics Code), a program that writes out the birth-death equations for a given set of molecules and clusters and solves them by numerical integration. Unlike in earlier implementations of ACDC where MATLAB was used, the birth-death equations were now integrated using the Fortran ordinary differential equation solver VODE (Brown et al., 1989). A detailed description of the code has been published elsewhere (McGrath et al., 2012; Olenius et al., 2013a), and only the main points and the differences to the earlier version are presented here.

3.1 ACDC simulations

To minimize the computational burden of solving the birth-death equations, only negatively charged clusters were considered. Both quantum chemical calculations and mass spectrometry measurements indicate that negatively charged clusters with three sulfuric acid molecules or less (including the bisulfate ion) do not take up ammonia molecules (Kirkby et al., 2011; Olenius et al., 2013b). Based on the main formation pathway in cluster formation simulations (Olenius et al., 2013a), the clusters $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_{0-2}$, $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_3 \cdot (\text{NH}_3)_{0-3}$ and $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_4 \cdot (\text{NH}_3)_{0-4}$ were chosen to form the simulated system in this study. The only electrically neutral species included in the simulation were the H_2SO_4 and NH_3 monomers.

Some of the negatively charged clusters could in principle result from collisions of neutral clusters with negative ions, but both experimental observations (Jen et al., 2014) and quantum chemical calculations (Olenius et al., 2013a) suggest that sulfuric acid–ammonia clusters are so weakly bound that their concentrations are orders of magnitude lower than the sulfuric acid monomer concentration at conditions corresponding to the experiments reported by Olenius et al. (2013b). Therefore, the contribution of neutral clusters was not taken into account in this study. Water molecules were not modeled explicitly, but the collision and evaporation coefficients should be interpreted as effective rates averaged over the hydrate distribution of each cluster type (see for instance Paasonen et al., 2012).

In addition to growing by collisions with monomers or decaying by monomer evaporations, the negative clusters can get neutralized by recombination with positively charged ions and clusters. To keep the situation simple, the distribution of positive clusters was not simulated explicitly, but the overall positive ion concentration was set to match the total negative ion concentration, and all negative ions were assumed to have the same recombination rate coefficient of $1.6 \times 10^{-6} \text{ cm}^3 \text{ s}^{-1}$ (Israël, 1970) with these generic positive ions. The formed neutral clusters were outside the system of interest, and their concentrations were not recorded.

The formation of negative ions was modeled similarly as was done by Almeida et al. (2013). Generic charger ions with the properties of O_2^- are first produced at a constant rate, and upon collisions with H_2SO_4 molecules they ionize these to form bisulfate ions. The charger ions can also be lost by recombination with positive ions. Finally, all clusters and charger ions can be lost on the chamber walls, and this was described by a size- and composition-independent wall loss coefficient.

To mimic the experimental conditions as closely as possible, each simulation was started from a situation with non-zero sulfuric acid and ammonia monomer concentrations and no ions. The charger ion source was switched on, and the time evolution of the cluster concentrations was simulated keeping the neutral monomer concentrations constant. The experimental cluster distributions correspond to steady-state conditions (Olenius et al., 2013b), and the lengths of the individual experiments were of the order of half an hour (Kirkby et al., 2011). The modeled cluster distribution was calculated as an average of the distributions at time $t_1 = 20 \text{ min}$ and $t_2 = 30 \text{ min}$ after the beginning of the run. The extent to which the simulation had reached a steady state was characterized by the ratio of the concentrations at t_2 and t_1 , calculated in each case for the cluster for which this ratio deviated most from unity. This convergence parameter was used together with the cluster concentrations to determine how well the simulations reproduced the experimental results.

3.2 Simulation parameters

As the measurement data consisted of steady-state concentrations, it was not possible to fit both the collision and evaporation rates – multiplying all rate constants by the same factor would only change the timescale of the process but not the steady-state concentrations. Collision frequencies between ions and polar or polarizable molecules can be approached theoretically by considering classical electrostatic interactions. While a closed-form analytical expression cannot be obtained even when neglecting quantum effects, theoretical estimates for collision rates are much more reliable than those for evaporation rates. In all the simulations presented in this study, the collision rate constants were computed using the parameterization of Su and

Chesnavich (1982) based on classical trajectory simulations. The values for the reactions in the studied system were between 10^{-9} and $4 \times 10^{-9} \text{ cm}^3\text{s}^{-1}$.

In principle, the evaporation rates might have any values, and there is no way to constrain even their order of magnitude based on earlier experimental evidence or simple theoretical considerations. However, the interval in which the evaporation rates are allowed to vary does not in practice need to be infinitely wide. If the length of the simulation is 30 minutes, it does not matter whether a cluster has a lifetime of one day or one week – it will in any case not evaporate. On the other hand, if a cluster collides with monomers on average once per second or once per minute, there is no effective difference whether it has an evaporation lifetime of one millisecond or one microsecond – it will almost certainly evaporate before it has a chance to grow further. Even so, the range of interest for the evaporation rates spans several orders of magnitude, and the base ten logarithms of the rates (used as the parameters to be varied by MCMC instead of the rates themselves) were sampled from the range of -10 to 10.

The simulations also involve a large number of experiment-related parameters whose values cannot be measured directly or estimated reliably based on any fundamental theory. These were also treated as free parameters and varied using MCMC. For some of the parameters, however, at least an order-of-magnitude estimate is available, and these estimates were used for constraining the range in which the parameters were allowed to vary.

A wall loss rate of $1.7 \times 10^{-3} \text{ s}^{-1}$ was determined for the electrically neutral H_2SO_4 monomer in the CLOUD chamber (Almeida et al., 2013). This rate decreases with increasing cluster size, but ions may have a higher loss rate. **The probability of an individual cluster being lost on a wall also varies with location inside the chamber, or in practice with time as the air is continuously circulated around the chamber by large fans. As the size, charge and composition dependence of the wall losses is not known, [-] all clusters were, for simplicity, assumed to have the same wall loss rate, and its value was sampled from the range 0 and 10^{-2} s^{-1} . The size-independence of the wall loss rate may cause some uncertainty to the results, but introducing even more free parameters in order to vary the value separately for each cluster would also be problematic.**

Based on measured ion concentrations and approximate loss rates of ions, the ion production rate due to natural ionization was estimated to be of the order of 3 ion pairs $\text{cm}^{-3}\text{s}^{-1}$ (Olenius et al., 2013b). In this study, it was sampled from the range of 0 to 10 ion pairs $\text{cm}^{-3}\text{s}^{-1}$.

In some experiments, no ammonia was added intentionally to the chamber. While its concentration was in these cases below the detection limit of 35 ppt, some trace amount must have been present as ammonia molecules were observed in the clusters. In the simulations, two approaches were used regarding the ammonia concentration: either a constant background ammonia mixing ratio of 5 ppt was used for all these experiments, or the mixing ratio was allowed to vary separately for each of these low-ammonia experiments, and the values were sampled between 0 and 50 ppt.

3.2.1 Fragmentation in the mass spectrometer

It is possible that some clusters fragment inside the instrument before detection. Weakly bound water molecules probably evaporate to a great extent (Ehn et al., 2011), and they are not taken explicitly into account in the cluster distribution. Also ammonia and sulfuric acid molecules may be detached from the clusters due to energetic collisions with gas molecules when the

clusters are accelerated inside the instrument. In some of the MCMC simulations, all clusters were allowed to fragment, and the fragmentation probabilities were sampled between 0 and 1, with the constraint that the sum of all fragmentation probabilities corresponding to the same cluster fragmenting to form different products could not be higher than one.

In an IMS-TOF (ion mobility spectrometer – time-of-flight mass spectrometer) experiment, detachment of sulfuric acid molecules was observed to be important at least for the pure trimers, $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_2$, which can lose either one or two H_2SO_4 molecules (Adamov et al., 2013). In the present study, each of the pure sulfuric acid clusters $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_i$ could fragment through i different processes with separate fragmentation probabilities, forming the products $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_{0,1,2,\dots,(i-1)}$.

On the other hand, in another IMS-TOF experiment, larger sulfuric acid–dimethylamine clusters $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_i \cdot ((\text{CH}_3)_2\text{NH})_i$ with $i = 3, 4, 5$ were observed not to fragment (Bianchi et al., 2014). The fragmentation patterns of larger clusters containing sulfuric acid and ammonia have not been determined experimentally, and it is possible that fragmentation is more important than for the above-mentioned dimethylamine-containing clusters. However, the larger the cluster, the more vibrational degrees there are to absorb any excess energy released in collisions, so the fragmentation probabilities can be expected to decrease with increasing cluster size Kurtén et al. (2010). For simplicity, detachment of sulfuric acid molecules from ammonia-containing clusters was not taken into account, although it might in reality occur to some extent, and the removal of ammonia molecules from the clusters was described by only four parameters: the probabilities of detecting $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_3 \cdot \text{NH}_3$ and $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_3 \cdot (\text{NH}_3)_{2-3}$ clusters as pure acid tetramers and of detecting $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_4 \cdot \text{NH}_3$ and $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_4 \cdot (\text{NH}_3)_{2-4}$ clusters as pure acid pentamers. This choice of fragmentation-related parameters is a trade-off between describing the processes as accurately as possible and keeping the number of free parameters reasonable.

3.3 Monte Carlo simulations

The effect of the above-mentioned unknown parameters (evaporation rates, ion production rate, wall loss coefficient, background ammonia concentrations, fragmentation probabilities) on the cluster distribution was studied by Bayesian analysis using Markov chain Monte Carlo (MCMC). (See e.g. Brooks et al., 2011, for an introduction to MCMC methods.) The aim of MCMC in parameter estimation is to find combinations of parameter values that reproduce the experimental data as well as possible. Instead of finding one best fit, the objective is to find a distribution of the most likely parameter values. This is accomplished by forming a chain \mathcal{Z} of parameter values that converges toward the desired distribution as the length of the chain increases.

3.3.1 The Metropolis algorithm

The parameters are varied using a random-walk approach, and at each step the new parameter values (denoted as the vector \mathbf{x}_{new} with length n_{coefs}) are used for running ACDC simulations corresponding to all experiments. In the Metropolis algorithm, the proposal density $q(\mathbf{x}_{\text{new}}, \mathbf{x}_{\text{old}})$ describing the probability of attempting a step from the old point \mathbf{x}_{old} to a new point \mathbf{x}_{new} is equal to the proposal density $q(\mathbf{x}_{\text{old}}, \mathbf{x}_{\text{new}})$ related to the reverse step from \mathbf{x}_{new} to \mathbf{x}_{old} . The difference between the modeled

and measured cluster distributions is quantified by the square sum

$$SS_{\text{new}} = \sum_{i=1}^{n_{\text{out}}} (\log_{10} y_{\text{exp},i} - \log_{10} y_{\text{new},i})^2 \quad (1)$$

where $n_{\text{out}} = n_e \times (n_c + 1)$ is the number of output values, $n_e = 22$ is the number of experiments, $n_c = 7$ is the number of cluster types whose concentrations are measured, \mathbf{y}_{new} is a vector of length n_{out} containing simulated cluster concentrations for all runs as well as one convergence parameter (see Sect. 3.1) for each run, and \mathbf{y}_{exp} is the corresponding vector for the experimental data with a value of 1 for the convergence parameter for all runs. The reason for including the convergence parameter here is to penalize low wall loss rates and ion source rates that would lead to an unrealistically slow time evolution of the cluster distribution.

Assuming that the experimental data contains measurement errors that can be described as uncorrelated multiplicative log-normal noise with the same variance σ^2 for each measured value $y_{\text{exp},i}$, the likelihood of observing the data \mathbf{y}_{exp} given the parameter values \mathbf{x}_{new} is

$$p(\mathbf{y}_{\text{exp}} | \mathbf{x}_{\text{new}}) = \frac{1}{(2\pi\sigma^2)^{n_{\text{out}}/2}} \exp\left(-\frac{1}{2\sigma^2} SS_{\text{new}}\right). \quad (2)$$

At each step of the random walk, the value SS_{new} is compared to the square sum SS_{old} saved at the previous step. If the new value is lower or equal to SS_{old} , that is if the new parameter values reproduce the experimental data at least as well as the previous ones, the point is accepted. If, on the other hand, $SS_{\text{new}} > SS_{\text{old}}$, the point may still get accepted, but only with probability

$$\frac{p(\mathbf{y}_{\text{exp}} | \mathbf{x}_{\text{new}})}{p(\mathbf{y}_{\text{exp}} | \mathbf{x}_{\text{old}})} = \exp\left[-\frac{1}{2\sigma^2} (SS_{\text{new}} - SS_{\text{old}})\right]. \quad (3)$$

The overall acceptance probability for both cases can then be written as $\alpha = \min\left(1, \exp\left[-\frac{1}{2}\sigma^{-2}(SS_{\text{new}} - SS_{\text{old}})\right]\right)$. If the new point is accepted, the parameter values \mathbf{x}_{new} are saved to the chain \mathbf{Z} and SS_{old} is replaced by SS_{new} . Otherwise the previous point \mathbf{x}_{old} is added again to the chain \mathbf{Z} .

3.3.2 DE-MC_Z algorithm for finding all local maxima of the distribution

Some parameters were found to have posterior distributions with more than one local maximum. Plotting two-dimensional posterior distributions of pairs of parameters showed in many cases L-shaped or otherwise non-convex regions of high probability that are hard to sample using traditional methods. In order to ensure that the random walk was able to find all the local maxima and converged to the correct distribution, the DE-MC_Z algorithm (Differential Evolution Markov Chain algorithm sampling the difference vectors from the past) introduced by ter Braak and Vrugt (2008) was employed. In DE-MC_Z, several chains are run in parallel, and each chain in turn takes a step $x_{\text{new}} = x_{\text{old}} + \gamma(x_1 - x_2) + \delta$, where γ is a scalar, x_1 and x_2 are two different randomly selected points from the joint history of all chains, \mathbf{Z} , and δ is a small additional term drawn from a normal distribution with a small variance compared to the width of the posterior distribution. Ter Braak and Vrugt (2008) found that three chains worked well in their test systems, but in this study, five chains were used as they were noted to ensure

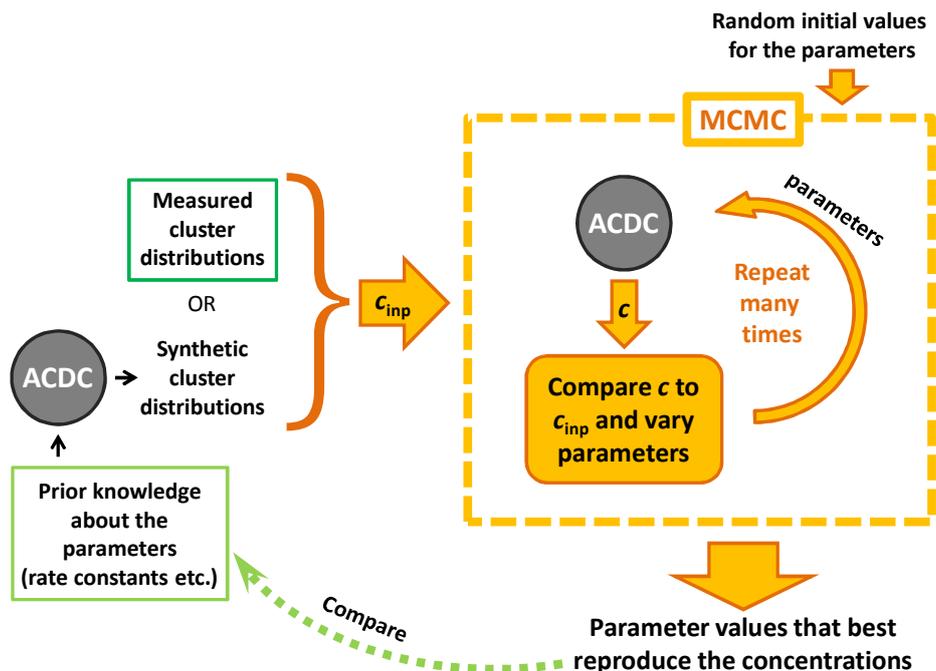


Figure 1. Schematic representation of the steps involved in the study. The green boxes show the two alternative starting points.

better mixing. Based on the recommendations of ter Braak (2006) and ter Braak and Vrugt (2008) and on test simulations, γ was set to 0.98 at every fifth step and $2.38/\sqrt{2 \times n_{\text{coefs}}}$ otherwise. The width of the distribution for sampling δ was based on an estimate of the width of the posterior distribution as discussed in the Supplementary Material. As the rule for proposing steps is symmetric with respect to x_{new} and x_{old} but depends on the history, the DE-MC_Z algorithm is an adaptive Metropolis algorithm and the acceptance probability is calculated like in the basic Metropolis algorithm.

Further details about the MCMC simulations are presented in the Supplementary Material.

3.4 Overview of the simulations

An overview of the simulation methods is presented in Fig. 1. The same MCMC procedure (shown in orange in the Figure) was used with two alternative sets of cluster distributions as input. In both cases, these cluster distributions corresponded to 22 individual experiments (or computer experiments) with varying sulfuric acid and ammonia vapor concentrations, and for each experiment the concentrations of seven cluster types were included in the distribution. In the MCMC simulation, all unknown parameters (evaporation rates etc.) were first given some random values, and these were used for running a set of 22 ACDC simulations with vapor concentrations corresponding to the input cluster distributions. The cluster concentrations obtained from the ACDC runs were compared with the input cluster concentrations, and the parameters were given new values. The new parameter values were again used to run a set of ACDC simulations, and the process was repeated over and over.

The starting point of the main part of the study (dark green box) were the 22 cluster distributions measured at CLOUD at varying sulfuric acid and ammonia vapor concentrations. These were used as input for an MCMC simulation, and the main output of the MCMC simulation were parameter values that reproduced most closely the measured cluster distributions. However, unlike traditional fitting procedures, MCMC gives a distribution of most likely parameter values (called the posterior distribution) and corresponding cluster distributions instead of one best fit.

The second part of the study focused on testing the performance of the MCMC data analysis method. First one possible set of parameter values was selected (light green box in Fig. 1). Quantum-chemistry based theoretical predictions were used for the cluster evaporation rates, and the other parameter values were estimated based on the experiment. These parameter values (referred to later in the paper as input parameter values) were used as input for a set of 22 ACDC simulations corresponding to the same sulfuric acid and ammonia vapor concentrations as in the experimental cluster distributions. The chosen input values of the fragmentation parameters were applied to the output concentrations from these ACDC runs to get a set of 22 cluster distributions. Some random noise (see Sect. 4.1) was added to these simulated cluster distributions to obtain synthetic 'measured' cluster distributions. These, in turn, were then used as input for an MCMC simulation, and the output was again a distribution of most likely parameter values as well as corresponding cluster distributions. Since in this case the 'correct answers', that is the input parameter values used to produce the synthetic cluster distribution, were known, the parameter distributions obtained as output from MCMC could be compared to the input values.

4 Results

Although the main result from the MCMC simulation are the distributions of likely parameter values, it is useful first to look at the cluster distributions corresponding to these output parameter values (referred to later as output cluster distributions) and check how accurately the input data is reproduced. Such comparisons are presented in Sect. 4.1 for the CLOUD data and two sets of MCMC simulations with a different set of free parameters. If the output cluster distributions are very far from the measured cluster distributions, it can be concluded that the model used in the simulations did not correspond closely enough to the actual processes determining the observed cluster distributions. In such a case, the fitted parameters do not necessarily correspond directly to the corresponding real parameters, or indeed have any clear physical interpretation.

The output values of the evaporation rates and fragmentation probabilities are discussed in detail in Sects. 4.2 and 4.3, respectively, only for cases where the output cluster distributions reproduce closely the measured concentrations. The results for the other parameters are presented in Sect. S3 of the Supplementary Material.

Even when the MCMC simulation finds a good fit to the observed distributions, the interpretation of the output parameter distributions is not always clear. **The number of input data points from the CLOUD experiment is so small that unambiguous values were not reached for most of the evaporation rates.** To get better insight into what conclusions can safely be drawn, Sect. S2 of the Supplementary Material presents test simulations for synthetic input cluster distributions with known evaporation rates and fragmentation probabilities.

4.1 Cluster distributions

Figure 2 presents the experimental cluster distributions from CLOUD together with the output cluster distributions from an MCMC simulation where only the evaporation rates are varied and fragmentation in the mass spectrometer is not taken into account. The background ammonia concentration is set to 5 ppt, and the values reported by Olenius et al. (2013b) are used for the ion production rate and wall losses. The medians of each concentration from the output of the MCMC simulations are presented as a horizontal line, and the vertical lines span between the 2.5th and 97.5th percentiles. Comparison of the measured and simulated concentrations shows that while overall the simulated concentrations are mostly of a correct order of magnitude, the MCMC fitting does not produce the correct precursor concentration dependence for all ion cluster types. In case of the bisulfate ion HSO_4^- and the charged dimer $\text{HSO}_4^- \cdot \text{H}_2\text{SO}_4$, the measured ion concentrations are notably lower in the experiments with a high ammonia concentration than in experiments with a similar acid concentration and no added ammonia, while the simulated concentrations show practically no ammonia dependence. For the larger clusters, on the other hand, the ammonia dependence is captured reasonably well. However, the sulfuric acid concentration dependence of the output cluster distributions differs from the observed dependence also for many of the larger clusters at low ammonia concentrations. This discrepancy is most prominent for $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_4 \cdot \text{NH}_3$ and HSO_4^- .

Using the ion production rate and wall loss constant as free parameters while still keeping a fixed background ammonia concentration does little to improve the fit. The same discrepancies remain also if the background ammonia concentrations are varied.

Figure 3 presents the output cluster distributions from an MCMC simulation where the fragmentation probabilities discussed in Sect. 3.2.1 are treated as free parameters. The ion production rate and wall loss constant are also varied, but all background ammonia concentrations are set to 5 ppt. Apart from a few outliers in the experimental concentrations, the agreement between the measured and modeled concentrations is remarkably good. This suggests that the poor fit in Fig. 2 may be explained by the concentrations observed by the mass spectrometer not corresponding directly to the ion concentrations in the CLOUD chamber, but instead to the concentrations after some of the clusters have fragmented in the inlet of the mass spectrometer. In fact, the acid and base monomer concentration dependence is very similar for the measured concentrations of the three smallest ions, $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_{0,2}$, which would be consistent with some of the trimers being detected as monomers and dimers after having fragmented inside the instrument.

4.2 Evaporation rates from the analysis

Figure 4 shows the posterior distributions of the coefficients corresponding to logarithms of the evaporation rates. The three sets of distributions correspond to different options for treating the background ammonia concentration. Either all below-detection-limit ammonia concentrations are varied separately as MCMC parameters (green), or they are all set to 1 ppt (blue) or 5 ppt (purple). In the MCMC simulation where the background ammonia concentration is fitted, the median values for these concentrations are between 7 and 20, although the values are spread from 0 to 30 or 40.

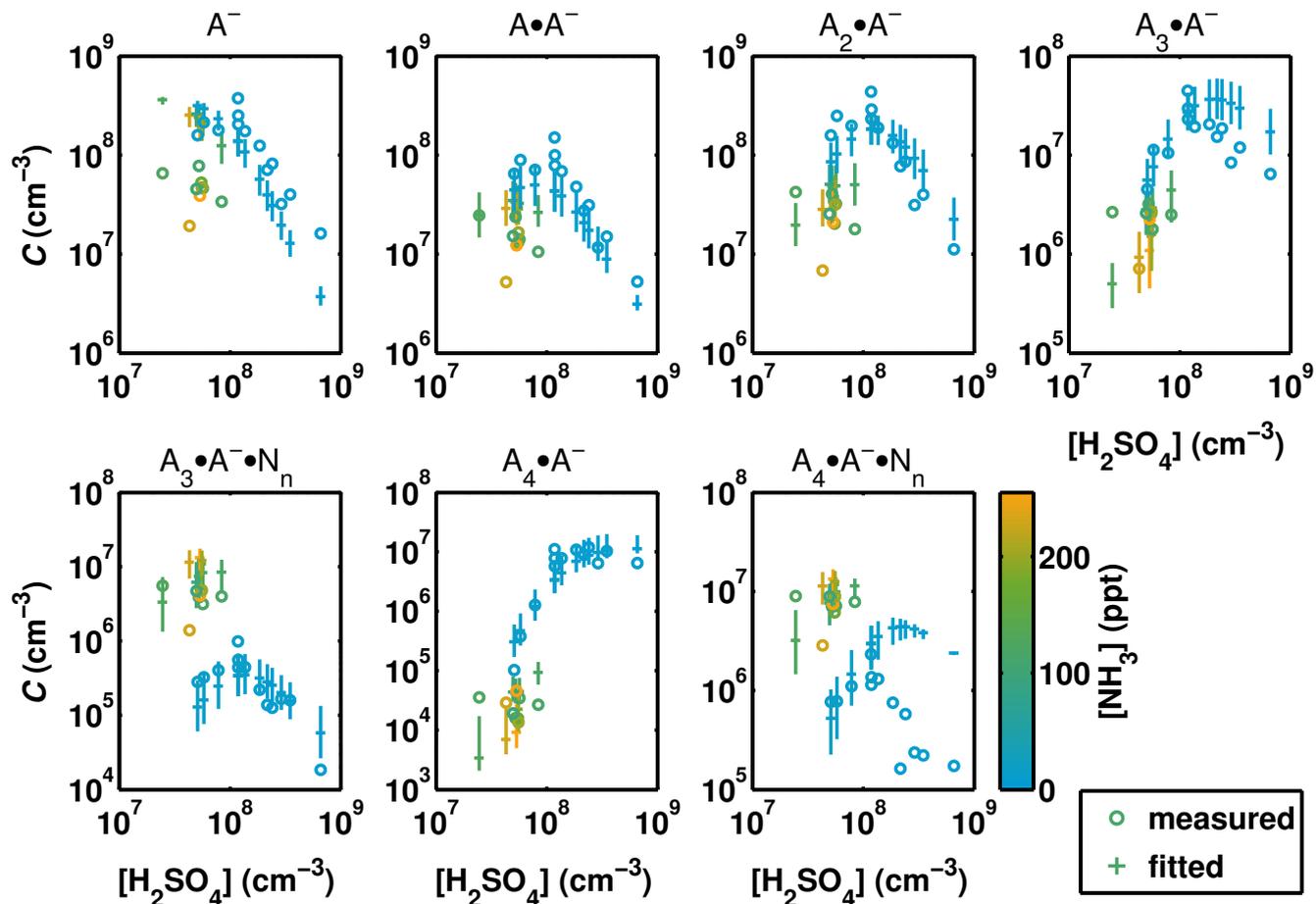


Figure 2. Cluster distributions measured at CLOUD and the corresponding modeled cluster concentrations from an MCMC simulation where only the evaporation rates are varied and no fragmentation is allowed. A stands for H_2SO_4 , A^- for HSO_4^- and N for NH_3 .

All sets of MCMC simulations give a similar result for parameters number 1, 2 and 4: the pure negatively charged sulfuric acid dimer $\text{HSO}_4^- \cdot \text{H}_2\text{SO}_4$, trimer $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_2$ and pentamer $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_4$ are stable, having evaporation rates below 1 s^{-1} . The reason for the uniform shape of these distributions at low evaporation rates is that once the evaporation rate is much lower than the rates of any competing processes, its exact value has no effect on the cluster distribution. As discussed in the

5 Supplementary Material, the peak seen in some of these distributions should not be interpreted as giving a good estimate for the evaporation rate – instead, the evaporation rate can have any value below the threshold where the probability density goes to zero.

The distributions of some of the other evaporation rates depend strongly on the ammonia concentration assumed for the low-ammonia experiments. For instance, an ammonia concentration of 10 or 20 ppt (corresponding to the case where the

10 ammonia concentrations were treated as free parameters) would require the $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_3 \cdot \text{NH}_3$ cluster to have an ammonia

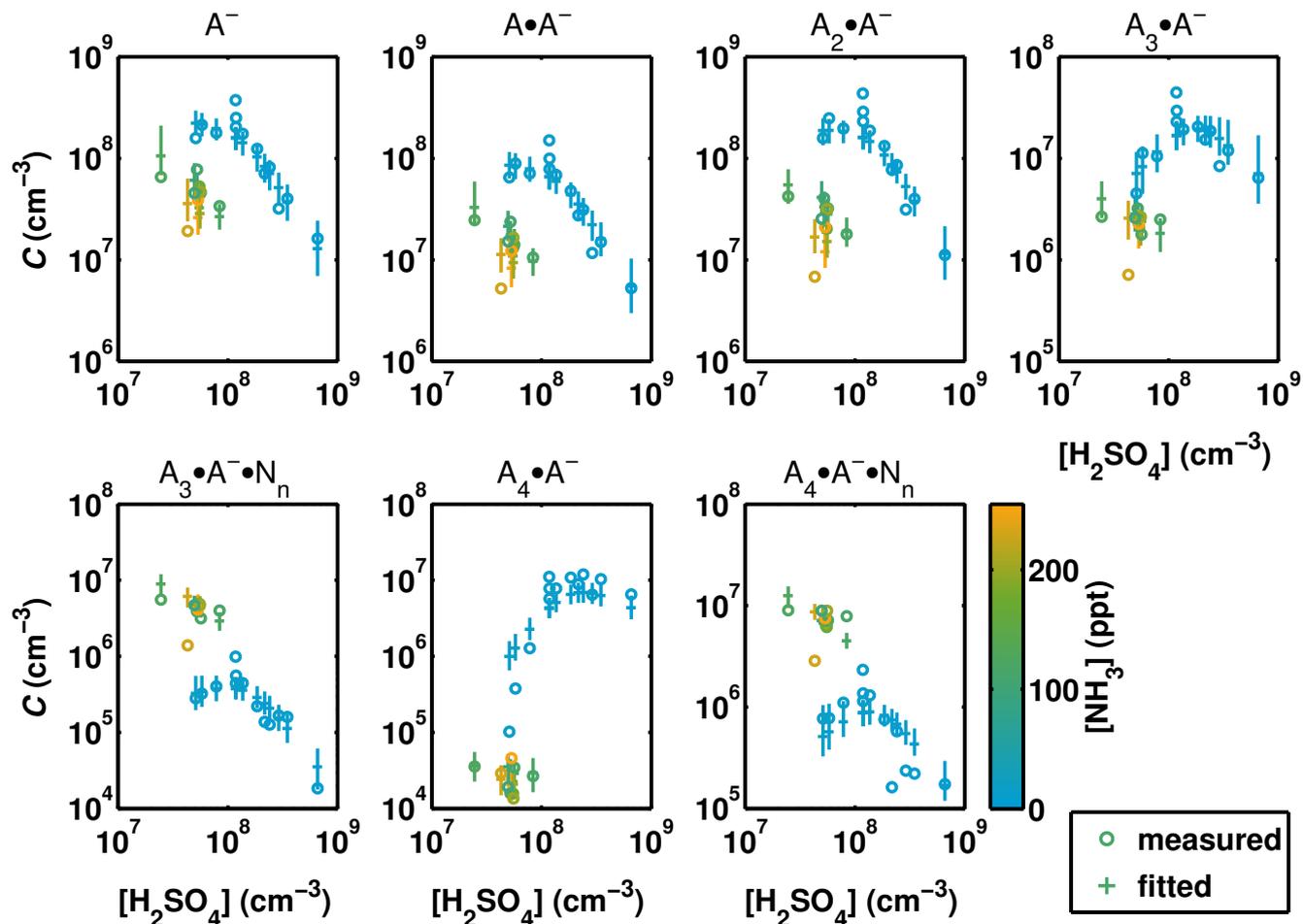


Figure 3. Cluster distributions measured at CLOUD and the corresponding modeled cluster concentrations from an MCMC simulation where evaporation rates, fragmentation probabilities, the ion production rate and the wall loss rate are varied. A stands for H_2SO_4 , A^- for HSO_4^- and N for NH_3 .

evaporation rate of about 200 s^{-1} in order for the enough pure sulfuric acid tetramers $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_3$ to be observed, while the evaporation rate would need to be well below 1 s^{-1} if the ammonia concentration was instead 1 ppt. A similar pattern is observed for some of the other ammonia evaporation rates, and interdependencies between the different evaporation rates lead to the posterior distributions of some sulfuric acid evaporation rates also depending on how the background ammonia concentration is treated

For the two cases where the background ammonia concentration is set to a fixed value, some of the posterior distributions consist of several peaks (see Fig. 4). As described in more detail in the Supplementary Material, the MCMC results can in fact be divided into two or three separate solutions, respectively, for the cases with background ammonia concentrations of 1 ppt

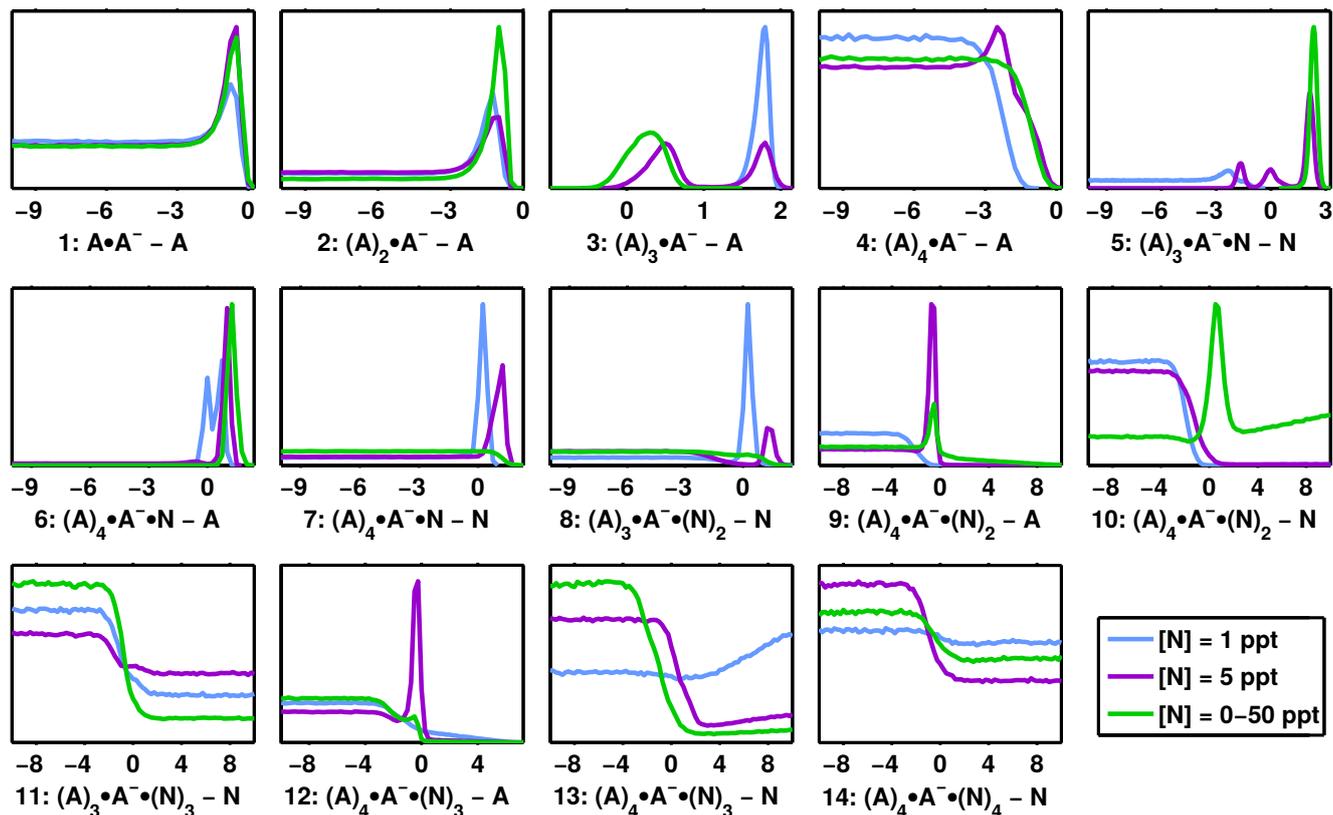


Figure 4. Posterior distributions of the base 10 logarithm of the evaporation rates (in units of s^{-1}) corresponding to the experimental cluster distributions and different options for treating the background ammonia concentration in the experiments where it was below the detection limit and therefore unknown. A stands for H_2SO_4 , A^- for HSO_4^- and N for NH_3 .

and 5 ppt. These alternative solutions correspond to different cluster types being stable and unstable, but they all still give an equally good fit to the measured cluster distributions. For instance, when assuming a background ammonia concentration of 5 ppt, the posterior distribution of parameter number 5 shows three separate peaks (purple line in Fig. 4). Looking only at the sets of parameter values in the right hand side peak, it can be noted that the value of parameter number 3 corresponds always to the left hand side peak of this distribution (see Fig. S14 of the Supplementary Material), and parameter number 8 always has a low value due to correlations between the different parameters. This set of ranges for the parameter values is denoted as solution (E), and similarly the two other peaks in the distribution of parameter number 5 correspond to solutions (C) and (D). The observation that the distributions can be divided into separate solutions in this way implies that, for instance, either an evaporation rate of $100 s^{-1}$ for ammonia from the $HSO_4^- \cdot (H_2SO_4)_3 \cdot NH_3$ cluster and an evaporation rate of $3 s^{-1}$ of the pure sulfuric acid tetramer or an evaporation rate of $0.2 s^{-1}$ for ammonia from the $HSO_4^- \cdot (H_2SO_4)_3 \cdot NH_3$ cluster and an evaporation rate of $60 s^{-1}$ of the pure sulfuric acid tetramer could produce a good fit to the experimental cluster distributions,

[NH ₃] in MCMC	1 ppt (A)	1 ppt (B)	5 ppt (C)	5 ppt (D)	5 ppt (E)	0–5 ppt	QC
1: A · A ⁻ – A	<1	<0.6	<1	<1	<1	<1	8 × 10 ⁻¹⁸
2: A ₂ · A ⁻ – A	<0.2	<0.2	<0.2	<0.2	<0.3	<0.3	2 × 10 ⁻⁴
3: A ₃ · A ⁻ – A	60 (20–90)	60 (30–90)	60 (20–100)	60 (8–100)	3 (0.5–20)	2 (0.4–7)	1
4: A ₄ · A ⁻ – A	<.01	<0.03	<0.03	<0.06	<0.6	<0.3	200
5: A ₃ · A ⁻ · N – N	<0.2	<0.03	0.02 (<0.1)	1 (0.1–20)	100 (20–600)	200 (20–800)	2
6: A ₄ · A ⁻ · N – A	1 (<2)	5 (2–10)	8 (<20)	<3	10 (<30)	20 (3–100)	0.08
7: A ₄ · A ⁻ · N – N	2 (<6)	<3	6 (1–30)	20 (6–60)	<20	<60	6 × 10 ⁻⁴
8: A ₃ · A ⁻ · N ₂ – N	<10	2 (<100)	20 (<200)	<30	<2	<60	0.5
9: A ₄ · A ⁻ · N ₂ – A	<0.1	<0.2	–	<1	<1	–	0.002
10: A ₄ · A ⁻ · N ₂ – N	<0.1	<0.3	–	<10	<0.6	–	0.01
11: A ₃ · A ⁻ · N ₃ – N	–	–	–	–	–	–	200
12: A ₄ · A ⁻ · N ₃ – A	< 10 ⁶	< 10 ⁶	<6 × 10 ⁵	<6 × 10 ⁵	<3	<3 × 10 ⁴	3 × 10 ⁻⁹
13: A ₄ · A ⁻ · N ₃ – N	–	–	–	–	–	–	3 × 10 ⁻⁴
14: A ₄ · A ⁻ · N ₄ – N	–	–	–	–	–	–	9 × 10 ⁸

Table 1. Evaporation rates corresponding to the three MCMC simulations presented in Fig. 4. The cases where the background ammonia is set to a fixed value of 1 ppt or 5 ppt are divided into two and three alternative solutions, respectively, denoted as (A)–(E). (See the Supplementary Material for more details.) For parameters that have a posterior distribution with a clear peak and practically zero probability density elsewhere, the location of the peak is given together with the range of possible values in parentheses. In many cases only an upper limit can be determined, and some rates cannot be determined at all (–). The last column presents quantum-chemistry based evaporation rates for comparison. In the reactions, A stands for H₂SO₄, A⁻ for HSO₄⁻ and N for NH₃.

but an evaporation rate of 100 s⁻¹ for ammonia from the HSO₄⁻ · (H₂SO₄)₃ · NH₃ cluster and an evaporation rate of 60 s⁻¹ of the pure sulfuric acid tetramer would not reproduce the data. For the simulations with an ammonia concentration of 1 ppt, the separate solutions (A) and (B) correspond to the two peaks in the distribution of coefficient number 6.[-:]

The estimates extracted for the evaporation rates from the MCMC simulations are presented in Table 1. As discussed above, only an upper limit can be determined for some evaporation rates, and it should be noted that the actual value could equally well be just below this limit or several orders of magnitude lower. For example, the rate at which HSO₄⁻ · H₂SO₄ dimers are lost through collisions with neutral sulfuric acid molecules is between 0.04 and 1.3 s⁻¹ in the different experiments. If the

evaporation rate of the dimer is lower than this, the dimers will practically never evaporate before colliding with an H_2SO_4 . If the evaporation process never happens, its rate cannot be expected to be determined based on the measurements. In order to constrain these low evaporation rates more tightly, experiments with very low but well quantified precursor concentrations would be needed, resulting in a lower rate for the competing growth process, but external losses and collisions with positive ions would still limit the range of evaporation rates that can be determined.

For certain evaporation rates, a distinct peak is observed in the posterior distribution. Also in this case it should be kept in mind that the true value could be anywhere within the width of the peak. As can be expected, all these well constrained evaporation rates are in the intermediate range, mostly between 1 and 100, where growth by collisions does not completely overwhelm the evaporation process, but the cluster is not so unstable that it would never collide and grow further. These clusters probably correspond to rate limiting steps on the main formation pathway.

Some of the parameters have posterior distributions with a non-zero probability density over the whole range. Some of these evaporation processes occur between clusters that are grouped together in the cluster distribution, and others are perhaps not on the main formation pathway. In any case, they do not have a strong impact on how well the modeled concentrations fit to the experimental data, and their values are therefore not constrained.

Also evaporation rates estimated from quantum chemical Gibbs free energies (Ortega et al., 2014; Almeida et al., 2013) are presented in Table 1 for comparison. The theoretical evaporation rates have an uncertainty of one or two orders of magnitude, as they depend exponentially on the stepwise cluster formation energies, which have an uncertainty of 1–2 kcal mol⁻¹. For the three smallest pure sulfuric acid clusters, $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_{1-3}$, the quantum chemistry-based evaporation rates are in good agreement with the values determined from analyzing the experimental data. The pure acid pentamer $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_4$, on the other hand, is predicted by quantum chemistry to have an evaporation lifetime of only 5 ms, while the analysis of the experimental data suggests that it has a very low evaporation rate (and hence a very long evaporation lifetime). In case of the ammonia-containing clusters, the MCMC simulations with different options concerning the background ammonia concentration, as well as the different alternative solutions from the simulations, give different ranges of most likely values of the evaporation rates, some of which agree better and some worse with the theoretical estimates.

4.3 Estimating fragmentation probabilities

The probabilities of fragmentation processes that might occur in the inlet of the mass spectrometer were varied separately from the evaporation rates, as the process involved is different: the evaporation rates discussed in the previous section correspond to molecules evaporating spontaneously from the cluster at atmospheric pressure and a temperature of 273 K, while fragmentation in the inlet occurs when the ionic clusters are accelerated and experience high-energy collisions with neutral carrier gas molecules. In reality, the two concepts are not totally unrelated, as both processes depend on the binding energy of the cluster, but the fragmentation probability is also likely to depend on the number of vibrational degrees of freedom that can absorb energy from the collision. As the different factors determining the fragmentation probability, and even the exact conditions inside the APi-TOF inlet, remain unclear, all fragmentation probabilities were varied freely.

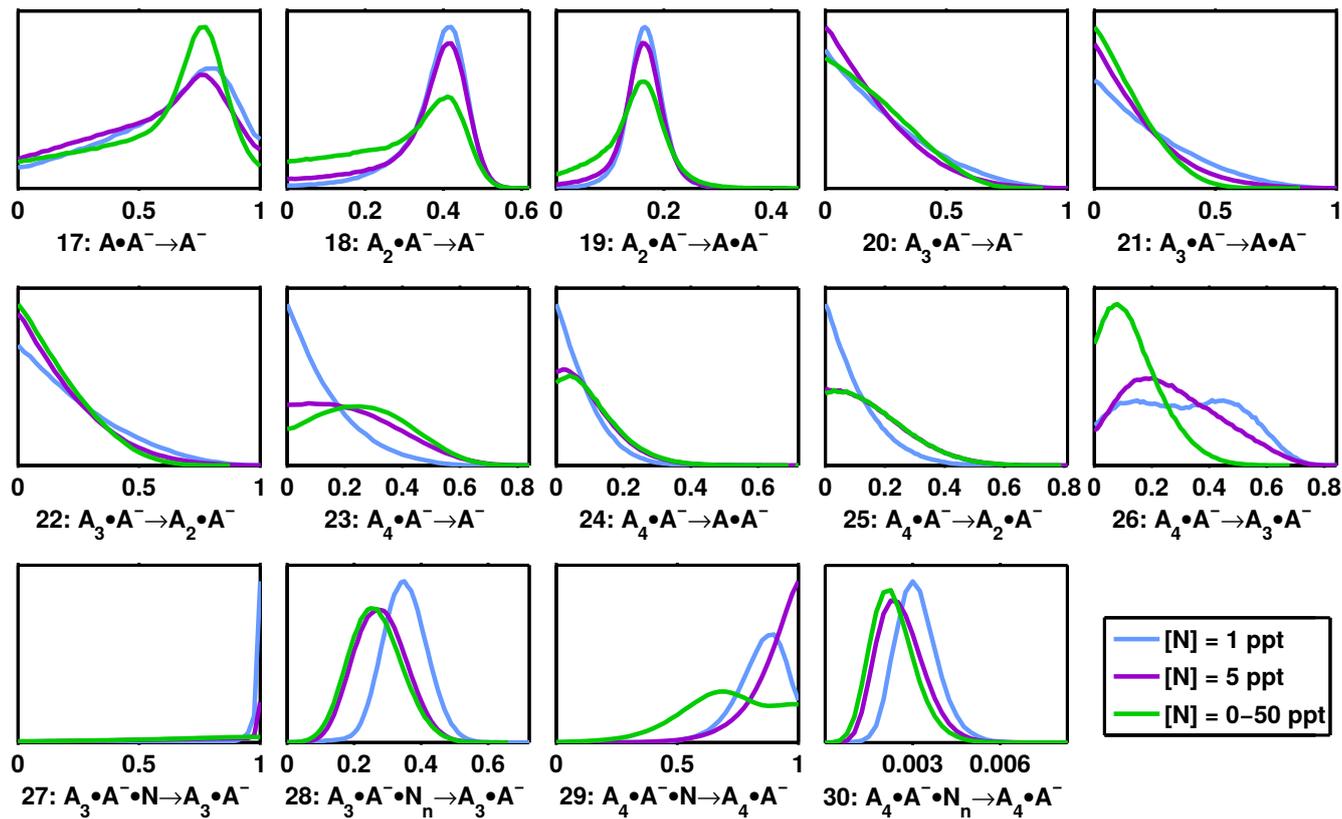


Figure 5. Posterior distributions of the fragmentation probabilities in the mass spectrometer inlet corresponding to the experimental cluster distributions and different options for treating the background ammonia concentration in the experiments where it was below the detection limit and therefore unknown. A stands for H_2SO_4 , A^- for HSO_4^- and N for NH_3 .

Figure 5 shows posterior distributions for the studied fragmentation probabilities. For the larger pure acid clusters $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_{2,4}$, several different fragmentation processes are considered, and their probabilities are presented separately in Fig. 5.

The **posterior distributions of the overall fragmentation probabilities (that is the sums of the probabilities of all fragmentation processes in which a given cluster can be lost)** of these clusters are shown in Fig. 6. For the MCMC simulations with a

5 fixed background ammonia concentration, the distributions corresponding to the alternative solutions (see previous Section) are shown in Sect. S3 of the Supplementary Material.

The posterior distribution of the dimer fragmentation probability is spread over the whole range from no fragmentation to 100% fragmentation. While there is a peak close to 70%, the possibility of dimers not fragmenting at all (which seems likely based on earlier experimental and theoretical evidence of the dimer being extremely stable) is not ruled out.

10 The trimers are found to fragment to some extent, producing both monomers and dimers. Assuming that both dimers and trimers have very low evaporation rates, but the tetramer is not very stable, the HSO_4^- ions that are formed from charging H_2SO_4

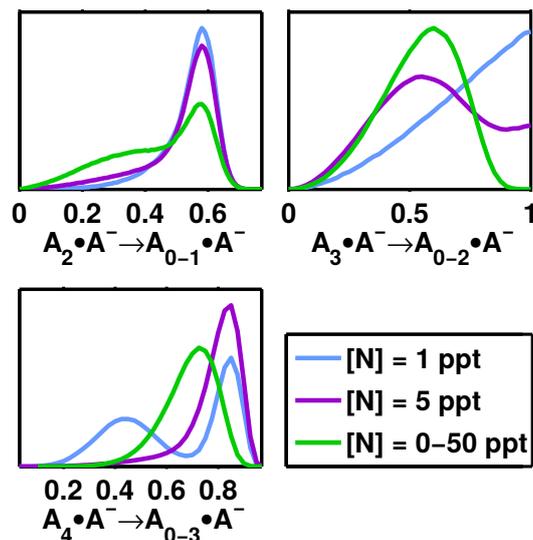


Figure 6. Posterior distributions of the total fragmentation probabilities of the $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_{2-4}$ clusters corresponding to the experimental cluster distributions and different options for treating the background ammonia concentration in the experiments where it was below the detection limit and therefore unknown. A stands for H_2SO_4 and A^- for HSO_4^- .

molecules will quickly gain first one and then a second acid molecule and, as the next growth step is slower, accumulate to form a high concentration of trimers. If a notable fraction of these trimers fragments both into monomers and dimers, most of the monomers and dimers that are detected may actually be fragmentation products from trimers, as was also observed experimentally by Adamov et al. (2013). This would mean that the actual concentrations of negatively charged monomers and dimers in the chamber cannot be measured, preventing the accurate determination of the dimer evaporation rate and fragmentation probability. This scenario is in good agreement with the observations that the dimer fragmentation probability cannot be determined and only a relatively high upper limit can be found for the dimer evaporation rate.

Also the pure acid tetramers and pentamers fragment, possibly even more than the trimers, but it cannot be determined which fragmentation pathways are most important. A large fraction of the $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_{3-4} \cdot \text{NH}_3$ clusters probably lose the ammonia molecule before detection, although the exact shape of the posterior distributions depends on how the low ammonia concentrations are treated in the MCMC simulation. The results for the probability of the clusters containing two or more ammonia molecules losing all of them, on the other hand, is almost independent of the simulation options, and only a small fraction of these clusters are detected as pure acid clusters.

5 Conclusions

A Markov chain Monte Carlo (MCMC) approach is presented for determining evaporation rates from measured cluster distributions. The time evolution of the cluster population is described by birth-death equations and solved numerically. The values

of the collision and evaporation rates are varied, and the obtained cluster distributions are compared to the measurements. In addition to the evaporation rates, also several other poorly known parameters related to the experimental setup are varied. The method is applied to concentration distributions of negatively charged sulfuric acid–ammonia clusters measured in the CLOUD chamber in CERN.

5 Of the pure sulfuric acid ion clusters $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_{1-4}$, the dimer, trimer and pentamer are found to be very stable, while the tetramer has a higher evaporation rate and may correspond to a rate limiting step in the cluster formation process. The stability of the dimer and trimer and the instability of the tetramer are consistent with cluster formation energies calculated with different quantum chemical methods (Ortega et al., 2014; Herb et al., 2013) and with semi-empirical estimates combining measurements and quantum chemistry (Lovejoy and Curtius, 2001; Curtius et al., 2001). However, the low evaporation rate
10 of the pure acid pentamer is in contradiction with the computational and semi-empirical cluster energies (Ortega et al., 2014; Lovejoy and Curtius, 2001; Curtius et al., 2001). On the other hand, these previously determined cluster energies correspond to dry clusters, and hydration is likely to stabilize clusters at least to some extent. **It is, in principle, possible that the pentamer could have a very stable hydrated structure, while the tetramer would only be moderately stabilized by hydration.** Furthermore, evaporation rates calculated based on cluster formation energies involve the assumption that the evaporation process proceeds
15 directly from the minimum energy configuration of the initial cluster to the minimum energy configuration of the product cluster. In reality, the process is likely to require some reorganization of the molecules and might have an energy barrier that slows down the evaporation. **Finally, the apparent stability of the pentamer might also be an artifact caused by the finite system size in the simulations.**

The results are more ambiguous for the ammonia containing clusters. The MCMC simulations produce several alternative
20 sets of evaporation rates that all provide an equally good fit to the experimental cluster distributions. This inconclusiveness stems at least partly from the choice of ammonia concentrations used in the set of experiments. In more than half of the experiments, the ammonia concentrations are in an unknown narrow range below the detection limit of 35 ppt, while the other runs have ammonia concentrations in a second narrow range from 100 to 250 ppt. Repeating the MCMC simulations with a new set of experimental cluster distributions measured at ammonia concentrations distributed evenly over a wide range would
25 most probably narrow down the estimates for many of the evaporation rates.

The observation that several alternative sets of parameter values can produce a good fit to the same experimental data highlights the risk in using a simplified cluster model with only one or two fitting parameters as was done by Jen et al. (2014) and Kürten et al. (2015). While the model may give a good fit to the observations, the corresponding set of evaporation rates may be only one out of several solutions, and does not necessarily correspond to the true evaporation rates.

30 Another important finding is that fragmentation in the inlet of a APi-TOF mass spectrometer may have a significant effect on the observed cluster distribution. The amount of fragmentation depends on the type of inlet that is used, and also the specific voltages and other settings that are used. However, if it is not possible to suppress fragmentation completely for some instrument type or experimental setup, it is important at least to gain some understanding of the fragmentation processes, and MCMC analysis appears to be a suitable tool for this. **In this study, the mass spectrometer was assumed to have been**

calibrated so that there was no mass discrimination, but in the future, also the mass dependent transmission efficiency of mass spectrometers could be studied using MCMC analysis.

While definitive values could not yet be obtained for all evaporation rates, the MCMC approach is shown to be a promising new tool for analyzing cluster concentration measurements. It can give valuable information about cluster evaporation processes that cannot be observed directly. However, enough experimental data measured over a wide range of all precursor concentrations are needed in order to draw clear conclusions. All details related to the experimental setup must be mimicked as closely as possible in the simulations in order for the fitting parameters to have a clear physical meaning. Furthermore, as cluster formation is inherently a dynamical process, the MCMC analysis would be more efficient for datasets of cluster concentrations as a function of time, instead of the steady-state distributions used here. This would also enable the fitting of collision rate constants in addition to evaporation rates.

Acknowledgements. I would like to thank CSC – IT Center for Science Ltd for computer resources, the Vilho, Yrjö and Kalle Väisälä Foundation and the European Research Council (project ERC-StG 257360-MOCAPAF) for funding, and Prof. Hanna Vehkamäki, Dr. Tinja Olenius and Prof. Heikki Haario for useful discussions and comments regarding the manuscript.

References

- Adamov, A., Junninen, H., Duplissy, J., Sipilä, M., Kulmala, M., and CLOUD collaboration: Cluster measurements at CLOUD using a high resolution ion mobility spectrometer-mass spectrometer combination, *AIP Conference Proceedings*, 1527, 350 – 353, doi:10.1063/1.4803275, 2013.
- 5 Almeida, J., Schobesberger, S., Kürten, A., Ortega, I. K., Kupiainen-Määttä, O., Praplan, A. P., Adamov, A., Amorim, A., Bianchi, F., Breitenlechner, M., David, A., Dommen, J., Donahue, N. M., Downard, A., Dunne, E., Duplissy, J., Ehrhart, S., Flagan, R. C., Franchin, A., Guida, R., Hakala, J., Hansel, A., Heinritzi, M., Henschel, H., Jokinen, T., Junninen, H., Kajos, M., Kangasluoma, J., Keskinen, H., Kupc, A., Kurtén, T., Kvashin, A. N., Laaksonen, A., Lehtipalo, K., Leiminger, M., Leppä, J., Loukonen, V., Makhmutov, V., Mathot, S., McGrath, M. J., Nieminen, T., Olenius, T., Onnela, A., Petäjä, T., Riccobono, F., Riipinen, I., Rissanen, M., Rondo, L., Ruuskanen, T., Santos, 10 F. D., Sarnela, N., Schallhart, S., Schnitzhofer, R., Seinfeld, J. H., Simon, M., Sipilä, M., Stozhkov, Y., Stratmann, F., Tomé, A., Tröstl, J., Tsagkogeorgas, G., Vaattovaara, P., Viisanen, Y., Virtanen, A., Vrtala, A., Wagner, P. E., Weingartner, E., Wex, H., Williamson, C., Wimmer, D., Ye, P., Yli-Juuti, T., Carslaw, K. S., Kulmala, M., Curtius, J., Baltensperger, U., Worsnop, D. R., Vehkamäki, H., and Kirkby, J.: Molecular understanding of sulphuric acid–amine particle nucleation in the atmosphere, *Nature*, 502, 359 – 363, doi:10.1038/nature12663, 2013.
- 15 Benson, D. R., Yu, J. H., Markovitch, A., and Lee, S.-H.: Ternary homogeneous nucleation of H₂SO₄, NH₃, and H₂O under conditions relevant to the lower troposphere, *Atmos. Chem. Phys.*, 11, 4755 – 4766, doi:10.5194/acp-11-4755-2011, 2011.
- Berndt, T., Stratmann, F., Sipilä, M., Vanhanen, J., Petäjä, T., Mikkilä, J., Grüner, A., Spindler, G., Lee Mauldin III, R., Curtius, J., Kulmala, M., and Heintzenberg, J.: Laboratory study on new particle formation from the reaction OH + SO₂: Influence of experimental conditions, H₂O vapour, NH₃ and the amine tert-butylamine on the overall process, *Atmos. Chem. Phys.*, 10, 7101 – 7116, doi:10.5194/acp-10-7101- 20 2010, 2010.
- Bianchi, F., Praplan, A. P., Sarnela, N., Dommen, J., Kürten, A., Ortega, I. K., Schobesberger, S., Junninen, H., Simon, M., Tröstl, J., Jokinen, T., Sipilä, M., Adamov, A., Amorim, A., Almeida, J., Breitenlechner, M., Duplissy, J., Ehrhart, S., Flagan, R. C., Franchin, A., Hakala, J., Hansel, A., Heinritzi, M., Kangasluoma, J., Keskinen, H., Kim, J., Kirkby, J., Laaksonen, A., Lawler, M. J., Lehtipalo, K., Leiminger, M., Makhmutov, V., Mathot, S., Onnela, A., Petäjä, T., Riccobono, F., Rissanen, M. P., Rondo, L., Tomé, A., Virtanen, A., Viisanen, Y., 25 Williamson, C., Wimmer, D., Winkler, P. M., Ye, P., Curtius, J., Kulmala, M., Worsnop, D. R., Donahue, N. M., and Baltensperger, U.: Insight into Acid–Base Nucleation Experiments by Comparison of the Chemical Composition of Positive, Negative, and Neutral Clusters, *Environmental Science & Technology*, 48, 13 675 – 13 684, doi:10.1021/es502380b, 2014.
- Brooks, S., Gelman, A., Jones, G. L., and Meng, X.-L., eds.: *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*, Chapman & Hall/CRC, Boca Raton, FL, 2011.
- 30 Brown, P., Byrne, G., and Hindmarsh, A.: VODE: A Variable-Coefficient ODE Solver, *SIAM J. Sci. and Stat. Comput.*, 10, 1038 – 1051, doi:10.1137/0910062, 1989.
- Bzdek, B. R., Ridge, D. P., and Johnston, M. V.: Amine exchange into ammonium bisulfate and ammonium nitrate nuclei, *Atmos. Chem. Phys.*, 10, 3495, doi:10.5194/acp-10-3495-2010, 2010.
- Chen, H., Ezell, M. J., Arquero, K. D., Varner, M. E., Dawson, M. L., Gerber, R. B., and Finlayson-Pitts, B. J.: New particle formation and 35 growth from methanesulfonic acid, trimethylamine and water, *Phys. Chem. Chem. Phys.*, 17, 13 699 – 13 709, doi:10.1039/C5CP00838G, 2015.

- Cox, R. A.: Some experimental observations of aerosol formation in the photo-oxidation of sulphur dioxide, *J. Aerosol Sci*, 4, 473 – 483, doi:10.1016/0021-8502(73)90139-0, 1973.
- Curtius, J., Froyd, K. D., and Lovejoy, E. R.: Cluster Ion Thermal Decomposition (I): Experimental Kinetics Study and ab Initio Calculations for $\text{HSO}_4^- (\text{H}_2\text{SO}_4)_x (\text{HNO}_3)_y$, *J. Phys. Chem. A*, 105, 10 867 – 10 873, doi:10.1021/jp0124950, 2001.
- 5 Doyle, G. J.: Self-Nucleation in the Sulfuric Acid-Water System, *J. Chem. Phys.*, 35, 795 – 799, doi:10.1063/1.1701218, 1961.
- Ehn, M., Junninen, H., Schobesberger, S., Manninen, H. E., Franchin, A., Sipilä, M., Petäjä, T., Kerminen, V.-M., Tammet, H., Mirme, A., Mirme, S., Hörrak, U., Kulmala, M., and Worsnop, D. R.: An Instrumental Comparison of Mobility and Mass Measurements of Atmospheric Small Ions, *Aerosol Sci. Technol.*, 45, 522 – 532, doi:10.1080/02786826.2010.547890, 2011.
- Eisele, F. L. and Hanson, D. R.: First Measurement of Prenucleation Molecular Clusters, *J. Phys. Chem. A*, 104, 830 – 836, doi:10.1021/jp9930651, 2000.
- 10 Herb, J., Xu, Y., Yu, F., and Nadykto, A. B.: Large Hydrogen-Bonded Pre-nucleation $(\text{HSO}_4^-)(\text{H}_2\text{SO}_4)_m(\text{H}_2\text{O})_k$ and $(\text{HSO}_4^-)(\text{NH}_3)(\text{H}_2\text{SO}_4)_m(\text{H}_2\text{O})_k$ Clusters in the Earth's Atmosphere, *J. Phys. Chem. A*, 117, 133 – 152, doi:10.1021/jp3088435, 2013.
- Israël, H.: Atmospheric Electricity, vol. I., Israel Program for Sci. Transl. & NSF, Jerusalem, 1970.
- 15 Jen, C. N., McMurry, P. H., and Hanson, D. R.: Stabilization of sulfuric acid dimers by ammonia, methylamine, dimethylamine, and trimethylamine, *J. Geophys. Res. Atmos.*, 119, 7502 – 7514, doi:10.1002/2014JD021592, 2014.
- Junninen, H., Ehn, M., Petäjä, T., Luosujärvi, L., Kotiaho, T., Kostianinen, R., Rohner, U., Gonin, M., Fuhrer, K., Kulmala, M., and Worsnop, D. R.: A high-resolution mass spectrometer to measure atmospheric ion composition, *Atmos. Meas. Tech.*, 3, 1039 – 1053, doi:10.5194/amt-3-1039-2010, 2010.
- 20 Kiang, C. S., Stauffer, D., Mohnen, V. A., Bricard, J., and Vigla, D.: Heteromolecular nucleation theory applied to gas-to-particle conversion, *Atmos. Environ.*, 7, 1279 – 1283, doi:10.1016/0004-6981(73)90137-6, 1973.
- Kirkby, J., Curtius, J., Almeida, J., Dunne, E., Duplissy, J., Ehrhart, S., Franchin, A., Gagné, S., Ickes, L., Kürten, A., Kupc, A., Metzger, A., Riccobono, F., Rondo, L., Schobesberger, S., Tsagkogeorgas, G., Wimmer, D., Amorim, A., Bianchi, F., Breitenlechner, M., David, A., Dommen, J., Downard, A., Ehn, M., Flagan, R. C., Haider, S., Hansel, A., Hauser, D., Jud, W., Junninen, H., Kreissl, F., Kvashin, A.,
- 25 Laaksonen, A., Lehtipalo, K., Lima, J., Lovejoy, E. R., Makhmutov, V., Mathot, S., Mikkilä, J., Minginette, P., Mogo, S., Nieminen, T., Onnela, A., Pereira, P., Petaja, T., Schnitzhofer, R., Seinfeld, J. H., Sipilä, M., Stozhkov, Y., Stratmann, F., Tomé, A., Vanhanen, J., Viisanen, Y., Vrtala, A., Wagner, P. E., Walther, H., Weingartner, E., Wex, H., Winkler, P. M., Carslaw, K. S., Worsnop, D. R., Baltensperger, U., and Kulmala, M.: Role of sulphuric acid, ammonia and galactic cosmic rays in atmospheric aerosol nucleation, *Nature*, 476, 429, doi:10.1038/nature10343, 2011.
- 30 Kupiainen-Määttä, O., Olenius, T., Kurtén, T., and Vehkamäki, H.: CIMS sulfuric acid detection efficiency enhanced by amines due to higher dipole moments - a computational study, *J. Phys. Chem. A*, 117, 14 109–14 119, doi:10.1021/jp4049764, 2013.
- Kupiainen-Määttä, O., Henschel, H., Kurtén, T., Loukonen, V., Olenius, T., Paasonen, P., and Vehkamäki, H.: Comment on 'Enhancement in the production of nucleating clusters due to dimethylamine and large uncertainties in the thermochemistry of amine-enhanced nucleation' by Nadykto et al., *Chem. Phys. Lett.* 609 (2014) 42–49, *Chem. Phys. Lett.*, 624, 107 – 110, doi:10.1016/j.cplett.2015.01.029, <http://www.sciencedirect.com/science/article/pii/S0009261415000391>, 2015.
- 35 Kürten, A., Münch, S., Rondo, L., Bianchi, F., Duplissy, J., Jokinen, T., Junninen, H., Sarnela, N., Schobesberger, S., Simon, M., Sipilä, M., Almeida, J., Amorim, A., Dommen, J., Donahue, N. M., Dunne, E. M., Flagan, R. C., Franchin, A., Kirkby, J., Kupc, A., Makhmutov, V., Petäjä, T., Praplan, A. P., Riccobono, F., Steiner, G., Tomé, A., Tsagkogeorgas, G., Wagner, P. E., Wimmer, D., Baltensperger, U.,

- Kulmala, M., Worsnop, D. R., and Curtius, J.: Thermodynamics of the formation of sulfuric acid dimers in the binary ($\text{H}_2\text{SO}_4\text{-H}_2\text{O}$) and ternary ($\text{H}_2\text{SO}_4\text{-H}_2\text{O-NH}_3$) system, *Atmospheric Chemistry and Physics*, 15, 10 701 – 10 721, doi:10.5194/acp-15-10701-2015, 2015.
- Kurtén, T., Kuang, C., Gómez, P., McMurry, P. H., Vehkamäki, H., Ortega, I. K., Noppel, M., and Kulmala, M.: The role of cluster energy nonaccommodation in atmospheric sulfuric acid nucleation, *J. Chem. Phys.*, 132, 024 304, doi:10.1063/1.3291213, 2010.
- 5 Loukonen, V., Bork, N., and Vehkamäki, H.: From Collisions to Clusters: First Steps of Sulfuric Acid Nanocluster Formation Dynamics, *Mol. Phys.*, 112, 1979 – 1986, doi:10.1080/00268976.2013.877167, 2014a.
- Loukonen, V., Kuo, I.-F. W., McGrath, M. J., and Vehkamäki, H.: On the stability and dynamics of (sulfuric acid) (ammonia) and (sulfuric acid) (dimethylamine) clusters: A first-principles molecular dynamics investigation, *Chem. Phys.*, 428, 164 – 174, doi:10.1016/j.chemphys.2013.11.014, 2014b.
- 10 Lovejoy, E. R. and Curtius, J.: Cluster Ion Thermal Decomposition (II): Master Equation Modeling in the Low-Pressure Limit and Fall-Off Regions. Bond Energies for $\text{HSO}_4^-(\text{H}_2\text{SO}_4)_x(\text{HNO}_3)_y$, *J. Phys. Chem. A*, 105, 10 874 – 10 883, doi:10.1021/jp012496s, 2001.
- McGrath, M. J., Olenius, T., Ortega, I. K., Loukonen, V., Paasonen, P., Kurtén, T., Kulmala, M., and Vehkamäki, H.: Atmospheric Cluster Dynamics Code: A flexible method for solution of the birth-death equations, *Atmos. Chem. Phys.*, 12, 2345 – 2355, doi:10.5194/acp-12-2345-2012, 2012.
- 15 Mirabel, P. and Katz, J. L.: Binary homogeneous nucleation as a mechanism for the formation of aerosols, *J. Chem. Phys.*, 60, 1138 – 1144, doi:10.1063/1.1681124, 1974.
- Olenius, T., Kupiainen-Määttä, O., Ortega, I. K., Kurtén, T., and Vehkamäki, H.: Free energy barrier in the growth of sulfuric acid-ammonia and sulfuric acid-dimethylamine clusters, *J. Chem. Phys.*, 139, 084 312, doi:10.1063/1.4819024, 2013a.
- Olenius, T., Schobesberger, S., Kupiainen-Määttä, O., Franchin, A., Junninen, H., Ortega, I. K., Kurtén, T., Loukonen, V., Worsnop, D. R., 20 Kulmala, M., and Vehkamäki, H.: Comparing simulated and experimental molecular cluster distributions, *Faraday Discuss.*, 165, 75–89, doi:10.1039/c3fd00031a, 2013b.
- Ortega, I. K., Olenius, T., Kupiainen-Määttä, O., Loukonen, V., Kurtén, T., and Vehkamäki, H.: Electrical charging changes the composition of sulfuric acid–ammonia/dimethylamine clusters, *Atmos. Chem. Phys.*, 14, 7995 – 8007, doi:10.5194/acp-14-7995-2014, 2014.
- Paasonen, P., Olenius, T., Kupiainen, O., Kurtén, T., Petäjä, T., Birmili, W., Hamed, A., Hu, M., Huey, L. G., Plass-Duelmer, C., Smith, J. N., 25 Wiedensohler, A., Loukonen, V., McGrath, M. J., Ortega, I. K., Laaksonen, A., Vehkamäki, H., Kerminen, V.-M., and Kulmala, M.: On the formation of sulphuric acid - amine clusters in varying atmospheric conditions and its influence on atmospheric new particle formation, *Atmos. Chem. Phys.*, 12, 9113 – 9133, doi:10.5194/acp-12-9113-2012, 2012.
- Su, T. and Chesnavich, W. J.: Parametrization of the ion-polar molecule collision rate constant by trajectory calculations, *J. Chem. Phys.*, 76, 5183 – 5185, doi:10.1063/1.442828, 1982.
- 30 ter Braak, C. J. F.: A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: Easy Bayesian computing for real parameter spaces, *Statistics and Computing*, 16, 239 – 249, doi:10.1007/s11222-006-8769-1, 2006.
- ter Braak, C. J. F. and Vrugt, J. A.: Differential Evolution Markov Chain with snooker updater and fewer chains, *Statistics and Computing*, 18, 435 – 446, doi:10.1007/s11222-008-9104-9, 2008.
- Zhao, J., Eisele, F. L., Titcombe, M., Kuang, C., and McMurry, P. H.: Chemical ionization mass spectrometric measurements of atmospheric 35 neutral clusters using the clusterCIMS, *J. Geophys. Res.*, 115, D08 205, doi:10.1029/2009JD012606, 2010.

Supplementary material for “A Monte Carlo approach for determining cluster evaporation rates from concentration measurements”

Oona Kupiainen-Määttä

University of Helsinki, Department of Physics, P.O. Box 64, FI-00014 University of Helsinki, Finland

This document contains a detailed description of the MCMC simulations (Sect. S1), a detailed discussion of the MCMC results for synthetic cluster distributions (Sect. S2), as well as supplementary figures related to the simulations discussed in the main text (Sect. S3).

S1. Details of the MCMC simulations

S1.1. Initialization

The DE-MC_Z algorithm (ter Braak and Vrugt, 2008) described in the main text requires an initial history chain \mathbf{Z}_0 . As the step sizes depend mostly on the history, \mathbf{Z}_0 must correspond roughly to the posterior distribution, which however is not known beforehand. The distribution of each parameter in the chain \mathbf{Z}_0 must be wide enough to encompass all local maxima in order for the DE-MC_Z algorithm to be able to jump from one maximum to another. In this study, this was accomplished by running a large number of short MCMC chains starting from random points in the allowed parameter space. On the other hand, the distribution must not be too wide, since then the proposed steps will be too long leading to a low acceptance probability. This was effected by discarding the chains that fit very poorly to the data.

Each simulation was started by running 40 separate MCMC chains of 5000 steps. The collision coefficients were sampled using the Metropolis algorithm as described in the main text, with the steps drawn from a normal distribution with a width σ_{step} of 0.05 times the width of the allowed range of each parameter.

At any step, only a fraction of the coefficients were varied. Each coefficient separately was chosen with a probability of 10 % to be varied at a given step, but if no coefficients were picked, the selection process was repeated. For a total of 43 coefficients, the average number of coefficients that were varied per step was ~ 4.3 , but at 6 % of the steps 8 or more coefficients were varied.

The acceptance probability was computed with $\sigma_0 = 0.2$. The first half of each chain was discarded, and during the second half the lowest value of the square sum, $SS_{\min,i}$, was recorded for each chain i . The second halves of the chains for which $SS_{\min,i}$ was lower than the geometric mean $(\prod_{j=1}^{40} SS_{\min,j})^{1/40}$ of all chains were combined and thinned to a total of 10000 points to form the initial history \mathbf{Z}_0 .

The overall lowest square sum SS_{\min} was also used for estimating the scatter in the measured data as $\sigma = (SS_{\min}/(n_{\text{out}} - n_{\text{coefs}}))^{1/2}$. This value was used for calculating the acceptance probability in the later stages of the simulation.

After combining the chains with lowest $SS_{\min,i}$ but before thinning the resulting chain, the variance $\sigma_{k,\text{ini}}^2$ of each parameter k was computed to give an estimate of the widths of the distributions. In the DE-MC_Z simulation, the term δ was drawn from a normal distribution with width $\sigma_{\delta,k} = 0.05 \times \sigma_{k,\text{ini}}$ for each parameter.

S1.2. Main simulation and convergence

The history \mathbf{Z} was initialized as the $10000 \times n_{\text{coefs}}$ matrix \mathbf{Z}_0 , and each of the five chains was started from a random point of the history. A varying number of coefficients were sampled simultaneously at each step as described above, now using the DE-MC_Z algorithm presented in the main text. At every 50th step, the parameter values of all chains were appended to the matrix \mathbf{Z} as five new rows.

The first 20000 steps of each chain were discarded as burn-in. After that, every 50th step was printed out. Monitoring of the convergence was started after the burn-in period. The mean and variance of all parameters were calculated separately for each chain and updated at every step. These were used to compute for each parameter i the \hat{R} statistic (Brooks and Gelman, 1998; Gelman and Rubin, 1992) $\hat{R}_i = \frac{n-1}{n} + \frac{m+1}{m} \frac{b_i}{W_i}$, where n is the number of steps in each chain, $m = 5$ is the number of chains, b_i is the variance of the means of parameter i in the different chains and W_i is the mean of the variances of parameter i in the different chains. A high value of \hat{R} means that the chains are far away from each other compared to the width of the distributions within each chain. Typically \hat{R} decreases with time as the chains cover the target distribution, but it can also increase again if different chains get stuck in different local maxima of the distribution. All simulations were run for a minimum of 10 000 000 steps per chain to ensure proper mixing and convergence. After that, the simulation was defined to have converged when $\hat{R} < 1.1$.

S2. Testing the method: analyzing synthetic cluster distributions

In order to test how reliably rate constants could be inferred from the MCMC simulation, artificial cluster distributions corresponding to known cluster evaporation rates were first produced using ACDC. The ion-molecule collision rates were computed according to the parameterization of Su and Chesnavich (1982), and evaporation rates were calculated based on quantum chemical cluster formation energies of dry clusters (Ortega et al., 2014; Almeida et al., 2013). The synthetic data set consisted of a total of 22 cluster distributions corresponding to the same sulfuric acid and ammonia vapor concentrations as in the experimental cluster distributions.

However, for the 13 experiments with no added ammonia, the ammonia concentration was not known. No quantitative estimates were available for the fragmentation probabilities, and also the estimates of the ion production rate and wall loss may not have been very accurate. For all of these parameters, the values corresponding to the peaks in the posterior distribution of the MCMC simulation analyzing the experimental data were used as input for producing the synthetic cluster

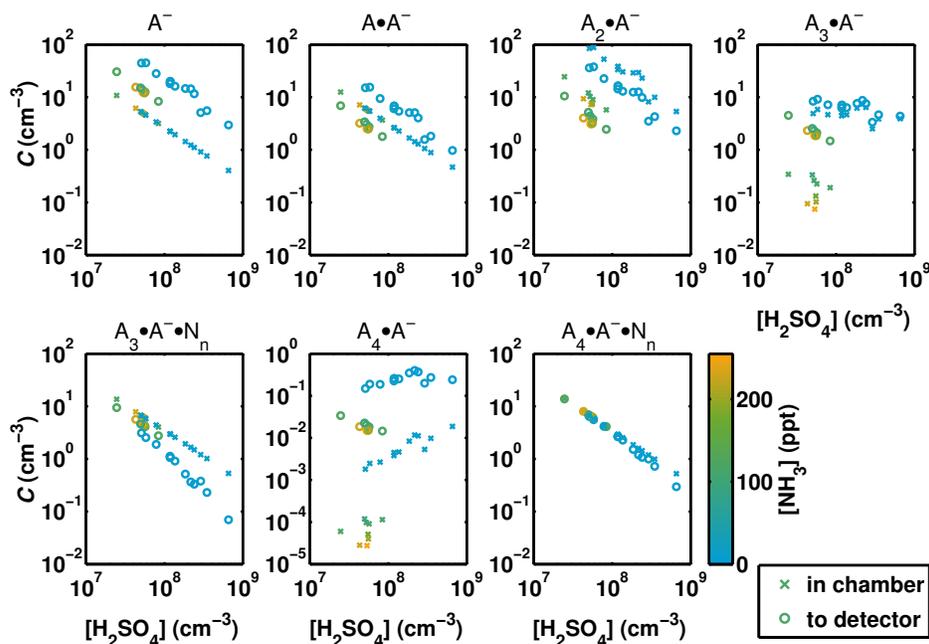


Figure S1: Synthetic cluster distributions before (x) and after (o) fragmentation assumed to happen in the APi-TOF. A stands for H_2SO_4 , A^- for HSO_4^- and N for NH_3 .

distribution. For consistency, all parameter values were taken from the same MCMC simulation where all possible parameters (evaporation rates, ion production rate, wall loss constant, fragmentation probabilities, background ammonia concentrations) were varied (Figs. 5, S11 and S12).

Figure S1 shows the synthetic cluster distributions before and after fragmentation. The distributions before fragmentation correspond to the actual simulated cluster concentrations in the CLOUD chamber, while the distributions after fragmentation correspond to how these concentrations would ideally be observed with the APi-TOF if there was no noise in the measurement.

To mimic measurement errors, random noise was added to the cluster distribution as $\log_{10} C_{\text{meas.}} = \log_{10} C_{\text{exact}} + \epsilon$, where C_{exact} is the exact concentration after fragmentation, $C_{\text{meas.}}$ is the 'measured' value and ϵ is drawn from a normal distribution with $\sigma_{\epsilon} = 0.2$. The 'measured' distribution is shown as crosses in Fig. S2 and is used as input in the MCMC simulations.

Two sets of MCMC simulations were performed for the synthetic 'measured' data: first varying all unknown parameters (evaporation rates, ion production rate, wall loss constant, fragmentation probabilities, background ammonia concentrations), and then setting the background ammonia concentrations to 5 ppt and only varying the other parameters. The output cluster distributions from these two cases are presented in Figs. S2 and S3, respectively.

The medians of each concentration from the output of the MCMC simulations are presented as a horizontal line, and the vertical lines show the 2.5th and 97.5th percentiles. The true model concentrations are well reproduced by the MCMC simulations, and especially the trends with respect to sulfuric acid and ammonia concentrations are correct. In cases where the 'measured' concentration contains a large 'measurement error' and is therefore far from the exact value, the MCMC simulation still mainly captures the true value rather than the inaccurate 'measured value' used as input for the MCMC simulation.

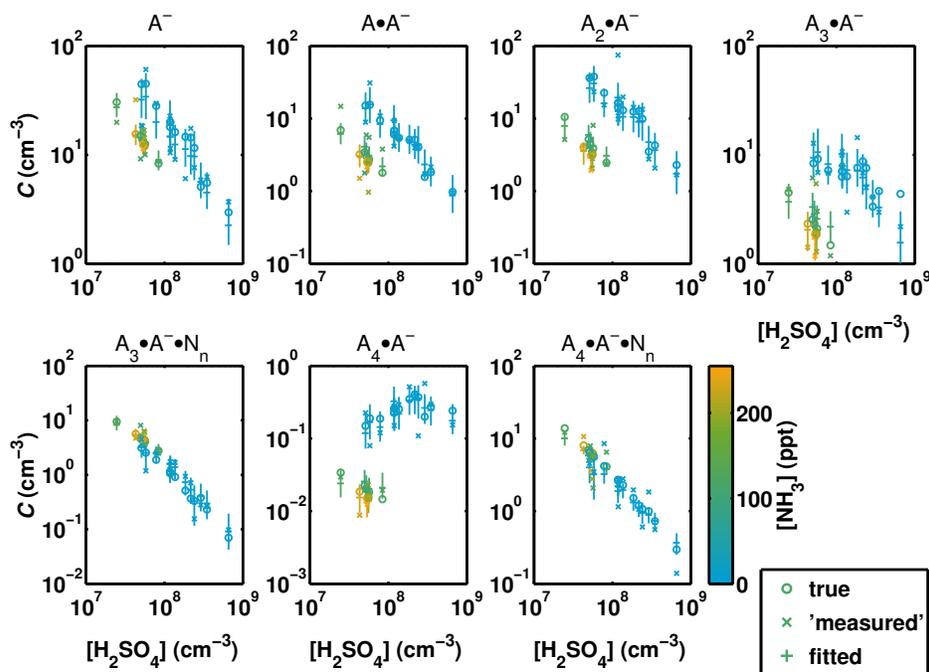


Figure S2: Synthetic cluster distributions after fragmentation (o) and with added noise (x), and the corresponding cluster concentrations from an MCMC simulation where evaporation rates, fragmentation probabilities, the ion production rate, the wall loss rate and the background ammonia concentrations are varied. A stands for H_2SO_4 , A^- for HSO_4^- and N for NH_3 .

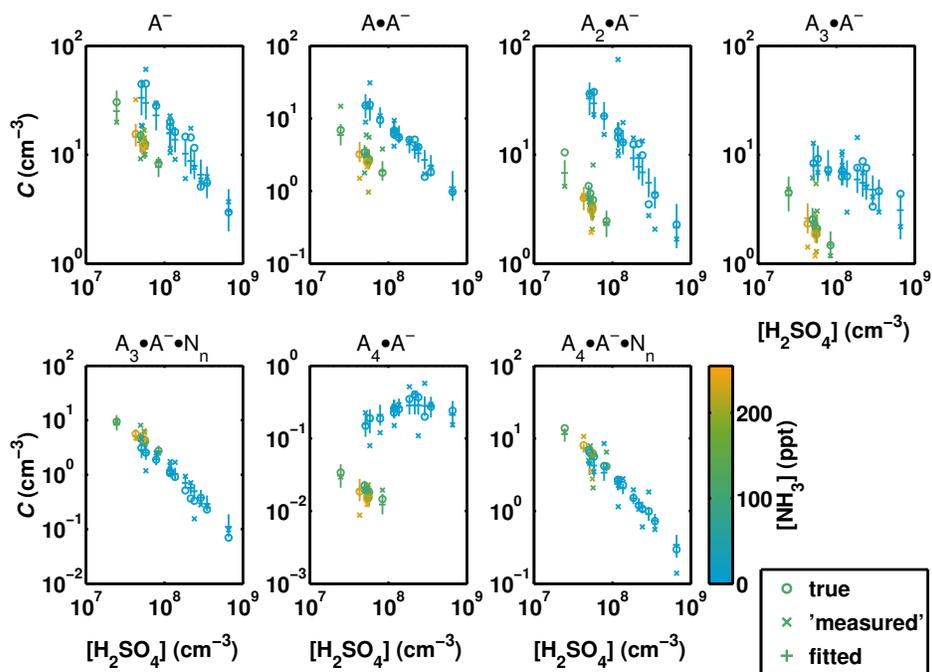


Figure S3: Same as Fig. S2, but all background ammonia concentrations are set to 5 ppt in the MCMC analysis instead of being varied.

S2.1. Extracting evaporation rates from the synthetic data

Figure S4 shows the posterior distributions of the coefficients corresponding to logarithms of the evaporation rates from the two sets of MCMC simulations. In order to interpret the results, it is useful first to consider how each evaporation rate affects the cluster distribution. In case of ammonia evaporation from clusters containing two or more ammonia molecules, both the evaporating cluster and the product contribute to the cluster distribution in the same way, as the clusters containing a given number of acid molecules are divided by ammonia content only into two groups, ammonia-free and ammonia-containing. Thus the value of the evaporation rate has no direct effect on the reported concentrations. On the other hand, ammonia evaporation from clusters containing one ammonia molecule converts the cluster from ammonia-containing to ammonia-free, and consequently alters the measured cluster concentration (unless the ammonia molecule would anyway be lost by fragmentation inside the instrument). Sulfuric acid evaporation rates also have a direct effect on the cluster distribution as clusters containing different numbers of acid molecules appear separately in the distribution. However, in case of acid evaporation from $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_4 \cdot (\text{NH}_3)_{1-3}$ clusters, both the evaporating cluster and the product appear in the measured distribution grouped with other clusters.

In case of the synthetic cluster distributions, the coefficients presented in Fig. S4 can roughly be divided into four categories according to the shape of the posterior distributions. A similar approach is also used in the main text for interpreting the MCMC simulations performed on the experimental data.

Coefficients number 3 and 5 have one or more clear peaks located near the input value of the parameter (shown as a vertical green line), and practically zero probability density elsewhere. Both of these evaporation coefficients have an input value close to 1 s^{-1} .

The second category is formed by the coefficients number 1, 2, 7 and, in the fixed background ammonia case, coefficient number 9. The posterior distributions of these coefficients have a constant non-zero probability density at low values, a constant practically zero probability density at high values, and between the two regions, close to the value 1 s^{-1} , either a sharp drop or a peak followed

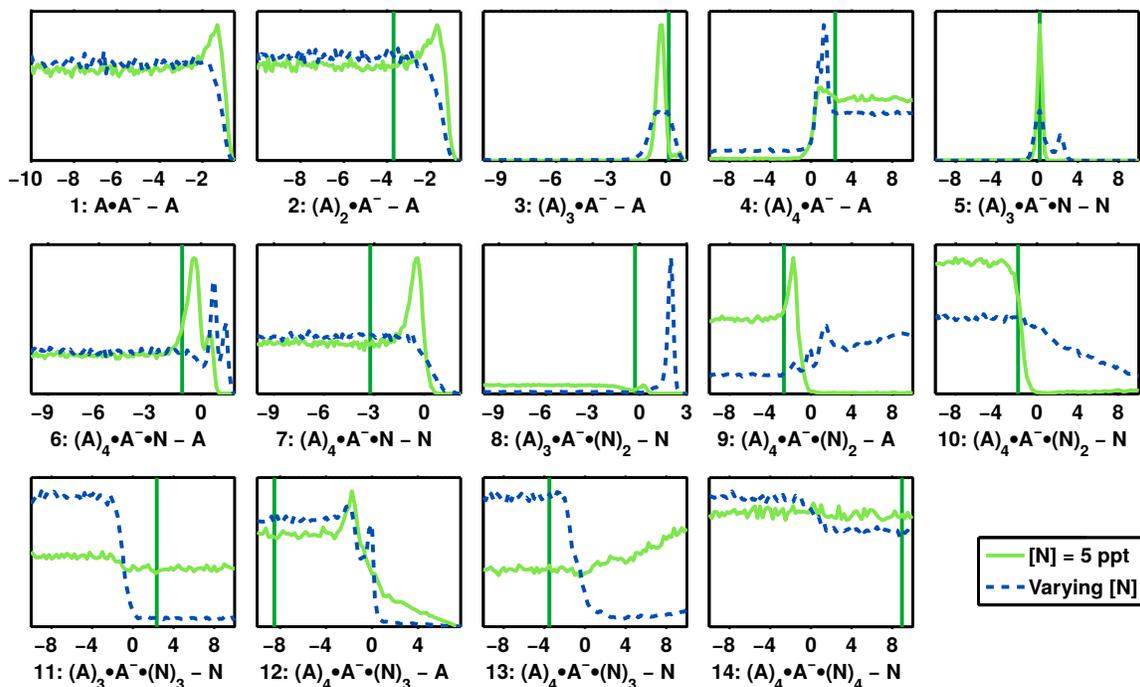


Figure S4: Posterior distributions of the base 10 logarithm of the evaporation rates (in units of s^{-1}) corresponding to the synthetic cluster distributions. The background ammonia concentrations have either been set to 5 ppt (green) or varied with MCMC (blue). The vertical green lines corresponds to the quantum chemistry-based estimates used as input parameter values for producing the synthetic cluster distributions. (For coefficient number 1, the input value -17.8 is outside the range sampled by MCMC.) A stands for H_2SO_4 , A^- for HSO_4^- and N for NH_3 .

by a sharp drop. The input values of these evaporation rates are well below 1 s^{-1} , but the precise values cannot be captured from the MCMC analysis. The uniform shape of the distributions at low evaporation rates is explained by the fact that if a cluster has a long enough lifetime with respect to evaporation, it will collide and grow before evaporating, and a further decrease of its evaporation rate does not change the situation. The peaks in the posterior distribution, when present, are mostly far from the corresponding input value of the coefficient, and the existence and prominence of the peaks also varies with the exact values of the random 'measurement errors' added to the cluster distributions. Therefore, the significance and interpretation of the peaks remain unclear.

Coefficients number 6, 8 and 12 also have a uniform non-zero probability density at low values and a zero probability at high values, but the distribution goes to zero at a higher evaporation rate value than in the previous category. In some cases there are also more than one peak. Similarly as in the second category, these evaporation rates have input values lower than 1 s^{-1} , but the exact value cannot be captured from the MCMC analysis.

The remaining coefficients form the fourth category. These have posterior distributions with a distinctly non-zero probability density over the whole range. These rate constants most probably do not have a strong impact on the cluster distribution, and can therefore have whatever value without interfering with the goodness of the fit. Some of the corresponding evaporation processes occur between clusters that are mutually indistinguishable in the cluster distribution, while others are perhaps not on the main formation pathway or at least do not correspond to rate-limiting processes.

S2.2. Extracting the ion production rate and wall loss constant from the synthetic data

The posterior distribution of the ion production rate (Fig. S5) peaks very close to the input value. However, in case of the synthetic data, the same value for the charging rate of sulfuric acid

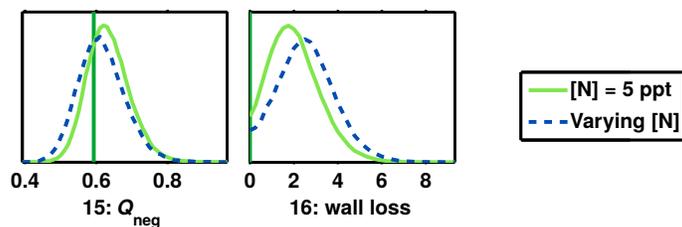


Figure S5: Posterior distributions of the ion production rate rates (in units of $\text{cm}^{-3}\text{s}^{-1}$) and wall loss constant (in units of 10^{-3}s^{-1}) corresponding to the synthetic cluster distributions. The background ammonia concentrations have either been set to 5 ppt (green) or varied with MCMC (blue). The vertical green lines corresponds to the MCMC estimates from the experimental data, which are used as input parameter values for producing the synthetic cluster distributions.

molecules by the produced charger ions is used both when producing the cluster distributions and analyzing them. For the experimental data, on the other hand, the true value is not known and the rate coefficient used in the MCMC simulation might not be accurate. Therefore, the promising result of Fig. S5, alone, does not prove how reliably the ion production rate can be determined from the experimental data.

The posterior distribution of the wall loss, on the other hand, does not peak at its input value. In fact, the MCMC result would seem to suggest a non-zero wall loss, as opposed to the input value of exactly zero. This serves as a reminder that the peak of the probability density cannot be directly interpreted as giving the true value of the corresponding parameter. Instead, the whole region of non-zero probability density should be interpreted as possible (and even likely) values for the parameter.

S2.3. Extracting the fragmentation probabilities from the synthetic data

The posterior distribution of the fragmentation probabilities are presented in (Fig. S6). For all parameters and both sets of simulations (either varying or not varying the background ammonia concentrations), the posterior distribution has a non-zero probability at the input parameter value. For the narrower distributions, the peak is mostly close to the input value, while especially for the simulations where the ammonia concentration is varied, some of the wider distributions peak quite far from the input value.

S2.4. Extracting the background ammonia concentrations from the synthetic data

Figure S7 shows the posterior distribution of the ammonia mixing ratios for the thirteen experiments where no ammonia was added intentionally and its concentration was below the detection limit. For each of these mixing ratios, the posterior distribution peaks at a value higher than the input mixing ratio. In most cases, the input value is at the very tail of the probability distribution. It must therefore be concluded that background ammonia concentrations cannot be reliably determined based on this kind of MCMC study.

S2.5. Options for treating the background ammonia concentrations

As the determination of the low ammonia concentrations was seen not to be reliable, the analysis of the experimental data in the main text concentrates mostly on MCMC simulations where the background ammonia concentration is not treated as a free parameter. In the case where the ammonia concentrations are varied with MCMC, the peak locations of the posterior distribution range between 6.4 and 18.3 ppt. Assuming that these values somewhat overestimate the background concentrations, a better estimate for the background concentration might be around 5 ppt or lower. In the MCMC simulations where the ammonia concentration is not allowed to vary, two fixed values are tested: either all background ammonia concentrations are set to 5 ppt, or in an other MCMC simulation they are all set to 1 ppt.

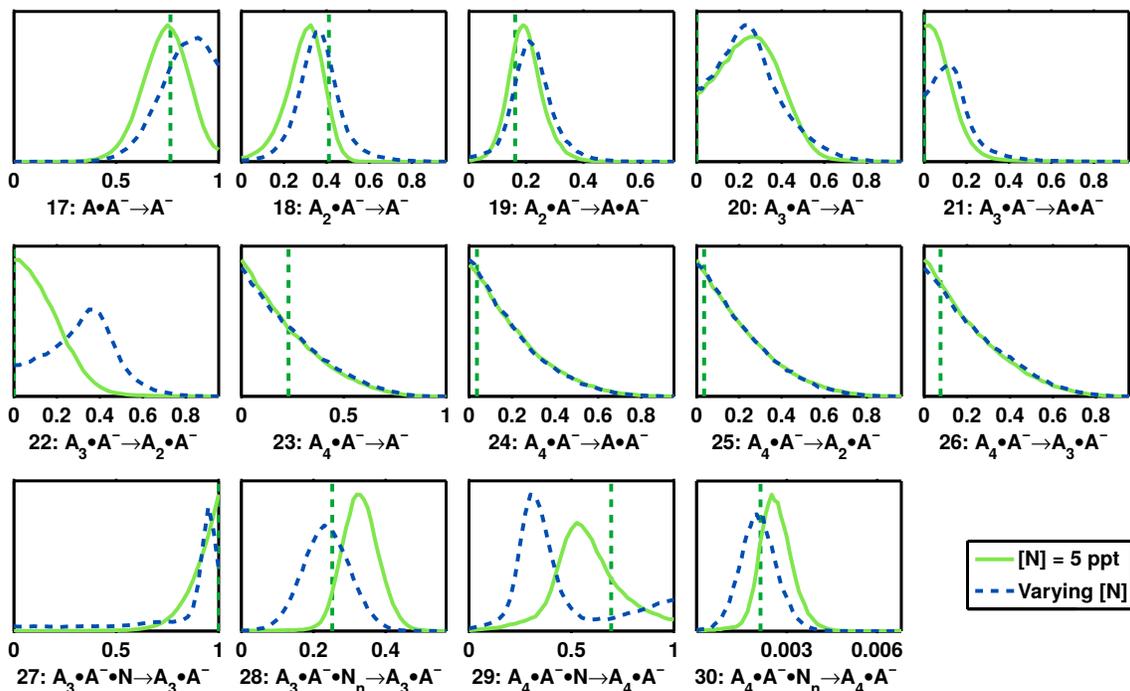


Figure S6: Posterior distributions of the fragmentation probabilities of clusters in the mass spectrometer inlet corresponding to the synthetic cluster distributions. The background ammonia concentrations have either been set to 5 ppt (green) or varied with MCMC (blue). The vertical green lines corresponds to the MCMC estimates from the experimental data, which are used as input parameter values for producing the synthetic cluster distributions. A stands for H_2SO_4 , A^- for HSO_4^- and N for NH_3 .

S2.6. Separating the posterior distributions into alternative scenarios

Ideally, the parameters varied in an MCMC simulation should all be uncorrelated. As a worst-case example of problems caused by correlated parameters, consider a case where the observed quantity depends on the ratio of two of the varied parameters. These might both have a uniform probability density over the whole allowed range, while the posterior distribution of their ratio might have a narrow peak. In this example case, the problem could easily be resolved by using as free parameters the ratio and product of the two parameters, instead of using the original parameters themselves.

With a large number of fitted parameters, also more complicated correlations involving several parameters can occur, and it may not, at least in practice, be possible to choose the parameters so as to eliminate all correlations. For a complicated enough system, there may even be several very different solutions that give an equally good fit to the data. In order to find and study the correlations between different parameters, the output from the MCMC simulations must be considered as a collection of *sets of parameter values* instead of seeing it as separate probability distribution for each parameter. The correlations can be found by grouping the sets of parameter values according to the value of one specific parameter, and comparing the posterior distributions of the other parameters for these groups. **If the groups of parameters form completely separate peaks also in the distributions of one or more of the other parameters, a correlation has been found.**

Figure S8 demonstrates the separation of the MCMC results into three scenarios for the synthetic cluster concentration data and a fixed background ammonia concentration of 5 ppt in the MCMC simulation. For clarity, the distributions of the scenarios are normalized to have the same overall probability.

First, it can be seen in Fig. S4 that the posterior distribution of coefficient number 3 has two peaks (light green line in Fig. S4). The points in the smaller peak, that is the sets of parameter

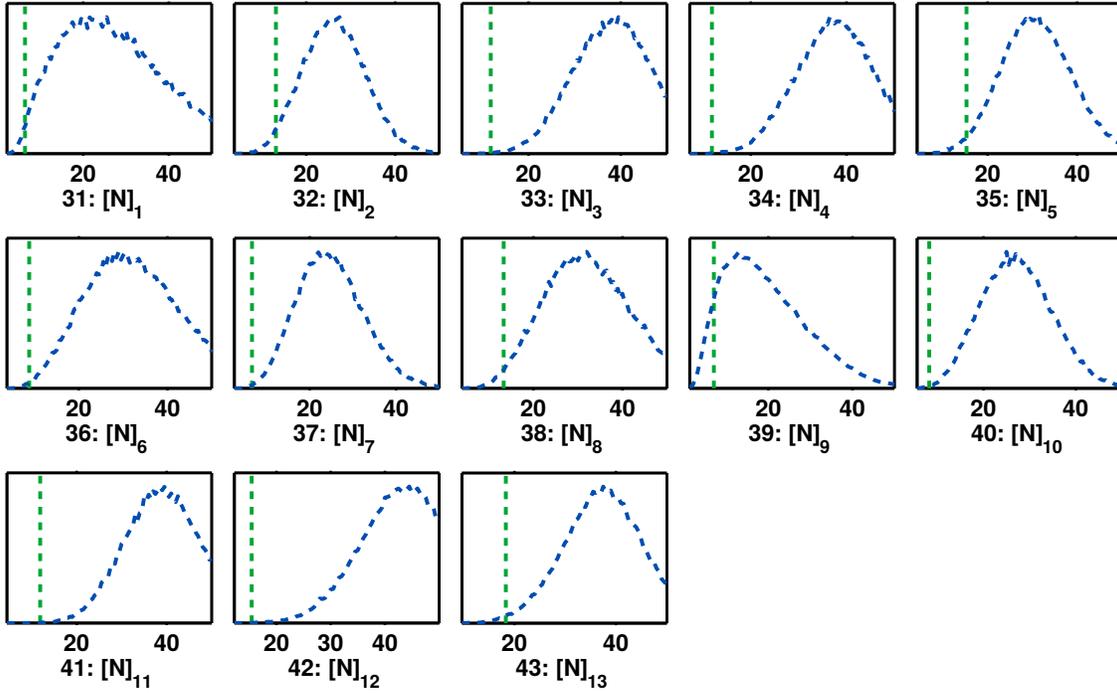


Figure S7: Posterior distributions of the background ammonia concentrations (as mixing ratios in ppt) corresponding to the synthetic cluster distributions and the MCMC simulation where the background ammonia concentrations were varied. The vertical green lines corresponds to the MCMC estimates from the experimental data, which are used as input parameter values for producing the synthetic cluster distributions.

values with coefficient number 3 being higher than 0.3, are chosen to form first group of points, denoted (a). The remaining points are further divided into two more groups according to the value of coefficient number 5. The points with coefficient number 3 below 0.3 and coefficient number 5 below 2 form group (b), and the points with coefficient number 3 below 0.3 and coefficient number 5 above 2 form group (c).

The grouping of the evaporation rates presented in Sect. S2.1 can now be performed separately for each of the scenarios (a)–(c). A vast majority, about 90%, of the points fall into category (b). For this scenario, the conclusions from Sect. S2.1 do not change much, except for coefficients number 6 and 8 now belonging to group 2 (evaporation rates below 1 s^{-1}).

Scenarios (a) and (c) correspond to two very different solutions that nevertheless give an equally good fit to the synthetic data. In all three scenarios, the pure dimer and trimer are very stable, the $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_4 \cdot \text{NH}_3$ cluster is stable with respect to ammonia evaporation, and coefficients number 10, 11 and 14 have distinctly non-zero probability density over the whole allowed range and their values cannot be determined. For the remaining coefficients, the parameters have different values (well below 1 s^{-1} , close to 1 s^{-1} , undefined) in the three different scenarios.

As the 'correct values' of the parameters are in this case known, it can be seen that scenario (b) is the 'right answer' that predicts the correct values for those parameters for which an estimate is possible. Scenario (b) also happens to have more points than the other two scenarios, but based on only one example, it is not clear that this would be a sufficient criterion for finding the correct scenario.

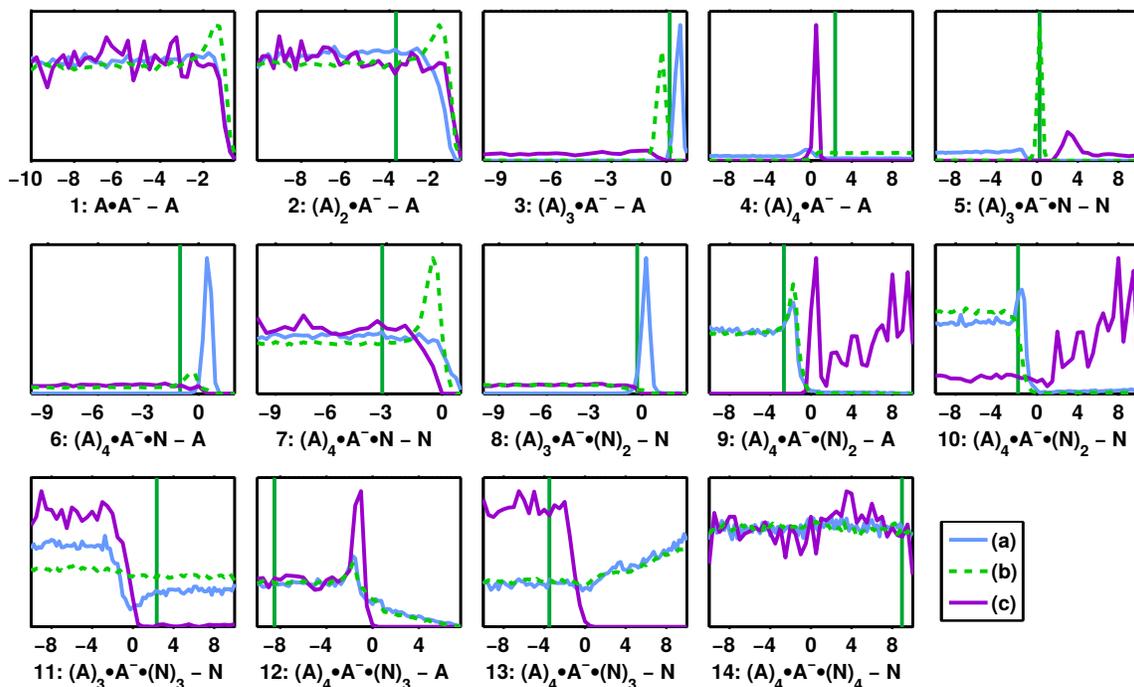


Figure S8: Posterior distributions of the base 10 logarithm of the evaporation rates (in units of s^{-1}) corresponding to the synthetic cluster distributions, separated into three scenarios. The background ammonia concentrations have been set to 5 ppt in the MCMC simulation. The vertical green lines corresponds to the quantum chemistry-based estimates used as input parameter values for producing the synthetic cluster distributions. (For coefficient number 1, the input value -17.8 is outside the range sampled by MCMC.) A stands for H_2SO_4 , A^- for HSO_4^- and N for NH_3 .

S3. Additional plots related to the analysis of the experimental data

S3.1. Output cluster distributions for different background ammonia options

The output cluster distributions from the MCMC simulation are presented in the main text for the case with an assumed background ammonia concentration of 5 ppt. Figs. S9 and S10 show similar plots for an assumed background ammonia concentration of 1 ppt and varying background ammonia concentrations, respectively. In all of these simulations, the output concentrations represent very well the main trends with respect to both ammonia and sulfuric acid vapor concentrations.

S3.2. Posterior distributions for the ion production rate, wall loss coefficient and background ammonia concentration

Figure S11 presents the posterior distributions of the ion production rate and wall loss coefficient from the MCMC simulations using the experimental cluster distributions and the three different options for treating the low ammonia concentrations. For the MCMC simulation where the background ammonia concentrations were varied, the corresponding posterior distributions of each ammonia concentration are shown in Fig. S12.

S3.3. Posterior distributions separated into different scenarios

Figures S13-S18 show the posterior distributions of Figs. 4, 5 and 6 of the main article corresponding to MCMC simulations with a fixed background ammonia concentration separated into the alternative solutions (A)–(E) of Table 1. The posterior distributions were split into the different solutions similarly as described in Sect. S2.6. For the simulation with 1 ppt, the solutions were separated based on the two peaks in the distribution of coefficient number 6 (blue line in Fig. 4 of

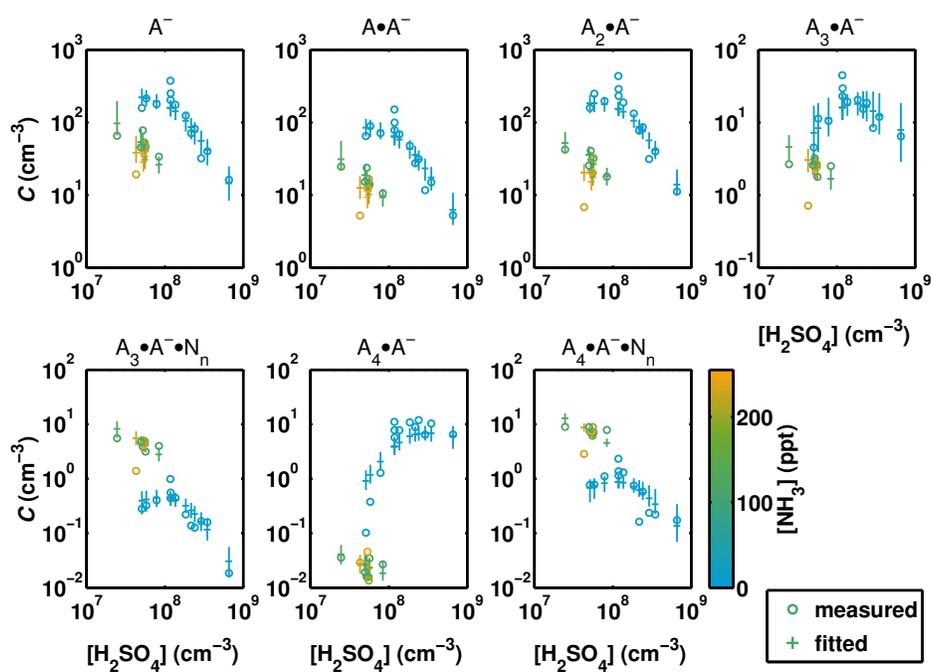


Figure S9: Cluster distributions measured at CLOUD and the corresponding modeled cluster concentrations from an MCMC simulation where evaporation rates, fragmentation probabilities, the ion production rate and the wall loss rate are varied, and the background ammonia concentration is set to 1 ppt. A stands for H_2SO_4 , A^- for HSO_4^- and N for NH_3 .

the main text), as this produced a neat separation also for coefficients number 7 and 8 (Fig. S13). For the case with 5 ppt ammonia the solutions were separated based on the three peaks in the distribution of coefficient number 5 (purple line in Fig. 4 of the main text), as this produced neat separation also for coefficients number 3, 4, 7 and 8 (Fig. S14).

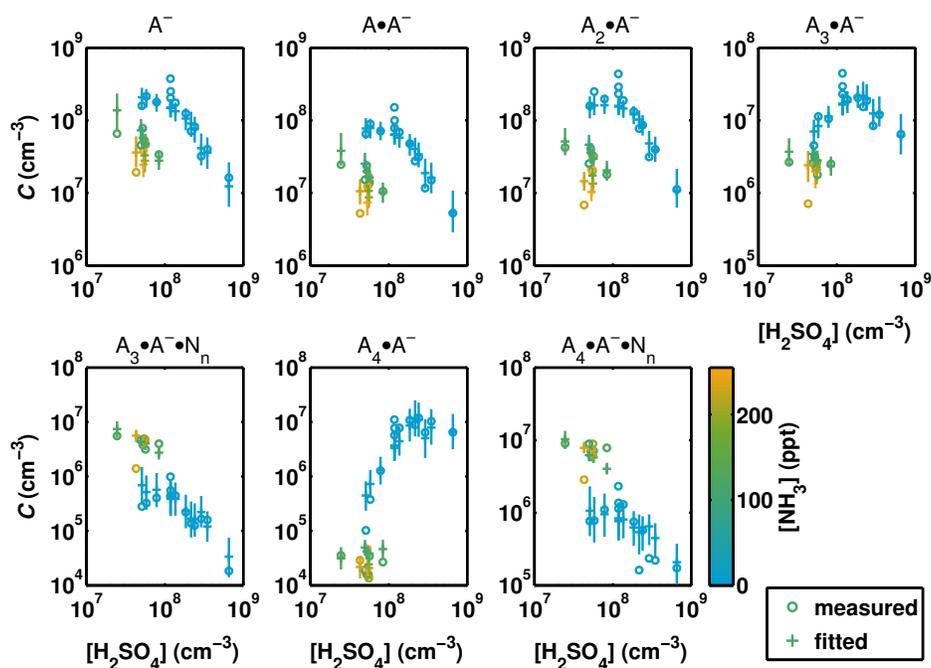


Figure S10: Cluster distributions measured at CLOUD and the corresponding modeled cluster concentrations from an MCMC simulation where evaporation rates, fragmentation probabilities, the ion production rate, the wall loss rate, and the all below-detection-limit ammonia concentrations are varied. A stands for H_2SO_4 , A^- for HSO_4^- and N for NH_3 .

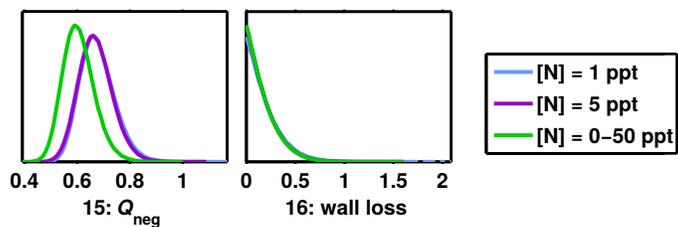


Figure S11: Posterior distributions of the ion production rate rates (in units of $\text{cm}^{-3}\text{s}^{-1}$) and wall loss constant (in units of 10^{-3}s^{-1}) corresponding to the experimental cluster distributions and different options for treating the background ammonia concentration in the experiments where it was below the detection limit and therefore unknown. A stands for H_2SO_4 , A^- for HSO_4^- and N for NH_3 .

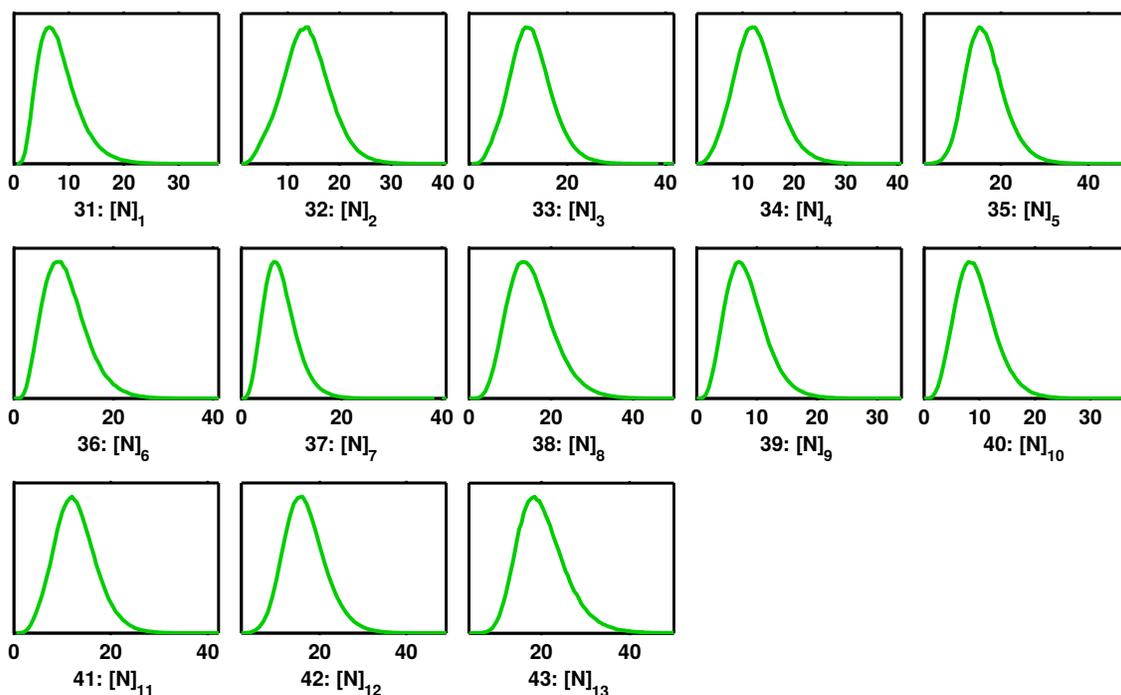


Figure S12: Posterior distributions of the background ammonia concentrations (as mixing ratios in ppt) corresponding to the experimental cluster distributions and the MCMC simulation where the background ammonia concentrations were varied. A stands for H_2SO_4 , A^- for HSO_4^- and N for NH_3 .

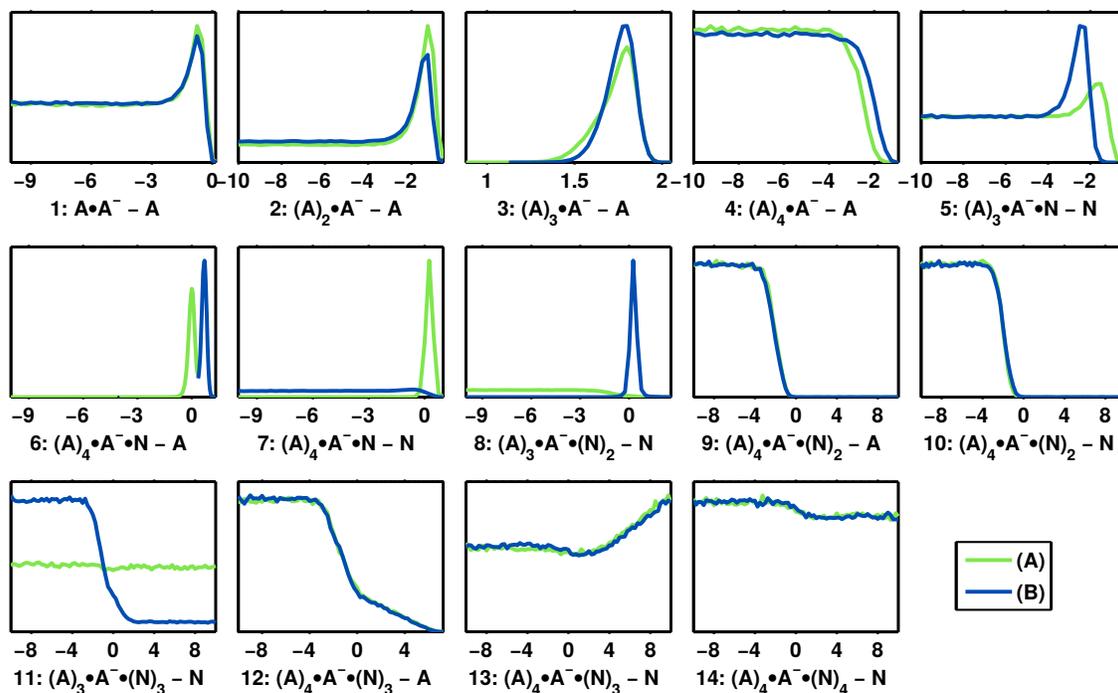


Figure S13: Posterior distributions of the base 10 logarithm of the evaporation rates (in units of s^{-1}) corresponding to the experimental cluster distributions and an assumed background ammonia concentration of 1 ppt, separated into two scenarios. A stands for H_2SO_4 , A^- for HSO_4^- and N for NH_3 .

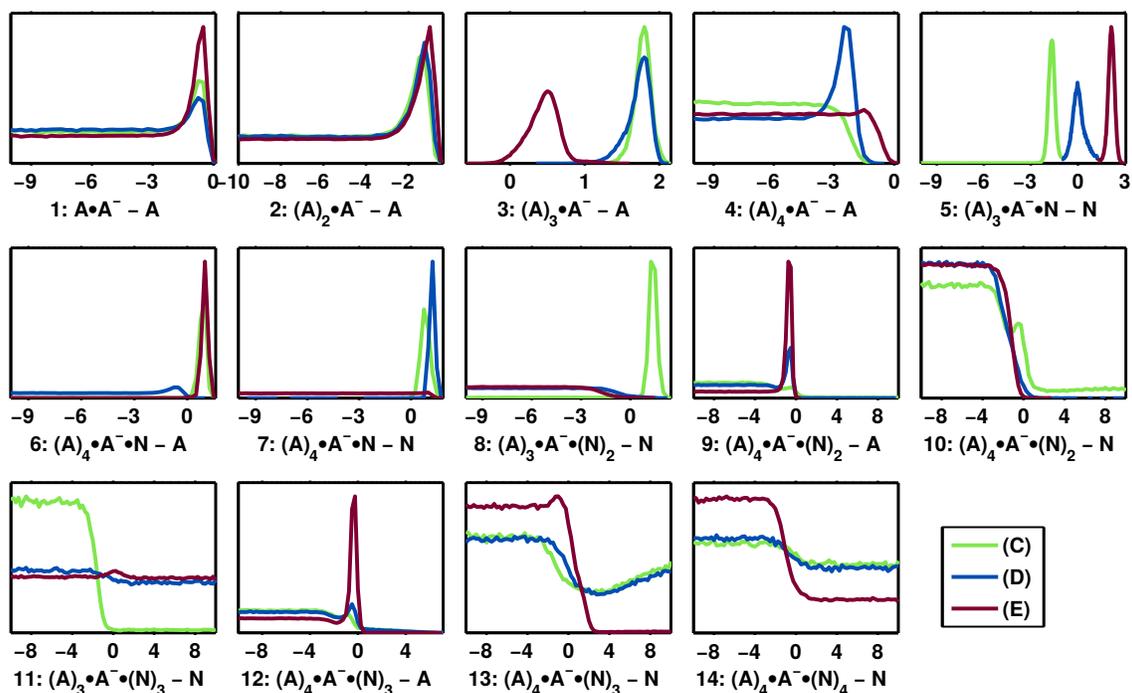


Figure S14: Posterior distributions of the base 10 logarithm of the evaporation rates (in units of s^{-1}) corresponding to the experimental cluster distributions and an assumed background ammonia concentration of 5 ppt, separated into three scenarios. A stands for H_2SO_4 , A^- for HSO_4^- and N for NH_3 .

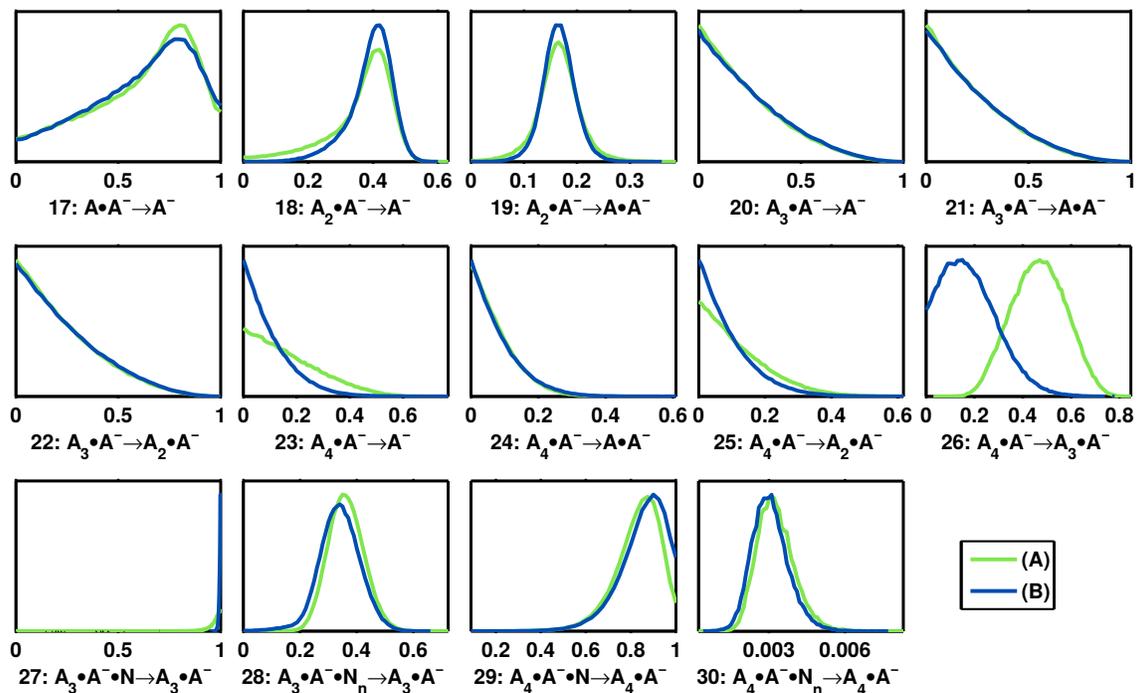


Figure S15: Posterior distributions of the fragmentation probabilities of clusters in the mass spectrometer inlet corresponding to the experimental cluster distributions and an assumed background ammonia concentration of 1 ppt, separated into two scenarios. A stands for H_2SO_4 , A^- for HSO_4^- and N for NH_3 .

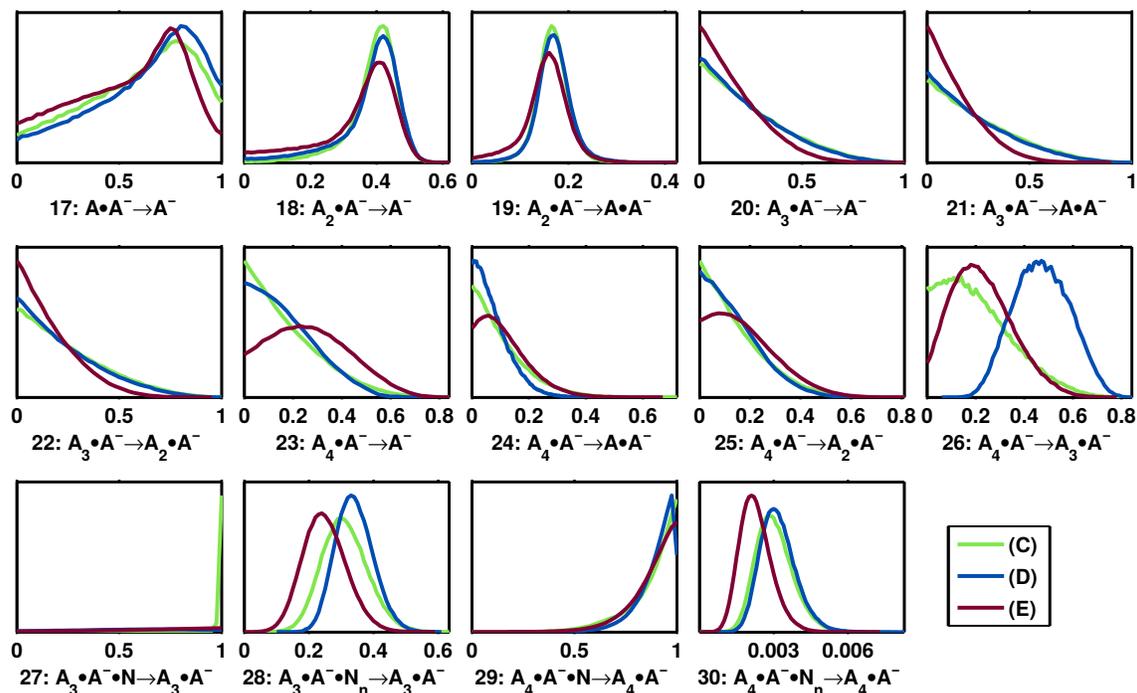


Figure S16: Posterior distributions of the fragmentation probabilities of clusters in the mass spectrometer inlet corresponding to the experimental cluster distributions and an assumed background ammonia concentration of 5 ppt, separated into three scenarios. A stands for H_2SO_4 , A^- for HSO_4^- and N for NH_3 .

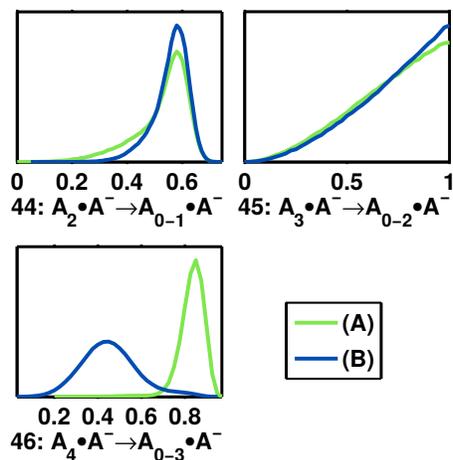


Figure S17: Posterior distributions of the total fragmentation probabilities of the $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_{2-4}$ clusters corresponding to the experimental cluster distributions and an assumed background ammonia concentration of 1 ppt, separated into two scenarios. A stands for H_2SO_4 and A^- for HSO_4^- .

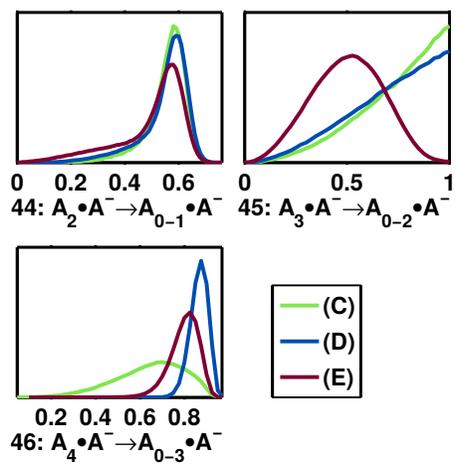


Figure S18: Posterior distributions of the total fragmentation probabilities of the $\text{HSO}_4^- \cdot (\text{H}_2\text{SO}_4)_{2-4}$ clusters corresponding to the experimental cluster distributions and an assumed background ammonia concentration of 5 ppt, separated into three scenarios. A stands for H_2SO_4 and A^- for HSO_4^- .

References

- Almeida, J., Schobesberger, S., Kürten, A., Ortega, I. K., Kupiainen-Määttä, O., Praplan, A. P., Adamov, A., Amorim, A., Bianchi, F., Breitenlechner, M., David, A., Dommen, J., Donahue, N. M., Downard, A., Dunne, E., Duplissy, J., Ehrhart, S., Flagan, R. C., Franchin, A., Guida, R., Hakala, J., Hansel, A., Heinritzi, M., Henschel, H., Jokinen, T., Junninen, H., Kajos, M., Kangasluoma, J., Keskinen, H., Kupc, A., Kurtén, T., Kvashin, A. N., Laaksonen, A., Lehtipalo, K., Leiminger, M., Leppä, J., Loukonen, V., Makhmutov, V., Mathot, S., McGrath, M. J., Nieminen, T., Olenius, T., Onnela, A., Petäjä, T., Riccobono, F., Riipinen, I., Rissanen, M., Rondo, L., Ruuskanen, T., Santos, F. D., Sarnela, N., Schallhart, S., Schnitzhofer, R., Seinfeld, J. H., Simon, M., Sipilä, M., Stozhkov, Y., Stratmann, F., Tomé, A., Tröstl, J., Tsagkogeorgas, G., Vaattovaara, P., Viisanen, Y., Virtanen, A., Vrtala, A., Wagner, P. E., Weingartner, E., Wex, H., Williamson, C., Wimmer, D., Ye, P., Yli-Juuti, T., Carslaw, K. S., Kulmala, M., Curtius, J., Baltensperger, U., Worsnop, D. R., Vehkamäki, H., and Kirkby, J. (2013). Molecular understanding of sulphuric acid–amine particle nucleation in the atmosphere. *Nature*, 502(7471):359 – 363.
- Brooks, S. P. and Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *J. Comp. Graph. Stat.*, 7(4):434 – 455.
- Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457 – 472.
- Ortega, I. K., Olenius, T., Kupiainen-Määttä, O., Loukonen, V., Kurtén, T., and Vehkamäki, H. (2014). Electrical charging changes the composition of sulfuric acid–ammonia/dimethylamine clusters. *Atmos. Chem. Phys.*, 14(15):7995 – 8007.
- Su, T. and Chesnavich, W. J. (1982). Parametrization of the ion-polar molecule collision rate constant by trajectory calculations. *J. Chem. Phys.*, 76(10):5183 – 5185.
- ter Braak, C. J. F. and Vrugt, J. A. (2008). Differential Evolution Markov Chain with snooker updater and fewer chains. *Statistics and Computing*, 18(4):435 – 446.