

Response to major revision comments for the paper "Identification of molecular cluster evaporation rates, enthalpies and entropies by Monte Carlo method"

September 9, 2020

1 Overview

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

We wish to thank the Referee for their insightful comments which we feel substantially increased the quality of the manuscript. We believe that we have addressed all of the major and minor comments made by the reviewer and, in so doing, have produced a paper that is more rigorous in structure and more clear in presentation.

Next, in Section 2 we list the Referee's comments. We also include our comment-by-comment responses. Each of the referee's comments are denoted with "**C**" and our responses to the referee's comments are denoted with "**R**". At the end of the document we supply a marked-up version of the paper which contains a detailed comparison of the previous and revised versions of the manuscript.

2 Referee comments and our responses

Recommendation to the editor

1. Scientific significance
Does the manuscript represent a substantial contribution to scientific progress within the scope of this journal (substantial new concepts, ideas, methods, or data)?
Outstanding **Excellent** Good Fair Low
2. Scientific quality
Are the scientific approach and applied methods valid? Are the results discussed in an appropriate and balanced way (consideration of related work, including appropriate references)?
Outstanding Excellent **Good** Fair Low
3. Presentation quality
Are the scientific results and conclusions presented in a clear, concise, and well structured way (number and quality of figures/tables, appropriate use of English language)?
Outstanding Excellent Good **Fair** Low

For final publication, the manuscript should be reconsidered after major revisions. I would be willing to review the revised paper, if the editor considers it necessary.

Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

Comment: Like the way the authors wrote the paper in a casual way, they seem to respond to the referees' comments in a similar way. It is hard to follow the response letter. There are so many errors, especially what were written in the letter are not the same as those in the revision. Although the authors addressed most of the comments and improvement has indeed seen in the revision, the authors will still need more efforts to improve the quality and the readability of the manuscript. There are still lot of errors/typos and those are really surprising. Below are some issues that need to be resolved before the paper can be publishable in ACP.

Response: Thank you for pointing out the typos, errors and inconsistencies that appear in the previous response letter. We strongly apologize for the inconvenience of reading and proof-checking our previous author responses and

tracking related manuscript changes. We further take your recommendations into account and thereby avoid confusions and inaccuracies here.

1. **Comment:** Section 2 is still hard to understand, although it is greatly improved after the revision. In addition, it is very lengthy and redundant. Would it be shortened to make it concise? Some of technical descriptions in my opinion can be moved to the Appendix or supplementary. In addition, the authors use a lot of very short paragraphs and the paper looks like a boring novel. Also, some languages used here are really awkward, given below are some examples:

Response: We have restructured and rewritten Section 2 to improve its quality and readability. We merged and reformulated the paragraphs to make the workflow more logical.

We first explain generation of synthetic data. Next, we place the section dedicated to Markov chain Monte-Carlo simulations which is subdivided into two parts: selection of minimum and maximum limits for unknown parameters, and overview of the MCMC runs. Both subsections have been made conciser. We have moved the technical details of Metropolis algorithm and its extended version (the DRAM method) to Appendix. These methods are given in A2 and A3, respectively. Additionally we reformulated the language in many of the sentences (see examples below). Here the lines from revised version of the manuscript are given in bold.

We have moved the first paragraph from "Discussions and future work" to Section 2.1 ("Generation of synthetic data"). Here we explain the sensitivity of MCMC parameter estimation to the quality and limitations of the synthetic data.

- (a) **Comment:** Line 108, by the following method. You really mean by the following procedures or steps, right?

Response: Section 2.1 has been substantially rewritten. We removed this sentence from the text. Instead, we explain the origin of synthetic data as follows: "We generated the birth-death equations using the ACDC code (McGrath et al., 2012), and then solved for the cluster concentrations using the Fortran ordinary differential equation solver VODE (N. Brown et al., 1989)." See **Lines 111-114**.

- (b) **Comment:** Line 118-119, what do you mean "each simulation was initialized with ... and no sulphuric acid"? You mean "without

sulphuric acid”?

Response: We reformulated this part into ”...the initial sulfuric acid was set to zero in each simulation.” See **Line 109**.

- (c) **Comment:** Line 122, what is “our particle system”? In particular, what is “particle” here?

Response: Thank you for highlighting this important issue which we believe makes our notions more consistent. Throughout the text, we have replaced the ‘particle system’ with ‘the simulated system of clusters’ or ‘the set of molecular clusters considered here’. See, e.g., **Lines 115, 122**.

- (d) **Comment:** Line 124, for time values less than the time at which. . . , do you mean “for time values before the system has attained the steady state”?

Response: Indeed, we intended to say “for time values before the system has attained the steady state”. We reformulated this part as follows: ”... measured at 1.5 min time intervals before the system reaches a steady state. This corresponded to a total of 41 time steps.” See **Lines 123-124**.

- (e) **Comment:** Line 126-127, “with time resolution comprising 1.5 minutes”, do you mean “with a time interval of 1.5 minutes”?

Response: We reformulated this part, as mentioned above.

- (f) **Comment:** Line 129, would it be “first...second”?

Response: Thank you for recommendation. We have changed the language accordingly. We explain two data sets generated for synthetic data (in **Line 123**). These are referred as ”..the fist set...” and ”In the second case...” (**Line 125**).

- (g) **Comment:** Line 132, reached not reached to and the sentence “The measure of how close . . .” is so complicated and awkward. Would it be modified for the sake of readers’ benefit?

Response: We modified this sentence as follows: ”Additionally, we include a convergence parameter for assessing the closeness of cluster concentrations to the steady state for every individual ACDC simulation.” See **Lines 127-128**.

- (h) **Comment:** Line 228-229, please rearrange the sentence.

Response: As it was mentioned above, the discussion related to Metropolis algorithm was moved to Appendix. We changed the sentences which describe the initial assumptions for parameter values as follows: ”We first select the flat prior distribution from which we will initially sample unknown parameters, as we wish

to generate physically reasonable parameter estimates. Therefore, we generate unknown parameters within the chosen minimum and maximum bounds where all the points are equally likely to be sampled” (**Lines 367-369**).

Comment: It is strongly recommended that this section should be completely rewritten.

Response: The section had been restructured, rewritten and shortened in accordance with the advice of the Referee, as mentioned above.

2. **Comment:** The Results and Discussion section looks better than section 2. However, there are still some improvements need to be made. A lot of sentences are quite redundant and need to be modified for conciseness. For example, Line 293, adding “(Table 2) after concentrations will serve the purpose; you don’t need to say listed earlier in Table 1, “(Table 1) will be the same. Line 295, “the steady-state” steady-state here I believe is adjective. There are several throughout the manuscript. Line 300, An example of one of, is it necessary to include “one of” here?

Response: Following recommendations from the Referee, we restructured and rewrote Section 3. Below we summarize the main structural changes as well as some minor edits.

Initially, we reformulated Sections 3.1-3.3 and thus removed the redundancies. According to the Referee’s advise, we omit the references to Tabs. 1 and 2.

Next, we deleted Sections 3.4 and 3.5. The last section (i.e. ”Discussion and Future Work”) has been redistributed over the manuscript as follows. First, we have moved the text in Lines 436-449 (related to quality of the computer-simulated cluster concentrations) to Section 2.1 (”Generation of synthetic data”). See **Lines 115-122**.

Second, we inserted Lines 450-455 into Conclusion (**Lines 311-317**). This part explains two general principles of inverse problems/Bayesian parameter estimation applied in our study.

The final part of Section 3.5 (Lines 456-465), which explains the correlations between formation enthalpies and entropies is moved to Section 3.3 (**Lines 288-297**). Here we explain the results of parameter estimation after we expressed evaporation rates as parametrized functions of the temperature, with the cluster formation enthalpies and entropies as the unknown parameters. Naturally, we discuss why re-parametrization has improved the results, and the reasons why thermodynamic data display correlations.

Following advise from the Referee, we distinguish between the adjective "steady-state" and the noun "steady state" in the revised version of the manuscript.

The authors decided to keep the sentence in Line 300 as in the previous version ("An example of one of the sampled chains ...".) We believe that the sentence is sufficiently comprehensive for the reader.

3. **Comment:** The ideas and the results are very interesting and the paper can benefit the community but the way the authors represent do really discouragement and the manuscript needs to be substantially improved in the next revision.

Response: Thank you for acknowledging the advantages and benefits of modelling the type accomplished in this study. We made a considerable effort which we hope has helped to improve the quality of the presentation at the major revision stage. The changes are reflected in updated version of the manuscript.

Identification of molecular cluster evaporation rates, cluster formation enthalpies and entropies by Monte Carlo method

Anna Shcherbacheva¹, Tracey Balehowsky², Jakub Kubečka¹, Tinja Olenius³, Tapio Helin⁴, Heikki Haario^{4,5}, Marko Laine⁵, Theo Kurtén^{6,1}, and Hanna Vehkamäki¹

¹Institute for Atmospheric and Earth System Research, P.O. Box 64 00014 University of Helsinki, Finland

²Department of Mathematics and Statistics Subunit, P.O. Box 64 00014 University of Helsinki, Finland

³Department of Environmental Science and Analytical Chemistry & Bolin Centre for Climate Research, Stockholm University, Svante Arrhenius väg 8, SE-11418 Stockholm, Sweden

⁴LUT School of Engineering Science, Lappeenranta-Lahti University of Technology, P.O.Box 20 FI-53851 Lappeenranta, Finland

⁵Finnish Meteorological Institute, P.O. Box 503, FI-00101 Helsinki, Finland Finland

⁶Department of Chemistry, P.O. Box 55 FI-00014 University of Helsinki, Finland

Correspondence: Anna Shcherbacheva (anna.shcherbacheva@helsinki.fi)

Abstract.

We address the problem of identifying the evaporation rates for neutral molecular clusters from synthetic (computer-simulated) cluster concentrations. We applied Bayesian parameter estimation using a Markov chain Monte Carlo (MCMC) algorithm to determine cluster evaporation/fragmentation rates from [\[..¹ \]synthetic cluster distributions generated by the Atmospheric Cluster Dynamic Code \(ACDC\) \[..² \]and based on gas kinetic collision rate coefficients and evaporation rates obtained using quantum chemical calculations \[..³ \]and detailed balances. The studied system consisted of electrically neutral \[\\[..⁴ \\]sulfuric acid and ammonia clusters with up to 5 of each type of molecules\]\(#\). We then treated \[\\[..⁵ \\]the concentrations generated by ACDC\]\(#\) as synthetic experimental data\[\\[..⁶ \\]\]\(#\). With the assumption that the collision rates are known, we tested two approaches for estimating the evaporation rates \[\\[..⁷ \\]from these data\]\(#\). First, we studied a scenario where \[\\[..⁸ \\]time-dependent cluster distributions are measured at a single temperature\]\(#\) before the system reaches a steady-state. In the second scenario, only steady-state cluster distributions are measured, but at several temperatures. \[\\[..⁹ \\]\]\(#\)Additionally, in the latter case the evaporation rates were represented in terms of cluster formation enthalpies and entropies\[\\[..¹⁰ \\]\]\(#\). This reparametrization reduced the number of unknown parameters, since several evaporation rates depend on the same cluster formation enthalpy and entropy values. We also estimated the evaporation rates using \[previously published\]\(#\) synthetic steady-state cluster concen-](#)

¹removed: known cluster distributions , assuming that the cluster collision rates are known. We used the

²removed: with evaporation rates based on

³removed: to generate cluster distributions for a set

⁴removed: sulphuric

⁵removed: these concentrations

⁶removed: , and

⁷removed: . First we have

⁸removed: at one single temperature

⁹removed: This allowed us to use multiple sets of concentrations at different temperatures.

¹⁰removed: which were considered to be free parameters

15 tration data at one temperature [..¹¹] and compared our two [..¹²] cases to this setting. Both the [..¹³] time-dependent and the two-temperature steady-state concentration data [..¹⁴] allowed us to estimate the evaporation rates with less variance than in the steady-state one temperature case.

We show that [..¹⁵] temperature-dependent steady-state data outperforms [..¹⁶] single-temperature time-dependent data for parameter estimation, even if only two temperatures are used. We can thus conclude that for experimentally determining 20 evaporation rates, cluster distribution measurements at several temperatures are recommended over time-dependent measurements at one temperature.

1 Introduction

The formation of molecular clusters, and their subsequent growth to aerosol particles, is an important yet poorly understood process in our atmosphere. Clusters and aerosols affect both climate, air chemistry (?), evapotranspiration in forest environ- 25 ments (?), and many other atmospheric processes (?).

Recent developments in mass spectrometers have enabled the detection, quantification, and chemical characterization of ionic clusters containing between one and some tens of molecules at atmospherically relevant mixing ratios ¹⁷ [..¹⁸] (??????). Molecular clusters in atmospheric conditions are predominantly electrically neutral, and must thus be charged prior to mass spectrometric detection. This may affect the measurement results, as only part of the sample molecules or clusters may be 30 charged (?), and the charging may also alter cluster compositions. For example, for sulfuric [..¹⁹] acid base clusters, negative charging tends to lead to a loss of base molecules, and positive charging to a loss of acid molecules (?). Modelling is thus needed to connect measured ion cluster distributions to the original neutral population.

Even when the atmospheric cluster [..²⁰] distributions can be accurately deduced from experimental data, [..²¹] these distributions do not quantify the individual kinetic parameters, such as the cluster collision and evaporation rates (?). [..²²] The 35 collision rates may be computed from kinetic gas theory or classical trajectory simulations with reasonable accuracy (?), although recent research has shown that long-range attractive interactions may enhance collision rates (?), for example by around a factor of 2-3 for H₂SO₄ – H₂SO₄ collisions (?). These relatively minor uncertainties in the collision rates are dwarfed by the error margins of cluster evaporation rates. In computational applications, evaporation rates are usually computed using the detailed balance assumption together with the free energies of cluster formation, which can in turn be computed using quantum

¹¹removed: (which has appeared in previous literature)

¹²removed: study

¹³removed: transient and

¹⁴removed: estimated

¹⁵removed: in the second setting, even if only two temperatures were used, the

¹⁶removed: the first setting

¹⁷around or below one part per trillion (ppt)

¹⁸removed: (??????)

¹⁹removed: acid-base

²⁰removed: distribution

²¹removed: this does

²²removed: Collision

40 chemical (QC) methods, (????). Unfortunately, the evaporation rates depend exponentially on the free energies[..²³], and typically observed variations of up to several kcal/mol between [..²⁴]the different applicable QC methods thus translate into orders of magnitude differences in evaporation rates [..²⁵](??).

Despite uncertainties involved in computational estimates of collision and evaporation rates, cluster population dynamic models based on Becker-Döring equations have been able to predict the [..²⁶]sulphuric acid concentration dependence of cluster concentrations (?), and even absolute particle formation rates (?) in [..²⁷]sulphuric acid-ammonia and [..²⁸]sulphuric acid-DMA systems, without empirical model calibration or parameter tuning. The Becker-Döring equations are a system of Ordinary Differential Equations (ODE), which account for cluster birth and death processes (which depend on the collision and evaporation rates), as well as external cluster sinks and sources. In both studies (? and ?), these equations were implemented through the Atmospheric Cluster Dynamic Code (ACDC) (?), using kinetic gas theory collision rates, and standard quantum chemistry techniques for computing cluster formation free energies (and thus evaporation rates).

In mathematical terms, the prediction of cluster concentrations using known collision and evaporation rates is called the forward problem. The associated inverse problem is to use known cluster concentrations to deduce the collision and evaporation rates. The inverse problem can be addressed with Bayesian approaches such as Markov [..²⁹]chain Monte Carlo (MCMC) methods. In a recent paper by ?, Differential Evolution (DE) MCMC (?) was applied to determine evaporation rates for negatively charged [..³⁰]sulphuric acid and ammonia clusters (containing up to five of each type of molecules, with the HSO_4^- ion here defined as an "acid"). This study used steady-state cluster concentrations measured in the CLOUD³¹ chamber experiment at constant temperature, with varying [..³²]sulphuric acid and ammonia concentrations (we refer to ? for details relevant to the experimental data). [..³³]The collision rates were computed from kinetic gas theory. ? concluded that these data were insufficient for estimation of all the evaporation rate coefficients. Another recent paper (?) reported thermodynamic data (cluster formation enthalpies and entropies) for 11 neutral [..³⁴]sulphuric acid and ammonia clusters. In the CLOUD experiment, these were deduced from new particle formation (NPF) rates measured at 5 different temperatures, over a wide range of [..³⁵]sulphuric acid and ammonia concentrations. Most of the thermodynamic parameters could not be narrowly constrained, as the ranges of cluster formation enthalpies and entropies that reproduced the measured NPF rates were quite wide. However, for each cluster only one monomer evaporation rate was taken into account (either acid or base). Furthermore, the NPF rates obtained using the fitted parameters were systematically lower than the measured ones for warmer temperatures (≥ 248 K).

²³removed: variations of

²⁴removed: different

²⁵removed: (??)

²⁶removed: sulphuric

²⁷removed: sulphuric

²⁸removed: sulphuric

²⁹removed: Chain

³⁰removed: sulphuric

³¹Cosmics Leaving OOutdoor Droplets

³²removed: sulphuric

³³removed: Collision rates were taken

³⁴removed: sulphuric

³⁵removed: sulphuric

In this study, we test which combinations of experimental data and fitted parameters lead to the best identification of the evaporation rates. As experiments are expensive and time-consuming to perform, we use synthetic cluster concentration data created from ACDC simulations to test if the use of time-dependent cluster distribution data would significantly improve the accuracy of the evaporation rates. [..³⁶]The use of synthetic data also allows us to know for sure if our inverse modelling
70 actually produces the correct kinetic parameters[..³⁷], which would not be possible with experimental concentration data. As in the ? study, we compute collision rates from kinetic gas theory, while the evaporation rates used to generate our synthetic data are calculated from Gibbs free energies published by [..³⁸]?. Note that the conclusions of this study are not sensitive to the accuracy of the quantum chemical data, as our focus is on the inverse problem of how to determine evaporation rates from known concentrations rather than on the forward problem.

75 For simplicity, we consider the case of neutral [..³⁹]sulfuric acid-ammonia clusters containing up to five of each type of molecules. Studying neutral clusters has the advantage that we can restrict ourselves to a smaller set of kinetic parameters, and ignore uncertainties related to charging and neutralization processes. In situations where a large fraction of the clusters are charged, accurate modelling would require at least three times as many parameters, as both the negative, positive and neutral cluster populations interact with each other. The downside of this simplification is that we lose the direct connection to potential
80 real-life experiments, as neutral atmospheric clusters cannot currently be measured without first charging them.

We investigate [..⁴⁰]three different scenarios for estimating evaporation rates. First, we use steady-state concentration measurements determined at a single temperature, similar to the approach used in ?. Next, we test the use of time-dependent cluster concentrations measured before the system has attained a steady state. This is motivated by the fact that [..⁴¹]time-dependent data should provide additional information about the speed of the processes, which is missing from
85 the steady-state data. [..⁴²]Third, we apply the approach of ?, and express the evaporation rates as parameterized functions of the temperature, with the cluster formation enthalpies and entropies (assumed here to be temperature-independent) as the unknown parameters. This reparametrization is useful for two reasons. First, since the formation enthalpies and entropies of the monomers can be set to zero, and since several evaporation rates depend on the same enthalpy and entropy values, the dimension of the unknown parameter space for our problem is actually reduced, despite the apparent doubling of the
90 number of parameters. Second, utilizing the temperature dependence allows us to produce and use arbitrarily many synthetic data sets at various temperatures, which mathematically has a regularizing effect on the problem. Note that unlike in ?, all possible evaporation processes, including cluster fissions into two daughter clusters, are taken into consideration. Also, while ? used steady-state new-particle formation rates measured at different temperatures to fit their data, we use cluster concentrations.

³⁶removed: Use

³⁷removed: or not

³⁸removed: ?

³⁹removed: sulphuric

⁴⁰removed: two

⁴¹removed: this transient

⁴²removed: Second

[⁴³]

2.1 Generation of synthetic data

[⁴⁴] We simulated the time evolution of cluster concentrations using collision rates computed from kinetic gas theory and evaporation rates computed from the Gibbs free energies reported by ?. To save computational time, we [⁴⁵] omitted clusters where the number of acid and base molecules [⁴⁶] differed by more than two. Based on both fundamental chemical principles and mass spectrometric data (????), these clusters are quite unstable, and thus have very high evaporation rates, leading to negligibly [⁴⁷] low concentrations. See Table 1 for a list of the considered clusters, 16 in total. We considered four different ammonia monomer mixing [⁴⁸] ratios between 5 and 200 ppt [⁴⁹], corresponding to concentrations between 1.3×10^8 and 5.0×10^9 molecules per cm^3 for the temperature ranges studied here [⁵⁰]. In each individual case, the ammonia mixing ratio was kept constant throughout the simulation. The source rate of sulfuric acid monomer [⁵¹] was kept constant at $Q = 6.3 \times 10^4 \text{ cm}^{-3} \text{ s}^{-1}$ [⁵²]. To reproduce experimental conditions in the CLOUD chamber as closely as possible, the initial sulfuric acid was set to zero in each simulation. See Table 2 for [⁵³]

[⁵⁴]

[⁵⁵] a summary of the concentration settings. Additionally, we [⁵⁶] considered the losses on the CLOUD chamber walls which depend on the cluster size [⁵⁷] (?) and a dilution loss of $S = 9.6 \times 10^{-5} \text{ s}^{-1}$. [⁵⁸] For simplicity, we omitted the effect of relative humidity. We generated the birth-death equations using the ACDC code (?), and then solved for

⁴³removed: In this section we describe the methods used to create data sets for the synthetic cluster concentrations. We also explain parameter estimation by Markov Chain Monte Carlo which was used to obtain the evaporation rates and thermodynamic parameters from synthetic data.

⁴⁴removed: The 16 cluster types included in our study are summarized in Table 1

⁴⁵removed: have excluded

⁴⁶removed: differs significantly from each other. Irrespective of the level of theory, quantum chemical data predict that these clusters will

⁴⁷removed: small concentrations. This is also supported by mass spectrometric measurements showing that the clusters with highest concentrations have roughly the same number of acid and base molecules (????). The

⁴⁸removed: ratio is assumed to remain constant in each individual simulation, and varied

⁴⁹removed: . (These correspond to concentrations of

⁵⁰removed: , respectively). The

⁵¹removed: source rate is

⁵²removed: in all simulations

⁵³removed: the summary of ammonia mixing ratio and the source of sulphuric acid monomer used for the ACDC simulations.

⁵⁴removed: Synthetic concentration data for such neutral clusters were generated by the following method.

⁵⁵removed: First, we computed the collision rates using the Eq. A3 from kinetic gas theory. Then, we used these values for the collision rates along with Eq. A4 and the Gibbs free energies computed from Eq. A5 to obtain the evaporation rates. Note that to compute the Gibbs free energies, we substituted the values for cluster formation enthalpies and entropies given by ? into Eq. A5.

⁵⁶removed: consider

⁵⁷removed: computed with Eq. A2

⁵⁸removed: These values for the rates and losses were substituted into the ACDC algorithm (?), which simulates the time evolution of molecular cluster concentrations. The ACDC code computes the first-order non-linear, ordinary differential system of cluster concentrations as given by Eq. A1. We then integrate the system produced by ACDC

the cluster concentrations using the Fortran ordinary differential equation solver VODE (?). [..⁵⁹] These equations and all related parameters are explained in Appendix A1.

115 [..⁶⁰] Our MCMC results are not specific to the set of molecular clusters considered here. This is supported by the fact that although the size of the system (the number of clusters, or more precisely the maximum size of the clusters, included in the simulations) has an impact on the particle formation rates at high temperatures (> 278 K), the particle formation rates and cluster concentrations produced using different cluster sets (e.g. 4x4, 5x5 and 6x6 sulfuric acid and ammonia molecules) are qualitatively similar (?). Thus, minor changes of the ACDC outputs due to the difference in the sets of considered clusters should not change the MCMC parameter estimation results. Additionally, the boundary conditions for the outgrowing clusters (the choice of the clusters that are considered as formed particles) have only minor influence on 120 the simulation results, as long as the simulated system of clusters is defined in a reasonable way (?).

[..⁶¹] Two data sets were created. In the first set, we generated time-dependent [..⁶²] concentrations for each cluster type, measured at 1.5 *min* time intervals before the system [..⁶³] reaches a steady state.

125 [..⁶⁴] This corresponded to a total of 41 time steps. The steady-state single-temperature data correspond to a subset of these data set. In the second case, we generated steady-state concentrations for all [..⁶⁵] cluster types at two temperatures (278 [..⁶⁶] and 292 K). In both [..⁶⁷] cases, the steady-state cluster concentrations [..⁶⁸] were calculated as the average of the concentrations [..⁶⁹] at $t_1 := 50$ min and $t_2 := 60$ min. [..⁷⁰] Additionally, we include a convergence parameter for assessing the closeness of cluster concentrations to the steady state [..⁷¹] for every individual ACDC simulation. This is computed as a ratio of concentrations taken at times t_2 and t_1 [..⁷²] in each case for the cluster for which this ratio deviated most from 130 unity [..⁷³] (?).

⁵⁹removed: A detailed description of this strategy for solving the forward-problem of finding the cluster concentration rates from Eq. A1 was published in ?. To reproduce the experimental conditions as realistically as possible, each simulation was initialized with non-zero concentration of ammonia monomer and no sulphuric acid. The source of sulphuric acid monomer was supplied at a constant rate as it was previously mentioned.

⁶⁰removed: The above method we used for producing synthetic concentration rates is similar to the one described in ?. We note that unlike ?, in this paper, our particle system is considered at various temperatures .

⁶¹removed: Using the above algorithm, model configuration and parameters, we generated two data sets . First, time evolution of the concentrations $Y_i(t)$ is computed for time values less than the time at which the system has attained the steady state. The maximum time we run is 60 minutes from beginning of the simulation, in the above model configurations. In this case, it is assumed that the concentrations for all the clusters are measured under constant temperature with time resolution comprising 1.5 minutes, which comprises overall 41

⁶²removed: concentration data for each of the cluster types i measured from beginning to the end of each simulation,

⁶³removed: has attained

⁶⁴removed: Secondly, we solve for time-independent

⁶⁵removed: the cluster types for two temperatures comprising

⁶⁶removed: K

⁶⁷removed: data configurations

⁶⁸removed: are

⁶⁹removed: determined for time instances

⁷⁰removed: The measure of how close the system has reached

⁷¹removed: is monitored by a convergence parameter, which is the ratio of the concentrations

⁷²removed: , taken

⁷³removed: ,

[..⁷⁴] Finally, we added measurement error (noise) to the cluster concentrations in both data sets. We call the resulting noisy cluster concentrations synthetic data. Our measurement error was sampled from a multivariate [..⁷⁵] Gaussian distribution, [..⁷⁶] with the variance depending on cluster type i , temperature [..⁷⁷] T , and time instance [..⁷⁸] t . We assume that the standard deviation of the [..⁷⁹] measurement error is 0.001 % of the original concentration.

135 [..⁸⁰]

Table 1. Neutral molecular clusters included [..⁸¹] in the model system (16 in total). The first column indicates the number of [..⁸²] sulfuric acid molecules, the second column stands for the number of ammonia in the cluster.

Number of H ₂ SO ₄ molecules	Number of NH ₃ molecules	Number of clusters
0	1	1
1	0-1	2
2	0-2	3
3	1-3	3
4	2-5	4
5	3-5	3

Table 2. Monomer concentrations used in simulations

[H ₂ SO ₄] monomer source	[NH ₃] concentration
$6.3 \times 10^4 \text{ cm}^{-3} \text{ s}^{-1}$	5 ppt
$6.3 \times 10^4 \text{ cm}^{-3} \text{ s}^{-1}$	35 ppt
$6.3 \times 10^4 \text{ cm}^{-3} \text{ s}^{-1}$	100 ppt
$6.3 \times 10^4 \text{ cm}^{-3} \text{ s}^{-1}$	200 ppt

2.2 Markov chain Monte-Carlo simulations

⁷⁴removed: In both data settings, the simulation outputs are amended with the measurement errors

⁷⁵removed: , non-correlated,

⁷⁶removed: where the variance of the distribution depends

⁷⁷removed: T

⁷⁸removed: t . While a simplification of noise characteristics of the real data obtained from a mass spectrometer, we impose

⁷⁹removed: noise comprises

⁸⁰removed: Note that apart from generation of synthetic data, we apply the ACDC as a kinetic model of cluster population in the MCMC simulations. The ACDC outputs are compared to the synthetic measurements and explained in Section 2.2.

[..⁸³][..⁸⁴] We used a Markov Chain Monte Carlo (MCMC) [..⁸⁵]

140 [..⁸⁶] based approach to estimate the evaporation rates which reproduce the synthetic cluster concentration data. Unlike optimization algorithms [..⁸⁷] which compute a single optimal parameter set, MCMC methods sample from a target distribution which contains the most-likely combinations of parameter values [..⁸⁸] for the given data. Multiple samples of possible parameter sets are taken along a random walk in the target distribution, and are saved as a parameter chain. As the length of the chain increases, the sampled sets converge to a probability (posterior) distribution of parameters, which estimates the likelihood of those parameters giving rise to the data. [..⁸⁹]

2.2.1 [..⁹⁰]

145 [..⁹¹]

[..⁹²]

[..⁹³]

[..⁹⁴]

[..⁹⁵]

⁸³removed: The evaporation rate coefficients $\gamma_{i+j \rightarrow i,j}$ appearing in the ACDC simulation of

⁸⁴removed: are treated as unknown parameters. Now we describe how we estimate the evaporation rates from the noisy synthetic data sets obtained by the method described in Section 2.1. We first give a general overview of the basic Metropolis algorithm (?), then describe a modification of the algorithm we implemented in this study, and finally, in Section 2.2.3 we apply this general framework to each of our study cases. Our purpose is to determine all the parameter sets that reproduce the synthetic data within their noise level (which is known). We do this using

⁸⁵removed: sampling.

⁸⁶removed: The objective of MCMC in parameter estimation is to identify possible parameter values which yield the best fit with the experimental

⁸⁷removed: that produce one best combination of parameter values, in the MCMC procedure all the most-probable

⁸⁸removed: are estimated given the

⁸⁹removed: To obtain these combinations, the values of parameters are generated and stored into the MCMC "chain". The MCMC chain will converge to the distribution containing all the most-likely combinations of parameter values as a number of sampled parameter sets (i. e., the chain length) increases. The distribution formed from the chain approximates a posterior probability density function which gives the likelihood of observing each of the parameters given the concentration data.

⁹⁰removed: The Metropolis algorithm

⁹¹removed: First, a prior distribution for the parameter values θ (represented in array form) is chosen and set to be the proposed "true" distribution from which possible parameters are sampled. The prior is typically selected based on the previous knowledge of the parameter values. Then an initial guess for parameter values (denoted as θ_0 or θ_{old}) is selected from the prior distribution.

⁹²removed: Starting from the initial guess, the algorithm samples candidate parameter values (denoted as θ_{new}) from a proposal distribution centred at the previous point (denoted as $q(\theta_{\text{old}}, \theta_{\text{new}})$). The proposal density $q(\theta_{\text{old}}, \theta_{\text{new}})$ is symmetric, which means that the probability of step taken from the 'old' θ_{old} to the 'new' point θ_{new} is same as the probability of the reverse step ($q(\theta_{\text{old}}, \theta_{\text{new}}) = q(\theta_{\text{new}}, \theta_{\text{old}})$).

⁹³removed: Then the candidate point θ_{new} is either accepted or rejected, according to the least-squares fit of the output to the data, which measures the difference between the modelled \mathbf{Y}_{mod} and measured \mathbf{Y}_{exp} cluster concentrations:

⁹⁵removed: where N stands for the number of measurements in synthetic data. We consider two sets of synthetic cluster concentrations: time-dependent, measured at $T = 278$ K and steady-state, measured for two temperatures (at $T = 278$ K and $T = 292$ K), as explained in Section 2.1. For the time-dependent synthetic data $N = N_C \times N_t$, where $N_C = 16$ stands for the number of cluster types included into simulations, while $N_t = 41$ stands for the number of time-step measurements available for each of the cluster types. For the second data set, $N = N_C \times N_T$, where $N_T = 2$ denotes the number of experiments

150 [..⁹⁶]

[..⁹⁷]

[..⁹⁸]

[..⁹⁹]The particular MCMC-based algorithm we use is Delayed Rejection Adaptive Metropolis (DRAM), [..¹⁰⁰]

2.2.1 [..¹⁰¹]

155 [..¹⁰²]

[..¹⁰³]

[..¹⁰⁴]

[..¹⁰⁵]

[..¹⁰⁶]

160 [..¹⁰⁷]

[..¹⁰⁸]

conducted at different temperatures. In the formula above we scale the squared residuals by the measurement error variance σ_i^2 to avoid overfitting to the larger concentration values. The error variance σ_i^2 is matched depending on cluster type, time instance and temperature. See A2 for more details.

⁹⁶removed: At each iteration of the Metropolis algorithm, the value $F(\theta_{\text{new}})$ is compared to the least-square sum from the previous step $F(\theta_{\text{old}})$. If the new value is lower (i.e., the candidate parameters fit the data at least as good as the old values), then the step is accepted. In the opposite case, when $F(\theta_{\text{new}}) > F(\theta_{\text{old}})$, the point will be accepted with the probability

⁹⁸removed: If the candidate point is accepted, the parameter combination θ_{new} is added to the chain, in the opposite case the old value is replicated in the chain. Finally, the value $F(\theta_{\text{old}})$ is replaced with $F(\theta_{\text{new}})$ and saved for the next iteration.

⁹⁹removed: In this paper we employ a variant of the Metropolis algorithm which is more efficient at parameter sampling when the parameter space is large (?). This variant is called the

¹⁰⁰removed: introduced in ?. We briefly explain our approach below.

¹⁰¹removed: The DRAM algorithm

¹⁰²removed: Similar to the basic Metropolis algorithm, the DRAM is initialized with a chosen prior distribution and initial guess for parameter values.

¹⁰³removed: We make our initial guess $\theta = \theta_{\text{old}}$, where θ_{old} is the flat distribution which obeys the estimates in Tabs. 3-4. The limits are explained in Section 2.2.3. We also assume that the conditional probability distributions for the parameters given the concentration data are of Gaussian type.

¹⁰⁴removed: Once initialized, the following iterative steps take place. From the likelihood probability distribution for θ_{old} , a new candidate for the unknown parameter values, θ_{new} , is sampled using the proposed Gaussian likelihood distribution. We then use the algorithm in Section 2.1 to obtain concentration outputs from the evaporation rates θ_{new} . In the first stage of DRAM, we chose to accept the new proposed values θ_{new} with probability

¹⁰⁶removed: where \mathbf{Y}_{exp} is the array of synthetic cluster concentration data, and $p(\mathbf{Y}_{\text{exp}}|\theta_{\text{old}})$, $p(\mathbf{Y}_{\text{exp}}|\theta_{\text{new}})$ denote the likelihood (conditional) probabilities for the old and new parameter values, respectively. These likelihood probabilities quantify how closely the kinetic model with parameters θ reproduce the data, as they depend on the sum of squared residuals (see Eqs. A8 and A7) between the given data and the concentrations obtained from the ACDC and VODE simulations with parameters θ_{old} and θ_{new} , respectively. This relationship is explained further in Appendix A1.

¹⁰⁷removed: In DRAM we allow for partial modification of the proposed parameters (the "delayed rejection" component of DRAM). This second stage of sampling improves the computational time needed to obtain an estimate for θ ; it is performed as follows. If the proposed θ_{new} is rejected, a nearby proposal is created, $\theta_{\text{new}2}$. We accept this second proposal keeping in mind the rejection probability of the first, according to

[..¹⁰⁹]

which is an extended variant of the classical Metropolis algorithm (?). We chose the DRAM algorithm as it is more efficient than the Metropolis regime at parameter estimation when the parameter space is large (?). The [..¹¹⁰]

165 [..¹¹¹]

[..¹¹²]

[..¹¹³]

[..¹¹⁴]two algorithms and their application to our cases are described in the Appendix.

[..¹¹⁵]

170 [..¹¹⁶]

[..¹¹⁷]

[..¹¹⁸]

2.2.1 Selection of minimum and maximum limits for unknown parameters

[..¹¹⁹][..¹²⁰] [..¹²¹][..¹²²]

175 2.2.2 [..¹²³]

[..¹²⁴]

¹⁰⁹removed: At the start of the MCMC simulations, the proposal covariances for both stages are initialized using arbitrary diagonal matrices with equal variances. It is assumed that the proposals of the form $p(\mathbf{Y}_{\text{exp}}|\cdot)$ and $p(\mathbf{Y}_{\text{exp}}|\cdot, \cdot)$ are Gaussian. They are updated at each successive iteration of the MCMC algorithm to improve the mixing of the chains.

¹¹⁰removed: first-stage proposal covariance is recomputed via the Adaptive Metropolis (AM) procedure (?). Let d be the dimension of the parameter space, and $\{\mathbf{X}_0, \dots, \mathbf{X}_n\} \subset \mathbb{R}^d$ be a set of d -dimensional vectors containing the sampled values of free parameters. Then the first-stage proposal is centred at the current position of the Markov chain \mathbf{X}_n , whereas the corresponding proposal covariance \mathbf{C}_n^1 is updated using the path of the previously sampled MCMC chain:

¹¹²removed: where \mathbf{C}_0 is the initial covariance assigned at the beginning of the MCMC runs, n_0 stands for the length of the initial non-adaptation period, $s_d = 2.4/d$ is the scaling parameter, and $\text{Cov}(\mathbf{X}_0, \dots, \mathbf{X}_{n-1})$ is the empirical covariance matrix for the vectors $\mathbf{X}_0, \dots, \mathbf{X}_{n-1}$:

¹¹⁴removed: where $\bar{\mathbf{X}}_{n-1}^T = \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{X}_i$ and $\mathbf{X}_i \in \mathbb{R}^d$ are column vectors. In our study and all runs therein, we set n_0 to be 100 iterations.

¹¹⁵removed: Simultaneously, the second-stage proposal covariance is computed as a scaled version of the first-stage proposal covariance:

¹¹⁷removed: with the scaling factor $\gamma = 5$ borrowed from ?. This value was chosen to increase the acceptance at the second stage.

¹¹⁸removed: Then, if both θ_{old} and θ_{new} are rejected at this stage, a new parameter candidate is sampled and the process is repeated. If the parameter candidate is accepted, the Markov chain is advanced one step and sampling as above is repeated. The process stops once the chain length is exhausted.

¹¹⁹removed: Parameter estimation is conducted using the

¹²⁰removed: 'mcmcstat'

¹²¹removed: toolbox implemented for FORTRAN (??). See the description and the examples of usage on the web page

¹²²removed: .

¹²³removed: Overview of the MCMC runs

¹²⁴removed: In our implementation of the DRAM algorithm, we impose upper and lower limits for the parameter values. We add such domain restrictions to exclude unphysical estimates for our parameters. These restrictions are encoded in our prior distribution, which we set to be a combination of so-called "flat priors", which are distributions that are proportional to a constant, (see Tabs. ??-3).

We emphasize that there are currently *no theoretical principles or experimental results* which [..¹²⁵] set sound restrictions for even the order of magnitude of the evaporation rates. However, [..¹²⁶] evaporation rates much lower than 10^{-10}s^{-1} are irrelevant in [..¹²⁷] practice, since the timescale for evaporation is then much longer than the cluster lifetime with respect to further growth. Similarly, when the evaporation rate is [..¹²⁸] much greater than 10^{+10}s^{-1} , [..¹²⁹] the cluster will [..¹³⁰] certainly evaporate before it has a chance to grow further. [..¹³¹] The base 10 logarithm of the evaporation rates was [..¹³²] therefore sampled in the interval of -12 to 12.

[..¹³³] For the cluster formation enthalpies, we chose an upper limit of 0 kcal/mol [..¹³⁴], as a positive ΔH would mean an absence of attractive interactions in the molecular cluster, which is physically incorrect for polar, H-bonding molecules such as H_2SO_4 and NH_3 . [..¹³⁵] This same argument also applies for each individual molecule, which gives rise to the requirement that the formation enthalpy of each cluster must be lower (more negative) than that of clusters with less acid and/or base molecules. See Table 3 for the full list of restrictions arising from this requirement. As a lower limit for the overall cluster formation enthalpies, we used $\Delta H = -400$ kcal/mol [..¹³⁶]. As our largest clusters contain 10 molecules, this would imply that, on average, each H_2SO_4 in all the studied clusters is bound substantially stronger than in the exceptionally strongly bound $\text{HSO}_4^- * \text{H}_2\text{SO}_4$ cluster, [..¹³⁷] (for which recent high-level computational studies indicate a binding enthalpy roughly around -40 kcal/mol, (??) [..¹³⁸]). This in turn implies that the evaporation rate is zero for all [..¹³⁹] practical purposes.

[..¹⁴⁰] The upper limit for the formation entropies was set to 0 cal/K/mol, [..¹⁴¹] as clustering must have a negative ΔS , [..¹⁴²] since the number of gas molecules is reduced (and translational and rotational degrees of freedom are thus converted into much more constrained vibrational degrees of freedom). [..¹⁴³] The lower limit of -400 cal/K/mol [..¹⁴⁴] can be justified by

¹²⁵removed: indicate possible

¹²⁶removed: we assume that the evaporation rates with orders of magnitude less

¹²⁷removed: practise, since such an evaporation event is highly improbable, and it is very likely that instead the cluster will grow further by collisions

¹²⁸removed: of the order of magnitude more

¹²⁹removed: it is reasonable to expect that

¹³⁰removed: most

¹³¹removed: With these assumptions, the prior distribution of the evaporation rates spans over several orders of magnitude, and the

¹³²removed: sampled from the range

¹³³removed: Next, we justify the limits selected for data setting 2, where we sample thermodynamic parameters. For the formation enthalpies

¹³⁴removed: is chosen by the fact that

¹³⁵removed: For the lower limit (

¹³⁶removed:) we mean that on average

¹³⁷removed: for which the most recent

¹³⁸removed: . Another motivation for the prior distribution selected for the cluster formation enthalpies comes from the fact that the largest cluster included into the system has 5 H_2SO_4 and 5 NH_3 , so 10 molecules, and -400 kcal/mol would give an enthalpy of -40 kcal/mol per molecule, which 1) corresponds to the strongest known cluster in the system and 2) which

¹³⁹removed: purposes of measurement (?)

¹⁴⁰removed: Next, we set the

¹⁴¹removed: since molecule

¹⁴²removed: as

¹⁴³removed: For the

¹⁴⁴removed: , we state

195 noting that the typical per-molecule ΔS for clustering is around -30 cal/K/mol, with a typical variation of up to +-10 cal/K/mol (?). [..¹⁴⁵]For a 10-molecule cluster this would imply a lower bound to ΔS of [..¹⁴⁶]around -400 cal/K/mol. [..¹⁴⁷] [..¹⁴⁸][..¹⁴⁹] [..¹⁵⁰][..¹⁵¹][..¹⁵²][..¹⁵³][..¹⁵⁴][..¹⁵⁵][..¹⁵⁶][..¹⁵⁷][..¹⁵⁸][..¹⁵⁹][..¹⁶⁰][..¹⁶¹][..¹⁶²][..¹⁶³][..¹⁶⁴][..¹⁶⁵][..¹⁶⁶]

Table 3. Additional [..¹⁶⁷]restrictions on the [..¹⁶⁸]cluster formation enthalpies arising from [..¹⁶⁹]the requirement that each individual molecule is bound The cluster formation enthalpy of the i -th cluster is denoted by ΔH_i . The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..¹⁷⁰] ammonia [..¹⁷¹] molecules.

$\Delta H_{2A} > \Delta H_{2A1N}$	$\Delta H_{3A2N} > \Delta H_{4A2N}$
$\Delta H_{1A1N} > \Delta H_{2A1N}$	$\Delta H_{4A2N} > \Delta H_{4A3N}$
$\Delta H_{2A1N} > \Delta H_{3A1N}$	$\Delta H_{4A3N} > \Delta H_{4A4N}$
$\Delta H_{2A2N} > \Delta H_{3A2N}$	$\Delta H_{4A4N} > \Delta H_{5A5N}$
$\Delta H_{3A1N} > \Delta H_{3A2N}$	$\Delta H_{4A4N} > \Delta H_{4A5N}$

¹⁴⁵removed: So for the largest clusters the upper limit corresponds to a per-molecule

¹⁴⁶removed: -40

¹⁴⁷removed: In this situation, all the new vibrational degrees of freedom formed in the product clusters are quite rigid, i.e. have very low entropy (?).

¹⁴⁸removed: h!

¹⁴⁹removed: Domain limitations for two data settings under consideration imposed to exclude non-physical parameters in parameter estimation procedure.

¹⁵⁰removed: Data settings

¹⁵¹removed: Estimated parameters

¹⁵²removed: Minimal value

¹⁵³removed: Maximal value

¹⁵⁴removed: Data setting 1

¹⁵⁵removed: Base 10 logarithms of

¹⁵⁶removed: -12

¹⁵⁷removed: 12

¹⁵⁸removed: evaporation rates (in s^{-1})

¹⁵⁹removed: Data setting 2

¹⁶⁰removed: Cluster formation

¹⁶¹removed: enthalpies ($kcal\ mol^{-1}$) and

¹⁶²removed: -400

¹⁶³removed: 0

¹⁶⁴removed: entropies ($cal\ K^{-1}\ mol^{-1}$)

¹⁶⁵removed: -400

¹⁶⁶removed: 0

[..¹⁷²][..¹⁷³]

2.2.2 Overview of the MCMC runs

200 We [..¹⁷⁴] first performed DRAM parameter estimation from both steady-state and time-dependent cluster concentrations at 278 K, treating evaporation rates as the unknown parameters θ [..¹⁷⁵][..¹⁷⁶]. For the time-dependent synthetic data, the number of output coefficients was $n_{out} = N_C \times N_t + 1$, where $N_C = 16$ is the number of cluster types included into simulations, and $N_t = 41$ is the number of time-step measurements available for each of the cluster types.

[..¹⁷⁷] Next, we performed parameter estimation based on steady-state [..¹⁷⁸] cluster concentrations at two temperatures, 278 K [..¹⁷⁹]

[..¹⁸⁰] and 292 K. The number of output coefficients in this case was $n_{out} = (N_C + 1) \times N_T$, where $N_T = 2$ denotes the number of experiments conducted at different temperatures. We use Eq. A4 and A5 to express the evaporation rates as functions of [..¹⁸¹] formation enthalpies, entropies and temperature:

$$\gamma_{i+j \rightarrow i,j} = f(T, \{\Delta H_k, \Delta S_k\}_{k \in \{i+j, i, j\}}). \quad (1)$$

210 In Eq. 1, we set $T = 278$ K or $T = 292$ K. We emphasize that the rates $\gamma_{i+j \rightarrow i,j}$ now depend on temperature and six other parameters: the [..¹⁸²] formation enthalpy ΔH_{i+j} and entropy ΔS_{i+j} of the evaporating/fragmenting cluster $i + j$, and the formation enthalpies $\Delta H_i, \Delta H_j$ and entropies $\Delta S_i, \Delta S_j$ of the product clusters i and j respectively. In this setting θ represents the array of quantities $\Delta H_{i+j}, \Delta S_{i+j}, \Delta H_i, \Delta H_j, \Delta S_i, \Delta S_j$ with $i + j \in \{1, 2, \dots, 16\}$. Similar approaches were applied for the inverse problem of chemical kinetics modelled by the Arrhenius equation, where chemical reaction rates are temperature-dependent (?).

[..¹⁸³] Many evaporation/fragmentation reactions have the same clusters as products, and thus several of the pairs $\Delta H_i, \Delta S_i$ appear in [..¹⁸⁴] Eq. 1 for the evaporation rates of [..¹⁸⁵] multiple different reactant clusters. The formation enthalpies and entropies of monomers are [..¹⁸⁶] defined in the context of molecular clustering to be zero [..¹⁸⁷]. The number

¹⁷²removed: An outline of the sampling procedure is illustrated in Figure 1 below.

¹⁷³removed: Schematic representation of the study methods.

¹⁷⁴removed: next explicitly describe what synthetic data (\mathbf{Y}_{exp}) and parameters (

¹⁷⁵removed:) which give the acceptance probability in

¹⁷⁶removed: represent in the two study cases.

¹⁷⁷removed: In the first study, the free parameters θ represent the evaporation rates. The data \mathbf{Y}_{exp} is either the time-independent

¹⁷⁸removed: or transient cluster concentrations measured at temperature

¹⁷⁹removed: .

¹⁸⁰removed: In the second study, we

¹⁸¹removed: thermodynamic data, parametrized by

¹⁸²removed: cluster

¹⁸³removed: At either temperature $T = 278$ K or $T = 292$ K, the smaller clusters for certain combinations of ammonia and sulphuric acid may arise from the evaporation of several larger clusters . This implies that

¹⁸⁴removed: expression

¹⁸⁵removed: different cluster types. Additionally, the Gibbs formation free energies

¹⁸⁶removed: fixed

¹⁸⁷removed: , and their associated

of distinct unknown formation enthalpies and entropies [..¹⁸⁸] is thus only 28 [..¹⁸⁹], compared to 39 unknown evaporation
220 rates. Furthermore, the cluster formation entropy and enthalpy values all lie within two orders of magnitude [..¹⁹⁰] [..¹⁹¹]
[..¹⁹²] [..¹⁹³]
[..¹⁹⁴], compared to the evaporation rates which span 24 orders of magnitude. This makes the MCMC method more
efficient.

[..¹⁹⁵] To create a reliable sample from the underlying parameter distribution, the length of the MCMC chain must be [..¹⁹⁶
225] "large enough" (??); that is, many different parameter combinations must be tested. [..¹⁹⁷] In our simulations, the MCMC
chain length typically comprised [..¹⁹⁸] 3 million samples. The MCMC acceptance probabilities (defined below) in each of the
cases were about 88.0%, which is a typical level of acceptance since the [..¹⁹⁹] forward ACDC model (in which the evaporation
and collision rates are known) is deterministic.

In [..²⁰⁰] the MCMC simulations, all sets of parameters which produce cluster concentrations within the allotted noise level
230 of the data (0.001%) are kept in the chain. [..²⁰¹] The sampling procedure is outlined in Figure 1 below. We tested that
the MCMC chains converge to the 'true' values (i.e., the reference parameter values from ?) when we start sampling the
chain from randomly selected initial guess.

3 RESULTS AND DISCUSSION

3.1 Identification of [..²⁰²] evaporation rate coefficients from steady-state data at a single temperature

¹⁸⁸removed: do not vary in our simulations. This imposes additional constraints on possible parameter values. One can calculate that of the 39 evaporations
that are involved in the dynamics of the neutral cluster system under consideration,

¹⁸⁹removed: distinct entropy and enthalpy values appear. Consequently, in this case the number of free parameters has been reduced from 39 to 28. This
information is summarized in Table ???. Moreover, from this table one can see that the

¹⁹⁰removed: . This feature of the cluster formation entropies and enthalpies has the effect of reducing the

¹⁹¹removed: stiffness

¹⁹²removed: of the differential system in

¹⁹³removed: (computed via ACDC) which allows for easier integration via VODE.

¹⁹⁴removed: For the setting above, the data Y_{exp} are the time-independent steady-state cluster concentrations measured at temperature 278 K or 292 K. We
note that several experiments conducted at different temperatures are needed to obtain state information concerning the specific evaporation rate associated
with each temperature level (?). In this work we consider two temperatures, which is one such minimal configuration that contains information sufficient for
determination of thermodynamic data. Similar approaches were applied for the inverse problem of chemical kinetics modelled by the Arrhenius equation,
where chemical reaction rates are temperature dependent (?)

¹⁹⁵removed: Note that to

¹⁹⁶removed: "large enough" in an appropriate sense (??),

¹⁹⁷removed: We remark here that in both our studies

¹⁹⁸removed: of

¹⁹⁹removed: "forward "

²⁰⁰removed: all simulations of the algorithm given in the previous section,

²⁰¹removed: Specifically, the sampled parameters of the posterior distribution represent the model evaluations which produce values within the noise level
of 0.001% of the data concentrations for each of the respective cluster types

²⁰²removed: the

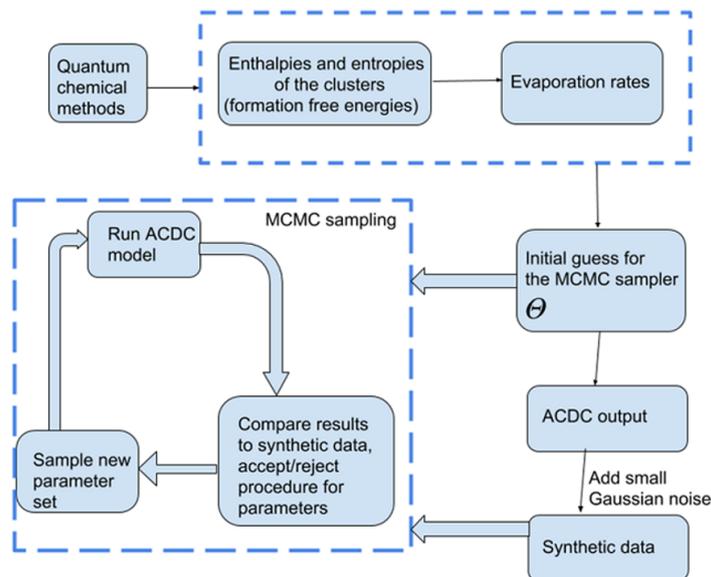


Figure 1. Schematic representation of the study methods.

235 [..²⁰³] A graphical representation of the steady-state [..²⁰⁴] cluster concentration data at 278 K, as a function of the number of acid molecules in the clusters, is given in Figure 2.

Next, [..²⁰⁹] we determine the base 10 logarithms of the evaporation rate coefficients from the synthetic data. Since the noise added to the cluster concentrations results in a random bias towards an increase (or decrease) from the original values produced from the ACDC, the estimates of parameters derived from synthetic data are likely to be biased. In order to average the effects attributed to [..²¹⁰] this random bias, we generated 3 sets of synthetic data by adding random increments to the original concentration measurements. Utilizing these data sets, three independent MCMC runs were conducted, each run containing 3 million parameter samples. An example of one of the sampled chains is depicted in Figs. B1-B2. We omit the initial one million samples, and plot the stationary²¹¹ parts of the chains. As we observe from the plots in Figs. B1-B2, all the parameter chains for the evaporation rates have values bounded above by an upper limit, which differs for different evaporation rates. However, 245 only 15 out of 39 evaporation rates are limited from below (see subfigures labelled 1-5, 7, 10, 12, 16, 18, 22, 27, 31, 33 and 35

²⁰³removed: First, we generate synthetic

²⁰⁴removed: data by the method in Section 2.1, for varying initial ammonia monomer concentrations, previously summarized in Table 2; the sulphuric acid monomer is supplied to the system at a constant rate comprising $6.3 \times 10^4 \text{ s}^{-1}$ at the temperature $T = 278 \text{ K}$. As an output, we obtain the concentrations for all cluster types considered (listed earlier in Table 1), measured when the system has attained the steady-state. A graphical representation of the data set is given above

²⁰⁹removed: from the steady-state data

²¹⁰removed: the

²¹¹Here stationary means that the probability of transitioning from the current state at position j to the new state at position $j + 1$ is independent of j .

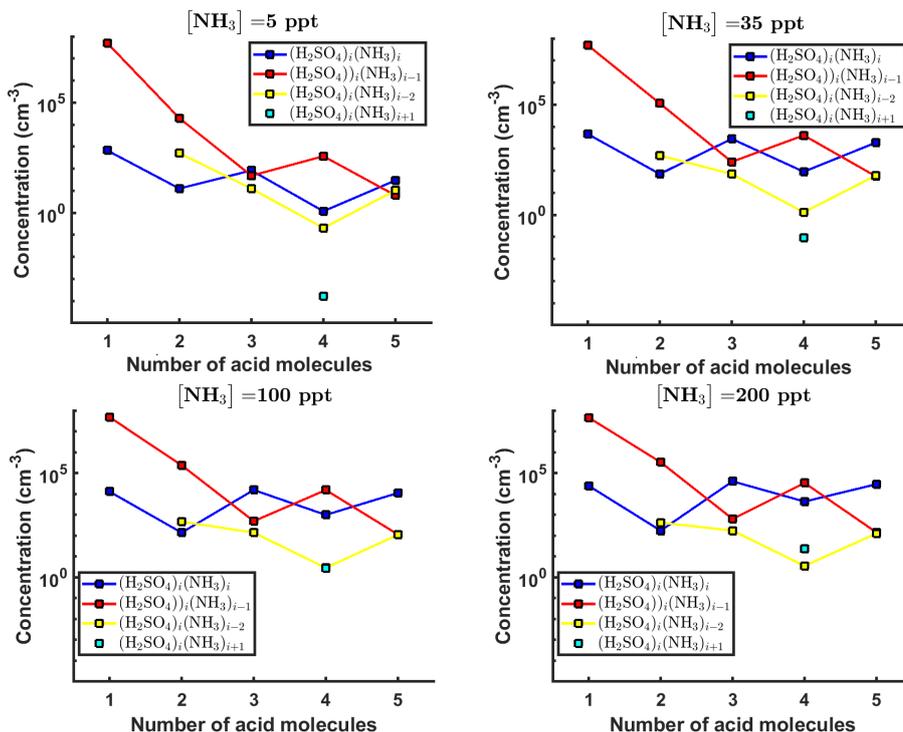


Figure 2. Steady-state cluster concentrations for the clusters containing $[..^{205}]$ sulfuric acid and a varying number of ammonia molecules, as a function of the number of acid molecules, for $[\text{NH}_3]$ $[..^{206}]$ mixing ratios of (a) 5 ppt, (b) 35 ppt, (c) 100 ppt and (d) 200 ppt at the temperature $T=278$ K. The concentrations have been amended with multivariate non-correlated Gaussian noise with standard deviation comprising 0.001% of the original cluster concentration. The source rate of $[..^{207}]$ sulfuric acid monomers is $[\text{H}_2\text{SO}_4] = 6.3 \times 10^4 \text{ s}^{-1}$ $[..^{208}]$.

in Figs. B1-B2). $[..^{212}]$ Notably, all monomer evaporation rates $[..^{213}]$ are bounded from below, except for some of the rates from the largest clusters: H_2SO_4 from $(\text{H}_2\text{SO}_4)_5(\text{NH}_3)_4$ and $(\text{H}_2\text{SO}_4)_5(\text{NH}_3)_5$, and $[..^{214}]$ NH_3 from $(\text{H}_2\text{SO}_4)_5(\text{NH}_3)_5$. $[..^{215}]$

For each evaporation $[..^{216}]$ rate, we calculate the one dimensional (that is, depending only on the evaporation rate) marginal posterior distribution as the position-wise average of the stationary parts of the three sampled chains. This procedure is needed to average the bias originating from random noise. The resulting distributions are given in Figs. 3-4. We use the maximum (also called the mode in the statistics literature) of the posterior marginal distribution function as our parameter estimate in the case

²¹²removed: This subset of evaporation parameters is comprised of the evaporation rates of monomers, with the exception of

²¹³removed: for

²¹⁴removed: the evaporation rate of

²¹⁵removed: These excluded parameters correspond to the evaporations of monomers from the largest and most stable clusters. Note that the estimated lower limits of monomer evaporations from all the clusters except for the most stable ones are far above the 10^{-10} s^{-1} as defined for complete growth.

²¹⁶removed: parameter

when the marginal posterior distributions have precisely one maximum value. In the cases where we have multiple estimators, we provide a range for the evaporation rate values.

255 All the evaporation rates larger than 10^{-3} s^{-1} are well-identified (see subfigures labelled 1, 2, 4, 5, 7, 10, 12, 16, 18, 22, 27, 31 and 35 in Figs. 3- 4), [..²¹⁷]as their estimated variances are well within our accepted error range of less [..²¹⁸]than one order of magnitude. The estimates for the remaining evaporation rates can take values within ranges spanning several orders of magnitude, and are thus uncertain. Also, [..²¹⁹]most of the marginal posterior distributions are non-uniform, except for the evaporation rate of $(\text{H}_2\text{SO}_4)_2(\text{NH}_3)_2$ from $(\text{H}_2\text{SO}_4)_5(\text{NH}_3)_5$. In five cases (refer to subfigures labelled 6, 21, 28, 32 and 36
260 in Figs. 3- 4), the estimated parameter values are not unique[..²²⁰]: the marginal posterior distributions feature multiple modes. The results of our parameter estimation are summarized in Tabs. C1- C2 and in subfigures labelled (a) and (b) in Figure 5.

The pairwise marginal posterior distributions for the estimated evaporation rates are illustrated in Figs. B3-B6. [..²²¹]The majority of the parameters are not correlated. However, the evaporation of monomers from $(\text{H}_2\text{SO}_4)_5\text{NH}_3$, $(\text{H}_2\text{SO}_4)_3(\text{NH}_3)_2$ and $(\text{H}_2\text{SO}_4)_5(\text{NH}_3)_4$ display non-linear inverse correlations. This implies that either H_2SO_4 rarely evaporates (at [..²²²]a
265 rate less than 10^{-4} s^{-1}) and that NH_3 evaporates often, or that the evaporation rates of H_2SO_4 and NH_3 are of comparable magnitude[..²²³]. Additionally, it can be seen from the pairwise posteriors that most of the estimated parameters are highly uncertain. [..²²⁴]

From a mathematical perspective, the existence of multiple distinct parameter estimates indicates that the problem of recovering evaporation rates from the synthetic steady-state concentration data is ill-posed. [..²²⁵]The general solution to this issue
270 is to regularize the problem[..²²⁶], either by adding more data or information to the model[..²²⁷], or by reducing the number of possible estimates.

Based on parameter estimation results, we conclude that a single-temperature steady-state cluster concentrations are not enough to estimate the evaporation rates with a reasonable accuracy (i.e., to obtain an upper and lower limits for the rates that reasonably restrict the cluster kinetics involved in the molecular-level process).

²¹⁷removed: in the sense that

²¹⁸removed: then

²¹⁹removed: notice that

²²⁰removed: ; that is

²²¹removed: From these plots one can see that the majority of

²²²removed: the rate less then

²²³removed: in these cases

²²⁴removed: Therefore, we conclude that in the situation where we determine parameters from the synthetic steady-state data, parameter estimation is not unique.

²²⁵removed: In these situations, one seeks to

²²⁶removed: ; that is, add

²²⁷removed: to reduce

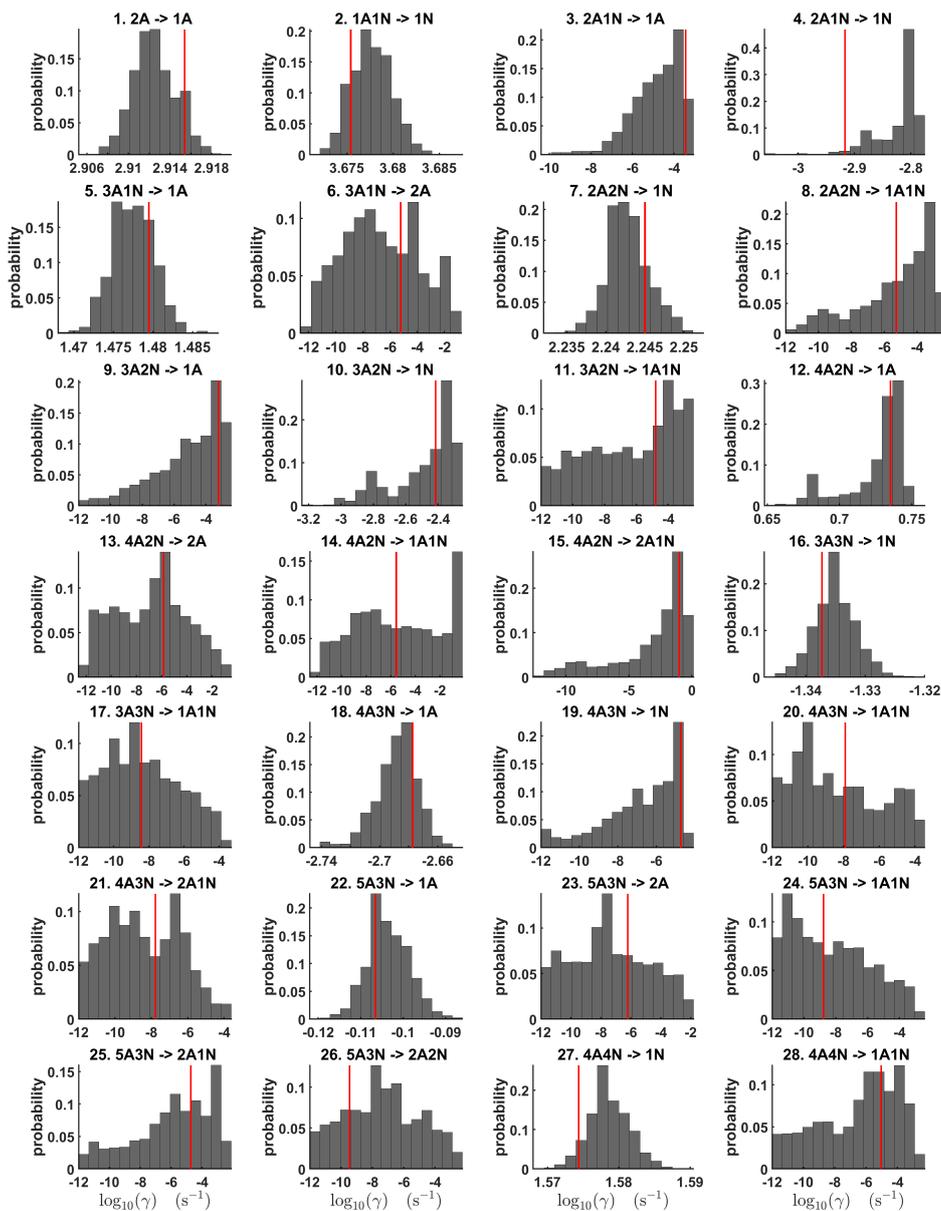


Figure 3. One-dimensional marginal posterior distributions (for parameter indexes ranging from 1 to 28) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) determined from steady-state cluster concentration measurements at the temperature 278 K. Red lines denote the baseline values from [..²²⁸] used to generate the synthetic data. [..²²⁹] The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..²³⁰] y ammonia molecules.

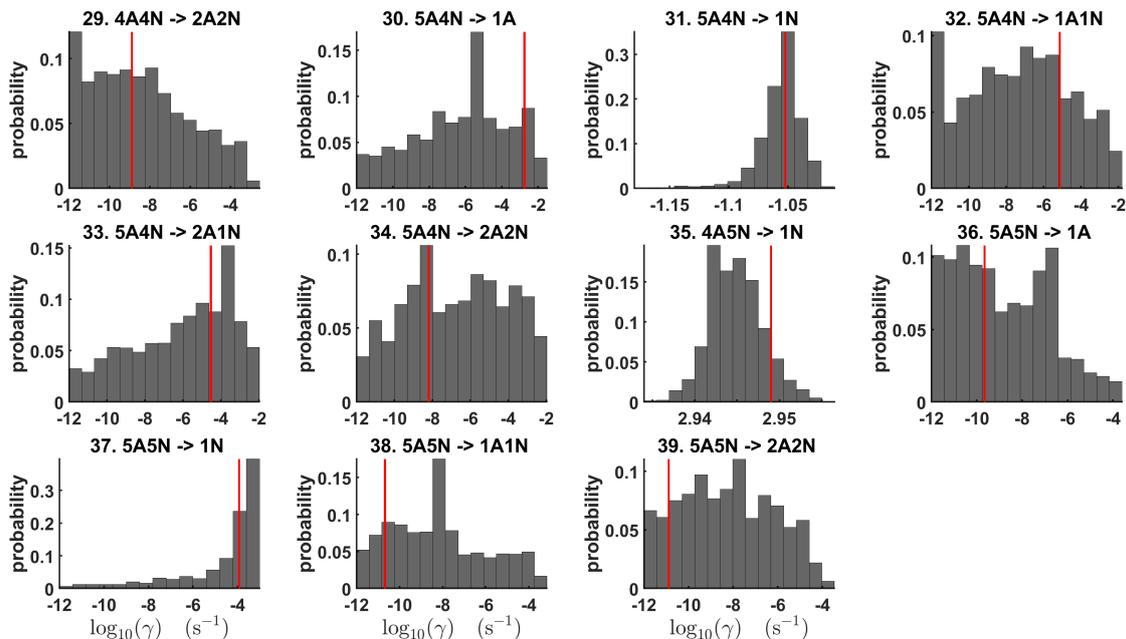


Figure 4. One-dimensional marginal posterior distributions (for parameter indexes ranging from 29 to 39) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) determined from steady-state cluster concentration measurements at the temperature 278 K. Red lines denote the baseline values from [..²³¹] used to generate the synthetic data. [..²³²] The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..²³³] y ammonia molecules.

275 3.2 Identification of [..²³⁴] evaporation rate coefficients from [..²³⁵] time dependent data at a single temperature

[..²³⁶]

The data set for [..²³⁷] time-dependent cluster concentrations is much larger than the data set for steady-state cluster concentrations [..²³⁸], as it contains the concentration values at multiple [..²³⁹] time instances. The time-dependent data also contain information about the [..²⁴⁰] time derivatives of the concentrations [..²⁴¹], (see C1.), which [..²⁴²] should contribute to

280 quantification of [..²⁴³] kinetic parameters (in this case evaporation rates). Our time-dependent cluster concentration data

²³⁵removed: transient

²³⁶removed: First, we extend the synthetic measurement data from steady state concentrations to transient concentrations.

²³⁷removed: transient cluster concentrations at one temperature is

²³⁸removed: at one temperature, as the transient data

²³⁹removed: times instances. Also the transient data

²⁴⁰removed: slope

²⁴¹removed: changing with time

²⁴²removed: contributes

²⁴³removed: the molecular-scale processes (such as collisions and evaporations). We thus expect that this larger data set will reduce the dimension of the solution space for the evaporation rates. Indeed, we will show that this is the case. We generate a synthetic transient

[..²⁴⁴] sets contain in total 656 concentration measurements (corresponding to 16 cluster types and 41 timesteps), for each of the four ammonia mixing ratios.

From this [..²⁴⁵] time-dependent cluster concentration data set, we then conduct [..²⁴⁶] MCMC runs as described in Section 2.2 [..²⁴⁷]. As in the steady-state setting, we conduct three independent MCMC runs to determine the base 10 logarithms of the evaporation rates. One of these runs is presented in Figs. C2-C3. Again, we omit the first one million samples [..²⁴⁸] and merge the stationary parts of the sampled chains to obtain the posterior distributions.

[..²⁴⁹] As seen in Figs. C2-C3, [..²⁵⁰] all the chains have [..²⁵¹] upper limits. Most of the chains are also bounded from below, with five exceptions. [..²⁵²] These exceptions, with arbitrarily large magnitudes, are the evaporation rates of $(\text{H}_2\text{SO}_4)_2(\text{NH}_3)_2$ from $(\text{H}_2\text{SO}_4)_4(\text{NH}_3)_4$ and $(\text{H}_2\text{SO}_4)_5(\text{NH}_3)_3$, and the evaporation rates of H_2SO_4 , [..²⁵³] $(\text{H}_2\text{SO}_4)(\text{NH}_3)$ and $(\text{H}_2\text{SO}_4)_2(\text{NH}_3)_2$ from $(\text{H}_2\text{SO}_4)_5(\text{NH}_3)_5$ [..²⁵⁴].

[..²⁵⁵] The one-dimensional marginal posterior distributions for the estimated parameters are shown in Figs. 6-7. [..²⁵⁶] Most of the estimates are close to the [..²⁵⁷] "true" values used in the generation of the synthetic data. However, the estimated evaporation [..²⁵⁸] rates still feature substantial uncertainties, as their marginal posterior distributions span several orders of magnitude (see subfigures 6, 8, 9, 11, 13, 14, 17, 21, 23-26, 30, 32-34, 37-39 in Figs. 6-7). [..²⁵⁹] The evaporation rate of $(\text{H}_2\text{SO}_4)_2(\text{NH}_3)_2$ from $(\text{H}_2\text{SO}_4)_5(\text{NH}_3)_3$ (which corresponds to subfigure 26) has a uniform posterior distribution, corresponding to an enormous uncertainty. Further, [..²⁶⁰] for the evaporation rates depicted in subfigures 20 and 36 [..²⁶¹], we can only determine upper limits of less than $1.96 \times 10^{-5} \text{ s}^{-1}$. However, [..²⁶²] the time-dependent data allows us conclude that the evaporation processes $(\text{H}_2\text{SO}_4)_4(\text{NH}_3)_3 \rightarrow (\text{H}_2\text{SO}_4)_4(\text{NH}_3)_2 + \text{NH}_3$ and $(\text{H}_2\text{SO}_4)_5(\text{NH}_3)_5 \rightarrow (\text{H}_2\text{SO}_4)_4(\text{NH}_3)_5 + \text{H}_2\text{SO}_4$ can be neglected, as they are relatively slow [..²⁶³] compared with competing evaporation processes.

²⁴⁴removed: set using the method in Section 2.1. The time resolution of our new synthetic data set is 1.5 minutes, which results in 656 total concentration measurements for all the cluster type measured for four different ammonia concentrations. These data sets are illustrated in C1.

²⁴⁵removed: transient

²⁴⁶removed: analogous MCMC runs (

²⁴⁷removed:)

²⁴⁸removed: , which are the samples before the chains have obtained their stationary

²⁴⁹removed: It is shown

²⁵⁰removed: that

²⁵¹removed: the

²⁵²removed: Specifically,

²⁵³removed: $\text{H}_2\text{SO}_4\text{NH}_3$

²⁵⁴removed: have arbitrarily large magnitude

²⁵⁵removed: We examine the

²⁵⁶removed: From these plots, one sees that most

²⁵⁷removed: baseline values used for

²⁵⁸removed: parameters

²⁵⁹removed: Three parameters (subfigures 20, 29 and 36 in Figs. 6-7) have multimodal marginal posterior distributions. We also note that the

²⁶⁰removed: we can only specify that the upper limits

²⁶¹removed: are

²⁶²removed: given the reliable upper estimates,

²⁶³removed: when compared with the other competing

Pairwise marginal posterior distributions for the evaporation rates are plotted in Figs. C4-C8. [..²⁶⁴]Most of the evaporation rates [..²⁶⁵]do not display substantial correlations. However, the evaporation rates of monomers from the cluster $(\text{H}_2\text{SO}_4)_2\text{NH}_3$ display a strong inverse linear relationship, [..²⁶⁶]indicated by the pairwise marginal posterior distribution of the coefficients

305 $(\text{H}_2\text{SO}_4)_2\text{NH}_3 \rightarrow (\text{H}_2\text{SO}_4)_2 + \text{NH}_3$ and $(\text{H}_2\text{SO}_4)_2\text{NH}_3 \rightarrow \text{H}_2\text{SO}_4\text{NH}_3 + \text{H}_2\text{SO}_4$, (see Figure C4). Also, the estimated rate coefficients $(\text{H}_2\text{SO}_4)_2 \rightarrow \text{H}_2\text{SO}_4 + \text{H}_2\text{SO}_4$ and $\text{H}_2\text{SO}_4\text{NH}_3 \rightarrow \text{H}_2\text{SO}_4 + \text{NH}_3$ exhibit linear correlation. [..²⁶⁷]The uncertainties in all the correlated parameters are relatively small (less [..²⁶⁸]than an order of magnitude). [..²⁶⁹]

In Tabs. C1-C2 we summarize the results of parameter estimation for the [..²⁷⁰]two data settings (steady-state and time-dependent) at a single temperature. Note that the estimated upper limits for some of the small evaporation rates (less than
310 10^{-5} s^{-1}) determined from the steady-state data can be as large as $1.55 \times 10^{-2} \text{ s}^{-1}$. This is a poor estimate, since the uncertainties in the synthetic data are small. For example, see the results for parameters shown in subfigures 32 and 34 of Figure 7. In these cases [..²⁷¹], the identification is improved when we [..²⁷²]extend the data set with time-dependent measurements. Overall [..²⁷³], the time-dependent data enabled us to determine the lower bounds for most of the parameters, with the exception of [..²⁷⁴]the parameters shown in subfigures numbered 26 and 29. Moreover, the additional [..²⁷⁵]time-dependent data
315 enabled us to reduce the uncertainties in the estimates of parameters in subfigures 15, 19 and 37. As a result, with the aid of time-dependent data we have improved the estimates of minimal and maximal values for the evaporation rate parameters (see comparison of the 95 % confidence intervals plotted in Figure 5).

[..²⁷⁶]

²⁶⁴removed: Notice that

²⁶⁵removed: of monomers for

²⁶⁶removed: which is

²⁶⁷removed: Additionally, the

²⁶⁸removed: then

²⁶⁹removed: We also remark that from these plots one can see that most of the evaporation rates do not display any substantial correlations.

²⁷⁰removed: above-discussed

²⁷¹removed: the identification has

²⁷²removed: extended

²⁷³removed: one observes that the transient

²⁷⁴removed: those

²⁷⁵removed: time dependent

²⁷⁶removed: In the case of the steady-state cluster concentrations we include only one value for each of the 16 cluster types considered in the study, which were taken when the system has attained a steady state (at the end of the ACDC simulation). The transient data contain the steady-state data as subset. Specifically, in this case we consider the concentrations measured when the system has attained the steady state together with the time-step concentration data measured from the starting point to the end of the ACDC simulation.

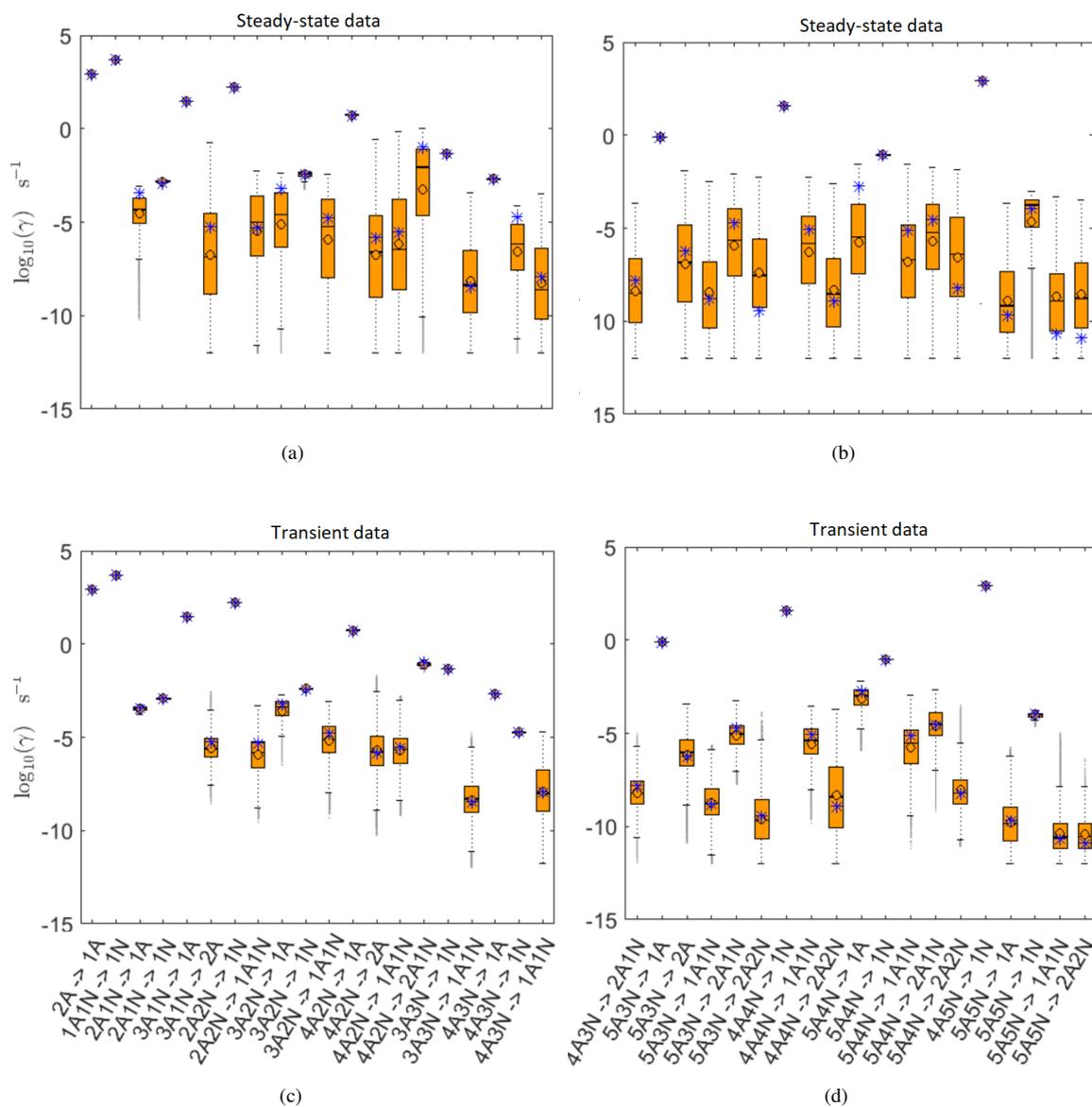


Figure 5. Comparison of 95 % confidence intervals (orange box plots) of base 10 logarithms of the evaporation rates determined from (a)-(b) steady-state and (c)-(d) time-dependent synthetic data measured at temperature 278 K. [..²⁷⁷] Here blue asterisks denote the baseline values used for creating the synthetic data (borrowed from ?). Black circle and horizontal line markers indicate the mode and the mean value of the distribution, respectively. The notation $xAyN$ corresponds to a cluster with x sulfuric acid and y ammonia molecules.

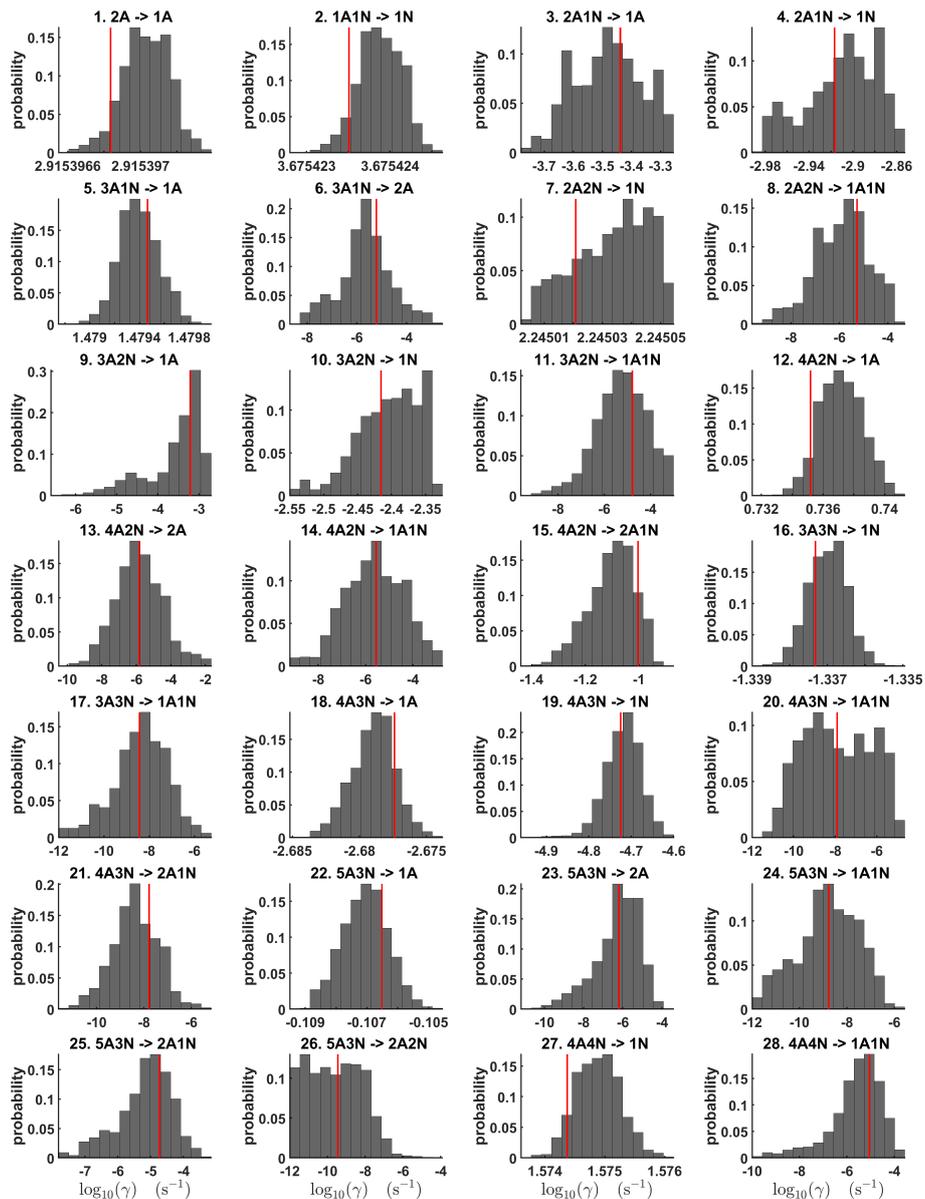


Figure 6. One-dimensional marginal posterior distributions (for parameter indexes ranging from 1 to 28) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) determined from [..²⁷⁸] time-dependent measurements of the cluster concentrations with time resolution comprising 1.5 minutes at the temperature 278 K. Red lines denote the baseline values from [..²⁷⁹] used to generate the synthetic data. [..²⁸⁰] The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..²⁸¹] y ammonia molecules.

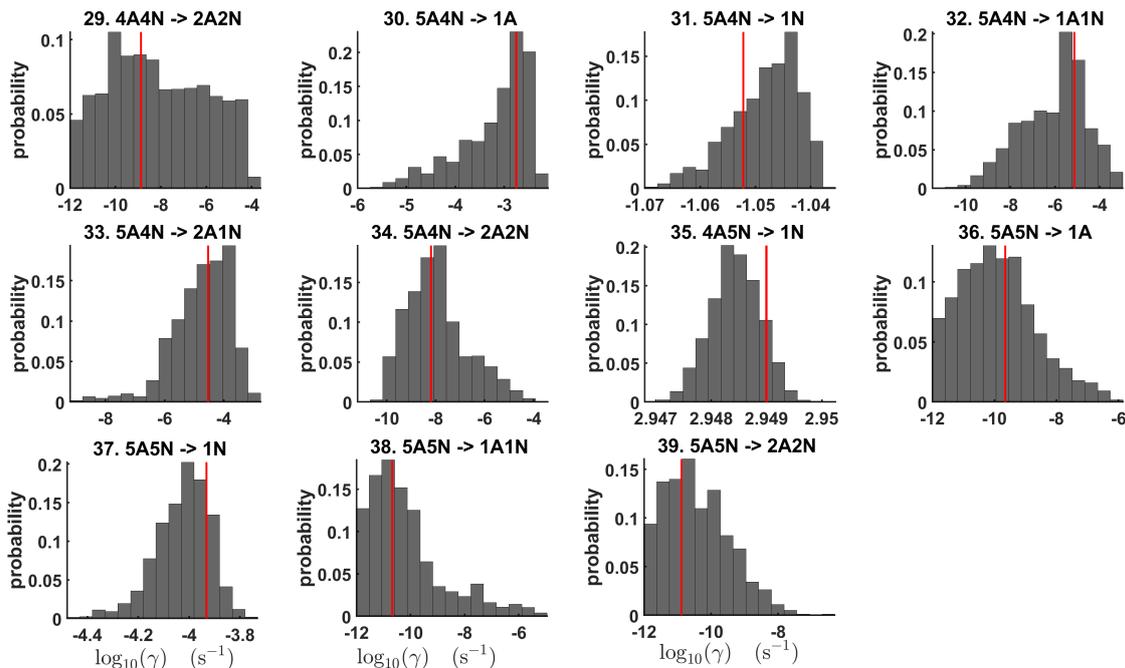


Figure 7. One-dimensional marginal posterior distributions (for parameter indexes ranging from 29 to 39) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) determined from [..²⁸²]time-dependent asurements of the cluster concentrations with time resolution comprising 1.5 minutes at the temperature 278 K. Red lines denote the baseline values from [..²⁸³] used to generate the synthetic data. [..²⁸⁴]The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..²⁸⁵] y ammonia molecules.

3.3 Estimating [..²⁸⁶] formation enthalpies and entropies from steady-state concentration measurements at multiple temperatures

320

[..²⁸⁷] We determined cluster formation enthalpies and entropies [..²⁸⁸] based on two sets of [..²⁸⁹] steady-state cluster concentrations, [..²⁹⁰] corresponding to two temperatures: 278 and 292 K. [..²⁹¹] These data sets are plotted in Figs. 2 and D1 for 278 K and 292 K, respectively.

[..²⁹²]

²⁸⁶removed: thermodynamic data

²⁸⁷removed: In this section we describe another method for regularizing our problem of estimating evaporation rates from steady-state concentration data.

We will determine the

²⁸⁸removed: from

²⁸⁹removed: synthetic,

²⁹⁰removed: now measured at

²⁹¹removed: This data set is

²⁹²removed: We will demonstrate that reparameterization (in terms of thermodynamic data) plus the extended data set transforms our parameter estimation problem from an ill-posed problem to a well-posed one. We use synthetic steady-state cluster concentrations generated for two temperatures to recover the

325 [..²⁹³] As in the previous sections, three MCMC runs were conducted to average the bias attributed to random noise [..²⁹⁴]. An example of [..²⁹⁵] the sampled chains is [..²⁹⁶] shown in Figure D2. It can be seen that all the chains are bounded, with the exception of the formation enthalpy and entropy of the [..²⁹⁷] largest cluster ($(\text{H}_2\text{SO}_4)_5(\text{NH}_3)_5$).

[..²⁹⁸] The one-dimensional [..²⁹⁹] marginal posterior distributions of [..³⁰⁰] the formation enthalpies and entropies, built from the stationary parts of the three sampled chains merged together, [..³⁰¹] are shown in Figure 9. [..³⁰²] For all the clusters except $(\text{H}_2\text{SO}_4)_5(\text{NH}_3)_5$, the variances of the [..³⁰³] estimated formation enthalpies are less than $0.46 \text{ kcal mol}^{-1}$, while the estimated formation entropies vary at most by $5.4 \text{ cal K}^{-1}\text{mol}^{-1}$. The estimated free parameters together with the [..³⁰⁴] "true" quantum chemistry-based values from [..³⁰⁵] ? used for generation of the synthetic data are summarized in Table D1.

Although the posterior distributions of [..³⁰⁶] the formation enthalpies and entropies of $(\text{H}_2\text{SO}_4)_5(\text{NH}_3)_5$ feature higher uncertainties in comparison to [..³⁰⁷] those of the smaller clusters, the evaporation rates [..³⁰⁸] from $(\text{H}_2\text{SO}_4)_5(\text{NH}_3)_5$, as calculated from the aforementioned posterior distributions, have low variances, see Table D3.

[..³⁰⁹]

Additionally, strong correlations are observed between formation enthalpies [..³¹⁰] and entropies of clusters containing the same number n of ammonia molecules [..³¹¹] when $n > 2$, except the case of $(\text{H}_2\text{SO}_4)_5(\text{NH}_3)_5$. [..³¹²]

[..³¹³]

thermodynamic parameters. This is done to improve the identification by using the temperature dependence of the Gibbs free energies (and the evaporation rates).

²⁹³removed: For each temperature choice, we use the methods described in Section 2 to obtain synthetic steady-state cluster concentration data. We summarize this data in Table 2; the data sets are plotted in Figure 2 for 278 K and D1 for 292 K. Three

²⁹⁴removed: added to the data, as discussed in the previous section

²⁹⁵removed: one of

²⁹⁶removed: illustrated

²⁹⁷removed: biggest

²⁹⁸removed: Next we consider the

²⁹⁹removed: (depending on the particular cluster formation entropy or enthalpy parameters)

³⁰⁰removed: free parameters

³⁰¹removed: see

³⁰²removed: It can be seen that for

³⁰³removed: variance for the

³⁰⁴removed: baseline

³⁰⁵removed: ?

³⁰⁶removed: sampled thermodynamic parameters for

³⁰⁷removed: the corresponding posterior distributions identified for the

³⁰⁸removed: for evaporations

³⁰⁹removed: Notice that the evaporation rates for all the molecular clusters calculated from a posterior distribution of sampled thermodynamic parameters for the temperature 278 K are close to the baseline values from ? used for generation of the synthetic data and their variances are less than one order of magnitude, see Figs. D6-D7.

³¹⁰removed: (entropies) of the clusters containing same number

³¹¹removed: larger then 2

³¹²removed: Since our parameters are strongly correlated, we may alternatively consider just cluster formation enthalpies or the ratios of cluster formation entropies and enthalpies as our free parameters.

³¹³removed: Pairwise marginal posterior distributions (for parameter indexes ranging from 1 to 8) of the cluster formation enthalpies and entropies determined from steady-state cluster concentration measurements at two temperatures $T=278 \text{ K}$ and $T = 292 \text{ K}$. Red rectangles denote the baseline values from

340 [..³¹⁴]

3.4 [..³¹⁵]

[..³¹⁶]

[..³¹⁷]

[..³¹⁸]

345 [..³¹⁹]

3.4 [..³²⁰]

[..³²¹]

[..³²²]

? used to generate the synthetic data. Here the symbols ΔH and ΔS stand for cluster formation enthalpies and entropies, respectively. Symbols "A", "N" denote H_2SO_4 and " NH_3 ", correspondingly.

³¹⁴removed: One-dimensional marginal posterior distributions of the cluster formation enthalpies (units given in kcal/mol) and entropies (units given in $\text{cal K}^{-1} \text{mol}^{-1}$) determined from steady-state cluster concentration measurements at two temperatures $T=278 \text{ K}$ and $T = 292 \text{ K}$. Red lines denote the baseline values from ? used to generate the synthetic data. Here the symbols ΔH and ΔS stand for cluster formation enthalpies and entropies, respectively. Symbols "A", "N" denote H_2SO_4 and " NH_3 ", correspondingly.

³¹⁵removed: Comparison to previous evaporation rate determinations

³¹⁶removed: The evaporation rates can be obtained either experimentally or computationally, when applying the Quantum Chemical (QC) methods, (?). Experimental detection was conducted from the measurements in a flow tube (???) and in the CLOUD chamber (????). The summary of thermodynamic parameters obtained from different methods has previously been published in ?. These parameters can be employed to calculate the evaporation rates at different temperatures.

³¹⁷removed: In this study we determine the evaporation rates and thermodynamic data from measurements of cluster concentrations. Supplementary to the methodology presented in ?, our first method enables to determine parameters from the time-dependent cluster concentrations measured before the system has attained the steady state. The transient data improved the estimates for all the evaporation rates.

³¹⁸removed: In the second method we identify thermodynamic parameters from the steady-state cluster concentrations measured at two different temperatures. This approach is similar to ?, but our model takes into account all the possible evaporation processes. In ? the thermodynamic parameters had been determined from the New Particle Formation Rates (NPFs) measured at different temperatures. Instead of the NPFs, we employ the measurements of cluster concentrations. By so doing, we find the combination of data and fitted parameters which enables to determine the evaporation rates with the variances comprising less than one order of magnitude.

³¹⁹removed: Although the transient data have improved the estimates, the temperature-dependent data have been demonstrated to yield the most accurate estimates of the evaporation rates, when we treat cluster formation enthalpies and entropies as free parameters.

³²⁰removed: Discussion and future work

³²¹removed: The MCMC results are not specific for the simulation box considered in the present study, but rather general. This is supported by the fact that although the size of the system (the number of clusters included into simulations) has impact on the particle formation rates at high temperatures ($> 278 \text{ K}$), the particle formation rates and cluster concentrations produced using different simulation boxes are qualitatively similar. Thus the changes of the ACDC outputs due to the difference in the simulation box does not change for MCMC parameter estimation results. In ? it was shown that the 5×5 simulation box (which is used for generation of the synthetic data) produces reasonable results with a good agreement with the measurements obtained from the CLOUD chamber experiment. Additionally, the boundary conditions for the outgrowing clusters (the choice of the clusters that are considered as formed particles) has only minor influence on the simulation results, given that the simulated system of clusters is defined in a reasonable way (?).

³²²removed: In general, the accuracy of the MCMC results increases when we include additional data. In particular, including more concentration data measured at different ammonia concentrations will yield better estimates for the evaporation rates. The sensitivity of the estimates to the number of ammonia

[..³²³]
350 [..³²⁴]These strong correlations are consistent with general principles of clustering thermodynamics. If a cluster has very strong bonds between [..³²⁵]its constituent molecules, then [..³²⁶]the formation enthalpy is very negative, and also the intermolecular vibrational frequencies corresponding in a broad sense to vibrations involving those bonds [..³²⁷]are fairly high, meaning that the entropy loss in forming the cluster is large. These intermolecular frequencies dominate the "variable part" of the formation entropy, as the entropy [..³²⁸]change from the loss of translational and rotational degrees of freedom is almost a constant factor[..³²⁹]. Thus, if the formation enthalpy of a cluster is very negative, so is also the formation entropy. Conversely, if the cluster is only quite weakly bound, the formation enthalpy is only slightly negative, and the intermolecular frequencies can be very low, leading to a less negative (though still negative[..³³⁰]) formation entropy[..³³¹]
[..³³²]. Evaporation rates for all the molecular clusters calculated from a posterior distribution of sampled formation enthalpies and entropies are close to the "true" values used for generation of the synthetic data at both temperatures
360 (278 K and 292K) and their variances are less than one order of magnitude, see Figs. D6-D7. Thus, reparametrization of evaporation rates in terms of formation enthalpies and entropies, and use of data at two different temperatures, thus transforms our parameter estimation problem from an ill-posed to a well-posed one.

concentrations will be considered in the future work. In the present study we rather focus on the question which combination of estimated parameters and concentration data will produce an accurate estimates for the evaporation rate.

³²³removed: The data of steady-state concentration with two temperatures allowed us to apply two general principles of inverse problems/Bayesian estimation to the problem of estimating evaporation rates. First, the two temperature data set enabled us to reformulate the problem in a numerically effective way (in terms of enthalpy and entropy) that reduced the number of unknown parameters we sought to estimate. Second, the reformulated differential equation describing the time evolution of the concentrations was more numerically stable than the original expression (the stiffness of the equation was reduced in the reformulated form). This made our estimates for the rates less sensitive to small perturbations/errors.

³²⁴removed: In addition, the fact that the formation entropies and enthalpies were strongly correlated made them an effective parametrization. The strong inverse correlations have a physical explanation. Firstly, both formation enthalpy and entropy follow from the partition function of the molecular complex, and their functional forms are partly similar (?). Practically, if a cluster has really

³²⁵removed: the

³²⁶removed: that means

³²⁷removed: (note that these

³²⁸removed: effect

³²⁹removed:) are fairly high, meaning that the entropy loss in forming the cluster is large. So

³³⁰removed: of course

³³¹removed: (?).

³³²removed: Note that experimental data can differ from the synthetic data in the sense that they contain noise which originate from measurement instruments and uncertainties associated with experimental conditions (e. g. , in CLOUD chamber experiments). Treating the noise inherent for experimental data will be the topic of our future studies

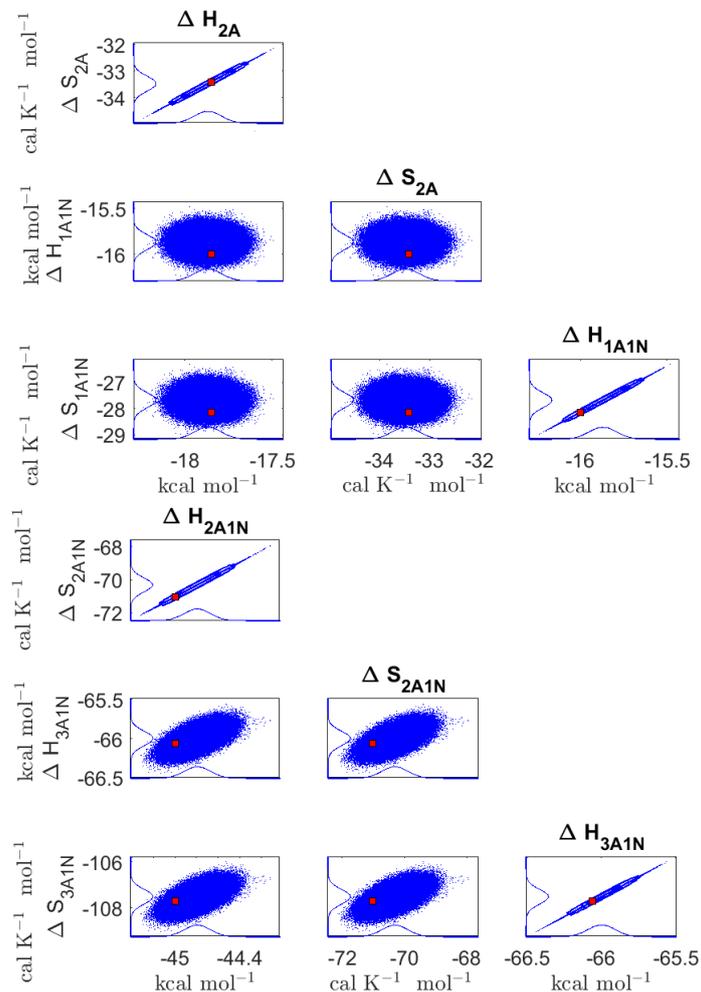


Figure 8. Pairwise marginal posterior distributions (for parameter indexes ranging from 1 to 8) of the cluster formation enthalpies and entropies determined from steady-state cluster concentration measurements at two temperatures $T=278$ K and $T = 292$ K. Red rectangles denote the baseline values from ? used to generate the synthetic data. Here the symbols ΔH and ΔS stand for cluster formation enthalpies and entropies, respectively. The notation $xAyN$ corresponds to a cluster with x sulfuric acid and y ammonia molecules.

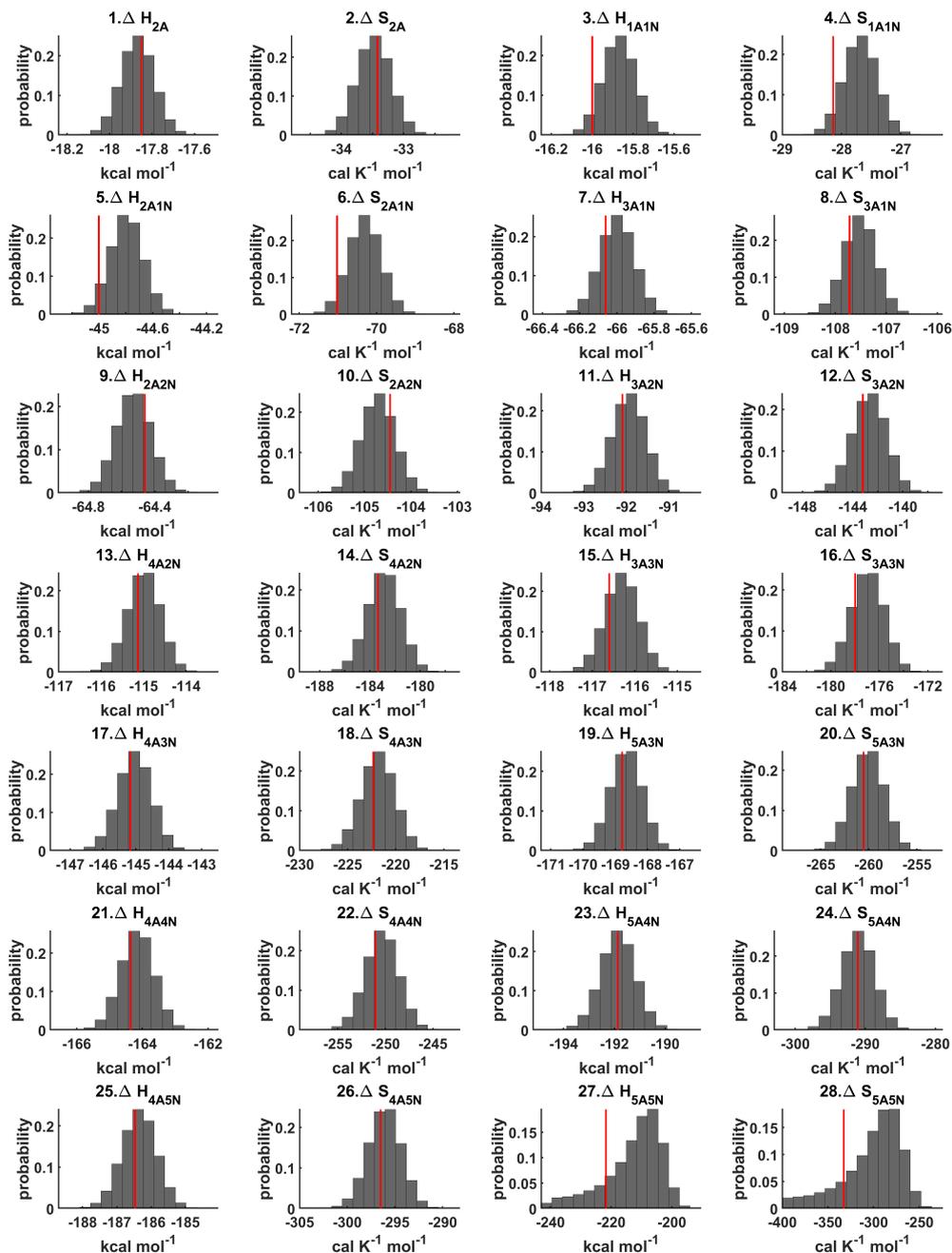


Figure 9. One-dimensional marginal posterior distributions of the cluster formation enthalpies (units given in kcal/mol) and entropies (units given in $\text{cal K}^{-1} \text{mol}^{-1}$) determined from steady-state cluster concentration measurements at two temperatures $T=278 \text{ K}$ and $T = 292 \text{ K}$. Red lines denote the baseline values from ? used to generate the synthetic data. Here the symbols ΔH and ΔS stand for cluster formation enthalpies and entropies, respectively. The notation $xAyN$ corresponds to a cluster with x sulfuric acid and y ammonia molecules.

4 Conclusions

[..³³³]

365 We applied Bayesian parameter estimation using a Markov chain Monte Carlo (MCMC) algorithm to identify cluster evaporation/fragmentation rates from [..³³⁴]synthetic cluster distribution data[..³³⁵], assuming that the cluster collision rates are known. We used the Atmospheric Cluster Dynamic Code (ACDC) [..³³⁶]together with evaporation rates based on quantum chemistry and detailed balance to generate synthetic data for the purpose of optimizing and validating the parameter estimation.

370 First, we sought to determine the cluster evaporation rates from both steady-state and time-dependent cluster concentration data at one temperature. [..³³⁷]We were only able to identify a subset of the free parameters (evaporation rates) from the available data [..³³⁸]using either of these approaches.

[..³³⁹]Next, we used [..³⁴⁰]steady-state concentration data [..³⁴¹]corresponding to two different temperatures. We introduced a reparametrization [..³⁴²]which expressed the evaporation rates in terms of temperature and cluster formation enthalpies and entropies[..³⁴³]. Using steady-state concentrations at two temperatures allowed us to apply two general principles of inverse problems/Bayesian estimation to the problem of estimating evaporation rates. First, the two-temperature data set enabled us to reformulate the problem in a numerically effective way (in terms of formation enthalpies and entropies), which reduced the number of unknown parameters. This reduced the number of parameters we sought to identify. [..³⁴⁴]Second, it also lessened the stiffness of the system, as the cluster formation enthalpies and entropies for our system [..³⁴⁵]span a much smaller range compared to the evaporation rates. We demonstrated that steady-state concentration data at two different temperatures could be used to determine all the unknown formation enthalpies and entropies, and thus the evaporation rates, to within acceptable accuracy.

In general, the accuracy of the MCMC results naturally increases when we include additional data. In particular, including more concentration data measured at different ammonia concentrations will yield better estimates for the evaporation rates. The sensitivity of the estimates to the number of ammonia concentrations, as well as different sulfuric acid source rates, will be considered in future work.

³³³removed: We applied a

³³⁴removed: known

³³⁵removed: and known

³³⁶removed: with quantum chemistry based evaporation rates

³³⁷removed: In this first scenario, we sought to determine the cluster evaporation rates from both steady-state and time-dependent cluster concentration data. Due to the mathematical stiffness of the ordinary differential equations describing the time evolution of the cluster concentrations, we

³³⁸removed: . This stiffness originates from the vastly different timescales of some of the key evaporation rates

³³⁹removed: In the second scenario

³⁴⁰removed: only

³⁴¹removed: but for

³⁴²removed: expressing

³⁴³removed: , and temperature

³⁴⁴removed: It

³⁴⁵removed: have comparable orders of magnitude

The approach presented here can also be applied to infer evaporation rates from mass spectrometric measurements of molecular cluster concentrations. This naturally requires accounting for the process of charging neutral clusters, with its associated instrumental and data-analysis-related uncertainties. A clear conclusion of our proof-of-concept study is that steady-state data at different temperatures is more useful for determining evaporation rates than time-dependent data at a single temperature. Moreover, reliable steady-state concentrations of clusters at various temperatures are generally easier to obtain experimentally (e.g. in chamber experiments) compared to time-dependent concentrations. This finding demonstrates the more general feature of modelling of the type performed here: it can be used to optimize planning of experiments, and thus save both time and resources. Determining very low (below 10^{-5} s^{-1}) evaporation rates may also require additional measurements at low vapor concentrations, which naturally require longer timescales to reach a steady state. Treating the uncertainties inherent in experimental data will be the topic of our future studies.

Code availability. The code is available via GitHub repository: <http://doi.org/10.5281/zenodo.3766925>

A Supplementary mathematical material

A1 Cluster ³⁴⁶kinetics

400 The kinetics of cluster formation is described by Becker-Döring equations (??), which model cluster birth and death which arises from collisions of the smaller clusters into larger ones and evaporations from the bigger clusters into smaller ones. Precisely, labelling the clusters by $i \in \{1, 2, \dots, N\}$, the time derivative of the i th cluster concentration Y_i is governed by

$$\frac{dY_i}{dt} = \frac{1}{2} \sum_{j < i} \beta_{i,(i-j)} Y_i Y_{i-j} + \sum_j \gamma_{i+j \rightarrow i,j} Y_{i+j} - \sum_j \beta_{i,j} Y_i Y_j - \frac{1}{2} \sum_{j < i} \gamma_{i \rightarrow j,i-j} Y_i + Q_i - S_i, \quad (\text{A1})$$

where $\beta_{i,j}$ is the collision coefficient of clusters i with j , and $\gamma_{i+j \rightarrow i,j}$ is the evaporation coefficient of cluster $i+j$ into clusters i and j , Q_i is an external source term of i , and S_i represents the total possible types of losses for the cluster of type i . These last two terms, which stand for external supply and destruction mechanisms, depend on the system under consideration.

We now specify the quantity and type of sinks and sources included in our studies. We assume that the concentration of ammonia monomers is constant, while ³⁴⁷sulfuric acid monomers are supplied to the system at a constant rate comprising $Q = 6.3 \times 10^4 \text{ cm}^{-3}\text{s}^{-1}$. This settings are selected to imitate the conditions inside of the CLOUD chamber, (??). Further, we include wall losses arising from clusters sticking on the walls of the experimental chamber, (?). These wall losses are parametrized by the size of the cluster

$$S_{\text{wall},i} = 10^{-12} / (2r_i + 0.3 \times 10^{-9}) \text{ s}^{-1}, \quad (\text{A2})$$

where r_i is the mass radius of the cluster (in cm). From Eq. A2, wall loss rates decrease with cluster size; in practise it also varies with respect to cluster position in the chamber and time. We neglect any uncertainties attributed to the wall losses. However, we do account for dilution losses, with size-independent value comprising $S_{\text{dil},i} = 9.6 \times 10^{-5} \text{ s}^{-1}$, which had previously been determined in the CLOUD chamber, (??).

Let T denote the temperature of the system of molecular clusters. Using classical kinetic gas theory, the collision rates $\beta_{i,j}$ in Eq. A1 obey

$$\beta_{i,j} = \sqrt{T} \left(\frac{3}{4\pi} \right)^{1/6} \left[6k_B \left(\frac{1}{m_i} + \frac{1}{m_j} \right) \right]^{1/2} \left(V_i^{1/3} + V_j^{1/3} \right)^2, \quad (\text{A3})$$

420 where m_i and V_i are respectively the mass and volume of cluster i , and k_B is Boltzmann's constant. In this paper, we assume that the masses and volumes are temperature-independent.

The cluster evaporation rates $\gamma_{i+j \rightarrow i,j}$ in Eq. A1 are given by the expression

$$\gamma_{i+j \rightarrow i,j} = \beta_{i,j} \frac{P_{\text{ref}}}{k_B T} \exp \left(\frac{\Delta G_{i+j} - \Delta G_i - \Delta G_j}{k_B T} \right), \quad (\text{A4})$$

³⁴⁶removed: kinematics

³⁴⁷removed: sulphuric

where P_{ref} is the reference pressure and ΔG_i is the Gibbs free energy of formation for cluster i . We may further describe the
 425 i th Gibbs free energy in terms of the cluster formation enthalpy ΔH_i and entropy ΔS_i :

$$\Delta G_i = \Delta H_i - T\Delta S_i. \quad (\text{A5})$$

We neglect here the weak temperature dependence of real cluster formation enthalpies and entropies.

A1.1 The Metropolis algorithm

We first select the flat prior distribution from which we will initially sample unknown parameters, as we wish to generate
 430 physically reasonable parameter estimates. Therefore, we generate unknown parameters within the chosen minimum
 and maximum bounds where all the points are equally likely to be sampled. Please see Section 2.2.3 and Tabs. 3-4 for
 more details. From the prior distribution, a starting guess for the parameters $\theta_{\text{old}} \in \mathbb{R}^{n_{\text{coef}}}$ is chosen (here n_{coef} is the
 total number of parameters).

A2 [..³⁴⁸]

435 [..³⁴⁹]The Metropolis algorithm then requires us to specify how to sample new parameter values θ_{new} . This is done by
 choosing a proposal distribution. We chose a multivariate Gaussian proposal density q , defined by:

$$[\text{..}^{350}]q([\text{..}^{351}]\theta_{\text{old}}, \theta_{\text{new}})[\text{..}^{352}] \simeq \exp[\text{..}^{353}] \left(-\frac{1}{2}[\text{..}^{354}][\text{..}^{355}] (\theta_{\text{new}} - \theta_{\text{old}})^T \Sigma^{-1} (\theta_{\text{new}} - \theta_{\text{old}}) \right), \quad (\text{A6})$$

where [..³⁵⁶] Σ is a covariance matrix (of dimensions $n_{\text{coefs}} \times n_{\text{coefs}}$) which specifies the scaling and spatial orientation of
 the Gaussian proposal distribution. As the normalization constants are cancelled out in Eq. A9, we do not take them into
 440 consideration.

Next, we run the ACDC and Fortran simulations with the parameter values θ_{new} . We collect the cluster concentration
 outputs in the column-vector $\mathbf{y}_{\text{mod}}(\theta_{\text{new}}) \in \mathbb{R}^{n_{\text{out}}}$, where n_{out} is the number of [..³⁵⁷]elements. The candidate vector of
 parameters θ_{new} is either accepted or rejected according to the least-squares fit of $\mathbf{y}_{\text{mod}}(\theta_{\text{new}})$ to the synthetic cluster
 concentrations \mathbf{y}_{exp} :

$$445 \quad SS(\theta_{\text{new}}) = \sum_{i=1}^{n_{\text{out}}} \frac{(y_{\text{exp},i} - y_{\text{mod},i}(\theta_{\text{new}}))^2}{\sigma_i^2}, \quad (\text{A7})$$

where n_{out} is the number of [..³⁵⁸]

³⁴⁸removed: Likelihood, data and cost function

³⁴⁹removed: The likelihood of observing the data \mathbf{Y}_{exp} given the parameter values θ is

³⁵⁶removed: n_{out} is the number of measurements and $F(\theta)$ is the cost function. We elucidate the cost function below. In our first study in which simulations
 are conducted with time-dependent data, the number of measurements is $n_{\text{out}} = 4 * (N_c * N_t + 1)$, where $N_c = 16$

³⁵⁷removed: cluster types whose concentrations are measured and $N_t = 41$

³⁵⁸removed: time-step measurements available for each of the cluster types. As explained in Section 2.1, after each VODE integration, a convergence
 coefficient is computed from the steady-state cluster concentrations to ensure that the system has attained the steady-state.

[..³⁵⁹] measurements in the [..³⁶⁰] / [..³⁶¹] / [..³⁶²] synthetic concentrations. [..³⁶³]

[..³⁶⁴]

450 By construction our synthetic data contains uncorrelated Gaussian measurement error, hence the likelihood of observing the data \mathbf{y}_{exp} given some parameter values θ is

$$p(\mathbf{y}_{exp}|\theta) \simeq \exp\left(-\frac{1}{2}SS(\theta)\right). \quad (\text{A8})$$

[..³⁶⁵] The value $SS(\theta_{new})$ is then compared to the least-square sum from the previous step $SS(\theta_{old})$ and accepted with the probability

$$p_{acc}(\theta_{old}, \theta_{new}) = \min \left\{ 1, \frac{p(\mathbf{y}_{exp}|\theta_{new})}{p(\mathbf{y}_{exp}|\theta_{old})} \right\} = \min \left\{ 1, \exp \left[-\frac{1}{2} (SS(\theta_{new}) - SS(\theta_{old})) \right] \right\}. \quad (\text{A9})$$

455 If θ_{new} is accepted, this parameter combination is added as the next element in the chain; else the old value is replicated in the chain. Finally, the value $SS(\theta_{old})$ is replaced with $SS(\theta_{new})$ and saved. This completes an iteration of the Metropolis algorithm.

[..³⁶⁶] We remark here that the likelihoods $p(\mathbf{y}_{exp}|\theta_{old})$ and $p(\mathbf{y}_{exp}|\theta_{new})$ in Eq. A9 characterize how closely the outputs of the ACDC simulations with the parameters θ_{old} and θ_{new} respectively fit the synthetic data. By definition of the acceptance probability $p_{acc}(\theta_{old}, \theta_{new})$ in Eq. A9, the candidate step is always accepted if the new parameters fit the data at least as good as the old values ($SS(\theta_{new}) \leq SS(\theta_{old})$).

A1.1 The DRAM algorithm for sampling from large parameter space

Our implementation of the Delayed Rejection Adaptive Metropolis (DRAM) (??) approach to MCMC parameter estimation modifies the above Metropolis algorithm in the following way.

465 First, we use the Adaptive Metropolis (AM) (?) method for updating the covariance matrix Σ of the proposal distribution $q(\theta_{old}, \theta_{new})$ in Eq. A6. That is, if we have generated samples $(\theta_0, \theta_1, \dots, \theta_{n-1})$, the next candidate set θ_{new} is proposed from $q(\theta_{n-1}, \theta_{new})$ using the empirical covariance $\Sigma = \text{Cov}(\theta_0, \theta_1, \dots, \theta_{n-1})$. Therefore the next candidate set is generated by taking a step with direction and size determined from the values of parameters previously sampled in the MCMC

³⁵⁹removed: In our first study,

³⁶⁰removed: parameter fit to the data was evaluated by the sum of squared residuals of the model outputs \mathbf{Y}_{mod} and the measurements, \mathbf{Y}_{exp} . The

³⁶¹removed: cost function

³⁶²removed: (sum of squared residuals) measures how far our model outputs are from the “true” experimental outputs

³⁶³removed: Precisely,

³⁶⁵removed: Since concentrations of molecular clusters span a large range (from 10^{-5} to 10^9 particles per cm^3), we normalize the residuals by the measurement error variance $\sigma_{j_i}^2$. Normalization in this way avoids overfitting to the larger concentration values. Note also that the error variance $\sigma_{j_i}^2$ is matched separately for each cluster type and every time instance. We assume that the instrument is capable of detecting all the cluster types represented in the system at arbitrary small levels of concentration. This simplification was considered in order to illustrate the proposed approach.

³⁶⁶removed: When parameter estimation is conducted with steady-state cluster concentrations (as is considered in our second study), we use the following cost function:

chain. This procedure is carried out after every 100 successive accept/reject iterations. To ensure computational stability, we also apply additional scaling and regularization for the proposal covariance (see ??); please see ? for a detailed explanation.

Second, we carry out local adaptation of the proposal distribution using the Delayed Rejection (DR) algorithm (?). It is implemented as follows: given n parameter sets $(\theta_0, \theta_1, \dots, \theta_n)$ generated by the AM method above, a candidate θ_{new} is proposed from the distribution $q(\theta_n, \theta_{\text{new}})$ in Eq. A6 and accepted with probability as in Eq. A9, as discussed before. However, if the proposed θ_{new} is rejected, instead of replicating the previous values in the MCMC chain (i.e., $\theta_{n+1} = \theta_n$), the algorithm tests a new candidate move $\theta_{\text{new},2}$ which is close to the current estimate θ_n . Then the second-stage proposal $\theta_{\text{new},2}$ is accepted with appropriately adjusted acceptance probability (see ?).

In summary, our application of the DRAM algorithm combines the AM procedure with a two-stage DR modification. In the first stage, our algorithm carries out the Metropolis regime with both AM adaptation. The proposal covariance at the initialization of DR (denoted as Σ) is computed as by AM method above, no matter at which stage of DR these points have been accepted in the sampling process. The covariance of the proposal for the second stage (denoted as Σ^2) is always computed as the scaled version of the first-stage proposal covariance:

$$[\dots]^{367} [\dots]^{368} \Sigma^2 = [\dots]^{369} \gamma \Sigma, \tag{A10}$$

with the scaling coefficient $\gamma = 1/5$ that was chosen to increase the number of accepted candidate steps at the second stage (?).

This DRAM parameter estimation was conducted using the 'mcmcstat' toolbox implemented for FORTRAN (?). See the description and the examples of usage on the web page helios.fmi.fi/~lainema/.

³⁷⁰removed: Now $N_T = 2$ denotes the number of steady state configurations at different

³⁷¹removed: temperatures

³⁷²removed: (not times!) and T_j stands for the measured

³⁷³removed: temperature

³⁷⁴removed: . In this study, the number of measurements for

³⁷⁵removed: likelihood given by Eq. . ?? is $n_{\text{out}} = 4 * (N_c * N_T + 1)$ (again $N_c = 16$ cluster types).

B Estimation of the evaporation rates from steady-state data

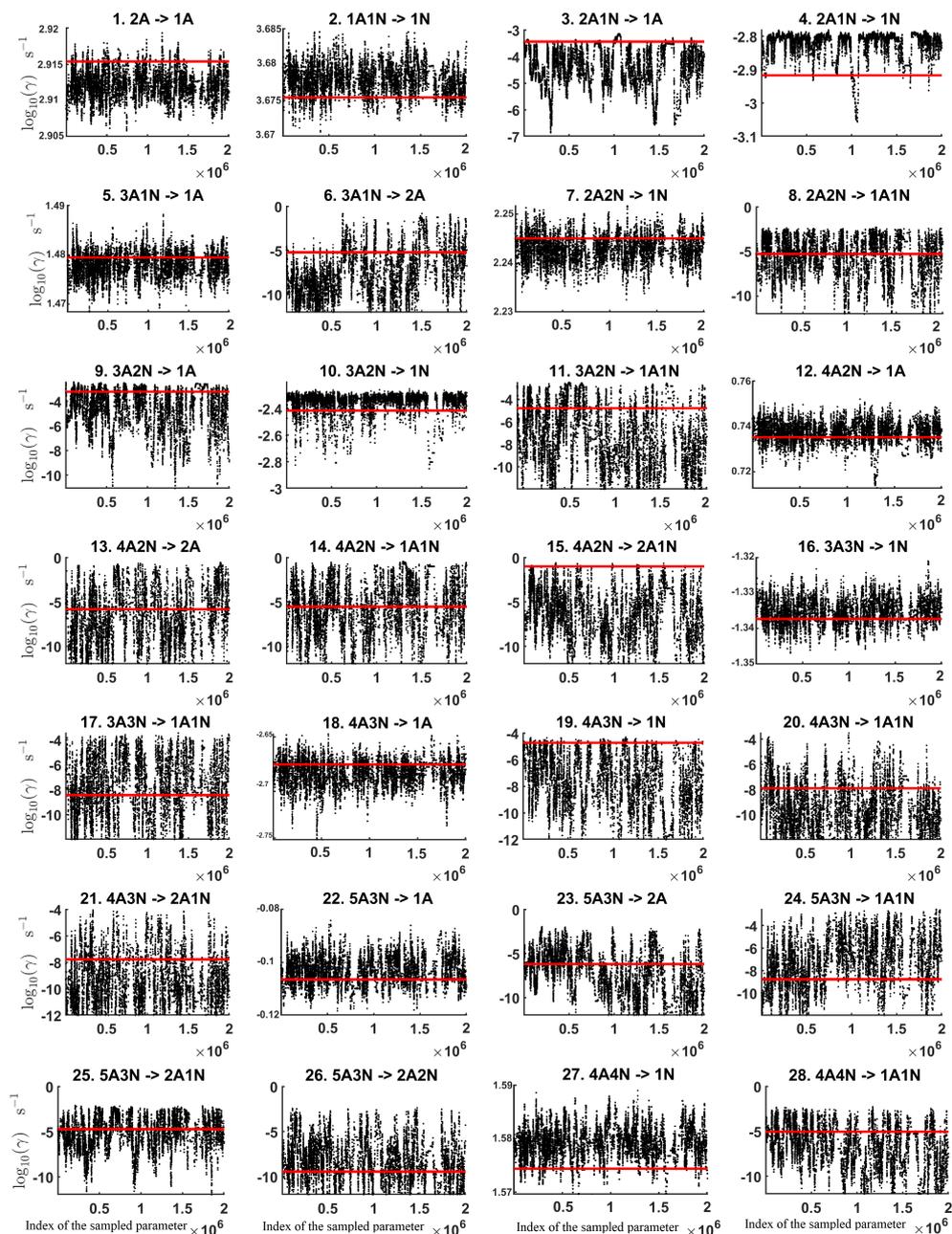


Figure B1. Parameter chains (for parameter indexes ranging from 1 to 28) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) determined from steady-state cluster concentration measurements at the temperature 278 K. Red lines denote the baseline values from [.,³⁷⁶] used to generate the synthetic data. [.,³⁷⁷] The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [.,³⁷⁸] y ammonia molecules.

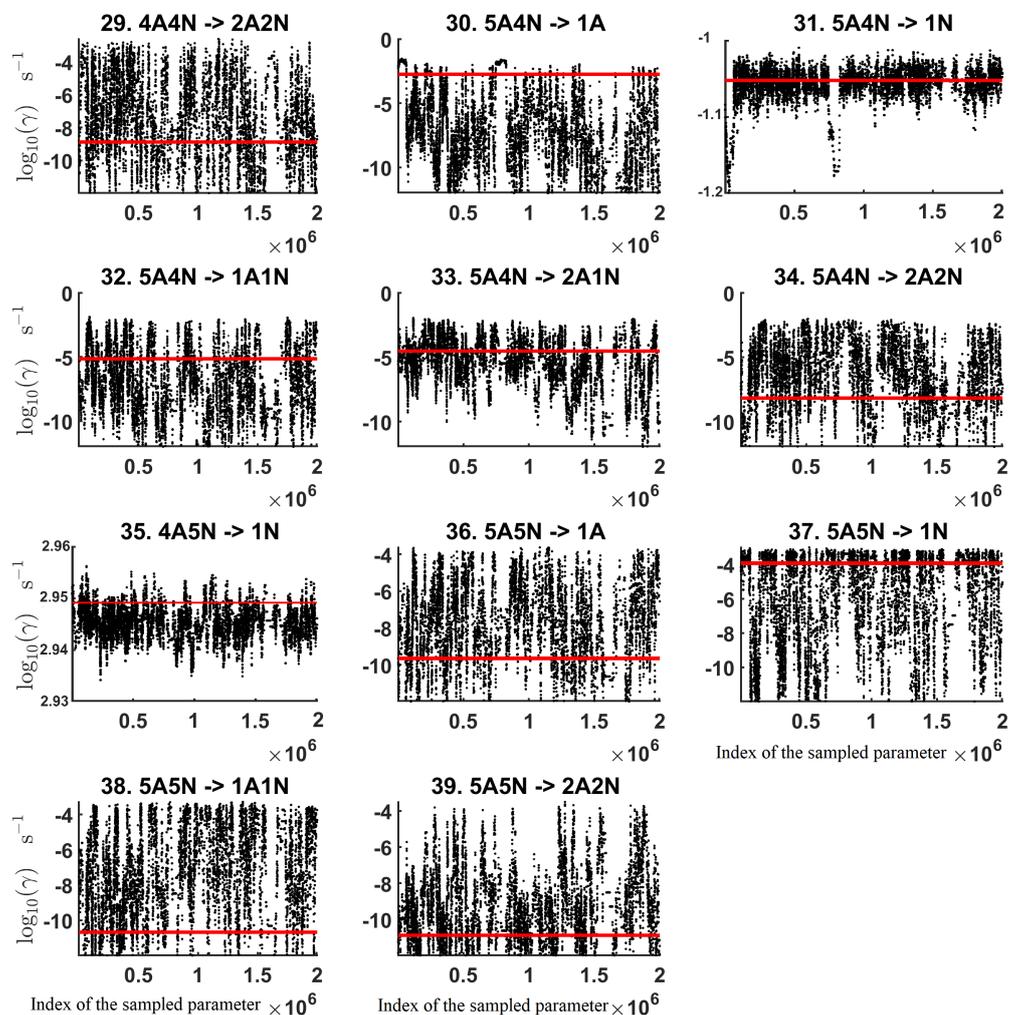


Figure B2. Parameter chains (for parameter indexes ranging from 29 to 39) of the base 10 logarithm of the evaporation rates (units given in s⁻¹) determined from steady-state cluster concentration measurements at the temperature 278 K. Red lines denote the baseline values from [..³⁷⁹] used to generate the synthetic data. [..³⁸⁰] The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..³⁸¹] y ammonia molecules.

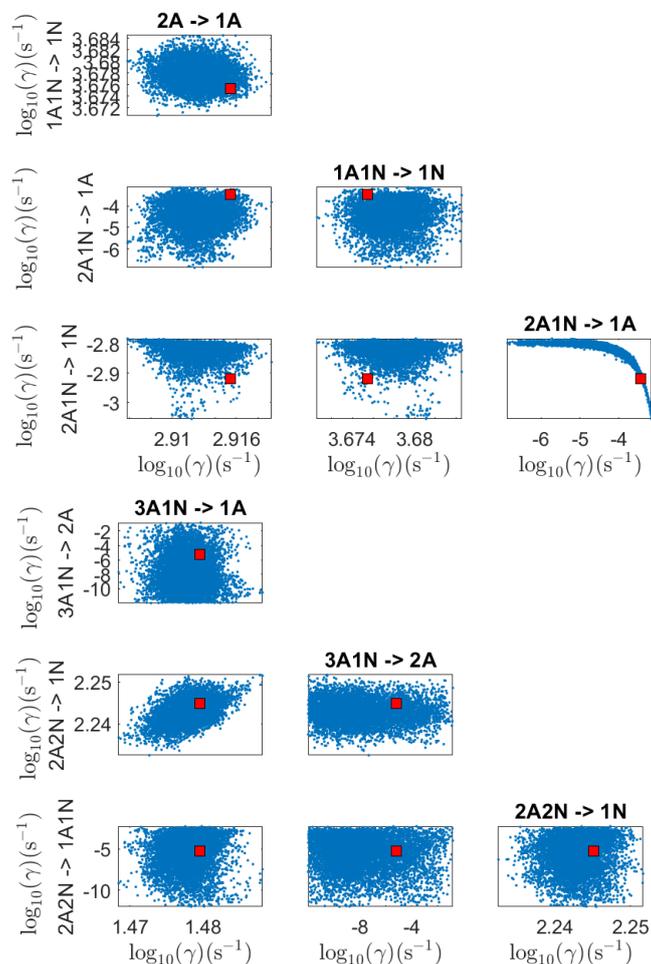


Figure B3. Pairwise marginal posterior distributions (for parameter indexes ranging from 1 to 8) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) determined from steady-state cluster concentration measurements at the temperature 278 K. Red rectangles denote the baseline values from [..³⁸²] used to generate the synthetic data [..³⁸³]. The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..³⁸⁴] y ammonia molecules.

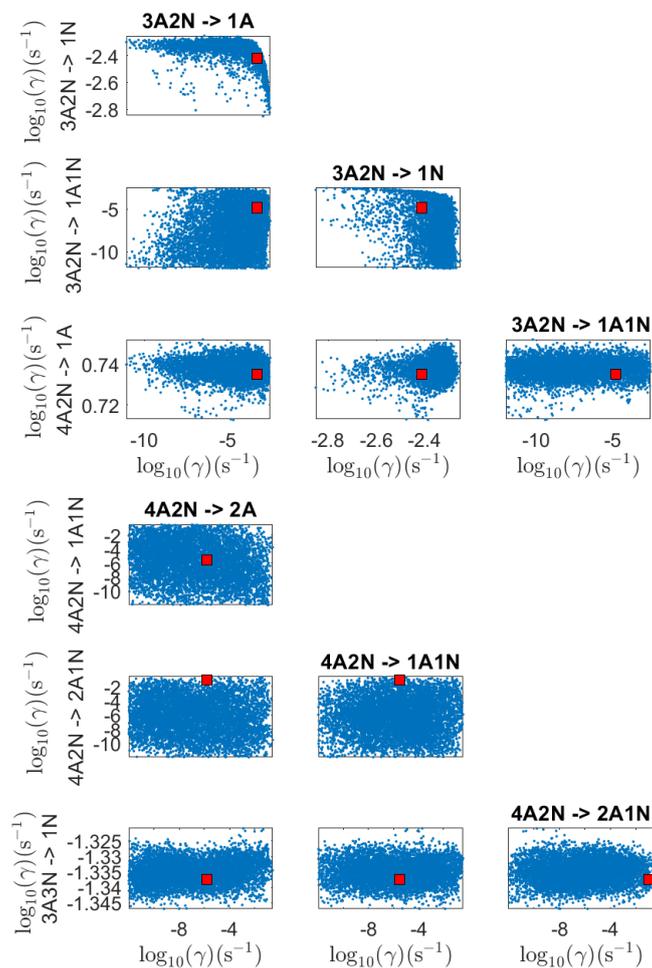


Figure B4. Pairwise marginal posterior distributions (for parameter indexes ranging from 9 to 16) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) determined from steady-state cluster concentration measurements at the temperature 278 K. Red rectangles denote the baseline values from [..³⁸⁵] used to generate the synthetic data. [..³⁸⁶] The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..³⁸⁷] ammonia molecules.

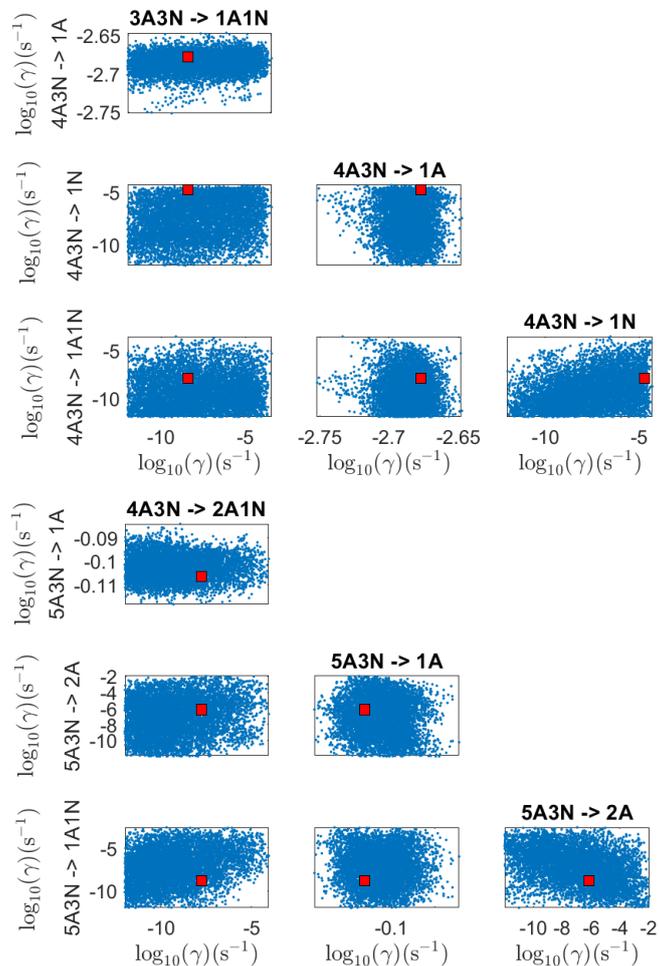


Figure B5. Pairwise marginal posterior distributions (for parameter indexes ranging from 17 to 24) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) determined from steady-state cluster concentration measurements at the temperature 278 K. Red rectangles denote the baseline values from [..³⁸⁸] used to generate the synthetic data. [..³⁸⁹] The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..³⁹⁰] y ammonia molecules.

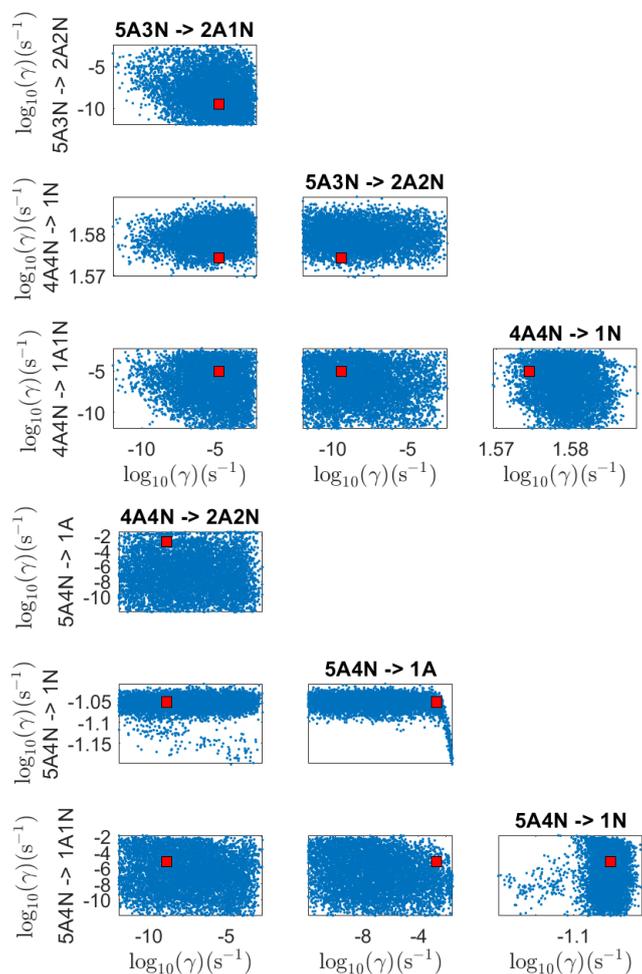


Figure B6. Pairwise marginal posterior distributions (for parameter indexes ranging from 25 to 32) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) determined from steady-state cluster concentration measurements at the temperature 278 K. Red rectangles denote the baseline values from [..³⁹¹] used to generate the synthetic data [..³⁹²]. The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..³⁹³] y ammonia molecules.

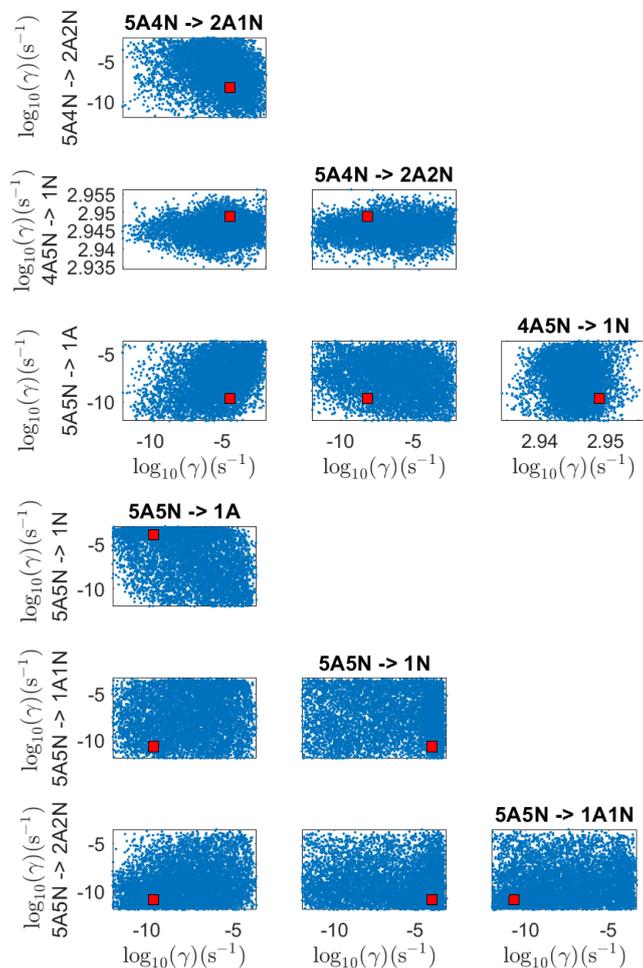


Figure B7. Pairwise marginal posterior distributions (for parameter indexes ranging from 33 to 39) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) determined from steady-state cluster concentration measurements at the temperature 278 K. Red rectangles denote the baseline values from [..³⁹⁴]? used to generate the synthetic data. [..³⁹⁵]The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..³⁹⁶] y ammonia molecules.

C Estimation of the evaporation rates from [..³⁹⁷]time dependent data

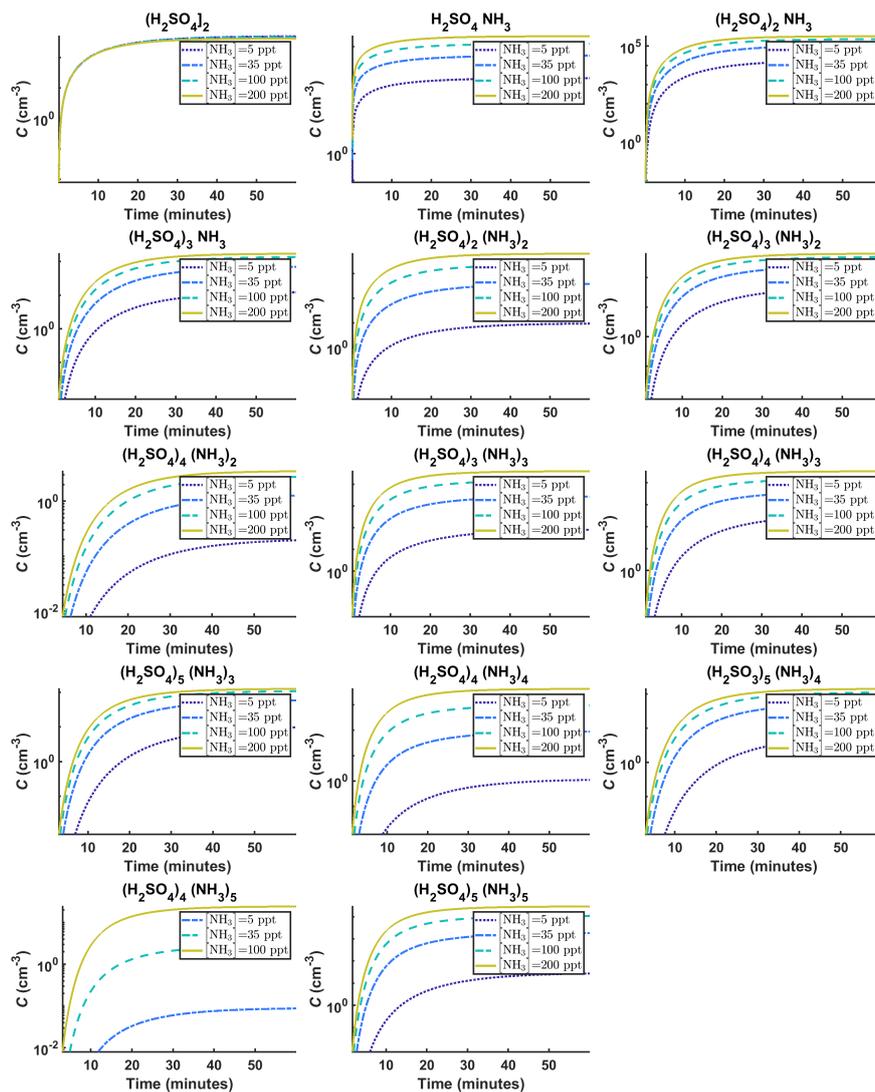


Figure C1. Time-dependent cluster concentrations. Simulated time evolution of concentrations for different cluster types at temperature $T=278$ K for varying $[\text{NH}_3]$ concentration: 5 ppt, 35 ppt, 100 ppt and 200 ppt (see the legend). All the model outputs are amended with multivariate non-correlated Gaussian noise with standard deviation comprising 0.001% of the original cluster concentration. Time resolution comprises 1.5 minutes. The source of [..³⁹⁸]sulfuric acid monomer is $[\text{H}_2\text{SO}_4] = 6.3 \times 10^4 \text{ s}^{-1}$ in all simulations. [..³⁹⁹]The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..⁴⁰⁰] y ammonia molecules.

³⁹⁷removed: transient

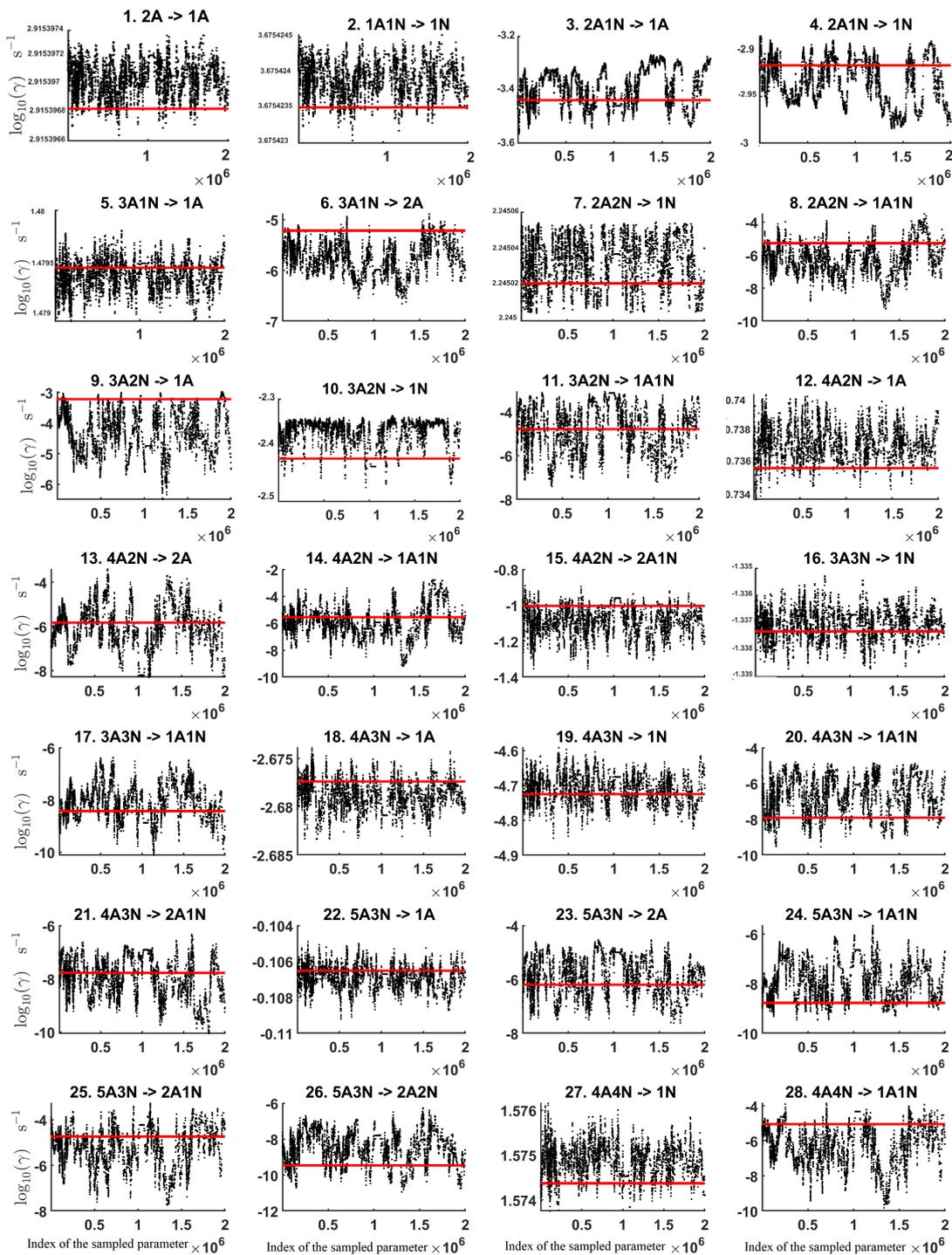


Figure C2. Parameter chains (for parameter indexes ranging from 1 to 28) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) determined from [..⁴⁰¹]time-dependent measurements of the cluster concentrations with time resolution comprising 1.5 minutes at the temperature 278 K. Red lines denote the baseline values from [..⁴⁰²] used to generate the synthetic data. Blue notation $xAyN$ corresponds to a cluster with x sulfuric acid and y ammonia molecules.

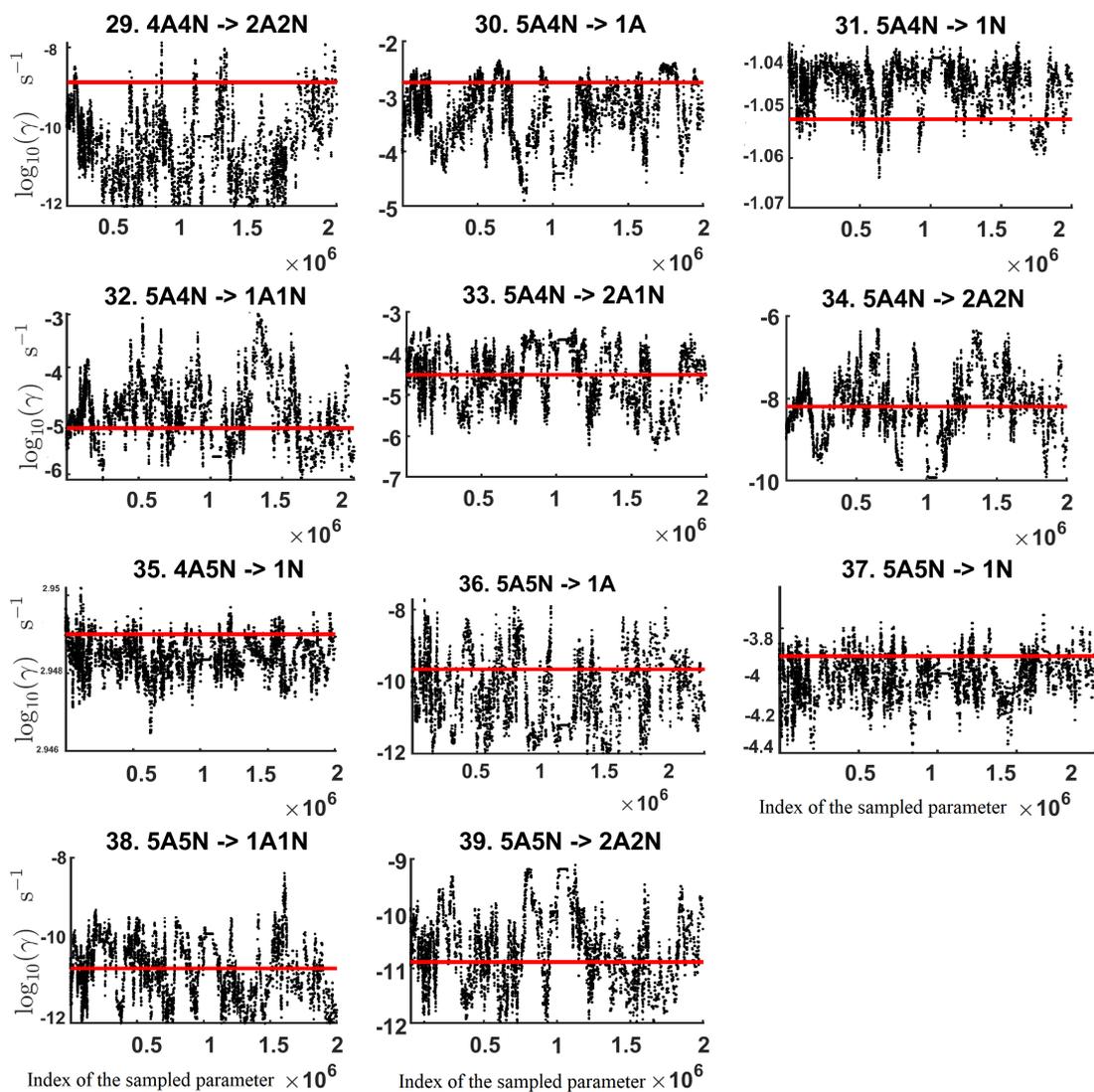


Figure C3. Parameter chains (for parameter indexes ranging from 29 to 39) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) determined from [..⁴⁰³]time-dependent measurements of the cluster concentrations with time resolution comprising 1.5 minutes at the temperature 278 K. Red lines denote the baseline values from [..⁴⁰⁴] used to generate the synthetic data [..⁴⁰⁵]. The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..⁴⁰⁶] y ammonia molecules.

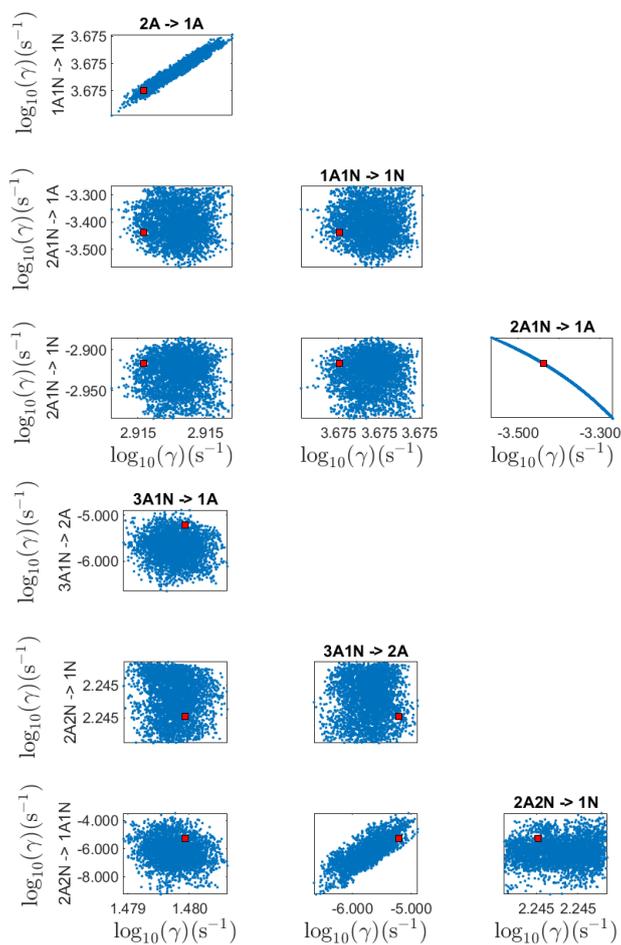


Figure C4. Pairwise marginal posterior distributions (for parameter indexes ranging from 1 to 8) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) determined from [..⁴⁰⁷]time-dependent measurements of the cluster concentrations with time resolution comprising 1.5 minutes at the temperature 278 K. Red rectangles denote the baseline values from [..⁴⁰⁸] used to generate the synthetic data. [..⁴⁰⁹]The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..⁴¹⁰] y ammonia molecules.

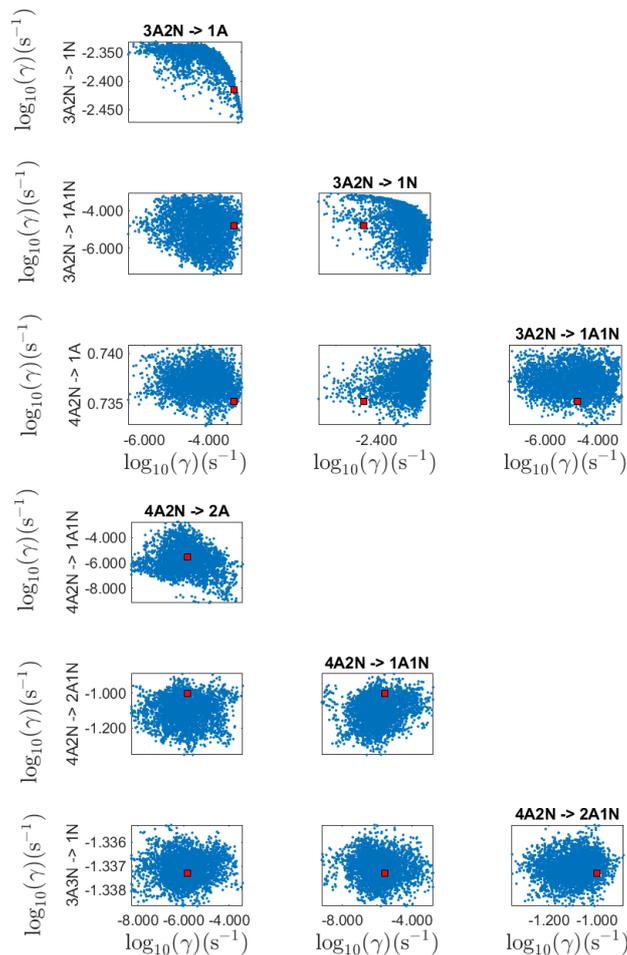


Figure C5. Pairwise marginal posterior distributions (for parameter indexes ranging from 9 to 16) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) determined from [..⁴¹¹]time-dependent measurements of the cluster concentrations with time resolution comprising 1.5 minutes at the temperature 278 K. Red rectangles denote the baseline values from [..⁴¹²] used to generate the synthetic data [..⁴¹³]The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..⁴¹⁴] y ammonia molecules.

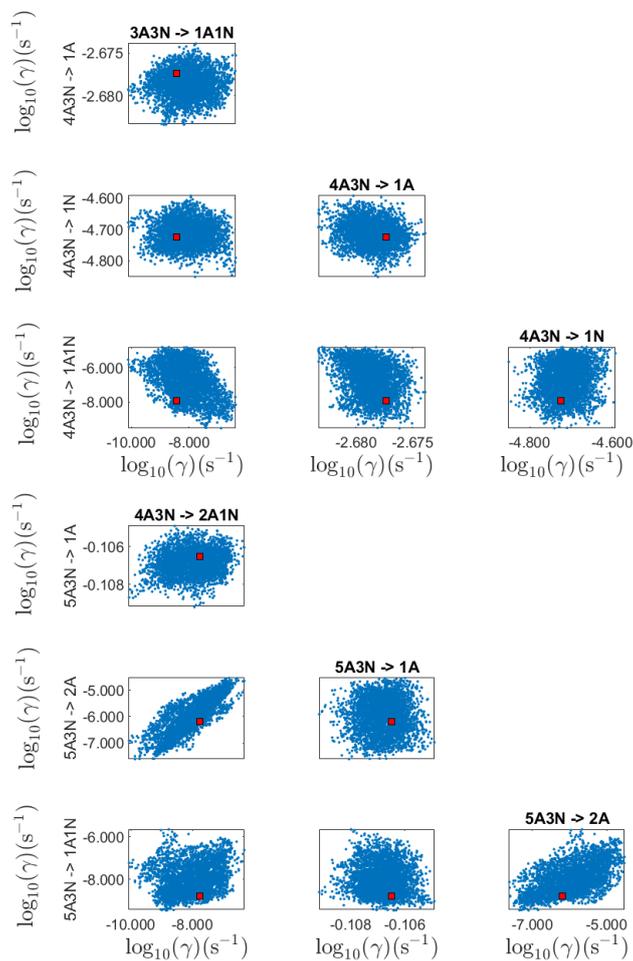


Figure C6. Pairwise marginal posterior distributions (for parameter indexes ranging from 17 to 24) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) from [..⁴¹⁵]time-dependent measurements of the cluster concentrations with time resolution comprising 1.5 minutes at the temperature 278 K. Red rectangles denote the baseline values from [..⁴¹⁶] used to generate the synthetic data [..⁴¹⁷]. The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..⁴¹⁸] y ammonia molecules.

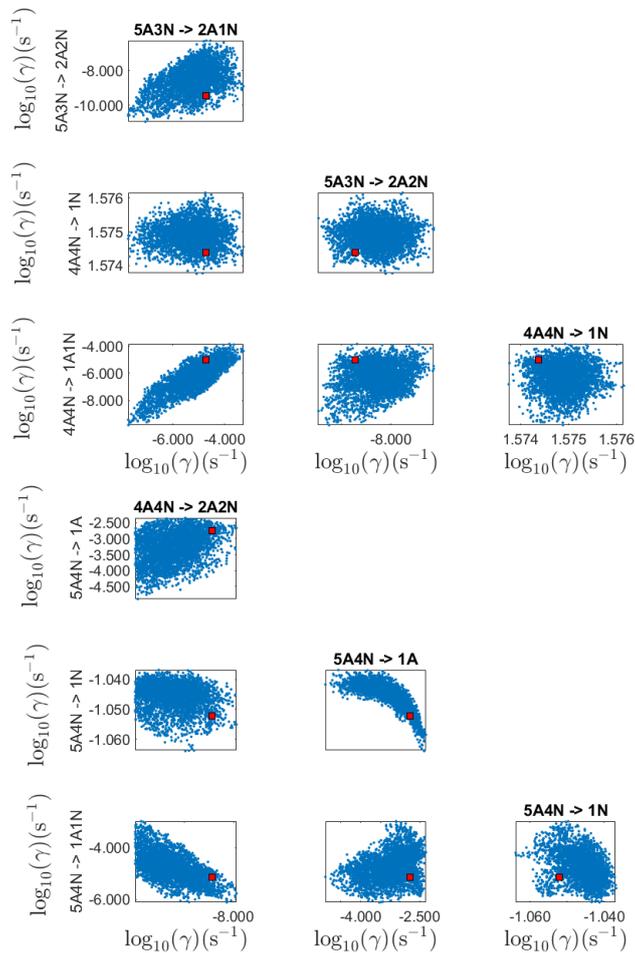


Figure C7. Pairwise marginal posterior distributions (for parameter indexes ranging from 25 to 32) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) from [..⁴¹⁹]time dependent measurements of the cluster concentrations with time resolution comprising 1.5 minutes at the temperature 278 K. Red rectangles denote the baseline values from [..⁴²⁰] used to generate the synthetic data. [..⁴²¹]The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..⁴²²] y ammonia molecules.

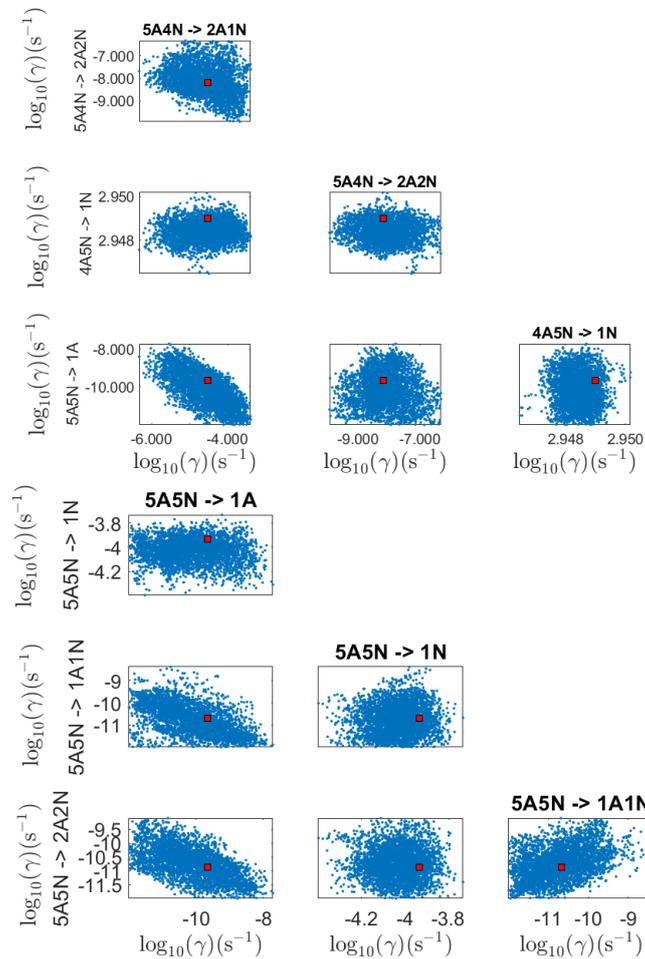


Figure C8. Pairwise marginal posterior distributions (for parameter indexes ranging from 33 to 39) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) from [..⁴²³] Time dependent measurements of the cluster concentrations with time resolution comprising 1.5 minutes at the temperature 278 K. Red rectangles denote the baseline values from [..⁴²⁴] used to generate the synthetic data. [..⁴²⁵] The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..⁴²⁶] y ammonia molecules.

Symbol	Steady-state data (s ⁻¹)	[.. ⁴²⁷]Time dependent data (s ⁻¹)	QC (s ⁻¹)
1: 2A → 1A	8.16 × 10² (8.05 × 10 ² , 8.31 × 10 ²)	8.23 × 10²	8.23 × 10 ²
2: 1A1N → 1N	4.75 × 10³ (4.69 × 10 ³ , 4.87 × 10 ³)	4.74 × 10³	4.74 × 10 ³
3: 2A1N → 1A	4.22 × 10⁻⁴ (5.92 × 10 ⁻¹¹ , 7.27 × 10 ⁻⁴)	3.30 × 10⁻⁴ (1.75 × 10 ⁻⁴ , 5.37 × 10 ⁻⁴)	3.64 × 10 ⁻⁴
4: 2A1N → 1N	1.56 × 10⁻³ (8.78 × 10 ⁻⁴ , 1.67 × 10 ⁻³)	1.33 × 10⁻³ (1.04 × 10 ⁻³ , 1.4 × 10 ⁻³)	1.21 × 10 ⁻³
5: 3A1N → 1A	2.99 × 10¹ (2.94 × 10 ¹ , 3.08 × 10 ¹)	3.02 × 10¹ (3.01 × 10 ¹ , 3.02 × 10 ¹)	3.02 × 10 ¹
6: 3A1N → 2A	— 1.50 × 10 ⁻¹	2.81 × 10⁻⁶ (2.86 × 10 ⁻⁹ , 2.76 × 10 ⁻³)	6.09 × 10 ⁻⁶
7: 2A2N → 1N	1.74 × 10² (1.71 × 10 ² , 1.79 × 10 ²)	1.76 × 10²	1.76 × 10 ²
8: 2A2N → 1A1N	5.52 × 10⁻⁴ < 5.16 × 10 ⁻³	2.11 × 10⁻⁶ (2.95 × 10 ⁻¹⁰ , 3.59 × 10 ⁻⁴)	5.33 × 10 ⁻⁶
9: 3A2N → 1A	3.30 × 10⁻⁴ < 2.91 × 10 ⁻³	7.51 × 10⁻⁴ (3.18 × 10 ⁻⁷ , 1.78 × 10 ⁻³)	6.07 × 10 ⁻⁴
10: 3A2N → 1N	4.47 × 10⁻³ (5.85 × 10 ⁻⁴ , 5.60 × 10 ⁻³)	4.16 × 10⁻³ (2.86 × 10 ⁻³ , 4.66 × 10 ⁻³)	3.84 × 10 ⁻³
11: 3A2N → 1A1N	9.79 × 10⁻⁵ < 3.88 × 10 ⁻³	1.00 × 10⁻⁵ (4.68 × 10 ⁻¹⁰ , 7.22 × 10 ⁻⁴)	1.64 × 10 ⁻⁵
12: 4A2N → 1A	5.50 × 10⁰ (4.50 × 10 ⁰ , 5.72 × 10 ⁰)	5.46 × 10⁰ (5.39 × 10 ⁰ , 5.51 × 10 ⁰)	5.43 × 10 ⁰
13: 4A2N → 2A	5.24 × 10⁻⁷ < 2.74 × 10 ⁻¹	1.03 × 10⁻⁶ (5.66 × 10 ⁻¹¹ , 1.88 × 10 ⁻²)	1.48 × 10 ⁻⁶
14: 4A2N → 1A1N	2.79 × 10⁻¹ < 6.92 × 10 ⁻¹	2.78 × 10⁻⁶ (6.50 × 10 ⁻¹⁰ , 1.66 × 10 ⁻³)	2.80 × 10 ⁻⁶
15: 4A2N → 2A1N	6.49 × 10⁻² < 1.02 × 10 ⁰	9.04 × 10⁻² (3.66 × 10 ⁻² , 1.33 × 10 ⁻¹)	9.94 × 10 ⁻²
16: 3A3N → 1N	4.62 × 10⁻² (4.50 × 10 ⁻² , 4.78 × 10 ⁻²)	4.61 × 10⁻² (4.58 × 10 ⁻² , 4.62 × 10 ⁻²)	4.60 × 10 ⁻²
17: 3A3N → 1A1N	1.37 × 10⁻⁹ < 3.58 × 10 ⁻⁴	6.32 × 10⁻⁹ (1.05 × 10 ⁻¹² , 4.91 × 10 ⁻⁶)	3.74 × 10 ⁻⁹
18: 4A3N → 1A	2.08 × 10⁻³ (1.79 × 10 ⁻³ , 2.27 × 10 ⁻³)	2.10 × 10⁻³ (2.07 × 10 ⁻³ , 2.12 × 10 ⁻³)	2.10 × 10 ⁻³
19: 4A3N → 1N	1.19 × 10⁻⁵ < 7.29 × 10 ⁻⁵	1.96 × 10⁻⁵ (1.11 × 10 ⁻⁵ , 2.50 × 10 ⁻⁵)	1.88 × 10 ⁻⁵
20: 4A3N → 1A1N	9.29 × 10⁻¹¹ < 2.65 × 10 ⁻⁴	— (1.81 × 10 ⁻¹² , 1.96 × 10 ⁻⁵)	1.23 × 10 ⁻⁸

Table C1. Part 1. Evaporation rates (units given in s⁻¹) determined from the steady-state and the [..⁴²⁸]time dependent data presented in Figure 5-6 and Figs. 16-17, respectively. For parameters that have a posterior distribution with the clear peak and practically zero probability density elsewhere, the mode of the distribution (bold face) is given together with the range of possible values in the parenthesis. In some of the cases only the limits can be determined. The last column presents the baseline values from [..⁴²⁹] used to generate the synthetic data. [..⁴³⁰]The notation *xAyN* corresponds to a cluster with x sulfuric acid and y ammonia molecules.

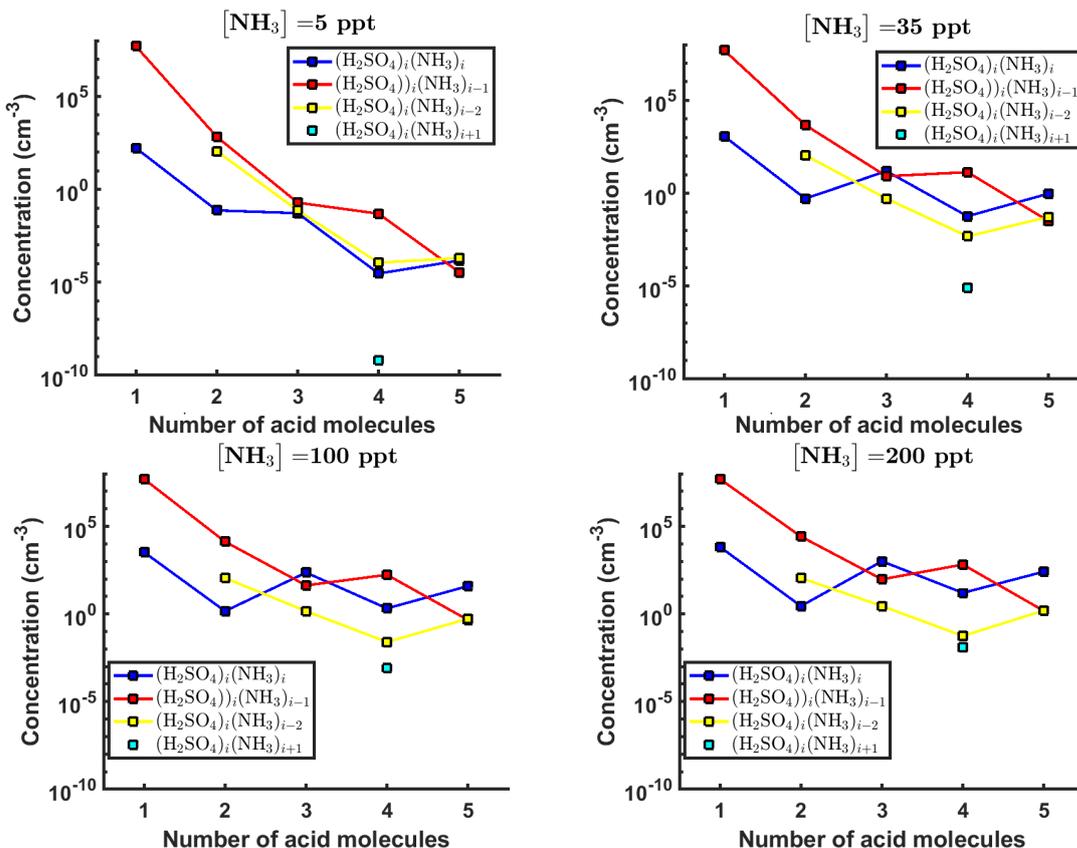


Figure D1. Steady-state cluster concentrations for the clusters containing [..⁴³⁵]sulfuric acid and a varying number of ammonia molecules as a function of the number of acid molecules for [NH₃] concentrations comprising (a) 5 ppt, (b) 35 ppt, (c) 100 ppt and (d) 200 ppt at temperature T=292 K amended with multivariate non-correlated Gaussian noise with standard deviation comprising 0.001% of the original cluster concentration. The source of [..⁴³⁶]sulfuric acid monomer comprises [H₂SO₄] = 6.3 × 10⁴ s⁻¹ in all the simulations. Here the symbols Δ*H* and Δ*S* stand for cluster formation enthalpies and entropies, respectively. [..⁴³⁷]The notation *xAyN* corresponds to a cluster with *x* sulfuric acid and [..⁴³⁸] *y* ammonia molecules.

Symbol	Steady-state data (s ⁻¹)	[.. ⁴³¹] Time dependent data (s ⁻¹)	QC (s ⁻¹)
21: 4A3N → 2A1N	— < 2.14 × 10 ⁻⁴	4.83 × 10⁻⁹ (3.36 × 10 ⁻¹² , 6.93 × 10 ⁻⁶)	1.66 × 10 ⁻⁸
22: 5A3N → 1A	7.88 × 10⁻¹ (7.56 × 10 ⁻¹ , 8.20 × 10 ⁻¹)	7.81 × 10⁻¹ (7.77 × 10 ⁻¹ , 7.86 × 10 ⁻¹)	7.83 × 10 ⁻¹
23: 5A3N → 2A	2.35 × 10⁻⁸ (< 1.21 × 10 ⁻²)	6.34 × 10⁻⁷ (1.26 × 10 ⁻¹¹ , 3.35 × 10 ⁻⁴)	6.37 × 10 ⁻⁷
24: 5A3N → 1A1N	9.12 × 10⁻¹² < 3.39 × 10 ⁻³	1.50 × 10⁻⁹ (1.02 × 10 ⁻¹² , 2.22 × 10 ⁻⁶)	1.70 × 10 ⁻⁹
25: 5A3N → 2A1N	7.22 × 10⁻⁴ < 6.95 × 10 ⁻³	1.24 × 10⁻⁵ (1.86 × 10 ⁻⁸ , 5.33 × 10 ⁻⁴)	1.85 × 10 ⁻⁵
26: 5A3N → 2A2N	1.52 × 10⁻⁸ < 4.49 × 10 ⁻³	— < 1.25 × 10 ⁻⁴	3.52 × 10 ⁻¹⁰
27: 4A4N → 1N	3.79 × 10¹ (3.70 × 10 ¹ , 3.88 × 10 ¹)	3.76 × 10¹ (3.75 × 10 ¹ , 3.77 × 10 ¹)	3.75 × 10 ¹
28: 4A4N → 1A1N	— < 5.38 × 10 ⁻³	9.05 × 10⁻⁶ (1.52 × 10 ⁻¹⁰ , 2.57 × 10 ⁻⁴)	9.06 × 10 ⁻⁶
29: 4A4N → 2A2N	2.07 × 10⁻¹² < 2.43 × 10 ⁻³	8.55 × 10⁻¹¹ < 1.90 × 10 ⁻⁴	1.33 × 10 ⁻⁹
30: 5A4N → 1A	3.87 × 10⁻⁶ < 2.52 × 10 ⁻²	2.51 × 10⁻³ (1.20 × 10 ⁻⁶ , 5.86 × 10 ⁻³)	1.77 × 10 ⁻³
31: 5A4N → 1N	8.92 × 10⁻² (6.68 × 10 ⁻² , 9.74 × 10 ⁻²)	9.03 × 10⁻² (8.52 × 10 ⁻² , 9.19 × 10 ⁻²)	8.87 × 10 ⁻²
32: 5A4N → 1A1N	— < 1.55 × 10 ⁻²	3.60 × 10⁻⁶ (6.48 × 10 ⁻¹² , 1.04 × 10 ⁻³)	7.33 × 10 ⁻⁶
33: 5A4N → 2A1N	2.28 × 10⁻⁴ < 1.06 × 10 ⁻²	1.32 × 10⁻⁴ (6.46 × 10 ⁻¹⁰ , 1.53 × 10 ⁻³)	2.97 × 10 ⁻⁵
34: 5A4N → 2A2N	— < 1.08 × 10 ⁻²	7.30 × 10⁻⁹ (1.51 × 10 ⁻¹¹ , 3.17 × 10 ⁻⁴)	6.42 × 10 ⁻⁹
35: 4A5N → 1N	8.75 × 10² (8.59 × 10 ² , 9.03 × 10 ²)	8.88 × 10² (8.85 × 10 ² , 8.92 × 10 ²)	8.89 × 10 ²
36: 5A5N → 1A	— < 2.32 × 10 ⁻⁴	— < 1.14 × 10 ⁻⁶	2.23 × 10 ⁻¹⁰
37: 5A5N → 1N	4.96 × 10⁻⁴ < 9.89 × 10 ⁻⁴	1.00 × 10⁻⁴ (3.48 × 10 ⁻⁵ , 1.85 × 10 ⁻⁴)	1.17 × 10 ⁻⁴
38: 5A5N → 1A1N	5.93 × 10⁻⁹ < 5.06 × 10 ⁻⁴	1.48 × 10⁻¹¹ < 1.06 × 10 ⁻⁵	2.11 × 10 ⁻¹¹
39: 5A5N → 2A2N	— < 3.09 × 10 ⁻⁴	2.06 × 10⁻¹¹ < 4.11 × 10 ⁻⁷	1.31 × 10 ⁻¹¹

Table C2. Part 2. Evaporation rates (units given in s⁻¹) determined from the steady-state and the [..⁴³²]time dependent data presented in Figure 5-6 and Figs. 16-17, respectively. For parameters that have a posterior distribution with the clear peak and practically zero probability density elsewhere, the mode of the distribution (bold face) is given together with the range of possible values in the parenthesis. In some of the cases only the limits can be determined. The last column presents the baseline values from [..⁴³³] used to generate the synthetic data. [..⁴³⁴]The notation *xAyN* corresponds to a cluster with *x* sulfuric acid and *y* ammonia molecules.

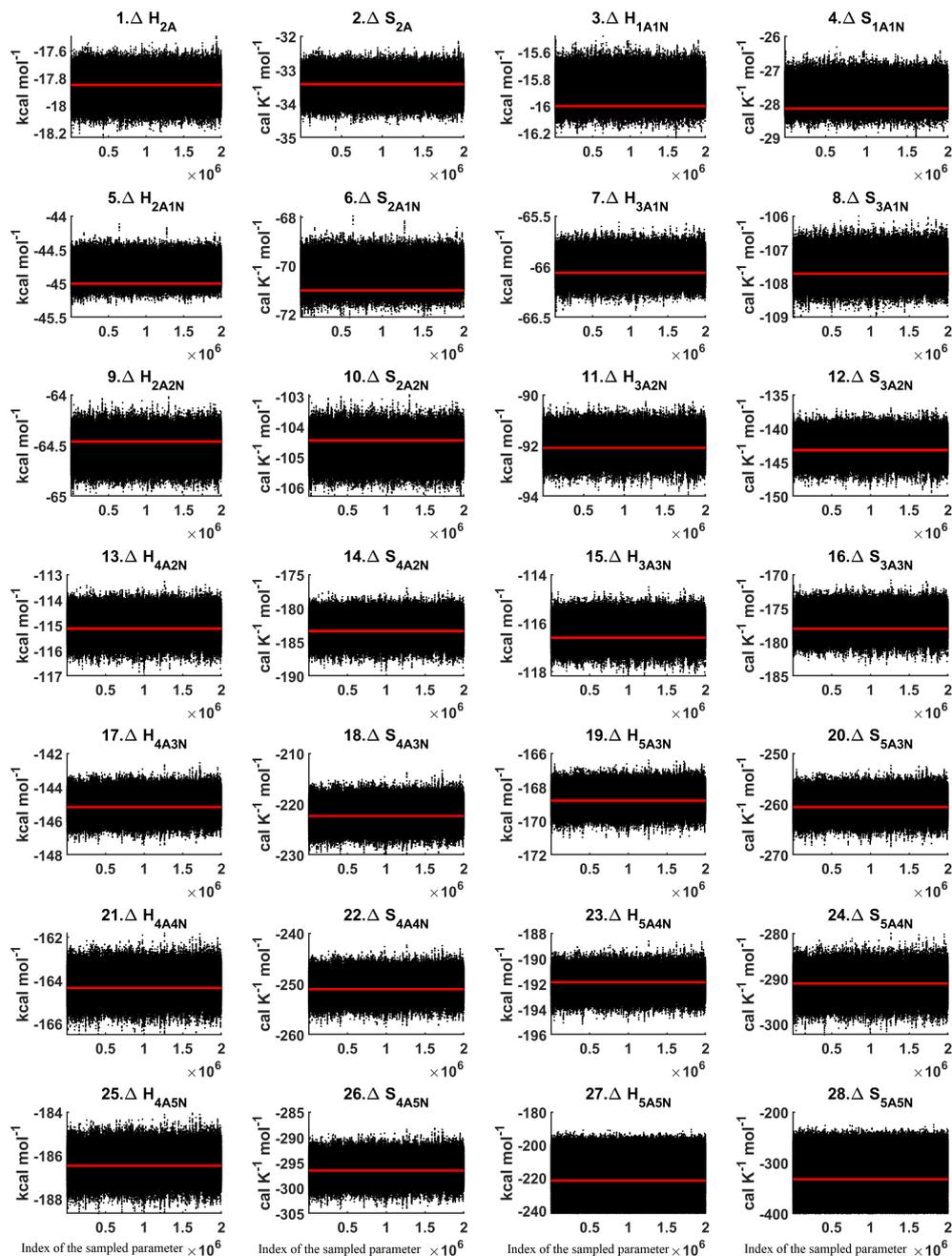


Figure D2. Parameter chains of the cluster formation enthalpies (units given in kcal/mol) and entropies (units given in cal K⁻¹ mol⁻¹) determined from steady-state cluster concentration measurements at two temperatures T=278 K and T = 292 K. Red lines denote the baseline values from [..⁴³⁹]? used to generate the synthetic data. Here the symbols ΔH and ΔS stand for cluster formation enthalpies and entropies, respectively. [..⁴⁴⁰]The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..⁴⁴¹] y ammonia molecules.

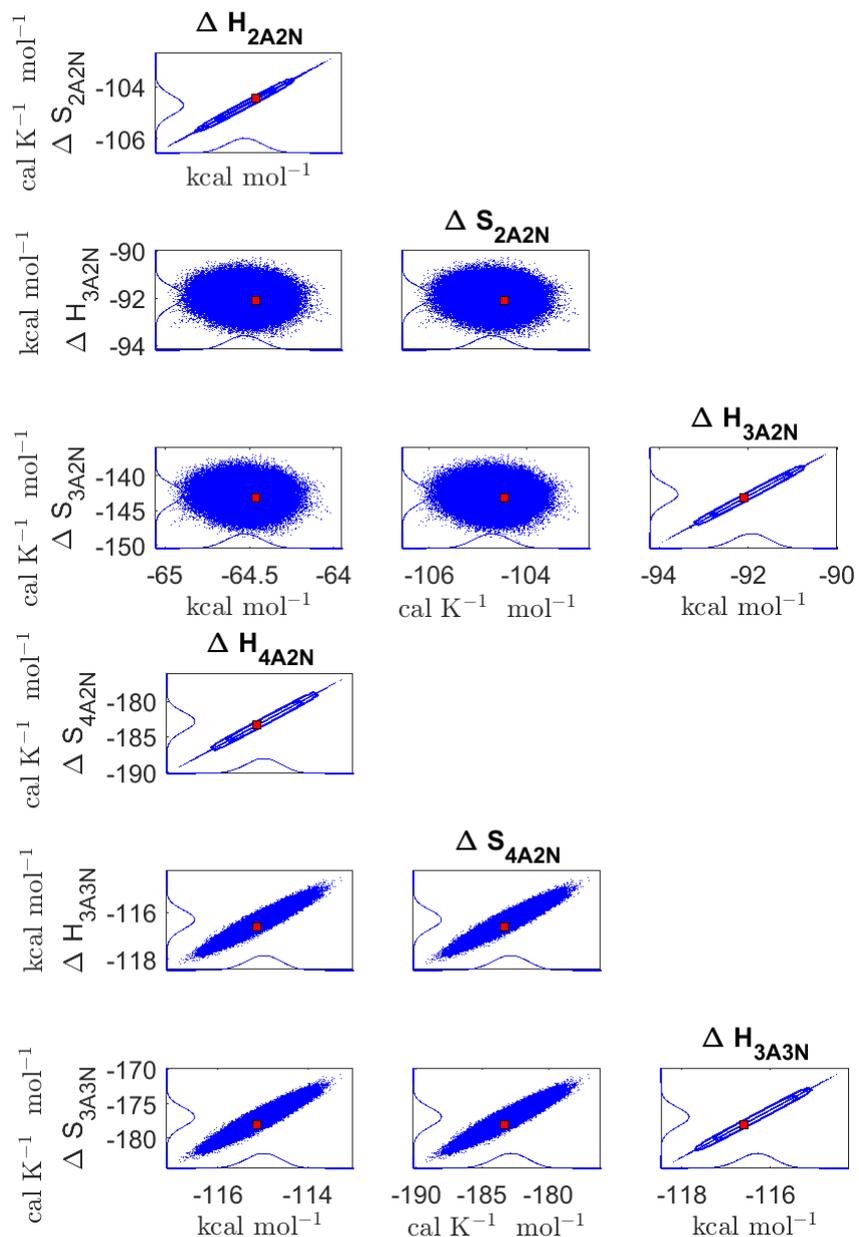


Figure D3. Pairwise marginal posterior distributions (for parameter indexes ranging from 9 to 16) of the cluster formation enthalpies and entropies determined from steady-state cluster concentration measurements at two temperatures $T=278$ K and $T = 292$ K. Red rectangles denote the baseline values from [..⁴⁴²] used to generate the synthetic data. Here the symbols ΔH and ΔS stand for cluster formation enthalpies and entropies, respectively. [..⁴⁴³] The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..⁴⁴⁴] y ammonia molecules.

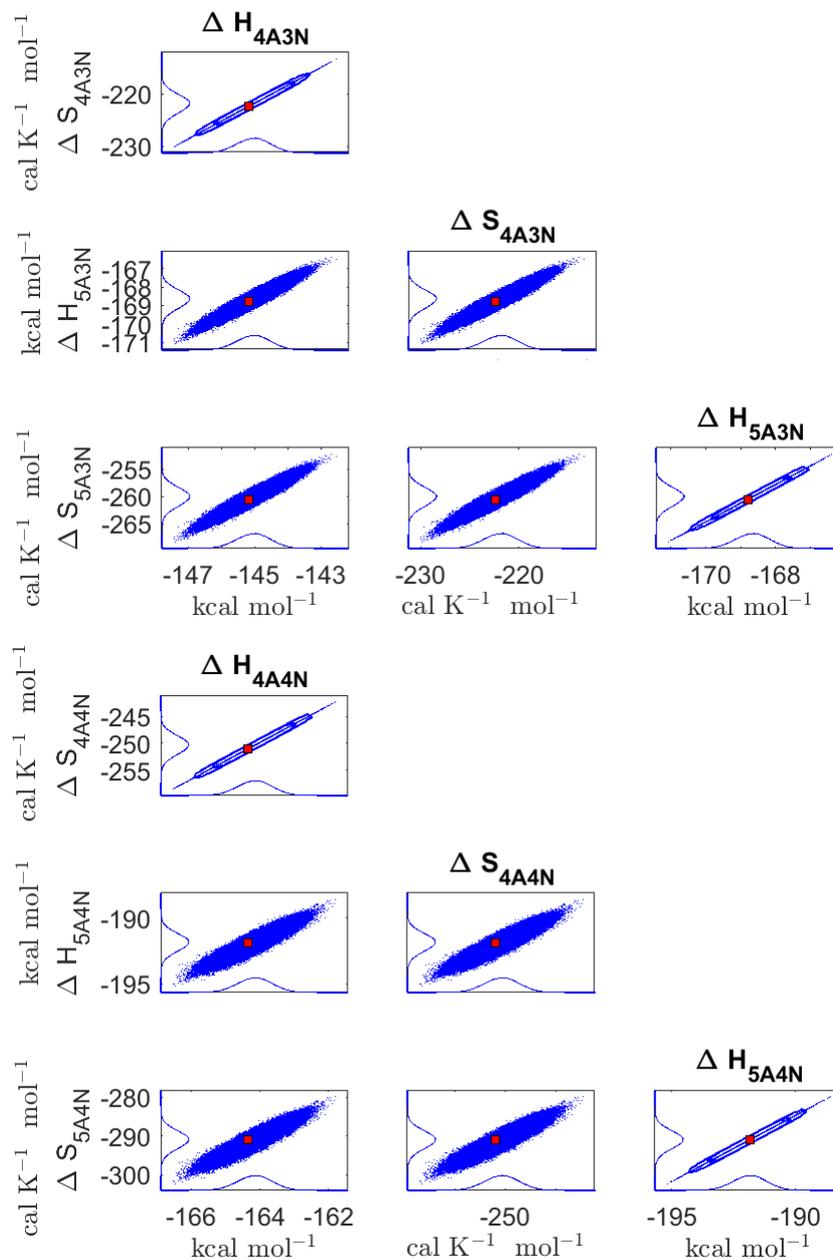


Figure D4. Pairwise marginal posterior distributions (for parameter indexes ranging from 17 to 24) of the cluster formation enthalpies and entropies determined from steady-state cluster concentration measurements at two temperatures $T=278$ K and $T = 292$ K. Red rectangles denote the baseline values from [..⁴⁴⁵] used to generate the synthetic data. Here the symbols ΔH and ΔS stand for cluster formation enthalpies and entropies, respectively. [..⁴⁴⁶]The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..⁴⁴⁷] y ammonia molecules.

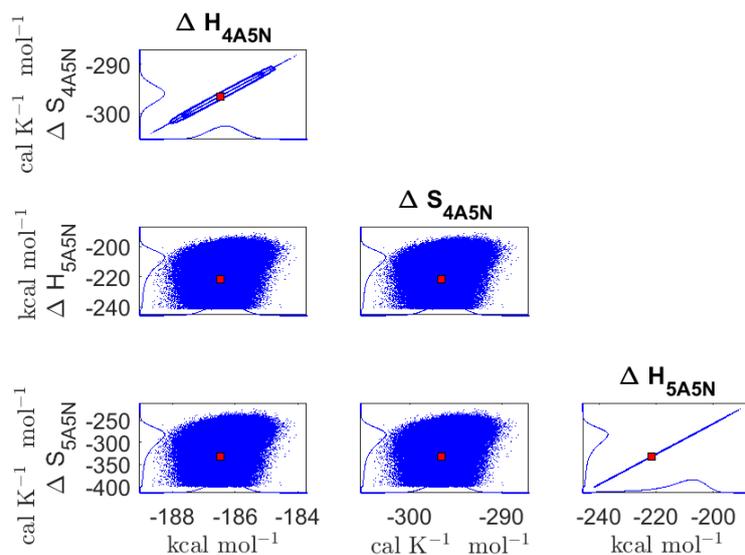


Figure D5. Pairwise marginal posterior distributions (for parameter indexes ranging from 25 to 28) of the cluster formation enthalpies and entropies determined from steady-state cluster concentration measurements at two temperatures $T=278$ K and $T = 292$ K. Red rectangles denote the baseline values from [..⁴⁴⁸]? used to generate the synthetic data. Here the symbols ΔH and ΔS stand for cluster formation enthalpies and entropies, respectively. [..⁴⁴⁹]The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..⁴⁵⁰] y ammonia molecules.

Symbol	Mode value	95% confidence interval	QC	Units
1: ΔH_{2A}	-17.8891	(-18.1913,-17.4941)	-17.85	kcal mol ⁻¹
2: ΔS_{2A}	-33.5475	(-34.6104,-32.1575)	-33.42	cal K ⁻¹ mol ⁻¹
3: ΔH_{1A1N}	-15.8751	(-16.2344,-15.5158)	-16	kcal mol ⁻¹
4: ΔS_{1A1N}	-27.6984	(-28.9594,-26.4374)	-28.14	cal K ⁻¹ mol ⁻¹
5: ΔH_{2A1N}	-44.8076	(-45.2922,-44.174)	-45	kcal mol ⁻¹
6: ΔS_{2A1N}	-70.3501	(-72.029,-68.1545)	-71.02	cal K ⁻¹ mol ⁻¹
7: ΔH_{3A1N}	-66.0006	(-66.428,-65.5732)	-66.06	kcal mol ⁻¹
8: ΔS_{3A1N}	-107.5233	(-109.0059,-106.0407)	-107.72	cal K ⁻¹ mol ⁻¹
9: ΔH_{2A2N}	-64.5005	(-64.9799,-64.021)	-64.46	kcal mol ⁻¹
10: ΔS_{2A2N}	-104.6181	(-106.2857,-102.9505)	-104.45	cal K ⁻¹ mol ⁻¹
11: ΔH_{3A2N}	-91.8512	(-93.9174,-90.2712)	-92.09	kcal mol ⁻¹
12: ΔS_{3A2N}	-142.3625	(-149.4438,-136.9474)	-143.18	cal K ⁻¹ mol ⁻¹
13: ΔH_{4A2N}	-115.0105	(-116.7515,-113.2696)	-115.13	kcal mol ⁻¹
14: ΔS_{4A2N}	-182.938	(-188.9067,-176.9693)	-183.34	cal K ⁻¹ mol ⁻¹
15: ΔH_{3A3N}	-116.3273	(-118.1437,-114.5108)	-116.6	kcal mol ⁻¹
16: ΔS_{3A3N}	-177.0462	(-183.2768,-170.8156)	-177.99	cal K ⁻¹ mol ⁻¹
17: ΔH_{4A3N}	-144.9757	(-147.3975,-142.554)	-145.17	kcal mol ⁻¹
18: ΔS_{4A3N}	-221.6575	(-229.9554,-213.3595)	-222.33	cal K ⁻¹ mol ⁻¹
19: ΔH_{5A3N}	-168.7305	(-171.0579,-166.4031)	-168.79	kcal mol ⁻¹
20: ΔS_{5A3N}	-260.3509	(-268.3225,-252.3794)	-260.55	cal K ⁻¹ mol ⁻¹
21: ΔH_{4A4N}	-164.1272	(-166.4394,-161.815)	-164.35	kcal mol ⁻¹
22: ΔS_{4A4N}	-250.2634	(-258.1819,-242.3449)	-251.03	cal K ⁻¹ mol ⁻¹
23: ΔH_{5A4N}	-191.7779	(-194.9426,-188.6133)	-191.86	kcal mol ⁻¹
24: ΔS_{5A4N}	-290.7782	(-301.6196,-279.9369)	-291.05	cal K ⁻¹ mol ⁻¹
25: ΔH_{4A5N}	-186.3473	(-188.639,-184.0557)	-186.47	kcal mol ⁻¹
26: ΔS_{4A5N}	-296.0839	(-303.9359,-288.2319)	-296.51	cal K ⁻¹ mol ⁻¹
27: ΔH_{5A5N}	-205.943	(-241.6193,-190.6532)	-221.65	kcal mol ⁻¹
28: ΔS_{5A5N}	-277.4	(-, -224.8575)	-332.49	cal K ⁻¹ mol ⁻¹

Table D1. Thermodynamic parameters identified from steady-state data measured at two temperatures (278 and 292 K). The last column presents the quantum-chemistry based values from [..⁴⁵¹](?) used to generate the synthetic data. Here the symbols ΔH and ΔS stand for cluster formation enthalpies and entropies, respectively. [..⁴⁵²]The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..⁴⁵³]y ammonia molecules.



Figure D6. One-dimensional marginal distributions (for parameter indexes ranging from 1 to 28) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) at temperature 278 K obtained from a posterior distribution of thermodynamic parameters (cluster formation enthalpies and entropies) determined from steady-state cluster concentration measured at temperatures 278 K and 292 K. Red lines denote the baseline values from [..⁴⁵⁴] used to generate the synthetic data. [..⁴⁵⁵] The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..⁴⁵⁶] y ammonia molecules.

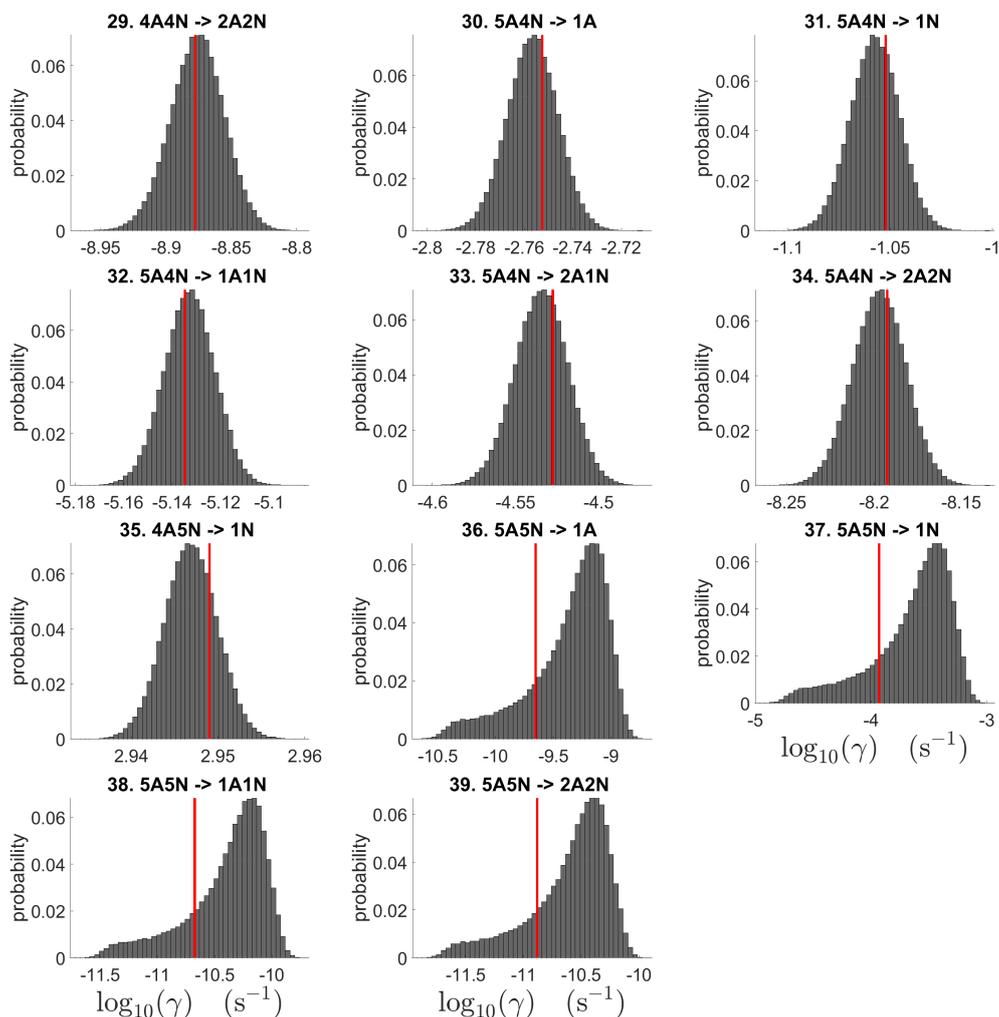


Figure D7. One-dimensional marginal distributions (for parameter indexes ranging from 29 to 39) of the base 10 logarithm of the evaporation rates (units given in s^{-1}) at temperature 278 K obtained from a posterior distribution of thermodynamic parameters (cluster formation enthalpies and entropies) determined from steady-state cluster concentration measured at temperatures 278 K and 292 K. Red lines denote the baseline values from [..⁴⁵⁷]? used to generate the synthetic data [..⁴⁵⁸]. The notation $xAyN$ corresponds to a cluster with x sulfuric acid and [..⁴⁵⁹] y ammonia molecules.

Symbol	Steady-state data for 278 K and 292 K (s ⁻¹)	QC (s ⁻¹)
1: 2A → 1A	8.17 × 10² (8.03 × 10 ² , 8.36 × 10 ²)	8.23 × 10 ²
2: 1A1N → 1N	4.76 × 10³ (4.66 × 10 ³ , 4.87 × 10 ³)	4.74 × 10 ³
3: 2A1N → 1A	3.64 × 10⁻⁴ (3.48 × 10 ⁻⁴ , 3.84 × 10 ⁻⁴)	3.64 × 10 ⁻⁴
4: 2A1N → 1N	1.23 × 10⁻³ (1.16 × 10 ⁻³ , 1.29 × 10 ⁻³)	1.21 × 10 ⁻³
5: 3A1N → 1A	3.01 × 10¹ (2.93 × 10 ¹ , 3.09 × 10 ¹)	3.02 × 10 ¹
6: 3A1N → 2A	6.12 × 10⁻⁶ (5.77 × 10 ⁻⁶ , 6.47 × 10 ⁻⁶)	6.09 × 10 ⁻⁶
7: 2A2N → 1N	1.77 × 10² (1.71 × 10 ² , 1.82 × 10 ²)	1.76 × 10 ²
8: 2A2N → 1A1N	5.33 × 10⁻⁶ (5.02 × 10 ⁻⁶ , 5.64 × 10 ⁻⁶)	5.33 × 10 ⁻⁶
9: 3A2N → 1A	6.09 × 10⁻⁴ (5.14 × 10 ⁻⁴ , 7.05 × 10 ⁻⁴)	6.07 × 10 ⁻⁴
10: 3A2N → 1N	3.89 × 10⁻³ (3.27 × 10 ⁻³ , 4.50 × 10 ⁻³)	3.84 × 10 ⁻³
11: 3A2N → 1A1N	1.65 × 10⁻⁵ (1.40 × 10 ⁻⁵ , 1.90 × 10 ⁻⁵)	1.64 × 10 ⁻⁵
12: 4A2N → 1A	5.45 × 10⁰ (5.25 × 10 ⁰ , 5.65 × 10 ⁰)	5.43 × 10 ⁰
13: 4A2N → 2A	1.49 × 10⁻⁶ (1.27 × 10 ⁻⁶ , 1.72 × 10 ⁻⁶)	1.48 × 10 ⁻⁶
14: 4A2N → 1A1N	2.82 × 10⁻⁶ (2.37 × 10 ⁻⁶ , 3.26 × 10 ⁻⁶)	2.80 × 10 ⁻⁶
15: 4A2N → 2A1N	1.01 × 10⁻¹ (8.35 × 10 ⁻² , 1.18 × 10 ⁻¹)	9.94 × 10 ⁻²
16: 3A3N → 1N	4.64 × 10⁻² (4.47 × 10 ⁻² , 4.81 × 10 ⁻²)	4.60 × 10 ⁻²
17: 3A3N → 1A1N	3.77 × 10⁻⁹ (3.19 × 10 ⁻⁹ , 4.36 × 10 ⁻⁹)	3.74 × 10 ⁻⁹
18: 4A3N → 1A	2.08 × 10⁻³ (1.86 × 10 ⁻³ , 2.29 × 10 ⁻³)	2.10 × 10 ⁻³
19: 4A3N → 1N	1.87 × 10⁻⁵ (1.69 × 10 ⁻⁵ , 2.05 × 10 ⁻⁵)	1.88 × 10 ⁻⁵
20: 4A3N → 1A1N	1.21 × 10⁻⁸ (1.09 × 10 ⁻⁸ , 1.33 × 10 ⁻⁸)	1.23 × 10 ⁻⁸

Table D2. Part 1. Evaporation rates (units given in s⁻¹) computed from a posterior distribution of the thermodynamic parameters (cluster formation enthalpies and entropies) which had previously been determined from the steady-state concentration measurements at temperatures 278 and 292 K. Here the mode of distribution (bold face) is given together with the range of possible values in the parenthesis. The last column presents the quantum-chemistry-based evaporation rates used for creating the synthetic data (borrowed from ?). The notation $xAyN$ corresponds to a cluster with x sulfuric acid and y ammonia molecules.^[..⁴⁶⁰]

Symbol	Steady-state data for 278 K and 292 K (s^{-1})	QC (s^{-1})
21: 4A3N \rightarrow 2A1N	1.65×10^{-8} ($1.30 \times 10^{-8}, 1.99 \times 10^{-8}$)	1.66×10^{-8}
22: 5A3N \rightarrow 1A	7.98×10^{-1} ($7.63 \times 10^{-1}, 8.43 \times 10^{-1}$)	7.83×10^{-1}
23: 5A3N \rightarrow 2A	6.40×10^{-7} ($5.76 \times 10^{-7}, 7.24 \times 10^{-7}$)	6.37×10^{-7}
24: 5A3N \rightarrow 1A1N	1.71×10^{-9} ($1.54 \times 10^{-9}, 1.88 \times 10^{-9}$)	1.70×10^{-9}
25: 5A3N \rightarrow 2A1N	1.87×10^{-5} ($1.66 \times 10^{-5}, 2.07 \times 10^{-5}$)	1.85×10^{-5}
26: 5A3N \rightarrow 2A2N	3.56×10^{-10} ($2.83 \times 10^{-10}, 4.30 \times 10^{-10}$)	3.52×10^{-10}
27: 4A4N \rightarrow 1N	3.82×10^1 ($3.69 \times 10^1, 3.95 \times 10^1$)	3.75×10^1
28: 4A4N \rightarrow 1A1N	8.97×10^{-6} ($8.13 \times 10^{-6}, 1.01 \times 10^{-5}$)	9.06×10^{-6}
29: 4A4N \rightarrow 2A2N	1.34×10^{-9} ($1.07 \times 10^{-9}, 1.62 \times 10^{-9}$)	1.33×10^{-9}
30: 5A4N \rightarrow 1A	1.76×10^{-3} ($1.56 \times 10^{-3}, 1.96 \times 10^{-3}$)	1.77×10^{-3}
31: 5A4N \rightarrow 1N	8.70×10^{-2} ($7.68 \times 10^{-2}, 1.00 \times 10^{-1}$)	8.87×10^{-2}
32: 5A4N \rightarrow 1A1N	7.42×10^{-6} ($6.59 \times 10^{-6}, 8.24 \times 10^{-6}$)	7.33×10^{-6}
33: 5A4N \rightarrow 2A1N	2.92×10^{-5} ($2.45 \times 10^{-5}, 3.40 \times 10^{-5}$)	2.97×10^{-5}
34: 5A4N \rightarrow 2A2N	6.40×10^{-9} ($5.40 \times 10^{-9}, 7.40 \times 10^{-9}$)	6.42×10^{-9}
35: 4A5N \rightarrow 1N	8.85×10^2 ($8.58 \times 10^2, 9.12 \times 10^2$)	8.89×10^2
36: 5A5N \rightarrow 1A	5.38×10^{-10} ($2.01 \times 10^{-11}, 2.24 \times 10^{-9}$)	2.23×10^{-10}
37: 5A5N \rightarrow 1N	2.77×10^{-4} ($1.09 \times 10^{-5}, 1.15 \times 10^{-3}$)	1.17×10^{-4}
38: 5A5N \rightarrow 1A1N	5.05×10^{-11} ($1.87 \times 10^{-12}, 2.10 \times 10^{-10}$)	2.11×10^{-11}
39: 5A5N \rightarrow 2A2N	3.07×10^{-11} ($1.16 \times 10^{-12}, 1.28 \times 10^{-10}$)	1.31×10^{-11}

Table D3. Part 2. Evaporation rates (units given in s^{-1}) computed from a posterior distribution of the thermodynamic parameters (cluster formation enthalpies and entropies) which had previously been determined from the steady-state concentration measurements at temperatures 278 and 292 K. Here the mode of distribution (bold face) is given together with the range of possible values in the parenthesis. The last column presents the quantum-chemistry-based evaporation rates used for creating the synthetic data (borrowed from ?). The notation $xAyN$ corresponds to a cluster with x sulfuric acid and y ammonia molecules. [...⁴⁶¹]

Author contributions. Author Shcherbacheva A. produced the codes and conducted all the computational experiments for generation of the synthetic data and the MCMC parameter estimation, prepared all the plots presented in the manuscripts. Authors Balehowsky T. and Shcherbacheva A., Kurtén T. and Vehkamäki H. and Haario H. are responsible for writing the manuscript. Author Olenius T. assisted with

generation of the synthetic data, performed sanity check of the results, gave valuable comments regarding the manuscript. Authors Helin T. and Balehowsky T. actively participated in development of the methodological approach. Author Laine M. provided technical assistance with the 'mcmcstat' toolbox which was used for MCMC simulations. Author Kubečka J. assisted with the code compilation and debug. Author Haario H. assisted with interpretation of the MCMC results and proper usage of the DRAM computational method. Authors Kurtén T. and Vehkamäki H. helped to interpret the outcomes of the study.

Competing interests. The authors declare that they have no conflict of interest

500 *Acknowledgements.* We thank the European Research Council project 692891-DAMOCLES, Academy of Finland (project number 307331), and University of Helsinki: Faculty of Science ATMATH project, for funding, and the CSC-IT Centre for Science in Espoo, Finland, for computational resources. We also thank Olli Pakarinen (Institute for Atmospheric and Earth System Research, University of Helsinki, Helsinki, Finland) for advise in plotting the synthetic data used in the present study.