Response to review #1

We thank Referee 1 for the positive and constructive comments on our manuscript. We have carefully considered all suggestions and have revised the manuscript accordingly. In the following, we provide a point-by-point response to each comment.

Referee comment: The paper is a little redundant in spots and thus could be shortened (e.g., line 362).

Authors' response: We have revised Section 2.2.1 and 2.2.2 to remove redundant phrasing and streamline the description of the surrogate selection methods.

Changes to the manuscript: In Sect. 2.2.1 (formerly lines 326–328) the sentence "Figure S1 illustrates the binary solution ..." was deleted since the former Fig. S1 has been removed (see also response to a comment by referee 2). Lines 345–347, the sentence "Water represents a highly polar ..." was deleted as it is redundant given similar statements prior to that description. In Section 2.2.2, we have removed the sentences on line 362 and revised the following sentence to read: "The 2D space is subdivided into a number of grid cells (or clusters) based on the targeted reduction in system complexity."

Referee comment: It would be helpful to provide the reader with an approximate conversion from saturation vapor pressures (in Pa) to the C* variable (in μg/m^3) that is common in literature for VBS. I know this depends on molecular weight. I assumed a 200 g/mol molecular weight and used gas law to create an equivalent scale to help me interpret the figures. Maybe the authors could consider labeling an upper x-axis with C* assuming an average molecular weight.

Authors' response: We have included a secondary x-axis to the affected figures showing the approximate C_j° values (in μg/m^3) at the top corresponding to the main x-axis stating pure-component vapour pressure. Instead of simply assuming an average molar mass in the scale conversion, given the relatively large volatility and molar mass ranges covered, we opted for a slightly more accurate conversion by fitting a linear molar mass to vapour pressure relationship specific to the system of interest. This approach is now described in the supplement and the code to produce such conversions is included in the code repository of the 2D polarity–volatility framework. We also note that the C_j° values of individual compounds are computed and part of the output of the 2D framework.

Changes to the manuscript: Figures 7–11 were updated to include a secondary x-axis showing C_j° (μg/m^3) as alternative volatility metric. The caption of Fig. 7 was updated to include a brief description of this secondary axis. A new section S2 was added to the revised supplementary material document, describing the details of the conversion from a vapour pressure to an approximate C_j° axis. That section also includes a new Fig. S1 (included below), which shows the relationships between molar mass and saturation vapour pressure and between component saturation vapour concentration and saturation vapour pressure at $T = 298.15$ K for the two example SOA systems introduced in the main text.

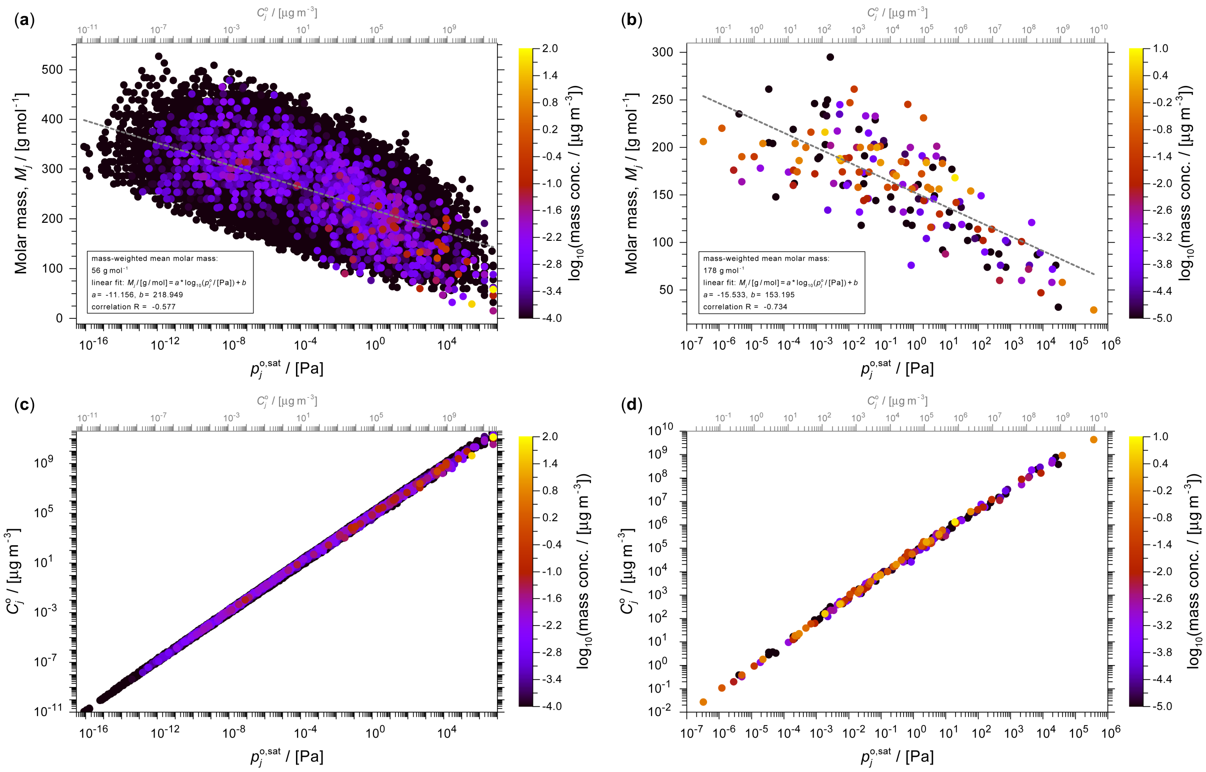

Figure S1: Relationships between molar mass and saturation vapour pressure (a, b) and between component saturation vapour concentration and saturation vapour pressure at $T = 298.15$ K. Panels (a) and (c) for the toluene-derived and panels (b) and (d) for α-pinene-derived systems of gas-phase reaction products. The linear fit in (a, b) is used in the conversion from the lower to upper x-axes in these and related figures.

Referee comment: The clustering algorithm seems like a novel approach with solid results. Even the simplified topologies give reasonable errors.

Authors' response: We thank the referee for this encouraging feedback and we agree that even relatively coarse resolutions give reasonable predictions. No changes were made in response to this comment.

Referee comment: I have one recommendation. I am wondering if the authors could create an additional table with results from a 1x3 parameter space. Where the 3 separates mass between water soluble, partially water soluble and water insoluble. Can the authors calculate the fraction of product mass and average ACR in these 3 bins for the two precursors (toluene, α-pinene)? Or if the readers feel more appropriate using 4 polarity bins to better resolve partial solubility space. I think this info would be very helpful to constrain chemical transport models. Most chemical transport models only resolve the 1D volatility space. Or maybe the mass concentration data from Figures 8b and 10b can be provided in tables so that readers could manipulate data to their model needs.

Authors' response: Rather than presenting a $1{\times}3$ polarity binning in the manuscript, to address this comment, we suggest that readers can analyze the output from, e.g. the $6{\times}3$ or $10{\times}5$ (volatility × polarity) grid for a system of interest. Data tables related to Figs. 8b and 10b have been added to the supplement. We add that the provided code repositories enable running the framework with a 1×3 grid space resolution, if desired. Furthermore, for the presented two systems (products from toluene or α-pinene oxidation), the lumping framework output for different resolutions are provided in the Zenodo repository at `https://doi.org/10.5281/zenodo.17088391` (mentioned in the *Code and data availability* section). For example, the toluene-derived data for the 6×3 grid and weighted medoid lumping are provided in that Zenodo repository under toluene > ACR_P0 > lumping_output > 6by3 > "LumpedConc_1263_Weighted_Medoid_06x03.txt" and "SystemCompProp_1263_Weighted_Medoid_06x03.txt". Those two files also provide additional information about the properties of these surrogate compounds, e.g. their molar masses and C_j° values.

Concerning the interpretation of the ACR metric in terms of water soluble, partially water soluble and water insoluble fractions, we caution that in case of a phase-separated aerosol, a polar organic compound may still predominantly partition into the organic-rich phase when the aqueous electrolyte-rich phase is of high ionic strength and/or when the organic-rich phase is at least moderately polar and greater in mass than the electrolyte-rich phase. Thus, while the ACR of a component is a pure-component property and indicator of phase polarity preference, the precise partitioning preference will also be affected by the abundance and properties of the other compounds in a specific (two-phase) aerosol system, as discussed by Zuend et al. (2010).

Changes to the manuscript: We have added Tables S5 and S6 in the Supplement, which summarize the surrogate mass concentrations and their vapour pressure and ACR values for the 10×5 resolution cases shown in Figs. 8b and 10b in tabular form. These additions are referenced in the related Figure captions of 8b and 10b as well as in Section 3.3 of the revised manuscript.

Referee comment: Please check the grammar on line 351 after words "activity coefficients ..."

Authors' response: The sentence has been revised; a few words were missing by accident.

Changes to the manuscript: In Section 2.2.1, we corrected the sentence to read: "Of note, for a system consisting of tens of thousands of organic components, computing activity coefficients of all species simultaneously using the related multicomponent mixtures (containing as many components) with AIOMFAC, is prohibitively slow due to the many functional groups present, all the possible group–group and molecule–molecule interactions that need to be summed over and the associated computer memory requirements."

References

Zuend, A., Marcolli, C., Peter, T., and Seinfeld, J. H.: Computation of liquid-liquid equilibria and phase stabilities: implications for RHdependent gas/particle partitioning of organic-inorganic aerosols, Atmos. Chem. Phys., 10, 7795–7820, https://doi.org/10.5194/acp-10- 7795-2010, 2010.

Response to review #2

We thank Referee 2 for the thoughtful and constructive comments on our manuscript. We appreciate the detailed feedback and suggestions, which have helped us improve the clarity and completeness of the paper. In the following, we provide a point-by-point response to each comment.

Referee comment: Sect. 2.1.1: The detailed description of the S2AS tool is appreciated. For people not familiar with AIOMFAC use, it could be useful to explain what is the improvement over existing methods. How were organic molecules decomposed into AIOMFAC subgroups before S2AS? I am afraid to learn it had to be done manually, in which case the authors could highlight the reduction in potential human error (and the time saved!) as an additional benefit from this tool.

Authors' response: We will clarify the novelty of S2AS relative to previous practice. Before S2AS, assigning AIOMFAC functional subgroups to an organic molecule was indeed often a manual or semi-manual process, feasible only for relatively small sets of compounds. The UManSysProp online tool (Topping et al., 2016) did include some automated subgroup assignment functionality via SMARTS matching, but as we noted (Sect. 2.1.2), the SMARTS patterns and logic in UManSysProp differ from ours and are not directly transferable. In particular, there was no publicly available general tool to parse any SMILES into AIOMFAC subgroups comprehensively. Thus, S2AS provides a significant improvement by automating this mapping for large numbers of molecules, with a carefully curated priority list of SMARTS patterns and consistent handling of exceptions (Sect. 2.1.1 and the discussion of findings in Sect. 3.2.1). This reduces the potential for human error and saves a substantial amount of time when dealing with thousands of species.

Changes to the manuscript: In Section 2.1.1, at the end of the first paragraph about S2AS, we added: "Previously, AIOMFAC subgroup assignments for organic molecules had to be determined either manually (for small sets of structures) or using limited tool-specific pattern matching (e.g., the UManSysProp facility; Topping et al., 2016). Our S2AS tool automates this process for arbitrary molecules. It can process tens of thousands of compounds in a consistent way, whereas manual assignment would be prohibitively laborious and prone to errors or inconsistencies"

Referee comment: Fig. S1: The general effort to illustrate this paper is appreciated, but I fail to see the purpose of this figure. Unless I missed something, it seems very trivial and does not benefit the understanding of Sect. 2.2.1.

Authors' response: We agree with the referee that Fig. S1 does not add substantial value to the understanding of Section 2.2.1. We have therefore removed this figure from the Supplement.

Changes to the manuscript: Figure S1 has been removed from the Supplement and related text in Sect. S1 adjusted. The reference to this figure in Sect. 2.2.1 of the main text has also been deleted.

Referee comment: Sect 3.1: I guess the lumping and AIOMFAC calculations are carried out for the mixture obtained at the end of each of the example simulations. It would be important to clarify this. With this remark come questions that may be addressed in the conclusion: how would this framework be applied when time-stepping through an atmospheric chemistry model? Would the surrogate species be recomputed at each call to AIOMFAC? How would the gas-aerosol partitioning of surrogate species be applied back to the explicit gas species to initialize the next gas chemistry timestep while ensuring mass conservation?

Authors' response: We confirm that the lumping and AIOMFAC calculations were performed on the product distributions at a specific (final) time of the simulations. We have clarified this in Section 3.1. We have added a new paragraph in the Conclusions section (see below) outlining several options for how the 2D framework could serve in offline or online applications of atmospheric chemistry and gas–particle partitioning models.

Changes to the manuscript: Section 3.1 (first paragraph) now includes a sentence clarifying that the lumping was applied to the data from the final time step of the mechanism simulations. The Conclusions section was expanded by the following paragraph.

We envision a few distinct options for future applications of this framework in different kinds of atmospheric chemistry models. (1) Within detailed chemical box or plume models, those that consider a large number of compounds and retain their molecular structure information, the computation of surrogates and subsequent gas–particle partitioning at each desired (output) time step may be the preferred option. (2) Alternatively, based on a separate offline calculation for a specific system, a fixed set of surrogate compounds could be determined with the 2D framework. Subsequently, at each time step, existing and newly formed compounds from the box model's chemical mechanism could be mapped to this conserved, predetermined set of surrogates using the closest normalized Euclidean distance to the various surrogates in the 2D space (similar to Eq. 6) to determine the surrogate to which a compound's mass will be lumped. (3) In the case of simplified chemical mechanisms, such as those often employed in large-scale chemical transport models, maintaining only a few organic aerosol surrogates or a 1D/2D VBS representation, the application differs since surrogate lumping during simulations is unnecessary. In that case, the 2D framework could serve in systematically generating sets of surrogate components after mechanism simulations (e.g. with GECKO-A) for targeted aerosol precursors (structure-resolved) or aid in generating 2D VBS bin-resolved (structure-agnostic) representations at desired polarity–volatility resolutions. In the latter case, the 2D lumping step may serve in assigning surrogates in the ACR vs. p° space and in translating the resulting surrogate mass concentrations into bin-based mass concentrations, e.g. in the O:C vs. C° coordinate space. In the case of atmospheric chemistry models that retain the molecular structure information of surrogates, we envision two options for invoking equilibrium gas–particle partitioning calculations during simulations. (i) Applying the gas–particle partitioning calculation offline at specific output times during a simulation while running the gas-phase chemical mechanism as if all material remained in the gas phase (no feedback from partitioning). (ii) Running the 2D lumping framework and the gas–particle partitioning method at every simulation time step, followed by treating the determined fractional surrogate amounts partitioned to the particle phase as partially or fully shielded from further gas-phase chemical reactions. The gas-phase fraction of a surrogate would then be applied to the list of associated compounds, updating their molecular gas-phase concentrations prior to the next chemical reaction step in the simulation. Optionally, reactions in the condensed phase could be treated separately by a distinct mechanism.

Referee comment: Conclusion (l.708–715): I agree with the authors that this new framework is a good step toward reducing the complexity of organic chemistry models for application in large-scale models. However, as it is described here, this framework cannot be applied in a large-scale model because it still relies on the explicit description of gas phase organic chemistry. Could the authors please expand on how they see their tools being used in the future? For instance, could this framework be treated as a reference when creating future 2D-VBS-like parametrisations? Is there any chance this type of approach could be used to simplify the gas component?

Authors' response: We have expanded the Conclusions section to discuss how our framework may serve for developing reduced SOA mechanisms, including for 2D-VBS applications (see response to the previous comment). Regarding simplifying the gas component, we assume that this question is about simplifying a gas-phase chemical mechanism. Such applications are outside the scope of this study. However, existing and new approaches for systematic chemical mechanism reduction, such as the GENOA (Wang et al., 2022) and AMORE (Wiser et al., 2025) approaches, could potentially benefit from including our 2D lumping when categorizing reactants during mechanism reduction steps.

Changes to the manuscript: The Conclusions section was revised to include the following paragraph.

A computationally effective use of near-explicit gas-phase chemical mechanisms in atmospheric chemistry models benefits often from a tunable reduction in the complexity of the mechanism itself, both in terms of number of explicit species and number of reactions covered. Methods such as the GENerator of reduced Organic Aerosol mechanism (GENOA) (Wang et al., 2022) and the Automated MOdel REduction (AMORE) algorithm based on graph theory (Wiser et al., 2025) serve this purpose. When targeting SOA formation applications, AMORE v2.0 employs a 2D categorization based on the saturation vapour pressures and Henry's law constants of organic components, which is similar to the polarity–volatility space of our 2D framework. Further development of such rule-based mechanism reduction methods may therefore benefit from considering also our 2D framework for potential application in compound classification.

Referee comment: Technical Corrections:

1. l. 320: dipol-dipol → dipole-dipole.
   Authors' response: Corrected to dipole–dipole" in Section 2.2.1.
2. l. 350 "Of note, ... AIOMFAC,": confusing sentence, missing a word?
   Authors' response: Revised the sentence for clarity as described in response to Referee #1.
3. l.611–612: this is repeating what is written in l. 608.
   Authors' response: Sentence on line 608 shortened to: "Additionally, the relatively small

modelled SOA mass concentrations contribute to the observed metric fluctuations, since minor absolute differences can result in larger relative errors."

4. l.716–719: a bit pompous, is this paragraph really needed?
   Authors' response: We deleted this final paragraph of the Conclusions section.

5. l.737: I am almost certain that Bernard Aumont is a professor, please check.
   Authors' response: This comment may reflect different norms. In the Acknowledgements, we had referred to "Dr. Bernard Aumont", which is consistent with common practice in acknowledgements. While Bernard Aumont holds a professorial position, we prefer to retain the current phrasing or simply refer to "Bernard Aumont" without a degree/title, whichever is preferred by the GMD journal. This aligns with typical conventions in scientific acknowledgements, where ranks/roles/titles such as "Professor" are often omitted or replaced with "Dr." to reflect earned academic degrees rather than job titles.

References

Wang, Z., Couvidat, F., and Sartelet, K.: GENerator of reduced Organic Aerosol mechanism (GENOA v1.0): an automatic generation tool of semi-explicit mechanisms, Geosci. Model Dev., 15, 8957–8982, https://doi.org/10.5194/gmd-15-8957-2022, 2022.

Wiser, F., Sen, S., Wang, Z., Lee-Taylor, J., Barsanti, K. C., Orlando, J., Westervelt, D. M., Henze, D. K., Fiore, A. M., Berman, A., Carter, R., and McNeill, V. F.: A graph theory-based algorithm for the reduction of atmospheric chemical mechanisms, PNAS Nexus, 4, 11, https://doi.org/10.1093/pnasnexus/pgaf273, 2025.