S2AS v1.0 and 2D polarity-volatility lumping framework v1.0: automated compound classification and scalable lumping for organic aerosol modelling (Amaladhasan et al., 2025)

General Comments

In this paper, Amaladhasan et al. introduce a new lumping framework to reduce the complexity of organic aerosol modeling, while preserving some molecular properties knowledge. They present a methodology aimed at processing the output of near-explicit atmospheric chemical schemes (e.g. MCM, GECKO-A) to obtain a simplified description of gas-particle partitioning and allow the prediction of SOA concentration and composition with thermodynamic equilibrium models such as AIOMFAC. In an approach inspired by the 2D-VBS, they develop a new two dimensional representation framework for organic compounds relying on several newly developed tools. They address the question of lumping explicit species onto the 2D framework with a statistically based approach.

The first tool of this framework is the SMILES to AIOMFAC subgroups (S2AS) tool. S2AS matches SMILES strings to a hierarchical list of SMARTS patterns designed to produce the subgroup descriptors used by AIOMFAC. If my understanding is correct, this is a clever systematization, and simplification, of the AIOMFAC molecular descriptions methods. The existing UManSysProp tool is also used here to calculate temperature-dependent pure-component vapour pressures with the EVAPORATION model.

A new polarity metric for the 2D-VBS based on the ratio of activity coefficients in two solvents, 1,2-hexanediol and water, is introduced. This metric is relatively simple to compute with AIOMFAC and should intrinsically contain more information about a given molecule's affinity with polar mixtures than the usual O/C ratio or average carbon oxidation state $\overline{\mathrm{OS}}_\mathrm{C}$.

Four different methods for lumping explicit species on the 2D-VBS grid are presented. These methods differ on the algorithm used to select an explicit compound representative of a section of the 2D domain. They present the advantage of being flexible with the size of the 2D-VBS grid.

The new framework is applied to two test cases of α-pinene and toluene oxidation respectively described with MCM and GECKO-A explicit schemes. These tests are used to evaluate the 2D-VBS model ability to predict SOA mass concentration and hygroscopicity (κ) when using different polarity axes, lumping algorithms and 2D-VBS grid resolutions.

This paper is providing convincing tools to address the question of reducing the complexity of the organic gas-particle problem in a systematic and rigorous way. The reader is only left hoping that the application of this framework in a real case scenario is coming soon. The descriptions of the various tools and methods used in this framework are clear, detailed, nicely illustrated and very well written.

I therefore recommend accepting this very nice paper in GMD after the authors address the following minor comments as they see fit.

Specific Comments

1. **Sect. 2.1.1:** The detailed description of the S2AS tool is appreciated. For people not familiar with AIOMFAC use (including myself), 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.

2. **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**.
3. **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 a 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?
4. **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 parametrizations? Is there any chance this type of approach could be used to simplify the gas component?

Technical Corrections

1. **l. 320:** dipol-dipol ⟶ dipole-dipole.
2. **l. 350 "*Of note, ... AIOMFAC,*"**: confusing sentence, missing a word?
3. **l.611–612**: this is repeating what is written in **l. 608**.
4. **l.716–719**: a bit pompous, is this paragraph really needed?
5. **l.737**: I am almost certain that Bernard Aumont is a professor, please check.