Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2019-495-RC2, 2019
© Author(s) 2019. This work is distributed under
the Creative Commons Attribution 4.0 License.

Atmospheric Chemistry and Physics
Discussions

Interactive comment on "RH-dependent organic aerosol thermodynamics via an efficient reduced-complexity model" *by* K. Gorkowski et al.

Anonymous Referee #2

Received and published: 12 August 2019

Gorkowski et al. present a modeling approach to predict the water content, CCN activity, liquid-liquid phase separation, and gas-particle partitioning of single component and mixed organic aerosol. The focus of the work is to produce reduced complexity models that have fast runtime while preserving the fidelity of the predictions. This is achieved by training the reduced complexity model using more computationally expensive modeling framework.

This manuscript is an ambitious attempt to contribute to the efficiency of modeling a wide range of organic aerosol processes. Constructing such comprehensive and fast models is technically demanding and the authors should be complimented for their often clever approaches. For example, conceiving and finding a suitable set of fitting coefficients that represent OA through Eq. (18) and (19) is impressive. A selected set of

validations is presented, and these validations appear to demonstrate that the reduced complexity models are adequate. However, I do have concerns about the stability and validation of the model. A detailed formal evaluation of the BAT and NN model that is independent of training data is needed. Furthermore, more systematic validation of the model predictions against experimental data is needed, especially against single-component CCN data. I anticipate that the paper will be acceptable for publication if formal, systematic, and independent validation is included.

Major comments

Both the BAT model and the NN model are trained. Figure 2a shows the training points for the BAT model. The standard approach in machine learning is to have a training set for which the model is optimized, and a validation set for compounds that the model has not been tuned to. This does not appear to have been done and one might seriously question the fidelity of the model outside the training set. Showing activity for citric acid is insufficient. I recommend that the authors test the model against 100 (or so) compounds that were not used in the optimization and show a scatter plot of AIOMFAC vs. BAT for activity coefficient at various RH, mole-fraction of the predicted miscibility gap, water activity of the miscibility gap, and predicted kappa CCN. Only through independent tests and systematic characterization of the error can one be confident that the BAT coefficients truly represent AIOMFAC. A similar case is to be made for the NN training. Systematic validation against with non-training data needs to be presented.

Related to this point: "Therefore, in an attempt to design a more general organic activity coefficient model, we made two important changes. First, we change the independent composition variable used in Eq. (15). Instead of mole fraction xorg , we introduce a scaled volume fraction (φorg) in the series expansion of GE/RT. Second, we introduce a parameterization of the scalar c0 n coefficients by means of multivariate functions, which are dependent on common characteristics of organic molecules."

(1) Please provide some rationale why switching to volume fraction was required. It is not clear to the referee or from the text.

(2) I don't understand why using Redlich-Kister was selected. The RK expansion can fit to arbitrary precision. While it is true that the model is thermodynamically consistent in the limit of x = 1 and x = 0, the polynomial can lead to maxima and minima in the excess function that may or may not be realistic. This is particularly concerning since the polynomial coefficients are themselves computed through a multivariate model. (It is impressive that the system converged). I am concerned that the BAT model coefficients are overfitted and not be representative of other compositions, especially with respect to a phase separation which represents a discontinuity and is highly sensitive to very small fluctuations in the excess function. This issue re-emphasizes the need to independently verify the fidelity of the BAT model. When addressing this concern, the authors should discuss why RK was selected instead of the Margules or van Laar model, which would be less sensitive to error from the use of polynomials by limiting the shape of the excess function.

There are a few comparisons to actual data. While it is clear that the model cannot be compared to every data point in the literature, the real-world validation appears not to be systematic. For example, it's not clear why single component data from Marsh et al. (2019) was selected for hygroscopic growth and various a-pinene SOA for CCN. The composition dependence of subsaturated water content on O:C for SOA (e.g. Pajunoja et al., 2015, doi:10.1002/2015GL063142) is far more revealing than the supersaturated data. Many data sets for single subsaturated water uptake of single component organic aerosol are available. Ideally a proper validation set would systematically probe O:C and functional group coverage, and would considers experimental error. A plot like Figure 11 should be made for available single component CCN data, including for cases where LLPS is known to control CCN activity. A validated dataset with comparison against UNIFAC/LLPS is available in Petters et al. (2016, GMD, 10.5194/gmd-9-111-2016).

Other comments

The tone of the hygroscopic growth and CCN section should be revised. For examples

"Over the past decade, the research community then progressed by characterizing (organic) aerosol hygroscopic growth measurements by a single κ value, with sometimes inconsistent distinction between a κ value at subsaturated and supersaturated humidity conditions."

"Our clear distinction between κCCN and the more general κHGF helps the community understand clearly the subsaturated and supersaturated behavior of organic aerosol"

While it is true that there has been a debate on κCCN and κHGF the authors should acknowledge that 100s of experimental and modeling papers were devoted to this subject, with many important individual contributions explaining the origin of the discrepancy and the composition dependence of κHGF. While the BAT model may capture some of these now very well understood effects, it does not really reveal anything new. Please rephrase the text and/or provide a more nuanced perspective on the topic.

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-495, 2019.