Supplement of

S2AS v1.0 and 2D polarity–volatility lumping framework v1.0: automated compound classification and scalable lumping for organic aerosol modelling

Dalrin Ampritta Amaladhasan[1], Dan Hassan-Barthaux[1], and Andreas Zuend[1]

[1]Department of Atmospheric Oceanic Sciences, McGill University, Montreal, Quebec, H3A 0B9, Canada

Correspondence: Andreas Zuend (andreas.zuend@mcgill.ca)

S1 Activity coefficients of binary solutions

Figure S1 shows an illustration of organic species dissolved separately in a polar solvent like water and a less polar solvent like 1,2-hexanediol as part of demonstrating the concept of using binary solutions to assess the solvent affinities and relative polarity of organic compounds. The activity coefficient ratio (ACR) is calculated by comparing the solute's activity coefficients in these two solvent environments, as predicted by the AIOMFAC model from the binary mixtures containing 0.01 mass fractions of solute. This approach provides a computationally efficient method to characterize the behaviour of organic components across a wide range of polarities. The use of water and 1,2-hexanediol as reference solvents establishes a robust basis for this characterization. Water represents a highly polar solution environment typical of aqueous aerosol phases, while 1,2-hexanediol serves as a proxy for moderately oxidized organic aerosol components that would preferentially partition into an organic-rich phase in the presence of liquid-liquid phase separation (LLPS). This binary solution setup allows for a rapid calculation of ACR values for the numerous organic components from near-explicit mechanism simulations or other rich data sources. The relatively inexpensive binary solution computations with AIOMFAC are indicative of the expected partitioning preferences of organics in atmospheric aerosols without the computational cost of multicomponent mixture calculations.

S2 Example: SOA mass concentrations calculated using lumped surrogates

The pure-component properties calculated by the S2AS and vapour pressure tools and the lumped mass concentrations obtained for the selected surrogate components can be used by the AIOMFAC-based equilibrium partitioning model to generate the secondary organic aerosol (SOA) prediction of the system. Figure S2 illustrates the predicted SOA mass concentrations as a function of water activity (i.e. equilibrium relative humidity) for the toluene SOA system. The water content itself is not shown in this figure. In this example, the surrogate components were selected via the grid cell mass-weighted medoid method on a 10×5 grid, with ACR as the polarity axis metric. This selection method emphasizes the influence of both the component

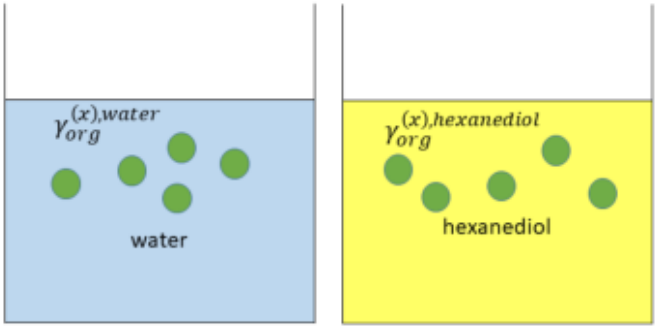

Figure S1. Schematic representation of the binary solution setup with small amounts of an organic solute component (green symbols) dissolved in water (left) or 1,2 -hexanediol (right) serving as more or less polar solvents. The ratio of the solute's activity coefficients predicted by AIOMFAC for both binary solutions provides the ACR metric.

Table S1. Initial precursor and oxidant concentrations as well as the MCM V3.3.1 (AtChem) simulation conditions used for the α-pinene ozonolysis system.

Species	Concentration ($\mathrm{molec\,cm^{-3}}$)
α-pinene	1.23×10^{17}
O_3	2.46×10^{12}
NO	4.43×10^{8}
NO_2	5.90×10^{8}
CO	4.92×10^{12}
H_2	1.23×10^{13}
H_2O	3.00×10^{17}
Temperature	298.15 K
Pressure	1013.25 hPa (1 atm)
Simulation duration	2.8 hours

positions in the 2D grid space and their relative mass contributions. Figure S2a shows that most of the gas-phase mass concentration is due to the special high-volatility surrogate. Panels (b) and (c) show the equilibrium compositions of the liquid phases – with two liquid phases coexisting here only at the highest water activity levels. Note that only the first (top) 30 organic surrogate components are shown in the legend. For this toluene-derived SOA case, the relatively low hygroscopicity is evident from the weak dependence of the condensed-phase mass concentration on the water activity.

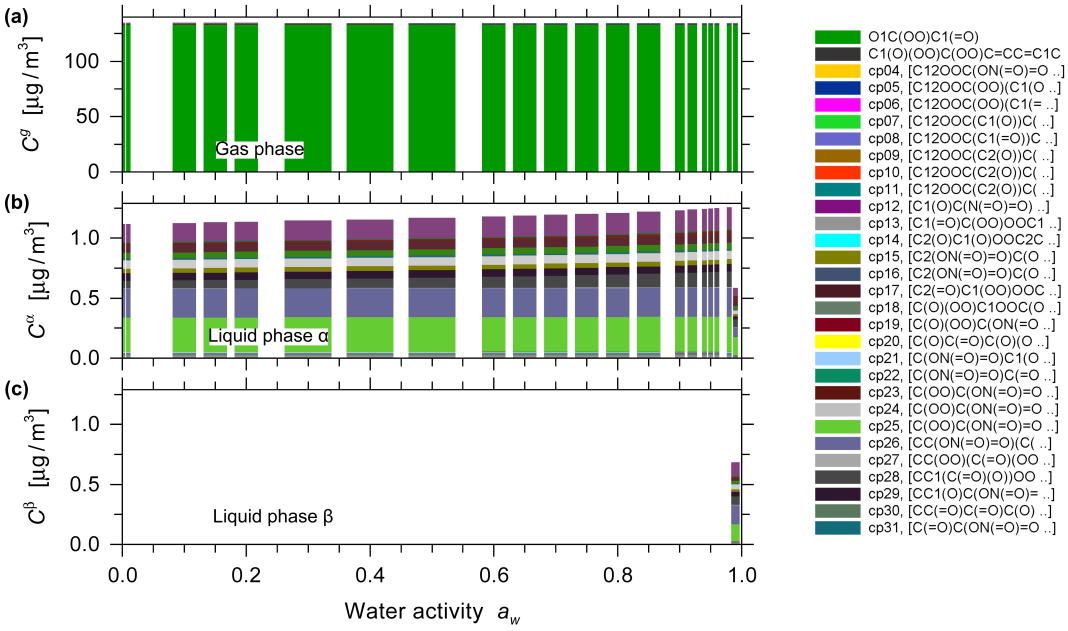

Figure S2. Predicted speciated gas phase (a) and liquid phase (b,c) SOA mass concentrations for the toluene SOA system (water content not shown). The shown surrogate species (legend limited to top 30) were obtained from the 2D lumping framework by means of the mass-weighted medoid surrogate selection method with a 10×5 grid resolution. Liquid–liquid phase separation was predicted to occur at high water activity, where amounts from panels (b) and (c) contribute to the total non-water aerosol mass.

Table S2. Initial precursor and oxidant concentrations as well as GECKO-A model simulation conditions for the toluene system.

Species	Concentration ($\mathrm{molec\,cm^{-3}}$)
Toluene	2.75×10^{11}
OH	5.11×10^{6}
HO_2	8.41×10^{7}
O_3	1.00×10^{12}
H_2O_2	2.22×10^{9}
NO_2	3.58×10^{11}
NO	1.42×10^{11}
NO_3	3.12×10^{6}
Temperature	298.15 K
Pressure	1013.25 hPa (1 atm)
Simulation duration	24 hours
Timestep size	5 minutes

Table S3. Comparison of predicted SOA mass concentrations and hygroscopicity parameter κ at 298 K for different surrogate selection methods and polarity metrics across four grid resolutions. Data shown are for the α-pinene system.

Surrogate selection	y-axis	4×2		6×3		8×4		10×5	
		SOA ($\mu g\,m^{-3}$)	κ (–)	SOA ($\mu g\,m^{-3}$)	κ (–)	SOA ($\mu g\,m^{-3}$)	κ (–)	SOA ($\mu g\,m^{-3}$)	κ (–)
Midpoint	$\log_{10}\left[\gamma_{j,\mathrm{hex}}^{(x)}/\gamma_{j,\mathrm{w}}^{(x)}\right]$	1.728	0.042	1.148	0.051	1.584	0.045	1.874	0.040
Medoid		1.613	0.051	1.162	0.049	1.170	0.047	1.834	0.041
Weighted medoid		1.234	0.045	1.586	0.043	1.600	0.043	1.624	0.043
Weighted k-means		1.581	0.042	1.560	0.042	1.544	0.043	1.549	0.044
Midpoint	O:C	2.350	0.076	1.608	0.050	1.856	0.052	1.953	0.045
Medoid		1.451	0.082	1.669	0.061	1.637	0.060	2.209	0.050
Weighted medoid		1.582	0.049	1.845	0.044	1.671	0.050	1.665	0.051
Weighted k-means		1.629	0.050	1.595	0.042	1.536	0.043	1.548	0.044
Midpoint	$\overline{\mathrm{OS}}_{\mathrm{C}}$	2.656	0.059	1.702	0.055	1.932	0.041	1.565	0.041
Medoid		1.569	0.058	1.148	0.045	1.781	0.053	1.578	0.041
Weighted medoid		1.182	0.049	1.542	0.042	1.534	0.041	1.596	0.040
Weighted k-means		1.570	0.040	1.557	0.041	1.563	0.042	1.559	0.043

Table S4. Comparison of predicted SOA mass concentrations and hygroscopicity parameter κ at 298 K for different surrogate selection methods and polarity metrics across five grid resolutions. Data shown are for the toluene-derived SOA system.

Surrogate selection	y-axis	4×2		6×3		8×4		10×5		25×10	
		SOA ($\mu g\,m^{-3}$)	κ (–)	SOA ($\mu g\,m^{-3}$)	κ (–)	SOA ($\mu g\,m^{-3}$)	κ (–)	SOA ($\mu g\,m^{-3}$)	κ (–)	SOA ($\mu g\,m^{-3}$)	κ (–)
Midpoint	$\log_{10}\left[\gamma_{j,\mathrm{hex}}^{(x)}/\gamma_{j,\mathrm{w}}^{(x)}\right]$	1.219	0.043	1.130	0.031	1.239	0.054	1.175	0.034	1.201	0.041
Medoid		1.256	0.039	1.103	0.062	1.179	0.054	1.168	0.031	1.194	0.045
Weighted medoid		1.300	0.047	1.234	0.065	1.184	0.045	1.188	0.040	1.202	0.038
Weighted k-means		1.230	0.030	1.188	0.039	1.205	0.039	1.210	0.037	1.161	0.043
Midpoint	O:C	1.307	0.103	1.038	0.059	1.176	0.032	1.136	0.046	1.131	0.024
Medoid		1.239	0.098	1.157	0.047	1.256	0.028	1.106	0.049	1.135	0.035
Weighted medoid		1.263	0.062	1.184	0.046	1.213	0.064	1.211	0.060	1.152	0.043
Weighted k-means		1.188	0.035	1.163	0.062	1.104	0.031	1.113	0.038	1.181	0.038
Midpoint	$\overline{\mathrm{OS}}_{\mathrm{C}}$	1.278	0.028	1.083	0.063	1.252	0.035	1.154	0.050	1.220	0.055
Medoid		1.140	0.029	1.112	0.052	1.249	0.081	1.165	0.032	1.221	0.048
Weighted medoid		1.293	0.033	1.151	0.040	1.151	0.024	1.140	0.025	1.234	0.049
Weighted k-means		1.202	0.072	1.185	0.035	1.160	0.055	1.195	0.053	1.185	0.047