Atmospheric
Chemistry
and Physics
Discussions

Open Access

Interactive comment on "Technical Note: Development of chemoinformatic tools to enumerate functional groups in molecules for organic aerosol characterization" *by* G. Ruggeri and S. Takahama

G. Ruggeri and S. Takahama

giulia.ruggeri@epfl.ch

We thank both reviewers for their extremely useful comments. Based on a general consensus that the focus of the Technical Note should be on the novelty of the tool itself and less on the example applications, we have made a few general changes:

- Methods (Section 2.1) describing the SMARTS pattern development is expounded

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

- Example Applications (Section 3.2) is made much more concise

- Conclusion (Section 4) is modified to summarize the chemoinformatic tools, more than interpretation from the specific applications presented

We address specific comments to the reviewer below:

1. **Comment**: I am unsure whether Atmospheric Chemistry & Physics is the most suitable journal for this manuscript. Further, in its current form, the manuscript needs to be simplified to be an effective Technical Note. I found the diversity of datasets and example applications distracting. For example, in Fig. 6, the authors are comparing functional group distributions derived from GC /MS measurements made during the late 1990's with functional group distributions derived from FTIR measurements in 2010; also included are estimates of measured/unmeasured fractions for each instrument based on additional (and separate) publications. The authors spend one paragraph describing each of the measurements, the dominant functional groups, and the reasons for the discrepancies between the wood burning samples. In the conclusions section, the authors return to this discrepancy between the GC/MS and FTIR measurements for wood smoke, which has little to do with the overall contribution of the Ruggeri and Takahama work. It would be sufficient to demonstrate that functional group distributions can be obtained using their approach. Further comments are provided below on the inclusion of likely extraneous information that diminishes the potential contribution of this work. It is recommended that the authors consider simplifying the applications presented, and focus more discussion on validation.

   Response: We propose this manuscript to Atmospheric Chemistry & Physics since its readers are the target audience for the new tools we introduce; as demonstrated in our example applications the tools can be used to analyze both

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

measurements and model simulations. We recently also submitted a companion paper to Atmospheric Chemistry & Physics to further demonstrate its application in the analysis of model simulations of secondary organic aerosol formation, which is a prime topic for this journal. Furthermore we have now restructured the Conclusions and the Section 3.2 in order to give more importance to the chemoinformatic patterns developed and validated in this work.

Section 3.2.2: "From this comparison, we find that the oxidized fraction is much higher in the biomass burning aerosol composition estimated by FTIR. The high abundance of alkane CH bonds in the compounds reported by GC-MS can be explained by the preference of this analytical method to characterize the least oxidized fraction of the collected aerosol. While high abundance of carbonyl groups are reported in FTIR measurements of biomass burning aerosol (Liu et al., 2009; Russell et al., 2009; Hawkins and Russell, 2010), more recent methods including advanced derivatization (Dron et al., 2010) are necessary for quantification of carbonyl containing compounds by GC-MS. In addition, neither amine compounds nor levogucosan were reported in this GC-MS study. Levoglucosan is a polysaccharide compound often used as a tracer for burning and decomposition of cellulose reported in modern GC-MS measurements (Simoneit, 1999). However, FTIR does not report high fraction of alcohol COH as levoglucosan near particular fuel sources may be found mostly in supermicron diameter particles (Radzi bin Abas et al., 2004) (submicron OA was analyzed by Hawkins and Russell, 2010), its degradation in the atmosphere is rapid (Hennigan et al., 2010; Cubison et al., 2011; Lai et al., 2014), and the overall mass contribution to biomass burning OA is small (less than 2% by mass, Leithead et al., 2006).

"Both estimation methods agree that more than 90% of OM mass is composed of alkane-CH for vehicle sources The fraction characterized by GC-MS and FTIR with PMF have associated uncertainties from derivatization and thermal separation in the chromatography column or in statistical separation, respectively, and

lead to different fractions of mass reported. However, the approximate consistency in FG abundances estimated by the two methods, suggest that the fraction not analyzed by the GC-MS may not vary significantly from the measured fraction by FTIR in these aerosol types."

Section 4: "We introduced the application of chemoinformatic tools that allow us to perform substructure matching in molecules to enumerate FGs present in compounds relevant for organic aerosol chemistry. We developed 50+ substructure patterns and validated them over a list of 125 compounds that were selected in order to account for all the functional groups (FGs) represented. We demonstrate how these tools can facilitate intercomparisons between GC-MS and FTIR measurements, and mapping of compounds onto the VBS space described by pure component vapor pressure and oxidation state.

"We further introduce a novel approach for defining a set of patterns which accounts for each atom in a chemical system once and only once (except for polyfunctional carbon atoms associated with multiple FGs); this condition is confirmed by an atomic-level validation scheme applied to chemically explicit α-pinene and 1,3,5-TMB degradation mechanisms. This validation scheme permits apportionment of quantities such as O:C, H:C, and N:C to contributions from individual FGs, which are conventionally analyzed as aggregate measures. We illustrate its application to the photochemical degradation of α-pinene from speciated simulations using MCMv3.2."

2. **Comment**: p 33638, line 20-23: The description of the partitioning calculations is confusing. It is not clear what parameters were of interest, or how they were determined. The calculated pure compound vapor pressures could be used to calculate partitioning coefficients, which then could be used (with total OA) to calculate the fraction of each compound in the gas vs. particle phase based on Pankow 1994. Alternatively, calculated particle phase fractions (with total OA) could be used to calculate pure component saturation concentrations based on

Donahue et al. 2006. It is suggested that the Greek symbol zeta not be used to define the partitioning coefficient, as that symbol traditionally has been used to define mole-fraction scale activity coefficients in partitioning calculations.

Response The phrase has been restructured as suggested. The Greek symbol used here is actually xi and not zeta, but as it is difficult to discern it in the font used by the journal we specified that it is a xi after the symbol. The phrase has been restructured as: "After diluting the total OA of a factor of 1000 the compounds were partitioned between the two phases based on the partitioning coefficient ξ_i (xi) calculated from the pure component saturation concentration (C_i^0) as described by Donahue et al. (2006)."

3. **Comment** Page 33639/33640, discussion of Fig. 5: It appears the authors are mapping compounds measured by Rogge et al. in vehicle OA and woodsmoke OA, as well as compounds predicted by MCM. If this is correct, the detailed discussion of PMF/AMS HOA and BBOA in the text does not add anything, and only confuses what is presented in Fig. 5. Additionally, the legend should make it clearer that compounds from woodsmoke POA (and not biomass burning) and vehicle POA are plotted. It does not make sense to me that the compounds in woodsmoke POA span a higher volatility range than the compounds predicted by MCM. Particularly given that MCM is typically unable to predict measured OA loadings without invoking particle phase or aqueous phase chemistry. The authors do reference the abundance of MCM compounds in the IVOC region and absence in the LVOC, as well as particle-phase chemistry. . .maybe the traces are mislabeled and the green trace is MCM?

Response We apologize for the mislabeling of the colors in the legend, which have now been corrected. In replotting, we also discovered an omission of four large, multifunctional compounds with log C^0 of -4 in the set of wood-burning compounds, which have also now been included. We have substantially shortened the discussion of the mapping in Section 3.2 to highlight the main points to

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

be communicated by Figure 5:

"The algorithm described has been used to project molecular composition of GC-MS and MCM compounds to 2D-VBS space delineated by carbon oxidation and pure component saturation concentration (C^0) (Figure 5). The properties of vehicle-related primary OA and wood combustion compounds measured by GC-MS are generally consistent with those reported for hydrocarbon-like OA and biomass burning OA, respectively, derived from PMF analysis of AMS spectra (Donahue et al., 2012). The low oxidation state is observed on account of more than 60% of carbon atoms being associated with methylene groups (-CH_2-, oxidation state of -2) in long-chain hydrocarbon compounds, and an association to lesser degree with CH groups in aromatic rings (oxidation state of -1) and methyl groups (-CH_3, oxidation state of -3).

"Most compounds in the MCMv3.2 system correspond to intermediate volatility organic compounds (IVOC), with only a small fraction with the semivolatile organic compound (SVOC) regime. When using of MCMv3.2 for simulation of secondary OA formation, additional mechanisms (e.g., in the condensed phase) are necessary to introduce low volatility organic compounds (LVOC) as observed in atmospheric and environmental controlled chamber observations (Ehn et al., 2014; Shiraiwa et al., 2014). Higher oxidation states than for compounds in the GC-MS set are observed on account of the larger number of functional groups containing electronegative atoms (oxygen and nitrogen) bonded to carbon."

4. **Comment** The caption for table 1 is long and contains important information that is not thoroughly discussed in the text. It is recommended that the discussion in the text be expanded to cover the inclusion of special patterns for formaldehyde and formic acid.

Response We added this part in Section 2.1 and rephrased the text as follows:

"Therefore, we provide an example for the aldehyde FG group to illustrate the

development process, with particular attention paid to the description of atoms returned in the matched set and how their bonding environments are defined. We first describe a formulation specific for fulfilling the atom-level validation which requires two patterns to account for all aldehyde groups in the system, and an alternate formulation for only enumerating FGs that requires only a single pattern.

"When applied to propionaldehyde, the set of atoms returned by matching the pattern for substructure 9 in Table 1 will be 3, 4, 10 (as labeled in Figure 1a). The first bracket `[CX3;$(C([#1])(=[O])[#6])]` describes the carbon atom to be matched and returned. `CX3` describes a carbon with 3 bonds (effectively sp^2); `$(C([#1])(=[O])[#6])` qualifies that it is bonded to hydrogen, oxygen, and another carbon. The expression `(=[O;!$([O][O])])` describes the double-bonded oxygen to this carbon atom; `!$([O][O])]` excludes preventing matching of $C=O^+\text{-}O^-$ (defined as a separate group, substructure 21 in Table 1) that are present in other molecules (an example is provided in Figure 1b). The last bracket `[H]` is included to explicitly include the hydrogen atom in the returned set. While the sp^3 carbon attached to the sp^2 is not returned in the set of matched atoms, this additional specificity is necessary to prevent double counting of the same aldehydic group in the formaldehyde molecule, which contains two hydrogen atoms bonded to sp^2 carbon. A separate SMARTS pattern is defined for formaldehyde (Table 1 substructure 15). (For similar reasons, a SMARTS pattern specific for formic acid has been specified alongside the carboxylic FG.)

"In this approach, all atoms in the aldehyde group are matched instead of just the identifying carbon, oxygen, or hydrogen. The advantage of this strict protocol is that we can devise a validation such that each atom in a molecule or chemical system is accounted for by one and only one group — except for polyfunctional carbon — for any proposed set of FGs (Appendix A). Fulfillment of this validation criterion provides a means for interpreting atomic ratios commonly used by the community (e.g., O:C, H:C, and N:C) through contributions of distinctly defined

FGs.

"Revisiting the aldehyde FG example, an alternative pattern specified only for the purposes of counting FGs for use in SIMPOL.1 is shown in Table 2. We only describe the bonding environment of the sp^2 carbon and count the number of its occurrence, so a single pattern can be used for both formaldehyde and other aldehyde compounds."

5. **Comment** Table 3 is not referenced in the text. Is it necessary?

   Response We thank the reviewer for pointing out this omission. This table is meant to offer a connection between the FTIR measurements and the functional groups we can harvest from GC-MS and MCM. We have added the reference to it in Section 2.1 as follows: "The regions of absorption in the IR spectrum associated with FGs patterns are reported in Table 3 as an additional reference"

6. **Comment** It is recommended that the authors list the groups of patterns in the same order throughout the manuscript. For example, in the abstract and the introduction, SIMPOL.1 appears as the first group, but in section 2.1 it is the second group.

   Response We have changed the text in order to keep the consistency as suggested.

7. **Comment** p. 33637, lines 15-19: Awkward. It is recommended that the authors reword this description of the test compounds. If I understand, it seems sufficient to say that compounds were selected either from and then give the list (without repeating either or selected).

   Response The phrase has been reworded as suggested and now they keep the same consistency as in the abstract, that has been rephrased as:

   "1) a complete set of functional groups that can entirely describe the molecules comprised in the α-pinene and 1,3,5-trimethylbenzene MCMv3.2 oxidation

schemes, 2) FGs that are measurable by Fourier transform infrared spectroscopy (FTIR), 3) groups incorporated in the SIMPOL.1 vapor pressure estimation model, and 4) bonds necessary for the calculation of carbon oxidation state."

8. **Comment** p. 33639, line 10: Replace *is* with *are* (. . .but are shown. . .).

   Response The text has been modified as suggested: "Matched FTIR FGs in Table 1 (substructures 33-57) are also identical to the true number of FGs in the set of compounds used for evaluation (Table B2), but are not shown as each group except alkane CH is matched at most once and a similar plot is uninformative"

9. **Comment** p. 33642, lines 17-19: The discussion of the contribution of peroxacyl nitrate mass vs. O:C is confusing as written. It is suggested that the authors reword these sentences to more clearly articulate that while the total mass contribution is only 26%, the total O:C contribution is 53% due to the high O:C ratio of these functional groups (or the dominant compound).

   Response The phrase has been rewritten as suggested: "A singular peroxyacyl nitrate compound (peroxyacetyl nitrate) accounts for 26% of the total gas phase mass. The peroxyacyl nitrate functional group furthermore accounts for the greatest fraction of the total O:C ratio after 20 hours of simulation (53% of the total O:C), as it contains five oxygen atoms per FG."

10. **Comment** Fig. 6 caption: Was the OA fraction in gray truly unmeasured? Or was it unresolvable using the technique specified?

    Response The distinction between "measurable" and "resolvable" in this context is not entirely clear, but the gray fraction is the estimated fraction of OA that was not quantified by both techniques. It is possible that with additional derivatization (Dron et al., 2010), for instance, that the GC-MS can resolve or measure a higher fraction of mass. We have changed our terminology to refer to this as the "unresolved" fraction. We have also changed Fig. 6 in order to let the gray area

represent the unresolved fraction by both FTIR and GC-MS, so now the gray area in the second column corresponds to 10% (mass fraction not resolved by FTIR). We have specified the type of biomass burning in the figure and changed the caption:

"Comparison of the FG distribution of the quantified fraction measured by GC-MS (a,b and c; Rogge et al., 1998; Rogge et al., 1993) and FTIR-PMF (d and e; Hawkins and Russell, 2010) in aerosol emitted by biomass burning (a and d) and vehicle emission (b,c and e) sources. The gray area is the unresolved OA fraction by the two different analytical techniques used (around 80% for GC-MS and around 10% for FTIR). The type of biomass burning is specified in the pie charts a and d."

References

Cubison, M. J., Ortega, A. M., Hayes, P. L., Farmer, D. K., Day, D., Lechner, M. J., Brune, W. H., Apel, E., Diskin, G. S., Fisher, J. A., Fuelberg, H. E., Hecobian, A., Knapp, D. J., Mikoviny, T., Riemer, D., Sachse, G. W., Sessions, W., Weber, R. J., Weinheimer, A. J., Wisthaler, A., and Jimenez, J. L.: Effects of aging on organic aerosol from open biomass burning smoke in aircraft and laboratory studies, Atmospheric Chemistry and Physics, 11, 12 049–12 064, doi:10.5194/acp-11-12049-2011, 2011.

Donahue, N. M., Robinson, A. L., Stanier, C. O., and Pandis, S. N.: Coupled partitioning, dilution, and chemical aging of semivolatile organics, Environmental Science & Technology, 40, 2635–2643, doi:10.1021/es052297c, 2006.

Donahue, N. M., Henry, K. M., Mentel, T. F., Kiendler-Scharr, A., Spindler, C., Bohn, B., Brauers, T., Dorn, H. P., Fuchs, H., Tillmann, R., Wahner, A., Saathoff, H., Naumann, K.-H., Moehler, O., Leisner, T., Mueller, L., Reinnig, M.-C., Hoffmann, T., Salo, K., Hallquist, M., Frosch, M., Bilde, M., Tritscher, T., Barmet, P., Praplan, A. P., DeCarlo, P. F., Dommen, J., Prevot, A. S. H., and Baltensperger, U.: Aging of biogenic secondary organic aerosol via gas-phase OH radical reactions, Proceedings of the National Academy of Sciences of the United States of America, 109, 13 503–13 508, doi:10.1073/pnas.1115186109, 2012.

Dron, J., El Haddad, I., Temime-Roussel, B., Jaffrezo, J.-L., Wortham, H., and Marchand, N.: Functional group composition of ambient and source organic aerosols determined by tandem mass spectrometry, Atmospheric Chemistry and Physics, 10, 7041–7055, doi:10.5194/acp-10-7041-2010, 2010.

Ehn, M., Thornton, J. A., Kleist, E., Sipilä, M., Junninen, H., Pullinen, I., Springer, M., Rubach, F., Tillmann, R., Lee, B., Lopez-Hilfiker, F., Andres, S., Acir, I.-H., Rissanen, M., Jokinen, T., Schobesberger, S., Kangasluoma, J., Kontkanen, J., Nieminen, T., Kurtén, T., Nielsen, L. B., Jø gensen, S., Kjaergaard, H. G., Canagaratna, M., Maso, M. D., Berndt, T., Petäjä, T., Wahner, A., Kerminen, V.-M., Kulmala, M., Worsnop, D. R., Wildt, J., and Mentel, T. F.: A large source of low-volatility secondary organic aerosol, Nature, 506, 476–479, doi:10.1038/nature13032, 2014.

Hawkins, L. N. and Russell, L. M.: Oxidation of ketone groups in transported biomass burning aerosol from the 2008 Northern California Lightning Series fires, Atmospheric Environment, 44, 4142–4154, doi:10.1016/j.atmosenv.2010.07.036, 2010.

Hennigan, C. J., Sullivan, A. P., Collett, Jeffrey L., J., and Robinson, A. L.: Levoglucosan stability in biomass burning particles exposed to hydroxyl radicals, Geophysical Research Letters, 37, L09 806, doi:10.1029/2010GL043088, 2010.

Lai, C., Liu, Y., Ma, J., Ma, Q., and He, H.: Degradation kinetics of levoglucosan initiated by hydroxyl radical under different environmental conditions, Atmospheric Environment, 91, 32–39, 2014.

Leithead, A., Li, S.-M., Hoff, R., Cheng, Y., and Brook, J.: Levoglucosan and dehydroabietic acid: Evidence of biomass burning impact on aerosols in the Lower Fraser Valley, Atmospheric Environment, 40, 2721–2734, doi:10.1016/j.atmosenv.2005.09.084, 2006.

Liu, S., Takahama, S., Russell, L. M., Gilardoni, S., and Baumgardner, D.: Oxygenated organic functional groups and their sources in single and submicron organic particles in MILAGRO 2006 campaign, Atmospheric Chemistry and Physics, 9, 6849–6863, doi:10.5194/acp-9-6849-2009, 2009.

Radzi bin Abas, M., Oros, D. R., and Simoneit, B. R. T.: Biomass burning as the main source of organic aerosol particulate matter in Malaysia during haze episodes., Chemosphere, 55, 1089–95, doi:10.1016/j.chemosphere.2004.02.002, 2004.

Rogge, W. F., Hildemann, L. M., Mazurek, M. A., Cass, G. R., and Simoneit, B. R. T.: Sources of Fine Organic Aerosol .2. Noncatalyst and Catalyst-equipped Automobiles and Heavy-duty Diesel Trucks, Environmental Science & Technology, 27, 636–651, doi:10.1021/es00041a007,

1993.

Rogge, W. F., Hildemann, L. M., Mazurek, M. A., Cass, G. R., and Simoneit, B. R. T.: Sources of fine organic aerosol. 9. Pine, oak and synthetic log combustion in residential fireplaces, Environmental Science & Technology, 32, 13–22, doi:10.1021/es960930b, 1998.

Russell, L. M., Bahadur, R., Hawkins, L. N., Allan, J., Baumgardner, D., Quinn, P. K., and Bates, T. S.: Organic aerosol characterization by complementary measurements of chemical bonds and molecular fragments, Atmospheric Environment, 43, 6100–6105, doi:10.1016/j.atmosenv.2009.09.036, 2009.

Shiraiwa, M., Berkemeier, T., Schilling-Fahnestock, K. A., Seinfeld, J. H., and Pöschl, U.: Molecular corridors and kinetic regimes in the multiphase chemical evolution of secondary organic aerosol, Atmospheric Chemistry and Physics, 14, 8323–8341, doi:10.5194/acp-14-8323-2014, 2014.

Simoneit, B. R. T.: A review of biomarker compounds as source indicators and tracers for air pollution, Environmental Science and Pollution Research, 6, 159–169, doi:10.1007/BF02987621, 1999.

Please also note the supplement to this comment:

http://www.atmos-chem-phys-discuss.net/15/C12885/2016/acpd-15-C12885-2016-supplement.pdf

---

Interactive comment on Atmos. Chem. Phys. Discuss., 15, 33631, 2015.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper