

Response to Reviewer #2

The present paper is a technical note to already-published model-measurement comparison SOA studies (Ruggeri et al., ACP 2016). The general scientific objectives are already explained in the previous publications. Specifically, the Authors aim to exploit organic functional group distributions to constrain explicit-chemistry models of secondary organic aerosol (SOA) formation. The idea, although not new, is sound, because functional groups provide direct information about organic reactivity (like in the Carbon Bond Mechanism developed long ago for gas-phase reactions) and keep track of the chemical mechanisms that govern the enrichment of SVOC in the particulate phase. This technical note, in particular, focuses on the derivation of carbon-based metrics (such as oxidation state, and O/C ratios) from measurable functional groups distributions, with the aim to support and inform the comparison between FG-based techniques (such as FTIR) and more established mass spectrometric methods (AMS). The methodology is discussed in detail, and for the first time O/C ratios from FTIR measurements are reported taking into account the possible biases due to the selectivity of FTIR spectroscopy for specific FGs.

We thank the reviewer for this lucid assessment.

Specific comments:

1. Simple examples, like the one shown in Figure 1, are essential for a chemist who is not familiar with linear algebra. I invite the Authors to comment such examples in the text, or in the Supplementary Information. For instance, it is not straightforward why negative values for ϕ can be obtained. This is important also to understand why values for λ lower than $1/3$ (the theoretical value for a tri-substituted carbon atom) are found in Table 2.

As the author correctly notes, lambda values should nominally take on values rational values of $\{1/4, 1/3, 1/2, 1\}$ at the level of individual carbon atoms. However, the coefficients reported represent a single set of values to be used across all carbon types, which leads to irrational values. A value of lambda below $1/3$ results from the empirical nature of regression methods, in which obtained coefficient values are insensitive to under-represented FGs (therefore could be replaced with $1/3$ with little impact on results), or eliminated (set to zero) in the case of redundant FGs (i.e. strongly correlated to another FG). We have included the following statements in Section 2.4 to address potential questions from other readers:

“Each of the solutions produces a series of irrational numbers (due to the multiplicitous configurations of FGs on carbon atoms) that may be overly precise for the data set used for estimation.”

and

“Direct fitting methods, on the other hand, may lead to insignificant coefficients from under-represented or redundant FGs [...].”

While lambda may have a physical interpretation in simple instances (not in complex mixtures with multiplicitous configurations of functional groups as noted above), negative

values for ϕ are permitted in that they are only required to satisfy the carbon type balance (equation 4). We have added the following statement to Section 2.4 to better connect readers with the illustration in Figure 1:

“The elements of Φ satisfy the carbon type balance (equation 4) but are not required to be non-negative, but their summation across rows (equation 6) yields values for λ_C that corresponds to the number of carbon atoms per FG associated with them.

[...]

“ λ_C may also not correspond to a physically interpretable quantity in such instances, as a single set of coefficients are insufficient to estimate the exact abundances of carbon atoms under these circumstances.”

2. The three methods for carbon abundance estimation discussed in Section 2.4 are compared for the examples of SOA formation considered in this study (Table 2, Figures 8, 9), but it is difficult to derive general conclusions on their applicability. How much details on the qualitative composition (in carbon types) must be known a priori? It would be important to expand the paragraph about the type and nature of the datasets which the three methods rely on.

It is yet difficult to recommend a best method for estimation in a general sense, but we hope that this exploratory study can initiate further inquiry into this topic. However, as noted by the reviewer, it is worth considering the nature of underlying data sets in more detail to describe their tradeoffs. We had previously written in Section 2.4:

“Numerical details aside, the main differences among the three are the data sets used for estimation. COUNT uses information from Θ only (defined for the FGs in the APIN mechanism), COMPOUND uses carbon type abundances in compounds (limited to SVOCs in the APIN mechanism), and MIXTURE uses mixture information of the condensed-phase (from different periods in the APIN simulation).”

However, the important difference which may have not been emphasized was the weighting of the estimates. We have added to Section 2.4 the following explanation:

“The resulting differences in estimates of $\hat{\lambda}_C$ are largely due to weighting of FGs associated with each carbon type: each type receiving equal weight (COUNT), by frequency of occurrence in SVOCs (COMPOUND), and by abundance in SOA formed in the APIN simulation (MIXTURE). While the COUNT method is physically significant at the level of individual carbon atoms, the representativeness of estimated values for use in mixtures can vary according to composition. Direct fitting methods, on the other hand, may suffer from errant coefficients from redundant or under-represented FGs, or be overly specific such that they cannot be generalized to other systems. Therefore, the results from all three methods are evaluated to explore the range of plausible values.”

Regarding the COUNT method, we have also added the following statement to aid the physical significance of this estimate:

“The main premise of this approach is to apportion fractional units of carbon to each measured FG such that their sum equals unity.”

3. The paper makes use of the notion of carbon types. These are exemplified for simple molecules in Figure 1, and are otherwise listed as numbers in the other Tables and Figures. I suggest to explicit the full list in the Supplementary Information. It will be important to

understand how many carbon types contain heteroatoms in functional groups (alcohols, carboxylic acids, nitro groups etc.), which seems to be the focus of the paper, instead of being included in the skeleton of the molecules (ethers, esters, etc.).

We thank the reviewer for this comment. We had discussed the case of anhydrides, esters, and organic peroxides in the supporting information but had not referred to it in the main text (ethers are currently out of the scope of our analysis as they are not included in mechanisms which we have studied). We have modified Section 2.3 to include the following statements:

“All elements in equation 3 can be known precisely for any set of molecules \mathcal{M} from the chemometric patterns and atom-level validation described by Ruggeri and Takahama (2016), and is summarized in Section S1. Furthermore, the FGs included in the APIN system are all those which are defined by association only to single carbon atoms (e.g., alcohol, carboxylic, methylene groups). Methods for extending this analysis to FGs containing multiple carbon atoms (e.g., anhydride, ester, and organic peroxide groups) are described in Section S2.”

While the APIN system does not include these multi-carbon functional groups, we have included additional analyses (Supporting Information, Section S2, and Tables S1–S4, and online code repository) to conclude that at least in terms of frequency of occurrence in α -pinene and 1,3,5-trimethylbenzene degradation systems studied by Ruggeri et al. (2016), our 41 carbon types presented in the main text encompasses 92% of the 2867 carbon atoms in the 441 molecules. The four tables can be viewed in the Supporting Information, but the text is copied below:

“Tables S1–S3 show carbon atom types associated with single-carbon FGs (conversely stated, each FG is uniquely associated with one carbon atom), two-carbon FGs (carbon atoms in these FGs share some heteroatoms with other carbon atoms), and carbon-only structures present in the combined set of molecules from the α -pinene and 1,3,5-trimethylbenzene degradation schemes. In this set of 441 molecules, there are 2867 carbon atoms that can be classified into one of 60 types (labeled in order of frequency, X1–X60, prefixed by character ‘X’ to prevent confusion with carbon type labels used in the APIN simulation) that differ in their association with 30 unique FGs. 46 of these types contain unique FGs (2557 / 2867 carbon atoms belong in this category), 11 of these types share FGs (116 / 2867 carbon atoms belong in this category), and 3 are bonded only to other carbon atoms (194 / 2867 carbon atoms belong in this category). 92% of the carbon atoms in this superset belong to the 41 carbon types (which includes two of the tertiary and quaternary carbon types) from the APIN simulation discussed in the main body of this manuscript, though this relative abundance is reported on a frequency basis and does not consider molecular abundances that might be typical in a SOA mixture. The correspondence of labels used in the main document (numbered by abundance of total carbon during the APIN simulation) and Tables S1–S3 (numbered by frequency of occurrence of in the 441 molecules) are listed in Table S4.”

The rest of Section S2 is dedicated to describing adjustments necessary to include carbon atoms associated with anhydrides, esters, and peroxides for estimation of carbon-centered metrics, and can be applied to any functional groups with which more than one carbon atom would be associated. The text has been modified to indicate that our framework is quite general in this sense (Section S2):

“The second generalization concerns FGs that contain skeletal heteroatoms. FGs of this

type — specifically in this case, anhydride, ester, and (organic) peroxide — are present in photooxidation products of 1,3,5-trimethylbenzene included in the MCMv3.2 mechanism (Bloss et al., 2005; Ruggeri et al., 2016), and corresponding SMARTS patterns were developed by Ruggeri and Takahama (2016) to match these structures. Equation S3 should accordingly permit two carbon atoms to be associated with each of these exceptional FGs. To accommodate such groups (and other FGs defined by membership of multiple carbon atoms) in our framework, the carbon type formulation can be a) extended to “carbon units” consisting of one or more carbon atoms and their bonded heteroatoms, or b) modified by the introduction of a correction factor.”