

Supplement of Atmos. Chem. Phys., 17, 4433–4450, 2017
<http://www.atmos-chem-phys.net/17/4433/2017/>
doi:10.5194/acp-17-4433-2017-supplement
© Author(s) 2017. CC Attribution 3.0 License.



Supplement of

Technical note: Relating functional group measurements to carbon types for improved model–measurement comparisons of organic aerosol composition

Satoshi Takahama and Giulia Ruggeri

Correspondence to: Satoshi Takahama (satoshi.takahama@epfl.ch)

The copyright of individual parts of the supplement might differ from the CC-BY 3.0 licence.

S1 Chemoinformatic tools

To construct \mathbf{Y} , Θ , and \mathbf{X} (Section 2.2), we use the APRL-SSP program with a minor modification. Ruggeri and Takahama (2016) showed that there can be a one-to-one correspondence between functional groups (FGs) and non-carbon atoms such that the mapping is unambiguous. Past constraints specified that all atoms must be accounted for by at least one FG:

$$5 \quad \bigcup_{j \in \mathcal{J}} \{a : a \in \mathcal{A}_i, a \in \mathcal{A}_j\} = \mathcal{A}_i \quad \forall i \in \mathcal{M}, \quad (\text{S1})$$

and that all non-carbon atoms $\mathcal{A}_i \setminus \mathcal{C}_i$ cannot be matched by more than one group:

$$\bigcap_{j \in \mathcal{J}} \{a : a \in \mathcal{A}_i \setminus \mathcal{C}_i, a \in \mathcal{A}_j\} = \emptyset \quad \forall i \in \mathcal{M}. \quad (\text{S2})$$

Polyfunctional carbon atoms were not included in their validation at the time, but we now impose an additional condition that each FG includes the definition for only one carbon atom (with exceptions noted in Section S2):

$$10 \quad \left| \{a : a \in \mathcal{C}_i, a \in \mathcal{A}_j\} \right| = 1 \quad \forall i \in \mathcal{M}, j \in \mathcal{J}. \quad (\text{S3})$$

$|\cdot|$ denotes the cardinality of the set. To satisfy this new condition, we split the $\text{C}=\text{O}-\text{O}$ group into three separate groups ($\text{R}_2\text{C}=\text{O}-\text{O}$, $\text{RHC}=\text{O}-\text{O}$, $\text{H}_2\text{C}=\text{O}-\text{O}$) as carbon was double counted otherwise. This step is inconsequential from a mass perspective, but important for fulfilling the relationship (equation 3) for the complete APIN MCM mechanism. The corresponding patterns have been updated in the APRL-SSP repository.

15 S2 Generalization of carbon types

We note two generalizations to the carbon type descriptions introduced in Section 2.2 that can be considered. First, these carbon types focus on the functionality of each carbon, but do not consider its complete bonding environment (e.g., configuration to other carbon atoms). For instance, carbon atoms defined by functionalization only by hydroperoxide, alkoxy radicals, and peroxy radicals can differ according to whether the carbon is sp^3 or sp^2 -bonded to other carbon atoms. Hydroxyl groups in phenols are differentiated from alcohols in similar instances, but we have not made this distinction for these three groups as nomenclature for them are not common and also does not affect our analysis. It is possible to define SMARTS patterns to make the differentiation in other applications where carbon type representations are useful.

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. In the latter approach, the carbon-group matrix θ_{kj} can be replaced by $\tilde{\theta}_{kj} = \theta_{kj}\gamma_j$ and group-carbon matrix ϕ_{jk} replaced by $\tilde{\phi}_{kj} = \phi_{kj}\gamma_j^{-1}$, where

γ is a coefficient is a correction factor to complete the FG and carbon balances of equations 3 and 4, respectively. $\gamma_j = 0.5$ for these two-carbon FGs, and $\gamma_j = 1$ for the rest (single-carbon FGs). All equalities expressed in this manuscript would hold exactly, except for carbon type oxidation state (equation 8) that will be only approximately true for ester groups (since one carbon atom is double-bonded to oxygen while the other is only singly bonded to another oxygen atom). However, the overall oxidation state estimate (equation 7) still holds when summed over each molecule that contains both carbon atoms of the ester group.

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.

20 S3 Supporting interpretations for chemical basis sets

When combined with information regarding the carbon skeleton, carbon types presented in this work can provide another origin for derivation of chemical basis sets. In Figure S1, molecular abundances for gas and aerosol phases in the APIN simulation are depicted using carbon types and n_C . Together with Figure 2 and definitions in Section 2.2, each of the common basis set dimensions (O, C, H, OS_C) used in the aerosol community can be derived. When neighboring interactions among groups are desired, these carbon types can form the basis of multi-carbon unit representations as hinted above.

References

- Bloss, C., Wagner, V., Jenkin, M. E., Volkamer, R., Bloss, W. J., Lee, J. D., Heard, D. E., Wirtz, K., Martin-Reviejo, M., Rea, G., Wenger, J. C., and Pilling, M. J.: Development of a detailed chemical mechanism (MCMv3.1) for the atmospheric oxidation of aromatic hydrocarbons, *Atmospheric Chemistry and Physics*, 5, 641–664, doi:10.5194/acp-5-641-2005, 2005.
- 30 Ruggeri, G. and Takahama, S.: Technical Note: Development of chemoinformatic tools to enumerate functional groups in molecules for organic aerosol characterization, *Atmospheric Chemistry and Physics*, 16, 4401–4422, doi:10.5194/acp-16-4401-2016, 2016.
- Ruggeri, G., Bernhard, F. A., Henderson, B. H., and Takahama, S.: Model-measurement comparison of functional group abundance in α -pinene and 1,3,5-trimethylbenzene secondary organic aerosol formation, *Atmospheric Chemistry and Physics*, 16, 8729–8747, doi:10.5194/acp-16-8729-2016, 2016.

Tables and Figures

Table S1: Functionalized carbon atoms that do not contain multi-carbon FGs (anhydride, ester, and peroxide). Rows are ordered according to OS_C (ascending) and number of occurrences n (out of 2867) (descending). z indicates the FG contribution to oxidation state (equation 7).

	$z \rightarrow$		-1	0	1										2	3									
			aCH	eCH	rCH	formaldehyde	H2C=O-O	aCOH	aldehyde	alkoxyl (*)	CONO2	hydroperoxide	nitro	peroxy nitrate	peroxyl (*)	phenol	RHC=O-O	formic acid	ketone	R2C=O-O	carbonyl peroxy acid	carbonyl peroxy acid (*)	carboxyl (*)	COOH	peroxyacyl nitrate
X58	-4	1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X1	-3	720	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X3	-2	366	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X49	-2	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X53	-2	1	3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
X54	-2	1	3	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
X55	-2	1	3	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X56	-2	1	3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X57	-2	1	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
X60	-2	1	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X4	-1	208	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X8	-1	78	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X10	-1	48	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X11	-1	48	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X28	-1	18	2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
X29	-1	18	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X30	-1	18	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
X35	-1	9	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X7	0	96	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X23	0	26	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X24	0	22	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X25	0	21	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
X26	0	21	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
X44	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X48	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X5	1	159	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X15	1	30	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X16	1	29	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X20	1	27	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X21	1	26	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
X22	1	26	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
X36	1	6	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
X42	1	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
X46	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
X2	2	368	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
X38	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
X39	2	3	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0

Table S1: Functionalized carbon atoms that do not contain multi-carbon FGs (anhydride, ester, and peroxide). Rows are ordered according to OS_C (ascending) and number of occurrences n (out of 2867) (descending). z indicates the FG contribution to oxidation state (equation 7).

		$z \rightarrow$	-1	0	1										2	3									
Label	OS_C	n	aCH	eCH	rCH	formaldehyde	H2C=O-O	aCOH	aldehyde	alkoxyl (*)	CONO2	hydroperoxide	nitro	peroxy nitrate	peroxyl (*)	phenol	RHC=O-O	formic acid	ketone	R2C=O-O	carbonyl peroxy acid	carbonyl peroxy acid (*)	carboxyl (*)	COOH	peroxyacyl nitrate
X40	2	3	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X41	2	3	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
X45	2	2	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X47	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
X9	3	54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
X17	3	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
X18	3	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
X19	3	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
X43	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

Table S2. Functionalized carbon atoms that contain multi-carbon FGs (anhydride, ester, and peroxide). Format follows that of Table S1.

		$z \rightarrow$	-1	1						2	4	6
Label	OS_C	n	aCH	aCOH	alkoxyl (*)	hydroperoxide	nitro	peroxyl (*)	ether	peroxide	ester	anhydride
X33	1	15	1	0	0	0	0	0	1	0	0	0
X14	2	31	0	0	0	0	0	0	0	1	0	0
X32	2	15	0	0	0	0	0	0	1	0	0	0
X34	3*	13	1	0	0	0	0	0	0	0	1	0
X37	3	3	0	0	0	0	1	0	0	1	0	0
X31	4*	17	0	0	0	0	0	0	0	0	1	0
X50	4*	1	1	0	0	0	0	1	0	0	1	0
X51	4*	1	1	0	1	0	0	0	0	0	0	1
X52	4*	1	1	0	0	1	0	0	0	0	1	0
X59	4*	1	1	1	0	0	0	0	0	0	1	0
X27	6	18	0	0	0	0	0	0	0	0	0	1

*Value is approximate since the current SMARTS pattern for ester groups defines membership of carbon atoms to the entire $-CO_2C-$ substructure, which does not differentiate between the carbon functionalized by ester carbonyl and the other which is not.

Table S3. Carbon atoms not functionalized by any heteroatoms (i.e., only bonded to other carbon atoms). Format follows that of Table S1.

			$z \rightarrow$		
			0		
			aromatic sp ² carbon	quaternary carbon	tertiary sp ² carbon
Label	OS _C	<i>n</i>			
X6	0	100	0	1	0
X12	0	47	1	0	0
X13	0	47	0	0	1

Table S4. Correspondence of carbon type labels used in the main document (molecules from APIN simulation, Figure 2) and Tables S1–S3 (molecules combined from α -pinene and 1,3,5-trimethylbenzene degradation schemes). Numbers indicate their relative abundance (in descending order) in APIN simulation (first column) and in combined set of molecules (right column). The character “X” is prepended to labels in the combined set only to prevent confusion with labels used for the APIN simulation.

Carbon type (α -pinene set)	Label (combined set)
1	X1
2	X3
3	X4
4	X2
5	X6
6	X22
7	X8
8	X7
9	X9
10	X17
11	X5
12	X15
13	X23
14	X16
15	X30
16	X26
17	X19
18	X35
19	X11
20	X13
21	X21
22	X47
23	X25
24	X57
25	X18
26	X28
27	X48
28	X60
29	X56
30	X58
31	X49
32	X53
33	X20
34	X54
35	X24
36	X29
37	X43
38	X36
39	X55
40	X44
41	X38

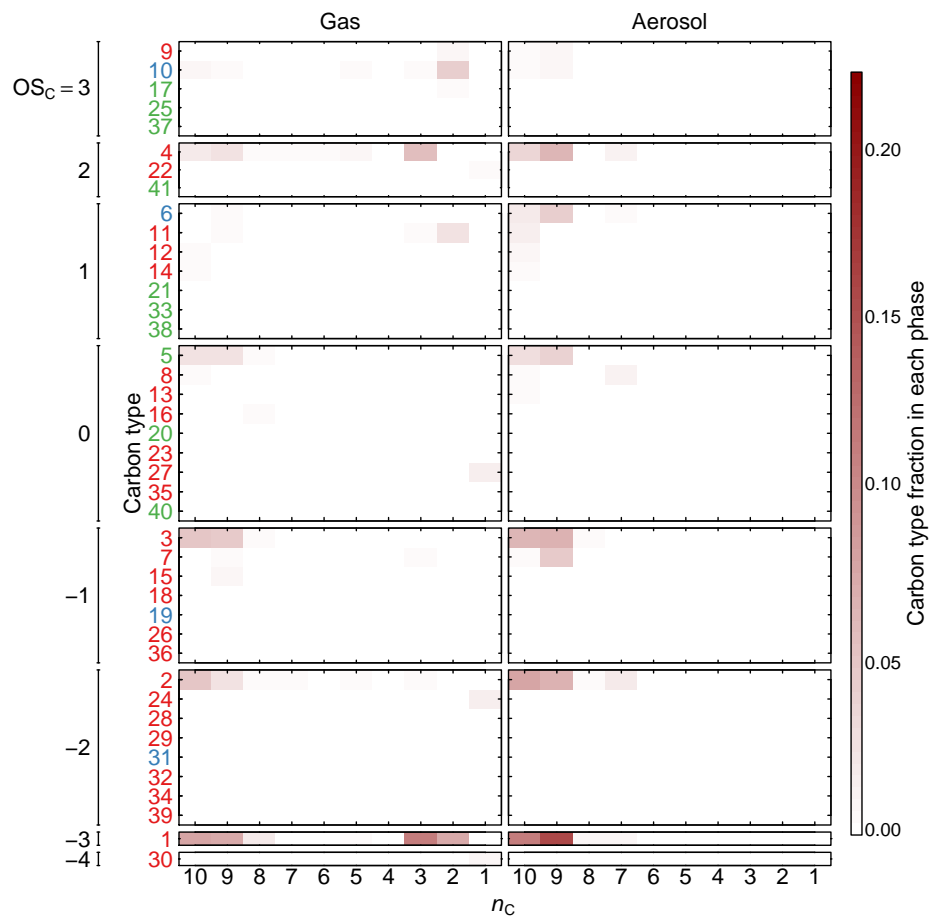


Figure S1. Molecular abundance at $t_{\max\text{SOA}}$ described in terms of their carbon types and number of carbon atoms. The carbon abundance in each grid cell is normalized by the total molar abundance of its phase (gas or aerosol). The colors for the carbon types are the same as in Figure 2.