Supplement of

Chemical Composition of Particulate Organics in October 2017 Northern California Wildfire Plumes

Yutong Liang, Coty N. Jen, Robert J. Weber, Pawel K. Misztal, Allen H. Goldstein

5

Supporting Information:

1. Procedure for compound identification and classification
2. Training the models for predicting saturation vapor pressures and average oxidation states
3. Supplementary figures and tables

10

1 Procedure for compound identification and classification

Compounds were first compared with the authentic standards in terms of mass spectra and retention times. For compounds not matched with standards, they were compared with mass spectral libraries mentioned in the main text. A forward match factor of 800 and a reverse match factor of 700 (999 is a perfect match) were used as the matching thresholds. This should result in an ~80% correct match (Worton et al., 2017). Since the GoAmazon, SOAS and FIREX libraries were created on the same instrument as this study, for compounds matched with those libraries, we further require the discrepancy in retention index to be within 10. If the compound was satisfactorily matched with an entry in the library, we confirmed its molecular ion in the VUV chromatogram. If the molecular ion was also confirmed, the compound was positively identified. Sugars and sugar derivatives still fragment substantially under 10.5 eV ionization, and the VUV confirmation step was skipped for most of those compounds.

To check the molecular ion, first, the possible molecular ion "M" (with the largest m/z in the VUV mass spectrum) was identified. For compounds with multiple -OH or -COOH groups, fragmentation could make the molecular ion small than 10% of the base ion even under VUV ionization. Therefore, the abundance of and m/z 73 and m/z 147 ions in the EI spectrum to find out whether the compound is derivatized. The m/z 73 ion comes from $[Si(CH_3)_3]^{\bullet+}$, a fragment of all TMS derivatized compounds. The m/z 147 ion $[(CH_3)_2Si=OSi(CH_3)_3]^{\bullet+}$ is a fragment of molecules containing at least 2 derivatized groups (Yu et al., 1998). For compounds with strong m/z 147 ion, we further looked for the abundance of "M" + 15 to see if the actual molecular ion was there. Many compounds after derivatization have molecular weight more than 200 g mol^{-1}. It is hard to calculate the molecular composition of those compounds directly because there could be more than one possible formula. Instead, once the actual molecular ion was determined, the exact number of derivatization groups were calculated from the abundance ratio of [M+2]/[M] because of the high natural abundance of ^{30}Si (3.1%). For example, for hydroquinone-2TMS, the mass abundance ratio [M+2]/[M] = 7.07%. However, for syringaldehyde-TMS, [M+2]/[M] = 4.60%, which suggests it only has one -TMS group. This step reduces the number of possible formulae. Using this approach, the molecular formulae of 57% of compounds were determined.

Chemical class of unidentified compounds were assigned by library match and the locations of the compounds in the second dimension. Alkanes, alcohols, and monocarboxylic acids all have unique second dimension retention times and were thus easily classified. The high abundance of m/z 147 was used as the criterion for (other) oxygenated species. Compounds having high abundance of m/z 204 $(Si_2C_8H_{20}O_2)^{\bullet+}$ and/or m/z 217 $(Si_2C_9H_{21}O_2)^{\bullet+}$ were assigned as sugars (Fabbri et al., 2002). Nitrogen-containing compounds should follow the nitrogen rule (i.e., uncharged molecules with an odd nominal mass also have an odd number of nitrogen atoms). Other compounds were classified by checking the top matches in the mass spectral libraries. Out of the 572 compounds, only 45 remained unclassified. If a compound contains nitrogen and other functional groups, it was classified as nitrogen-containing compound group. Aromatic acids, aromatic alcohols and oxygenated aromatic

compounds were grouped into the aromatic class. Sugar-like compounds, if partially derivatized, could be misclassified into the oxygenated compound group. However, due to the complexity of the data, we only corrected for the partial derivatization products of levoglucosan.

2. Training the models for predicting saturation vapor pressures (v_P) and average carbon oxidation states $\overline{OS_c}$

GC×GC measurement gives the retention times (indices), and mass spectra of compounds. With the classification method described, the functionality (class) of most compounds could be found. The mass spectra of compounds are highly related to the functionality. As described above, the molecular formulae of only 57% of compounds could be determined. Therefore, the d-alkane first-dimension retention index (numerical), second dimension retention time (numerical) and functionality (categorical) were selected as the input parameters. We trained the saturation vapor pressure model and the average carbon oxidation state model using the parameters of standard compounds as inputs, and their $\log_{10}(v_P)$ in the EPI Suite and EVAPORATION Model (for dicarboxylic acids) and $\overline{OS_c}$ as outputs, respectively. For 64% of the compounds, the $\log_{10}(v_P)$ value from experimental studies were used as input. Five-fold cross validation were used for the training. That means the data were randomly split into 5 groups with similar numbers of observations. The data in 4 groups were used to train the model and the data in the fifth group were used for evaluation. Then the training and evaluation groups were shuffled to make sure every group served as the evaluation group. The models giving the highest R^2 between the predicted response with the observed response were chosen. The models generated are included in the supplement .zip archive. Figure S2 shows the performance of the model in predicting the saturation vapor pressure $\log_{10}(v_P)$ and $\overline{OS_c}$ in the 5-fold cross validation. We checked the residuals of predictions for different classes of compounds Fig. S2b and found the volatility of sugars were systematically underpredicted. Therefore, we manually added 0.285 [the difference between actual $\log_{10}(v_P)$ of levoglucosan and the predicted $\log_{10}(v_P)$ of levoglucosan] to the $\log_{10}(v_P)$ predictions for sugars.
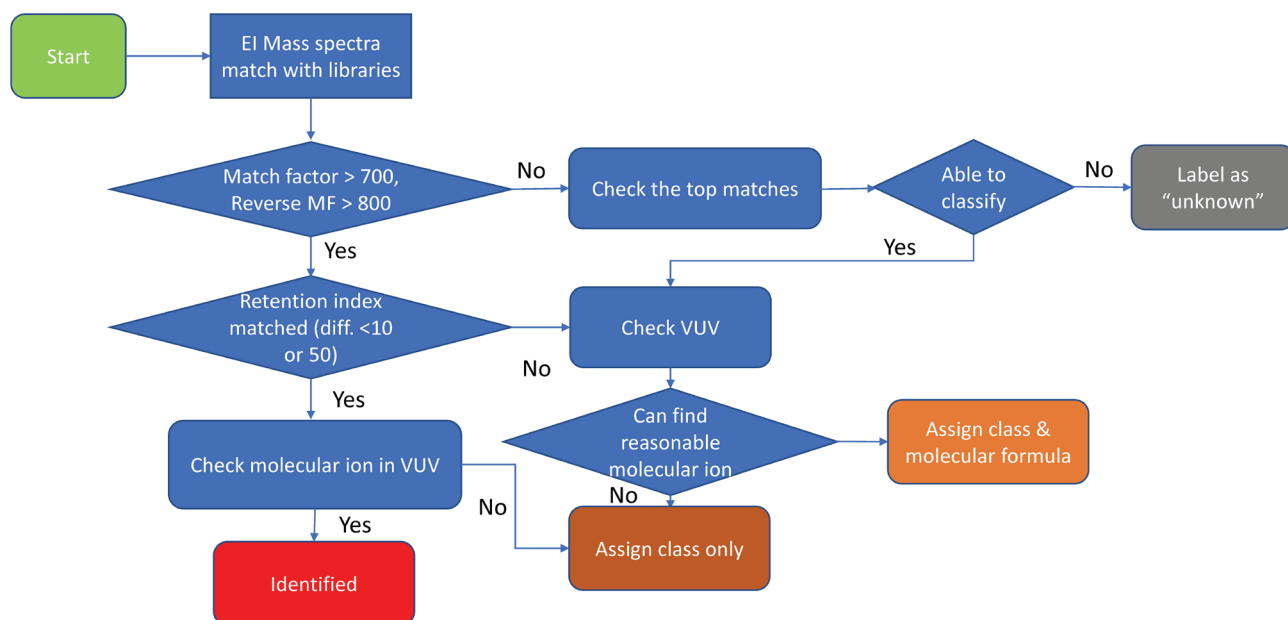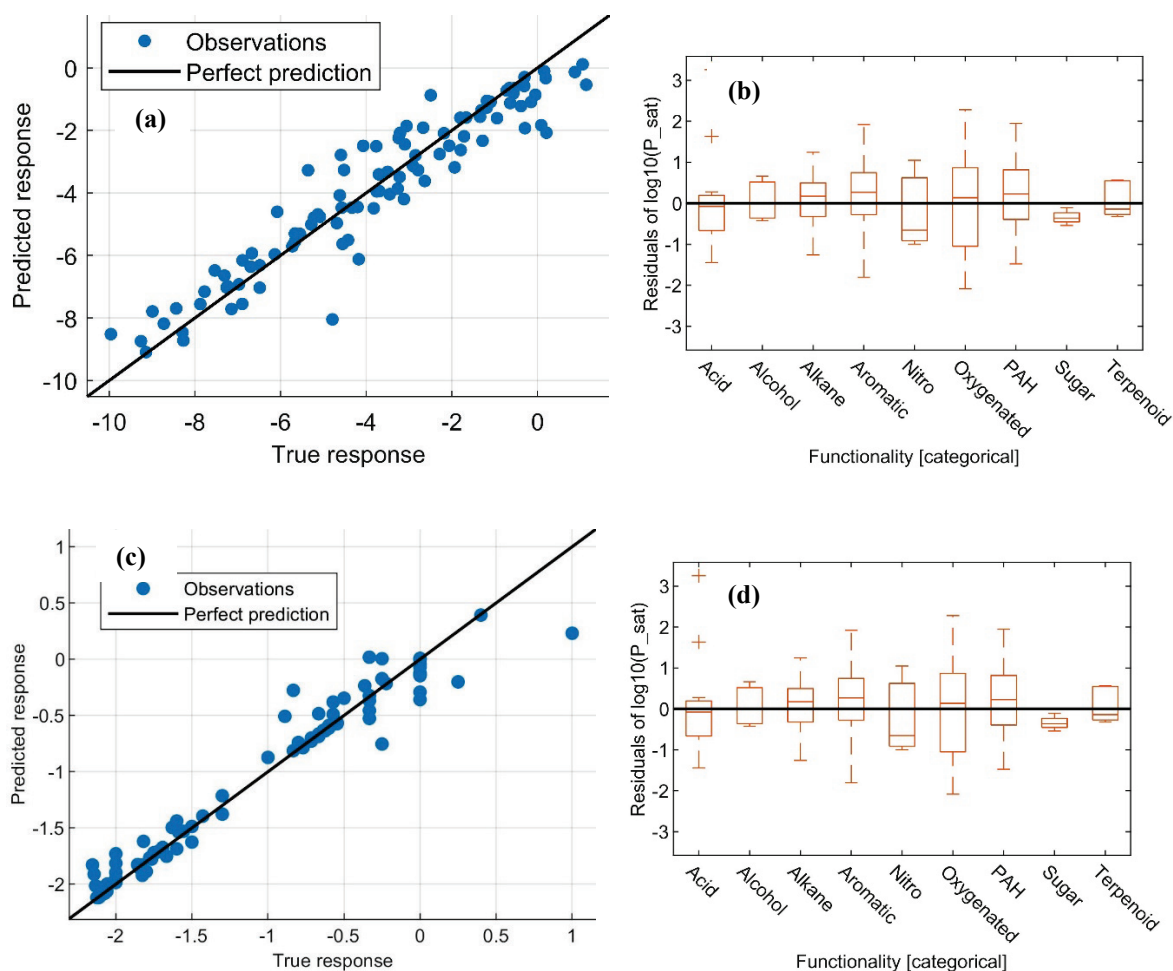
3. Supplementary Figures

Figure S1 Workflow for identifying compounds not matched with compounds in the standard and assigning chemical classes.

80   **Figure S2 (a)** Scatter plot of the predicted vapor pressure $\log_{10}(v_P)$ vs the true vapor pressure of the standard compounds. **(b)** Boxplot of the residuals of the $\log_{10}(v_P)$ for compounds in different classes. **(c)** Scatter plot of the predicted $\overline{OS_c}$ vs the true $\overline{OS_c}$ of the standard compounds. **(d)** Boxplot of the residuals of the $\overline{OS_c}$ for compounds in different classes.
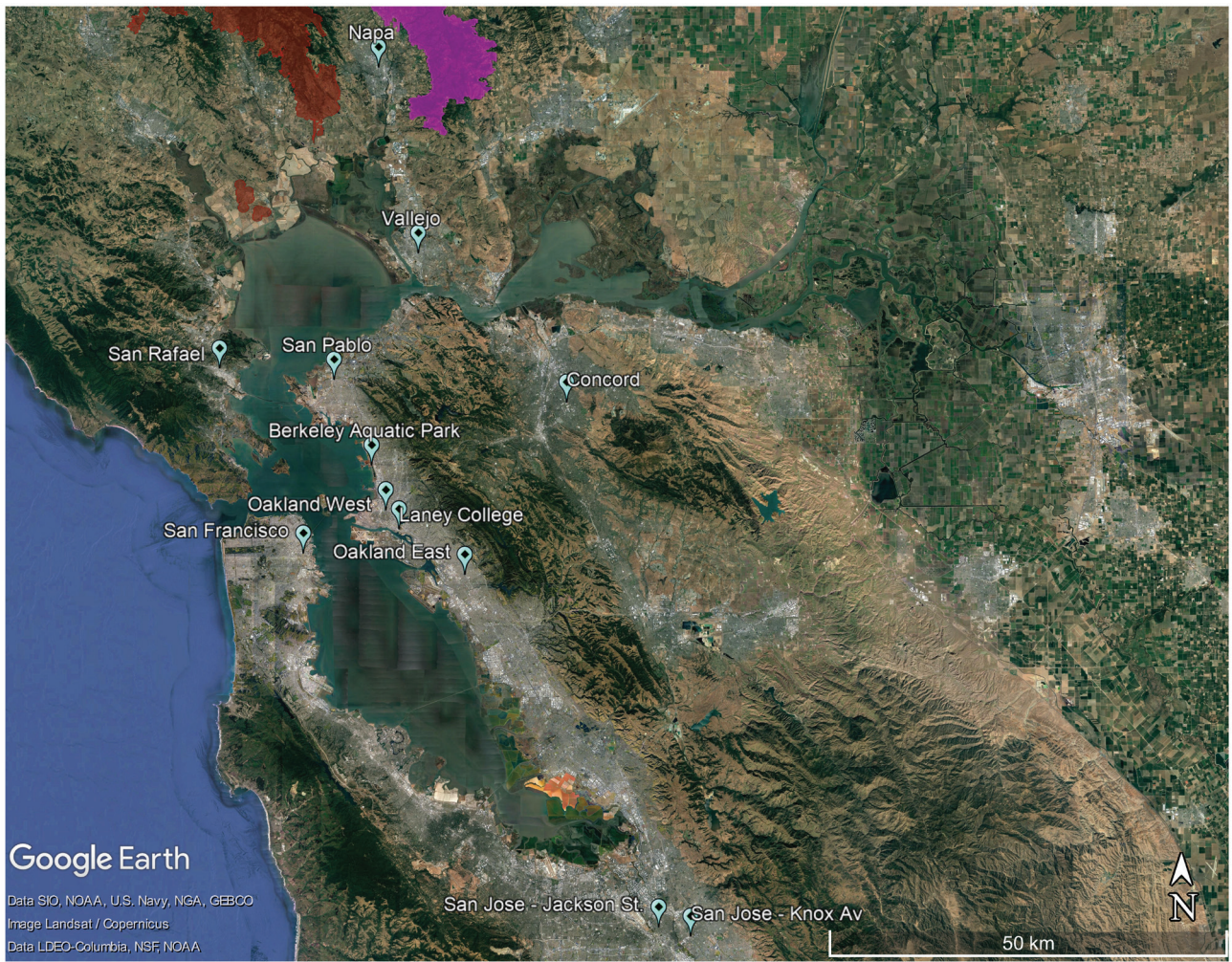
85

90

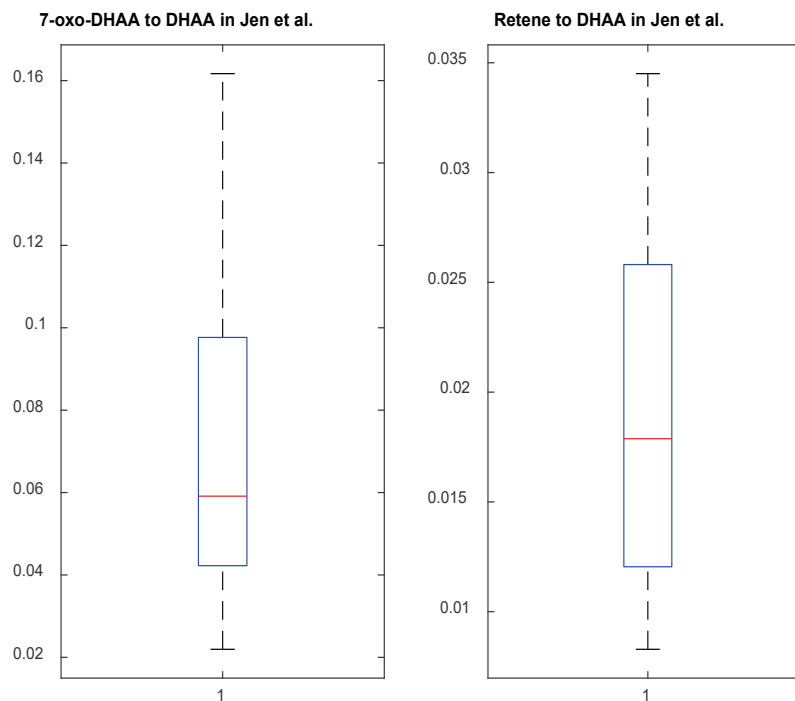**Figure S3**. Locations of BAAQMD monitoring stations in Figure 2. Map © Google Earth 2020

Figure S4 Oxo-DHAA to DHAA mass ratio and retene to DHAA mass ratio reported by Jen et al. (2019). The ratio was calculated from the emission factors of these compounds from fuels relevant to the Napa/Sonoma Fires. The top and bottom edges of the box indicate the 75th and 25th percentiles, respectively. The red central mark is the median. Whiskers extend to 1.5×interquartile ranges out of the boxes.
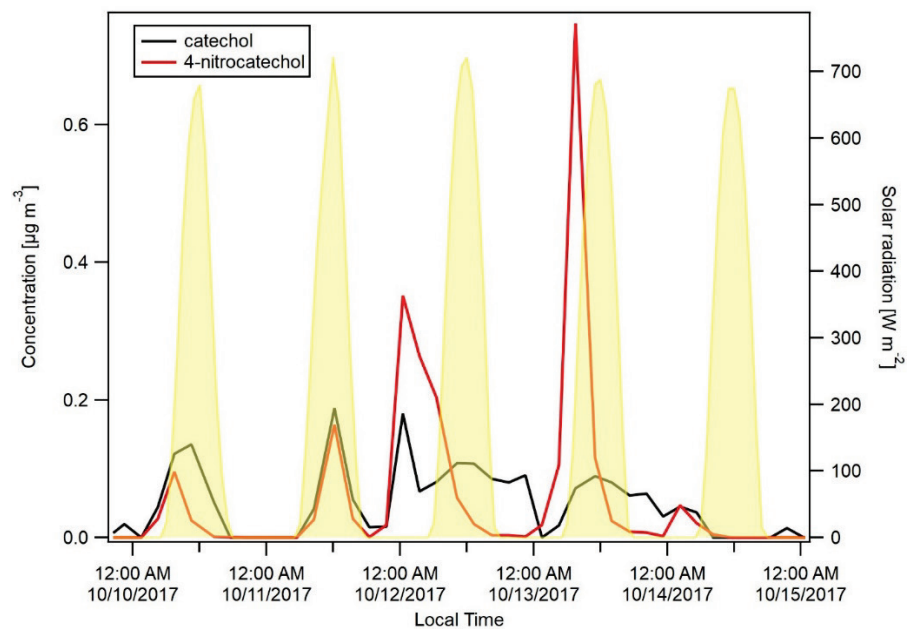
Figure S5 Particle phase catechol and 4-nitrocatechol concentration timelines. 4-Nitrocatechol to catechol ratio was lower in daytime biomass burning plumes. Solar radiation measured at Bethel Island is shown (in yellow, right axis).

Table S1 Positively identified BB-related compounds with highest abundances and PAHs observed at Berkeley.

Compound	CAS Number	Max Conc. [µg m⁻³]	GHS Hazard Class[1]
levoglucosan	644-76-8	3.84	H312, H315, H319, H335
dehydroabietic acid[2]	1740-19-8	0.84	H302
4-nitrocatechol	3316-09-4	0.75	H302, H305, H319, H335
4-methyl-5-nitrocatechol	68906-21-8	0.41	no data
tetracosanoic acid	557-59-5	0.37	H315, H319, H335
eicosanoic acid	506-30-9	0.35	H315, H319, H335
docosanoic acid[3]	112-85-6	0.33	not classified
galactosan	644-76-8	0.30	H312, H315, H319, H335
myo-inositol	87-89-8	0.28	H315, H319, H335
D-pinitol	484-68-4	0.26	no data
catechol	120-80-9	0.19	H301, H311, H315, H319, H341, H350
1,4-dihydroxy-2,6-dimethoxybenzene	15233-65-5	0.19	H315, H319, H335
3-methyl-5-nitrocatechol	5378-76-7	0.18	no data
hexacosanoic acid	506-46-7	0.17	H315, H319, H335
pyrogallol	87-66-1	0.15	H302, H312, H332, H341, H412
hexanedioic acid	124-04-9	0.15	H319
4-hydroxybenzoic acid	99-96-7	0.13	H315, H318, H319, H335
vanillic acid	121-34-6	0.11	H315, H319, H335
PAHs			
Compound	**CAS Number**	**Max Conc. [µg m⁻³]**	**GHS Hazard Class**
retene	483-65-8	0.042	H410
1,3-dihydroxynaphthalene	132-86-5	0.006	H315, H319, H335
1-(10-methylanthracen-9-yl)ethanone	36778-18-4	0.005	no data
9H-fluoren-9-one	486-25-9	0.004	H319, H411
1,8-dihydroxynaphthalene	569-42-6	0.001	H318
1-acenaphthenone	2235-15-6	0.001	no data

[1]Globally Harmonised System (GHS) classifies the hazards of chemicals. Data were obtained from European Chemicals Agency (ECHA) via PubChem. The hazard codes for each compound were reported by companies (e.g. manufacturer's safety datasheets). Only hazard codes mentioned in more than 10% of all reports for each compound are shown. Code H301: Toxic if swallowed; H302: Harmful if swallowed; H311: Toxic in contact with skin; H312: Harmful in contact with skin; H315: Cause skin irritation; H318: Causes serious eye damage; H319: Cause serious eye irritation; H332: Harmful if inhaled; H335: May cause respiratory irritation; H341: Suspected of causing genetic defects; H350: May cause cancer; H400: very toxic to

aquatic life (acute); H410: Very toxic to aquatic life with long lasting effects; H411: Toxic to aquatic life with long lasting effects; H412: Harmful to aquatic life with long lasting effects; H351: Suspected of causing cancer.

[2]Dehyaroabietic acid was classified as H301, H400 and H410 by 39 notifiers (less than 10% of all notifiers).

[3]This chemical has been verified to be of low concern by US Environmental Protection Agency's Safer Choice program.

References

Fabbri, D., Chiavari, G., Prati, S., Vassura, I. and Vangelista, M.: Gas chromatography/mass spectrometric characterisation of pyrolysis/silylation products of glucose and cellulose, Rapid Commun. Mass Spectrom., 16(24), 2349–2355, doi:10.1002/rcm.856, 2002.

Jen, C. N., Hatch, L. E., Selimovic, V., Yokelson, R. J., Weber, R., Fernandez, A. E., Kreisberg, N. M., Barsanti, K. C. and Goldstein, A. H.: Speciated and total emission factors of particulate organics from burning western US wildland fuels and their dependence on combustion efficiency, Atmos. Chem. Phys., 19(2), 1013–1026, doi:10.5194/acp-19-1013-2019, 2019.

Worton, D. R., Decker, M., Isaacman-VanWertz, G., Chan, A. W. H., Wilson, K. R. and Goldstein, A. H.: Improved molecular level identification of organic compounds using comprehensive two-dimensional chromatography, dual ionization energies and high resolution mass spectrometry, Analyst, 142(13), 2395–2403, doi:10.1039/c7an00625j, 2017.

Yu, J., Flagan, R. C. and Seinfeld, J. H.: Identification of products containing -COOH, -OH, and -C=O in atmospheric oxidation of hydrocarbons, Environ. Sci. Technol., 32(16), 2357–2370, doi:10.1021/es980129x, 1998.