## **Response to the comments of Anonymous Referee #2**

## Referee Comment Summary:

This paper furthers previous efforts to model the volatility of organic compounds based on their elemental composition. This is of great interest to members of the atmospheric community coupling soft ionization and high-resolution mass spectrometry techniques, which allows for precise molecular formula determination, but does not provide insight into molecular structure. To accomplish this, the authors utilized the Estimation Programs Interface (EPI) Suite to estimate the vapor pressure of more than 31,000 organic molecules contained in the National Cancer Institute open database. Each compound was grouped into one of six categories (CH, CHO, CHN, CHON, CHOS, CHONS) based upon its molecular formula. Multi-linear least squares analysis for each category was used to determine the proper coefficients to estimate saturation concentrations based on Eq.1 in the text. This is an extension of saturation concentration estimates presented by Donahue et al. (2011), which now includes contributions from nitrogen and sulfur-containing functional groups. Reasonable agreement is demonstrated for the volatilities predicted by Eq. 1 compared to both the EPI and EVAPORATION models. With the new volatility estimates, over 9,000 compounds observed in chamber or field measurements are presented within the molecular corridor framework outlined by Shiraiwa et al. (2014). This provides further evidence that molecular weight and volatility are key parameters that can effectively constrain reaction pathways for organic compounds in a variety of locations and oxidation regimes. This study is worthy of publication in ACP once the following comments have been addressed.

## **Responses:**

We thank Anonymous Referee #2 for the review and the positive evaluation of our manuscript. Based on your constructive suggestions for improvement, we will expand discussions in the revised manuscript as detailed below. Referee General Comment 1:

1) The EPI suite reports vapor pressures calculated in multiple ways (Antoine method, modified Grain method, Mackay method, mean of Antoine and Grain methods) with the MPBPWIN model. Which of these calculated values were used for the compounds in the NCI database? Additionally, it should be explicitly shown how these vapor pressures are converted to the saturation concentrations used for comparison with Eq. 1.

Response: Thanks for this helpful comment. Following your suggestion, we will include the below discussion in the revised manuscript.

"The EPI Suite reports vapor pressure calculated in three separate methods, including the Antoine method (Lyman et al., 1990), the modified Grain method (Lyman, 1985), and the Mackay method (Lyman, 1985). In this study, the modified Grain estimate was used for solids and the average of the Antoine and the modified Grain estimates was used for liquids and gases, following the suggestions of the EPI Suite. The vapor pressure was then converted to the saturation mass concentration ( $C_0$ in µg m<sup>-3</sup>) through the relationship:

$$C_0 = \frac{M10^6 p_0}{760RT}$$
(2)

where *M* is the molar mass (g mol<sup>-1</sup>),  $p_0$  is the saturation vapor pressure (mm Hg), R is the ideal gas constant (8.205 × 10<sup>-5</sup> atm K<sup>-1</sup> mol<sup>-1</sup> m<sup>3</sup>) and *T* is the temperature (K)."

2) While the new parameters derived for use with Eq. 1 are very useful in the extension to nitrogen- and sulfur-containing organic compounds, this model still does not capture differences in volatility due to the number of hydrogens in the molecule. For example, an aldehyde with the molecular formula CxH2xO and a primary alcohol (CxH2x+2O) would erroneously be predicted to have the same saturation concentration. Even though this was also true for the equation reported by Donahue et al., 2011, it may be worth noting if efforts have been made to account for the number of hydrogen atoms in a given molecule.

Response: Thanks for this helpful comment. Following your suggestion, we have

tried to account for the number of hydrogen atoms, leading to improvement in the CH class slightly, but have little impact on the prediction for compounds in other composition classes. As the effect of inclusion of hydrogen is limited, further analysis of ambient samples was kept as is. We will expand discussions in the revised manuscript as detailed below.

"Considering the number of hydrogen atoms in Eq. 2 could improve the prediction for compounds in the CH class slightly, but have little impact on the prediction for compounds in other composition classes. Detailed information including an estimated  $b_{\rm H}$  (the contribution of hydrogen atoms to  $\log_{10}C_0$ ) can be found in the Supplement. As the effect of inclusion of hydrogen is limited, further analysis was conducted using Eq. (2) for simplicity."

Following discussion and figures will be added in the Supplement:

"Table S1 shows the best-fit parameters for the following equation including the contribution of the number of hydrogen atoms to the saturation mass concentration. Comparing Fig. 3 and Fig. S1 as well as Fig. 4 and Fig. S2, we found that inclusion of the number of hydrogen atoms (Eq. S1) could improve the prediction for compounds in the CH class but has only slight impact on the prediction for compounds in other composition classes.

$$\log_{10}C_0 = (n_{\rm C}^0 - n_{\rm C})b_{\rm C} - n_{\rm O}b_{\rm O} - n_{\rm H}b_{\rm H} - 2\frac{n_{\rm C}n_{\rm O}}{n_{\rm C} + n_{\rm O}}b_{\rm CO} - n_{\rm N}b_{\rm N} - n_{\rm S}b_{\rm S}$$
(1)"

**Table S1.** Composition classes and the  $n_c^0$  and *b* values for saturation mass concentration parameterizations (Eq. S1) obtained by least-squares optimization using the NCI database.

Classes	$n_{\rm C}^0$	$b_{\mathrm{C}}$	$b_{ m H}$	b <sub>O</sub>	$b_{\rm CO}$	$b_{ m N}$	$b_{\mathrm{S}}$
СН	17.95	0.5742	-0.1417				
CHO	15.77	0.6238	-0.1387	1.735	-0.8592		
CHN	23.01	0.4307	-0.02110			0.9528	
CHON	21.12	0.4139	-0.03760	0.8092	-0.1174	1.1010	
CHOS	16.07	0.5348	-0.1507	1.354	-0.4175		0.8993
CHONS	19.20	0.5469	-0.1368	1.183	0.07310	1.0289	1.323



Figure S1. Saturation mass concentration ( $C_0$ ) of organic compounds predicted by Eq. S1 using the parameters in Table S1 plotted against  $C_0$  determined by the EPI Suite software in the NCI database for elemental composition classes of (**a**) CH, (**b**) CHO, (**c**) CHN, (**d**) CHON, (**e**) CHOS, and (**f**) CHONS.



**Figure S2.** (a) Saturation mass concentration ( $C_0$ ) of CHO compounds predicted by Eq. S1 with the coefficients in Table S1 and with the coefficients from Donahue et al. (2011) plotted against  $C_0$  computed by the EVAPORATION model (Compernolle et al., 2011). The data comprise 704 SOA oxidation products from biogenic (isoprene,  $\alpha$ -pinene, limonene, glyoxal) and anthropogenic precursors (C12 alkanes) as

presented in Shiraiwa et al. (2014). (b) Comparison of  $C_0$  predicted by Eq. S1 with the coefficients in Table S1 and experimental values taken from PHYSPROP database (http://esc.syrres.com/interkow/EpiSuiteData.htm)."

3) While a thorough statistical analysis is presented for the comparison of volatilities from the EPI suite and Eq. 1, a similar analysis is lacking for the comparison of the EVAPORATION model with volatilities from Eq.1 and Donahue et al. (Figure 4). It is reported "our newly developed parameterization also agrees well with EVAPORATION predictions and shows better agreement than Donahue et al. (2011)." This claim should be demonstrated from a statistical perspective. Additionally, from a visual perspective the new parameters appear to result in saturation concentrations biased high relative to the EVAPORATION estimates. Potential reasons for this discrepancy should be discussed.

Response: Thanks for this helpful comment. Following your suggestion, statistical analysis will be added. In the revised manuscript, additional discussions will be added to show the validation of our parameterization using observed data, which are more reliable than the EVAPORATION model.

We will expand discussions in the revised manuscript as detailed below.

"As shown in Fig. 4(a), our newly developed parameterization also agrees well with EVAPORATION predictions, showing better agreement than Donahue et al. (2011), as shown by smaller values of MB and MAGE. Figure 4(b) shows the comparison between  $C_0$  estimated by our parameterization and the experimental data. The experimental  $C_0$  values of 1729 organic compounds with heteroatoms including N or S are taken from the PHYSPROP database that is part of the EPI Suite (<u>http://esc.syrres.com/interkow/EpiSuiteData.htm</u>). Our method gives a good performance with *R* of 0.84, MB of -0.41, and MAGE of 1.24.

As shown in Fig. 4(b), the estimation error increases as the saturation mass concentration decreases. An accurate prediction of low vapor pressure is difficult due, in large part, to the limitation caused by measurement challenges. For example, the measured vapor pressure of dicarboxylic acid, a low-volatility compound commonly found in atmospheric aerosols, disagreed by up to two orders of magnitude between different measurement techniques (Krieger et al., 2012; Huisman et al., 2013; Bilde et al., 2015). The effects of functionalization, phase states (crystalline, amorphous, (subcooled) liquid), and molecular structure need to be considered in future experimental studies (Huisman et al., 2013; Bilde et al., 2015). The vapor pressure estimation methods could give very divergent predictions for the same compound (Clegg et al., 2008; O'Meara et al., 2014). Another difficulty in predicting low vapor pressure is that most estimation methods, including the EPI Suite, are constrained by databases heavily biased toward mono-functional compounds with saturation vapor pressures in the range of  $\sim 10^3 - 10^5$  Pa (Barley and McFiggans, 2010; Krieger et al., 2012). The EPI Suite uses the PHYSPROP database as its training data set. When comparing with the PHYSPROP database, the estimation error of EPI Suite increases as the vapor pressure decreases, especially when the vapor pressure decreases below  $10^{-4}$  Pa (http://www.epa.gov/sab/pdf/sab-07-011.pdf).

Despite of some limitations as described above, the new estimation parameterization derived from a large dataset in this study is sufficiently good to predict  $C_0$  for various structural organic classes as shown by overall good correlation. In the next section, the saturation mass concentration of ambient OA (e.g., the compounds with elemental composition measured by high-resolution mass spectrometry techniques) is estimated using Eq .(2)."





**Figure 4.** (a) Saturation mass concentration ( $C_0$ ) of CHO compounds predicted by Eq. (2) with the coefficients from this study and with the coefficients from Donahue et al. (2011) plotted against  $C_0$  computed by the EVAPORATION model (Compernolle et al., 2011). The data comprise 704 SOA oxidation products from biogenic (isoprene,  $\alpha$ -pinene, limonene, glyoxal) and anthropogenic precursors (C12 alkanes) as presented in Shiraiwa et al. (2014). (b) Comparison of  $C_0$  predicted by Eq. (2) in this study and experimental values taken from PHYSPROP Database (http://esc.syrres.com/interkow/EpiSuiteData.htm).

4) For the figures showing multiple molecular corridor plots (Figures 5, 6, 7, and especially Figure 1), the larger symbols (representing the average properties for a given subclass of compounds) are indistinguishable unless the figure is magnified by a factor of 4-5. This muddles many of the discussion points focused on these reported averages (e.g. lines 10-12, page 27882 discussing trends in the CHN molecular weight, volatility relationship). Either these larger symbols need to be more pronounced over the smaller, individual compound symbols, or they should be plotted separately. Either way, Figure 1 especially needs to be increased in size. Also, figure 1 axes are switched from the remaining plots, why not just start with the axes plotted the way you will display throughout the remainder of the paper?

Response: Following you suggestions, the larger symbols have been enlarged and the size of Fig. 1 has been increased. We would like to keep the axes in Fig. 1 to keep in accordance to the original representation of "molecular corridors" (Shiraiwa et al., 2014). We displayed saturation mass concentration as a function of molar mass in Figs. 5-9, which appears more straightforward for direct comparisons to mass spectra.

5) In Figure 8, there exists a homologous series of organosulfates with molecular weights between 400 - 600 g mol<sup>-1</sup> and log10(C0) < -10, which appear to have distinct, lower dlogC0/dM values. Any characteristics of these compounds that differentiate them from the organosulfate and organonitrate compounds contained in

the orange oval should be mentioned.

Response: Thanks for this suggestion. The following discussion will be added in the revised manuscript.

"Note that there exists a homologous series of organosulfates with molar mass between 400 – 600 g mol<sup>-1</sup> and log10( $C_0$ ) < -10, which appear to have lower dlog $C_0/dM$  values. These compounds, e.g.,  $C_{17}H_{18}O_{16}S_1$ ,  $C_{18}H_{20}O_{16}S_1$ , and  $C_{19}H_{22}O_{16}S_1$ , were found in the cloud water and their formation may be due to esterification of hydroxyl groups with sulfuric acid or acid-catalyzed reactions of epoxides (Zhao et al., 2013)."

6) In Table S1, the log10(C0) values are being reported with units of  $\mu$ g m-3 when they should be dimensionless. The same issue exists with the reported mean bias and mean absolute gross error values. These statistics were either calculated with the dimensionless saturation concentrations, and should be reported as dimensionless, or should be recalculated as such.

Response: Thanks for pointing it out. The units of statistical measures will be changed to be dimensionless all through the manuscript.

## Specific comments:

1) Page 27879, Lines 26-27: Reword to read "Volatility is a consequence of the molecular characteristics of molar mass, chemical composition, and structure."

Response: The sentence has been revised following your suggestion.

2) Page 27880, Line 15: These references are not all soft ionization techniques. Please check these references, or be more generic to how these studies determined elemental composition. This is an important point though. Methods not using soft ionization will contain fragments and not necessarily original molecules. Please describe the measurement methods used to obtain atmospheric data (section 4).

Response: Following your suggestion, we will clarify this point in the revised manuscript. Information of measurement methods used to obtain atmospheric data is

3) Page 27882, Lines 6-9: The sentence concerning the oxidation state of nitrogen and sulfur-containing compounds should be moved to the end of the paragraph after the introduction of Figures 1d-f.

Response: Following your suggestion, the sentence has been moved to the end of the paragraph.

4) Page 27882, Lines 10-12: At this point in the text it has not been mentioned that the large symbols in these molecular corridor plots correspond to averages for the different compound sub-classes. Doing so would help drive home the point for the trends with the CHN compounds (along with addressing general comment #4).

Response: Following your suggestion, the following sentence will be added in Page 27882:

"The surrogate compounds with the mean values of M,  $C_0$ , and O:C ratio computed for each of the structural sub-classes are indicated by larger symbols with error bars indicating standard deviations."

5) Page 27884, Lines 11-13: Are the reported statistics (R, MAGE, MB) for each major class (CH, CHO, etc...) limited to compounds with molecular weights below 500 g mol-1? If so, why restrict this analysis to the lower molecular weight compounds when this was not apparently done in Figure 3? Please clarify what statistics are being reported here. Additionally, see general comment #6 concerning the use of units for these statistics.

Response: Following your suggestion, the sentence has been revised as "As shown in Fig. 3, our new method performs well with R above 0.8 and MAGE less than 1.8 for all the composition classes."

6) Page 27886, Line 1: Change "events" to "event".

Response: It will be corrected following your suggestion.

7) Figure 1: Number of compounds in each class only shown in panels (a) and (b). Please add this information to panels c-f.

Response: Number of compounds will be added for panels c-f.

8) Figure 3: Change x-axes label from "EPI suit" to "EPI suite".
 Response: Thanks. It will be corrected.