Modeling the influence of precursor volatility and molecular structure on secondary organic aerosol formation using evaporated fuel experiments

S. H. Jathar\textsuperscript{1,*}, N. M. Donahue\textsuperscript{1}, P. J. Adams\textsuperscript{1}, and A. L. Robinson\textsuperscript{1}

\textsuperscript{1}Center for Atmospheric Particle Studies, Carnegie Mellon University, Pittsburgh PA, USA
*present address: Department of Civil and Environmental Engineering, University of California, Davis CA USA

Received: 3 August 2013 – Accepted: 2 September 2013 – Published: 16 September 2013

Correspondence to: A. L. Robinson (alr@andrew.cmu.edu)

Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

We use SOA production data from an ensemble of evaporated fuels to test various SOA formation models. Except for gasoline, traditional SOA models focusing exclusively on volatile species in the fuels under-predict the observed SOA formation. These models can be improved dramatically by accounting for lower volatility species, but at the cost of a large set of free parameters. In contrast, a SOA model based only on the volatility of the precursor, starting with the volatility distribution of the evaporated fuels and optimized for the volatility reduction of first-generation products, reasonably reproduces the observed SOA formation with relatively few free parameters. The exceptions are exotic fuels such as Fischer-Tropsch fuels that expose the central assumption of the volatility based model that most emissions consist of complex mixtures displaying reasonably average behavior. However, for the vast majority of fuels, the volatility based model performs well.

1 Introduction

Secondary organic aerosol (SOA) is aerosol mass formed from the oxidation of gas-phase organic species emitted by natural and anthropogenic sources. Most SOA models under-predict SOA formation both in the laboratory and in the atmosphere (Heald et al., 2005; Vutukuru et al., 2006; Johnson et al., 2006; Morris et al., 2006; Dzepina et al., 2009, 2010; Grieshop et al., 2009; Jathar et al., 2012). This indicates that there are large gaps in understanding the numerous precursors and pathways to SOA formation (gas-phase oxidation, multi-generational aging, heterogeneous chemistry, condensed-phase reactions, cloud processing).

Gas-phase oxidation of organic compounds is a major source of SOA and hence, has been extensively studied using smog chamber experiments. However, these experiments have focused on a handful (tens) of organic compounds, which are a small subset of the thousands of organic compounds found in the atmosphere (Goldstein and...
Galbally, 2007). SOA models based on those data mostly include emissions and subsequent SOA formation from high-flux volatile organic compounds (VOC) like isoprene, terpenes, single-ring aromatics, and alkanes and alkenes that have a carbon number less than or equal to 12. Multiple studies have shown that these SOA models are able to explain only a small fraction of the SOA measured from gasoline, diesel and aircraft exhaust, wood smoke and urban plumes (De Gouw et al., 2005; Gordon et al., 2013a, b; Dzepina et al., 2009; Volkamer et al., 2006; Robinson et al., 2007; Miracolo et al., 2011, 2012; Grieshop et al., 2009). That is probably because combustion sources emit substantial amounts of lower volatility organics (e.g. hydrocarbons with more than 12 carbons), that are difficult to speciate and therefore not commonly included in models (Robinson et al., 2007). Hence, more research is needed to test these SOA models against data for complex mixtures that contain both small and large organics.

Single-compound smog chamber studies have demonstrated that SOA formation depends both on the precursor's volatility (vapor pressure) and molecular structure. In general, SOA formation increases with decreasing volatility but has a complex dependence on molecular structure. For example, SOA formation from alkanes increases as the carbon-number increases (Lim and Ziemann, 2009a, b; Presto et al., 2010; Tkacik et al., 2012) increasing the carbon-number reduces the volatility of the precursor. The importance of molecular structure is illustrated by experiments with different classes of alkanes. For the same carbon-number or volatility, cyclic alkanes form the most SOA followed by \( n \)-alkanes followed by branched alkanes (Lim and Ziemann, 2009a, b; Tkacik et al., 2012). Single-ring aromatics (benzene, toluene and xylenes) form much more SOA than similar sized (C\(_6\) to C\(_9\) \( n \)-alkanes or alkenes (Ng et al., 2007; Song et al., 2007; Hildebrandt et al., 2009). However, multi-ring aromatics and \( n \)-alkanes with similar carbon numbers form similar amounts of SOA (Chan et al., 2009; Shakya and Griffin, 2010; Presto et al., 2010). Existing chemical mechanisms account for some differences in volatility and molecular structure. For example, the gas-phase mechanism SAPRC (Carter et al., 2007) has different model species to account for differences in volatility, i.e. ALK4 (C\(_5\)-C\(_7\)) vs. ALK5 (C\(_7\)+) and molecular structure, i.e. alkanes (ALK)
vs. alkenes (OLE) vs. aromatics (ARO). So SOA models built using SAPRC can account for some differences in volatility and molecular structure. However, SAPRC and other reduced chemical mechanisms were developed to simulate ozone formation and therefore lump different species by reaction rates rather than SOA mass yields. Even fully explicit mechanisms such as the Master Chemical Mechanism (Saunders et al., 2003) emphasize low carbon-number VOCs that dominate the hydrocarbon flux in most locations, due to a focus on ozone formation and fast photochemistry. Therefore, it is unclear how any of these mechanisms are suitable for predicting SOA formation from complex mixtures.

Another challenge is incorporating unspeciated organics into SOA models. Low volatility organics (semi-volatile and intermediate volatility organic compounds; SVOC and IVOC) are thought to be important classes of SOA precursors (Grieshop et al., 2009; Robinson et al., 2007; Shrivastava et al., 2008) but are difficult to speciate (Schauer et al., 1999a, b, 2001, 2002a, b). Gas-particle partitioning (Robinson et al., 2007; Shrivastava et al., 2006; May et al., 2013a, b) and gas-chromatography (Presto et al., 2012) measurements provide some information on the volatility distribution of these emissions; therefore, aerosol models have adopted a volatility-based approach to model SOA formation from unspeciated S/IVOCs. For example, Robinson et al. (2007) assumed that unspeciated SVOCs and IVOCs react with the hydroxyl radical (OH) to form products that were one order of magnitude lower in volatility than their precursor. Jathar et al. (2012) extended this approach by distributing the reaction products across a range of volatilities. Pye and Seinfeld (2010) proposed a single-step mechanism for SVOC where the products of oxidation were two orders of magnitude lower in volatility than the precursor and used SOA mass yield data for naphthalene as a surrogate for unspeciated IVOCs. Although they adopted somewhat different approaches, all of these studies have shown that including unspeciated organics in SOA models helped close large gaps between predicted and measured SOA mass concentrations (Shrivastava et al., 2008; Tsimpidi et al., 2009; Dzepina et al., 2010; Pye and Seinfeld, 2010; Jathar et al., 2011). These new models have used a lump-and-yield scheme to model
SOA formation from speciated emissions and a volatility-based scheme to model SOA formation from unspeciated S/IVOCs. Given the success of volatility-based schemes, it could be worthwhile to explore their capabilities for modeling all of SOA, not just SOA from S/IVOCs.

In this work, we compare predictions of different SOA models to data from smog-chamber experiments conducted with evaporated fuels to investigate the effects of precursor volatility and molecular structure on SOA formation. We suggest modifications and propose new (volatility-based) schemes to model SOA formation. Since smog-chamber data have less than a day of oxidant exposure, the data represent relatively fresh SOA formed under urban-like conditions. This work does not consider multigenerational oxidation of SOA products.

2 Materials and methods

2.1 Experimental data

Jathar et al. (2013) presented data from twenty-three high-NOX photo-oxidation experiments in the Carnegie Mellon University smog chamber, constraining SOA formation from eleven different fuels (gasoline: 3 experiments; Fischer Tropsch derived from coal: 2 experiments; Fischer Tropsch derived from natural gas: 2 experiments; Jet Propellent-8: 6 experiments; multiple diesels: 10 experiments). The Fischer-Tropsch fuels are synthetic substitutes for Jet Propellent-8 fuel. The diesels are part of a test fuel matrix (Fuels for Advanced Combustion Engines) designed around three properties of importance to combustion engines: ignition quality, fuel chemistry and volatility (Alnajjar et al., 2010). Here we provide a brief summary of the experiments; details can be found in Jathar et al. (2013).

First, ammonium sulfate seed was added to the chamber to facilitate condensation and prevent nucleation of SOA products. Second, nitrous acid (HONO) was bubbled into the chamber. Third, the precursor (fuel) was introduced into the chamber using...
a heated septum. Finally, photo-oxidation was initiated by turning on the chamber UV black lights, which photolyzed HONO to produce OH radicals. NO and NO$_2$ formed as by-products of HONO irradiation resulted in a low VOC/NO$_x$ ratio that was consistent with ratios found in urban polluted regions. The experiment was performed at low relative humidity (<5%) and a temperature of around 298 K. Concentrations of gas phase organic compounds were tracked using a GC-MS (Logue et al., 2009) and a proton-transfer reaction mass spectrometer (PTR-MS, Ionicon Analytik) and were used to calculate OH concentrations and OH exposure during the experiment. Particle-phase measurements were made using a scanning mobility particle sizer (SMPS, TSI Inc.) and a quadrupole or high resolution aerosol mass spectrometer (Q-AMS or HR-AMS, Aerodyne Research Inc.) to measure non-refractory aerosol mass, size and composition. Both SMPS and Q-AMS/HR-AMS results were wall-loss corrected to calculate a lower and upper bound on the total SOA formation (Weitkamp et al., 2007; Hildebrandt et al., 2009).

Jathar et al. (2013) found that, for a unit mass of fuel reacted, evaporated diesel formed the most SOA followed by Jet Propellent-8, Fischer-Tropsch from natural gas, gasoline and Fischer-Tropsch from coal. More simply, fuels with higher carbon-numbers and more aromatics formed more SOA. Fuels with lower-carbon numbers and substituted alkanes formed less SOA. In contrast to gasoline (Odum et al., 1997), the SOA formation from evaporated diesel was not a strong function of the aromatic content. In fact, all diesels had similar SOA mass yields even though they varied in volatility and had different proportions of $n$-alkanes, branched/cyclic alkanes and aromatics.

### 2.2 Fuels

The fuels used in Jathar et al. (2013) spanned a wide range of volatility and molecular structure. In Fig. S1 we plot the fuel composition data using the volatility basis set (VBS) (Donahue et al., 2006), color-coded by molecular structure. The VBS is a modeling framework that classifies organics into logarithmically spaced bins of effective
saturation concentration ($C^*$). $C^*$ (inverse of the partitioning coefficient, $K_p$ (Pankow, 1994) is proportional to the saturation vapor pressure; it is a semi-empirical property describing the gas-particle partitioning of an organic mixture at 298 K.

To represent a fuel in the VBS, one needs to know the volatility (or $C^*$) of each individual species in the fuel. However, the granularity of the composition data varied substantially across the fuels. The most comprehensive data were available for the diesel fuels (Alnajjar et al., 2010) for which we had finely-resolved data both by organic class and carbon number. For gasoline, we had alkane and aromatic data by carbon number but no information about the specific species and no resolution in the alkene data. For Fischer-Tropsch from natural gas and Jet Propellent-8, we had speciated data for $n$-alkanes ranging from $n$-heptane to $n$-nonadecane but lumped data for branched alkanes, cyclic alkanes and aromatics. We assumed the hydrocarbons other than $n$-alkanes had the same carbon number distribution as the $n$-alkanes in that fuel. Finally, the Fischer-Tropsch from coal is mostly composed of branched and cyclic alkanes that are difficult to speciate. Here, we assumed that the distribution of branched and cyclic alkanes was similar to the distribution of $n$-alkanes in Fischer-Tropsch from natural gas (Corporan et al., 2011).

To overcome the limitations in the fuel composition data, we developed a mathematical relationship for $C^*$ as a function of carbon number of a hydrocarbon. The relationship was derived using vapor pressure data from NIST for $n$-alkanes, pure cyclic alkanes, simple branched alkanes, single ring aromatics, naphthalene, straight and cyclic alkenes, isoprene, and common terpenes (NIST, 2012). We found aromatics ($C^* = e^{(22.3-\text{carbon#})/0.806}$) to have a slightly different relationship than alkanes and alkenes ($C^* = e^{(24.5-\text{carbon#})/0.899}$).

### 2.3 SOA models

In this work, we compared the measured SOA formation from Jathar et al. (2013) to SOA predictions/fits from a number of different models. In all models, fuel species react
with the OH radical to form SOA, which is semi-empirically represented using a set of five semi-volatile surrogate products in the VBS (Donahue et al., 2006).

Fuel Species + OH → \( \sum \alpha_i P_i \)  

(1)

where \( \alpha_i \) is the mass yield of product \( P_i \), which has a \( C^* \) of 0.1, 1, 10, 100 and 1000 \( \mu g m^{-3} \). The amount of SOA at equilibrium is defined by the gas-particle partitioning of the surrogate products, \( P_i \)

\[
\zeta_i = \left(1 + \frac{C^*_i}{C_{OA}}\right)^{-1}
\]

\[C_{OA} = \sum_{i=1}^{N} \zeta_i \times M_i |_{g+p}
\]

where, \( \zeta_i \) is the fraction of mass of \( P_i \) in the particulate phase, \( C^*_i \) is the effective saturation concentration of \( P_i \) in \( \mu g m^{-3} \), \( C_{OA} \) is the total particulate OA concentration in \( \mu g m^{-3} \), \( M_i \) is the total organic concentration (gas + particle) of \( P_i \) in \( \mu g m^{-3} \).

The production of semi-volatile species is described using the following equations:

\[
\frac{d[X_j]}{dt} = -k_{OH,X_j}[OH][X_j]
\]

(3)

\[
\frac{d[M_i |_{g+p}]}{dt} = \sum \alpha_{i,j} k_{OH,X_j}[OH][X_j]
\]

(4)

Equation (3) represents the first-generation oxidation of an SOA precursor where \( k_{OH,X} \) is the reaction rate constant between the oxidant [OH] and SOA precursor [X]. The index \( j \) indicates different precursors, either individual fuel species or volatility bins of SOA.
the fuel. The $k_{\text{OH}}$ for each fuel species is listed in Tables S1 through S5 (Supplement). Equation (4) tracks the total (gas+particle) concentration of the semivolatile products, \(M_{i/g+p}\); gas-particle partitioning is calculated using Eq. (2).

We do not explicitly consider the multi-generational oxidation of semi-volatile products because the smog-chamber experiments used in this work had relatively low OH exposure; average of $1.7 \times 10^7$ molecules-h cm\(^{-3}\). For a typical precursor with a reaction rate of $1 \times 10^{-11}$ cm\(^{-3}\) molecule\(^{-1}\) s\(^{-1}\), this OH exposure corresponded to 1.5 e-folding lifetimes. If the products oxidized at a similar rate, less than 12% of the first generation products and less than 1% of the second generation products would react. Therefore the SOA formed in the smog chamber was relatively fresh and dominated by first-generation products.

We used three different types of models in this work, which are described in more detail in the subsequent sections. First, an empirical model where the data were fit to develop a unique SOA parameterization for each fuel. Second, a traditional model that used volatility- and molecular structure-resolved schemes found in SOA modules in chemical transport models (CTM). Third, a volatility-based model that related SOA production only to the precursor’s volatility and ignores molecular structure.

### 2.3.1 Empirical

Figure 1a shows the schematic for the Empirical model. Here, we assumed that each species in the fuel reacted with the OH radical as per Eq. (3) but had the same mass yield in Eq. (4) as all the other species in that fuel, i.e. $\alpha_{i,j} = \alpha_j$. Each fuel was treated separately and therefore has a different set of $\alpha_i$. $\alpha_j$ for each fuel were fit across a 5-bin VBS by minimizing the sum of the squared error between the model predictions of SOA using Eqs. (2)–(4) and the measurements of SOA. The Empirical model represented a “best fit” to the experimental data, against which other models could be compared.
2.3.2 Traditional

Most SOA models used in CTMs account for SOA formation from speciated VOCs like isoprene, terpenes, single-ring aromatics and smaller alkanes and alkenes. We refer to these models as Traditional models, which is shown schematically in Fig. 1b. In this work, the Traditional model was based on the SOA model in PMCAMx (Lane et al., 2008), which used the SAPRC gas-phase mechanism to lump SOA precursors into a model species, i.e. \( n \)-decane is lumped as ALK5, benzene is lumped as ARO1 and so on (Carter et al., 2007). We used mass yields \( \alpha_{ij} \) published by Murphy and Pandis (2010) for each SAPRC model species to predict SOA formation; these values \( \alpha_{ij} \) are listed in Table S6 (Supplement). These were determined by fitting data from smog-chamber experiments conducted on single compounds.

The Traditional model accounted for some differences in both the precursor volatility and molecular structure. For example, SAPRC differentiates between alkanes (ALK), alkenes (OLE), aromatics (ARO), isoprene (ISOP) and terpenes (TERP). (There are no terpenes or isoprene in any of the fuels, so we do not discuss them further.) For each molecular structure (ALK, OLE and ARO), there are two or more SAPRC species to account for differences in reactivity with the OH radical. For a given molecular structure, reactivity partially correlates with volatility so there is some differentiation with precursor volatility. However, SAPRC and most other gas-phase mechanisms were designed to simulate hydrocarbon-NO\(_x\)-ozone photochemistry, not SOA formation. So, they do not account for differences in SOA formation that recent smog-chamber experiments have demonstrated. For example, SAPRC does not differentiate between branched and straight alkanes even though branched alkanes have much lower SOA yields (Lim and Ziemann, 2009b; Tkacik et al., 2012). SAPRC also lumps all alkanes greater than \( C_7 \) into ALK5 even though there is a wide variation in SOA formation across this set of species (Lim and Ziemann, 2009b; Presto et al., 2010; Tkacik et al., 2012).

Traditional models failed to account for the SOA formation potential of large organic compounds (organics that have a carbon number greater than 12) and organics that
remain unspeciated. The reason why larger and unspeciated organics are neglected is because it is hard to speciate them using traditional GC-MS techniques. These organics, therefore, are rarely found in emission profiles and emission inventories (Robinson et al., 2007; Goldstein and Galbally, 2007). For example, in EPA’s SPECIATE database (version 4.3) – a cornerstone database used exclusively to build emission inventories for anthropogenic sources in the US – the emission profiles of the top six combustion sources of VOCs (on- and off-road gasoline, on- and off-road diesel, wood burning and open burning; they accounted for 80% of all VOC combustion emissions in the US in 2008) had a negligible fraction (< 1%) of organics with a carbon number greater than 12. Similarly, Simon et al. (2010) found that there were no organics with a carbon number greater than 12 in the top-50 organic-compound list by emissions in the US. Contrastingly, Schauer and coworkers (Schauer et al., 1999a, b, 2001, 2002a, b) and recently Robinson and coworkers (Presto et al., 2011; Shrivastava et al., 2008) and Goldstein and coworkers (Gentner et al., 2012; Worton et al., 2012) have demonstrated that combustion sources emit substantial mass of organics with a carbon number greater than 12 and one-third or more of the organic emissions cannot be speciated. These organics are capable of forming SOA and characterizing and quantifying them is an ongoing area of research. In light of this finding, the Traditional model was run in two configurations:

**Traditional (specified):** In this configuration, the model considered SOA formation from speciated organic compounds less than or equal to a carbon number of 12. This configuration was representative of the SOA treatment in most aerosol models. The SAPRC lumping for the Traditional (specified) model is listed in Tables S1 through S5.

**Traditional (all):** In this configuration, the model included SOA formation from speciated organics accounted for by the Traditional (specified) model and from other larger (greater than C\(^{12}\)) and unspeciated organics. The SAPRC lumping for the Traditional (base) model is listed in Tables S1 through S5.
To implement the Traditional models in an aerosol transport model, one needs 9 SAPRC precursors. Each precursor required a 5-bin VBS parameterization. Therefore, the Traditional model had 45 free parameters.

### 2.3.3 Volatility-based model

In the Volatility-based model, we assumed that SOA formation depended only on the precursor volatility. The Volatility-based model is shown schematically in Fig. 1c. The precursor (fuel) mass was distributed using the VBS. Each precursor $C^*$ bin (corresponding to $X_j$ in Eq. 3) formed a certain product distribution with mass yields $\alpha_{i,j}$. Each higher (or lower) precursor $C^*$ bin formed the same product distribution but shifted by one $C^*$ bin based on the work of Presto et al. (2010) and Lim and Ziemann (2009b), who found that for $n$-alkanes, the addition of 2 carbon atoms to an $n$-alkane shifted its corresponding SOA product distribution, on average, by one $C^*$ bin.

A key input to the Volatility-based model was the fuel volatility distribution, which was constructed using composition data. Figure S1 shows the volatility distribution for each fuel. Another key input to the Volatility-based model was the precursor $C^*$ bin’s reaction rate with the OH radical ($k_{OH,X}$). Using Atkinson and Arey (2003), we developed a mathematical relationship between $C^*$ of a hydrocarbon and $k_{OH}$. Alkanes, alkenes and aromatics had different relationships; alkanes: $k_{OH} = -1.84 \times 10^{-12} \log(C^*) + 4.27 \times 10^{-11}$, alkenes: $k_{OH} = 4.0 \times 10^{-11}$, and aromatics: $k_{OH} = -5.7 \times 10^{-12} \log(C^*) + 1.14 \times 10^{-10}$.

To implement the Volatility-based model in a CTM, one needs 8 lumped precursor species, one for each $C^*$ bin from $10^2 \mu g m^{-3}$ to $10^9 \mu g m^{-3}$. Each precursor had the same 5-bin VBS parameterization but shifted in volatility space. Therefore, the VBS only had 5 free parameters.
3 Modeling results

3.1 Empirical model

The Empirical model was fit to the measured SOA data; the quality of fit is shown in Fig. 2a. The parameters ($\alpha_i$) were fit using the entire dataset, but only the end-of-experiment SOA data are shown in Fig. 2 for visual clarity. We used statistical metrics of fractional error and fractional bias to quantitatively evaluate model performance, which are calculated using all of the chamber data.

\[
\text{Fractional Bias} = \frac{1}{N} \sum_{i=1}^{N} \frac{P - M}{\frac{P+M}{2}}
\]

\[
\text{Fractional Error} = \frac{1}{N} \sum_{i=1}^{N} \frac{|P - M|}{\frac{P+M}{2}}
\]

The fractional error and fractional bias are plotted in Fig. 4.

The Empirical model represented the best fit to the data and therefore its statistical metrics presented an upper limit on the performance of other models. The Empirical model produced a good model-measurement comparison based on the criteria of Morris et al. (2005). The high fractional error (50%) was likely, in part, due to the experiment-to-experiment variability and uncertainty in the measurements. The experiment-to-experiment variability and uncertainty probably arose from the SOA formation being sensitive to modest changes in the wall-loss rates, oxidant or radical concentrations and VOC-to-NO\(_x\) ratios. The extent of variability and/or uncertainty was comparable to previous SOA smog chamber studies. To cite a few examples, Hildebrandt et al. (2009) found that the estimated SOA mass yields for toluene from half a dozen experiments varied by a factor of two. When compared against Ng et al. (2007), the SOA mass yields were a factor of two higher. Similarly, SOA mass yields for similar
experiments on naphthalene from Chan et al. (2009) and Shakya and Griffin (2010) varied by at least a factor of two.

### 3.2 Traditional models

SOA predictions from the Traditional (speciated) model are compared against measurements in Fig. 2b. The model severely under-predicted SOA formation with both high fractional error (−170 %) and fractional bias (−130 %). The only exception were the gasoline experiments, for which the model over-predicted the measured SOA by more than a factor of 5. The under-prediction occurred because the Traditional (speciated) model only considered SOA formation from speciated organics with carbon numbers less than 12. Almost 100 % of the gasoline mass was considered in the Traditional (speciated) model, but for the other fuels, only 30–50 % of the mass was considered in the Traditional (speciated) model because the majority of the fuels could not be speciated (Fischer-Tropschs) and/or contained significant amounts of high carbon number species (Jet Propellent-8 and diesels).

The Traditional (all) model performed substantially better (fractional error = 76 %, fractional bias = 26 %) than the Traditional (speciated) model but not quite as well as the Empirical model; model-measurement comparison is shown in Fig. 2c. The Fischer Tropsch from natural gas, Jet Propellent-8 and diesel SOA data are predicted within a factor of two but the SOA data from gasoline and Fischer Tropsch from coal are over-predicted. The improved performance occurred because the Traditional (all) model includes 100 % of the organic precursor mass (speciated and unspeciated). This implies that SOA model performance can be improved by simply accounting for SOA formation from higher carbon number (> 12) and unspeciated organic compounds.

Both Traditional models (speciated and all) over-predicted the gasoline data by a factor of 5. These models used the Murphy and Pandis (2010) mass yields for aromatics, which are some of the highest published. Figure S2(a, c) compares the Murphy and Pandis (2010) yields to other published parameterizations. We find that they are a factor of 3 or 4 larger than those measured by Odum et al. (1997b) for various single-ring
aromatics and a factor of 2 higher than those measured by Ng et al. (2007) for toluene and m-xylene. When we used the SOA mass yields of Odum et al. (1997b) instead of those from Murphy and Pandis (2010) for the SAPRC species ARO1 and ARO2, we achieved much better model-measurement comparison for gasoline (but not for the other fuels). This suggests that SOA mass yields (especially those of Murphy and Pandis, 2010) for aromatics in SOA models might be biased high and may need to be corrected.

The Traditional (all) model over-predicted the Fischer-Tropsch from coal data by a factor of 5. This fuel is mostly composed of branched alkanes (88% by mass), which have been shown to have much smaller SOA mass yields than straight and cyclic alkanes (Lim and Ziemann, 2009a; Tkacik et al., 2012). Therefore, it is important to differentiate between branched, straight and cyclic alkanes when modeling SOA formation. However, the Traditional models assumed that the SOA formation does not change with alkane structure (all alkanes are lumped into the same set of SAPRC species). Since the Murphy and Pandis (2010) alkane SOA mass yields used in this work have been parameterized using n-alkane data (Lane et al., 2008), the Fischer-Tropsch from coal data are over-predicted.

Based on the model performance for gasoline and Fischer-Tropsch from coal, the performance of the Traditional (all) model can be improved if the SOA mass yields for aromatics are lowered and branched alkanes are treated separately from straight and cyclic alkanes. However, lowering the aromatic SOA mass yields and adding a branched alkane model species would change model predictions for the other fuels. Therefore, we used a genetic optimization algorithm to determine a set of SOA mass yields for the branched alkane model species, lower SOA mass yields for single-ring aromatics and nudge mass yields for the other model species (ALK4, ALK5, OLE1, OLE2) all at the same time. The objective of the algorithm was to minimize the fractional error. One solution (a genetic optimization code does not guarantee absolute minimization or a unique solution) requires the Traditional (all) model to undergo two major changes: (1) addition of a branched alkane model species (BALK) that included...
only branched alkanes that are C7 and higher and (2) lumping single-ring aromatics under ARO1 and lumping multi-ring aromatics under ARO2. The mass yields for BALK and nudged mass yields for ARO1, ARO2, ALK4, AL5, OLE1 and OLE2 are listed in Table 1. We call this new model Traditional (extended).

The mass yields from Murphy and Pandis (2010) are compared to mass yields used with the Traditional (extended) model in Fig. S2(b, d.) The optimization slightly increased mass yields for ALK4 and ALK5; they are now closer to mass yields for n-dodecane than n-decane. The mass yields for ARO1 are significantly lowered and are much closer to those measured by Odum et al. (1997b) and Ng et al. (2007) for toluene and m-xylene while those for ARO2 remain about the same and closer to those measured by Chan et al. (2009) for naphthalene.

Since the Traditional (extended) model are fit to the data, Fig. 2d compares the goodness-of-fit. The Traditional (extended) model performed better than the Traditional (all) model as most of its predictions lied within a factor of two of the measurements (green diamond in Fig. 4). As expected, most of the improvement stemmed from better (lower) predictions for gasoline and Fischer Tropsch from coal.

### 3.3 Volatility-based model

For the Volatility-based model, all of the SOA data were fit to determine a SOA mass-yield matrix \((\alpha_{i,j})\) for the precursor mass distributed across a set of \(C^*\) bins; \(\alpha_{i,j}\) is listed in Table 2. The goodness-of-fit is shown in Fig. 3a and the fractional error and fractional bias are plotted in Fig. 4. Most predictions were within a factor of 2, which implies that a model based on precursor volatility alone can explain a lot of the variability in the measured SOA data. However, there were several instances where the Volatility-based model’s performance was poor. For the diesel data at lower \(C_{OA}\), the Volatility-based model predicted roughly the same SOA mass despite variability in the measured SOA mass because all the diesel fuels had similar volatility distributions. The Volatility-based model predicts the same SOA formation from Fischer Tropsch from coal and Fischer Tropsch from natural gas because they had the same volatility.
profile but different molecular structures; Fischer Tropsch from coal is mostly composed of branched alkanes while Fischer Tropsch from natural gas is an equal mix of straight and branched alkanes. Therefore, the Volatility-based model over-predicted the Fischer Tropsch from coal SOA and under-predicted the Fischer Tropsch from natural gas SOA.

In Fig. 3b we plot the SOA mass yields at a $C_{OA}$ of 5 µg m$^{-3}$ from the Volatility-based model as a dashed line, with the yellow band capturing uncertainty. On the same figure, we also plot SOA mass yields for various SOA precursors from the literature; the error bars represent an uncertainty of a factor of two. The SOA mass yields derived by fitting the Volatility-based model to the data compared favorably with the single-compound data. For example, the fit for the Volatility-based model was able to correctly predict the range and drop in SOA mass yield for $n$-alkanes. Similarly, the fit was able to predict the right SOA mass yield range for the lower carbon number alkenes and biogenics. It did that despite having no a-priori information about what the SOA mass yields might be for species in a given volatility range. This suggests that the SOA mass yields interpreted using the Volatility-based model are able to mirror the SOA mass yield dependency observed with the volatility of single compounds.

One could extend the Volatility-based model to also incorporate molecular structure by having a different set of mass yields for different molecular structures. However, SOA data used in this work were insufficient to determine a molecular structure-resolved Volatility-based scheme.

4 Discussion

The evaporated fuel experiments underscore the importance of including all speciated and unspeciated organic precursor mass in SOA models. Historically traditional SOA models have only considered SOA formation from VOCs with carbon number less than or equal to 12. The data also indicate that the current ozone-focused VOC lumping schemes are inadequate for modeling SOA formation. The predictions of these models
can be improved by accounting for the effects of precursor molecular structure and size on SOA formation (branched vs. straight/cyclic alkanes and single-ring vs. multi-ring aromatics).

An SOA model based purely on the precursor volatility is able to capture the variability in SOA formation from different fuels that like combustion emissions contain complex mixtures of alkanes, alkenes and aromatics and their sub-types (straight, branched, cyclic, single-ring, multi-ring). In addition, the SOA yields of the lumped, volatility-based species are similar to data from single compound experiments.

The SOA model based on volatility alone performs poorly when tested with SOA data from synthetic fuels like Fischer-Tropsch from coal and natural gas. We hypothesize that this is due to these fuels having a much simpler composition, dominated by one class of species, than the other fuels. For example, the Fischer-Tropsch from coal is mostly composed of branched alkanes. For these types of systems, it becomes important to account for the effects of molecular structure on SOA formation.

An advantage of the volatility-based approach is that traditional SOA models have many more free parameters (~50) than volatility-based models (5 parameters) but their performance is only marginally better for the evaporated fuels (Fig. 4). In addition, Traditional SOA models require many more SOA precursors (12 to 21, Shrivastava et al., 2008; Jathar et al., 2011; Tsimpidi et al., 2009; Dzepina et al., 2009; Pye and Seinfeld, 2010) than the volatility-based model alone (8). So, although molecular structure influences SOA formation, given the results of this work, SOA formation as a function of volatility may be sufficient for use in CTMs.

Supplementary material related to this article is available online at http://www.atmos-chem-phys-discuss.net/13/24405/2013/acpd-13-24405-2013-supplement.pdf.

Acknowledgements. We would like to thank Eric Hittinger for sharing the genetic code used for model fitting. Funding was provided by the U.S. Department of Defense Strategic Environ-
mental Research and Development Program (SERDP) under project WP-1626 and by the U.S. Environmental Protection Agency National Center for Environmental Research (NCER) through the STAR program (R833748).

References


Dzepina, K., Volkamer, R. M., Madronich, S., Tulet, P., Ulbrich, I. M., Zhang, Q., Cappa, C. D., Ziemann, P. J., and Jimenez, J. L.: Evaluation of recently-proposed secondary or-


Table 1. SOA mass-yields ($\alpha_i$) for model precursors in the Traditional (extended) model.

<table>
<thead>
<tr>
<th>Group</th>
<th>$C^*$ (µg m$^{-3}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1</td>
</tr>
<tr>
<td>BALK</td>
<td>0.001</td>
</tr>
<tr>
<td>ALK4</td>
<td>0.001</td>
</tr>
<tr>
<td>ALK5</td>
<td>0.001</td>
</tr>
<tr>
<td>ARO1</td>
<td>0.001</td>
</tr>
<tr>
<td>ARO2</td>
<td>0.007</td>
</tr>
</tbody>
</table>

Note: OLE1, OLE2, ISOP, SESQ and TERP are the same as in Murphy and Pandis (Murphy and Pandis, 2010).
Table 2. SOA mass yields for model precursors in the Volatility-based model.

<table>
<thead>
<tr>
<th>Precursor $C^*$ ($\mu g m^{-3}$)</th>
<th>Product $C^*$ ($\mu g m^{-3}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1</td>
</tr>
<tr>
<td>$10^2$</td>
<td>0.431</td>
</tr>
<tr>
<td>$10^3$</td>
<td>0.210</td>
</tr>
<tr>
<td>$10^4$</td>
<td>0.142</td>
</tr>
<tr>
<td>$10^5$</td>
<td>0.097</td>
</tr>
<tr>
<td>$10^6$</td>
<td>0.011</td>
</tr>
<tr>
<td>$10^7$</td>
<td>0.000</td>
</tr>
<tr>
<td>$10^8$</td>
<td>0.000</td>
</tr>
<tr>
<td>$10^9$</td>
<td>0.000</td>
</tr>
<tr>
<td>$10^{10}$</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Fig. 1. Schematics that demonstrate the (a) Empirical, (b) Traditional and (c) Volatility-Based SOA models.
Fig. 2. SOA predictions/fits from the (a) Empirical, (b) Traditional (speciated), (c) Traditional (all) and (d) Traditional (extended) models compared to measurements. The fractional error (f.e.) and fractional bias (f.b.) are mentioned in parentheses. FT: Fischer-Tropsch, JP-8: Jet-Propellant 8.
**Fig. 3.** (a) SOA fits from the volatility-basis-set model compared to measurements. (b) SOA yield presented as a function of precursor $C^*$ at a COA of 5 µg m$^{-3}$. The yellow band represents fits for the volatility-basis-set model. $n$-alkane data is from Presto et al. (2010), c-alkane and iso-alkane data is from Tkacik et al. (2012), biogenic data is from Farina et al. (2010), alkene data is from Forstner et al. (1997), Na et al. (2006) and Keywood et al. (2004) and aromatic data is from Ng et al. (2007), Song et al. (2007), Hildebrandt et al. (2009), Chan et al. (2009) and Shakya et al. (2010). $C^*$ values are determined either from the NIST database or EPA’s Estimation Program Interface suite.
Fig. 4. Fractional error and fractional bias plotted for different models used in this study.