



## 32 **1 Introduction**

33 Driven by the need to simulate secondary organic aerosol (SOA) formation for  
34 Chemical Transport Models (CTMs), the results of “smog” chamber studies of SOA  
35 formation have been extensively parameterized. Early smog chamber experiments were  
36 translated into aerosol mass fractions (yields) for models by assuming a constant aerosol  
37 mass fraction for each precursor (Hatakeyama et al., 1989) or assuming independent  
38 condensation of each SOA component when its concentration exceeded its saturation  
39 value (Pandis et al., 1992). With additional chamber experiments and theoretical  
40 investigation, it was demonstrated that SOA aerosol mass fractions of the various organic  
41 precursors are not constant, but depended on aerosol concentration in a manner consistent  
42 with equilibrium gas-particle partitioning (Odum et al., 1996; Pankow, 1994a; Pankow,  
43 1994b).

44 With the completion of more chamber studies and the wider acceptance of  
45 absorptive partitioning theory for organic aerosol formation, a database of empirical  
46 concentration-dependent aerosol mass fractions was developed, and implemented in  
47 CTMs (Griffin et al., 1999; Lurmann et al., 1997; Odum et al., 1997; Strader et al., 1999).  
48 SOA studies have been reviewed by (Kanakidou et al., 2005; Seinfeld and Pankow,  
49 2003).

50 Prior to about 2004, the main type of chamber SOA formation experiment gave  
51 one aerosol mass fraction (aerosol formed / hydrocarbon reacted) after several hours of  
52 oxidation and aerosol formation. These “final aerosol mass fraction” experiments were  
53 limited in time resolution, usually by the time required to measure the precursor  
54 concentration by GC (aerosol concentration could be monitored at minute time scales

55 using Scanning Mobility Particle Sizers). While real-time monitoring of VOC levels  
56 during SOA formation was available using long path FTIR, this required high  
57 concentrations of reactants. Now that Proton Transfer Reaction Mass Spectrometry  
58 (PTRMS) is in use in several chambers, dynamic concentrations can be measured  
59 generating dozens of data points for each chamber experiment (Lee et al., 2006; Ng et al.,  
60 2006; Presto et al., 2005). The applicability of dynamic aerosol mass fractions (AMF) can  
61 be evaluated by comparison with fixed AMF (Pathak et al., 2007). Other experiments or  
62 series of experiments have been conducted to establish the relationship between  
63 temperature and SOA formation. These include series of experiments with similar  
64 hydrocarbon concentrations, but different temperatures (Pathak et al., 2007; Takekawa et  
65 al., 2003) as well as formation at a fixed temperature followed by heating or cooling of  
66 the chamber inside a temperature-controlled room or enclosure (Stanier et al., 2007).  
67 This work attempts to synthesize all of these types of data to inform the SOA formation  
68 parameterization.

69 Smog chamber experiments done at one temperature give no information about  
70 the temperature dependence of SOA AMF. Product yields may vary as a function of  
71 temperature and the vapor pressures of individual products will certainly vary with  
72 temperature. In the latter case, lower temperatures should favor higher yields.  
73 Temperature dependence is typically included in CTMs by calculating temperature-  
74 dependent saturation concentrations (or partitioning coefficients) according to the  
75 Clausius-Clapeyron equation (Pankow, 1994a; Pankow, 1994b), using enthalpy of  
76 evaporation  $\Delta H$  for representative semivolatile species. This leads to a strong link  
77 between temperature dependence of SOA concentrations and the  $\Delta H$  values used.

78 Measurements of the temperature dependence of SOA AMF are now becoming available.  
79 One goal of this work is to present a method for including the temperature dependence  
80 results into parameter fitting, and thus reduce CTM uncertainty due to lack of knowledge  
81 of the appropriate  $\Delta H$  (Pun et al., 2003; Tsigaridis and Kanakidou, 2003).

82 Two other methods for the determination of partitioning coefficients and AMF  
83 should be noted. The first is the direct measurement of gas and particle phase  
84 concentrations through filter/denuder GC/MS techniques (Kamens and Jaoui, 2001; Yu et  
85 al., 1999). The other is a combined kinetic and thermodynamic approach where the time  
86 series of aerosol mass and individual products is used to constrain a combination of rate  
87 constants, stoichiometric yields, and gas-particle partitioning parameters (Kamens and  
88 Jaoui, 2001). To date, these methods have been useful at elucidating gas-phase kinetic  
89 mechanisms and product distributions, but have not been used to inform CTM modeling  
90 of gas-aerosol partitioning. As emphasized in recent overview articles on SOA  
91 formation, representation of the full complexity of SOA formation and aging will require  
92 a combination of thermodynamic approaches (as emphasized in this work) as well as  
93 kinetic descriptions of SOA processes (Donahue et al., 2006; Kroll et al., 2007).

94 Figure 1 illustrates many of the concepts to be explored in this paper. The figure  
95 was created by simulating the oxidation of a VOC producing two semivolatile  
96 compounds, with temperature independent stoichiometric yields. The absorptive  
97 partitioning equations were solved to determine the fraction of product in the gas and  
98 aerosol phases. The aerosol phase portion contributes to the aerosol mass fraction shown  
99 as the z axis. The goal of this work is to determine how to parameterize the AMF surface  
100 from smog chamber experimental data, including traditional final AMF, fixed-

101 temperature experiments, variable temperature experiments, and dynamic chamber  
102 results. Also noted in Figure 1 is the possibility that experimental data is not available  
103 across the concentration and temperature regime where the models operate. Uncertainty  
104 bounds are critical for meaningful extension of the experimental results to CTM  
105 applications. The uncertainty bounds need to be valid for interpolation within the  
106 experimental data, and as much as possible, for extrapolation outside of the experimental  
107 regime.

108         The goal of this work is to refine the estimation of parameters from chamber  
109 experiments using absorption partitioning theory. In section 2.1, absorptive partitioning  
110 is reviewed. In section 2.2, the parameterization algorithm is presented. Section 2.3  
111 reviews the types of chamber data that are available for fitting. Section 2.4 defines  
112 several terms needed for discussion of the fitting and describes four sets of pseudodata  
113 used in testing the fitting algorithm. Sections 3.1-3.4 describe the application of the  
114 fitting procedure to the pseudodata sets, while section 3.5 applies it to a combined  $\alpha$ -  
115 pinene ozonolysis dataset including multiple chambers and multiple authors.

116         This work does not consider the parameterization of aging, heterogeneous  
117 reactions, or oligomerization. It is limited to “prompt” aerosol mass formation from  
118 chamber experiments, which capture chemistry and partitioning occurring over  
119 approximately a 10-300 minute timescale. The choice to leave these more recently  
120 discovered processes out of this work is not due to a belief that they are negligible or that  
121 they do not need to be eventually incorporated into CTMs. However, the choice is  
122 justifiable on at least two grounds: (1) a sound procedure for parameterization of prompt  
123 SOA formation and its uncertainty is a necessary building block for more advanced

124 schemes; and (2) widely used CTMs (e.g. CMAQ, CAMx, and GEOS-CHEM) share the  
125 absorptive partitioning framework, and thus can be modified without changing their basic  
126 structure for increased accuracy and for prediction of uncertainties on SOA formation due  
127 to limitations in the underlying chamber data.

128

## 129 **2 Methods**

### 130 *2.1 Theoretical Basis for Semi-Empirical Fitting Equations*

131 Absorptive SOA partitioning is used throughout this work, with the following  
132 nomenclature. The effective saturation concentration  $c_i^*$  is defined through the  
133 expression:

$$134 \quad y_i c_i^* \equiv c_{i,gas} \quad 1$$

135 where  $c_{i,gas}$  is the gas phase mass concentration of species  $i$  and  $y_i$  is its mass fraction in  
136 the absorbing aerosol phase ( $y_i = c_{i,aer} / c_{OA}$ ). The variables  $c_{i,aer}$  and  $c_{OA}$  are the aerosol  
137 mass concentration of species  $i$ , and the total absorbing organic aerosol concentration,  
138 respectively. This is equivalent to relating effective saturation concentration to pure  
139 component vapor pressure using:

$$140 \quad c_i^* = \gamma_i c_i^o \frac{M_{OA}}{M_i} = \frac{\gamma_i p_{L,i}^o}{RT} M_{OA} \quad 2$$

141 where  $M_i$  is the molecular weight of species  $i$ ,  $M_{OA}$  is the average molecular weight of the  
142 absorbing aerosol phase,  $c_i^o$  is the saturation concentration and  $p_{L,i}^o$  is the subcooled  
143 liquid vapor pressure of  $i$ . We note two advantages for this treatment of partitioning for  
144 CTM parameterizations: first, the activity coefficient is included in the effective  
145 saturation concentration, and therefore the mixture is treated as pseudoideal; and second,

146 the equations do not require explicit terms for molecular weight ratios of the precursor  
 147 and products, and of the various products and the average organic aerosol molecular  
 148 weight.

149 The mass fraction of species  $i$  partitioned to the aerosol phase,  $\xi_i$ , for a single  
 150 semivolatile compound in absorption partitioning is:

$$151 \quad \xi_i \equiv \frac{c_{aer,i}}{c_{tot,i}} = \frac{1}{1 + \frac{c_i^*}{c_{OA}}} \quad 3$$

152 where  $c_{tot,i}$  is the total concentration (gas + aerosol) of species  $i$ . Therefore, for a mixture  
 153 of semivolatile compounds, the total organic aerosol concentration is (Donahue et al.,  
 154 2006):

$$155 \quad c_{OA} = \sum c_{tot,i} \xi_i = \sum \frac{c_{tot,i}}{1 + \frac{c_i^*}{c_{OA}}} \quad 4$$

156 In equations 1-4, we have not made any distinction as to the source of organic material  
 157 (either  $c_{tot,i}$  or  $c_{OA}$ ) – it can be either preexisting in the atmosphere, or it can be generated  
 158 from reactions or from combustion emissions. Smog chambers are a special case, where  
 159 typically all the organic mass (for all species) is generated from oxidation of the parent  
 160 reactive organic gas (ROG). Some chamber experiments are conducted with preexisting  
 161 semivolatiles in the gas or aerosol phase, but they are the minority and would require  
 162 minor changes to the fitting equations and parameterization algorithm. In cases where  
 163  $c_{tot,i}$  is solely the result of a reaction of a precursor gas:

$$164 \quad c_{tot,i} = \alpha_{i,molar} \Delta ROG M_i / M_{ROG} = \alpha_{i,mass} \Delta ROG \quad 5$$

165 In this work, all  $\alpha_i$  values are mass-based yields. Expressing equations 3-5 for a  
 166 smog chamber experiment with no preexisting organic aerosol (such that  $\Delta c_{OA} = c_{OA}$ ):

167 
$$\zeta \equiv \frac{\Delta c_{OA}}{\Delta ROG} = \sum \frac{\alpha_i}{1 + \frac{c_i^*}{c_{OA}}} \quad 6$$

168 where  $\zeta$  is the overall aerosol mass fraction. Equation 6 suggests that the aerosol mass  
 169 fraction can exceed unity, owing to increase in molecular weight upon reaction. The  
 170 partitioning equations (1-4) are applicable to all atmospheric absorptive partitioning  
 171 calculations, while the smog chamber equations (5-6) should be reserved for cases where  
 172 the assumption of zero preexisting adsorbing aerosol is met.

173 The term aerosol mass fraction ( $\zeta$ ) is used throughout instead of the term yield  
 174 (yield  $\equiv \Delta c_{OA}/\Delta ROG$ ) common in SOA literature. This is done to reserve the term yield  
 175 for product yields ( $\alpha_i$ ) and avoid confusion between product yields  $\alpha_i$  and aerosol mass  
 176 fraction (AMF or  $\zeta$ ). For comparison to previous works, the substitution yield  $\equiv$   
 177  $\Delta c_{OA}/\Delta ROG = \zeta$  can usually be made. As pointed out by Presto and Donahue (2006)  
 178 there is uncertainty on the density of organic aerosols. Aerosol mass concentration is  
 179 often measured by combining the measured aerosol number distribution by a scanning  
 180 mobility particle sizer with the assumption of sphericity and an assumed density (often  
 181 unity). Thus, application of equation 6 to both atmospheric prediction and to  
 182 experimental data reduction requires careful treatment of density. The relevant density  
 183 correction equations are in the appendix. In this work, all yields are calculated using  
 184 assumption of aerosol density of  $1 \text{ g cm}^{-3}$ .

185 The partition coefficient  $K_{om,i}$  (units  $\mu\text{g}^{-1}\text{m}^3$ ) can be used interchangeably with the  
 186 effective saturation concentration, with the equation for interconversion as:

187 
$$K_{om,i} \equiv \frac{c_{aer,i}}{c_{gas,i} c_{OA}} = \frac{1}{c_i^*} \quad 7$$



188 A typical procedure (Cocker et al., 2001; Griffin et al., 1999; Odum et al., 1996) for data  
189 reduction of chamber experiments is to use equation 6 to regress smog chamber data  
190 (pairs of  $c_{OA}$  and  $\zeta$ ) to fit either 2 parameters (e.g., a single  $\alpha_i$  and saturation concentration  
191  $c_i^*$ ) or 4 parameters (e.g., a pair of  $\alpha_i$  values with corresponding saturation  
192 concentrations).

193 When aerosol mass fractions are needed at different temperatures, the Clausius-  
194 Clapeyron equation is used to adjust the effective saturation concentrations.

$$195 \quad c_i^*(T) = c_i^*(T_{ref}) \frac{T_{ref}}{T} \exp \left[ \frac{\Delta H_i}{R} \left( \frac{1}{T_{ref}} - \frac{1}{T} \right) \right] \quad 8$$

196 where  $T_{ref}$  is a reference temperature where a reference effective saturation is defined, and  
197  $\Delta H_i$  is the enthalpy of evaporation. There is considerable uncertainty regarding the value  
198 of  $\Delta H$  to use in this equation (Bian and Bowman, 2002; Stanier et al., 2007; Strader et al.,  
199 1999; Tsigaridis and Kanakidou, 2003). One approach (which usually gives to strong of  
200 a temperature response) is to use the few known  $\Delta H$  values of identified SOA  
201 components.

202

## 203 2.2 *Parameterization Algorithm*

204 The parameterization strategy is to determine values of  $c_i^*$ ,  $\alpha_i$ , and  $\Delta H_i$  so that  
205 aerosol mass fractions (AMF) predicted by equations 6 and 8 match experimental values  
206 as closely as possible. This is done by nonlinear least squares regression. The actual  
207 regression is straightforward, but there are several challenging issues related to the  
208 parameterization: (1) selecting the number of basic set compounds ( $n$ ) to use, (2) deciding  
209 whether to use fixed or variable effective saturation values, (3) estimating the

210 temperature sensitivity accurately (or at all), and (4) estimating confidence intervals  
211 across the range of  $c_{OA}$  and temperature values anticipated in the CTM application. An  
212 overview of the procedure is shown in Figure 2.

213

### 214 2.2.1 Selection of extreme values of $c^*$

215 Equations for selecting extreme values of  $c^*$  are derived in the appendix. They depend  
216 on the accuracy goals, organic aerosol concentrations, and temperatures of the expected  
217 CTM application. For  $c^*$  values referenced to 298 K (25°C) the equations are:

$$218 \quad c_{\min,298K}^* \approx \delta \exp\left[\frac{\Delta H}{R}\left(\frac{1}{T_{\max}} - \frac{1}{298}\right)\right] \quad 9$$

$$219 \quad c_{\max,298K}^* \approx \frac{5}{f} c_{OA,\max} \exp\left[\frac{\Delta H}{R}\left(\frac{1}{T_{\min}} - \frac{1}{298}\right)\right] \quad 10$$

220 where  $\delta$  is the allowable error ( $\mu\text{g m}^{-3}$ ) in the semivolatile partitioning calculation under  
221 clean conditions,  $T_{\min}$  and  $T_{\max}$  are the temperature limits of CTM modeling application,  
222 and  $f$  is the maximum fractional error in the parameterization at high concentration (e.g.  
223 0.1 for  $\pm 10\%$  allowable organic aerosol concentration). Some values from equations 9  
224 and 10 are given in Table 1 for  $\Delta H / R$  of 12,000 K. For the  $\alpha$ -pinene cases examined in  
225 the results and discussion section, it turns out that this choice is overly conservative, and  
226 a value of 4,000 K will suffice.

227

### 228 2.2.2 Selection of number of $c^*$ values

229 A priori specification of the number of  $c^*$  values (including  $c_{\min}^*$  and  $c_{\max}^*$ )  
230 required for a suitable fit and for confidence interval characterization is difficult. A  
231 practical approach is to space the  $c^*$  values on a lognormal basis (e.g. 0.01, 0.1, 1, ...  $10^3$ ,

232  $10^4, 10^5 \mu\text{g m}^{-3}$ ) and thus the number of elements in the basis vector is determined by the  
 233 spacing between  $c^*_{min}$  and  $c^*_{max}$ . The regression calculation is quick and can be repeated  
 234 easily with a different basis, so trial and error selection of the  $c^*$  basis set is feasible.  
 235 From doing the fits for this work, we recommend:

$$236 \quad n \approx a \log_{10} \left( c^*_{max} / c^*_{min} \right) + 1 \quad 11$$

237 where  $a$  is from 0.5 to 1.0. There exist cases where the *experimental data* can be fit with  
 238 a small basis set (e.g.  $n = 1$  or  $2$ ). This will be true especially when the data have a small  
 239 dynamic range in terms of  $c_{OA}$  values. When equation 11 gives a large recommended  
 240 basis set but one or two well selected  $c^*$  values are sufficient to fit the experimental data,  
 241 the additional parameters are still necessary for calculating uncertainty when the  
 242 parameterization is extrapolated outside the range of the underlying experimental data.

243 Although the current work is limited to fitting AMF (and not a full mass balance  
 244 for the SOA formation chemical reaction) the use of a suitably large basis set may have  
 245 further utility in tracking the mass balance, secondary reactions, aging, and atmospheric  
 246 fate of biogenic and anthropogenic ROG emissions (Donahue et al., 2006; Robinson et  
 247 al., 2007).

248

### 249 2.2.3 Regression

250 The objective function minimized during regression is

$$251 \quad OF = \sum_{i=1..m} \left[ \xi - \hat{\xi}(\alpha_i, \Delta H, c^*_i(T_{ref}), T_{ref}, c_{OA}, T) \right]^2 \quad 12$$

252 where  $\xi$  are the measured AMF ( $m$  data points) and  $\hat{\xi}$  are the modeled AMF calculated  
 253 using equations 6 and 8. The modeled AMF are calculated using fixed parameters ( $T_{ref}$ ),

254 and experimental variables  $c_{OA}$  and  $T$ . The adjustable parameters are  $\alpha_i$ , a single  $\Delta H$   
 255 (instead of  $n \Delta H_i$  values). Fixed saturation concentrations  $c_i^*$  are used in this work, but  
 256 they can also be adjustable parameters if  $n$  is small.

257 The fact that the AMF is linear in the adjustable parameters  $\alpha_i$  (Equation 6)  
 258 greatly speeds up the calculations. At any step in the regression (and for any value of  
 259  $\Delta H$ ), the best set of  $\alpha_i$  can be computed by solving the following equation subject to yield  
 260 non-negativity constraints:

$$261 \quad \begin{bmatrix} \frac{c_{OA,1}}{c_{OA,1} + c_1^*} & \frac{c_{OA,1}}{c_{OA,1} + c_2^*} & \dots & \frac{c_{OA,1}}{c_{OA,1} + c_n^*} \\ \vdots & \vdots & & \vdots \\ \frac{c_{OA,m}}{c_{OA,m} + c_1^*} & \frac{c_{OA,m}}{c_{OA,m} + c_2^*} & \dots & \frac{c_{OA,m}}{c_{OA,m} + c_n^*} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_m \end{bmatrix} \quad 13$$

$$262 \quad \alpha_i \geq 0$$

263 Because of this feature, nonlinear minimization is performed for only  $\Delta H$ . In this  
 264 work, the Levenberg-Marquardt method implemented in MATLAB's *lsqnonlin* function  
 265 (MathWorks, 2002) is used for the nonlinear minimization and the MATLAB function  
 266 *lsqnonneg* is used for the nonnegative least squares problem. The regression calculation  
 267 is not sensitive to initial guesses and takes a few seconds for fixed  $c_i^*$  values, with  $n=8$   
 268 and hundreds of datapoints. The goodness of fit will be characterized using the metrics  
 269 shown in Table 2.

270 Initial confidence intervals are calculated using the asymptotic confidence interval  
 271 method (Seber and Wild, 2003) as implemented in MATLAB's *nlpredci* function. The  
 272 uncertainty bounds are calculated *without* using the constraints that there are physical

273 upper and lower limits to AMF. Therefore, far enough from the data, the inherent  
274 multicollinearity in the basis set manifests itself as unreasonably high and low values of  
275 the confidence intervals. These are truncated using the logic that AMF cannot decrease  
276 with increasing  $c_{\text{OA}}$  (at constant temperature), that AMF cannot increase with decreasing  
277  $c_{\text{OA}}$  (at constant temperature), and that AMFs have both lower (zero) and upper ( $\sim 1.5$ )  
278 limits.

279 This approach often leads to wide nonphysical confidence intervals when  
280 extrapolating outside the experimental range. Therefore, a second method of calculating  
281 confidence intervals is employed: repeated random (Monte Carlo) sampling of selected  $\alpha_i$   
282 values. The basic idea is that the experimental data may poorly constrain certain yields  
283 associated with low volatility (low  $c^*$ ) and high volatility (high  $c^*$ ) elements in the basis  
284 set. Therefore, the yields for these compounds are selected randomly while the  
285 remaining  $\alpha_i$ 's are fit to the experimental data. The algorithm for the Monte-Carlo  
286 confidence intervals is included in an appendix. Random sampling is done in large  
287 ( $\sim 1000$ ) simultaneous trials with either 1, 2 or 3 random  $\alpha_i$ 's. Random sampling is  
288 stopped when change in the confidence intervals is smaller than a preset tolerance. This is  
289 the slowest part of the overall parameterization code. 1000 trials require  $\sim 1$  second using  
290 MATLAB's *lsqnonneg* routine on 2.1 MHz PC with 2 GB RAM (this is while estimating  
291 confidence intervals with 300 points in the  $c_{\text{OA}}$  vs. T space). The trials need to be  
292 repeated from 2-20 times for each value of  $\Delta H$  explored, so exploration of the confidence  
293 intervals using 10 values of  $\Delta H$  can take 1-2 CPU minutes. The combination of these  
294 techniques for different cases will be discussed in subsequent sections.

295

296 2.3 *Description of Data Used in Fitting*

297 Data used in this work to parameterize SOA AMF fall into three categories. The  
298 first category is the traditional fixed-temperature smog chamber AMF experiment. In  
299 these experiments, the AMF of secondary organic aerosol (defined as the mass of created  
300 aerosol divided by the mass of reactant consumed) is measured after the reactive organic  
301 gas is oxidized in a large chamber. The relative humidity and concentration of  
302 preexisting aerosol is controlled. Temperature may be controlled or uncontrolled,  
303 although a single temperature is usually recorded corresponding to the temperature at the  
304 time when the AMF estimate is made. A series of experiments is usually completed  
305 according to a design to provide variation in reactant concentration,  $c_{OA}$ , and or  
306 temperature. Oxidation is by OH, O<sub>3</sub>, or NO<sub>3</sub>. O<sub>3</sub> can be injected directly from a corona  
307 discharge source, or photochemical oxidation can be initiated from NO<sub>2</sub> photolysis. If  
308 ozonolysis is the targeted process, an OH scavenger such as 2-butanol can be included in  
309 the chamber in excess. As discussed in the review by Kanakidou et al. (2005) (Kanakidou  
310 et al., 2005), there is a large body of literature data based on experiments of this type for  
311 various precursor-oxidant combinations. These types of experiments are referred to as  
312 static, final AMF, or traditional smog chamber experiments.

313 A second category of data available for fitting is from chamber experiments  
314 where VOC concentration can be monitored rapidly using PTRMS and aerosol  
315 concentrations are measured at minute time scale by SMPS (Presto and Donahue, 2006;  
316 Presto et al., 2005). Therefore, a time series of  $c_{OA}$ ,  $\Delta ROG$ , AMF, and temperature is  
317 generated.

318 A third category of data available for fitting is from chamber experiments that  
319 include a change in temperature following aerosol formation. This is done solely for the  
320 purpose of assessing the temperature sensitivity of aerosol partitioning. Rather than  
321 compare two experiments at different temperatures (each with independent uncertainties  
322 that complicate attributing changes in AMF to  $\Delta T$ ), the ramp or step change in  
323 temperature allows direct observation of the change in partitioning. The technique is  
324 called TREVA (Temperature-Ramped Equilibrium Volatility Analysis). Results are  
325 available only for the  $\alpha$ -pinene-ozone and  $\beta$ -pinene-ozone systems (Pathak et al., in press,  
326 2007; Pathak et al., 2007; Stanier et al., 2007).

327 Data from all three of these experiment types is used in the final section of the  
328 results and discussion to parameterize AMF and uncertainty for  $\alpha$ -pinene ozonolysis.  
329 The specific datasets are listed and referenced in Table 3. Individual datapoints used in  
330 fitting are listed in a supplemental data section.

331

#### 332 2.4 Nomenclature and Description of Synthetic Data Used in Fitting

333 Synthetic data, also called pseudodata, is useful for demonstrating that the fitting  
334 procedures work and assessing specific strengths and weaknesses of the parameterization  
335 strategy. Values of  $\alpha_i$ ,  $c^*_i$ , and  $\Delta H_i$  are used to simulate AMF experiments, generating  
336 synthetic (pseudo) data as datapoints with values of  $c_{OA}$ ,  $\xi$ , and  $T$ . These can be  
337 simulated with random error and bias if desired. We will use the following terminology:

- 338 • *Underlying model*: the  $\alpha_i$ ,  $c^*_i$ , and  $\Delta H_i$  used to generate synthetic data
- 339 • *Ideal AMF*:  $\xi$  calculated from the underlying model
- 340 • *Synthetic data or pseudodata*: data ( $\xi$ , temperature, and  $c_{OA}$ ) used in regression

- 341 • *Predictions or regression predictions*: AMF ( $\hat{\xi}$ ) calculated according to the  
 342 regression output. For pseudodata cases, regression predictions are compared to  
 343 synthetic data, where agreement is expected, and to ideal data, where agreement is  
 344 not necessarily expected and will depend on many factors.
- 345 • *Regression model*: the structure (equation 6) and parameters for  
 346  $\hat{\xi}(\alpha_i, \Delta H, c_i^*(T_{ref}), T_{ref}, c_{OA}, T)$ . The adjustable parameters are  $\alpha_i$  ( $n$  values) and  $\Delta H$ .  
 347  $T_{ref}$  is a fixed parameter (set to 298 K for this work).  $c_{OA}$  and  $T$  are the  
 348 independent experimental variables.
- 349 • *Basis set or basis vector*: the fixed values of  $c_i^*$  used in a regression model.
- 350 • *Parameterization applicability domain (also model domain)*: the expected range  
 351 of  $c_{OA}$  and temperatures where the parameterization will be employed.
- 352 • *Experimental domain*: range of  $c_{OA}$  and temperatures in the experimental or  
 353 synthetic data.
- 354 • *Temperature sensitivity*: the change in aerosol concentration with temperature due  
 355 to semivolatile partitioning, expressed as  $\partial \ln(c_{OA}) / \partial T$  with units of  $K^{-1}$ . For SOA  
 356 from  $\alpha$ -pinene ozonolysis, this is expected to be  $\sim -0.004$  to  $-0.036 K^{-1}$  (Stanier et  
 357 al., 2007). For consistency, all temperature sensitivity values in this work are  
 358 calculated at  $40 \mu g m^{-3}$  and  $20^\circ C$ .

359 The ideal models and synthetic data experiments used in this work are described in Table  
 360 4. Two underlying models are used to generate four sets of pseudodata.

361 A 2-product underlying model is used to generate sets A, B1, and B2. In  
 362 pseudodata set A, the sequence of hypothetical experiments includes 19 chamber final  
 363 AMF experiments, with experimental  $c_{OA}$  values from  $7-525 \mu g m^{-3}$  and temperatures



364 from 26 to 37°C. Experiments are assumed to have very low independent random error  
365 in determination of  $\Delta$ ROG such that the coefficient of variation ( $CV=\sigma/\text{mean}$ ) on  $\Delta$ ROG  
366 from repeated smog chamber experiments would be 0.007.

367 In pseudodata set B1, 21 final AMF experiments are simulated and they are  
368 assumed to come from 3 separate series of chamber experiments. The first series of  
369 experiments is 10 experiments, all at 26°C. They are assumed to have random error such  
370 that  $CV=0.02$ , but they are also assumed to have a series-specific bias of +10% in  
371 measured AMF (e.g. true AMF of 0.2 would be sampled as 0.22). The second series of  
372 experiments is 6 experiments, all at 37°C. They are assumed to have random error such  
373 that  $CV=0.02$ , but they are also assumed to have a series-specific bias of -10% in  
374 measured AMF. A third series of experiments is 5 experiments at the same fixed  $\Delta$ ROG,  
375 but each at a different temperature, from 17-37°C. They are assumed to be free of  
376 systematic error but have random error such that  $CV=0.02$ .

377 In pseudodata set B2, the pseudodata experiments from B1 are supplemented with  
378 5 temperature ramp experiments, where the AMF is measured at 5 different temperatures  
379 as the chamber is heated or cooled. The data is assumed to be noisy, and to have  
380 correlated errors. Specifically, each of the 5 experiments is assumed to have a bias  
381 shared by all 5 datapoints, associated with the determination of the  $\Delta$ ROG for that  
382 experiment. This is generated randomly such that the  $CV=0.06$ . The details of the  
383 hypothetical experiments are that each experiment gives 5 AMF datapoints at  
384 temperatures ranging from 22 to 39°C. Each datapoint was generated with an additional  
385 independent error, corresponding to a  $CV=0.03$ .

386 In pseudodata set C, a much more complex physical mixture is assumed, with 14  
387 products with  $c^*$  values ranging from  $\sim 10^{-2} \mu\text{g m}^{-3}$  to  $\sim 10^4 \mu\text{g m}^{-3}$ . No single SV  
388 component was dominant, but the most prominent three had yields ( $\alpha_i$ ) of 0.09, 0.042,  
389 and 0.03 at  $c^*$  values of  $10^{-0.2}$ ,  $10^{0.7}$  and  $10^{2.6} \mu\text{g m}^{-3}$ , respectively. Two of the fourteen  
390 had yields of less than 0.005. A full list is given in table 4. The pseudodata comprising  
391 dataset C are sampled in a very similar sequence of experiments as in set A. 19 simulated  
392 experiments, with low random error ( $CV=0.007$ ), no systematic error, spanning min/max  
393  $c_{OA}$  values of 7 and  $525 \mu\text{g m}^{-3}$ , and including temperatures from 26 to  $37^\circ\text{C}$ .

394

### 395 **3 Results and Discussion**

#### 396 *3.1 Ability to Fit Synthetic Data*

397 Following the flowchart in figure 2, there are 3 separate steps before any  
398 calculations begin. The first requirement is experimental data. For this demonstration,  
399 pseudodata set A is used, where 19 pseudo-experimental data points have been sampled  
400 from an underlying 2 product model (see also section 2.4).

401 The second requirement is a basis vector of  $c^*$  values to be used in the fit; the  
402 basis vector needs to be consistent with the CTM model domain, as described in section  
403 2.2. For this demonstration, we assume the model domain is:  $0-40^\circ\text{C}$ ;  $c_{OA,max}$  of  $25 \mu\text{g m}^{-3}$ ,  
404 acceptable error ( $\delta$ ) of  $0.07 \mu\text{g m}^{-3}$  in SOA partitioning under very clean conditions,  
405 and an acceptable error of  $1.25 \mu\text{g m}^{-3}$  ( $f \cdot c_{OA,max}$  with  $f=0.05$ ) in SOA partitioning under  
406 very polluted conditions. This set of values (listed in Table 1) requires (equation 9 and  
407 10)  $c^*_{min}$  of  $0.01 \mu\text{g m}^{-3}$  and  $c^*_{max}$  of  $10^5 \mu\text{g m}^{-3}$ . Equation 11 puts  $n$  at 4-8. The choice of  
408 8 corresponds to a decadal spacing ( $0.01, 0.1, 1, 10, 100, 10^3, 10^4, 10^5 \mu\text{g m}^{-3}$ ).

409           The third requirement is to form vectors to represent the model domain as a  $p \times j$   
410 grid of  $c_{OA}$  and temperature values. The grid needs to have enough points so that  
411 curvature in the AMF surface can be suitably represented. In this work, we use 20 values  
412 of  $c_{OA}$  ( $p=20$ ) and repeated all calculations at 5 different temperatures (0, 10, 20, 30 and  
413 40°C). Therefore, an output of the algorithm will be AMF estimates and confidence  
414 intervals at the  $p \times j$  points. Finally, a vector of  $\Delta H$  values is required for probing the  
415 uncertainty limits. In this work, we test values ranging from 10 to 120  $\text{kJ mol}^{-1}$  in 5-15  $\text{kJ}$   
416  $\text{mol}^{-1}$  increments. For most datasets, several of the extreme values cannot fit the data  
417 under any choice of  $\alpha_i$ . Therefore, for a final fit, a narrower range of more tightly spaced  
418  $\Delta H$  vector may give the widest uncertainty bounds.

419           With those three items set, the optimization and the determination of the  
420 confidence intervals can proceed. The result for the basis  $c^*$  vector (0.01, 0.1, 1, 10, 100,  
421  $10^3$ ,  $10^4$ , and  $10^5 \mu\text{g m}^{-3}$ ) is shown in Figure 3. The regression lines (thick solid lines) go  
422 through the data very well (the mean absolute fractional error is less than 1%). The  
423 uncertainty bounds (grey) are narrowest where there is the most data (30°C) and become  
424 wider away from the data. The underlying model included two SOA species with  $\Delta H$   
425 values of 80 and 60  $\text{kJ/mol}$ . The regression estimated a  $\Delta H$  value of 77  $\text{kJ mol}^{-1}$ . The  
426 fitted  $\alpha$  values were (0; 0; 0.027; 0.097; 0; 0.069; 0.25; 0.25). The ideal AMF (narrow  
427 dotted lines) are encompassed within the confidence intervals for nearly the entire  
428 parameterization applicability domain (88 percent of the ideal AMF fall within the  
429 confidence intervals). The uncertainty bounds are output not as a parameterized function,  
430 but rather as an array with a value for the upper and lower CI at each of the  $p \times j$  aerosol-  
431 temperature combinations. The confidence intervals are small in parts of the model

432 domain, reflecting the high precision in the pseudodata. In Figure 3, and in all  
433 subsequent figures of the same style, the experimental data is shifted to the nearest  
434 temperature line using a preset temperature sensitivity of  $-0.015 \text{ K}^{-1}$ . For fitting, the data  
435 is fit at its actual temperature – the shift is only so that the data can be summarized in one  
436 AMF vs.  $c_{OA}$  plot.

437

### 438 *3.2 Exploring Goodness of Fit vs. Basis Set Selection*

439 To explore the relationship between the choice of the basis set and the quality of  
440 the parameterization, we now repeat the fit of pseudodata set A with different basis  
441 vectors, varying the number of elements in the basis vector, their spacing, and the  
442 extreme values of the basis set ( $c^*_{min}$  and  $c^*_{max}$ ). The pseudodata are unchanged; only the  
443 basis vectors are changed. The choices are then compared on the basis of goodness-of-  
444 fit, fraction of data within confidence intervals, and confidence interval width.

445 The results are shown in Figures 4 and 5. Five different fits with basis vector  
446 length 2 (n=2 fits) were done. These are denoted in Figure 4 as entries 2a-2e. Four  
447 different n=3 fits were included in the comparison (3a-3d), as well as 2 n=4 fits (4a, 4b).  
448 Single fits were done at n=5, 6, 7, 8, and 9, respectively. The n=5-7 cases had lognormal  
449 spacing between 0.01 and  $1,000 \mu\text{g m}^{-3}$ . The n=8 fit is repeated from section 3.1. The  
450 n=9 fit has lognormal spacing from 0.01 to  $c^*_{max}$  to  $10^5 \mu\text{g m}^{-3}$ . Figure 4 has performance  
451 statistics for all 16 of these cases. Figure 5 shows AMF vs.  $c_{OA}$  plots for 4 of the 16  
452 cases.

453 On the y axis of figure 4a is a goodness-of-fit of the regression prediction relative  
454 to the synthetic data (black bars) and to the ideal AMF (white bars). The plotted error is

455 the mean error fraction,  $\text{err}_{\text{frac}}$  (see table 3). For example, a prediction of 0.18 AMF  
456 versus a data point of 0.20 is expressed as a 10% error in Figure 4a. Figure 4b graphs the  
457 percentage of the true AMF across the entire domain that fall within the confidence  
458 intervals. Figure 4c shows the width of the confidence intervals at 4 specific points in  
459 the system. One point ( $30^{\circ}\text{C}$ ,  $26 \mu\text{g m}^{-3}$ , triangle symbols) is right in the heart of the  
460 pseudodata. One would expect the narrowest CI for that portion of the regression.  
461 Another, ( $30^{\circ}\text{C}$ ,  $7 \mu\text{g m}^{-3}$ , circle symbols) is within the data, but just barely. Two points  
462 represent the extrapolation in temperature ( $10^{\circ}\text{C}$ ,  $26 \mu\text{g m}^{-3}$ ) and in temperature and  
463 concentration ( $10^{\circ}\text{C}$ ,  $7 \mu\text{g m}^{-3}$ ).

464 Several conclusions can be drawn from Figures 4 and 5. First, low  $n$  (2-3  
465 components in the basis vector) fits are variable in goodness-of-fit relative to the data.  
466 As shown in Figure 4a (fits 2a-3d), some basis sets lead to good fits, while other fail  
467 badly. On the other hand, using an  $n=2$  basis vector for estimating confidence intervals  
468 on the regression gives poor results regardless of the  $c^*$  values. The fraction of ideal  
469 model points falling in the confidence intervals (Figure 4b, fits 2a-2e) is fairly low, while  
470 the size of the confidence intervals is highly dependent on the goodness of fit, varying  
471 from unreasonably narrow (2a) to very wide (2c). The variability in the regression results  
472 is reinforced by the  $n=2$  and  $n=3$  panels of Figure 5; the goodness-of-fit and the width of  
473 the confidence intervals show extreme variability, depending on the exact choice.

474 In case 2a, the  $c^*$  values selected correspond exactly to the values in the  
475 underlying model. Therefore, the data are fit very well, giving the lowest value of the  
476 objective function of all the cases in figures 4 and 5. However, the excellent fit coupled

477 with the fact that there are no poorly constrained regression parameters gives very narrow  
478 confidence intervals.

479 The second point to draw from figures 4 and 5 is that the regression results  
480 improve in quality and consistency when a basis vector with  $n \geq 6$  and appropriate  $c^*_{\min}$   
481 and  $c^*_{\max}$  are used. The exact choice of individual  $c^*$  values becomes much less  
482 important as  $n$  increases to values of 6 and higher. In all the  $n \geq 6$  cases, the errors  
483 relative to the data and pseudodata are small. The fraction of ideal points within the  
484 confidence intervals are reasonable, and the uncertainty in the prediction is smallest in the  
485 heart of the data, and grows larger as the degree of extrapolation increases. AMF vs.  $c_{\text{OA}}$   
486 plots can be seen with uncertainty bounds for the  $n=6$  case (Figure 5) and the  $n=8$  case  
487 (Figure 3).

488

### 489 *3.3 Temperature Sensitivity and $\Delta H$*

490 In some fitting applications the data will constrain the temperature sensitivity of  
491 aerosol concentrations. In other words, in a plot such as Figure 3, the AMF will be  
492 equally well predicted at a range of different temperatures. Pseudodata set A is, for the  
493 most part, able to constrain the change in AMF with temperature. For example, the  $n=7$   
494 regression recovers a temperature sensitivity of  $-0.022 \text{ K}^{-1}$ . The temperature sensitivity  
495 of the underlying model at the same point is  $-0.021 \text{ K}^{-1}$ . The reason for this good  
496 agreement is that the pseudodata includes data from a range of temperatures, there is no  
497 correlation between error and temperature, and random errors are small. The fitted  $\Delta H$   
498 value ( $77 \text{ kJ mol}^{-1}$ ) is in agreement with the underlying model (60 and  $80 \text{ kJ mol}^{-1}$ ), as  
499 explained in section 3.1. Furthermore, in pseudodata set A, the  $\Delta H$  value is tightly

500 constrained, with a standard error of  $\sim 3 \text{ kJ mol}^{-1}$  and visible deterioration of the goodness  
501 of fit when  $\Delta H$  is reduced from 77 to 60.

502 In still other scenarios the data will *not* constrain temperature sensitivity. This  
503 can occur when all data were sampled within a narrow temperature range. In that case,  
504 the confidence intervals will be narrow at temperatures with data, and increasingly wide  
505 at other temperatures. Or the cause can be excessive random error in the data masking  
506 any temperature signal. This should lead to wide confidence intervals at all temperatures.  
507 Another possibility is that correlations between error and temperature lead to an *incorrect*  
508 apparent temperature sensitivity in the data.

509 To illustrate the case of incorrect apparent temperature sensitivity, a pseudodata  
510 set B1 was generated. In pseudodata B1, a hypothetical series of experiments at 29 °C  
511 has AMF 10% higher than their true values (e.g. 0.11 instead of 0.10). Another  
512 hypothetical series of experiments at 37 °C reports AMF 10% lower than their true  
513 values. The underlying model has SOA  $\Delta H$  values at 80 and 60 kJ/mol, and the true  
514 underlying model temperature sensitivity is  $-0.021 \text{ K}^{-1}$ . The regression result is shown in  
515 Figure 6a. As expected, the regression predictions are incorrect in terms of temperature  
516 sensitivity; the recovered  $\Delta H$  value is too low ( $40 \text{ kJ mol}^{-1}$ ) and the temperature  
517 sensitivity of the regression prediction is  $-0.019 \text{ K}^{-1}$ .

518

### 519 3.3.1 Using “Temperature Ramp” Chamber Experiments

520 One method of overcoming poorly constrained temperature sensitivity is through  
521 repeated smog chamber experiments at different temperatures (Pathak et al., 2007).  
522 However, the number of experiments may need to be fairly large ( $>15$ ) to overcome

523 random error between chamber experiments. Another type of experiment design is a  
524 temperature ramp design, where a temperature controlled smog chamber is heated or  
525 cooled quickly (minutes) after initial SOA formation (Stanier et al., 2007). This has the  
526 advantage of minimizing the effect of experiment-to-experiment noise in measuring the  
527 temperature sensitivity. This type of experiment was included in pseudodata set B2. Set  
528 B2 is identical to B1 except that 5 temperature ramp experiments, each generating 5  
529 datapoints, are added to the 19 experiments already in set B1. Figure 6b shows how the  
530 regression result when set B2 is fit. The temperature ramp experiments, although noisy  
531 compared to the original data, meet the objective of improving the estimation of the  
532 temperature sensitivity. The regression shown in figure 6b has a temperature sensitivity  
533 of  $-0.022 \text{ K}^{-1}$  and a regressed  $\Delta H$  value of 67 kJ/mol.

534         Similar results to figure 6 were obtained by reversing the biases in the 29 and  
535 37°C data series. In the above paragraphs and in Figure 6, the lower temperature data  
536 was biased high while the higher temperature data was biased low, leading to an apparent  
537 low temperature sensitivity. If the biases are reversed, the apparent temperature  
538 sensitivity becomes large compared to the ideal temperature sensitivity of the underlying  
539 model. For example, with the true temperature dependence the same as above ( $-0.021 \text{ K}^{-1}$ ),  
540 the initial regression's value is  $-0.034 \text{ K}^{-1}$ . Once the temperature ramp data is added,  
541 the sensitivity goes to  $-0.024 \text{ K}^{-1}$ .

542         Volatility TDMA experiments (An et al., 2007; Offenbergl et al., 2006; Philippin  
543 et al., 2004) may also be suitable for this type of analysis, as long as equilibrium is  
544 | achieved in the flow setup.

545



### 546 3.3.2 User Specified Limits on Temperature Sensitivity

547 If the CTM modeling results are sensitive to temperature effects on partitioning,  
548 then realistic temperature sensitivity and confidence intervals are highly desirable. If no  
549 experimental data is available as a constraint, then suggested limits are (based on  $\alpha$ -  
550 pinene SOA)  $\partial \ln(c_{OA})/\partial T$  of 0 to  $-0.04 \text{ K}^{-1}$  (Stanier et al., 2007). Bracketing temperature  
551 sensitivity is preferred over the simpler approach of just limiting the  $\Delta H$  value used  
552 equation 6 to a preset range. The reason for this is that the value of  $\Delta H$  required for a  
553 certain temperature sensitivity depends on the basis vector selection and aerosol  
554 concentrations. At least two methods exist for using artificial limits on temperature  
555 sensitivity. One is to simply check confidence intervals for adherence to the user selected  
556 temperature sensitivity, and to narrow confidence intervals appropriately. The other is to  
557 include artificial data in the fitting itself.

558

### 559 *3.4 Pseudodata Set C – Complex Mixture*

560 The underlying model for sets A and B was relatively simple, with only 2 SOA  
561 products. Actual organic aerosols, even for single precursors, are much more complex.  
562 In case C, the SOA mixture is assumed to have 14 components with  $c^*$  values ranging  
563 from  $\sim 10^{-2} \mu\text{g m}^{-3}$  to  $\sim 10^4 \mu\text{g m}^{-3}$ . No changes to the algorithm or procedures are  
564 required. The result of the fitting algorithm is shown in Figure 7. A basis vector of 8  
565 lognormally space  $c^*$  values (from 0.01 to  $10^5 \mu\text{g m}^{-3}$ ) was used to perform the  
566 regression. The fit results (error relative to data, relative to underlying model,  
567 temperature sensitivity, and confidence intervals) are quite good. The mean absolute  
568 fractional error for prediction of the datapoints is less than 0.02. All of the ideal model

569 values are within the confidence intervals. The underlying model temperature sensitivity  
570 is  $-0.026 \text{ K}^{-1}$  and the regression temperature sensitivity is  $-0.022 \text{ K}^{-1}$  (with a fitted  $\Delta H$  of  
571  $56 \text{ kJ mol}^{-1}$ ).

572 One-to-one mapping of the ideal to the fitted yields ( $\alpha$  values) is not expected at  
573 the individual basis vector level. Nevertheless, broad agreement is seen when ideal and  
574 fitted yields are compared. When mapped to the mapped to the nearest  $c^*$  value, the true  
575 yields would be: [ 0.002 0 0.04 0.07 0.04 0.11 0.02 0 ]. The corresponding fitted values  
576 are: [ 0 0 0.09 0.03 0.09 0.14 0 0 ]. Grouping into four bins instead of eight, the  
577 comparison is closer: true yields at [ 0.002 0.11 0.15 0.02 ] and fitted yields at [ 0 0.12  
578 0.14 0 ].

579

### 580 *3.5 Fitting $\alpha$ -pinene ozonolysis data*

581 We now demonstrate the fitting algorithm on real data. The fitting is very similar  
582 to the pseudodata cases presented above. The main differences are that the dataset is  
583 larger, that the data somewhat noisier, and that new types of possible input data are  
584 included (final AMF experiments, dynamic AMF experiments using PTRMS, and  
585 temperature ramp experiments).

586

#### 587 3.5.1 Demonstration algorithm on large $\alpha$ -pinene ozonolysis dataset

588 The data used for fitting is summarized in Table 3, and individual datapoints are  
589 listed in the appendix. For the model applicability domain the “general” model case from  
590 Table 1 was selected ( $0.1\text{-}60 \mu\text{g m}^{-3}$  with a temperature range from  $0\text{-}40^\circ\text{C}$ ). While table  
591 1 (based on equations 9-11) recommends a  $c^*_{\text{min}}$  and  $c^*_{\text{max}}$  of  $1 \times 10^{-3}$  and  $2.4 \times 10^5$ ,

592 respectively, these are calculated based on a single component with  $\Delta H$  of  $100 \text{ kJ mol}^{-1}$   
593 (see appendix for derivation), and the real aerosol does not have nearly that temperature  
594 sensitivity. Therefore, a narrower range might be employed. This necessary range was  
595 tested by fitting with basis vectors of length 9 ( $10^{-3}$  to  $10^5$ ), length 7 ( $10^{-2}$  to  $10^4$ ), and  
596 length 5 ( $10^{-1}$  to  $10^3$ ). The fits and confidence intervals were nearly identical for the  $n=7$   
597 and  $n=9$  cases. However, the confidence intervals change somewhat in the  $n=5$  case.  
598 Specifically, the confidence intervals narrowed at low and high concentrations from the  
599 removal of the  $0.01$  and  $10,000 \mu\text{g m}^{-3}$  entries in the basis set. Therefore, all of the fits in  
600 section 3.5 utilize an  $n=7$  basis set with decadal  $c^*$  values from  $10^{-2}$  to  $10^4 \mu\text{g m}^{-3}$ .  
601 Because many of the experimental  $c_{OA}$  values are greater than the upper limit of the  
602 model applicability domain, the plots for this section will be extended to  $200 \mu\text{g m}^{-3}$ .

603 The fitting result for the  $n=7$  case is shown in Figure 7. The mean absolute  
604 fractional error is 0.08. In other words, the average error on an AMF of 0.2 is  $\pm 0.016$ .  
605 From the plot, it can be seen that the source of this fitting error is the variability in the  
606 data itself. There is obvious variability in the experiments, and some bias between the  
607 different series of experiments. The yield values (for  $c^* = 10^{-2}$  to  $10^4 \mu\text{g m}^{-3}$ ) are [0 0  
608 0.073 0.072 0.087 0.29 0.29] where the 0.29 values at  $\alpha_6$  and  $\alpha_7$  are user-selected upper  
609 limits (the data is not sufficient to constrain the most volatile component yields). Fitted  
610  $\Delta H$  is  $33 \text{ kJ mol}^{-1}$  and temperature sensitivity was  $-0.02 \text{ K}^{-1}$ . The largest confidence  
611 intervals occur under two conditions: at less than  $1 \mu\text{g m}^{-3}$ , and also at temperatures  
612 below  $20^\circ\text{C}$ . The cause of both of these areas of large uncertainty is a relative scarcity of  
613 data.

614 3.5.2 Removing correlated data in highly time-resolved data

615 The use of PTRMS to track the reacting VOC concentration, and the use of SMPS  
616 or aerosol mass spectrometers can generate multiple data points per experiment. The two  
617 PTRMS experiments fit in this section have 38, and 89 datapoints, respectively. The  
618 Durbin-Watson test, a test for serial correlation in the residuals, is performed by first  
619 fitting a basis vector of length 7 to the data for the PTRMS experiment in question only.  
620 The residuals from this fit are used in the Durbin-Watson test (Hamilton, 1992). In one  
621 of the 2 cases, the 38 residuals were determined to not be correlated. In the other, the  
622 Durbin-Watson test did indicate excessive correlation. This was removed by averaging  
623 successive data points (starting with those contributing most to the Durbin-Watson score)  
624 until either the test was passed, or the number of data points was reduced to 15.

### 625 3.5.3 Response of confidence intervals to key data

626 Figure 9 shows the result when only a subset of data is fit. The data that is used is  
627 that of Cocker, 2001. This data was selected because all of the samples are at  $29\pm 1^\circ\text{C}$ .  
628 Therefore, not surprisingly, extrapolation to different temperatures gives significant error.  
629 The best fit  $\Delta H$  and temperature sensitivity for this dataset are  $10 \text{ kJ mol}^{-1}$  and  $-0.010 \text{ K}^{-1}$ .  
630 The uncertainty bounds are determined by the range of allowed  $\Delta H$  values in the Monte  
631 Carlo sensitivity study. Those shown are for limits of 5 and  $120 \text{ kJ mol}^{-1}$ . A wider  
632 allowed range would give even wider confidence intervals for the temperatures other than  
633  $29\pm 1^\circ\text{C}$ . For example, setting  $\Delta H$  to  $120 \text{ kJ mol}^{-1}$  gives a temperature sensitivity of in  
634 this case at  $20^\circ\text{C}$  and  $40 \mu\text{g m}^{-3}$  of  $-0.046 \text{ K}^{-1}$ .

635 Two conclusions can be drawn from figure 9. First, the confidence interval  
636 algorithms used here respond in the correct qualitative fashion as key data are  
637 added/subtracted to the dataset. Second, extending the temperature range and/or

638 temperature sensitivity of AMF measurements is just as necessary for reduction in overall  
639 uncertainty as is extending the concentration range. Figures 8 and 9 are directly  
640 comparable and show the effect of the increase in the range of data on the uncertainty  
641 bounds.

642

### 643 *3.6 Discussion of application to Chemical Transport Modeling*

644 The paper and algorithm have focused on leaving the basis set selection flexible, to be  
645 varied as needed according to the demands of the expected model application (e.g.  $T_{\min}$ ,  
646  $T_{\max}$ ,  $c_{\text{OA},\min}$ ,  $c_{\text{OA},\max}$ , and allowable errors in partitioning). If followed literally, that idea  
647 could lead to a proliferation of data reductions, each tailored to a specific model domain;  
648 that is not a direction advocated by the authors. Rather, the key concepts from this work  
649 that should translate to improved modeling are (1) mismatches in the fitting basis vector  
650 and model organic aerosol concentrations and temperatures can have undesirable  
651 consequences; (2) realistic confidence intervals on aerosol mass fraction are possible; (3)  
652 multiple experiment types can be integrated to give improved parameterization; (4)  
653 comparisons of aerosol mass fractions (measurements, predictions, and confidence  
654 intervals) and temperature sensitivity are much more informative than comparisons of  
655 regression parameters ( $\alpha_i$ ,  $c^*_i$ , and  $\Delta H$ ); and (5) high  $n$  fixed  $c^*$  basis vector fits are  
656 feasible and useful.

657 A practical method of achieving the improved modeling results may be for wider  
658 adoption of a decadal basis set including "nonvolatile" ( $c^* < 0.1 \mu\text{g m}^{-3}$ ), "semivolatile"  
659 (SVOC;  $0.1 \mu\text{g m}^{-3} < c^* < 1000 \mu\text{g m}^{-3}$ ), and "intermediate-volatility" (IVOC;  $1000 \mu\text{g}$   
660  $\text{m}^{-3} < c^* < 100,000 \mu\text{g m}^{-3}$ ) organic compounds. In other words, the  $n=8$  basis set with

661  $c^*$  from  $10^{-2}$  to  $10^5 \mu\text{g m}^{-3}$ . In this work, the  $n=8$  set proves more than suitable for fitting  
662 the existing  $\alpha$ -pinene ozonolysis data, and for spanning simulated SOA partitioning data  
663 across realistic ranges of temperature and organic aerosol concentration.

664 For modeling applications where a reduced  $n$  is desirable, (e.g. a global model)  
665 the  $n=8$  regression results can be translated to a reduced basis set, to polynomial fits, or to  
666 lookup tables. For example,  $n=4$  may be an attractive treatment for online partitioning  
667 calculations in a large-scale CTM. With  $n=4$ , the choice of  $c^*_{\text{min}}$  and  $c^*_{\text{max}}$  values will be  
668 important in determining the performance of the partitioning fit, and use of equations 9  
669 and 10 is recommended for basis vector selection. Figure 4 showed that  $n=4$  basis  
670 vectors were not acceptable, but the specific  $n=4$  basis sets were not optimized for best  
671 performance in section 3.2. The basis vectors of  $[0.1 \ 1 \ 10 \ 100]$ ,  $[1 \ 10 \ 100 \ 10^3]$ , and  $[0.1$   
672  $2.2 \ 46 \ 1000]$  should be considered if an  $n=4$  representation is desired. It should be  
673 emphasized that an  $n=4$  basis set may be rich enough for fitting the experimental data and  
674 for the online partitioning calculations, but a basis vector with  $n>4$  may be needed for  
675 quantifying uncertainty limits during the data reduction step.

676

## 677 **Summary and Conclusion**

678 A method for fitting of smog chamber data for CTM applications was presented.  
679 The method uses a volatility basis set of compounds with fitted AMF and user-specified  
680 temperature-dependent saturation concentrations. A single effective enthalpy is fitted to  
681 match experimental and modeled temperature sensitivity of AMF. Important aspects of  
682 the method include: selection of minimum and maximum  $c^*$  values, selection of the

683 number of basis set compounds, and estimation of uncertainty bounds for the regression  
684 prediction. Main conclusions include:

- 685 • When fitting parameters to experimental smog chamber data, the range of  
686 effective saturation concentration ( $c^*$ ) values used must be consistent with the  
687 expected temperature and organic aerosol concentration range in the final model  
688 application. Equations for  $c^*_{min}$  and  $c^*_{max}$  are presented in section 2.2.
- 689 • The length of the basis vector (number of  $c^*$  values used,  $n$ ) can influence both  
690 the quality of the data fitting, and the ability to compute confidence intervals.  
691 Careful selection of  $c^*$  values is required with  $n \leq 5$  to insure a good fit and  
692 suitable confidence intervals. With 6 or more lognormally spaced  $c^*$  values, the  
693 best fit line and confidence intervals become insensitive to the exact  $c^*$  basis  
694 vector used.
- 695 • A method for determination of confidence intervals is demonstrated. It is useful  
696 to divide the confidence interval problem for AMF prediction from SOA chamber  
697 experiments into two different regimes: (1) confidence intervals for *interpolation*  
698 *within* experimental data; and (2) extrapolation *outside the data*. Interpolation CIs  
699 depend on the scatter in the experimental data itself.
- 700 • Although parts of the absorption SOA parameterization are nonlinear and/or  
701 implicit, the calculation of AMF is linear in the stoichiometric coefficients  $\alpha_i$ .  
702 Therefore, fitted  $\alpha_i$  values can be calculated using nonnegative linear least squares  
703 if the  $c_{OA,tot}$  and  $\Delta H$  are fixed. This makes the regression computationally  
704 efficient.

- 705 • In addition to traditional final AMF smog chamber experiments, newer dynamic  
706 aerosol formation experiments (e.g. PTRMS) and experiments with varied  
707 temperatures can be useful in constraining AMF parameterizations. Highly time-  
708 resolved data, such as a 1-min time series of AMF vs.  $\Delta$ ROG from a PTRMS and  
709 SMPS, may require averaging to reduce serial correlation between datapoints  
710 (discussed in section 3.5).
- 711 • Errors in apparent temperature dependence can be generated by correlated errors  
712 in measurements and temperature. This can be partially corrected with  
713 temperature ramp experiments that isolate the temperature sensitivity.
- 714 • Temperature-dependence of semivolatile aerosols is best dealt with by matching  
715 experimental and modeled temperature sensitivity ( $\partial \ln c / \partial T$ ). For the  $\alpha$ -pinene  
716 ozonolysis, the fitted temperature sensitivity at  $40 \mu\text{g m}^{-3}$  and  $20^\circ\text{C}$  is  $-0.02 \text{ K}^{-1}$   
717 and the associated  $\Delta H$  is  $33 \text{ kJ mol}^{-1}$ . Fitted  $\Delta H$  values depend on the basis vector  
718 (but temperature sensitivities do not); therefore, comparisons should be made of  
719 temperature sensitivity and not of  $\Delta H$ .
- 720 • In order to reduce uncertainty in CTM modeling, additional experimental data is  
721 needed at both lower concentrations *and* at a range of temperatures.
- 722 • Although a large number value of  $n$  ( $\sim 6$  or more) may be required for an initial  
723 characterization of confidence intervals, this large number of parameters does not  
724 necessarily have to be used in the CTM. In other words, the items of value to the  
725 CTM modeler are the aerosol mass fraction and uncertainty bounds themselves,  
726 not the parameters, data, and error-correlation matrices used to generate them.



727 • While each CTM modeling application may have a different optimal basis set, a  
728 practical method of achieving improved CTM modeling results may for wider  
729 reduction of experimental data to a basis set including "nonvolatile" ( $c^* < 0.1 \mu\text{g m}^{-3}$ )  
730  $\text{m}^{-3}$ ), "semivolatile" (SVOC;  $0.1 \mu\text{g m}^{-3} < c^* < 1000 \mu\text{g m}^{-3}$ ), and "intermediate-  
731 volatility" (IVOC;  $1000 \mu\text{g m}^{-3} < C^* < 100,000 \mu\text{g m}^{-3}$ ) organic compounds. In  
732 other words, the  $n=8$  basis set with  $c^*$  from  $10^{-2}$  to  $10^5 \mu\text{g m}^{-3}$ . The parameters  
733 can be reduced to a smaller set as needed for specific CTM applications as  
734 needed.

735

## 736 **Acknowledgements**

737 Although the research described in this manuscript has been funded wholly or in  
738 part by the United States Environmental Protection Agency through grant RD-83108101,  
739 it has not been subjected to the Agency's required peer and policy review and therefore  
740 does not necessarily reflect the views of the Agency and no official endorsement should  
741 be inferred.

742 **REFERENCES**

- 743 An, W.J., Pathak, R.K., Lee, B.H. and Pandis, S.N., 2007. Aerosol volatility  
744 measurement using an improved thermodenuder: Application to secondary  
745 organic aerosol. *Journal Of Aerosol Science*, 38(3): 305-314.
- 746 Bian, F. and Bowman, F.M., 2002. Theoretical method for lumping multicomponent  
747 secondary organic aerosol mixtures. *Environmental Science & Technology*,  
748 36(11): 2491-2497.
- 749 Cocker, D.R., Clegg, S.L., Flagan, R.C. and Seinfeld, J.H., 2001. The effect of water on  
750 gas-particle partitioning of secondary organic aerosol. Part I: alpha-pinene/ozone  
751 system. *Atmospheric Environment*, 35(35): 6049-6072.
- 752 Donahue, N.M., Robinson, A.L., Stanier, C.O. and Pandis, S.N., 2006. Coupled  
753 partitioning, dilution, and chemical aging of semivolatile organics. *Environmental*  
754 *Science & Technology*, 40(8): 2635-2643.
- 755 Edgerton, E.S. et al., 1999. Particulate Air Pollution in Mexico City: A Col laborative  
756 Research Project. *Journal Of The Air & Waste Management Association*, 49(10):  
757 1221.
- 758 Griffin, R.J., Cocker, D.R., Flagan, R.C. and Seinfeld, J.H., 1999. Organic aerosol  
759 formation from the oxidation of biogenic hydrocarbons. *Journal Of Geophysical*  
760 *Research-Atmospheres*, 104(D3): 3555-3567.
- 761 Hamilton, L.C., 1992. *Regression with Graphics*. Wadsworth, Inc., Belmont, CA.
- 762 Hatakeyama, S., Izumi, K., Fukuyama, T. and Akimoto, H., 1989. Reactions Of Ozone  
763 With Alpha-Pinene And Beta-Pinene In Air - Yields Of Gaseous And Particulate  
764 Products. *Journal Of Geophysical Research-Atmospheres*, 94(D10): 13013-  
765 13024.
- 766 Kamens, R.M. and Jaoui, M., 2001. Modeling aerosol formation from alpha-pinene plus  
767 NO<sub>x</sub> in the presence of natural sunlight using gas-phase kinetics and gas-particle  
768 partitioning theory. *Environmental Science & Technology*, 35(7): 1394-1405.
- 769 Kanakidou, M. et al., 2005. Organic aerosol and global climate modelling: a review.  
770 *Atmospheric Chemistry And Physics*, 5: 1053-1123.
- 771 Kroll, J.H., Chan, A.W.H., Nga, N.L., Flagan, R.C. and Seinfeld, J.H., 2007. Reactions of  
772 Semivolatile Organics and Their Effect on Secondary Organic Aerosol Formation.  
773 *Environmental Science & Technology*, 41(10): 3545-3550.
- 774 Lee, A. et al., 2006. Gas-phase products and secondary aerosol yields from the  
775 photooxidation of 16 different terpenes. *Journal Of Geophysical Research-*  
776 *Atmospheres*, 111(D17).
- 777 Lurmann, F.W. et al., 1997. Modelling urban and regional aerosols.2. Application to  
778 California's South Coast Air Basin. *Atmospheric Environment*, 31(17): 2695-  
779 2715.
- 780 MathWorks, 2002. *MATLAB*.
- 781 Ng, N.L. et al., 2006. Contribution of first- versus second-generation products to  
782 secondary organic aerosols formed in the oxidation of biogenic hydrocarbons.  
783 *Environmental Science & Technology*, 40(7): 2283-2297.
- 784 Odum, J.R. et al., 1996. Gas/particle partitioning and secondary organic aerosol yields.  
785 *Environmental Science & Technology*, 30(8): 2580-2585.

786 Odum, J.R., Jungkamp, T.P.W., Griffin, R.J., Flagan, R.C. and Seinfeld, J.H., 1997. The  
787 atmospheric aerosol-forming potential of whole gasoline vapor. *Science*,  
788 276(5309): 96-99.

789 Offenberg, J.H., Kleindienst, T.E., Jaoui, M., Lewandowski, M. and Edney, E.O., 2006.  
790 Thermal properties of secondary organic aerosols. *Geophysical Research Letters*,  
791 33, L03816, doi:10.1029/2005GL024623.

792 Pandis, S.N., Harley, R.A., Cass, G.R. and Seinfeld, J.H., 1992. Secondary Organic  
793 Aerosol Formation And Transport. *Atmospheric Environment Part A-General*  
794 *Topics*, 26(13): 2269-2282.

795 Pankow, J.F., 1994a. An Absorption-Model Of Gas-Particle Partitioning Of Organic-  
796 Compounds In The Atmosphere. *Atmospheric Environment*, 28(2): 185-188.

797 Pankow, J.F., 1994b. An Absorption-Model Of The Gas Aerosol Partitioning Involved In  
798 The Formation Of Secondary Organic Aerosol. *Atmospheric Environment*, 28(2):  
799 189-193.

800 Pathak, R.K. et al., in press, 2007. Ozonolysis of  $\alpha$ -pinene: parameterization of secondary  
801 organic aerosol mass fraction. *Atmospheric Chemistry And Physics*.

802 Pathak, R.K., Stanier, C.O., Donahue, N.M. and Pandis, S.N., 2007. Ozonolysis of alpha-  
803 pinene at atmospherically relevant concentrations: Temperature dependence of  
804 aerosol mass fractions (yields). *Journal Of Geophysical Research-Atmospheres*,  
805 112(D03201, doi:10.1029/2006JD007436).

806 Philippin, S., Wiedensohler, A. and Stratmann, F., 2004. Measurements of non-volatile  
807 fractions of pollution aerosols with an eight-tube volatility tandem differential  
808 mobility analyzer (VTDMA-8). *Journal Of Aerosol Science*, 35(2): 185-203.

809 Presto, A.A. and Donahue, N.M., 2006. Investigation of alpha-pinene plus ozone  
810 secondary organic aerosol formation at low total aerosol mass. *Environmental*  
811 *Science & Technology*, 40(11): 3536-3543.

812 Presto, A.A., Hartz, K.E.H. and Donahue, N.M., 2005. Secondary organic aerosol  
813 production from terpene ozonolysis. 2. Effect of NO<sub>x</sub> concentration.  
814 *Environmental Science & Technology*, 39(18): 7046-7054.

815 Pun, B.K. et al., 2003. Uncertainties in modeling secondary organic aerosols: Three-  
816 dimensional modeling studies in Nashville/Western Tennessee. *Environmental*  
817 *Science & Technology*, 37(16): 3647-3661.

818 Robinson, A.L. et al., 2007. Rethinking organic aerosols: Semivolatile emissions and  
819 photochemical aging. *Science*, 315(5816): 1259-1262.

820 Salcedo, D. et al., 2006. Characterization of ambient aerosols in Mexico City during the  
821 MCMA-2003 campaign with Aerosol Mass Spectrometry: results from the  
822 CENICA Supersite. *Atmospheric Chemistry And Physics*, 6: 925-946.

823 Seber, G.A.F. and Wild, C.J., 2003. *Nonlinear Regression*. John Wiley and Sons.

824 Seinfeld, J.H. and Pankow, J.F., 2003. Organic atmospheric particulate material. *Annu.*  
825 *Rev. Phys. Chem.*, 54: 121-140.

826 Stanier, C.O., Pathak, R.K. and Pandis, S.N., 2007. Measurements of the Volatility of  
827 Aerosols from alpha-Pinene Ozonolysis. *Environmental Science and Technology*,  
828 41(8): 2756-2763.

829 Strader, R., Lurmann, F. and Pandis, S.N., 1999. Evaluation of secondary organic aerosol  
830 formation in winter. *Atmospheric Environment*, 33(29): 4849-4863.

831 Takekawa, H., Minoura, H. and Yamazaki, S., 2003. Temperature dependence of  
832 secondary organic aerosol formation by photo-oxidation of hydrocarbons.  
833 Atmospheric Environment, 37(24): 3413-3424.  
834 Tsigaridis, K. and Kanakidou, M., 2003. Global modelling of secondary organic aerosol  
835 in the troposphere: a sensitivity analysis. Atmospheric Chemistry And Physics, 3:  
836 1849-1869.  
837 Yu, J.Z., Cocker, D.R., Griffin, R.J., Flagan, R.C. and Seinfeld, J.H., 1999. Gas-phase  
838 ozone oxidation of monoterpenes: Gaseous and particulate products. Journal Of  
839 Atmospheric Chemistry, 34(2): 207-258.  
840  
841

842  
843  
844

**Table 1** Minimum and maximum  $c^*$  values for some scenarios

Scenario	$c_{OA,max}$ ( $\mu\text{g m}^{-3}$ )	$\delta$ ( $\mu\text{g m}^{-3}$ )	f	$T_{min}-T_{max}$ ( $^{\circ}\text{C}$ )	$c^*_{min}(25^{\circ}\text{C})$ ( $\mu\text{g m}^{-3}$ )	$c^*_{max}(25^{\circ}\text{C})$ ( $\mu\text{g m}^{-3}$ )	recommended n
General	60	0.1	0.05	0 – 40	$1 \times 10^{-3}$	240,000	5-10
Remote with high altitude	5	0.05	0.05	-30 – 40	$7 \times 10^{-4}$	$4.5 \times 10^6$	6-11
Polluted urban	120	0.5	0.05	0 – 40	0.07	480,000	4-8
Proposed decadal range	25	0.07	0.05	0 – 40	0.01	100,000	4-8

845  
846  
847  
848

**Table 2.** Error Formulas Used for Goodness of Fit

Name & Symbol	Formula
Mean Squared Error ( $err_{mse}$ ) vs. Data	$\frac{1}{m} \sum (\xi_i - \hat{\xi}_i)^2$
Average Bias vs. Data ( $err_{mean}$ )	$\frac{1}{m} \sum (\xi_i - \hat{\xi}_i)$
Mean Absolute Fractional Error vs. Data ( $err_{frac}$ )	$\frac{1}{m} \sum \left( \frac{ \xi_i - \hat{\xi}_i }{\xi_i} \right)$

849  
850

**Table 3** List of data sources for  $\alpha$ -pinene SOA AMF

Study	Experiments Used for Fitting in This Work	$\alpha$ -pinene Concentration Range	Temperature Range	No. Experiments (for this analysis)
Griffin et al., 1999	Outdoor (dark) chamber reaction with O <sub>3</sub> in excess, OH scavenger, inorganic seed, and dry conditions	15-65 ppb	32-37 °C	6
Hoffmann et al., 1997	Outdoor (dark) chamber reaction with O <sub>3</sub> in excess and OH (no OH scavenger), inorganic seed, and dry conditions	88-154 ppb	46-50 °C	5
Cocker et al., 2001	Indoor chamber reaction with O <sub>3</sub> in excess or limiting, OH scavenger, RH < 20%; some seeded experiments and some unseeded	23-163 ppb	28-30 °C	15
Pathak et al., 2007	Indoor chamber reaction with O <sub>3</sub> in excess, OH scavenger. Seeded and unseeded experiments at a many temperature / concentration combinations	4-50 ppb	0-40 °C	31
Presto and Donahue, 2006	Indoor chamber reaction with O <sub>3</sub> in excess, low NO <sub>x</sub> . Experiments 6/14/05 and 6/28/05.	1.5-138 ppb	22 °C	2

851  
852  
853  
854

855  
856  
857

**Table 4.** Description of synthetic data experiments

Chemical and Physical Characteristics of Underlying SOA Model				Pseudodata Sets Generated from Each Underlying Model		
Description		Components				
		1-5	6-10	11-14		
2 SOA Products	$\log_{10}c^*$	0.699; 2.90			Pseudodata A, B1, B2 (see descriptions below)	
	$\alpha \times 100$	10; 15				
	$\Delta H$ (kJ mol <sup>-1</sup> )	80; 60				
14 SOA products	$\log_{10}c^*$	-1.75; -1.54; -0.436; -0.218; +0.60		0.702; 1.04; 1.56; 1.84; 2.05	2.58; 3.02; 3.24; 3.83	Pseudodata C
	$\alpha \times 100$	0.12; 0.06; 0.6; 3; 1.8	4.2; 1.2; 0.9; 2.4; 0.6	9.0; 1.2; 0.6; 2.4		
	$\Delta H$ (kJ mol <sup>-1</sup> )	120; 110; 90; 80; 60	45; 72; 45; 83; 100	40; 35; 40; 50		

858  
859  
860  
861  
862  
863  
864  
865  
866

Pseudodata:

- A. Hypothetical sampling by 19 chamber experiments. Very low sampling error and no bias. Experimental  $c_{OA}$  values from 7-525  $\mu\text{g m}^{-3}$  and temperatures from 26 to 37°C.
- B1. Hypothetical sampling by 3 series of chamber experiments. Series one: 10 experiments at 26°C with low error but 10% AMF bias high. Series 2: 6 experiments at 37°C with low error but 10% AMF bias low. Series 3: 5 experiments with constant reactant concentration, but with different temperature in each case (17-37°C); low error; no bias.
- B2. Same as B1 but adding a 4th series: 5 temperature ramp experiments so that 5 AMF measurements are taken during each experiment, with temperatures ranging from 22-39°C.
- C. Hypothetical sampling by 19 chamber experiments. Very low sampling error and no bias. Experimental  $c_{OA}$  values from 7-525  $\mu\text{g m}^{-3}$  and temperatures from 26 to 37°C. (Identical to Pseudodata A).

867 **Appendix A –AMF (yield) equations with density and preexisting organic aerosol**  
868 **considerations**

869  
870 Equation 6 is the commonly encountered fitting equation. Here we show that equation 6 is a  
871 special case of a more general set of mass balances for the semivolatile aerosol system. The  
872 equations in this appendix are necessary when accounting for experimental and simulation cases  
873 where the organic aerosol density is not assumed to be unity, and for cases with preexisting  
874 organic mass.

875 Starting from the following definition:

$$876 \quad c_{i,gas} \equiv y_i c_i^* = \frac{c_{aeri,i}}{c_{OA}} c_i^* \quad (A1)$$

877 the following expression can be derived by mass balance:

$$878 \quad \xi_i \equiv \frac{c_{aeri,i}}{c_{tot,i}} = \frac{1}{1 + \frac{c_i^*}{c_{OA}}} \quad (A2)$$

879 This can be used with the following definitions of AMF and reaction mass yield  $\alpha_i$ .

$$880 \quad \Delta c_{tot,i} \equiv \alpha_i \Delta ROG \quad (A3)$$

$$881 \quad AMF = \xi \equiv \frac{\Delta c_{OA}}{\Delta ROG} = \sum \frac{\alpha_i}{1 + \frac{c_i^*}{c_{OA}}} \quad (A4)$$

882 Using the above equations, we can write expressions for prediction of organic aerosol mass  
883 ( $c_{OA}$ ), species-specific partitioning values  $\xi_i$ , and smog chamber AMF as function of parameters  
884 such as  $\alpha_i$ ,  $\Delta ROG$ , and  $c_i^*$ . To keep the expressions general, we assume the density of the  
885 organic aerosol is unknown, and with preexisting organic aerosol.

886



887 We denote the values of variables corresponding to the real organic aerosol density as in the  
 888 above equations (no subscript). We denote values of variables under the case of assumed unit  
 889 organic aerosol density with a subscript  $\rho=1$  for physical variables such as  $\xi$  and  $c_{OA}$ . Preexisting  
 890 aerosol is denoted by the superscript P.

891 Expression for  $c_{OA}$ ,  $c_{OA,\rho=1}$ ,  $\xi$  and  $\xi_{\rho=1}$  are then:

$$892 \quad c_{OA} = c_{OA,\rho=1} \frac{\rho}{\rho_1} = \sum \xi_i c_{i,tot} = \sum \frac{c_{tot,i}}{1 + \frac{c_i^*}{c_{OA}}} = \sum \frac{c_{tot,i}}{1 + \frac{c_i^* \rho_1}{c_{OA,\rho=1} \rho}} = \sum \frac{c_{tot,i}}{1 + \frac{c_i^{*'}}{c_{OA,\rho=1}}} = \sum \frac{c_{tot,i}}{1 + \frac{c_i^{*'} \rho}{c_{OA} \rho_1}} \quad (A5)$$

893 where the definitions  $c_i^{*'} = c_i^* \rho_1 / \rho$  and  $\alpha_i' = \alpha_i \rho_1 / \rho$  are used. The 4 alternate versions of the  
 894 denominator in equation A5 can be made in all the cases below, and the different expressions are  
 895 not repeated. Separating the organic aerosol ( $c_{OA}$ ) and the total concentration of species i into  
 896 preexisting and reaction-generated parts, gives the equations for  $c_{OA}$  and AMF in the case of  
 897 preexisting aerosol and assumed unit density. An equation for AMF in the case of preexisting  
 898 aerosol is not written, since the organic aerosol concentration can change from partitioning of  
 899 preexisting and reaction-generated species.

$$900 \quad c_{OA} = \sum \xi_i c_{i,tot} = \sum \frac{c_{tot,i}^P + \Delta c_{tot,i}}{1 + \frac{c_i^*}{c_{OA}}} = \sum \frac{c_{tot,i}^P + \alpha_i \Delta ROG}{1 + \frac{c_i^*}{c_{OA,\rho=1} \rho / \rho_1}} = \sum \frac{c_{tot,i}^P + \alpha_i' \rho / \rho_1 \Delta ROG}{1 + \frac{c_i^{*'} \rho / \rho_1}{c_{OA}}} \quad (A6)$$

901 In the case of no preexisting aerosol:

$$902 \quad \frac{c_{OA,\rho=1}}{\Delta ROG} = \xi_{\rho=1} = \frac{\rho_1}{\rho} \sum \frac{\alpha_i}{1 + \frac{c_i^* \rho_1}{c_{OA,\rho=1} \rho}} = \sum \frac{\alpha_i'}{1 + \frac{c_i^{*'}}{c_{OA,\rho=1}}} \quad (A7)$$

903 and 
$$\xi = \frac{\rho}{\rho_1} \xi_{\rho=1} = \frac{\rho}{\rho_1} \sum \frac{\alpha'_i}{1 + \frac{c_i^*}{c_{OA, \rho=1}}} = \frac{\rho}{\rho_1} \sum \frac{\alpha'_i}{1 + \frac{c_i^* \rho}{c_{OA} \rho_1}} \quad (\text{A8})$$

904 and 
$$c_{OA} = \Delta ROG \frac{\rho}{\rho_1} \sum \frac{\alpha'_i}{1 + \frac{c_i^* \rho}{c_{OA} \rho_1}} \quad (\text{A9})$$

905 In this work, all aerosol concentrations and AMFs are calculated using unit density  
 906 conversions from SMPS aerosol volume, so the fitting really recovers the normalized values  
 907  $c_i^* = c_i \rho_1 / \rho$  and  $\alpha'_i = \alpha_i \rho_1 / \rho$  according to equation A7. To make concentration predictions for  
 908 organic aerosols with densities other than 1.0 requires the use of equation A9. At low aerosol  
 909 mass fractions, the terms cancel and correct AMF and  $c_{OA}$  predictions will be made by regressed  
 910 values of  $\alpha$ , regardless of the density assumption made. At high aerosol loadings, AMF and  $c_{OA}$   
 911 concentration predictions will be low for actual aerosol specific gravities  $> 1$ .

912

913 **Appendix B – Derivation of formulae for  $c^*_{\min}$  and  $c^*_{\max}$**

914

915 For any expected parameterization applicability domain (defined by  $c_{OA,\min}$ ,  $c_{OA,\max}$ ,  $T_{\min}$ ,  
 916 and  $T_{\max}$ ) the extreme values required of the effective saturation concentration ( $c^*_{\min}$  and  $c^*_{\max}$ )  
 917 are calculated as follows. For a single condensing component in the presence of  $c_{NV,\min}$   
 918 (nonvolatile, organic, and solution forming) preexisting aerosol, the mass balance *for the*  
 919 *semivolatile species* is:

$$920 \quad c_{aer,SV} = c_{tot,SV} - c_{gas,SV} = c_{tot,SV} - y_{SV}c^*_{SV} \quad (B1)$$

921 where  $c_{gas,SV}$  and  $c_{aer,SV}$  refer to the gas- and particle-phase concentrations of the semivolatile in  
 922  $\mu\text{g m}^{-3}$ . Adding in the preexisting nonvolatile organic, the total organic aerosol concentration is:

$$923 \quad c_{OA} = c_{aer,SV} + c_{NV,\min} = c_{tot,SV} - y_{SV}c^*_{SV} + c_{NV,\min} \quad (B2)$$

924 To determine  $c^*_{\min}$ , the lowest  $c^*$  necessary to achieve the required modeling accuracy goal  $\delta$ ,  
 925 we construct a scenario where a truly non-volatile species is represented in the model as having a  
 926 finite saturation concentration  $c^*_{\min}$ , therefore creating an error. The true aerosol concentration  
 927 in this case will be:

$$928 \quad c_{OA,true} = c_{tot,SV} + c_{NV,\min} \quad (B3)$$

929 while the modeled aerosol concentration will be:

$$930 \quad c_{error} = y_{SV}c^*_{\min} \leq \delta \quad (B4)$$

931 to keep this error less than a user-selected goal  $\delta$  :

$$932 \quad c^*_{\min} \leq \frac{\delta}{y_{SV}} \quad (B5)$$

933 If the organic aerosol consists only of the nonvolatile and semivolatile fractions, then we can  
 934 eliminate  $y_{SV}$  to give:

935 
$$c_{\min}^* \leq \delta \left( 1 + \frac{c_{NV,\min}}{c_{tot,SV}} \right) \quad (B6)$$

936 For a conservative estimate of  $c_{\min}^*$ , we assume  $c_{tot,SV} \gg c_{NV,\min}$ , to keep the right hand side as  
 937 small as possible. Furthermore, the calculation is assumed to apply at the highest temperature  
 938 applied in the model ( $T_{\max}$ ).

939 
$$c_{\min}^*(T_{ref}) = \delta \frac{T_{\max}}{T_{ref}} \exp \left[ \frac{\Delta H}{R} \left( \frac{1}{T_{\max}} - \frac{1}{T_{ref}} \right) \right] \quad (B7)$$

940 Putting in 100 kJ/mol, neglecting  $T_{\max}/T_{ref}$ , and with temperature in Kelvin:

941 
$$c_{\min,298K}^* \approx \delta \exp \left[ \frac{12000}{T_{\max}} - 40.27 \right] \quad (B8)$$

942 With  $\delta$  of  $1 \mu\text{g m}^{-3}$  and  $T_{\max}$  at  $40^\circ\text{C}$ ,  $c_{\min}^*$  is  $0.14 \mu\text{g m}^{-3}$ .

943 To derive  $c_{\max}^*$  we consider a case with non-volatile solution-forming aerosol (secondary or  
 944 primary)  $c_{NV,\max}$  and one condensing species with saturation concentration  $c_{\max}^*$ . The goal is  
 945 then to select  $c_{\max}^*$  sufficiently high so that material in this bin is not in the aerosol phase.

946 Donahue, Robinson et. al (2006) suggest 1000  $c_{NV,\max}$  which ensures that less than about 0.1% of  
 947 material in the  $c_{\max}^*$  bin partitions to the aerosol phase. A slightly more flexible approach is to  
 948 include the CTM modeler's tolerance for error in semivolatile partitioning under high

949 concentrations and low temperature. The mass balance for the semivolatile species is given by

950 B1. A relative accuracy goal, denoted by  $f$ , is defined. At the high end of the expected

951 concentration range in the model, the accuracy goal ( $\mu\text{g m}^{-3}$ ) is  $\pm f c_{oa,\max}$  (e.g.,  $f=0.1$  for  $\pm 10\%$

952 accuracy).

953 
$$c_{aer,SV} \leq f (c_{NV,\max} + c_{aer,SV}) \quad (B9)$$

954 If we assume that  $c_{aer,SV}$  is a small fraction of  $c_{NV,\max}$  then:

955 
$$c_{aer,SV} = \frac{c_{tot,SV}}{1 + \frac{c_{max}^*}{c_{NV,max}}} \leq f c_{NV,max} \quad (B10)$$

956 which can be solved for  $c_{max}^*$  yielding:

957 
$$c_{max}^* \geq \frac{c_{tot,SV}}{f} - c_{NV,max} \quad (B11)$$

958 The  $c_{tot,SV}$  term is difficult to estimate. It represents the total pool of semivolatile compounds  
 959 that need to be modeled as in the gas phase. One reasonable assumption is to assume for any  
 960 given location, that gas phase semivolatile concentrations are proportional to organic aerosol (or  
 961 vis versa), such that  $c_{tot,SV} = k c_{NV,max}$ . This simplifies B11 to:

962 
$$c_{max}^* \geq c_{NV,high} \left( \frac{k}{f} - 1 \right) \quad (B12)$$

963 If we further assume that k is at least unity (the mass of semivolatiles is at least the organic  
 964 aerosol concentration) and the desired fractional error is small, then:

965 
$$c_{max}^* \geq \frac{k}{f} c_{NV,high} \quad (B12)$$

966 where k is an unknown ratio of the semivolatiles gases to the organic aerosol. Taking values  
 967 from Mexico City, we find a total VOC concentration of  $\sim 2000 \mu\text{g m}^{-3}$  (Edgerton et al., 1999)  
 968 and an organic aerosol concentration of  $\sim 20 \mu\text{g m}^{-3}$  (Salcedo et al., 2006). If 5% of the VOC is  
 969 SV, then  $k = 5$  for Mexico City. Applying a temperature correction similar to the one for  $c_{min}^*$   
 970 (but in the other direction)

971 
$$c_{max}^*(T_{ref}) \geq \frac{k}{f} c_{NV,max} \frac{T_{min}}{T_{ref}} \exp \left[ \frac{\Delta H}{R} \left( \frac{1}{T_{min}} - \frac{1}{T_{ref}} \right) \right] \quad (B13)$$

972 Using  $T_{ref}$  of 25 °C,  $\Delta H$  of 100 kJ mol<sup>-1</sup>,  $k=5$ , and ignoring  $T_{min}/T_{ref}$ , equation B13 reduces to:

973 
$$c_{\max,298K}^* \approx \frac{5}{f} c_{NV,\max} \exp\left[\frac{12000}{T_{\min}} - 40.27\right] \quad (\text{B14})$$

974 If the fractional accuracy goal is 10%,  $c_{NV,\max}$  is  $100 \mu\text{g m}^{-3}$ , and  $T_{\min}$  is  $0^\circ\text{C}$  then  $c_{\max}^*(298 \text{ K}) \approx$   
975  $200,000 \mu\text{g m}^{-3}$ , or  $200 \text{ mg m}^{-3}$ . Thus the total multiplier to  $c_{NV,\max}$  in this case is 2000,  
976 comparable to the multiplier of 1000 in Donahue et al. (2006).

977 **Appendix C. Algorithm for the Monte-Carlo confidence intervals**

978 1. Create a vector  $\mathbf{c}_{OA,model}$  (length  $p \sim 20$ ) of aerosol concentrations, logarithmically spaced  
979 from  $c_{OA,min}$  to  $c_{OA,max}$  (the maximum and minimum expected aerosol concentrations in  
980 the modeling application).

981 2. Create a vector (length  $j$ ) of temperatures from  $T_{min}$  to  $T_{max}$ . In this work,  $j$  is selected for  
982 spacing between elements of  $10^\circ\text{C}$ .

983 3. Let the experimental yields be a vector  $\xi$ , and the fitted predictions at the experimental  
984 values of  $c_{OA}$  and  $T$  are the vector  $\hat{\xi} \pm \delta$  where the values  $\delta$  are from the confidence  
985 intervals of the asymptotic method (e.g., MATLAB `nlpredci` function). All these vectors  
986 ( $\xi$ ,  $\hat{\xi}$ , and  $\delta$ ) will have length of  $m$  corresponding to number of experiments.

987 4. The experimental data are shifted by  $\delta$  such that:  $\xi_{high} = \xi + \delta$  and  $\xi_{low} = \xi - \delta$ .

988 5. A reference error value, corresponding to how well the fit goes through the data, is noted:

989 
$$MSE_{ref} = \sum (\xi_i - \hat{\xi}_i)^2 / m$$

990 6. Initialize  $j$  upper confidence intervals (CI) vectors (each of length  $p$ ), one for each  
991 temperature. Set each element to zero.

992 7. Initialize  $j$  lower confidence interval vectors (each of length  $p$ ), one for each temperature.  
993 Set each element to a large value (e.g. 999).

994 8. Create a vector (length  $k$ ) of  $\Delta H$  values. In this work,  $k$  was usually selected for a  
995 separation of  $5 \text{ kJ mol}^{-1}$  and a range from 10 to  $120 \text{ kJ mol}^{-1}$ .

996 The outer loop (steps 9-16) is repeated  $4 \cdot k$  times. For each value of  $\Delta H$ , the four runs are  
997 for upper CI (allowing random selection of low volatility yields); lower CI (allowing  
998 random selection of low volatility yields); upper CI (allowing random selection of high  
999 volatility yields); and lower CI (allowing random selection of high volatility yields).

1000            OUTER LOOP

1001            9. Since for any iteration  $\Delta H$  is fixed, the left hand matrix of equation 13 (C) is calculated

1002            and stored. This speeds up the subsequent calculation of yields given any vector  $\alpha$ .

1003            10. The  $n$  yields  $\alpha$  are divided into two groups, those that will be selected randomly, and

1004            those that will be fit. If the iteration involves randomized low volatility yields, then some

1005            number of the lowest volatility species (e.g.  $\alpha_1$  and  $\alpha_2$ ) are selected randomly. If the

1006            iteration involves randomized high volatility yields, then some of the highest volatility

1007            species (e.g.,  $\alpha_{n-1}$  and  $\alpha_n$ ) are chosen randomly.

1008            In this work, the number of low and high volatility yields to be randomized were selected

1009            by the user on a case by case basis, with the goal of only selecting yields that are poorly

1010            constrained by the experimental data. The typical  $n=7$  case required randomization of  $\alpha_1$ ,

1011             $\alpha_2$  and  $\alpha_6, \alpha_7$ . Inspection of the error covariance matrix and the standard errors on the

1012            original fitted yields will help in selecting which yields to randomize. In the

1013            demonstration case of Figure 3 (section 3.1), the  $\alpha$ 's with the 5 largest standard errors

1014            were selected, or elements 1, 2 (uncertain low volatility compounds) and 6, 7 and 8

1015            (uncertain high volatility compounds). The range of experimental data can serve as a

1016            guide for which terms are poorly constrained. If a given  $c^*$  value is outside the range of

1017            experimental data (in this case from 7-525  $\mu\text{g m}^{-3}$ ), then the corresponding  $\alpha$  value should

1018            probably be randomized. Erring on the conservative side, and trying to select  $\alpha$  values

1019            randomly that are in fact well constrained by the data does not effect the quality of the

1020            result. Rather it adds run time as calculations are performed that are of no subsequent use

1021            in determining the CIs.



1022 11. Let the number of random  $\alpha$  values be  $q$  and the number to be fit equals  $n-q$ . Let the  
 1023 fixed yields be denoted as  $\mathbf{\alpha}_q$  and the variable yields as  $\mathbf{\alpha}_{n-q}$ . Also let the  $\mathbf{C}$  matrix (from  
 1024 equation 13) be split into an  $m \times q$  matrix  $\mathbf{C}_q$  corresponding to the columns of the fixed  
 1025 yields, and a matrix  $\mathbf{C}_{n-q}$  [size  $m \times (n-q)$ ] corresponding to the columns of the adjustable  
 1026 yields.

1027 INNER LOOP

1028 12. 1000 random combinations of the random  $\alpha$  values are chosen.

1029 13. For each of the 1000 cases, the remaining  $n-q$  yields are calculated by solving a  
 1030 modified version of equation 13. The adjustable yields  $\mathbf{\alpha}_{n-q}$  are determined by  
 1031 solving  $\mathbf{C}_{n-q}\mathbf{\alpha}_{n-q} = \xi^* - \mathbf{C}_q\mathbf{\alpha}_q$  with  $\mathbf{\alpha}_{n-q} \geq 0$  where  $\xi^*$  is  $\xi_{high}$  when calculating an  
 1032 upper CI and  $\xi^*$  is  $\xi_{low}$  when calculating a lower CI.

1033 14. For each of the 1000 cases, the mean squared error is calculated

$$1034 \quad MSE = \sum (\xi_i - \hat{\xi}_i(\mathbf{\alpha}_q, \mathbf{\alpha}_{n-q}))^2 / m.$$

1035 If  $MSE \leq MSE_{ref}$  then it is a suitably good fit. Fitted values  $\hat{\xi}$  are calculated at all  $j \times$   
 1036  $p$  combinations of temperature and organic aerosol concentration. For upper CI,  
 1037  $CI_{high} = \max(CI_{high}, \hat{\xi})$ , and for the lower CI,  $CI_{low} = \min(CI_{low}, \hat{\xi})$  where the  
 1038 comparison is on an element by element basis.

1039 15. If the confidence intervals moved more than a present tolerance, repeat the inner loop by  
 1040 going to step 12.

1041 16. Loop back to step 9 until all  $4k$  cases are completed.

1042 17. For each temperature of interest, upper and lower confidence interval vectors are now  
1043 calculated using the Monte-Carlo method. They should be inspected graphically and  
1044 compared to the limits from the asymptotic method.  
1045

Author's Identifier	$\Delta$ ROG (ppb)	$\Delta$ ROG ( $\mu\text{g m}^{-3}$ )	O <sub>3</sub> initial (ppb)	Temperature (°C)	RH	OH-Scavenger	Seed	Result	
								$\Delta$ M ( $\mu\text{g m}^{-3}$ )	AMF $\Delta$ M/ $\Delta$ ROG (%)
Single-temperature final AMF chamber experiments									
Griffin et al., 1999									
6/5/98a	16.7	89		36.8		2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	7.4	8.3
6/5/98b	18.2	97		36.8		2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	8.5	8.7
6/7/98a	31.0	169		30.2		2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	30.3	17.9
6/7/98b	45.5	249		30.2		2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	46.0	18.4
6/9/98a	57.0	307		34.8		2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	52.3	17.0
6/9/98b	65.0	350		34.8		2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	65.1	18.6
Hoffmann et al., 1997									
9/27A	88.0	454	210	48.1		none	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	82.0	18.0
9/27B	38.0	196	220	48.9		none	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	29.9	15.2
9/29A	101.9	795	320	47.2		none	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	80.4	23.0
9/29B	154.1	527	327	48.1		none	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	183.5	15.2
10/02A	132.9	690	210	46.1		none	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	94.9	13.7
Cocker et al., 2001									
11/29a	74.7	411	280	28.0	<2	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	82.0	20.0
11/29b	144.1	792	360	28.0	<2	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	190.0	24.0
12/03a	40.6	223	370	28.0	<2	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	34.0	15.2
12/03b	72.3	398	315	28.0	<2	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	79.0	19.8
12/06b	22.9	126	280	28.0	<2	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	10.0	7.9
01/08a	72.4	399	270	27.8	<2	2-butanol	none	80.0	20.1
01/10a	41.0	225	278	28.9	<2	2-butanol	none	35.0	15.5
01/10a	93.4	512	499	28.9	<2	2-butanol	none	110.0	21.5
01/13a	72.3	396	244	29.5	14.1	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	79.0	19.9
01/13b	118.4	648	352	29.5	14.4	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	145.0	22.4
01/14a	48.4	265	402	29.6	14.6	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	48.0	18.1
01/14b	98.8	541	313	29.6	16.6	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	115.0	21.3
01/18a	116.1	637	130	29.1	<2	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	140.0	22.0
01/22a	162.3	888	175	29.6	<2	2-butanol	none	220.0	24.8
01/22b	139.3	763	380	29.6	<2	2-butanol	none	185.0	24.3
Pathak et al., 2007									
1	14.3	76	250	40.0	<10	2-butanol	none	7.0	9.2
2	42.0	242	250	15.0	<10	2-butanol	none	50.7	21.0
3	38.3	217	250	20.0	<10	2-butanol	none	33.7	15.6
4	50.0	273	250	30.0	<10	2-butanol	none	48.7	17.8
5	50.0	265	250	40.0	<10	2-butanol	none	34.7	13.1
7	7.3	41	1500	20.0	<10	2-butanol	none	2.8	6.8
9	3.7	21	3100	20.0	<10	2-butanol	none	0.7	3.3
10	7.3	39	1500	40.0	<10	2-butanol	none	2.1	5.4
11	7.3	42	1500	15.0	<10	2-butanol	none	2.6	6.2
12	7.3	40	1500	30.0	<10	2-butanol	none	2.5	6.3
13	38.3	217	250	20.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	37.6	17.4
14	7.3	41	1500	20.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	2.9	7.0

Author's Identifier	$\Delta$ ROG (ppb)	$\Delta$ ROG ( $\mu\text{g m}^{-3}$ )	O <sub>3</sub> initial (ppb)	Temperature (°C)	RH	OH-Scavenger	Seed	Result	
								$\Delta$ M ( $\mu\text{g m}^{-3}$ )	AMF $\Delta$ M/ $\Delta$ ROG (%)
15	8.0	45	250	20.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	2.6	5.7
19	8.0	46	1500	15.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	3.7	8.0
20	7.3	39	1500	40.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	1.4	3.6
21	8.5	46	1500	30.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	3.4	7.3
22	14.3	82	250	15.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	7.2	8.8
23	14.3	82	300	15.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	9.3	11.3
24	14.3	81	250	20.0	<10	2-butanol	none	9.8	12.1
25	14.3	78	250	30.0	<10	2-butanol	none	10.0	12.8
26	38.3	220	250	15.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	33.3	15.1
27	38.3	203	250	40.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	37.0	18.3
28	29.0	167	250	15.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	18.2	10.9
29	7.3	41	750	20.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	3.1	7.5
30	33.0	180	250	30.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	26.5	14.7
31	14.3	78	250	30.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	7.6	9.7
32	15.0	91	250	0.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	16.1	17.7
33	17.0	98	250	15.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	10.5	10.7
34	14.3	78	250	30.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	10.8	13.8
35	24.0	127	250	40.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	16.0	12.6
36	42.0	255	250	0.0	<10	2-butanol	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	96.2	37.7
PTRMS Experiments									
Presto and Donahue, 2006									
6/14/2005			280	22.0		none	none		
	1.50	8.4						0.001	0.017
	2.47	13.9						0.057	0.41
	2.87	16.1						0.313	1.94
	4.01	22.5						0.717	3.18
	3.95	22.2						1.12	5.06
	4.45	25.0						1.43	5.71
	4.78	26.9						1.71	6.37
	4.87	27.4						1.91	6.99
	5.52	31.0						2.07	6.69
	5.17	29.0						2.21	7.63
	5.49	30.8						2.32	7.53
	5.85	32.8						2.40	7.30
	5.69	31.9						2.49	7.78
	6.04	33.9						2.56	7.56
	6.00	33.7						2.62	7.77
	6.06	34.0						2.63	7.72
	6.18	34.7						2.68	7.73
	6.22	35.0						2.70	7.74
	6.06	34.0						2.74	8.05
	6.23	35.0						2.72	7.79
	6.21	34.9						2.74	7.85
	6.06	34.0						2.76	8.10
	6.33	35.5						2.75	7.74

Author's Identifier	$\Delta$ ROG (ppb)	$\Delta$ ROG ( $\mu\text{g m}^{-3}$ )	O <sub>3</sub> initial (ppb)	Temperature (°C)	RH	OH-Scavenger	Seed	Result	
								$\Delta$ M ( $\mu\text{g m}^{-3}$ )	AMF $\Delta$ M/ $\Delta$ ROG (%)
	6.28	35.3						2.73	7.75
	6.07	34.1						2.75	8.07
	6.25	35.1						2.73	7.79
	6.21	34.9						2.76	7.91
	6.46	36.3						2.74	7.56
	6.16	34.6						2.74	7.92
	6.25	35.1						2.74	7.82
	6.29	35.3						2.73	7.75
	6.31	35.4						2.73	7.72
	6.38	35.8						2.74	7.66
	6.30	35.4						2.71	7.66
	6.28	35.3						2.72	7.70
	6.20	34.8						2.71	7.80
	6.29	35.3						2.71	7.67
	6.29	35.3						2.71	7.66
6/28/2005			390	22.0		none	none		
	24.1	135						15.3	11.3
	50.2	282						46.3	16.4
	71.1	399						73.7	18.4
	84.7	476						95.1	20.0
	96.8	543						112.9	20.8
	113.1	635						143.7	22.6
	129.5	727						164.9	22.7
	128.8	723						167.6	23.2
	129.2	726						170.3	23.5
	129.6	728						174.0	23.9
	134.1	753						185.0	24.6
	138.1	775						191.1	24.7
	138.8	779						190.7	24.5
	138.9	780						192.7	24.7
	139.8	785						191.7	24.4

1048