

1 **Reconciling discrepancies in the source characterization of VOCs**  
2 **between emission inventories and receptor modeling**

3

4 Jiamin Ou<sup>a</sup>, Junyu Zheng<sup>b,\*</sup>, Zibing Yuan<sup>c</sup>, Dabo Guan<sup>a</sup>, Zhijiong Huang<sup>b</sup>, Fei Yu<sup>b</sup>, Min  
5 Shao<sup>b</sup>, Peter K.K. Louie<sup>d</sup>

6

7 <sup>a</sup> School of International Development, University of East Anglia, Norwich, NR4 7TJ,  
8 U.K.

9 <sup>b</sup> Institute of Environment and Climate Research, Jinan University, Guangzhou 510006,  
10 China

11 <sup>c</sup> School of Environment and Energy, South China University of Technology, Guangzhou  
12 510006, China

13 <sup>d</sup> Hong Kong Environmental Protection Department, Revenue Tower, 5 Gloucester Road,  
14 Wanchai, Hong Kong, China

15

16

17

18 **AUTHOR INFORMATION**

19 **Corresponding Author**

20 \*J.Zheng; Phone: 86-20-39380021; fax: 86-20-39380021; e-  
21 mail:zheng.junyu@gmail.com

22

23

24

25

26

27 **ABSTRACT**

28 Emission inventory (EI) and receptor model (RM) are two of the three source apportionment  
29 (SA) methods recommended by Ministry of Environment of China and used widely to provide  
30 independent views on emission source identifications. How to interpret the mixed results they  
31 provide, however, were less studied. In this study, a cross-validation study was conducted in one  
32 of China's fast-developing and highly populated city cluster- the Pearl River Delta (PRD) region.  
33 By utilizing a highly resolved speciated regional EI and a region-wide gridded volatile organic  
34 compounds (VOCs) speciation measurement campaign, we elucidated underlying factors for  
35 discrepancies between EI and RM and proposed ways for their interpretations with the aim to  
36 achieve a scientifically plausible source identification. Results showed that numbers of species,  
37 temporal and spatial resolutions used for comparison, photochemical loss of reactive species,  
38 potential missing sources in EI and tracers used in RM were important factors contributed to the  
39 discrepancies. Ensuring the consensus of species used in EIs and RMs, utilizing a larger spatial  
40 coverage and longer time span, addressing the impacts of photochemical losses, and  
41 supplementing emissions from missing sources could help reconcile the discrepancies in VOC  
42 source characterizations acquired using both approaches. By leveraging the advantages and  
43 circumventing the disadvantages in both methods, the EI and RM could play synergistic roles to  
44 obtain robust SAs to improve air quality management practices.

45

46 **Key Words:** Source characterization; VOCs; Emission inventory; Receptor models;

47 Discrepancy

48

49

## 50 **1. Introduction**

51 Air pollution has emerged as a serious concern for many rapidly developing nations  
52 such as China. The premise of effective air pollution control is reliable source  
53 identification (Zhong *et al.*, 2013). To guide the emission source characterization,  
54 Ministry of Environment of China issued the guideline for emission source  
55 apportionment and recommended three methods, i.e., emission inventory (EI), receptor  
56 model (RM) and numerical methods such as the Comprehensive Air-quality Model with  
57 extensions (CAMx) (MEP, 2012). These three methods track down the emission sources  
58 from different perspectives and subsequently, provide varied results on source  
59 identification. In particular, EI and RM are widely used due to the complexities and  
60 higher data demand of numerical methods. The question on how to interpret the mixed  
61 results EI and RM generate and leverage the advantages of both methods, therefore,  
62 emerges with practical importance to support the policy-making process for air pollution  
63 control.

64 EI and RM quantify the emission sources from different angles. Typically, EI follows  
65 the bottom-up methodology to estimate the primary emissions for a given averaging time  
66 and geographic area. It involves estimation using statistics of activity level data (e.g.,  
67 fuel consumption, product output, population of vehicles) and emission factors. For each  
68 source category, the bulk inventory (total emissions of a pollutant) for a given pollutant is  
69 typically based on the product of an emission factor and an activity factor. For some  
70 pollutants comprised by multiple species such as PM<sub>2.5</sub> (particulate matter with dynamic  
71 diameter less than 2.5 μm) and volatile organic compounds (VOCs), speciated inventory  
72 will be developed combining the bulk inventory with speciated source profiles (The

73 percentage of different species in a source's emission). The established bulk or speciated  
74 inventory can be further refined to different temporal and spatial resolutions using  
75 surrogate data such as the temporal profiles of traffic flow, industrial activities and the  
76 spatial location of enterprise, road network and population density. The overall  
77 uncertainty in the inventory depends on the precision, accuracy, and representativeness of  
78 activity level data, emission factor, source profiles as well as the temporal and spatial  
79 surrogates (NARSTO, 2005; Miller *et al.*, 2006).

80 RM, by contrast, generally follows the top-down based methodology. It statistically  
81 apportions the measured ambient air pollutant concentrations, for multiple time periods at  
82 one or multiple monitoring sites, to the emission sources according to some pre-  
83 knowledge of their emission characteristics (primarily their chemical characteristics). The  
84 site- and time-specific ambient VOC species measurements are subject to sampling and  
85 analytical errors and to meteorological variability (Karagulian & Belis, 2012; Belis *et al.*,  
86 2015). Uncertainties in statistical inferences regarding source apportionment arise as a  
87 result of tracer elements that are common to multiple sources, lack of locally-  
88 representative emission source profiles, adoption of different receptor models, effects of  
89 reaction loss on different species, and judgment regarding interpretation of results (Hopke,  
90 1991; Watson *et al.*, 2001; Morino *et al.*, 2011; Yuan *et al.*, 2012; Ling & Guo, 2014; Ou  
91 *et al.*, 2015a).

92 Due to their different ways to track down emissions and sources of uncertainty, it is  
93 not surprising that they can produce mixed results for source apportionment (Leuchner &  
94 Rappenglück, 2010; Morino *et al.*, 2011; Wang *et al.*, 2014). For example, quite a few of  
95 studies compared VOCs emission source identification results between the emission

96 inventory and receptor models, and found significant inconsistencies in source  
97 contributions, especially for solvent use, Liquefied Petroleum Gas (LPG) uses, and  
98 biogenic sources (e.g., Fujita *et al.*, 1995; Scheff *et al.*, 1996; Watson *et al.*, 2001;  
99 Morino *et al.*, 2011; Wang *et al.*, 2014). In China, Zheng *et al.* (2009a) compared the  
100 source apportionment by EI with receptor modelling by Liu *et al.* (2008). General  
101 consistency was gained on the high contributions from gasoline vehicles, coating and  
102 solvents, but large discrepancies were observed in the contribution of LPG, and some  
103 specific areas featured by high local emission loadings. Despite of the observed  
104 discrepancies, the question of how to interpret the mixed and sometimes conflicting  
105 answers for source identification remains less studied. Reasons responsible for the  
106 discrepancies scattered in different studies and they were proposed and studied in a  
107 somewhat biased way with the underlying assumption that one of the methods is more  
108 reliable and the discrepancies are mainly attributed by the limitation or flaw of the other.  
109 For studies focused on EI, representativeness of sampling time and sites, photochemical  
110 loss and the tracers used in RM were questioned (Zheng *et al.*, 2009a). As for studies  
111 based on RM, they argued that EI may fall short of the data quality of activity level data,  
112 emission factor and potentially missing sources that lead to under- or over-estimations  
113 (Wang *et al.*, 2014). The variations between EI and RM results call for a more systematic  
114 study with an objective view on the limitations and uncertainties of both methods in order  
115 to leverage their advantages to support robust air pollution control policy formulation.

116 By cross-validating SAs of EI and RM, it is possible to elucidate factors contributing  
117 to the discrepancies, to figure out ways for better interpretation, and finally to have a  
118 scientifically more plausible source apportionment. In this study, the SA cross-validation

119 was demonstrated in the Pearl River Delta (PRD) region of China. The region is chosen  
120 for case study since it has a routinely updated regional-scale highly resolved EI for VOC  
121 species and a speciated ambient VOC measurement dataset generated from a region-wide  
122 gridded sampling campaign. The EI in the PRD provides hourly emission information for  
123 every  $3\times 3\text{km}^2$  grid. The RM is based upon VOCs measurement simultaneously  
124 conducted at 84 stations across the PRD, which overcomes limited spatial representation  
125 by a small number of monitoring sites that most of previous RM works were based upon  
126 (Louie *et al.*, 2013). The high-resolution speciated EI and gridded VOC measurement  
127 campaign for the exactly same area provided us a good opportunity to systematically  
128 examine the underlying factors responsible for the discrepancies between two methods  
129 and to further explore ways to negate their impacts for more scientifically plausible  
130 determinations of SAs.

## 131 **2. Data and Methods**

### 132 **2.1 Study region**

133 As mentioned above, the PRD was selected as the study region due to the availability  
134 of highly resolved speciated VOC EIs and an extensive VOC ambient measurement  
135 campaign (the location of the study area is shown in Fig. S-1 in the Supporting  
136 Information (SI)). As one of the fastest developing regions in China, the PRD suffers  
137 from air pollution problems characterized by haze and serious photochemical smog  
138 episodes due to its dense population and intense economical activities (Zheng *et al.*,  
139 2009ab; Louie *et al.*, 2013). Additionally, a variety of VOC emission sources in this  
140 region, e.g., vehicle exhaust, industrial and household solvent use, combustion, fuel  
141 evaporation, marine vessels, biomass burning, and biogenic emissions (Lau *et al.*, 2010;

142 Ou *et al.*, 2016), leads to further complexity in the source characterization. Hot and  
143 humid weather favors photochemical reactions over the PRD, and the atmospheric  
144 chemistry also plays an important role in shaping the VOC loadings and compositions in  
145 the atmosphere therein. Such complex emission and atmospheric chemistry  
146 characteristics in conjunction with the available high-resolution EIs and gridded ambient  
147 VOC measurements make the PRD an ideal place to conduct an SA cross-validation  
148 study to elucidate the possible factors causing the inconsistencies between the EI and RM.

## 149 **2.2 Emission inventory**

150 High-resolution VOC bulk EIs (i.e., aggregated emissions of VOC species) in the  
151 PRD region from 2008 and 2009 were adopted in this study (Zheng *et al.*, 2009b). These  
152 bulk EIs from 2008 and 2009 were employed because a gridded ambient VOC  
153 measurement campaign (see details in Section 2.3) was conducted over those two years.  
154 These bulk EIs include anthropogenic emissions from 7 categories in addition to over 40  
155 sub-categories and biogenic emissions estimated by the Model of Emissions of Gases and  
156 Aerosols from Nature (MEGAN) (Guenther, 2006). With source-specific temporal and  
157 spatial allocation surrogates (CUHK, 2015), these bulk EIs have a spatial resolution of  
158  $3 \times 3$  km<sup>2</sup> and an hourly temporal resolution. On the basis of the bulk EIs, the speciated  
159 VOC EIs with approximately 300 VOC species were developed using local VOC source  
160 profiles (Zheng *et al.*, 2009a; Ou *et al.*, 2015b; HKPU& SCUT, 2016); meanwhile, only  
161 30 VOC species were used in previous RMs. The reliabilities of the bulk and speciated  
162 EIs have been demonstrated in previous studies (Zheng *et al.*, 2009ab; Ou *et al.*, 2015b),  
163 and they are widely used in air quality modeling endeavors (Liu *et al.*, 2008; Ou *et al.*,  
164 2016).

### 165 **2.3 VOC grid study and receptor modeling**

166 A gridded VOC sampling campaign (hereafter as “VOC grid study”) with eight  
167 sampling events was conducted at 5 am and 10 am on four days (29 October 2008 and 1  
168 March, 26 September and 5 December 2009). The PRD was equally divided into 100  
169 grids with the grid size of 20×20 km<sup>2</sup>, and both VOC and oxygenated VOC (OVOC)  
170 samples were simultaneously collected at 84 on-land grids. A total of 672 samples were  
171 collected and analyzed using gas chromatography (GC) with a multi-detector system and  
172 high-pressure liquid chromatography (HPLC) with a photodiode array detector for VOCs  
173 and OVOCs, respectively. Details of the sampling and analysis methods utilized in the  
174 VOC grid study can be found in Louie *et al.* (2013).

175 Because the SA for this VOC speciated dataset has been reported previously (Yuan *et*  
176 *al.*, 2013), only a brief description is given here. The SA was performed using a positive  
177 matrix factorization (PMF) model (version 3.0). The uncertainties were determined  
178 following the protocol of Polissar *et al.* (1998). Twenty base runs and 100 bootstrap runs  
179 were performed to select the best solution and estimate the stability and uncertainty of the  
180 SA. Nine factors were identified, after which they were mapped onto the emission  
181 sources according to the abundances of various tracers, i.e., combustion, diesel exhaust,  
182 gasoline exhaust, gasoline evaporation, liquefied petroleum gas (LPG)-related sources,  
183 mixed solvents, industrial emissions, biogenic emissions and secondary and aged air  
184 masses. Details about the PMF modeling procedure can be found in Yuan *et al.* (2013).

### 185 **2.4 Methods for comparison**

186 To enable a comparison, the results obtained from both the EI and the PMF  
187 approaches were unified in terms of their source classification, sampling time, and



188 temporal and spatial resolutions. Regarding the source classification, the bottom-up EI  
189 method incorporates a much more detailed source classification system, while the RM  
190 technique provides a general delineation of multiple sources based on the similarities  
191 among source profiles. Therefore, the deliberately classified sources in the EIs were  
192 grouped to match the 8 RM-based source categories, including combustion, gasoline  
193 exhaust, diesel exhaust, industrial processes, mixed solvents, LPG-related sources,  
194 gasoline evaporation and secondary and aged air masses. While secondary and aged air  
195 mass sources were classified within the RM, no primary emission source in the EIs was  
196 assigned to this category. The source mapping between the EI and RM and the unified  
197 source classification (see Table S-1) are detailed in the SI.

198 Hourly VOC emissions of the 8 sampling periods, i.e., 5 am and 10 am on 29 October  
199 2008 and on 1 March, 26 September and 5 December 2009, were extracted from the EIs  
200 of 2008 and 2009 for comparison. To unify the spatial scale, the spatial surrogates used in  
201 the 3×3 km<sup>2</sup> EI were used to develop the 20×20 km<sup>2</sup> spatial factors for the emission  
202 allocation. The source characterization results acquired using the EIs and the PMF  
203 therefore had the same sampling time and spatial resolution and were ready for  
204 comparison in terms of source contribution percentages at both different temporal  
205 variations (i.e., hourly and annual) and different spatial scales (i.e., 20×20 km<sup>2</sup>, 40×40  
206 km<sup>2</sup>, and 200×200 km<sup>2</sup>).

207

### 208 **3. Results and Discussion**

#### 209 **3.1 Comparison of SAs using the EI and RM**

210 Fig. 1 a-d show the source contribution percentages derived from the bulk EI and RM.

211 For the bulk EI with an aggregated emission of approximately 300 VOC species, mixed  
212 solvent usage was the dominant contributor (42%), followed by industrial processes  
213 (15%), biogenic emissions (15%), and gasoline exhaust (14%). In comparison, gasoline  
214 exhaust accounted for the largest fraction of VOCs (19%) in the RM, though it was only  
215 marginally higher than the contributions from mixed solvent (16%), industrial processes  
216 (15%) and LPG-related sources (15%). Gasoline evaporation (9%), combustion (8%),  
217 diesel exhaust (8%), and secondary and aged air masses (7%) also accounted for notable  
218 contributions. Biogenic emissions only contributed 3% on average.

219 A good agreement was found between the gasoline exhaust contributions in the EI  
220 and RM. The bulk EI generally provided much higher estimates of the mixed solvent and  
221 biogenic emission contributions and lower value for LPG-related sources. The  
222 contributions from mixed solvents and biogenic emissions in EI were 3 and 5-6 times  
223 those in RM, respectively. On contrast, the LPG-related source contribution estimated via  
224 the RM was 7 times the contribution in EI.

225 The diurnal variations of gasoline exhaust, combustion, diesel exhaust, gasoline  
226 evaporation and biogenic emission sources exhibited consistencies between EI and RM  
227 with higher contributions at 10 am than at 5 am (Fig. 1 a, c and Fig. S-2 in the SI).  
228 However, the RM results were also associated with significant seasonal variations (Fig.  
229 1c) due to the large impacts of meteorological conditions and non-local transport  
230 mechanisms. Located along the southern coast of China, the PRD receives greater non-  
231 local contributions during the winter and spring with polluted continental air masses  
232 brought by the northeasterly monsoon; meanwhile, lower contributions are introduced  
233 during the summer and autumn with clean oceanic air masses brought by the

234 southwesterly monsoon (Zheng *et al*, 2009b). In contrast, the EI reflected only local  
235 emissions within the PRD with slight seasonal variations. This constitutes an inherent  
236 cause of the discrepancies discussed below in Section 3.2.

237

## 238 **3.2 Factors contributing to the inconsistencies between the EI and RM**

239 To investigate the possible reasons leading to the notable discrepancies between the  
240 RM and EI approaches, we perform a SA cross-validation in this section to identify the  
241 major factors that impact the source characterization results by EI and RM and explore  
242 possible ways to reconcile those discrepancies.

### 243 **3.2.1 Number of VOC species**

244 The bulk EIs utilized in this study contained approximately 300 VOC species, while  
245 only 30 VOC species were consistently detectable by the instrument for the RM. The  
246 number of VOC species detectable within the ambient air is commonly much less than  
247 the number of VOC species detectable from emission sources because of their lower  
248 concentrations and the chemical loss of some reactive species. To eliminate the impacts  
249 of differences in the number of species, we constructed a speciated EI that contained the  
250 same species as those in the RM. Fig. 1 e-f show the source contribution percentages  
251 derived from the speciated EI. The largest source in the speciated EI was industrial  
252 processes (23%), followed by gasoline exhaust (21%), mixed solvents (20%), and  
253 biogenic emissions (17%). The predominant contribution from mixed solvents in the bulk  
254 EI was significantly weakened relative to that in the speciated EI. The contributions from  
255 industrial processes, gasoline exhaust and mixed solvents to VOCs in the speciated EI  
256 were comparable to those in the RM.

257 The differences in the contribution patterns in the bulk EI and speciated EI can be

258 attributed to the proportions of the 30 RM species within the total VOC mass from a  
259 particular source. As shown in Fig. 2, the 30 RM species only accounted for 29% of the  
260 mixed solvent emissions, while they explained 85%, 76% and 74% of the emissions from  
261 combustion, gasoline exhaust and industrial processes, respectively. As a result, the  
262 contribution of mixed solvents dropped significantly in the speciated EI, thereby  
263 becoming more similar to those in the RM. Larger differences (i.e., reaching an order of  
264 magnitude) between the mixed solvent contributions from the bulk EI and RM were also  
265 reported in previous studies (Wang *et al.*, 2014; Watson *et al.*, 2001). These differences  
266 might be significantly reduced if an EI with the same species as a RM was used for  
267 comparison. This finding highlights the importance of species consensus when comparing  
268 SAs determined via EIs and RMs. Therefore, for the remainder of this study, comparisons  
269 were conducted between the speciated EI and the RM based upon the same 30 VOC  
270 species.

### 271 **3.2.2 Temporal and spatial resolutions**

272 After reconciling the number of species that used for comparison, different scales of  
273 temporal and spatial resolutions also explained part of the discrepancies between the two  
274 methods.

275 Starting with the finest temporal and spatial resolution, i.e., instant sample for the  
276 20×20 km<sup>2</sup> grid, we illustrated the SA results by EI and RM in two sampling periods (2<sup>nd</sup>:  
277 10am, 1 March 2009; 6<sup>th</sup>: 10am, 26 September 2009) for the selected cells to represent  
278 urban (Grid 74), industrial (Grid 66), suburban (Grid 54), upwind rural (Grid 99) and  
279 downwind rural (Grid 14) areas in PRD in Fig. 3. Noteworthy is that, the two sampling  
280 periods were of the same local time but different seasons (March for spring, and

281 September for late summer in PRD).

282 As shown in Fig. 3, SA by EI was generally stable during these two seasons, and its  
283 source contributions generally coincided with the local emission characteristics.  
284 Industrial processes, mixed solvents and gasoline exhaust dominated in urban and  
285 suburban grids. Industrial processes contributed the most significantly in the industrial  
286 grid, while biogenic emissions overwhelmed in the two rural grids. By contrast, SA by  
287 RM was much sampling event-specific, and the resolved contributions deviated  
288 significantly to their EI counterparts. For example, industrial processes contributed to  
289 40% of VOCs in the urban grid in the 2<sup>nd</sup> event and 14% in the 6<sup>th</sup> event, both differed  
290 much from 30% and 23% in the EI. In two rural grids, RM resolved much lower biogenic  
291 contributions. The upwind rural grid was associated with higher contribution of  
292 secondary & aged air mass from long-range transport. Its source contribution patterns  
293 also varied as the prevailing wind changed from weak north-easterly in the 2<sup>nd</sup> event to  
294 moderate easterly in the 6<sup>th</sup> event. The downwind rural grid was dominated by high  
295 contributions of mixed solvents and remained stable despite the shift of prevailing winds.  
296 Significant mixed solvents may come from nearby upwind emissions, as Grids 24 and 25  
297 to the north and northeast of the downwind rural grid were characterized by high  
298 emissions of mixed solvents.

299 The above selected cells showed how the SA results from EI and RM might be  
300 different from cell to cell and period to period, if we summarized their discrepancies for  
301 all the 84 grids over the 8 sampling periods, we found that the discrepancies were quite  
302 large. As shown in Fig. 4a, 57% of the results between two methods varied more than 3  
303 times, i.e., the EI result was more than 3 times of RM or the other way around (RM result

304 was more than 3 times of EI). 24% of the estimations even had differences greater than  
305 15-fold, and almost all source categories contributed to these extreme values. Therefore,  
306 such disagreements were independent of source categories.

307 If comparisons were made in larger temporal and spatial resolutions, e.g., combined  
308 the 8 sampling periods as annual average or combined every 4 grids to a bigger grid of  
309  $40 \times 40 \text{ km}^2$ , the discrepancies between the two methods seemed to be smoothed in some  
310 degrees. As Fig. 4b illustrated, the percentage for those with variations more than 3 times  
311 decreased to 46% when comparison was made for annual average in  $20 \times 20 \text{ km}^2$   
312 resolution. Similarly, when the grid size was increased to  $40 \times 40 \text{ km}^2$ , the percentage  
313 dropped to 46% (Fig. 4c). If both spatial and temporal enhancements were adopted, only  
314 38% of the results remained in the range of more than 3 times, i.e., 62% of the results fell  
315 in the range between 1/3 and 3 (Fig. 4d). If the grid size further increased to cover the  
316 entire PRD region and samples in all eight events averaged, 78% of the percentage ratios  
317 (7 out of 9 sources) fell in the range between 1/3 and 3 (Fig. 4e). Only biogenic emission  
318 and LPG-related sources still had percentage ratios greater than 3, implying other factors  
319 may contribute larger to discrepancies for the two categories. We shall address them in  
320 later sections.

321 The reasons why the EI and RM results can be reconciled by spatial and temporal  
322 averaging enhancements might due to the fact that RM results were sampling event-  
323 specific. As discussed in Fig. 3, the source contributions of EI were generally more stable  
324 while the results by RM tended to fluctuate from different sampling times and were  
325 subject to meteorological conditions and the upwind sources. For example, under  
326 different wind patterns, hotspots of toluene concentration were at different grids (as

327 shown in Fig. S3-a&b of SI) and sometimes could be 40-60 kilometers away from the in-  
328 situ emission hot spots (Fig. S3c&b of SI). As inter-grid transport can be reflected by RM  
329 but cannot by EI, averaging in a larger area over a longer period of time is a possible way  
330 to ease this inherent discrepancy. Hence, in order to make a scientifically sound cross-  
331 validation between EI and RM, sampling should be conducted over a large area during a  
332 long period of time, preferentially a year, to minimize the impact by meteorology and  
333 inter-grid transport.

### 334 **3.2.3 Chemical loss**

335 Even after reconciling the spatiotemporal resolution as discussed above, substantial  
336 disagreements still existed for biogenic emissions and LPG-related sources. Though  
337 similar diurnal variations were presented by EI and RM, the biogenic emission  
338 contribution resolved by the EI was 5.7 times that by RM. Intriguingly, the biogenic  
339 emission estimates from the EI and RM had a high correlation ( $r=0.77$ ), suggesting that  
340 both estimates were robust in terms of their temporal variations. The significant  
341 difference in the absolute mass contributions is therefore believed to be largely due to the  
342 high reactivity of isoprene, the tracer used to identify biogenic emissions in an RM. An  
343 inherent assumption of an RM is that all samples in a source profile are constant during  
344 the analysis, and the high reactivity of isoprene certainly violates this assumption (Harley  
345 and Cass, 1995). Since the reactivity of isoprene is one to two orders higher than those of  
346 other species in an RM (Harley and Cass, 1995), it was removed at a faster rate in the  
347 atmosphere, thereby distorting the ambient VOC profiles for interpretation using the RM  
348 and resulting in much lower biogenic contributions.

349 The discrepancy between the SAs acquired using the EI and RM correlated well with

350 the source reactivity. By combining the reactivity ( $k_{OH}$ ) of different VOC species with  
351 their proportions in a source, the source reactivity can be estimated (details are provided  
352 in the “Estimation of source reactivity” material of the SI). As shown in Fig. 5, biogenic  
353 emissions (point 9) in the upper-right corner constituted the most reactive source, and it  
354 was associated with the largest difference between the EI and RM. As the source  
355 reactivity decreased, the relative differences of source contributions by EI and RM  
356 declined as well, with the exception of LPG-related sources and secondary and aged air  
357 masses. Secondary sources exhibited the lowest reactivity, as it is composed of long-lived  
358 species. Since a secondary source cannot correspond to any source in an EI, its associated  
359 discrepancy was expected to be high. If LPG-related sources and secondary and aged air  
360 masses were removed, the source reactivity showed a positive relationship with the  
361 relative differences of two methods ( $r^2=0.59$ ). Therefore, chemical loss constituted the  
362 single most important factor in the disagreement between EI and RM. Accordingly, some  
363 adjustment methods have been developed to account for the chemical losses of VOC  
364 species in the atmosphere to reconcile the SAs acquired using EIs and RMs (Na and Kim,  
365 2007; Yuan *et al.*, 2012).

366

#### 367 **3.2.4 Potential missing sources in the EI and tracers used in the RM**

368 The above factors failed to explain the significant discrepancy in the contribution of  
369 LPG-related sources. Its source contribution estimated using the RM was 15 times that by  
370 EI. The disagreement in the LPG-related source contributions from EI and RM has been  
371 reported elsewhere in China (Zhang *et al.*, 2009; Zhao *et al.*, 2012), Japan (Morino *et al.*,  
372 2011) and North America (Blake and Rowland, 1995; Fujita *et al.*, 1995), suggesting that



373 this discrepancy is globally pervasive. Two reasons might be blamed for it.

374 One is the usage of propane and *i/n*-butane in the RM as unique tracers of LPG  
375 sources. Propane and *i/n*-butane are ubiquitous in the atmosphere and generally make up  
376 large portions of the measured VOCs. For a long time, these species were treated as  
377 tracers of LPG sources due to their higher percentages in the source profiles (percentage  
378 of a species in a source's emission). Propane and *i/n*-butane each comprises 40%, 4% and  
379 9% of the VOC emitted from LPG exhaust (Ou *et al.*, 2015b; HKPU&SCUT, 2016),  
380 much higher than their percentages in other sources. With the measurements of high  
381 concentrations of propane and *i/n*-butane in ambient samples and the underlying  
382 assumption that propane and *i/n*-butane came dominantly from LPG sources, LPG was  
383 constantly apportioned with high source contribution by RM. However, if the emission  
384 intensity was considered, industrial processes, which dominated the emissions in PRD,  
385 would contributed 47%, 29% and 54% of the total propane and *i/n*-butane emissions in  
386 the PRD according to the EI (as shown in Fig. S-4 in SI). Regardless of whether these  
387 percentages were accurate or not, we need to be cautious on treating propane and *i/n*-  
388 butane as the tracers of LPG. More efforts are needed to measure the local source profiles,  
389 especially the presence of propane and *i/n*-butane, in a wide variety of industrial  
390 processes.

391 Another possibility is underestimation of LPG emissions in the current EI. A previous  
392 study suspected that usage of LPG might result in significant leakage (Blake & Rowland,  
393 1995), with leakage rate of 1-5% depending on the boundary conditions. Evaporative  
394 emissions from LPG usage and gasoline evaporation during vehicle movement and  
395 parking were absent in the current EI. A recent study in China highlighted that vehicular

396 evaporative emissions (predominantly from gasoline) constituted a missing yet significant  
397 part of VOC emissions in EI, and estimated that one vehicle in China emitted 1.6 kg of  
398 VOC emissions per year (Liu *et al.*, 2015). If these two potentially missing sources were  
399 taken into account in EI, the source contributions by EI would change as those shown in  
400 Fig.6. It was noted that the large discrepancy in LPG-related sources was reconciled by  
401 inclusion of vehicular evaporative emissions and 2% LPG leakage rate. This highlighted  
402 that the need to review and improve emission estimations from evaporative sources in EI.  
403 Nevertheless, cross-validation on SAs between EI and RM revealed the necessity to  
404 improve both methods.

405

### 406 **3.3 Implications for reconciling the inconsistencies between the EI and RM**

407 The characterization of VOC sources is challenging due to the complexity in emission  
408 sources and species. The EI and RM represent two widely used VOC SA techniques that  
409 approach VOC from different perspectives, i.e., EIs focus on emission sources (bottom-  
410 up) while RMs emphasize pollutant levels in the environment (top-down); therefore, it is  
411 natural that these two methods may generate intrinsically different SA results. Given their  
412 inherent limitations, the SAs determined using either the EI or the RM could deviate from  
413 the ‘real’ source contributions, implying that control strategies based on either the EI or  
414 RM separately can be easily biased. This study identified the factors contributing to these  
415 discrepancies and provided ways to address these factors. The work shown here could  
416 help negate their impacts to obtain a more plausible SA, thereby generating a more robust  
417 control strategy.

418 The first implication of this study was that a consensus of species in cross-validating

419 the results from the EI and RM is a prerequisite for reducing their discrepancies. The  
420 species included within RM generally present with high concentrations in the ambient air  
421 and low instrumental detection limits, while species in EIs are detectable in source profile  
422 measurements. As emission sources have high loadings of various species, higher  
423 quantities of species are more often detected and thus included in speciated EIs than in  
424 RMs. Inconsistencies in such species generally lead to lower estimations of some source  
425 contributions such as mixed solvent in RMs. These discrepancies can be even larger if the  
426 SA by an RM is simply compared with the EI of a total pollutant amount (e.g., bulk  
427 VOCs). It is therefore improper to use RM results based on a limited number of species to  
428 infer the control measures for a bulk amount of pollutants.

429 This study also highlighted the importance of reconciling the EI and RM SA results in  
430 a larger spatial coverage and longer time span. As pollutants in the air may originate from  
431 both local and non-local (i.e., areas outside the EI coverage) sources, an enhanced spatial  
432 coverage could downgrade the impact of non-local sources, thereby improving the  
433 consistency between the SAs from EIs and RMs. Meanwhile, longer time spans could  
434 improve the SAs determined by both methods. In an EI, temporal allocation surrogates  
435 are often used to allocate the total annual emission into months, days and hours; however,  
436 this can introduce additional uncertainties into these surrogates and therefore impact the  
437 emission estimates at any specific time. The influences of such uncertainties can be  
438 reduced if SAs are constructed with a longer time span, e.g., seasonally or annually. In  
439 RM, a longer time span would cover a wider range of synoptic conditions and thereby  
440 increase the temporal variability in the source contributions. In particular, a sampling  
441 interval spanning at least one year is suggested in areas dominated by a monsoonal

442 climate to avoid any biases in SAs produced from limited directions of prevailing winds.

443 Third, addressing the impacts of chemical losses by employing receptor models for  
444 SAs could help reconcile the discrepancies between EIs and RMs. Chemical losses might  
445 be the single most important factor in SA discrepancies with regard to reactive pollutants,  
446 e.g., VOCs. Reactive species decay rapidly in the atmosphere and can drop to low or even  
447 undetectable levels if the receptors are far from their emission sources. Therefore, RMs  
448 most likely underestimate the contributions of sources with high abundances of reactive  
449 species and accordingly overestimate the contributions of sources with low reactivity. The  
450 significant decay of a reactive species could therefore distort the conclusions and  
451 implications for emission controls, reactive-based controls or health-based controls  
452 derived from an RM. For example, many studies have used maximum incremental  
453 reactivity (MIR)-weighted factor loadings from RM to indicate the ozone forming  
454 potentials (OFPs) of different VOC sources (Carter, 2008; Ou *et al.*, 2015b). However,  
455 such OFPs only account for the post-receptor ozone production capacity; meanwhile,  
456 they neglect ozone that was produced by the decay of VOCs prior to reaching the receptor.  
457 As a result, the OFPs for sources with high abundances of reactive species are  
458 underestimated, and the degree of underestimation depends on the reactivity of the  
459 species and the proximity of the receptor to its source. Indeed, although they are used  
460 widely, MIR-weighted VOC factors in RM are scientifically inaccurate when estimating  
461 source-specific OFPs, as they generate misleading scientific conclusions and control-  
462 oriented implications. Hence, the impacts of chemical losses must be quantitatively  
463 addressed before they are used for cross-validation and policy formulation.

464 There are several types of RMs, in which PMF and the Chemical Mass Balance

465 (CMB) are the two most popular ones. In this study, only PMF is used for comparison  
466 with EI. This is due to the higher requirement of input data in CMB, such as the actual  
467 number of emissions sources and their source profiles, which requires a comprehensive  
468 pre-knowledge of all the significant sources and their speciated characteristics (Na and  
469 Kim, 2007; Lau *et al.*, 2011; Teixeira *et al.*, 2015). Previous studies showed that different  
470 RMs have general consistencies in the major emission sources. Specific variations in  
471 certain sources vary from sites and the VOC species in calculation. For example, Song *et al.*  
472 (2008) found the contribution of gasoline-related sources using CMB was higher than  
473 PMF for 31 VOC species (not including polycyclic aromatic hydrocarbons (PAHs)) in  
474 Beijing. On contrast, Teixeira *et al.* (2015) reported that PMF attributed a slightly greater  
475 amount of PAHs to the gasoline and diesel sources in Brazil. Differences in the  
476 theoretical approaches of RMs, the site-specific atmospheric chemistry and reactivity of  
477 VOC species might contribute to the variations of model performances. Though the  
478 variations between different RMs were much less than the discrepancies between PMF  
479 and EI in this study, cross-validation between different RMs and emission inventories is  
480 recommended to better understand the source characteristics of a region.

481 Last but not least, speciated EIs of reactive species are associated with significant  
482 uncertainties. Studies have shown that uncertainties in bulk VOC emission estimates  
483 sometimes reach or exceed 100% (Wei *et al.*, 2008; Zheng *et al.*, 2009b), and these  
484 estimates may be even higher for reactive species (Simpson *et al.*, 1995). These  
485 uncertainties could be enhanced when a speciated EI is inferred by multiplying bulk  
486 emissions with speciated source profiles that are either measured or borrowed from an  
487 established source profile database such as SPECIATE from the U.S. Environmental

488 Protection Agency (EPA). Considering significant uncertainties in the SA of a reactive  
489 species using both EI and RM, cross-validation and reconciliation can shed light on  
490 directions for further improvement in both methods. For example, the large discrepancies  
491 in LPG-related sources indicate one or both of the methods call for critical review.  
492 Vehicular evaporation emissions and LPG leakage might be underestimated or missing in  
493 the current EI, while the tracers of propane and *i/n*-butane in RMs need to be reviewed to  
494 address the issue of overlapped tracers. In this regard, the EI and RM both play as  
495 synergistic roles in improving the confidence of SA results by both methods.

#### 496 **4. Summary and Conclusions**

497 Source identification and apportionment are fundamental for the formulation of air  
498 pollution control measures. The two most widely used source characterization approaches  
499 are the bottom-up-based EI and the top-down-based RM; however, these two techniques  
500 often provide inconsistent SA results due to their inherent differences and limitations. In  
501 this study, we utilized the high-resolution, speciated VOC emission inventory and the  
502 region-wide VOC gridded measurement campaign for the PRD as a case study to perform  
503 a detailed examination of the factors leading to SA discrepancies and to explore possible  
504 ways to reconcile those discrepancies. The consensus of species included in the analysis,  
505 the temporal and spatial resolutions of the data for comparison, chemical losses of  
506 reactive species, potential missing sources in EIs and some tracers used in RMs were  
507 identified as the important factors responsible for the SA discrepancies between EIs and  
508 RMs. Therefore, ensuring a consensus of the species used in the EI and RM, employing a  
509 larger spatial coverage and longer time span, addressing the impacts of chemical losses  
510 by using receptor models, supplementing emissions from missing sources and selecting

511 proper source tracers will help to reconcile the discrepancies among VOC source  
512 characterizations generated using EIs and RMs.

513 This study also highlighted the fact that high-resolution, speciated EIs and large-scale  
514 gridded monitoring campaigns are essential for a plausible source comparison between an  
515 EI and an RM. By leveraging the advantages and circumventing the disadvantages in  
516 both methods, EIs and RMs could play as synergistic roles in producing reliable source  
517 characterizations, especially those for reactive species such as VOCs. The work shown in  
518 this study could be used to more accurately characterize the sources of PM<sub>2.5</sub> emissions.  
519 The approach outlined herein could thus be promulgated to other regions in China and  
520 other developing countries with intense and complex emissions sources to more  
521 accurately characterize source impacts on the ambient pollution and ultimately to  
522 improve the efficacies of pollution control strategies.

523

## 524 **ACKNOWLEDGMENTS**

525 This work was supported by the National Distinguished Young Scholar Science Fund  
526 (grant number 41325020) and the Key Program (grant number 91644221) of the National  
527 Natural Science Foundation of China, the National Major Scientific Instruments  
528 Development Project of China (grant number 41627809), and the National Key Research  
529 and Development Program of China (grant number 2016YFC0202201). The opinions  
530 expressed in this paper are those of the authors and do not necessarily reflect the views or  
531 policies of the Governments of Guangdong Province and the Hong Kong Special  
532 Administrative Region, nor does any mention of a trade name or commercial product  
533 constitute an endorsement or recommendation of their use.

534

535 **SUPPORTING INFORMATION AVAILABLE**

536 Four figures, one table and additional information are available for the following: (1) the  
537 locations of the PRD region and the VOCs grid sampling campaign; (2) the source  
538 categories in the EI and their grouping relative to the RM sources; (3) the diurnal  
539 variations in the RM and EI region-wide estimates; (4) the spatial distribution of ambient  
540 toluene under different wind patterns and the spatial locations of its major sources-  
541 industrial solvent and industrial processes; (5) the proportions of propane and *i/n*-butane  
542 in the source profiles and the total emission budget; and (6) the estimation of the source  
543 reactivity.



544 **References**

- 545 Blake, D. R.; Rowland, F. S. Urban leakage of liquefied petroleum gas and its impact on Mexico city  
546 air quality. *Science*, 1995, 269, 953–6.
- 547 Carter WP. Reactivity estimates for selected consumer product compounds. Final Report to the  
548 California Air Resources Board Contract 2008: 06-408.
- 549 Karagulian, F.; Belis, C.A. Enhancing source apportionment with receptor models to foster the air  
550 quality directive implementation. *Int. J. Environment and Pollution* 2012, 50, 190-198.
- 551 Belis, C.A.; Pernigotti, D.; Faragulian, F.; Pirovano, G.; Larsen, B.R.; et al. A new methodology to  
552 assess the performance and uncertainty of source apportionment models in intercomparison  
553 exercises. *Atmos. Environ.* 2015, 119, 35-44.
- 554 Hong Kong Polytechnic University and South China University of Technology (HKPU& SCUT).  
555 Characterization of VOC sources and Integrated Photochemical Ozone Analysis in Hong Kong  
556 and the Pearl River Delta Region; Final Report, 2016.
- 557 Fujita, E. M.; Watson, J. G.; Chow, J. C.; Magliano, K. L. Receptor Model and Emissions Inventory  
558 Source Apportionments of Nonmethane Organic Gases in California’S San Joaquin Valley and  
559 San Francisco Bay Area. *Atmos. Environ.* 1995, 29, 3019–3035.
- 560 Guenther, A.; Karl, T.; Harley, P.; Wiedinmyer, C.; Palmer, P.I.; Geron, C. Estimates of global  
561 terrestrial isoprene emissions using MEGAN ( Model of Emissions of Gases and Aerosols from  
562 Nature ). *Atmos. Chem. Phys.*, 2006, 6, 3181-3210.
- 563 Lau, A.K.H.; Yuan, Z.; Yu, J.Z.; Louie, P.K.K. Source apportionment of ambient volatile organic  
564 compounds in Hong Kong. *Sci. Total Environ.* 2010, 408, 4138–49.
- 565 Leuchner, M.; Rappenglück, B. VOC source–receptor relationships in Houston during TexAQS-II.  
566 *Atmos. Environ.* 2010, 44, 4056-4067.
- 567 Li, Y.; Lau, A.K.H.; Fung, J.C.H.; Zheng, J.Y.; Zhong, L.J.; Louie, P.K.K. Ozone Source  
568 Apportionment (OSAT) to Differentiate Local Regional and Super-regional Source  
569 Contributions in the Pearl River Delta Region, China. *J. Geophys. Res.*, 2012, 117, D15305.
- 570 Ling, Z.H.; Guo, H. Contribution of VOC sources to photochemical ozone formation and its control  
571 policy implication in Hong Kong. *Environ. Sci. & Policy* 2014, 38, 180-191.
- 572 Liu, H.; Man, H.Y., Tschantz, M., Wu, Y., He, K.B., Hao, J.M. VOC from Vehicular Evaporation  
573 Emissions: Status and Control Strategy. *Environ. Sci. Technol.* 2015, 49, 14424–14431.
- 574 Liu, Y.; Shao, M.; Lu, S.H.; Chang, C.C.; Wang, J.L.; Fu, LL. Source apportionment of ambient  
575 volatile organic compounds in the Pearl River Delta, China: Part II. *Atmos. Environ.* 2008, 42,  
576 6261-6274.
- 577 Louie, P.K.K., Ho, J.W.K.; Tsang, R.C.W.; Blake, D.R.; Lau, A.K.H.; Yu, J.; Yuan, Z.B.; Wang, X.M.;  
578 Shao M.; Zhong, L.J. VOCs and OVOCs distribution and control policy implications in Pearl  
579 River Delta region, China. *Atmos. Environ.* 2013, 76, 125–135.
- 580 Ministry of Environmental Protection of the People's Republic of China (MEP). Guidelines on source  
581 apportionment techniques of particulate matters (trail), 2012; Available via:

582 [http://www.zhb.gov.cn/gkml/hbb/bwj/201308/t20130820\\_257699.htm](http://www.zhb.gov.cn/gkml/hbb/bwj/201308/t20130820_257699.htm).

583 Miller, C.A.; Hidy, G.; Hales, J.; Kolb, C.E.; Werner, A.S.; Haneke B, et al. Air Emission Inventories  
584 in North America: A Critical Assessment. *J. Air & Waste Manage. Assoc.* 2006, 56:1115–1129

585 Morino, Y.; Ohara, T.; Yokouchi, Y.; Ooki, A. Comprehensive source apportionment of volatile  
586 organic compounds using observational data, two receptor models, and an emission inventory in  
587 Tokyo metropolitan area. *J. Geophys. Res.* 2011, 116, D02311.

588 Na, K.; Pyo Kim, Y. Chemical mass balance receptor model applied to ambient C2-C9 VOC  
589 concentration in Seoul, Korea: Effect of chemical reaction losses. *Atmos. Environ.* 2007, 41,  
590 6715–6728.

591 North American Research Strategy for Tropospheric Ozone (NARSTO). Improving Emission  
592 Inventories for Effective Air Quality Management Across North America: A NARSTO  
593 Assessment, 2005.

594 Ou, J.; Guo, H.; Zheng, J.Y.; Cheung, K.; Louie, P.K.K.; Ling, Z.H.; Wang, D. Concentrations and  
595 sources of non-methane hydrocarbons (NMHCs) from 2005 to 2013 in Hong Kong: A multi-  
596 year real-time data analysis. *Atmos. Environ.*, 2015a, 103, 196–206.

597 Ou, J.; Zheng, J.Y.; Li, R.R.; Huang, X.B.; Zhong, Z.M.; Zhong, L.J.; Lin, H. Speciated OVOC and  
598 VOC emission inventories and their implications for reactivity-based ozone control strategy in  
599 the Pearl River Delta region, China. *Sci. Total Environ.* 2015b, 530–531, 393–402.

600 Ou, J.; Yuan, Z.B.; Zheng, J.Y.; Huang, Z.J.; Shao, M.; Li, Z.K.; Huang, X.B.; Guo, H.; Louie, P.K.K.  
601 Ambient Ozone Control in a Photochemically Active Region: Short-Term Despiking or Long-  
602 Term Attainment? *Environ. Sci. Technol.* 2016, 50, 5720–5728.

603 P.K. Hopke. Receptor Modelling for Air Quality Management. Elsevier Science Ltd, 1991, pp. 213-  
604 249.

605 Polissar, A. V.; Hopke, P. K.; Paatero, P.; Malm, W. C. & Sisler, J. F. Atmospheric aerosol over Alaska:  
606 2. Elemental composition and sources. *J. Geophys. Res.* 1998, 103, 19045.

607 Scheff, P.A., Wadden, D.M., Kenski, J., Chung J., Wolff, G. Receptor model evaluation of the  
608 Southeast Michigan Ozone Study ambient NMOC measurements. *J. Air & Waste Manage.*  
609 *Assoc.* 1996, 46(11): 1048-1057.

610 Simpson, D, Guenther A, Hewitt CN, Steinbrecher R. Biogenic emissions in Europe: 1. Estimates and  
611 uncertainties. *Journal of Geophysical Research: Atmospheres*, 1995; 100: 22875-22890.

612 Song, Y., Dai, W., Shao, M., Liu, Y., Lu, S.H., Kuster, W., Goldan, P. Comparison of receptor models  
613 for source apportionment of volatile organic compounds in Beijing, China. *Environ. Pollut.*  
614 2008, 156: 174-183.

615 Teixeira, E.C., Agudelo-Castañeda, D.M., Mattiuzi, C.D.P. Contribution of polycyclic aromatic  
616 hydrocarbon (PAH) sources to the urban environment: A comparison of receptor models. *Sci.*  
617 *Total Environ.* 2015, 538: 212-219.

618 Wang, M.; Shao, M.; Chen, W.; Yuan, B.; Lu, S.; Zhang, Q.; Zeng, L.; Wang, Q. A temporally and  
619 spatially resolved validation of emission inventories by measurements of ambient volatile

620 organic compounds in Beijing, China. *Atmos. Chem. Phys.* 2014, 14, 5871–5891.

621 Watson, J. G.; Chow, J. C.; Fujita, E. M. Review of volatile organic compound source apportionment  
622 by chemical mass balance. *Atmos. Environ.* 2001, 35, 1567–1584.

623 Yuan, B.; Shao, M.; Gouw, J.; Parrish, D.D.; Lu, S.H.; Wang, M.; Zeng, L.M.; Zhang, Q.; Song, Y.;  
624 Zhang, J.B.; Hu, M. Volatile organic compounds (VOCs) in urban air: How chemistry affects  
625 the interpretation of positive matrix factorization (PMF) analysis. *J. Geophys. Res. Atmos.* 2012,  
626 117, 1–17.

627 Yuan, Z.; Zhong, L.J.; Lau, A.K.H.; Yu, J.Z.; Louie, P.K.K. Volatile organic compounds in the Pearl  
628 River Delta: Identification of source regions and recommendations for emission-oriented  
629 monitoring strategies. *Atmos. Environ.* 2013, 76, 162–172.

630 Zhao, B.; Wang, P.; Ma, J.Z.; Zhu, S.; Pozzer A.; Li, W. A high-resolution emission inventory of  
631 primary pollutants for the Huabei region, China. *Atmos. Chem. Phys.* 2012, 12, 481–50.

632 Zheng, J.; Shao, M.; Che, W.W.; Zhang, L.J.; Zhong, L.J.; Zhang, Y.H.; Streets, D. Speciated VOC  
633 emission inventory and spatial patterns of ozone formation potential in the Pearl River Delta,  
634 China. *Environ. Sci. Technol.* 2009a, 43, 8580–8586.

635 Zheng, J.; Zhang, L.; Che, W.; Zheng, Z.; Yin, S. A highly resolved temporal and spatial air pollutant  
636 emission inventory for the Pearl River Delta region, China and its uncertainty assessment.  
637 *Atmos. Environ.* 2009b, 43, 5112–5122.

638 Zhong, L.J.; Louie, P.K.K.; Zheng, J.Y.; Yuan, Z.B.; Yue, D.L.; Ho, J.W.K. Science–policy interplay:  
639 Air quality management in the Pearl River Delta region and Hong Kong. *Atmos. Environ.* 2013,  
640 76, 3–10.

641