Response to Referee #2's comments

This paper presents an interesting analysis for understanding missing VOC sources in urban areas. The science and the methodology are sound; however, the paper itself lacks depth. It seems as if the authors were hesitant to list out too much information within the manuscript. It also assumes any reader has the exact background knowledge to follow all arguments so the authors don't elaborate. While it is true readers can check the references themselves, and should, it would be nice if the authors could share tangible evidence from the sources that support their work. A paper should not only introduce a new idea or method but also be written in a way that subsequent papers can test the method for themselves and apply it to other data sets. As it is currently written that is not possible. This work definitely should be published but the authors need to go over the manuscript and hone the message. I suggest this paper is accepted with minor revisions because no additional analysis needs to be done but share a clearer message for what has already been done.

Reply: Many thanks for your comprehensive and valuable comments. According to your suggestion, we have extended the discussion in the manuscript. In the introduction, we provided more information about the current shortcoming of VOC measurements to highlight the importance of VOC_R measurements. In the method, we provide more detailed introduction of GC-MS/FID, PTR-ToF-MS and the correction of NO interference on R_{OH} measurement. In addition, we provided the detailed information about how to consider the missing VOC_R in the box model. In the results of discussion, we provided the fitted coefficients of the MLR and improve relevant figures.

Specific Comments

Lines 90-91: You mention that missing VOC_R is a large part of total VOC_R but you don't give numbers or examples for the reader. You mention suburban sources in reference to R_{OH} (Line 76) but not for VOC_R

Reply: Thanks for our suggestions. In different regions, the missing VOC_R accounted for 10-75% of total R_{OH}. Given that total VOC_R is one part of total R_{OH}, missing VOC_R would account for a larger percentage of total VOC_R (>10%-75%).

Lines 94-95: Given that total VOC_R is one part of total R_{OH}, missing VOC_R would account for a larger percentage of total VOC_R (>10%-75%).

Line 127: Could you show how a NO_x correction is applied? Perhaps in the supplement. What does "carefully applied" mean?

Reply: Thanks for your suggestion. We have depicted how the NO_X correction curve is obtained and how it is applied to correct measured R_{OH} in the supplement.

Lines 145-147: A correction curve was acquired from these NO interference experiments, which can be used to correct the R_{OH} thanks to the simultaneous measurement of ambient NO concentrations (Supplementary information S1; Fig. S1).

Lines 28-38 in supplementary information:

S1 The correction of NO interference on R_{OH} measurements

The NO-correction experiments were conducted by introducing given amounts of VOC standard gases into the reactor. Different levels of NO were injected into the reactor and the difference between "measured" R_{OH} and true R_{OH} increased as the NO concentration increased. Here, the difference between "measured" R_{OH} and true R_{OH} is defined as δR_{OH}. Then, a correction curve was fitted between the δR_{OH} and NO concentrations. Several standard gases (propene and PAMS mixture) and different levels of base reactivity (from 30 to 90 s^{-1}) have been tried and the curve was quite consistent for all tested gases, as shown in Fig. S. According to this correction curve and ambient NO concentrations, we calculated the δR_{OH} which was used to correct the measured R_{OH}.

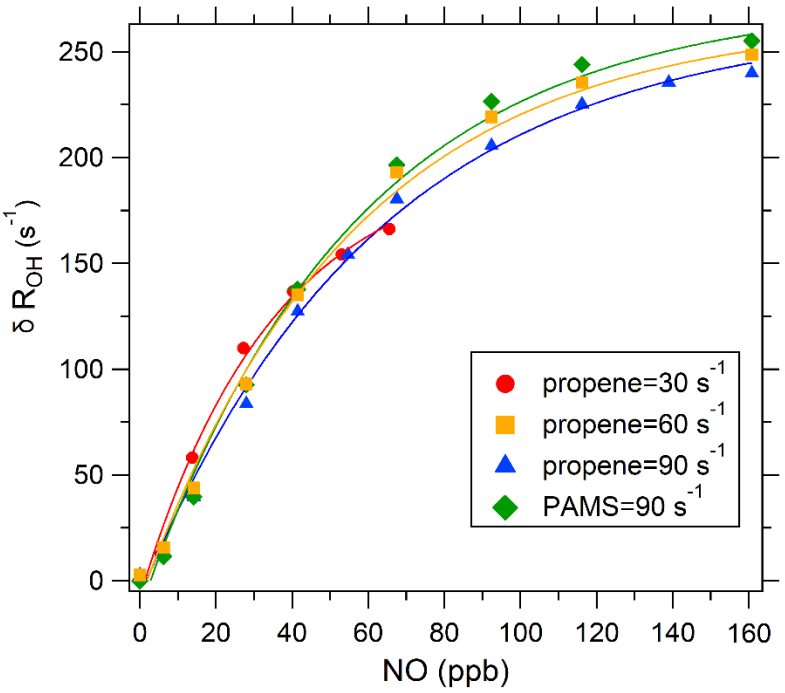

Figure S1. NO-correction experiments and fitting curves in Guangzhou in 2018 at different R_{OH} of propene standard gas and mixture standard gas. The mixture standard gas used is the mixture PAMS (photochemical assessment monitoring stations) of 56 non-methane hydrocarbons (NMHCs; SpecialGas Ltd, USA).

Line 177 – 178: Why don't use you include values for a, b, c and $C_{background}$ in your paper? A main point in your conclusion is how this is new method to use the CO with VOC_R to get at anthropogenic missing fraction but then you don't show any concrete numerical examples using this new method.

Reply: Thanks for your suggestion. We have provided the values for a, b, c and $C_{background}$ in the manuscript accordingly.

Lines 329-330: The fitted coefficient a is 0.031 s^{-1} ppb^{-1}, b is 0.012 s^{-1} ppb^{-1}, c is 1.8 s^{-1} ppb^{-1} and $C_{background}$ is 1.3 s^{-1}.

Lines 211-213: Did you ever have negative VOC_R? In other words, observations higher than the calculated missing VOC_R? It looks like you have some periods no visible gray in the figure for missing VOC, if negligible but still some 'missing' also good to point out when that happens as well as the general 20% since that is impressive given the issues you introduce in the beginning of the paper from previous work.

Reply: Thanks for your suggestions. Actually, the missing VOC_R was negative during some periods. This is probably due to the uncertainty in the measurements of R_{OH} and reactive gases. The negative missing VOC_R primarily occurred in the afternoon (12:00–17:00) when the photochemistry was most active. Most of the negative missing VOC_R values were larger than -5 s^{-1}.

Lines 273-276: In some periods the missing VOC_R was negative, which is probably due to the uncertainty in the measurements of R_{OH} and reactive gases. The negative missing VOC_R primarily occurred in the afternoon (12:00–17:00) when the photochemistry was most active.

Line 256: Include the reaction rate constants used

Reply: Thanks for your suggestion. We have included the reaction rate constants.

Lines 319-322: A higher ratio of ethylbenzene to m,p-xylene corresponds to a higher aging degree of air masses as the m, p-xylene has a larger reaction rate constant (18.9×10^{-12} cm^3 $molecule^{-1}s^{-1}$) than ethylbenzene (7.0×10^{-12} cm^3 $molecule^{-1}s^{-1}$) when reacting with the major oxidant - OH radicals.

Line 293: By what factor did you increase all the NMHC? Were they all increased the same amount? What was the process here?

Reply: Under the base scenario, the measured VOC_R of all 56 NMHCs are 4.6 s^{-1}. To consider the missing VOC_R (13 s^{-1}) in the model, concentrations of the 56 NMHCs were increased by a factor of 1.9, leading to an additional increase in VOC_R of both NMHCs and unconstrained secondary products, which exactly compensated for the missing VOC_R. Given that the VOC_R of unconstrained secondary products increases with the increase in the concentration of NMHCs, several attempts of different values are needed to determine the increasing factor.

Lines 249-260: The box model was used to evaluate the impact of missing VOC_R on the O_3 production rate. In the base scenario, the box model was constrained by all measured inorganic and organic gases but the missing VOC_R was not considered. To consider the missing VOC_R in the box model, we additionally increased the concentration of NMHCs to exactly compensate for the missing VOC_R by multiplying

a factor, on the basis of measured NMHC concentrations. We simulated four scenarios by increasing the concentration of: (1) n-pentane, (2) ethylene, (3) toluene, (4) all measured 56 NMHCs. For the scenario of increasing all 56 NMHCs, concentrations of 56 NMHC species were increased by multiplying the same factor. Given that the VOC_R of unconstrained secondary products increases with the increase in the concentration of NMHCs, several attempts of different values are needed to determine the increasing factor.

Lines in 350-357: The setting of model simulations for different scenarios are depicted in Section 2.6. Under the base scenario, on average the measured VOC_R of n-pentane, ethylene, toluene and all 56 NMHCs are 0.14 s^{-1}, 0.53 s^{-1}, 0.60 s^{-1} and 4.6 s^{-1} respectively. To consider the missing VOC_R (on average of 13 s^{-1}) in the model, four scenarios were simulated by additionally increasing n-pentane, ethylene, toluene and 56 NMHCs by a factor of 70, 16, 13.3 and 1.9, respectively. These increasing factors led to an additional increase in VOC_R of both NMHCs and unconstrained secondary products, which exactly compensated for the missing VOC_R.

Section 3.3: The writing is unclear about the sensitivity studies. Were the individual VOC species (represented by the 3 examples) and the "all measured NMHC" done together or 4 different sensitivity studies? I'm assuming it was 4 different model runs but as written that isn't apparent and it sounds like they were done all together. It wasn't until looking at the figure it seemed like 4 runs. In particular, line 293 "and the one considering" suggests one model run which means you didn't look at the impact of each species. Were the individual species taken from the measured results? Where did those numbers come from?

Reply: Thanks for your suggestions. (1) We simulated four different scenarios, rather than they were done all together. (2) "and the one considering" has been changed into "and the scenarios considering". (3) To consider the missing VOC_R in the box model, we additionally increased the concentration of NMHCs (individual species or all NMHCs) to exactly compensate for the missing VOC_R by multiplying a factor, on the basis of measured NMHC concentrations.

We have modified the corresponding sentences in the manuscript to make the meaning more clearly.

Lines 249-260: The box model was used to evaluate the impact of missing VOC_R on the O_3 production rate. In the base scenario, the box model was constrained by all measured inorganic and organic gases but the missing VOC_R was not considered. To consider the missing VOC_R in the box model, we additionally increased the concentration of NMHCs to exactly compensate for the missing VOC_R by multiplying a factor, on the basis of measured NMHC concentrations. We simulated four scenarios by increasing the concentration of: (1) n-pentane, (2) ethylene, (3) toluene, (4) all measured 56 NMHCs. For the scenario of increasing all 56 NMHCs, concentrations of 56 NMHC species were increased by multiplying the same factor. Given that the VOC_R of unconstrained secondary products increases with the increase in the concentration of NMHCs, several attempts of different values are needed to determine the increasing factor.

Lines in 350-357: The setting of model simulations for different scenarios are depicted in Section 2.6. Under the base scenario, on average the measured VOC_R of n-pentane, ethylene, toluene and all 56 NMHCs are 0.14 s^{-1}, 0.53 s^{-1}, 0.60 s^{-1} and 4.6 s^{-1} respectively. To consider the missing VOC_R (on average of 13 s^{-1}) in the model, four scenarios were simulated by additionally increasing n-pentane, ethylene, toluene and 56 NMHCs by a factor of 70, 16, 13.3 and 1.9, respectively. These increasing factors led to an additional increase in VOC_R of both NMHCs and unconstrained secondary products, which exactly compensated for the missing VOC_R.

Line 358: **Figure 4** shows the simulated $P(O_3)$ for the base scenario and the scenarios considering missing VOC_R.

Lines 347-348: in regards to the parametric equation “developed here” do you mean being able to separate them all out? It isn’t just “versus CO” according to equation 4 so this is misleading.

Reply: Thanks for your suggestion. We have modified this sentence to avoid misleading.

Lines 411-413: In addition, the parametric equation of missing VOC_R derived from MLR method (Eq (4)) here can be used to estimate missing VOC_R according to measurements of CO, O_X and isoprene.

Lines 348-350: "are also expected" doesn't make sense in the context

Reply: Thanks for your suggestion. We have removed this sentence.

Figure Comments:

Figure 1: Why are a and b blue but c red? If for primary vs secondary that isn't referenced in the caption so is irrelevant since it doesn't match the color scheme in c or d. For example in d, missing VOC is red but in c it would be secondary sources. In e, what are the green squares? You don't reference them anywhere. Why are the circles magenta/pink? Not necessary and detracting.

Reply: Thanks for your suggestion. We have changed the color to make consistent. In e, we changed the green squares to red squares, which represents the mean values of missing VOC_R in different ranges of ethylbenzene/m,p-xylene with classification width of 0.1.

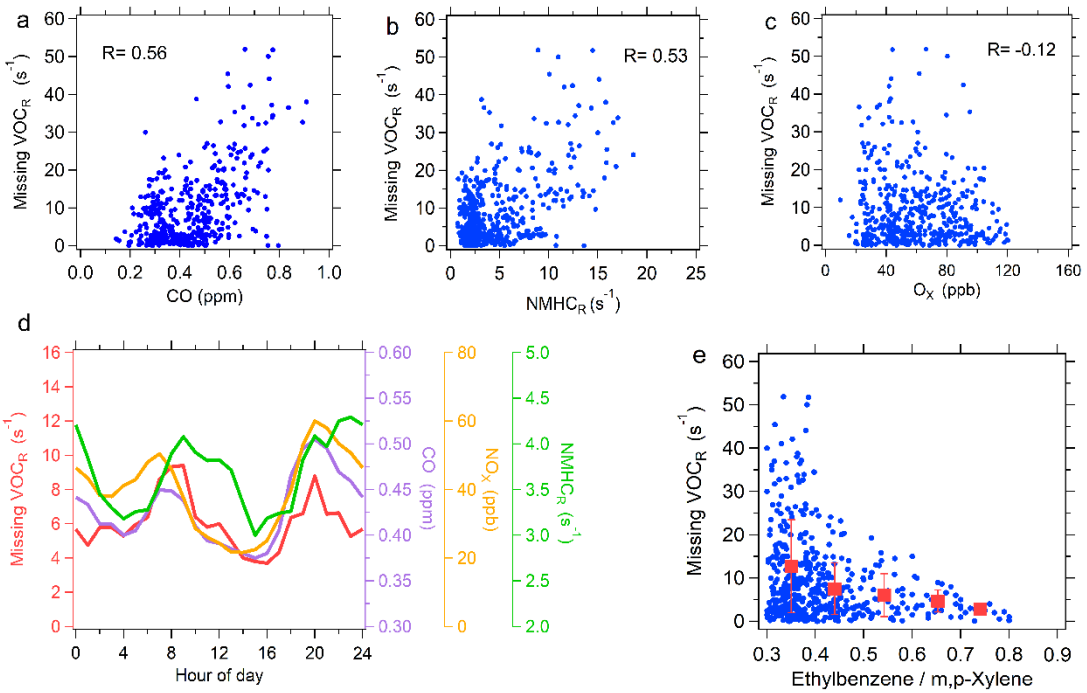

Figure 2. Correlation of missing VOC_R with major tracers during the whole measurement period. (a-c) Correlation of missing VOC_R with CO, OH reactivity of NMHCs ($NMHC_R$) and O_X. Each point represents hourly data. (d) Diurnal variations

in missing VOC_R, CO, NO_X and NMHCs. (e) The dependence of missing VOC_R on ethylbenzene to m, p-xylene ratio. The red squares indicate the mean values of missing VOC_R in different ranges of ethylbenzene/m,p-xylene with classification width of 0.1, and the error bars represent standard deviation.

Figure 2: Again the different colors for a and b seem unnecessary and then don't match c and are in fact opposite. CO = anthropogenic but it is blue and red in a and c, respectively.

Reply: Thanks for your suggestion. We think it is no need to provide the panel b (missing VOC_R vs O_X) because the correlation between missing VOC_R and O_X is poor and this information has been provided in Fig. 2c. Thus, we have removed the panel b in the updated manuscript. In addition, we change the color of the first bar in panel c to be blue to make it consistent with panel a.

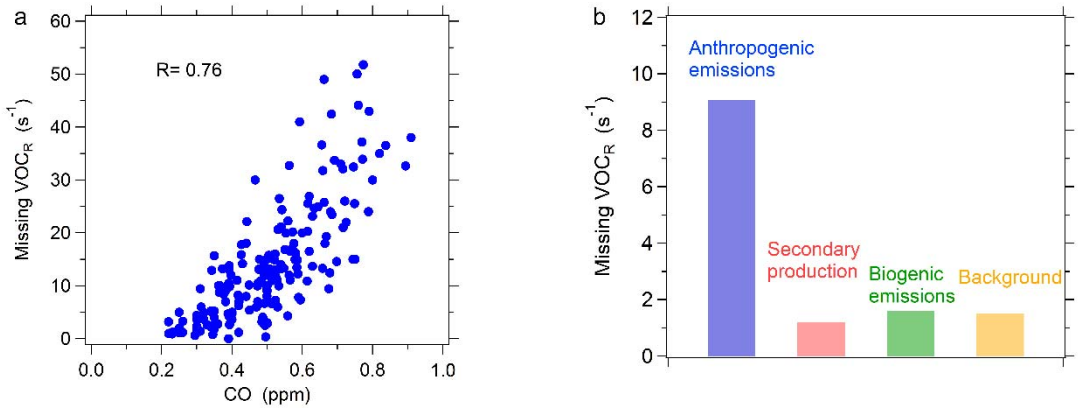

Figure 3. The source apportionment of missing VOC_R in high missing-VOC_R days. (a) Correlation of missing VOC_R with CO. Each point represents hourly data. (b) Contributions of different sources to missing VOC_R according to the MLR.

Figures 4 and 5: Nice use of colors here that tie the idea together.

Figure 5: there is no blue bar a but that is referenced in the caption. Also for a, it would be nice to note the dashed line represents the NO_x vs VOC limited regimes.

Reply: Thanks for your suggestion. We have modified it accordingly.

The caption of Fig. 5: (a) Diurnal variations in L_N/Q for the base scenario and the scenarios considering missing VOC_R. The missing VOC_R is considered by adding individual species (n-pentane, ethene or toluene) or increasing all measured NMHCs to

fill the missing VOC_R. The dashed line represents the threshold value of L_N/Q that distinguishes VOC-limited and NO_X-limited regimes.

Technical Corrections

Line 57: Indented but shouldn't be

Reply: Thanks. We have deleted the indent.

Line 68 and elsewhere: Perhaps a personal preference but the oxford comma can be very useful with complicated lists in sentences

Reply: Thanks. We have used the oxford comma for complicated lists in sentences.

Line 120: "thank to the" should be "thank**s** to the"

Reply: Thanks. We have revised it.

Line 131 and elsewhere: Be sure to have a space between a) and the next word

Reply: Thanks. We have revised it.

Line 140: "cannot" is the more common spelling for this

Reply: Thanks. We have revised it.

Line 166: "The multiple linear regression (MLR) have" is awkward and incorrect tense. Perhaps something like "Multiple Linear Regression (MLR) has been"

Reply: Thanks. We have revised it.

Line 228: "As the" should be "As a"

Reply: Thanks. We have revised it.

Line 247: no period between 22:00 and)

Reply: Thanks. We have revised it.

Line 256: The wording of "higher aging degree of air masses" is awkward. Perhaps "higher degree of aging air masses" or "higher degree of air mass aging" based on what you write below at 259

Reply: Thanks. We have revised it.

Line 287: and elsewhere: Be consistent with – and spacing, sometimes an extra space and sometimes not

Reply: Thanks. We have revised it.