Reply to Reviewer’s comments

General Comments

The paper by Yang et al. report CRM OH reactivity measurements conducted in two Chinese cities: Beijing in August 2013 over a 17 day period and in Heshan between 19 Oct – 22 Nov 2014. Using the measured datasets and a zero dimensional box model, the authors investigate the source of missing OH reactivity in Beijing (21%) and Heshan (32%) and also calculated the ozone production efficiency (defined as ratio of ozone production rate/NOx production rate). They conclude that the ozone production efficiencies (OPE) would be significantly underestimated at both sites (by 27% in Beijing and 35% in Heshan), if the OPE were not constrained by the measured OH reactivity. Hence they conclude that OH reactivity measurements are necessary for accurate determination of ozone production efficiencies.

The paper is interesting and has attempted deployment of the CRM technique in very challenging high NOx concentration environments. Such a study would certainly be of great interest to the ACP leadership from the perspective of fundamental process based understanding of OH reactivity and ozone production at two important sites in China. The efforts of the authors ought to be appreciated from this perspective and the study could be a valuable addition to the literature. However, there are some major technical concerns concerning the quality of the measurements and analyses, which need to be clarified/addressed/corrected by the authors before one can have confidence in the datasets and conclusions. The presentation and language also needs to be improved before it can be considered suitable for publication in ACP.

Response: Appreciate your comments. As you have said above, there’re a lot of problems in our previous manuscript, whether in technical part or language part. Wish these answers and latest version of the paper could help with it.

Major Concerns:

1) OH reactivity measurements: The authors have provided a good qualitative description of the CRM measurement system, but this description is generic for the CRM technique. In order to assess the quality of the measurements, relevant technical information pertaining to the operating conditions must also be provided. For example: What was the pyrrole/OH concentration ratio inside the CRM reactor during these measurements? Did it change between the deployments? What were the typical pyrrole concentrations for the C2, C0 and C1 stages? What was the dilution factor for ambient air inside the reactor (i.e. flow of syn air/amb air to total flow)? What was the residence time of air inside the reactor?

It appears that numerical simulations for calculating deviations from the first order conditions were neither carried out nor applied. Why? While it is good that the authors tested the accuracy of the system using propene, propane and a hydrocarbon mixture, it can be seen from the results (Figure 2) that the slopes obtained are rather different from propene (1.31) Vs propane (0.93 to 1.04). If the pyrrole (C1 concentration)/OH

(C1-C2 concentration) ratio was different in these calibration experiments, it could explain the same and then correcting for the deviations from first order conditions for the relevant pyrr/OH ratio, would ensure better accuracy and take into account correction factors for each type of standard.

Response: Thanks for asking. This is a mistake and in the latest version of manuscript, we have included some fundamental parameters of the CRM system. In the last paragraph of 2.1.1, we re-phrase it as below:
Ambient air or synthetic air was introduced at 160 -170 ml min^{-1} with the total flow 320 – 350 ml min^{-1}(The typical dilution factor was about 2-2.15 depending on the situation). The residence time of air inside the reactor was less than 30 s before they were pumped by the Teflon pump. The typical c1 mixing ratio for pyrrole in Beijing and Heshan measurements were about 60 ppbV and 55 ppbV, while the mixing ratios of OH radicals generated by mercury lamp were about 35 ppbV and 28 ppbV. The mixing ratios were quite consistent for either of the campaigns, respectively.
Also we have tried the FACSIMILE model for the deviation to achieve the equations 2-2 and 2-3.
Also your suggestion may be right, the pyrrole/OH ratios from propene and propane were different due to the slightly different ratios used in two periods of time. This would introduce uncertainty in our measurements.

2) NO-correction experiments: Figure 3: The results are rather strange. From the paper it is not clear whether the NO concentration shown on the x axis are the ambient NO concentrations or the NO concentrations inside the CRM reactor after dilution. The authors state that they have used the former (Lines 177-179; Page 6). This would be inappropriate to use considering the non-linearity of NOx-VOC chemistry for OH formation. Instead corrections are valid only for NO concentrations inside the reactor.

I am also puzzled that the concentration factor (delta s-1 on y axis) is the same in magnitude for 60 s-1 of OH reactivity as well 120 s-1 of OH reactivity at the same NO concentration. The NO correction should logically be lower at 120 s-1, as the additional reactivity would compete with NO more efficiently for the HO2 resulting in lowering the OH formation due to NO+HO2. A simple numerical simulation would reveal the same. At the large NO concentrations observed during their deployment (>10 ppb NO), it is difficult to trust that the highly non-linear secondary chemistry effects arising from NO+RO2 and NO+HO2 or even HONO photolysis inside the reactor can be accurately corrected. Note that the secondary chemistry is also occurring in a mixture containing ambient air that has several reactive compounds that have different chemistry from the standard mixtures used and some of the compounds are even unknown. In such scenario, relying on calibrations involving just propane and propene or a mixture of few compounds cannot yield robust corrections factors at high NO concentrations. For the high NO concentrations and OH reactivity conditions encountered in the present study, the authors should have used the flexibility of the CRM *method in terms of adjusting the dilution factor of ambient air inside the CRM reactor. They could have then ensured that the NO concentrations were at more few ppb inside the CRM reactor, keeping the effect of NO induced secondary chemistry and its magnitude much lower. In the context of the present study, applying corrections that maybe at times 100% or even higher in magnitude relative to the uncorrected OH reactivity measurements is rather disconcerting. It renders most of the reported OH reactivity a function of the correction factor! This effect is apparent in Fig 5a and 5b, when one looks at the correlation between the time series of NO and the corrected measured OH reactivity shown in the graph. The authors need to discuss this issure more comprehensively. In this regard, correlation plots of the measured OH reactivity (y axis) Vs the NO concentration (x axis) and the missing OH reactivity Vs NO concentration at both sites using the temporal data (5 or even 10 min averages would do), would help much to shed light on the above point. Also, it would be good to add the time series of calculated OH reactivity to Fig 5a and 5b for an idea of temporal variations in both the measured OH reactivity and calculated OH reactivity. At present only few days of time series data from both sites have been presented in Fig 11 and Fig12.*

The authors may also wish to consider excluding the ambient OH reactivity data for periods when the NO concentrations lead to corrections of the same magnitude as the uncorrected OH reactivity (100% or more). In that case, the present findings of unexplained OH reactivity may need to be revisited. At the very least, suggestions on how to perform the CRM experiments better in high NOx environments can be an outcome of the present study.

Response: Thanks for the advice. This is the key question to the CRM technique in this work. For the first question, the NO concentrations we used in the calculation were "ambient" NO concentrations, which were introduced into the system with synthetic air. We assumed this was the starting situation for the reaction and calculation. We did not have the NO measurements in the CRM reactor and could not quantify the concentrations well, thus we could not calculate the coefficients for the equations.

The second question remains unsolved even to us. We would expect the same situation as the referee did. The difference in quantities of reactants should have a different influence on the NO recycling and thus a different delta reactivity. However, after we've tried the experiments several times with different standard gases, we've found the correction curve were very close. We are still pursuing the principles underlying but we've not reached a persuasive one.

For the method, most of the colleagues suggest that we should deploy the advantages of flow adjustment. However, that advantage is very useful when the observation site is changed or the forecasts predict the coming of a heavy polluted air masses. However, like in Peking University Site, which was close to vehicular emissions, high levels of NO always come in several minutes and not enough time is left to flow adjustment and system stabilization. If we set the ratio of nitrogen to ambient air too high, maybe the OH reactivity of diluted samples will be too low to reach detection limits. But with our experiments, we believe in our corrections and employ the correction curve to obtain the measured reactivity. And for these two observations, the corrections in the manuscript presented the best we can do at that time.

For the relationship between measured reactivity and NO, we found similar diurnal patterns between measured reactivity and NO_x, which were dominated by NO_2 in both sites. For the comparison between measured and calculated reactivity, we would like to focus on the missing reactivity in some pollution episodes. However, if the referee insists, we could supply with the similar figures as Fig 5a and Fig 5b with calculated reactivity included.

Thank you very much for these three paragraphs discussion.NO interference is really the most important and annoying problem concerning CRM technique. For further researches, more effort will be put in the experimental and analyzing work. For experimental part, the experiments should be undertaken more carefully and systematically to figure out the key factors influencing the NO interference. For analytic part, we need to find one way to find the correction curve and explain it in a principle way.

3) NO_2 measurements and ozone production efficiency calculation in model using calculation NO_2: The authors mention (Lines 221-223; page 8) that the NO_x measurements were performed using the chemi-luminescence technique (Instrument Model 42i, Thermo Fischer Inc., U. S.). If this is the case then obviously their NO2 measurements would suffer from a positive bias (see for e.g. Reed et al. 2016; Atmos. Chem. Phys., 16, 4707-4724, 2016) due to interference from organic nitrates, nitrous acid (which the authors report was quite high during their study without showing the actual temporal data) and nitric acid. The magnitude of such differences in inter-comparison studies involving more specific NO_2 instruments have been shown to overestimate the NO_2 concentration by up to 400% during the daytime (see Chapter 2 by Kleffman et al. in Disposal of Dangerous Chemicals in Urban Areas and Mega Cities, NATO Science for Peace and Security Series C: Environmental Security, I. Barnes and K. J. Rudzin Ski (eds,), DOI 10.1007/978-94-007-5034-0_2, # Springer Science + Business Media Dordrecht 2013. The authors mention that they used that measured NO_2 to constrain the model (lines 246-248; page 9). If the measured NO_2 is significantly in error (say 50%), how would it affect the results of their analysis keeping in mind the implications for the unexplained OH reactivity and the calculation of the ozone production efficiency using $P(NO_z)$ (Equation 2-3). This could affect their conclusions majorly and the authors should address their concern.

Response: Thanks for the advice. This recent piece of literature really brings a big question about the reliability of NO_2 measurements. Especially in our calculation and analysis, the contribution from NO_2 to total OH reactivity is very large. However, we could only supply with one piece of supportive evidence for our manuscript. In Heshan, we have two set of NO_x analyzers. One was equipped with a home-build photolytic converter (NO_x-PL) and the other one was equipped with a catalytic converter (NO_x-Mo), both were similar as the setups of the instruments in Tan et al (2017). This comparison could help us to figure how great the interference could NO_z have on the measurements. From the data comparison, we found that the datasets were close within 10% for most of the time but could be over 20% when in some nights or morning time. However, we have no similar inter-comparison in Beijing measurements. However, this could be a great interference and should call our attentions in later research.

Other Comments:
Abstract:
(Line 22; Page1) and elsewhere in the paper: The authors should report the measured OH reactivities and other concentration measurements by rounding off to significant figures.

Response: Thanks for the advice. Modified as suggested.

(Line 34; Page 2): English is wrong: "… was presumably attribute to oxidized species…"

Response: Sorry for this mistake and revised as suggested.

Introduction:
Equation 1-1: The concentration is expressed as small $[x_i]$*. Please replace by capital* X_i *as in subscript here and elsewhere.*

Response: Thanks for the advice and changed as suggested.

Line 46, Page2: should be "reactive" instead of "reductive"

Response: Thanks for asking. However, we think the species here are the species could react with OH radicals, which should be reductive species. Appreciate your suggestion anyway for this clarification.

Lines 50-61: The authors should include other more recent relevant measurements of OH reactivity from another suburban site in Asia in Table 2, report by Kumar, V et al., Int. J. of Mass. Spectrom., 374, 55-63, 2014.

Response: Thanks for the advice about this latest results and included in Table 2 in last manuscript.

Line 108; Page 4: Although Yang et al. 2016 is cited, Sinha et al., 2012 were the first to outline this approach and use it for determining ozone production regimes.

Response: Thanks for advice. Modified as suggested.

Line 125-126: Page 5: Mention the inlet residence time. What sorts of inlet filters were used? How often were they changed under such polluted conditions?

Response: Thanks for these details. The last sentence of this paragraph was modified as below: Ambient air was sampled after a teflon filter and then pumped through a 14.9m Teflon 3/8 inch (outer diameter) inlet at about 7 L·min-1 rate, with a 5 - 6 s residence time.

Line 159-164; Page 6: Humidity adjustments to synthetic air for matching ambient humidity changes: It is difficult to imagine how the simple need valve contraption can dynamically track and adjust to the ambient humidity. Some data showing the m39 water cluster concentrations of the PTR-MS for the C2 and C3 stages would be helpful.

Response: Thanks for asking. This is really an important part for the work. The needle valve we used in the experiments could not adjust itself to certain humidity. However, once we set it to certain position and it could remain the relative humidity for quite a long time. However, as it was tuned manually rather than automatically, it could not track the ambient humidity and could only be adjusted by the operator when he was free to check the humidity and adjust it. That was also why the correction equation was needed for further corrections. The figure for one example is supplied in latest version of supplement information.

Line 167; Page6: Spelling error: "generated"

Response: Thanks for the careful check and modified as suggested.

Lines 185-201: HONO interference: The authors do not show or mention anywhere the actual HONO concentration measurements. Considering that the OH concentration inside the CRM reactor is typically several tens of ppb, it is difficult to understand how typical ambient HONO concentrations of few hundred ppt can cause a significant interference, through the mechanism outlined by the authors. At this stage I am beginning to wonder if the authors have a fair assessment of what goes on inside the CRM reactor.

Response: Thanks for asking. First, HONO results ranged from several tens of pptV to several ppbV in Heshan campaign. However, the data was under-preparation for publications now so I did not present the figures in this paper. Second, while the HONO was just 100 pptV or so, the influence was not very significant. However when the ambient HONO mixing ratio was over 5 ppbV, which was observed in the campaign, the influence should not be omitted. Moreover, the HONO peak came out several times with low reactivity periods, which called for more careful corrections.

In general, the figures, their captions and legends need to be improved.

Response: Thanks for suggestions and modified.

In the supplement Table S3 and S4 are identical except for the captions!
In the supplement Tables S5 and S6 are identical except for the captions!

Response: Thanks for the check and very sorry for this careless mistake. Modified in the new manuscript.

Section 2.3.1: How is the dilution accounted for in the model? What is the model estimated concentration of PAN and its production rate? The latter would have bearing on their assumptions of approximating P(NOz) as P(HNO3)

Response: Thanks for this question very much. Referee 3 also brought this question, which was a big problem in previous calculation. In the new version, we include the organic nitrate part in the calculation and the results could be a little different.

Section 3.1: lines 270-273, Figure 5-a: During some periods (e.g. evening of 17-08-2013 and 24-08-2013) CO mixing ratios are close to zero in Beijing. How can one have trust in such measurements?

Response: Thanks for asking. We've checked the data. There're some points too low which we think the data was influenced by the zero-calibration. However, for most of the data, we think they're valid. I think some of the points appeared very low in the figure due to the large axis. When the CO mixing ratio was about 100 ppbV, you could not find the point in the figure. However, the suggestion is important for us.

Section 3.1: Figure 5-a: There seems to negative mass concentration of PM2.5 on 17, 19, 23, 25 and 26 October evening which are masked because the axis starts from zero. Please explain.

Response: Thanks for asking. We're using the TEOM for PM2.5 measurements, and our instrument could not measure precisely when the PM2.5 concentration was very low. So there're some points below zero but that makes no sense so we delete the data below zero and then we got the figure as you saw.

Figures 6: I cannot make our much as the it is difficult to read as the Figure legends are hardly legible. There does not seem to be much difference between the peak NO/NO2 ratios between Beijing and Heshan. Then how can one be sure that Heshan is influenced more strongly by transported air masses (see Lines 301-303)? Please clarify.

Response: Thanks for the question. The figures were modified in the latest version of manuscript. NO/NOx ratio was lower in Heshan compared to Beijing, and the other factor was the time when NO went down to near zero. With these two factors, and also other explanation listed after this paragraph, we assume that Heshan is influenced by transported air masses compared to Beijing.

L317 and 319: The unit of OH concentration should be molecules cm-3 (and not moel cm-3). Similarly in L314 and 316, mole should be replaced by molecules.

Response: Thanks for the advice and revised as suggested.

Figure 7a and b: Please mention what the red dotted lines and shaded regions signify?

Response: Thanks for the suggestion and modified in the figures.

Figure 8: Use appropriate legends? Not clear what is meant by "reac_mea_median"

Response: Thanks for the suggestion and modified as k_{mea}_median and explained as the median results of measured reactivity.

L338-340: For Heshan, another pollution episode can be observed on 5th November despite lower CO, NOx and other pollutants. Please mention that too.

Response: Thanks for mentioning that. This was correct but we select the two episodes for the accumulation of secondary pollutants.

L348-352: Reasoning is not clear: This contradicts the previous statement that Heshan receives more aged air masses. Aged air masses should also show more contribution from OVOCs. Moreover if faster photochemical production during August in Beijing, is really the reason, then one would expect less contribution from primary hydrocarbons in Beijing and more contribution of primary hydrocarbons in Heshan.

Response: Thanks for the question. I think the expression in previous manuscript was not good. So we re-phrase it with no expression comparing the OVOC percentages in both sites. Your suggestion is right that the measuring seasons were different in both sites which would significantly influence the secondary species production rate. However, there was another possibilities lying in the missing reactivity. However, the previous expression was inappropriate. Appreciate the advice.

L425: Please provide the major contributors among unmeasured aldehydes. L437 and Table S5 and S6: Please provide full name of "ALD", "GLY+MGLY", "ISOP" and "DCP" at least once. As mentioned in the text, HCHO is a measured species. Please provide the value of k_{cal} due to HCHO in table S5 and S6 also for comparison.

Response: Thanks for the advice. In the latest version, we explained the species concerning these four categories. For the HCHO part, we've included HCHO as an input for the model, so the k_{cal} and k_{mod} was almost the same and thus we did not present the k_{cal} in the table. However, appreciate the question.

Section 4.3: Please mention the range of OPE or an average OPE in the text also to *provide an estimate of number of ozone molecules produced per molecule of NOx consumed.*

Response: Thanks for the suggestions and modified as advised.

Lines 464-469: The argument is circular: First the authors assume that the missing reactivity is due to alkenes and OVOCs which have high ozone production potential. Next the model is constrained by the "assumed" species in the second scenario. Obviously these species will provide extra ozone production.

Response: Thanks for asking. I am not sure about the logical question here. We have mentioned two probable explanations in 4.2 and 4.3 and then take them into consideration in 4.4 to calculate the OPE here. If we exclude these alkenes and OVOCs, then the calculation would appear un-constrained. However, I am not sure whether we should take other species into consideration.

The LOD reported for the CO, NOx and O_3 *instruments in the supplement are much less than what is claimed by the instrument manufacturers in their manuals (40 ppt for NOx, 1 ppb for* O_3*, and 40 ppb for the 48i trace level enhanced CO analyzer). Please provide the details of how you obtained a lower LOD for these instruments or correct the Table in the supplement.*

Response: Thanks for asking. However, after checking with our engineers, these were typical running conditions for these instruments for several campaigns since CAREBEIJING-2006 and PRIDE-PRD2006 (Lu et al 2012; 2013).