Response to Referee #2:

Thanks very much for your comments, suggestions and recommendation with respect to improve our paper. The response to all your comments are listed below. There was an extensive discussion among the authors regarding how to revise the content, and this paper is subjected to a major revision including an update of all retrievals using new inputs (e.g., S_a based on standard deviation of a dedicated WACCM run from 1980 to 2020), re-plot all figures, condense/reorganize the content and focus more on the scientific topics. Thus, the response is delayed, and we are sorry for this.

Summary:

The authors report on solar absorption FTIR measurements of tropospheric columns of O3, CO, and HCHO at a candidate NDACC IRWG observation station in Hefei, China. High spectral resolution measurements were conducted between 2014 and 2017 and fill a data gap within the NDACC observation network. The data shows higher tropospheric O3, also with higher variability, in spring and summer. The authors compare these O3 measurements to OMI satellite O3 (PROFOZ product), as well as GEOS-Chem (2 x 2.5 deg) and WRF-Chem (20 x 20 km) model O3 outputs. Comparisons are done in both profile and tropospheric partial column form.

Ozone FTS vs. GEOS-Chem model differences (481 coincidences) are attributed to uncertainties in GEOS-Chem input files ("ozone production loss rates and emission inventory"), it is concluded that GEOS-Chem is biased 13% lower (along profile), with r=0.5 for tropospheric column correlation plots.

Ozone FTS vs. WRF-Chem model differences (481 coincidences) are attributed to uncertainties in WRF-Chem input files ("ozone production and loss rates and MEIC inventory"), it is concluded that WRF-Chem is biased 12% lower (along profile), with r = 0.65 for tropospheric column correlation plots.

Comparisons to coincident OMI ozone profiles and partial (tropospheric) OMI columns were done on 53 coincident measurements after filtering for 0.7 ° spatial coincidence. Coincident FTS profiles were averaged in a 3 hour window around the OMI overpass at 13:30. OMI profiles were smoothed with FTIR averaging kernels. The OMI profiles are biased 2-13% lower than FTIR profiles, with r=0.73 for

tropospheric column correlation plots, in which most OMI points sit below the 1:1 line, indicating also a low bias of OMI w.r.t. FTS.

Both sets of model ozone data are described as "smoother" than FTIR data and are "bias corrected" by adding a constant offset to the tropospheric O_3 columns throughout the year to shift the model data towards FTIR partial column values. GEOS-Chem partial columns are increased by ~100% while WRF-Chem partial columns are increase by ~33% to increase agreement with FTS. Finally, OMI ozone partial column data were increased by ~20% and only then were monthly mean ozone partial column differences calculated.

24-hour back trajectories were calculated arriving at Hefei at 3000 m.a.s.l. from 2014-2017, presumably for those days with FTS observations (?), and they were grouped into spring/summer (presumably MAM/JJA?) and autumn/winter (presumably SON/DJF?). Summer transport is less vigorous and more varied than winter transport, as expected, bringing more air from highly polluted areas, e.g., east China, and broadly accounting for higher O3 and higher O3 variability in the data in spring/summer.

Finally, the O3 production regime is analyzed by describing correlations to meteorological variables recorded at Hefei, as well as looking at O3 vs. CO, O3 vs NO2 (for ratios of HCHO/NO2 > 2.8, assumed to correspond to NOx-ltd O3 production) and O3 vs HCHO (for ratios of HCHO/NO2 < 1.3, assumed to correspond to VOC-ltd O3 production). The ratios to indicate the O3 production regime were found iteratively until the correlation between O3 and NO2 or O3 and HCHO was > 0.6. 106 days of observations (O3, HCHO, CO from FTS; NO2 from OMI) were identified and of those 60% were NOx-ltd, 11% were VOC-ltd, and the remainder were mixed.

Major comments:

The paper is generally well written and presents a thorough error budget and sensitivity analysis of FTIR retrievals (O_3, CO, HCHO) from a new candidate station in the NDACC network. The methods used here are well known and figures 2-5 should also move to the appendix, along with the Rodgers & Connor formulation,

unless the authors highlight how their averaging kernels and error budget profiles differ from other similar published results. The paper presents a valuable new and growing observational dataset, however, this reviewer recommends major revisions in order to meet the ACP criteria of scientific significance and quality.

Response: This paper has been subjected to a major revision based on the comments from three referees. All your comments are appreciated and have been addressed in the revised version. Main changes/improvements are listed as follows:

1) We have updated all retrievals with new S_a deduced from standard deviation of a dedicated WACCM run from 1980 to 2020, which should be more close to actual natural variation compared to the previous version. This improvement doesn't change the results of this paper.

2) We have reorganized the paper's structure, with less focus on known results and more describing about what is scientifically new. The objectives of the paper are clarified and listed in a concise way. The number of figures is reduced to focus more on the main scientific results. We have condensed quite a lot the descriptions of site/instrument, retrieval, theoretical basis but added many discussions/explanations regarding the observed results and photochemical regime. The figures and descriptions that are useful for understanding this paper but not scientific new are now shifted to the supplement (e.g., previous figures 2 - 5).

3) After an extensive discussion among the authors, we deleted all paragraphs and figures regarding comparisons with the correlative data, i.e., OMI, GEOS-Chem and WRF-Chem data, due to the following reasons:

a) The scientific topic of our manuscript is the investigation of the ozone seasonal evolution, source and photochemical production regime in polluted eastern China. The main interesting message we would like to present is the application of the FTS tools to determine if the tropospheric O_3 is produced by NOx or VOC, and give a recommendation about what could be done to mitigate the high O_3 levels. This can not only improve the understanding of regional photochemical O_3 production regime, but also contributes to the evaluation of O_3 pollution controls. In the revised version, we leads straightly to this recommendation. For things which are not important for the

main message, especially the deviation or something which probably misleads a potential reader, are removed. Accordingly, we removed the comparison with the models and the satellite.

b) This topic regarding comparisons with the correlative data, i.e., OMI, GEOS-Chem and WRF-Chem data, is interesting, but it cannot be clarified clearly within a few sentences or paragraphs and is basically a separate paper. Considering that this paper is already very long (referee's comments), we keep the intention of investigating the ozone seasonal evolution, source and photochemical production regime and removed all comparison with the correlative data.

4) We have responded to all referees' comments point-by-point and revised the manuscript accordingly.

Related change: The changes/improvements listed above have been done in the revised paper.

The FTS O_3 measurements are higher than both models (global and regional) and the OMI measurements. The FTS measures a total column through a particular atmospheric slant column, and is expected to be less sensitive to local O3 events than an in situ sensor. We expect generally good agreement with downward-looking OMI, although coincidences are always a challenge. We also expect differences in the FTS vs. model comparisons because of different representativeness offered by a 20x20 km model vs. as 2.5 °x 2 °model. This is not discussed in the paper.

Also, for the 20x20 km WRF-Chem model, the profile up to 10 km could extend over two horizontal grid boxes for most SZAs > 45 °,depending on the location of Hefei within a model grid box. Has this been considered?

Without discussing representativeness, the authors attribute FTS vs. model differences to model "input files", e.g., "ozone production loss rates and emission inventory" which is superficial. As a consequence, we learn little, if anything, about specific model processes and emission inventories that may be responsible.

Also, why is the data from this candidate station considered as "truth" in the comparison to OMI and the models? The total errors are estimated as 10% but they are dominated by smoothing error and based on very tight Sa values for O_3 (10%), so

(as the authors note), they are an underestimate.

If the authors plotted OMI vs. FTS trop O_3 column data with both data sets' error bars they would still not overlap, but presumably OMI data has been validated – is it generally found to be low compared to other data?

The addition of a simple offset to model O3 values before looking at fractional monthly mean differences w.r.t. FTS is problematic because it is evident in figures 9 and 11 that such a simple manipulation does not bring the data points onto the 1:1 line. Instead, we have the highest O3 values below the FTS measurements and the lowest values above. This is even more dramatic in GEOS-Chem data, presumably because of lower model resolution, which homogenizes high O_3 values over a large grid cell, while raising the background O3 values.

Since the highest values occur in spring/summer and the lowest in autumn/winter, the bias is seasonally dependent and therefore not just due to spatial representativeness. Is it due to incorrect emissions or chemistry?

What are the main chemistry and emissions differences between the two models being compared to FTS? WRF-Chem is running with the MEIC inventory, presumably optimized for China, as well as biogenic emissions from MEGAN – why does it only do a little bit better than GEOS-Chem?

About smoothing the OMI profile by the FTIR averaging kernels, this method is meant to be applied to high vertical resolution correlative data, which OMI is not. It has about ~1 DOF in the troposphere itself. This may explain why there is still a lot of "shape" left in the fractional difference between FTIR and smoothed OMI profiles. What do OMI kernels look like and where is its peak of sensitivity – is it the same as for FTS?

Response: After an extensive discussion among the authors, we deleted all paragraphs and figures regarding comparisons with the correlative data, i.e., OMI, GEOS-Chem and WRF-Chem data. Now all these problems don't exist in the revised version. Please check above clarification (page 4) for the reason.

Related change: Please check the revised version for details.

The trajectory cluster analysis is difficult to follow without familiarity with China's

geography. That can easily be fixed by adding the major city or region names referred to in the discussion to Figure 13. Without this information, it is hard to quickly judge if 1-day trajectories are long enough for transport to occur to Hefei. It is also not clear how the trajectories are clustered and the mean cluster trajectories (in color) are hard to see. Another way to represent this data would be to count trajectory elements crossing, e.g., 0.5 ° x 0.5 ° grid boxes. Also, why 3000 m? That seems much higher than the typical boundary layer height in winter, and probably also in summer. This choice will influence strongly both the speed and footprint of the pollution regions influencing Hefei. Have the authors tried 1500 m?

Response: In the revised version, all your comments regarding coincident trajectory cluster analysis have been addressed. Now we used 1500 m a.s.l. While the relative contribution/direction of each trajectory changes a little bit, the main point is still the same.

Related change: Now the height is 1500m, and China's geography is included. Please check figure 2 in the revised version for details.

Finally, regarding O3 production regimes, ratios of HCHO/NO2 were varied until the correlation was > 0.6 in plots of O3 vs. HCHO and O3 vs. NO2. The outcome is that the correlation for the NOx-ltd plot of O3 vs. NO2 is 0.66 (moderate) while the correlation for the VOC-ltd plot of O3 vs. HCHO is 0.92, with far fewer points remaining in the fit. This seems rather arbitrary and needs justification. Also of the 106 days available for this analysis, which are from spring/summer and which are from autumn/winter? Are all VOC-ltd days in winter?

Response: a) The previous figure (R = 0.919 with 8 points) was only used to demonstrate that PO_3 is more sensitive to VOC within VOC-limited region. Actually, the transition occurs close to about 0.6 (not 0.919). At the transition ratio, there are many more points than 8. In the revised version, a detailed description of obtaining the transition threshold is presented, this kind of subfigures (only used for demonstration) are all removed. Briefly, we iteratively altered the column $HCHO/NO_2$ ratio threshold and judged whether the sensitivities of tropospheric O_3 to HCHO or NO_2 changed abruptly. For example, in order to estimate the VOC-limited

threshold, we first fitted tropospheric O_3 to HCHO that lies within column $HCHO/NO_2$ ratios < 2 (an empirical start point) to obtain the corresponding correlation/slope, and then we decreased the threshold by 0.1 (an empirical step size) and repeated the fit, i.e., only fitted the data pairs with column $HCHO/NO_2$ ratios < 1.9. This has been done iteratively. Finally, we sorted out the transition ratio which shows an abrupt change in correlation/slope, and regarded this as the VOC-limited threshold. Similarly, the NO_x-limited threshold was determined by iteratively increasing the column $HCHO/NO_2$ ratio threshold till the sensitivity of tropospheric O_3 to NO_2 changed abruptly.

The transition threshold estimation using this scheme exploits the fact that O_3 production is more sensitive to VOCs if it is VOCs-limited and is more sensitive to NO_x if it is NO_x limited, and it exists a transition point near the threshold (Martin et al., 2004). Su et al. (2017) used this scheme to investigate the O_3-NOx-VOCs sensitivities during the 2016 G20 conference in Hangzhou, China, and argued that this diagnosis of PO_3 could reflect the overall O_3 production conditions.

b) Table 4 and the last paragraph of section 5.3.2 present detailed description of classification for these 106 days measurements. Not all but ~ 75% VOC-ltd days are in winter.

Related change: This problem has been addressed in the revised version. Please check section 5.3 in the revised version for details.

When I look at the full O3 data in Figure 12, I wonder why there isn't a stronger signature of JJA O3 enhancements in Hefei? (Is it related to filtering out days affected by haze, App B?) Many high values seem to be in May, although the x-axis is hard to read and should really be changed to, Jan 1, June 1,etc., throughout the paper where dates are shown. Or possibly at boundaries between MAM, JJA, SON, DFJ, if these are the groupings for the seasons in the paper.

Response: a) Compared to other high resolution FTS sites, the O_3 measurement in Hefei in JJA are very high, and we observed higher day-to-day variations in summer than other seasons. Vigouroux et al. (2015) studied O_3 trends and variability with

eight NDACC FTS stations that have a long-term time series of O_3 measurements, namely, Ny-Ålesund (79 ° N), Thule (77 ° N), Kiruna (68 ° N), Harestua (60 ° N), Jungfraujoch (47 ° N), Izaña (28 ° N), Wollongong (34 ° S) and Lauder (45 ° S). All these stations were located in non-polluted or relatively clean areas. The results showed a maximum tropospheric column in spring at all stations except at Jungfraujoch which extended into summer. This is because the stratosphere troposphere exchange (STE) is most effective during late winter and spring (Vigouroux et al. 2015). We don't think there isn't a stronger signature of JJA O_3 enhancements in Hefei is related to filtering criteria which are used to guarantee the data quality. It is most probably because the STE process is weaker in summer, though photochemical O_3 production is higher. Thus, tropospheric O_3 (STE fraction plus photochemical production fraction) in JJA is not the highest.

b) "June" and "MAM, JJA, SON, DFJ" have been used in the revised version.

Related change: "June" and "MAM, JJA, SON, DFJ" have been used in the revised version. Please check figure 1(b) in the revised version for details.

Have the FTS partial columns been compared to in situ O_3 monitors in Hefei to see if they also show enhancements in May/June 2015 and 2016? What about the low values in Jan 2015 and 2017 vs. the higher ozone in Jan 2016?

Response: We did not compared the FTS to in situ O_3 data due to a lack of co-existing in situ O_3 measurements. The O_3 variations in Jan 2016 are higher compared to Jan 2015 and 2017, most probably because of higher air pollution.

Related change: None

Finally, the Pearson coefficient of 0.35 – 0.6 was taken to mean "moderately correlated" in this work. Typically moderate correlation is associated with values of 0.5 – 0.8, since the lower bound would mean that the model fit to the data explains only 25% of the variations in the data. At 0.35 that drops to only 12%.

Response: In previous version, we regard it as good correlation if the correlation is larger than 0.6, and regard it as moderate correlation if the correlation lies in between 0.4 and 0.6. However, in the revised version, we only present the numbers and don't use a description such as "good" or "moderate" or "poor".

Related change: All these statements have been removed. Please check section 5 in the revised version for details.

Further detailed technical comments:

Fig. 1a: most names in this figure are illegible. Use a cleaner map to reduce clutter.

Fig. 1b: no red hexagons are visible, but I assume the red arc is the azimuth and the un-described yellow circles are the SZA.

Response: In order to focus on the main objectives, the content has been shortened quite a lot. We removed this figure in the revised version. Detailed site/instrument descriptions can be found in our previous paper (Yuan et al.,2017; Wei et al., 2017).

Related change: This figure has been removed. Please check section 2 in the revised version for details.

Fig. 2: what does "with measured ILS" mean in this caption? Is the ILS characterized with linefit and then fixed in the retrievals, or are some ILS parameters still being retrieved? Why is there a loss of sensitivity to HCHO right at the surface? Is this a priori related?

Response: a) In sfit4, the ILS can be treated with three options: one is assuming an ideal ILS, two is retrieving the ILS together with the trace gas retrieval, and three is using the measured ILS. We regularly used a low-pressure HBr cell to monitor the instrumental line shape (ILS) of the instrument, and included the measured ILS in the retrieval.

b) It is not a priori related but a characteristic of the HCHO retrieval. The sensitivity at the ground is low because of the very weak absorption feature of HCHO. The spectral signature at the ground is very broad, thus in the presence of noise very indistinguishable from the features created by the interfering species. The previous figures 2 and 3 have been shifted to supplement. Now is figures S2 and S3.

Related change: A statement " We regularly used a low-pressure HBr cell to monitor the instrument line shape (ILS) of the instrument and included the measured ILS in the retrieval." has been included in the revised version. Please check section 3.1 in the revised version for details.

Fig. 3: the HCHO trop column AK seems unhealthy for growing so far past 1 quickly

above ~3 km, even if there is little HCHO there. What is the reason for this shape?

Response: We find a bug in our previous plotting script. In the revised version, we fixed this bug and now this problem doesn't exist. This bug only for PAVK plotting, and has no influence on retrieval. Thus, every deduction is the same as before. Thanks for point out this bug. Please check figure S3 for details.

Related change: It has been shifted to supplement.

Fig. 4: What is the explanation for the peak in the CO error at around ~3 km?

Response: This is due to smoothing at around 3 km.

Related change: It has been shifted to supplement.

Fig. 5: Legend seems reversed for total random error and z shift for CO.

Response: We plot the three gases (O_3, CO, HCHO) using the same script, and after a careful check with our plotting script, we find there is no problem between the total random error and z shift for CO.

Related change: It has been shifted to supplement.

Fig. 9 and 11: it's hard to judge seasons with the date labels as presented. Also, why do these figures not have the identical number of O3 data points if they are derived from the same data filtering applied to FTS data that is described in App B?

Fig. 14: is based on Fig 12, not 13 as the caption says. Again, what are the model process and inventory differences leading to this? Panel a) says smoothed model, but is OMI not also smoothed in this figure?

Response: After an extensive discussion among the authors, we deleted all paragraphs and figures regarding comparisons with the correlative data, i.e., OMI, GEOS-Chem and WRF-Chem data. Now all these problems don't exist in the revised version. Please check above clarification (page 3) for the reason.

Related change: Please check the revised version for details.

Fig. 15: The wind sensor appears to be installed in a poor location as the wind speed never exceeds 0.3 m/s or ~1 km/h! If that is the case, then the wind direction data is also spurious. That's too bad, because I wanted to see a plot of Hefei O_3 vs. wind direction to see if O_3 is higher when winds blow from the city.

Response: The weather station gives an output every 10 seconds, but the previous

figure 15 only presents the daily average data that coincident with O_3. The wind direction and wind speed are **vector**s, thus, the averages are quite different compared to the short term data. The changing wind direction is the reason why the daily averaged wind speed seems never exceeds 0.3 m/s or ~1 km/h, and not because the wind sensor in a poor location. The figure 3 in the revised version, which presents minutely, hourly, daily, and monthly averaged data, illustrates the features better. For minutely- averaged data, the wind speed can exceed 6 m/s.

Wind direction is also important because it affects pollution transport, giving rise to high O_3 in downwind locations (Wang et al., 2016). The city downtown locates in eastern of the observation site and the majority of the Chinese population lives in the eastern part of China, easterly winds (direction less than 180˚) could generally transport more pollutants to the observe area than westerly winds (direction larger than 180˚), resulting in a higher O_3 level.

Related change: Minutely, hourly, daily, and monthly averaged data are included. Please check figure 3 for details.

Fig 16: In spite of problems above, the highest O_3 values occur for the lowest of the low wind speeds, pointing to the accumulation of local pollution. There is a "moderate" negative correlation between O3 and RH – why? We could learn more if these data were colored according to spring/summer and autumn/winter.

Response: The data are now color coded into spring/summer (MAM/JJA) and autumn/winter (SON/DJF) groups in figure 4 in the revised paper. We have fitted the minutely, hourly and daily average data with the coincident O_3, and all of them showed weak negative correlation between O_3 and RH. Elevated O_3 concentrations generally occurs on days with dry condition, low pressure and low winds in Hefei probably because these conditions favor the accumulation of O_3 and its precursors.

Related change: The data are color coded into spring/summer (MAM/JJA) and autumn/winter (SON/DJF) groups in figure 4 in the revised paper.

Fig. 19: hard to judge seasons with x-axis labels. Panel b is based on data in panel a that does not seem to sample seasons evenly. This should be discussed.

Response: In the revised paper, we only present the time series of column

HCHO/NO_2 ratios (figure 7), and the detailed discussion for PO_3 limitation is listed in table 4. The HCHO and O_3 are not retrieved within the same spectra, which means a measurement day that has a robust HCHO retrieval does not always has a robust O_3 retrieval, vice versa. The previous figure 19 (a) presents all days that have robust HCHO and NO_2, and figure 19 (b) presents the days that have robust HCHO, O_3 and NO_2. The criteria in figure 19 (b) is more stricter than figure 19 (a), and thus seems don't sample seasons evenly.

Related change: The previous figure 19 (b) is removed in the revised version and the detailed discussion for PO_3 limitation is listed in table 4. Please check figure 7 in the revised version for details.

Table 1: retrieved interfering gases →as columns, I presume, except for H2O, as noted? Also, WM →MW. I'm not sure what footnote b means, please clarify.

Response: The rows for H_2O has been deleted, WM is changed to MW. All footnotes has been removed because we think they are not necessary.

Related change: Please check table 1 for details.

Manuscript:

P1L74: sun spectra → solar absorption spectra

Response: This paragraph focuses on descriptions of the NDACC network. In the revised version, I removed the whole paragraph since it doesn't have much contributions to the main point of this paper.

Related change: This paragraph has been removed.

P1L3: replace wiki reference with something from the many, many refereed papers on Chinese modernization and growing air pollution problems.

Response: This sentence has been removed in the revised version.

P4L89: what are China's AQ standards in ppb for long- and short-term exposure?

Response: Tropospheric O_3 was already included in the new air quality standard as a routine monitoring component (http://www.mep.gov.cn, last access on 23 May 2018), where the limit for the maximum daily 8 h average (MDA8) O_3 in urban and industrial areas is 160μg/m^3 (~ 75 ppbv at 273 K, 101.3 kPa). Please check line 120 in the revised version for details.

P4L95: greatly contribute to ozone pollution controls →contribute to the evaluation of O3 pollution controls

Response: This has been done in the revised version. Please check line 127 in the revised version for details.

P4L117: … after it is itself validated as an NDACC site and it moves from candidate to regular status.

P5L129: then increases →then SZA increases

P5L129-133: what region influences the measurements depends on the azimuth of observation, yes, but also on the direction and wind speed pushing air masses above Hefei, especially for the lowest parts of the atmosphere. This could be significant when local pollution events are occurring as some events can be completely swept away from the FTS obs path.

Response: We agreed with your comment but we have removed these descriptions when condensing the paper. Detailed site/instrument descriptions can be found in our previous paper (Yuan et al.,2017; Wei et al., 2017).

P6L173: cited references missing from references section

Response: This problem has been addressed in the revised version.

P6L178: please explain deweighting more clearly. What are instrument SNR levels without deweighting?

Response: This sentence has been removed when condensing this paper. The standard deweighting stuff can be found in NDACC network or on request from the co-author Mathias Palm who is one of the SFIT4 developer. The instrument SNR level without deweighting is around 200 to 600. Please check section 3 for details.

P7L187: how are the Sa diagonal element magnitudes chosen? WACCM?

Response: In the revised version, Sa diagonal element is based on standard deviation of WACCM simulations from 1980 to 2020.

P7L191: is the ILS retrieved in all retrievals or is it done with LINEFIT and then held constant?

Response: We normally perform cell measurement once per month. For all measurements within this month, it is done with LINEFIT and then held constant. We

included this clarification in the revised version. Please check section 3 for details.

P11L315-317: tagged O3 runs are mentioned, which would be nice and would allow the attribution of pollution to various source regions, but these 3 lines are very unclear (i.e., also about restart files)

P14L393-4: basically reproduced … but with slight shifts in timing July is wrong in both models; why are they low in August? When is the local Hefei smog season?

Response: In the revised version, these problems don't exist because we removed all comparisons with correlative data. By the way, in my impression, most smog occurs in winter season. Please check section 4 for details.

P14L403: Logan (1985) "observed" →I presume this is a model study?

Response: This reference which based on both model simulation and observation has been replaced by some newly references in the revised version.

P16L448: basically consistent throughout all seasons →it really doesn't look like that to me; would be easier to think about if time series started with MAM as opposed to JFM.

P16L457-63: this really is a shallow explanation of what may be causing the differences, from which we learn nothing concrete. Also, how does larger air pollution increase uncertainty in either emission inventories or the photochemical regime? Isn't the latter, especially, something that is diagnosed from the emission rates and relative abundances of NOx and VOCs?

Response: After an extensive discussion among the authors, we deleted all paragraphs and figures regarding comparisons with the correlative data, i.e., OMI, GEOS-Chem and WRF-Chem data. Now all these problems don't exist in the revised version. Please check above clarification (page 3) for the reason.

Related change: Please check the revised version for details.

P17L495/6: which emissions are being discussed here: biogenic? anthropogenic? What are the expected magnitudes and timing of each?

Response: In the revised version, this problem doesn't exist because this sentence has been changed to "Pronounced tropospheric CO and NO_2 variations were observed but the seasonal cycles are not evident probably because of air pollution which is not

constant over season or season dependent". Please check section 5.2 for details.

P18L528: straightly applied →straight forwardly applied

Response: This has been done in the revised version.

P19L554: "validate OMI" →that's a strong statement given the unproven nature of these particular FTIR measurements, and given there's no reference to other OMI validation efforts and what they have typically revealed.

P19L560: WRF-Chem agreement is "better" →it has a lower "bias" but greater summer differences. It's not clear if that is better given it is a high res model using optimized emissions for China

Response: After an extensive discussion among the authors, we deleted all paragraphs and figures regarding comparisons with the correlative data, i.e., OMI, GEOS-Chem and WRF-Chem data. Now all these problems don't exist in the revised version. Please check above clarification (page 3) for the reason.

Related change: Please check the revised version for details.

P22 L651: would not screening out hazy days eliminate a lot of JJA O_3 pollution days? Haze isn't a problem for FTIR as much as non-constant intensity (e.g., clouds floating by during a ~20 minute observation time).

Response: This criterion that is used to eliminate bad spectra requires that the solar intensity variation (SIV) is less than 10%. Empirically, most of the variations are caused by floating clouds and some of them may be caused by other objects such as smog or unknown opaque object. The 10% empirical threshold keeps a reliable retrieval. We don't think it eliminated a lot of JJA O_3 pollution days. Haze is not a key factor that cause the variation and we have removed the word "haze" in the revised version. Please check supplement for details.