Response to referee comments on "Spatial distribution and temporal trend of ozone pollution in China observed with the OMI satellite instrument, 2005–2017"

We thank the referees for their careful reading of the manuscript and the valuable comments. This document is organized as follows: the Referee's comments are in *italic*, our responses are in plain text, and all the revisions in the manuscript are shown in blue. The line numbers in this document refer to the updated manuscript.

Referee #1

This paper seeks to quantify surface ozone across China using the SAO OMI tropospheric column ozone product. While I appreciate this effort to quantify such a relationship, the current analysis has not demonstrated a clear and convincing link between the lower/mid-tropospheric OMI retrievals and day-to-day ozone variability at the surface. (1) I realize that the authors are trying to find some signal in OMI that reflects ozone at the surface, but the degrees of freedom are so small, and the sensitivity to surface ozone is so weak, that there's no real way to distinguish between the signal that comes from the surface and that which comes from 800 or 700 hPa. (2) As presented, the relationship is more likely due to weather pattern variability causing ozone at the surface and in the free troposphere to vary in tandem. (3) Far more work is required, including a thorough evaluation of the OMI product against extensive IAGOS aircraft observations across mainland China, South Korea, Taiwan and Hong Kong. The additional analysis required to convince me that OMI can provide a meaningful evaluation of surface ozone across China goes beyond a standard major revision. My recommendation to the editor is that the paper be rejected to allow the authors adequate time to conduct additional product evaluation. If the expanded analysis can indeed demonstrate sensitivity of OMI to surface ozone then the authors will have the basis for a new manuscript which will make a valuable contribution to ozone monitoring across East Asia.

Response. Thanks for the careful reading of our manuscript and raising so many good points. The feedback has significantly improved our work. We also wish to draw the reviewer's attention to that we have revised the title, discussed the limitations of this work and also validated the OMI inferred ozone trends with surface observations (Figure 6).

New title. Ability of the OMI satellite instrument to observe surface ozone pollution in China: application to 2005-2017 ozone trends.

Since this is a long comment, we decompose it into three parts and answer each part one by one.

(1) Yes, the reviewer made a very good point here that the OMI cannot distinguish the signal that comes from the surface and from the lower troposphere (e.g. 800 and 700 hPa), given the relatively low vertical resolution of the retrievals. We have made these changes in the text to reduce the confusion.

P5 L18. The correlation of OMI with the MEE surface ozone data likely does not reflect a direct sensitivity of OMI to surface ozone, which is very weak, but rather a sensitivity to boundary layer ozone extending up to a certain depth and correlated with surface ozone.

The DOFS depends on what error is assumed in the prior estimate. Since the prior from McPeters et al. (2007) has low boundary layer ozone with low associated error then the DOFS would underestimate the ability to observe the polluted boundary layer. Now we say this.

P3L19. Even though a DOFS of 0.3 is still low, it is based on the prior estimate of low boundary layer ozone in the McPeters et al. (2007) zonal mean climatology.

Also the ozone sonde observations show that the ozone variability in the boundary layer is 80-100% higher than in the free troposphere. In more polluted regions like in mega city clusters, this difference should be even larger. These results indicate that the boundary layer ozone is more likely to drive the daily variability of OMI 850-400 hPa retrievals in regions like South China. We have added the following figure.

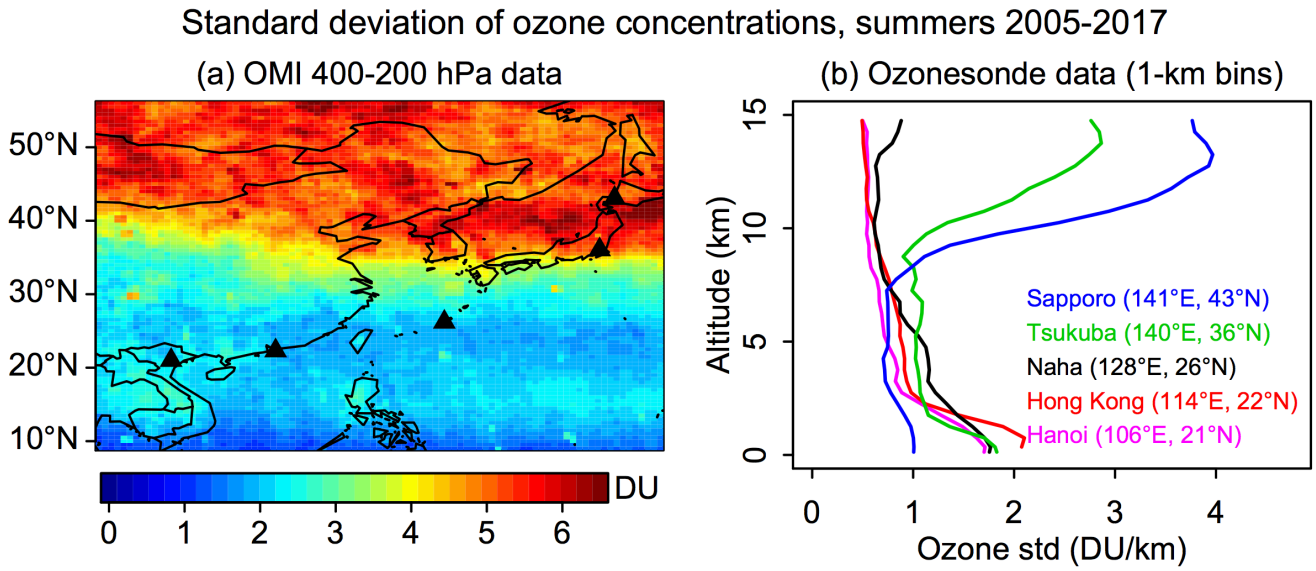

Figure 4. (a) Standard deviation of daily OMI 400-200 hPa ozone in East Asia during 2005-2017 summers. The triangles are the locations of ozonesonde sites with observations during this period. (b) Vertical profiles of daily ozone standard deviation in 1-km bins (DU/km) in the ozonesonde data for the 2005-2017 summers.

P6 L26. We find that the low correlation of OMI with boundary layer ozone in the northern ozonesonde data is due not only to the low DOFS but also to a large variability of ozone in the upper troposphere. Figure 4 (left panel) shows the standard deviation of daily OMI 400-200 hPa ozone during 2005-2017 summers, indicating that upper tropospheric ozone has much higher variability in the north (> 34°N) than in the south. This is related to the location of the jet stream and more active stratospheric influence (Hayashida et al., 2015). Figure 4 (right panel) displays the vertical profiles of ozone standard deviations for the five ozonesonde sites. For the two sites north of 34°N, the ozone variability becomes very large above 8 km. Since the OMI 850-400 hPa retrieval also contains information from above 400 hPa, this upper tropospheric variability causes a large amount of noise that masks the signal from boundary layer variability. For the three sites south of 34°N, the ozone variability in the boundary layer is much higher than in the free troposphere and the upper tropospheric ozone variability still remains low even above 8 km. In the rest of this paper we focus our attention on ozone episodes and the long-term trends in southern China (south of 34°N).

(2) We thank the reviewer for pointing out this issue. But based on our analysis, the weather patterns are insufficient to explain the observed relationship. We have conducted a sensitivity experiment and made these changes in the text.

P 6 L15. The correlation between boundary layer ozone pollution and the OMI ozone retrievals could be due in part to correlation between boundary layer and mid-tropospheric ozone, considering that both tend to be driven by the same weather systems. We used the ozonesonde data to examine what correlation with boundary layer (950-850 hPa) ozone would be observed if OMI were sensitive only to the free troposphere at ~500 hPa (where its sensitivity is maximum, Figure 3c) and not to the boundary layer. In that case the correlation coefficient $R_{1,3}$ of boundary layer ozone and the OMI 850-400 hPa retrievals would be given by (Vos, 2009):

$$R_{1,3} = R_{1,2}R_{2,3} \pm \sqrt{(1-R_{1,2}^2)(1-R_{2,3}^2)} \quad (2)$$

where $R_{1,2}$ is the correlation coefficient between boundary layer and 500 hPa ozone in the ozonesonde data, and $R_{2,3}$ is that between 500 hPa ozone and the OMI 850-400 hPa retrievals. As seen from Figure S4, $R_{1,3}$ at the five sonde sites is only ~0.2, implying that direct sensitivity to the boundary layer dominates the correlation of OMI with surface ozone at least in southern China. Further evidence for this is the ability of OMI to detect the ozone enhancements in megacity clusters (Figure 1).

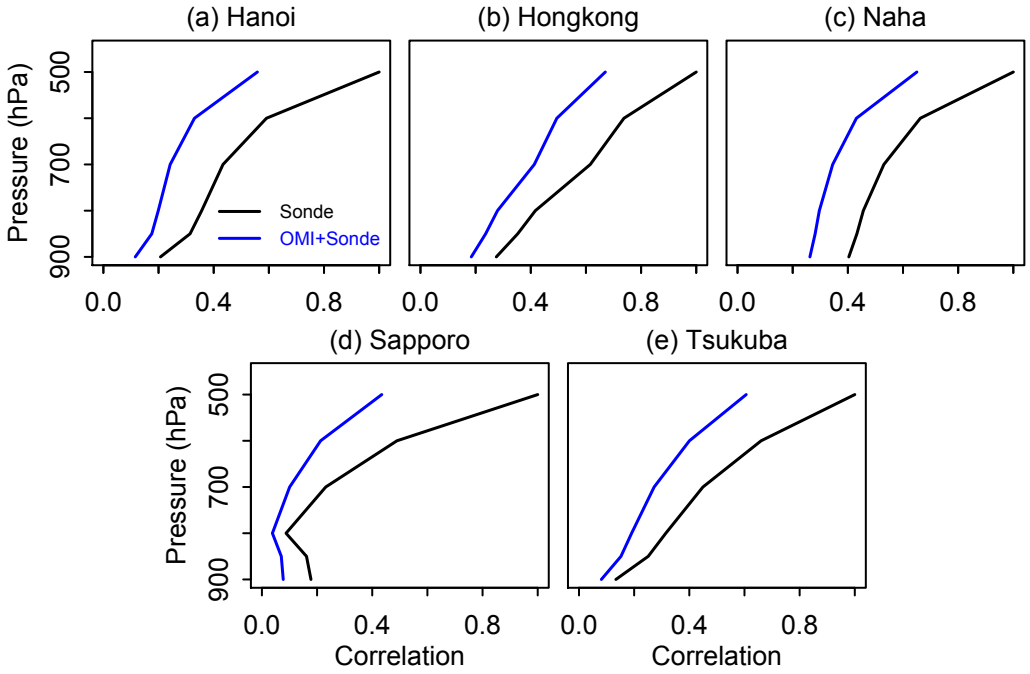

Figure S4. Correlation of OMI 850-400 hPa ozone and the boundary layer ozone assuming that the OMI were sensitive only to the free troposphere at ~500 hPa (where its sensitivity is maximum, Figure 3c) and not to the boundary layer. The black line is the correlation of ozone at different pressure levels with 500 hPa ozone in the sonde observations. The blue line is the estimated correlation of OMI 850-400 hPa with ozone at different layers if the satellite can only detect the signal at 500 hPa but not from other layers, as calculated using Equation 2. See text for more details.

(3) We have processed all IAGOS aircraft observations in East Asia during 2005-2017 summers and we only find 54 profiles in 8 airports that can be used to validate the OMI. Given so few profiles, it is unlikely to evaluate the long-term correlation of surface and mid-tropospheric ozone at each airport. But combining all profiles together, the temporal correlation coefficients of the 950 hPa ozone and 850-400 hPa OMI ozone is 0.59. These results are also consistent with what we find in the Hong Kong sonde site (Figure 2). We have made these changes.

P6 L9. We applied the same daily correlation analysis to the other ozonesonde datasets and IAGOS aircraft measurements during 2005-2017 summers. For the 54 IAGOS vertical profiles coincident with OMI observations, the correlation coefficient of the 950 hPa in situ ozone and 850-400 hPa OMI ozone is $R = 0.59$ ($p<0.05$) (Figure S2).

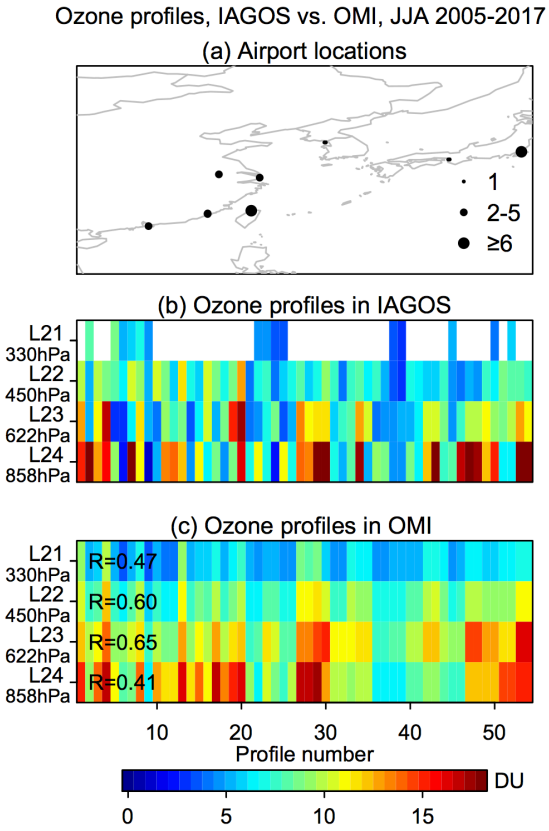

Figure S2. (a) Location and the number of tropospheric profiles at each airport in IAGOS that are coincident with OMI retrievals. We only select these profiles between 12-15 local time. (b) Ozone profiles from IAGOS, but mapped to OMI layers. The missing data at L21 is in part because we only select pixels that are within 200 km from the airport on the flight path. (c) OMI ozone profiles coincident with the ozonesondes. The correlations of unsmoothed 950 hPa ozone data in IAGOS with the OMI retrievals for different levels are shown inset. The correlation with 850-400 hPa OMI ozone is 0.59 ($p<0.05$)

Further comments:

When I read the title and abstract I was under the impression that the authors had made a new

breakthrough regarding the detection of surface ozone using OMI. It seemed like the instrument could actually detect ozone at the surface and the detection was so good that daily ozone variability at any given surface site could be determined with a precision of +/- 10.7 ppb. But when I read the full paper I learned that this is not the case.

The premise that lower/mid-tropospheric OMI ozone retrievals are closely associated with surface ozone is not shown in a convincing manner. The initial correlation of: [O3]= 8.9 ΔΩ + 15.8 ± 10.7 appears to be driven entirely by the latitudinal gradient of ozone at the surface and in the mid-troposphere. Just because surface and OMI ozone have similar latitudinal gradients, when averaged over several years, does not mean that the mid- to lower troposphere can tell us how ozone varies at the surface from day to day. A better test of the relationship is to focus on a narrow latitude range. This is done in Figure 3, for daily observations, where we can see that the correlation is very low above five urban regions. For the region of Beijing the correlation is only R=0.27 which corresponds to an r-squared value of 0.07, which means that the variation in OMI only explains 7% of the ozone variability at the surface. The best case is made by the Wuhan region, but even here R=0.53, which means OMI only explains 28% of the surface ozone variability. Figure 3 shows that OMI is only weakly correlated with surface ozone and provides no convincing argument that the retrieval is sensitive to surface ozone. The weak correlation is probably just due to weather patterns causing surface and lower to mid- tropospheric ozone to vary in tandem.

Response. Yes, the reviewer has made a very good point here that the satellite has a lot of noise. This is a common problem when we use satellite data. But noisy data still contains information, and we can reduce the noise by either temporally averaging the data over multiple years, or training the model with a lot of data together. In our study, we tried both ways. When we average the data over a five-year period, the OMI 850-400 hPa ozone displays a strong correlation (R=0.73) with the surface observations. When we fit all the daily data in eastern China together, we find the extreme value model can well simulate the distribution of high ozone concentrations. We also find the resulting model can accurately estimate the probability of higher thresholds, which is a strong signal that our extreme value model is well fitted.

The reviewer is correct that our old manuscript indeed overpromised the value of OMI data, especially in the northern China. Due to the stronger jet wind activities and more active stratosphere-troposphere exchange in the north, the upper tropospheric ozone there has much higher daily variability, making OMI 850-400 hPa ozone less reliable in predicting the daily episode and inferring the long-term trends. The strong jet wind also means these regions are very sensitive to global background. At the same time, OMI has relatively low sensitivity in the north. So in our revised manuscript, we don't fit the extreme value model or predict the trends in the north. We have added a new figure 4 and more discussions at P6L26-P7L6. Based on reviewer's suggestion, we have changed the title (see new title) and also discussed the limitations of this work (see many blue highlighted text in the revised manuscript).

But the OMI data should still be useful in south China for these reasons. First, OMI has higher sensitivity in the South, and the daily correlation of surface ozone and OMI 850-400 retrievals is

statistically significant for most sites (Figure 1). Second, the upper tropospheric ozone variability is much smaller and the boundary ozone variability can be a strong modulator of OMI 850-400 hPa ozone variability (new figure 4). Third, the inferred long-term trends of surface ozone from OMI fairly agree with these from TOAR sites (Figure 6).

We also make it clear we have removed the latitude-dependent background.

P3 L23. To remove this gradient and also any long-term uniform drift in the data, we subtract the monthly mean Pacific background (150°E-150°W) for the corresponding latitude and month

P4 L22. After subtracting the North Pacific background for the corresponding latitude in month, we obtain the OMI ozone enhancements shown in Figure 1d.

P4 L23. The spatial correlation coefficient between the OMI ozone enhancements and the MEE surface network is $R = 0.73$ over eastern China. The correlation is driven in part by the latitudinal gradient but also by the enhancements in the large megacity clusters identified as rectangles in Figure 1b. Thus the correlation coefficient is $R = 0.55$ for the 26-34°N latitude band including YRD, SCB, and Wuhan.

In a related comment, do the authors think that any correlation between ozone at the surface and ozone in the lower/mid troposphere is linked because of similar photo- chemical processes, or is the correlation just a coincidence due to meteorology? For example we know that in southern China the ozone at the surface varies strongly with the strength of the summertime Asian monsoon. When transport is from the south then the relatively clean air masses from the tropical Pacific bring air that is low in ozone, both at the surface and in the lower-mid troposphere. But when the monsoon winds weaken, mid-latitude air is allowed to move back into the region of southern China, bringing higher ozone to the lower and mid-troposphere. At the same time, the flow of clean air from the south also ceases at the surface, allowing ozone to build up in the polluted air masses from mainland China. Under this scenario ozone in the mid- troposphere is correlated with ozone at the surface even if the two layers are isolated from each other by strong temperature inversions.

Response. Thanks for pointing out this issue. But meteorology is insufficient to explain the observed correlation of surface ozone and OMI 850-400 hPa ozone. Please check L75-100 of this response letter for more details (or P5 L15-25 in the main text).

OMI ozone could be compared to long-term ozone monitoring sites in rural areas which would be a better comparison than the urban data from the new Chinese monitoring network. It would be very helpful to see time series of daily OMI values (when avail- able) and corresponding surface observations from the following sites: Mt Tai – data can be obtained from Prof. Likun Xue, Shandong University [Sun et al., 2016] Hok Tsui – located on the south coast of Hong Kong, data can be obtained from Prof. Tao Wang, Hong Kong Polytechnic [Wang et al., 2017] Shangdianzi – see Ma et al., 2016 LongFengShan – located in northeastern China. Contact Dr. Xiaobin Xu at the China Meteorological Administration: xiaobin_xu@189.cn LinAn – Near Shanghai, Contact Dr. Xiaobin Xu at the China Meteorological Administration: xiaobin_xu@189.cn Xi- angGeLiLa – in south central China, Contact Dr. Xiaobin Xu at the China Meteorologi- cal Administration: xiaobin_xu@189.cn

Response. We have obtained the Hong Kong site observations from Prof. Tao Wang from the above list. And we have also used the TOAR dataset to validate our model. We find that the OMI inferred ozone trends are fairly consistent with these long term surface records.

P1 L21. Comparison of 2005-2009 and 2013-2017 OMI data indicates that mean summer afternoon surface ozone in southern China (including urban and rural regions) has increased by 3.5 ppb over the 8-year period and the number of episode days per summer has increased by 2.2 (as diagnosed by an extreme value model), fairly consistent with the few long-term surface records.

P9 L6. We compared the OMI trends in Figure 6 to the trends of MDA8 ozone and number of high-ozone days reported by the long-term TOAR sites (Schultz et al., 2018) and our own analysis for the Hok Tsui station in Hong Kong (Wang et al., 2009). For Lin'an, Hong Kong, and the 5 sites in Taiwan, the changes of mean ozone concentrations from 2005-2009 to 2013-2017 are 1.1, 2.3, and -0.18±2.2 ppbv (standard deviation among the 5 sites) as estimated from OMI, compared to 0.7, 5.6 (or 5.8 in Hok Tsui station), and -0.75±3.4 ppbv for MDA8 ozone at the TOAR sites. The changes in the number of ozone episodes per summer are 1.2, 1.9, and -0.17±0.74 days in OMI, compared to 2.1, 1.8 (or 2.1 in Hok Tsui station), and -3.5±3.9 days at the TOAR sites. These OMI inferred trends are fairly consistent with the long-term records available from surface sites.

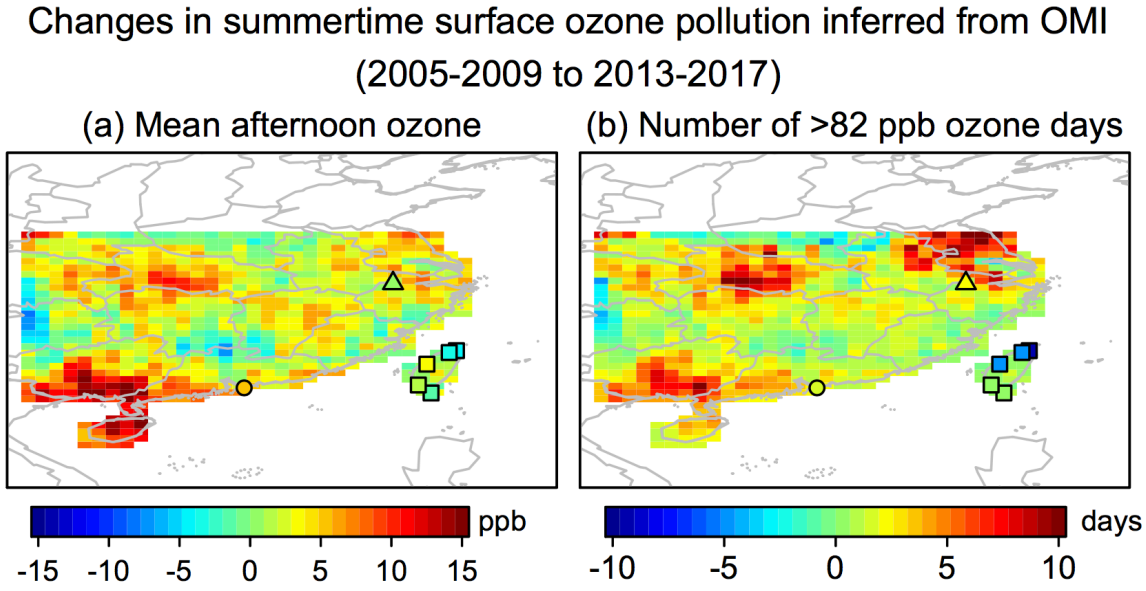

Figure 6. Changes in surface ozone pollution in China between 2005-2009 and 2013-2017 as inferred from OMI afternoon observations at around 13:30 local time. (a) Change in mean summer afternoon concentrations, obtained from the difference in the mean OMI enhancements at 850-400 hPa and applying equation (1). Also shown with symbols are observed changes in mean MDA8 ozone from in situ observations in Lin'an, Hong Kong, and Taiwan reported by TOAR (Schultz et al., 2018). Because the TOAR observations are only reported for 2005-2014, we estimate the changes from 2005-2009 to 2013-2017 on the basis of the reported linear trends during 2005-2014 (ppb a^{-1}). The change of 12-15 LT ozone at the Hok Tsui station in Hong Kong is 5.8 ppb. (b) Change in the number of high-ozone days (> 82 ppb) per summer, calculated by applying the probability of exceeding 82 ppb (equation 8) to the daily OMI enhancements. Also shown with symbols are observed changes of the number of days with MDA8 ozone exceeding 80 ppb at the TOAR sites, similarly adjusted as the change from 2005-2009 to 2013-2017. The change in the

number of days with 12-15 LT ozone exceeding 82 ppbv at the Hok Tsui station in Hong Kong is 2.1 days.

Ozone at the surface and in the mid-troposphere varies greatly with transport path- way and abrupt changes in air masses, and recent studies have shown that ozone in China varies with meteorology [Pu et al., 2017; Zhao et al., 2018]. The authors are aware of this phenomenon as their previous work has explored the impact of climate variability on ozone. Therefore I'm surprised that the authors didn't first explore how surface ozone across China varies with meteorology, such as surface temperatures (or temperature at 850 hpa) [Pusede et al., 2015], or with the height of the 500 hPa surface [Reddy et al., 2016], both of which correlate quite well with surface ozone. The authors should first determine the correlation between surface ozone and meteorology, and then compare these results to what they find from OMI ozone. Does OMI give more information on surface ozone than basic meteorology? Given that reanalysis data are available for all of China under all weather conditions (no cloud screening) I would think that the meteorology would perform better than OMI. If OMI performs less well than meteorology, is there any reason to use OMI to try to predict surface ozone, when meteorological analyses are available everywhere and at all times?

Response. Our focus is to evaluate the OMI observation capability, not to analyze the correlation of ozone with meteorological variables which has been done before and would not capture the variability in ozone driven by emissions. The reviewer makes a good point that mid-tropospheric and surface ozone may respond similarly to meteorological conditions, which may in turn contribute to the correlation of OMI with surface ozone. We now address this point in Section 4 by analysis of the ozonesonde data (P6 L15-25 with new Figure S4). We also mention this in other parts of the manuscript.

P5L18. The correlation of OMI with the MEE surface ozone data likely does not reflect a direct sensitivity of OMI to surface ozone, which is very weak, but rather a sensitivity to boundary layer ozone extending up to a certain depth and correlated with surface ozone.

P9L25. To better understand the correlation of OMI with surface ozone we examined vertical ozone profiles from Hong Kong and other ozonesondes, and from the IAGOS commercial aircraft program. Some of the correlation is driven by similar meteorology influencing ozone in the mid-troposphere (where OMI sensitivity is maximum) and the boundary layer, but most of the correlation is driven by direct sensitivity to the boundary layer.

Another necessary analysis is to see if in situ observations of ozone in the mid- troposphere are correlated with surface ozone. I realize that the authors did look at ozonesonde profiles above Hong Kong, but they are not very frequent and they don't tell us anything about ozone in other parts of China, especially in the highly polluted North China Plain. The IAGOS program has hundreds of profiles above East Asia since 1995. As shown by Ding et al. [2005] and by Gaudel et al. [2018] ozone in sum- mertime in the boundary layer is much greater than ozone in the mid-troposphere. The difference is due to very strong ozone production in the boundary layer, versus distant source regions for ozone in the mid-troposphere. If the authors conducted a trans- port study for ozone in the mid-troposphere they would find that very little of the air in this layer comes from the surface of China. Probably 80-90% of the mid-tropospheric above China air has either been in the mid-troposphere

for days, or it comes from the boundary layer far upwind of China. The authors can freely access hundreds of com- mercial aircraft profiles of ozone and carbon monoxide above mainland China, Hong Kong, Taiwan and South Korea from the IAGOS database. They can then apply the OMI averaging kernel to the profiles and determine the relationship between IAGOS ozone in the mid- and lower troposphere to ozone at the surface. Does IAGOS ozone in the mid-troposphere correlate with ozone at the surface? Is the correlation any better than when surface ozone is correlated with meteorology? Then compare the IAGOS relationship to the OMI relationship. Does OMI perform any better than IAGOS?

Response. We only find 54 profiles that can be used to validate the OMI data. This is because the OMI crossing time is 13:30 local time and we have to use observations close to this time window. We have added more discussion in the text.

P6 L9. We applied the same daily correlation analysis to the other ozonesonde datasets and IAGOS aircraft measurements during 2005-2017 summers. For the 54 IAGOS vertical profiles coincident with OMI observations, the correlation coefficient of the 950 hPa in situ ozone and 850-400 hPa OMI ozone is $R = 0.59$ ($p<0.05$) (Figure S2).

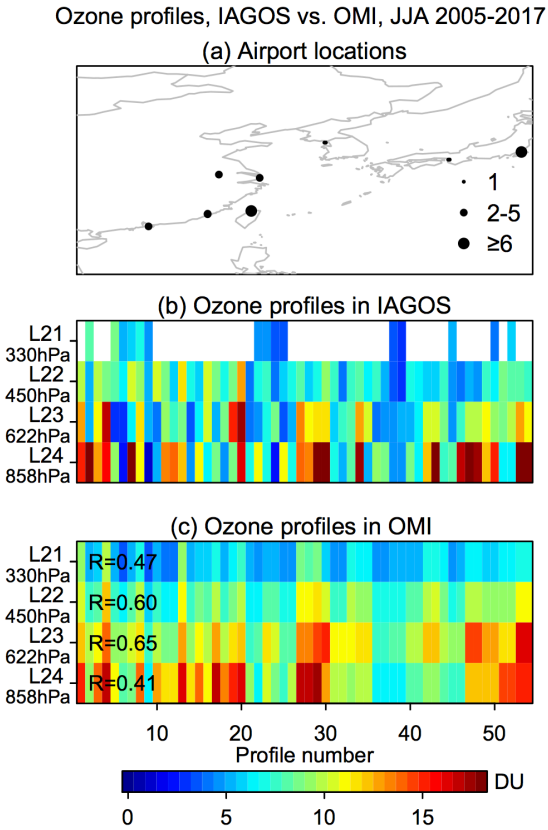

Figure S2. (a) Location and the number of tropospheric profiles at each airport in IAGOS that are coincident with OMI retrievals. We only select these profiles between 12-15 local time. (b) Ozone profiles from IAGOS, but mapped to OMI layers. The missing data at L21 is in part because we only select pixels that are within 200 km from the airport along the flight path. (c) OMI ozone profiles coincident with the ozonesondes. The correlations of unsmoothed 950 hPa ozone data in IAGOS with the OMI retrievals for different levels are shown inset. The correlation with 850-400 hPa OMI ozone is 0.59 ($p<0.05$)

Figure 5 shows surface ozone trends across China which were derived from the OMI ozone product. The strongest trends are in the far north of China and in the far south of China. Based on the summer OMI trends (2005-2015) reported by the Tropospheric Ozone Assessment Report in supplementary Figure S-24 of Gaudel et al. [2018], OMI has a strong trend across southern China but no trend across northern China. Therefore I don't understand how Figure 5 can show trends across northern China. It would be helpful to include a map that shows the OMI trends across China.

Response. Thanks for the careful reading. The authors who plotted Figure S24 in the TOAR report (Gaudel et al., 2018) are also coauthors of this work. The difference of trends in northern China arises from these reasons. First, we use the 850-400 hPa ozone but Gaudel et al. (2018) uses the tropospheric column ozone. Second, we use different methods to remove the background. Third, we have removed the low quality L2 data but Gaudel et al. (2018) kept all of them. Now we have this new figure in the supplement.

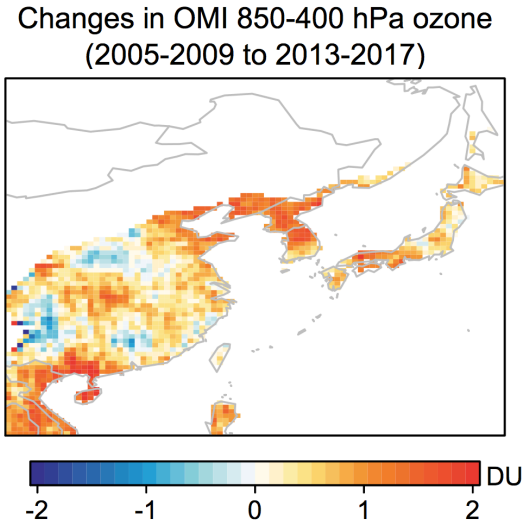

Figure S5. Difference of the mean OMI enhancements at 850-400 hPa from 2005-2009 to 2013-2017 after correcting the Pacific background. Data are only shown for regions with DOFS below 400 hPa (Figure 1a) greater than 0.30.

Referee #2

This is a nice study that explores the potential of OMI observations of tropospheric ozone to detect the ozone pollution over China. While it's unrealistic to use OMI data to capture the day-to-day variability of ozone pollution, the authors show extreme ozone pollution may be detectable by aggregating long-term observations using statistical methods. Overall I think this is an important study to the field, which opens up the possibility to use satellite observations to detect surface ozone pollution, but I think the authors overpromise the value of satellite data. I have several major concerns:

Response. Thanks for raising these good points. This feedback has significantly improved the manuscript. Now we have a new Figure 4 showing that OMI 850-400 retrievals have limited skill in predicting the daily ozone variability in the north and we only predict the trends of ozone pollution in southern China (south of 34°N). We have new in-situ observations to validate the trends inferred from the OMI, which are shown in Figure 6.

1. *My major concern is that the authors seem to overpromise the value of OMI data for characterizing the spatial and temporal trend of ground-level ozone. The title and the abstract leave me an impression that OMI satellite data can capture the spatial distribution and the long-term trend in ground-level ozone, but the results only suggest OMI may be able to detect high ozone pollution and capture the large-scale or latitudinal variations. I suggest the authors consider revising the title, otherwise it'd be misleading to readers. The authors need to be more careful with the wording. I think this work would actually be much more valuable if the authors can clarify the limitations of OMI data, which will also be useful for preparation of next-generation satellites.*

Response. Thanks for making such a good point. Now we revised the title and also discussed the limitations in many parts of the main text.

New title. Ability of the OMI satellite instrument to observe surface ozone pollution in China: application to 2005-2017 ozone trends

P1 L18. OMI is much more successful at capturing the day-to-day variability of surface ozone at sites in southern China <34°N (R = 0.3-0.6) than in northern China (R = 0.1-0.3) because of weaker retrieval sensitivity and larger upper tropospheric variability in the north.

P5 L7. This implies that OMI can only provide statistical rather than deterministic temporal information on ozone pollution episodes, and may be more useful in South than in North China. We return to this point in Section 4.

P5 L18. The correlation of OMI with the MEE surface ozone data likely does not reflect a direct sensitivity of OMI to surface ozone, which is very weak, but rather a sensitivity to boundary layer ozone extending up to a certain depth and correlated with surface ozone.

P6L27. We find that the low correlation of OMI with boundary layer ozone in the northern ozonesonde data is due not only to the low DOFS but also to a large variability of ozone in the upper troposphere. Figure 4 (left panel) shows the standard deviation of daily OMI 400-200 hPa ozone during 2005-2017 summers, indicating that upper tropospheric ozone has much higher variability in the north (> 34°N) than in the south. This is related to the location of the jet stream and more active stratospheric influence (Hayashida et al., 2015). Figure 4 (right panel) displays the vertical profiles of ozone standard deviations for the five ozonesonde sites. For the two sites north of 34°N, the ozone variability becomes very large above 8 km. Since the OMI 850-400 hPa

retrieval also contains information from above 400 hPa, this upper tropospheric variability causes a large amount of noise that masks the signal from boundary layer variability. For the three sites south of 34°N, the ozone variability in the boundary layer is much higher than in the free troposphere and the upper tropospheric ozone variability still remains low even above 8 km. In the rest of this paper we focus our attention on ozone episodes and the long-term trends in southern China (south of 34°N).

2. Is the point process model you used to predict ozone exceedance probability site specific? If so, how can you apply this method widely to areas without ground-based sites (as you promised in the conclusion)? The authors present the surface ozone pollution and exceedance probability only at ground-based sites, but why not show the distribution across China? For example, MEE network mainly consists of urban sites. Can you use OMI data to tell the spatial patterns of ozone pollution over rural/remote areas? If not, what's the added value of OMI data to existing ground-based network?

Response. Thanks. The point process model makes use of all the data. Now we show the trends of ozone for all rural and remote regions in south China.

P7 L14. We fit the model to all daily concurrent observations of surface ozone and OMI ozone enhancements for the ensemble of eastern China sites in Figure 1 (90,601 observations for summers 2013-2017).

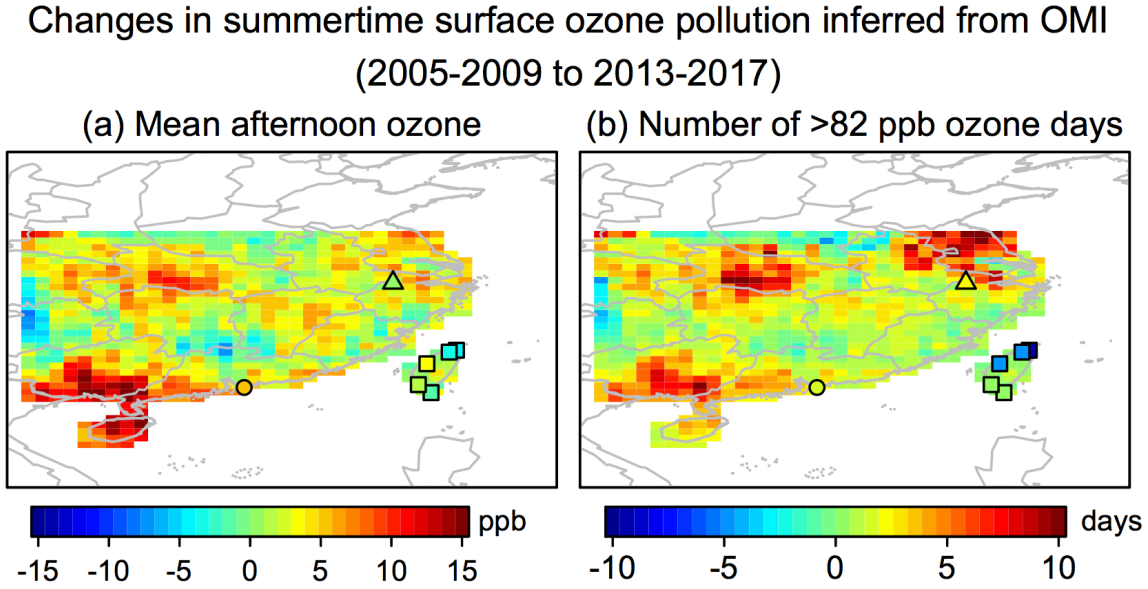

Figure 6. Changes in surface ozone pollution in China between 2005-2009 and 2013-2017 as inferred from OMI afternoon observations at around 13:30 local time. (a) Change in mean summer afternoon concentrations, obtained from the difference in the mean OMI enhancements at 850-400 hPa and applying equation (1). Also shown with symbols are observed changes in mean MDA8 ozone from in situ observations in Lin'an, Hong Kong, and Taiwan reported by TOAR (Schultz et al., 2018). Because the TOAR observations are only reported for 2005-2014, we estimate the changes from 2005-2009 to 2013-2017 on the basis of the reported linear trends during 2005-2014 (ppb a^{-1}). The change of 12-15 LT ozone at the Hok Tsui station in Hong Kong is 5.8 ppb. (b) Change in the number of high-ozone days (> 82 ppb) per summer, calculated by applying the probability of exceeding 82 ppb (equation 8) to the daily OMI enhancements. Also shown with symbols are observed changes of the number of days with MDA8 ozone exceeding 80 ppb at the TOAR sites, similarly adjusted as the change from 2005-2009 to 2013-2017. The change in the

number of days with 12-15 LT ozone exceeding 82 ppbv at the Hok Tsui station in Hong Kong is 2.1 days.

3. Figure 5: While OMI data may be able to detect the sign of the change in ground- level ozone, the magnitude of the change is less convincing to me. The authors suggest a 0.67 ppb /year increase in mean ozone over China, which seems to be lower than previous studies. The point process model is trained with ground-based observations in 2013-2017, but it's unknown how the model performs for early years 2005 - 2009. I'd suggest the authors use available long-term ground-based ozone observations to verify the long-term change. I understand long-term ground-based observations are not generally available over China, but since the OMI data are global, it's possible to extend the analysis to wider regions (e.g. Hong Kong, Japan) where long-term sites are available for evaluation.

Response. Thanks. We have new in-situ observations from TOAR and also from a Hong Kong site to validate the trends inferred from the OMI, which are shown in Figure 6. We find the OMI inferred trends are fairly consistent with the long-term records available from surface sites. We also add discussion in the main text.

P4 L11. For evaluating the long-term surface ozone trends inferred from OMI, we use 2005-2014 trend statistics for maximum daily 8-hour average (MDA8) ozone from the Tropospheric Ozone Assessment Report (TOAR) (Schultz et al., 2018). We also have 2005-2017 JJA 12-15 LT mean ozone at the Hok Tsui station in Hong Kong (Wang et al., 2009).

P9 L5. We compared the OMI trends in Figure 6 to the trends of MDA8 ozone and number of high-ozone days reported by the long-term TOAR sites (Schultz et al., 2018) and our own analysis for the Hok Tsui station in Hong Kong (Wang et al., 2009). For Lin'an, Hong Kong, and the 5 sites in Taiwan, the changes of mean ozone concentrations from 2005-2009 to 2013-2017 are 1.1, 2.3, and -0.18±2.2 ppbv (standard deviation among the 5 sites) as estimated from OMI, compared to 0.7, 5.6 (or 5.8 in Hok Tsui station), and -0.75±3.4 ppbv for MDA8 ozone at the TOAR sites. The changes in the number of ozone episodes per summer are 1.2, 1.9, and -0.17±0.74 days in OMI, compared to 2.1, 1.8 (or 2.1 in Hok Tsui station), and -3.5±3.9 days at the TOAR sites. These OMI inferred trends are fairly consistent with the long-term records available from surface sites.

Referee 3

This paper explored the capability of OMI ozone columns to represent the surface O3. I feel the satellite data is over-interpreted based on the evidence provided in the paper. However, I do believe it will be big news if substantial improvements are made to prove that the conclusion is solid.

Response. We thank the reviewer for raising so many good points, which have significantly improved our work. Now we have a new Figure 4 showing that OMI 850-400 retrievals have limited skill in predicting the daily ozone variability in the north and we only predict the trends of ozone pollution in southern China (south of 34°N). We have new in-situ observations to validate the trends inferred from the OMI, which are shown in Figure 6. And we have revised the title.

New title. Ability of the OMI satellite instrument to observe surface ozone pollution in China: application to 2005-2017 ozone trends

General comments:

1. The sensitivity of OMI O3 to the lower troposphere is very low. I feel that is the reason why no quantitative comparison to surface observations has so far been done. I'm wondering is there any improvements that have been made to make the quantitative comparison robust? Why does not the quantitative comparison work for other regions, but work for China?

Response. We now explain this better in the Introduction.

P2L23. However, no quantitative comparison of the satellite data to surface observations has so far been done. Surface ozone network data are available in the US and Europe but levels are generally too low to enable statistically meaningful validation. Ozone levels in China are much higher (Lu et al., 2018). The high density of the MEE network, combined with vertical profile information from ozonesondes and aircraft, provides a unique opportunity for evaluating quantitatively the ability of OMI to observe ozone pollution.

2. *The robustness of the residual. How large is the temporal and spatial variations of the background? Is it likely that such variations bring significant uncertainties to the subtraction?*

Response. Thanks for making this good point. We have tested different approaches to correct the background and the results are consistent with what have presented in the paper. In the text, we make it simple by saying this.

P4L1. We examined different spatial and temporal averaging domains for the North Pacific background and found little effect on the residual.

The uncertainty related to the background correction can be found in these two figures.

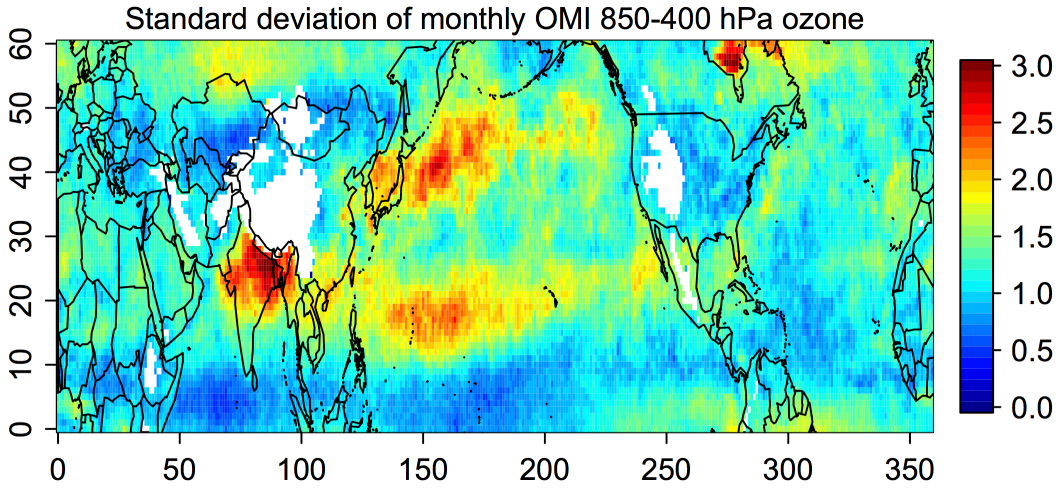

Figure SX. Standard deviation of monthly OMI 850-400 hPa ozone during 2005-2017 summers.

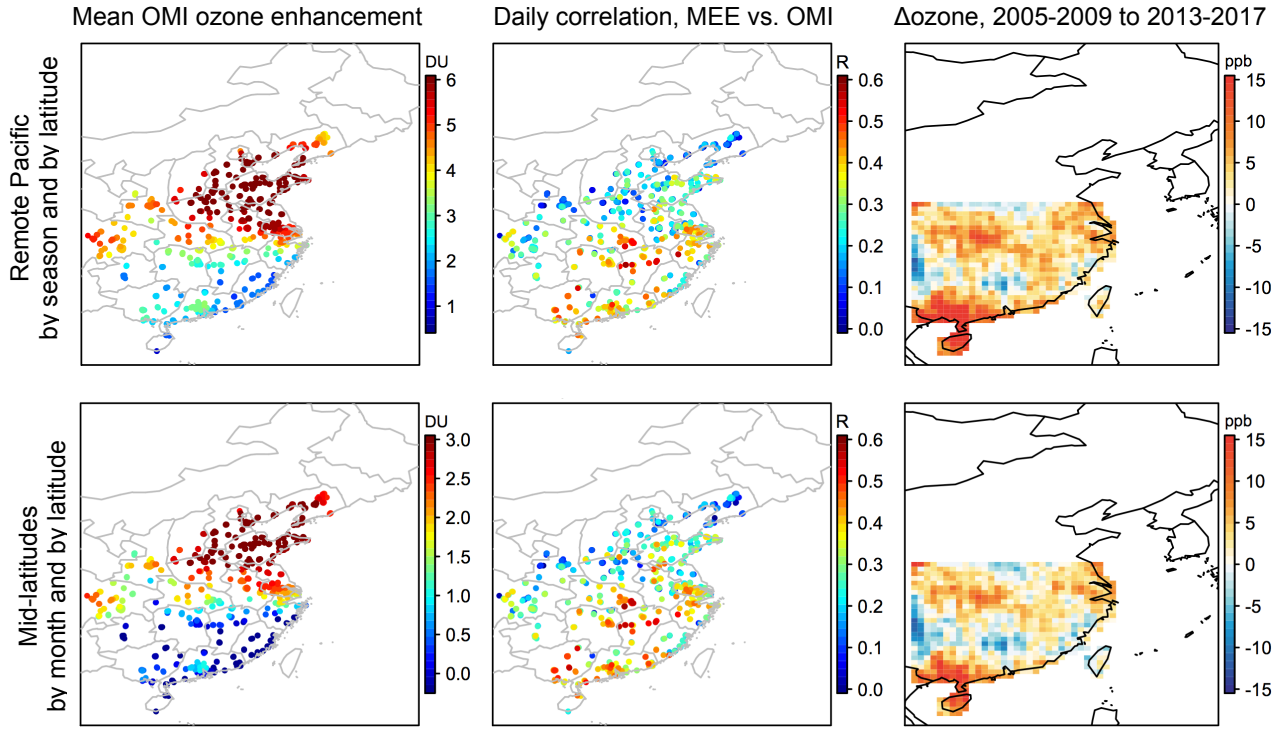

Figure SX. The mean ozone enhancement (left panel), daily correlation of OMI and MEE ozone (mid panel), and the OMI inferred changes of mean ozone concentrations from 2005-2009 to 2013-2017 using different approaches correcting the OMI drift. In the top panel, we subtract the monthly mean Pacific background (150°E-150°W) for the corresponding latitude and season. In the bottom panel, we subtract the monthly mean mid-latitude ozone for the corresponding latitude and month.

3. *The correlation between MEE and OMI. The correlation seems to be related with the dependence of O3 on latitude. I suggest additional analysis here to prove that is not the case.*

Response. We are not sure if the reviewer is referring to Figure 1f here. In Figure 1f, the correlation

is higher in the south and lower in the north. This is because in the northern China, OMI 850-400 hPa ozone has lower sensitivity in the boundary layer, more likely to be influenced by the upper tropospheric ozone variability and stratosphere-troposphere exchange. We have added new discussion in the text.

P6 L27. We find that the low correlation of OMI with boundary layer ozone in the northern ozonesonde data is due not only to the low DOFS but also to a large variability of ozone in the upper troposphere. Figure 4 (left panel) shows the standard deviation of daily OMI 400-200 hPa ozone during 2005-2017 summers, indicating that upper tropospheric ozone has much higher variability in the north (> 34°N) than in the south. This is related to the location of the jet stream and more active stratospheric influence (Hayashida et al., 2015). Figure 4 (right panel) displays the vertical profiles of ozone standard deviations for the five ozonesonde sites. For the two sites north of 34°N, the ozone variability becomes very large above 8 km. Since the OMI 850-400 hPa retrieval also contains information from above 400 hPa, this upper tropospheric variability causes a large amount of noise that masks the signal from boundary layer variability. For the three sites south of 34°N, the ozone variability in the boundary layer is much higher than in the free troposphere and the upper tropospheric ozone variability still remains low even above 8 km. In the rest of this paper we focus our attention on ozone episodes and the long-term trends in southern China (south of 34°N).

If the reviewer refers to Figure 1d, we now make it clear that we have corrected the background that is dependent on latitudes.

P3 L23. To remove this gradient and also any long-term uniform drift in the data, we subtract the monthly mean Pacific background (150°E-150°W) for the corresponding latitude and month

P4 L22. After subtracting the North Pacific background for the corresponding latitude in month, we obtain the OMI ozone enhancements shown in Figure 1d.

P4 L23. The spatial correlation coefficient between the OMI ozone enhancements and the MEE surface network is $R = 0.73$ over eastern China. The correlation is driven in part by the latitudinal gradient but also by the enhancements in the large megacity clusters identified as rectangles in Figure 1b. Thus the correlation coefficient is $R = 0.55$ for the 26-34°N latitude band including YRD, SCB, and Wuhan.

Specific comments:

1. "We exclude outliers with over 35 Dobson Units (DU) at 850- 400 hPa (>99th percentile in eastern China) and exclude July 2011 when the retrievals are anomalously high." Please give the reference to the exclusion. Otherwise, please quantify the influence of the exclusion.

Response. Thanks for pointing this out. We delete this because we don't use the July 2011 data and not excluding the extremely high data has little effect on our result.

2. "We see that high-ozone episodes in the 950-850 hPa sonde data are systematically associated with high OMI values, though the converse does not always hold." Additional explanation for the reason is expected.

Response. Thanks. Now we say

P6L4. We see that high-ozone episodes in the 950-850 hPa sonde data are systematically associated with high OMI values, though the converse does not always hold because free tropospheric enhancements affecting OMI

526 can also occur.
527
528
529