Author responses to Anonymous Referee # 2 RC#1

Referee comments in **boldface**, author comments in normal typeface, locations where revisions are made are prefaced with an underlined heading. All edits are prefaced with their PXXLYY location relative to the original discussions manuscript.

Thank you for the detailed and excellent comments and suggestions which have significantly strengthened the paper. We address them as responses to (1) your general comments and (2) the specific comments, below. We direct attention to specific places in the text or supplement where clarifications exist or are being included in future versions of this manuscript. Overall, we find the reviews do not quite recognize the difficulty in obtaining observational data from areas where access is limited for various reasons. We would like to point out that despite lacking a dense observational data set, scientific studies must continue to occur (with the appropriate caveats, as we think we provided in this study) to guide future efforts to fill these obvious data gaps. We feel there is significant value to studies like this, *particularly* in light of the fact that there is such a dearth of observationally-constrained analyses.

General Comments.

1. **The authors state that evaluating all existing CO2 inventories is outside of the scope of this paper, so they compare ZHAO to two global inventories – CDIAC and EDGAR, which primarily use population as a proxy data to distribute emissions. There is another key proxy that is ignored in this work, satellite observations of nighttime lights. ODIAC (Oda et al., 2108) and FFDAS (Asefi-Najafabady et al., 2014) primarily use this proxy (along with some other data). A strength of ODIAC is that it first uses a point source inventory (CARMA, which is no longer available) then distributes the remaining emissions according to the night lights. One can expect that population would not be a good proxy for a large country with regional variations in wealth, industry and climate. This has already been demonstrated in a comparison of CDIAC and ODIAC over Canada in Nassar et al. (2013) (which the authors have cited in another context), where CDIAC did not accurately represent the provincial distribution, yet ODIAC was much closer. By ignoring ODIAC and FFDAS, I don't think the authors have demonstrated that a regional inventory is generally better than a global inventory for this region of China, just that a regional inventory is better than a population-proxy-based global inventory. In fact, production of the CDIAC 1x1 gridded inventory based on a population proxy has been discontinued. It is my understanding that it has effectively merged with ODIAC such that CDIAC national emission totals are distributed spatially using the ODIAC method, hence the author composition of Oda et al. (2018).**

   Thank you for raising these points – we will clarify this to a greater extent in the text. As noted in the introduction (Page 3 Line 28 to Page 4 Lines 1-3): "As evaluating all existing inventories is outside the scope of our analysis, we focus on investigating the performance of three bottom-up anthropogenic inventories that *represent the dominant methods* currently employed for estimating China's CO_2 emissions." The similarities between CDIAC and ODIAC imply that CDIAC and ODIAC methodology are comparable for these purposes. As you point out, the CDIAC totals are redistributed according to the ODIAC proxies and, this is also clearly stated on the ODIAC website: "Odiac emissions are based on estimates made by Carbon Dioxide

Information and Analysis Center (CDIAC), US Department of Energy, Oak Ridge National Laboratory (ORNL).” The authors of the ACP paper by Oda et al. mentioned by the reviewer note the high correlation between nightlight and population proxies. In addition, they compare the ratio of CDIAC/ODIAC emissions at the 1x1 degree level in Figure 5 of their paper. For the Northern China region relevant to our study, we note that the comparison between the two is either identical (greens), or that ODIAC is either 2-5x higher (yellows) or equivalently 2-5x lower (blues). We also note that in China where data is sparse, downscaling to the very high resolution of 1kmx1km is not synonymous with equivalent increase in information.

In terms of our particular choice of CDIAC over ODIAC, we note this on Page 11, Lines 11-13: “We include the CDIAC inventory here due to its historical prevalence as a benchmark inventory for global indicators, including evaluations of carbon intensity provided by the World Bank (World Bank, 2017).” One of our research goals was to assess, to the best of our ability given the spatial and temporal limitations of the observational data set, China’s goals emerging from the Paris Agreements. We specifically note that one of China’s main goals—reduction in carbon intensity by 60-65% relative to 2005—begged the question of how CDIAC performs at least in the northern China region where our observations restrict us. International carbon intensity calculations currently rely on CDIAC’s estimates, and CDIAC estimates are still used extensively by organizations including the Global Carbon Project, CarbonTracker, and as mentioned previously, ODIAC itself. ODIAC also includes estimates of emissions based on nightlights from satellite retrievals, and a power plant profile data set. For China, this is complicated as the power for the nightlights is often generated in remote areas away from the consumption. As ODIAC is a global model, we again note that it lacks access to China-specific emissions factors from power plants, in addition to lacking information on small-scale power plants that are too small to appear in their data set. (The CARMA power plant data set Oda et al. used specifically provides publicly accessible power plant data, and note on their website that the data was drawn from “the United States, European Union, Canada, India, and South Africa as well as the International Atomic Energy Agency. For facilities lacking publicly-disclosed data, estimates are generated using a new suite of statistical models.” China is not shown to be included in CARMA. In our analysis we specifically note that access to power plant data from China adds significant value over modeled estimates based on global defaults, and is a dominant source of uncertainty, with a greater effect than the choice of spatial distribution proxies. Your comment notes this, in fact: “A strength of ODIAC is that it first uses a point source inventory [CARMA] …” but as we pointed out earlier in this response, CARMA like other global data sources lacks information on China.

We also would like to note the Turnbull et al. (2011) analysis that was brought to our attention following the publication of this study in ACPD, and we will reference their analysis in our discussion. The authors of that study examined China’s fossil fuel and biological emissions using observations from two NOAA/ESRL weekly flask sites dominated by Northern China emissions (Shangdianzi, Tae-ahn Peninsula) and a bottom-up emissions estimate of fossil fuel emissions that combined CDIAC and EDGAR estimates redistributed according to provincial emissions (rather than nightlights/population). We feel that our use of CDIAC (and EDGAR) provides important continuity to the existing (yet sparse) body of scientific literature on this topic.

All that being said, one of our major points here was to show that inventories that are *currently* being extensively used need to be re-assessed in terms of the weight placed on them, at least for the subset that includes China. We reiterate that our goal was to assess the inventories that are *dominant* in our *current* understanding of China's past, present, and future anthropogenic CO_2 emissions and span *typical* approaches. Specifically spotlighting the performance of ODIAC or any other individual inventory for China is certainly valuable, but would be the work for future studies. There are many inventories out there—and unfortunately we will always be missing some. The best we can do is highlight the *typical* ones and the methodology they *represent* both in terms of choice of spatial allocation and in terms of emissions processes themselves.

<u>We have revised the manuscript to be clearer in the following locations, updating references as required:</u>

P23L2. "Future studies can examine this impact by using ODIAC data (Oda et al., 2011) instead of CDIAC, where CDIAC emissions are spatially allocated by nightlights rather than population. While these proxies are highly correlated, Oda et al. (2011) demonstrate improved performance over CDIAC in many parts of the world. Furthermore, the ODIAC inventory is on a very high resolution grid (1km x 1km) more suitable for top-down emissions optimizations study but we caution that in regions where bottom-up observations are not readily accessible, downscaling does not bring with it an equivalent increase in information."

P29L10-13. Conclusions, changing the wording from "require" to "likely to benefit from" to better reflect the uncertainties associated with a single site and subsampling of inventories: "Our results, backed by a robust high-resolution time series of CO_2 observations, show that assessments of China's CO_2 emissions are likely to benefit from regional inventories with a methodology such as that employed in ZHAO, where China-specific field and facility-level data are used with increased reliance on provincial energy statistics."

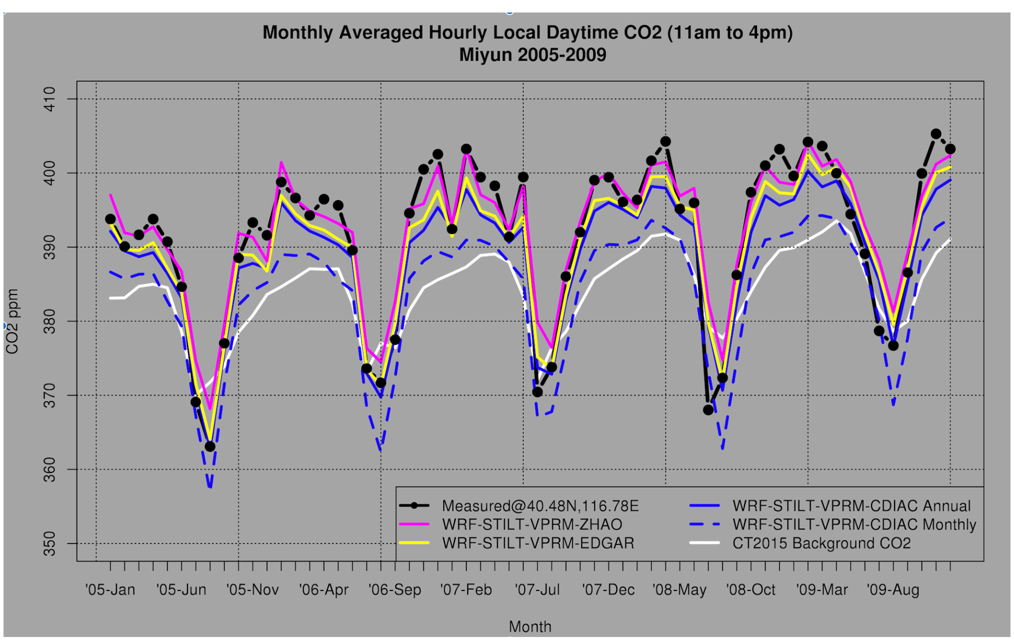
2. **Furthermore, CDIAC 1x1 gridded data have a seasonal cycle (monthly) but the authors state that emissions are invariant over the course of a year. Why was the seasonal cycle ignored? Gregg et al. (2008) discuss the seasonal cycle and show that the amplitude of the seasonal cycle is not negligible. In fact, China has a unique seasonal cycle with a peak in December and a minimum in January, which differs from the standard seasonal cycle of other countries in its latitude range. Most other countries peak during the cold months due to heating or more recently show two peaks due to heating and air-conditioning use in the coldest and warmest months of the year. Due to the use of a single observation station and the changing wind direction with season that the authors have demonstrated, I think the issue of seasonality becomes even more important for assessing inventory biases.**

Thank you for highlighting this. We evaluated seasonality early on in the study, and we found the CDIAC monthly data sets for China consistently led to underestimated CO_2 relative to observations even more so that the annual CDIAC data set. An early graphic from this analysis is displayed below. In the absence of seasonal anthropogenic emissions of sufficient reliability and quality, unfortunately we could not address this issue in the present analysis. Also, as we noted in SI Sect. S4, applying the Nassar et al. temporal scaling factors (time of day/day of week) for China made no significant difference to the results of this study. The point you make is certainly true – seasonality in anthropogenic emissions is very important. The difficulty is finding and accessing seasonal and other temporal scaling factors that are China-specific. Future studies can add that as an important additional dimension to this baseline analysis that used the best information that was readily accessible. We have added a sentence in the text to clarify this.

In addition, we noticed that this was partially addressed in the SI Sect. S4, but it would be more appropriately brought to the main text.

<u>We have revised the manuscript accordingly:</u>

P11L25: "The original inventories do not embed or provide estimates of intra-annual variability. Previous work has found that temporal variations in CO_2 can be significant, and surface CO_2 can be perturbed from 1.5-8ppm based on time of day and/or day of week (Nassar et al., 2013). However, in this study we assume anthropogenic CO_2 fluxes are temporally invariant on intra-annual timescales as the effect of applying the weekly and diurnal scaling factors were not statistically significant. This is ascribed to the difficulty in acquiring appropriate data for establishing reasonable temporal scaling factors for China. No seasonal scaling factors were directly available. While CDIAC does provide monthly gridded inventories with seasonality embedded, these data were found to consistently perform lower relative to observations than the annual CDIAC gridded inventory and are therefore not included in this assessment."



3. **I am also not convinced that the authors have demonstrated that observation-model discrepancies cannot be attributed to their biospheric model fluxes, initial CO_2 fields, or transport from outside of the regional model domain or transport errors. In fact, all of this is very difficult to do (maybe impossible) with a single observation station.**

Information ruling out some of these factors may actually be buried in the supplementary information (35 pages), which I should note is the most extensive that I have ever experienced in many years of reviewing.

This is certainly true. However, the single-site limitation is partially overcome by the long-term (5-year) high resolution (hourly) nature of the dataset. The error analysis is described in the text as well as the supplementary information and involves treating all errors as embedded within the model-measurement residuals. We again would like to point out that a more extensive error analysis cannot be undertaken with a single measurement site. We further note that a study conducted by Turnbull et al. (2011) (which we will include in the discussions section of the revised version of our manuscript) examined the ~weekly flask data from the NOAA/ESRL/WMO sampling network referenced in this study. For the Turnbull et al. study, the greater number of sites was offset by the much lower temporal resolution. In their conclusions, they also note the difficulty of sufficiently assessing bias with sparse data: "Although it is tempting to conclude that the CO_2 [fossil fuel] emission flux we use in the model and reported emissions are accurate, we are at this time not able to sufficiently assess the magnitude of biases in model transport to confirm this. Potential biases in the model transport include the underlying meteorology, particularly wind speed and boundary layer height, and the parameterization of vertical mixing in the model. Nevertheless, this result indicates the promise of top-down atmospheric observations to constrain urban and regional fluxes, as modeled transport is improved, and the observational network becomes denser."

The purpose of the (extensive) SI was to provide as transparent a guide as possible for reproducing and consistently extending this analysis in the event of increased availability of ground CO_2 measurement stations. Not including it would likely have raised more questions than including it.

<u>We have revised the manuscript to caveat the error analysis both in the Uncertainty Analysis section and in the Conclusions:</u>

P14L8: "Absent a dense network of observations, a more sophisticated and extensive error analysis cannot be conducted with meaningful results. Turnbull et al. (2011) faced a similar issue, where weekly flask data collected between 2004 and 2010 from two sites in the NOAA/ESRL/WMO sampling network were used to evaluate a bottom-up fossil inventory based on CDIAC and EDGAR estimates. Turnbull et al. (2011) note the difficulty in assessing the transport error given the paucity of regional observations but also demonstrate the power of top-down assessments given improvements in regional transport modeling and density of observations."

P29L19: "In particular, access to a spatially dense network of measurements will allow for a sophisticated error analysis that can more readily assess uncertainty in key model components such as transport, flux fields, and background concentrations. However, past studies and studies such as this one provide key information that is necessary to guide and motivate more extensive future studies."

4. **It is difficult to predict the global climate policy implications of finding a large error in China's reported CO_2 emissions. For this reason, any scientific studies that suggest our understanding of China's emissions may be incorrect, need to have very solid evidence to support the finding. In the present version of this manuscript, there are some weak spots in the evidence presented.**

   It is not new information that there is large uncertainty in China's reported CO_2 emissions. This is a well-known issue, and we note this in the introduction of the paper beginning on Page 3 Line 11: "China's emissions inventories for CO_2 have a large uncertainty, as indicated by differences in data reported at national and provincial levels. In 2012 this discrepancy was approximately half of China's 2020 emission reduction goals (EIA, 2017; NDRC, 2015; Guan et al., 2012; Zhao et al., 2012)." Our point here is to demonstrate a way forward to reducing the large uncertainty. Namely, we show how even a few observations can help guide emissions analysis and that ultimately a network of strategically placed sites can go a long way to providing an observational basis for reducing the large uncertainty in China's emissions which is already known to exist.

   We agree that there needs to be solid evidence to support findings related to emissions estimates. However, the solid evidence base needs to be built from somewhere and there are currently very few and sparse studies outside of this study (Wang et al., 2010 and Turnbull et al., 2011 being others) that use the best available observations to begin to address that dearth of solid scientific evidence. We note this in the introduction (Page 3, Line 14): "Our study addresses the critical need for independent and observational testing of emissions estimates to enable China to successfully achieve its policy targets. We do this with the best information we have access to.

   <u>We have edited/reworded the manuscript accordingly, to make it clearer without over-reaching:</u>
   P3L14: "Our study is a first step toward addressing the critical need for independent and observational testing of emissions estimates to enable China to successfully achieve its policy targets."

   P29L19: "Absent data from a dense network of high temporal resolution measurements, there will constantly be a tradeoff between drawing conclusions using low-temporal resolution measurements from a few sites like the NOAA/ESRL flask network and continuous data from a single location. Future efforts can use OCO-2 satellite CO_2 data to fill that gap, but would ideally include observations from more ground based sites."

Specific Comments.

P11, line 11-16: CDIAC national emission numbers are a benchmark as stated, but the gridded and spatially distributed data are really a separate dataset.

The gridded data spatially allocates those national emission numbers as described in the documentation associated with the gridded data set so the numbers are certainly linked.

To be clear, however, we have revised the manuscript and references list accordingly:
P4L17: "…and the CDIAC national total (Boden et al., 2016)…"

P17, line 13: Dayalu et al., 2018 is not listed in the references. Do they mean Dayalu et al. 2017, or a different manuscript?

Thank you for noticing that. Apologies, that should be Dayalu et al., 2018 which was recently accepted and published in Biogeosciences. We have changed that in all locations and in the references accordingly.

P17, Fig 4: There is no discussion on why the VPRM signal relative to CarbonTracker is so small for June, when biospheric uptake is near its maximum.

We will expand upon this in the discussion and refer to the Dayalu et al., 2018 paper. Briefly, the low uptake here is associated with the fact that the northern China winter wheat/corn dual cropping region is at the winter wheat/corn transition period in June. While biospheric uptake in other ecosystem would indicate close-to-peak uptake around June, this is not the case for the region influencing the observation station. We note that this pattern is also in agreement with the Turnbull et al. (2011) analysis of the biosphere for a larger region (encompassing our analysis region). In Figure 5 of their paper, the biosphere peak uptake occurs around the July/August time frame with a relatively small drawdown occurring in June.

We have revised the paper accordingly:
P17L11: "As noted in Sect. 3, the regional growing season does not have a typical pattern in that peak uptake occurs around July/August with the onset of the corn growing season. The atypical lower uptake during June represents the winter wheat/corn transition period. These results are consistent with the biological component estimated by Turnbull et al. (2011)."

References:

Boden, T.A., G. Marland, and R.J. Andres. 2016. Global, Regional, and National Fossil-Fuel CO_2 Emissions. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A. doi 10.3334/CDIAC/00001_V2016

Dayalu, A., Munger, J. W., Wofsy, S. C., Wang, Y., Nehrkorn, T., Zhao, Y., McElroy, M. B., Nielsen, C. P., and Luus, K.: Assessing biotic contributions to CO2 fluxes in northern China using the Vegetation, Photosynthesis and Respiration Model (VPRM-CHINA) and observations from 2005 to 2009, Biogeosciences, 15, 6713-6729, https://doi.org/10.5194/bg-15-6713-2018, 2018.

Turnbull, J.C., Tans, P.P., Lehman, S.J., Baker, D., Chung, Y., Gregg, J.S., Miller, J.B., Southon, J.R., Zhao, L., 2011. Atmospheric observations of carbon monoxide and fossil fuel CO_2 emissions from East Asia. Journal of Geophysical Research Atmospheres 116. DOI:10.1029/2011JD016691