General Overview:

The main conclusion of the paper is that a "Chinese specific inventory based on subnational data and domestic field-studies (20% higher than global inventories) agrees significantly better with observations than the global inventories at all timescales." The target of the paper is timely and interesting and certain to draw attention to the policy community outside of the academic trace gas community.

While I believe the comparisons between the inventories to be reasonable, *relatively*, I don't see how the authors can have faith in the overall accuracy of either of the inventories. The total national Chinese fossil fuel emissions are being estimated from a single 6 meter high tower on the outskirts of Beijing. The height of the tower is not even mentioned in the manuscript (that has to be added in main text and certainly highlighted as a caveat in Discussions/Conclusions).

To this reviewer, at a minimum, this paper needs to explore the uncertainty of their result, in particular, with respect to atmospheric transport. At 6 meters CO2 can even vary in a "well mixed column", let alone in shoulder transition times or as a function of PBL scheme. The authors mention McKain et al 2015 for uncertainty references. That paper itself is far from clear about the averaging up of errors in their "bootstrapping" technique and also appears to provide some evidence of robustness of the results to "configurations" of the system. That paper used 4 towers and multiple species, INFLUX (Indianapolis experiment) used a dozen towers and multiple species, and fossil projects in Paris and Los Angeles have similar constraints. If the authors are going to be bold enough to say that national inventories can be measured with a single tower located near the surface on the outskirts of town, they must provide far more evidence on the uncertainties and robustness of their results.

Detailed Comments:

I see little value in including Supp Figures: S6, S12, S13, S14, S15

Optimization method: You can succinctly say that you are "scaling" the whole pattern of fossil up or down based on the mean difference between modeled CO2 and observed CO2, instead of pushing readers to the supplemental information. Please just summarize this in Sect 3.6. Unless I'm misunderstanding the technique.

Background Inflow (Fig S17): It concerns me that your comparisons of CT2016 bgd from the south (LLN), "urban impact" appear to show a couple ppm bias (model > obs). The difference between fossil inventories, as seen in the site CO2, have almost the same pattern (or lack of) as the bgd inflow. These would seem difficult to distinguish.

Supp Sect 8: "At annual scales, the dominant contributor to the CO2 signal are anthropogenic emissions; optimization at annual scales is therefore applied only to the anthropogenic emissions inventories. "

I'm confused here. At annual scales the dominant contributors to CO_2 are *fossil* as stated, but other significant contributors are biological/ocean carbon sink as well as any interannual variability. Please revise this statement.

Supp Sect 8: "The heavily cropped L_0.90 influence region implies rapid turnaround of vegetation carbon stocks at the annual scale, justifying this assumption (18)."

First of all, saying that the L_0.90 is "all crops" seems like quite a stretch in order to justify quick turnover of C. Do you have information to support this? Even if it was 100% crops, the nature of crop-C is a redistribution laterally and seasonally, which is likely very similar from year to year, could easily induce a "low bias" in the croplands (where the tower is), and a high bias in the cities and livestock areas, where the C is utilized, as well as seasonal effects.

Again, these biases should preserve the relative differences between the inventories but affect the overall accuracy of them taken together.

FigS18: The bio signal contribution looks very odd. Max drawdown in PBL in middle of night? Or is there a strong diurnal component to fossil from local sources? This is not what one normally considers a diurnal time series of biological CO_2 to look like in the summer. The "observations" show the classic local drawdown impact on CO_2 in spring/summer w/ flatter diurnal cycle in periods of low GPP, but the models don't seem to reproduce this at all. I'm very concerned that uncertainty in this biological component, whose mean diurnal cycle is much larger than the spread of fossil estimates, could swamp the contrast between the fossil inventories. Furthermore, this figure doesn't seem to match Fig S19, which seems to show the standard respiration buildup (presumably at night).

Section S2: This section would benefit from

1) Summary stats for the wind speed differences
2) An analysis across more than one WRF pixel. For example, you should be able to pull the matching WRF pixel for a dozen or so CMA stations. It's hard to tell whether this cell was just cherry-picked as it isn't the tower site.
3) There is no wind direction information, which in my mind would be much more useful than most of the met data analysis in this section. Windroses by season for example. Mostly we're interested in differences in wind speed, PBL height, and direction across the WRF domain.

L25, Pg28: " .. so this conclusion applies strictly to the other three seasons." What does this mean? Please clarify this statement.

Conclusions: There is no discussion of limitations or caveats anywhere that I could see. The obvious ones include what appear to be a (locally?) poorly modeled biological flux term, biased low respiration (in general) from VPRM, and a fossil "signal" whose pattern may be difficult to distinguish from errors in background inflow. Furthermore, there is no characterization of transport error. This shouldn't always be required but the degree to which the authors are trying to extract information from a single 6m tower seems like it should require caveats as well as some quantitative characterization of uncertainty.