Summary

This paper analyzes different methodologies for calculating background values for CO_2 and CH_4 using a tower network in the Baltimore/DC area. The paper is very thorough, using various tower-based and model based methods combined with a Lagrangian approach to calculate the background. Results for the CO_2 portion behave as one might expect, with reasonable explanations provided throughout the paper and an ideal background methodology specific to that region determined by the end of the paper. Results for CH_4 also behave as expected, so long as expectations are built on a foundation of understanding that bottom-up CH_4 inventories are dreadful. While ultimately the results of this work are not surprising, they are **absolutely essential** in understanding the background variability for the DC/Baltimore region, and a necessary prerequisite to performing any sort of more complex emission quantification analysis using the tower network. To quote the final sentence of this paper "We recommend evaluation of background methods for a given urban domain, as the same background methodology may not be the best-suited for a different network design, region, or trace gas of interest." This paper does just that, and it does it well. **Publish with *very* minor revisions.**

Line 49: The abbreviation CO_2 is used here before it's defined on line 61.

Line 166: *"[CH_4] inventories have been show to disagree significantly with measurements in the region upwind of our domain (Barkley et al., 2019), possibly due to the fact that the inventories are for different years than our study"* While the increase of unconventional activity in the region since 2012 would create an underestimation of emissions, an updated EPA inventory would be nearly as wrong due to their flawed bottom-up inventory methods, as it thinks unconventional wells in that region have an average emission rate of 0.1% (hint: it's larger). Additionally, the EDGAR inventory you use is only about a year off of your analysis, so time isn't too much of an issue there. Personally I'd prefer a stronger statement on why the inventories are off. "*CH4 Inventories disagree with measurements, likely due to underestimations in oil and gas emissions inventories*". Skating around the fact feels like a disservice to the billions of studies screaming into the void that bottom-up oil and gas inventories are too low. We know the inventory is wrong.

Section 2.5.2 Afternoon tower method: May also be worth mentioning as a concern that on days with more complex wind patterns, the upwind tower may not even represent the same airmass as the downwind tower (i.e. frontal crossings or stagnant winds).

Line 216: *"First, each particle is tracked back to its exit location from the domain, and the nearest background station is determined by comparing the exit angle and the angle between the background site and the urban station. If the nearest station does not have observations for the time that the particle exited, the next nearest is used. Until May 2017, only one background site was operational, BUC, meaning that backgrounds constructed using any of the upwind-observation-based 220 methods always use BUC until May 2017, when TMD was established. SFD was established in July 2017, so after that period all three stations were options. Note that in the synthetic data study, we use all three sites for the entire year as the* *ideal case, and then investigate the effect of using only one site without filtering for particular wind directions, as other studies have done."* I may just be confused here, but I would think that all of this is relevant to all your tower-reliant background methods, not just your upwind column one. If so, it feels out of place as the end of the previous paragraph (Line 214) makes it sound like this process and missing tower sites are exclusive to the upwind column method. I figured it out eventually, but it could probably be arranged better.

Lines 239-end of section: Not having gotten to the results yet, I just want to say I hope the column method is the worst because I don't want to have to replicate it on all my tower studies. But it's going to be the best, isn't it, or else I wouldn't be reading about it?

Line 267: "*whereas for CH4, we find large differences between model estimates and observations*". Wetlands would definitely be a problem for BUC, but was it still unsalvageable when winds have a westerly component and BUC would be irrelevant?

Line 416: "*Unlike for CO2, using the Upwind Afternoon observations (green) performs just as well as (even slightly better than, in terms of bias) the Upwind Column (red)"* Thank goodness

Line 424: *"even though the model-based backgrounds might be assumed to better capture the spatial variability of incoming air, that does not seem to be the case, because the poor quality of the emissions products used here negates this advantage".*

So one rather significant problem with your model background approach for CH4 is that neither EDGAR nor EPA inventories contain anything for wetlands. So that's dangerous, and perhaps part of the reason for the poor spatial correlation with the model-based backgrounds, as some wetland maps show wetland emissions in your large domain to play a substantial role in the concentration field, both with emissions on the east coast and emissions from Canada. I could make your life hell and say "redo this using one of the 250 WetCHART ensemble members included in the flux", but honestly I've never seen any of them actually produce anything resembling improvement to model vs obs comparisons. So in the end, it might be best to just mention that your EPA and EDGAR approaches are missing wetland emissions, which could in part explain why the upwind tower approach seems to do better, but there's little that can be done about it (and maybe even an argument that an upwind tower is necessary for CH4 since we can't adequately model a major source of CH4 spatially or temporally).