

Atmos. Chem. Phys. Discuss., referee comment RC1
<https://doi.org/10.5194/acp-2022-281-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on acp-2022-281

Anonymous Referee #1

Referee comment on "Understanding greenhouse gas (GHG) column concentrations in Munich using the Weather Research and Forecasting (WRF) model" by Xinxu Zhao et al., Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2022-281-RC1>, 2022

This paper describes a modelling framework based around WRF that will be used with a network of spectrometers for top-down monitoring of greenhouse gases in Munich. The methodology based around WRF-CHEM and STILT with other datasets for emissions, land cover, boundary conditions etc. and data from the spectrometer network was introduced in a previous paper by the authors. The modelling framework is clearly outlined and some best practices for top-down emission monitoring are described. In particular, the authors describe some qualitative checks for determining when a gradient method may be appropriate for top-down measurements. These concerns can be of interest to other groups working on urban or regional carbon observations. The paper also makes some comparisons of their meteorological fields and modelled greenhouse gas signals to measurements from their networks on a few days in August, 2018. The paper also suggests explanations for discrepancies they observe between their measurements and model. However, some of these explanations are not strongly supported and the authors may need to consider some alternative explanations to make this aspect of the paper stronger.

Additionally, although the authors describe two models in their paper, there is no discussion contrasting the approaches or explaining the differences in results that the authors find with them. The authors also find that their gradient method is unable to isolate signals from their region of interest during their measurement days. I think that a discussion of why this is and how such a network could be improved or the challenges creating a network at a 10's of km scale in a complex source region would be very useful.

The paper is well-written without many typos and well structured for the most part.

Comments:

Figure 1: include a scale bar please. I also suggest you remove extraneous land use categories (eg tundra) from your legend as it is confusing and use a different colour scale (eg greens for vegetation etc.). Jet could be interpreted as a continuous scale but land cover is discrete and categorical. As the figure is drawn now, it is difficult to tell what land covers are what colour on your map. I would also indicate your meteorological reference stations (sondes and surface).

Section 3: Your footprints suggest that you are often sensitive to emissions outside of Munich and you refer to discrepancies in the background concentrations. Therefore a comparison between two stations in Munich to your meteorology may not be appropriate for evaluating model bias. Including more stations within your domain may be better (2 would be appropriate if you were isolating signals in Munich but your initial results suggest you aren't?).

I also think that this section should be combined with your discussion of the radiosonde data and you do a more holistic comparison of the meteorology based on the two.

This section may also be better placed after the section introducing your study area (4.1) so the reader is situated in your field site before discussing the comparison.

181: Since your measurements are not available on every day in August, it might be appropriate to also report the biases on the days when you have measurements as well. Otherwise errors in the wind specific to the measurement days and relevant to your modelling in this paper may be masked.

235: The FTS retrieval typically uses a profile based on NCEP for the pressure and temperature profile and surface pressure from the instrument. Here you use the WRF meteorology to define the model pressure levels and surface pressure from WRF. I wonder if this discrepancy might introduce a bias in your model-measurement comparison. It may not but it is worthwhile to consider and discuss briefly.

260: Here you refer to the WACCM a-priori profile. However, in your section on the EM27, you say you use the retrieval methodology of Dietrich et al (2021). They say they use GGG2014 for their retrieval and this software typically calculates its own a-priori profiles during the retrieval. For the calculation of the averaging kernel correction, you use the a-priori that you use in your retrieval so can you clarify what you used in your retrieval in your section describing your EM27 measurement? Is it WACCM or the GGG-2014 profile?

286: Note that Galkowski et al. are about in-situ concentrations. The size of the bias or variability expected for an in-situ measurement would generally be larger than a column error. Additionally, their analysis would ignore flaws in your stratospheric column so I think it is appropriate to note in your text that the comparison is not like to like.

290: I don't follow why data from the early part of the month aren't included? Please explain why here.

318 and 342: Please clarify why your model XCO₂ is consistently lower than your measurements. Since you removed an overall bias here I would expect the mean bias would be around 0. I also don't understand why the MB isn't 0 since you said above you subtracted it out.

Section 4.2.3: It is not clear if the model comparison you're making in this section has a mean bias removed. Please clarify near the start of the section how exactly the data is treated.

Re nighttime columns: It is not clear to me how you define a nighttime slant column concentration. Since the sun is below the horizon, obviously your zenith angles are negative. Additionally, your averaging kernel matrix is based on retrievals from the day time so it is not clear how you extrapolate it, and I don't think it makes sense to do that if you did. I would clarify what exactly you did here and I note that line 298 ("pressure weighted as a proxy") is very unclear.

398: I don't follow what you mean by "pressure weighted column concentrations" here. I also don't understand how you define the slant column when the sun is below the horizon. If you switch to vertical columns for the night time I would note that here.

422: The upwind site can't be used as a relative background if there are sources upwind of it that are still unmixed. In that case the upwind will see a stronger signal from those sources than the downwind one and they wouldn't be removed in a gradient. This seems to be what you are seeing with your data so you should mention this caveat here.

423: STILT footprints have a magnitude and the footprints generally fall off as you move away from the instrument. So two footprints can overlap within their 90% contours but the instruments can have different measurements since one will be more strongly influenced by some sources (and see a higher peak) than the other. I would keep this in mind when discussing footprint overlap and gradients and think that looking at the difference in magnitude of the footprints could be more quantitative.

489: You should include a time series of XCH₄ gradients similar to figure 9. Without one, it is quite difficult to follow your logic in this section.

499: It is also good to note that your network does not isolate emissions from Munich which is what it was designed to. I think it would be useful to discuss how performing gradients on a 10's of km scale and in a complex source region can complicate isolating signals from an area of interest. This type of limitation is important for other groups.

519: Your paper discusses two distinct modelling frameworks under WRF-GHG and STILT but you do not compare the approaches in terms of their benefits and drawbacks and when one might be more appropriate than the other. I think incorporating this will add a lot to your paper and be useful for other groups.

Re your comments about unknown emissions: The TNO inventory you use seems to include point sources, agriculture and waste according to your van der Gon et al. . (2019) reference (retrived here <https://www.che-project.eu/sites/default/files/2019-02/CHE-D2-3-V1-1.pdf>). Additionally, it seems like the sources you plot on figure 10 fall on regions with emissions. So it might not be correct to say there are missing or unknown sources (eg line 365, 394 and in the abstract) that are responsible for your discrepancies without work to ensure that the sources in the inventory are not responsible for the mismatch you see. Additionally, if there are issues with the multinational inventory, it may be good to reference and discuss work by groups in Toronto (Pak et al., 2021) and California (Cararnza et al., 2018 & Mareklein et al., 2021) to create fully resolved and detailed methane inventories in preparation for top-down monitoring. Otherwise, see my comment about considering other explanations for your discrepancies.

Re statistics: the authors report the quality of their model results using the mean bias between the measurements and model, the root mean square error of the model and the

Pearson's correlation coefficient. However, in a few places in the manuscript, the authors discuss the variability in their measurements and model (eg l 310, 342 and say to the effect that the model captures the variability in the measurements. However in these instances they refer to the mean bias or rmse. This is somewhat misleading. Because the range of variation in their data is on the order of 1-4 ppm for XCO₂ for example, a bias or rmse of 1 ppm in the model is actually quite significant and wouldn't necessarily represent the model capturing the variation or performing well. In those cases the r² which is scaled by the variance in the variables is more appropriate. I suggest that the authors rework their discussion of the fit of their model to reflect that.

Re considering alternative explanations: with the preliminary data and model results the authors suggest reasons for discrepancies they see in the model. However, in some instances it seems like the authors do not justify why they settled on a particular explanation rather than another. For example, the authors attribute a bias in their xco₂ model to the cams model used to provide boundary and initial conditions. However they also mention that results at a subdiurnal scale suggest a too high net ecosystem exchange in their model and the authors suggest there is a flaw in their emission inventory with CH₄. How would the authors know if the cams model is responsible for the mean bias they observe in the daily data? For their analysis of the methane gradients, the authors also conclude that there are missing species on their inventory but they do not consider alternatives that can affect a column gradient. For example as elevation offsets (Hedilius et al., 2017), variability in the background signal at each instrument due to time lags in air reaching the boundary (Jones et al., 2021) or meteorological errors (cf Wu et al., 2018). I think the authors should consider these possibilities and others as well as the idea of missing data in their measurements

Additionally at line 302, can you quantify the background variability your model predicts? That information is useful for interpreting your gradients as any background variability between sites would remain in a gradient.

Re negative gradients in both directions: it is quite surprising that you find negative gradients in CH₄ in both directions. Does that indicate that emissions from Munich itself are small on the days you measure or just that it isn't captured by your network. What are the implications of its emissions being missed during "good" measurement days in terms of your network design.

Re Meteorological validation: Wu et al. (2018) found that the boundary layer height can have an impact on column measurements in addition to the 3d winds. Are you able to assess how your model performs in terms of boundary layer height?

Minor Comments:

5: suggest you change "and interpret" to "attempt preliminary interpretation" to reflect that your conclusions on the model measurement mismatch is still preliminary

10: I would alter or clarify the phrase "1 to 30 August" to clarify that you are only able to use measurements on a limited subset of days rather than the full month

11: Based on your discussion, it seems that the CO₂ signals are not well captured by the model so I suggest you change this line to reflect the different results between CO₂ (poor match) and CH₄ (good match) in your WRF GHG analysis.

12: Because the variability in your measured XCO₂ is on the order of 1 ppm, I would note that the 3.7 ppm bias you observe is quite large and could be taken to indicate a poor model fit for your initial WRF-GHG fields

12: In your write up you say that some of the error in CO₂ may be attributable to flaws in your biogenic CO₂ flux but here you say it's due to the initial and background conditions. Please clarify this?

12: Your abstract is missing a discussion of your CH₄ results with WRF-GHG or STILT.

13: In your write up you say that you can't interpret the XCO₂ gradients because of the biogenic fluxes so you may want to add that in your abstract as the way it is written now makes it sounds like you do interpret them?

22: approx. should be approximately here and in other places

25: adaption should be mitigation

39: The sentence beginning with "Using" is awkward, especially its final phrase. I suggest you break it up or restructure it

59: I think that you should also note that top-down emissions have uncertainty due to their own spatial and temporal representativeness (cf Vaughn et al., 2018) in addition to the other reasons you list. Those issues need to be carefully treated to interpret emissions from top-down.

82: Jones et al. (2021) doesn't do a bayesian inversion using biogenic signals so I don't understand your reference here. They do an inversion for CH₄ in an urban area.

115: not sure what you mean by morphological, you just used specific land use correct?

119: refer to a specific part of the supplement.

120: Later in your write up you refer to the background signals as CAM so I would strongly suggest you introduce and use that acronym here for clarity. As written now, I was confused if IFS and CAM later on were different.

188: How do you treat wind direction differences that occur across the cut in wind directions (ie 179 and -179 are only 2 degrees apart but would be 358 degrees with a simple difference). Additionally, since you mention the standard deviation in the wind direction and since this is a proxy for stability which is important for transport modelling, it might be appropriate to include that as a panel.

Figure 2: I would suggest you adopt an unambiguous date time format in your x axis as the mm/dd/yy is the standard in america but not elsewhere.

204: The reference to Hedilus et al. (2016) is incorrect. I would suggest Gisi et al. (2012) as the paper to refer to for the EM27/sun operating principles.

205: the final product of the standard retrieval for the EM27/SUN is typically a total column average dry air mole fraction not an abundance so it might be better to change the word "abundance".

205: Do you use GGG or PROFIT to fit your spectra? I would include that information and cite the relevant software.

218: I don't think that Vogel et al. is the appropriate citation for saying the FTS is influenced by meteorology. Their work is more of an application rather than an error or bias assessment and Gisi et al. (2012) or Hediulus et al. (2016) may be more appropriate. If you meant that you're adopting their measurement screening it might be better to say "We screened the measurement days following Vogel et al. (2018)."

219: I think changing the phrase "measurement performance" to something like measurement quantity may be a good idea because the number of spectra is the only aspect of performance that you seem to assess

246: I don't understand why you cite Borsdorff et al here. If I understand your write up correctly you use the methodology in Zhao et al. and they don't mention using methods explicitly from this paper.

263: I'm not sure if the reference to Tu et al. is right, it seems like they used a different cut off and if you just meant to refer to the fact that the retrieval becomes worse at higher air mass, Gisi et al. (2012) may be more appropriate.

Figure 3: For clarity I think it would be better to use different colour maps for the different locations and to represent dates in your scatter plots. Someone glancing at the plots could be confused as to what your scatter plot colours refer to (time rather than site).

269: It is unclear what you mean by "We have considered the limited measurement period"

292: Do you mean the unselected days?

308: I would note here that the period you refer to is the time you are actually measuring.

310: If you're talking about variability, an r^2 might be more appropriate than the mean bias. Since you already removed some of the bias with your diurnal average analysis, it is not very surprising that the bias is low.

310: Additionally, given that the variability in your measurements is on the order of 1 ppm, a bias of .8 ppm could be taken as being quite large.

316: It would be useful and add to your argument to quantify the size in terms of ppm that this respiration effect might have on your data.

375: if you mean accounting for differences in the wind vectors with height I think you should say accounting "for wind shear" rather than "differences in wind shear".

408: In your write up about WRF-GHG, you note that you were careful to use realistic release heights for your sources. In STILT, sources are assumed to be in the lower half of the boundary layer so saying surface emissions could be a little confusing.

417: Jones et al. used NCEP pressure weights. Do you do the same or use your WRF field pressure weights?

Figure 8: red curve missing in panel c.

460: See note about consolidating your wind comparisons.

485: The underestimate is only in some sites in figure 9 so specify that.

Figure 10: include natural gas pipeline in legend.

601: Is there a link to this reference

641: Link to reference?

715: link to reference and DOI for dataset?

SI Table S4: you mention the wind speed variability in your text but don't include it here.

SI Figure S11: It appears to me that the 18th and 19th (for some instruments) also have a visual overlap in the footprints but are excluded. Can you explain your reasoning for the exclusion more? Additionally, why are the early month days excluded from the footprint analysis

References

Carranza, Valerie, et al. "Vista-LA: Mapping Methane-Emitting Infrastructure in the Los Angeles Megacity." *Earth System Science Data*, vol. 10, no. 1, 28 Mar. 2018, pp. 653–676, essd.copernicus.org/articles/10/653/2018/, 10.5194/essd-10-653-2018. Accessed 19 Oct. 2022.

Dietrich, Florian, et al. "MUCCnet: Munich Urban Carbon Column Network." *Atmospheric Measurement Techniques*, vol. 14, no. 2, 11 Feb. 2021, pp. 1111–1126, [10.5194/amt-14-1111-2021](https://amt.copernicus.org/articles/14/1111/2021/). Accessed 19 Oct. 2022.

Hedelius, Jacob K., et al. "Emissions and Topographic Effects on Column CO₂ (XCO₂) Variations, with a Focus on the Southern California Megacity." *Journal of Geophysical Research: Atmospheres*, vol. 122, no. 13, 11 July 2017, pp. 7200–7215, [10.1002/2017jd026455](https://doi.org/10.1002/2017jd026455).

Jones, Taylor S., et al. "Assessing Urban Methane Emissions Using Column-Observing Portable Fourier Transform Infrared (FTIR) Spectrometers and a Novel Bayesian Inversion Framework." *Atmospheric Chemistry and Physics*, vol. 21, no. 17, 6 Sept. 2021, pp. 13131–13147, acp.copernicus.org/articles/21/13131/2021/acp-21-13131-2021.html, [10.5194/acp-21-13131-2021](https://doi.org/10.5194/acp-21-13131-2021). Accessed 19 Oct. 2022.

Marklein, Alison R., et al. "Facility-Scale Inventory of Dairy Methane Emissions in California: Implications for Mitigation." *Earth System Science Data*, vol. 13, no. 3, 22 Mar. 2021, pp. 1151–1166, essd.copernicus.org/articles/13/1151/2021/essd-13-1151-2021.pdf, [10.5194/essd-13-1151-2021](https://doi.org/10.5194/essd-13-1151-2021). Accessed 6 July 2021.

Mostafavi Pak, Nasrin, et al. "The Facility Level and Area Methane Emissions Inventory for the Greater Toronto Area (FLAME-GTA)." *Atmospheric Environment*, vol. 252, May 2021, p. 118319, [10.1016/j.atmosenv.2021.118319](https://doi.org/10.1016/j.atmosenv.2021.118319). Accessed 24 Aug. 2021.

Wu, Dien, et al. "A Lagrangian Approach towards Extracting Signals of Urban CO₂ Emissions from Satellite Observations of Atmospheric Column CO₂ (XCO₂): X-Stochastic Time-Inverted Lagrangian Transport Model ("X-STILT V1")." *Geoscientific Model Development*, vol. 11, no. 12, 4 Dec. 2018, pp. 4843–4871, 10.5194/gmd-11-4843-2018. Accessed 20 Oct. 2020.