Review of Hu et al, ACPD, 2020. "Anthropogenic and natural controls on atmospheric $\delta^{13}C$-CO_2 variations in the Yangtze River Delta: Insights from a carbon isotope modeling framework"

General Comments

As the authors point out, this is the first time (that I'm aware of) that CO_2 and $\delta^{13}C$ have been both modeled and measured for an urban area. I'm glad that they've attempted to tackle this issue, because applying multiple data streams (in this case CO_2 and $\delta^{13}C$) within an atmospheric modeling framework should help us better understand urban CO_2 sources and sinks. Although the inverse modeling methodology used is somewhat simplistic, I think it's a good start that can later be made more sophisticated. The forward modeling skill is only 'moderate' in my opinion ($R^2 < 0.2$ for CO_2 and marginally better for $\delta^{13}C$), but this is perhaps unsurprising given the high noise urban environment and the relative proximity of the sampling site to local sources. Beyond the simulations, the authors carry out numerous interesting analyses using a combination of model results and observations. The topic they are addressing is important, the study is ambitious in scope, and is completely appropriate for ACP. There is no doubt that this paper represents a great deal of hard work in terms of both measurements, modeling, and analysis. All this said, I have some significant concerns and questions that need to be addressed before the paper can be published. I will outline my concerns immediately below and then provide detailed line by line questions and comments.

1. $\delta^{13}C$ data

My biggest concern is with the $\delta^{13}C$ measurements presented. In particular, I am having a hard time understanding $\delta^{13}C$ values that are greater (more positive) than -6 per mil, at times (Fig. 6), and more generally, summer afternoon values that appear to be close to -7 per mil (Fig. 7h). Looking at well-established background sites in the Northern Hemisphere such as Mauna Loa (from https://scrippsco2.ucsd.edu/data/atmospheric_co2/mlo.html) and La Jolla, $\delta^{13}C$ values for July 2015 are around -8.4 and -8.3 per mil, respectively. $\delta^{13}C$ from the NOAA network for Dec. 2014 at Mauna Loa (the last month available on their website) is -8.4 vs. -8.6 per mil for Dec. 2014 from the Scripps measurements, strongly suggesting that there are not significant offsets in the Scripps data. Given a rough starting point of ~ -8.4 per mil, it's very difficult to understand how a heavily polluted urban region where, as the authors say, the biosphere is a relatively minor component of fluxes, could raise a broadly representative background value of -8.4 to something close to -7 per mil. In principle, this could occur only with a large removal of CO_2 from the atmosphere by net photosynthesis (leaving the atmosphere more enriched). To put a rough number on this, C3 plants fractionate approximately at a ratio of -0.05 per mil/ppm. This means that the CO_2 levels seen at this study's measurement site would need to be roughly 28 ppm lower than La Jolla (for example) in July. While it's hard to tell exactly what the July 2015 daytime CO_2 is, from Figures 3a and 7g it appears to be right around 400 ppm. For comparison, La Jolla CO_2 for the same month is 397 ppm (the CarbonTracker backgrounds mentioned in the paper are very similar to this value). Another possible explanation for high $\delta^{13}C$ would be a source of CO_2 with an isotopic signature heavier than the atmospheric value of ~ -8 per mil. Cement production ($\delta^{13}C \sim 0$), which is discussed extensively in the text, is of course such a source.

However, even with the relatively large fraction of anthropogenic emissions as cement in the study region, the flux-weighted mean isotopic signature of anthropogenic emissions will still be much lower than the atmospheric values (~ -24 per mil, Fig. 12a). One could also ask the question if air from higher in the atmosphere might have higher $\delta^{13}C$ levels (and be more appropriate to consider as a reference than La Jolla, e.g.). To answer that, we can look at data from 4 km asl from collected aboard aircraft (the NOAA site CAR at around 40 deg. N; see https://www.esrl.noaa.gov/gmd/dv/iadv/), where in July 2015, the $\delta^{13}C$ is around -8.2 per mil, not much different than either La Jolla or Mauna Loa. The NOAA site LEF (a continental forested site with little industrial/urban influence) did record a $\delta^{13}C$ of ~ -7.4 per mil in July, 2015. However, this isotopic enrichment was associated with a CO_2 level of ~370 ppm, much lower than the hemispheric mean background. The only time prior to this paper that I've seen such enriched $\delta^{13}C$ values have been in ice core samples (e.g. Francey et al., Tellus, 1999).

So the question is, why is this happening? I am not an expert in optical $\delta^{13}C$ measurements, but reading the referenced paper Xu et al, ACP, 2017 as well as Ghasemifard et al, 2019 (Atmosphere) and Ghasemifard et al, 2019 (Aerosol and Air Quality Research), it is clear that the Picarro instrument used in this study requires significant corrections due to, among other things, water vapor. The fact that the very high values of $\delta^{13}C$ are seen mainly in summer, while in winter (e.g. Fig. S4) the values seem much more reasonable, makes me wonder if the water corrections (which will be much more significant in summer) are playing a role here. It is worth noting here that even with frequent calibrations with reference air of well-assigned CO_2 and $\delta^{13}C$, the fact that the reference gas is bone dry while the sample air is moist will pose a problem. I'm not saying here that water vapor is the explanation for the unreasonably enriched values during summer but rather suggesting this as a candidate for investigation. Almost more important than the fact that the data appears to be biased, I'm worried that there may be a seasonally varying bias in the data. Because the seasonality of the signal is an important part of the analysis in the study, it's important to make sure that, at the very least, any biases/offsets in the data are constant. I don't know why the data are as enriched as they are, but it's the authors' responsibility to convincingly explain the high $\delta^{13}C$ values they observe.

2. Daytime analysis

The vast majority of both forward and inverse model analyses have focused on afternoon data. The main reason for this is that atmospheric transport models generally have a much harder time simulating shallow nighttime boundary layers often with strong vertical gradients (where fewer model levels are available to capture vertical gradients) than they do simulating higher mid-day boundary layers, when the PBL tends to be well-mixed. Another big advantage of focusing on mid-day data is that the daytime turbulence in the PBL serves to integrate fluxes over a much larger upwind region (in time and space). Thus, conclusions about sources and sinks, especially when using data from just a single site, are much more likely to be spatially representative. I would like to see the model-data comparisons and other analyses using afternoon selection criteria (12-16 hr, e.g.). Even if model-data comparison statistics do not radically improve, other analyses, such as the enhancement proportions of different sectors could change by minimizing the influence of very local sources. With model-data comparisons, in particular, it could be that the model performs similarly for the full record as it does for just the daytime part. If so, this would be an interesting finding.

3. Equations

As detailed below, the details of the equations are hard to follow, especially in terms of what is simulated and what is measured. I have made some suggestions to clarify the notation.

Specific Comments

L46 change v432 to v4.3.2

L47-48. "and constrained the anthropogenic CO_2 emission categories." This is misleading. The scaling factor approach only constrained the total anthropogenic emissions. The isotopic data were used constrain the cement fraction to some extent.

L50. "performed well" This is debatable and subjective. The R^2 values for fits to CO_2 data were less than 0.2. If you want to comment on WRF performance, please quantify instead of saying "well".

L54. Delta_s has not been defined at this point, so you need to say what it means.

L58. Change 'plants' to 'plant'

L78. Change 'by' to 'from'

L85. I don't think this is the correct IPCC reference. What you want to cite here are the IPCC guidelines on emissions calculations, not the IPCC report on the science of climate change.

L88. Change 'into the inversion of global biological CO_2 flux' to 'into the estimation of biological fluxes in atmospheric inversions.' It's not the biological fluxes that are being inverted. Also global is not appropriate here because you are talking about high fossil uncertainties at local scales.

L122. Change 'have recently be' to 'have recently been'

L123. Change 'inversion has been' to 'inversions have been'

L129. Change 'power' to 'the power'

L151. The NOAA/ESRL lab you refer to should now be referred to as NOAA/GML (NOAA Global Monitoring Laboratory).

L176. A) does "ms" in CO_2_ms refer to measured? Or is the left-hand-side of eq. 1 just a simulated quantity. This seems to be the case from the text above, but to make that clearer, I suggest adopting more intuitive notation such as 'CO_2_sim'. B) Delta_CO2 on the right hand side of eq. 1 might be better written as Sum{i=i,n}[Delta_CO2]i to be consistent with eq. 2.

L177. Change ‘hands’ to ‘hand sides’

L179. Again, I think the notation should be clarified. The left hand side of eq. 2, as I understand it, is the simulated value of atmospheric $\delta^{13}C$. Make that notation consistent with that for the simulated value of atmospheric CO_2, e.g. $\delta^{13}C\_sim$.

L183. As mentioned above, in eq. 3, instead of Delta_CO2 use Sum{i=i,n}[Delta_CO2]i. Then it becomes very clear what the definition of delta_s is: the enhancement-weighted mean isotopic value of all sources/sinks.

L184. Change ‘the mixture’ to ‘the enhancement-weighted mean’, which is more precise.

L194-195. I don’t agree with: “background air masses should originate from the free atmosphere at heights of 1000 m or higher above the ground”. In general there shouldn’t be a specific altitude requirement for background air. A more general definition when conducting regional studies would be that background is the concentration or isotope ratio of air when the air enters the regional study domain, which is often determined using back trajectories. The back trajectories (often an ensemble as in the case of STILT) will exit the domain at a variety of altitudes, including possibly below 1000 m. Also, with regard to WLG in particular, this is a remote high altitude site that would be expected to be sampling free tropospheric air most of the time. While it is true that WLG is significantly to the west of the domain 1 border, given the size of the CO_2 enhancements (and $\delta^{13}C$ depletions) at the observation site, I would expect WLG to be a reasonable background site. As mentioned above, I think a more likely reason why WLG doesn’t appear to be a good background site (more so in summer) has less to do with WLG and more to do with potential bias in the dataset. One quick experiment you can do is compare the CO_2 values you extracted from CarbonTracker with the CarbonTracker values for WLG. I doubt there will be a substantial difference. If true, this would suggest that WLG should also be a reasonable background for $\delta^{13}C$.

L198-199. “can cause a high bias in the $\boldsymbol{\delta}^{13}C$-$CO_2$ background when using this approach.” Apologies if I am misinterpreting something, but using WLG is much lower (more negative) than the other background approaches. Why ‘high bias’?

L200. ‘second approach’. It seems that the second approach is a very limited approach whose main purpose is to validate the ‘third approach’. Thus, I would move this after the third approach and maybe not call it an ‘approach’ but say ‘in order to test the second approach, we…’.

L203. Here $\delta^{13}C\_a$ is referred to as an observed quantity, whereas in eq. 2 the same notation was used to refer a simulated quantity.

L205-206. ‘minimize simulated CO_2 enhancement errors’ Are you referring to errors coming from NEE here? Is that why you chose the wintertime? If so, state this more explicitly.

L209. Change ‘equations’ to ‘equation’

L209. Is Delta_CO2 here derived from CO_2_obs – CO_2_bg, or is it simulated? The first usage of Delta_CO2 in eq. 1 implies that Delta_CO2 is a simulated quantity, because you write about eq. 1: "CO_2 was simulated as the sum of background (CO_2_bg) and the contribution from all regional sources/sinks (ΔCO_2)". Perhaps eq. 1 is meant to describe that observed CO_2 can be decomposed into a background component and the contribution from all sources. (Thus implying that Delta_CO2 is not simulated using footprints, but rather can be derived from observations and the background estimate.) However, on line 210 you write that in the third approach you do not need to simulated Delta_CO2_i. Please clarify.

L213. 'Similar methods…' I wouldn't say the studies referred to use similar approaches. For one, they don't involve isotopes. Second, in most of the approaches referenced, back-trajectories and information from remote sites were combined to determine background. In contrast, and very importantly, the $\delta^{13}C$ background determined using method 3 is not independent of the observations themselves. This is a an important point because this is what allows you to define a background that fits so closely to the upper envelope of the observations, despite the fact that the smooth curve fit through the background (Fig. 6) is close to -6 per mil in the summer of 2015. As mentioned earlier, such values are not physically reasonable.

L221-222. '1000 m above ground' Why wouldn't you use the concentrations from CarbonTracker at the altitudes where the back-trajectories exited the domain, instead of 1000 m agl? This may be a mis-interpretation or mis-reading of the Hu et al. 2019 methodology. As mentioned above, the background concentration should be taken from the lats, lons, and alts at which the ensemble of trajectories exit the domain. These values (500 in the case of STILT?) can then be averaged (and some aspect of their variance, perhaps the std. error of the mean, used to compute an uncertainty.)

L223. 'hourly footprint function' say how these were calculated, or provide a reference here. Also, does this imply that footprints and back-trajectories were calculated for every hour of the data record? If so, state this.

L233. When convolving the fluxes with the footprints to produce Delta_CO2 (here it's usage is clearly as a simulated quantity!), were hourly footprints for a single measurement convolved with hourly NEE to account for covariances in the diurnal patterns of both the footprints and the NEE? In other words, was there just a single footprint per measurement, summed over the 7 days, or were there 7x24 footprints saved? When focusing on simulations of biospheric CO_2 this factor is very important. I suspect in your case, where anthropogenic fluxes without a significant diurnal cycle are dominant, neglecting this covariance between NEE and transport is reasonable. However, neither eq. 5 nor the text contain any of this information.

L251. Here, you say that you used CarbonTracker values above 1000 m, which is different than what you said on line 221, where you implied that you said you used values at 1000 m. Please clarify. However, I'll repeat that there is nothing special about 1000 m. You should use the CO_2 value from CarbonTracker at whatever altitude the trajectories left the domain.

L254. Say immediately which version of EDGAR you used.

L258. Modify ‘the most up to date global inventory’ to ‘the most up to date global inventory with sectoral detail’.

L264. It is not quite true that EDGAR v4.3.2 is only available for 2010. This is only true for the version with monthly resolution. Please add this qualifier.

L266. Please explain more about how the factor of 1.145 was calculated. For example, was it based just on CT or did it use EDGAR for 2010 and CarbonTracker for 2014 and 2015? The former would be much better for consistency.

L269. Change ‘posteriori’ to ‘a posteriori’.

L275. Again, with notation, please distinguish simulated (or bottom-up) delta_s as in eq. 6 with delta_s determined by a Keeling Plot, e.g.

L280. How much of non-metallic mineral production is cement? 60%, 90%, 99%? If nearly all, then I recommend stating this here and saying that from here on you will just refer to it as cement.

L308. Change ‘it’ to ‘them’

L310. At least for Hu et al., 2019, the inverse modeling approach was very different than that described here. It is a bit misleading to cite it.

L313. Again, confusing notation. Earlier, CO_2_obs, was referred to as [CO_2]. The “ms” subscript is still a mystery to me. In eq. 1. CO_2_ms was total CO_2 including all sources (and background). Here, Delta_CO2_ms is only the simulated anthropogenic enhancement. Why not make the subscript more intuitive, such as ‘Delta_CO2_anth’?

L313. Are these terms monthly means? Or is a scaling factor calculated at high time frequency and then averaged to a monthly mean SF? Sargent et al., e.g., calculated the mean of all afternoon modeled and observed enhancements. Also, subsetting to afternoon data only, when modeled PBL heights are likely most reliable, is a necessary test.

L329. As mentioned in the general comments, evaluating model performance with only afternoon data is strongly advised.

L331. As mentioned above, I don’t agree that R^2 < 0.2 equals “good” performance.

L348. I’m confused as to how you can compare 2014 and 2015, if 2014 was a full year but 2015 was a partial year. Are you only comparing the months that are common to both years?

L349. Here you say that the emissions are the same for both years. This should be mentioned explicitly in the methods section.

L353. I don't think you can conclude that meteorological differences between years were relatively small based on comparison of annual values. More analysis would be needed. For example, there could be a lot of seasonal cancellation that still results in similar annual averages.

L355. In terms of comparing NEE year to year, I think much more analysis is needed for this to be meaningful. First, NEE changes significantly on a seasonal basis. Second, because you are presumably looking at the full record, nighttime respiration is over-represented. Daytime values on the other hand incorporate both daytime NEE (GPP and ER) and respiration from the previous night (and probably more day/night cycles depending on the size of the domain and the windspeeds).

L362. Note that the GPP product used is derived from the SIF product used. They are not really that independent. This can be seen in the shape of their seasonal cycles.

L370. While you may be able to ignore GPP during the winter, it doesn't mean you can ignore respiration. There will still likely be some NEE.

L383. Change Washington to Washington D.C. (to distinguish from Washington state.)

L385. Change 'Eastern' to 'the eastern'

L389. "indicating greater…' Most likely, yes, but also trapping of emissions in the PBL will play a role. You cannot immediately transfer enhancements to emissions. So you need to qualify this by saying something like 'assuming similar windspeeds and PBL heights…'

L392. Say briefly how the percentages were calculated. I assume by convolution of each regions emissions with the footprints.

L401. Change 'oil refinery' to 'oil refineries'

L412. How were the annual scaling factors calculated? As the unweighted (or weighted?) mean of the monthly ones? Also, the MSFs are never presented. What is their seasonality? This seems like a major result of the CO_2 part of your analysis (and the 'inversion'). The monthly results should be discussed and/or presented.

L415. Change posteriori to 'a posteriori' and 'for YRD area' to 'for the YRD area'.

L416. The last sentence on cement seems unnecessary and out of place, even if it is true.

L429. Using 'discrimination' to describe the isotope ratio of cement production is not appropriate, because it is not a fractionating process like photosynthesis is, e.g. You could instead say 'the isotopic signature associated with cement production…', e.g.

L431. Change 'plants' to 'plant'

L432. Change 'than observed' to 'than the observed'

L438. Regarding Chen et al., 2006, the vertical gradients in that study are based on models. Observations were generally only at 20 m agl. In at least one example (Fig. 1) The vertical gradient looks to just about 0.1 per mil at mid-day. Moreover, at least for the summer months, the simulated values in the surface layer are more enriched due to photosynthetic drawdown.

L439. Ghasemifard is not in the reference list.

L440. Saying that Zugspitze values were around -7 per mil for winter 2013 is not an accurate characterization of the Ghasemifard results. Looking closely at Fig. 2 of Ghasemifard et al. 2019, the average $\delta^{13}C$ for DJF of 2012/2013 is at least 1 per mil lower than -7.0 per mil (even excluding pollution events). This extraordinarily high value of ~ -7.0 is only reached in two cases, once in October and once in January. Also note that -7.0 per mil is an unrealistically high value for Oct. 2012 in relation to similar high altitude data like that from the Scripps Mauna Loa record.

L442. Saying that a clean air event pulls the $\delta^{13}C$ down from -7 to -7.5 does not make sense. Pollution events ($\delta^{13}C$ of fossil fuels average around -30 per mil) will make $\delta^{13}C$ more negative. Here you are saying a *clean* air event makes the $\delta^{13}C$ more negative.

L449. As with CO_2, doing an afternoon hours-only analysis of model performance would be valuable.

L468. 'and generally caused by…' you could test this hypothesis by correlating the model minus obs. residuals for CO2 and $\delta^{13}C$. That is, if $\delta^{13}C$ simulation errors are caused by CO_2 errors, there should be a strong correlation and the slope should be related to the mean isotopic signature of the sources.

L471. In Figure 9, by focusing only on daytime data you may minimize the impact of large night time NEE enhancements (Fig. 7b) and get a better correlation.

L477. After 'was observed in December and July' reference the relevant figure.

L484. Regarding delta_s, presumably this was calculated using equation 6 and the emission proportions listed on line 397. And are these the all-day values shown in Figure 12? Or are these data derived delta_s values from Keeling Plots. As mentioned earlier, different notation for simulated and data-based delta_s would be helpful.

L496. The seasonal cycle attribution is very confusing to me. To start with, Fig. 10 shows that about two thirds of the seasonal cycle is from the $\delta^{13}C$ background. This seems straightforward. However, then you say that the background and regional source terms are 59% and 41% of the seasonal cycle. And then you say that total CO_2 enhancement and CO_2 enhancement components further contribute another 20% each. This seems to add up to 140%.

L499. I don't think your approach to separately investigate photosynthesis and respiration is valid. Negative NEE instances will still have a substantial influence from respiration. And some positive NEE will have substantial photosynthesis. I highly recommend simplifying the analysis

(and make it more accurate) and only analyze the effect of NEE. You just don't have enough information to estimate the effects separately (unless perhaps you use the GPP estimates to partition NEE).

L506-507 'via respiration' I would change this to 'via net respiration', or 'positive NEE', because photosynthesis might be active at some points.

L525. Eliminate 'a' prior to 0.40 per mil.

L526. Why is the impact of cement expressed as a range?

L542. Another notation question: what exactly is delta_ms? It has not been defined.

L561. It's not clear that earlier in the paper you stated a hypothesis.

L565. I disagree that the cement isotopic signature is the most distinct. In fact it is only ~ 8 per mil away from the atmospheric value, while biological sources and other fossil sources are ~ 20 per mil away and thus should exert more leverage on the atmospheric value. I also disagree that its emission is large. It is large for cement compared to other parts of the world, but it is still only ~ 10%.

L567. Provide a reference for the statement that the YRD is the largest cement producing region in the world.

L588. Regarding the 16.85 %, why is there such a large seasonal cycle in the cement enhancement proportion. Can you discuss this or offer any explanation? Seasonal changes in wind direction? Also, I am a bit confused about the a and b superscripts in Table S2. I thought the cement enhancements were simply calculated by convolving the cement emissions and the footprints. Do a and b just refer to whether the proportion is relative to total flux or only anthropogenic flux? If so, I would not use 'considering' in Table S2 and instead say explicitly that a is ratio of cement/anthro and b is a ratio of cement/total.

L609. I don't think you can conclude that delta_s is more sensitive to cement emissions than other emission categories without testing those categories!

L616. 'contributed 0.40 per mil' to the seasonal cycle? Please say this explicitly.

L631. Change 'write' to 'wrote'

L665. You shouldn't really cite a 'Discussions' paper from 2016. Please replace with the final paper (or remove if not available).

L684. The journal title seems to appear twice.

L845. What is [J]?

L901. It would help to link panel B to panel A so one can see which of the cement locations are in which domain. Maybe draw some of the domain boxes in panel B?

L931. Are the units really nmol/m2/s? This implies a *maximum* flux of 1e-14 mol/m2/s, based on the colorbar.

L932. The colorbars/legends are very hard to see in the population density maps. Also, panel B colorbar needs units. Also for panel B, which domain in Figure 1 does this concentration footprint correspond to. It's not clear.

L959. Consider using something besides 'Delta_CO2' as the y-axis label, because this has already been defined differently elsewhere in the paper. Also, in the figure caption be more clear about what 'Delta' means, which (I think) are differences between the same monthly averages in two different years. Panels b,c, and d are referred to after Fig. 5a and are not strongly related to panel a. I recommend separating them from this figure. Also what does 'distance xx' in the legend mean?

L971. In panel a should 'model' really be 'anthropogenic'? b and c are switched in the caption.

L1019. Do the individual data points in the plots represent daily means? Please clarify.

L1051. Relabel panel a y-axis as delta_s.

Note. There are some language erros/typos in the supplement that need to be fixed.