Review of Baier et al. (2025), „Mid-Atlantic U.S. observations of radiocarbon in CO_2: fossil and biogenic source partitioning and model evaluation"

This study investigates and analyses atmospheric observations made during four aircraft campaigns over the northwestern U.S. in different seasons between 2017 and 2019, with a particular focus on the collected $\Delta^{14}CO_2$ flask samples. In a first part, the $\Delta^{14}CO_2$ flask observations were used to interpret the variability of the CO_2 enhancements in the boundary layer by separating the fossil from the biogenic contributions. The second part of the study focuses on model evaluation and provides a comparison between modelled and observed CO_{2ff} concentrations along the flight paths, the latter derived from high-frequency CO measurements and R_{CO} ratios. In a case study of a single afternoon flight, the authors show how the detection of urban CO_{2ff} plumes could help to identify transport model deficits.

In my opinion, this is a relevant study that nicely demonstrates how $\Delta^{14}CO_2$ flask observations from aircraft missions can be used to estimate CO_{2ff} enhancements and to interpret CO_2 signals in the boundary layer over large continental regions. The results are presented in a clear way and the associated uncertainties are carefully calculated. However, what I'm missing a bit is a concluding discussion of what can be learned from the findings of the case study for potential future aircraft missions aimed at evaluating transport model performance, e.g. in different regions. Should such missions focus on urban regions so that the uncertainty in the pseudo CO_{2ff} is small enough for model evaluation purposes? Furthermore, since it is difficult to separate the effects of transport model errors and potential flux biases on modelled CO_{2ff} concentrations, should only plumes from regions with well-known CO_{2ff} fluxes be sampled? Regarding the flask sampling approach, could a (realistic) increase in the $\Delta^{14}CO_2$ data collected during a flight (or a different ratio between ABL and FT flasks) significantly reduce the uncertainty of the pseudo CO_{2ff} record used for model evaluation? If you could elaborate a little on these (exemplary) points, I think it would add to the relevance and applicability of this study.

Overall, this study fits well within the scope of ACP and I recommend it for publication after addressing my (minor) comments below.

Specific comments:

p. 1, l. 17: Please explain the abbreviation "ACT".

p. 2, l. 64: "CO_{2ff}" has not yet been introduced.

p. 3, l. 68: It seems that the study from Graven et al. (2018) is not based on "aircraft $\Delta^{14}CO_2$ observations". They used $\Delta^{14}CO_2$ observations from nine measurement sites in California.

p. 3, l. 75: From what is given in Sect. 2.1., it seems that - apart from the first ACT campaign, which you have excluded for this analysis anyway - the campaigns lasted 2-3 weeks rather than 6 weeks. Please clarify.

p. 3, l. 82: Briefly saying that/why OCS can be used as a tracer for photosynthesis might be helpful here.

p. 3, l. 84: Please cite the final revised version of this manuscript instead:
Maier, F., Rödenbeck, C., Levin, I., Gerbig, C., Gachkivskyi, M., and Hammer, S.: Potential of 14C-based vs. ΔCO-based $\Delta ffCO_2$ observations to estimate urban fossil fuel CO_2 ($ffCO_2$) emissions, Atmos. Chem. Phys., 24, 8183–8203, https://doi.org/10.5194/acp-24-8183-2024, 2024.

p. 4, l. 111: How long was the flask filling time, i.e. for what time interval are the flask samples representative?

p. 5, l. 137: What about CO_2 contributions from the ocean? Fig. 1 indicates that there may be an oceanic influence in spring and fall.

p. 7, l. 196ff: It would be interesting to have the contribution of the CO_{2corr} also in relative numbers of the total CO_{2ff} signal.

p. 7, l. 198-200: To be consistent with what is written in l. 197, change “CO_2” into “CO_{2ff}”.

p. 7, l. 200: Perhaps you could add that the lack of a clear seasonal cycle in the nuclear CO_{2corr} is consistent with the fact that you assumed constant nuclear emissions.

p. 7, l. 206: Please explain what PSUWRF stands for. What is the vertical resolution of the Eulerian model?

p. 7, l. 216-219: This is not clear to me. Why did you create a single flux product only for model runs in 2019? What are you using for 2017-2018? Can this lead to inconsistencies in the CO_{2ff} simulations for the different years? Please clarify.

p. 8, l. 239-241: It is unclear to me how you've calculated R_{CO}. From what is written in this paragraph, I would assume that you've used the median CO enhancement from the continuous measurements and not the CO enhancements from the flask measurements to calculate R_{CO}. Is this true? Is the argument, that the variability of the CO enhancement within one flight is larger than the variability of the CO_{2ff} enhancement. And that the CO enhancements of the flasks are therefore less representative for the whole flight than the CO_{2ff} from the flasks, which is why you’ve used the median CO enhancement from the continuous measurements and the CO_{2ff} from the flasks to calculate R_{CO}? However, in contrast to that is the caption of Fig. 10: “ΔCO from flasks is used to calculate R_{CO}…”. Please motivate and explain the calculation of R_{CO}. It would be helpful to have an additional equation for the calculation of R_{CO}.

Fig. 1: Change “white” -> “black” in the caption.

p. 10, l. 281ff: From Fig. 2b, it appears that the winter 2017 FT samples are not “slightly higher” than the 3-month averaged GGGRN measurements (only the fall and spring samples seem to be higher).

Fig. 2: Does the shaded area of the 3-month averaged GGGRN data indicate the 1-sigma standard deviation?

p. 11, l. 294-295: One could add here that negative CO_{2ff} values could also be explained by an inappropriate $\Delta^{14}CO_2$ background or by underestimated nuclear/bio masking of CO_{2ff}.

p. 11, l. 295ff: Have those studies also used the FT as a $\Delta^{14}CO_2$ background to calculate the CO_{2ff} concentrations?

p. 13, l. 327: Can part of the weak correlations between CO_{2tot} and CO_{2ff} be explained by the small CO_{2ff} signals?

p. 13, l. 330: The spring campaign took place in May, not in March-April, right?

p. 14, l. 332: The fraction of $CO_{2bio}<0$ is similar in fall and winter (8% and 9%, respectively). Could the $CO_{2bio}<0$ data in winter be explained by the fact that the winter campaign took place in March when biosphere is already starting to be active or is it due to observational uncertainties (i.e. is the small fraction of $CO_{2bio}<0$ significant in winter)?

p. 14, l. 341-342: Do you mean: “when combining all seasonal ACT data except fall”?

Fig. 6: Maybe you could briefly mention in the caption of Fig. 6 that you have changed the ordering of the seasons (compared to the previous plots in Fig. 4 and 5).

p. 16, l. 377ff: Another possibility for the large variability in R_{CO} are measurement uncertainties. In Maier et al. (2024b, see the appendix A1 of the study), we have seen that the large relative uncertainty of small CO_{2ff} signals can lead to spuriously high R_{CO}, even in the absence of natural CO sources:

Maier, F., Levin, I., Conil, S., Gachkivskyi, M., Denier van der Gon, H., and Hammer, S.: Uncertainty in continuous ΔCO-based $\Delta ffCO_2$ estimates derived from ^{14}C flask and bottom-up $\Delta CO / \Delta ffCO_2$ ratios, Atmos. Chem. Phys., 24, 8205–8223, https://doi.org/10.5194/acp-24-8205-2024, 2024b.

Fig. 8: Is the distribution of the model-data differences in hourly resolution?

p. 20, l. 414: Is the 20% uncertainty in the fossil fuel CO_2 fluxes also appropriate for the temporal scale of individual hours of the day? I'm wondering if an inappropriate diurnal profile in the CO_{2ff} fluxes could contribute to the observed model-data differences? For a better interpretation of the results, it would be helpful to show the time axis in Fig. 11 in local time instead of UTC, or to indicate in the caption of Fig. 11 what 16:30 UTC is in local time.

p. 20, l. 415ff: Could a slight misalignment in the modelled wind direction be the reason for the temporal shift in the maximum of modelled CO_{2ff} in Washington?

Technical corrections:

p. 2, l. 43: Delete ", however" or connect it with the next sentence.

p. 8, l. 225: "Maier et al., 2023" -> "Maier et al., 2024"

p. 10, l. 288: Delete „or"?

p. 12, l. 314: "four years" -> "three years"?

p. 16, l. 381: „Spring" -> „spring"

Fig. 11: „Psuedo" -> „Pseudo" (in the caption)

p. 20, l. 435: "Maier et al., 2023" -> "Maier et al., 2024"