We want to thank the anonymous referee for the review of our manuscript. Our replies are marked in blue.

The manuscript by Maier et al. presented measurements of 14CO2 and CO in flasks collected at two sites, and the purpose of these measurements is to assess the ratio and the associated uncertainties of CO and fossil-fuel sourced CO2 (ΔffCO2). The latter, proposed by the authors, can be further used to estimate ΔffCO2 based on CO measurements which is easy doing. The idea behind this study appears to be that since C-14 analysis is difficult and there is a need of alternatives to estimate ΔffCO2, while CO that is produced mainly from the same fossil fuel combustion source can serve as a tracer to estimate ΔffCO2 as long as the ratio of CO to ΔffCO2 can be constrained.

I would say this is a fair theory, however, the manuscript is more like a measurement report rather than a research article. This is my first impression of reading this manuscript. The second, as the authors indicated in the manuscript, there are many issues regarding the ratio based on flask measurements, so as the ability of this ratio used to predict ΔffCO2 from CO measurements. This whole idea has a few confusions or flaws that leads to little confidence for its applications. I list my major concerns as follows:

We regret that we have not succeeded in explaining the relevance of our study to the broader community. As we state in the introduction, the basic idea of using CO as a proxy for $ffCO_2$ is more than 20 years old. Even older is the realisation that there will be no semi-continuous measurements of the direct $ffCO_2$ tracer $\Delta^{14}CO_2$ in atmospheric observing networks for the foreseeable future. As we discuss in the paper, CO is not a perfect $ffCO_2$ proxy for many reasons. Therefore, we consider it the primary goal of our study to emphasise the shortcomings of the CO-proxy approach clearly. This includes showing that the use of the CO proxy requires calibration by ^{14}C measurements to achieve the necessary precision. Despite all the difficulties and deficits of the CO-proxy-based $ffCO_2$ estimation, we demonstrate in this paper and in the companion paper by Maier et al. (2023) the great potential of this method. We understand that due to our comprehensive and in-depth analysis of the problems of the CO-proxy, the impression of the measurement report may arise. However, it is precisely this detailed and dedicated investigation of the CO-proxy approach that sets this publication apart from its predecessors and makes future users of the CO-proxy approach, whether for in-situ or remote sensing approaches, aware of its potential and shortcomings.

1). The authors point out the disadvantage of using C-14 measurement to assess ΔffCO2, high cost, low coverage, and etc., and thus there is a need of independent tracers that are easy monitoring. The proposed tracer is CO. But in order to use CO to estimate ΔffCO2 in a site or regions, one has to first measure enough C-14 data to build a robust CO/ΔffCO2 ratio? This is sort of a loop, if there is already C-14 measurements, isn't that can be directly used to calculate ΔffCO2? Although CO measurements can be continuous, so as the derived ΔffCO2 with a known CO/ΔffCO2 ratio, would continuous ΔffCO2 record be offering significantly more valuable information than discrete (e.g., daily or hourly average) ΔffCO2 measurements?

As mentioned above, using $^{14}CO_2$ measurements is the most direct way to estimate $\Delta ffCO_2$ concentrations. However, the sparse $^{14}CO_2$ observations lead to a small amount of $\Delta ffCO_2$ estimates, which could be used in inverse models to infer $ffCO_2$ emissions. The question is then: is the amount of $\Delta ffCO_2$ data enough for getting robust and data-driven inversion results? This is exactly what we investigated in the companion paper (Maier et al., 2023), which followed up this study. It turned out, that we indeed get no robust $ffCO_2$ emission estimates for the surroundings of Heidelberg if we use the ^{14}C-based $\Delta ffCO_2$ estimates from the Heidelberg flasks. This can mainly be explained by the very heterogenic distribution of the $ffCO_2$ sources around Heidelberg (including several large point sources) and the shortcomings of the transport model in accurately simulating $ffCO_2$ concentrations for individual (afternoon) hours. In contrast, the continuous Heidelberg CO-based $\Delta ffCO_2$ estimates, which we derived in this study, yield robust and data-driven inversion results that could be used to investigate the effect of the Corona restrictions and to validate the seasonal cycle of the TNO $ffCO_2$ emission inventory in the main footprint of Heidelberg. Thus, the continuous CO-based $\Delta ffCO_2$ record indeed yields more valuable information than the discrete ^{14}C-based $\Delta ffCO_2$ estimates from flasks for this urban region around Heidelberg and at the present time when the traffic and heating sectors show here similar $CO/ffCO_2$ emission ratios. Of course, if the ^{14}C-based $\Delta ffCO_2$ had a similar temporal resolution like the CO-based $\Delta ffCO_2$, the information content of the ^{14}C-based $\Delta ffCO_2$ would be higher than that of the CO-based $\Delta ffCO_2$. However, as this is not (yet) the case, we conclude that the usage of CO-based $\Delta ffCO_2$ to infer $ffCO_2$ emissions could be a valuable approach for other (urban) sites.
For such a CO-based $\Delta ffCO_2$ inversion, it is essential to have a reliable quantification of the uncertainties of the CO-based $\Delta ffCO_2$ estimates and an investigation of potential seasonal or diurnal cycles in the $\Delta CO/\Delta ffCO_2$ ratios. Both aspects are treated in detail in our present study, which is why it forms the basis for CO-based $\Delta ffCO_2$ inversions.

2) I am not sure if this is the authors' idea: once a robust CO/ΔffCO2 ratio was available, then perhaps this ratio is able to apply to other locations or the same location but at a different time? This is my impression of reading the manuscript, and this is perhaps the most (if not only) significance of establishing the ratio. However, as the measurement reports from the two sites indicates, the CO/ΔffCO2 ratio are different among these sites, and there are also large temporal variations. This causes doubts on the ability of the ratio to be applied to period or regions without C-14 but only CO observations. In fact, due to spatial and temporal variations in the use of fossil fuel energy, as well as the spatial and temporal variations in other sources/factors that could influence the production and atmospheric removal of CO, there is a doubt that whether the CO/ΔffCO2 ratio can stay relatively constant with time and space.

We agree, the $\Delta CO/\Delta ffCO_2$ can show large spatial and temporal differences due to the spatiotemporal variability of the $CO/ffCO_2$ emission ratios and/or non-fossil CO contributions. This is also what we found in our study: the average $\Delta CO/\Delta ffCO_2$ ratio in Heidelberg is lower compared to that at OPE. That's why we do not recommend to apply a single $\Delta CO/\Delta ffCO_2$ ratio for different sites or times, if it is not validated by ^{14}C measurements at the respective sites. In our study, we discuss how many ^{14}C measurements should be performed to get robust $\Delta CO/\Delta ffCO_2$ ratios, which can then be used to construct a CO-based $\Delta ffCO_2$ record with high temporal resolution.

Moreover, our comparison of the observed ratios with inventory-based ratios from TNO clearly reveals that the $\Delta CO/\Delta ffCO_2$ ratios should be calibrated with ^{14}C measurements to avoid large biases in the CO-based $\Delta ffCO_2$ estimates. Ongoing ^{14}C measurements are therefore a prerequisite for monitoring changes in future $\Delta CO/\Delta ffCO_2$ ratios and continuing to apply the CO proxy approach. This information is also relevant for studies combining satellite-based CO observations and inventory-based $CO/ffCO_2$ emission ratios to derive $ffCO_2$ emissions.

The first concern emphasizes that no matter what one needs to measure C-14, and the second indicates the established ratio is probably not able to fulfil the intended tasks. As such, the scientific significance of these measurements is somewhat weak, and this is why I suggested that this manuscript is more like a measurement report.

Unfortunately, it is not straightforward to calculate $\Delta ffCO_2$ concentrations, which then could be used to derive top-down $ffCO_2$ emission estimates. If there were continuous ^{14}C measurements with high spatiotemporal coverage and low uncertainty, we would hardly need to think about using CO as a tracer for $ffCO_2$. However, this is at least not yet the case. Thus, the CO proxy approach is a way to bridge the gap between more reliable but sparse ^{14}C-based $\Delta ffCO_2$ estimates with temporal information derived from CO.
With this in mind, we are convinced that our present study is important and has a scientific significance. We want to summarize here the main points, why we do not think that this study is just a measurement report:

1. In addition to quantifying the uncertainties of the CO-based $\Delta ffCO_2$, we conducted a synthetic data experiment to characterize and interpret the cause of these uncertainties. From this study, we were able to draw conclusions about the impact of the spatiotemporal variability of the $CO/ffCO_2$ emission ratios on the CO-based $\Delta ffCO_2$ uncertainties at an urban and a rural site.
2. We used the STILT model to also simulate $\Delta CO/\Delta ffCO_2$ ratios for both sites. We carefully compared the simulated ratios with our measurements and demonstrate that there can be substantial biases. Our study suggests that the TNO inventory may underestimate the $CO/ffCO_2$ emission ratios in the Rhine Valley. A more detailed investigation (in the revised version of the manuscript) reveals that especially the $CO/ffCO_2$ emissions from the TNO heating sector might be too low in the main footprint of Heidelberg. Such studies are important to improve the emission inventories.
3. By performing a bootstrapping approach, we provide a rough estimate on how many ^{14}C flasks should be collected at other sites to obtain robust $\Delta CO/\Delta ffCO_2$ ratios, which can then be used to derive CO-based $\Delta ffCO_2$ concentrations at those sites. This is an essential information if CO-based $\Delta ffCO_2$ data is to be used in inverse models to estimate $ffCO_2$ emissions. In our follow-up study, we show that the CO-based $\Delta ffCO_2$ data from Heidelberg lead to robust $ffCO_2$ emission estimates, which could be used to validate the seasonal cycle of the TNO emission inventory.

From that we argue that our study has enough scientific significance to be published as a research article.

In addition, I have some technical suggestions that the authors can consider to use or not:

1) in the method part, please state more clearly how other sources especially the biomass burning source of CO is treated to get ΔCO; This includes the modeling of inventory based ratio.

The TNO inventory includes CO emissions from agricultural waste burning, i.e. the burning of crop residues based on a 10 year climatology. However, CO emissions from forest fires are not included. These emissions are erratic and can be obtained from e.g. the Copernicus Global Fire assimilation system (GFAS). For the central European region, forest fires are not a dominant source of CO. We clarified this in the manuscript. (p. 7, l. 205-207)

2) In 2.2.1., the defined Rflask is in fact not used but instead ΔCO/ΔffCO2 in follow-up results and discussion. I suggest to keep constant throughout the manuscript.

Thank you for this hint. We replaced Rflask by $\Delta CO/\Delta ffCO_2$ in our manuscript.

3) Figure 3b, could not understand how real flask measurements can be plotted in the figure, the X-axis is measurements, while the Y-axis is synthetic data which is not scaled to real measurements.

The real and synthetic data share the same X-axis and Y-axis. We labelled the X-axis with "^{14}C-based $\Delta ffCO_2$" because it refers to the ^{14}C-based $\Delta ffCO_2$ observations (in case of the real observations, i.e. the black and grey dots) and to the error-prone synthetic $\Delta ffCO_2$ data (in case of the synthetic data, i.e. the red and orange dots), which should mimic the variability of the ^{14}C-based $\Delta ffCO_2$ observations. Similarly, the Y-axis with label "ΔCO-based $\Delta ffCO_2$" refers to the ΔCO-based $\Delta ffCO_2$ observations as well as to the synthetic ΔCO-based $\Delta ffCO_2$ data, which were constructed by dividing the synthetic ΔCO data by a constant $\Delta CO/\Delta ffCO_2$ ratio of 8.44 ppb/ppm. We tried to make this clearer in the caption of Fig. 3.

References:

Maier, F. M., Rödenbeck, C., Levin, I., Gerbig, C., Gachkivskyi, M., and Hammer, S.: Potential of ^{14}C-based versus ΔCO-based $\Delta ffCO_2$ observations to estimate urban fossil fuel CO_2 ($ffCO_2$) emissions, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-1239, 2023.