We would like to thank reviewer 3 for the effort, interesting questions, and constructive comments. Below we address the points one by one. The reviewer's comments are listed in cursive, with our answers in blue and excerpts from the revised text in red.

Review of Efficacy of high-resolution satellite observations in inverse modeling of carbon monoxide emissions using TM5-4dvar (r1258) by Johann Rasmus Nüß et al.
The article is quite comprehensive with many tests and evaluations. It is somehow frustrating that a lot of efforts has been put on looking at fire emissions with different priors and a setup attributing more errors to fire emissions, but in the end, the results not being presented (sect. 4.3.3). The paper would read better either with the inclusion of the fire emissions results, eventually with some explanation about why it does not work as intended to guide future studies, or by removing the sensitivity to prior fire emissions to make the paper more concise.
We thank the reviewer for this comment. While we share the frustration, an in-depth analysis of the biomass burning emissions for the current setup is not meaningful, as pointed out in Sec 4.3.3. This focus on fire emissions was removed from the already lengthy manuscript for the sake of brevity at an early stage. TM5-4DVAR has been used for the analysis of biomass-burning events in the past, but some parts of the setup are now outdated. This paper introduces various improvements over earlier configurations, one of which is using the modern FINN2.5 as prior emissions. Since the GFED4.1s inventory is well-tested in TM5-4DVAR, the comparison is necessary to showcase the applicability of FINN2.5.

The CO budget is not well presented with a lack of references to other CO inversions studies and other publications that compare inversion results, for instance Elguindi et al. (2020) Intercomparison of magnitudes and trends in anthropogenic surface emissions from bottom-up inventories, top-down estimates, and emission scenarios. Earth's Future, 8, e2020EF001520. https://doi.org/10.1029/2020EF001520
We thank the reviewer for the valuable input and have extended the comparison to the budgets from other top-down CO emission studies, which can be found in L743ff of the revised manuscript and reads:
"Overall, the a posteriori secondary CO source is lower than the a priori production flux in all experiments, as can be seen in the global budgets provided in Table 3, where the posterior masses at the end of the inversion period (final masses) are consistently lower than the prior final masses. Naus et al. (2022), who used a similar setup, also found too high secondary CO production. All fluxes were extrapolated to annual budget terms in Tg CO yr^{-1}. This extrapolation may lead to misleading results when compared to budget terms published elsewhere, because the inversion period of the main inversions includes the biomass burning season, but excludes the increased anthropogenic emissions due to heating during most of the northern hemispheric winter. With this caveat in mind, we compare our prior and posterior budget terms with values from other inversion studies with different setups, namely to Jiang et al. (2017), who assimilated MOPITT CO and methyl chloroform surface measurements in the GEOS-Chem model, Müller et al. (2018), who assimilated IASI CO in the IMAGES model, and Zheng et al. (2019), who assimilated MOPITT CO in the LMDz-SACS model. A detailed comparison of these three studies can be found in Elguindi et al, (2020). Compared to the results of either of those studies, our extrapolated annual a priori budget terms for secondary CO production and chemical loss of CO to OH are far too large. However, our posterior chemical loss falls between the values found in Müller et al. (2018) and Zheng et al. (2019) and our posterior secondary CO production, while still larger, is much closer to what those studies found than our prior. This improved agreement implies that our a posteriori terms are more realistic than the a pirori ones. Note that our secondary production implicitly includes ocean and biogenic CO. While the total production and loss terms show reasonably good agreement with the aforementioned studies, the partitioning by source category of our emission terms differs slightly. Our anthropogenic/fossil fuel a posteriori CO is close to that found by Müller et al. (2018) and Jiang et al. (2017), but significantly lower than that reported by Zheng et al. (2019). In contrast, our biomass burning estimate is close to the multi-year mean of

Zheng et al. (2019). However, due to the high year-to-year variability in biomass burning emissions, as shown by both Müller et al. (2018) and Zheng et al. (2019), this result is difficult to interpret, especially since neither study covers 2018."

Minor comments:
The title is misleading as most of the work is done with large scale setup for monthly emission inversions, and is focus on the comparison of TROPOMI inversions with global in-situ network.
We agree with this assessment. The title has been revised to:
"Top-down CO emission estimates using TROPOMI CO data in the TM5-4DVAR (r1258) inverse modeling suit"

Abstract: "Compared to the bottom-up estimates, all experiments result in strong (by up to 75%) broad-scale emission reductions in China and India. In part, the reduction over China can be attributed to policy changes." Explain the time period, the differences can be in absolute sense, or because of the change in emissions over time, please clarify.
The abstract has been updated to reflect the requested clarification. The sentence now reads:
"Compared to the bottom-up estimates, all experiments result in strong (by up to 75 %) broad-scale emission reductions in China and India throughout the entire inversion period."

L81: "spatial sampling of IASI (up to about 25×25km2; Clerbaux et al., 2009)." The citation indicates that IASI has footprints with diameters of 12 km diameter footprint. It is confusing because it looks like the MOPITT pixels are of similar size compared to IASI
The spatial sampling of a satellite is not always the same as its footprint size. Unlike TROPOMI and MOPITT, where the spatial sampling and footprint size are the same, i.e. two adjacent footprints touch, for IASI the footprints are much smaller, than the distance between their center points. Every 50km view of IASI has four 12km footprints loosely arranged in a square (compare Fig 1 in Clerbaux et al., 2009). This leads to the distance between the center points of two adjacent footprints, the spatial sampling, to be around 25km. This quantity is more relevant for global scale inversions because it informs on the amount of data to be analyzed and the size of structures that can be resolved. The information on the footprint size has been added to the relevant places in the manuscript to reduce confusion, which now reads:
"Furthermore, TROPOMI procures CO observations at a high spatial resolution of up to $7 \times 7\,\mathrm{km}^2$ (Veefkind et al., 2012), which is roughly 10 times higher than the resolution of MOPITT (up to about $22 \times 22\,\mathrm{km}^2$; Drummond et al., 2010) and the spatial sampling of IASI (up to about $25 \times 25\,\mathrm{km}^2$ with 12 km diameter footprints; Clerbaux et al., 2009)."
and:
"For example, an empirically chosen variance inflation of 2 was used in Chevallier (2007) for Orbiting Carbon Observatory (OCO) CO_2 observations gridded to $3.75° \times 2.5°$, an inflation of 50 was used in Hooghiemstra et al. (2012a) and Naus et al. (2022) for MOPITT V4 (gridded to $1° \times 1°$) and V8 CO observations, respectively, and an inflation of again 50 was used in both Krol et al. (2013) and Nechita-Banda et al. (2018) for IASI CO observations at their native sampling resolution of up to about $25 \times 25\,\mathrm{km}^2$, with footprints of at least 12 km diameter."

L142: 2.2 4DVAR approach: I am sorry if I missed it, but what is the assimilation window ?
We may have used a different terminology here. The inversion periods (the time spans for which the model runs and data are considered) are 2018/01/01-2018/07/01 and 2018/06/01-2019/01/01 for the spin-up and main inversions, respectively (see Sec 2.3.3). The resolution of the state (the time step at which the model is allowed to scale the emissions in each category) is 1 month for both secondary CO production and fossil fuel CO and 1 day for biomass burning CO (see Sec 2.3.1).

L227: "the a prior error is set to zero over the ocean" Chose between a priori or prior
Fixed.

L248: "Therefore, we use an exponentially decreasing correlation time of 9.5 months x for the secondary CO production at different times from the same cell." This seems to be a long time as CO itself has an average lifetime of 2 to 3 months.

The correlation time denotes how long changes in the overall production patterns are expected to persist in time. This is independent of the lifetime of CO and more related to the lifetime of CO's precursor species (e.g. methane) and the time scales at which we expect the sources of those precursors to change.

L265: "and the fairly up to date inventory (with historical data up to 2014 and projected data from 2015 onwards)," It is not really up to date. But what matters in the end is how the scenario matches the observations.

For the time period investigated in this study, the year 2018, and at the time of initial creation of this study, CMIP6 was one of the most up-to-date global anthropogenic emission inventories. Alternatives, such as the CAMS-GLOBAL-ANT_v2.3 inventory, had the same limitation of providing only projected data after 2014. Considering that changes in anthropogenic emissions on a global scale are usually fairly predictable over relatively short time spans such as three years, the CMIP6 emissions should be adequate for inversions because they correctly predict where countries are and on what order of magnitude their emissions are. The inversion will correct for any potential mismatches between emissions and TROPOMI CO columns.

L728: "Regardless, our extrapolated annual a posteriori budget terms are much closer to the ones found in literature (e.g. Zheng et al., 2019) than the a priori terms, implying that the a posteriori terms are more realistic." Please explain and clarify, maybe cite more paper about CO budgets, it is an interesting part of the paper that is not really complete at the moment.

As mentioned in response to the second comment above, the comparison to the budgets from other top-down CO emission studies has been extended, and relevant literature has been added. Further investigations of the implications of this comparison are the subject of a future study.