Reviewer #1

Review of The CO anthropogenic emissions in Europe from 2011 to 2021: insights from the MOPITT satellite data by Audrey Fortems-Cheiney et al.

This study could provide interesting insights on an emission inversion framework for CO at the European scale. Posterior emissions are used to verify trends.

We wish to thank the referee for his/her helpful comments. His/her full comments are copied hereafter in normal black font, and our responses are inserted in between in bold font.

There is a lack of bibliographic references, which leads to misleading statements in the introduction and probably for some of the design of the study and the choice of parameters used in the inversions. There has been CO emission inversions at the regional scales, for instance Jiang et al. (2015) and Qu et al., (2022) performed regional MOPITT inversions at the grid cell level at 0.5° × 0.667°.

We added references for the study of CO at the global scale in the introduction: « Over the last two decades, the space-borne Measurement of Pollution in the Troposphere ... have revolutionized our ability to map CO concentrations and to understand the trends and the spatio-temporal variability of its concentrations and emissions(Arellano et al., 2006;Chevallier et al., 2009; Jones et al., 2009; Kopacz et al., 2010; Jiang et al., 2011; Fortems-Cheiney et al., 2011; Hooghiemstra et al., 2012; Miyazaki et al., 2015; George et al., 2015; Yin et al., 2015; Jiang et al., 2017; Zheng et al., 2018; Buchholz et al., 2021; Gaubert et al., 2023)."

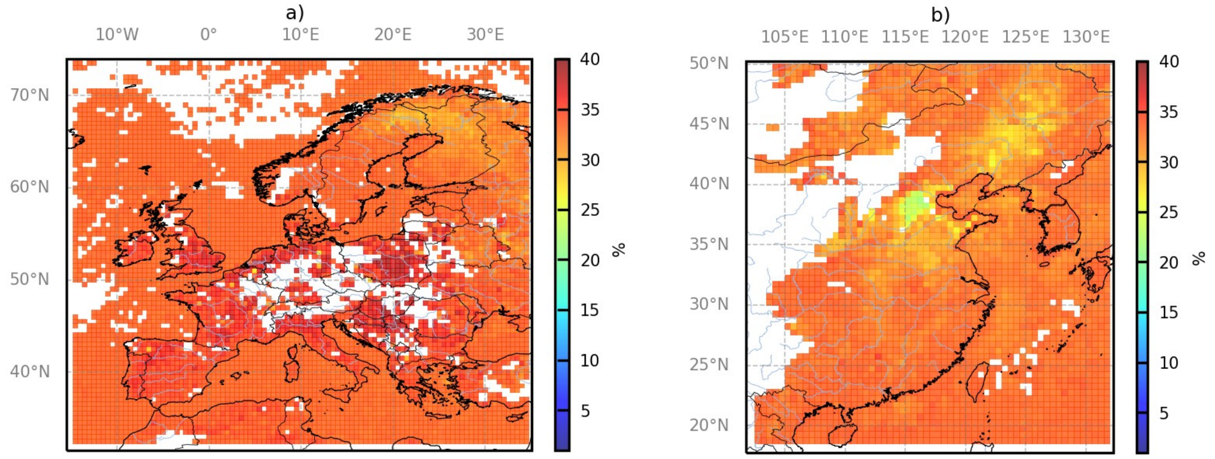
We have also added the references for the study of CO at the regional scale: « In the past decade, CO regional scale inversions based on the MOPITT data covered CO emissions in North America (Jiang et al., 2015) and East Asia (Qu et al., 2022). To our knowledge, there has been only a few studies covering the European CO emissions based on satellite observations (Konovalov et al., 2016; Fortems-Cheiney et al., 2021), this continent being more challenging for regional scale inversions of the CO anthropogenic emissions owing to weaker CO signal (Konovalov et al., 2016).»

We have also changed the sentence: « The ability of regional inverse systems to quantify CO anthropogenic emission budgets at the national and monthly to annual scales in Europe from the MOPITT TIR-NIR satellite observations has not been assessed yet.»

There are two main concerns:

1. MOPITT errors reported on Fig. 3c (30 to 40 %) seems to be larger than usual.

Indeed, the MOPITT errors appear to be larger over Europe than over other parts of the globe, such as over Eastern China (see the figure below).



Averages of the errors associated with theCO MOPITT super observations a) over Europe in January 2015 as in Figure 5b and b) over Eastern China in January 2019, in %.

Regarding this topic, it is important to note that studies in the literature often use MOPITT "super-observations" within relatively large model grid, i.e. aggregated observations based on the average of all the observation within the model grid cells. The corresponding observation error, in principle, should be smaller than that of the individual observations (typically by a factor of 1/sqrt(nobs)when deriving super-observations as the mean of the observations, and if assuming that the errors in the different retrieval within a grid cell are totally independent from each other). However, here, when defining such super-observations as the observation whose value is the median of the ensemble of observation within a 0.5°x0.5° grid cell of the CTM, and within the CTM physical time steps, we assign the observation error associated to this observation to the super-observation, which is implicitly a conservative observation error derivation. We assume that a large part of the errors on the individual retrievals bear spatial correlations between the different retrievals at the scale of the CHIMERE grid cells (Deeter et al., 2019).

In any case, the typical number of MOPITT observations per CTM grid cell is often about 1 or 2or 3 observations. Cases when it is 4 to 6 are relatively rare, so that the choice to decrease the error on the super-observation as a function of the number of observation or not do not imply large changes in most of the cases.

We have changed the sentences: « In order to associate the super-observations to a real AK, the super-observations have been taken as the individual observation corresponding to the value of the median of the MOPITT concentrations within the 0.5°x0.5°grid-cell of the CTM and within the CTM physical time steps (about 5-10 min). The AK and the uncertainty associated to this individual value are then used to define the AK and uncertainty for the « super-observation ». In principle, the observation error associated to such a median value should be smaller than the error associated to individual observation, but, here, we keep the error for the individual observation used to define the super-observation as a conservative estimate of the super-observation error. The super-observations therefore do not have a smaller error than the individual observations. »

This is concerning because the main result of the paper indicates a lack of convergence of the emission optimization “The posterior simulation still presents positive biases compared to the observations, which can be partly explained by large errors in the MOPITT observations”.

Yes, we agree that the lack of convergence can be partly explained by the large errors in the MOPITT observations. This is now better detailed in the conclusion. However, we cannot push the system to overfit these data and we must follow the best knowledge of the level of errors in the Bayesian inversion framework.

There are techniques to adjust the errors in order to reach convergence and at least an evaluation of the model data mismatch could be presented.

We already indicate numbers demonstrating that our system reaches convergence (reduction of the norm of the gradient of the cost function by more than 90%)and we have checked that the reduction of the cost function (and its observational component) hardly evolve in the last iterations of the inversions (i.e. that the algorithm has converged well). The model data mismatch is also already presented in Figure 3.

The MOPITT data reports both the measurement error covariance and the smoothing error covariance in addition to the often use total retrieval error covariance. In this study, the averaging kernels (smoothing) are applied in the observation operator (as it should be) to estimate the columns, only the measurement error covariance should then be included, as done in Gaubert et al., (2023).

We acknowledge that we are not familiar with this type of treatment for the retrieval error. We are accustomed to following the recommendations of the MOPITT user’s guides and to our knowledge, there is no mention of using only measurement error in the recommendations of these user’s guides so far. Here, we have used the estimated errors (i.e., uncertainties) available in the error field (second element) of the "Retrieved CO Surface Mixing Ratio" variables of the MOP02 files, including a smoothing error and the instrumental noise, as indicated in a MOPITT user’s guide (MOPITT V7 Level 2 Data Quality Summary, 2016).
Nevertheless, we take this information into account. We have added this information in the conclusion: “The robustness of the inversions would still benefit from a refinement of our configuration of the R matrix which would lead to a better fit to the observations. First, we should probably investigate the components of the retrieval errors which are distributed along with the MOPITT product. Gaubert et al., (2023) indicate that when applying the averaging kernel, the smoothing error could be ignored and that the weight of this component is significant.»

The authors should at least consider showing some statistics on the convergence (e.g., chi-square) of the assimilation for the entire run and adjust the errors accordingly.

As mentioned above, we already show some statistics on the convergence with the reduction of the norm of the gradient of the cost function by more than 90%, which indicates a robust mathematic behavior of the system.

The uncertainties in the observations and the uncertainties in the prior estimate of the control vector are characterized by their covariance matrices R and B. The matrices B and

R are generally derived from expert knowledge based on studies on the performances of the atmospheric and process models (Fortems-Cheiney et al., 2021). The derivation of the B covariance matrix is a complicated task and there is still a critical lack of knowledge on the amplitude and spatio-temporal patterns of uncertainties in anthropogenic emissions of pollutants and GHG (Super et al., 2020).

We have already detailed above our rationale and computations for the derivation of the R matrix. The error standard deviations assigned to the prior CO emissions in B at 1-day and 0.5° resolution are 100 %. We have added sentences in the text explaining this choice: « This value of 100 % has already been chosen in the literature (Pétron et al. 2002 ; Yumimotoa and Uno 2006, Kopacz et al. 2010 ; Fortems-Cheiney et al. 2011 ; Fortems-Cheiney et al. 2012 ; Fortems-Cheiney et al., 2021).Even though annual CO emissions in western Europe may be well known, with uncertainties of 6 % according to Super et al. (2020), larger uncertainties could affect eastern Europe. Moreover, large uncertainties still affect bottom-up emission inventories at the 0.5° resolution: spatial disaggregation of the national-scale estimates to provide gridded estimates causes a significant increase in the uncertainty for CO (Super et al., 2020) ».

The $\chi 2$ value is used to diagnose balance between actual errors and estimated errors. Our $\chi 2$ is of about 0.2 for January 2015.When the $\chi 2$ value is smaller than the ideal value of 1, it suggests overestimated background error covariance or observation errors. This $\chi 2$ may therefore confirm that our set-up of the uncertainties in the covariance matrices R are too large (see the discussion above regarding the assignment of the observation error for super-observations). However, the typical factor of decrease of the observation error for super-observation when accounting for the noise component of the individual retrieval errors may hardly explain that the $\chi 2$ is about 0.2. Actually, we also have an important level of prior uncertainty at the model grid cell and 1-day scale (100%), i.e. the balance between the prior and observation uncertainties is not necessarily misrepresented. The robustness of the interpretation of the $\chi 2$ diagnostic can thus hardly be used to correct the observation errors, for which we have tried to assign a sensible level based on the reported retrieval errors.

As mentioned above, we now indicate in the conclusion that our derivation of the error associated with each super-observation could be conservative, and we discuss the potential and challenges associated to the improvement of the assignment of observation errors for individual observations (Gaubert et al., 2023) and for super-observations:« The posterior simulation still presents positive biases compared to the observations. The minimization algorithm of the inversion appears to converge correctly with the constraints used in practice. Therefore, these residual positive biases can mainly explained by the large errors associated to the observations in our inversion framework. As discussed in Section 2.2, our derivation of the error associated with each super-observation is conservative. Other indices support this assumption. In particular, the $\chi 2$ diagnostic (Ménard and Chang, 2000) is significantly lower than 1. This indicates that the B and R matrices used here to characterize the prior and observation errors likely overestimate~~s~~ the amplitude of these errors (Ménard and Chang, 2000). However, even if assuming that the observation errors would only consist in random noise uncorrelated in space and setting the error on the super-observation as that of the average of the number of observations nbobs in the model

grid cells, i.e. of the order of 1/sqrt(nbobs) times the observation error for individual observations, the impact would be moderate since nbobs is generally equal to 1 to 3. Actually, the set-up of the B matrix is also rather conservative, and the balance between the two errors in the set-up of the B and R matrices may be relatively good. Therefore, the lack of fit to the observations in these inversions could be associated to the large retrieval error corresponding to the MOPITT product. The robustness of the inversions would still benefit from a refinement of our configuration of the R matrix which would lead to a better fit to the observations. First, we should probably investigate the components of the retrieval errors which are distributed along with the MOPITT product. Gaubert et al., (2023) indicate that when applying the averaging kernel, the smoothing error could be ignored and that the weight of this component is significant. The revision of our conservative assignment of the observation errors to the super-observation would be more challenging. It would require a good knowledge of the respective weight of the random noise (without spatial correlation) and of the systematic errors (with spatial correlations) in the total retrieval errors, as well as a good knowledge of the typical correlation length scales of the systematic errors, while we could lack of insights regarding this. The use of notional assumptions (as for the characterization of the model error) may still represent a sensible trade-off and allow for an improved assimilation of the observations. »

Fortems-Cheiney, A., F. Chevallier, I. Pison, P. Bousquet, S. Szopa, M. N. Deeter, and C. Clerbaux (2011), Ten years of CO emissions as seen from Measurements of Pollution in the Troposphere (MOPITT), *J. Geophys. Res.*, 116, D05304, doi:10.1029/2010JD014416.

Fortems-Cheiney, A., Chevallier, F., Pison, I., Bousquet, P., Saunois, M., Szopa, S., Cressot, C., Kurosu, T. P., Chance, K., and Fried, A.: The formaldehyde budget as seen by a global-scale multi-constraint and multi-species inversion system, Atmos. Chem. Phys., 12, 6699–6721, https://doi.org/10.5194/acp-12-6699-2012, 2012.

Kopacz, M., Jacob, D. J., Fisher, J. A., Logan, J. A., Zhang, L., Megretskaia, I. A., Yantosca, R. M., Singh, K., Henze, D. K., Burrows, J. P., Buchwitz, M., Khlystova, I., McMillan, W. W., Gille, J. C., Edwards, D. P., Eldering, A., Thouret, V., and Nedelec, P.: Global estimates of CO sources with high resolution by adjoint inversion of multiple satellite datasets (MOPITT, AIRS, SCIAMACHY, TES), Atmos. Chem. Phys., 10, 855–876, https://doi.org/10.5194/acp-10-855-2010, 2010.

Ménard, R., and L. Chang, 2000: Assimilation of Stratospheric Chemical Tracer Observations Using a Kalman Filter. Part II: χ2-Validated Results and Analysis of Variance and Correlation Dynamics. *Mon. Wea. Rev.*, **128**, 2672–2686, https://doi.org/10.1175/1520-0493(2000)128<2672:AOSCTO>2.0.CO;2.

Pétron, G., C. Granier, B. Khattatov, J.-F. Lamarque, V. Yudin, J.-F. Müller, and J. Gille, Inverse modeling of carbon monoxide surface emissions using Climate Monitoring and Diagnostics Laboratory network observations, *J. Geophys. Res.*, 107(D24), 4761, doi:10.1029/2001JD001305, 2002.

Yumimotoa K, Uno I. Adjoint inverse modeling of CO emissions over Eastern Asia using four-dimensional variational data assimilation. Atmos. Chem. Phys. 2006; 40: 6836–6845. 10.3402/tellusb.v64i0.19047.

2. Using a spatial correlations e-folding length of 50 km is effectively forcing the system to constrain hotspots only. Qu et al., (2022) used correlation lengths that varied by sectors with a 100 km to 200 km range. Only point sources from the energy sectors were considered to be at scales smaller than 100 km.

With our set-up, the inversions system is free, in principle, to homogeneously correct large regions to match large scale gradients in the satellite data, which cover the European domain relatively well; putting larger spatial correlations in the prior error covariance matrix would smooth the corrections of the inversion but it has to be consistent with the actual structures of error correlations in the gridded inventories and we are not aware of analysis demonstrating that these structures have long typical lengths.

On the opposite, we think that isotropic correlations decreasing with distance are not good representations of the error on anthropogenic emissions, even for large sectors of activity; considering the heterogeneity of anthropogenic emissions favors a more conservative view by limiting extrapolation via the matrix B (see Super et al., 2023 for an analysis of the uncertainties in the gridded emissions of CO_2).

Our results show that the inversions do not really filter large scale variations of CO in the data associated to the emissions, the MOPITT data providing constraints over hotspots only. Such a result should not be blurred by artificially extrapolating the information on the emission hotspot to the rest of the countries via spatial correlations in the prior error covariances.

Super, I., Choulga, M. and Hohenberger, T.: PED uncertainty 2018and uncertainties based on Monte Carlo simulation using the emission model from D2.5 v2, D2.7 COCO2 Deliverable, https://www.coco2-project.eu/sites/default/files/2023-11/CoCO2-D2-7-V1-0.pdf

Ma et al. (2019) and Gaubert et al. (2020) considers larger correlation lengths of 600 and 500 km on the basis that emission inventories are constructed at the province level (China) or at the scale of entire countries.

In general, the emission inventories are indeed often constructed at the scale of country and the robustness of the budgets are considered maximum at the national scale. However, such national budgets are then disaggregated in space and this spatial disaggregation, in principle, creates uncertainty and negative correlations at sub-national scale. Furthermore, the disaggregation follows average activity maps which generally lack of granularity, generating complex and heterogeneous spatial structures of the uncertainties. Errors on average emission factors by detailed sectors of activity, create positive correlations across the countries and provinces, but the correlations within the countries associated to the actual local emission factors are complex, the errors in average emission factors generate varying correlations across countries due to highly varying situations between different pairs of countries, and the combination between these correlations and the error structure associated to the spatial disaggregation of the emissions results in complex error correlation features, especially when working with aggregated sectors. Again, this pushes for a conservative modeling of the spatial correlation to avoid abusive extrapolation of the

information from the constraint of the satellite data on emission hotspots(e.g. from traffic in large urban areas to major road axis or to traffic in small towns).

While they are not inverting the CO emissions, Inness et al., 2023 show a global mean horizontal correlation length of 125 km at the surface level (Figure 1). This is important when the objective of the paper is to assessed "the ability of regional inverse systems to quantify CO budgets at the national scale from the MOPITT TIR-NIR satellite observations" and that the corrections of the hot spots are more "convincing".

Inness et al. 2023 do not derive this correlation length from an analysis of the structures of errors in emission inventories. The use of isotropic error correlations decreasing with distance with long length-scales is a kind of tradition from atmospheric transport inversion cases with poor observation coverage and targeting relatively diffuse flux fields (such as natural greenhouse gases fluxes), but it may not be appropriate to tackle the anthropogenic emissions and to spatialized inventories, disaggregated from national budgets. This explains why "fossil fuel data assimilation" systems where the inversions controls parameters of emission models rather than emission maps become increasingly appealing(Kaminski et al., 2022).Furthermore, the good coverage from satellite observation should allow not relying on such simple correlation model to ensure that the top down constraint in inversions is significant. Working with short correlations emphasize the direct observational constraint in the inversion, and the spatial coverage of the MOPITT observations could have allowed, in principle, a direct control of the national scale budgets, even though the results demonstrate it is currently not the case.

Kaminski, T *et al* : Assimilation of atmospheric CO_2 observations from space can support national CO_2 emission inventories, *Environ. Res. Lett.*17 014015, DOI 10.1088/1748-9326/ac3cea, 2022.

Minor comments:

Check that the figures appear in the same order in the text.

This has been done.

Figure 5's colorbar indicates "ppb" while the maps show the number of observations.

This has been corrected.

References:

Gaubert, B., Emmons, L. K., Raeder, K., Tilmes, S., Miyazaki, K., Arellano Jr., A. F., Elguindi, N., Granier, C., Tang, W., Barré, J., Worden, H. M., Buchholz, R. R., Edwards, D. P., Franke, P., Anderson, J. L., Saunois, M., Schroeder, J., Woo, J.-H., Simpson, I. J., Blake, D. R., Meinardi, S., Wennberg, P. O., Crounse, J., Teng, A., Kim, M., Dickerson, R. R., He, H., Ren, X., Pusede, S. E., and Diskin, G. S.: Correcting model biases of CO in East Asia: impact on oxidant distributions during KORUS-AQ, Atmos. Chem. Phys., 20, 14617–14647, https://doi.org/10.5194/acp-20-14617-2020, 2020.

Gaubert, B.; Edwards, D.P.; Anderson, J.L.; Arellano, A.F.; Barré, J.; Buchholz, R.R.; Darras, S.; Emmons, L.K.; Fillmore, D.; Granier, C.; et al. Global Scale Inversions from MOPITT CO and MODIS AOD. Remote Sens. 2023, 15, 4813. https://doi.org/10.3390/rs15194813

Ma, C., Wang, T., Mizzi, A. P., Anderson, J. L., Zhuang, B., Xie, M., & Wu, R. (2019). Multiconstituent data assimilation with WRF-Chem/DART: Potential for adjusting anthropogenic emissions and improving air quality forecasts over eastern China. Journal of Geophysical Research: Atmospheres, 124, 7393–7412. https://doi.org/10.1029/2019JD030421

Inness, A., Aben, I., Ades, M., Borsdorff, T., Flemming, J., Jones, L., Landgraf, J., Langerock, B., Nedelec, P., Parrington, M., and Ribas, R.: Assimilation of S5P/TROPOMI carbon monoxide data with the global CAMS near-real-time system, Atmos. Chem. Phys., 22, 14355–14376, https://doi.org/10.5194/acp-22-14355-2022, 2022.

Jiang, Z., Jones, D. B. A., Worden, J., Worden, H. M., Henze, D. K., and Wang, Y. X.: Regional data assimilation of multi-spectral MOPITT observations of CO over North America, Atmos. Chem. Phys., 15, 6801–6814, https://doi.org/10.5194/acp-15-6801-2015, 2015.

Qu, Z., Henze, D. K., Worden, H. M., Jiang, Z., Gaubert, B., Theys, N., & Wang, W. (2022). Sector-based top-down estimates of NOx, SO2, and CO emissions in East Asia. Geophysical Research Letters, 49, e2021GL096009. https://doi. org/10.1029/2021GL096009