Referee #1

Overall comments:

The paper is packed with useful information, and it seems the authors have invested tremendous effort. While some details are missing, making certain parts challenging to follow, the prior simulation itself is commendable. I believe this paper is suitable for publication in EGUsphere once the authors address the comments.

I would like to suggest that the authors dedicate some time to refining the sentence structures for a smoother reading experience. Additionally, as mentioned below, I recommend relocating certain paragraphs from the Results section to the Methodology section or the supplementary materials, as the two sections appear to be mixed.

Furthermore, I have a specific request regarding Figure S5: It would be beneficial to include a scatter plot comparison that depicts "local" enhancements by subtracting the background. I am curious about how the background estimation was carried out and affects the scatter plot comparison. Additionally, I am curious to know whether the inversion was performed after the background subtraction.

I hope that the authors will thoroughly address the detailed comments below.

We thank you for your careful reading of our paper and for providing your valuable comments. We have refined the sentence structures, clarified the treatment of background, and revised our manuscript according to your comments. Please see our specific responses below.

Regarding Figure S5 (Figure 9 in the revised manuscript), a scatter plot between the simulated XCO_2 enhancements (ΔXCO_2) and the observed ΔXCO_2 after subtracting the background is shown in Figure R1a, together with a scatter plot for XCO_2 (Figure R1b, same as Figure 9 in the main text). The background is defined as the Tsukuba TCCON XCO_2 data minus the simulated Tsukuba XCO_2 enhancements.

In the initial manuscript, the differences in XCO_2 measurements between the urban and Tsukuba sites were considered observational data and the corresponding XCO_2 differences were simulated (as represented by Equation (7) in the initial manuscript). However, in Figures 8 and 9 (in the initial manuscript), the observed and simulated "XCO_2" were shown by transforming the equation. This may have caused some confusion. In the revised manuscript, the XCO_2 measurements at the urban sites have been considered observational data (as represented by Equation (2) in the revised manuscript) to be consistent with what those figures show. We note that this change is mathematically identical in the inverse analysis (just movement of a few terms in the equation) and does not affect the inversion

results at all.

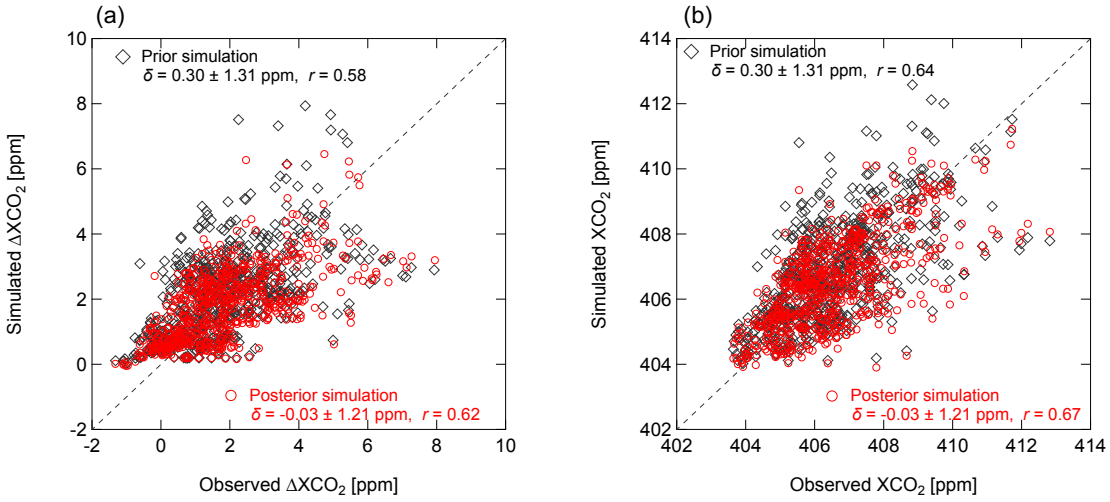

Figure R1. (a) Scatter plot between the XCO_2 enhancements (ΔXCO_2) simulated from the prior (black) and posterior (red) emission fluxes and the observed ΔXCO_2. (b) Scatter plot between the simulated and observed XCO_2 values. The mean difference between the simulations and observations (simulation minus observation) with the standard deviation ($\pm 1\sigma$) is denoted as δ, and r is the correlation coefficient.

Detailed comments:

L28-29: The following statement is subjective because it depends on the a priori assumption. For example, if the prior is assigned with large uncertainty, the percentage of uncertainty reduction in the posterior will be larger, e.g., even larger than a factor of 3. So, the author needs to clarify this sentence: "In addition, the inverse analysis reduced the uncertainty in total CO2 emissions in the TMA by a factor of ~2."

We have revised the last two sentences of the abstract as follows: "The prior and posterior total CO_2 emissions in the TMA are 1.026 ± 0.116 and 1.037 ± 0.054 Mt-CO_2 d^{-1} at the 95% confidence level, respectively. The posterior total CO_2 emissions agreed with emission inventories within the posterior uncertainty, demonstrating that the EM27/SUN spectrometer data can constrain urban-scale monthly CO_2 emissions."

L30 – 31: Instead of the current conclusion, I recommend the authors use a statement, e.g., the posterior emissions are X+/-Y times the prior emissions (at the 95% CI). This way, the readers get more information, e.g., how tightly the measurements constrain the emissions.

We have revised the sentence as mentioned above.

L76 – 78: I strongly recommend that the authors add a couple of sentences describing this work's unique contribution in addition to the previous work for the TMA.

We have added the following sentences: "We constructed CO_2 emission inventories with more accurate information on both the locations and emissions of large point sources. Anthropogenic CO_2 emissions from area sources and large point sources were estimated separately using this inventory as the prior. In addition, the area source emission estimates with higher spatial resolution allow verification of the emissions reported by each administrative division."

L90: I would recommend that the authors add a map of Japan as an inset to show the relative location of the study area. The elevation map is good, but it is hard for those unfamiliar with the area to make sense of the study area relative to the entire country.

We have added a map of Japan to Figure 1. In addition, we have added the following sentence to the caption of Figure 1: "The upper right figure shows the location of the study area relative to Japan as a whole."

L142: Was the footprint normalized? The unit for footprint should be "ppm/flux" or, specifically, "ppm /(umol/m2/s)"? It seems that clarification is needed.

We have corrected the unit for the footprint to "ppm/(μmol/m^2/s)".

L258: Are the authors referring to the GVF data from VIIRS? It would be useful to add the exact VIIRS product name.

We have revised the sentence as follows: "we spatially downscaled the hourly VISITc NEE data using GVF data from the Visible Infrared Imaging Radiometer Suite (VIIRS) sensor onboard the Suomi National Polar-orbiting Partnership satellite (VIIRS Global Green Vegetation Fraction). The GVF data are produced with an approximately 4-km spatial resolution on a daily basis from the past 7 days of VIIRS observations (Ding and Zhu, 2018)."

L263: As written, it is not clear. Was the ratio of the interpolated GVF versus the original GVF applied to the NEE data at 1 km? Or something else?

We have revised the sentence as follows: "The ratio of the original GVF to the interpolated GVF was multiplied by the interpolated NEE data to produce the downscaled NEE data (Fig. 4c)."

Section 3.3: Overall, I think the authors did a good job of making the prior fluxes more accurate!

Thank you for your positive feedback.

L264 – 265: Suggestion for rewriting to improve clarity: “The downscaling process was conducted in a manner that ensured all original sums of the NEE data from the TMA were preserved following the downscaling.”

We have made this revision.

L268: “forward” seems wrong. First, WRF-STILT is not a physical “forward” model in this setting, although it can be used for forward simulation. Second, this is a linear or nonlinear model, statistically speaking.

As you pointed out, the WRF-STILT simulations were performed in “backward” mode to trace back the origin of the observed airmasses. However, in the present study, the terms “forward simulation” and “forward model” mean the process for calculating XCO_2 values from surface fluxes via atmospheric transport as opposed to an “inverse analysis” or “inverse model” that infers surface fluxes from XCO_2 values. In the revised manuscript, we have explicitly written the forward model (Equations (2) to (4)). In addition, we note that the term “forward simulation” is also used in other similar studies on top-down emission estimates (e.g., Cusworth et al., 2020; Huang et al., 2019; Maksyutov et al., 2021; Pisso et al., 2019).

Cusworth, D. H. et al.: Synthesis of methane observations across scales: Strategies for deploying a multitiered observing network, Geophys. Res. Lett., 47, e2020GL087869, https://doi.org/10.1029/2020GL087869, 2020.

Maksyutov, S. et al.: Technical note: A high-resolution inverse modelling technique for estimating surface CO_2 fluxes based on the NIES-TM–FLEXPART coupled transport model and its adjoint, Atmos. Chem. Phys., 21, 1245–1266, https://doi.org/10.5194/acp-21-1245-2021, 2021.

Huang, Y. et al.: Seasonally resolved excess urban methane emissions from the Baltimore/Washington, DC Metropolitan region, Environ. Sci. Technol., 53, 11285–11293, https://doi.org/10.1021/acs.est.9b02782, 2019.

Pisso, I. et al.: Assessing Lagrangian inverse modelling of urban anthropogenic CO_2 fluxes using in situ aircraft and ground-based measurements in the Tokyo area, Carbon Balance Manage., 14, 6, https://doi.org/10.1186/s13021-019-0118-8, 2019.

L269: I suggest that the authors present H(x, b) more explicitly, e.g., by writing out the Jacobian matrix and x together. That way, the reader can understand the nonpoint and point source inversion more easily. This is related to Eq. (3), where “K” is introduced. Showing how “K” is associated with “b” should be useful (unless it is presented in the supplemental; I don’t see it).

We have added the following sentences: “The forward model simulates XCO_2 values at the urban sites (Saitama or Sodegaura) as follows:

$$\boldsymbol{H}(\boldsymbol{x},\boldsymbol{b}) = \Delta\mathrm{XCO_2}^{\mathrm{urban}}_{\mathrm{STILT}}(\boldsymbol{x},\boldsymbol{b}) + \mathrm{XCO_2}^{\mathrm{BG}}(\boldsymbol{x},\boldsymbol{b}), \quad (2)$$

where $\Delta\mathrm{XCO_2}^{\mathrm{urban}}_{\mathrm{STILT}}$ is the XCO_2 enhancement at the urban sites simulated by the pressure-weighted footprint and the surface fluxes, and $\mathrm{XCO_2}^{\mathrm{BG}}$ is the background value. We calculated the ΔXCO_2 values as follows:

$$\Delta\mathrm{XCO_2}^{\mathrm{urban}}_{\mathrm{STILT}}(\boldsymbol{x},\boldsymbol{b}) = \boldsymbol{F}^{\mathrm{urban}}_{\mathrm{aggr}}\,\boldsymbol{x}_{\mathrm{area}} + \boldsymbol{F}^{\mathrm{urban}}_{\mathrm{fine}}\,\boldsymbol{b}_{\mathrm{point}}\,x_{\mathrm{point}} + \boldsymbol{F}^{\mathrm{urban}}_{\mathrm{fine}}\,\boldsymbol{b}_{\mathrm{bio}}, \quad (3)$$

where $\boldsymbol{F}_{\mathrm{fine}}$ and $\boldsymbol{F}_{\mathrm{aggr}}$ are the original and the spatially aggregated footprints, respectively. $\boldsymbol{x}_{\mathrm{area}}$ and x_{point} are the emission flux vector for area sources and the (scalar) scaling factor for large point sources, respectively. $\boldsymbol{b}_{\mathrm{point}}$ and $\boldsymbol{b}_{\mathrm{bio}}$ are the emission flux vectors for large point sources and biogenic sources, respectively."

"We therefore obtained the background XCO_2 values by subtracting the simulated ΔXCO_2 values at the Tsukuba site ($\Delta\mathrm{XCO_2}^{\mathrm{Tsukuba}}_{\mathrm{STILT}}$) from the Tsukuba TCCON XCO_2 values ($\mathrm{XCO_2}^{\mathrm{Tsukuba}}_{\mathrm{TCCON}}$):

$$\begin{aligned}\mathrm{XCO_2}^{\mathrm{BG}}(\boldsymbol{x},\boldsymbol{b}) &= \mathrm{XCO_2}^{\mathrm{Tsukuba}}_{\mathrm{TCCON}} - \Delta\mathrm{XCO_2}^{\mathrm{Tsukuba}}_{\mathrm{STILT}}(\boldsymbol{x},\boldsymbol{b}) \\ &= \mathrm{XCO_2}^{\mathrm{Tsukuba}}_{\mathrm{TCCON}} - \left(\boldsymbol{F}^{\mathrm{Tsukuba}}_{\mathrm{aggr}}\,\boldsymbol{x}_{\mathrm{area}} + \boldsymbol{F}^{\mathrm{Tsukuba}}_{\mathrm{fine}}\,\boldsymbol{b}_{\mathrm{point}}\,x_{\mathrm{point}} + \boldsymbol{F}^{\mathrm{Tsukuba}}_{\mathrm{fine}}\,\boldsymbol{b}_{\mathrm{bio}}\right), \quad (4)"\end{aligned}$$

Also, it is not clear at which temporal resolution the authors solve for "x." Are you solving for sub-daily emissions for each pixel? Yes, is it also solved for each pixel as well? If so, how the "b" matrix is constructed? I am asking this question because the authors use hourly emissions, at least for NEE and anthropogenic. Then the "b" matrix should be extensive. As it is written, many things are not clear.

The state vector **x** was optimized as a single average during the entire campaign period. The temporal variation of anthropogenic emissions (weekly and diurnal correction factors from the TIMES model) was taken into account in summing the hourly footprints over the STILT run time. On the other hand, **x** consists of only "average" area source emission fluxes from ODIAC for each pixel, and such a single set of fluxes were optimized. The hourly biogenic fluxes are all included in $\boldsymbol{b}$.

We have revised the two descriptions on the application of the TIMES model as follows:

(1) "The hourly footprints calculated over the STILT run time (24 h) at a given time were weighted by temporal correction factors of CO_2 emissions (described in Sect. 3.3) and aggregated in each grid cell." (Section 3.1)

(2) "Because we applied weekly and diurnal correction factors from the TIMES model to the hourly footprints in summing them over the STILT run time, we optimized one static emission distribution during the campaign period, assuming that the temporal variation of the emissions followed the TIMES model." (Section 3.5)

In addition, we have added the following sentence in Section 3.5: "Similarly, a single average scaling factor for the large point sources was optimized from the data over the entire campaign period."

L270: I would not recommend using "state" in the fixed quantity as in the sentence "b is the fixed state vector"; "State" is typically suitable for parameters (please change accordingly if "state" was used for

"b." in other places)

We have revised the sentence as follows: "$\boldsymbol{b}$ is the vector consisting of fixed physical quantities."

L272-273: Based on "the state vector x includes spatially resolved nonpoint source emissions and a scaling factor of the large point source emissions," the reader may be confused about how the inversion was done. Are you solving for the "flux" directly for the nonpoint source but the "scaling factor" for the point source? If it is the case, it is ok. But it needs clarification. Maybe, the authors did this way, but it is not clear from the writing.

We did it the way you suggest. To clarify, we have revised the sentence as follows: "the state vector x includes spatially resolved fluxes for the area source emissions and a scaling factor for the large point source emissions."

L288: How is "the Levenberg–Marquart parameter" estimated? Or prescribed?

We have revised the sentence as follows: "γ is the Levenberg–Marquart parameter fixed at 10 (Chen et al., 2022)."

Chen, Z., Jacob, D. J., Nesser, H., Sulprizio, M. P., Lorente, A., Varon, D. J., Lu, X., Shen, L., Qu, Z., Penn, E., and Yu, X.: Methane emissions from China: a high-resolution inversion of TROPOMI satellite observations, Atmos. Chem. Phys., 22, 10809–10826, https://doi.org/10.5194/acp-22-10809-2022, 2022.

L314-315: I am curious how the authors matched the vertical profiles between CarbonTracker (CT) and EM27 to get the background for EM27. A weighting scheme was used? Ideally, the particle trajectory for each receptor (at different locations and vertical levels) of EM27 should be computed and then averaged using a kernel (or a set of weights, likely based on pressure distributions) compatible with EM27. To sample values from CT (using particle trajectories), the same method should be used to match the vertical profile between the two. I wonder if the authors did that or something else.

We did not use the CO_2 profile product of CarbonTracker, but rather the XCO_2 product (CT2019B.xCO2), so we did not perform any weighing by the column averaging kernel. We have added the product name (CT2019B.xCO2), and these sentences have been moved to Section 3.4.

L310: By "XCO2 differences", do the authors mean "enhancement" above the background? The phrase "XCO2 differences (XCO2Diff) from daily background values" needs to be revised for clarification.

We have revised the sentence as follows: "To characterize the diurnal variation in XCO_2 at each observation site, we examined the diurnal variation in XCO_2 enhancements (XCO_2^{Enh}) above the daily

XCO_2 baseline." We note that the "5 percentile value of the Tsukuba TCCON measurements" has been referred to as the "baseline" in the revised manuscript, not to be confused with the "background" defined and used in the simulations and inverse analyses.

L313: How did the author account for the background uncertainty based on this "5 percentile" assumption?

We have added the following sentences: "When the 2 (10) percentile values of the Tsukuba TCCON measurements were used as the daily XCO_2 baseline, the maximum XCO_2^{Enh} values were 9.6 (9.4) ppm at Saitama and 9.5 (8.9) ppm at Sodegaura. These changes had little effect on the standard deviations of the mean XCO_2^{Enh} values and the pattern of the diurnal variation."

L318: Please add "diurnal" so that it reads "The average diurnal XCO2Diff." By the way, I think "ΔXCO2" is more informative to represent the local signal (I find both are used). Some people use "XCO2" to describe the local mixing ratio (after subtracting background). I suggest the authors review the notation a bit more to avoid confusion. In fact, what is the difference between "XCO2Diff" and "ΔXCO2" in Line 277? I may have misunderstood, but further clarification would help. Thank you.

We have revised the sentence as follows: "The average diurnal XCO_2^{Enh} value per 15-min bin was calculated for each site using the entire field campaign dataset (Fig. 6)."

"ΔXCO_2" is used only to represent the simulated local enhancements. "XCO_2^{Enh} (XCO_2^{diff} in the initial manuscript)" represents the observed XCO_2 enhancements above the daily 5-percentile value of the Tsukuba TCCON measurement. We have added the following sentence: "We note that the XCO_2^{Enh} values were calculated using only the observed XCO_2 values, whereas the ΔXCO_2 values represent the simulations of local XCO_2 enhancement."

L321: I am a bit confused to see that there is a moderate-level effect of biogenic fluxes while the authors said, "the biogenic flux was allocated to the state vector b" in Line 276; it was assumed negligible there. Any clarification?

The biogenic effect due to photosynthesis is not so large and the value is expected to be (relatively) similar at the different sites. This is seen both for the observations (Figure 6) and simulations (Figure S5). However, the biogenic fluxes are not small enough to be negligible, so they are included in the forward calculation of ΔXCO_2 (as the vector $\boldsymbol{b}$). We have revised the sentences in Line 276 as follows: "the biogenic flux was allocated to the fixed vector $\boldsymbol{b}$. Note that the contribution of biogenic flux to the simulated ΔXCO_2 was small compared to that of anthropogenic flux and the differences among ΔXCO_2 calculated from four different biogenic flux products are also small (Sect. 4.2)."

L323: It is unclear what the authors mean by "the high early morning values at Saitama may reflect

an airmass-dependent bias."

We have added the following sentences: "The airmass-dependent variation in XCO_2 is caused by the effects of inaccurate spectroscopic parameters on the retrievals, which vary with the depth of the absorption lines (i.e., airmass) (Wunch et al., 2015). Although this effect is corrected in the GGG2014 software, the error may remain for a large airmass."

L330-334: The sentence sound awkward. Please revise.

We have revised the sentences as follows: "the XCO_2 enhancement (ΔXCO_2) was calculated from the column-averaged footprint and the surface fluxes from area sources, large point sources, and biological activity. The ΔXCO_2 values resulting from the large point source emissions and biogenic fluxes were calculated from the original footprints with a spatial resolution of approximately 1 km × 1 km (0.0083° × 0.0083°). For area source emissions, however, we re-gridded the original footprints to a spatial resolution of 0.025° × 0.025° to degrade the spatial resolution for the inverse analysis. First, the area source emissions were summed for each 0.025° × 0.025° grid cell. Then, individual footprints for the 0.025° × 0.025° grid were derived by dividing the sum of the nine XCO_2 contributions for the 0.0083° × 0.0083° grid by the emissions for the 0.025° × 0.025° grid."

L330-345: I would recommend that the authors move this particular paragraph to the Methodology section or possibly to the supplementary materials. As it stands, the Results section seems a bit extensive, and this adjustment could help with maintaining focus and flow.

We have moved L331-342 to the Methodology section.

L350-355: Here, the authors describe the background again, which I thought was done in Section 4.1. Given that both mention "Tsukuba," I understood that site measurements were used as the background common to the other sites. What's surprising to me is that the authors subtract the simulations at "Tsukuba" from the "Tsukuba" measurements to remove the local enhancements for the background site. It is possible, but it adds more uncertainty to the background because the simulated quantity itself is uncertain. Typically, using the particle trajectories from the STILT model, we would sample 4-D background data (over the ocean) simulated from a global model. The method used here is somewhat convenient but adds uncertainty.

We believe that a method that takes the background from measurements away from the emissions is as typical as the method that combines the trajectory with the global model. In the present study, the Tsukuba measurements were considered background due to their distance from the main emission sources. However, as demonstrated by Babenhauserheide et al. (2020), the Tsukuba measurements can sometimes be impacted by emissions in the central area of the TMA. Therefore, the simulated enhancements (from anthropogenic and biogenic emissions) at Tsukuba were subtracted from the

Tsukuba measurements. As you pointed out, the simulated enhancements at Tsukuba added uncertainty to the background. However, optimizing the anthropogenic emission fluxes in the inversion analyses would reduce the uncertainties.

Also, I suggest this paragraph be merged this the relevant paragraph in Section 4.1. Otherwise, the manuscript gets longer, and the reader is distracted/confused.

We have refined the structure; this paragraph and the first paragraph of Section 4.3 have been combined and moved to Section 3.4.

L358-359: I suggest the authors add a scatter plot for predicted versus measured, corresponding to Figure 8, only for the 15-min average. I think the figures are already many, but Figure 2 and Figure 4 (maybe more) can be moved to the supplemental.

Such a scatter plot was shown as Figure S5 in the supplemental material. In the revised manuscript, we have moved Figure S5 to the main text (Figure 8). In addition, Figures 2 and 4 in the initial manuscript have been moved to the supplemental section.

L360: By "the sum of the WRF–STILT ΔXCO2 value every 15 min at each site and the background XCO2 value", I assume "ΔXCO2" is the local enhancement. It needs to clarify between "ΔXCO2" and "XCO2diff."

"ΔXCO_2" is used only to represent the simulated local enhancements. "XCO_2^{Enh} (XCO_2^{diff} in the initial manuscript)" represents the observed XCO_2 enhancements above the daily 5-percentile value of the Tsukuba TCCON site. In the revised manuscript, these have been clarified. We note, however, that this sentence itself has been removed in the refinement of the sentence structure.

L361: "forward"? STILT back trajectories were used.

Although the STILT model was used in "backward" mode to calculate footprints, "forward" simulation means the process to calculate XCO_2 values from the footprints and the surface CO_2 fluxes.

We have added a description of forward simulation in Section 4.2: "We compared the XCO_2 data for the forward simulations, which correspond to the XCO_2 simulations from the footprints and the surface CO_2 fluxes based on Eqs. (2) to (4), with the EM27/SUN observations at Saitama and Sodegaura (Figs. 7 and 8)."

L363: What kind of point source? Is it identifiable, e.g., a power plant?

As shown in Figure 3 in the revised manuscript, there are several point sources near the Sodegaura site, including steel plants as well as power plants, so it would not be possible to identify the source. We have revised the sentence as follows: "which were likely caused by the plume from large point

sources such as the power plants and steel plants located near the Sodegaura site."

L369: I don't necessarily agree with the statement: "Therefore, we attribute this large model–observation discrepancy to errors in the WRF-STILT model rather than to the emission data." First, I don't expect ERA5 to perform better than WRF because it is a much coarse resolution model product (I also see that in this work's Figure S3). From my experience, it can be much worse than WRF, depending on the region. I would say that the authors only considered a limited set of meteorology, not exploring a broader set of meteorological data. So, it is possible that the limited meteorology didn't capture the temporal variation. However, as the author said, it is still possible that the short-term local source not included in the prior fluxes is associated with this discrepancy between measurements and predictions. To summarize, although it is likely that the transport source is the primary source of the discrepancy, I don't see evidence for the strong statement above.

Since the simulations using the prior emission fluxes were able to reproduce the diurnal variation well, except for 3 March 2016, we thought that the modeling error on specific meteorological conditions might be the dominant cause of the discrepancy on that day. However, as you pointed out, a short-term local source not included in the prior fluxes could be the cause of the mismatch between the prior simulations and the observations on 3 March 2016.

We have revised the sentences as follows: "However, we cannot rule out the possibility that short-term local sources not included in the prior fluxes may cause the discrepancy between the prior simulations and the observations. Therefore, we attribute this large model–observation discrepancy to errors in the WRF-STILT model, or to the short-term local sources not included in the prior fluxes, or both."

L380: Equation 7 is confusing. What is the purpose of this equation? If this should be included, it should be presented in the section (e.g., 4.1) where the background is described. Based on the earlier description, wasn't "XCOBG " derived from XCOTsukuba_TCCON? As pointed out, this whole paragraph should be in the Method section, not the Result section.

In the revised manuscript, Equation (7) has been removed, and new equations that provide a detailed description of the forward model (Equations (2) to (4)) have been added. These equations make it clear that background is defined as the difference between the Tsukuba TCCON measurements and the Tsukuba STILT simulations (i.e., Tsukuba TCCON minus Tsukuba STILT). In addition, this paragraph has been moved to Section 3.4.

L387-445: This should be included in the Method section for the abovementioned reason. There is no meaningful result described or discussed. They would agree with me if the authors read similar inverse modeling papers.

We have moved the description on the construction of the prior error covariance matrix and

measurement error covariance matrix to the Methodology section (i.e., L387-412 in the initial manuscript). Because the remaining part (L413-446 in the initial manuscript) discusses the uncertainties in our model–observation system based on the simulation results, it has been moved to Section 4.2.

L417-418: This work differs from the system in Turner et al., where they have a dense measurement network. I cannot offer any temporal correlation length scale for this work, but I am not quite sure about adopting the 1-hr length scale.

Since Turner et al. (2020) have dense measurement data, a spatial correlation length scale and a temporal correlation length scale are imposed on the off-diagonal components of the measurement error covariance matrix. The effect of the dense measurement data is taken into account by including the spatial correlation length.

Meanwhile, as you note, the temporal correlation length is uncertain. We have added inversion analyses using different temporal correlation lengths to the sensitivity analysis (in Section 4.3).

L448: Which period does Figure 12a represent? Is it the average of the hourly posterior fluxes during the study period?

Figure 12a (Figure 11a in the revised manuscript) represents the single average emission fluxes optimized using all data during the campaign period. As described above, this has been clarified in Section 3.5 in the revised manuscript.

L470-471: Related to Equation 1, how many scaling factors were used/solved? Is this value of "0.856" just the average of many scaling factors? A simple average of many scaling factors would not work, though.

The scaling factor and the spatially resolved anthropogenic emission fluxes were each solved as single averages during the campaign period. We have added the following sentence to Section 3.5: "Similarly, a single average scaling factor for the large point sources was optimized from the data over the entire campaign period."

L512-513: Can the author offer further discussion on the difference between this study and Pisso et al.?

We have added the following sentences: "Pisso et al. (2019) and this study use comparable Lagrangian transport models to calculate atmospheric transport; however, there are several differences, including the type of observational data (in-situ vs. column), the prior emission fluxes (EDGAR vs. ODIAC), the meteorological fields for driving the transport model (ERA-Interim vs. WRF based on GPV-MSM), and the spatial resolution of emission estimates (20 km × 20 km vs. 3 km × 3 km). Our sensitivity

analysis shows that changing the prior fluxes, meteorological field, and emission estimation resolution to roughly match Pisso et al. (2019) did not produce a result substantially different from the emission estimation result of the reference inversion. We thus concluded that the improved accuracy of emission estimates in our study may be due to the use of columns as observational data. Column data are less susceptible to the effect of PBL height changes that are difficult to simulate in transport models and have information on a larger area of emissions due to the difference in wind direction at each altitude."

L535: With "forward simulation," as pointed out above, how is footprint-based (backward is assumed unless explicitly stated) inversion possible?

In the revised manuscript (Section 4.2), we have added an explanation that the forward simulations correspond to calculating the XCO_2 values from the footprints and the surface CO_2 fluxes using Equations (2) to (4).

L540: The mismatch between predictions and observations could be due to local sources not included in the prior, not necessarily due to transport error. Do you have evidence that there was a clear transport error? For CH4, EDGAR is generally not as good as regional inventories. I see both CO2 and CH4 measurements are significantly higher later in the afternoon (from Figure S3). It seems that the CO2 and CH4 sources are correlated. It may be the transport model didn't capture the afternoon winds. Any evidence for that?

We have no clear evidence to suggest that there was an error in the transport (e.g., wind speed and direction do not substantially differ from the measurements; simulated PBL heights do not take extreme values). On the other hand, as you pointed out, the mismatch between the prior simulations and the observation on 3 March 2016 may be attributable to a short-term local source not included in the prior fluxes.

We have revised the sentence as follows: "As described in Sect. 4.2, in some cases, the simulations failed to reproduce the diurnal variation and to capture the plume from nearby large point sources, possibly because of the transport modeling error or the short-term local sources not included in the prior fluxes (Figs. 8d and S6)."

Referee #2

The authors developed an inversion scheme to infer the anthropogenic carbon dioxide emissions in the Tokyo Metropolitans Area from observations of three ground based remote sensing sites. One of which is a TCCON site.

The authors obtained the background by subtracting the simulated CO2 enhancement (from the footprint and surface flux) from the observed XCO2 values at the Tsukuba COCCON site for the forward modeling. To assess the biosphere, they spatially downscaled the terrestrial biospheric model VISITc to simulate the biogenic influence and found the influence to the enhancements to be small. The authors infer the meteorological surface interaction using WRF-STILT with a spatial resolution of 1km. The Bayesian inversion scheme inverts for spatially resolved emissions, separated into point and area-sources for the more than two months period with a total of approximately 6.5 degrees of freedom. The authors also compared 12 different model configurations.

The authors report total carbon dioxide emissions for the study area and compare it to several literature reports and find good agreements within the reported uncertainties. The scientific value is to be rated as high, since emission estimates from observations still remain a tough challenge and needed to confirm or refine reported emission inventories. The paper is written in a clear, structured style. However, some details need improvements.

We thank you for your careful reading of our paper and for providing many valuable comments. We have added descriptions related to the potential weaknesses that you raise and revised our manuscript according to your comments. Please see our specific responses below.

Potential weaknesses are:

1. The authors conduct inversion in log-space, and therefore negative emissions are suppressed, which is not very realistic. The biogenic model needs to be perfect, so that we can be sure that there are no "negative emissions".

This study does not optimize total (anthropogenic + biogenic) fluxes, but only anthropogenic fluxes. Because the magnitude of the biogenic fluxes (negative fluxes during the daytime) in the Tokyo Metropolitan Area (TMA) in February and March is more than an order of magnitude smaller than the anthropogenic fluxes and their differences among four models are small (with a standard deviation of 0.09 ppm), the biogenic fluxes were fixed at the prior values. Therefore, it is reasonable to constrain the anthropogenic fluxes (nonpoint or area sources) to positive values by the inversions in log-space. In the revised manuscript, we have made it clear that only "anthropogenic" emissions are optimized.

2. DOFS of 6.49 implicates that solution tend to stick to the a-priori, given that the dimension of the state vector is rather large (m = 1921 or 481 or 121).

As you suggested below, we have investigated how the degrees of freedom for signal (DOFS) change when the prior uncertainties are increased by a factor ~1.5 and 2. Although the DOFS increase with the prior uncertainty, we found that their changes are not very large (please see our response below). However, a DOFS of ~6.5 would be useful for evaluating emissions from administrative divisions. In the present study, the focus was on emissions for each administrative division rather than smaller-scale individual emissions, and we compared the estimated emissions aggregated with the administrative boundaries with the reported administrative emissions.

3. The authors assumed that all the sites have the same background air. It is not always true, when considering the transport time that the air needed to travel from upwind to downwind especially when the distance between the sites are big (~ 60 km).

The background values used in the simulation and the inverse analysis are specified as the XCO_2 measurements at Tsukuba minus the STILT-calculated XCO_2 enhancements (ΔXCO_2) at Tsukuba. These background values correspond to the concentrations at the boundary of the TMA defined in this study, and we think it is appropriate to consider the background to be common to the observation sites within the relatively small TMA. The XCO_2 values for urban sites other than Tsukuba are represented as the sum of the background and the ΔXCO_2 calculated in consideration of fluxes and atmospheric transport within the TMA.

4. The definition of background is confusing. The authors have two definitions of background in the paper, i.e. 5 percentile value of TCCON station at Tsukuba and observed XCO2 from Tsukuba COCCON site subtracted with simulated CO2 enhancement.

In the revised manuscript, the 5-percentile value of the Tsukuba TCCON site has been referred to as the “baseline”. The “background” is now used only in the simulations. We note that the Tsukuba COCCON data were not used to estimate emissions but used only to correct XCO_2 values observed by the other spectrometers.

I would appreciate if the authors could comment on their thoughts on the potential weaknesses and/or discuss it in the paper, before the acceptance.

We have added the discussion regarding the limitations and possible improvements from both the measurement and simulation sides in Section 5 (L601-605 and L579-583, respectively).

Briefly stated, from the measurement side, one limitation is the number of measurements. More instruments and longer time series would probably increase our sensitivity and thus the DOFS. More instrument locations would also help to constrain the background. Another limitation from the

simulation side is the difficulty of accurately modeling the wind fields. As we saw for 3 March 2016, we had mismatches possibly due to imperfect wind fields. To better constrain wind fields and PBL, additional wind lidar observations would be useful.

Detailed comments:

L 15: Suggestion: "We conducted ..." --> "In order to infer a top down emission estimate, we conducted..."

We have made this revision.

L17: I thought that you deployed 3 EM27SUN spectrometers, please clarify.

As described in Section 2, the SN63 EM27/SUN arrived in Tsukuba in the middle of the campaign, and sunlight measurements were not performed during the entire campaign period (i.e., only from March to April 2016). To avoid any misunderstanding that the three EM27/SUNs were used for emission estimates, we would like to keep this description here. For clarification, we have added the following sentence in Section 3.4: “In the following simulations and inverse analyses, only the TCCON data were used as the measurement data at Tsukuba, since the SN63 EM27/SUN measurements started in the middle of the campaign (as described in Sect. 2).”

L 22: "nonpoint source" --> I would suggest the term "area source" (29 occurrences)

We have made these revisions.

L 26: "emission fluxes at > 3km" To my understanding, the WRF-STILT resolution is 1km, please clarify.

We have added the following description in Section 3.4: “For area source emissions, however, we re-gridded the original footprints to a spatial resolution of 0.025° × 0.025° to degrade the spatial resolution for the inverse analysis. First, the area source emissions were summed for each 0.025° × 0.025° grid cell. Then, individual footprints for the 0.025° × 0.025° grid were derived by dividing the sum of the nine XCO_2 contributions for the 0.0083° × 0.0083° grid by the emissions for the 0.025° × 0.025° grid.”

L 31: Please add your final emission number for the study area, or at least the scaling factor with the according uncertainty to the abstract and if feasible, compare it to the literature references.

We have revised the last two sentences of the abstract as follows: “The prior and posterior total CO_2 emissions in the TMA are 1.026 ± 0.116 and 1.037 ± 0.054 Mt-CO_2 d^{-1} at the 95% confidence level, respectively. The posterior total CO_2 emissions agreed with emission inventories within the posterior uncertainty, demonstrating that the EM27/SUN spectrometer data can constrain urban-scale monthly

CO_2 emissions."

L 87: Suggestion: "when the daily sunshine duration in this region is high" --> "during the high-insolation period" for clarity and specificity.

We have revised the sentence as follows: "when the proportion of clear days is high"

L 93: "city center" --> "city-center"

We have made this revision.

L101: "ASL" --> "a.s.l." (standard abbreviation, multiple occurrences)

We have made this revision.

L107: " and is now continuously operated " --> " and has since been continuously operated "

We have made this revision.

L116: "interval of approximately 1 min" --> "interval of about 1 minute"

We have made this revision.

L130: What is the integration time for determining sigma? If it is 1 min, it might be useful to also report the 15 min values as you did for comparing the observations with the forward simulations.

An integration time of 15 min was used. We have added the following sentence: "Each of the EM27/SUN data points was averaged per 15-min bin."

L132: you scale the TCCON to EM27, would it not make more sense to scale EM27 to TCCON, since TCCON is considered as standard.

It is certainly common to scale EM27/SUN data to TCCON data. However, the Tsukuba TCCON XCO_2 data have a slightly larger scatter than the other EM27/SUN data used in this study, and this made the variation in the TCCON XCO_2 data at a high solar zenith angle somewhat ambiguous. To derive an airmass-dependent correction factor (ADCF) for the SN44 EM27/SUN, we used the SN63 EM27/SUN data as the reference, which were validated using co-located aircraft measurements (Ohyama et al., 2020).

We note that, in analyses where measurements at one site are used as part of the background for measurements at other sites, the differences between them (enhancements above the background) rather than the absolute values of the concentration are particularly important. Which instrument is used as the reference has little effect on the emission estimates. In fact, in the case where the SN44 EM27/SUN XCO_2 data corrected for the airmass dependence and the SN38 EM27/SUN XCO_2 data

were scaled to the original TCCON data, the relative change in the total TMA CO_2 emissions is less than 0.1%.

Ohyama, H., Morino, I., Velazco, V. A., Klausner, T., Bagtasa, G., Kiel, M., Frey, M., Hori, A., Uchino, O., Matsunaga, T., Deutscher, N. M., DiGangi, J. P., Choi, Y., Diskin, G. S., Pusede, S. E., Fiehn, A., Roiger, A., Lichtenstern, M., Schlager, H., Wang, P. K., Chou, C. C.-K., Andrés-Hernández, M. D., and Burrows, J. P.: Validation of XCO_2 and XCH_4 retrieved from a portable Fourier transform spectrometer with those from in situ profiles from aircraft-borne instruments, Atmos. Meas. Tech., 13, 5149–5163, https://doi.org/10.5194/amt-13-5149-2020, 2020.

L135: Maybe mention the altitudes of the stations somewhere in the text.

The altitudes of each station are described in the first paragraph of Section 2.

L142: unit wrong → ppm/(mol/m2/s)

We have made this revision.

L145: Since you use the exact same altitudes as T.S.Jones et al., 2021 uses, you can add a citation here.

The paper by Jones et al. (2021) has been cited here.

L154: "multiplied by anthropogenic and biogenic fluxes" --> "multiplied with spatially resolved emission inventories for anthropogenic and biogenic fluxes separately"

We have made this revision.

L155ff: "The change ... over all grid cells." --> "The change ... over all grid cells and serves for the forward modeling."

We have made this revision.

L150ff: "We then aggregated the footprints in each grid over the STILT run time." It is not clear what you mean by "aggregate". If it is meant as an introduction into the following sentence I would suggest to move the line break before this sentence.

We have moved the position of the line break and revised the sentences as follows: “The hourly footprints calculated over the STILT run time (24 h) at a given time were weighted by temporal correction factors of CO_2 emissions (described in Sect. 3.3) and aggregated in each grid cell. From the summed footprints at each altitude, we then calculated the pressure-weighted column-average footprint, taking account of the column-averaging kernel of the EM27/SUN spectrometer (Rodgers and Connor, 2003; Jones et al., 2021).”

L243: please break down this long sentence into at least two shorter ones

We have revised the sentence as follows: “Specifically, hourly net ecosystem exchange (NEE) data from the Vegetation Integrative SImulator for Trace gases (VISIT) model, referred to as VISITc, were adopted as the biogenic CO_2 flux data. The NEE data were combined with green vegetation fraction (GVF) data to downscale them.”

L248: "the original VISIT" --> "the initial VISIT"

We have made this revision.

L251: Gaussian T382 Grid --> please explain shortly, give reference or just state something like "operate on exactly the same grid" in order to make your point.

We have revised the sentence as follows: “The VISITc model operates on the same grid as the CFSR data (i.e., approximately 0.31° × 0.31°).”

Section 3.3 in general: Multiple sentences are very long; consider breaking them into shorter pieces for a better understanding.

We have made this revision.

L257: How you can downscale VISITc product from 0.31 * 0.31 deg. to 1 km x 1 km using 4 km resolution GVF data? I am not sure whether you have the high-resolution information necessary to achieve this goal.

The effective spatial resolution of the downscaled biogenic fluxes is about 4 km, although the biogenic flux data were generated on a 1 km x 1 km grid to be consistent with the footprints. To avoid misunderstanding, we have revised the sentence as follows: “to better characterize the spatial distribution of biogenic CO_2 fluxes, we spatially downscaled the hourly VISITc NEE data using GVF data from the Visible Infrared Imaging Radiometer Suite (VIIRS) sensor onboard the Suomi National Polar-orbiting Partnership satellite (VIIRS Global Green Vegetation Fraction). The GVF data are produced with an approximately 4-km spatial resolution on a daily basis from the past 7 days of VIIRS observations (Ding and Zhu, 2018).”

In addition, we have added the following sentence: “We note that the effective spatial resolution of the downscaled biogenic fluxes is about 4 km, although they were generated on a 1 km x 1 km grid.”

L267: "DXCO2 values measured" This statement is confusing, since DXCO2 values are derived from the forward model as described in the referenced section 3.1

In the revised manuscript, we have modified the forward model that calculates XCO_2. We have revised

the sentence as follows: "XCO_2 measurements at a given location are quantitatively related to the presumed surface CO_2 fluxes via the forward model $\boldsymbol{H}$"

L268: "H, representing atmospheric transport" --> To my understanding it is the forward model.

We have removed "representing atmospheric transport".

L273: Is it correct, that you have the logarithmic of a scaling factor (unitless) as well as an emission value (in mole/area/time) in the state vector x? Please clarify.

The scaling factor is linear, not logarithmic. We have added the following sentence: "On the other hand, the scaling factor for the large point source emissions was optimized at linear scale."

L279: Inverting in the log-space introduces a strong bias to positive emissions. Negative emissions are not necessarily non-physical, especially in case of CO2, because biospheric activity might be stronger than assumed. Did you try to invert in linear space? Negative emissions could serve as a sanity check here.

The inversion in linear space was only tried at the initial stage. In the present study, we do not estimate total (anthropogenic + biogenic) fluxes, but only anthropogenic fluxes. The negative emissions for the anthropogenic sources could cause large uncertainty in their emission estimates. In addition, in February and March in the TMA, the magnitude of the biogenic fluxes (negative fluxes during the daytime) is more than an order of magnitude smaller than the anthropogenic fluxes (Table 4 and Figure S5) and their differences among four models are small (with a standard deviation of 0.09 ppm) (Section 4.2). Therefore, it is reasonable to constrain the anthropogenic fluxes to positive values by the inversions in log-space.

In the revised manuscript, we have revised the sentence to make it clear that "anthropogenic area source" emissions are optimized as follows: "because the area source emissions from each grid cell differ by a couple of orders of magnitude, and the optimization of area source emissions at linear scale might lead to unphysical negative posterior emissions."

L331: "sources, large point sources" --> "sources, strong point sources" to separate from the spatial meaning of "large"

In the revised manuscript, we have defined "large point sources" as point sources with large emissions (the first paragraph of Section 2).

L311ff: It is not very clear in the text what XCO2^{Diff} means and how it separates from (DXCO2).

We have revised the sentence as follows: "To characterize the diurnal variation in XCO_2 at each observation site, we examined the diurnal variation in XCO_2 enhancements (XCO_2^{Enh}) above the daily

XCO_2 baseline."

Additionally, in the revised manuscript, "ΔXCO_2" is used for only representing XCO_2 enhancements calculated from the forward model.

L315: It is a bit confusing here, because you define another background (5 percentile value of the Tsukuba TCCON site) than the one you use for the forward modeling. What XCO2^{Diff} is actually used for? Just to look into temporal fluctuation?

To avoid confusion with another "background" used in the simulations, the 5-percentile value of the Tsukuba TCCON site has been referred to as the "baseline" in the revised manuscript. XCO_2^{Enh} (XCO_2^{Diff} in the initial manuscript) values were calculated using only the observed XCO_2 values to examine the temporal fluctuation at each site. These XCO_2^{Enh} confirmed that using XCO_2 measurements at Tsukuba as background in the simulations would be valid (please see also the next response).

L312: Please explain the reasons to use Tsukuba as a background site.

We have added the following sentence in Section 3.4: "We assumed that the XCO_2 values at Tsukuba approximately represent background air, as there are lower CO_2 emissions around Tsukuba (Fig. 3) and the XCO_2 values observed at Tsukuba were systematically lower than those at the other urban sites, which can be seen from the XCO_2 values in Fig. 2a."

L314ff and L355: How many days (or observations of the n=654 observations) were replaced by CarbonTracker?

We have added the following description: "For days when measurements at Tsukuba were not available (16, 17, 27, and 28 February and 23 March)"

L359: you averaged the data in 15 mins. Why is it optimal or in another word, why no drift of the sensor is integrated? You could refer to: https://acp.copernicus.org/articles/16/8479/2016/acp-16-8479-2016.pdf, section 3.1, where the optimal integration time is determined by using Allan analysis.

We have added the following sentences in Section 2: "Chen et al. (2016) derive an optimal integration time of 10 to 20 min, based on the Allan variance of two sets of EM27/SUN data from side-by-side measurements. However, they used a shorter integration time of 5 min to derive the EM27/SUN differences between upwind and downwind of local emission sources. In the present study, we found that it is difficult for the XCO_2 simulation to accurately reproduce the times at which point source plumes are observed (Sect. 4.2), and a comparison of the simulations and observations at short time intervals is not beneficial. Thus, we adopted an integration time of 15 min for the EM27/SUN data."

L370: You talked about the model-observation discrepancy, forward modeling vs. observation is mainly given by the errors in the WRF-STILT, what about the background error?

In this case, the background is represented as the Tsukuba TCCON XCO_2 data minus the ΔXCO_2 simulations at Tsukuba (i.e., $XCO_2{}^{\text{Tsukuba}}_{\text{TCCON}} - \Delta XCO_2{}^{\text{Tsukuba}}_{\text{STILT}}$). If this background value in the late afternoon became larger by ~4 ppm, the simulation would agree with the observation. Considering the uncertainty in the $XCO_2{}^{\text{Tsukuba}}_{\text{TCCON}}$ data and the magnitude of the $\Delta XCO_2{}^{\text{Tsukuba}}_{\text{STILT}}$ data (Figure S5c), we believe that the effect of the background on the model-observation discrepancy would be small. Meanwhile, as pointed out by Referee #1, short-term local sources not included in the prior fluxes could contribute to the discrepancy. Therefore, we have revised the sentence as follows: "However, we cannot rule out the possibility that short-term local sources not included in the prior fluxes may cause the discrepancy between the prior simulations and the observations. Therefore, we attribute this large model–observation discrepancy to errors in the WRF-STILT model, or to the short-term local sources not included in the prior fluxes, or both."

L445: you are looking into the model-observation mismatch for the inverse modeling framework. However, in your inversion you assume the same background for all sites. The background influence is canceled out in the forward model. Why you need to take the uncertainty of the background into account?

Indeed, Equation (7) in the initial manuscript seems to indicate that the background is canceled out. In the revised manuscript, the equation has been modified to represent how the urban XCO_2 measurements are simulated. Equations (2) to (4) in the revised manuscript make it clear that background (i.e., Tsukuba TCCON minus Tsukuba STILT) is included in the simulation.

We note that this change is mathematically identical (with just a movement of the $XCO_2{}^{\text{Tsukuba}}_{\text{TCCON}}$ term), resulting in the same inversion results.

L455: It is not exactly clear what the authors mean with "upward" and "downward"

We have revised the sentence as follows: "the emissions from the central TMA region became smaller than the prior values, and the emissions from the other regions became larger than the prior values."

L470: With 6.5 degrees of freedom the model has not enough freedom to scale the sources individually. What happens if you provide an intentionally much uncertain a-priori (e.g. Factor 2 higher).

When the prior uncertainties are increased by a factor ~1.5 and 2 (i.e., 120% and 170% of the prior emissions, respectively), the degrees of freedom for signal (DOFS) are 8.35 and 10.18, respectively. Although the DOFS increase with the prior uncertainty, they still seem insufficient to resolve the sources individually. We note that the case with 120% uncertainty is included in the sensitivity analysis. In addition, the DOFS for all sensitivity analyses have been added to Table 5.

L475ff: Table 5: Please add the degrees of freedom and the Bayesian Information Criterion (BIC) to this list. The latter is a helpful number to tell which of the models could be a better choice.

The DOFS have been added to Table 5. In addition, we calculated the BIC according to Rayner (2020) (Table R1). With coarser spatial resolution (cases #7a and #7b), the BIC becomes smaller (i.e., a better model) due to the substantial decrease in the $m \log(n)$ term of the BIC. According to this parameter, the worse the spatial resolution, the better the inverse model. We acknowledge that there are a variety of ways to optimize the grids for spatially resolved emission flux estimates, and we intend to consider this in future studies.

Rayner, P.: Data assimilation using an ensemble of models: a hierarchical approach, Atmos. Chem. Phys., 20, 3725–3737, https://doi.org/10.5194/acp-20-3725-2020, 2020.

Table R1. Bayesian information criterion (BIC) for the different meteorological data, prior emission data, prior uncertainty (σ_a), spatial correlation length of $\mathbf{S_a}$ (l_s), temporal correlation length of $\mathbf{S_\varepsilon}$ (l_t), and spatial resolution of the inversion domain (r_s).

Case	Meteorological data + prior emission data	σ_a (%)	l_s (km)	l_t (h)	r_s (°)	BIC
#0	WRF/MYJ + ODIAC	85	10	1	0.025	11932
#1	WRF/MYJ + ODIAC (LPS fixed)	85	10	1	0.025	11958
#2a	WRF/MYNN25 + ODIAC	85	10	1	0.025	11912
#2b	WRF/YSU+topo + ODIAC	85	10	1	0.025	11890
#2c	ERA5 + ODIAC*	85	10	1	0.025	11719
#3a	WRF/MYJ + ODIAC	50	10	1	0.025	11937
#3b	WRF/MYJ + ODIAC	120	10	1	0.025	11927
#4a	WRF/MYJ + ODIAC	85	5	1	0.025	11933
#4b	WRF/MYJ + ODIAC	85	20	1	0.025	11932
#5a	WRF/MYJ + ODIAC	85	10	0.5	0.025	11854
#5b	WRF/MYJ + ODIAC	85	10	2	0.025	12161
#6	WRF/MYJ + EDGAR	95	14	1	0.025	12752
#7a	WRF/MYJ + ODIAC	75	16	1	0.05	3324
#7b	WRF/MYJ + ODIAC	65	25	1	0.1	1138

*Data from Sodegaura on 23 March 2016 were excluded.

L494ff: The statement appears reasonable. However, referenced Fig. S6 does not appear to have a connection to this statement.

The reference to Figure S6 (Figure S7 in the revised manuscript) has been changed to the sentence describing EDGAR as follows: “case #5, EDGAR version 6 (0.1° × 0.1° spatial resolution) without large point source correction used as the prior estimate (Fig. S7)”

L497ff: The sentence is very long. Please reformulate.

We have revised the sentence as follows: “Although the number of grid cells with a spatial resolution of 0.05° and 0.1° was equivalent to or lower than the number of measurement data points, respectively, the total DOFS slightly decreased (to 5.84 for 0.05° and 5.05 for 0.1°). This was due to the changes in the prior uncertainty and the spatial correlation length.”

L526: I thought a third EM27/SUN is also deployed at Tsukuba site.

As described in our response above, since the solar measurements with the SN63 EM27/SUN were not used for emission estimates, we would like to keep this description here. For clarification, we have added the following sentence in Section 3.4: “Note that in the following simulations and inverse analyses, only the TCCON data were used as the measurement data at Tsukuba, since the SN63 EM27/SUN measurements started in the middle of the campaign (as described in Sect. 2).”

L553: Again here is 3km resolution mentioned. To my understanding it is 1km. If not correct please explain the reasons.

We have added the following description in Section 3.4: “For area source emissions, however, we re-gridded the original footprints to a spatial resolution of 0.025° × 0.025° to degrade the spatial resolution for the inverse analysis. First, the area source emissions were summed for each 0.025° × 0.025° grid cell. Then, individual footprints for the 0.025° × 0.025° grid were derived by dividing the sum of the nine XCO_2 contributions for the 0.0083° × 0.0083° grid by the emissions for the 0.025° × 0.025° grid.”

L915: “sigma_a” for prior uncertainty instead of “sigma_e”.

We have made this revision.

General model description:

lack of overview and strict separation of description of the inversion methodology, model setup details and results

We have refined the sentence structures in the revised manuscript. Specifically, we have moved the description of the simulation conditions in Section 4.2 (L331-342 in the initial manuscript) and the description of the construction of the prior error covariance matrix and measurement error covariance matrix in Section 4.2 (L387-412) to the Methodology section. Additionally, the descriptions of the

background (L349-357 and L375-382) have been combined and moved to the Methodology section. The remaining part of Section 4.3 (L413-446) has been merged with Section 4.2, and Section 4.3 has been removed.