

The authors developed an inversion scheme to infer the anthropogenic carbon dioxide emissions in the Tokyo Metropolitan 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 CO<sub>2</sub> enhancement (from the footprint and surface flux) from the observed XCO<sub>2</sub> 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.

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".
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  $121$ ).
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).
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 XCO<sub>2</sub> from Tsukuba COCCON site subtracted with simulated CO<sub>2</sub> enhancement.

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.

Detailed comments:

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

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

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

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

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.

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

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

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

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

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

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.

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

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

L142: unit wrong → ppm/(mol/m<sup>2</sup>/s)

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

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

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

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.

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

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

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.

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

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.

L267: "DXCO<sub>2</sub> values measured" This statement is confusing, since DXCO<sub>2</sub> values are derived from the forward model as described in the referenced section 3.1

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

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.

L279: Inverting in the log-space introduces a strong bias to positive emissions. Negative emissions are not necessarily non-physical, especially in case of CO<sub>2</sub>, 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.

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

L311ff: It is not very clear in the text what  $XCO_2^{\text{Diff}}$  means and how it separates from (DXCO<sub>2</sub>).

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  $XCO_2^{\text{Diff}}$  is actually used for? Just to look into temporal fluctuation?

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

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

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.

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?

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?

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

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).

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.

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

L497ff: The sentence is very long. Please reformulate.

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

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

L915: "sigma\_a" for prior uncertainty instead of "sigma\_e".

General model description:

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