

Response to Reviewer 1

We thank the two reviewers for their efforts and constructive comments (<https://www.geosci-model-dev-discuss.net/gmd-2018-123/#discussion>). Each reviewer's comments are shown below in *italics*, followed by our point-by-point responses in blue.

Anonymous Referee #1

This paper describes a new modification and application of the STILT model for total column measurements. This will allow X-STILT to be used to interpret satellite (and ground-based) total column abundances, and is a timely contribution, given the rapidly increasing number of satellite greenhouse gas total column measurements. The manuscript is thorough and technical, generally clear and well written, and suitable for this journal. I would recommend its publication after the following comments are addressed.

We thank anonymous referee #1 for the positive feedback and have attempted to address these comments and made several clarifications and changes to the manuscript. Here are major changes made during the review process including

- 1) the quantification of XCO₂ errors due to vertical mixing errors within X-STILT (**Sect. 2.6.2**);
- 2) discussions on results using observations from b8 Lite files (**Sect. 4.4**) and X-STILT's potential for broader applications and expected changes (**Sect. 4.3**);
- 3) details on the wind bias correction within the model (**Appendix C**); and
- 4) updated simulations and codes built on the STILT-R version 2 (Fasoli et al., 2018) with updated description in **Sect. 2.1.2**.

General Comments

"Is X-STILT restricted to XCO₂, or could it be applied to any total column tracer (e.g., XCH₄, XCO, XN₂O, etc.)? Could it also be applied to ground-based total column measurements? My understanding from reading this paper is that X-STILT could be applied more generally, and that you were showing a rigorous example of its functionality with OCO-2 XCO₂. If this is true, its generality should be made more clear – perhaps with a more general title and introduction."

In general, we agree with the reviewer on X-STILT's potentials for wider applicability. We did not study other species than CO₂ or incorporate profiles from other sensors. The codes we currently modified only aim at XCO₂ and incorporate OCO-2 satellite profiles. However, we do foresee that X-STILT could be applied more generally. Still, we are inclined to retain the current title and majority of the manuscript, as our methods (in particular our definition of background) are built on our particular focus over urban areas. Continuous work is ongoing towards a more flexible model framework that can be more easily applied to other column measurements.

We have now clarified the model's generality and expected changes for other tracers within a new subsection (Sect. 4.3) of the main text:

"4.3 X-STILT's potential for broader applications

In theory, X-STILT can be applied to other column measurements and other species. The underlying Lagrangian atmospheric model (STILT) has been applied to simulate other atmospheric species, such as CO, CH₄ and N₂O (Mallia et al., 2015; Kort et al., 2008). One of the key modifications to X-STILT from STILT is the column weighting of STILT footprint values (Sect. 2.1.2). Specifically, X-STILT interpolates the OCO-2 AK and PW onto each modeled level and then applies weighting of the trajectory-level footprints before generating a horizontal footprint map. The X-STILT code can be easily modified to apply sensor-specific vertical profiles of AK and PW from other satellites or ground-based column measurements.

Lastly and more importantly, background may need to be derived differently according to different applications, e.g., local urban emissions versus regional fluxes. The overpass specific background (M3) aims at isolating the citywide emissions, so it makes use of the measurements outside the city, but still are quite closed to the city (within the few

degrees latitude). However, if the study focus is to look at emissions over a much broader region (e.g., statewide emissions), background region should be defined farther away from the target region, e.g., taking the advantages of measurements from available upstream overpasses.”

Specific Comments

1. P2L30-36.

“I don’t think the current suite of CO2 satellites will completely fill in the gaps of the surface in situ networks, especially over specific locations such as cities. Certainly the future looks bright with OCO-3’s “city mode”, and the geostationary missions on the horizon, but I think you are overstating the impacts over cities as our satellite observing system currently stands.”

We agree with the reviewer that the surface in situ networks have their vital and irreplaceable role in the CO₂ measurement system. We might overstate the impacts over cities and be over-optimistic about the current suite of CO₂ satellites. Thus, we reworded the previous text (P2L34-36) as:

“Although most carbon-observing satellites have revisit times of multiple days (e.g., 3 days for GOSAT and 16 days for OCO-2), their global coverage, large number of retrievals and multi-year observations may further complement the current surface observing networks. Space-borne CO₂ measurements, in combination with surface CO₂ networks, may help reduce emission uncertainties and benefit urban emissions analysis, especially over regions with no surface observations (Duren and Miller, 2012; Houweling et al., 2004; Rayner and O’Brien, 2001)”

2. Section 2.1.

“I’m having trouble with your definition of the sensitivity of the satellite sensor: it seems incomplete. The column averaging kernel represents the change in the retrieved total column with respect to a perturbation in the abundance at a particular altitude. When the column averaging kernel is 0, the measurement is insensitive to changes at that altitude, and thus relies completely on the information in the a priori profile to construct the column. (Ref: OCO-2 ATBD, P58: https://co2.jpl.nasa.gov/static/docs/OCO-2%20ATBD_140530%20with%20ASD.pdf)

Weighting functions, which you mention on L33, at least to the retrieval community, refer to the Jacobian matrices, and while these are related to the averaging kernel matrices, they are not the same as the column averaging kernels (see P54 on the ATBD document above). I’d ask that this section is clarified further.

We apologize for the many confusions caused by the lack of clarity and a few mistakes made in Sect. 2.1 and Eq. (1). We modified the definition of AK as: “the sensitivity of the change in retrieved XCO₂ due to CO₂ anomaly at each retrieved grid” (P5L3). We reframed the entire Sect. 2.1 to clear up these confusions. And, here are some responses and clarifications:

- 1) In terms of representing the atmospheric column, we mislabeled n in Eq. (1) in the previous version, and corrections have been made to P5L10-13:

$$XCO_{2, sim, ak} = \sum_{n=1}^{n_{level}} (AK_{norm, n} PW_n CO_{2, sim, n} + (I - AK_{norm, n}) PW_n CO_{2, prior, n}) \quad (1)$$

I is the identity vector and n stands for the combined vertical levels of X-STILT plus OCO-2. Specifically, we replaced OCO-2 levels with denser model release levels for the lower part of the troposphere (red circles from the surface in Fig. 2), while kept OCO-2 levels for upper part (blue circles in Fig. 2). To reduce computational cost, the air column is only simulated up to the maximum release height (MAXAGL in meters above the ground level, mAGL; Fig. 2).”

- 2) Background definitions can be quite different among studies depending on their applications (e.g., examined spatiotemporal scales). In this study, because our focus is about the urban emissions from a target city, we defined the background XCO₂ as the portion that is not affected by urban emissions and naturally broke the total modeled XCO₂ down to two components – XCO_{2, ff} and XCO_{2, bg}.

- When we solely rely on model to estimate $XCO_{2,bg}$ (the trajectory-endpoint method, M1), the background XCO_2 is the sum of AK-weighted modeled biospheric and oceanic perturbations on top of CO_2 boundary conditions using CarbonTracker and a portion from OCO-2 a priori profiles (Eq. (2)).
- When the other background methods (M2H, M2S, M3) are being examined, the background XCO_2 per overpass is simple one number derived from statistical methods or the forward plume. Thus, no simulations on biospheric or oceanic anomalies involving OCO-2 prior or CarbonTracker is used in these cases.

- Clarifications have been made to the relevant text on **P5L23-28**:

“Eq. (1) can further be rewritten as Eq. (2), since the simulated CO_2 profiles in Eq. (1) is comprised of CO_2 boundary condition plus CO_2 anomalies due to sources/sinks (FFCO₂, biospheric and oceanic fluxes):

$$XCO_{2, sim.ak} = XCO_{2, sim.ff} + XCO_{2, sim.bio} + XCO_{2, sim.ocean} + XCO_{2, sim.bound} + XCO_{2, prior} = XCO_{2, sim.ff} + XCO_{2, bg}. \quad (2)$$

Given our focus, we defined background value as the XCO_2 portion not “contaminated” by urban emissions. Thus, $XCO_{2, sim.ak}$ is the sum of the XCO_2 enhancement due to FFCO₂ ($XCO_{2, sim.ff}$) and estimated background value ($XCO_{2, bg}$). Estimates of XCO_2 anomalies are further explained in Sect. 2.2.2 and four ways to estimate background values ($XCO_{2, bg}$) are proposed in Sect. 2.3.”

- 3) Yes—the “weighting functions” we referred to in Sect. 2.1 is the column averaging kernel and pressure weighing functions, instead of the Jacobian matrices in the retrieval. We have corrected our word choices (**P5L6-7**) to “the satellite’s column averaging kernels”.

On P5, you mention the interpolation of the measurement onto the model levels. Why wouldn’t you do the reverse: interpolate the model onto the retrieval grid? Your method requires that you make several assumptions that seem to complicate your analyses (i.e., these “scaling factors” you mention). Is there a compelling reason not to interpolate the model instead of the measurements? Please explain these “scaling factors” in more detail to walk the reader through Fig. 2.”

Here are explanations to the Reviewer’s questions on the scaling factors and interpolation of measurements’ profiles:

- 1) *PW* function is primarily estimated based on the air mass or pressure difference (dp) between two layers. Fig. 2b shows that OCO-2 levels have relatively **constant pressure difference** (dp_{oco2}) and *PW* values (in black dots, except for the very top and bottom level). On the contrary, model levels are finer (in orange circles in Fig. 2b) with two different vertical spacings in **altitudes** (100 m vs. 500 m) below and above 3 km. So, those scaling factors are to adjust the *PW* function according to difference in dp_{oco2} vs. dp_{stilt} .

For example, from 0 to 3 km, dp_{stilt} ranges from ~8-10 mb, while dp_{oco2} are mostly ~52 mb (red circles vs. black dots in Fig. 2b). Thus, we further scaled down the interpolated *PW* (red circles in Fig. 2b) by the ratio of dp_{stilt} over dp_{oco2} (e.g., $sf = 10 \text{ mb}/52 \text{ mb}$), because of less air between model levels than initial retrieval grids. Thus, final *PW* after scaling has value of ~0.01 (orange circles in Fig. 2b).

Comparing air below 3km, fewer model levels are placed from 3-6 km (~650 to 450 mb), which gives larger dp_{stilt} , larger *PW* scaling factors and resultant larger scaled *PW* of ~0.03–0.04 (orange circles in Fig. 2b).

Since no model level placed above MAXAGL, *PW* stay the same as the initial OCO-2 *PW* values (blue circles in Fig. 2b). Finally, we made sure that the sum of vertical *PW* profile ends up being 1 for each sounding/receptor.

- 2) We agree with the reviewer that we could construct the ‘redistribution matrix’ and interpolate the model values on retrieval grid, as done in Basu et al. (2013). However, we would like to keep the current method for the following reasons (with reasons explained on **P5L14-16** in the revised version as well):

- It seems that the construction of redistribution matrix may also need to deal with or resolve the mismatch between model levels versus retrieval grids, unless model levels perfectly agree with retrieval grids.
- Most importantly, as our intention is to preserve finer variations in modeled CO₂ benefit from placing denser model levels within the PBL, we foresee the reversed interpolation (from model levels back to retrieval grids) could potentially involve some averaging or smoothing.

3) Clarifications have been made to the main text in **Sect. 2.1 (P5L14-22)** to explain above comments:

“Interpolations are further needed to resolve the mismatch between prescribed OCO-2 retrieval grids and model levels for the lower part of the troposphere. Our intention is to preserve the finer modeled CO₂ variations by performing interpolations of satellite profiles from retrieval grids to model levels. Vertical profiles of AK_{norm} , PW and $CO_{2,prior}$ are treated as continuous functions and interpolated linearly to model grids (red circles in Fig. 2). Note that the initial OCO-2 PW functions have steady value of ~ 0.052 (except for the very bottom and top levels; black dots in Fig. 2b), which results from constant pressure spacings (dp_{oco2}) between two adjacent OCO-2 levels. However, X-STILT levels are much denser with smaller pressure spacings (dp_{stilt}) or less airmass between their two adjacent levels. Therefore, the linearly interpolated PW (red circles in Fig. 2b) needs an additional scaling via a set of “scaling factors” representing the ratios of pressure spacings in STILT versus OCO-2 retrieval (dp_{stilt}/dp_{oco2}), to arrive at the correct PW for each finer model grid (orange circles in Fig. 2b).”

3. Section 2.4.

“A recent paper by Nassar et al. (2017) would be relevant to cite in Section 2.4. They use OCO-2 data to quantify power plant emissions, and they choose an overpass-dependent background that would be interesting to compare with your method. Ref: Nassar, R., T. G. Hill, C. A. McLinden, D. Wunch, D. B. A. Jones, and D. Crisp (2017), Quantifying CO₂ emissions from individual power plants from space, Geophys. Res. Lett., doi:10.1002/2017GL074702.”

We appreciate the reviewer in pointing out this paper. We have added this relevant paper when discussing different ways to determine the background in **Sect. 2.3.3 (P8L27-29)**:

“Nassar et al. (2017) derived overpass-dependent background and its uncertainty based on the averaged OCO-2 observations within four different tested background latitudinal ranges.”

as well as discussing the challenges in defining background and associated uncertainties when dealing with column measurements in **Sect. 1 (P4L17-19)**:

“Lastly but more importantly, recent column studies (Nassar et al., 2017; Fischer et al., 2017) studied the impact of potential errors/biases in background values on their emission or fluxes estimates.”

In general, both studies derived the background from OCO-2 measurements over the “clean” region and accounted for background uncertainties. However, as Nassar et al. (2017) focuses more on emissions from individual power plants over different regions, we did not conduct direct comparisons against our background values.

4. Section 2.6.

My (admittedly simplistic) understanding of the transport error issue is that it is still important for models to get the vertical transport right when assimilating or inverting total column measurements, because the vertical transport sets the altitude at which advection occurs, and thus the distribution of the gas around the planet. So while the total column measurements themselves are insensitive to the altitude of the molecule within the column, it is not necessarily the case that the models are better able to reproduce the column. Indeed, the abstract of Lauvaux and Davis cited in this paragraph seems to confirm that vertical transport errors are very important for calculating fluxes from column-integrated measurements. Please address this issue further.

We agree with the reviewer that while likely small, we may not completely be able to neglect the impact of vertical transport on column simulations. To quantify this error, we have now conducted another set of transport analysis to quantify XCO₂ errors due to vertical mixing. Sect. 2.6 in previous version has now been divided into two subsections for horizontal (Sect. 2.6.1) and vertical transport errors (Sect. 2.6.2, newly added). Section 2.6.2 has been added as:

“2.6.2 Vertical transport errors

Vertical turbulent mixing dominates the vertical transport of air parcels and control the dilution of surface emissions within the PBL (e.g., Gerbig et al., 2008). Uncertainties in the vertical mixing or PBL height can affect both the footprint magnitude and the its spatial distribution via different horizontal advections at each altitude. Although column-integrated measurements may be less sensitive to vertical distribution of air particles than in situ measurements, vertical transport errors can have some impacts on column simulations nonetheless, due to wind shear and its interaction with vertical redistribution of air parcels (Lauvaux and Davis, 2014). Comprehensive quantifications of the vertical transport errors in a column sense are performed in Lauvaux and Davis (2014) using ensemble of surface and planetary BL parameterizations involving a regional inverse modeling framework.

Instead, we made use of the stochastic nature of STILT and propagated typical PBL height errors in the model. Changes in STILT-modeled mixed layer height modify the vertical profiles of turbulent statistics that directly control the stochastic motions of the Lagrangian air parcels (Lin et al., 2003). Thus, we obtained different air parcel trajectories with rescaled PBL heights. The resultant vertical transport error in XCO₂ space is calculated as the root-mean-squared errors (RMSEs) between two sets of XCO₂ enhancements among different receptors for each overpass. Due to this calculation, vertical transport errors are only provided at the overpass level (results in Sect. 3.5). Gerbig et al. (2008) reported typical relative PBL errors in the range of $\pm 20\%$. Thus, we rescaled the PBL heights higher and lower by 20% and evaluated the scaling's impact on XCO₂ enhancements. Because of our focus on the urban emissions and potential small XCO₂ enhancements contributions beyond one day backwards in time, we only rescaled PBL within the first 24 hours of transport before arrival of the air parcels at the column receptors.”

We further briefly mentioned the XCO₂ errors due to vertical mixing error as in **Sect. 3.5**:

“XCO₂ errors solely resulted from vertical mixing errors are in general $< 15\%$ of the modeled signal for each overpass, whereas XCO₂ errors due to horizontal wind errors dominate the overall XCO₂ transport error (Table 1).”

Furthermore, there is little discussion about the atmosphere above MAXAGL (~450 hPa). While this is unlikely to be important for CO₂ emissions over regional or smaller scales, it may be important for other tracers (e.g., CH₄). Can you comment on how important the tropopause altitude, for example, might impact this work? Over what spatial and temporal scales would X-STILT properly represent the total column?

We strongly agree with the reviewer that atmosphere above MAXAGL can be less important for XCO₂ emissions, but can be important for other tracers like methane, due to chemical productions and losses along the atmospheric transport. In theory, the vertical profile of methane can be sensitive to the tropopause altitude since it rapidly decreases in the stratosphere.

We admit that no chemical reaction is considered in X-STILT at this point and the model levels are only placed over the lower troposphere, given our focus on urban-scale CO₂ surface emissions (that only get mixed within the PBL). As particles released from higher levels can hardly get entrained and make contact with the near-field land surface, the X-STILT column particles may properly represent the total column over the urban scale for less than one day. Due to our specific focus on urban CO₂, we decided not to add related content in the main text, but to address the reviewer's question here.

However, we may expect small impact on our work with reasoning showed below.

- Although variations for atmosphere above MAXAGL are not explicitly modeled or represented by X-STILT particles given the way we release air parcels, those variations are part of the defined background.
- Take methane as an example. Recall that the background air defined by the overpass-specific method (M3) are atmospheric columns located outside (but not far away from) the city plume. The background value is derived from measured XCH₄ over those background atmospheric columns. In fact, those measured XCH₄ are the resultant XCH₄ after being produced or destructed along their way (backward in time) and have contained information about the upper tropospheric and stratospheric variations over upwind regions. Since the background air and the plume-elevated air are not far away from each other, we may assume no big difference in their tropopause altitudes. Thus, both background and plume-elevated XCH₄ contains information about the XCH₄ over atmosphere over MAXAGL. And, the difference between the two gives us the methane enhancements due to urban emissions.
- One note on the background definitions: When using the trajectory-endpoint method (M1), model trajectories are used to model the XCO₂ boundary condition. However, when using the other three background methods (M2H, M2S and M3), model trajectories are only used to estimate the XCO₂ anomalies due to different sources and sinks. The choice of max release height may impact the modeled XCO₂ enhancements. While as shown in the MAXAGL sensitivity test, almost no changes in XCO₂ enhancement due to increase in MAXAGL.

5. OCO-2 v7.

You are using v7 of the OCO-2 data in this paper, but v8 is available and v9 will be available soon. V8 has significant improvements in the treatment of aerosols and throughput, which may be important for work over polluted urban regions. V9 will have improved pointing, especially important over topography. Please comment on whether your results will be robust against these changes to the OCO-2 data.

We thank the reviewer in pointing out different OCO-2 versions and look forward to the upcoming b9 product to help future studies over cities with complex terrain. We have now performed few more simulations and analysis using version 8 Lite product and briefly depicted the changes in observed and simulated urban XCO₂ signals in **Sect. 4.4 (P19L23-33)**. As ongoing improvements made in OCO-2 retrievals are modifying the observations, it may not be that fair for us to comment the robustness in our results, especially given model evaluation against observations.

Text has been added:

“Emission evaluations for different regions can be different and affected by different observational constraints. Even changes in different versions of the retrieval (Lite b7 vs. b8) may slightly affect the model-data comparisons and simple inversion results in this work. Modeled XCO₂ enhancements using the newer b8 differ slightly from those using b7 (purple dots in Fig. S8 vs. in Fig. S13) due to changes in the locations of the receptors, column averaging kernels, and data filtering (QF) for measurements around Riyadh. Specifically, observations from b8 may yield more overpasses with sufficient screened soundings than those from b7 (black and red bars in Fig. S1). However, much larger differences in observed enhancements are found and caused by the changes in total observed XCO₂ and estimated background values. Specifically, background uncertainty decreases by up to 0.1 ppm primarily attributed to smaller spread (smaller SD) of the observed XCO₂. Positive shifts in the total observed XCO₂ for b8 from b7 are found over most overpasses (Fig. S11). The M3-derived observed enhancements may be less affected by positive shifts in total observations, given similar positive shift associated with the overpass-specific background near the target urban region (dark green dashed lines in Fig. 6e vs. Fig. S12).”

Technical Comments

6. Section 2.1.1

I found myself wondering which version of OCO-2 data you were using, given the discussion about quality flags and albedo cutoffs, which is version-specific. I realize this is answered later in Section 2.2. I'd suggest either not mentioning the specifics of the quality filtering in section 2.1.1, or mentioning the data version in 2.1.1.

We realize the confusion caused by the introduction on quality flags in Sect. 2.1.1. We have removed this discussion (the data filtering on observations, i.e., QF and aerosols cutoffs) in Sect. 2.1.1.

7. P10L14-7

I'm having trouble understanding these sentences, and I believe this wind correction may be an important step in X-STILT. Please explain in more detail.

We apologize for the confusion and clarify that this attempt to correct wind biases can be an important part in X-STILT, in particular over places with large systematic wind error, less complex terrain and denser wind observations. We briefly mentioned this correction and discussed its limitation in **Sect. 2.6.1 (P11L33- P12L9)** and added an **Appendix C** for details of this bias-correction:

“Appendix C: Correcting for wind biases within X-STILT

While we did not apply the wind bias correction for the overpasses analyzed in this paper due to the biases being generally small (previously explained in Sect. 2.3.3), X-STILT has the capability to account for biases, if necessary. The basic idea is to correct the near-field wind biases in both forward- and backward- time trajectories. Because wind error at each observed pressure level can be quite different, vertically-weighted u- and v- wind biases were calculated by fitting logarithmic mean wind profiles based on available near-fields observed and simulated wind speeds and directions. We then calculated the deviations in latitude and longitude directions (dx, dy, with conversion from distance to degrees) given estimated u- and v- wind biases. These deviations accumulate as air parcels travel further backward or forward in time and are used to correct the location of each particle. After fixing the particle locations, Fig. S6b shows the general distribution of backward trajectory being clockwise rotated, compared to initial trajectory distribution in Fig. S6a. Air parcels in Fig. S6b appear to be “noisier” than those in Fig. S6a, due to inclusion of the random wind error component. Then the new bias-corrected set of column trajectory is used to generate spatial footprint. This correction can also be performed to forward-time trajectory to reduce wind bias impact on best-estimated background value using the M3 method.”

Unfortunately, we ended up not performing bias correction to model particles, because the wind observation sites are not perfect around Riyadh to estimate a robust wind error to rotate model particles. We mentioned these limitations and other more comprehensive methods in **Sect. 2.6.1 (P12L1-9)**:

“Unfortunately, only 2 radiosonde stations around Riyadh with 3 vertical pressure levels within the PBL (and sometimes with missing data) may be insufficient to correctly interpolate the near-field vertical wind biases. However, cities with meteorological profiles sampling more levels within the PBL and higher temporal frequency in reporting observed vertical winds will be more suitable sites to retrieve the near-field wind errors. Other methods include rotation and stretching of urban plumes derived from WRF-Chem (Ye et al., 2017), similar to the rotation of X-STILT air parcels, to quantify errors in wind directions and speeds. Deng et al. (2017) sought correction of wind biases in a sophisticated manner via data assimilation. Yet, the near-field correction within X-STILT can be potentially utilized in the future as a quick bias correction to the near-field wind in LPDMs, given denser wind observations and relatively flat terrains. Therefore, we decided to reduce the potential impact of wind bias on model-data comparisons using a latitudinal integration (further in Sect. 3.5).”

8. P13L3-6:

I believe you are saying that M2H is in general lower than M3. But can you say definitively that M3 is correct, and thus M2H has the bias? Also, 0.56 ppm is not small! This can be 25-50% of the enhancement.

We agree that a mean difference of 0.56 ppm (between M2H- and M3-derived background values) is actually quite large, given small column enhancement. And, we cannot definitively say that our method is correct since the “true” background is more of an unknown value. So, we have reworded some relevant sentences (e.g., “bias” to “mean difference”).

However, we would argue that M2H may not be suitable for local/urban studies and have rewritten the latter two paragraphs in **Sect. 3.3 (P13L36-P14L17)** to objectively discuss the pros and cons of M2H and M3 methods (also pasted below):

“We now focus on the comparison between M3 and M2H with objectively analyzing their advantages and limitations. On average, M2H derived background is lower than our localized “overpass-specific” background by 0.55 ppm (Fig. 6e), which can primarily be attributed to different defined background regions. M3 defined the background region from the same track as the one over Riyadh, which guarantees that the background air contains variations due to long-term atmospheric transport, natural sources/sinks and FFCO₂ emissions except for local emissions (e.g., from Riyadh). Whereas the enhanced air contains the enhancements due to local emissions on top of all the information included in the background air. Therefore, the subtraction between M3-defined background and enhanced air correctly represent the XCO₂ portion enhanced by the local emissions. On the contrary, M2H use a fairly broad background region (0° N–60° N, 15° W–60° E in Fig. S7) to estimate gridded anomalies over all places in Europe, Middle East and North Africa. Although may yield more data, this broad spatial region may misrepresent the correct upwind region, because the wind regime can be quite different among different overpass dates or seasons.

We admit M3-defined background range and background value can be affect by potential large wind bias over cities other than Riyadh. However, the impact on background may be small and is implicitly considered in the background uncertainty (previously discussed in the last paragraph of Sect. 2.3.3). As for M2H, all regional OCO-2 measurements are lumped into its background calculation. For example, some measurements on the east-most overpass in Fig. S7 are affected by Riyadh’s emissions, whereas atmospheric columns at soundings along the west two overpasses in Fig. S7 may not necessarily be the background air that eventually arrives at region around Riyadh. Thus, the regional median of XCO₂ may not physically indicate the accurate background that is supposed to isolate local-scale fluxes. Therefore, our localized overpass-specific background is designed and more suitable for extracting local-scale XCO₂ anomalies. Given relatively small urban enhancements around our study site, this 0.55 ppm difference may lead to large differences in estimated observed urban signals and emission evaluations (Sect. 4.2).”

Lastly, we may admit that our definition (using one mean value to represent the background) is not perfect and are aware of the potential wind bias impact on background definition. So, we made several attempts in this work:

- 1) by widening the polluted latitude range by bringing a wind error component (Sect. 2.3.3);
- 2) by trying to correct the wind bias on forward plume via a near-field correction (Appendix C, better for regions with denser wind observations); and
- 3) by introducing background uncertainty as the combined error impacts from retrieval error and natural variation (SD) of observed XCO₂ over background latitude range (Sect. 2.3.3). Retrieval errors of measurements over the background range have now been included in the background error as well, based on a comment from the Reviewer #2.

Relevant discussions based on above point 1) to point 3) have been made to **Sect. 2.3.3** on **P9L17-28** (also pasted here):

“In addition to random errors (that are resolved by the inclusion of the aforementioned wind error component and broadening of the city plume), potential large bias in near-field wind direction may lead to mismatch in modeled and observed background regions and may bring relatively higher XCO₂ values into background XCO₂. However, we do not explicitly account for the potential near-field wind bias’s impact on forward-trajectories defined urban plume with following considerations. Firstly, we attempted to propagate a near-field wind bias into the modeled plume by rotating

forward trajectories, whereas the robustness of this near-field bias can be affected by the very few wind measurements near Riyadh (further explained in Sect. 2.6.1). Secondly, the background latitude range defined by M3 with the broadening effect (blue lines in Fig. 5b) in general matches well with that observed from OCO-2 for most overpasses, which implies that the overall wind bias around our study site is not significant. Lastly, even if potential wind bias may result in less accurate background range and bring elevated XCO₂ into the background, the background uncertainty implicitly contains information about the spatial variation in background measurements (green ribbon in Fig. 5b). In addition, the M3-derived background is the mean value of mostly hundreds of background observations (numbers in Fig. 6e), which may not be greatly affected by a few potential urban-enhanced measurements.”

9. There are several instances of omitted definite and indefinite articles, and a few typos here and there, but I assume that once this paper has been accepted, the copy editor will find and correct them more thoroughly than I have. However, I will list the ones I caught here. Between ** are the edits I suggest.

P2L17: top-down constrain*ts* P3L6-7: shed light** on CO2 emission** monitoring network*s*. P4L18: Riyadh*,* with *a* population P4L20: Saudi Arabia has the largest CO2 emission*s* among... P4L32: “apple*s*-to-apple*s*” P4L33: weighted using *the* satellite’s *column averaging kernels*... P5L17: at the same lat/lon as *the* satellite... P5L21: compare ** overall modeled P5L28: *The l*onger the time an air parcel..., *the* higher its footprint value... P6L5: FFCO2 *are* derive*d* from... P6L13: we binned ** the observed... P6L16: estimate *the* increase in observed... P6L22: 1x1 km resolution on ** monthly scale*s*... emission estimates by fuel type** from the... specific ODIAC emission categories on *a* monthly basis... P6L31: line sources and diffuse** sources... P8L32: which are more straightforward** and efficient** than solely *relying* on... P9L7: boundary of *the* city... P10L19: more suitable sites *to* retrieve... P10L25: get around ** the impact on... P11L6: we fit *an* exponential variogram... P12L11: which results in *an* overall smaller footprint... Yet, column foot- print*s* cover**... P12L14: an air column can be one or *a* few orders... P12L18: regardless *of the* adopted meteorological fields. P12L33: Here we emphasi*ze*... P13L29: sensible → sensitive P14L21: according to *the* OCO-2 Lite file... P14L24: scatter*ed* P15L36: latitudinal** integration ←this happens in other locations as well P16L30: exceeding *a* certain averaged... P17L8: emissions of *a* target city P18L2: even large impact*s* on *the* posterior... can be caused by using *a* background derived from simplistic statistic*s* P18L13: hampered ** due to... P18L24: improves biospheric flux** estimation... P20L10: for *a* few levels... P21L17: These small changes *show* that our *latitude band integration*... a second peak or miss** large XCO2 enhancements. P21L21: *widths* P21L27: The word “benefit” seems out of place here. P21L30: Based on three simpl*e* tests...

We sincerely thank the reviewer in thoroughly reading the manuscript and pointing out these issues listed above. We have corrected them all in the relevant text.