Evaluating the Impact of Storage-and-Release on Aircraft-based Mass-Balance Methodology Using a Regional Air Quality Model

Sepehr Fathi *et al.*

ACPD – Responses to Referee Comments

Manuscript number: ACP-2021-343

Response to Referee Dr. Wayne Angevine

Please note the different font/colors in this document for:

1. Referee Comments (RC) in black,
2. Author Comments (AC) in light blue,
3. **Revised and/or added text from the revised manuscript in bold dark blue font**.

RC1: 'Comment on acp-2021-343', Wayne Angevine, 12 Jul 2021
This paper is a useful contribution to an important and neglected topic, the estimation of uncertainty in top-down emissions estimates. It is reasonably clear, comprehensive, and will provide an important reference for further work on the topic. I have a few suggestions for possible minor improvements, but generally find the paper suitable for publication.

AC.1

We thank the referee Dr. **Wayne Angevine** for providing positive and constructive feedback and for raising valid and important points (https://doi.org/10.5194/acp-2021-343-RC1). Below we provide specific responses to each comment.

General comments:

1. The analysis assumes a certain framework for the observations, that is, repeated passes ("screens") at different altitudes, at relatively short distances downwind of the source. This is one common observation strategy, particularly applicable to compact sources. The other common strategy is to fly single legs at longer downwind distances in well-mixed conditions (see for example the works by Jeff Peischl and coauthors). That strategy is better for large area sources. Some comments about the applicability (or not) of this analysis to the alternative flight strategy would be useful. In particular, how does the uncertainty found here depend on downwind distance?

AC.1.1

Dr. Angevine makes a good point that different aircraft flight path strategies have been employed for emissions retrievals aside from the box flights considered here. Based on our analysis however, we feel that the effects of storage and release events may become more significant with increasing distance from the sources. We have added the following text and Figure C1, in Appendix C, acknowledging this issue in the revised manuscript"

Revised version: Section 4.4, 2nd paragraph

"**We have examined storage and release events in the specific context of "box" flights around an emitting facility, however note that other strategies have been put forward in the literature (e.g. near-field downwind transects, Peischl *et al*., 2010). We note: (1) The heterogeneous nature of meteorology even at the scales employed for the box flights here suggests that the impact of storage and release will likely be greater as the distance between screen(s) and emission source(s) increases (for example see Appendix C, Fig. C1, for case 8, where the contribution of the storage term increases as a function of downwind distance): short distance "box" flights such as examined here would reduce the storage and release impact; (2) Our $\Delta_t Ri$ and $\Delta_t \alpha$ parameters may nevertheless be applied to "single-leg" downwind transects retrievals, and repeat flights (e.g. by a second aircraft following behind the first aircraft) may be one way of reducing storage and release uncertainties in these flights; (3) Our ϕ parameter relies on the availability of upwind information and hence is not applicable for "single-leg" (e.g. downwind transects) retrievals. (4) Our ϕ parameter provided the strongest correlation between storage and release and retrieval outcomes and therefore, where possible, should be estimated from additional upwind observations. These findings suggest that, while some aspects of storage and release can be identified with "single-leg" downwind flights, the impacts of storage and release for these flight patterns may be higher than for "box" flight patterns such as studied here.**"

"**Appendix C**

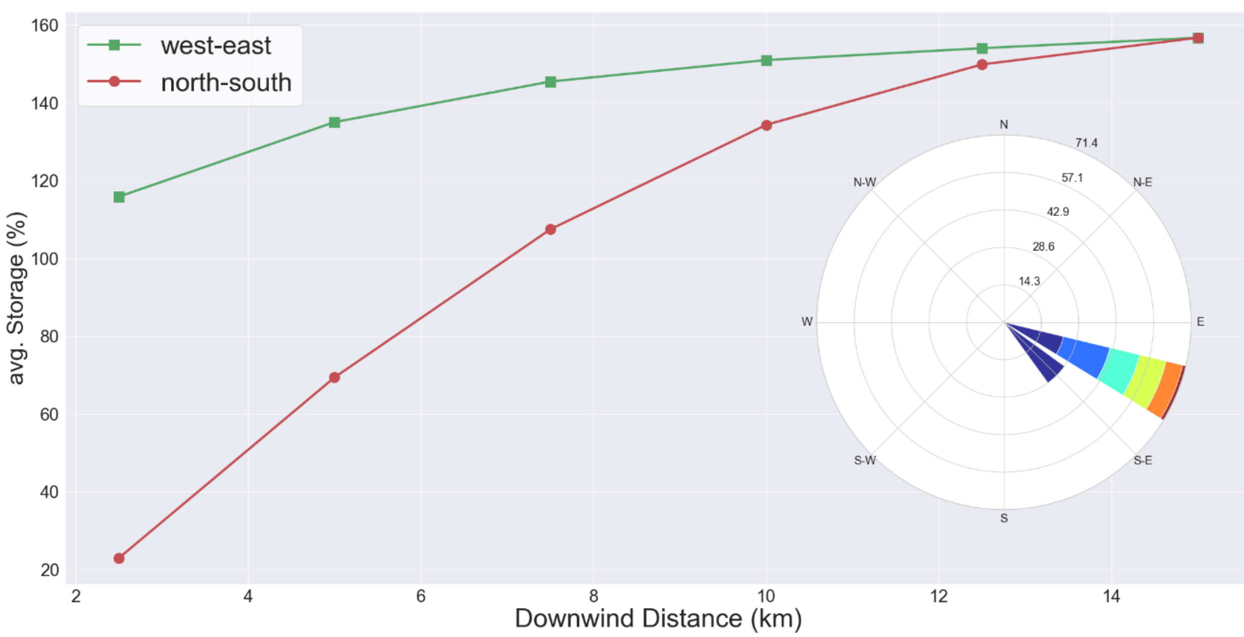

Figure C1. The contribution of the storage rate term ($E_{C,S}$) to the net retrieved emission rates (E_C) as a function of downwind distance for case 8. As shown in the wind-rose diagram, the wind direction for this case was towards south-east (S-E), and therefore the downwind dependence is shown in two directions west-east (green squares) and north-south (red circles). The contribution of storage increases as a function of downwind distance. This suggests that the impact of storage-release may increase as

the distance between the downwind screen (sampling location) and the emission sources increase. A similar dependence of the magnitude of the storage-release with downwind distance was observed for the other cases."

2. I am somewhat skeptical of the claim that the Revised Terra Retrieval method used in the model can be replicated in the real world by duplicate flights. For that to be true, the duplicate flights would have to provide a robust estimate of the time derivative of the observed concentration. For the concentration variations shown in figure 7, for example, two randomly chosen points in time would not be likely to produce a robust estimate. Two are certainly better than one, but maybe not good enough. Please comment.

AC.1.2

We thank the referee for raising this valid and important point. Tandem aircraft flying a short distance apart along the same flight path would reduce the time duration issue – however, this would effectively double the cost of a study. However, in discussing the Reviewers' comment, we realized that a refinement of our suggested approach would alleviate some of these issues. A single aircraft equipped with a high spatial resolution vertical scanning instrument such as a LIDAR, along with repeat flights at a single level, could provide a similar highly time-resolved set of column data as the regional model used in our work as a proxy for observations. In response to this comment we therefore added the following paragraphs to the revised version and the discussion in the manuscript:

Revised version: Section 2.2.3, 2nd paragraph

"**Considering the potentially prohibitive operational cost of conducting repeat flights, and the fact that few (e.g. 2-3) time-consecutive measurements may not provide enough information for the estimate of temporal trends in within-box tracer mass budget, the following alternative is proposed. As we discuss later in Sect. 3.3, observed temporal trends of tracer concentrations ($\partial\chi_C/\partial t$) downwind of the emission source are also in similar strong correlation ($P_r = 0.9$) with the corresponding DR estimates (using volumetric time-series). Furthermore, our studied cases suggest that downwind vertical-profiling of tracer concentrations can provide similar estimates of the temporal tends required for the estimate of the storage rate $E_{C,S}^{tr*}$ and can be substituted in Eq. (15) for $\partial\overline{\chi_C}(t,z)/\partial t$. In observational applications, this can be achieved through ground-based or, preferably, aircraft-based vertical profiling remote measurements (e.g. LIDAR measurements, Aggarwal *et al*., 2018), where the sampling aircraft with a remote measurement setup would collect column data while flying around or downwind of the emission source. The use of an aircraft-based vertical profiling instrument for a small number of chemical fields, along with a repeated, single-loop flight path around a facility (10-15 min) would generate fields similar to those generated by the air-quality model used here, in turn allowing highly time-resolved estimates of the storage term. This alternative approach can be more cost (operational) efficient compared to the strategy of repeat flights (which would require a second aircraft travelling in tandem behind the first to achieve the same time resolution), while providing more time-consecutive data points for the study of the temporal trends in the tracer mass budget, and further reducing the relatively small emissions biases associated with extrapolation to the ground of observed fields.**"

Revised version: Section 3.3, mid-paragraph

"**As discussed in Sect. 2.2.3, an alternative to multiple aircraft-based measurements is to estimate the temporal trends in tracer concentrations ($\partial\chi_C/\partial t$) through remote vertical-profiling (e.g. LIDAR measurements). Our analysis of different flight cases show that downwind trends are also in strong correlation with DR method volumetric time-series estimates with the same correlation coefficient of $P_r = 0.9$ (with different $m = 0.57$ and $b = 0.03$). The advantage of remote (downwind and upwind) vertical-profiling over multiple aircraft or UAV measurements, in addition to operational cost efficiency, would be the collection of more temporal measurements over the sampling time, which in turn can result in improved estimates of $\partial\chi_C/\partial t$ and the storage term ($E^{tr*}_{C,S}$); as opposed to estimates based on few time-consecutive measurements (e.g. 2-3) via multiple or in tandem aircraft.**"

Revised version: Section 3.3, end of paragraph

"Again, for **observational applications: (1)** a repeat box flight procedure can be used, where an aircraft carries out a second box flight immediately after the first flight, or two aircraft (or UAVs) follow the same flight path in tandem, one positioned a fixed distance behind the other**, or (2) remote vertical-profiling (e.g. remote LIDAR measurements) on a single aircraft can be employed to collect time-consecutive measurements of relevant fields, in the column, around the emitting facility.**"

Revised version: Conclusions section, end of last paragraph

"We have also devised a methodology to reduce the impact of this form of emissions retrieval error by estimating the storage rate ($E^*_{C,S}$) through **(a)** the use of repeat flights, around the same facility, either with a single or multiple aircraft(s), **or (b) aircraft-based remote vertical-profiling of relevant fields during the sampling period.**"

Minor comment:

1. The marker colors in all the figures should match. For example, I think the colors in figure 8 don't match figure 3.

AC.1.3

In response to this comment figures 3, 5, 8, 10, 12 have been revised so that the markers and colors representing results from the three methods would match in all the figures. We thank the Reviewer for catching this!

Citation: https://doi.org/10.5194/acp-2021-343-RC1

AC.1 - Additional Author Comment

Please also note that we have corrected a minor mistake in the original manuscript, in the information presented in Figure 6, Section 3.3, and the manuscript text referring to this Figure, as we show below. Note that this correction presents the more accurate (and stronger) correlation between the 4D and 3D estimates of temporal trends in tracer mass. Note that this correction has no impact on the results presented in the manuscript.

In original manuscript, section 3.3, Figure 6 caption:

"… The least-squares fit line has slope m=2.98 and intercept b=0.09. Note that case 8 estimates are in low correlation due to relatively high upwind emissions during the time of that particular flight (August 28th), rendering the 3D estimates less representative of the 4D estimates. Excluding Case 8 from this analysis, results in P_r=0.88. $E^{*}_{C,S}$ can be estimated from multiple consecutive 2D screens, by assuming 3D estimations of $\partial \chi_C / \partial t$ as representative of the rate of change in species mass within the flux box."

In revised manuscript, section 3.3, Figure 6 caption:

"… The least-squares fit line has slope m=**1.36** and intercept b=0.09. Note that case 8 estimates are in low correlation due to relatively high upwind emissions during the time of that particular flight **case** (August 28th), rendering the 3D estimates less representative of the 4D estimates. Excluding Case 8 from this analysis, results in P_r=0.88**, m=1.07 and b=0.03**. $\boldsymbol{E^{tr*}_{C,S}}$ can be estimated from multiple consecutive 2D screens, by assuming 3D estimations of $\partial \chi_C / \partial t$ as representative of the rate of change in species mass within the flux box. **The same can also be estimated from downwind vertical profiling of tracer concentrations over the sampling time.**"

In original manuscript, section 3.3, 1st paragraph:

"$\partial \chi_C / \partial t$ range is higher for 4D estimates (2.52 ppbv/hr) than for 3D estimates (0.98 ppbv/hr), but the two variables are correlated. The correlation coefficient $P_r = 0.7$ (with slope $m = 2.98$ and $b = 0.09$ intercept for the least-square fit line), indicates strong correlation between 4D and 3D estimates. The exception being rejected case 8, where the correlation is low due to relatively high upwind emissions during the time period of this case (on August 28th). Excluding case 8 from this analysis, increases the correlation between 3D and 4D estimates to $P_r = 0.9$"

In revised manuscript, section 3.3, 1st paragraph:

"**The** $\partial \chi_C / \partial t$ range is **slightly** higher for 4D estimates (2.52 ppbv/hr) than for 3D estimates (**2.17** ppbv/hr), **and** the two variables are **strongly** correlated. The correlation coefficient $P_r = 0.7$ (with slope $m = \mathbf{1.36}$ and $b = 0.09$ intercept for the least-square fit line), indicates strong correlation between 4D and 3D estimates. **An** exception **is** rejected case 8, where the correlation is low due to relatively high upwind emissions during the time period of this case (on August 28th). Excluding case 8 from this analysis, increases the correlation between 3D and 4D estimates to $P_r = 0.9$ **(with $\boldsymbol{m = 1.07}$ and $\boldsymbol{b = 0.03}$).**"

Evaluating the Impact of Storage-and-Release on Aircraft-based Mass-Balance Methodology Using a Regional Air Quality Model

Sepehr Fathi *et al.*

ACPD – Response to Referee Comments

Manuscript number: ACP-2021-343

Response to Anonymous Referee #1

Please note the different font/colors in this document for:

1. Referee Comments (RC) in black,
2. Author Comments (AC) in light blue,
3. **Revised and/or added text from the revised manuscript in bold dark font**.

RC2: 'Comment on acp-2021-343', Anonymous Referee #1, 21 Jul 2021
The authors validate the accuracy of the aircraft-based top-down emission rate retrieval algorithm. They investigated the impact of meteorology and surrounding emissions on the derived emission rates and gave suggestions to design aircraft campaign. The investigation looks sound to me, but the manuscript is not very easy to follow. I would suggest polishing the paper. I would recommend revision before publication.

AC.2

We thank **Anonymous Referee #1** for providing constructive feedback and suggestions for improving the manuscript (https://doi.org/10.5194/acp-2021-343-RC2). Below we provide specific responses to each comment.

General comments:

1. I would recommend rephrasing the introduction. The authors have conducted a good literature review by listing several studies in the first and second paragraph. However, it is a pity that the literatures have not been well organized. It is hard for readers to tell why those literatures are chosen as representativity and how our knowledge has been gained gradually. For example, the pro and con of current methods are missing. I'm not very sure about the major point of the 2nd paragraph. Please try to better organize those two paragraphs in the revised manuscript. Additionally, is the term of "Storage-and-Release" invented by this study? If so, I'm wondering how this effect has been described by previous studies? What is the motivation of giving it a new name here?

AC.2.1

We thank the Reviewer for this comment. In order to address the issue raised, the text in the introduction section has been reorganized and revised (we have included several additional references to give better representation across the literature).The changes are in four places in the revised manuscript as follows:

Original manuscript text, Introduction, 1st paragraph:

"Top-down estimates can also be used to provide highly resolved emissions data for input into air quality models. The mass-balance technique and Gauss's divergence theorem can be employed to infer point and area source emission rates from aircraft measured data. Several studies have utilized aircraft-based mass balance emission rate retrievals in the past (Kalthoff *et al*., 2002; Mays *et al*., 2009; Turnbull *et al*., 2011; Karion *et al*., 2013, 2015; Cambaliza *et al*., 2014; Gordon *et al*., 2015; Nathan *et al*., 2015; Tadic *et al*., 2017; Conley *et al*., 2017; Ryoo *et al*., 2019). Flight patterns from these studies include single-height transects, single-screen flights and box flights. Box flight refers to flight patterns with enclosed shapes (polygonal, cylindrical) and is accomplished by flying in closed paths (loops) at multiple altitudes around the emission source while making measurements of meteorological fields and pollutant concentrations."

Revised manuscript text, Introduction, 1st paragraph:

"Top-down estimates can also be used to provide highly resolved emissions data for input into air quality models. **Several studies have utilized aircraft-based top-down emission rate retrievals in the past (Kalthoff *et al*., 2002; Mays *et al*., 2009; Turnbull *et al*., 2011; Karion *et al*., 2013, 2015; Cambaliza *et al*., 2014; Gordon *et al*., 2015; Nathan *et al*., 2015; Tadic *et al*., 2017; Conley *et al*., 2017; Ryoo *et al*., 2019). Flight patterns from these studies include single-height transects, single-screen flights and box flights. Box flights, which expand upon "single-leg" (mainly downwind) flights, refers to flight patterns with enclosed shapes (polygonal, cylindrical) and is accomplished by flying in closed paths (loops) at multiple altitudes around the emission source while making measurements of meteorological fields and pollutant concentrations (Gordon *et al*., 2015, Tadic *et al*., 2017). The mass-balance technique and Gauss's divergence theorem — where the net mass flux exiting a control volume (enclosed by the box flight path) is equated to the within-volume emission rates — can be employed to infer point and area source emission rates from aircraft measured data.**"

The following text has been added to the revised manuscript, Introduction, 2nd paragraph:

"**The potential for measurement error to affect retrieval accuracy has been discussed in the literature. Turnbull *et al*. (2011) noted that observation errors in wind-speed and planetary boundary layer height led to a factor of two estimated error in CO_2 emissions retrievals. Ryoo *et al*. (2019) suggested that errors associated with wind speed errors and assumptions about background concentrations for greenhouse gases for city-scale emissions sources could contribute emissions retrieval errors for a factor of 1.5 to 7. However, Ryoo *et al*. (2019) also noted that these errors stem largely from the use of flight patterns consisting of downwind screens, and that flight patterns which enclose the sources, (closed shape flight patterns) such as the box flights used in our current work, greatly reduce these errors. Tadic *et al* (2017) noted that cylindrical flight patterns around facilities (similar to the "box" flights described in Gordon *et al*., 2015) provide superior retrieval performance relative to downwind screens. The strategy for determining the maximum height of aircraft flights also varies between references, some recommending sampling up to the PBL height. However, Tadic *et al*. (2017) and Gordon et al. (2015) integrate over the entire vertical extent of the plume (as in the work described here) rather than limit the analysis to below the planetary boundary layer height. Tadic *et al* (2017) estimated the uncertainties in retrievals, calculated from flight patterns using these procedures,**

associated with different factors. Interpolation uncertainties accounted for 21% of the determined emissions value, while wind speed errors of 0.1 m s^{-1} were attributed to -1.2% to +1.5% of the retrieval error, respectively. The latter suggests that the impact of wind speed error when box or cylinder flight patterns such as we use in our current work are relatively small."

Original manuscript text, Introduction, original 2nd paragraph:

"Uncertainties in aircraft-based top-down retrievals may arise from various sources (e.g. measurements, data interpolation and extrapolation), most of which have been studied in the past (e.g. Cambaliza *et al*., 2014; Gordon *et al*., 2015; Angevine *et al*., 2020). Here we focus on investigating the underlying assumption of time-invariant meteorological conditions (during observation time), which is shared among most (if not all) top-down retrieval methodologies, in particular the impact of localized variations in meteorology (e.g. change in atmospheric stability or wind direction) on the application of the mass-balance technique in box flight emission rate retrievals."

Revised manuscript text, Introduction, new 3rd paragraph:

"Uncertainties in aircraft-based top-down retrievals may arise from various sources (e.g. measurements, data interpolation and extrapolation), most of which have been studied in the past (e.g. Cambaliza *et al*., 2014; Gordon *et al*., 2015; Angevine *et al*., 2020). **Prior studies have also noted the potential contribution of temporal trends in tracer mass budget to mass-balance retrievals (e.g. Conley *et al*., 2017), but mainly relied on the assumption of steady-state conditions.** Here we focus on investigating the underlying assumption of time-invariant meteorological conditions (during observation time), which is shared among most (if not all) top-down retrieval methodologies, in particular the impact of localized variations in meteorology (e.g. change in atmospheric stability or direction of the transport) on the application of the mass-balance technique in box flight emission rate retrievals."

Original manuscript text, Introduction, original 3rd paragraph:

"Herein, we refer to these circumstances as "storage-and-release" events."

Revised manuscript text, Introduction, new 4th paragraph:

"Herein, we refer to these circumstances as "storage-and-release" events. **To our knowledge, all mass-balance techniques to date assume steady-state conditions. Here we investigate the uncertainty associated with that assumption, and we introduce the term storage-release (used interchangeably with storage-and-release throughout the paper) events for transient non-steady-state conditions. This term is chosen to emphasize the cyclical nature of the events, during which material within the control volume accumulated and is then subsequently released.**"

2. Equations in Section 2.2.1. There are many equations in this subsection. It is very easy for me to get lost after reading all the equations. For example, some variables in Eq.(6). can be derived from previous equations, some not. I would recommend a flow chart to demonstration the calculation.

AC.2.2

We thank the Reviewer for this comment. Additional clarifications on math notations have been included in the revised version in Sect. 2.2 to address this comment, such as the addition of superscripts "tr" and "tr*" in Equations (6)-(16) to distinguish between the terms needed for the DR method estimates and terms for TR and TR* methods. Also a new table (Table 1, see below) was added to summarize the terms needed for each retrieval method, as a guideline for their use with observation data.

3. The fundamental differences between the three methods have not been clearly stated. I would appreciate a figure or chart to compare them. Additional text to summary the figure/chart is also appreciated. I noticed that there are some texts about this in the conclusion part. But I feel it would be much easier for readers to follow the content if such description appears earlier.

AC.2.3

We thank the referee for the suggestion. In response to this and similar comments (e.g. the reviewer's second point above), a new table (Table 1 in Section 2.2.3) summarizing the main features of the three retrieval methods is added to the revised manuscript.

"**Table 1 summarizes the main features of the three retrieval methods (DR, TR and TR*), along with possible practical/observational applications for each. Meta-data-set types, terms for estimating the source emission rates, descriptions for each method, and their relevant equation numbers are also noted on the table. The horizontal flux and deposition terms $E_{C,H}$ and $E_{C,VD}$ are shared among all three methods, with the other terms approximated for the TR and TR* methods. Note that here $E_{C,VD}$ is "directly" extracted from the model output, but in observational methods this term is also approximated based on the other measured fields (Table 1).**

Table 1. Summary of the three retrieval methods: Direct Retrieval (DR), TERRA Retrieval (TR) approximation and Revised TERRA Retrieval (TR*). See Eqs. (4 –15) and equations in Appendix B for descriptions of each term.

Method	Meta data	Directly derived terms	Approx. terms	Equation numbers	Model-based applications (methods developed here)	Potential observational applications
Direct Retrieval (DR): E_C	4D	$E_{C,H}$, $E_{C,V}$, $E_{C,VD}$, $E_{C,M}$, $E_{C,S}$	—	(B1), (B3), (4), (5), (7)	Comprehensive mass-balance analysis using model output volumetric time-series data.	Aircraft-based retrievals with air-quality model estimates of temporal trends and the impact of storage-release.
TERRA approx. (TR): E_C^{tr}	2D	$E_{C,H}$, $E_{C,VD}$	$E_{C,V}^{tr}$, $E_{C,M}^{tr}$	(B1), (10), (11), (12)	Simulating observation-based mass-balancing by limiting the analysis to data along box walls (screens), no estimate of storage.	Aircraft measurements (standard methods).
Revised TERRA (TR*): E_C^{tr*}	3D	$E_{C,H}$, $E_{C,VD}$	$E_{C,V}^{tr}$, $E_{C,M}^{tr*}$, $E_{C,S}^{tr*}$	(B1), (14), (15), (16)	Improving upon the TR method by estimating the storage term from time-consecutive screen data.	Aircraft measurements with estimate of storage via (a) repeat flights (e.g. in tandem aircrafts), (b) remote vertical profiling (e.g. lidar).

"

Specific comments:

1. Is emission rate retrieval over- and under-estimation a common word to use? It sounds uneasy to follow to me. Would you suggest emission biases? Additionally, the last sentence of the abstract is unclear.

AC.2.4

While the term 'emission rate under/overestimate' is admittedly a bit awkward, we use the phrase since it exactly describes the resulting effect – and distinguishes it from measurement or computational errors or biases. We note that in emission rate retrieval methods, the estimation of the emission rate is the primary goal. The terminology is thus more appropriate than similar terms such as bias since the main interest of the user is to determine the extent to which emission rates have been over or under estimated.

Also, we have decided to remove the last sentence of the originally submitted abstract in the revised manuscript.

2. Page 2, line 35. What does the mass-balance technique refer to here? It is difficult for readers without good knowledge of this specific topic to catch.

AC.2.5

To address this comment the manuscript text was revised as follows:

Original text: "The mass-balance technique and Gauss's divergence theorem can be employed to infer point and area source emission rates from aircraft measured data."

Revised text: "The mass-balance technique and Gauss's divergence theorem **— where the net mass flux of the species of interest exiting a control volume (enclosed by the box flight path) is equated to the within-volume emission rates of that species —** can be employed to infer point and area source emission rates from aircraft measured data."

3. Page 2, line 46. Which model is used here? air quality model?

AC.2.6

Here the authors used the KAMM/DRAIS air-quality model. In response to the Reviewer's comment the manuscript text was revised as follows:

Original text: "Through model simulations, Panitz *et al*. (2002) analyzed …"

Revised text: "Through **air-quality model (KAMM/DRAIS)** simulations, Panitz *et al*. (2002) analyzed …"

4. Page 2, line 50. Please clarify which species has been considered in Gordon *et al*. (2015).

AC.2.7

In response manuscript text was revised as follows:

Original text: "Gordon *et al*. (2015) conducted top-down aircraft-based emission rate retrievals by flying polygonal (e.g. rectangular) box flights around individual oil sands facilities in Alberta, Canada."

Revised text: "Gordon *et al*. (2015) **assessed uncertainties in the top-down aircraft-based emission rate retrievals using** polygonal (e.g. rectangular) box flights measuring SO_2 and CH_4 around individual oil sands facilities in Alberta, Canada."

5. Page 3, line 59. Please clarify the definition of mass transfer.

AC.2.8

In response manuscript text was revised as follows:

Original text: "These and other similar studies, employed either mass transfer or a mass-balance approach (box flights) to infer point and area source emission rates from aircraft measured data."

Revised text: "These and other similar studies, employed either mass transfer **(e.g. the single-screen flight approach where the horizontal mass flux through a downwind vertical plane is equated to the upwind source emission rates)** or a mass-balance approach (box flights) to infer point and area source emission rates from aircraft measured data."

6. Fig 1. The color recognized as dark green in the caption looks like blue from my screen. I recommend using thicker white lines to identify the flight tracks.

AC.2.9

In response Fig. 1 was revised to show flight tracks with thicker white lines and the figure caption was revised as follows:

Original text: "…, shown as a dark green shaded area,"

Revised text: "…, shown as a **dark shaded** area,"

7. Page 6, line 139. Shall "their use" be replaced by "our use" considering the reference is from the same authors?

AC.2.10

In response to this comment the following revision was made to the manuscript text.

Original text: "Earlier work (Fathi, 2017) suggests that the use of regional air-quality models as proxies for observations must be taken with care, with three main considerations in their use for this purpose:"

Revised text: "Earlier work (Fathi, 2017) suggests that **there are three main considerations which should be taken into account when using regional air-quality models as proxies for observations:**"

8. Page 6, line 152. Please clarify the definition of virtual sampling.

AC.2.11

In response the manuscript text was revised as follows:

Original text: "… in virtual sampling with model predicted data"

Revised text: "… in **extraction of emissions retrieval algorithm inputs from 4D air-quality model predicted output**"

9. Page 6, line 158. Please consider changing studied cases to case studies.

AC.2.12

We thank the reviewer for the suggestion. The manuscript text was changed to "**case studies**"

10. Page 21, line 513. The citation of AMS, 2020 is unclear to me.

AC.2.13

The citation was revised to "**(American Meteorological Society - AMS, 2021)**"

Citation: https://doi.org/10.5194/acp-2021-343-RC2

AC.1 - Additional Author Comment

Please also note that we have corrected a minor mistake in the original manuscript, in the information presented in Figure 6, Section 3.3, and the manuscript text referring to this Figure, as we show below. Note that this correction presents the more accurate (and stronger) correlation between the 4D and 3D estimates of temporal trends in tracer mass. Note that this correction has no impact on the results presented in the manuscript.

In original manuscript, section 3.3, Figure 6 caption:

"… The least-squares fit line has slope m=2.98 and intercept b=0.09. Note that case 8 estimates are in low correlation due to relatively high upwind emissions during the time of that particular flight (August 28th), rendering the 3D estimates less representative of the 4D estimates. Excluding Case 8 from this analysis, results in P_r=0.88. $E_{C,S}^{*}$ can be estimated from multiple consecutive 2D screens, by assuming 3D estimations of $\partial\chi_C/\,\partial t$ as representative of the rate of change in species mass within the flux box."

In revised manuscript, section 3.3, Figure 6 caption:

"… The least-squares fit line has slope m=**1.36** and intercept b=0.09. Note that case 8 estimates are in low correlation due to relatively high upwind emissions during the time of that particular flight **case** (August 28th), rendering the 3D estimates less representative of the 4D estimates. Excluding Case 8 from this analysis, results in P_r=0.88**, m=1.07 and b=0.03**. $\boldsymbol{E_{C,S}^{tr*}}$ can be estimated from multiple consecutive 2D screens, by assuming 3D estimations of $\partial\chi_C/\,\partial t$ as representative of the rate of change in species mass within the flux box. **The same can also be estimated from downwind vertical profiling of tracer concentrations over the sampling time.**"

In original manuscript, section 3.3, 1st paragraph:

"$\partial\chi_C/\,\partial t$ range is higher for 4D estimates (2.52 ppbv/hr) than for 3D estimates (0.98 ppbv/hr), but the two variables are correlated. The correlation coefficient $P_r = 0.7$ (with slope $m = 2.98$ and $b = 0.09$ intercept for the least-square fit line), indicates strong correlation between 4D and 3D estimates. The exception being rejected case 8, where the correlation is low due to relatively high upwind emissions during the time period of this case (on August 28th). Excluding case 8 from this analysis, increases the correlation between 3D and 4D estimates to $P_r = 0.9$"

In revised manuscript, section 3.3, 1st paragraph:

"**The** $\partial \chi_C / \partial t$ range is **slightly** higher for 4D estimates (2.52 ppbv/hr) than for 3D estimates (**2.17** ppbv/hr), **and** the two variables are **strongly** correlated. The correlation coefficient $P_r = 0.7$ (with slope $m = \mathbf{1.36}$ and $b = 0.09$ intercept for the least-square fit line), indicates strong correlation between 4D and 3D estimates. **An** exception **is** rejected case 8, where the correlation is low due to relatively high upwind emissions during the time period of this case (on August 28th). Excluding case 8 from this analysis, increases the correlation between 3D and 4D estimates to $P_r = 0.9$ **(with $\boldsymbol{m = 1.07}$ and $\boldsymbol{b = 0.03}$).**"

Evaluating the Impact of Storage-and-Release on Aircraft-based Mass-Balance Methodology Using a Regional Air Quality Model

Sepehr Fathi *et al.*

ACPD – Response to Referee Comments

Manuscript number: ACP-2021-343

Response to Anonymous Referee #3

Please note the different font/colors in this document for:

1. Referee Comments (RC) in black,
2. Author Comments (AC) in light blue,
3. **Revised and/or added text quoted from the revised manuscript in bold dark font**

RC3: 'Comment on acp-2021-343', Anonymous Referee #3, 22 Jul 2021
Manuscript number: ACP-2021-343

Review of the manuscript entitled "Evaluating the Impact of Storage-and-Release on aircraft-based Mass-Balance Methodology Using a Regional Air Quality Model" by Sepehr Fathi *et al*.

General Comments:

This paper introduces the storage-release events in the mass-balance aircraft top-down methods. Using Global Environmental Multiscale-Modeling Air-Quality and CHemistry (GEM-MACH) model, the authors present the storage-release events (The author shows the transient storage of the emitted mass within box volume and its later release) can contribute to the overestimation or underestimation of emission rate, up to (-29 – +156%). When certain meteorological conditions are set up (e.g. when temporal and spatial variations in meteorological conditions, so the source emission rate result in mass imbalance through box top and lateral walls and the deposition (storage) to ground.

They also introduced three productive parameters to forecast the storage-release events, such as 1) Richardson number (atmospheric stability/shear), 2) Plume Shift in the transport direction, and 3) Weighted upwind to downwind concentration ratio, emphasizing the potential practical use for both aircraft and model-based forecasts of their parameters.

The paper reads well and designs well based on the thorough consideration of the current issues associated with the mass-balance approach, and potential sources of the uncertainties.

However, there are a few points that would have been considered.

AC.3: We thank **Anonymous Referee #3** for providing constructive feedback and suggestions for improving the manuscript (https://doi.org/10.5194/acp-2021-343-RC3). Below we provide specific responses to each comment.

1. First, while the authors kept mentioned that the emission estimate can be highly affected by the meteorological conditions, they didn't show the actual wind conditions for each case. Furthermore, the authors also didn't consider how the emission estimate can be significantly changed depending on how to handling wind data [Turnbull *et al*., 2011; Tadic *et al*., 2017; Ryoo *et al*., 2019]. The fluxes can be also varied depending on what planetary boundary layer height (PBLH) is used. Model PBLH is known to be incorrect, so please specify how you handle this PBLH for flux calculation in your GEM-MACH data.

AC.3.1

The Reviewer raises several points within main point 1, we address each of these in turn.

Re: "while the authors kept mentioning that the emission estimate can be highly affected by the meteorological conditions, they didn't show the actual wind conditions for each case":

In this study we referred to the TERRA method (Gordon *et al*. 2015) as an example retrieval method (where handling measured wind data is described in detail) and considered cases from the 2013 JOSM campaign. The cases we considered were "chosen *a priori* for relatively constant wind speeds at each height" (see Sect. 4.2) and the meteorological conditions for our nine studied cases are summarized in Table 3. We also added average wind speed data to Table 3 in the revised manuscript.

"

Table 3. A summary of the meteorological conditions and source emission scenarios by the three oil sands facilities for the period of the nine JOSM 2013 box flight cases in our GEM-MACH model simulations. Flight date-time and duration is provided for each case, LT stands for local time (UT−6hr). The flight time average $\overline{R_i}$, wind speed and direction at plume height are shown. The three forecast parameters are also provided: $\Delta_t R_i$, $\Delta_t \alpha$ and (ϕ). DR, TR and TR* method estimates are compared to model input emissions (MIE). Storage (S) and/or release (R) events, their order of occurrence (for the cases where both were present) and their relative magnitude (represented by font-size) are given. The flight time average storage rate as a fraction (%) of the total retrieved emission rates are also provided for each case. The performance of the three retrieval methods against MIE, are shown in terms NM and NRMS errors.

Facility	Syncrude			CNRL			Suncor		
Case ID	1	2	3	4	5	6	7	8	9
Flight Date &	Aug 14	Aug 24	Sep 3	Aug 20	Aug 26	Sep 2	Aug 22	Aug 28	Aug 29
Start Time (LT)	10:12	11:50	14:58	10:30	13:43	11:43	11:45	11:18	15:01
Duration (hh:mm)	1:18	2:06	2:07	2:10	1:52	1:45	2:58	2:30	2:13
Atmos. Conditions	stable	unstable	unstable	unstable	stable	unstable	unstable	unstable	unstable
$\overline{R_i}$	0.26	-0.03	-0.66	-0.14	1.01	-0.26	-0.19	-0.44	-0.58
avg. Wind Direction	200°	264°	215°	259°	243°	100°	246°	297°	55°
avg. Wind Speed (m/s)	4.3	2.0	4.9	10.9	0.8	6.9	11.4	3.2	2.0
$\Delta_t R_i$ (h^{-1})	-0.38	-0.13	0.19	-0.25	0.62	-0.28	-0.26	-0.34	0.15
$\Delta_t \alpha$ (h^{-1})	3.3°	6.5°	1.3°	1.8°	45.8°	33.9°	2.4°	3.3°	11.8°
ϕ (%)	0.244	2.766	0.167	0.004	10.97	6.725	5.992	13.04	0.137
avg. Model Input	3423	4050	6923	2979	120	83	1467	1273	1013
Emissions (MIE)	kg/hr	kg/hr	kg/hr	kg/hr	kg/hr	kg/hr	kg/hr	kg/hr	kg/hr
Storage and/or Release	S-R	S	R-S	R-S	S	R	R-S	S	R
Net Contribution	17 %	20 %	2 %	-3 %	43 %	-29 %	13 %	156 %	-27 %
Norm. Mean Error:									
DR (E_C) vs. MIE	-0.07	0.04	-0.00	-0.01	0.11	-0.00	-0.14	0.22	-0.08
TR (E_C^{tr}) vs. MIE	-0.22	-0.16	-0.02	0.02	-0.36	0.24	-0.25	-1.66	0.17
TR* (E_C^{tr*}) vs. MIE	-0.11	-0.07	-0.01	-0.00	-0.26	0.05	-0.14	-0.96	0.11
Norm. RMS Error:									
DR (E_C) vs. MIE	0.10	0.04	0.01	0.02	0.12	0.07	0.14	0.23	0.08
TR (E_C^{tr}) vs. MIE	0.23	0.18	0.04	0.06	0.39	0.25	0.28	1.66	0.17
TR* (E_C^{tr*}) vs. MIE	0.16	0.08	0.02	0.03	0.34	0.07	0.17	1.14	0.15

"

Also, as we explain in the revised paper, section 2.1, "**We note that this work is not intended as an evaluation of the GEM meteorological model and its components, which have been extensively evaluated elsewhere in the literature (Belair *et al*. 2003a,b; Cote *et al*. 1998; Fillon *et al*. 2010; Girard *et al*. 2014; Li and Barker 2005; Milbrandt *et al*. 2005a,b; Milbrandt and Morrison 2016). A recent detailed evaluation of GEM-MACH's meteorological performance is also provided in Makar *et al.* (2021). Rather, we use the model as a proxy for observations since it can provide instantaneous 3D values for chemical concentrations and meteorological variables, which in turn allow us to construct parameters that can be used to predict storage-and-release events in observation-based applications.**"

Re: "the authors also didn't consider how the emission estimate can be significantly changed depending on how to handling wind data [Turnbull *et al*., 2011; Tadic *et al*., 2017; Ryoo *et al*., 2019]".

We have included a discussion of the potential impact of measurement errors on our revised paper's Introduction, making use of the Reviewer's references:

In revised manuscript, Introduction, 2nd paragraph:

"**The potential for measurement error to affect retrieval accuracy has been discussed in the literature. Turnbull *et al*. (2011) noted that observation errors in wind-speed and planetary boundary layer height led to a factor of two estimated error in CO_2 emissions retrievals. Ryoo *et al*. (2019) suggested that errors associated with wind speed errors and assumptions about background concentrations for greenhouse gases for city-scale emissions sources could contribute emissions retrieval errors for a factor of 1.5 to 7. However, Ryoo *et al*. (2019) also noted that these errors stem largely from the use of flight patterns consisting of downwind screens, and that flight patterns which enclose the sources, (closed shape flight patterns) such as the box flights used in our current work, greatly reduce these errors. Tadic *et al* (2017) noted that cylindrical flight patterns around facilities (similar to the "box" flights described in Gordon *et al*., 2015) provide superior retrieval performance relative to downwind screens. The strategy for determining the maximum height of aircraft flights also varies between references, some recommending sampling up to the PBL height. However, Tadic *et al*. (2017) and Gordon *et al.* (2015) integrate over the entire vertical extent of the plume (as in the work described here) rather than limit the analysis to below the planetary boundary layer height. Tadic *et al* (2017) estimated the uncertainties in retrievals, calculated from flight patterns using these procedures, associated with different factors. Interpolation uncertainties accounted for 21% of the determined emissions value, while wind speed errors of 0.1 m s^{-1} were attributed to -1.2% to +1.5% of the retrieval error, respectively. The latter suggests that the impact of wind speed error when box or cylinder flight patterns such as we use in our current work are relatively small.**"

Re: "The fluxes can be also varied depending on what planetary boundary layer height (PBLH) is used. Model PBLH is known to be incorrect, so please specify how you handle this PBLH for flux calculation in your GEM-MACH data"

Please note that, as recommended by Tadic *et al* (2017) and Gordon *et al* (2015), the box flight patterns used for flux calculations extend far above the PBL height in both model and observations in order to reduce retrieval error. The ability of the meteorological model used here as a proxy for observations to simulate the PBL height accurately, thus becomes irrelevant to the accuracy of the retrieval algorithm.

2. Second, the storage and release terms seem to be determined by the rate of compound mixing ratio change with time and the temporal variations in air density with time. The assumption underneath here is that the gas concentration change with time during flux measurement (normally 30-2 hours), and the values are significant enough to change the emission rate. But this may be scale-dependent. For example, for the local scale, the variation of the mixing ratio can be small. It is great for authors to suggest "universal" methods to consider all the factors to influence the emission estimate in all possible meteorological conditions. But does this "storage" term really include the surface-emission flux which can be highly varied with extrapolation methods when we use aircraft data in practice [Gordon *et al*., 2015]?

AC.3.2

Re: "Second, the storage and release terms seem to be determined by the rate of compound mixing ratio change with time and the temporal variations in air density with time. The assumption underneath here is that the gas concentration change with time during flux measurement (normally 30-2 hours), and the values are significant enough to change the emission rate."

It would be more accurate, and as already described in the manuscript, to say that storage and release is characterized by three factors: changes in atmospheric stability (via the gradient Richardson number), changes in direction of the transport, and the standard deviation weighted upwind to downwind concentration ratio.

Please note that the storage rate term $E_{C,S}$, as defined in our Eq. (4), refers to the net change in the accumulated (tracer) mass within the control volume. This change in mass and therefore change in compound mixing ratio, is due to a transient imbalance between the addition of mass due to source emissions and the removal of mass due to all the other processes (e.g. advection, deposition). Therefore, $E_{C,S}$ will be non-zero only during non-steady state conditions when the addition of mass due to source emissions is not balanced out with the removal of mass due to processes such as advection and deposition. This is not to be confused with changes due to chemistry ($E_{C,X}$) which is negligible relative to the source emission rates considered here.

Further, we have revised the sentence leading to Eq. (4) in the revised manuscript as follows,

Original manuscript text: "The first integral in Eq. (3) represents the rate of change in compound mass due to variations in compound mixing ratio $\chi_C(t, V)$ independent of the changing air density"

Revised manuscript text: "The first integral in Eq. (3) represents the rate of change in compound mass **characterized as** variations in compound mixing ratio $\chi_C(t, V)$ independent of the changing air density"

Re: "But this may be scale-dependent. For example, for the local scale, the variation of the mixing ratio can be small."

Our third metric of storage and release (the standard deviation weighted upwind to downwind concentration ratio) provides an indication of whether the magnitude and variation in mixing ratio both entering and leaving the box containing the source is sufficient to influence emission retrieval accuracy.

Re: "It is great for authors to suggest "universal" methods to consider all the factors to influence the emission estimate in all possible meteorological conditions. But does this "storage" term really include the surface-emission flux which can be highly varied with extrapolation methods when we use aircraft data in practice [Gordon *et al.*, 2015]?"

As we describe in the manuscript, one of the advantages of our use of a 3D regional transport air-quality model as a proxy for observations is that all of the terms influencing emissions flux estimation are fully characterized. The emissions from the facility of interest are known in advance as well as the meteorological influence of storage and release – so the surface emissions flux is definitely included in the calculations, as are the meteorological terms which may hamper its accurate retrieval.

The uncertainties in determining the emission rate associated with extrapolation methods are discussed in section 3.2 of Gordon *et al*. (2015). Errors associated with the extrapolation using different techniques – with the choice of which technique to use being dependent on the concentration profiles and the type of source – ranged from 0.25 to 19.2%, with the error associated with the use of an extrapolation method without knowledge of the prior behavior of the plume estimated at most 20%.

The storage term includes the surface-emission flux in the sense that these emitted fluxes might then gather within the volume and then get released later under non-steady-state conditions. The determination of the surface emission flux using extrapolation methods can then lead to the above mentioned uncertainties of up to 20%; however, these uncertainties are independent of the uncertainties associated with storage-release events. Hence the uncertainties associated with storage-release events must be added to all other measurement uncertainties in real-life field campaigns.

We note for context that the range of uncertainties associated with extrapolation is smaller than the range of uncertainties derived for the significant storage and release cases (5 and 8) in the current work.

3. Third, the actual data processing is not well described. While mathematical terms are well described, it is questionable whether this can help interpreting field data since all the terms used in this study are obtained directly from the model (e.g. surface deposition rate). Furthermore, the model data may not need to do additional interpolation/extrapolation as air-craft data does, so the uncertainty associated with interpolation can be lower. However, emission estimate methods based on mass balance using aircraft data often adapt the interpolation (e.g. kriging or exponential) and extrapolation, and the values are highly varying depending on what method you use. Since model data is gridded, and somewhat averaged data, data uncertainty is already included in model output. In this way, how can we trust the storage or surface deposition rate from the model is correct? Comparing the model input emission input with the data methods are good enough for trust model-based calculation of storage value? In addition, how do you know the storage and release terms can be obtained "along" the lateral walls of box-shape (i.e. closed shape) flight over measurement time, not throughout the bottom area of the box?

AC.3.3

Re: "Third, the actual data processing is not well described. While mathematical terms are well described, it is questionable whether this can help interpreting field data since all the terms used in this study are obtained directly from the model (e.g. surface deposition rate)."

As we noted in the manuscript, the main advantage and reason for the use of the regional air-quality model as a proxy for observations is that all of the factors affecting the concentration can be described exactly, hence allowing the creation of metrics allowing the *a priori* prediction of the extent to which storage and release may influence emissions prediction. The resulting three metrics may be easily determined from aircraft observations.

Re: "Furthermore, the model data may not need to do additional interpolation/extrapolation as air-craft data does, so the uncertainty associated with interpolation can be lower. However, emission estimate methods based on mass balance using aircraft data often adapt the interpolation (e.g. kriging or exponential) and extrapolation, and the values are highly varying depending on what method you use"

Please see our above response regarding the reference to Gordon *et al* (2015), and Tadic *et al* (2017) – interpolation and extrapolation errors have been investigated in past work and have been found to have an impact of at most 20% and 21%, and that specifically with respect to surface (area)-based sources (Gordon *et al*., 2015), with the error for the emissions from high temperature plumes lofted above the minimum aircraft flying altitude being much smaller. Similarly, assumptions regarding surface deposition rates were found to have a minimal impact on predicted emissions in that previously published work. In response to another reviewer's suggestion, we have also noted in the revised text that vertical profiling instrumentation such as air-borne LIDAR can be used to provide very similar and near-instantaneous data to that provided by the air-quality model, in flights enclosing a facility.

See Introduction, end of 1st paragraph: "**Gordon *et al*. (2015) estimated uncertainty in aircraft-based retrievals as 2%-30%, with the source of this uncertainty attributed mainly to extrapolation of aircraft measurements below the lowest flight track (20%).**", and the added text in the new 2nd paragraph in revised version. In the revised manuscript, we have also noted that the use of vertical profiling instrumentation such as LIDAR observations is one means by which column measurements may be obtained via a single flight around a facility, creating near-instantaneous enclosing screens similar to that derived from the 3D model used here (see sections 2.2.3, 3.3 and Conclusions) – please see our response to the 4th point raised by the Reviewer (AC.3.4) for the text mentioning this approach.

Re: "Since model data is gridded, and somewhat averaged data, data uncertainty is already included in model output."

Yes, this is one of the reasons why the model may be used to determine the relative impact of storage and release and generate metrics for its estimation from observation data – since all of the terms are known exactly.

Re: "In this way, how can we trust the storage or surface deposition rate from the model is correct?"

As noted above, the interpolation or extrapolation methods were shown in previous work to have a small impact (0.2 to 20%) on estimated emissions, while we have shown here (cases 5 and 8) that the impact of storage and release may be considerably larger than the upper limit of 20% The same previously published work showed that the impact of deposition ranged from 2.6 to 6.7% of the estimated emissions, depending on the species examined (also, for elevated, non-fumigating stack sources, the deposition rate can be assumed to be zero). The influence of interpolation, extrapolation and deposition has already been determined, in previously published work, to be minimal, relative to the magnitude of the storage and release as calculated here.

It is noted again that these uncertainties are all independent of the uncertainties associated with storage and release. In an actual field campaign, the storage and release uncertainties would have to be considered in addition to these other uncertainties.

Re: "Comparing the model input emission input with the data methods are good enough for trust model-based calculation of storage value?"

Based on our ability to completely characterize the factors influencing emissions using a 3D regional transport model, and on the relatively small magnitude of interpolation, extrapolation, and deposition compared to both recovered emissions and storage and release, our answer is 'Yes".

Re: "In addition, how do you know the storage and release terms can be obtained "along" the lateral walls of box-shape (i.e. closed shape) flight over measurement time, not throughout the bottom area of the box?"

The objection the Reviewer is raising here is unclear. The use of observations to generate lateral wall fluxes and hence estimate emissions has been demonstrated in previous papers both referenced by ourselves and by the Reviewer (e.g. Tadic *et al*., 2017; Gordon *et al* 2015). Extrapolation for the bottom of the box (we assume that the Reviewer means "below flight level" here by "bottom area") has been previously demonstrated to have a 0.2 to 20% impact on emissions results, smaller than the impacts of storage and release events demonstrated here. The uncertainties of the storage and release term are independent of these other uncertainties.

4. Finally, the suggestion of multiple flights or several loops with one flight is great and sounds ideal. However, given aircraft constraints, is this realistically feasible for most cases? Some redesign will be necessary to satisfy this.

AC.3.4

We thank the Reviewer for this positive comment. The feasibility will of course depend on study budgets. The addition of a second aircraft would of course double study costs, assuming the same instrumentation is used on both aircraft. However, alternative strategies are possible, as we outline in the revised manuscript, including the use of column observations such as ground-based or airborne LIDAR aboard a single aircraft to estimate storage and release for specific chemical constituents. The text has been modified as follows:

Revised version: Section 2.2.3, 2nd paragraph

"**Considering the potentially prohibitive operational cost of conducting repeat flights, and the fact that few (e.g. 2-3) time-consecutive measurements may not provide enough information for the estimate of temporal trends in within-box tracer mass budget, the following alternative is proposed. As we discuss later in Sect. 3.3, observed temporal trends of tracer concentrations ($\partial \chi_C / \partial t$) downwind of the emission source are also in similar strong correlation ($P_r = 0.9$) with the corresponding DR estimates (using volumetric time-series). Furthermore, our studied cases suggest that downwind vertical-profiling of tracer concentrations can provide similar estimates of the temporal tends required for the estimate of the storage rate $E_{C,S}^{tr*}$ and can be substituted in Eq. (15) for $\partial \overline{\chi_C}(t,z) / \partial t$. In observational applications, this can be achieved through ground-based or, preferably, aircraft-based vertical profiling remote measurements (e.g. LIDAR measurements, Aggarwal *et al.*, 2018), where the sampling aircraft with a remote measurement setup would collect column data while flying around or downwind of the emission source. The use of an aircraft-based vertical profiling instrument for a small number of chemical fields, along with a repeated, single-loop flight path around a facility (10-15 min) would generate fields similar to those generated by the air-quality model used here, in turn allowing highly time-resolved estimates of the storage term. This alternative approach can be more cost (operational) efficient compared to the strategy of repeat flights (which would require a second aircraft travelling in tandem behind the first to achieve the same time resolution), while providing more time-consecutive data points for the study of the temporal trends in the tracer mass budget, and further reducing the relatively small emissions biases associated with extrapolation to the ground of observed fields.**"

Revised version: Section 3.3, mid-paragraph

"**As discussed in Sect. 2.2.3, an alternative to multiple aircraft-based measurements is to estimate the temporal trends in tracer concentrations ($\partial\chi_C/\partial t$) through remote vertical-profiling (e.g. LIDAR measurements). Our analysis of different flight cases show that downwind trends are also in strong correlation with DR method volumetric time-series estimates with the same correlation coefficient of $P_r = 0.9$ (with different $m = 0.57$ and $b = 0.03$). The advantage of remote (downwind and upwind) vertical-profiling over multiple aircraft or UAV measurements, in addition to operational cost efficiency, would be the collection of more temporal measurements over the sampling time, which in turn can result in improved estimates of $\partial\chi_C/\partial t$ and the storage term ($E_{C,S}^{tr*}$); as opposed to estimates based on few time-consecutive measurements (e.g. 2-3) via multiple or in tandem aircraft.**"

Revised version: Section 3.3, end of paragraph

"Again, for **observational applications: (1)** a repeat box flight procedure can be used, where an aircraft carries out a second box flight immediately after the first flight, or two aircraft (or UAVs) follow the same flight path in tandem, one positioned a fixed distance behind the other**, or (2) remote vertical-profiling (e.g. remote LIDAR measurements) on a single aircraft can be employed to collect time-consecutive measurements of relevant fields, in the column, around the emitting facility.**"

Revised version: Conclusions section, end of last paragraph

"We have also devised a methodology to reduce the impact of this form of emissions retrieval error by estimating the storage rate ($E_{C,S}^{tr*}$) through **(a)** the use of repeat flights, around the same facility, either with a single or multiple aircraft(s), **or (b) aircraft-based remote vertical-profiling of relevant fields during the sampling period.**"

Specific comments:

- Line 84: "Storage-and Release" events sound new. Is this new terminology used in this study?

  AC.3.5

  Yes, "Storage-and-Release events" is new terminology introduced in this study. The following text was added to the revised manuscript in the Introduction section, 4th paragraph:

  "Herein, we refer to these circumstances as "**storage-and-release**" events. **To our knowledge, all mass-balance techniques to date assume steady-state conditions. Here we investigate the uncertainty associated with that assumption and we introduce the term storage-release (used interchangeably with storage-and-release throughout the paper) events for transient non-steady-state conditions. This term is chosen to emphasize the cyclical nature of the events, during which accumulation of material within the control volume is then subsequently released.**"

- Line 187: Good approach. However, can't we just say that the "strorage" term just represents the surface flux and entrainment through the top and bottom flux?

AC.3.6

No. The point here is that the surface and top fluxes are already characterized and that storage and release constitutes a previously unquantified potential source of emissions retrieval error. Note that, surface flux and entrainment are included in $(E_{C,out} - E_{C,in})$ in Eq. (2).

- Figure 7: This is a nice illustration. However, this storage term cannot be measured by aircraft if we only consider "along-track change". How to validate the model storage term? Comparing MIE with estimates from three different methods are enough? I am not sure how applicable this method will be in a practical way.

AC.3.7

The strategies of repeat flights and/or downwind vertical-profiling have been discussed as potential practical/observational approaches for estimating the storage term (Sect. 2.2.3). Model-based estimates of the storage term may be validated with field observations such as controlled tracer release (see Sect. 2.2.3, last paragraph before Eq. 16)

- The mathematical formulas in general: The mathematical notations are useful, but ironically these are also very confusing as well. I liked the detailed description of the mathematical formulas, but I often lost and forgot what each abbreviation referred to. Could you please be more specific in the notation? Maybe adding more words in the notation will be helpful for readers to understand what each term refers to, keeping them from looking back at the previous pages.

AC.3.8

This was a good point – to improve the clarity of the manuscript and make it more accessible to readers, we added superscripts tr and tr* in sections 2.2.2 and 2.2.3 and throughout to address this comment. Also added a new table (Table 1, see below) which summarizes the computational steps, and equations needed for emission rate retrieval for each method.

"**Table 1 summarizes the main features of the three retrieval methods (DR, TR and TR*), along with possible practical/observational applications for each. Meta-data-set types, terms for estimating the source emission rates and descriptions for each method are also noted on the table. The horizontal flux and deposition terms $E_{C,H}$ and $E_{C,VD}$ are shared among all three methods, with the other terms approximated for the TR and TR* methods. Note that here $E_{C,VD}$ is "directly" extracted from the model output, but in observational methods this term is also approximated based on the other measured fields (Table 1).**

Table 1. Summary of the three retrieval methods: Direct Retrieval (DR), TERRA Retrieval (TR) approximation and Revised TERRA Retrieval (TR*). See Eqs. (4 –15) and equations in Appendix B for descriptions of each term.

Method	Meta data	Directly derived terms	Approx. terms	Equation numbers	Model-based applications (methods developed here)	Potential observational applications
Direct Retrieval (DR): E_C	4D	$E_{C,H}$, $E_{C,V}$, $E_{C,VD}$, $E_{C,M}$,$E_{C,S}$	—	(B1), (B3), (4), (5), (7)	Comprehensive mass-balance analysis using model output volumetric time-series data.	Aircraft-based retrievals with air-quality model estimates of temporal trends and the impact of storage-release.
TERRA approx. (TR): E_C^{tr}	2D	$E_{C,H}$, $E_{C,VD}$	$E_{C,V}^{tr}$, $E_{C,M}^{tr}$	(B1), (10), (11), (12)	Simulating observation-based mass-balancing by limiting the analysis to data along box walls (screens), no estimate of storage.	Aircraft measurements (standard methods).
Revised TERRA (TR*): E_C^{tr*}	3D	$E_{C,H}$, $E_{C,VD}$	$E_{C,V}^{tr}$, $E_{C,M}^{tr*}$, $E_{C,S}^{tr*}$	(B1), (14), (15), (16)	Improving upon the TR method by estimating the storage term from time-consecutive screen data.	Aircraft measurements with estimate of storage via (a) repeat flights (e.g. in tandem aircrafts), (b) remote vertical profiling (e.g. lidar).

”

Citation: https://doi.org/10.5194/acp-2021-343-RC3

AC.3 - Additional Author Comment

Please also note that we have corrected a minor mistake in the original manuscript, in the information presented in Figure 6, Section 3.3, and the manuscript text referring to this Figure, as we show below. Note that this correction presents the more accurate (and stronger) correlation between the 4D and 3D estimates of temporal trends in tracer mass. Note that this correction has no impact on the results presented in the manuscript.

In original manuscript, section 3.3, Figure 6 caption:

“... The least-squares fit line has slope m=2.98 and intercept b=0.09. Note that case 8 estimates are in low correlation due to relatively high upwind emissions during the time of that particular flight (August 28th), rendering the 3D estimates less representative of the 4D estimates. Excluding Case 8 from this analysis, results in P_r=0.88. $E_{C,S}^{*}$ can be estimated from multiple consecutive 2D screens, by assuming 3D estimations of $\partial\chi_C/\partial t$ as representative of the rate of change in species mass within the flux box.”

In revised manuscript, section 3.3, Figure 6 caption:

"… The least-squares fit line has slope m=**1.36** and intercept b=0.09. Note that case 8 estimates are in low correlation due to relatively high upwind emissions during the time of that particular flight **case** (August 28th), rendering the 3D estimates less representative of the 4D estimates. Excluding Case 8 from this analysis, results in P_r=0.88**, m=1.07 and b=0.03**. $\boldsymbol{E_{C,S}^{tr*}}$ can be estimated from multiple consecutive 2D screens, by assuming 3D estimations of $\partial\chi_C/\,\partial t$ as representative of the rate of change in species mass within the flux box. **The same can also be estimated from downwind vertical profiling of tracer concentrations over the sampling time.**"

In original manuscript, section 3.3, 1st paragraph:

"$\partial\chi_C/\,\partial t$ range is higher for 4D estimates (2.52 ppbv/hr) than for 3D estimates (0.98 ppbv/hr), but the two variables are correlated. The correlation coefficient $P_r = 0.7$ (with slope $m = 2.98$ and $b = 0.09$ intercept for the least-square fit line), indicates strong correlation between 4D and 3D estimates. The exception being rejected case 8, where the correlation is low due to relatively high upwind emissions during the time period of this case (on August 28th). Excluding case 8 from this analysis, increases the correlation between 3D and 4D estimates to $P_r = 0.9$"

In revised manuscript, section 3.3, 1st paragraph:

"**The** $\partial\chi_C/\,\partial t$ range is **slightly** higher for 4D estimates (2.52 ppbv/hr) than for 3D estimates (**2.17** ppbv/hr), **and** the two variables are **strongly** correlated. The correlation coefficient $P_r = 0.7$ (with slope $m = \mathbf{1.36}$ and $b = 0.09$ intercept for the least-square fit line), indicates strong correlation between 4D and 3D estimates. **An** exception **is** rejected case 8, where the correlation is low due to relatively high upwind emissions during the time period of this case (on August 28th). Excluding case 8 from this analysis, increases the correlation between 3D and 4D estimates to $P_r = 0.9$ **(with $\boldsymbol{m = 1.07}$ and $\boldsymbol{b = 0.03}$).**"