We are very grateful to the three reviewers for their insightful comments and suggestions. Many identical comments were made by all three reviewers, which demonstrates the necessity of addressing these concerns. We have attempted to incorporate each comment and suggestion into a revised manuscript. The most significant changes include:

- Inclusion of more information on measurement uncertainty and frequency.
- Recommendations in the conclusions section.
- A reorganization of the discussion of uncertainties, including the removal of the old Table 7 and 8 and the addition of a new Table 7.

The comments below make many references to the revised manuscript, using the new section, table, and figure numbers. We have avoided using page and line numbers, which will change in the revised manuscript.

#### Anonymous Referee #1

Received and published: 8 June 2015

#### **General Comments**

The authors demonstrate an aircraft mass-balance algorithm based emission quantification technique that accounts for advective and turbulent flux in three dimensions. Two sample flights are analyzed using the proposed technique. The authors present a useful comparison of previous techniques as reported in the literature and the relative strengths and weaknesses of each technique. The subject and novelty of this article is appropriate for this journal.

Overall this manuscript is clearly written and provides a useful framework towards a consistent and rigorous aircraft emission quantification sampling and calculation routine. The technique provides clear benefits as illustrated by the disaggregation of the flux into 3-D components highlighting flux components that are not usually captured with some current techniques. My concern is that this technique as described is not as directly translatable to other emission quantification operations as may be implied by the paper. The authors analyze relatively ideal case studies to showcase the improvements possible from their technique, which is useful, but do less to identify the specific situations (in terms of both types of emission sources and meteorology) this technique is most suitable for and how this may be adapted to non-ideal situations.

This point was picked up by all reviewers (R1.6, R2.3, R2,8, R3.3), so a new paragraph is added to the end of the Conclusions section (now Conclusions and Recommendations). This paragraph provides suggestions to apply the method to future projects and how to reduce uncertainties.

#### Specific Comments

4772 Line 26: The computation of the contribution of advective transport may, but does not necessarily indicate lower uncertainty compared to other approaches, though it certainly indicates a source of error that is not usually included. Furthermore, author derived uncertainties are not calculated uniformly and do not necessarily include natural limitations of each sampling method. Some of these estimates are purely on the basis of measured variability and do not assess the influence of the assumptions made to complete the calculation. For instance Cambaliza et al. 2014 showed that individual single transect results from transects

at different heights collected on the same day could differ by over 100% downwind of a source. Truly lower uncertainty can only be accomplished by improved sampling methods, regardless of how others report uncertainty. Please clarify the uncertainties reported are author derived and do not follow consistent and necessarily comparable protocols.

R1.1) All the uncertainties are now described as "author-derived". The following text is added to the end of the second paragraph: "In comparing method uncertainties it is noted that different authors use inconsistent methodologies to estimate total uncertainties, and some estimates are more conservative than others. Hence the relative values of author-derived uncertainties in this section is considered a qualitative comparison only."

4774 line 24: Are results from these two flights expected to be representative of all flights? Was meteorology a consideration for the flights chosen for this analysis? Please include a summary of meteorological conditions (wind speed, direction, stability) in the text.

R1.2) A paragraph is also added at the end of Section 2.2 with a summary of measured wind speed, direction, and estimated stability for each flight (as determined from the WBEA towers). A paragraph is also added to the Conclusions Section (second last paragraph) with suggestion for future studies.

4778 line 22: Interpolation resolution should be determined by the physical resolution of the instrumentation. Please include the sampling frequency on your instruments and the airspeed of the aircraft. I believe this Picarro model has 5s sampling frequency and the cruising speed of this aircraft is \_125 m/s which, if correct, would mean the best resolution possible for CH4 is 625 m. Variations smaller than this (or whatever the true resolution is) cannot be observed by the aircraft.

R1.3) Section 2.1 is rewritten to clear up some misleading information and to add measurement frequency (SO<sub>2</sub> instrument is 1 Hz, Picarro is 0.3 Hz). A new paragraph is added: "To consolidate the various measured parameters, high-frequency data (wind and state parameters) were averaged to a frequency of 1 Hz, while low-frequency data (CH<sub>4</sub>) were linearly interpolated to a frequency of 1 Hz. The Convair has a cruising speed of 90 m s<sup>-1</sup>, which gives a spatial resolution of 90 m for SO<sub>2</sub> and 270 m for CH<sub>4</sub>." (See also comment R3.5).

While it is true that this measurement resolution is much greater than  $\Delta s = 40$  m and  $\Delta z = 20$  m, the measurements are not in a perfect grid shape and there is a large amount of deviation of the flight paths from regular grid coordinates (e.g. Fig's 2, 3, 4, 5). This means that some measurement locations are much closer together than 90 m or 270 m (especially in the vertical coordinate). It is also standard practice when interpolating to choose a resolution smaller than the sampling interval.

4782 Section 3.1: The methods to interpolate near the surface for some of these parameters (wind speed and air density) would theoretically be valid for the entire mixed boundary layer and potentially provide a better estimation than any interpolation method. Did the authors consider using this technique for the entire mixed boundary layer? Were the results different?

R1.4) Although it is not included in the manuscript, the wind speed measurements demonstrate some degree of wind shear, which would not be modelled by Eq. 8 (log wind profile). Further, there is significant terrain variability in the area (see grey region of Fig 2c). Hence, it is unlikely that a parameterization of wind speed

would give more accurate values than interpolated measurements. For air density, using a parameterization ( $\rho_{air} = a + bz$ ) as opposed to interpolation results in the following change in horizontal advective flux: SO<sub>2</sub>, Aug 20: 8.3%, Sep 17: 4.4%, CH<sub>4</sub>, Aug 20: -9.2%, Sep 17: -7.0%. However, the interpolated measurements are considered more accurate than a bulk parameterization based on all measurements, so these data are not included in the manuscript.

4791 Discussion: The authors do not address growth of the boundary layer, or uncertainty in the boundary layer depth, as a source of uncertainty. Please include discussion of how the boundary layer is identified (showing plots of the pertinent variable is the spirals are used would be appropriate) and how it affect the results. It is unclear whether the spirals are conducted before and after the box transect or during the experiment.

R1.5) The sequence of flight events (spirals, boxes, transects) is added to Section 2.3. A subsection is added at Section 4.1.6. to discuss the various uncertainties associated with boundary-layer height. However, it is noted that boundary-layer height is not a variable in the mass-balance equations. There is a discussion of the "inversion step change" in this section and the errors associated with its location, as this will affect the vertical turbulent fluxes into and out of the box-top. This inversion step is determined from the spirals, as the reviewer suggests. We have added a sentence (5<sup>th</sup> para, Section 3.4) to clarify that boundary-layer depth is being addressed: "For the calculation of the vertical turbulent fluxes ( $E_{C,VT}$ ), an inversion step change of concentration (determined from the flight spirals) is used as a proxy for the boundary layer depth."

4796 Conclusions: This analysis focuses on the relatively ideal case of an isolated point source in flat terrain. Please comment on the applicability of this in other arenas (i.e. isolating point sources from complex emission systems or topography). Physically, many sources of interest are very large and to actually sample a box in a reasonable amount of time an aircraft would likely have to limit the number of transects used. The number of transects available can greatly affect the interpolation output. Furthermore, this analysis provides useful information as to which components carry most of the flux and thus, which would be most appropriate to measure to reduce sampling bias which the authors should comment on.

R1.6) It is noted that CH<sub>4</sub> is a surface emission with a potentially (but unknown) large emission area. We have added a paragraph to the end of the paper to include recommendations for future measurement projects.

## Anonymous Referee #2

Received and published: 6 June 2015

## General comments

The authors propose an algorithm to retrieve emissions of pollutants by a mass balance approach using aircraft measurements of winds and chemical concentrations. The proposed methodology and the related implementation choices are clearly described through example calculations for two different flights made during an intensive monitoring field campaign. Results are discussed in an attempt of assessing model uncertainties and a comparison to industry reported emission for one of the pollutants is also made. To my knowledge, this manuscript is a very useful contribution to develop new approaches for quantifying air pollutants emissions. The authors also present, in the introduction, a good critical review of the main techniques addressing similar tasks in the recent literature, highlighting differences among these

methodologies; in particular, their algorithm is described as an improvement of the box method due to the better quantification of uncertainties and to a modified treatment for near surface data. This claim seems to be confirmed in the specific circumstances described in the reported experiments, but indeed it should be better motivated, on more theoretical ground.

Specific comments

#### Interpolation and extrapolation

At which frequency are the data (wind and chemical species) measured? Are they averaged before interpolation, and how?

R2.1) More information regarding sampling rate and temporal interpolation is added to Section 2.1 (as addressed in R1.3 above).

Table 4 statistics refer to interpolated values of concentration, of wind, or to the final emission results?

R2.2) A new paragraph is added (Section 3.2,  $3^{rd}$  para). It is explained that each interpolated and extrapolated mixing ratio screen is being compared to the mixing ratio screen derived from Equation 10. The mixing ratio variable  $\chi(s,z)$  is also added to the Table 4 caption to make this clearer.

Methods to extrapolate pollutant mixing ratios seem to highly depend on the considered chemical species, on the flight conditions, on the specific sources in the box area. This is confirmed also by authors' comments (page4787 line8: "The other cases require a proper choice of extrapolation technique based on knowledge of the mixing ratio behavior in this region"). Is this choice subjective and left to the researcher's experience or can something more be suggested by the authors?

R2.3) As discussed in R1.6, we have added a recommendation paragraph at the end of the manuscript to attempt to address this.

## Emissions algorithm

It is not clear to me how the box top height is chosen. While for the case study of SO2 emissions this choice seems to be not relevant, on the contrary for CH4 emissions the vertical advection contribution, as reported in Table 6 (third row), is of the same order of the net horizontal advection (difference of the first two rows) for both days. The authors should clarify how the vertical flux is affected by different box top heights.

R2.4) If the reviewer is asking how the vertical extent of the screen is chosen, it is determined by the highest flight point (rounded up to 100m resolution for simplicity). If the reviewer is asking how the highest flight path was chosen during the actual project, it was informed by a measure of the boundary-layer height during the spiral flight pattern. Generally, during box flights, laps of the facility would continue until it was clear from measurements that the aircraft was above the plume.

To address this issue we have added a subsection to the uncertainty analysis (4.1.6. Boundary-Layer Height), in which emission rate is recalculated for a reduced box height (100 m reduction). The reduction of the box-top height has minimal effect, except in the case of the Sep 2 flight for CH4, in which there is a highly elevated plume (Fig. 5b) which is not fully captured by the reduced box. For this specific case, this results in a reduction in the estimated emission of ~16%. This is added to Table 7.

Similar considerations concern the turbulent flux, as suggested by the authors on page 4790 line12:"there is a large uncertainty in this EC,VT estimation and it is unclear from these measurements if the inversion step change occurs near enough to the box top to necessitate inclusion in the calculated emissions".

R2.5) We have added another component to the uncertainty analysis (Table 7 and Subsection 4.1.6) in which the emission rate is calculated without the  $E_{C,VT}$  term (Section 4.1 second paragraph and Table 8, second row). The error introduced by this term is estimated as zero for SO<sub>2</sub> and 2% or 7% for CH<sub>4</sub>.

Finally, also on page 4791 line2 they state "For CH4, the resulting values of EC,M are small relative to the horizontal flux term EC,H, but are large compared to the final calculated emission rate"

R2.6) This effect is already included in the uncertainty analysis (Table 7: Density Change). To help make this clear the following sentence is added at Section 4.1.5.: "These ratios are used to determine the magnitude of the change in air density within the box, which defines the  $E_{air,M}$  and  $E_{C,M}$  terms."

#### Discussion

In general, it seems to me that the proposed methodology is well suited for single and concentrated sources, while for surface-based or diffused emission sources and low altitude plumes the improvement with respect to the other emissions algorithms reported in the literature is not so relevant.

R2.7) We agree. However it is often the case that both types of plumes are contained in the same region, and we demonstrate that both types of plumes can be captured simultaneously by multiple instruments on the same aircraft (with different degrees of uncertainty).

Uncertainty quantification: the considerations on the relative importance of the uncertainty terms are very interesting but are they general or case-specific?

R2.8) This is discussed in the added Recommendations paragraph (see R1.6).

Finally, the authors should say something on the actual applicability of the TERRA algorithm on different study areas and different flight path designs: for example did the authors try to estimate the emissions for the same study area by using less horizontal data to reproduce a shorter flight path?

R2.9) While it is impossible to reduce the horizontal extent of the flight path (since the flight circuit would no longer be a closed loop), we have added analysis of a reduced vertical extent (Section 4.1.6) and have demonstrated what effect that has on the results (see R2.4).

#### Anonymous Referee #3

Received and published: 3 June 2015

Review of "Determining air pollutant emission rates based on mass balance using airborne measurement datas over the Alberta oil sands operations" by Gordon et al., submitted to AMT.

This paper describes and outlines a specific methodology for calculating emissions of pollutants from point sources (for example, specific facilities) using aircraft concentration and wind measurements. To my knowledge, this is the first attempt at standardizing this method, which is often used but with slight differences between different research groups. It is a suitable subject matter for an AMT publication, as it focuses on the method developed for this estimation.

Overall, this is a well written and clear paper that outlines a method for calculating emissions rates using aircraft measurements of pollutants and is appropriate for publication in AMT after some specific comments are addressed. My one concern is the treatment of uncertainties which is not as rigorous as it could be. The uncertainty of each component is generally estimated based on several cases (trying different extrapolation methods for example) and always stated as approximate (using approximately equals signs). The authors should be more definitive, even though they do state that the uncertainties are being "estimated" and not really calculated. I felt that a paper specifically describing a method should go the extra step and define uncertainties in a more systematic way, and I would like to see an attempt at this. Or, if the method is retained, some overall description: "for the TERRA method, we estimate uncertainties for each term by calculating each term using different assumptions and using the largest difference between two different results from the different methods" or something to that effect, to show at least methodological consistency.

R3.1) We agree. To our knowledge there is no way to calculate exact uncertainty estimates given the method and the set of measurements that we have. We believe that, for these circumstances, the only possible way to estimate uncertainties is to recalculate the emission rate using difference assumptions (as the reviewer states). Hence, we retain the method used, and add the following paragraph to the start of Section 4.1:

"To calculate uncertainties in the final emission rate, we attempt to identify and estimate each source of uncertainty....In some cases, such as wind speed measurement error, wind speed extrapolation, and box-top height, the uncertainty affects multiple variables simultaneously. For this reason all uncertainties are expressed as a fraction of the base case emission rate ( $E_C$ ) and also in units of t h<sup>-1</sup>. Each uncertainty is assumed to be independent and they are added in quadrature to give the total estimated emission rate uncertainty as..." Eq. 11 is given here.

## Specific comments:

Line 25: In addition to Wratt et al., others have used this budgeting approach, for example see Gatti et al., Nature, 2014 for using aircraft profiles upwind and downwind to estimate fluxes in the Amazon.

R3.2) We thank the author for bringing this recent article to our attention. It has been included in the summary. Although flux uncertainties are presented in absolute units in the Gatti et al. study, we use an estimate of uncertainty of ~40%, based on the scaled total 2010 flux (their Table 1).

Introduction does a nice job of outlining the different similar but slightly different methods that have been used in the past - this is a nice overview for the reader. See comment at the end of this review on perhaps differentiating between estimates of point source emissions vs. regional emissions over a larger area and measurements farther downwind.

R3.3) See new paragraph at the end of the manuscript, which addresses this differentiation (also see R1.6).

p4774 L1: define IMU - is altitude not measured via GPS, but rather only converted from pressure altitude? What is the accuracy of this, usually pressure altitude conversion assumes some ideal atmosphere, etc.? How do you have lat/lon without GPS?

R3.4) We have expanded (and corrected) Section 2.1 to explain: "Aircraft state parameters (latitude, longitude, and ellipsoid height altitude) are measured by GPS and Honeywell HG1700 Inertial Measurement Unit (IMU). Kalman filtering of the integrated IMU data is combined with the GPS to provide state parameters at a rate of 100 Hz.".

p4774 L1-5 I would also be interested in the frequency of the measurements on the aircraft.

R3.5) The measurement frequency of all instruments has been added to Section 2.1. Also added is: "To consolidate the various measured parameters, high-frequency data (wind and state parameters) were averaged to a frequency of 1 Hz, while low-frequency data (CH<sub>4</sub> were linearly interpolated to a frequency of 1 Hz." (See also comment R1.3).

p4774 L15 - I cannot find the Picarro G2204 on their website - what is the frequency of measurements and is this a flight-ready analyzer?

R3.6) The G2204 is currently a specialized instrument, not yet listed on the Picarro web-site. It is not considered flight-ready, as it has trouble operating at very low pressures, and the instrument flow rate is weakly dependent on sampling pressure. However, for the height range of the box flights, the instrument functioned well and the change in flow rate over that range of pressure difference results in a delay of < 0.5 s. Further, a G2401-m (flight ready) was also used on the flight and the CH<sub>4</sub> measurements correlate very well (R2 = 0.987 over all 22 flights). The frequency of the measurements is effectively 0.28 Hz.

By time delay, are you referring to how long it takes for air to go through the inlet line on the aircraft, or actually internal to the analyzer? This seems quite long - what is your flow rate on the Picarro unit?

R3.7) It is the sum of both delays (text is added to explain this). The delay was determined by timed pulsing tests, measuring the average delay between a change at the inlet and the value recorded by the instrument. The flow rate is  $\sim$ 0.35 L/min, and some of the 6 m length of inlet was shared with a higher flow rate instrument, which implies that most of the delay is internal to the instrument.

Regarding the precision, I understand why you want to define it based on five different calibrations, but it seems like the short-term precision might be more important given your application. You are differencing upwind and downwind, so your uncertainty in that difference is only tied to the drift of the analyzer within a flight, not across the whole project. I would hypothesize that will be much smaller than 1.3%, which is very high. I need to read on as to how you fold this into an uncertainty on the emission flux, because 1.3% is not actually the precision of the methane measurement, but of the slope. How different are actual mole fractions you measure for each time you calibrate? At 1900 ppb, 1.3% is nearly 25ppb error - I doubt your calibration standards were measuring 25 ppb different from one calibration to another.

R3.8) We also include the standard deviation of the measurements during the two calibrations nearest to each flight. For CH4, this gives 0.014%, and for SO2, this gives 5.6%. Some text is added to explain the representation of short-term variability versus long-term drift. (Section 2.1)

How does this uncertainty propagate into your final uncertainty? It is never addressed how this or other measurement uncertainty (in winds for example) affects the final uncertainty.

R3.9) We have added a Monte Carlo simulation in which the emission rate is recalculated (approximately 1000 times) with modified wind speed ( $U_x$  and  $U_y$ ), SO<sub>2</sub>, and CH<sub>4</sub> (standard deviations of 0.4 m s<sup>-1</sup> for wind speeds, 5.6% for SO<sub>2</sub>, and 0.014% for CH<sub>4</sub>). This simulation is used to determine the uncertainty due to measurement uncertainty. (Section 4.1.1)

p4774 Line 10 - since this is a methods paper, how do you run those 5 calibrations for both gases? What range of concentrations (mole fractions) are used as your references?

R3.10) The ranges are also included in Section 2.1 (2 to 3 ppm for CH4 and 0 to 400 ppb for SO2).

p4779: what is the frequency of your wind measurement?

R3.11) As R3.5, all frequencies are added in Section 2.1. Wind is sampled at 32 Hz (by a Rosemount 858), but is averaged to 1 Hz to align with GPS data.

p4779, prior to interpolation using any of the 3 methods, what is the native resolution of the measurements along S? Do you average your wind measurements prior to interpolating? My understanding of wind measurements is that there is a lot of "noise", i.e. short term variability that would need to be averaged before interpolating (real atmospheric variability due to turbulence).

R3.12) See R3.5. High frequency turbulence is averaged out of the wind measurements, which necessitates the addition of turbulent mixing terms into Equation 1.

p4781 line 21 and 23: deposition velocity is sometimes Vd and sometimes VD, keep consistent.

R3.13) Two occurrences of  $V_d$  (page 4781 and Table 2) are changed to  $V_D$ .

It seems to me that separating the turbulent and advective fluxes means that you are averaging the concentrations you have measured (and the winds) so that you are not resolving turbulent eddies, and only capturing mean advection. Otherwise you are double counting. Am I understanding that?

R3.14) This is a correct interpretation. From Reynolds decomposition  $(U = \overline{U} + U', \rho = \overline{\rho} + \rho')$ , the advective flux term captures  $\overline{\rho} \overline{U}$ , while the turbulent flux term approximates  $\overline{u'\rho'}$  (which can't be measured at our 1 Hz frequency). If we have interpreted the reviewer's comments correctly, this seems to be a request for clarification in the discussion only and not a suggested revision – so no text is modified.

p4781 L28 what is the Ea,V term?

R3.15) This is corrected to read " $E_{air,V}$ ".

p4785. It would clarify the different methods for extrapolating the concentration measurements down to ground level if you had a figure, or refer to Figure 6 here.

# R3.16) Added to the end Section 3.1.3: "The constant, linear-fit, and exponential fit are compared in a later Section (Fig. 6)."

p4786, 20-25. Can you clarify how you compare the three interpolation methods to the original synthetic data to evaluate them? (rms error and correlation coefficent are defined how exactly?). The caption in Table 4 explains a little bit more, but I'm not sure if it says - you are comparing actual emissions calculations (which take the wind and the concentration into account both?). What is the interpolated average? (average emission, or average concentration?), A little more information would be good here, perhaps in the text.

R3.17) (See R2.2) A new paragraph is added (Section 3.2,  $3^{rd}$  para). It is explained that each interpolated and extrapolated mixing ratio screen is being compared to the mixing ratio screen derived from Equation 10. The mixing ratio variable  $\chi(s,z)$  is also added to the Table 4 caption to make this more clear.

p4786 If you can add a figure, I would like to see the concentration screens after the interpolation based on the flight data locations for all three interpolations and cases (i.e. the corresponding figure to Figure 3 but after you've re-sampled and interpolated the data) to get a qualitative feel for how this works. Since you ran so many cases, even just showing it for one case (perhaps Kriging which you determined was the best fit) would be nice.

## R3.18) We have added three panels to Figure 3 (d,e,f) to include kriged and extrapolated data. The difference is barely perceptible, but of enough interest to include (as the review suggests).

p4787 L26: the figure caption and the earlier text indicate that you are referring to the zero-to-constant extrapolation, not the linear extrapolation here. Otherwise, this is confusing.

## R3.19) Text is corrected.

p4787 section 3.3 - Which interpolation method is used in Figure 4? is this the Kriging method, and is that what is also plotted in Figure 5 and 6, because it was the best in your evaluation of the three methods? Please note this in the text and captions.

## R3.20) All three figure captions now state "krig interpolated". Section 3.3 now begins with "The kriging method discussed in Section 3.2 is used for these and all subsequent interpolations."

p4793 Line 27, this is only true if n are \*independent\* measurements of the same quantity. At 1 Hz, these are not independent because you have correlation length scales in the atmosphere much longer than your measurement frequency. I am guessing that the mean measurement of the upwind side of the top box is much different than of the downwind. I would argue that N might only be equal to 4 (for the four sides of the box top). Otherwise, the time traces of the measurements will tell you the correlation length/time scale, and then one can calculate N. This is what you would do if you had a circle for example. The 95% confidence interval in this mean is larger than you state.

R3.21) This is an excellent suggestion which we have incorporated into the text (Section 4.1.4). Autocorrelation of the box top mixing ratio values demonstrates an independent length scale between 1.5 km and 3 km. For the lap distance of nearly 59 km, this gives N between 20 and 40. For our error estimation we conservatively use the lower value of 20, which increases the confidence interval by a factor of 5.5.

p4794 L9: What highest uncertainties? you mean for each delta? In the text, values are given for each delta but not the highest or any range, so it's not clear what is meant by "highest" here.

R3.22) This has been removed as part of the restructuring of Section 4.1. We now list all the uncertainties (for each flight and each species) separately in Table 7.

p4795 section 4.2 - how do these wind shifts affect your uncertainty estimate? (should uncertainties be higher on days when the plume may shift?).

R.23) We add the following text at the end of the section: "These results suggest that wind shifts do not affect our measurements for these two flights; however, this demonstrates the necessity of stationarity testing such as this for mass-balance flight emissions calculations. This is especially the case when weather conditions change during the box flight duration."

p4797 general comment: One thing should be mentioned either in the introduction or here, which is that other methods (e.g. single transect) that assume perfect mixing in the vertical direction (and thus, the extrapolation to the surface is constant), make that assumption only far enough downwind of the source so that the plume has had time and distance to mix vertically within the PBL. The TERRA method could also certainly be used farther away from a source or for calculating emissions from a larger region, but it is most useful when emissions are being calculated from a single source whose location is known and surrounded by the flight path. Some of the references in Table 1 use this type of method for significantly greater source regions (Turnbull, Peischl, Karion, Mays and Cambaliza).

## R2.24) This is now briefly discussed the in the last paragraph Conclusions and Recommendations section.

Another general comment: What kind of uncertainties do you claim for the TERRA method in general? Is it dependent on the specific flight - i.e. would the same uncertainty analysis you have conducted (multiple methods comparisons, generally) need to be conducted every time TERRA is used to calculate emissions, or can it be generalized somehow?

R2.25) It is necessary for the reader to compare our uncertainties with their own particular circumstances and determine which uncertainties are applicable to each particular emission type and flight pattern. We hesitate to provide a generalized claim as this may be over-reaching.