7-27-2022

Comparing airborne algorithms for greenhouse gas flux measurements over the Alberta oil sands

Broghan M. Erland

Cristen Adams

Andrea Darlington

Mackenzie L. Smith

Andrew K. Thorpe

See next page for additional authors

Authors

Broghan M. Erland, Cristen Adams, Andrea Darlington, Mackenzie L. Smith, Andrew K. Thorpe, Gregory R. Wentworth, Steve Conley, John Liggio, Shao-Meng Li, Charles E. Miller, and John A. Gamon

Atmos. Meas. Tech., 15, 5841–5859, 2022
https://doi.org/10.5194/amt-15-5841-2022

Atmospheric Measurement Techniques
Open Access
EGU

Comparing airborne algorithms for greenhouse gas flux measurements over the Alberta oil sands

Broghan M. Erland[1], Cristen Adams[2], Andrea Darlington[3], Mackenzie L. Smith[4], Andrew K. Thorpe[5], Gregory R. Wentworth[2], Steve Conley[4], John Liggio[3], Shao-Meng Li[3], Charles E. Miller[5], and John A. Gamon[1,6]

[1]Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, AB, T6G 2R3, Canada
[2]Resource Stewardship Division, Alberta Environment and Parks, Edmonton, AB, T5J 5C6, Canada
[3]Air Quality Research Division, Environment and Climate Change Canada, Toronto, ON, M3H 5T4, Canada
[4]Scientific Aviation, Inc., Boulder, CO 80301, USA
[5]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA
[6]School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE 68583, USA

Correspondence: Broghan M. Erland (erland@ualberta.ca)

Received: 11 April 2022 – Discussion started: 4 May 2022
Revised: 27 July 2022 – Accepted: 29 August 2022 – Published: 14 October 2022

Abstract. To combat global warming, Canada has committed to reducing greenhouse gases to be (GHGs) 40 %–45 % below 2005 emission levels by 2025. Monitoring emissions and deriving accurate inventories are essential to reaching these goals. Airborne methods can provide regional and area source measurements with small error if ideal conditions for sampling are met. In this study, two airborne mass-balance box-flight algorithms were compared to assess the extent of their agreement and their performance under various conditions. The Scientific Aviation's (SciAv) Gaussian algorithm and the Environment and Climate Change Canada's top-down emission rate retrieval algorithm (TERRA) were applied to data from five samples. Estimates were compared using standard procedures, by systematically testing other method fits, and by investigating the effects on the estimates when method assumptions were not met. Results indicate that in standard scenarios the SciAv and TERRA mass-balance box-flight methods produce similar estimates that agree (3 %–25 %) within algorithm uncertainties (4 %–34 %). Implementing a sample-specific surface extrapolation procedure for the SciAv algorithm may improve emission estimation. Algorithms disagreed when non-ideal conditions occurred (i.e., under non-stationary atmospheric conditions). Overall, the results provide confidence in the box-flight methods and indicate that emissions estimates are not overly sensitive to the choice of algorithm but demonstrate that fundamental algorithm assumptions should be assessed for each flight. Using a different method, the Airborne Visible InfraRed Imaging Spectrometer – Next Generation (AVIRIS-NG) independently mapped individual plumes with emissions 5 times larger than the source SciAv sampled three days later. The range in estimates highlights the utility of increased sampling to get a more complete understanding of the temporal variability of emissions and to identify emission sources within facilities. In addition, hourly on-site activity data would provide insight to the observed temporal variability in emissions and make a comparison to reported emissions more straightforward.

1 Introduction

Global warming is on the pathway to a minimal projected global temperature increase of 3.3–5.7 °C by 2100 unless meaningful change is enacted to reduce anthropogenic greenhouse gas (GHG) emissions (Le Quéré et al., 2018; Friedlingstein et al., 2020; IPCC, 2021). Anthropogenic carbon dioxide (CO_2) and methane (CH_4) emissions are the first and second largest contributors to climate change, respectively (Friedlingstein et al., 2020). Accurate quantification of GHG emissions is an essential foundation for emissions reductions.

Regional, national, and global CH_4 and CO_2 emissions are estimated using a combination of bottom-up and top-down methods. In general, bottom-up methods aggregate and extrapolate source-specific data to estimate emissions at a larger scale, whereas top-down methods measure atmospheric GHG concentrations at a larger scale and infer point and area source emissions (National Academies of Sciences, Engineering, and Medicine, 2018). Anthropogenic CO_2 and CH_4 emissions are estimated using (a) ground-based methods, (b) airborne methods, or (c) satellite methods (e.g., Frankenberg et al., 2016; Conley et al., 2017; National Academies of Sciences, Engineering, and Medicine, 2018; Lauvaux et al., 2022; Irakulis-Loitxate et al., 2022). Large differences between bottom-up aggregated inventory estimates and top-down atmospheric budget estimates need to be reconciled to reduce the uncertainty in estimating global and regional GHG emissions (Nisbet and Weiss, 2010; Dlugokencky et al., 2011; Allen, 2014; Kort et al., 2014; Johnson et al., 2017; Alvarez et al., 2018; Liggio et al., 2019; Saunois et al., 2020; Friedlingstein et al., 2020; MacKay et al., 2021). Improved emissions estimates facilitate the best climate change policy (Le Quéré et al., 2018), allowing us to adopt pathways for lower global warming increases. This paper focuses on airborne approaches, as they are intermediate in spatial scale between proximal and satellite sampling methods and therefore improve estimation by providing essential validation between top-down and bottom-up methods (National Academies of Sciences, Engineering, and Medicine, 2018; Friedlingstein et al., 2020; Cusworth et al., 2020; Nisbet et al., 2020).

While some airborne methods utilize eddy covariance measurements (Yuan et al., 2015; Wolfe et al., 2018), methods for sampling anthropogenic GHG emissions using aircraft tend to fall into two major categories: (i) mass-balance methods (O'Shea et al., 2014; Gordon et al., 2015; Conley et al., 2017; France et al., 2021; Foulds et al., 2022) and ii) spectral imaging methods (Duren et al., 2019; Tyner and Johnson, 2021; Krautwurst et al., 2021; Cusworth et al., 2022). These methods capture atmospheric fluxes using varying approaches that are affected by different biases and are complementary when creating emission budgets. Mass-balance methods quantify the mass flux, or change, in the mixing ratio of a species due to emissions from a known source area. Sampling schemes for mass-balance flights range from flying a single transect downwind of a source, to multiple stacked transects creating a vertical "screen" to catch the plume at various altitudes, or flying a "box-flight" around a facility – or specific intra-facility source area – to constrain a plume (Gordon et al., 2015; Conley et al., 2017; Baray et al., 2018; Liggio et al., 2019; France et al., 2021). Remote spectral imaging methods fly above potential sources and use absorption spectroscopy of reflected solar radiance or thermal emissions to capture regional or facility emissions (Frankenberg et al., 2016; Bartholomew et al., 2017). Currently, mass-balance box-flight methods can attain a lower uncertainty from a single sample of emission estimates ($\sim 2\,\%$) (Gordon et al., 2015; Conley et al., 2017) than the spectral methods ($< 30\,\%$) (Duren et al., 2019; Thorpe et al., 2020) due to smaller background and wind measurement uncertainties. However, they require an understanding of plume sources to know where to fly, and they can take longer; therefore they can be more costly.

Mass-balance box-flights involve sampling in stacked, often cylindrical, flight laps, typically surrounding a known source or set of sources, at altitudes varying from the minimum safe-flight altitude to the atmospheric boundary layer capping an emission plume (Gordon et al., 2015; Conley et al., 2017). Due to minimum flight height restrictions a gap between the surface and the flight box is inevitable. Concurrent surface sampling is ideal but often unavailable, so operators aim to fly at a distance from the source where the plume has risen enough but has not dispersed to the degree that it cannot be detected to capture the plume inside the box (Conley et al., 2017). For mass-balance box-flights, extrapolation to the ground is often the largest error source, nearing $\sim 30\,\%$ when the bottom of the plume is not captured (Gordon et al., 2015; Conley et al., 2017). Airborne mass-balance box-flight methods depend on the assumption of a stable boundary layer and that the emission plume is captured at the top of the box and does not change during sampling (i.e., that conditions are stationary) (Fathi et al., 2021).

Methods applying mass-balance equations to aircraft measurements have been developed and refined over the last two decades (Kalthoff et al., 2002; Alfieri et al., 2010; Karion et al., 2013; Gordon et al., 2015; Conley et al., 2017; Gordon et al., 2018; Krings et al., 2018; France et al., 2021). In this work, two box-flight mass-balance sampling methods, a top-down emission rate retrieval algorithm (TERRA) developed by Environment and Climate Change Canada (ECCC) (Gordon et al., 2015) and a Gaussian theorem algorithm developed by Scientific Aviation (SciAv) (Conley et al., 2017), provide two approaches to evaluate mass fluxes from aircraft measurements. To our knowledge, a detailed comparison of the two methods has not yet been conducted. If algorithm comparisons indicate agreement, then emissions estimates from multiple campaigns using mass-balance methods can be aggregated, which will improve the certainty in GHG budgets.

A complementary method to airborne mass-balance are airborne spectral methods, which can be considered top-down methods that produce results similar to satellite data but with higher accuracy (Kort et al., 2014; Frankenberg et al., 2016). Single flights are used to sample and estimate emissions from sources, and repeated sampling can determine source persistence to infer regional emission budgets (Duren et al., 2019). Stationarity of an emission plume occurs when the source of emission is consistent and meteorological conditions such as the boundary layer and wind are stable throughout the time of sampling. Remote spectral sampling provides quick "snapshots" of features and

therefore avoids the stationarity requirement inherent to airborne mass-balance methods, which have lengthy sampling times ranging from less than one hour to multiple hours, depending on the region measured. Remote spectral imaging methods are being advanced by the NASA Jet-Propulsion Laboratory, which has had success in mapping, inferring wind vectors, and estimating emissions over large areas using their Airborne Visible InfraRed Imaging Spectrometer – Next Generation (AVIRIS-NG) (Frankenberg et al., 2016; Duren et al., 2019; Jongaramrungruang et al., 2019; Thorpe et al., 2020; Cusworth et al., 2021). Coincident sampling using the SciAv and AVIRIS-NG methods at facilities has indicated that the methods tend to agree, within errors (Duren et al., 2019; Thorpe et al., 2020).

Reducing GHG emissions in Canada has become a national and provincial priority (Johnson and Tyner, 2020; Government of Canada, 2021). Airborne and ground-based campaigns suggest that the inventories used to facilitate national and provincial policy are under-reporting GHG emissions (Brandt et al., 2014; Gordon et al., 2015; Johnson et al., 2017; Baray et al., 2018; Liggio et al., 2019; Chan et al., 2020; MacKay et al., 2021; Baray et al., 2021; Tyner and Johnson, 2021). For example, a recent study aggregated thousands of mobile, ground-based emission rate estimates taken without notice to operators from upstream Canadian oil and gas and found that inventories underestimated methane emissions (Atherton et al., 2017; MacKay et al., 2021). Using tower data, methane emissions estimates over 8 years from oil and gas operations in Western Canada were estimated to be nearly twice those reported in Canada's National Pollution Release Inventory (Chan et al., 2020). Airborne campaigns by ECCC measuring carbon dioxide and methane have also estimated emissions to be 13 %–123 % (Liggio et al., 2019) and 40 %–56 % higher (Baray et al., 2018), respectively, than national inventories. In a comparable campaign by Scientific Aviation, industrial upstream oil and gas CH_4 emissions estimated in two regions in Alberta were 5 and 17 times higher than values reported to the Alberta Energy Regulator (Johnson et al., 2017). Greater certainty in top-down emissions estimates helps flag under-reporting in bottom-up inventories and better informs GHG policy makers of emissions, allowing them to enact meaningful GHG reductions. As part of the Oil Sands Monitoring (OSM) program mandate to advance the understanding of Alberta's emissions, a collaborative study was initiated in 2017 by Alberta Environment and Parks (AEP) and the U.S. National Oceanic and Atmospheric Administration (NOAA), contracted to Scientific Aviation. The goal was to use airborne measurements to quantify facility- and activity-specific GHG emissions from mineable and in situ oil sands developments in northern and east-central Alberta. Between August 2017 and October 2018, sampling was conducted for various facilities and repeated over several days to assess both temporal and inter-facility variability in GHG emissions rates.

In this study, we compared emissions estimated using the same data from five box-flights from the 2017–2018 campaign using two airborne mass-balance algorithms (TERRA and SciAv). Our intention was to assess the comparability of emissions estimates from past campaigns flown by ECCC and SciAv, which may provide greater certainty of GHG emissions from the Alberta oil sands and other regions where these methods are used. The main objective was to test if emissions estimates from the TERRA and SciAv algorithms agreed with uncertainty and then to assess the sensitivity of emissions estimates to surface extrapolation using a variety of schemes. The cause of any differences between the algorithms was assessed. Since mass-balance flights are typically flown with the knowledge of and permission from facilities operators, these methods, while they may be accurate, may not necessarily reflect typical operating conditions or GHG emissions. Consequently, a secondary research objective was to examine the potential of utilizing complementary spectral imaging methods, such as AVIRIS-NG, to supplement mass-balance box-flights by providing contextual information to capture the spatial and temporal variability of oil sands GHG emissions.

2 Methodology

Both mass-balance methods involve flying around a known source in a box pattern to fully capture an emission plume for estimation, but they differ in their approaches (Table 1). TERRA evaluates the entire dynamic system with terms to quantify the horizontal advective and turbulent flux through the box walls and box top, deposition of flux to the ground, chemical mass changes, and air density changes (Gordon et al., 2015). TERRA applies a simple kriging of the raw data to spatially interpolate between the raw lap data, then estimates the dynamic terms to solve mass-balance equations and derive an overall total emission rate (Gordon et al., 2015). In contrast, the SciAv algorithm simplifies the system to a single horizontal flux through the box and estimates the flux divergence from the box by evaluating a mass-balance equation derived from Gauss's theorem for relating flux through a closed surface to a divergence from a volume integral (Conley et al., 2017). The SciAv and TERRA theory, including equations, are described in further detail in the Supplement – Sect. S1.1 and S1.2, respectively.

2.1 Box-flight aircraft measurements

The AEP-NOAA-Scientific Aviation 2017–2018 Alberta oil sands flight campaign conducted 150 flight segments at 16 different facilities across Alberta. Many of these facilities included multiple source areas, such as a plant, a mine, and/or tailings ponds. The aircrafts flew in laps around either the entire perimeter of a facility or around specific source areas. The data were collected and processed by Scientific Avi-

Table 1. Characteristics of the mass-balance SciAv and TERRA algorithm for box-flights.

	SciAv	TERRA
Parameterization of flux	Simplified to one horizontal flux term.	Quantifies the dynamic system using several flux terms.
Conceptual algorithm steps	First step: for each lap, solves a single mass-balance integral equation, derived from a Gaussian theorem, using flight measurements decomposed into one single horizontal flux vector, to estimate the flux divergence due to an emission source within the box.	First step: applies simple kriging to interpolate flight lap measurements to a spatially resolved screen.
	Second step: bin lap flux divergence estimates by altitude ranges, estimate an average flux divergence for each bin, then integrate the bins across the total flight height to produce total emission rate estimates.	Second step: simultaneously solves two mass-balance equations with multiple integrals to fully constrain the system to evaluate a total emission rate estimate. The first equation quantifies emission flux using seven integral terms and the second has three air flux integrals to account for air flow.
Surface extrapolation	Extends lowest bin average flux divergence as a constant to the surface.	Spatially resolved screen of mixing ratios extended to the surface using one of five extrapolation options depending on plume character.
Output	Emission rate from point or area source.	Emission rate from point or area source.
Fundamental assumptions	Stationary plume, stable meteorological conditions, and full plume capture at the top of the plume.	Stationary plume, stable meteorological conditions, and full plume capture at the top of the plume.
Error terms	Three broad terms.	Seven specific terms.

ation (Boulder, CO, USA) on contract to NOAA. Flights were performed using two fixed-wing, single-engine aircraft, a Mooney M20R (Aircraft N617DH) and a Mooney M20M (Aircraft N2132X), equipped with monitoring equipment. Concentrations of CO_2 and CH_4 were measured using a cavity ring-down spectrometer (Picarro 2401m or 2210m, Picarro Inc., Sunnyvale, CA, USA) in its precision mode at $\sim$ 0.5 Hz, as described by Crosson (2008). Other variables used in the analysis were measured using the airplane primary flight information system and GPS, including wind speed components ($m\,s^{-1}$), pressure (mb), temperature (K), heading (degree), altitude (m), and latitude and longitude. Each flight segment was screened to assess whether (i) sufficient altitude was reached to capture the entire plume, (ii) winds were sufficiently strong and consistent, and (iii) upwind sources were negligible relative to emissions inside the box. If these criteria were not met, then the algorithm results (i.e., emissions estimates) were considered unreliable. It was challenging to determine a quantitative threshold for adequate wind conditions or negligible upwind sources, since the relative impact on the calculated emission rate depends on the magnitude of emissions. Nonetheless, any flight segments with average wind speeds below $5\,m\,s^{-1}$ were flagged (Gordon et al., 2015), as were flight segments with upwind mixing ratios above background, and were then assessed further using professional judgement. It is important to note that light and/or variable winds will increase the uncertainty of the SciAv algorithm by increasing the variability between laps.

Five flights from three facilities were selected for the algorithm comparison and are summarized in Table 2, with sample codes (F01 to F05) assigned for comparison purposes. The three facilities from the Athabasca oil sands region included in the study were: Mildred Lake and Aurora North plant sites (Syncrude), Horizon Oil Sands Processing Plant and Mine (CNRL), and Suncor Energy Inc. oil sands (Suncor). Flight paths around the facilities are shown in Fig. 1. These five flights were chosen to capture a range of possible sample types given varying profile shapes, number of laps, boundary layer height, season, and whether the flight was around a facility perimeter or plant. Four of the five flights selected were considered ideal samples during preliminary flight screening. One flight, F05, was chosen as a poor-quality sample, rejected during flight screening as having "not fully captured" the top of the emission plume, and used to assess how the methods compare when a fundamental assumption of the method is not met. Boundary layer height was estimated by Scientific Aviation by assessing profile changes in potential temperature gradients before and after flights. Through the flight screening process, all five flights were judged to have consistent, stationary winds and stable boundary layers. F02 had normal operating conditions and no flaring events reported by CNRL Horizon. Facilities were informed before sampling. Operating conditions at the oil

sands facilities for F01, F03, F04, and F05 were not shared at the time of writing.

Figure 2 shows an example of a close view of a flight path for F04 CH_4. Blue dots indicate background levels in ppm, and enhanced mixing ratios within the plume are in a gradient of cyan-yellow-red, with the largest enhancements in red. A large plume can be seen on the north-east section of the flight path in Fig. 2, and a few enhancements were measured elsewhere along the flight path. There is evidence that the top of the plume is captured, as dots at the highest altitude show background concentrations, and the flight path goes above the estimated boundary layer.

2.2 Box-flight emissions estimate algorithms

2.2.1 TERRA algorithm

Environment and Climate Change Canada (ECCC) provided the TERRA algorithm, ran portions of the analysis, and detailed instructions on how to produce estimates using the algorithm with commercial plotting software (IGOR Pro 8, Wavemetrics, Lake Oswego, OR, USA). The first step of the TERRA algorithm creates the screen of spatially interpolated lap data by applying simple kriging to the campaign data collected by Scientific Aviation. The spatially resolved screen of mixing ratios is a 2-dimensional unraveling of the lap data by altitude over the length of sampling and is commonly referred to as the "box" (Gordon et al., 2015).

For the second step, TERRA has 5 options for extrapolating emission concentrations from the lowest flight layer to the surface to account for fluxes below the flight path: (1) a background extrapolation fills all data below the flight path with a background concentration (applied when there is a fully captured, elevated plume and there is a reason for choosing one single mixing ratio to be extrapolated to the surface); (2) a constant extrapolation uses the concentration at the bottom of the screen and assumes that this remains constant to the surface (best used in the general case of a fully captured, elevated plume, as it avoids the assumption of a background value); (3) the linear method fits a line through the lowest points on the screen up to an altitude of 300 m above ground level (preferred in the scenarios when emissions occur from the surface, such as a low plume that was not fully captured or a mixed plume with ground sources such as a tailings pond); (4) the interpolation between the concentration at the lowest altitude of the screen and the background concentration at the surface (ideally used when there is evidence of decreasing emissions with only a trace of the plume at the bottom of the flight path); and (5) exponential extrapolation calculates a Gaussian fit through the lowest points on the screen (largely avoided unless there is a strong argument that it best fits the plume behaviour).

Surface extrapolation was essential for this study, as all flights had emission plumes that were not fully captured at the lowest flight track. For the TERRA standard estimates, a linear fit was used for F01, F03, and F05 for both CH_4 and CO_2 due to their low position on the screen and their likelihood of having an increasing emission towards the surface. An interpolation to background fit (i.e., Option 4 described in the paragraph above) was applied to F02 for both CH_4 and CO_2, and F04 for CO_2, as these cases largely captured plumes with low mixing ratios at the bottom of the flight path. A constant extrapolation (Option 2) was fit to F04 for CH_4 to avoid an assumption about the background concentration, as it was the one flight with a very large plume dispersion where plume behaviour was unknown (Sect. S1.3). All extrapolation outcomes were produced to calculate the surface extrapolation error, which accounts for potential differences when choosing the best surface extrapolation (Gordon et al., 2015), and to compare it with the range of possible outcomes from the SciAv method by running a bootstrap analysis. The settings for the standard TERRA emissions estimate were chosen by assessing the plume location, boundary layer conditions, and plume source information to determine the appropriate surface extrapolation (Gordon et al., 2015; Baray et al., 2018).

The TERRA total uncertainty estimate was calculated by adding seven error terms in quadrature (i.e., by taking the square root of the sum of squares). Four of the seven TERRA error terms were evaluated. The wind and measurement error had been previously determined to each be $< 1\,\%$, and the vertical turbulence term had been functionally removed from TERRA analysis (Gordon et al., 2015; Baray et al., 2018). The surface extrapolation error was calculated as the maximum percent change amongst the plausible surface extrapolation estimates. For example, background extrapolations (Option 1), which assume no mixing ratio enhancement below the flight path, were not considered for standard estimates when a flight had increasing emissions at the bottom of the screen. A description of the calculation of the box-top mixing ratio, air density, and box-top height error terms is given in Sect. S1.4.

2.2.2 SciAv algorithm

Scientific Aviation provided results from the first step of estimating the flux divergence for each lap. They applied their algorithm to the flight data and provided output that could be used to address the research objectives. Although the algorithm itself is proprietary, the concepts and formulae underpinning the algorithm are described in detail in Conley et al. (2017). The algorithm output included standard emissions estimates and uncertainties using the SciAv preferred settings. It also included profiles of flux divergence and uncertainty for each lap versus altitude and preferred bin altitude ranges. These were used in the second step analysis of binning lap estimates and integration of the flight profiles to test cases such as extrapolation to the surface in MATLAB 2020a (The MathWorks, Inc., Natick, Massachusetts, United States). Figure 3 provides an example profile of the average

Table 2. Information on the five box-flight samples used in the comparative analysis.

	F01	F02	F03	F04	F05
Facility code	Syncrude	CNRL	Suncor	Syncrude	Suncor
Area sampled	Perimeter	Plant	Plant	Plant	Perimeter
Date	24 April 2018	19 July 2018	19 April 2018	14 August 2017	6 September 2018
Season	Spring	Summer	Spring	Summer	Fall
Min. altitude (m)	168	173	139	150	157
Max. altitude (m)	1057	1246	775	1043	563
Boundary layer height (m)	1100 ± 150	900 ± 200	600 ± 100	900 ± 50	500 ± 100
# Laps	8	14	19	25	7
Start time (GMT)	20:17:06	20:48:38	17:14:36	19:11:37	17:36:56
End time (GMT)	22:47:42	21:43:31	18:12:12	20:09:09	19:55:00

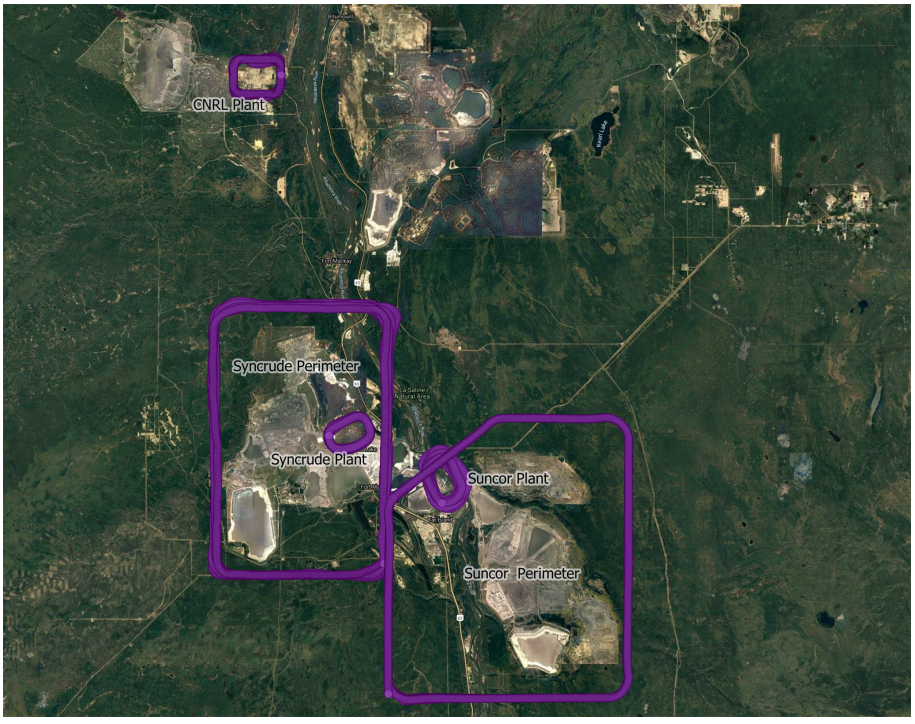


Figure 1. The flight paths (one second intervals) are shown for each facility sample used in the study. Perimeter flights are the large polygons, and plant flights the smaller ovals. Map layer data © Google Satellite Hybrid 2017.

flux divergence per lap estimate calculated in the first step of the SciAv algorithm. A flight is classified as "fully capturing" the emission plume when the mean flux divergence of the highest flight laps approaches zero. The standard SciAv emissions estimates assume that the flux divergence profile is constant below the lower flight altitude.

For this study, the use of different surface extrapolations for SciAv were developed and tested. To obtain a greater range of possible outputs from the algorithm, SciAv was fit using differing surface extrapolation methods: (1) constant (SciAv's standard of extending the lowest bin to the surface); (2) background (estimating flux divergence as zero by applying no extrapolation below the lowest profile point); (3) linear (estimating the linear trend of the profile points at a specific height that was chosen given the profile shape, location of the plume enhancement, and sparsity of points, and extrapolating the trend to the surface); (4) linear weighted (estimating the linear trend of the profile points, but weighted by the calculated flux divergence uncertainty for each lap estimate); (5) linear interpolation to background (fitting a line between the lowest profile bin), and (6) the average of the surface points calculated in methods (1)–(5) (Sect. S1.5).

The data used for both the TERRA and SciAv methods were identical, but due to different approaches to assessing conditions and analysis of the data, different error estimates were produced. Conley et al. (2017) found that binning by lap

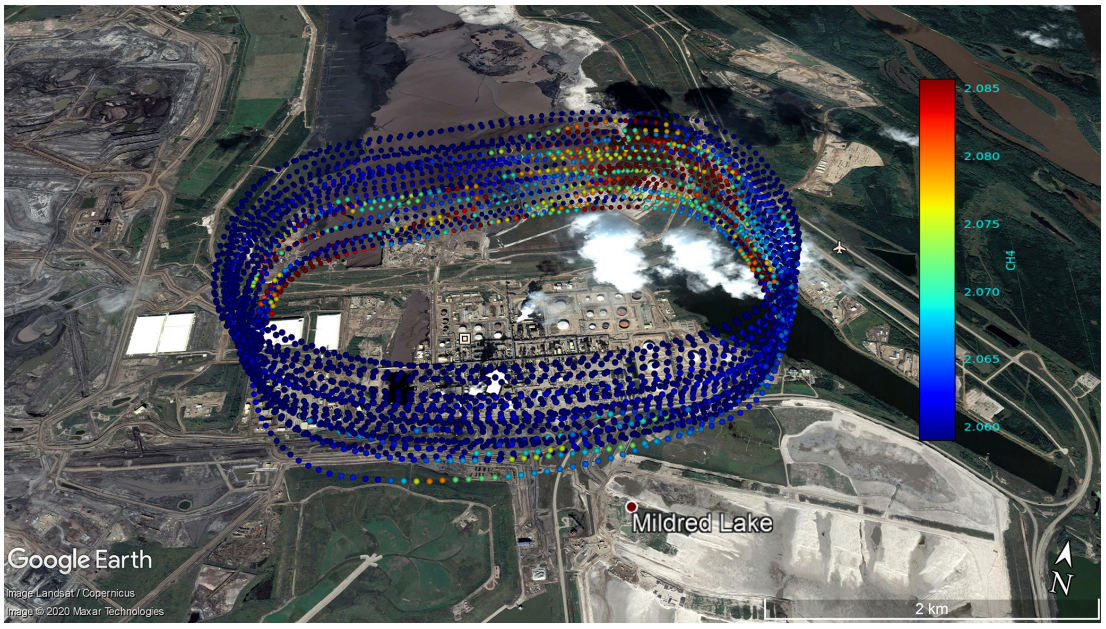


Figure 2. The flight path for F04 depicts the mixing ratio of CH_4 (ppm) measured at 1 Hz intervals for each of the 25 laps around the Syncrude plant. A Google Earth historical image from September 2016 was used, as it shows an emission plume with wind conditions similar to the 2017 F04 flight. A KML file containing mixing ratios provided by Scientific Aviation was overlaid on the image. Each measurement of a mixing ratio is depicted as a dot, and the layout traces the 25-lap flight path for sampling during F04. Satellite imagery © 2020 Maxar Technologies, Google Earth.

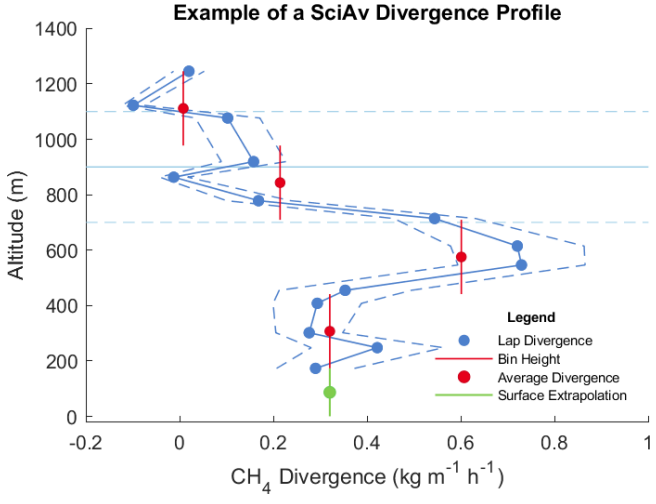


Figure 3. An example of a SciAv profile used in the second step of the method. The blue points are the estimated flux divergence for each lap, which are connected to show profile shape with the associated uncertainty (a dashed blue line). Red points are bin averages, and the vertical red bar is the bin height range. The boundary layer height is drawn in light blue with error bars (light blue dashed lines). The standard SciAv surface extrapolation method of extending the lowest red bin to the surface is shown in green.

and using a constant extrapolation produces a stable estimate when 20–25 laps are flown around an emission source (Conley et al., 2017). However, with larger area samples, such as perimeter flights, fewer than 10 laps are often flown. These types of samples may be better suited to a different type of integration as well as surface extrapolation. A potential method of improving estimation in the SciAv methods was investigated by using trapezoidal integration rather than binning to estimate a total emission estimate from the lap flux divergence points. Figure 4 depicts the two methods of integrating the SciAv flux divergence profiles as well as the different types of SciAv surface extrapolations, using the profile for F01 CH_4 as an example. The same surface extrapolation estimation procedure was used for both the binning and trapezoidal methods (Sect. S1.7). Surface extrapolation methods were fit to the lowest flux divergence lap point for the trapezoidal method to remain consistent with the SciAv method.

2.3 AVIRIS-NG aircraft emissions estimates

Three days prior to the F04 flight, a NASA Jet Propulsion Laboratory (JPL) AVIRIS-NG flight covered the Syncrude plant. This measurement was part of a larger Arctic-Boreal Vulnerability Experiment (ABoVE), which included flight lines flown over the Alberta oil sands region. AVIRIS-NG measures ground-reflected solar radiation (380–2500 nm) with a 34° field of view and a spectral resolution of 5 nm to map CH_4 plumes by utilizing absorption features in the

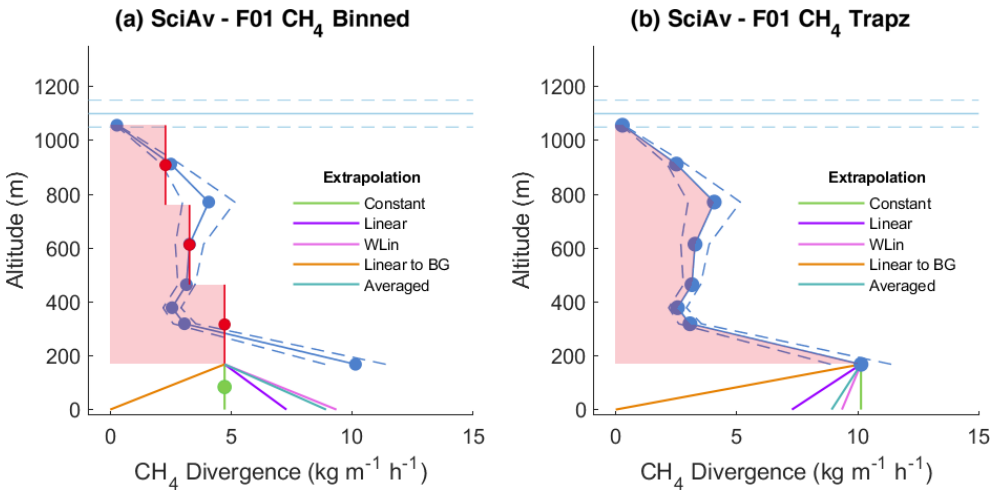


Figure 4. The two different integration methods applied to the SciAv F01 CH4 profile are depicted as the red area. Figure **(a)** shows the standard binning method of estimating an average flux divergence for each bin and integrating by altitude over the area as rectangular boxes. Figure **(b)** shows the trapezoidal (Trapz) method of estimating an average area under the curve by connecting the flux divergence lap points. The boundary layer height is drawn in light blue with error bars (light blue dashed lines). Both figures do not include the extra emissions that would be included using surface extrapolation.

shortwave infrared (Thorpe et al., 2017, 2020). As described in Duren et al. (2019), emissions estimates are calculated by combining the integrated mass enhancement (IME) and wind speed, as demonstrated in a number of recent studies throughout the United States (Cusworth et al., 2021, 2022) as well as a controlled release experiment (Thorpe et al., 2021). Previous studies have shown that the AVIRIS-based estimates of methane emissions agree well with box-flight emissions (Frankenberg et al., 2016; Duren et al., 2019; Thorpe et al., 2020). Figure 5 shows the AVIRIS-NG plume imagery that was captured over the Syncrude plant site with the Scientific Aviation KML lap data (shown in Fig. 2) overlaid. By measuring emissions at the same facility within a few days, this independent sample using a different method provides a contrast to the box-flight, mass-balance data.

AVIRIS-NG data were collected at 5.3 km (17.5 kft), and the Syncrude facility was not informed prior to sampling. NASA-JPL provided CH_4 emissions calculated using AVIRIS-NG data and three sources for hourly estimation of the wind: ECCC meteorological towers 3062696 and 3062697, and MERRA2 (Modern-Era Retrospective analysis for Research and Applications, version 2) reanalysis. MERRA2 is an atmospheric reanalysis method produced by NASA that utilizes numerous satellite observations to produce a global time series of atmospheric data (Gelaro et al., 2017). To estimate variability in wind speed, an average over a 3 h window was used for the met tower data, and nine kernels centred on the plume latitude and longitude were used for the MERRA2 analysis. The magnitude of the AVIRIS-NG estimates was then compared to the SciAv estimate as an independent way of evaluating the temporal consistency of emissions.

3 Results

3.1 Box-flight emissions estimate comparisons

Standard Scientific Aviation (SciAv) emission results were compared to the estimates produced by applying TERRA to the same flight data. A constant extrapolation to the surface was used for all SciAv samples, whereas the extrapolation for TERRA varied by the flight profile and source. The standard estimate results from both algorithms are shown in Table 3 and Fig. 6. RStudio (RStudio Team, PBC, Boston, Massachusetts, United States) was used for comparison analysis, and comparison figures were created using the ggplot2 package (Wickham, 2016). Standard emissions estimates for four of the five flights agree within their uncertainties. Confidence intervals for the estimates were not produced, as there is only one estimate for each flight, and therefore the error bars are simply the range for each estimate. In Fig. 6, the error bars for each estimate overlap with each other, aside from F04, which has a large gap between estimates. The uncertainty for the TERRA estimates are consistently smaller than for SciAv (averaging $\sim 8\,\%$ smaller).

Algorithm agreement is implied when the range for each estimate overlaps. For all flights except F04, the differences between the algorithms are in the range of the estimate uncertainties. For F04 the emissions estimates disagreed, as there is a large gap between the estimates with no overlap of the ranges. To compare the estimates using the uncertainty range, the relative mean percentage difference and propagated percentage uncertainty of the two estimates were calculated. The whole set of results from five flights was formally tested for differences between the SciAv and TERRA estimates using

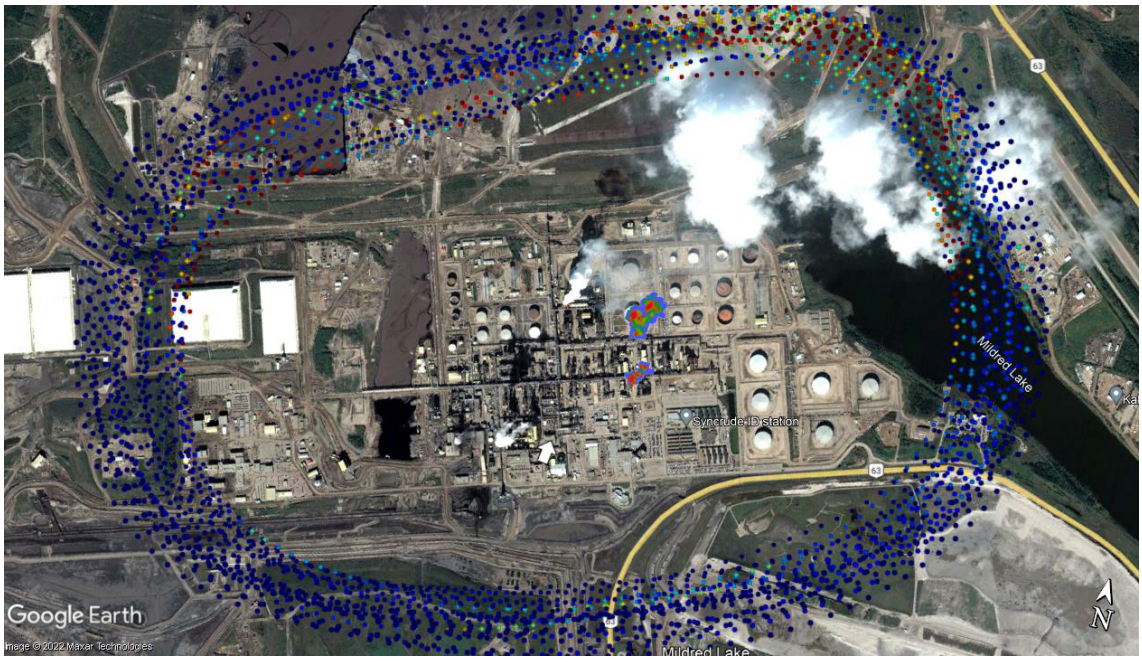


Figure 5. AVIRIS-NG-captured CH_4 column enhancements are shown from 11 August 2017 inside the F04 raw CH_4 lap data around the Syncrude plant from SciAv for 14 August 2017 using the Google Earth historical image from September 2016. AVIRIS-NG data and imagery provided by the NASA-JPL and satellite imagery © 2020 Maxar Technologies, Google Earth. Large CH_4 enhancements are depicted in red. The wind direction is shown by the white arrow, as measured by SciAv.

Table 3. Results of the CH_4 (in kilograms per hour) and CO_2 (in tonnes per hour) standard fit estimates from each algorithm, with their uncertainty as a percentage ± % and the range derived from the percentage uncertainty of each standard estimate.

Estimate type	F01	F02	F03	F04	F05
SciAv CH_4 ($kg\,h^{-1}$)	$3840 \pm 18\,\%$	$362 \pm 21\,\%$	$497 \pm 17\,\%$	$349 \pm 11\,\%$	$3470 \pm 29\,\%$
Range ($kg\,h^{-1}$)	(3150–4540)	(287–437)	(415–579)	(310–387)	(2480–4470)
TERRA CH_4 ($kg\,h^{-1}$)	$4810 \pm 11\,\%$	$395 \pm 6\,\%$	$476 \pm 5\,\%$	$125 \pm 18\,\%$	$3910 \pm 15\,\%$
Range ($kg\,h^{-1}$)	(4310–5300)	(373–418)	(452–501)	(102–148)	(3330–4490)
SciAv CO_2 ($t\,h^{-1}$)	$1040 \pm 22\,\%$	$563 \pm 18\,\%$	$526 \pm 11\,\%$	$1170 \pm 11\,\%$	$850 \pm 34\,\%$
Range ($t\,h^{-1}$)	(807–1270)	(464–662)	(469–583)	(1040–1300)	(561–1140)
TERRA CO_2 ($t\,h^{-1}$)	$1340 \pm 10\,\%$	$515 \pm 6\,\%$	$467 \pm 4\,\%$	$569 \pm 7\,\%$	$877 \pm 26\,\%$
Range ($t\,h^{-1}$)	(1200–1470)	(486–545)	(451–483)	(561–60)	(650–1110)

a weighted *t* test and Wilcoxon signed rank test. As a collective, the differences between the algorithms were found to be insignificant for both CH_4 and CO_2 (Sect. S1.6).

Large, anomalous differences between the SciAv and TERRA estimates occurred for F04. During screening of the flights, no issues with F04 were flagged (Sect. S1.8). This flight can be used as an example for improving SciAv flight screening and for assessing the implications when assumptions of a stationary plume and stable meteorological conditions are violated. No other flights had non-stationary conditions. The flight that was intentionally included as a poor-quality sample (F05), due to the large emission plume occurring at the highest altitude transects flown, has very good agreement between the two algorithms.

During the study design, F04 was considered an ideal sample, and the non-stationarity of the emission plume was not flagged until more in-depth analysis was applied to discern the reason for the large disagreement between the two algorithms. Over the course of F04, the concentrations of CH_4 and CO_2 changed, both within the plume (downwind of the plant) as well as in background air masses (upwind of the plant). Based on available information, it is unknown whether changes in facility emissions contributed to the observed changes in the plume's non-stationarity. It was noted that during sampling, facility operators instructed researchers that future flights could only sample the boundary of the facility. Operating conditions were not provided by the industry. NO_2 and SO_2 emissions are often used as tracer data for

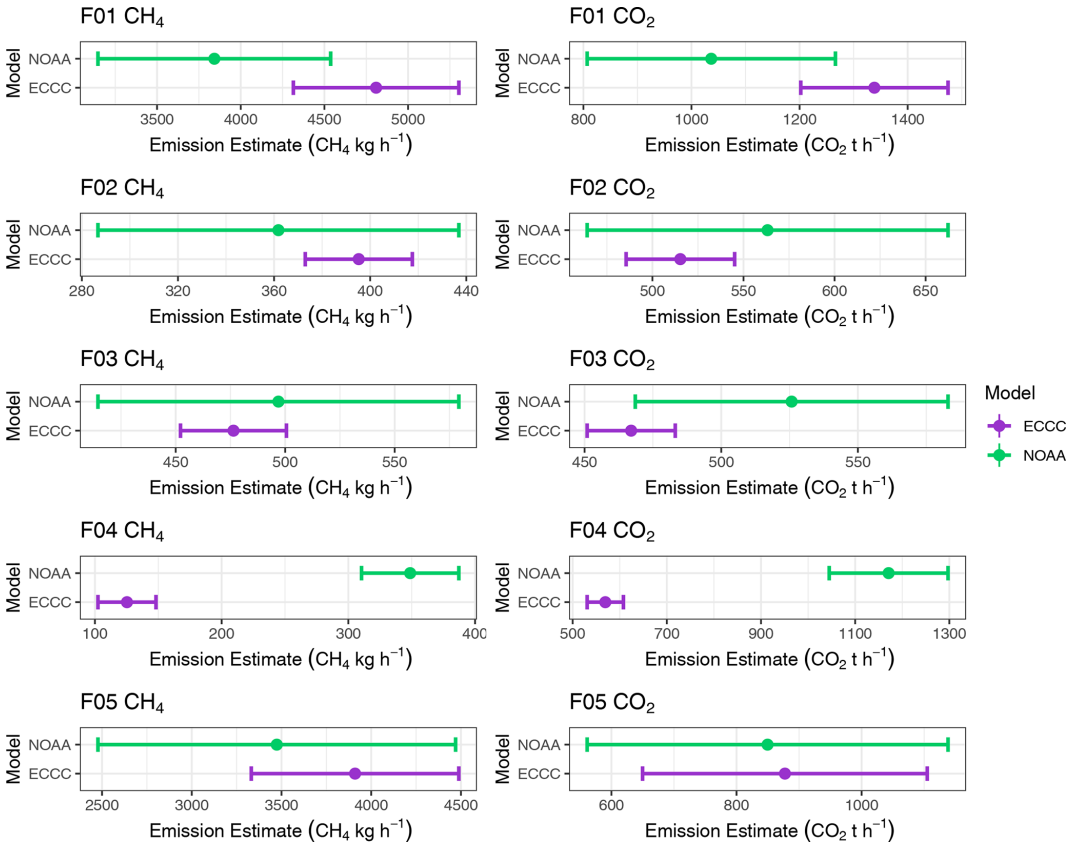


Figure 6. Flight emission estimates for CH_4 (left) and CO_2 (right) derived for each algorithm are plotted as points, along with the range of each estimate, as error bars. TERRA standard estimates are shown in green and SciAv in purple. Samples of F01, F02, F03, and F05 produce estimates that coincide within each method's error bars.

upscaling CH_4 emissions (Baray et al., 2018; Li et al., 2017; Liggio et al., 2019). Continuous emissions monitoring system (CEMS) facility stack emissions during F04 were consistent with typical operations and showed that NO_2 emissions spiked at the beginning of the day, before the aircraft measurements, and then continuously decreased. Furthermore, flaring data, including the volume of gas flared and SO_2 emissions, did not suggest unusual operations at the plant on the day of the flight. The disagreement between the two algorithm estimates for F04 arise from the non-stationary emission plume, which affected the box-flight mass-balance algorithms.

To test the effect of assumptions associated with plume shape below the lowest flight lap, various surface extrapolations were applied to the lowest bin of the SciAv flux divergence profiles for all five flights. The set of all results, based on the differing surface extrapolations, was compiled (Sect. S1.3 and S1.5), and estimates are plotted together in Fig. 7. Estimates that clustered together for both methods indicate a good agreement, with little difference between the varying surface extrapolation estimates. F04 has large disagreement between algorithms for both CH_4 and CO_2 (Fig. 7). The mean emission estimate and standard deviation of each method's various surface extrapolations were calculated (Sect. S1.6). The larger the spread in estimates, the more sensitive the flight was to the choice in extrapolation. Systematic bias is not evident in the differences between algorithms, as emissions estimates intersect, and no one method produces consistently larger or smaller estimates for all flights. To remove the effect of the choice in surface extrapolation, estimates were produced by background mixing ratios below the lowest flight path. These estimates were compared, and the SciAv and TERRA methods were still found to agree (Sect. S1.6).

To assess the sensitivity of emissions estimates using different surface extrapolations, the differences between each algorithm were calculated for the same four surface extrapolations (Sect. S1.6). For most flights, the choice in surface extrapolation had only a small effect on the difference between the estimates (≤ 3 %). The choice in surface extrapolation is a source of large variation between the algorithms for F01, the one flight with large emissions at the lowest flight path. The average of the estimates using the same four surface extrapolations was also computed. There is no evidence that agreement changes when removing the effect of surface

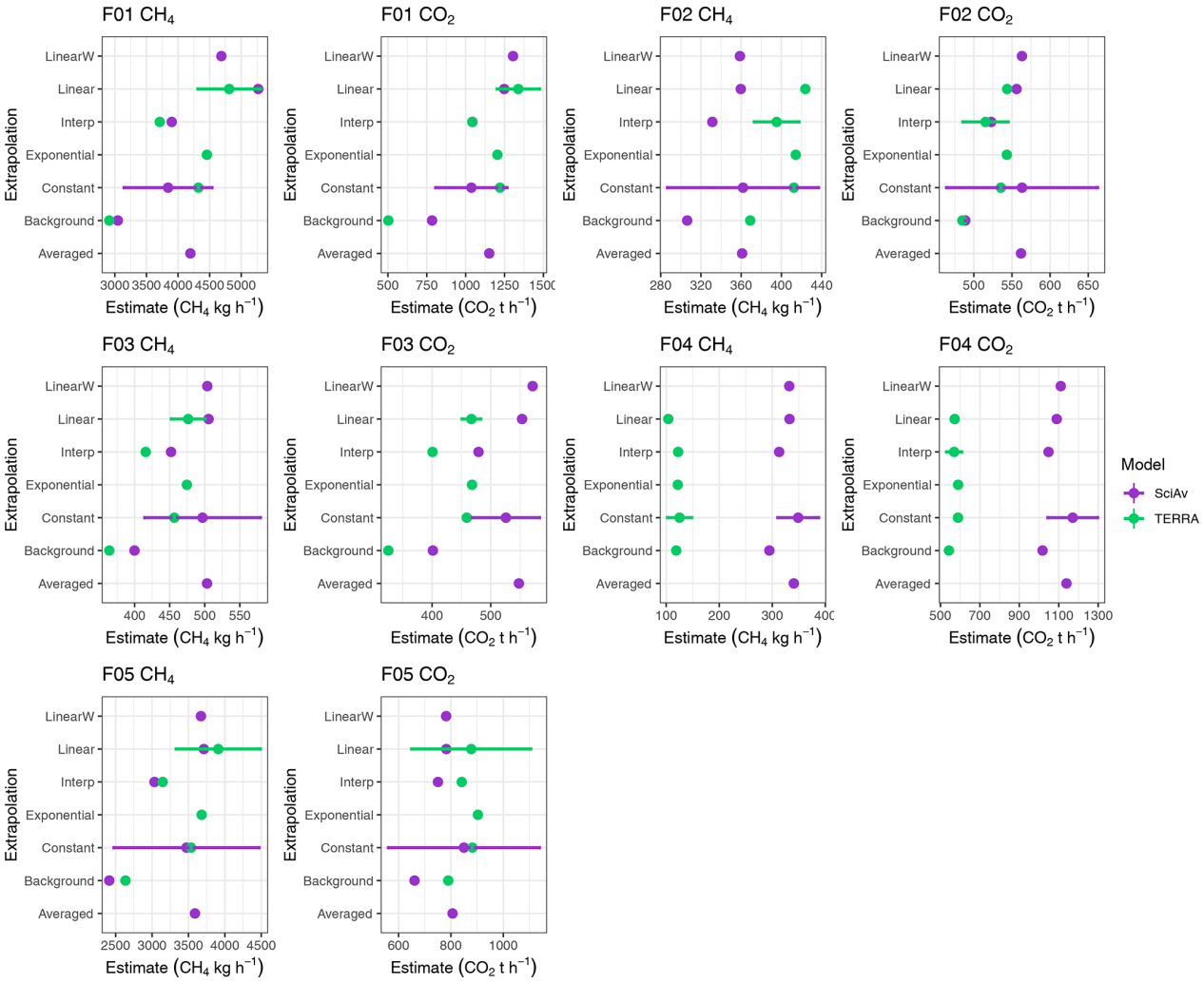


Figure 7. CH_4 and CO_2 estimates, based on various surface extrapolation fits, are plotted in purple for the SciAv estimate and green for TERRA. Error bars are drawn onto each algorithm's standard estimate.

extrapolation, and there is consistent agreement between the algorithm estimates (Sect. S1.6).

For each flight, algorithm estimates for the whole set of varying surface extrapolations were resampled using the bootstrap method (described below) to estimate the range in the difference between estimates based on the various surface extrapolations for each flight. The difference between the two algorithms, based on the various surface extrapolation estimates, was computed and contrasted with the standard error of each estimate. In Fig. 8, the distribution of the randomly sampled mean difference was calculated using the bootstrap method and plotted along with the propagated uncertainty range for each flight and gas (CH_4 and CO_2). A value of zero implies that there is no difference between the methods. Aside from F04, the distributions all either include zero, or the uncertainty of the standard estimates include zero, indicating that there is good agreement between the algorithms in most cases.

3.2 AVIRIS-NG aircraft emissions estimates

CH_4 enhancements were imaged at the Syncrude plant site within the perimeter of the Scientific Aviation box-flight path of F04 at 21:17:24 UTC on 11 August 2017 (Fig. 9), 3 d prior to the SciAv flight. There appeared to be two separate source plumes on that day, both of which were well inside the mass balance transects flown by Scientific Aviation during F04. The F04 flight also appeared to capture these two plumes (see Fig. S9). NASA-JPL provided data, analysis, and plume imagery using the methods described in Duren et al. (2019) over the F04 site to help provide additional context for the aircraft measurements (Table 4). The average instantaneous CH_4 emission rate of estimates derived using three different wind speed and direction datasets was 1665 ($kg\,h^{-1}$), with an average uncertainty of 707 ($kg\,h^{-1}$). This average emission rate likely reflects day-to-day emissions variability, as it was approximately 5 times larger than emissions measured

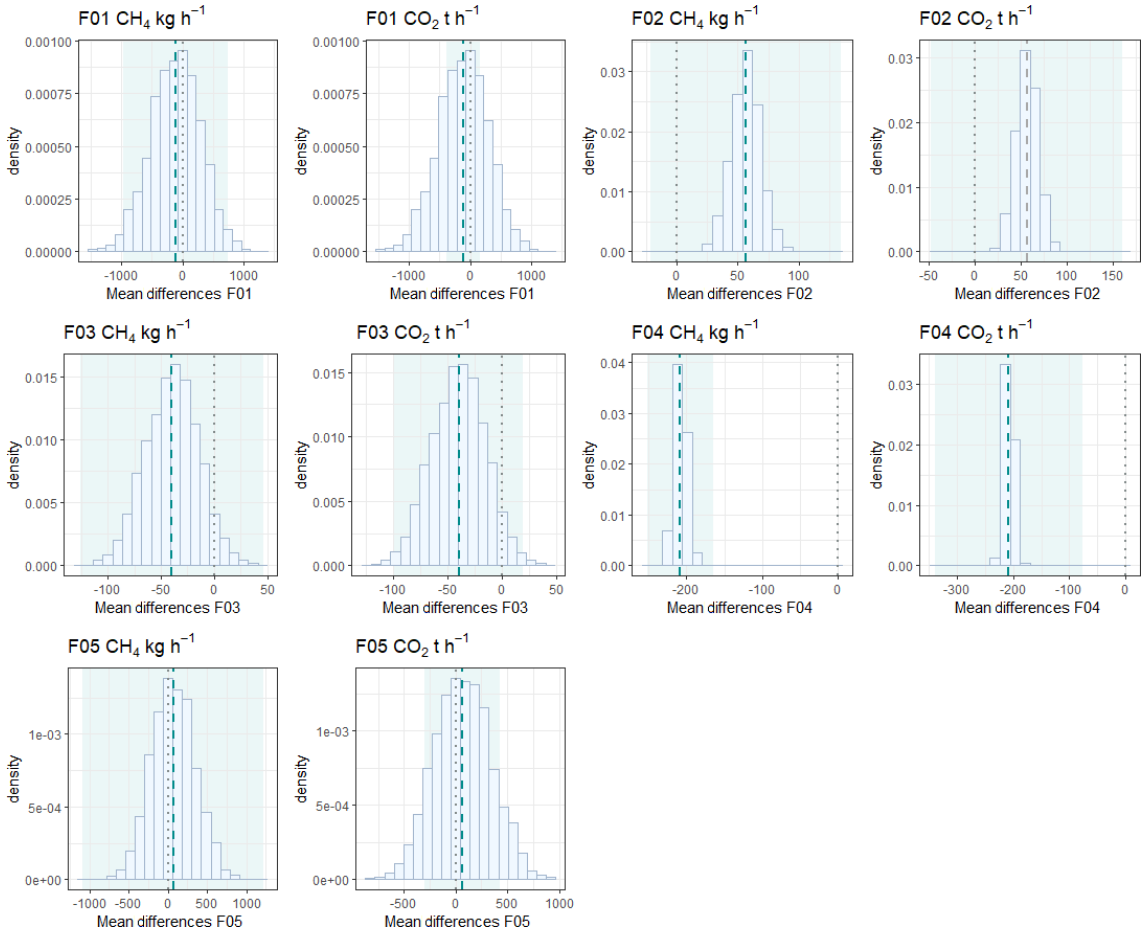


Figure 8. Distributions of the mean difference between all fits of CH_4 and CO_2 for the SciAv and TERRA algorithms are shown as a light blue histogram. The mean difference between the standard estimates is plotted as a teal dashed lined, and the range in the difference between standard estimates is shown as a light teal box. A grey dot dashed line is drawn at zero as a reference point for the location of exact agreement between the algorithms.

using the SciAv method 3 d later (349 kg h^{-1}). However, the AVIRIS-NG-derived emission rate was significantly less than the SciAv Syncrude perimeter estimate (F01, 3840 kg h^{-1} and F05, 3470 kg h^{-1}).

4 Discussion

4.1 Box-flight emissions estimate comparisons

In general, when fundamental assumptions were met, the SciAv and TERRA algorithms produced similar results. For the average flight scenario, the algorithm estimates derived using various surface extrapolations tended to agree regardless of how the surface extrapolation was fit. This consistency between estimates provides larger certainty in the estimates and in the top-down regional budgets that are inferred from them and also implies that emissions estimates between studies using different algorithms can be compared.

The SciAv and TERRA estimates were also compared when no surface extrapolation was applied (the background surface extrapolation scenario). These estimates also agreed, which suggests that the first steps in the core mass-balance algorithm produce similar outputs. Results from applying multiple surface extrapolations indicate that a potential difference between the methods may occur in the second algorithm step, due to the different methods of extrapolating to the surface, when an emission plume increases towards the surface.

Plots of the SciAv lap flux divergence estimates tend to follow three profile types. Examples of the three profile types for measured CH_4 lap enhancements (with uncertainties), as shown in Fig. 10, are (1) an emission plume with constant enhancements persisting at the lowest flight track (type I); (2) an elevated emission plume where enhancements approach zero at lower altitudes (type II); and (3) a plume that has enhancements increasing towards the surface (type III). All the profiles shown in Fig. 10 fully capture the top of each emission plume. Of the five sample flights compared, three

Table 4. AVIRIS-NG data captured on 14 August 2017, 3 d prior to the F04 flight, is estimated using three sources of wind data.

Estimate wind source	Average wind speed ($m\,s^{-1}$)	Wind speed uncertainty ($m\,s^{-1}$)	CH_4 estimate ($kg\,h^{-1}$)	CH_4 estimate uncertainty ($kg\,h^{-1}$)
Met 3062696	3.52	0.42	1767	744 (42 %)
Met 3062697	3.80	0.64	1907	834 (44 %)
MERRA2 reanalysis	2.62	0.235	1320	543 (41 %)
Average			1665	707 (41 %)

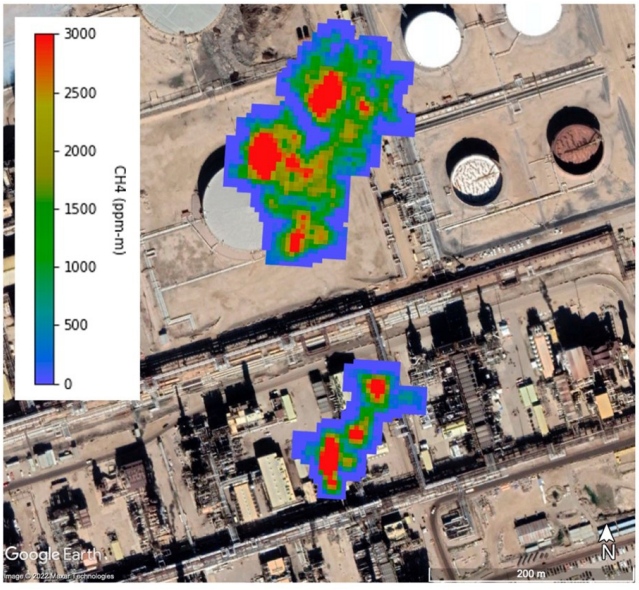


Figure 9. AVIRIS-NG observed CH_4 enhancements were imaged at the Syncrude plant site on 21:17:24 UTC. There appear to be two distinct sources located in close proximity. Satellite imagery © 2020 Maxar Technologies, Google Earth.

had clear profile shapes. The SciAv method of extrapolating as a constant is the most appropriate choice unless a type III pattern of increasing emissions at the lowest flight path is evident.

F01 was the only flight with a definitive SciAv profile type III and highlights the difference in the approaches to surface extrapolation when emissions increase at the lowest flight track. An "increasing-to-surface" fit for profile type III shapes would likely improve estimate accuracy for the SciAv algorithm. SciAv fits a constant surface extrapolation regardless of the behaviour of emissions at the lowest flight path, whereas TERRA chooses the surface extrapolation from various fits by assessing the plume. When a flight has large emissions at the bottom of the plume, the SciAv and TERRA methods agree more when some form of increasing-to-surface extrapolation is applied to SciAv compared to the use of the standard, constant extrapolation. This indicates that the second step of each algorithm, the choice of an appropriate surface extrapolation, is likely a significant factor contributing to any differences between the methods for type III profiles (see Fig. 10).

For the five flights compared, the uncertainty from TERRA is lower than that from the SciAv by an average of 8 %. The smaller uncertainty for the TERRA estimates may be due to how each algorithm quantifies error. TERRA calculates seven specific error terms to address the error of these assumptions. Increases in the error of one assumption do not directly increase the error of others (see Sect. S1.4). On the other hand, SciAv uses two main broad terms: a temporal error term to capture the extent of stationarity and a flux divergence error term to estimate capture of the plume. The flux divergence error of SciAv can become very large when only a few flight laps are flown. The uncertainty in the surface extrapolation from the SciAv algorithm may be reduced if it is decoupled from the current flux divergence error term and calculated following TERRA methods (i.e., using the maximum percentage change between probable fits).

The mean of the estimates derived using the six different surface extrapolations were calculated for each integration method's set of results, and differences were tested using a pairwise t test and Wilcoxon signed rank test. There was no evidence of a difference in mean estimates for each flight between the two integration methods evaluated (binning vs. trapezoidal; Sect. S1.7). This indicates that, for this adaption of the SciAv algorithm, choosing a different surface extrapolation is more important than the type of integration method. To further assess the effects of applying different surface extrapolation options, surface measurements at the time of sampling and more information about the behaviour of a plume are the most likely path towards further reducing the uncertainty for both algorithms.

4.2 Box-flight algorithm assumptions investigated

During the initial screening of F04 by members of Scientific Aviation and AEP, the non-stationary plume was not identified, as focus was placed on assessing meteorological conditions and plume capture. Meteorological conditions are often the most likely source of non-stationarity; as such both SciAv and TERRA methods apply a set of criteria to screen samples. TERRA also assesses conditions using explicit error terms (Sect. S1.4). Prior to the ad hoc analysis of splitting

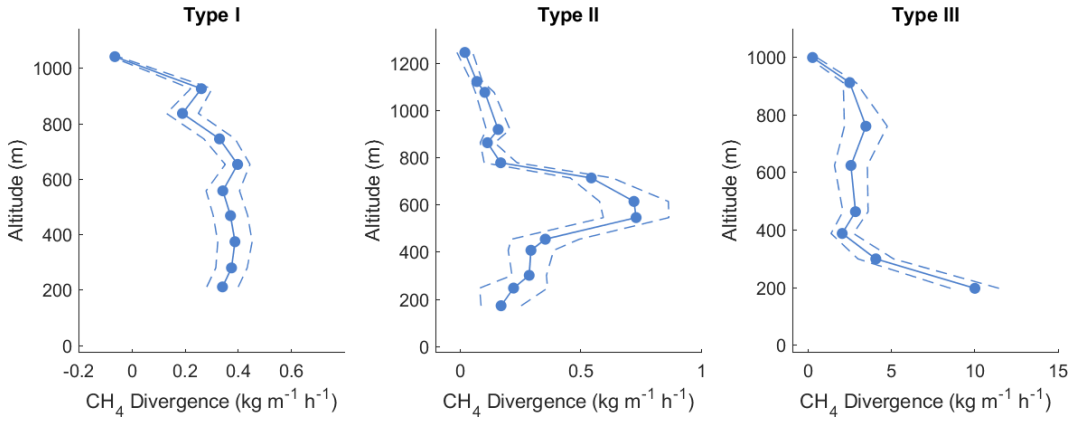


Figure 10. Sample data representing the three common types of flux divergence profiles for the SciAv method. Blue dots represent the flux divergence of each lap, with the associated uncertainty drawn as blue dashes.

the flight apart, the only measurement that might have indicated atmospheric instability was an air density error term calculated in TERRA. This produced an uncertainty estimate (4 %–6 %) that was noticeably larger than the other four flights (1 %–0.01 %) but was not an unusually high value for the method in general (Sect. S1.4).

The change in the emission plume during sampling was apparent in each method when the data were separated into the ascending and descending flight periods. In SciAv, emission enhancements for the flight laps going up in altitude differ noticeably from those flying laps going down (Fig. 11). In TERRA, ECCC split the flight data into the upward and downward portions, which were noticeably different (Sect. S1.8). For samples with many laps (≥ 20), separating the SciAv flux divergence profile into upward and downward flight components during the screening process would help identify the non-stationarity of an emission plume.

4.3 AVIRIS-NG aircraft emissions estimates

Methods based on imaging spectroscopy (e.g., AVIRIS-NG, GHGSat) provide a unique opportunity for emissions estimation, validation of ground-based measurements, and to help develop satellite monitoring techniques, while also providing leak detection (Cusworth et al., 2019; Frankenberg et al., 2016; Tyner and Johnson, 2021). CH_4 emissions from the oil and gas industry are sporadic, with higher emissions only captured 20 %–35 % of the time when sampling (Duren et al., 2019). While the SciAv and TERRA methods have lower uncertainties than the AVIRIS-NG method, they require prior knowledge of presumed sources. Therefore, the mass-balance methods are unlikely to identify unknown sources located outside sampling boundaries. The persistence of large, sporadic emissions, along with their relation to sampling with or without operator notice, should be studied further. While AVIRIS-NG estimates are dependent on estimates of the wind speed and direction as well as repeated sampling to assess source trends, they avoid the mass-balance requirements of a stationary source and the need to extrapolate emissions to the surface (Duren et al., 2019). This is because AVIRIS-NG effectively samples the entire atmospheric column between the ground and the sensor as a "snapshot".

In our study, the SciAv and TERRA algorithms yielded similar results when proper sampling conditions were met. While non-stationarity occurred for F04, the change in estimates between the upwards and downwards segments does not account for the large difference between the SciAv and AVIRIS-NG measurements of the Syncrude plant. Discrepancies between estimates due to missing, potentially large emissions below the lowest SciAv flight path are highly unlikely, as the F04 plume was "fully captured" with emissions decreasing towards the surface (Sect. S1.5). The substantially larger AVIRIS-NG estimate may be due to industrial operations, such a flaring or venting events. It is possible that only one of the two plumes observed by AVIRIS-NG were present when SciAv sampled, or that large day-to-day variability exists in CH_4 emissions. However, unlike the SciAv flight, operators were not informed before sampling, and operating conditions were not shared by the facility, so the underlying reason for this difference could not be evaluated. Due to the sporadic tendency of CH_4 emissions and the different sampling date, the AVIRIS-NG result is not directly comparable to the mass-balance algorithm's results for F04, but it can provide an idea of the range of potential emissions and specific source locations within the Syncrude plant. The SciAv and AVIRIS-NG methods have been independently compared and have been shown to provide consistently similar emissions estimates when employed under similar conditions (Frankenberg et al., 2016; Duren et al., 2019; Thorpe et al., 2020). Given the results of these studies, the AVIRIS-NG data captured on 14 August 2017 may not be anomalously high but could instead represent independent information on the variability of emissions from the region. Further work

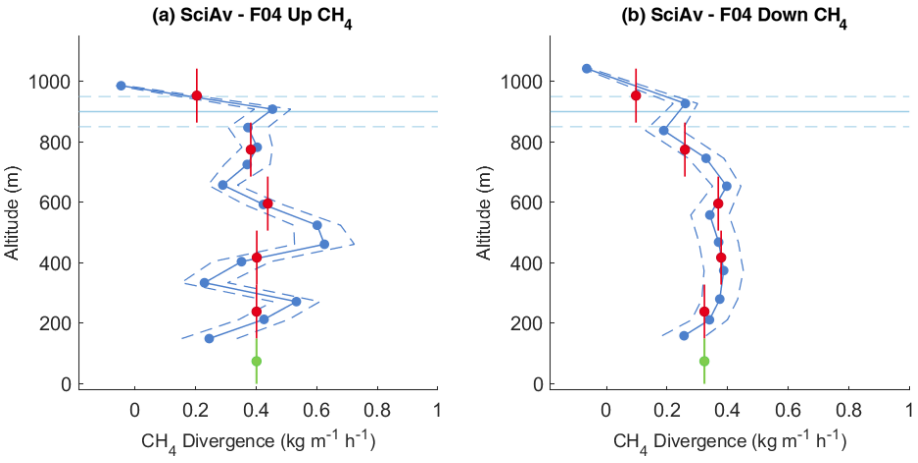


Figure 11. F04 CH_4 flux divergence profiles for the flight up **(a)** versus down **(b)**. The profile shapes differ, with a much more variable flux divergence in the up profile.

comparing these methods under similar conditions, along with greater transparency in facilities operations, could help confirm these conclusions and support more accurate emission budgets. The large emissions estimates from the same site outlines the importance of repeated sampling and the benefit of using multiple methods to characterize source behaviour, estimate the distribution of emissions from facilities, and estimate regional, national, and global emission budgets.

5 Conclusion

We found that emissions estimates were consistent between two top-down, mass-balance methods, providing confidence in these methods at a time when emissions reductions are needed. This finding is important, because airborne methods are used to validate top-down and bottom-up GHG emissions and to develop emissions inventories. When fundamental assumptions were met, the airborne mass-balance algorithms, SciAv and TERRA, produced similar estimates that agreed (3 %–25 %) within algorithm uncertainties (4 %–34 %). The two algorithms disagreed when the fundamental assumption of a stationary emission plume was not met (F04). Having increased confidence in estimates from the two mass-balance airborne methods provides a more certain foundation for policy and regulatory decisions. Including airborne imaging spectrometer emissions estimates in top-down regional budgets can provide additional information about emissions by capturing unknown sources or sporadic emissions. The ideal approach for characterizing and estimating GHG budgets would include repeated measurements, using a combination of airborne methods (in conjunction with new spectroscopic measurements from satellites for larger, continuous regional estimates), and by using ground-based equipment for small-scale point source quantification (Hardwick and Graven, 2016; National Academies of Sciences, Engineering, and Medicine, 2018; Saunois et al., 2020; Nisbet et al., 2020; Rutherford et al., 2021; Cusworth et al., 2022). Observations that combine and cross-validate multiple monitoring methods at varying scales of sampling will provide the most accurate modelling, improve GHG estimation, and help reconcile the often-reported gap between top-down and bottom-up estimates. Continued advances in developing more accurate inventories will allow for more effective policy and regulatory decisions that target the contribution of CO_2 and CH_4 to climate change.

Data availability. The data files for the five flights can be accessed through the Government of Alberta Portal: http://ckandata01.canadacentral.cloudapp.azure.com/dataset/aep-noaa-greenhouse-gas-measurement-flights (last access: 4 December 2021; Alberta Environment and Parks et al., 2021).

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/amt-15-5841-2022-supplement.

Author contributions. BME, JAG, and CA conceptualized the study with help from MLS. MLS, GRW, CA, and AD curated the data. BME did the primary formal analysis with contributions by MLS, AD, and AKT. Funding acquisitions were attained by BME. Investigations were conducted by CA, AD, GRW, and BME. Methodology was developed by BME and CA. Resources were provided by JAG, CEM, JL, SML, and SC. Software and programming were provided by SC, AD, AKT, and BME. JAG and CA primarily supervised the study with support by SC and JL. BME, CA, AD, GRW, SC, and JAG verified aspects of the research. BME, AD, and AKT contributed to the visualizations. BME created the original draft. JAG, CA, AKT, AD, GRW, JL, and SML reviewed and edited the work.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. The authors wish to thank the pilots of Scientific Aviation for conducting the field work, Quamrul Huda for contributing to the project design and coordination, and Bill Donahue for contributing to the project conceptualization and initiation of the Alberta Environment and Parks AEP-NOAA-Scientific Aviation 2017–2018 Alberta Oil Sands Flight Campaign. Thank you to Environment and Climate Change Canada's Mark Gordon for his advice on calculating the TERRA error terms, and Julie Narayan, who created the flight path boxes for TERRA. The AVIRIS-NG flights were supported by NASA's Terrestrial Ecology Program's Arctic-Boreal Vulnerability Experiment (ABoVE). Portions of this research were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004).

Financial support. Broghan M. Erland was supported by scholarships from Alberta Innovates (grant no. GSS – 1584401) and the University of Alberta, and through grants from Natural Sciences and Engineering Research Council (NSERC; grant no. CGS-M 767059), and Alberta Environment and Parks (grant no. 19GRAEM06). The work has been partially funded under the Oil Sands Monitoring Program. It is independent of any position of the OSM Program, or of the authors' institutions. Funding was provided to PI Charles E. Miller from the Terrestrial Ecology Program via ABoVE for AVIRIS-NG flights.

Review statement. This paper was edited by Huilin Chen and reviewed by two anonymous referees.

References

Alberta Environment and Parks (AEP), NOAA, Scientific Aviation, and UC Irvine: AEP-NOAA Greenhouse Gas Measurement Flights, Oil Sands Monitoring and Alberta Environment and Parks (OSM) and AEP [data set], http://ckandata01.canadacentral.cloudapp.azure.com/dataset/aep-noaa-greenhouse-gas-measurement-flights, last access: 4 December 2021.

Alfieri, S., Amato, U., Carfora, M. F., Esposito, M., and Magliulo, V.: Quantifying trace gas emissions from composite landscapes: A mass-budget approach with aircraft measurements, Atmos. Environ., 44, 1866–1876, https://doi.org/10.1016/j.atmosenv.2010.02.026, 2010.

Allen, D. T.: Methane emissions from natural gas production and use: reconciling bottom-up and top-down measurements, Curr. Opin. Chem. Eng., 5, 78–83, https://doi.org/10.1016/j.coche.2014.05.004, 2014.

Alvarez, R. A., Zavala-Araiza, D., Lyon, D. R., Allen, D. T., Barkley, Z. R., Brandt, A. R., Davis, K. J., Herndon, S. C., Jacob, D. J., Karion, A., Kort, E. A., Lamb, B. K., Lauvaux, T., Maasakkers, J. D., Marchese, A. J., Omara, M., Pacala, S. W., Peischl, J., Robinson, A. L., Shepson, P. B., Sweeney, C., Townsend-Small, A., Wofsy, S. C., and Hamburg, S. P.: Assessment of methane emissions from the U.S. oil and gas supply chain, Science, 361, 186–188, https://doi.org/10.1126/science.aar7204, 2018.

Atherton, E., Risk, D., Fougère, C., Lavoie, M., Marshall, A., Werring, J., Williams, J. P., and Minions, C.: Mobile measurement of methane emissions from natural gas developments in northeastern British Columbia, Canada, Atmos. Chem. Phys., 17, 12405–12420, https://doi.org/10.5194/acp-17-12405-2017, 2017.

Baray, S., Darlington, A., Gordon, M., Hayden, K. L., Leithead, A., Li, S.-M., Liu, P. S. K., Mittermeier, R. L., Moussa, S. G., O'Brien, J., Staebler, R., Wolde, M., Worthy, D., and McLaren, R.: Quantification of methane sources in the Athabasca Oil Sands Region of Alberta by aircraft mass balance, Atmos. Chem. Phys., 18, 7361–7378, https://doi.org/10.5194/acp-18-7361-2018, 2018.

Baray, S., Jacob, D. J., Maasakkers, J. D., Sheng, J.-X., Sulprizio, M. P., Jones, D. B. A., Bloom, A. A., and McLaren, R.: Estimating 2010–2015 anthropogenic and natural methane emissions in Canada using ECCC surface and GOSAT satellite observations, Atmos. Chem. Phys., 21, 18101–18121, https://doi.org/10.5194/acp-21-18101-2021, 2021.

Bartholomew, J., Lyman, P., Weimer, C., and Tandy, W.: Wide area methane emissions mapping with airborne IPDA lidar, in: Proc. SPIE, San Diego, California, United States, 30 August 2017, https://doi.org/10.1117/12.2276713, 2017.

Brandt, A. R., Heath, G. A., Kort, E. A., O'Sullivan, F., Pétron, G., Jordaan, S. M., Tans, P., Wilcox, J., Gopstein, A. M., Arent, D., Wofsy, S., Brown, N. J., Bradley, R., Stucky, G. D., Eardley, D., and Harriss, R.: Methane Leaks from North American Natural Gas Systems, Science, 343, 733–735, https://doi.org/10.1126/science.1247045, 2014.

Chan, E., Worthy, D. E. J., Chan, D., Ishizawa, M., Moran, M. D., Delcloo, A., and Vogel, F.: Eight-Year Estimates of Methane Emissions from Oil and Gas Operations in Western Canada Are Nearly Twice Those Reported in Inventories, Environ. Sci. Technol., 54, 14899–14909, https://doi.org/10.1021/acs.est.0c04117, 2020.

Conley, S., Faloona, I., Mehrotra, S., Suard, M., Lenschow, D. H., Sweeney, C., Herndon, S., Schwietzke, S., Pétron, G., Pifer, J., Kort, E. A., and Schnell, R.: Application of Gauss's theorem to quantify localized surface emissions from airborne measurements of wind and trace gases, Atmos. Meas. Tech., 10, 3345–3358, https://doi.org/10.5194/amt-10-3345-2017, 2017.

Crosson, E. R.: A cavity ring-down analyzer for measuring atmospheric levels of methane, carbon dioxide, and water vapor, Appl. Phys. B, 92, 403–408, https://doi.org/10.1007/s00340-008-3135-y, 2008.

Cusworth, D. H., Jacob, D. J., Varon, D. J., Chan Miller, C., Liu, X., Chance, K., Thorpe, A. K., Duren, R. M., Miller, C. E., Thompson, D. R., Frankenberg, C., Guanter, L., and Randles, C. A.: Potential of next-generation imaging spectrometers

to detect and quantify methane point sources from space, Atmos. Meas. Tech., 12, 5655–5668, https://doi.org/10.5194/amt-12-5655-2019, 2019.

Cusworth, D. H., Duren, R. M., Yadav, V., Thorpe, A. K., Verhulst, K., Sander, S., Hopkins, F., Rafiq, T., and Miller, C. E.: Synthesis of Methane Observations Across Scales: Strategies for Deploying a Multitiered Observing Network, Geophys. Res. Lett., 47, e2020GL087869, https://doi.org/10.1029/2020GL087869, 2020.

Cusworth, D. H., Duren, R. M., Thorpe, A. K., Olson-Duvall, W., Heckler, J., Chapman, J. W., Eastwood, M. L., Helmlinger, M. C., Green, R. O., Asner, G. P., Dennison, P. E., and Miller, C. E.: Intermittency of Large Methane Emitters in the Permian Basin, Environ. Sci. Tech. Let., 8, 567–573, https://doi.org/10.1021/acs.estlett.1c00173, 2021.

Cusworth, D. H., Thorpe, A. K., Ayasse, A. K., Stepp, D., Heckler, J., Asner, G. P., Miller, C. E., Chapman, J. W., Eastwood, M. L., Green, R. O., Hmiel, B., Lyon, D., and Duren, R. M.: Strong methane point sources contribute a disproportionate fraction of total emissions across multiple basins in the U.S., P. Natl. Acad. Sci. USA, 119, e2202338119, https://doi.org/10.1073/pnas.2202338119, 2022.

Dlugokencky, E. J., Nisbet, E. G., Fisher, R., and Lowry, D.: Global atmospheric methane: budget, changes and dangers, Philos. T. Roy. Soc. A, 369, 2058–2072, https://doi.org/10.1098/rsta.2010.0341, 2011.

Duren, R. M., Thorpe, A. K., Foster, K. T., Rafiq, T., Hopkins, F. M., Yadav, V., Bue, B. D., Thompson, D. R., Conley, S., Colombi, N. K., Frankenberg, C., McCubbin, I. B., Eastwood, M. L., Falk, M., Herner, J. D., Croes, B. E., Green, R. O., and Miller, C. E.: California's methane super-emitters, Nature, 575, 180–184, https://doi.org/10.1038/s41586-019-1720-3, 2019.

Fathi, S., Gordon, M., Makar, P. A., Akingunola, A., Darlington, A., Liggio, J., Hayden, K., and Li, S.-M.: Evaluating the impact of storage-and-release on aircraft-based mass-balance methodology using a regional air-quality model, Atmos. Chem. Phys., 21, 15461–15491, https://doi.org/10.5194/acp-21-15461-2021, 2021.

Foulds, A., Allen, G., Shaw, J. T., Bateson, P., Barker, P. A., Huang, L., Pitt, J. R., Lee, J. D., Wilde, S. E., Dominutti, P., Purvis, R. M., Lowry, D., France, J. L., Fisher, R. E., Fiehn, A., Pühl, M., Bauguitte, S. J. B., Conley, S. A., Smith, M. L., Lachlan-Cope, T., Pisso, I., and Schwietzke, S.: Quantification and assessment of methane emissions from offshore oil and gas facilities on the Norwegian continental shelf, Atmos. Chem. Phys., 22, 4303–4322, https://doi.org/10.5194/acp-22-4303-2022, 2022.

France, J. L., Bateson, P., Dominutti, P., Allen, G., Andrews, S., Bauguitte, S., Coleman, M., Lachlan-Cope, T., Fisher, R. E., Huang, L., Jones, A. E., Lee, J., Lowry, D., Pitt, J., Purvis, R., Pyle, J., Shaw, J., Warwick, N., Weiss, A., Wilde, S., Witherstone, J., and Young, S.: Facility level measurement of offshore oil and gas installations from a medium-sized airborne platform: method development for quantification and source identification of methane emissions, Atmos. Meas. Tech., 14, 71–88, https://doi.org/10.5194/amt-14-71-2021, 2021.

Frankenberg, C., Thorpe, A. K., Thompson, D. R., Hulley, G., Kort, E. A., Vance, N., Borchardt, J., Krings, T., Gerilowski, K., Sweeney, C., Conley, S., Bue, B. D., Aubrey, A. D., Hook, S., and Green, R. O.: Airborne methane remote measurements reveal heavy-tail flux distribution in Four Corners region, P. Natl. Acad. Sci. USA, 113, 9734–9739, https://doi.org/10.1073/pnas.1605617113, 2016.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S., Aragão, L. E. O. C., Arneth, A., Arora, V., Bates, N. R., Becker, M., Benoit-Cattin, A., Bittig, H. C., Bopp, L., Bultan, S., Chandra, N., Chevallier, F., Chini, L. P., Evans, W., Florentie, L., Forster, P. M., Gasser, T., Gehlen, M., Gilfillan, D., Gkritzalis, T., Gregor, L., Gruber, N., Harris, I., Hartung, K., Haverd, V., Houghton, R. A., Ilyina, T., Jain, A. K., Joetzjer, E., Kadono, K., Kato, E., Kitidis, V., Korsbakken, J. I., Landschützer, P., Lefèvre, N., Lenton, A., Lienert, S., Liu, Z., Lombardozzi, D., Marland, G., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Niwa, Y., O'Brien, K., Ono, T., Palmer, P. I., Pierrot, D., Poulter, B., Resplandy, L., Robertson, E., Rödenbeck, C., Schwinger, J., Séférian, R., Skjelvan, I., Smith, A. J. P., Sutton, A. J., Tanhua, T., Tans, P. P., Tian, H., Tilbrook, B., van der Werf, G., Vuichard, N., Walker, A. P., Wanninkhof, R., Watson, A. J., Willis, D., Wiltshire, A. J., Yuan, W., Yue, X., and Zaehle, S.: Global Carbon Budget 2020, Earth Syst. Sci. Data, 12, 3269–3340, https://doi.org/10.5194/essd-12-3269-2020, 2020.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J. Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Gordon, M., Li, S.-M., Staebler, R., Darlington, A., Hayden, K., O'Brien, J., and Wolde, M.: Determining air pollutant emission rates based on mass balance using airborne measurement data over the Alberta oil sands operations, Atmos. Meas. Tech., 8, 3745–3765, https://doi.org/10.5194/amt-8-3745-2015, 2015.

Gordon, M., Makar, P. A., Staebler, R. M., Zhang, J., Akingunola, A., Gong, W., and Li, S.-M.: A comparison of plume rise algorithms to stack plume measurements in the Athabasca oil sands, Atmos. Chem. Phys., 18, 14695–14714, https://doi.org/10.5194/acp-18-14695-2018, 2018.

Government of Canada: Canada's Climate Actions for a Healthy Environment and a Healthy Economy, Gatineau QC, Environment and Climate Change Canada, Report no. EC21125, 44 pp., 2021.

Hardwick, S. and Graven, H.: Satellite observations to support monitoring of greenhouse gas emissions, Grantham Institute, Imperial College London, Briefing paper No. 16, 16 pp., https://www.imperial.ac.uk/grantham/publications/briefing-papers/satellite-observations-to-support-monitoring-of-greenhouse-gas-emissions.php (last access: 11 August 2021), 2016.

IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., May-

cock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, in press, 2021.

Irakulis-Loitxate, I., Guanter, L., Maasakkers, J. D., Zavala-Araiza, D., and Aben, I.: Satellites Detect Abatable Super-Emissions in One of the World's Largest Methane Hotspot Regions, Environ. Sci. Technol., 56, 2143–2152, https://doi.org/10.1021/acs.est.1c04873, 2022.

Johnson, M. R. and Tyner, D. R.: A case study in competing methane regulations: Will Canada's and Alberta's contrasting regulations achieve equivalent reductions?, Elementa: Science of the Anthropocene, 8, 7, https://doi.org/10.1525/elementa.403, 2020.

Johnson, M. R., Tyner, D. R., Conley, S., Schwietzke, S., and Zavala-Araiza, D.: Comparisons of Airborne Measurements and Inventory Estimates of Methane Emissions in the Alberta Upstream Oil and Gas Sector, Environ. Sci. Technol., 51, 13008–13017, https://doi.org/10.1021/acs.est.7b03525, 2017.

Jongaramrungruang, S., Frankenberg, C., Matheou, G., Thorpe, A. K., Thompson, D. R., Kuai, L., and Duren, R. M.: Towards accurate methane point-source quantification from high-resolution 2-D plume imagery, Atmos. Meas. Tech., 12, 6667–6681, https://doi.org/10.5194/amt-12-6667-2019, 2019.

Kalthoff, N., Corsmeier, U., Schmidt, K., Kottmeier, C., Fiedler, F., Habram, M., and Slemr, F.: Emissions of the city of Augsburg determined using the mass balance method, Atmos. Environ., 36, 19–31, https://doi.org/10.1016/S1352-2310(02)00215-7, 2002.

Karion, A., Sweeney, C., Pétron, G., Frost, G., Michael Hardesty, R., Kofler, J., Miller, B. R., Newberger, T., Wolter, S., Banta, R., Brewer, A., Dlugokencky, E., Lang, P., Montzka, S. A., Schnell, R., Tans, P., Trainer, M., Zamora, R., and Conley, S.: Methane emissions estimate from airborne measurements over a western United States natural gas field, Geophys. Res. Lett., 40, 4393–4397, https://doi.org/10.1002/grl.50811, 2013.

Kort, E. A., Frankenberg, C., Costigan, K. R., Lindenmaier, R., Dubey, M. K., and Wunch, D.: Four corners: The largest US methane anomaly viewed from space, Geophys. Res. Lett., 41, 6898–6903, https://doi.org/10.1002/2014GL061503, 2014.

Krautwurst, S., Gerilowski, K., Borchardt, J., Wildmann, N., Gałkowski, M., Swolkień, J., Marshall, J., Fiehn, A., Roiger, A., Ruhtz, T., Gerbig, C., Necki, J., Burrows, J. P., Fix, A., and Bovensmann, H.: Quantification of CH_4 coal mining emissions in Upper Silesia by passive airborne remote sensing observations with the Methane Airborne MAPper (MAMAP) instrument during the CO_2 and Methane (CoMet) campaign, Atmos. Chem. Phys., 21, 17345–17371, https://doi.org/10.5194/acp-21-17345-2021, 2021.

Krings, T., Neininger, B., Gerilowski, K., Krautwurst, S., Buchwitz, M., Burrows, J. P., Lindemann, C., Ruhtz, T., Schüttemeyer, D., and Bovensmann, H.: Airborne remote sensing and in situ measurements of atmospheric CO_2 to quantify point source emissions, Atmos. Meas. Tech., 11, 721–739, https://doi.org/10.5194/amt-11-721-2018, 2018.

Lauvaux, T., Giron, C., Mazzolini, M., d'Aspremont, A., Duren, R., Cusworth, D., Shindell, D., and Ciais, P.: Global Assessment of Oil and Gas Methane Ultra-Emitters, Science, 375, 557–561, https://doi.org/10.1126/science.abj4351, 2022.

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., Pickers, P. A., Korsbakken, J. I., Peters, G. P., Canadell, J. G., Arneth, A., Arora, V. K., Barbero, L., Bastos, A., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Doney, S. C., Gkritzalis, T., Goll, D. S., Harris, I., Haverd, V., Hoffman, F. M., Hoppema, M., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Johannessen, T., Jones, C. D., Kato, E., Keeling, R. F., Goldewijk, K. K., Landschützer, P., Lefèvre, N., Lienert, S., Liu, Z., Lombardozzi, D., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S., Neill, C., Olsen, A., Ono, T., Patra, P., Peregon, A., Peters, W., Peylin, P., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rocher, M., Rödenbeck, C., Schuster, U., Schwinger, J., Séférian, R., Skjelvan, I., Steinhoff, T., Sutton, A., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Laan-Luijkx, I. T., van der Werf, G. R., Viovy, N., Walker, A. P., Wiltshire, A. J., Wright, R., Zaehle, S., and Zheng, B.: Global Carbon Budget 2018, Earth Syst. Sci. Data, 10, 2141–2194, https://doi.org/10.5194/essd-10-2141-2018, 2018.

Li, S.-M., Leithead, A., Moussa, S. G., Liggio, J., Moran, M. D., Wang, D., Hayden, K., Darlington, A., Gordon, M., Staebler, R., Makar, P. A., Stroud, C. A., McLaren, R., Liu, P. S. K., O'Brien, J., Mittermeier, R. L., Zhang, J., Marson, G., Cober, S. G., Wolde, M., and Wentzell, J. J. B.: Differences between measured and reported volatile organic compound emissions from oil sands facilities in Alberta, Canada, P. Natl. Acad. Sci. USA, 114, E3756–E3765, https://doi.org/10.1073/pnas.1617862114, 2017.

Liggio, J., Li, S. M., Staebler, R. M., Hayden, K., Darlington, A., Mittermeier, R. L., O'Brien, J., McLaren, R., Wolde, M., Worthy, D., and Vogel, F.: Measured Canadian oil sands CO_2 emissions are higher than estimates made using internationally recommended methods, Nat. Commun., 10, 1863, https://doi.org/10.1038/s41467-019-09714-9, 2019.

MacKay, K., Lavoie, M., Bourlon, E., Atherton, E., O'Connell, E., Baillie, J., Fougère, C., and Risk, D.: Methane emissions from upstream oil and gas production in Canada are underestimated, Scientific Reports, 11, 8041, https://doi.org/10.1038/s41598-021-87610-3, 2021.

National Academies of Sciences, Engineering, and Medicine: Improving Characterization of Anthropogenic Methane Emissions in the United States, The National Academies Press, Washington, DC, https://doi.org/10.17226/24987, 2018.

Nisbet, E. G. and Weiss, R.: Top-down versus bottom-up, Science, 328, 1241–1243, https://doi.org/10.1126/science.1189936, 2010.

Nisbet, E. G., Fisher, R. E., Lowry, D., France, J. L., Allen, G., Bakkaloglu, S., Broderick, T. J., Cain, M., Coleman, M., Fernandez, J., Forster, G., Griffiths, P. T., Iverach, C. P., Kelly, B. F. J., Manning, M. R., Nisbet-Jones, P. B. R., Pyle, J. A., Townsend-Small, A., Al-Shalaan, A., Warwick, N., and Zazzeri, G.: Methane Mitigation: Methods to Reduce Emissions, on the Path to the Paris Agreement, Rev. Geophys., 58, e2019RG000675, https://doi.org/10.1029/2019RG000675, 2020.

O'Shea, S. J., Allen, G., Gallagher, M. W., Bower, K., Illingworth, S. M., Muller, J. B. A., Jones, B. T., Percival, C. J., Bauguitte, S. J.-B., Cain, M., Warwick, N., Quiquet, A., Skiba, U., Drewer, J., Dinsmore, K., Nisbet, E. G., Lowry, D., Fisher, R. E., France, J. L., Aurela, M., Lohila, A., Hayman, G., George, C., Clark, D. B., Manning, A. J., Friend, A. D., and Pyle, J.: Methane and carbon dioxide fluxes and their regional scalability for the European Arctic wetlands during the MAMM project in summer 2012, Atmos. Chem. Phys., 14, 13159–13174, https://doi.org/10.5194/acp-14-13159-2014, 2014.

Rutherford, J. S., Sherwin, E. D., Ravikumar, A. P., Heath, G. A., Englander, J., Cooley, D., Lyon, D., Omara, M., Langfitt, Q., and Brandt, A. R.: Closing the methane gap in US oil and natural gas production emissions inventories, Nat. Commun., 12, 4715, https://doi.org/10.1038/s41467-021-25017-4, 2021.

Saunois, M., Stavert, A. R., Poulter, B., Bousquet, P., Canadell, J. G., Jackson, R. B., Raymond, P. A., Dlugokencky, E. J., Houweling, S., Patra, P. K., Ciais, P., Arora, V. K., Bastviken, D., Bergamaschi, P., Blake, D. R., Brailsford, G., Bruhwiler, L., Carlson, K. M., Carrol, M., Castaldi, S., Chandra, N., Crevoisier, C., Crill, P. M., Covey, K., Curry, C. L., Etiope, G., Frankenberg, C., Gedney, N., Hegglin, M. I., Höglund-Isaksson, L., Hugelius, G., Ishizawa, M., Ito, A., Janssens-Maenhout, G., Jensen, K. M., Joos, F., Kleinen, T., Krummel, P. B., Langenfelds, R. L., Laruelle, G. G., Liu, L., Machida, T., Maksyutov, S., McDonald, K. C., McNorton, J., Miller, P. A., Melton, J. R., Morino, I., Müller, J., Murguia-Flores, F., Naik, V., Niwa, Y., Noce, S., O'Doherty, S., Parker, R. J., Peng, C., Peng, S., Peters, G. P., Prigent, C., Prinn, R., Ramonet, M., Regnier, P., Riley, W. J., Rosentreter, J. A., Segers, A., Simpson, I. J., Shi, H., Smith, S. J., Steele, L. P., Thornton, B. F., Tian, H., Tohjima, Y., Tubiello, F. N., Tsuruta, A., Viovy, N., Voulgarakis, A., Weber, T. S., van Weele, M., van der Werf, G. R., Weiss, R. F., Worthy, D., Wunch, D., Yin, Y., Yoshida, Y., Zhang, W., Zhang, Z., Zhao, Y., Zheng, B., Zhu, Q., Zhu, Q., and Zhuang, Q.: The Global Methane Budget 2000–2017, Earth Syst. Sci. Data, 12, 1561–1623, https://doi.org/10.5194/essd-12-1561-2020, 2020.

Thorpe, A. K., Frankenberg, C., Thompson, D. R., Duren, R. M., Aubrey, A. D., Bue, B. D., Green, R. O., Gerilowski, K., Krings, T., Borchardt, J., Kort, E. A., Sweeney, C., Conley, S., Roberts, D. A., and Dennison, P. E.: Airborne DOAS retrievals of methane, carbon dioxide, and water vapor concentrations at high spatial resolution: application to AVIRIS-NG, Atmos. Meas. Tech., 10, 3833–3850, https://doi.org/10.5194/amt-10-3833-2017, 2017.

Thorpe, A. K., Duren, R. M., Conley, S., Prasad, K. R., Bue, B. D., Yadav, V., Foster, K. T., Rafiq, T., Hopkins, F. M., Smith, M. L., Fischer, M. L., Thompson, D. R., Frankenberg, C., McCubbin, I. B., Eastwood, M. L., Green, R. O., and Miller, C. E.: Methane emissions from underground gas storage in California, Environ. Res. Lett., 15, 45005, https://doi.org/10.1088/1748-9326/ab751d, 2020.

Thorpe, A. K., O'Handley, C., Emmitt, G. D., DeCola, P. L., Hopkins, F. M., Yadav, V., Guha, A., Newman, S., Herner, J. D., Falk, M., and Duren, R. M.: Improved methane emission estimates using AVIRIS-NG and an Airborne Doppler Wind Lidar, Remote Sens. Environ., 266, 112681, https://doi.org/10.1016/j.rse.2021.112681, 2021.

Tyner, D. R. and Johnson, M. R.: Where the Methane Is–Insights from Novel Airborne LiDAR Measurements Combined with Ground Survey Data, Environ. Sci. Technol., 55, 9773–9783, https://doi.org/10.1021/acs.est.1c01572, 2021.

Wickham, H.: ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag New York [code], ISBN: 978-3-319-24277-4, https://ggplot2.tidyverse.org (last access: 11 April 2022), 2016.

Wolfe, G. M., Kawa, S. R., Hanisco, T. F., Hannun, R. A., Newman, P. A., Swanson, A., Bailey, S., Barrick, J., Thornhill, K. L., Diskin, G., DiGangi, J., Nowak, J. B., Sorenson, C., Bland, G., Yungel, J. K., and Swenson, C. A.: The NASA Carbon Airborne Flux Experiment (CARAFE): instrumentation and methodology, Atmos. Meas. Tech., 11, 1757–1776, https://doi.org/10.5194/amt-11-1757-2018, 2018.

Yuan, B., Kaser, L., Karl, T., Graus, M., Peischl, J., Campos, T. L., Shertz, S., Apel, E. C., Hornbrook, R. S., Hills, A., Gilman, J. B., Lerner, B. M., Warneke, C., Flocke, F. M., Ryerson, T. B., Guenther, A. B., and de Gouw, J. A.: Airborne flux measurements of methane and volatile organic compounds over the Haynesville and Marcellus shale gas production regions, J. Geophys. Res.-Atmos., 120, 6271–6289, https://doi.org/10.1002/2015jd023242, 2015.

Supplement of Atmos. Meas. Tech., 15, 5841–5859, 2022
https://doi.org/10.5194/amt-15-5841-2022-supplement

Atmospheric Measurement Techniques
Open Access
EGU

Supplement of

Comparing airborne algorithms for greenhouse gas flux measurements over the Alberta oil sands

Broghan M. Erland et al.

Correspondence to: Broghan M. Erland (erland@ualberta.ca)

1 Mass-balance Box-flight Algorithm Theory Overview

The two mass-balance box-flight algorithms fundamentally differ in their parameterization of the emission flux and the mass-balance system of equations: SciAv simplifies the process to one flux and one mass-balance equation, whereas TERRA includes several flux terms to solve two mass-balance equations. Both algorithms require the same fundamental assumptions, but their approaches to quantifying the uncertainty in meeting those assumptions and estimating emissions differ. TERRA evaluates the entire dynamic system and estimates integral terms of a mass-balance equation used to derive an overall total flux (Gordon et al. 2015).

1.1 Top-Down Emission Rate Retrieval Algorithm (TERRA)

A strength of the TERRA model is its ability to capture and account for all the flux dynamics such as the emission vertical flux transport, chemistry, and deposition. It evaluates individual integral terms for the system of fluxes then adds them collectively. TERRA parameterizes a system of equations to derive flux and solves two mass-balance equations to estimate emissions (Gordon et al. 2015). Two common flux motions include advective transport following flow, and turbulent quasi-random, swirling traverses (Vinuesa and Galmarini 2009). The first mass balance equation for TERRA constrains the emission mixing ratio concentrations within the sampling box (the control volume), for each integral *E* term:

$$E_C = E_{C,H} + E_{C,HT} + E_{C,V} + E_{C,VT} + E_{C,VD} - E_{C,m} - E_{C,X} \quad (1)$$

Where *C* is the emission of interest (in this study CH_4 or CO_2), *H* the horizontal advective flux through the box walls, *HT* the horizontal turbulent flux through the box walls, *V* the advective flux through the box top, *VT* the turbulent flux through the box top, *VD* the deposition of flux to the ground, *m* the mass increase within the box due to air density change, and *X* the mass increase within the box due to chemical changes. Figure S 1 depicts the flux directions of each *E* integral term in Equation 1 given a cylindrical flight path to create the 'box'. The walls of the box are the boundaries of the flight laps and arrows depict the flux that is occurring given a western wind (from the left).

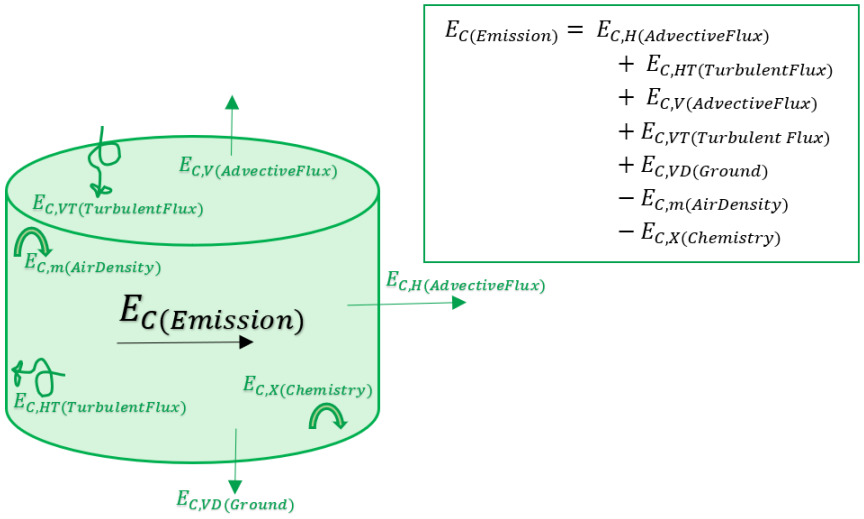


Figure S 1: The TERRA integral E terms of Equation 1 with arrows depicting their contribution to estimating emission flux through the 'box', or volume.

The second mass balance TERRA equation accounts for air flow in the control volume:

$$0 = E_{air,H} + E_{air,V} - E_{air,M} \quad (2)$$

Where the three terms are: *H* the horizontal advective flux of air through the sides of the box, *V* the advective flux through the box top, and *M* the change in the airmass within the box respectively. The integrals are estimated using input variables derived directly from the raw data collected, or through functional equations as described in the methodology paper published by Gordon et al. (2015).

There are seven error terms (δ) defined in methodology of TERRA that relate to the calculation of the integral terms in Equation 1 which are added in quadrature to estimate the total uncertainty (δ^2_{TERRA}) in the emission rate estimate (Gordon et al. 2015).

$$\delta^2_{TERRA} = \delta^2_M + \delta^2_{Ex} + \delta^2_{Wind} + \delta^2_{Top} + \delta^2_{dens} + \delta^2_{VT} + \delta^2_{BH} \quad (3)$$

The seven error terms in Equation 3 pertain to the uncertainty in: measurement (δ^2_M), near surface extrapolation of the mixing ratio (δ^2_{Ex}), near surface extrapolation of the wind (δ^2_{Wind}), the box-top mixing ratio (δ^2_{Top}), the change in air density (δ^2_{dens}), the vertical turbulent flux (δ^2_{VT}), and the boundary layer height (δ^2_{BH}). The largest error term of the method is the extrapolation of the mixing ratio from the lowest flight to the near-surface (Gordon et al. 2015). Mixing ratio surface extrapolation is chosen as either background, constant, background to constant, linear, or as an exponential fit depending on the location and dispersion of the plume and boundary layer conditions (Gordon et al. 2015). The contribution of the error of the vertical turbulent mixing term has been found to be functionally negligible and has been dropped from the overall calculation in TERRA (Baray et al. 2018; Gordon et al. 2015). The other uncertainty terms of TERRA are calculated from measuring the wind and mixing ratios, estimating the range in the box-top mixing ratio, calculating the uncertainty due to the height flown to capture the plume, and assessing atmospheric stability from the temperature and pressure ratios (Gordon et al. 2015).

1.2 SciAv Gaussian Plume Inversion Algorithm

The SciAv model simplifies the dynamics within the box to estimate flux due to emissions from within the box. It assumes that the vertical flux is zero and that there is no flux deposition to the ground, so that only the horizontal flux needs to be estimated. As the name would suggest the Gaussian theorem algorithm utilizes Gauss's theorem, also known as the Divergence Theorem, for relating a volume integral to an integral of a surface (the laps) enclosing the volume (the box). Conley at al. (2017) use the measurements to directly evaluate the balance budget of Equation 4 as described in their methods:

$$Q_c = \langle \frac{\partial m}{\partial t} \rangle + \int_0^{z_{max}} \oint c' u_h \cdot \hat{n} dl dz \quad (4)$$

The definition of terms and diagram of the SciAv method given a cylindrical 'box', derived from the laps of the flight path, is shown in Figure S 2. Equation 4 is used to calculate the mean flux divergence for each lap. An enhancement of the mixing ratio occurs when the emission plume is captured within a lap and a positive divergence is calculated. Bins are created for ranges of altitudes and the lap estimates are aggregated into bins according to altitude and average flux values are estimated for each bin as described by Conley et al 2017. The bin averages are multiplied by the height of their respective bin then summed together. The lowest bin average is extrapolated to the ground as a constant to estimate the emission to the surface, multiplied by its bin height, then added to the other bin estimates to produce the total emission. The final flight emission estimate is the result of this discrete integration of bins of averaged lap estimates over the entire flight altitude.

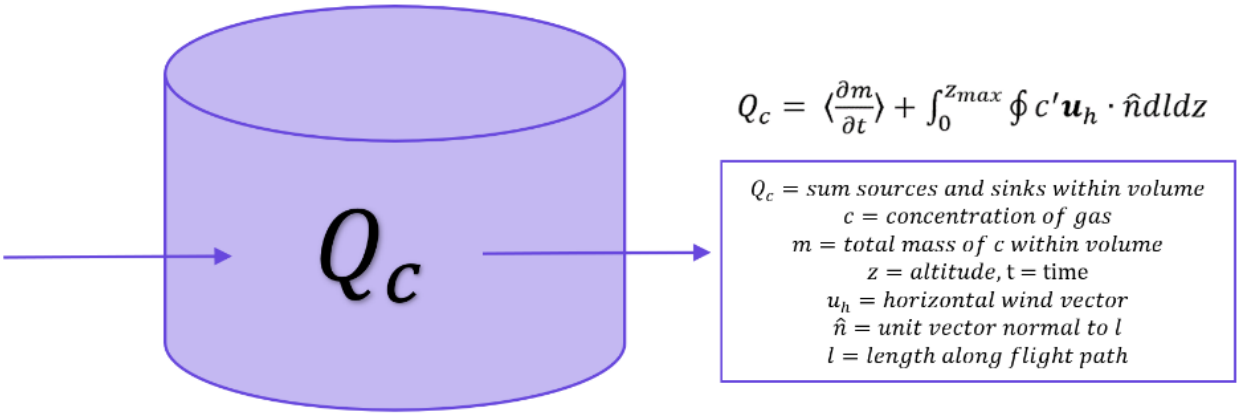


Figure S 2: The SciAv method simplifies the mass-balance equation to one instantaneous flux to calculate the overall horizontal sum of the source and sinks within the 'box', or volume.

SciAv defines three error terms that are added in quadrature to estimate the uncertainty (δ^2_{SciAv}) in the final emission estimate:

$$\delta^2_{\text{SciAv}} = \delta^2_{\text{M}} + \delta^2_{\text{flux divergence}} + \delta^2_{\text{temporal}} \quad (5)$$

The first term is the measurement error (δ^2_{M}) which contributes the smallest amount to the final estimate error (Conley et al. 2017). The second is the flux divergence error ($\delta^2_{\text{flux divergence}}$) and is calculated by summing the variance of the flux divergence estimates within each bin. The error of the surface extrapolation is estimated as twice the error of the lowest bin

and included in the flux divergence term (Conley et al. 2017). The flux divergence error accounts for the stochastic variation in plume capture for each lap and is often the largest component of the total error. The third error term is the temporal variance ($\delta^2_{\text{temporal}}$) and is a measurement of the stationarity of the plume. The error due to the time rate of change is extracted from a regression of the emission density over altitude and time (Conley et al. 2017).

Use of the TERRA Algorithm in Further Detail

To use TERRA (top-down emission rate retrieval algorithm), the appropriate surface extrapolation needed to be chosen and the emission screens assessed, the error terms calculated, and a background value calculated to use specific fits for each flight. This section outlines the processes to evaluate those four requirements as instructed by Environment and Climate Change Canada (ECCC).

1.3 TERRA Flight Screens

The emission screens produced by TERRA after fitting the chosen surface extrapolation are shown in Figure S 3 - 12. The left y-axis gives the altitude above ground level in meters, the right hand the colour scale of CH_4, or CO_2 in ppm. The length along each lap, denoted as s(m), is plotted for each s location in meters along the x-axis with the direction of sampling overlaid. The surface is shown as grey and the gap between sampling has been filled in by the surface extrapolation. Most figures show a concentrated plume surrounded by a blue of background mixing ratio concentrations. The extent of the dispersion of the F04 CH_4 is noticeable in Figure S 9. The surface extrapolation sometimes estimates a decreasing emission plume towards the surface as per Figure S 4 and S 12. This leads to a larger range in the mixing ratio and a change from the typical royal blue background colour to a lighter shade such as cyan, or even light green to adjust for the lower scale. The background values are not affected by the change in colour. Aside from F04, the flight data used in the comparison analysis represent standard emissions screens for the TERRA method.

To use the interpolate, or background fit options, TERRA requires estimates of background concentrations of the desired gas present in the atmosphere, unrelated to the emission source. As part of ECCC's methodology for TERRA, ideally independent samples are gathered to estimate the background mixing ratio value of each gas for a box-flight. Background values are used in the surface extrapolation for the "background" and "linear interpolate to background" fits. The data gathered by Scientific Aviation for this analysis did not have independent samples, so the background mixing ratios were determined by inspecting the histogram of concentrations, removing the tail of enhanced emissions, then fitting a normal distribution to the values and estimating the background value as the mean of the distribution. Results are given in Table S 1. The background values used are given in Table S 2.

100

Table S 1: TERRA emission rate estimates using the method's five surface extrapolation fits in kilograms per hour for CH_4 and tons per hour for CO_2.

Flight	Surface Extrapolation	CH_4 Emission ($kg\ h^{-1}$)	CO_2 Emission ($t\ h^{-1}$)
F01	Constant	4320	1220
F01	Linear	4810	1340
F01	Interpolate	3710	1040
F01	Background	2910	503
F01	Exponential	4460	1200
F02	Constant	412	535
F02	Linear	424	544
F02	Interpolate	395	515
F02	Background	369	485
F02	Exponential	414	543
F03	Constant	457	459
F03	Linear	476	467
F03	Interpolate	416	401
F03	Background	364	326
F03	Exponential	475	468
F04	Constant	125	589
F04	Linear	104	572
F04	Interpolate	123	569
F04	Background	119	543
F04	Exponential	122	589
F05	Constant	3540	882
F05	Linear	3910	877
F05	Interpolate	3150	8410
F05	Background	2630	790
F05	Exponential	3680	1220

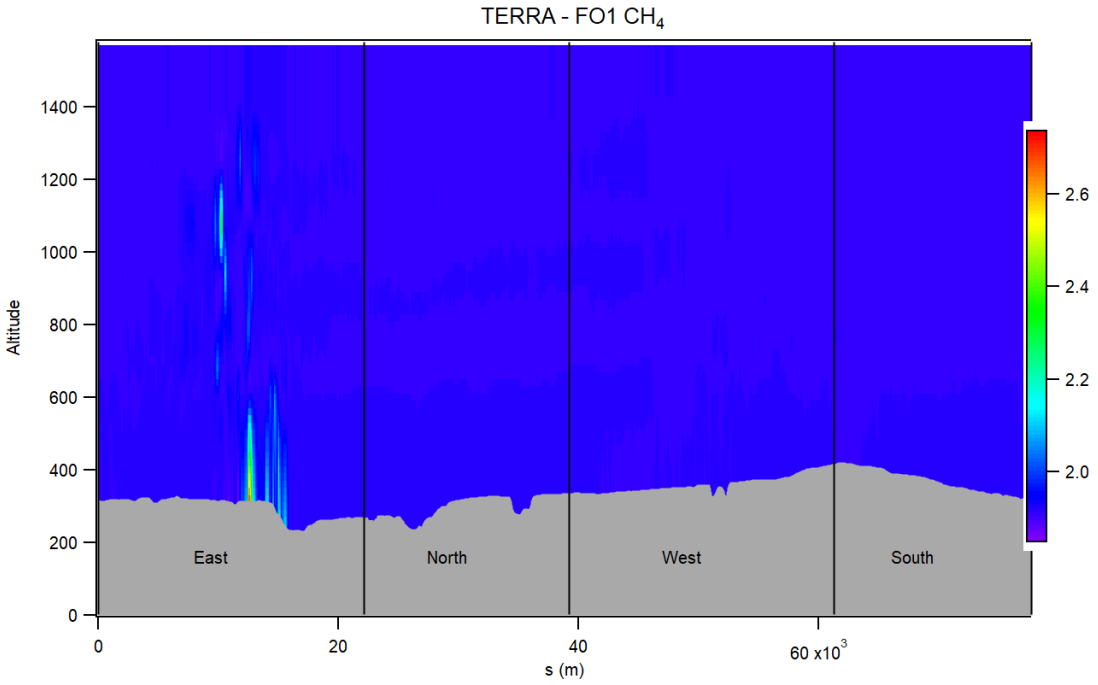


Figure S 3: The TERRA screen for F01 CH_4 in ppm. Altitude measured in meters is shown along the left y-axis and the colour bar on the right depicts the mixing ratio gradient in ppm. The s location along each lap in meters, all in scientific notation, is plotted along the x-axis as s(m) with the direction of sampling overlaid. The ground is shown in grey.

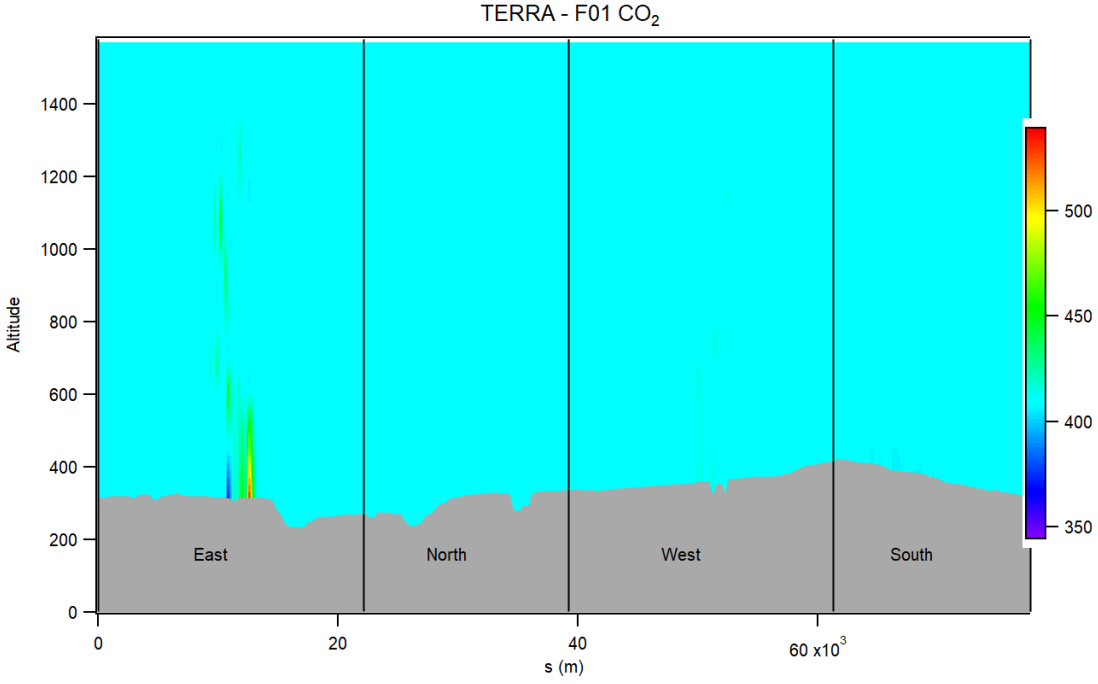


Figure S 4: Same as Figure S 3 for flight F01 CO_2 (ppm).

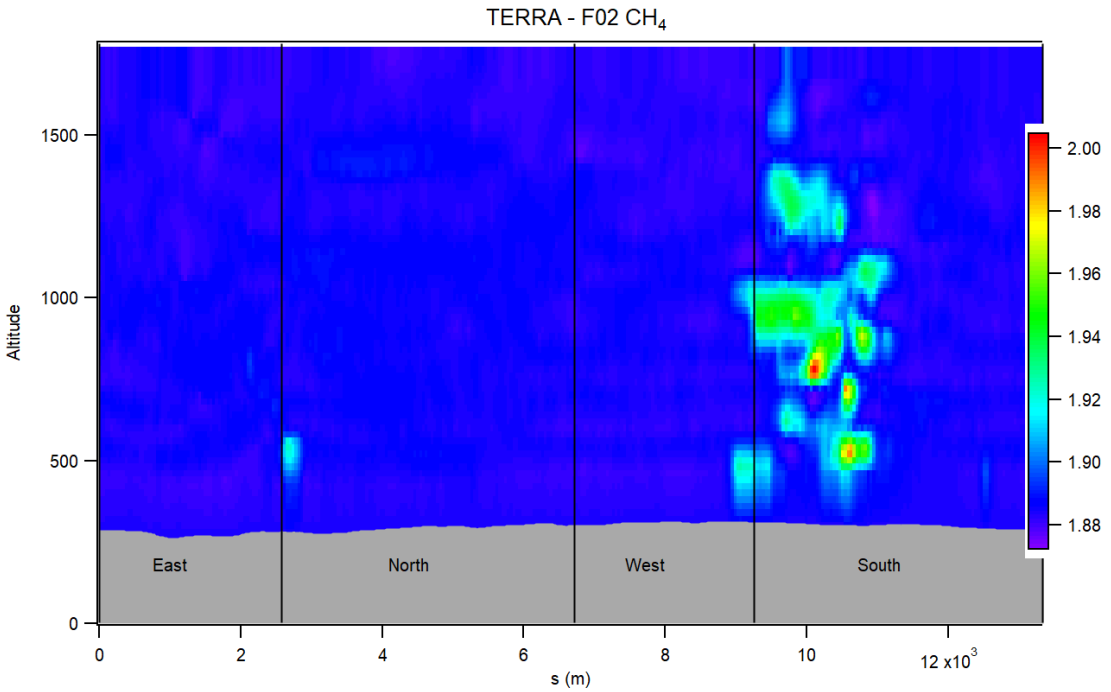


Figure S 5: Same as Figure S 3 for flight F02 CH_4 (ppm).

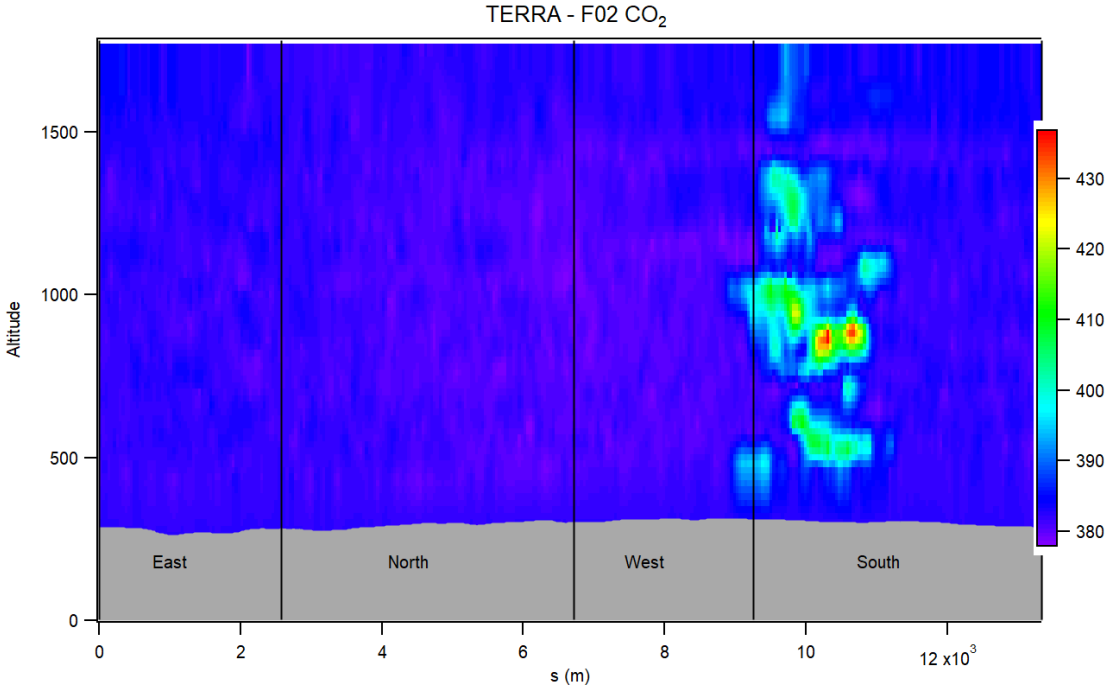


Figure S 6: Same as Figure S 3 for flight F02 CO_2 (ppm).

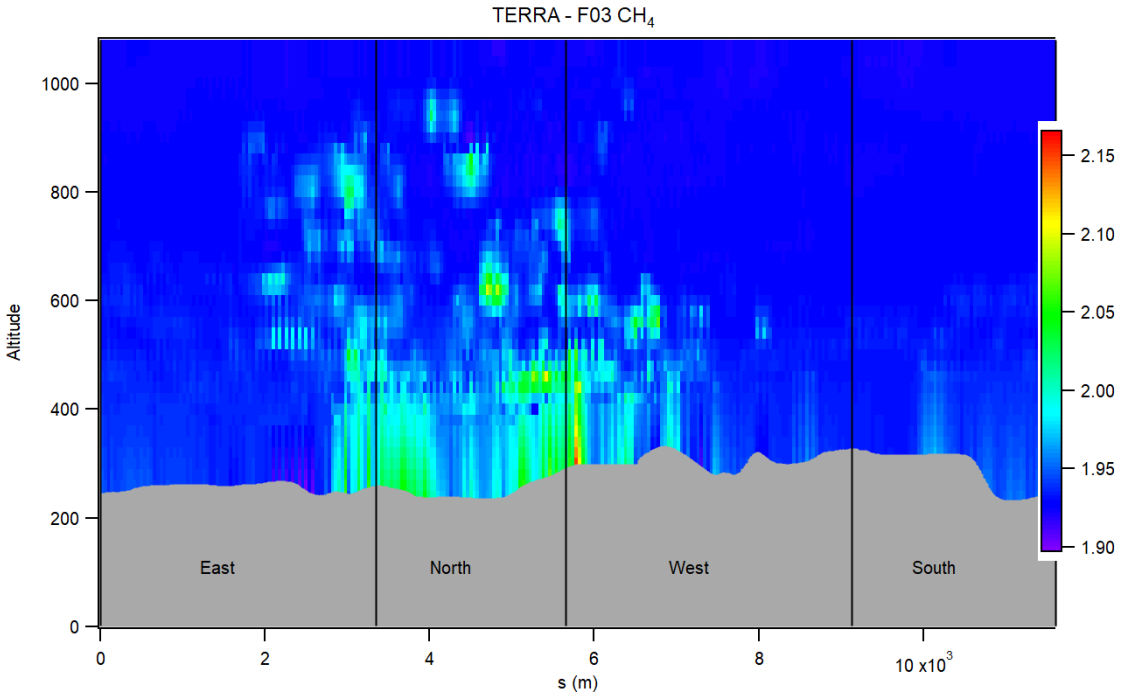
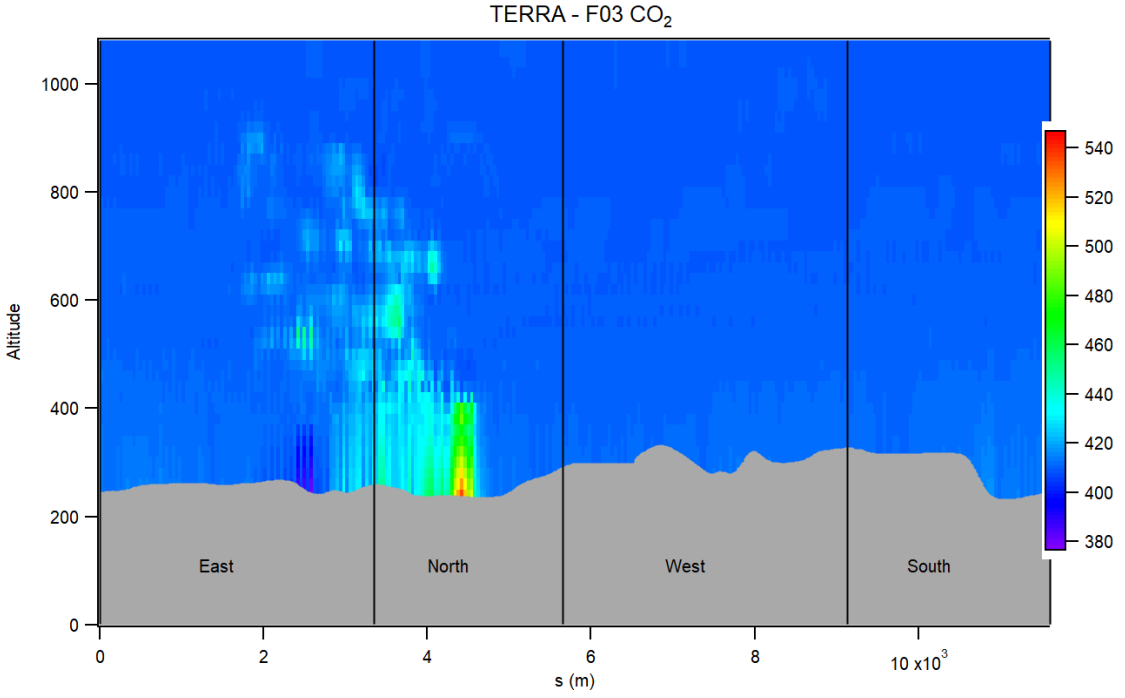


Figure S7: Same as Figure S 3 for flight F03 CH_4 (ppm).



115

Figure S 8: Same as Figure S 3 for flight F03 CO_2 (ppm).

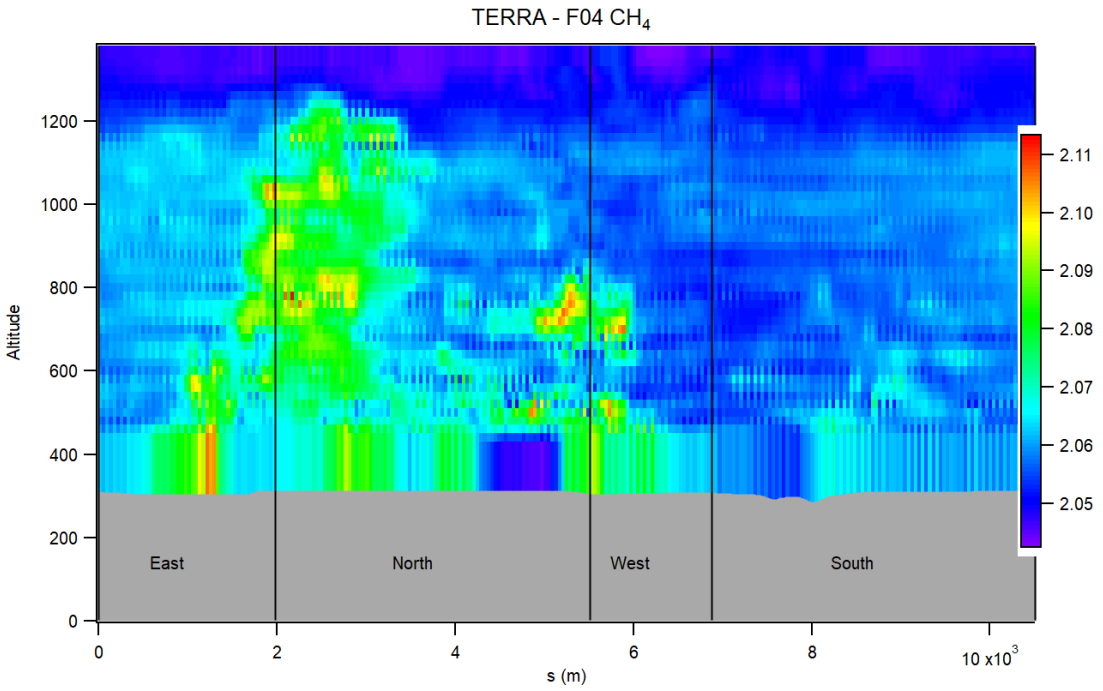


Figure S 9: Same as Figure S 3 for flight F04 CH_4 (ppm).

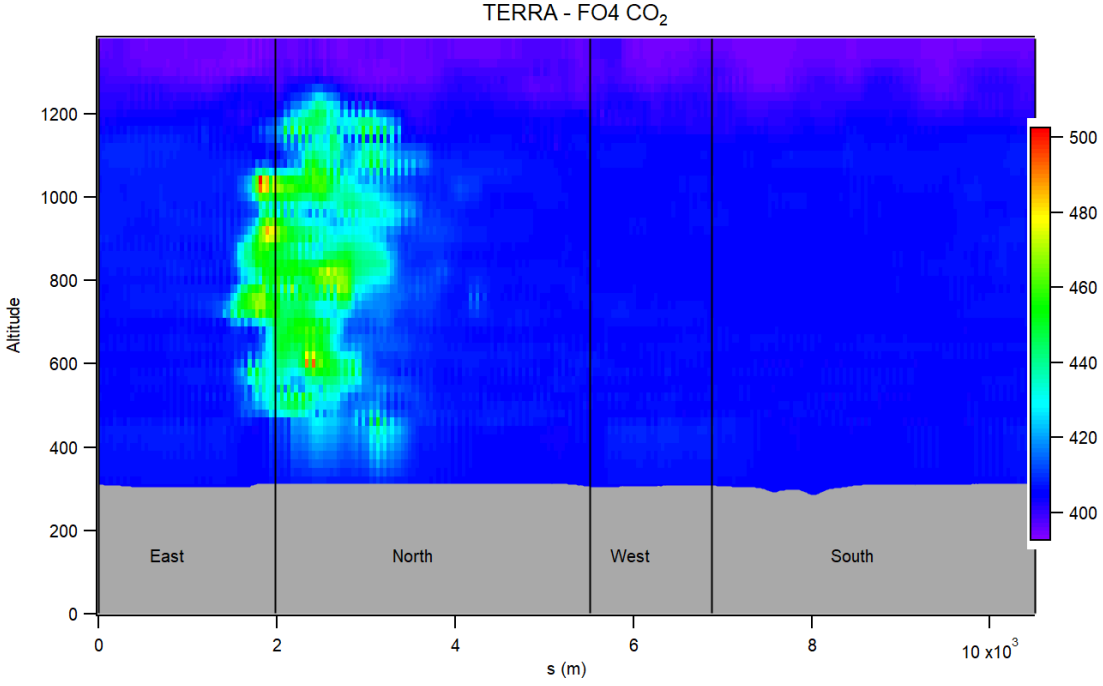


Figure S 10: Same as Figure S 3 for flight F04 CO_2 (ppm).

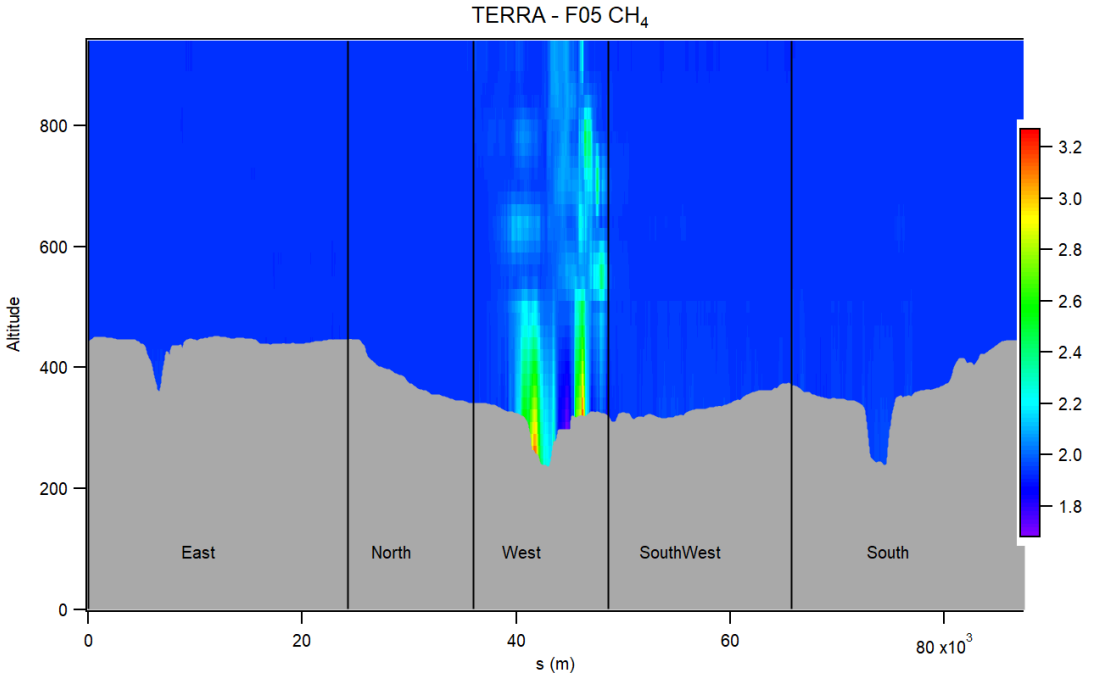


Figure S 11: Same as Figure S 3 for flight F05 CH_4 (ppm).

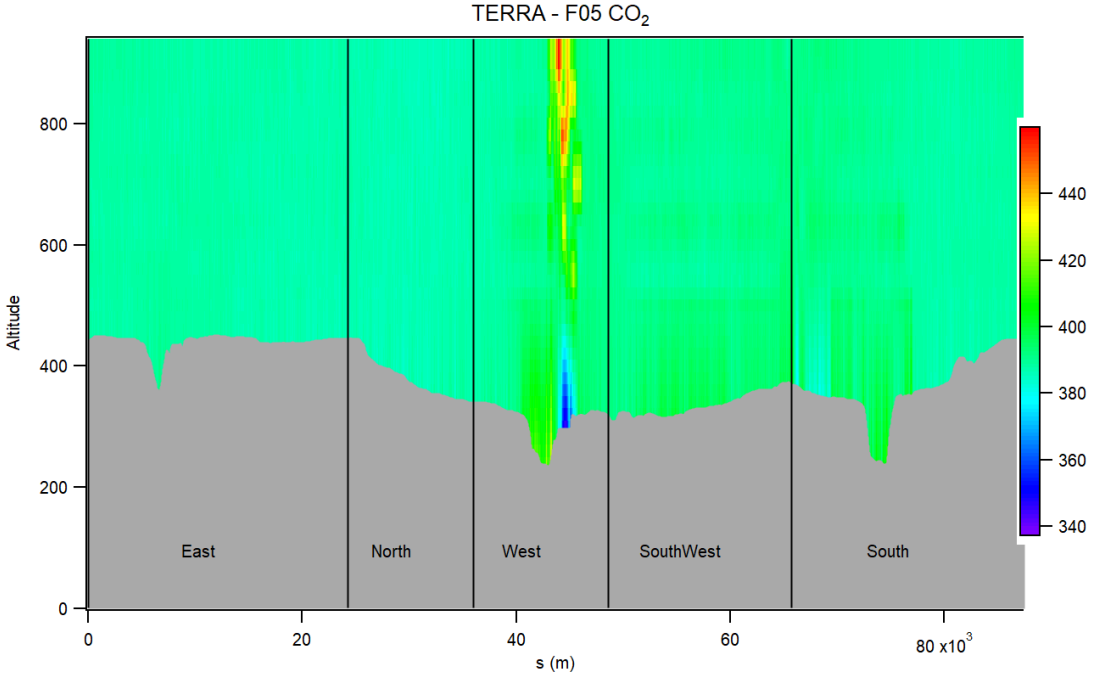


Figure S 12: Same as Figure S 3 for flight F05 CO_2 (ppm).

Table S 2: Background mixing ratio values used in the surface extrapolation of each flight.

Flight	CH_4 (ppm)	CO_2 (ppm)
F01	1.91613	409.1896
F02	1.88481	382.3219
F03	1.93007	410.2672
F04	2.05365	405.8721
F05	1.93103	389.4703

1.4 TERRA Error Terms

The calculated TERRA error terms are provided in Table S 3 and descriptions of how the terms were calculated are detailed in this section.

The box-top mixing ratio error is assumed to be normally distributed and is calculated as the percent change in the 95% confidence interval of the mean mixing ratio at the box top. The confidence interval was computed as the mean value +/- $2\sigma/\sqrt{n}$, where σ is the standard deviation of measurements and n the number of independent samples (Gordon et al., 2015). The value of n is determined by the length of a single lap divided by the length scale which was conservatively set to 3km as a maximum of the distance needed for the autocorrelation of the mixing ratio series to approach zero (Gordon et al., 2015).

To calculate the density, change error air pressure and temperature measurements were extracted from independent towers near the sampling area. Four towers were identified near the flight locations: two at the Fort McMurray Airport (A and CS at 56.65 °N, 111.22° W, http://climate.weather.gc.ca) and the Wood Buffalo Environmental Association's JP104 (57.12 °N, 111.43° W) and JP311 (56.56°N, 111.95° W) meteorological towers (http://wbea.org). Average, maximum, and minimum changes in the ratio difference of pressure and temperature ($\Delta p_i/p_i - \Delta T_i/T_i$) among the four stations were calculated then used to estimate total emissions using those values (Gordon et al., 2015). The uncertainty percentage was then calculated as the uncertainty range between *Emax/E* to *Emin/E*.

The box top height error was calculated in IGOR 8 using TERRA by estimating the percent change in the final emission estimate when the screen produced from kriging is redrawn 100 m lower. ECCC was only able to provide two flights with a redrawn box for analysis, therefore two examples of the average, and extreme case for plume capture were used. Redrawn screens were provided by ECCC for two of the flights representing the change given a 'fully captured' plume (F01), and a sample when the plume still had large enhancements at the top of the flight (F05). As part of a preliminary analysis of flights

from the larger Alberta Campaign, AEP and Scientific Aviation deemed F02, F03, and F04 to have 'fully captured' the top of the emission plume. Therefore, the F01 error was rounded up to the nearest integer and used to estimate the box top height errors for F02, F03, and F04.

Table S 3: TERRA emission estimate uncertainties as a percentage of the total emission estimate. Error terms are added in quadrature to give the total percentage uncertainty (δ). Wind and measurement error were added in quadrature as values of 1 and the vertical turbulence term was dropped from the calculation.

Flight	Measurement Error δ_M	Mixing ratio Extrapolation δ_{Ex}	Wind Extrapolation δ_{Wind}	Box-top mixing ratio δ_{Top}	Density Change δ_{dens}	Vertical Turbulence δ_{VT}	Box-top height δ_{BH}	Total Uncertainty δ
F01 - CH_4	<1%	10.13	<1%	0.89	0.03	N/A	0.56	10.28
F01 - CO_2	<1%	10.04	<1%	0.89	0.02	N/A	0.35	10.18
F02 - CH_4	<1%	4.80	<1%	2.15	1.06	N/A	1	5.64
F02 – CO_2	<1%	5.43	<1%	0.38	0.89	N/A	1	5.78
F03 – CH_4	<1%	4.16	<1%	2.30	0.48	N/A	1	5.08
F03 – CO_2	<1%	1.71	<1%	2.30	0.83	N/A	1	3.45
F04 – CH_4	<1%	17.08	<1%	2.42	5.79	N/A	1	18.28
F04 – CO_2	<1%	3.53	<1%	2.43	4.93	N/A	1	6.76
F05 – CH_4	<1%	9.54	<1%	1.00	0.04	N/A	11.17	14.79
F05 – CO_2	<1%	7.33	<1%	1.11	0.10	N/A	24.82	25.94

Adaptation of the SciAv Method in Further Detail

The raw data files for the five flights can be accessed through the Government of Alberta Portal:
http://ckandata01.canadacentral.cloudapp.azure.com/dataset/aep-noaa-greenhouse-gas-measurement-flights
The calculated average divergence for each lap, and associated errors provided by Scientific Aviation are provided in Table S 4. The lap bin bounds also provided by Scientific Aviation are given in Table S 5. Together these data can be used to recreate the SciAv profiles and emission calculations for each flight.

Table S 4: The average flux divergence and uncertainty for each flight lap.

Flight	Lap Number	Altitude (m)	CH_4 Average Flux Divergence	CH_4 Uncertainty	CO_2 Average Flux Divergence	CO_2 Uncertainty
F01	1	168.3	10.1355	1.2505	3088.564	572.941
	2	319.1	3.0669	0.4684	1596.195	359.556
	3	614.8	3.2761	0.5843	722.717	185.185
	4	912.6	2.5053	0.3873	470.83	104
	5	1056.9	0.2748	0.0572	-65.667	26.51
	6	771.2	4.0707	1.0834	862.677	280.279
	7	464.3	3.1503	0.3544	706.667	97.554
	8	378.6	2.5528	0.3681	614.653	182.681
F02	1	173	0.2896	0.0831	392.011	81.71
	2	301.6	0.2762	0.0713	428.288	107.227
	3	454.7	0.3529	0.1409	439.68	76.177
	4	614.6	0.7203	0.1425	805.679	149.198
	5	778.3	0.1679	0.0682	379.872	81.875
	6	919.3	0.1575	0.0689	286.82	81.201
	7	1076.7	0.1021	0.0688	227.823	82.947
	8	1245.6	0.0194	0.0327	120.558	44.357
	9	1122.5	-0.0992	0.0231	-26.271	24.561
	10	863.2	-0.013	0.0299	224.649	84.076
	11	714.1	0.5437	0.0871	786.83	129.209
	12	546.2	0.7292	0.1357	1355.712	261.852
	13	407.8	0.2934	0.0943	169.528	60.786
	14	247.7	0.4213	0.1408	726.796	146.438
F03	1	217	0.64	0.264	0.981	0.16
	2	190	0.896	0.142	1.358	0.293
	3	146	0.713	0.151	1.065	0.17
	4	163	0.572	0.218	0.513	0.088
	5	189	0.71	0.329	0.732	0.117
	6	238	0.651	0.166	0.708	0.121
	7	273	0.635	0.132	0.687	0.141
	8	305	0.232	0.159	0.698	0.105
	9	351	0.952	0.14	0.729	0.1
	10	411	1.216	0.133	1.404	0.165
	11	479	0.534	0.126	0.789	0.085
	12	760	-0.06	0.021	-0.034	0.007
	13	743	-0.086	0.014	-0.022	0.005
	14	685	0.394	0.095	0.089	0.006
	15	630	0.363	0.054	0.233	0.027
	16	541	1.693	0.137	0.366	0.052
	17	599	0.854	0.083	0.467	0.07
	18	377	1.166	0.272	0.439	0.064
	19	270	-0.383	0.282	1.454	0.173
F04	1	149.6	0.2452	0.0907	410.132	101.242

	2	213.1	0.426	0.0818	542.878	100.554
	3	271.9	0.533	0.0776	1481.367	243.59
	4	334.2	0.2294	0.0768	705.472	176.109
	5	404	0.3517	0.0959	1265.809	247.591
	6	461.4	0.6256	0.0987	1888.899	358.211
	7	524.4	0.6017	0.0773	2187.546	330.685
	8	593.1	0.4238	0.0633	1784.319	299.135
	9	657.3	0.29	0.0487	1041.434	164.144
	10	725.1	0.3704	0.0755	1836.424	375.014
	11	782.3	0.4037	0.0505	1402.95	202.694
	12	847.4	0.3731	0.0652	1396.115	239.28
	13	908.1	0.4542	0.0599	1293.877	176.586
	14	985.6	-0.0459	0.0167	-102.74	37.426
	15	1042.2	-0.0654	0.013	114.457	15.781
	16	926.7	0.2604	0.0403	861.268	146.277
	17	837.5	0.1892	0.0601	708.29	207.149
	18	745.7	0.3288	0.0589	968.067	194.864
	19	653.6	0.3971	0.0467	1426.952	186.054
	20	558.2	0.3413	0.0629	1543.505	230.08
	21	468.5	0.3703	0.0649	1323.173	180.568
	22	374.6	0.387	0.0648	1197.24	255.098
	23	280.2	0.3741	0.0591	1448.465	316.375
	24	211.6	0.3401	0.0593	1123.024	248.665
	25	159	0.2561	0.0724	1143.119	212.599
F05	1	157.2	7.8943	4.2509	1134.799	542.43
	2	306.6	7.9906	2.7143	1929.259	686.358
	3	455.8	8.117	3.0318	2176.205	1094.005
	4	562.7	3.6854	1.5554	2296.309	1159.373
	5	508.5	3.6561	1.4827	1717.746	1048.646
	6	359.7	6.0673	2.7987	1119.228	812.077
	7	211.7	5.0426	3.222	611.016	578.053

Table S 5: The profile altitude bin bounds provided by Scientific Aviation.

Flight	Bin Lower Limit (m)	Bin Middle (m)	Bin Upper Limit (m)
F01	0	84.1	168.3
	168.3	316.4	464.5
	464.5	612.6	760.7
	760.7	908.8	1056.9
F02	0	86.5	173
	173	307	441.1
	441.1	575.2	709.3
	709.3	843.3	977.4
	977.4	1111.5	1245.6
F03	0	69.59	139.18
	139.18	202.76	266.34
	266.34	329.92	393.5
	393.5	457.08	520.66
	520.66	584.23	647.81
	647.81	711.39	774.97
F04	0	74.8	149.6
	149.6	238.9	328.1
	328.1	417.4	506.6
	506.6	595.9	685.2
	685.2	774.4	863.7
	863.7	952.9	1042.2
F05	0	78.6	157.2
	157.2	258.6	359.9
	359.9	461.3	562.7

175

1.5 Fitting the SciAv Surface Extrapolation

The following section provides the Scientific Aviation (SciAv) algorithm surface extrapolation for all five different fits: Constant, Linear, Interpolate, Background, and Linear Weighted. The algorithm's integration method of binning the estimates and add adding in summation the average of each bin was applied for the measured part of the profile (Conley et al., 2017). For each surface extrapolation the lowest bin was extrapolated to the surface using the given fit. Certain profile shapes can be used to determine the fitting height depending on the number of points, minimum height, and shape of the bottom of the profile. The chosen fitting heights were: 1100m for F01, 400m for F02, 250m for F03, 400m for F04, and 600m for F05. The final estimate was determined by combining the emissions calculated from binned profile added to the trapezoidal integration of area between the extrapolation fit, the surface, and a flux enhancement of zero (main text, Figure 4). The estimates produced by fitting the different surface extrapolations to SciAv are given in Table S 6. Figure S 13A shows that both emission profiles for F01 follows a type III shape with fits for linear, weighted linear, and averaged increasing towards the surface. The F01 surface extrapolation and resulting emission estimate would be larger and more closely follow plume behaviour if the lowest divergence point, rather than the lowest bin estimate (the method standard), was used (Figure S 13 A). The F02 SciAv profile follows a Type II shape (Figure S 13 B) with very little variation in the extrapolation fits, which are clustered around the standard constant fit. The F03 SciAv profile has a lot of between-lap variation with a profile shape that largely follows a Type I shape (Figure S 13 C). The F03 surface extrapolation was the most unstable fit due to the larger variation between laps at the bottom of the profiles. F05 was the only flight with varying profile shapes for and CH_4 and CO_2. The shape of the emission profile in Figure S 13 E for CH_4 shows some constraining of the plume at the highest altitude; however, the divergence is still larger than 2 kg m^{-1} hr^{-1} which is larger than the peak of divergence from plant samples. The profile for F05 in Figure S 13 E shows incomplete sampling for CO_2 as the divergence points are still increasing at the highest altitude indicating that the plume was not fully captured.

Table S 6: SciAv emission rate estimates using six surface extrapolation fits.

Flight	Surface Extrapolation	CH_4 Emission (kg h^{-1})	CO_2 Emission (t h^{-1})
F01	Constant	3840	1040
F01	Linear	5270	1250
F01	Interpolate	3900	1040
F01	Background	3050	784
F01	Linear Weighted	4690	1300
F01	Averaged	4200	1150
F02	Constant	362	5630
F02	Linear	360	556
F02	Interpolate	331	523

F02	Background	306	489
F02	Linear Weighted	359	563
F02	Averaged	361	562
F03	Constant	497	526
F03	Linear	506	553
F03	Interpolate	452	479
F03	Background	400	401
F03	Linear Weighted	504	571
F03	Averaged	503	548
F04	Constant	349	1170
F04	Linear	332	1090
F04	Interpolate	313	1050
F04	Background	295	1020
F04	Linear Weighted	332	1110
F04	Averaged	341	1140
F05	Constant	3470	850
F05	Linear	3710	783
F05	Interpolate	3030	751
F05	Background	2410	661
F05	Linear Weighted	3670	782
F05	Averaged	3590	1040

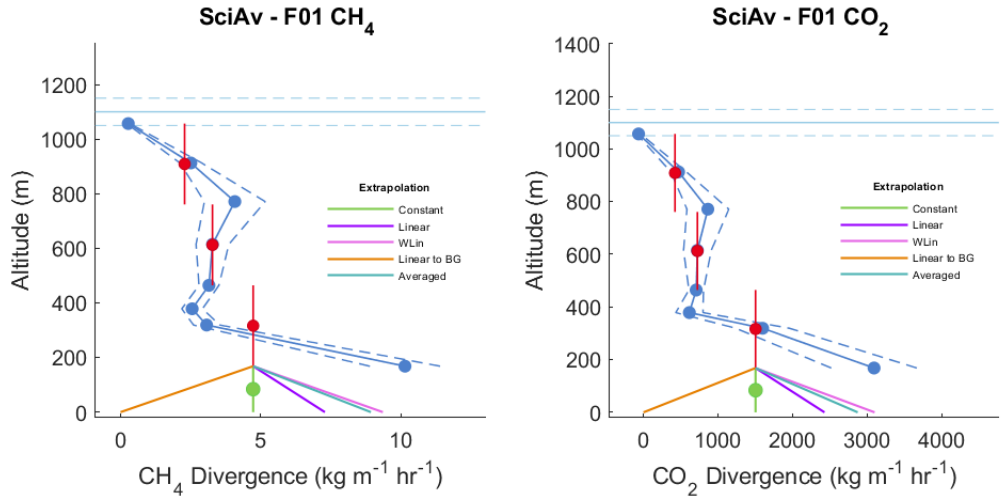


Figure S 13 A

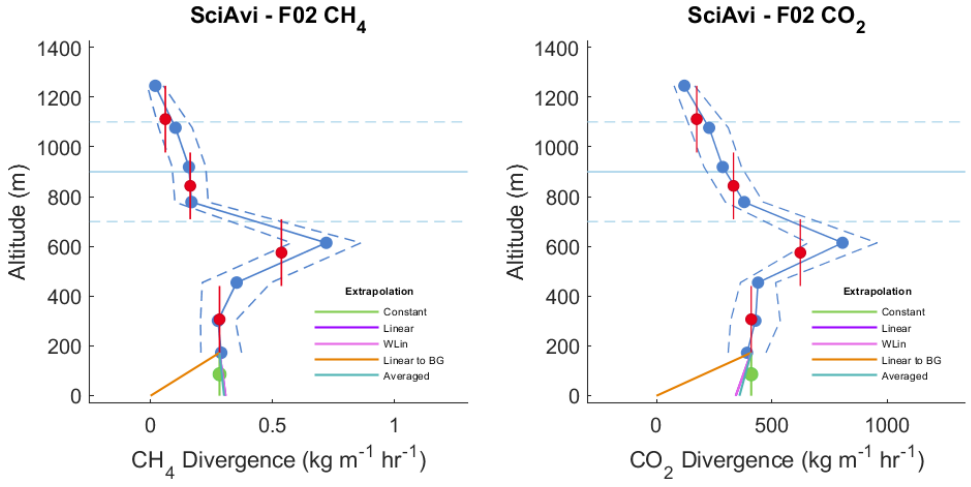


Figure S 13 B

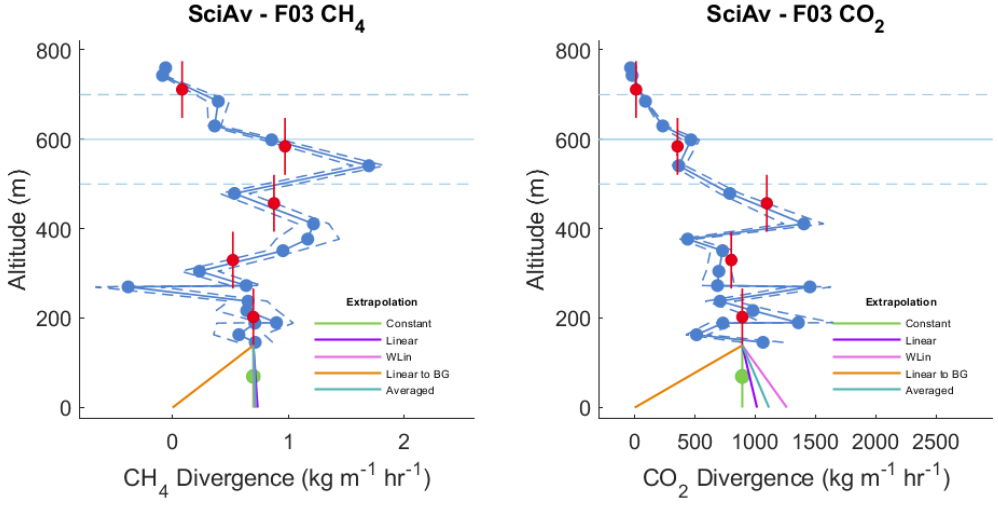


Figure S 13 C

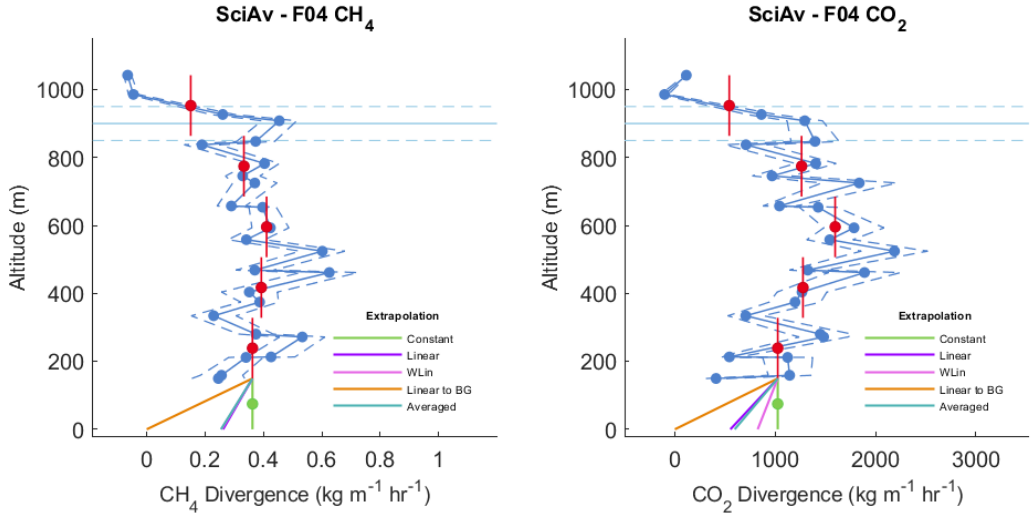


Figure S 13 D.

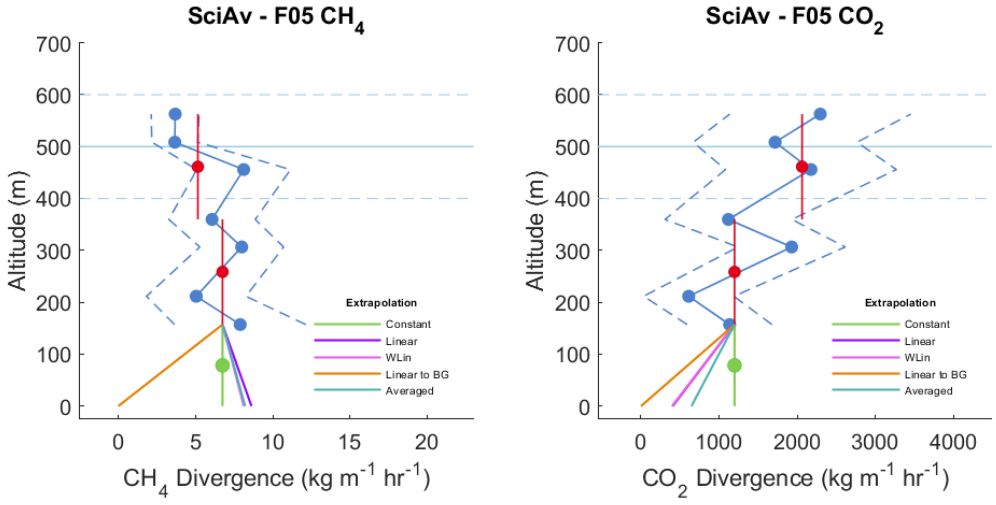


Figure S 13 E.

Figure S 13: The SciAv profile is shown for F01-F05 in Figures S 13A – 13E. The blue points are the estimated flux divergence for each lap which are connected to show profile shape with the associated error (a dashed blue line). Red points are bin averages and the vertical red bar is the bin height range. The boundary layer height is drawn in light blue with error bars (light blue dashed lines). The five surface extrapolations are drawn from the bottom of the lowest red bin.

1.6 Significance Testing and Further Comparison

The difference between estimates, relative mean percentage difference, and the propagated error in percentage for the average estimate of the two algorithms are shown in Table S 7. The whole set of results from five flights were formally tested for differences between the SciAv and TERRA estimates using a weighted t-test and Wilcoxon signed rank test. As a collective, the differences between the estimates were found to be insignificant for both CH_4 and CO_2 (Table S 8).

Table S 7: Differences between algorithms as SciAv estimate – TERRA estimate for each flight, relative percentage differences and the propagated percentage error of the estimates. A mean percentage difference that is approximately equal to, or smaller than the propagated percentage error of the estimates indicates agreement within the uncertainty of the estimates.

	F01	F02	F03	F04	F05
CH_4: Difference (kg h^{-1})	-967	-33.5	20.5	224	-436
CH_4: Relative Percentage Difference	22%	8%	4%	94%	12%
CH_4: Propagated Percentage Error	21%	22%	17%	18%	29%
CO_2: Difference (kg h^{-1})	-302000	47900	58700	602000	-27700
CO_2: Relative Percentage Difference	25%	9%	12%	69%	3%
CO_2: Propagated Percentage Error	24%	19%	12%	13%	43%

Table S 8: Results of parametric (weighted t-test) and non-parametric (Wilcoxon signed-rank test) significance testing of differences between the two box-flight algorithms using the set of all five flight estimates.

	Weighted t-test: p-value	Weighted t-test: t-value	Wilcoxon signed-rank test: p-value	Wilcoxon signed-rank test: V
CH_4	0.306	-1.09	0.438	4
CO_2	0.366	-0.96	0.625	10

To remove the effect of the surface extrapolation the two algorithms were compared when using each method's "background" surface extrapolation fit. For TERRA this meant fitting the chosen background mixing ratio value below the lowest flight lap, and for SciAv calculating zero divergence below the lowest flight lap. Neither algorithm would choose background extrapolation for the standard estimate to the flights as increasing or trace emissions were present at the bottom of each the flight track.

The results of calculations using the assumption of no emission plume below the lowest flight track were compared to the standard fit in Figure S 12. The estimate uncertainties were only calculated for each algorithm's standard fit, therefore estimates using other fits have no error bars. The mean of the surface extrapolation estimates and the difference between the

mean estimates is given in Table S 9. Results of parametric (weighted t-test) and non-parametric (Wilcoxon signed-rank test) significance testing of differences between the two box-flight algorithms using the set of all five flights given each the standard fit, background fit, and average of the fits estimates is given in Table S 9.

Emission estimates using a background fit are systematically lower than the standard estimates (Figure S 14), but no clear pattern is identifiable that would indicate systematic disagreement. There is a considerably larger TERRA estimate for F01 between the standard fit compared to the background. The choice of surface extrapolation influences the extent of agreement. The algorithms tend to agree more when the effect of a different surface extrapolation is removed as much as possible. This indicates good agreement between the methods. When fundamental assumptions are met, variation between the estimates can largely be attributed to different surface extrapolation methods.

Table S 9: The mean differences between the two algorithm outputs amongst the five flights and results from significance testing were calculated for comparison between the standard, background and average of the algorithms fit.

	CH_4 Standard	CH_4 Background	CH_4 Average	CO_2 Standard	CO_2 Background	CO_2 Average
Mean Difference (kg h^{-1})	-239	12.4	47.8	75800	141000	115000
Pairwise t-test: p-value	0.857	0.989	0.969	0.724	0.321	0.525
Pairwise t-test: t-value	-0.187	0.0145	0.0404	0.367	1.07	0.665
Wilcoxon signed-rank test: p-value	0.438	1	0.625	0.625	0.313	0.313
Wilcoxon signed-rank test: V	4	8	10	10	12	12

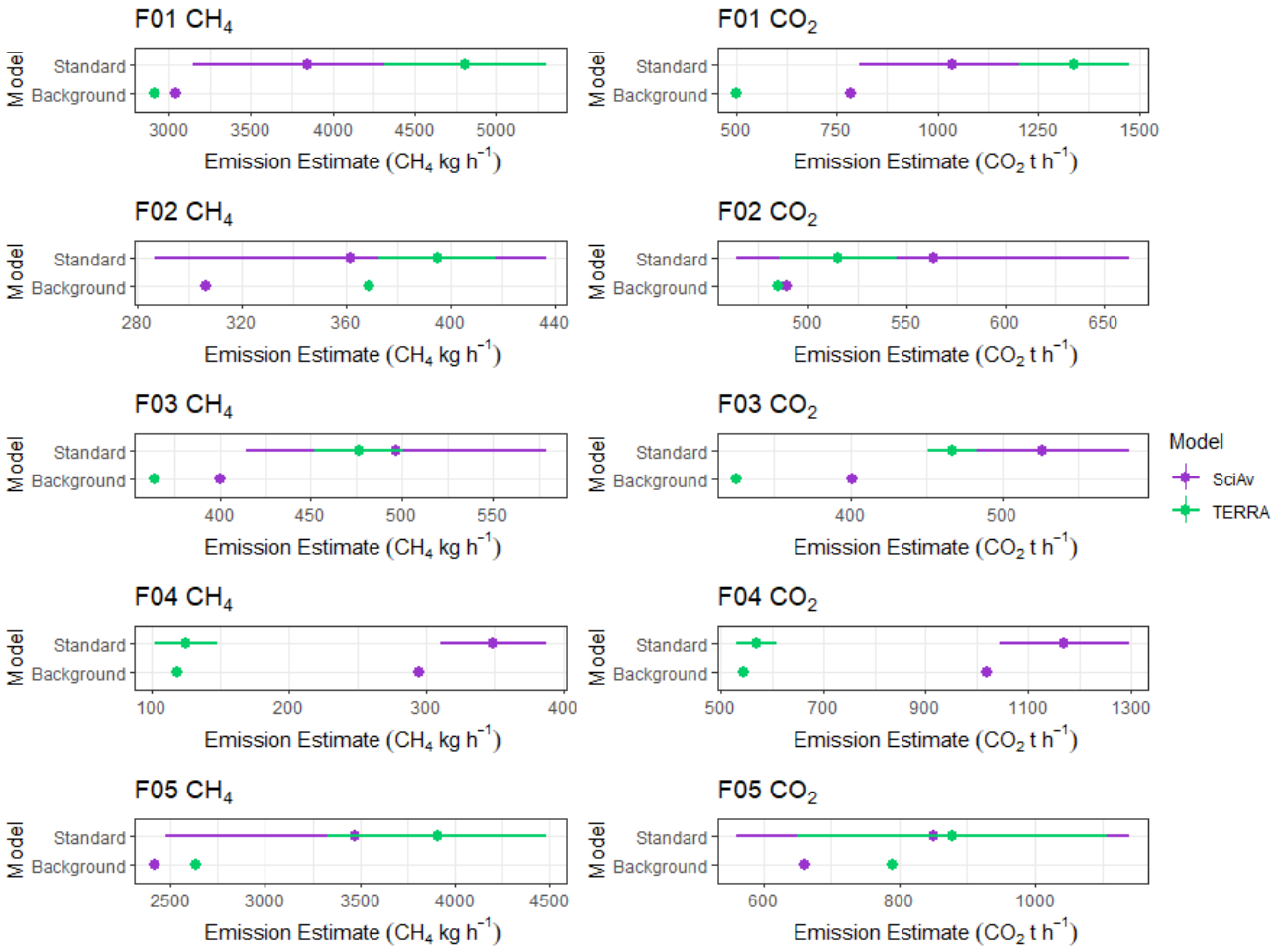


Figure S 14: Algorithm estimates for CH_4, and CO_2 given the background mixing ratio fit of no emission plume extrapolation to the surface compared to the standard estimates.

Table S 10 gives the percentage relative differences between each algorithm, and the standard deviation of the percentages for each surface extrapolation, calculated for the four surface extrapolations used for each method. In general, the choice in surface extrapolation varied less than the relative percentage difference between the algorithms. Fitting extrapolations with increasing emissions towards the surface increased agreement for F01 CO_2 from a percentage relative difference of 16% for the standard constant fit, to 7% using the linear fit.

Table S 10: Percentage relative difference of estimates between algorithms, calculated as SciAv – TERRA, for each of the linear, Interpolate, constant and background surface extrapolations. The standard deviation between the percentages rounded to the nearest integer for each flight is given in the last column.

	Linear	Interpolate	Constant	Background	Standard Deviation
F01 CH_4 (%)	9	5	-12	5	9
F02 CH_4 (%)	-16	-18	-13	-19	3
F03 CH_4 (%)	6	8	8	9	1
F04 CH_4 (%)	105	87	94	85	9
F05 CH_4 (%)	-5	-4	-2	-9	3
F01 CO_2 (%)	-7	0	-16	44	26
F02 CO_2 (%)	2	1	5	1	2
F03 CO_2 (%)	17	18	14	21	3
F04 CO_2 (%)	62	59	66	61	3
F05 CO_2 (%)	-11	-11	-4	-18	6

To obtain a sense of the range of estimates from each algorithm a distribution of randomly sampled mean differences was created. For each flight and gas, a bootstrap analysis was applied by randomly sampling an estimate given one of the extrapolations for each method individually and the difference computed as:

$$difference = mean(Estimate_{SciAv}) - mean(Estimate_{TERRA}) \quad (6)$$

Analysis was run in Rstudio using the two.boot function from the simpleboot package (Peng, 2019). 5000 replications were used for each bootstrap to obtain the distributional shape of the results. These distributions were used as a proxy for statistical confidence intervals (main text, Figure 8). They were used to estimate testing of the null hypothesis that estimates agree, and if the distributions do not overlap with zero then there is evidence that the algorithms differ.

1.7 Comparison Results and Significance Testing

The current SciAv method uses a binning method for integrating the profile points over altitude to estimate the overall emission estimate. This method can produce large flux divergence error terms when a small number of laps have been flown and divergence points differ greatly between laps, which can inaccurately estimate error if the variation is due to a trend such as emissions increasing to the surface. In these scenarios it may be beneficial to use trapezoidal integration rather than binning to estimate emissions. Rather than grouping into altitude bins the trapezoidal integration estimates the area under the curve by a trapezoidal fit between each point. Trapezoidal integration fits the area by altitude between a divergence of zero and the positive divergence calculated for each lap using the lines connecting the points. This method parallels the calculation of the SciAv divergence error term whilst avoiding the potential bias from a small number of flight laps.

Table S 11 gives the SciAv trapezoidal integration method estimates. Figures S 15 A – 15 E show the SciAv profiles given the trapezoidal integration method. Surface extrapolations are fit from the lowest mean divergence lap estimate. The estimated end point for the surface extrapolation does not change between the two integration methods as it is determined by fitting curves by the position of the points and fitting height. As the surface extrapolation points were the same for both profiles which caused the F01 trapezoidal profiles to decrease despite their noticeable increasing to surface trend (Figure S 15 A). A less conservative fitting height procedure would likely produce a large range in emission estimates, and should be explored if the trapezoidal method is adapted, as it could produce an extrapolation truer to the nature of the divergence profile.

Table S 11: SciAv emission rate estimates using trapezoid integration.

Flight	Surface Extrapolation	CH_4 Emission (kg h^{-1})	CO_2 Emission (t h^{-1})
F01	Constant	4840	1350
F01	Background	3130	831
F01	Linear	4600	1300
F01	Interpolate	3980	1090
F01	Linear Weighted	4770	1350
F01	Averaged	4730	1330
F02	Constant	352	550
F02	Background	302	483
F02	Linear	355	550
F02	Interpolate	327	517
F02	Linear Weighted	354	557

F02	Averaged	354	552
F03	Constant	530	530
F03	Background	426	375
F03	Linear	532	527
F03	Interpolate	478	453
F03	Linear Weighted	530	545
F03	Averaged	531	534
F04	Constant	338	1120
F04	Background	301	1060
F04	Linear	339	1130
F04	Interpolate	320	1090
F04	Linear Weighted	339	1150
F04	Averaged	339	1130
F05	Constant	3780	797
F05	Background	2540	619
F05	Linear	3830	740
F05	Interpolate	3160	708
F05	Linear Weighted	3800	739
F05	Averaged	3800	759

290

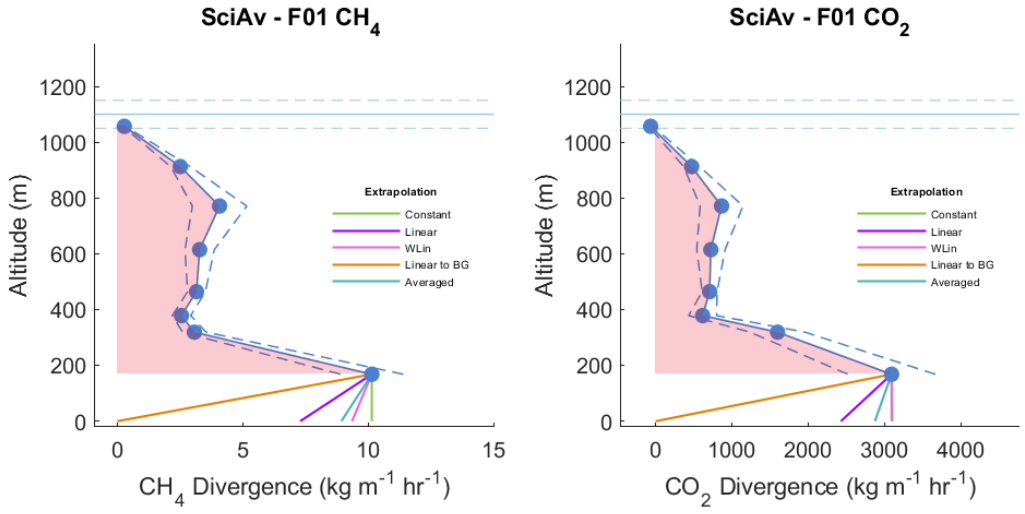


Figure S 15 A.

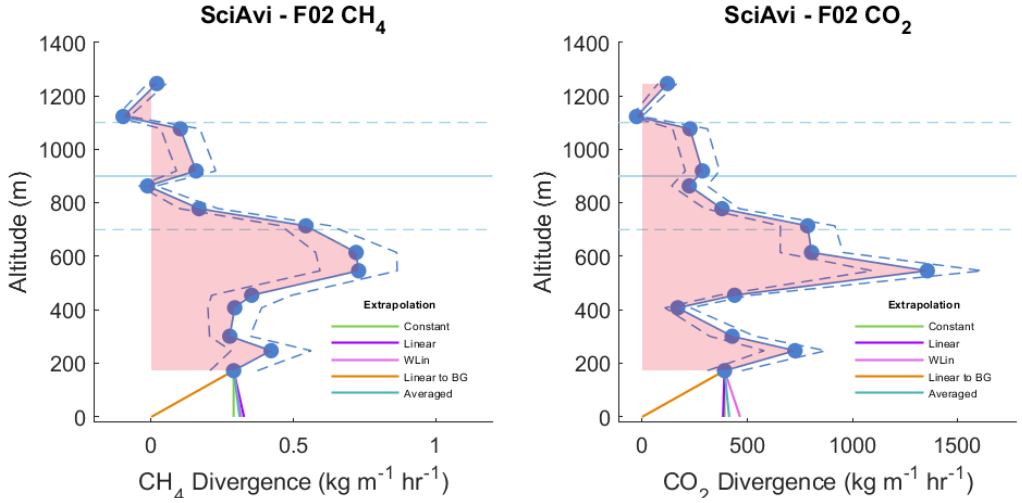


Figure S 15 B.

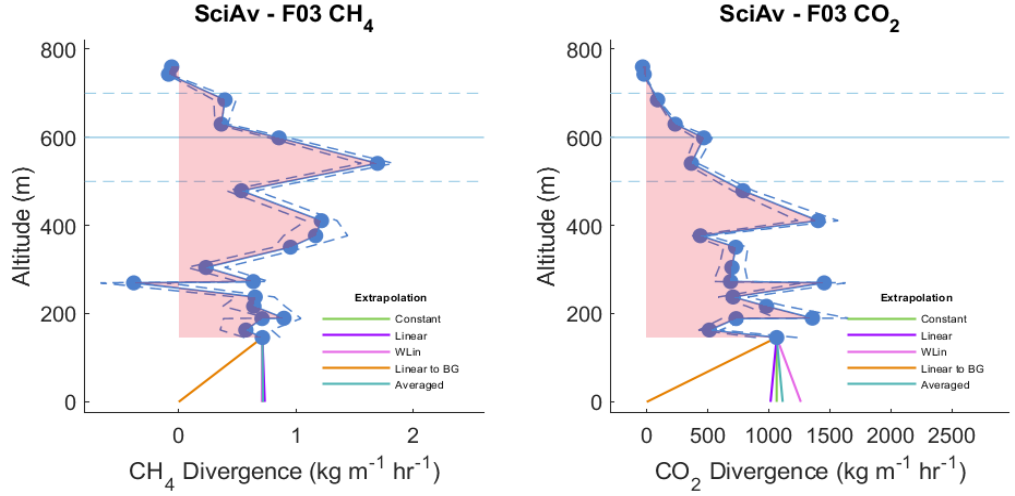


Figure S 15 C.

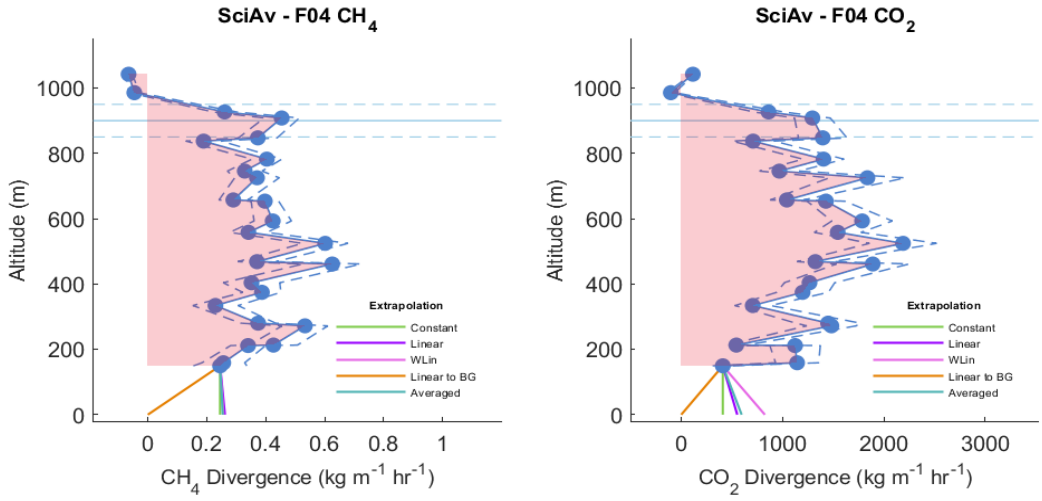


Figure S 15 D.

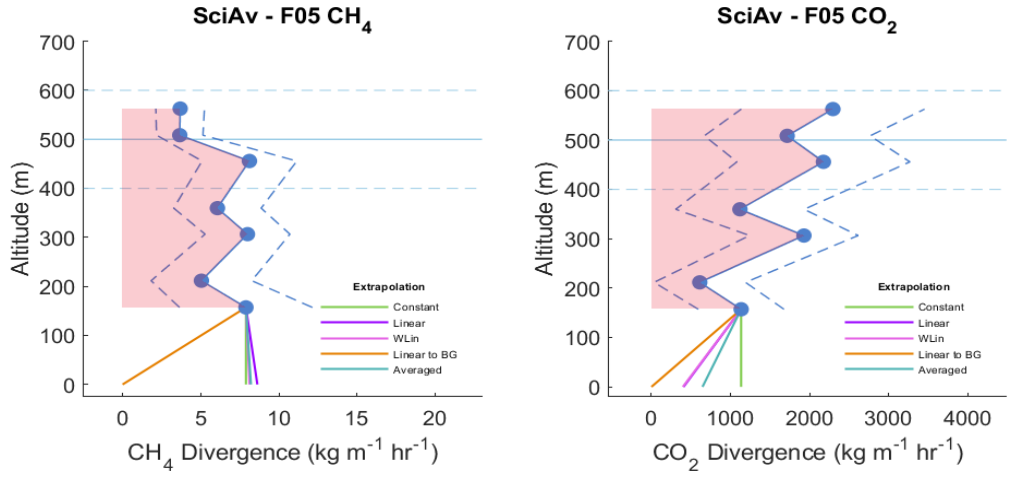


Figure S 15 E.

295

Figure S 5: A - E of the SciAv profiles for flights F01 through F05. The blue points are the estimated flux divergence for each lap which are connected to show profile shape with the associated error (a dashed blue line). The red area is the trapezoidal integrated area of the blue divergence points excluding the surface extrapolation. The boundary layer height is drawn in light blue with error bars (light blue dashed lines). The five surface extrapolations are drawn from the bottom of the lowest blue divergence point.

Figure S 16 show the binned estimates compared to the trapezoidal (Trapz). No systematic pattern is evident that would indicate that one integration method produces consistently larger or smaller emissions estimates. There is a systematic gap between estimates that could be attributed to the use of different extrapolation points. The binned method uses the lowest bin for extrapolation, whereas the trapezoidal method uses the lowest divergence point.

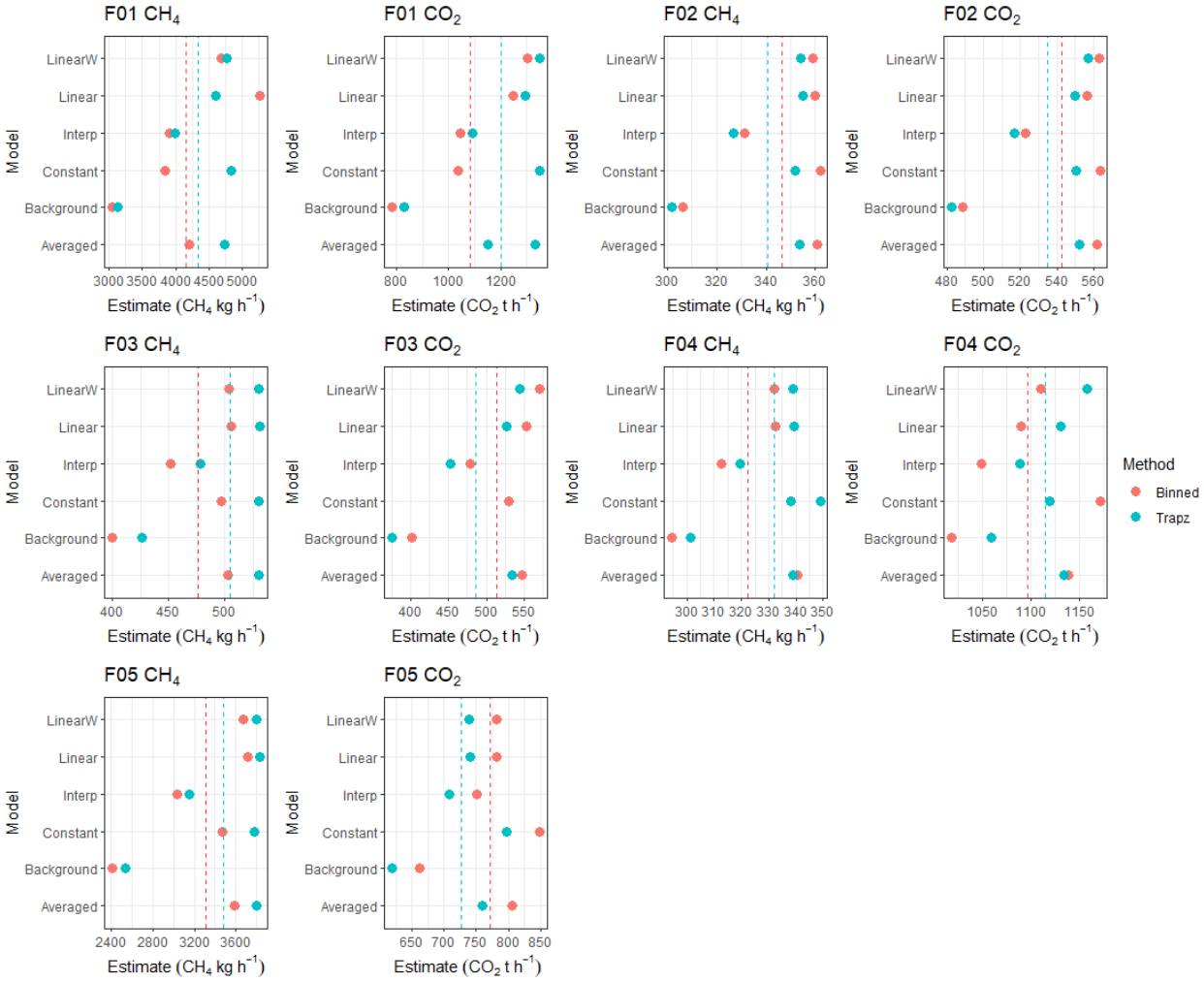


Figure S 6: SciAv emission estimates of CH_4 (kg h^{-1}) and CO_2 (t h^{-1}) derived using both a binning (red) and trapezoidal (blue) integration method. Mean values for each method are shown as a dotted line.

Inspecting Figure S 16 there is a larger variation between estimates due to the chosen surface extrapolation than integration method. There was one exception, there is a larger difference between the integration types when a constant extrapolation is chosen for F01. This flight has an increasing to surface profile point, and because the lowest divergence point differs greatly from the location of the lowest bin estimate, the extrapolation fits differ. Applying an appropriate surface extrapolation that follows the shape of the profile remedied the difference between integration methods.

Percentage relative standard deviations (%RSD) were calculated (Table S 12) to estimate the deviation in the surface extrapolation estimates given the binning and trapezoidal approaches to integrating the SciAv lap divergence data. These were compared to the %RSD of the difference between the two integration methods. Table S 13 gives the standard deviations of the surface extrapolation estimates for each integration method and the standard deviation for the difference between integration methods for all five fives. Table S 14 gives the results from significance tests on the differences between the estimates. The choice of surface extrapolation leads to a greater standard deviation than the choice of integration method. Results showed that the range in emission output was a factor of ~ 3 times larger for the range in surface extrapolation estimates than the difference between using a binned or trapezoidal integration. The choice in the surface extrapolation affected the range of emission outputs more than the integration type.

Table S 12: Percentage relative standard deviations (RSD) calculated from the set of seven SciAv surface extrapolations given a binned or trapezoidal integration method and for the difference between the two. The mean of each %RSD is calculated amongst all flights.

	F01 CH_4	F02 CH_4	F03 CH_4	F04 CH_4	F05 CH_4	F01 CO_2	F02 CO_2	F03 CO_2	F04 CO_2	F05 CO_2	Mean
Binned %RSD	18	7	9	6	15	17	6	12	5	8	10
Trapezoid % RSD	15	7	9	5	15	17	5	14	3	8	10
Difference % RSD	12	1	1	2	2	9	1	3	3	0	3

Table S 13: Standard deviation of the surface extrapolation estimates for each integration and the difference between integration method extrapolation estimates (calculated as: binned estimate – trapezoidal estimate).

	SD Binned Extrapolation Estimates	SD Trapezoidal Extrapolation Estimates	SD Difference in Methods Extrapolation Estimates
CH_4 (kg hr $^{-1}$)	1740	1810	581
CO_2 (t hr $^{-1}$)	274	315	139

Table S 14: Results of parametric (weighted t-test) and non-parametric (Wilcoxon signed-rank test) significance testing of differences between the two SciAv integration method, by calculating the mean of the estimates for each flight and tested over the whole set of five flights.

	Weighted t-test: p-value	Weighted t-test: t-value	Wilcoxon signed-rank test: p-value	Wilcoxon signed-rank test: V
CH_4	0.952	-0.062	0.188	2
CO_2	0.952	-0.062	1	8

To further compare the two methods of integrated a distribution of the mean differences for each flight was computed by bootstrapping the mean difference between randomly sampled estimates. The mean difference was computed as:

$$difference = mean(SciAvEstimate_{Binned}) - mean(SciAvEstimate_{Trapz}) \quad (7)$$

As shown in Figure S 17, the distributions for all flight samples of each emission type cover zero indicating that the two integration methods produce similar estimates. The trapezoidal integration may be a useful method for profiles with a small number of laps, but the choice of an appropriate surface extrapolation has a larger effect on the emission estimate than the type of integration used. The SciAv divergence error term could be calculated by using the lap data and bootstrapping the trapezoidal integration over thousands of profile points that are randomly sampled within the uncertainty of each point to estimate the random sampling error between each lap by altitude. A surface extrapolation term could be derived following methods used in TERRA by assessing the maximum percent change between plausible extrapolations.

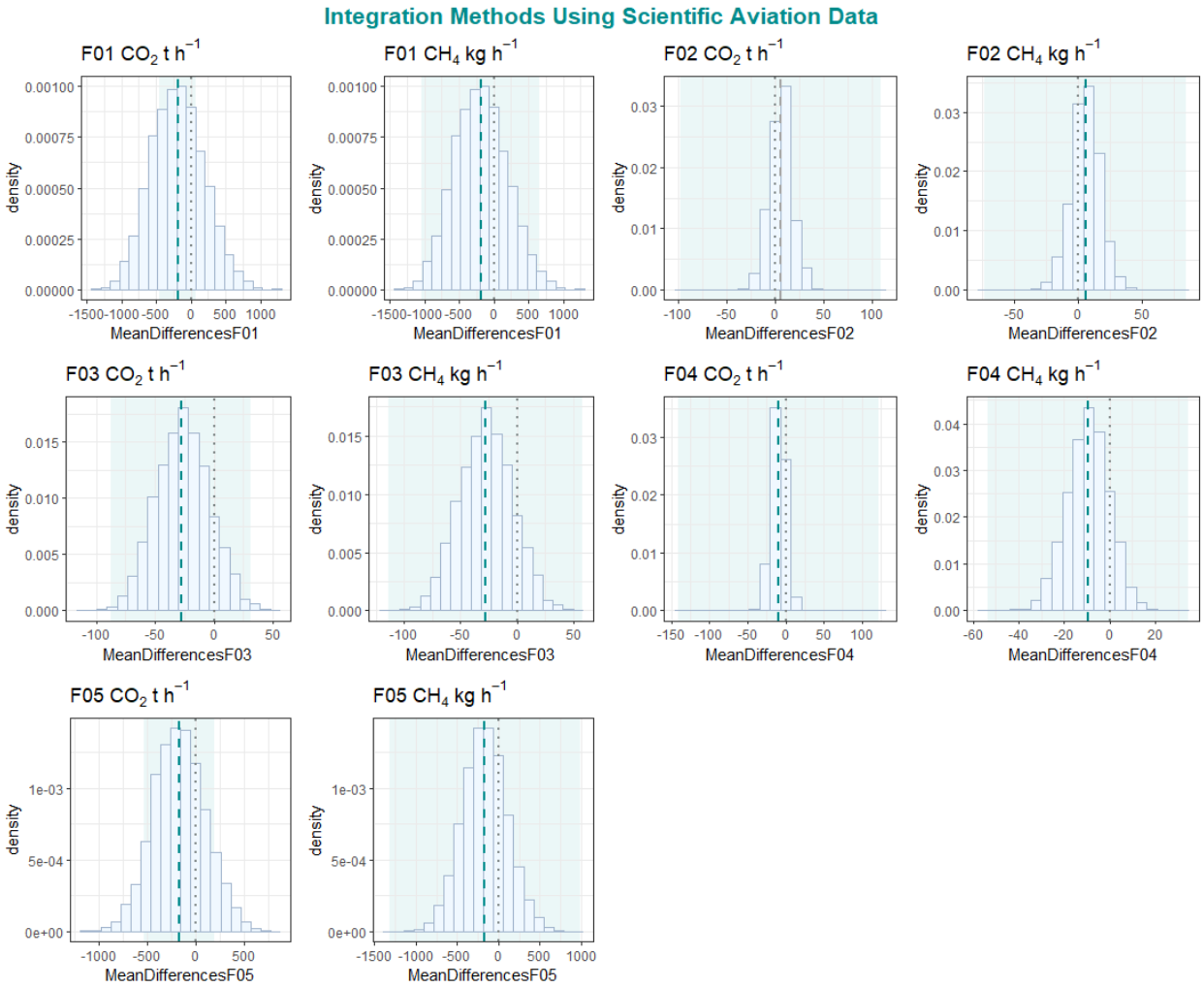


Figure S 17: Distributions of the mean difference between all SciAv fits of CH_4 and CO_2, given the binning compared to the trapezoidal integration methods, in light blue. The mean difference between the standard estimates is plotted as a teal dashed lined and the range in the estimate as the light teal box behind the distribution. A grey dot dashed line is drawn at zero as a reference point for the location of exact agreement between the methods.

Box-flight Algorithm Assumptions Investigated

1.8 Flight Screening and Analysis of Meteorological Conditions During Flight F04

The potential cause of the non-stationarity of F04 was thoroughly investigated. Analysis of the raw data and emission profiles indicates that the emission plume itself was non-stationary as the mixing ratios enhancements noticeably differed between the upwards and downwards portion of the flight, but other potential causes such as changing meteorological conditions were inspected. Along with the flight screening provided by Scientific Aviation and Alberta Airshed Stewardship, conditions were assessed using the full data and then split into upwards (ascending) and downwards (descending) sections for comparison. The various analysis provided in this document supports the conclusion that meteorological conditions were not the cause of the FO4 non-stationarity.

Prior to this study, Scientific Aviation and Alberta Airshed Stewardship performed a screening of all flights in the AEP-NOAA-Scientific Aviation 2017-2018 Alberta Oil Sands Flight Campaign. This process includes looking at 'circle' files and assessing the wind direction and mixing ratio measurement for each lap. The circle files help detect plume capture, non-stationarity of the wind, and any upwind flux that is not coming from the intended emission source. The SciAv lap divergence profiles are also inspected for plume capture and the shape of the plume. Stationarity and consistency of the wind data was deemed good, the plume was assessed as fully captured, and there was negligible upwind flux entering the box.

The non-stationarity of F04 was not confirmed until post-hoc analysis was applied by Environment and Climate Change Canada. The TERRA step of kriging the data uses the lap data to produce spatially gridded emission mixing ratios. Figure S 20 shows the middle step of gridded screen for F04 CH_4. Some boxes contained a high (red) emission and a low (blue) emission enhancement (Figure S 18) which is evidence of non-stationarity as the plume and background mixing ratios changed over time. The change in mixing ratios may have occurred due to changing operating conditions at or near the facility, but operations conditions were not shared by the facility. The non-stationarity of the emission plume biased the TERRA estimates as the large mixing ratio differences in each square led to lower estimates from the kriging. For TERRA, there is a noticeable difference in the CH_4 mixing ratios in the upward flight compared to the downward, with higher mixing ratios (red) detected in the ascending portion of the flight (Figure S 19).

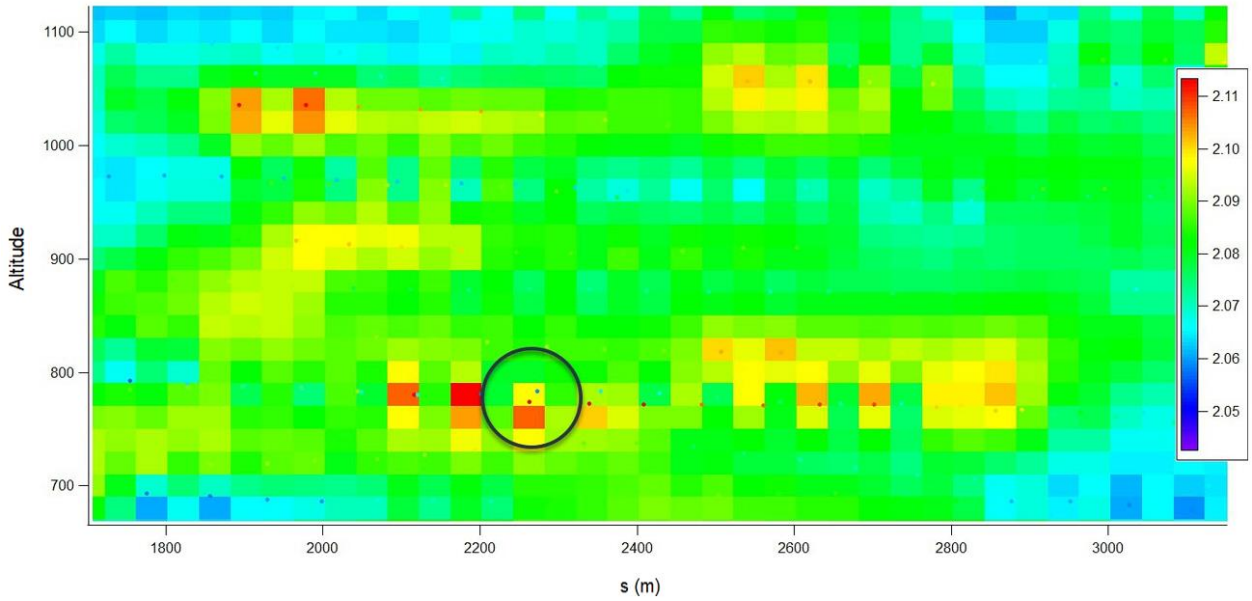


Figure S 18: A zoomed in look at the final F04 CH_4 box of kriging-interpolated data. A black circle is drawn around a gridded square that contained a large red enhancement at the same location as a low blue mixing ratio. The square was averaged to a yellow CH_4 value. Imagery provided by Environment and Climate Change Canada.

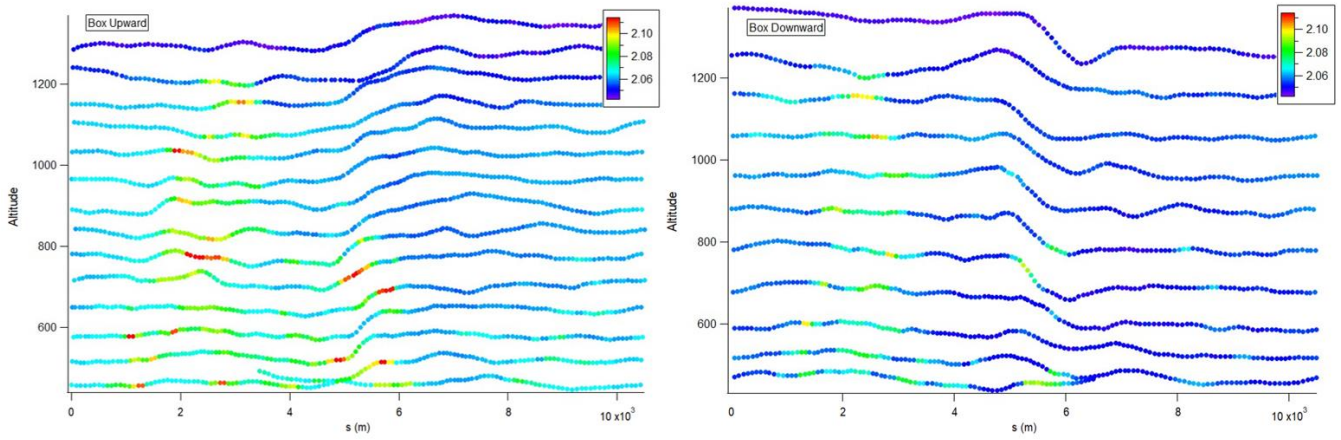


Figure S 19: Per lap CH_4 mixing ratio screen data for the box upward (left) and downward (right) prior to kriging of the data. Imagery provided by Environment and Climate Change Canada.

As part of the meteorological conditions' analysis, the stability of the boundary layer was assessed. To thoroughly investigate conditions multiple variables were calculated to produce multiple estimates of the boundary layer. Following previous calculations for estimating the boundary layer from aircraft flight data (Dai et al., 2014) variables were computed: gradient Richardson number, bulk Richardson number, change in virtual temperature, change in potential temperature, and the change in virtual potential temperature. The boundary layer height was estimated from the maximum of each of these variables to assess any changes during the flight up versus down. These values were also compared to the SciAv estimate that was provided which was estimated the day of the flight. The calculated maximum values used to estimate the boundary layer are plotted in Figure S 20 for the full, upwards, and downwards F04 flight segments. Analysis by Dai et al. 2014 suggest the best practice is to estimate the boundary layer at maximum change in the virtual potential temperature (maxVPT/dz). For this flight, the boundary layer estimate is near SciAv's estimate and appears stable. If the meteorological conditions caused a significant shift in the boundary layer this would likely show up during the reanalysis. The consistency in boundary layer estimates for the flight up versus down further indicates that the boundary layer was stationary during the flight.

Thorough analysis of the boundary layer, and the wind components supports the conclusion that meteorological conditions were not the source of the non-stationarity of the emission plume for F04. Ruling out changing meteorological conditions leaves a change in the background mixing ratio, or a change in the facility emissions during the flight as potential causes of the non-stationarity. However, the cause of the non-stationarity of the emission plumes for F04 remains unknown as there was no independent measurement of the background mixing ratio during sampling, and the operating conditions near and at the facility were not shared.

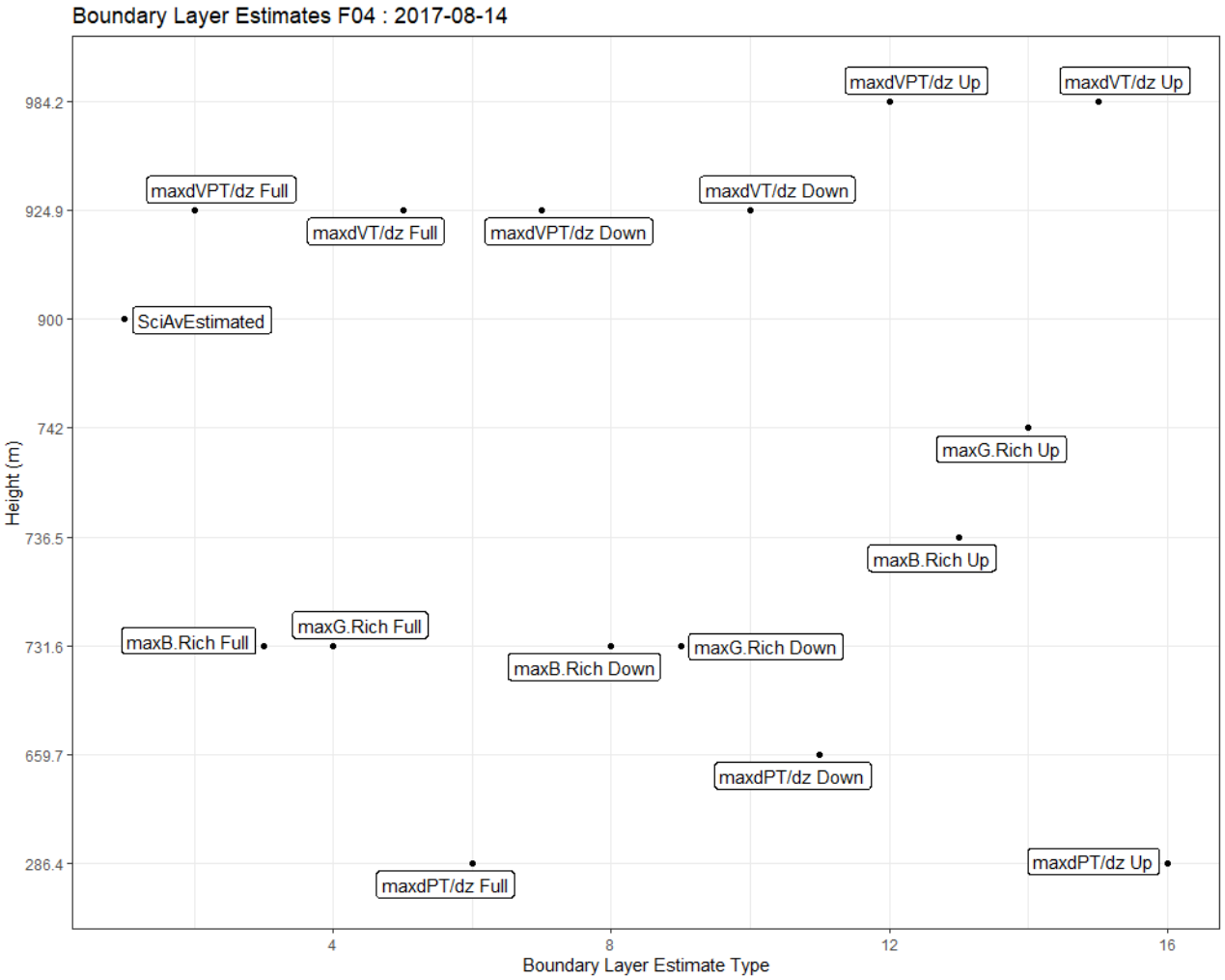


Figure S 20: The five boundary layer estimates for the full, up and down F04 flight segments plotted as the maximum: gradient Richardson number (maxG.Rich), bulk Richardson number (maxB.Rich), change in virtual temperature (dVT/dz), change in potential temperature (dPT/dz), and change in virtual potential temperature (dVPT/dz). The acronym dz is the change in altitude above ground level. The y-axis gives the estimated boundary layer in meters above ground level.

References

Conley, S., Faloona, I., Mehrotra, S., Suard, M., Lenschow, D. H., Sweeney, C., Herndon, S., Schwietzke, S., Pétron, G., Pifer, J., Kort, E. A., and Schnell, R.: Application of Gauss's theorem to quantify localized surface emissions from airborne measurements of wind and trace gases, Atmos. Meas. Tech., 10, 3345–3358, https://doi.org/10.5194/amt-10-3345-2017, 2017.

Dai, C., Wang, Q., Kalogiros, J. A., Lenschow, D. H., Gao, Z., and Zhou, M.: Determining Boundary-Layer Height from Aircraft Measurements, Boundary-Layer Meteorol., 152, 277–302, https://doi.org/10.1007/s10546-014-9929-z, 2014.

Gordon, M., Li, S. M., Staebler, R., Darlington, A., Hayden, K., O'Brien, J., and Wolde, M.: Determining air pollutant emission rates based on mass balance using airborne measurement data over the Alberta oil sands operations, Atmos. Meas. Tech., 8, 3745–3765, https://doi.org/10.5194/amt-8-3745-2015, 2015.

Peng, R. D.: simpleboot: Simple Bootstrap Routines, [code], https://github.com/rdpeng/simpleboot, 2019.