

Boxed-in: Comparing Algorithms for Box-flight Mass-Balance Greenhouse Gas Flux
Measurements from Mineable and In Situ Oil Sands Developments

by

Broghan Erland

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Earth and Atmospheric Sciences
University of Alberta

© Broghan Erland, 2022

Abstract

To combat global warming, Canada has committed to reducing greenhouse gases (GHGs) 40-45% below 2005 emission levels by 2025. Monitoring emissions and deriving accurate inventories are essential to reaching these goals. GHGs can be measured at a small scale, often using ground measurements which are extrapolated to estimate regional emissions, or at larger scales using airborne or satellite methods to infer area sources. Airborne methods can provide regional and area source measurements if ideal conditions for sampling are met. For the first time, 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, SciAv Gaussian algorithm and the Environment and Climate Change Canada 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 model errors (4-34%). Implementing a surface extrapolation procedure for the SciAv model may improve emission estimation. Models disagreed when non-ideal conditions occurred (under non-stationary conditions). Results from both models were largely unusable when a flight sample did not capture the top of an emission plume. 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 model assumptions should be assessed for each flight. The box-flight mass-balance airborne methods provide in-depth measurements by flying around a source to box in an emission plume. They can attain a low uncertainty ($\sim 2\%$), but they require known sources, often require advance coordination with industry, and are quite expensive, so cannot always be easily applied. A different airborne

method, airborne imaging spectroscopy, takes a snapshot from above to estimate emissions, can sample large areas quickly, and can capture unknown emissions leaks and sources, but currently has larger uncertainty in emission estimates than mass-balance methods (<20%). Using a spectral remote instrument, the Airborne Visible InfraRed Imaging Spectrometer - Next Generation (AVIRIS-NG), an independent sample captured a large sporadic emission 4-7 times larger than the SciAv sample three days later. The range in estimates highlights the utility of supplementing airborne mass balance methods with other methods to get a more complete understanding of emissions. Both methods are limited by intermittent sampling and cannot be solely used to characterize day-to-day emissions operations, or seasonal variability. Current national emissions inventories could be improved by incorporating independent airborne samples for validation of independent repeated and, or continuous ground, and satellite measurements with industry-reported inventories to provide better estimates. Advances in developing inventories using a combination of defensible methods at varying scales of measurement will enable more accurate and complete estimation of GHG fluxes, while holding nations and industry accountable to their emissions. These changes would be key steps to reducing emissions and addressing climate change goals.

Preface

This thesis was achievable due to the level of collaboration with the Government of Alberta, Environment and Climate Change Canada, Scientific Aviation, and the NASA Jet Propulsion Laboratory. While Chapter 2 is new unpublished research primarily written by me, it compares two algorithm methods already published by Environment and Climate Change Canada and Scientific Aviation. Emissions estimates, uncertainty, and raw data were provided by Scientific Aviation through a collaboration with the Government of Alberta. Environment and Climate Change Canada processed the data through the first step of their algorithm, and then shared the code for processing and evaluating the second half. The NASA Jet Propulsion Laboratory analyzed and shared estimates that had already been gathered over areas of interest. Chapters 1 and 3 are my original work.

Dedication

This thesis is dedicated to Nessarose Everdene; to my parents who take on the world for me, so that I may try and conquer it; to my sister, I hope to continue to be just what you always wanted; to my my gramma for all the cookies and support, and my babcia for her humour and perspective; and to Bill Donahue for his creative problem solving and encouragement without which this thesis could not have happened.

“Loyalty and obedience to wisdom and justice are fine; but it is still finer to defy arbitrary power, unjustly and cruelly used – not on behalf of ourselves, but on behalf of others more helpless.”

– Elizabeth Gaskell

Acknowledgments

Thank you to my friends and family for smiling and nodding whenever I discuss statistics.

I would also like to acknowledge my supervisor Dr. John Gamon for his kindness, support throughout my program, and encouragement to always consider the bigger picture. His guidance has allowed me to flourish and explore new possibilities of research. Thank you to Dr. Cristen Adams who has served as a role model and mentor, is always willing to help with the little questions, and supports my aspirations. To my committee member Dr. Alberto Reyes, thank you for your calm guidance. I am grateful to my Exam Chair Dr. Jeffrey Kavanaugh for his time and interest. Thank you to Dr. Mark Lewis for starting and supporting this project. I would like to acknowledge Dr. Andrea Darlington and Environment and Climate Change Canada for their time, collaboration, and work supporting this research. I would like to acknowledge the Alberta Airshed Sciences team for their support and collaboration. Thank you to Dr. Steve Conley and Dr. Mackenzie Smith of Scientific Aviation for their time analyzing the original data and explaining methods. Thank you to the NASA-Jet Propulsion Laboratory's Dr. Andrew Thorpe for your interest, honesty, and expanding the scope of this thesis. Without the collaboration and contribution from these organizations this thesis work could not have been done. Thank you to the authors who allowed adaptation of their figures for Chapter 1.

Lastly, I would like to thank Alberta Innovates, the Natural Science and Engineering Research Council of Canada, the Government of Alberta, and the University of Alberta Faculty of Graduate Student Research and Faculty of Science for the financial support they provided me for my research. This work was funded under the Oil Sands Monitoring Program and is a contribution to the program, but does not necessarily reflect the position of the program.

Table of Contents

Abstract	ii
Preface.....	iv
Dedication.....	v
Acknowledgements.....	vi
Table of Contents	vii
List of Tables	x
List of Figures.....	xii
Chapter 1: Introduction	1
1.1 Introduction	1
1.2 Top-down and Bottom-up Methods	7
1.2.1 Overview of Top-down and Bottom-up Approaches.....	7
1.2.2 Ground-based Methods.....	11
1.2.2.1 Core Samples	12
1.2.2.2 Chamber Methods.....	12
1.2.2.3 Proximal Spectral Imaging	13
1.2.2.4 External Tracer and Dispersion Models	13
1.2.2.5 Micrometeorological Methods.....	14
1.2.3 Airborne Methods	15
1.2.3.1 Mass-balance Methods.....	16
1.2.3.2 Remote Spectral Imaging.....	17
1.2.4 Satellite Methods	18
1.3 Conclusion.....	22

Chapter 2: Main Chapter	25
2.1 Introduction	25
2.2 Methodology	30
2.2.1 Algorithm Comparison Research Objectives.....	30
2.2.2 Campaign Data Used in Comparison Analysis	31
2.2.2.1 AEP - NOAA - Scientific Aviation Campaign Data	31
2.2.2.2 NASA - JPL Campaign Data	33
2.2.3 The Use and Adaptation of the Two Mass-balance Algorithms.....	34
2.2.3.1 Use of the TERRA Algorithm	34
2.2.3.2 Adaptation of the SciAv Method	36
2.2.4 Model Comparison	38
2.2.4.1 Objective 1: Compare Emission Estimates Between Models	38
2.2.4.2 Objective 2: Investigate Model Assumptions	42
2.2.4.3 Objective 3: Examine Potential for Validation of Mass-balance Emissions Estimate Methodology Through Comparison to the Spectral AVIRIS-NG Method.....	43
2.3 Results	43
2.3.1 Objective 1: Compare Emission Estimates Between Models.....	43
2.3.2 Objective 2: Investigate Model Assumptions	57
2.3.3 Objective 3: Examine Potential for Validation of Mass-balance Emissions Estimate Methodology Through Comparison to the Spectral AVIRIS-NG Method.....	59
2.4 Discussion	60
2.4.1 Objective 1: Compare Emission Estimates Between Models	60
2.4.2 Objective 2: Investigate Model Assumptions	64
2.4.3 Objective 3: Examine Potential for Validation of Mass-balance Emissions Estimate Methodology Through Comparison to the Spectral AVIRIS-NG Method.....	66
2.5 Conclusion	67

Chapter 3: Conclusion.....	69
Bibliography	72
Appendix A: Summary of Relevant Methods from the Literature	88
Appendix B: Use of the TERRA Algorithm in Further Detail	92
Appendix C: Adaptation of the SciAv Method in Further Detail.....	107
Appendix D: Comparison of SciAv and TERRA Methods using Background Surface Extrapolation Fits.....	122
Appendix E: QA/QC and Meteorological Conditions Analysis During Flight F04.....	125

List of Tables

Table 1. Descriptions, strengths, and limitations of the main methods for ground-based, airborne- and satellite-based anthropogenic carbon dioxide and methane emission estimation illustrated in the taxonomy (Figure 2).....	9
Table 2. Examples of instruments for measuring carbon dioxide and methane using the satellite methods depicted in Figure 4.	21
Table 3. Characteristics of the mass-balance SciAv and TERRA models for box flights.....	28
Table 4. Information on the five flight samples used in the comparative analysis.....	32
Table 5. Results of the CH ₄ and CO ₂ standard fit estimate from each model with their error as a percentage ± % and the range derived from the percentage error of each model’s standard estimate.....	44
Table 6. Differences between models as SciAv estimate – TERRA estimate for each flight, relative percentage differences and the propagated percentage error of the estimates.	46
Table 7. Results of parametric (weighted t-test) and non-parametric (Wilcoxon signed-rank test) significance testing of differences between the two box-flight models using the set of all five flight estimates.	46
Table 8. Percentage relative difference of estimates between models as SciAv – TERRA for the linear, interpolate, constant and background surface extrapolations.	52
Table 9. 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.....	56
Table 10. F04 CH ₄ and CO ₂ standard estimates, ranges, and the emission estimates for the SciAv model split into the flight upwards and downwards.	58
Table 11. AVIRIS-NG data captured on August 14, 2017, three days prior to the F04 flight is estimated using three sources of wind data.	60
Table 12. TERRA emission estimate uncertainties as a percentage of the total emission estimate.	99
Table 13. Background mixing ratio values used in the surface extrapolation of each flight.....	100
Table 14. Summary statistics of the raw flight data in ppm	101

Table 15. Summary statistics of the raw flight data excluding emission enhancements in ppm..	101
Table 16. Percentage change in the emission estimate when using the min and max background values for the two surface extrapolations that require setting a background value in TERRA.....	105
Table 17. F02 CH ₄ (kg hr ⁻¹) emissions estimates for each surface extrapolation fit given varying background mixing ratio values (ppm).	106
Table 18. F02 CO ₂ (kg hr ⁻¹) emissions estimates for each surface extrapolation fit given varying background mixing ratio values (ppm).	106
Table 19. F04 CH ₄ (kg hr ⁻¹) emissions estimates for each surface extrapolation fit given varying background mixing ratio values (ppm).	106
Table 20. F02 CO ₂ (kg hr ⁻¹) emissions estimates for each surface extrapolation fit given varying background mixing ratio values (ppm).	106
Table 21. The average divergence and uncertainty for each flight lap.....	107
Table 22. The profile altitude bin bounds provided by Scientific Aviation	112
Table 23. SciAv method estimates.....	110
Table 24. Fitting heights used to test the stability of the surface extrapolation of each flight. ..	112
Table 25. SciAv model emissions estimates for differing fitting heights.....	112
Table 26. Emission estimates from an adapted SciAv method. Trapezoidal integration is used to estimate emissions over the profile rather than the binning method.....	118
Table 27. Standard deviation of the surface extrapolation estimates for each integration and the difference between integration method extrapolation estimates.	119
Table 28. Results of parametric (weighted t-test) and non-parametric (Wilcoxon signed-rank test) significance testing of differences between the two SciAv integration method	120
Table 29. Mean estimates and the standard deviation of each model's surface extrapolation emission estimate.	124
Table 30. The mean differences between the two model outputs amongst the five flights and results from significance testing.....	124

List of Figures

Figure 1. Map of the 1,700 Canadian facilities reporting to the Greenhouse Gas Reporting Program for 2019, coloured by emission range	4
Figure 2. Taxonomy of methods for estimating point-source anthropogenic atmospheric GHG emissions.	8
Figure 3. Taxonomy of main ground-based methods for measuring anthropogenic carbon dioxide and methane emissions.....	11
Figure 4. Taxonomy of main airborne methods for measuring anthropogenic carbon dioxide and methane emissions.....	15
Figure 5. Taxonomy of main satellite methods for measuring anthropogenic carbon dioxide and methane emissions. Figures adapted from: (Jacob et al. 2016; Cusworth et al. 2019; Nalli et al. 2020).....	19
Figure 6. CH ₄ (left) and CO ₄ (right) mixing ratios for every one second are shown for the flight path of each facility sample used in the study.....	32
Figure 7. The flight path for F04 depicts the mixing ratio of CH ₄ (ppm) measured at 1Hz intervals for each of the 25 laps around the Syncrude plant.	33
Figure 8. AVIRIS-NG captured CH ₄ column enhancements are shown inside the F04 raw CH ₄ lap data around the Syncrude plant.	34
Figure 9. An example of a SciAv profile used in the second step of the method.....	37
Figure 10. The two different integration method applied to the SciAv F01 CH ₄ profile is depicted as the red area.....	41
Figure 11. Flight emission estimates for CH ₄ (left) and CO ₂ (right) derived for each model are plotted as points along with the range of each estimate as error bars	43
Figure 12. Sample data representing the three common types of divergence profiles for the SciAv method.....	47
Figure 13. The SciAv profile is shown for F01-F05 in Figures 13A – 13E	48
Figure 14. CH ₄ estimates given various surface extrapolations are plotted in purple for the SciAv estimate and green for TERRA	50
Figure 15. CO ₂ estimates given various surface extrapolation fits are plotted in purple for the SciAv estimate and green for TERRA	51

Figure 16. Distributions of the mean difference between all fits of CH ₄ for the SciAv and TERRA models are shown as a light blue histogram.....	53
Figure 17. Distributions of the mean difference between all fits of CO ₂ for the SciAv and TERRA models are shown as a light blue histogram.....	54
Figure 18. SciAv CH ₄ (kg hr ⁻¹) emission estimates derived using both a binning (red) and trapezoidal (blue) integration method	55
Figure 19. SciAv CO ₂ (t hr ⁻¹) emission estimates derived using both a binning (red) and trapezoidal (blue) integration method.	56
Figure 20. F04 CH ₄ divergence profiles for the flight up (left) versus down (right). The profile shapes differ with a much more variable divergence in the up profile.	58
Figure 21. F02 CH ₄ divergence profiles for the flight up (left) versus down (right). The profile shapes do not differ.	58
Figure 22. Conceptual diagram outlining when to apply an increasing towards surface extrapolation, or the SciAv standard constant surface extrapolation based on the shape of the SciAv profile for a given flight.	61
Figure 23. Conceptual diagram outlining when to use a differing integration method (such as Trapz), or the SciAv standard binning method given the number of laps in the SciAv profile for a given flight.	63
Figure 24. The TERRA integral E terms of Equation 3 with arrows depicting their contribution to estimating emission flux through the ‘box’, or volume.	89
Figure 25. 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.	91
Figure 26. The TERRA screen for F01 CH ₄ in ppm.....	93
Figure 27. The TERRA screen for flight F01 CO ₂ (ppm).....	93
Figure 28. The TERRA screen for flight F02 CH ₄ (ppm)..	94
Figure 29. The TERRA screen for flight F02 CO ₂ (ppm).....	94
Figure 30. The TERRA screen for flight F03 CH ₄ (ppm).	95
Figure 31. The TERRA screen for flight F03 CO ₂ (ppm).....	95
Figure 32. The TERRA screen for flight F04 CH ₄ (ppm)..	96
Figure 33. The TERRA screen for flight F04 CO ₂ (ppm).....	96
Figure 34. The TERRA screen for flight F05 CH ₄ (ppm).	97

Figure 35. The TERRA screen for flight F05 CO ₂ (ppm).....	97
Figure 36. TERRA produced screens of the measured emission plume for F02 CH ₄	102
Figure 37. TERRA screens for F04 CH ₄ when using 2.07 ppm (left) and 2.05365 ppm (right) as the background value.	102
Figure 38. The histogram of F02 CH ₄ raw data collected every second during the flight	104
Figure 39. The histogram of F04 CH ₄ raw data collected every second during the flight	104
Figure 40. A - E of the SciAv profiles for flights F01 through F05.	116
Figure 41. Distributions of the mean difference between all fits of CH ₄ (top panel) and CO ₂ (bottom panel) given the SciAv model for the binning and trapezoidal integration methods are drawn in light blue.....	121
Figure 42. Model 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.....	123
Figure 43. Summary of the QA/QC of flight F04 provided by the Alberta Environmental Monitoring Division and Scientific Aviation.....	126
Figure 44. Mixing ratio values of CH ₄ (top panel) and CO ₂ (bottom panel) for F04 in ppm, plotted by altitude over time.....	127
Figure 45. Wind profiles for the u and v components and change in the wind components (d) for the whole F04 flight.	128
Figure 46. Wind profiles for the u and v components and change in the wind components (d) for the flight up (top panel) and the flight down (bottom panel)	129
Figure 47. A zoomed in look at the final F04 CH ₄ box of kriging-interpolated data	130
Figure 48. Per lap CH ₄ mixing ratio screen data for the box upward (left) and downward (right) prior to kriging of the data.....	131
Figure 49. The gradient potential temperature, gradient virtual potential temperature, bulk Richardson number, and gradient Richardson numbers by altitude over time.	132
Figure 50. F04 profiles for the calculated potential temperature, virtual potential temperature, gradient (d) virtual potential temperature, gradient virtual temperature, gradient Richardson number, and bulk Richardson number.	133
Figure 51. Same as for Figure 50, but for the F04 flight up (top panel) and flight down (bottom panel).....	134

Figure 52. The five boundary layer estimates for the full F04 flight are 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 ($dVPPT/dz$) 136

Figure 53. Same as for Figure 52 given the F04 split into the upward lap data for the flight 137

Figure 54. Same as for Figure 52 given the F04 split into the downwards lap data for the flight
..... 138

Chapter 1

1.1 Introduction

The rate of warming of the Earth is greater now than at any other period over the last 22,000 years and scientific consensus is that this warming is an unequivocal consequence of human activity since the Industrial Revolution (IPCC, 2021; IPCC, 2018; Etheridge et al. 1998; Dlugokencky et al. 2011; Ciais et al. 2014; Falkowski et al. 2019; Schuur et al. 2015; Saunio et al. 2020; Le Quéré et al. 2018). Carbon dioxide and methane emissions are the first and second greatest contributors to anthropogenic climate warming (Saunio et al. 2016). Emissions since the Industrial Revolution have increased the concentration of carbon dioxide in the atmosphere from approximately 277 parts per million (ppm) in 1750 to an estimated 410 ppm in 2019 and continue to increase unabated (Friedlingstein et al. 2020; IPCC, 2021). At that rate, in 15 to 30 years, carbon dioxide concentrations in the atmosphere will exceed levels required to limit global warming below a 2 °C increase from pre-industrial levels (Rogelj et al. 2016). Global methane concentrations also continue to rise and have more than doubled from an approximate average of 695 ppb between 1000 to 1800 to an estimate of 1866 ppb in 2018 (IPCC 2021; Saunio et al. 2020; Etheridge et al. 1998). GHG emissions must be immediately reduced to have a 66% chance of staying on track to avoid drastic increases in global temperature ($> 8.5^{\circ}\text{C}$) by the end of this century (Rogelj et al. 2016; Friedlingstein et al. 2020).

Methane is the most abundant of the reactive trace gases in the troposphere (Schuur et al. 2015; Le Quéré et al. 2018); and has 20 – 80 times the warming potential of carbon dioxide depending on the timeframe considered (Balcombe et al. 2018). While 82% of carbon dioxide emissions between 1959-2017 have been from fossil fuels, decreases in methane emissions are the most efficient way to combat global warming due to the larger methane warming potential (Dlugokencky et al. 2011; Le Quéré et al. 2018). The study and estimation of methane fluxes, sources and sinks has become increasingly common despite the small concentrations in the atmosphere (Dlugokencky et al. 2011). It is estimated that 60% of global methane emissions are from anthropogenic sources including fossil fuels (19%), agriculture and landfills (38%) and biomass and biofuel burning (5%), with the remaining 40% accounted for by natural sources such as wetland emissions (31%) and geological sources, permafrost thaw, wild animals, and termites (7% combined) (Saunio et al. 2020). Atmospheric methane is largely removed through

oxidation reactions with the atmosphere and within the soil within 10 years (Saunois et al. 2020). However, even if net-zero emissions were attained, atmospheric methane concentrations are estimated to continue to increase for three decades before stabilizing due to chemical feedbacks and the potential declining oxidative capacity of the global atmosphere (Turner, Frankenberg, and Kort 2019; Heimann et al. 2020).

Independent estimates of fluxes used to derive the sources and sinks of the global carbon dioxide and methane budgets do not always agree with each other (Saunois et al. 2020; Le Quéré et al. 2018). Bottom-up methods aggregate estimates of numerous point source emissions to extrapolate an emission budget, whereas top-down methods measure the atmosphere at larger scales to estimate emissions from point and area sources (National Academies of Sciences, Engineering 2018). The largest atmospheric budget uncertainties arise from the differences in anthropogenic bottom-up inventory estimates and top-down budget estimates (Johnson et al. 2017). This disagreement has led to a yearly increase in the uncertainty of the atmospheric carbon dioxide and methane budgets in 2018 of 2.7% and 3% respectively (Saunois et al. 2020; Le Quéré et al. 2018). Reducing uncertainties of methods that contribute to the disagreement in these greenhouse gas (GHG) estimates will be an essential step towards providing better estimation of local, regional, and global atmospheric GHG changes.

International climate accords such as Kyoto (1997) and Paris (2005) have seen nations enthusiastically sign and then back out as governments' commitments shift. Despite thirty years of collective agreements, carbon dioxide emissions have continued to rise at increasing rates, illustrating the failure of current initiatives (Maslin 2020). National inventories are essential to creating a global picture of climate change and generating international pressure for action. Nations that signed the United Nations Framework Convention on Climate Change (UNFCCC) agreed to use comparable methodologies to compile national inventories of anthropogenic emissions. As of 1996, the Intergovernmental Panel on Climate Change (IPCC) has developed and refined guidelines for creating national greenhouse gas inventories (IPCC 2008). The IPCC defines levels of assessment, and the first and simplest for calculating a bottom-up estimate of emissions involves combining human activity data with the extent of related emissions (IPCC 2008). Inventory estimates that follow IPCC protocols combine tiers (1-3) of data spanning all methods of emission estimation to produce as complete an inventory as possible (IPCC 2008).

Industry inventories, environmental agencies, and academic research are relied upon in the protocol as sources of data. In North America, most inventory estimates tend to be on a smaller regional scale using the most accurate IPCC Tier 3 approaches, focusing heavily on industries and relying on ground-based methods in combination with emission intensity calculations using Continuous Emissions Monitoring System data to report to provincial, or state regulators (National Academies of Sciences, Engineering 2018, Liggio et al. 2019). Industry reporting derived from these inventories and indexes such as Canada's National Pollutant Release Inventory (NPRI) (Figure 1) can be compiled and are relied upon in many fields of study for accurate spatial ground observations (Johnson et al. 2017). For example, industry inventories are often used as prior knowledge for modeling of satellite measurements for quantifying point source atmospheric emissions (Cusworth et al. 2019). Industry inventories provide vital summarized information used as a foundation for regulation.

Independent measurements provide an essential check on industry reported values. Industry estimates utilizing solely ground-based measurements are often criticized as underestimating regional emissions (Liggio et al. 2019; Johnson et al. 2017; Liu et al. 2018; Lyon et al. 2015; Brandt et al. 2014). In a study systematically comparing 20 years of top-down and bottom-up estimates of anthropogenic methane emissions from the US natural gas and oil sectors, official bottom-up derived inventories were found to consistently under-report methane emissions (Brandt et al. 2014). Many factors contribute to this underestimation. The distribution of regional emissions is often mischaracterized so that the extent of emissions from 'super-emitters' are under-represented, and potentially biased. Large confidence intervals of estimates can further contribute to under-reporting (Brandt et al. 2014). Inventories can also be incomplete and contradictory due to methodological issues. For example, the exclusion of abandoned wells, unknown sources, and emitters below detection limits as well as a lack of modern sampling technologies can all contribute to an under-estimation of methane emissions (Riddick et al. 2019; Brandt et al. 2014; Johnson et al. 2017). Currently some heavy oil production sites report zero methane emissions despite measurable sporadic releases still occurring, as reporting is only necessary when emissions are greater than a given threshold (Johnson et al. 2016). Upscaling using bottom-up methods that do not account for sporadic emissions are bound to produce inventories that underestimate emissions.

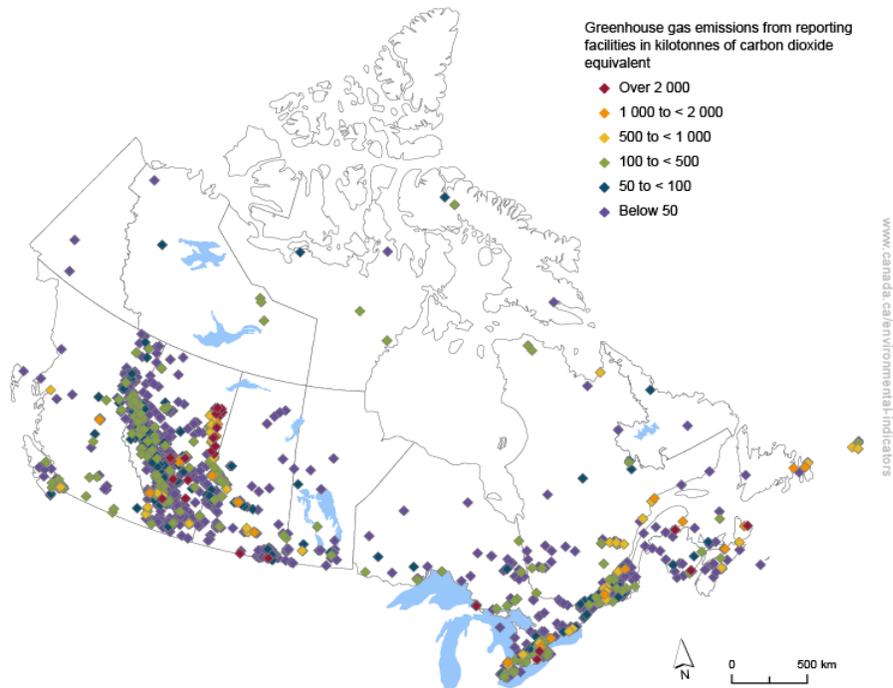


Figure 1. Map of the 1,700 Canadian facilities reporting to the Greenhouse Gas Reporting Program for 2019, coloured by emission range. Source: Environment and Climate Change Canada.

Addressing the climate crisis has become a National and Provincial priority in Canada with the adaptation of carbon pollution pricing systems, improvements in the transport sectors’ fuel efficiency, and investment in renewable resources research and adaptation (Canada, 2016; Environment and Climate Change Canada, 2021). Canada contains the world’s third largest oil reserves, and of all nations has the fourth highest volume-weighted average crude oil upstream GHG intensities, just behind Algeria, Venezuela, and Cameroon (Masnadi et al. 2018).

The province of Alberta emits approximately 54% of the nation’s GHG emissions and, as shown in Figure 1, contains 696 of the 1700 facilities reporting to the national greenhouse gas emissions program (Government of Canada 2021c). Approximately 35% of national GHG emissions are from oil and gas operations (Government of Canada 2021c). Canada has committed to reducing greenhouse gases (GHGs) 40-45% below 2005 emission levels by 2030, and the Government of Alberta has committed to a 45% reduction in methane emissions from the oil and gas sector by 2025 (Government of Alberta 2015; Government of Canada 2021b). Further, at the COP26 summit in Glasgow in the Fall of 2021, Canada committed to an emissions cap on the domestic oil and gas sector, with the goal of achieving net zero by 2050 (Government of Canada 2021a).

A critical part of achieving these goals is the development and implementation of robust GHG emissions monitoring programs (Cronmiller and Noble 2018). The Joint Oil Sands Monitoring Program (JOSM) was created in 2012 to deliver scientifically defensible environmental monitoring and cumulative impacts assessment studies, under the shared leadership of the Governments of Alberta and Canada, and was renamed the Oil Sands Monitoring (OSM) Program in 2015. Since then, the program has produced rigorous scientific publications focusing on mineable oil sands operations in the Athabasca Oil Sands Region (e.g., Liggio et al. 2019; Baray et al. 2018; Liggio et al. 2016; Li et al. 2017; Small et al. 2015; Zhang et al. 2018; Shephard et al. 2015). OSM has an approximate \$50 million per year budget funded under Alberta's Oil Sands Environmental Monitoring Program, which dictates that oil sands companies in Alberta contribute to funding in return for deemed compliance with conditions on regulatory approvals that relate to such activities (Government of Alberta 2013). In 2013, the Government of Alberta legislated the creation of an independent monitoring agency to perform environmental monitoring in Alberta (Alberta Environmental Monitoring, Evaluation and Report Agency; AEMERA), which was established and received its initial budget in 2015. However, the agency was dissolved in 2016, less than a year after its independence was functionally established, and reincorporated into the Government of Alberta as the Environmental Monitoring and Science Division (EMSD), a decision that has been criticized by the scientific community (Cronmiller and Noble 2018). Alberta's EMSD has since been further integrated into the Government of Alberta, and monitoring work is now done under the Resource Stewardship Division. During the COVID-19 pandemic, in addition to the cancellation of a number of OSM monitoring programs and research studies, Alberta suspended many industrial reporting obligations when environmental monitoring was reduced across Canada (Goodday 2021).

This history of discontinuity in Oil Sands monitoring has led to concerns about credibility and transparency (Cronmiller and Noble 2018). The lack of stability in monitoring programs has presented both scientific and management challenges. Given the goal of reducing methane emissions by 45% by 2025, it is critical to have a well coordinated emissions reduction plan. However, the Government of Canada and the Province of Alberta have adopted different and potentially conflicting emissions reduction regulations and targets, with increasing public disagreement between ministers, and while federal policies are found to be stronger, neither government is currently poised to achieve the 2025 target (Johnson and Tyner 2020; Canadian

Press 2021). The Government of Canada has pledged to end federal financing for foreign fossil fuel projects in 2022 (The Globe and Mail 2021). In 2019, the Government of Alberta initiated a “fight-back” policy which included creating the Canadian Energy Centre to promote Alberta's O& G sector, and a public inquiry into “anti-Alberta energy campaigns” and foreign investment in environmental initiatives (Government of Alberta 2019; Allan 2021). These recent developments suggest that environmental monitoring in the Oil Sands will continue to be a complex and contentious issue for the foreseeable future.

Scientific autonomy from political mandates is essential to the credibility of monitoring and science programs. Instability, lack of scientific leadership, lack of clarity of purpose, and what can be complex technical issues are a challenge for any long-term monitoring program (Cronmiller and Noble 2018), let alone one that appears to be at odds with a government's stated public priorities and policies. Despite this complex history, the Government of Alberta remains engaged in implementing monitoring and emissions reductions plans (Boyd 2019). Ultimately, the success of any monitoring program will depend on the rigour of the science underpinning it, the use of appropriate technology enabling it, and adoption of a clear purpose and mandate.

Comparisons of top-down and inventory-based estimates have suggested that the Alberta Oil Sand's carbon dioxide and methane emissions are considerably under-estimated (Liggio et al. 2019; Johnson et al. 2017; Baray et al. 2018). Increased monitoring for identifying and mitigating unknown methane emission leaks and large anomalous emissions is one of the most efficient approaches to reducing global GHG emissions (Frankenberg et al. 2016). The recent development and ongoing refinement of top-down methods make them an increasingly popular tool for independent estimation of point-source anthropogenic carbon dioxide and methane emissions.

This review aims to provide an overview of current top-down, and bottom-up methods for estimating point source and regional anthropogenic carbon dioxide and methane emissions with the goal of assessing effective methods that might help improve monitoring accuracy. Three branches of point, and area source estimation methods are reviewed: a) relatively small-scale, ground-based methods, b) intermediary airborne methods, and c) regionalized satellite methods. Characterizing the strengths and weaknesses of methods and the potential for synergies between them is an essential step towards creating exhaustive estimations of anthropogenic emissions.

1.2 Top-down and Bottom-up Methods

1.2.1 Overview of Top-down and Bottom-up Approaches

The current uncertainty in the GHG fluxes that comprise emission budgets is epitomized by the gap between top-down and bottom-up estimates (Saunois et al. 2020; Le Quéré et al. 2018). Bottom-up and top-down measurements can have varying definitions depending on the spatial scale and context of estimation (Gordon et al. 2015). In general bottom-up methods can be described as measurements to obtain component or site-specific emission data that are then extrapolated by some emission factor to estimate regional emissions (National Academies of Sciences, Engineering 2018; Karion et al. 2013). Top-down measurements attempt to constrain the overall budget of atmospheric GHG concentrations by sampling at larger spatial scales and utilizing modeling tools to infer point and area source emissions estimates (Gordon et al. 2015; National Academies of Sciences, Engineering 2018). National inventories are compiled from data using both methods (IPCC 2008). Top-down estimates provide crucial validation for how well external and industry bottom-up derived inventories estimate regional budgets; often estimates do not agree well and top-down estimates tend to be higher (Johnson et al. 2017; Liggio et al. 2019; Li et al. 2017; Brandt et al. 2014). By attaining greater knowledge of emission budgets, top-down sampling can help plan for and achieve GHG reduction goals (Johnson and Tyner 2020; Johnson et al. 2017).

Top-down and bottom-up approaches to measure gas fluxes can be further defined as ground-based, airborne, or satellite approaches, which are examined in this review. Airborne methods are the primary focus of this study. Top-down methods can include measurements from global surface networks of large tower data, aircraft campaigns and satellites depending on the context and spatial scale (Conley et al. 2017). Examples of bottom-up, ground based methods include proximal spectral imaging, external tracer, and flux chamber measurements (Fox et al. 2019; National Academies of Sciences, Engineering 2018). Data from both approaches are increasingly being included in multi-scale models (Zhang et al. 2018; Nambiar et al. 2020; Varon et al. 2018). Spectral imaging methods are unique in their ability to capture unknown, or anomalous point sources at the proximal, remote, or orbital level. A taxonomy (Figure 2) illustrates methods for quantifying atmospheric anthropogenic GHG emissions at the varying spatial scales. The ground-based heading is broken into stationary, mobile or tower sampling

methods and lower branches provide examples of techniques that utilize one, or multiple methods. Airborne methods sit in the middle between bottom-up and top-down approaches and can be classified as either depending on the scale, and context of sampling (National Academies of Sciences, Engineering 2018). Airborne methods are broken into the two main approaches of estimating point and area source emissions by flying and sampling through or around emissions to apply mass-balance algorithms (Conley et al. 2017; Gordon et al. 2015), or flying above potential emission sources and utilizing remote spectral imaging methods (Frankenberg et al. 2016). Satellite spectral imaging methods are split into Lidar, thermal infrared, and shortwave infrared instruments. Almost all methods utilize direct, proximal, or remote spectroscopic methods to quantify the concentration of methane and, or carbon dioxide in their given sample. Table 1 discusses the main methods from Figure 2, provides coordinating references and summarizes their strengths and weaknesses.

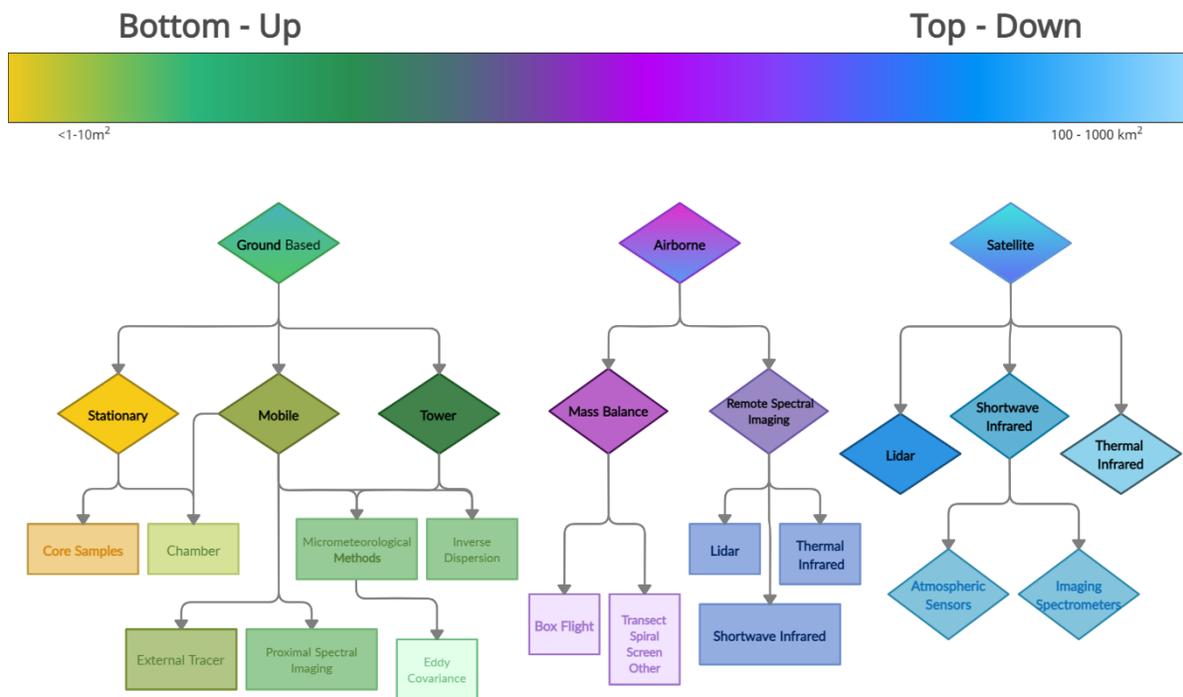


Figure 2. Taxonomy of methods for estimating anthropogenic atmospheric GHG emissions. A left to right, yellow-green-purple-blue gradient generalizes the spatial range of measuring of pixel size from less than 1-10 m² to the upper range of 100 - 100 km², and maps each method in the given order for bottom-up and top-down classification. More colour saturated upper branches depict main method subcategories, whilst the lightly coloured lower branches provide examples of techniques, data types, and algorithms.

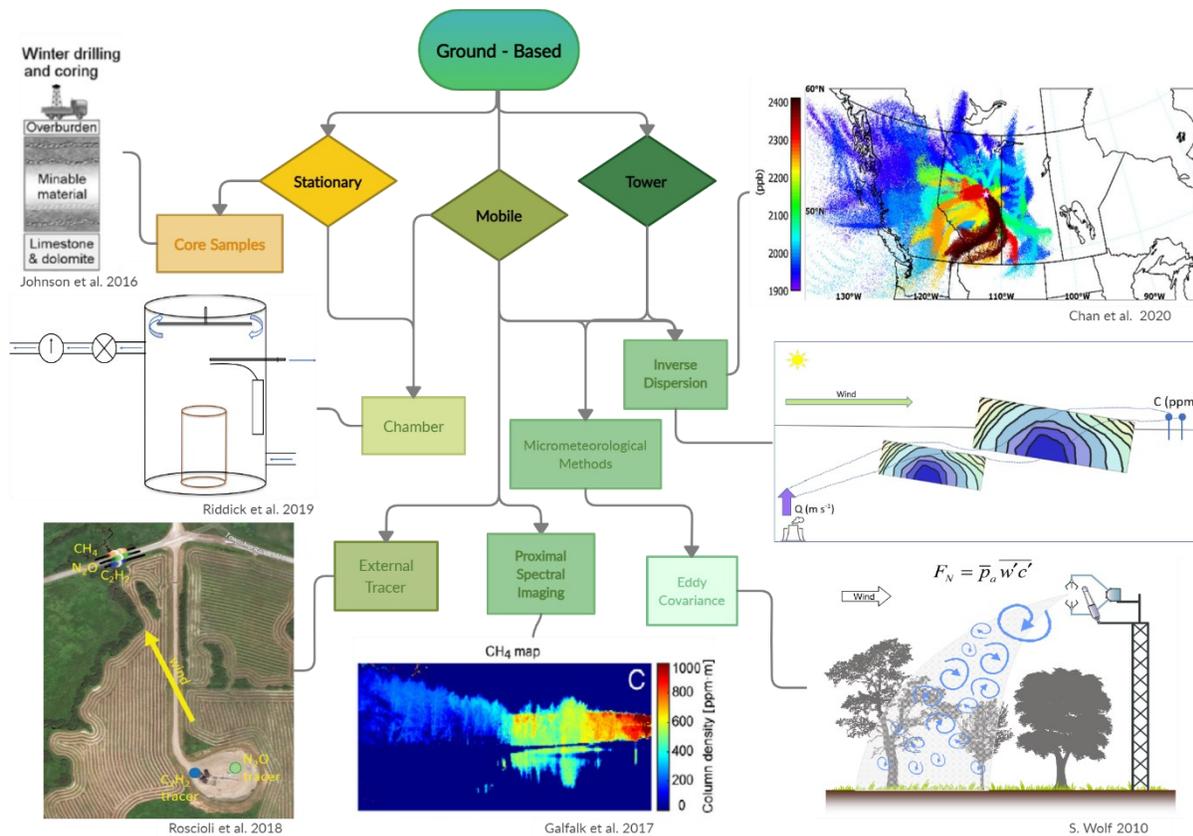
Table 1. Descriptions, strengths, and limitations of the main methods for ground -based, airborne- and satellite-based anthropogenic carbon dioxide and methane emission estimation illustrated in the taxonomy (Figure 2).

Method	Discussion	Strengths	Limitations	Main Reference(s)
Ground: Core Samples	Measured methane concentrations from core samples are used to approximate the release budget at a facility and extrapolate emission release by the extent of oil processing.	Provides a unique approach to estimating emissions from below the ground. Used to estimate the regional methane emission budget constraints from rate of extraction and processing.	Limited by the correlation between concentrations measured in samples and emissions released during the processes. Requires multiple samples from sites. Currently only used to estimate methane emissions.	(Johnson, Crosland et al. 2016)
Ground: Chamber Based	Directly measures emission concentration and flux rates from a small diffuse plume.	Low uncertainty of emission estimation within the sample. Method is independent of atmospheric modeling methods. Can sample at night.	Limited to small scales of sample and is very reliant on well-developed sampling schemes. Experiences bias from chamber artifacts. criticized for missing sources and variability in regional estimates. Difficultly capturing sporadic emissions.	(Jeong et al. 2019; Chaichana et al. 2018)
Ground: Proximal Spectral Imaging	Proximal ground-based imaging spectroscopy to quantify and characterize plumes.	High spatial and spectral resolution to identify samples at near ambient levels. Does not rely on external sources for wind measurements. More cost-effective than most airborne sampling.	Uncertainties increase as temperature contrast of GHG to the background decreases. Reliant on isolating the absorption lines from background noise. Cannot quantify large areas and is restricted by the distance of sampling.	(Gålfalk et al. 2017)
Ground: External Tracer	Models emission concentrations from the downwind distribution of known concentration of a tracer gas from a plume source.	Well developed, simplified method. Can handle complex sources and measure total emissions from a source area.	Labour intensive. Requires consistent, stationary wind and stable atmospheric conditions. Method breaks down if the desired gas does not follow the same dispersion as the tracer gas.	(Roscioli et al. 2018; National Academies of Sciences, Engineering 2018)

Ground: Inverse Dispersion	Directly measures concentrations of downwind plumes to model the upwind emission rate.	More cost-effective sampling than airborne methods and more robust than tracer or chamber-based methods. Can be used to quantify temporal trends.	Larger errors can occur from non-stationary wind and plume sources. Difficult to quantify and isolate complex, mixed sources. Reliant on externally modelled atmospheric conditions that may not match sampling conditions.	(Flesch et al. 2005; National Academies of Sciences, Engineering 2018)
Ground: Micrometeorological measurements [e.g., “eddy covariance”]	Models flux from direct measurements of plume transfer in the appropriately scaled area of sampling.	Largest ground-based scale of atmospheric sampling. Ideal for capturing temporal trends. Measures uptake as well as atmospheric flux. More cost-effective method of sampling large areas than airborne sampling.	Constrained by the need for homogenous terrain and stable atmospheric conditions for low error in estimates. Requires extremely rapid-response sampling device, which is expensive.	(Chaichana et al. 2018; Davidson et al. 2002; National Academies of Sciences, Engineering 2018)
Airborne: Mass Balance	Airborne measurements for quantification of localized, or regional plumes.	Can attain very low uncertainty (~ 2 %) in estimating fully captured stationary plume sources. Provides ideal in-depth modeling of regional emissions. Used for validation of both ground-based and satellite methods.	Costly sampling that requires known sources. Limited by boundary layer height. Issues with external sources, shifting plumes and widely dispersed ground sources. Dependent on good extrapolation from lowest flight to the ground.	(Conley et al. 2017; Gordon et al. 2015)
Airborne: Remote Spectral Imaging	Absorption imaging spectroscopy of reflectance or thermal emissions to capture regional or facility emissions.	Quick sampling. Best for large mapping of plumes in a region. Can identify unknown sources such as leaks. Avoids temporal issues inherent to mass-balance methods.	Limited by meteorological conditions due to spectral interference. Requires multiple sampling to determine persistence of a source.	(Frankenberg et al. 2016; Bartholomew et al. 2017)
Satellite: Remote Spectral Imaging^a	Quantification and imaging spectroscopy of reflectance or absorption to capture global, regional, and local estimates of emissions.	Repeatedly samples global and regional emissions. Provides independent monitoring.	Currently has coarse spatial resolution. Some methods are restricted by spectral interference and difficult sampling of dark scenes, or high reflectance scenes such as snow, or water.	(Cusworth et al. 2019; Kort et al. 2014; Jacob et al. 2016)

^aSee Table 2 for a further breakdown of the satellite methods.

1.2.2 Ground-based Methods



**Figure 3. Taxonomy of main ground-based methods for measuring anthropogenic carbon dioxide and methane emissions. Figures adapted from: (Chan et al. 2020; Gålfalk, Olofsson, and Bastviken 2017^a; Johnson et al. 2016; Riddick et al. 2019; Roscioli et al. 2018; Wolf 2010).
^aReprinted with permission from Elsevier.**

Ground-based methods are currently the most direct route for continuous estimation of point source and regional anthropogenic emissions. Figure 3 illustrates the types of methods that can be applied through sampling on stationary towers, fixed sensors, portable handheld devices, or automobile laboratories (Fox et al. 2019). Tall towers provide continuous observations for estimating regional emissions and networks of towers can provide ground-based, top-down atmospheric estimates (Chan et al. 2020). Mobile laboratories are often used to provide numerous spatially explicit eddy covariance, flux chamber, or tracer gas samples to scale up to site and regional estimates (Johnson et al. 2017; Roscioli et al. 2018). Site features, such as tank vents, well pads, pump stations or shutoff valves, are often best estimated by ground-based measurements for pin-pointing source emissions due to their small spatial scale (Roscioli et al.

2018). Ground-based, bottom-up methods can be more challenging than airborne methods to quantify an regional area of emission as they are dependent on good extrapolation (Flesch, Wilson, and Harper 2005; Roscioli et al. 2018). Ground-based measurements provide validation data for top-down methods; tower data are essential for satellite validation and surface measurements provide validation for extrapolation calculations used in mass-balance airborne methods. (Gordon et al. 2015; Janssens-Maenhout et al. 2020; Nalli et al. 2020).

1.2.2.1 Core Samples

A novel approach to quantifying point source, fugitive methane emissions from heavy oil mine sites has been invented by Johnson et al. 2016. Fugitive emissions can be estimated by quantifying the amount of methane present *in situ* and subtracting the remaining gas after processing (Johnson et al. 2016). Measurements are limited by the reliance on industry to obtain the core samples and currently have high rates of uncertainty for estimates (34- 69%) (Johnson et al. 2016). This method presents a new way of constraining a budget for atmospheric methane emissions from below the ground, rather than from on or above the ground, and estimates the potential emissions of proposed mining sites (Johnson et al. 2016).

1.2.2.2 Chamber Methods

Chamber measurements are a simple way of measuring fluxes with few modeling assumptions and as such are commonly used for estimates by industry (National Academies of Sciences, Engineering 2018). Chamber sampling methods provide a direct measurement of emissions from sources and when instruments are mobile they can more completely cover an ecosystem (Chaichana et al. 2018); however, their spatial sampling range is very small. To address this, Jeong et al. 2019 assessed interpolation methods of point sampling to determine an efficient sampling scheme that maximizes distribution while minimizing the number of samples for a given spatial interpolation (Jeong et al. 2019). Poor design of stratified sampling, and or using a limited number of samples can result in large over, or under estimation of fluxes (Chaichana et al. 2018; You et al. 2021). Chamber artifacts must be minimized when placing the chamber as the act of measuring can bias the flux concentrations (Davidson et al. 2002). Closed chamber methods can be successful in accurately measuring emissions from well-known temporally distributed emissions (Johnson et al. 2016). Flux chambers are ideally suited to measuring small, contained sources (Riddick et al. 2019; You et al. 2021). Dynamic flux chambers were

successfully used in a campaign to estimate fugitive methane emissions from abandoned wells by fitting the chambers over the wells to estimate leakage (Riddick et al. 2019). Issues arise in extrapolating samples from complex sites, with large sporadic emissions, so chamber methods are not well suited for estimating annual emissions from an entire oil production site (Johnson et al. 2016; You et al. 2021).

1.2.2.3 Proximal Spectral Imaging

Recent proximal imaging spectroscopy methods are becoming increasingly popular as a cost-effective way to capture unknown, fugitive emissions, produce imagery to characterize sources, as well as estimate emissions (Gålfalk, Olofsson, and Bastviken 2017; Kobayashi et al. 2010; Wojcik et al. 2015). A recent hyperspectral method can capture both known and unknown sources by simultaneously sampling meteorological variables, and methane and nitrous oxide in the 1.0 – 5.5 μm midwave range and the 7.7-9.5 μm longwave range to estimate emission flux from a single instrument (Gålfalk, Olofsson, and Bastviken 2017). Handheld, and mobile spectral devices are also being developed that utilize differential absorption lidar (DIAL) to detect, locate, and quantify carbon dioxide and methane emissions from oil and gas facilities (Wojcik et al. 2015; Johnson, Tyner, and Szekeres 2021). Both methods are used to scan scenes, take ‘snapshots’ to locate fugitive leaks, and assess source emissions. Ideal samples are captured when conditions allow for a distinct emission enhancement that can be easily separated from the background (Gålfalk, Olofsson, and Bastviken 2017). Methods by Gålfalk, Olofsson, and Bastviken 2017 produce near-ground measurements of the distribution of GHGs from various environments, sources and processes images compiled as movies to characterize plume behaviour.

1.2.2.4 External Tracer and Dispersion Models

The tracer-gas ratio method has been used to measure emissions from fossil fuel extraction sites for over 20 years (Roscioli et al. 2018). It applies the working assumption that a known gas has a dispersion distribution identical to the desired emission gas and estimates emissions based on a known ratio of concentrations (Roscioli et al. 2018). Dispersion models can be applied to, or combined with tracer measurements, eddy covariance, or spectral laser measurements to extrapolate given samples to the distribution of emissions from a site (Flesch, Wilson, and Harper 2005; Jeong et al. 2019). Inverse dispersion modeling works backwards sampling a

known concentration to estimate emission information from an upwind source (Flesch, Wilson, and Harper 2005). A recent study aggregated 6650 mobile ground-based measurements using external tracer and inverse dispersion model to produce estimates from the upstream Canadian oil and gas and found that inventories underestimated methane emissions by a factor of 1.5 (MacKay et al., 2021). Methods for up-scaling to larger regional methods are still being developed, and methods are best used for well-known consistent concentrations (National Academies of Sciences, Engineering 2018). Both external tracer and dispersion modeling benefit from averaging, therefore the greater the spatial distribution and number of samples, the lower the error (Roscioli et al. 2018; Chan et al. 2020; Zavala-Araiza et al. 2018).

1.2.2.5 Micrometeorological Methods

Flux towers can directly measure methane and carbon dioxide fluxes as they occur using eddy covariance techniques. They are able to detect small changes in net ecosystem exchange, or the flux between the atmosphere and vegetation, by measuring eddies to calculate the covariance between GHG mixing ratios and vertical wind velocities (Baldocchi et al. 2001). Measurements can be analyzed independently or combined with samples acquired using other ground-based methods to model flux within the area of sampling. Regional Bayesian inverse dispersion modeling, using existing inventory estimates, was recently successful in modeling eight years of observations from four tall tower sites to estimate annual methane emissions from Alberta and Saskatchewan (Chan et al. 2020). Flux towers are often constrained by the scale of sample, a lack of homogeneous terrain and a stable atmosphere. For example, rugged mountainous areas with volatile atmospheric boundary layers make for inopportune flux tower locations. In spite of these limitations, flux towers are emerging as a fundamental source of GHG emission data as this technology becomes more available (AmeriFlux 2019). There is an increasing global abundance of flux tower sites providing continuous large scale measurements, and international networks such as FLUXNET are enhancing the value of using the eddy covariance method for validating airborne and satellite data (Baldocchi et al. 2001, AmeriFlux 2019).

1.2.3 Airborne Methods

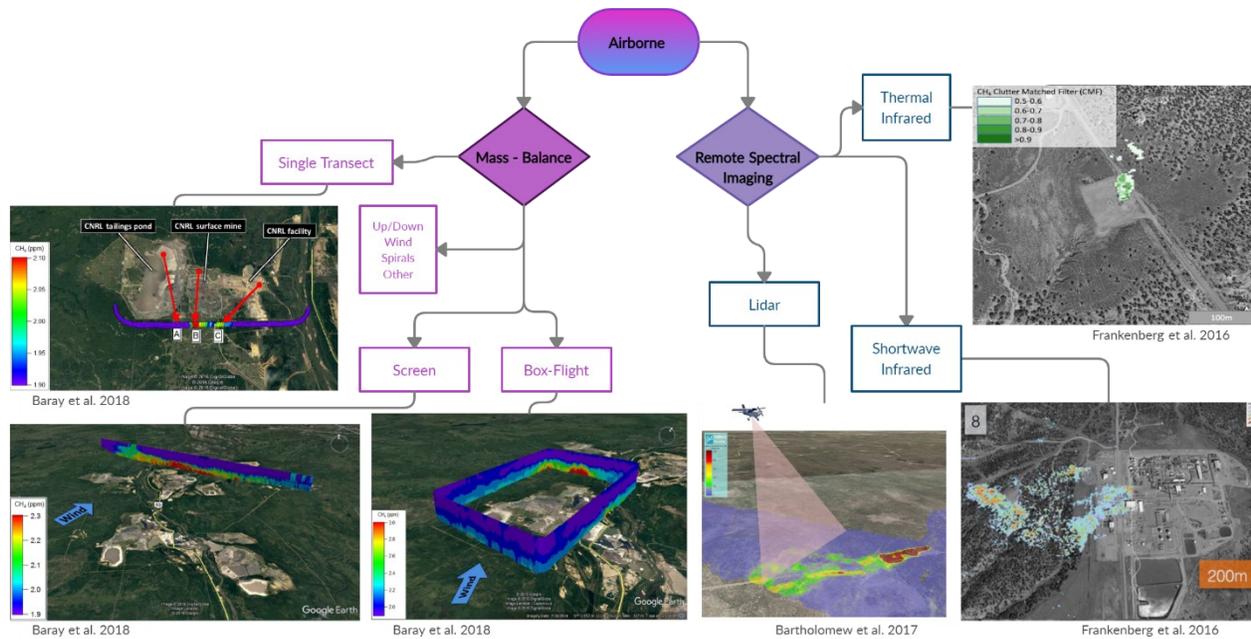


Figure 4. Taxonomy of main airborne methods for measuring anthropogenic carbon dioxide and methane emissions. Figures adapted from: (Baray et al. 2018; Bartholomew et al. 2017^a; Frankenberg et al. 2016). ^a Reprinted with permission from SPIE.

Airborne measurements are used to characterize and validate emissions from sites (Johnson et al. 2017). They provide an intermediary sampling scale that can attain detailed quantification of emission budgets from sites and sample large regional areas (Karion et al. 2013). Methods for sampling atmospheric GHGs from the air tend to fall into two major categories: i) remote spectral imaging methods and ii) mass-balance spectroscopic methods. Figure 4 illustrates the difference between the sampling schemes and imagery produced from the two approaches. Accurate wind measurements are paramount to all airborne methods. Airborne methods are limited by airspace restrictions and can be costly and time intensive which makes them ineffective as continuous monitoring methods; but they provide in depth emissions estimates for supplementing and auditing emissions inventories, as well as follow up to emission hotspots flagged by satellite data (Varon et al. 2019). Airborne methods are weaker at assessing the temporal variability of emissions and rely on ground-based, or satellite measurements to upscale hourly emission rates to monthly and yearly estimates for inventories (Liggio et al. 2019, Thorpe et al. 2020). Drones have become popular in some fields to sample methane and carbon dioxide

emissions hotspots from the air, and progress is being made in emissions estimation, however they currently sample at too small a scale to get meaningful data for monitoring oil and gas facilities (Burgués and Marco 2020). As they avoid some ground access and safety issues, and to the extent that they can sample large areas over facilities, drones and airborne methods are often ideal tools for monitoring pipelines for leaks (Burgués and Marco 2020; Frankenberg et al. 2016). Along with satellite methods, certain spectral airborne methods often complement ground measurements, particularly for the flux tower network (FLUXNET); this combination of direct flux assessment from the ground with remote sensing methods creates a particularly powerful synergy and is central to the emerging spectral tower networks (SpecNet) (Gamon et al. 2006; Baldocchi et al. 2001).

1.2.3.1 Mass-balance Methods

Mass-balance methods most commonly capture point and area source plumes by flying in transects, screens, or box-flights that constrain the emission plume as it traverses the flight path around the source (Gordon et al. 2015; Frankenberg et al. 2016; Jongaramrungruang et al. 2019; Baray et al. 2018; Karion et al. 2013; Kalthoff et al. 2002). Airborne mass-balance flights can sample following several sampling schemes. The simplest are single transects and screens which are faster, but have large errors (25 - 60%), or there are more complex time-intensive box-flights which attain lower estimate error (~2%) (Gordon et al. 2015). Two box flight mass-balance algorithms, the Top-down Emission Rate Retrieval Algorithm (TERRA) and SciAv (Scientific Aviation) model, have recently been applied to box-flight patterns when an aircraft encircles a source to estimate emissions by calculating the flux through the boxed-in source (Gordon et al. 2015; Conley et al. 2017; Johnson et al. 2017; Frankenberg et al. 2016; Baray et al. 2018; Liggio et al. 2019; Duren et al. 2019). Full capture at the top of the box is often attained by operators flying laps up to the top of the (stable) atmospheric boundary layer, which typically caps the top of an emission plume (Gordon et al. 2015). Due to minimum flight height restrictions, a gap between the surface and the flight box is inevitable. Extrapolation to the ground has been shown to often be the largest error source, nearing ~30% in both models when the bottom of the plume is not captured (Conley et al. 2017; Gordon et al. 2015). Currently, mass-balance box flight methods can attain a lower uncertainty in emission estimates than the remote spectral imaging due to smaller background and wind measurement uncertainties, but require prior understanding

of plume sources, stationary conditions, take longer, and are more costly (Thorpe et al. 2020; National Academies of Sciences, Engineering 2018).

1.2.3.2 Remote Spectral Imaging

Remote spectral imaging from the air can be used to quickly cover broad areas providing snapshots of plumes for estimation, catch the large sporadic leaks that are common to fossil fuel extraction sites, and depict a distribution of plumes to characterize an area of interest (Thorpe et al. 2017; Bartholomew et al. 2017; Frankenberg et al. 2016; Sherwin et al. 2021; Johnson, Tyner, and Szekeres 2021; Cusworth et al. 2021). This method samples within seconds and therefore avoids the temporal limitations inherent to airborne mass-balance methods, and its ability to quickly capture unknown leaks and emission sources provides a cost-effective method for reducing GHG emissions over large regions (Thorpe et al. 2020, Schwietzke et al. 2019, Sherwin et al. 2021). By rapidly addressing methane leaks, a recent study of the Southern Midland Basin using an airborne hyperspectral imaging instrument recovered costs of their first flight campaign within 5 days (Johnson et al. 2021). However, remote spectral sampling is still collected on a “campaign” basis, and is not typically used for routine emissions monitoring (Duren et al. 2019).

Passive airborne methods utilizing shortwave or thermal infrared spectroscopy have been developed to map carbon dioxide and methane concentrations, and detect and estimate emission rates over large areas (Duren et al. 2019; Frankenberg et al. 2016; Thorpe et al. 2017, Sherwin et al. 2021, Johnson, Tyner, and Szekeres 2021). A shortwave infrared instrument, the Airborne Visible InfraRed Imaging Spectrometer - Next Generation (AVIRIS-NG; Hamlin et al. 2011), (NASA Jet Propulsion Laboratory, Pasadena, CA, USA), can attribute and estimate point source emission rates as small as 1-3 m with a detection limit of 2 kg h⁻¹ to 5 kg h⁻¹ depending on wind speed with uncertainties of ~ 30% (Thorpe et al. 2020, Duren et al. 2019). The Kairos LeakSurveyor (Kairos Aerospace, Mountain View, CA, USA), is a smaller, relatively inexpensive instrument designed for commercial deployment, rather than research purposes of the AVIRIS-NG (Schwietzke et al. 2019). It calculates a wind-adjust methane emission rate with a minimum detection level of 5 kg per hour per meter per second of wind (kg h mps⁻¹) with a method error of ~ 30 - 40 % uncertainty (Sherwin et al. 2021). Shortwave infrared airborne sampling can be confounded by spectral inference from the surface (Thorpe et al. 2013), or reflectance from features such as tailings ponds (Krings et al. 2018). Thermal infrared airborne

sampling using instruments such as the Hyperspectral Thermal Emission Spectrometer (HyTes), (NASA Jet Propulsion Laboratory, Pasadena, CA, USA), provide a useful complement to shortwave infrared as they can sample when spectral interference impedes infrared sampling, but have coarser spatial resolution making them less ideal for emission estimation (Frankenberg et al. 2016). Several airborne imaging campaigns in California, Colorado, Texas, and New Mexico have been successful at mapping large regional methane plumes and estimating point sources and are increasingly becoming a popular method (Thorpe et al. 2020, Johnson, Tyner, and Szekeres 2021, Duren et al. 2019, Sherwin et al. 2021).

Airborne Lidar methods are active methods that are ideal for quickly sampling large areas for unknown sources and to characterize regional patterns in background and anthropogenic emissions (Rashid et al. 2020; Li et al. 2010; Bartholomew et al. 2017). The IPDA Lidar (Ball Aerospace & Technologies Corp, Boulder, CO, USA) provided one of the first unique opportunities to the measure regional and temporal variability in background concentrations, which had largely been restricted to ground-based methods and modeling (Bartholomew et al. 2017). Aircraft vibration can create stripes in the Lidar data, and determining the ideal beam diameter to measure at, and dependence on accurate geolocation data are weaknesses of the method (Bartholomew et al. 2017). The Gas Mapping LidarTM (Bridger Photonics, Bozeman, MT, USA) has a spatial resolution of 2 m, and can estimate methane emissions with an uncertainty range of 31 – 68 % (Johnson, Tyner, and Szekeres 2021). In 2019, the Bridger Photonics lidar completed a campaign across 167 geographically distinct sites in British Columbia and estimated 80 methane sources with a range in emission averages from 0.5 to 399 kg hr⁻¹ (Tyner and Johnson 2021). While estimate uncertainties are on average higher for airborne lidar than shortwave infrared, lidar methods can currently attain a lower detection limit of emission plumes (~ 0.6 kg hr⁻¹ versus 2 kg hr⁻¹) (Johnson, Tyner, and Szekeres 2021, Thorpe et al. 2020).

1.2.4 Satellite Methods

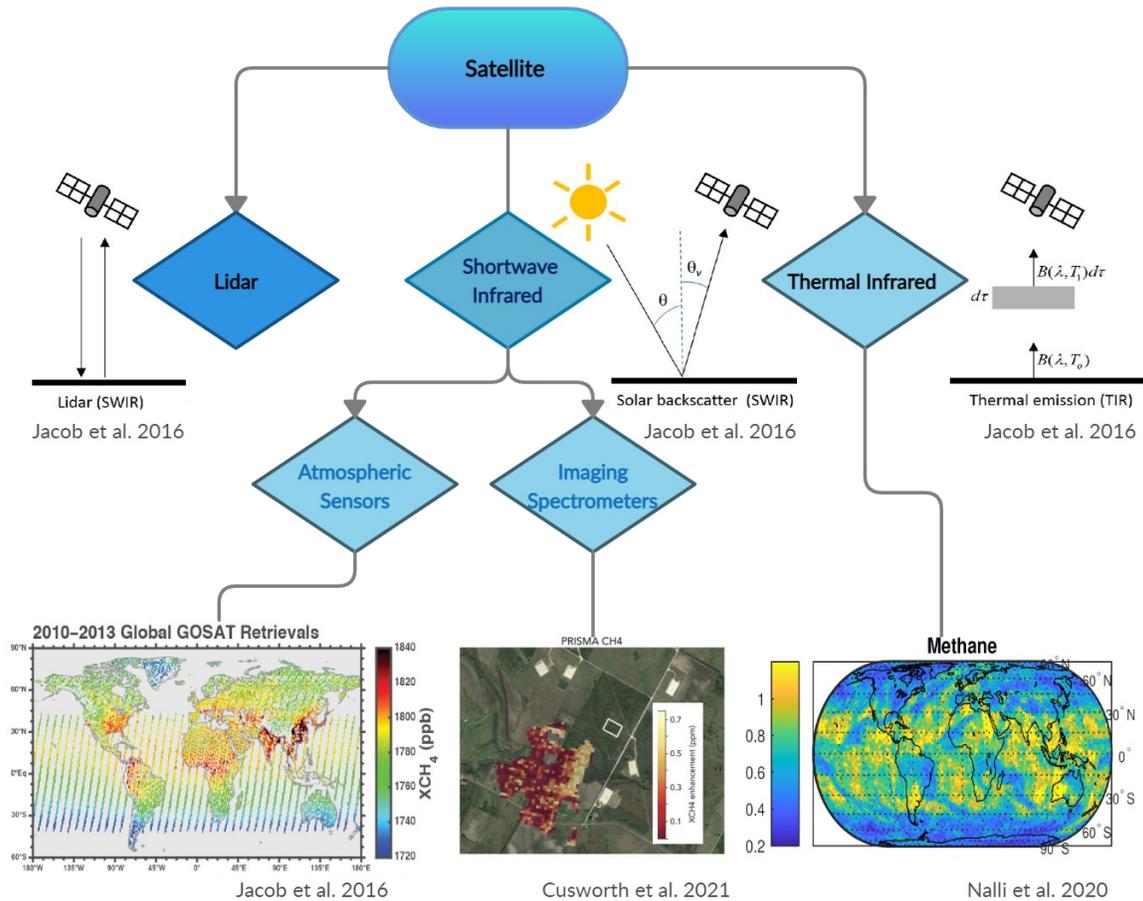


Figure 5. Taxonomy of main satellite methods for measuring anthropogenic carbon dioxide and methane emissions. Figures adapted from: (Jacob et al. 2016; Cusworth et al. 2021; Nalli et al. 2020).

Satellite measurements are ideal for identifying regional differences in methane emissions (Kort et al. 2014). Naturally, instruments on satellites provide the largest spatial top-down atmospheric carbon dioxide and methane measurements (Fox et al. 2019). They allow for repeated sampling over regional to global scales and can produce estimates where it is not feasible to sample using other methods (National Academies of Sciences, Engineering 2018). Figure 5 depicts current methods for satellite observations of atmospheric methane and carbon dioxide by measuring the absorption of radiation by solar backscatter in either the shortwave infrared range (SWIR), or thermal emissions in the thermal infrared (TIR) wavelengths (Jacob et al. 2016). While each GHG has different measurement bands, each satellite instrument also measures absorption features in slightly differing spectral ranges called fitting windows. In

general, carbon dioxide is measured in SWIR and near infrared ranges (NIR) around 0.76 μm , 1.61 μm and 2.06 μm and in TIR around 15 μm (Nassar et al. 2017). Methane is measured in the SWIR spectrum around 1.65 μm or 2.3 μm wavelengths, or in TIR wavelengths around 8 μm (Jacob et al. 2016).

Table 2 summarizes some of the instruments used in the methods illustrated in Figure 5 for measuring carbon dioxide and, or methane emissions from space. Atmospheric sensing satellites passively sample using large wavelength ranges and at a high spectral resolution to capture absorption features. Satellite imaging spectrometers aim to address the desire for fine pixel resolution to resolve point sources while increasing the spectral resolution (Cusworth et al. 2019). TIR methods are not restricted by spectral interference and can sample dark sources, but tend to have poor sensitivity of lower tropospheric emissions which can make point source estimation difficult (Jacob et al. 2016). Satellite Lidar provide active measurements of emissions using its own laser light source, freeing it from sun illumination, and can therefore measure during the day or night (Wührer et al. 2019). This means satellite Lidar methods using Integrated Path Differential Absorption will uniquely provide the ability to measure methane emissions nearer the poles when daylight is limited (Jacob et al. 2016).

The next decade of planned imaging satellites (ex. EnMap, PRISMA) are being developed to sample in more than one SWIR range. Hyperspectral, or even ultraspectral, sampling of the entire SWIR range combined with thermal infrared, using both solar backscatter and thermal emissions methods, will increase the accuracy and improve the precision of point source estimation by increasing the independent measurements within the given spectral window and allowing for better resolution of absorption features (Cusworth et al. 2019; Nalli et al. 2020). As most SWIR measurements rely on the quantification of solar backscatter they are greatly affected by cloud coverage and are sensitive to interfering reflectance from features such as bodies of water (Jacob et al. 2016).

Table 2. Examples of instruments for measuring carbon dioxide and methane using the satellite methods depicted in Figure 5^a.

Method	Instrument	Gas	Discussion	Resolution^b Spectral (nm ^c) ^d Spatial (km ²)	Main Reference
SWIR: Atmospheric Sensor	SCIAMACHY	CO ₂ ,	Large spatial coverage and achieved first global mapping. (2002 - 2012)	0.4, 1.4, 0.2	(Buchwitz et al. 2005)
		CH ₄		30 x 60	
	GOSAT	CO ₂ ,	Fine spectral resolution, measures select global locations. (2009 - present)	0.02, 0.06, 0.10	(Kataoka et al. 2017)
		CH ₄		10 x 10	
	GHGSat	CO ₂ ,	Constructed with a very fine pixel resolution to measure facility scale emissions. Claire (2016 - present)	0.1	(Jervis et al. 2021)
		CH ₄		0.05 x 0.05	
TROPOMI	CH ₄	Can produce daily estimates of surface emissions of methane. (2017 - present)	0.25 7 x 7	(Hu et al. 2018)	
OCO - 2	CO ₂	Focuses on better quantification of the carbon cycle at regional scales. (2014 - present)	0.04, 0.08, 0.10 1.29 x 2.25	(Reuter et al. 2017)	
SWIR: Imaging Spectrometer	PRISMA	CO ₂ ,	The first two main satellites launched using hyperspectral sampling. Attains fine pixel resolution. (2019 - present)	10	(Cusworth et al. 2019)
	EnMap	CH ₄		0.03 x 0.03	
TIR	TES	CH ₄	Achieved the smallest pixel size TIR instrument with fine precision (2004-2011).	0.8 5 x 8	(Worden et al. 2012)
	IMG	CO ₂ ,	Provided the first satellite sensing of methane (1996 - 1997)	0.1 cm ⁻¹	(Shimoda and Ogawa 2000)
		CH ₄		8 x 8	
CrIS	CO ₂ ,	Hyperspectral method with the best vertical resolution through the troposphere. (2011 - present)	0.625 cm ⁻¹	(Nalli et al. 2020)	
CH ₄	14 x 14				
Lidar	MERLIN	CH ₄	Intended to measure in dark conditions at a fine resolution. (Not launched yet.)	N/A 200 μm	(Wührer et al. 2019)

^a Full names of abbreviations: SCanning Imaging Absorption spectroMeter for Atmospheric Cartography (SCIAMACHY), Greenhouse gases Observing SATellite (GOSAT), Greenhouse Gas Satellite (GHGSat), TROPOspheric Monitoring Instrument (TROPOMI), Orbiting Carbon Observatory-2 (OCO-2), PRecursive IperSpettrale della Missione Applicativa (PRISMA), The Environmental Mapping and Analysis Program (EnMAP), Technology Experiment Satellite (TES), Interferometric Monitor for Greenhouse gases (IMG), Cross-track Infrared Sounder (CrIS), Methane Remote Sensing Lidar Mission (Merlin). ^b When the instrument measures both CO₂ and CH₄ at varying spectral resolutions, resolution is given as band1, band2, and band3. ^cExcept where otherwise noted. ^dGiven as the Full Width at Half Maximum (FWHM).

As methods improve, the trade-offs between spectral and pixel resolution, and global sampling versus point source attribution are becoming more relaxed (Cusworth et al. 2020). OCO-2 was built with the specific intention of better quantification of the global carbon cycle, but measurements have even been used to estimate power plant emissions that agree within 1-17% of reported daily values (Nassar et al. 2017). Over the last decade several measurements have been made by satellites to estimate point-source emissions with increasing pixel and spectral resolution (Varon et al. 2019; Kort et al. 2014; Nassar et al. 2017; Varon et al. 2020; Jervis et al. 2021). In the fall of 2020 using a controlled release, GHGSat (Iris) accurately measured the smallest methane emission plume from space which was also validated by a simultaneous aircraft measurement (260 kg hr^{-1}) (GHGSat 2020). Satellite technology will provide the remarkable capability to produce independent point source GHG measurements from space of international sites that may lack transparency of emissions (Nassar et al. 2017). At the time of writing, geostationary methods have not been launched, but they will allow for continual monitoring of a chosen region, rather having a predictable overpass time for each orbit (Sheng et al. 2018). The refinement of satellites in the next decade lends itself to the ability for temporally and spatially continuous point source emission quantification and attribution. Continuous monitoring and time-averaged satellite emissions estimates will be a unique tool for estimating annual emissions from facilities for monitoring and reporting, and validation (Varon et al. 2020).

1.3 Conclusion

This review has summarized ground-based, airborne, and satellite methods of estimating point and area source anthropogenic carbon dioxide and methane emissions. Methods of measuring anthropogenic GHGs can be used to complement each other to fill measurement gaps, clarify assumptions, and reduce uncertainty in overall regional inventories.

Bottom-up, ground-based methods are ideal over small spatial scales with small emissions. These methods are often the most cost effective and accessible of emissions monitoring methods, provide essential data for temporal interpolation of top-down methods, but can be difficult to extrapolate to regional emissions. Chamber-based methods have been commonly used by industry to estimate from the smallest spatial scale and produce emission inventory estimates. They are starting to be replaced by proximal imaging methods that provide leak-detection capabilities as well as instantaneous variable measurement and plume

characterization (Fox et al. 2019). Towers using micrometeorological methods such as eddy covariance techniques and inverse dispersion modeling can provide continuous regional measurements of emissions to assess seasonal variability. Ground-based methods are ideal for targeting near ground plumes and assessing emissions from small facility components.

Top-down methods provide vital atmospheric carbon dioxide and methane budget checks as industry inventories have been found to under-report atmospheric anthropogenic GHG emissions (Brandt et al. 2014). Airborne methods are an integral source of validation data between top-down and bottom-up estimates. Airborne mass-balance methods provide detailed point and area source estimates with a low uncertainty under ideal conditions. Remote spectral imaging can quickly map an area to characterize emissions and avoid the temporal issues inherent to mass-balance methods. As remote spectral imaging methods advance, and error is reduced, this method of sampling heralds the possibility for quick, exhaustive regional sampling to identify both known and unknown sources. Annual estimates are often derived from limited airborne and ground-based methods to compare to industry estimates, and the methods for extrapolating from a limited number of samples are still being developed. Satellite data can be an ideal source for correlating sporadic measurements to an upscaled yearly extrapolation and have been proven as an essential tool to highlight larger emitters for follow up in-depth assessment through airborne campaigns (Kort et al. 2014). Future satellite instruments will approach airborne sampling precision and accuracy for source attribution and estimation of emission plumes (GHGSat 2020).

International climate agreements have a history of poor implementation as nations sign and then back out as governmental priorities change (Maslin 2020). The development and continual refinement of a global GHG monitoring system, with point source emission attribution estimation, will increase global pressure on nations when emissions targets are not being met. Increased transparency using satellite monitoring will improve certainty, and address under-estimation of emissions inventories whilst holding industry more accountable to their emissions. As anthropogenic methane emissions have a higher radiative greenhouse gas effect than carbon dioxide, and a short atmospheric lifetime, decreasing atmospheric methane emissions is an effective route towards combating global warming (Saunio et al. 2016; Balcombe et al. 2018). Reducing the uncertainty in quantifying atmospheric anthropogenic methane emissions is an

essential step towards producing the most effective policy for reducing climate change. Advances in obtaining regular coverage using spectral airborne and satellite sampling methods will make them ideal for quantifying process emissions and sporadic flaring events to monitor changes and capture leaks, enabling more accurate and complete estimates of carbon dioxide and methane fluxes.

Chapter 2

2.1 Introduction

Accurate quantification of GHG emissions, at a range of spatial scales, is an essential foundation for emissions reductions. Regional, national, and global CH₄ and CO₂ emissions are estimated using a combination of bottom-up and top-down methods. In general, bottom-up methods aggregate component, or site specific data, and extrapolate to estimating 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 2018). Anthropogenic carbon dioxide and methane emissions are estimated using a) ground-based methods, b) airborne methods, or c) satellite methods (e.g., Conley et al., 2017; Frankenberg et al., 2016; National Academies of Sciences, Engineering, 2018). 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 greenhouse gas (GHG) emissions (Allen 2014; Le Quéré et al. 2018; Saunio et al. 2020; Kort et al. 2014; Dlugokencky et al. 2011; Liggio et al. 2019; Johnson et al. 2017; Nisbet and Weiss 2010; National Academies of Sciences, Engineering 2018). Improved emission estimates facilitate the best climate change policy, allowing us to adopt pathways for lower global warming increases (Le Quéré et al. 2018; Tian et al. 2016). This thesis focuses on airborne approaches as they are intermediate in spatial scale between proximal and satellite sampling methods, and are essential for the validation and synthesis of top-down to bottom-up estimates (e.g., Conley et al., 2017; Cusworth et al., 2020).

Methods for sampling atmospheric fluxes from the air tend to fall into two major categories: i) spectral imaging methods and ii) mass-balance methods. These methods capture emissions using varying approaches that are affected by different biases and are complementary when creating emission budgets. Remote spectral imaging methods fly above potential sources and use absorption spectroscopy, or thermal emissions, to capture regional or facility emissions (Frankenberg et al., 2016). Spectral imaging can identify unknown sources, providing a unique, invaluable tool for monitoring sites for leaks of emissions, which can lead to cost savings for companies and reductions in global emissions if appropriate action is taken (Frankenberg et al., 2016). 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 transects creating a vertical “screen” to catch the plume at various altitudes, or flying a “box-flight” using multiple (over 20) laps around a facility, or facility component, to constrain a plume (Conley et al. 2017). Currently, mass-balance box flight methods can attain a lower uncertainty from a single sample of emission estimates ($\sim 2\%$) than the spectral methods ($< 30\%$), due to smaller background and wind measurement uncertainties, but require understanding of plume sources to know where to fly, take longer, and are therefore more costly (Conley et al. 2017; Thorpe et al. 2020; Duren et al. 2019).

Airborne spectral methods can be considered as top-down methods that produce results similar to satellite data, but with higher precision and accuracy (Frankenberg et al. 2016; Kort et al. 2014). Repeated sampling is needed to infer regional emission budgets; however single flights are used as a sample to estimate emissions from sources. 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 issues inherent to airborne mass-balance methods, which have lengthy sampling time of one to nearly three hours. However, these methods are dependent on the uncertainty and accuracy of the nearest wind measurements available at the time of sampling (Jongaramrungruang et al., 2019). Remote spectral imaging methods are being advanced by the NASA Jet-Propulsion Lab 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) (Duren et al., 2019; Frankenberg et al., 2016; Jongaramrungruang et al., 2019; Thorpe et al., 2020). While routine sampling is essential for reliable monitoring, flight data is currently collected on a “campaign” basis, which cannot be easily used for routine monitoring (Duren et al., 2019; Thorpe et al., 2020).

For mass-balance flights, the location of emission sources must be known. 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 (Conley et al. 2017; Gordon et al. 2015). The intention is to fully capture an emission plume within the ‘box’ created from the flight laps

and estimate the emission flux across the box (Gordon et al. 2015; Conley et al. 2017). Industry is often notified before sampling, so that site access can be coordinated. 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, yet has not dispersed, to capture the plume inside the box (Conley et al. 2017). Extrapolation to the ground is often the largest error source, nearing ~30% when the bottom of the plume is not captured (Conley et al. 2017; Gordon et al. 2015). Mass-balance airborne methods depend on the assumption that the emission plume is captured at the top of the box and does not change during sampling (i.e., that conditions are stationary). A moving boundary layer becomes a significant challenge when sampling and can directly affect the extent of plume capture and stationarity. Failure to capture the top of the plume can occur for two main reasons: sampling did not occur all the way to the top of the flight path, or the boundary layer was unstable, and emissions were not contained. Extrapolation to the top of the boundary layer can address this problem and avoid the inherent estimate bias when the plume is not fully captured. Unlike surface extrapolation, upward extrapolation is only occasionally applied. Upward extrapolation is appropriate when the emission plume extends beyond the top of the box and when the plume behaviour and height can be reasonably approximated.

Methods applying mass-balance equations to aircraft measurements have been developed and refined over the last two decades (Alfieri et al., 2010; Kalthoff et al., 2002; Karion et al., 2013). Recently, two box-flight mass-balance sampling methods, a Top-down Emission Rate Retrieval Algorithm (TERRA) developed by Environment and Climate Change Canada (ECCC) and a Gaussian theorem algorithm (SciAv) derived by Scientific Aviation, have been developed and provide two approaches to evaluate mass sampling from aircraft (Gordon et al. 2015; Conley et al. 2017).

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 modeling approaches (Table 3). 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 divergence, or flux density, from the box by evaluating a mass balance equation derived from Gauss’s Theorem for relating flux through a closed surface, to a flux density from a volume integral (Conley et al. 2017). The SciAv and TERRA methods are described in further detail in Appendix A.

Table 3. Characteristics of the mass-balance SciAv and TERRA models 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	<p>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 divergence due to an emission source within the box.</p> <p>Second step: Bin lap divergence estimates by altitude ranges, estimate an average divergence for each bin, then the integrate the bins across the total flight height to produce total emission rate estimates.</p>	<p>First step: Applies simple kriging to interpolate flight lap measurements to a spatially resolved screen.</p> <p>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.</p>
Surface Extrapolation	Extends lowest bin average 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.

While both TERRA and the SciAv model collect data using the same box-flight processes, results from the two algorithms have not previously been compared. Coincident sampling using the SciAv and AVIRIS-NG methods at facilities indicates the methods tend to agree, within model errors, when conditions such as the wind speed are similar (Duren et al., 2019; Frankenberg et al., 2016; Thorpe et al., 2020). If model comparisons indicate agreement, then emission estimates from multiple campaigns using mass-balance and spectral imaging can be aggregated, which will improve the certainty in top-down GHG budgets.

Reducing GHG emissions in Canada has become a National and Provincial priority (Canada 2016; Johnson and Tyner 2020). Under the Specified Gas Emitters Regulation, Alberta committed to 45% reductions in CH₄ emissions by 2025, and has set GHG emissions reduction targets for major GHG emitters while imposing per-tonne fees for those who do not comply (Government of Alberta 2015). 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; Johnson et al. 2017; Thorpe et al. 2020; Chan et al. 2020; MacKay et al. 2021). 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 JOSM 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, to use instrumented high-performance aircraft 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, a campaign was completed 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). We had three algorithm comparison research objectives. The first objective was to produce and compare emissions estimates following standard model protocols and then assess model sensitivity using a varying surface extrapolation scheme. The second objective was to assess the importance of fundamental model assumptions. Our third objective was to examine the

potential for validation of mass-balance emission estimates using complementary AVIRIS-NG spectral imaging data. In addressing these objectives, our intention was to assess the comparability of emissions estimates from 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.

2 Methodology

2.2.1 Algorithm Comparison Research Objectives

The approach to achieving the research objectives of this study is summarized in the outline below:

Objective 1: Compare emission estimates between methods

Using 5 sample flights, determine if there are substantial differences in emissions estimates between the two box-flight mass balance methods.

- a. Evaluate overall final emission estimates using the standard methods of the two models given the same data set.
- b. Systematically test the effects of different surface extrapolation methods for each model.
- c. Assess SciAv methods used to integrate divergence and characterize uncertainties.

Objective 2: Investigate model assumptions

Using flights from the comparison analysis, investigate implications of violating the model assumptions of:

- a. a stationary plume and stable meteorological conditions.
- b. full capture of the plume.

Objective 3: Examine potential for validation of mass-balance emissions estimate methodology through comparison to the spectral AVIRIS-NG method

2.2.2 Campaign Data Used in Comparison Analysis

2.2.2.1 AEP - NOAA - Scientific Aviation Campaign Data

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 include multiple components, 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 components. The data were collected and processed by Scientific Aviation (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 carbon dioxide and methane were measured using a cavity ring down spectrometer (Picarro 2301f, Picarro Inc., Sunnyvale, CA, USA) in its precision mode at 1 Hz as described by Crosson (2008). Other variables used in the analysis were measured using airplane primary flight information system and GPS, including wind speed components (m s^{-1}), pressure (mb), temperature (K), heading (deg), altitude (m), and latitude and longitude.

Five flights from three facilities were selected for the algorithm comparison, and are summarized in Table 4, with sample codes (F01 to F05) assigned for the purpose of the comparisons. The three facilities from the Athabasca Oil Sands Region included for the study were: Mildred Lake and Aurora North Plant Sites (Syncrude), and Horizon Oil Sands Processing Plant and Mine (CNRL), and Suncor Energy Inc. Oil Sands (Suncor). Flight paths around the facilities are shown in Figure 6, with a colour gradient depicting measured background mixing ratios of CH_4 and CO_2 in blue and enhancements from emission plumes in red. The five flights were chosen for the comparative analysis due to their varying profile shapes, type of flight, time of collection, and number of laps. Four of the five flights were selected as ideal samples due to their having ‘fully captured’ the top of the plume. One flight, F04 was chosen as a poor-quality sample, rejected during QA/QC as having ‘not fully captured’ the top of the emission plume, and was used to assess how the methods compare when fundamental assumptions of the method are not met. Boundary layer height was estimated by Scientific Aviation by assessing profile changes in potential temperature gradients before and after flights. All five flights were judged to have consistent, stationary winds and stable boundary layers through QA/QC done by Scientific Aviation and AEP. F02 had normal operating conditions and no flaring events reported by

CNRL Horizon. Operation conditions for F01, F03, F04, and F05 were not shared at the time of writing.

Table 4. Information on the five 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	2018-04-24	2018-07-19	2018-04-19	2017-08-14	2018-09-06
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 6. CH₄ (left) and CO₂ (right) mixing ratios for every one second are shown for the flight path of each facility sample used in the study. Perimeter flights are the large polygons, and plant flights the smaller ovals. Mixing ratios are shown in a gradient from background concentrations in blue to larger concentrations in yellow and red indicating the location of an emission plume. The colour scale is unique to each flight’s range of measured mixing ratios (see Appendix B). The highest laps are most visible as data are drawn on top of each other. Map layer data © 2017 Google.

Figure 7 shows an example of a close view of a flight path for F04 CH₄. A GoogleEarth historical image from 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. 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 Figure 7. There is also 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.

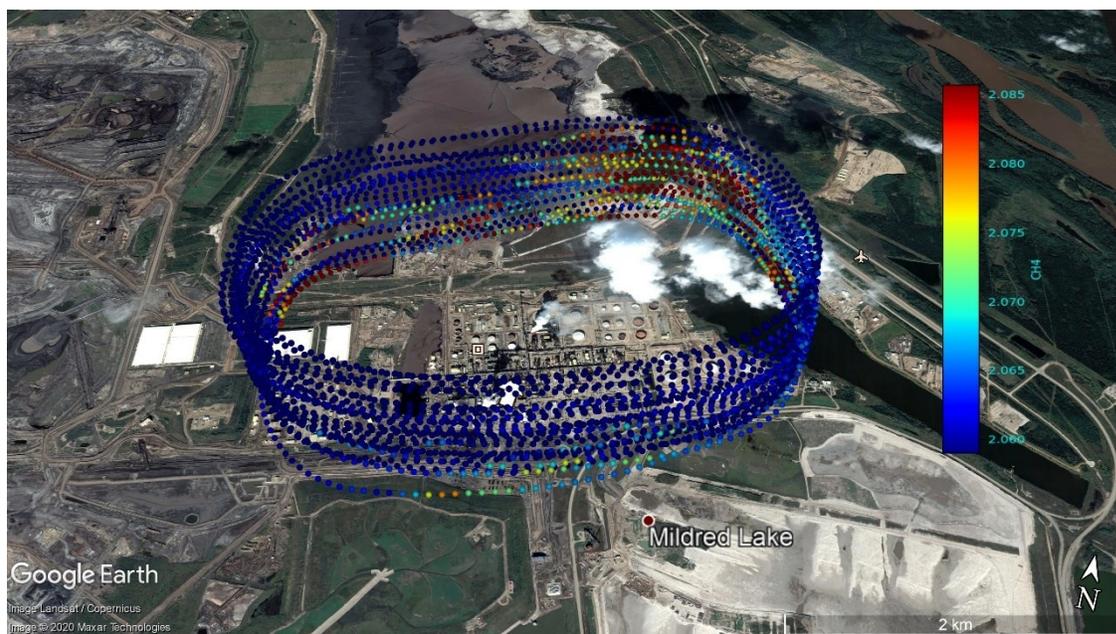


Figure 7. The flight path for F04 depicts the mixing ratio of CH₄ (ppm) measured at 1 Hz intervals for each of the 25 laps around the Syncrude plant.

2.2.2.2 NASA - JPL Campaign Data

Three days prior to the F04 flight a NASA – Jet Propulsion Laboratory (JPL) AVIRIS-NG flight sampled the exact same CH₄ emission source. This sample was part of a larger Arctic-Boreal Vulnerability Experiment (ABOVE) which included flight lines flown over the Alberta Oil Sands region. The AVIRIS-NG measures ground-reflected solar radiation (380 – 2500 nm) with a 34° field of view, and a spectral resolution of 5 nm to estimate emission plumes from

absorption features (Thorpe et al. 2020; Duren et al. 2019). Figure 8 shows the AVIRIS-NG plume data that was captured over the Syncrude plant site with the Scientific Aviation KML lap data shown in Figure 7 overlaid. By measuring a plume at the same location within a few days, this independent sample using a different method provides a contrast to the box-flight, mass-balance data.

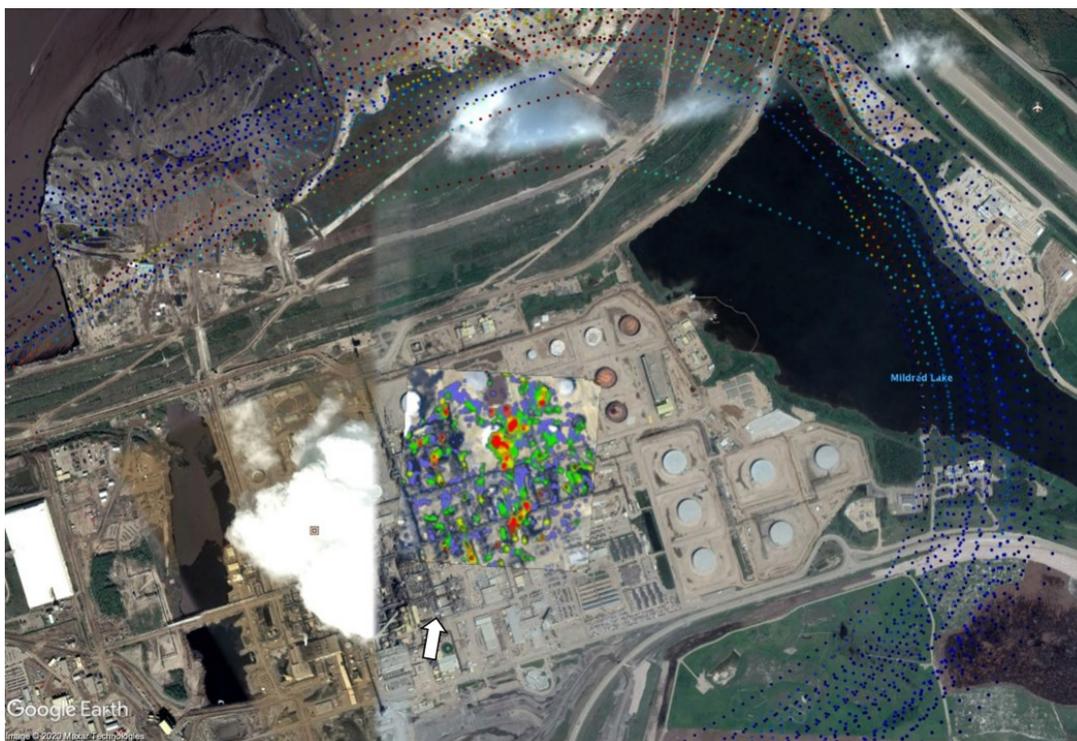


Figure 8. AVIRIS-NG captured CH₄ column enhancements are shown inside the F04 raw CH₄ lap data around the Syncrude plant. AVIRIS-NG data and imagery provided by the NASA-JPL and satellite imagery © 2020 Maxar Technologies, Google Earth. Large CH₄ enhancements are depicted in red. The processed plume imagery is overlaid on a different day causing the ‘cut cloud’. The wind direction is shown by the white arrow.

2.2.3 The Use and Adaptation of the Two Mass-balance Algorithms

2.2.3.1 Use of the TERRA Algorithm

Environment and Climate Change Canada (ECCC) provided the TERRA algorithm and detailed instructions on how to produce estimates using the algorithm with commercial plotting software (IGOR Pro 8, Wavemetrics, Lake Oswego, OR, USA). ECCC performed the first step of the TERRA algorithm, which provided 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). The second step was completed as part of this study.

TERRA has options for extrapolating emission concentrations from the lowest flight layer to the surface to account for fluxes below the flight path. The TERRA surface extrapolation fit options are *background*, *constant*, *linear*, *linear interpolation to background*, and *exponential*. A background extrapolation is used when there is a fully captured, elevated plume and there is a benefit to choosing one single mixing ratio extrapolated to the surface. A constant extrapolation is best in the general case of a fully captured, elevated plume as it avoids the assumption of a background value. The linear extrapolation is preferred in the scenarios when emissions occur towards the surface such as a low plume that was not fully captured, or a mixed plume with ground sources such as a tailings pond. The interpolate fit is ideal when there is evidence of decreasing emissions with only a trace of plume at the bottom of the flight path. Exponential extrapolation is largely avoided unless there is a strong argument that it best fits the plume behaviour as it can be unstable.

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 CH₄ and CO₂ due their low position on the screen and likelihood of having an increasing emission towards the surface. An interpolate to background fit was applied to F02 for both CH₄ and CO₂, and F04 for CO₂ as these cases largely captured plumes with low mixing ratios at the bottom of the flight path. A constant extrapolation was fit to F04 for CH₄ 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 (Appendix B).

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. ECCC often flies separate samples to independently measure background concentrations, then removes background concentrations prior to running TERRA (Baray et al. 2018). As these measurements were not collected, 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 (Appendix B). Extrapolations are fit from the mixing ratio screen data at a height of 300 m. All extrapolation outcomes were produced to calculate the surface extrapolation error and to compare with the range of possible outcomes from the SciAv method. The settings for the standard TERRA emissions estimate were chosen in consultation with ECCC by assessing the plume location, boundary layer conditions, and plume source information to determine the appropriate surface extrapolation (Baray et al. 2018; Gordon et al. 2015).

The TERRA model total uncertainty estimate was calculated by adding seven error terms in quadrature (i.e., by taking the square root of the sum of squares). In consultation with ECCC four of the seven TERRA error terms were evaluated. The wind and measurement error have been previously determined to each be <1%, and the vertical turbulence term has been functionally removed from analysis (Baray et al. 2018; Gordon et al. 2015). The surface extrapolation error was calculated as the maximum percent change amongst the plausible surface extrapolation estimates produced. For example, background extrapolations were not considered for standard estimates when a flight has 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 Appendix B.

2.2.3.2 Adaptation of the SciAv Method

Scientific Aviation provided results from the first step of estimating the divergence for each lap. They applied their algorithm to the flight data and provided output that could be used to address the research objectives. This output included standard emissions estimates and uncertainties using the SciAv preferred settings. It also included profiles of 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 method for applying the second step of the SciAv model was redeveloped and coded in MATLAB 2020a (The MathWorks, Inc., Natick, Massachusetts, United States) as part of this study, since Scientific Aviation did not provide their algorithm code. For this reason, the method of plotting the profile, averaging the data by the given bin, extrapolating to the surface, and summing the results all had to be recoded. To verify the accuracy of the MATLAB 2020a code, the SciAv emission estimates were reproduced from

the divergence profiles and binning information. Uncertainty estimates were not reproducible as the description of the error estimates (Conley et al. 2017) was ambiguous and clarification was not obtained within the time frame of this study (Equation 7, Appendix A).

Figure 9 provides an example profile of the average divergence per lap estimate calculated in the first step of the SciAv model. In the second step, the lap estimates were binned by altitude and averaged within each bin, as shown in red. Extrapolation to the surface is shown in green using the SciAv method of extending the lowest bin as a constant to the surface. The red and green estimates were integrated over altitude, binned estimates were multiplied by the height of each bin then summed up, to produce the methods final emission estimate. The estimated range of the atmospheric boundary layer is shown as light blue horizontal dashes.

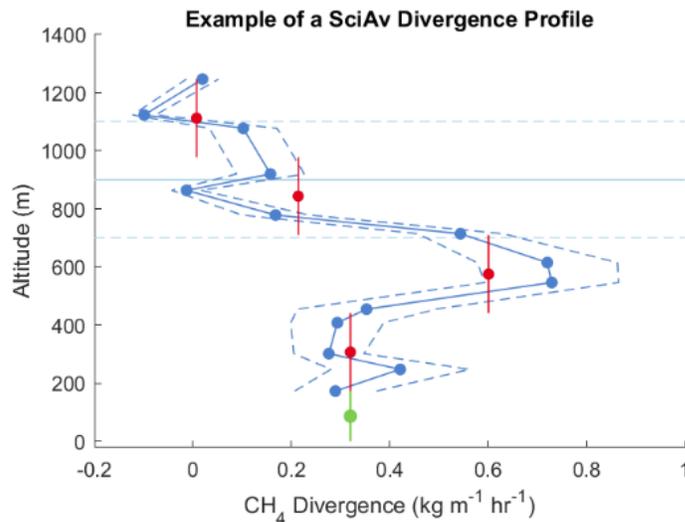


Figure 9. An example of a SciAv profile used in the second step of the method. The blue points are the estimated divergence for each lap which are connected to show profile shape with the associated error (a dashed blue line). Red points are bin averages and the vertical red bar is the bin height range. The boundary layer height is drawn in light blue with error bars (light blue dashed lines). The standard SciAv surface extrapolation method of extending the lowest red bin to the surface is shown in green.

SciAv uses a constant extrapolation of the average divergence in the lowest bin, yet this may not always be the most appropriate fit when a sample is increasing towards the surface. For this study, the use of different surface extrapolations for the SciAv model were developed and tested. Differing surface extrapolation methods of constant (SciAv’s standard), background (no extrapolation), linear, linear weighted, linear interpolation to background, and the average, were applied to obtain a greater range of possible outputs from the model. Extrapolations were fit to

the profile points below a specific height that was chosen given the profile shape, location of the plume enhancement, and sparsity of points (Appendix C). Conley et al. (2017) found that binning by lap and using a constant extrapolation produces a stable estimate when 20-25 laps are flown (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. Therefore, trapezoidal integration of the divergence points over the profile was applied in MATLAB 2020a and the various surface extrapolations were fit to this estimate. A method of producing uncertainty in the trapezoidal integration method was devised by using the uncertainty of each lap divergence point provided by Scientific Aviation.

2.2.4 Model Comparison

2.2.4.1 Objective 1: Compare Emission Estimates Between Models

a) *Evaluate overall final emission estimates using the standard methods of the two models given the same data set.*

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 estimates, whereas the extrapolation for TERRA varied by the flight profile and sources.

Model agreement was assessed by comparing the range of error for each estimate and the extent of overlap. The range in the SciAv estimate was provided by Scientific Aviation, calculated from the three uncertainty terms of the model. The range in the TERRA estimate was calculated from the total uncertainty as described in Appendix B. The difference between model estimates was calculated as: $Estimate_{SciAv} - Estimate_{TERRA}$. The relative percentage difference was calculated as the difference divided by the average of the two models:

$$Percentage_{Relative} = \frac{(Estimate_{SciAv} - Estimate_{TERRA})}{\left(\frac{Estimate_{SciAv} + Estimate_{TERRA}}{2}\right)} \times 100 \quad (1)$$

The relative percentage difference was compared to propagated percentage error of the difference which is the two models' estimated errors added in quadrature. Model agreement

occurs when the percentage differences are near, or less than the models' propagated error percentages. Statistically significant differences between model estimates for each flight and emission were first tested using a weighted t-test to include the error of each estimate. Nested model testing could not be run due to the small number of repetitions (non-singularity issues). The data are not normally distributed as estimates differ in scales between the boundary and plant flights, and have a very small sample size. Therefore, the non-parametric paired Wilcoxon signed-rank test was then used to test differences between the model outcomes (Rey and Neuhäuser 2011). Differences were chosen to be significant at a power of 0.1 given a sample size of $n=5$, when the absolute t-value is greater than 1.476 for the t-test and when the critical value (V) is equal to, or less than 3 for the Wilcoxon signed-rank test (Chernick and Friis 2003; James et al. 2013).

b) Systematically test the effects of different surface extrapolation methods for each model.

The standard SciAv emissions estimates assume that the divergence profile is constant below the lower flight altitude. However, in some cases the profile shape at the lower altitudes does not appear constant (see Sect. 2.3.1). As a new method, various surface extrapolations were applied to the SciAv divergence profiles. This approach was slightly different than for TERRA, which fits extrapolations to the screen interpolated mixing ratios. Furthermore, TERRA fits consistently from a height of 300 m, but the SciAv profile has fewer points to fit from so the height was selected to produce the most stable, sensible extrapolation fits. Along with the standard TERRA fits of constant, background, and linear, additional fits of a linear weighted, linear interpolated to background, and an average of the extrapolation were applied to the SciAv divergence profiles. An exponential extrapolation could not be fit due to the small number of points in the profile. Emissions were estimated using the SciAv binning method for the portion of the known divergence profile and then the emission estimate from the surface extrapolation was calculated as the area under the curve using trapezoidal integration and added to the binned estimate. Extrapolations were fit from the lowest bin estimate, rather than the lowest divergence point, to remain consistent with the SciAv method. The process of determining the fitting height is detailed in Appendix C.

Model estimates of emissions along the measured flight path were compared by assuming no emission plume below the lowest lap measurement to remove the effect of the differing

surface extrapolation methods. For the SciAv method, background emissions were fit by applying no extrapolation and assuming the divergence was zero below the lowest flight path. The TERRA method has a fit to background option that runs the emission estimate algorithm assessing the integrals assuming a constant background of emissions below the lowest flight track. These two approaches are not exactly equal, but give a sense of how the models compare without the added step of choosing an appropriate surface extrapolation. Comparing the two models without surface extrapolation allows for better assessment of potential systematic bias in each model. Model estimates for each surface extrapolation were also compared for the four surface extrapolations fit by both models: linear, interpolate, constant, and background. Percentage differences between model estimates for each flight were calculated as per Equation 1 and differences between models were assessed using a pairwise t-test and Wilcoxon signed-rank test.

For each flight, algorithm estimates for the varying surface extrapolations were resampled using the bootstrap method (described below) to estimate the range in the difference between model estimates given various surface extrapolations for each flight. The difference between the two models given the various surface extrapolation estimates was computed and contrasted with the standard estimate model's error. To obtain a sense of the range of estimates from each model 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 model individually and the difference computed as:

$$difference = mean(Estimate_{SciAv}) - mean(Estimate_{TERRA}) \quad (2)$$

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. They were used to estimate testing of the null hypothesis that model estimates agree, and if the distributions do not overlap with zero then there is evidence that the models differ.

c) Assess SciAv methods used to integrate divergence and characterize uncertainties.

The flight data and associated uncertainties used for both the SciAv and TERRA models were identical, but the other error terms differed due to different approaches to assessing conditions and analysis of the data. A potential method of reducing error in the SciAv methods was

investigated by using trapezoidal integration rather than binning to estimate a total emission estimate from the lap divergence points. Figure 10 depicts the two methods of integrating the SciAv divergence profiles using the profile for F01 CH₄ as an example. The same surface extrapolation estimation procedure was used for both the binning and trapezoidal methods (Appendix C). Surface extrapolation methods were fit to the lowest divergence lap point for the trapezoidal method, and the binning method fitted from the lowest bin. The SciAv temporal error term could not be reproduced with available information, so total error for estimates were not derived.

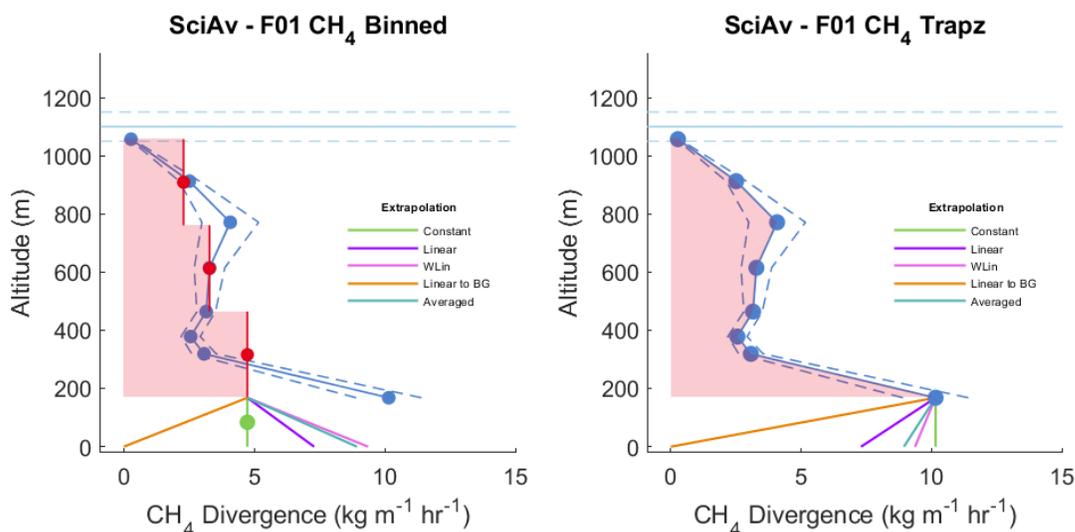


Figure 10. The two different integration method applied to the SciAv F01 CH₄ profile is depicted as the red area. The left figure shows the standard binning method of estimating an average divergence for each bin and integrating by altitude over the area as rectangular boxes. The right figure shows the trapezoidal method of estimating an average area under the curve by connecting the divergence lap points. Both figures do not include the extra emission that would be included using surface extrapolation.

The percentage relative difference between integration methods given estimates using the various surface extrapolation was calculated using Equation 1 as $SciAv_{Binned}$ minus $SciAv_{Trapezoid}$. The standard deviation in the percentage relative difference between integration method estimates was calculated and compared to the percentage relative standard deviation (%RSD) of the emission estimate for each set of estimates derived using the two integration methods. The mean estimate was derived for each method and tested using a pairwise t-test and Wilcoxon signed-rank test for statistically significant differences between the two sets of SciAv estimates.

2.2.4.2 Objective 2: Investigate Model Assumptions

Large, anomalous differences between the SciAv and TERRA estimates occurred for F04. This flight was used as an example for assessing the implications when model assumptions of a stationary plume and stable meteorological conditions are violated. The effects of violating the assumption of a fully captured plume were assessed using F05. Exploratory analysis of flight features was performed in RStudio and MATLAB R2020a.

a) Investigate implications of violating the model assumptions of a stationary plume and stable meteorological conditions.

Sampling of the five flights used in this analysis took, on average, an hour each. The laps collected were aggregated together by applying simple kriging to produce a screen of mixing ratios (TERRA), or calculating lap averages of the divergence and integrating over the profile of the lap estimates (SciAv). Initial QA/QC of the data used was performed by SciAv and methods involve inspecting the divergence profile over the whole flight to assess the model assumption of plume capture, but do not assess plume stationarity. Flight segments were examined for extent of capture of the plume, sufficient wind speed and consistent wind direction, and for significant upwind emissions affecting background estimates.

The SciAv data did not include the explicit error term for the extent of stationarity of the wind and therefore error terms were not able to be reproduced. The TERRA error terms were calculated to assess stationarity. Additional analysis of the meteorological conditions for F04 during the flight upward and downward was assessed in this study and is detailed in Appendix E.

A novel approach of separating lap data into upward and downward flight components was developed in this study and applied in MATLAB R2020a to evaluate plume stationarity by comparing profile shapes and variation of the emission estimate. Results from a flight with a stationary plume (F02) and non-stationary plume (F04) were compared. TERRA estimates were not reproduced for the flight upward vs downward as time did not permit ECCC to run the necessary first step of TERRA.

b) Investigate implications of violating the model assumptions of full capture of the plume.

Aircraft-based emissions estimates are only successful if the boundary layer confines the plume and measurements, or extrapolation can capture the full extent of the plume. Some flights were immediately rejected from the AEP-NOAA-Scientific Aviation Flight Campaign due to a collapsed boundary layer, as there was nothing to confine the emission plume. When emission plumes are not fully captured at the highest flight paths TERRA has the option to apply an extrapolation of the emission mixing ratios at the top of the box to the predicted boundary layer rather than discard the flight data. To highlight the implications when the top of an emission plume is not captured, the height of the box was increased by 300 m and emission mixing ratios were extrapolated upward for F05. F05 was chosen for this study as it provides an example of a flight that misses the top of the emission plume when sampling.

2.2.4.3 Objective 3: Examine Potential for Validation of Mass-balance Emissions Estimate Methodology Through Comparison to the Spectral AVIRIS-NG Method

AVIRIS-NG data were collected at 17.5 kft and industry was not informed before sampling. NASA-JPL provided CH₄ 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 the wind speed and direction, an average over a three-hour 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 compared to the SciAv estimates.

2.3 Results

2.3.1 Objective 1: Compare Emission Estimates Between Methods

a) Evaluate overall final emission estimates using the standard methods of the two models given the same data set.

The standard estimate results, following the standard protocol of each method, from both models are shown in Table 5 and Figure 11. Standard emission estimates for four of the five flights agree within their errors. The error bars for each estimate are the calculated range of each estimate not a confidence interval. The error for the SciAv estimates are those provided by Scientific Aviation. The error estimates were not reproducible as the published description of the methods is ambiguous (Conley et al. 2017). SciAv was contacted for clarification, but did not reply within the time frame of this study. The error for the TERRA estimates was calculated as part of this study and is detailed in Appendix B. In Figure 11 the error bars for each model estimate overlap with each other, aside from F04 which has a large gap between estimates. The errors for the TERRA estimates are consistently smaller than for SciAv (averaging at ~ 8 % smaller). The ranges for each estimate as well as the estimate error as a percentage are stated for comparison in Table 5. Model agreement is implied when the ranges for each estimate overlap. For all flights except F04, the differences between the models are in the range of the estimate errors. For F04 the emissions estimates disagreed as no overlap of the estimate range occurred.

Table 5. Results of the CH₄ and CO₂ standard fit estimate from each model with their error as a percentage ± % and the range derived from the percentage error of each model’s standard estimate.

Estimate Type	F01	F02	F03	F04	F05
SciAv CH ₄ (kg hr ⁻¹)	3840 ± 18%	362 ± 21%	497 ± 17%	349 ± 11%	3470 ± 29%
Range	(3150 - 4540)	(287 - 437)	(415 - 579)	(310 - 387)	(2480 - 4470)
TERRA CH ₄ (kg hr ⁻¹)	4810 ± 11%	395 ± 6%	476 ± 5%	125 ± 18%	3910 ± 15%
Range	(4310 - 5300)	(373 - 418)	(452 - 501)	(102 - 148)	(3330 - 4490)
SciAv CO ₂ (t hr ⁻¹)	104 ± 22%	563 ± 18%	526 ± 11%	1170 ± 11%	850 ± 34%
Range	(807 - 1270)	(464 - 662)	(469 - 583)	(1040 - 1300)	(561 - 1140)
TERRA CO ₂ (t hr ⁻¹)	1340 ± 10%	515 ± 6%	467 ± 4%	569 ± 7%	877 ± 26%
Range	(1200 - 1470)	(486 - 545)	(451 - 483)	(561 - 60)	(650 - 1110)

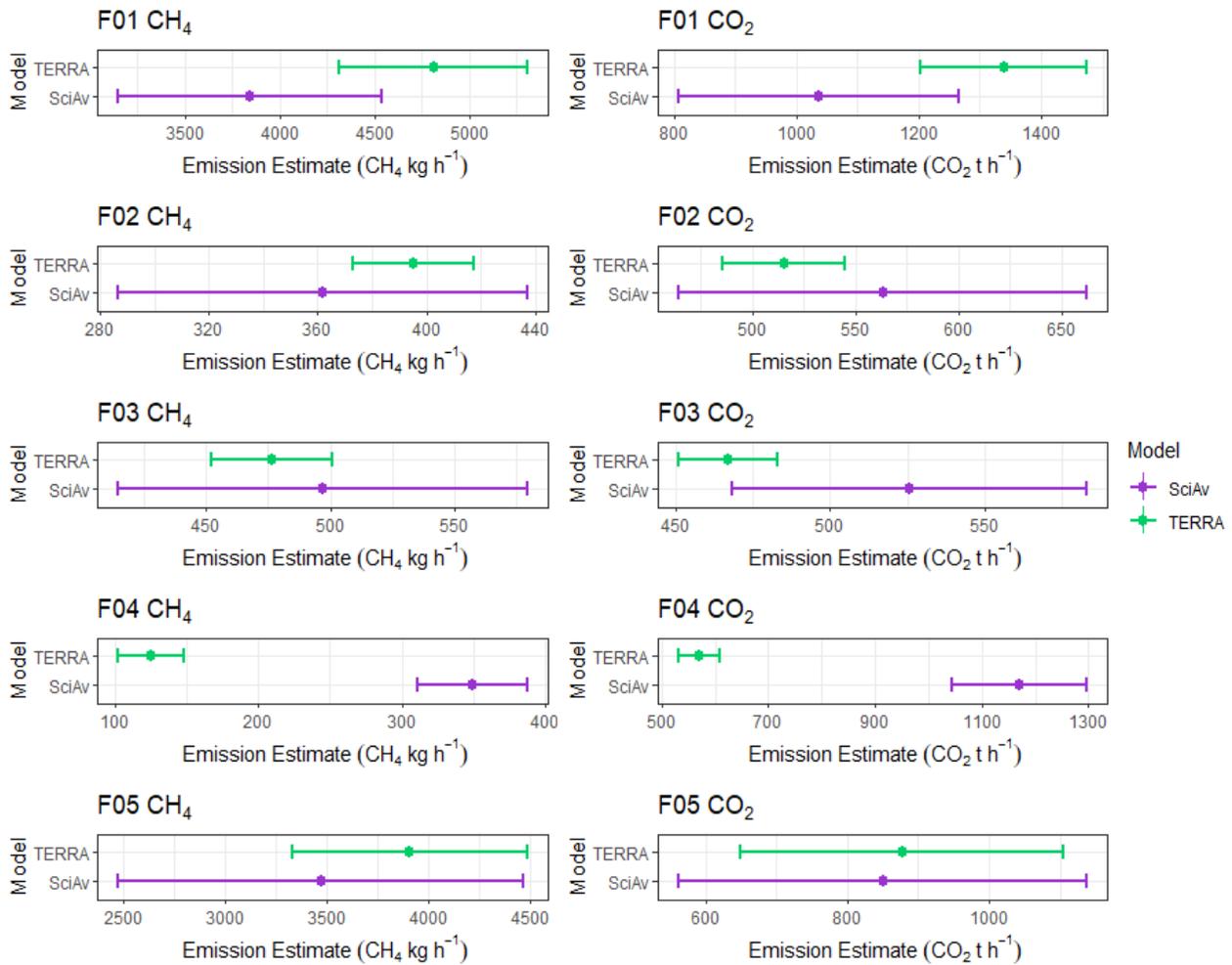


Figure 11. Flight emission estimates for CH₄ (left) and CO₂ (right) derived for each model are plotted as points along with the range of each estimate as error bars. TERRA standard estimates are shown in green and SciAv in purple. Samples of F01, F02, F03, and F05 produce estimates that coincide within each model's error bars.

The difference between estimates, relative mean percentage difference, and the propagated error in percentage for the average estimate of the two models are shown in Table 6. Confidence intervals for the model 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. To compare the model estimates using the error range the relative mean percentage difference and propagated percentage error of the two estimates were calculated. A mean percentage difference that is approximately equal to, or smaller than the propagated percentage error of the model estimates indicates model agreement within the uncertainty of the estimates. 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 model estimates were found to be insignificant for both CH₄ and CO₂ (Table 7).

Table 6. Differences between models as SciAv estimate – TERRA estimate for each flight, relative percentage differences and the propagated percentage error of the estimates.

	F01	F02	F03	F04	F05
CH ₄ : Difference (kg hr ⁻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 hr ⁻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 7. Results of parametric (weighted t-test) and non-parametric (Wilcoxon signed-rank test) significance testing of differences between the two box-flight models 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

b) Systematically test the effects of different surface extrapolation methods for each model.

To test the effect of assumptions about plume shape below the lowest flight lap various surface extrapolations were applied to lowest bin of the SciAv divergence profiles for all five flights. Plots of the lap divergence estimates tend to follow three profile types. Examples of the three profile types for measured methane lap enhancements with error shown in Figure 12 are: an emission plume with constant enhancements persisting at the lowest flight track (Type I), an elevated emission plume where enhancements approach zero at lower altitudes (Type II), and a plume that has enhancements increasing towards the surface (Type III). A flight is classified as ‘fully-capturing’ the emission plume when the mean divergence of the highest flight laps

approach zero. All the profiles shown in Figure 12 fully capture each emission plume. Of the five sample flights, three had clear profile shapes.

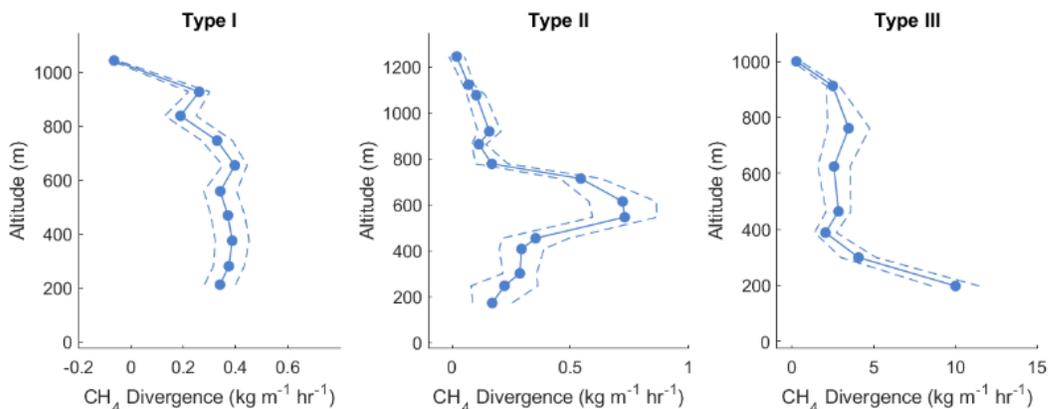


Figure 12. Sample data representing the three common types of divergence profiles for the SciAv method. Blue dots represent the divergence of each lap calculated in Equation 6 with the associated error drawn as blue dashes.

Figure 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 13A). The F02 SciAv profile follows a Type II shape (Figure 13B) 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 13C). 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 CH₄ and CO₂. The shape of the emission profile in Figure 13E for CH₄ 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 13E shows incomplete sampling for CO₂ as the divergence points are still increasing at the highest altitude indicating that the plume was not fully captured.

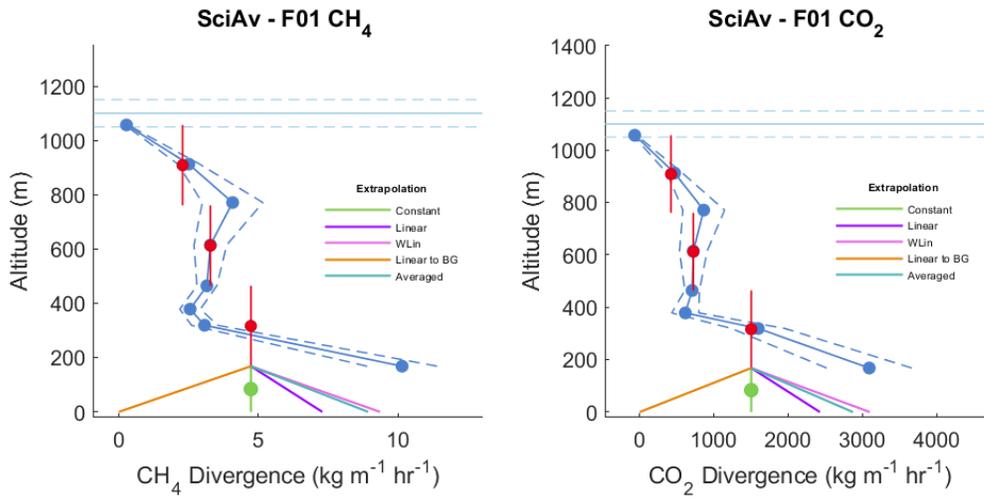


Figure 13A.

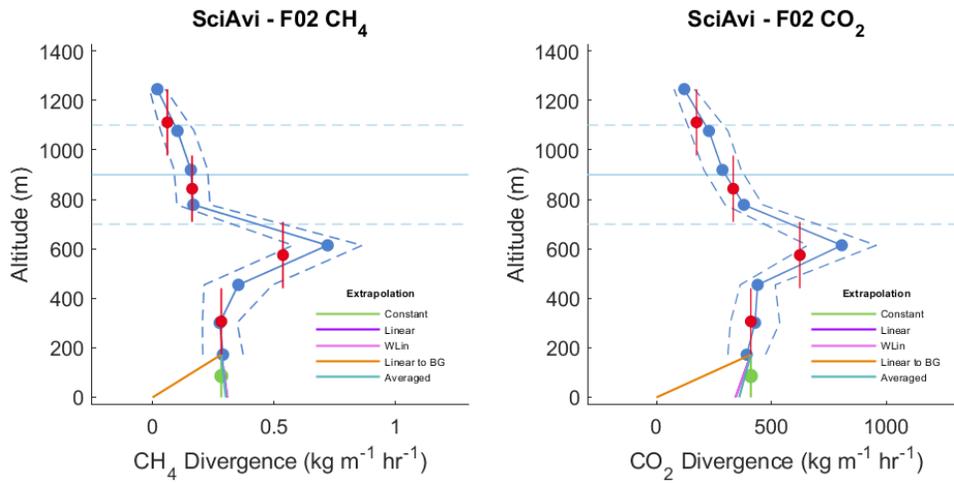


Figure 13B.

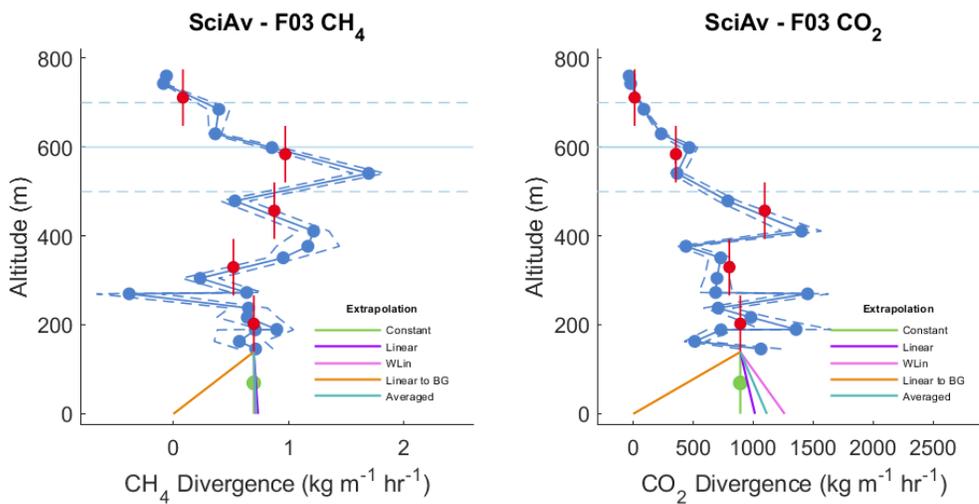


Figure 13C.

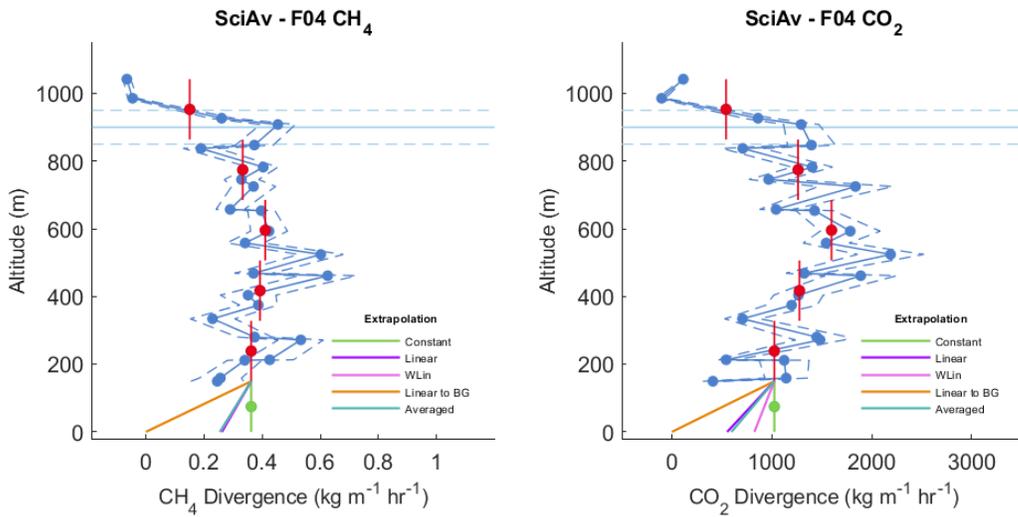


Figure 13D.

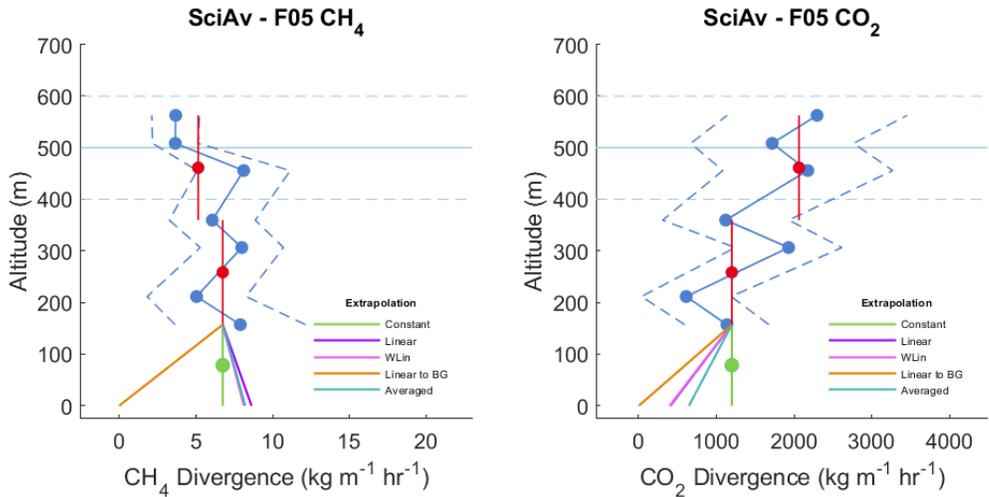


Figure 13E.

Figure 13. The SciAv profile is shown for F01-F05 in Figures 13A – 13E. The blue points are the estimated 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.

The set of all results given the differing surface extrapolations was compiled (Appendix C) and estimates plotted together in Figures 14 and 15. In Figure 15, the F01 CO₂ interpolated fits are so similar that the TERRA estimate overlaps the SciAv estimate. Estimates that clustered together for both models, such as F02 CO₂, indicate a good agreement with little difference between the varying surface extrapolation estimates. F03 CO₂ has the second most consistent gap between model estimates; however, the standard estimates still agree within their errors. F04 has large disagreement between models for both CH₄ and CO₂ (Figure 14 & 15). The mean emission estimate and standard deviation of each model's various surface extrapolations were calculated (Appendix C). 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 models as emission estimates intersect and no one model produces consistently larger, or smaller estimates for all flights.

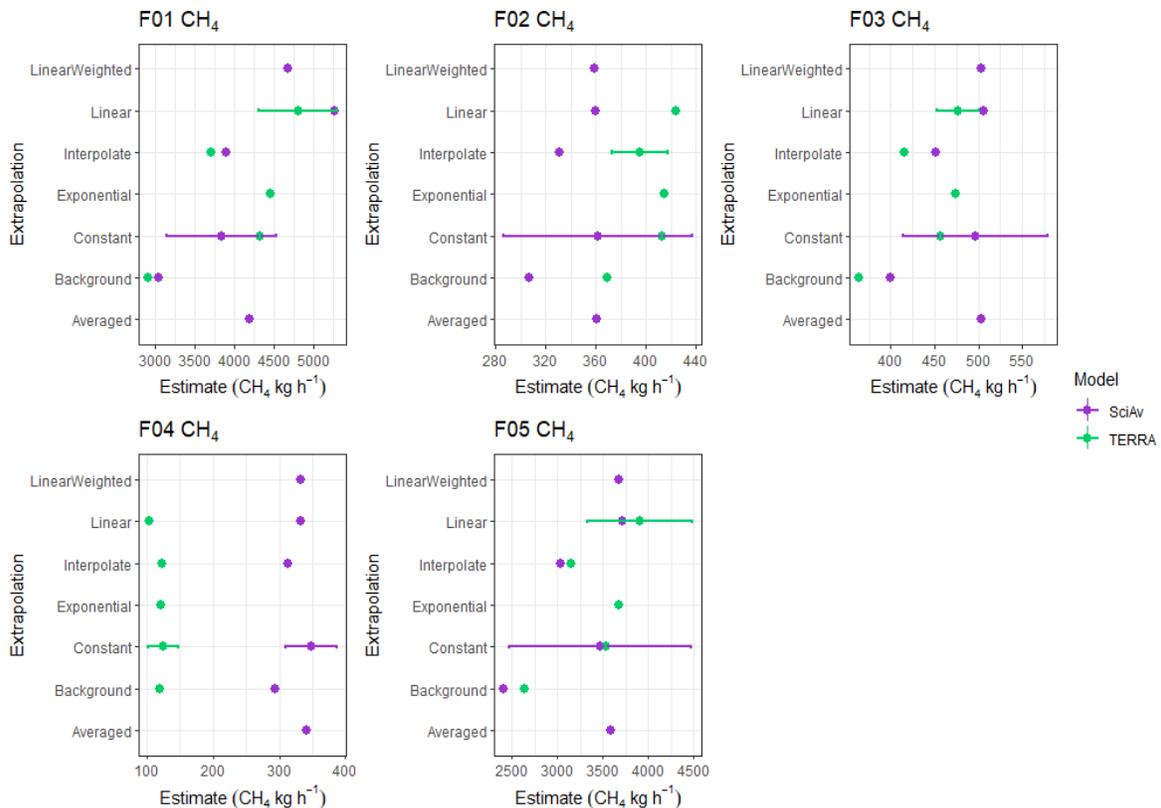


Figure 14. CH₄ estimates given various surface extrapolations are plotted in purple for the SciAv estimate and green for TERRA. Error bars are drawn onto each model's standard estimate.

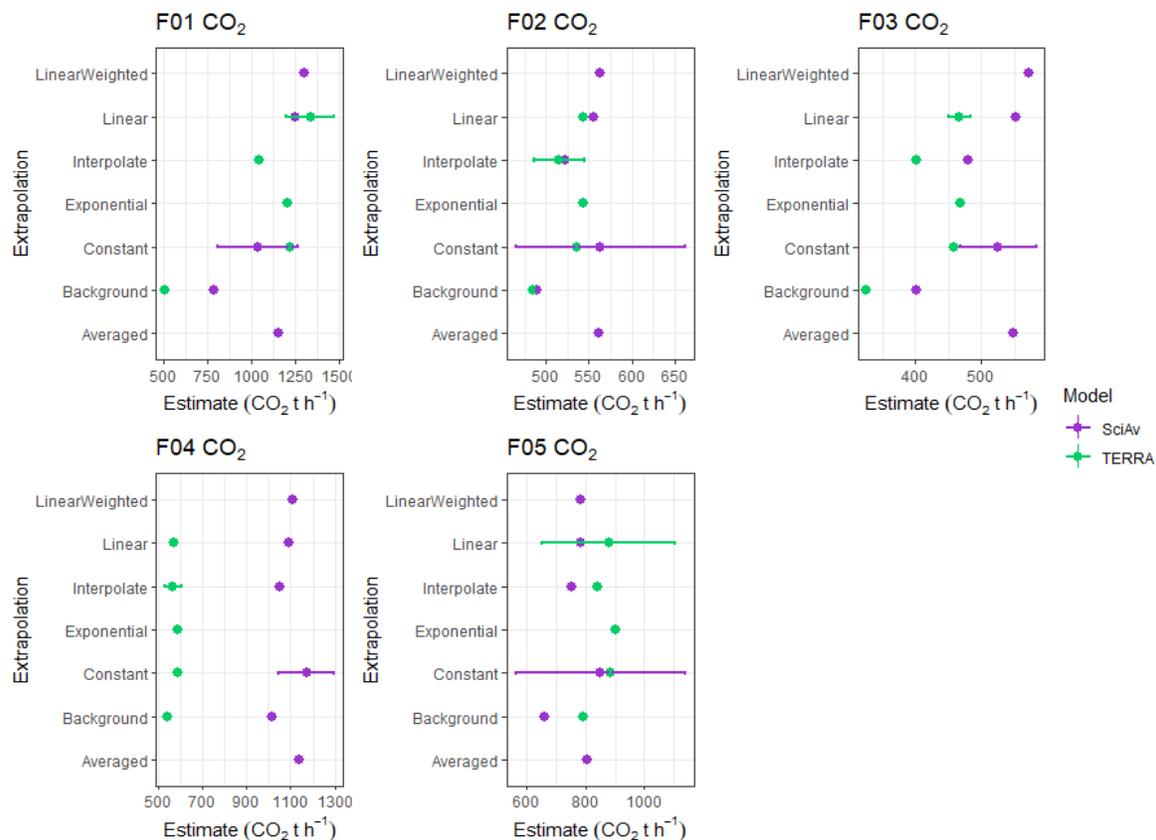


Figure 15. CO₂ estimates given various surface extrapolation fits are plotted in purple for the SciAv estimate and green for TERRA. Error bars are drawn onto each model’s standard estimate.

To remove the effect of the choice in surface extrapolation, model estimates were produced by fitting background concentrations below the lowest flight path. These estimates were compared and the SciAv and TERRA models were still found to agree. Emission estimates using a background fit were systematically lower than the standard estimates, but no clear pattern was identifiable that would indicate systematic model disagreement. For the five flights, the CH₄ background estimates agreed more than the standard estimates while the CO₂ background estimates agreed less. The average of the model estimates using the same four surface extrapolation extrapolations was also computed. To compare significance between the standard, average, and background estimate, a Wilcoxon signed-rank test and pairwise t-test were run. There is no evidence that model agreement changes when removing the effect of surface extrapolation, and that there is consistent agreement between the model estimates (Appendix D).

To assess the sensitivity of emission estimates using different surface extrapolations, the differences between each model were calculated for the four surface extrapolations used for each model (Table 8). For most flights, the choice in surface extrapolation had only a small effect on the difference between the model estimates ($\leq 3\%$, Table 8). The choice in surface extrapolation is a source of large variation between the models for F01, the one flight with large emissions at the lowest flight path. Fitting extrapolations with increasing emissions towards the surface increased model agreement for F01 CO₂ from a percentage relative difference of 16% for the standard constant fit, to 7% using the linear fit.

Table 8. Percentage relative difference of estimates between models, 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

In Figures 16 and 17, the distribution of randomly sampled mean difference was plotted along with the propagated error range (see Table 6) for each flight and gas (CH₄ and CO₂). A value of zero implies that there is no difference between the models. Aside from F04, the distributions all either include zero, or the error of the standard estimates includes zero, indicating that there is good agreement between the models in most cases. While the distribution

for F03 CO₂ just barely includes zero, the standard estimate uncertainty is small, and the models still agree within the errors (Figure 17). The extent of the differences in F04 is apparent from Figures 16 and 17. F04 is the only distribution where the error range of the mean difference of the standard estimate and the distribution of the mean differences of all the estimates are far from including zero.

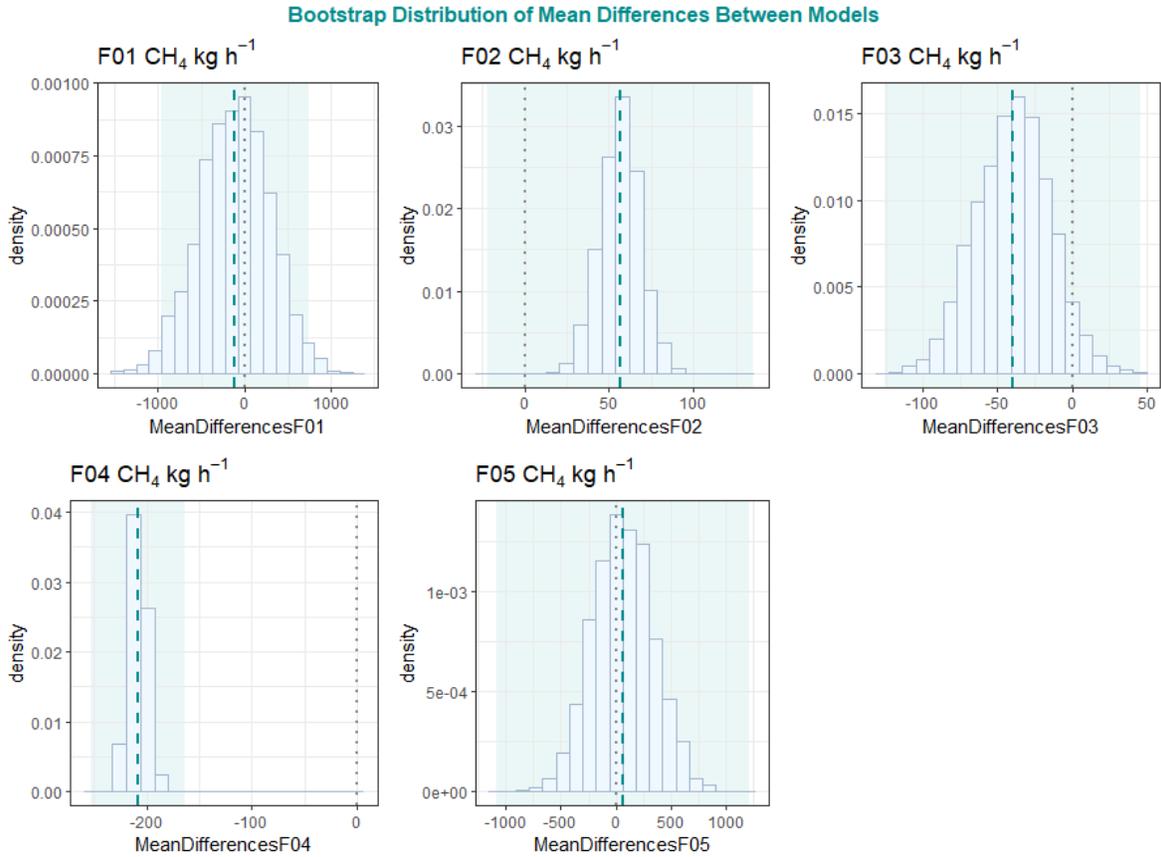


Figure 16. Distributions of the mean difference between all fits of CH₄ for the SciAv and TERRA models are shown as a light blue histogram. The mean difference between the standard estimates of the models is plotted as a teal dashed lined and the range in the difference between standard model estimates 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 models.

Bootstrap Distribution of Mean Differences Between Models

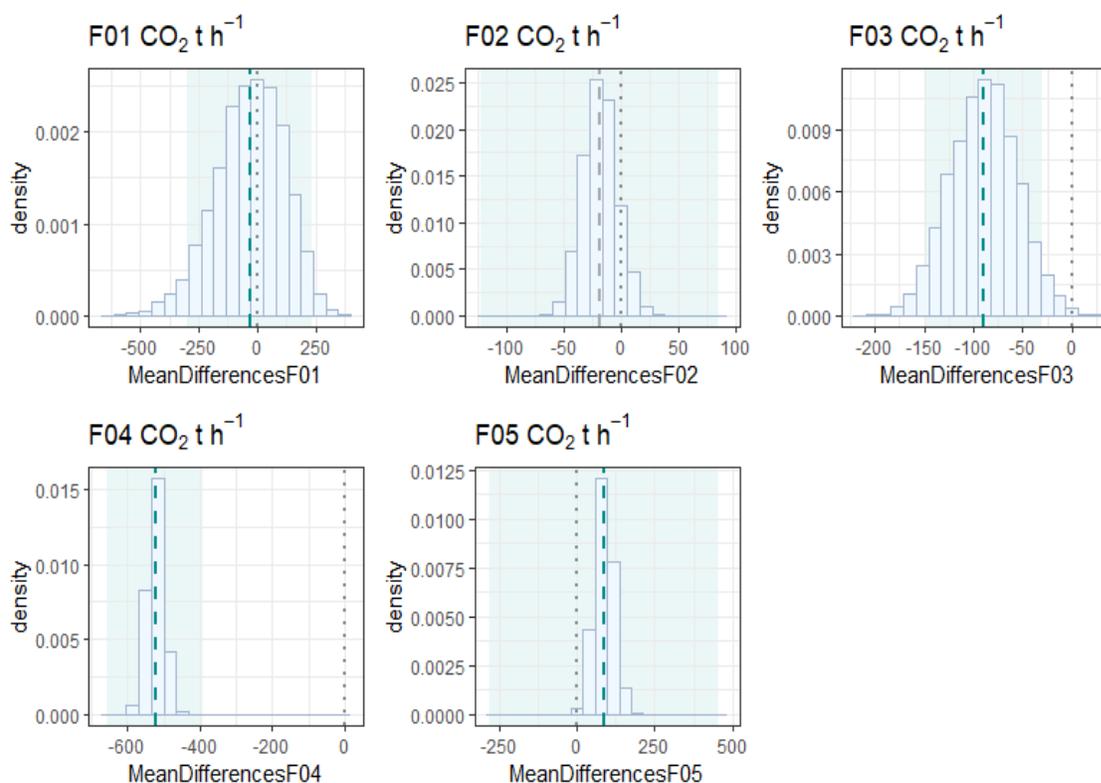


Figure 17. Distributions of the mean difference between all fits of CO₂ for the SciAv and TERRA models are shown as a light blue histogram. The mean difference between the standard estimates of the models is plotted as a teal dashed lined and the range in the difference between standard model estimates 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 models.

c) Assess SciAv methods used to integrate divergence and characterize uncertainties.

For the five flights compared, the error from TERRA is lower than that from the SciAv model by an average of 8%. Conley et al. (2017) developed a ‘flux divergence’ term which is the largest uncertainty term of the SciAv method. The largest calculated TERRA uncertainty term was the surface extrapolation. The potential of using trapezoidal integration, rather than binning by altitude, for the SciAv model was examined. The resulting emissions estimates produced by comparing the two integration methods given varying extrapolations are provided in Appendix C.

Model estimates did not significantly change when the trapezoidal integration method was used. Figures 18 and 19 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 (Figures 18 and 19). The

binning method uses the lowest bin for extrapolation, whereas the trapezoidal method uses the lowest divergence point. The mean of the estimates derived using the six different surface extrapolations were calculated for each integration methods set of results, and differences were tested using a pairwise t-test and Wilcoxon signed-rank test. Results indicate no evidence that there is a difference in mean estimates for each flight between the two integration methods (Appendix C).

Inspecting Figures 18 and 19, there is a larger variation between estimates due to the chosen surface extrapolation than integration method. This indicates that the choice in surface extrapolation is more important than the type of 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 (Figures 18 and 19).

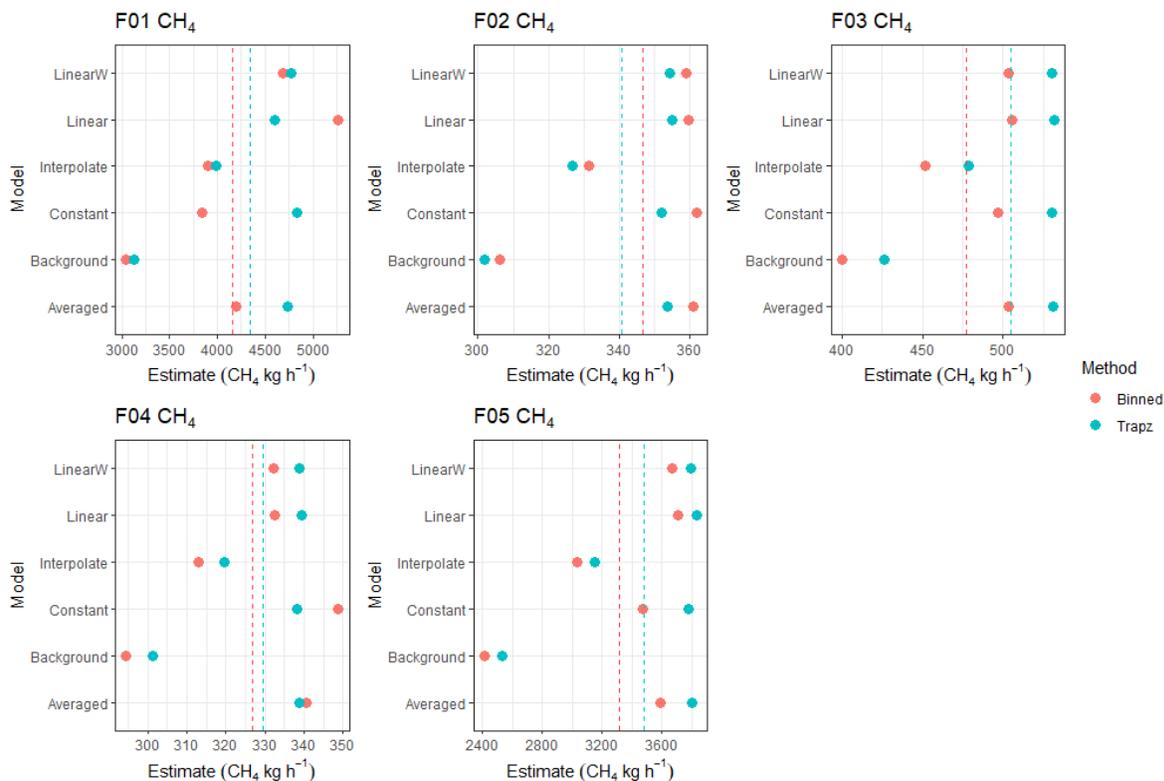


Figure 18. SciAv CH₄ (kg hr⁻¹) emission estimates derived using both a binning (red) and trapezoidal (blue) integration method. Mean values for each method are shown as a dotted line.

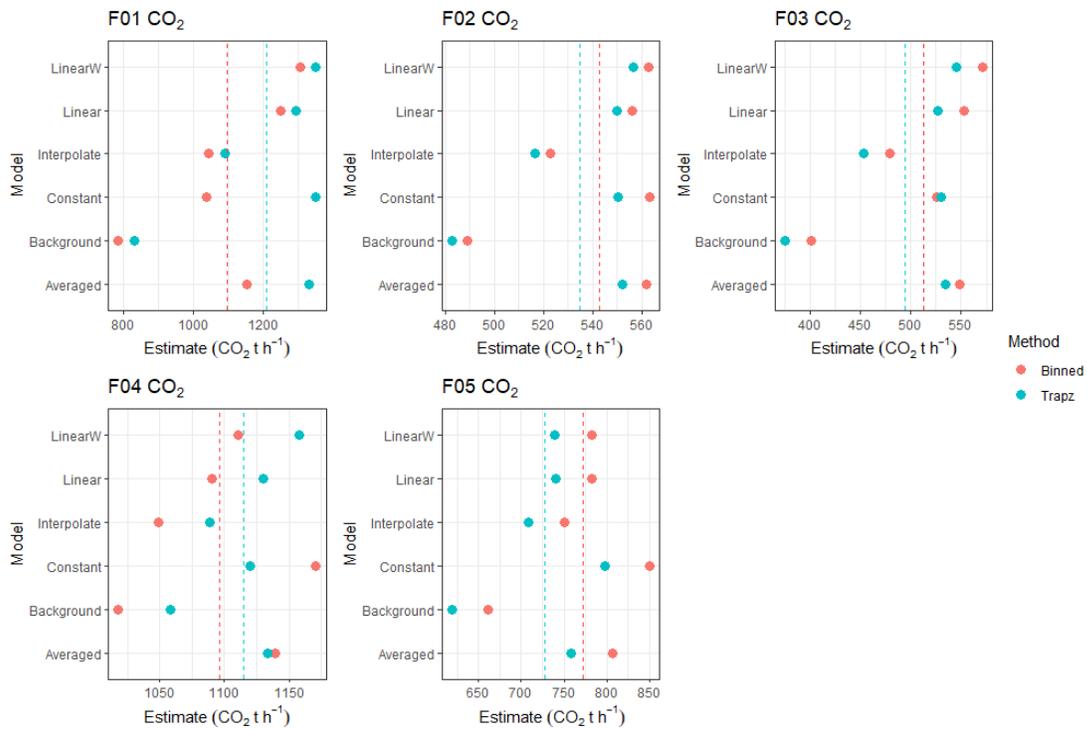


Figure 19. SciAv CO₂ (t hr⁻¹) emission estimates derived using both a binning (red) and trapezoidal (blue) integration method.

Percentage relative standard deviations (%RSD) were calculated (Table 9) 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. 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 (Appendix C, Table 25). The choice in the surface extrapolation affected the range of emission outputs more than the integration type.

Table 9. 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

2.3.2 Objective 2: Investigate Model Assumptions

a) Investigate implications of violating the model assumptions of a stationary plume and stable meteorological conditions.

To assess the effect on each method when model assumptions are violated, we investigated flights with instances of the fundamental model assumptions not being met. For F04, the fundamental assumption of a stationary plume was violated. The change in the emission plume during sampling is apparent in each model when the data was separated into the ascending and descending flight periods. In the SciAv model, emission enhancements for the flight laps going up in altitude noticeably differ from those flying laps going down. In the TERRA model, ECCC split the flight data into the upward and downward portions which were noticeably different (Appendix E). ECCC reported that when running the first step of TERRA no meteorological or other factors indicated non-stationarity. No other flights had non-stationary conditions.

Prior to the ad-hoc analysis of splitting the flight apart, the only measurement that indicated atmospheric instability was an air density error term calculated in TERRA. This produced an error estimate (4 - 6%) that was noticeably larger than the other four flights (1% - 0.01%), but not an unusually high value for the method in general (Appendix B).

The SciAv F04 profile contains two different profile shapes for the up (ascending) versus down (descending) flight (Figure 20 and 21). The up and down profiles for F02 remain similar, whereas the F04 profiles differ noticeably. The location of the red bin averages for F04 do not significantly change and the emission estimates produced from the up and down profiles remain near, or within the estimate range of the standard estimate for the whole flight (Table 10).

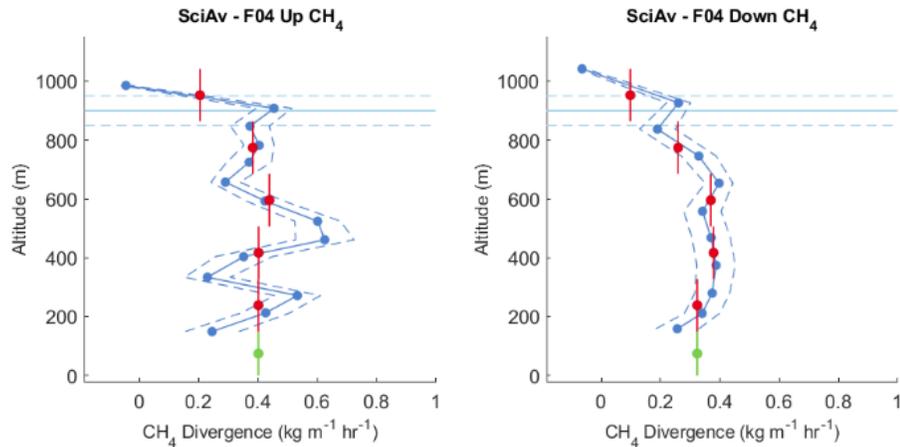


Figure 20. F04 CH₄ divergence profiles for the flight up (left) versus down (right). The profile shapes differ with a much more variable divergence in the up profile.

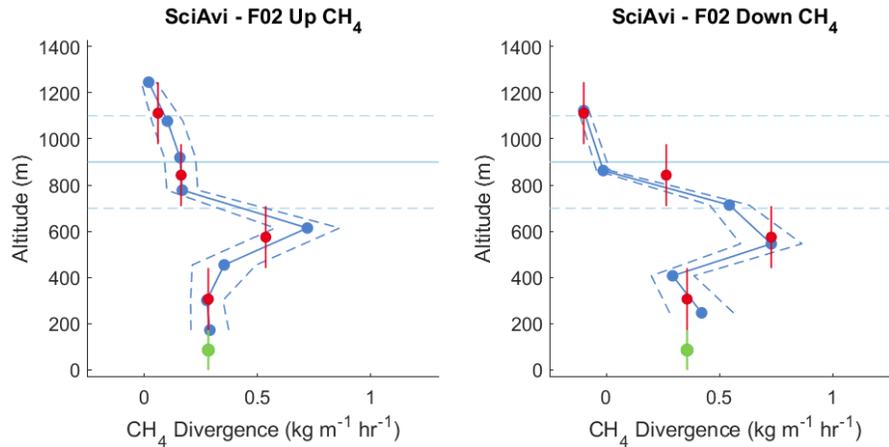


Figure 21. F02 CH₄ divergence profiles for the flight up (left) versus down (right). The profile shapes do not differ.

Table 10. F04 CH₄ and CO₂ standard estimates, ranges, and the emission estimates for the SciAv model split into the ascending and descending flight periods.

	SciAv Standard Estimate	SciAv Estimate Range	SciAv Ascending Estimate	SciAv Descending Estimate	TERRA Standard Estimate	TERRA Estimate Range
F04 CH ₄ (kg hr ⁻¹)	349	(310,387)	387	303	125	(102, 148)
F04 CO ₂ (t hr ⁻¹)	1170	(1050,13000)	1180	1130	570	(561, 608)
F04 CH ₄ (kg hr ⁻¹)	362	(287, 437)	329	398	395	(373, 418)
F02 CO ₂ (t hr ⁻¹)	563	(464, 662)	484	690	5153	(486, 545)

b) Investigate implications of violating the model assumptions of full capture of the plume.

To explore the extent of bias when a plume is not fully captured, upward extrapolation of the emission plume was applied to F05, the flight that did not capture the top of the plume, in TERRA. When doing this, the standard estimate changed from 3910 kg hr⁻¹ of CH₄ to 20641 kg hr⁻¹ and 877381 kg hr⁻¹ of CO₂ to 10340730 kg hr⁻¹. This tenfold increase exhibits why the method may not produce sensible estimates when the emission plume is increasing at the top of the box. With these methods, flights that fail to capture the top of the emission plume can rarely be used due to the large uncertainty that can occur.

2.3.3 Objective 3: Examine Potential for Validation of Mass-balance Emissions Estimate Methodology Through Comparison to the Spectral AVIRIS-NG Method

Two intersecting CH₄ plumes were captured at the Syncrude Plant site within the perimeter of the Scientific Aviation box-flight path of F04 at 21:17:24 on 08/11/2017 (111.6122455, 57.04661388). The NASA-JPL provided data, analysis, and imagery calculated from their sample, as described in Duren et al. (2019), over the F04 site to help provide context for the aircraft measurements (Table 11). The average instantaneous CH₄ emission rate from the three methods of calculating was 1,989 (kg hr⁻¹) with an averaged uncertainty of 950 (kg hr⁻¹). AVIRIS-NG estimates currently have on average larger uncertainties in emission estimates than the box-flight mass balance methods while avoiding the requirement of a stationary source and the need to extrapolate emissions to the surface. The estimated instantaneous emission rate measured by the monitoring of the NASA-JPL was approximately 4-7 times larger than emissions measured using the SciAv method three days later (349 kg hr⁻¹). This may be due to industrial operations, such a flaring or venting events. It is possible only one of the two plumes observed by AVIRIS-NG were present when SciAv sampled. However, operating conditions were not shared by the facility, so the underlying reason for this difference could not be evaluated.

Table 11. AVIRIS-NG data captured on August 14, 2017, three days 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 hr ⁻¹)	CH ₄ Estimate Uncertainty (kg hr ⁻¹)
Met 3062696	4.44	0.735	2230	973 (44%)
Met 3062697	4.81	1.81	2420	1330 (55%)
MERRA2 reanalysis	2.62	0.235	1320	543 (41%)

2.4 Discussion

2.4.1 Objective 1: Compare Emission Estimates Between Methods

a) Evaluate overall final emission estimates using the standard methods of the two models given the same data set.

The difference between models was computed (SciAv – TERRA) and compared as a relative percentage of the mean emission between the two models to the models’ errors added in quadrature (Table 6). Results support the assessment that four of the five flights produce estimates that agree between the two models. The considerably larger relative percentage difference of F04 compared to the percentage error in the estimates indicates that the difference between the model estimates is due to violated model assumptions, rather than error in the methods. The larger percentage difference is illustrated by the considerable gap between the F04 estimates in Figure 11. In general, when model assumptions were met the models produced similar results.

b) Systematically test the effects of different surface extrapolation methods for each model.

If the first steps in each model’s algorithm produced estimates that disagreed, then it would be most evident when the second step of extrapolating emissions to the surface is removed as a possible confounding factor. The SciAv and TERRA model estimates were compared when no surface extrapolation was applied. Estimates agreed which suggests the first, core mass-balance algorithm steps produce similar outputs.

The analysis did show one exception of a discernable pattern for flight F01, which indicates a source of bias between the models, due to the different methods of extrapolating to the surface. F01 was the only flight with a definitive SciAv profile type III and is an example of the difference in the model approaches to surface extrapolation when emissions are increasing at the lowest flight track. A linear weighted fit for profile III shapes, the scenario when emissions noticeably increase towards the surface, is recommended for SciAv to use as it fits the emission profile and is weighted by the uncertainty in each lap estimate (Figure 22).

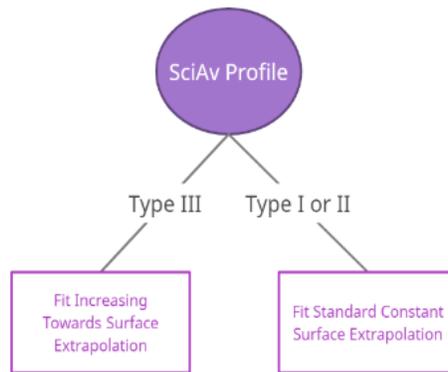


Figure 22. Conceptual diagram outlining when to apply an increasing towards surface extrapolation, or the SciAv standard constant surface extrapolation based on the shape of the SciAv profile for a given flight.

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 models agree more when some form of increasing to surface extrapolation is applied to the SciAv model compared to 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 models.

Results from assessing multiple surface extrapolations indicates that the SciAv method of extrapolating as a constant is the most appropriate choice unless a specific pattern of increasing emissions at the lowest flight path is evident. The SciAv method of applying a constant extrapolation avoids the bias introduced assuming either increasing or decreasing emission estimates when the behaviour is unknown. In the scenario when large variation occurs between lap estimates, such as F03, the SciAv method of extrapolating as a constant seems the most

appropriate as it avoids assumptions about the ambiguous profile behaviour at the bottom of the flight. F05 has an ambiguous emission profile, and while laps were flown to the height of the estimated boundary, emissions increase at the top of the profile, which indicates that the plume was breaking through the boundary layer. With so few sample laps to fit and a lack of capture at the top of the plume, extrapolating the emission as a constant to the surface likely introduces the least bias. F04 (the flight with non-stationarity) has large disagreement between the models, but the choice in the surface extrapolation has very little effect on the disagreement (3-9%). For the average flight scenario, the model estimates derived using various surface extrapolations tend to agree regardless of how the surface extrapolation is fit. This consistency between estimates provides larger certainty in the estimates and in the top-down regional budgets that are inferred from them.

c) Assess SciAv methods used to integrate divergence and characterize uncertainties.

Assumptions of the airborne, mass-balance algorithms include a stationary emission source, stationary meteorological conditions, and a fully captured emission plume. The smaller error for the TERRA estimates compared to the SciAv error estimates may be due to how each model quantifies the error in assuming these conditions are met. TERRA calculates seven specific error terms to address the error of these assumptions. Increases in the error of one assumption does not directly increase the error of others (see Appendix B). Whereas, SciAv uses two main broad terms, a temporal error term to capture the extent of stationarity, and a divergence error term to estimate capture of the plume. The divergence error of the SciAv method can become very large when few flight laps are flown. With a small number of laps (≤ 10) fewer bins are used than the six proposed in the Conley et al. (2017) methodology. This can lead to large variation amongst lap estimates within each bin (see F01 profile in Figure 13A). This error is often augmented, as the lowest-altitude bin error is extended to the surface, then doubled, to estimate the uncertainty in the surface extrapolation. Applying a different integration method and divergence error calculation, such as the trapezoidal integration method detailed in Appendix C, could reduce the estimate error by reducing the influence that a small number of laps, capturing different trends within a single bin, has on the calculation of the standard deviation of each bin (Figure 23). A proposed method of calculating an estimate of the error is to randomly sampling 5,000 times within the confidence interval of each lap point and integrating

across the differing profiles to determine a percent change over altitude amongst the sampled profiles. This term could be added to an extrapolation error term assessing the percent change in plausible fits to produce an error term akin to the SciAv's divergence term.

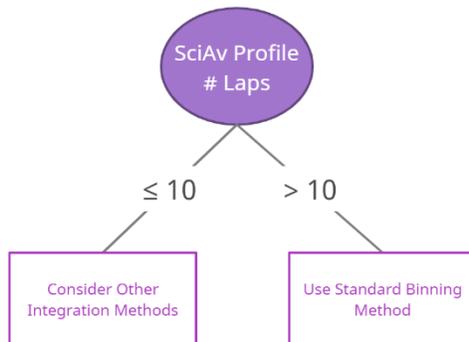


Figure 23. Conceptual diagram outlining when to use a differing integration method (such as Trapz), or the SciAv standard binning method given the number of laps in the SciAv profile for a given flight.

The flight found to have non-stationarity, F04, has 25 laps to average out the increased variability in the upward and downward profiles. The error calculated by SciAv for F04 was smaller than the error of the other four flights whereas CH₄ the error for the TERRA estimates was the largest for F04. This implies that current SciAv uncertainty terms are more influenced by the number of laps than by the degree of stationarity. More detailed knowledge of SciAv methodology, and analysis of a large sample of flights would be needed to determine the overlap, and influence model assumptions have on the temporal variability and divergence error terms described by Conley et al (2017).

Previous studies using both the SciAv and TERRA methods report that a large component of uncertainty for each method is due to the extrapolation to the surface. Choosing appropriate surface extrapolations reduces bias and increases the accuracy of the emission plume estimate, but the extent of error in the choice remains dependent on the extent of uncertainty in assuming the behaviour towards the surface. The uncertainty in the surface extrapolation from the SciAv model could be reduced if it is decoupled from the current divergence error term and calculated following TERRA methods of using the maximum percentage change between probable fits. Coinciding surface measurements at the time of sampling, and more information about the behaviour of a plume is the most likely path towards further reducing the uncertainty in the surface estimates for both models.

2.4.2 Objective 2: Investigate Model Assumptions

a) Investigate implications of violating the model assumptions of a stationary plume and stable meteorological conditions.

The difference between the models for F04 emerged due to conditions of a non-stationary emission plume. The contrasting approaches of kriging the data to produce spatially gridded data (TERRA) versus creating profiles of the divergence for each lap separately (SciAv) have varying sensitivities to the non-stationarity of the plume. The TERRA estimates are drastically lower than the SciAv estimates. This may be due to TERRA interpolating the screen of mixing ratios with grids averaging over the data from the flight up and down in one box, whereas SciAv may be more robust to non-stationarity as it averages each lap to estimate a divergence and then averages these over altitude bins.

During the initial QA/QC 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 (Appendix E). The SciAv F04 flight profile is an example of how a profile can look when many laps are sampled. In the scenario when the ideal 20-25 laps are sampled, averaging the laps over binned altitudes is necessary to smooth out the random variability in sampling. The whole (unsplit) F04 profile alone did not indicate non-stationarity of the emission plume (Figure 13D), but with post-hoc analysis (Figures 20 & 21), the consistent pattern of zig-zagging points might indicate a temporally varying profile and could be used as a flag for further investigation during QA/QC. For samples with many laps (≥ 20) separating the SciAv divergence profile into upward and downward flight components would help identify non-stationarity of an emission plume.

Over the course of F04, the concentrations of methane and CO₂ 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 emissions contributed to the observed changes in concentration. Comments were reported by study operators (AEP) on this flight that during sampling, facility operators instructed researchers to only sample the boundary of the facility for future flights, but operating conditions were not provided by industry. NO₂ and SO₂ emissions are often used as tracer data for upscaling CH₄ 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 NO₂ emissions spiked at the beginning of the day, before the aircraft measurements, and then continuously decreased. Furthermore, flaring data, including volume of gas flared and SO₂ emissions, did not suggest unusual operations at the plant on the day of the flight. The disagreement between the two model estimates for F04 arise from the non-stationary emission plume and how this effects the differing algorithms.

Meteorological conditions are often the most likely source of non-stationarity and as such are thoroughly assessed by explicit error terms in TERRA and through a QA/QC applied by Scientific Aviation that assesses whether conditions have been met. As the effects of violating the assumption of constant meteorological conditions are well assessed by both models, and due to the limited number of flights able to be compared, a flight with non-stable meteorological conditions was not intentionally included in the comparison analysis.

b) Investigate implications of violating the model assumptions of full capture of the plume.

When the top of an emission plume is not fully captured, the resulting emissions estimates can have uncertainties so large that the flight is unusable. F05 lacks plume capture at the top of the flight, but analysis of the boundary layer height indicated that sampling occurred to the top of the boundary layer. A moving, or unstable boundary layer would allow the plume to seemingly ‘break-through’ the layer. By extrapolating the emission screens in F05 upwards in TERRA, emissions estimates increased by approximately a factor of 10 for CH₄ and a factor of 100 for CO₂. If the bottom of a plume is not captured the extrapolation distance is restricted by the ground, but plume behaviour above the flight path is only approximately constrained by an estimate of the atmospheric boundary layer (Conley et al. 2017; Gordon et al. 2015). A ‘fully captured’ plume is essential to avoid the sampling bias that would lead to under-estimating emissions. Capture at the top of the plume by flying beyond the atmospheric boundary layer is essential for producing emission estimates with small uncertainties for mass-balance airborne flights.

2.4.3 Objective 3: Examine Potential for Validation of Mass-balance Emissions Estimate Methodology Through Comparison to the Spectral AVIRIS-NG Method

While the SciAv and TERRA methods have lower uncertainties than the AVIRIS-NG method, they require prior knowledge of presumed sources and approval from facilities to sample. Therefore, the mass-balance methods are unlikely to identify unknown sources located outside sampling boundaries, and may not capture the higher sporadic emissions that spectral sampling can. Due to the sporadic tendency of CH₄ emissions, and the different sampling date, the AVIRIS-NG results are not directly comparable to the mass-balance algorithms results for F04, but can provide an idea of the range of potential emissions. Unlike previously published comparisons that showed good agreement between spectral imaging and mass-balance methods (Frankenberg et al. 2016; Duren et al. 2019), the AVIRIS-NG data measured emissions 4-7 times larger than the SciAv estimates three days later. The large range in emission 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.

Currently, the Alberta and Canadian governments differ in their regulatory approaches to reducing methane emissions. It was estimated that Federal policy reduces emissions ~26% more than provincial regulations (Johnson and Tyner 2020). One of the key differences in policy was the increased number of oil and gas inspections required by the federal policy per year, 60,465 versus 26,531, reducing emissions by 66% (Johnson and Tyner 2020). Methane emission from the oil and gas industry are sporadic, with higher emissions only captured 20-35% of the time when sampling using remote spectral methods (Duren et al. 2019). Methods based on imaging spectrometry (e.g., AVIRIS-NG, GHGSat) provide a unique opportunity for validation of ground-based measurements and to help develop satellite monitoring techniques, while also providing leak detection.

Industrial upstream oil and gas CH₄ emissions estimated using the SciAv method in two regions in Alberta were 5 and 17 times higher than values reported to the Alberta Energy Regulator (Johnson et al. 2017). In a comparable campaign by ECCC, TERRA derived CO₂ emissions rates from the Alberta Oil Sands indicated intensities were 13-123% larger than emission estimates presented in publicly available, bottom-up inventories (Liggio et al. 2019).

The SciAv and AVIRIS-NG methods have been independently tested and have shown to provide consistently similar emission estimates when conducted under similar conditions (Frankenberg et al. 2016; Thorpe et al. 2020). Given the results of these studies, the AVIRIS-NG data captured on August 14, 2017, may not be anomalously high, but could instead represent independent information on the variability of emissions from the region. Further work comparing these methods under comparable conditions, along with greater transparency in facilities operations, could help confirm these conclusions and support accurate emission budgets.

2.5 Conclusion

The results of this study indicate that when fundamental assumptions are met the airborne, mass-balance algorithms, SciAv and TERRA, produce similar estimates that agree (3-25%) within model errors (4-34%). The implementation of a new method of applying different surface extrapolations to the SciAv estimates for flights with large increasing emissions towards the surface is recommended. The condition under which the models disagreed was when the fundamental model assumption of a stationary emission plume was not met. Stationarity has shown to be a more important factor for model agreement than plume capture, despite the surface extrapolation being the largest source of uncertainties in the model estimates. Introducing a more explicit procedure of screening for non-stationary emission plumes would improve confidence in the mass-balance models. Capturing the emission plume is a fundamental assumption of both models; lack of capture can produce large errors, yet model estimates closely agreed, despite producing equally under-estimated values for the flight that was missing capture of the top of the plume. This implies that estimates from these two methods can be combined to form a more comprehensive, cohesive dataset of airborne emission estimates.

Airborne methods provide validation for top-down and bottom-up estimates, and determining that estimated emissions are consistent between two top-down mass-balance methods increases the confidence in GHG emissions estimates. Since samples can only be obtained on an intermittent campaign basis, airborne methods are dependent on being representative of normal facility operations and require complementary data to capture the range in sporadic emissions when upscaling to yearly emissions estimates (Liggio et al. 2019). Airborne mass-balance methods provide detailed point-source estimates with a low uncertainty that are essential to the validation of inventory estimates. Having increased

confidence in estimates from the two mass-balance airborne methods provides a more certain foundation for regulatory decisions. Including airborne spectral estimates in top-down regional budgets improves characterization of emissions, provides more robust estimates, and increases the accuracy of bottom-up inventories. The current development of satellites specifically for monitoring GHG emissions has exciting potential to provide independent, top-down GHG budgets (Jacob et al. 2016).

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 2018; Saunio et al. 2017). Observations that combine and cross validate multiple monitoring methods at varying scales of sampling will provide the most accurate modeling, improve GHG estimation and help reconcile the gap between top-down and bottom-up estimates. Continued advances in developing more accurate inventories will allow for more effective policy decisions that target the contribution of carbon dioxide and methane to climate change.

Chapter 3

3.1 Conclusion

There are many methods for estimating GHG emissions using top-down and bottom-up approaches which often result in estimates that are inconsistent with emissions inventory estimates. Ground-based methods are cost effective and therefore an accessible approach for continuous monitoring, but due to their small scale of sampling often fail to produce regional estimates. Airborne methods provide more exhaustive sampling of emissions for sites and at a regional scale. They are ideal for validation between ground-based and satellite measurements, but can be costly, time intensive, but they cannot provide continuous monitoring. There are two main airborne methods, spectral remote imaging, and mass-balance methods. Spectral imaging can quickly sample large areas and can monitor for unknown emissions such as leaks, whereas mass-balance methods can measure emissions with a very low error rate and are more commonly used to capture a known emission plume when sampling. Satellite methods can provide continuous, independent global and regional emissions estimates. The method of estimating point source emissions estimates using satellite methods is still somewhat in its infancy. As satellite methods develop and improve accuracy, they will be pivotal in developing a network of global GHG monitoring systems. All top-down and bottom-up methods have unique strengths that are important to contributing to the understanding of GHG emissions.

Assessing how different estimation methods compare is vital for increasing credibility of greenhouse gas inventories and informing regulatory decisions. Numerous studies have shown that independent emissions estimates are larger than national inventories using industry estimates. Using tower data, methane emission estimates over eight 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). A 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, and were significantly underestimated in Alberta (MacKay et al., 2021). Airborne campaigns measuring carbon dioxide and methane have also estimated emissions to be larger than national inventories (Liggio et al. 2019; Baray et al. 2018). The development of spatially resolved top-down and

bottom-up inclusive inventories will be crucial for regulation and in holding nations accountable to their climate commitments.

In this study, for the first time two airborne mass-balance box flight methods, Scientific Aviation's Gaussian theorem algorithm (SciAv) and Environment Canada's Top-down Emission Rate Retrieval Algorithm (TERRA), were compared. While this comparison was limited by a small sample size ($n = 5$), a small sample size is not unusual as flights are costly, and analysis is time intensive. When fundamental model assumptions were met, the two methods produced results that agreed within model errors. SciAv model accuracy and uncertainty may be improved by applying an increasing surface extrapolation when an emission plume is increasing towards the surface. Meeting the model assumptions is essential to producing consistent and accurate emissions estimates with low uncertainty. Flights that indicate non-stationarity, or lacking plume capture should be thoroughly investigated before being used. A complementary, and independent AVIRIS-NG flight measured a sporadic emission nearly 4-7 times larger than a SciAv flight collected at the same facility three days earlier. This range in estimates from a single facility highlights the importance of accounting for large, sporadic emissions, which are inherent to the oil and gas industry, but seldom captured using the box-flight method, when producing yearly inventory estimates from airborne methods. Thus, to obtain a complete and accurate emissions budget, a mix of approaches involving validated, external monitoring will be needed to supplement the airborne methods used in this study.

Airborne, mass-balance box-flights are ideal for assessing facility level emissions as they can attain small error, and the consistency between the SciAv and TERRA methods gives confidence when using these methods, but using them alone would not produce sufficient inventory estimates. Instead, these methods provide a sharper tool to assess problem areas, and will be essential to validate and connect the scale of sampling between satellite and ground modeling. The sample estimate using airborne spectral imaging demonstrates the range in estimates found from a single site when sampling on different days, using different methods, and independent of facility operations and notification. This temporal variability in emissions must be considered and included in any emissions budget if it is to be complete and accurate. Increasing the consistency, and agreement when comparing methods moves us closer to

achieving routine monitoring at different scales and will provide more complete, transparent measurements that are not constrained by industry practices.

The ideal inventory would be built using industry reported ground-based data, in-depth airborne measurements, and independent, continuous monitoring using satellite methods. Using an array of emission estimation methods provides a check on accuracy depending on the conditions of operating. The resulting agreement between the SciAv and TERRA algorithms increases the certainty in regulatory decisions made from results when using these methods. Greater confidence in estimating GHG emissions from point and area sources using mass-balance, box-flights will help target anomalous carbon dioxide and methane emissions and will help meet regulatory and policy goals to address the global climate crises.

Bibliography

- Alfieri, S, U Amato, M F Carfora, M Esposito, and V Magliulo. 2010. “Quantifying Trace Gas Emissions from Composite Landscapes: A Mass-Budget Approach with Aircraft Measurements.” *Atmospheric Environment* 44 (15): 1866–76.
<https://doi.org/https://doi.org/10.1016/j.atmosenv.2010.02.026>.
- Allan, J. Stephens. 2021. “Report of the Public Inquiry into Anti-Alberta Energy Campaigns.”
<https://open.alberta.ca/dataset/3176fd2d-670b-4c4a-b8a7-07383ae43743/resource/a814cae3-8dd2-4c9c-baf1-cf9cd364d2cb/download/energy-report-public-inquiry-anti-alberta-energy-campaigns-2021.pdf>.
- Allen, David T. 2014. “Methane Emissions from Natural Gas Production and Use: Reconciling Bottom-up and Top-down Measurements.” *Current Opinion in Chemical Engineering* 5: 78–83. <https://doi.org/https://doi.org/10.1016/j.coche.2014.05.004>.
- AmeriFlux. 2019. “Overview—Year of Methane.” 2019. <https://ameriflux.lbl.gov/year-of-methane/year-of-methane/>.
- Balcombe, Paul, Jamie F Speirs, Nigel P Brandon, and Adam D Hawkes. 2018. “Methane Emissions: Choosing the Right Climate Metric and Time Horizon.” *Environ. Sci.: Processes Impacts* 20 (10): 1323–39. <https://doi.org/10.1039/C8EM00414E>.
- Baldocchi, Dennis, Eva Falge, Lianhong Gu, Richard Olson, David Hollinger, Steve Running, Peter Anthoni, et al. 2001. “FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities.” *Bulletin of the American Meteorological Society* 82 (11): 2415–34.
[https://doi.org/10.1175/1520-0477\(2001\)082<2415:FANTTS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2).
- Baray, Sabour, Andrea Darlington, Mark Gordon, Katherine L. Hayden, Amy Leithead, Shao Meng Li, Peter S.K. Liu, et al. 2018. “Quantification of Methane Sources in the Athabasca Oil Sands Region of Alberta by Aircraft Mass Balance.” *Atmospheric Chemistry and Physics* 18 (10): 7361–78. <https://doi.org/10.5194/acp-18-7361-2018>.
- Bartholomew, Jarett, Philip Lyman, Carl Weimer, and William Tandy. 2017. “Wide Area Methane Emissions Mapping with Airborne IPDA Lidar.” In *Proc.SPIE*. Vol. 10406.

<https://doi.org/10.1117/12.2276713>.

- Borchardt, Jakob, Konstantin Gerilowski, Sven Krautwurst, Heinrich Bovensmann, Andrew K. Thorpe, David R. Thompson, Christian Frankenberg, Charles E. Miller, Riley M. Duren, and John Philip Burrows. 2021. "Detection and Quantification of CH₄ Plumes Using the WFM-DOAS Retrieval on AVIRIS-NG Hyperspectral Data." *Atmospheric Measurement Techniques* 14 (2): 1267–91. <http://10.0.20.74/amt-14-1267-2021>.
- Boyd, Brendan. 2019. "A Province under Pressure: Climate Change Policy in Alberta." *Canadian Journal of Political Science* 52 (1): 183–99. <https://doi.org/DOI:10.1017/S0008423918000410>.
- Brandt, A R, G A Heath, E A Kort, F O'Sullivan, G Pétron, S M Jordaan, P Tans, et al. 2014. "Methane Leaks from North American Natural Gas Systems." *Science* 343 (6172): 733 LP – 735. <https://doi.org/10.1126/science.1247045>.
- Buchwitz, M, R de Beek, J P Burrows, H Bovensmann, T Warneke, J Notholt, J F Meirink, et al. 2005. "Atmospheric Methane and Carbon Dioxide from SCIAMACHY Satellite Data: Initial Comparison with Chemistry and Transport Models." *Atmos. Chem. Phys.* 5 (4): 941–62. <https://doi.org/10.5194/acp-5-941-2005>.
- Burgués, Javier, and Santiago Marco. 2020. "Environmental Chemical Sensing Using Small Drones: A Review." *Science of the Total Environment* 748 (December). <http://10.0.3.248/j.scitotenv.2020.141172>.
- Canada, Environment and Climate Change. 2016. *Pan-Canadian Framework on Clean Growth and Climate Change : Canada's Plan to Address Climate Change and Grow the Economy*. Gatineau: Government of Canada. http://publications.gc.ca/collections/collection_2017/eccc/En4-294-2016-eng.pdf.
- Canadian Press, The. 2021. "The Politics and Reality of Capping Alberta's Oil and Gas Emissions." *The Canadian Broadcasting Corporation*, November 4, 2021. <https://www.cbc.ca/news/canada/calgary/capping-emissions-alberta-oil-gas-ottawa-1.6237432>.
- Chaichana, Nongpat, Sonoko Dorothea Bellingrath-Kimura, Shujiro Komiya, Yoshiharu Fujii,

- Kosuke Noborio, Ottfried Dietrich, and Tiwa Pakoktom. 2018. "Comparison of Closed Chamber and Eddy Covariance Methods to Improve the Understanding of Methane Fluxes from Rice Paddy Fields in Japan." *Atmosphere* 9 (9): 356. <http://10.0.13.62/atmos9090356>.
- Chan, Elton, Douglas E J Worthy, Douglas Chan, Misa Ishizawa, Michael D Moran, Andy Delcloo, and Felix Vogel. 2020. "Eight-Year Estimates of Methane Emissions from Oil and Gas Operations in Western Canada Are Nearly Twice Those Reported in Inventories." *Environmental Science & Technology* 54 (23): 14899–909. <https://doi.org/10.1021/acs.est.0c04117>.
- Chernick, Michael R., and Robert H. Friis. 2003. "Appendix C: Quantiles of the Wilcoxon Signed-Rank Test Statistic." *Introductory Biostatistics for the Health Sciences*. Wiley Online Books. <https://doi.org/10.1002/0471458716.app3>.
- Ciais, P, C Sabine, G Bala, L Bopp, V Brovkin, et al., and Joanna Isobel House. 2014. "Carbon and Other Biogeochemical Cycles." In *Climate Change 2013*, edited by O Edenhofer, R Pichs-Madruga, Y Sokona, E Farahani, S Kadner, K Seyboth, A Adler, et al., 465–570. United Kingdom: Cambridge University Press.
- Conley, Stephen, Ian Faloon, Shobhit Mehrotra, Maxime Suard, Donald H. Lenschow, Colm Sweeney, Scott Herndon, et al. 2017. "Application of Gauss's Theorem to Quantify Localized Surface Emissions from Airborne Measurements of Wind and Trace Gases." *Atmospheric Measurement Techniques* 10 (9): 3345–58. <https://doi.org/10.5194/amt-10-3345-2017>.
- Cronmiller, Joshua G, and Bram F Noble. 2018. "The Discontinuity of Environmental Effects Monitoring in the Lower Athabasca Region of Alberta, Canada: Institutional Challenges to Long-Term Monitoring and Cumulative Effects Management." *Environmental Reviews* 26 (2): 169–80. <http://10.0.4.115/er-2017-0083>.
- Crosson, E R. 2008. "A Cavity Ring-down Analyzer for Measuring Atmospheric Levels of Methane, Carbon Dioxide, and Water Vapor." *Applied Physics B* 92 (3): 403–8. <https://doi.org/10.1007/s00340-008-3135-y>.
- Cusworth, Daniel H., Riley M. Duren, Vineet Yadav, Andrew K. Thorpe, Kristal Verhulst,

- Stanley Sander, Francesca Hopkins, Talha Rafiq, and Charles E. Miller. 2020. “Synthesis of Methane Observations Across Scales: Strategies for Deploying a Multitiered Observing Network.” *Geophysical Research Letters* 47 (7): e2020GL087869.
<https://doi.org/10.1029/2020GL087869>.
- Cusworth, Daniel H, Riley M Duren, Andrew K Thorpe, Winston Olson-Duvall, Joseph Heckler, John W Chapman, Michael L Eastwood, et al. 2021. “Intermittency of Large Methane Emitters in the Permian Basin.” *Environmental Science & Technology Letters* 8 (7): 567–73. <https://doi.org/10.1021/acs.estlett.1c00173>.
- Cusworth, Daniel H, Daniel J Jacob, Daniel J Varon, Christopher Chan Miller, Xiong Liu, Kelly Chance, Andrew K Thorpe, et al. 2019. “Potential of Next-Generation Imaging Spectrometers to Detect and Quantify Methane Point Sources from Space.” *Atmospheric Measurement Techniques Discussions*, May, 1–29. <http://10.0.20.74/amt-2019-202>.
- Dai, C, Q Wang, J A Kalogiros, D H Lenschow, Z Gao, and M Zhou. 2014. “Determining Boundary-Layer Height from Aircraft Measurements.” *Boundary-Layer Meteorology* 152 (3): 277–302. <https://doi.org/10.1007/s10546-014-9929-z>.
- Davidson, E A, K Savage, L V Verchot, and Rosa Navarro. 2002. “Minimizing Artifacts and Biases in Chamber-Based Measurements of Soil Respiration.” *Agricultural and Forest Meteorology* 113 (1): 21–37. [http://10.0.3.248/S0168-1923\(02\)00100-4](http://10.0.3.248/S0168-1923(02)00100-4).
- Diamond, Jared M. n.d. *Guns, Germs, and Steel : The Fates of Human Societies*. New York : Norton, [2005] ©2005. <https://search.library.wisc.edu/catalog/9910007689602121>.
- Dlugokencky, Edward J, Euan G Nisbet, Rebecca Fisher, and David Lowry. 2011. “Global Atmospheric Methane: Budget, Changes and Dangers.” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 369 (1943): 2058–72. <https://doi.org/10.1098/rsta.2010.0341>.
- Duren, Riley M, Andrew K Thorpe, Kelsey T Foster, Talha Rafiq, Francesca M Hopkins, Vineet Yadav, Brian D Bue, et al. 2019. “California’s Methane Super-Emitters.” *Nature* 575 (7781): 180–84. <https://doi.org/10.1038/s41586-019-1720-3>.
- Etheridge, D M, L P Steele, R J Francey, and R L Langenfelds. 1998. “Atmospheric Methane

- between 1000 A.D. and Present: Evidence of Anthropogenic Emissions and Climatic Variability.” *Journal of Geophysical Research: Atmospheres* 103 (D13): 15979–93. <https://doi.org/10.1029/98JD00923>.
- Falkowski, Author P, R J Scholes, E Boyle, J Canadell, D Canfield, J Elser, K Hibbard, et al. 2019. “The Global Carbon Cycle : A Test of Our Knowledge of Earth as a System” 290 (5490): 291–96.
- Flesch, T. K., J. D. Wilson, and L. A. Harper. 2005. “Deducing Ground-to-Air Emissions from Observed Trace Gas Concentrations: A Field Trial with Wind Disturbance.” *Journal of Applied Meteorology* 44 (4): 475–84. <https://doi.org/10.1175/jam2214.1>.
- Fox, Thomas A, Thomas E Barchyn, David Risk, Arvind P Ravikumar, and Chris H Hugenholtz. 2019. “A Review of Close-Range and Screening Technologies for Mitigating Fugitive Methane Emissions in Upstream Oil and Gas.” *Environmental Research Letters* 14 (5): 53002. <https://doi.org/10.1088/1748-9326/ab0cc3>.
- Frankenberg, Christian, Andrew K. Thorpe, David R. Thompson, Glynn Hulley, Eric Adam Kort, Nick Vance, Jakob Borchardt, et al. 2016. “Airborne Methane Remote Measurements Reveal Heavy-Tail Flux Distribution in Four Corners Region.” *Proceedings of the National Academy of Sciences* 113 (35): 9734–39. <https://doi.org/10.1073/pnas.1605617113>.
- Friedlingstein, Pierre, Michael O’Sullivan, Matthew W. Jones, Robbie M. Andrew, Judith Hauck, Are Olsen, Glen P. Peters, et al. 2020. “Global Carbon Budget 2020.” *Earth System Science Data* 12 (4): 3269–3340. <http://10.0.20.74/essd-12-3269-2020>.
- Gålfalk, Magnus, Göran Olofsson, and David Bastviken. 2017. “Approaches for Hyperspectral Remote Flux Quantification and Visualization of GHGs in the Environment.” *Remote Sensing of Environment* 191: 81–94. <https://doi.org/10.1016/j.rse.2017.01.012>.
- Gamon, J. A., A. F. Rahman, J. L. Dungan, M. Schildhauer, and K. F. Huemmrich. 2006. “Spectral Network (SpecNet)-What Is It and Why Do We Need It?” *Remote Sensing of Environment* 103 (3): 227–35. <https://doi.org/10.1016/j.rse.2006.04.003>.
- Gelaro, Ronald, Will McCarty, Max J Suárez, Ricardo Todling, Andrea Molod, Lawrence Takacs, Cynthia A Randles, et al. 2017. “The Modern-Era Retrospective Analysis for

- Research and Applications, Version 2 (MERRA-2).” *Journal of Climate* 30 (14): 5419–54. <https://doi.org/10.1175/JCLI-D-16-0758.1>.
- GHGSat. 2020. “SMALLEST METHANE EMISSION EVER DETECTED BY SATELLITE.” 2020. <https://www.ghgsat.com/smallest-methane-emission-ever-detected-by-satellite/>.
- Goodday, Victoria. 2021. “Environmental Regulation and the COVID-19 Pandemic: A Review of Regulator Response in Canada.” *The School of Public Policy Publications* 14 (10). <https://doi.org/10.11575/sppp.v14i.71913>.
- Gordon, M., S. M. Li, R. Staebler, A. Darlington, K. Hayden, J. O’Brien, and M. Wolde. 2015. “Determining Air Pollutant Emission Rates Based on Mass Balance Using Airborne Measurement Data over the Alberta Oil Sands Operations.” *Atmospheric Measurement Techniques* 8 (9): 3745–65. <https://doi.org/10.5194/amt-8-3745-2015>.
- Government of Alberta. 2013. *Oil Sands Environmental Monitoring Program Regulation, Reg 226/2013*.
- Government of Alberta. 2015. *Climate Change and Emissions Management Act: Specified Gas Emitters Regulation*. Edmonton, Canada. <http://canlii.ca/t/83nf>.
- Government of Alberta. 2019. “Public Inquiry into Anti-Alberta Energy Campaigns.” 2019. <https://www.alberta.ca/public-inquiry-into-anti-alberta-energy-campaigns.aspx>.
- Government of Canada. 2021a. “Canada’s Achievements at COP26.” <https://www.canada.ca/en/services/environment/weather/climatechange/canada-international-action/un-climate-change-conference/cop26-summit/achievements-at-cop26.html>.
- Government of Canada. 2021b. “Canada’s Climate Actions for a Healthy Environment and a Healthy Economy.” Gatineau QC. <https://doi.org/EC21125>.
- Government of Canada. 2021c. “Overview of 2019 Reported Emissions.” Gatineau, QC. <https://doi.org/2371-1035>.
- Hamlin, L, R O Green, P Mouroulis, M Eastwood, D Wilson, M Dudik, and C Paine. 2011. “Imaging Spectrometer Science Measurements for Terrestrial Ecology: AVIRIS and New

Developments.” In *2011 Aerospace Conference*, 1–7.
<https://doi.org/10.1109/AERO.2011.5747395>.

Hardwick, Stephen, and Heather Graven. 2016. “Satellite Observations to Support Monitoring of Greenhouse Gas Emissions.” *Grantham Briefing Papers*, no. 16.
<https://workspace.imperial.ac.uk/grantham/Public/publications/Satellite-observations-to-support-monitoring-of-greenhouse-gas-emissions-Grantham-BP-16.pdf>.

Heimann, I, P T Griffiths, N J Warwick, N L Abraham, A T Archibald, and J A Pyle. 2020. “Methane Emissions in a Chemistry-Climate Model: Feedbacks and Climate Response.” *Journal of Advances in Modeling Earth Systems* 12 (10): e2019MS002019-e2019MS002019. <https://doi.org/10.1029/2019MS002019>.

Hu, Haili, Jochen Landgraf, Rob Detmers, Tobias Borsdorff, Joost de Brugh, Ilse Aben, André Butz, and Otto Hasekamp. 2018. “Toward Global Mapping of Methane With TROPOMI: First Results and Intersatellite Comparison to GOSAT.” *Geophysical Research Letters* 45 (8): 3682–89. <https://doi.org/10.1002/2018GL077259>.

IPCC. n.d. “2018: Summary for Policymakers.” In *In: Global Warming of 1.5°C. An IPCC Special Report on the Impacts of Global Warming of 1.5°C above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change*, edited by P.R. Shukla, Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, M.I. Gomis, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, and T. Waterfield (eds.) E. Lonnoy, T. Maycock, M. Tignor. In Press.

IPCC 2008. n.d. “2006 IPCC Guidelines for National Greenhouse Gas Inventories – A Primer.” Edited by Srivastava N. and Tanabe K Prepared by the National Greenhouse Gas Inventories Programme, Eggleston H.S., Miwa K. *IGES, Japan*.

Jacob, D J, A J Turner, J D Maasackers, J Sheng, K Sun, X Liu, K Chance, I Aben, J McKeever, and C Frankenberg. 2016. “Satellite Observations of Atmospheric Methane and Their Value for Quantifying Methane Emissions.” *Atmos. Chem. Phys.* 16 (22): 14371–96.
<https://doi.org/10.5194/acp-16-14371-2016>.

- James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning : With Applications in R*. Springer Texts in Statistics. 103. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-7138-7_2.
- Janssens-Maenhout, G, Bernard Pinty, Mark Dowell, H Zunker, E Andersson, Gianpaolo Balsamo, J.-L Bézy, et al. 2020. “Toward an Operational Anthropogenic CO2 Emissions Monitoring and Verification Support Capacity.” *Bulletin of the American Meteorological Society* 101 (February). <https://doi.org/10.1175/BAMS-D-19-0017.1>.
- Jeong, Sangjae, Jeryang Park, Yeong Min Kim, Man Ho Park, and Jae Young Kim. 2019. “Innovation of Flux Chamber Network Design for Surface Methane Emission from Landfills Using Spatial Interpolation Models.” *Science of The Total Environment* 688: 18–25. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2019.06.142>.
- Jervis, D, J McKeever, B O A Durak, J J Sloan, D Gains, D J Varon, A Ramier, M Strupler, and E Tarrant. 2021. “The GHGSat-D Imaging Spectrometer.” *Atmospheric Measurement Techniques* 14 (3): 2127–40. <https://doi.org/10.5194/amt-14-2127-2021>.
- Johnson, Matthew R., Brian M. Crosland, James D. McEwen, Darcy B. Hager, Joshua R. Armitage, Mojgan Karimi-Golpayegani, and David J. Picard. 2016. “Estimating Fugitive Methane Emissions from Oil Sands Mining Using Extractive Core Samples.” *Atmospheric Environment* 144 (November): 111–23. <https://doi.org/10.1016/j.atmosenv.2016.08.073>.
- Johnson, Matthew R., David R. Tyner, Stephen Conley, Stefan Schwietzke, and Daniel Zavala-Araiza. 2017. “Comparisons of Airborne Measurements and Inventory Estimates of Methane Emissions in the Alberta Upstream Oil and Gas Sector.” *Environmental Science and Technology* 51 (21): 13008–17. <https://doi.org/10.1021/acs.est.7b03525>.
- Johnson, Matthew R, and David R Tyner. 2020. “A Case Study in Competing Methane Regulations: Will Canada’s and Alberta’s Contrasting Regulations Achieve Equivalent Reductions?” *Elementa: Science of the Anthropocene* 8 (1). <https://doi.org/10.1525/elementa.403>.
- Johnson, Matthew R, David R Tyner, and Alexander J Szekeres. 2021. “Blinded Evaluation of Airborne Methane Source Detection Using Bridger Photonics LiDAR.” *Remote Sensing of*

- Environment* 259: 112418. <https://doi.org/https://doi.org/10.1016/j.rse.2021.112418>.
- Jongaramrungruang, Siraput, Christian Frankenberg, Georgios Matheou, Andrew Thorpe, David R. Thompson, Le Kuai, and Riley Duren. 2019. “Towards Accurate Methane Point-Source Quantification from High-Resolution 2D Plume Imagery.” *Atmospheric Measurement Techniques Discussions*, no. May: 1–22. <https://doi.org/10.5194/amt-2019-173>.
- Kalthoff, N, U Corsmeier, K Schmidt, Ch. Kottmeier, F Fiedler, M Habram, and F Slemr. 2002. “Emissions of the City of Augsburg Determined Using the Mass Balance Method.” *Atmospheric Environment* 36: 19–31. [https://doi.org/https://doi.org/10.1016/S1352-2310\(02\)00215-7](https://doi.org/https://doi.org/10.1016/S1352-2310(02)00215-7).
- Karion, Anna, Colm Sweeney, Gabrielle Pétron, Gregory Frost, R Michael Hardesty, Jonathan Kofler, Ben R Miller, et al. 2013. “Methane Emissions Estimate from Airborne Measurements over a Western United States Natural Gas Field.” *Geophysical Research Letters* 40 (16): 4393–97. <https://doi.org/10.1002/grl.50811>.
- Kataoka, Fumie, David Crisp, Thomas Taylor, Christopher O’Dell, Akihiko Kuze, Kei Shiomi, Hiroshi Suto, et al. 2017. “The Cross-Calibration of Spectral Radiances and Cross-Validation of CO₂ Estimates from GOSAT and OCO-2.” *Remote Sensing* 9 (November): 1158. <https://doi.org/10.3390/rs9111158>.
- Kobayashi, N, G Inoue, M Kawasaki, H Yoshioka, M Minomura, I Murata, T Nagahama, et al. 2010. “Remotely Operable Compact Instruments for Measuring Atmospheric CO₂ and CH₄ Column Densities at Surface Monitoring Sites.” *Atmos. Meas. Tech.* 3 (4): 1103–12. <https://doi.org/10.5194/amt-3-1103-2010>.
- Kort, Eric A, Christian Frankenberg, Keeley R Costigan, Rodica Lindenmaier, Manvendra K Dubey, and Debra Wunch. 2014. “Four Corners: The Largest US Methane Anomaly Viewed from Space.” *Geophysical Research Letters* 41 (19): 6898–6903. <https://doi.org/10.1002/2014GL061503>.
- Krings, Thomas, Bruno Neininger, Konstantin Gerilowski, Sven Krautwurst, Michael Buchwitz, John P. Burrows, Carsten Lindemann, Thomas Ruhtz, Dirk Schüttemeyer, and Heinrich Bovensmann. 2018. “Airborne Remote Sensing and in Situ Measurements of Atmospheric

- CO₂ to Quantify Point Source Emissions.” *Atmospheric Measurement Techniques* 11 (2): 721–39. <https://doi.org/10.5194/amt-11-721-2018>.
- Li, Shao-Meng, Amy Leithead, Samar G. Moussa, John Liggio, Michael D. Moran, Daniel Wang, Katherine Hayden, et al. 2017. “Differences between Measured and Reported Volatile Organic Compound Emissions from Oil Sands Facilities in Alberta, Canada.” *Proceedings of the National Academy of Sciences* 114 (19): E3756–65. <https://doi.org/10.1073/pnas.1617862114>.
- Li, Steve, Xiaoli Sun, Kenji Numata, Stewart Wu, Michael Krainak, James Abshire, and Haris Riris. 2010. “Lidar Measurements of Methane and Applications for Aircraft and Spacecraft.” United States: NASA Center for Aerospace Information (CASI). <https://login.ezproxy.library.ualberta.ca/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edsnas&AN=edsnas.20110015176&site=eds-live&scope=site>.
- Liggio, John, Shao Meng Li, Katherine Hayden, Youssef M. Taha, Craig Stroud, Andrea Darlington, Brian D. Drollette, et al. 2016. “Oil Sands Operations as a Large Source of Secondary Organic Aerosols.” *Nature* 534 (7605): 91–94. <https://doi.org/10.1038/nature17646>.
- Liggio, John, Shao Meng Li, Ralf M. Staebler, Katherine Hayden, Andrea Darlington, Richard L. Mittermeier, Jason O’Brien, et al. 2019. “Measured Canadian Oil Sands CO₂ Emissions Are Higher than Estimates Made Using Internationally Recommended Methods.” *Nature Communications* 10 (1). <https://doi.org/10.1038/s41467-019-09714-9>.
- Liu, Peter S. K., Sabour Baray, Mengistu Wolde, Amy Leithead, Shao-Meng Li, Richard L. Mittermeier, Doug Worthy, et al. 2018. “Quantification of Methane Sources in the Athabasca Oil Sands Region of Alberta by Aircraft Mass Balance.” *Atmospheric Chemistry and Physics* 18 (10): 7361–78. <https://doi.org/10.5194/acp-18-7361-2018>.
- Lyon, David R, Daniel Zavala-Araiza, Ramón A Alvarez, Robert Harriss, Virginia Palacios, Xin Lan, Robert Talbot, et al. 2015. “Constructing a Spatially Resolved Methane Emission Inventory for the Barnett Shale Region.” *Environmental Science & Technology* 49 (13): 8147–57. <https://doi.org/10.1021/es506359c>.

- MacKay, Katlyn, Martin Lavoie, Evelise Bourlon, Emmaline Atherton, Elizabeth O’Connell, Jennifer Baillie, Chelsea Fougère, and David Risk. 2021. “Methane Emissions from Upstream Oil and Gas Production in Canada Are Underestimated.” *Scientific Reports* 11 (1): 8041. <https://doi.org/10.1038/s41598-021-87610-3>.
- Maslin, Mark A. 2020. “The Road from Rio to Glasgow: A Short History of the Climate Change Negotiations.” *Scottish Geographical Journal* 136 (1–4): 5–12. <http://10.0.4.56/14702541.2020.1853873>.
- Nalli, Nicholas R, Changyi Tan, Juying Warner, Murty Divakarla, Antonia Gambacorta, Michael Wilson, Tong Zhu, et al. 2020. “Validation of Carbon Trace Gas Profile Retrievals from the NOAA-Unique Combined Atmospheric Processing System for the Cross-Track Infrared Sounder.” *Remote Sensing* . <https://doi.org/10.3390/rs12193245>.
- Nambiar, Manoj K, Françoise R Robe, Alison M Seguin, Matthew Endsinn, and Amir A Aliabadi. 2020. “Diurnal and Seasonal Variation of Area-Fugitive Methane Advective Flux from an Open-Pit Mining Facility in Northern Canada Using WRF.” *Atmosphere* . <https://doi.org/10.3390/atmos11111227>.
- Nassar, Ray, Timothy G Hill, Chris A McLinden, Debra Wunch, Dylan B A Jones, and David Crisp. 2017. “Quantifying CO2 Emissions From Individual Power Plants From Space.” *Geophysical Research Letters* 44 (19): 10,10-45,53. <https://doi.org/https://doi.org/10.1002/2017GL074702>.
- National Academies of Sciences, Engineering, and Medicine. 2018. “Improving Characterization of Anthropogenic Methane Emissions in the United States,” March. <https://doi.org/10.17226/24987>.
- Nisbet, Euan, and Ray Weiss. 2010. “Top-Down Versus Bottom-Up.” *Science* 328 (5983): 1241 LP – 1243. <https://doi.org/10.1126/science.1189936>.
- Peng, Roger D. 2019. “Simpleboot: Simple Bootstrap Routines.” <https://github.com/rdpeng/simpleboot>.
- Quéré, Corinne Le, Robbie M Andrew, Pierre Friedlingstein, Stephen Sitch, Judith Hauck, Julia Pongratz, Penelope A Pickers, Jan Ivar Korsbakken, Glen P Peters, and Josep G Canadell.

2018. “Global Carbon Budget 2018,” 2141–94.
- Rashid, Kashif, Andrew Speck, Timothy P Osedach, Dominic V Perroni, and Andrew E Pomerantz. 2020. “Optimized Inspection of Upstream Oil and Gas Methane Emissions Using Airborne LiDAR Surveillance.” *Applied Energy* 275: 115327. <https://doi.org/https://doi.org/10.1016/j.apenergy.2020.115327>.
- Reuter, Maximilian, Michael Buchwitz, Oliver Schneising, Stefan Noël, Heinrich Bovensmann, and John P Burrows. 2017. “A Fast Atmospheric Trace Gas Retrieval for Hyperspectral Instruments Approximating Multiple Scattering—Part 2: Application to XCO₂ Retrievals from OCO-2.” *Remote Sensing* . <https://doi.org/10.3390/rs9111102>.
- Rey, Denise, and Markus Neuhäuser. 2011. “Wilcoxon-Signed-Rank Test BT - International Encyclopedia of Statistical Science.” In , edited by Miodrag Lovric, 1658–59. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_616.
- Riddick, Stuart N, Denise L Mauzerall, Michael A Celia, Mary Kang, Kara Bressler, Christopher Chu, and Caleb D Gum. 2019. “Measuring Methane Emissions from Abandoned and Active Oil and Gas Wells in West Virginia.” *Science of The Total Environment* 651: 1849–56. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2018.10.082>.
- Rogelj, Joeri, Michiel Schaeffer, Pierre Friedlingstein, Nathan P. Gillett, Detlef P. Van Vuuren, Keywan Riahi, Myles Allen, and Reto Knutti. 2016. “Differences between Carbon Budget Estimates Unravelling.” *Nature Climate Change*. Nature Publishing Group 2016-02-24. <https://doi.org/10.1038/nclimate2868>.
- Roscioli, Joseph R, Scott C Herndon, Tara I Yacovitch, W Berk Knighton, Daniel Zavala-Araiza, Matthew R Johnson, and David R Tyner. 2018. “Characterization of Methane Emissions from Five Cold Heavy Oil Production with Sands (CHOPS) Facilities.” *Journal of the Air & Waste Management Association* 68 (7): 671–84. <https://doi.org/10.1080/10962247.2018.1436096>.
- Saunio, Marielle, Philippe Bousquet, Ben Poulter, Anna Peregon, Philippe Ciais, Josep G. Canadell, Edward J. Dlugokencky, et al. 2016. “The Global Methane Budget 2000-2012.” *Earth System Science Data* 8 (2): 697–751. <https://doi.org/10.5194/essd-8-697-2016>.

- Saunois, Marielle, Philippe Bousquet, Ben Poulter, Anna Peregon, Philippe Ciais, Josep G. Canadell, Edward J. Dlugokencky, et al. 2017. “Variability and Quasi-Decadal Changes in the Methane Budget over the Period 2000-2012.” *Atmospheric Chemistry and Physics*. Copernicus GmbH. <https://doi.org/10.5194/acp-17-11135-2017>.
- Saunois, Marielle, Ann R Stavert, Ben Poulter, Philippe Bousquet, Joseph G Canadell, Robert B Jackson, Peter A Raymond, et al. 2020. “The Global Methane Budget 2000–2017.” *Earth System Science Data* 12 (3): 1561–1623. <https://doi.org/10.5194/essd-12-1561-2020>.
- Schuur, E A G, A D McGuire, C Schädel, G Grosse, J W Harden, D J Hayes, G Hugelius, et al. 2015. “Climate Change and the Permafrost Carbon Feedback.” *Nature* 520 (April): 171. <https://doi.org/10.1038/nature14338>.
- Schwietzke, Stefan, Matthew Harrison, Terri Lauderdale, Ken Branson, Stephen Conley, Fiji C George, Doug Jordan, et al. 2019. “Aerially Guided Leak Detection and Repair: A Pilot Field Study for Evaluating the Potential of Methane Emission Detection and Cost-Effectiveness.” *Journal of the Air & Waste Management Association* 69 (1): 71–88. <https://doi.org/10.1080/10962247.2018.1515123>.
- Sheng, J.-X., D J Jacob, J D Maasackers, Y Zhang, and M P Sulprizio. 2018. “Comparative Analysis of Low-Earth Orbit (TROPOMI) and Geostationary (GeoCARB, GEO-CAPE) Satellite Instruments for Constraining Methane Emissions on Fine Regional Scales: Application to the Southeast US.” *Atmos. Meas. Tech.* 11 (12): 6379–88. <https://doi.org/10.5194/amt-11-6379-2018>.
- Shephard, M. W., C. A. McLinden, K. E. Cady-Pereira, M. Luo, S. G. Moussa, A. Leithead, J. Liggio, et al. 2015. “Tropospheric Emission Spectrometer (TES) Satellite Observations of Ammonia, Methanol, Formic Acid, and Carbon Monoxide over the Canadian Oil Sands: Validation and Model Evaluation.” *Atmospheric Measurement Techniques* 8 (12): 5189–5211. <https://doi.org/10.5194/amt-8-5189-2015>.
- Sherwin, Evan D, Yuanlei Chen, Arvind P Ravikumar, and Adam R Brandt. 2021. “Single-Blind Test of Airplane-Based Hyperspectral Methane Detection via Controlled Releases.” *Elementa: Science of the Anthropocene* 9 (1). <https://doi.org/10.1525/elementa.2021.00063>.

- Shimoda, H, and T Ogawa. 2000. “Interferometric Monitor for Greenhouse Gases (IMG).” *Advances in Space Research* 25 (5): 937–46. [https://doi.org/https://doi.org/10.1016/S0273-1177\(99\)00926-6](https://doi.org/10.1016/S0273-1177(99)00926-6).
- Small, Christina C., Sunny Cho, Zaher Hashisho, and Ania C. Ulrich. 2015. “Emissions from Oil Sands Tailings Ponds: Review of Tailings Pond Parameters and Emission Estimates.” *Journal of Petroleum Science and Engineering* 127: 490–501. <https://doi.org/10.1016/j.petrol.2014.11.020>.
- The Globe and Mail. 2021. “Ottawa Pledges to End Financing for Foreign Fossil-Fuel Projects in 2022,” November 4, 2021. <https://www.theglobeandmail.com/world/article-ottawa-pledges-to-end-financing-for-foreign-fossil-fuel-projects-in/>.
- Thorpe, A. K., C Frankenberg, D.~R. Thompson, R.~M. Duren, B.~D. Bue, and R.~O. Green. 2017. “Mapping Methane Emissions Using the Airborne Imaging Spectrometer AVIRIS-NG.” In *Mapping Methane Emissions Using the Airborne Imaging Spectrometer AVIRIS-NGAGU Fall Meeting Abstracts*, 2017:A43N-07.
- Thorpe, Andrew K., Dar A. Roberts, Eliza S. Bradley, Christopher C. Funk, Philip E. Dennison, and Ira Leifer. 2013. “High Resolution Mapping of Methane Emissions from Marine and Terrestrial Sources Using a Cluster-Tuned Matched Filter Technique and Imaging Spectrometry.” *Remote Sensing of Environment* 134 (July): 305–18. <https://doi.org/10.1016/j.rse.2013.03.018>.
- Thorpe, Andrew K, Riley M Duren, Stephen Conley, Kuldeep R Prasad, Brian D Bue, Vineet Yadav, Kelsey T Foster, et al. 2020. “Methane Emissions from Underground Gas Storage in California.” *Environmental Research Letters* 15 (4): 45005. <https://doi.org/10.1088/1748-9326/ab751d>.
- Tian, Hanqin, Ray Langenfelds, Brett F. Thornton, Thomas Kleinen, Robin Locatelli, Giuseppe Etiope, Christine Wiedinmyer, et al. 2016. “The Global Methane Budget 2000–2012.” *Earth System Science Data* 8 (2): 697–751. <https://doi.org/10.5194/essd-8-697-2016>.
- Turner, Alexander J, Christian Frankenberg, and Eric A Kort. 2019. “Interpreting Contemporary Trends in Atmospheric Methane.” *Proceedings of the National Academy of Sciences of the*

- United States of America* 116 (8): 2805–13. <https://doi.org/10.1073/pnas.1814297116>.
- Tyner, David R, and Matthew R Johnson. 2021. “Where the Methane Is—Insights from Novel Airborne LiDAR Measurements Combined with Ground Survey Data.” *Environmental Science & Technology* 55 (14): 9773–83. <https://doi.org/10.1021/acs.est.1c01572>.
- Varon, Daniel J., Daniel J. Jacob, Jason McKeever, Dylan Jervis, Berke O.A. Durak, Yan Xia, and Yi Huang. 2018. “Quantifying Methane Point Sources from Fine-Scale Satellite Observations of Atmospheric Methane Plumes.” *Atmospheric Measurement Techniques* 11 (10): 5673–86. <https://doi.org/10.5194/amt-11-5673-2018>.
- Varon, Daniel J, Daniel J Jacob, Dylan Jervis, and Jason McKeever. 2020. “Quantifying Time-Averaged Methane Emissions from Individual Coal Mine Vents with GHGSat-D Satellite Observations.” *Environmental Science & Technology* 54 (16): 10246–53. <https://doi.org/10.1021/acs.est.0c01213>.
- Varon, Daniel, J McKeever, D Jervis, J D Maasackers, Sudhanshu Pandey, S Houweling, I Aben, T Scarpelli, and D J Jacob. 2019. “Satellite Discovery of Anomalously Large Methane Point Sources From Oil/Gas Production.” *Geophysical Research Letters* 46 (October). <https://doi.org/10.1029/2019GL083798>.
- Vinuesa, Jean François, and Stefano Galmarini. 2009. “Turbulent Dispersion of Non-Uniformly Emitted Passive Tracers in the Convective Boundary Layer.” *Boundary-Layer Meteorology* 133 (1): 1–16. <https://doi.org/10.1007/s10546-009-9416-0>.
- Wojcik, Michael, Blake Crowther, Robert Lemon, Prasad Valupadas, Long Fu, Bonnie Leung, Zheng Yang, Quamrul Huda, and Allan Chambers. 2015. “Development of Differential Absorption Lidar (DIAL) for Detection of CO₂, CH₄ and PM in Alberta.” In *Proc.SPIE*. Vol. 9486. <https://doi.org/10.1117/12.2176984>.
- Wolf, Sebastian. 2010. “Carbon Dioxide and Water Vapour Fluxes of Tropical Pasture and Afforestation: Seasonal Variations of Net Ecosystem Exchange and Carbon Sequestration Potentials.” ETH Zurich. <https://doi.org/DISS>. ETH NO. 19415.
- Worden, J, S Kulawik, C Frankenberg, V Payne, K Bowman, K Cady-Peirara, K Wecht, J.-E. Lee, and D Noone. 2012. “Profiles of CH₄, HDO, H₂O, and N₂O with Improved Lower

- Tropospheric Vertical Resolution from Aura TES Radiances.” *Atmos. Meas. Tech.* 5 (2): 397–411. <https://doi.org/10.5194/amt-5-397-2012>.
- Wührer, Christian, Christopher Köhl, Stefano Lucarelli, and M Bode. 2019. “MERLIN: Overview of the Design Status of the Lidar Instrument.” In *Proc.SPIE*. Vol. 11180. <https://doi.org/10.1117/12.2535999>.
- You, Y, R M Staebler, S G Moussa, J Beck, and R L Mittermeier. 2021. “Methane Emissions from an Oil Sands Tailings Pond: A Quantitative Comparison of Fluxes Derived by Different Methods.” *Atmospheric Measurement Techniques* 14 (March): 1879–92. <https://doi.org/10.5194/amt-14-1879-2021>.
- Zavala-Araiza, Daniel, Scott C Herndon, Joseph R Roscioli, Tara I Yacovitch, Matthew R Johnson, David R Tyner, Mark Omara, and Berk Knighton. 2018. “Methane Emissions from Oil and Gas Production Sites in Alberta, Canada.” *Elementa: Science of the Anthropocene* 6 (1). <https://doi.org/10.1525/elementa.284>.
- Zhang, Junhua, Michael D. Moran, Qiong Zheng, Paul A. Makar, Pegah Baratzadeh, George Marson, Peter Liu, and Shao Meng Li. 2018. “Emissions Preparation and Analysis for Multiscale Air Quality Modeling over the Athabasca Oil Sands Region of Alberta, Canada.” *Atmospheric Chemistry and Physics* 18 (14): 10459–81. <https://doi.org/10.5194/acp-18-10459-2018>.

Appendix A: Summary of Relevant Methods from the Literature

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:

$$E_C = E_{C,H} + E_{C,HT} + E_{C,V} + E_{C,VT} + E_{C,VD} - E_{C,m} - E_{C,X} \quad (3)$$

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 24 depicts the flux directions of each E integral term in Equation 3 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).

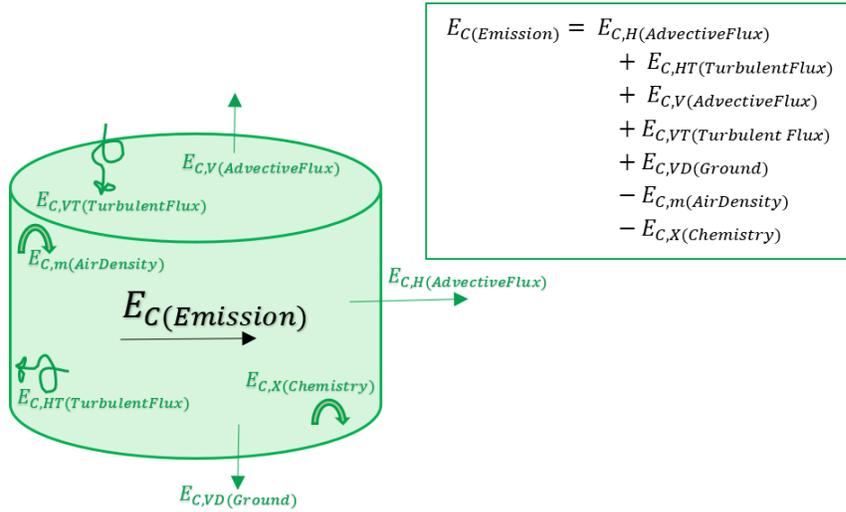


Figure 24. The TERRA integral E terms of Equation 3 with arrows depicting their contribution to estimating emission flux through the ‘box’, or volume.

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

$$0 = E_{\text{air},H} + E_{\text{air},V} - E_{\text{air},M} \quad (4)$$

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 air mass 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 3 which are added in quadrature to estimate the total uncertainty (δ_{TERRA}^2) in the emission rate estimate (Gordon et al. 2015).

$$\delta_{\text{TERRA}}^2 = \delta_M^2 + \delta_{\text{Ex}}^2 + \delta_{\text{Wind}}^2 + \delta_{\text{Top}}^2 + \delta_{\text{dens}}^2 + \delta_{\text{VT}}^2 + \delta_{\text{BH}}^2 \quad (5)$$

The seven error terms in Equation 5 pertain to the uncertainty in: measurement (δ_M^2), near surface extrapolation of the mixing ratio (δ_{Ex}^2), near surface extrapolation of the wind (δ_{Wind}^2), the box-top mixing ratio (δ_{Top}^2), the change in air density (δ_{dens}^2), the vertical turbulent flux (δ_{VT}^2), and the boundary layer height (δ_{BH}^2). 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 Scientific Aviation: 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 et al. (2017) use the measurements to directly evaluate the balance budget of Equation 6 as described in their methods:

$$Q_c = \left\langle \frac{\partial m}{\partial t} \right\rangle + \int_0^{z_{\max}} \oint c' u_h \cdot \hat{n} dl dz \quad (6)$$

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 25. Equation 6 is used to calculate the mean divergence for each lap. Gas expanding outwardly from the box, due to a larger positive than negative flux through the box, is an example of positive divergence, and gas collapsing would be negative divergence. 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 (Chapter 2, Figure 9).

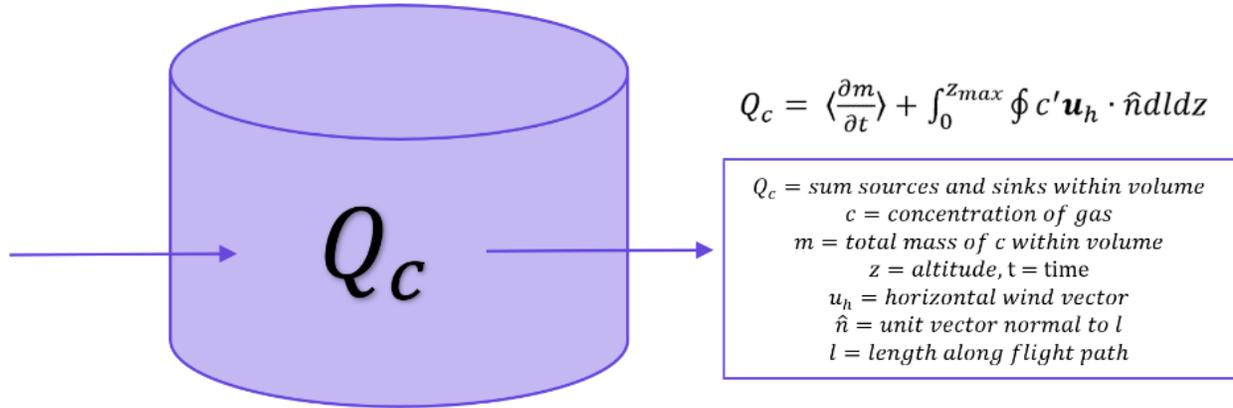


Figure 25. The SciAv method simplifies the mass-balance equation to one instantaneous flux to calculate the overall horizontal sum of the source and sinks within the ‘box’, or volume.

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

$$\delta_{\text{SciAv}}^2 = \delta_M^2 + \delta_{\text{flux divergence}}^2 + \delta_{\text{temporal}}^2 \quad (7)$$

The first term is the measurement error (δ_M^2) which contributes the smallest amount to the final estimate error (Conley et al. 2017). The second is the flux divergence error ($\delta_{\text{flux divergence}}^2$) and is calculated by summing the variance of the 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 ($\delta_{\text{temporal}}^2$) 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).

Appendix B: 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 certain fits for each flight. This document outlines the processes to evaluate those four requirements as instructed by Environment and Climate Change Canada (ECCC).

Section 1.1: TERRA Flight Screens

Chapter 2 section 2.3.1, details the surface extrapolation fitting process. The emission screens produced by TERRA after fitting the chosen surface extrapolation are shown in Figures 26 - 35. The left y-axis gives the altitude above ground level in meters, the right hand the colour scale of CH₄, or CO₂ in ppm. The length along each lap s(m) is plotted 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 CH₄ is noticeable in Figure 32. The surface extrapolation sometimes estimates a decreasing emission plume towards the surface as per Figures 27 and 35. 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 (see section 1.3). Aside from F04, the flight data used in the comparison analysis represent standard emissions screens for the TERRA method.

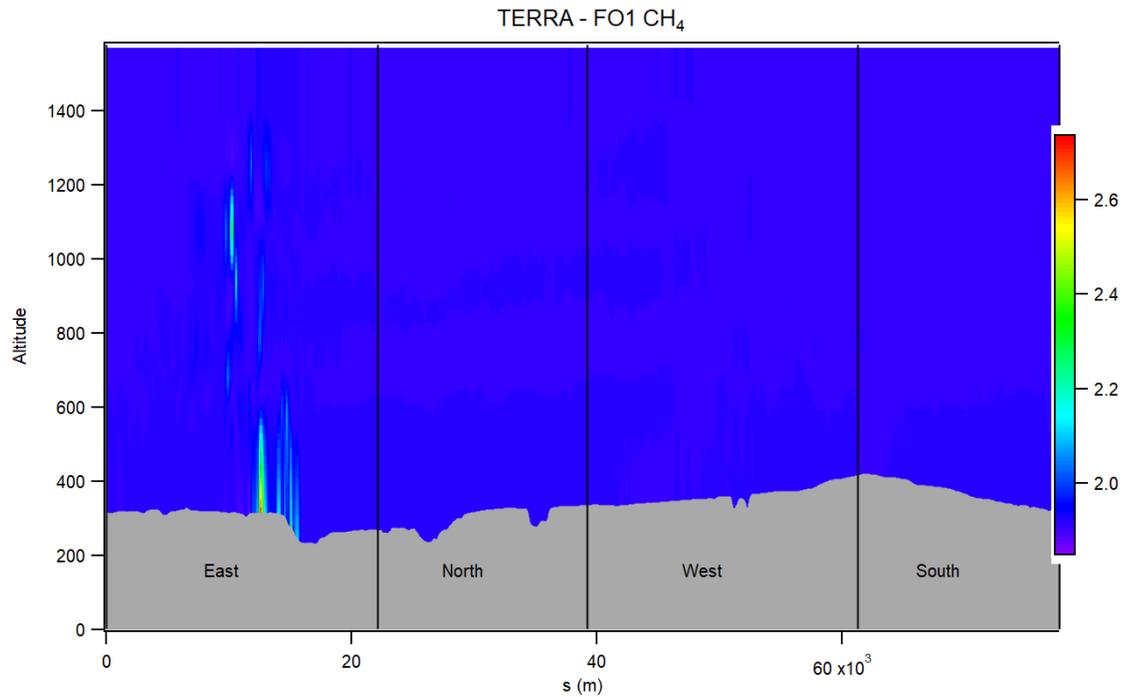


Figure 26. The TERRA screen for F01 CH₄ in ppm. Altitude measured in meters is shown along the left y-axis and the colour bar on the right depicts the mixing ratio gradient in ppm. The length along each map (s) is plotted along the x-axis with the direction of sampling overlaid. The ground is shown in grey.

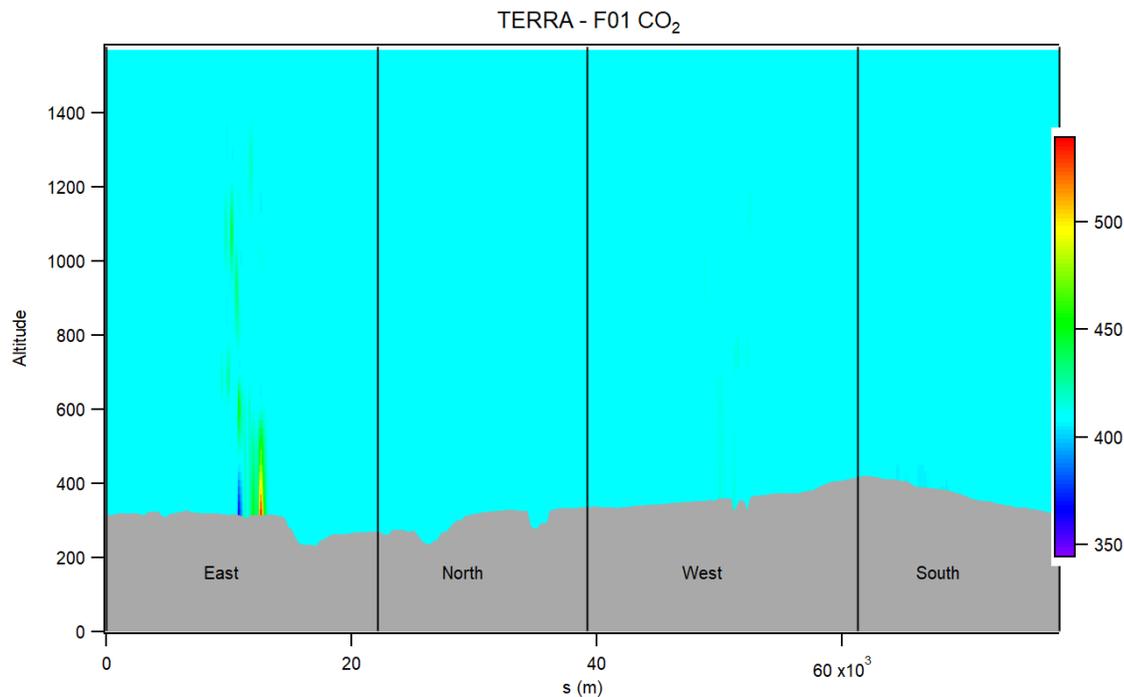


Figure 27. Same as Figure 26 for flight F01 CO₂ (ppm).

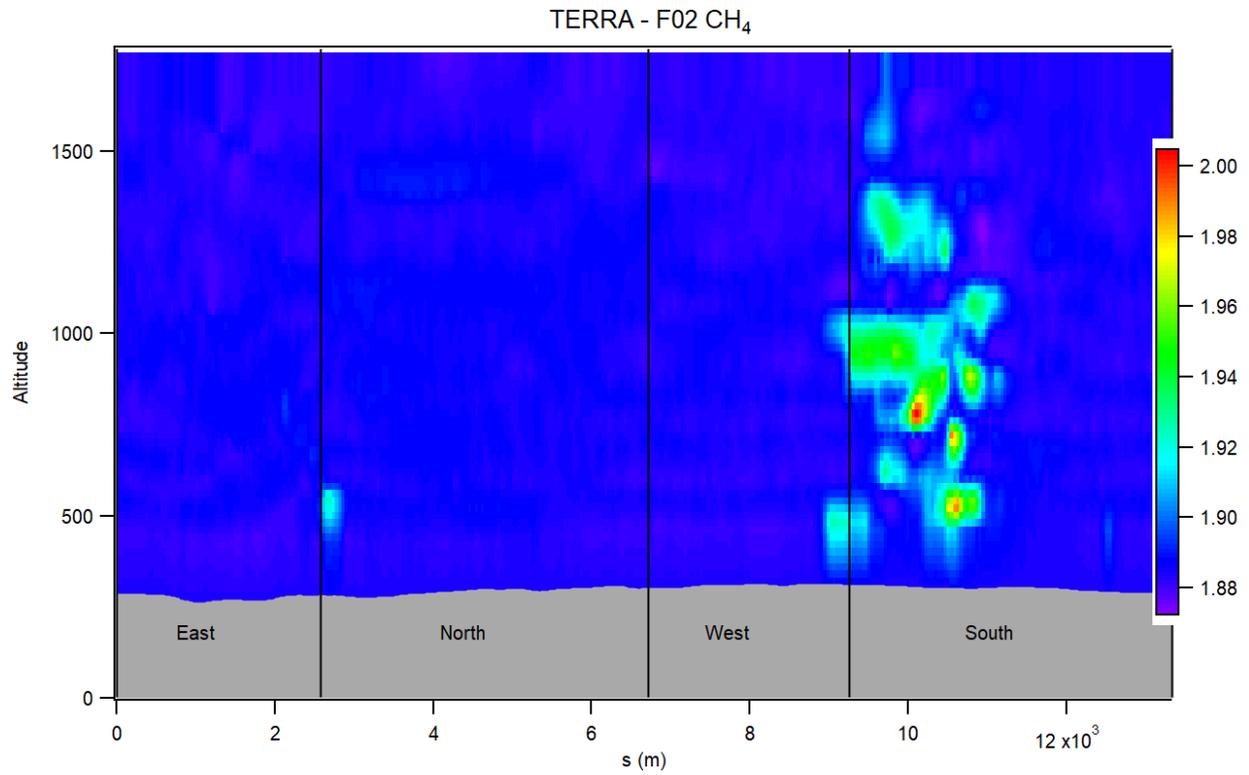


Figure 28. Same as Figure 26 for flight F02 CH₄ (ppm).

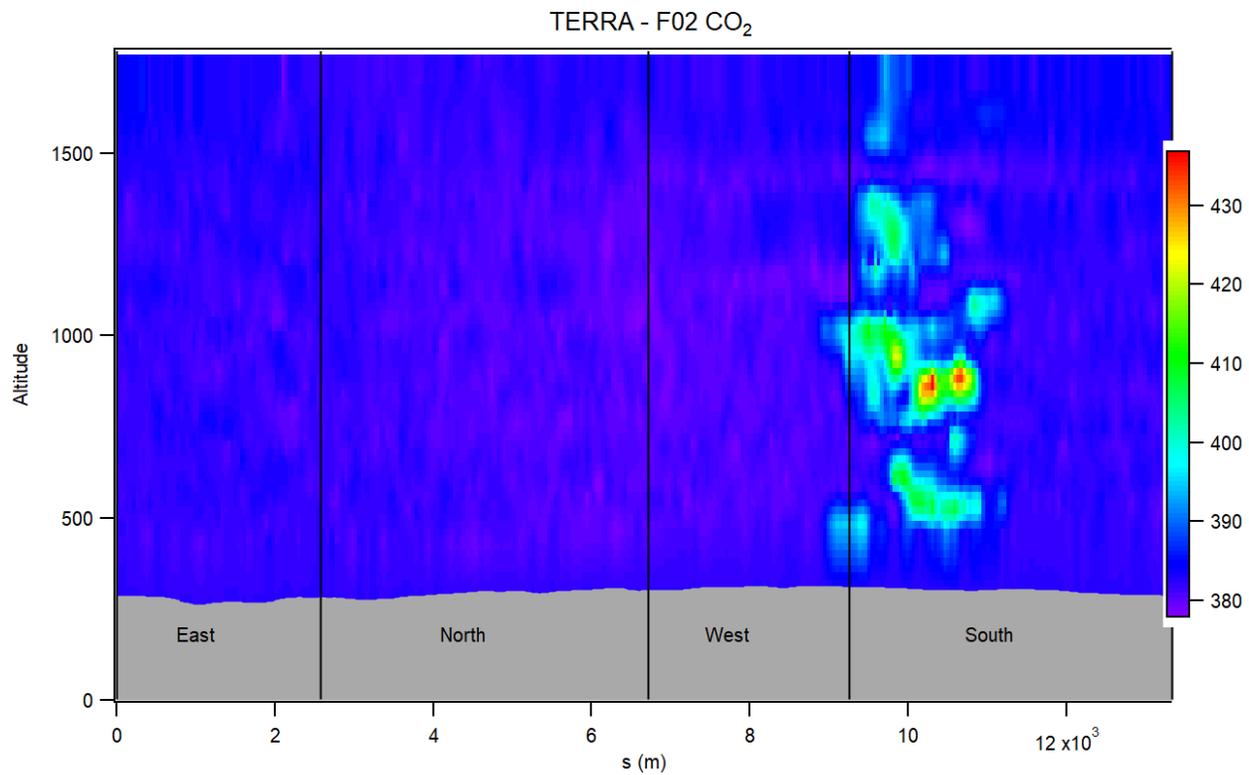


Figure 29. Same as Figure 26 for flight F02 CO₂ (ppm).

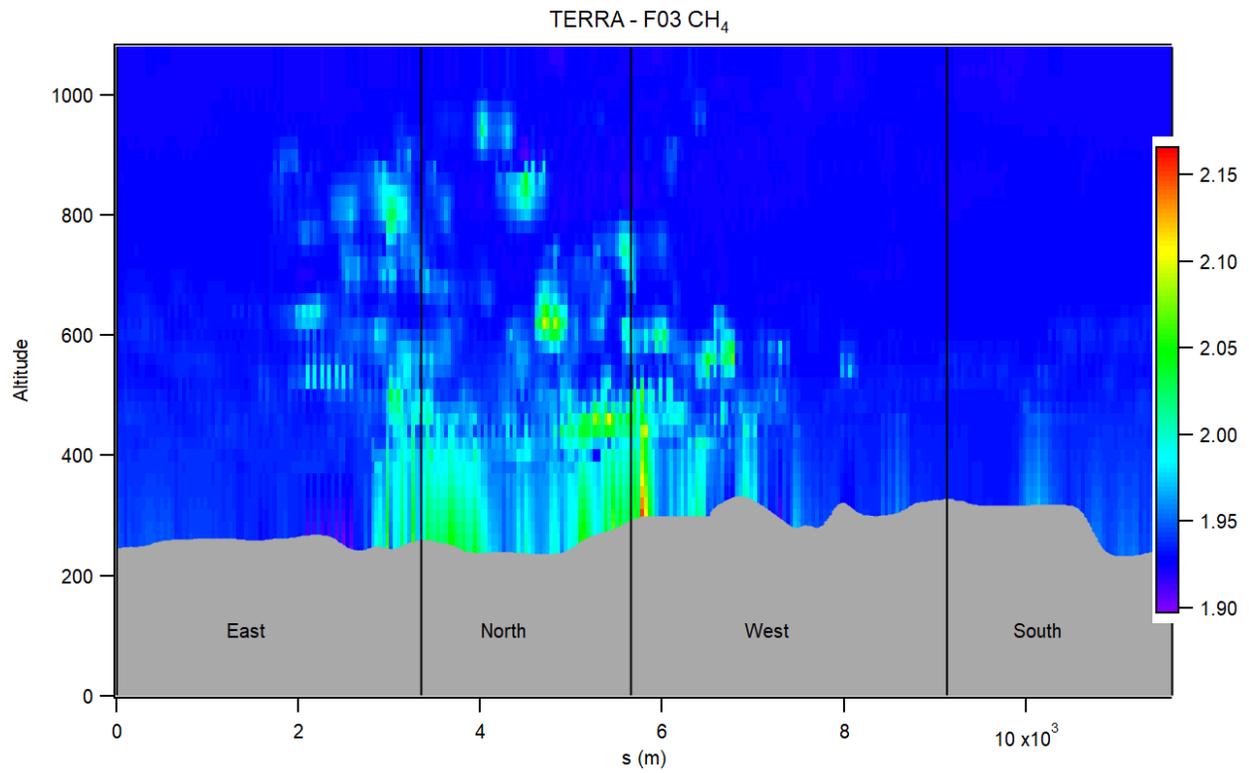


Figure 28. Same as Figure 26 for flight F03 CH₄ (ppm).

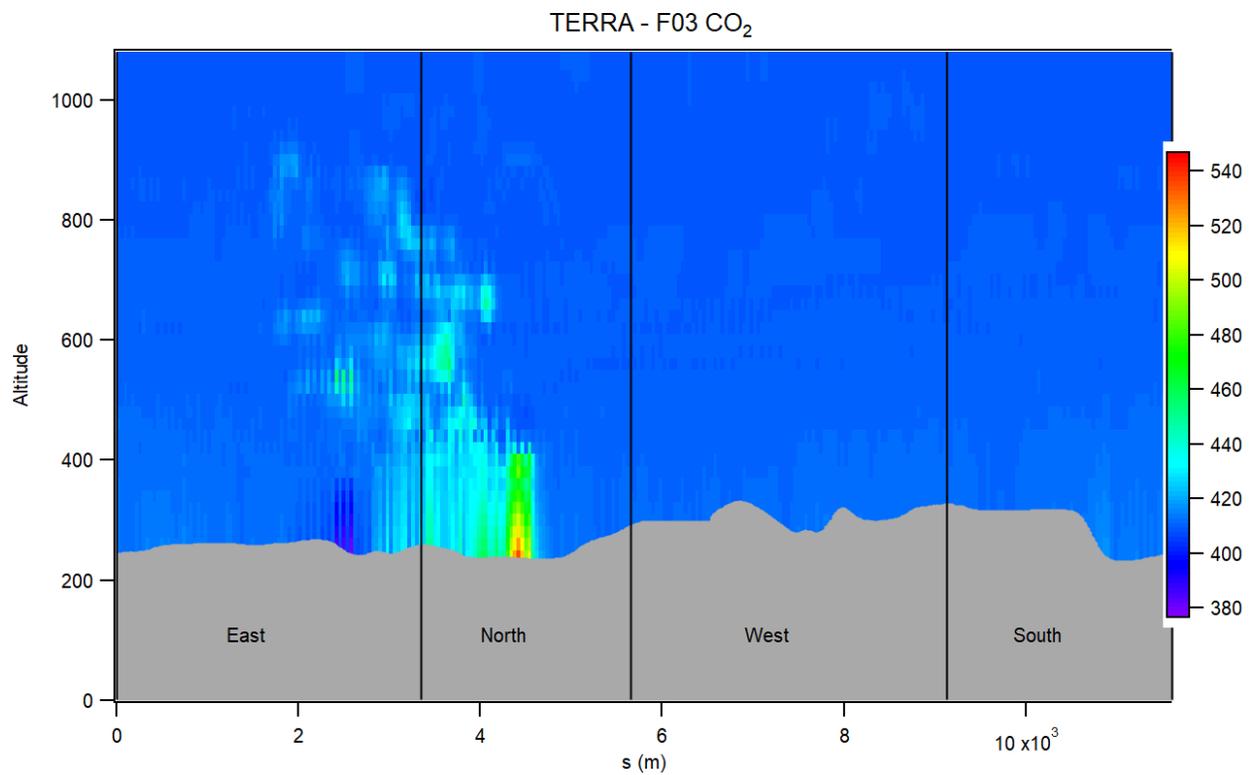


Figure 29. Same as Figure 26 for flight F03 CO₂ (ppm).

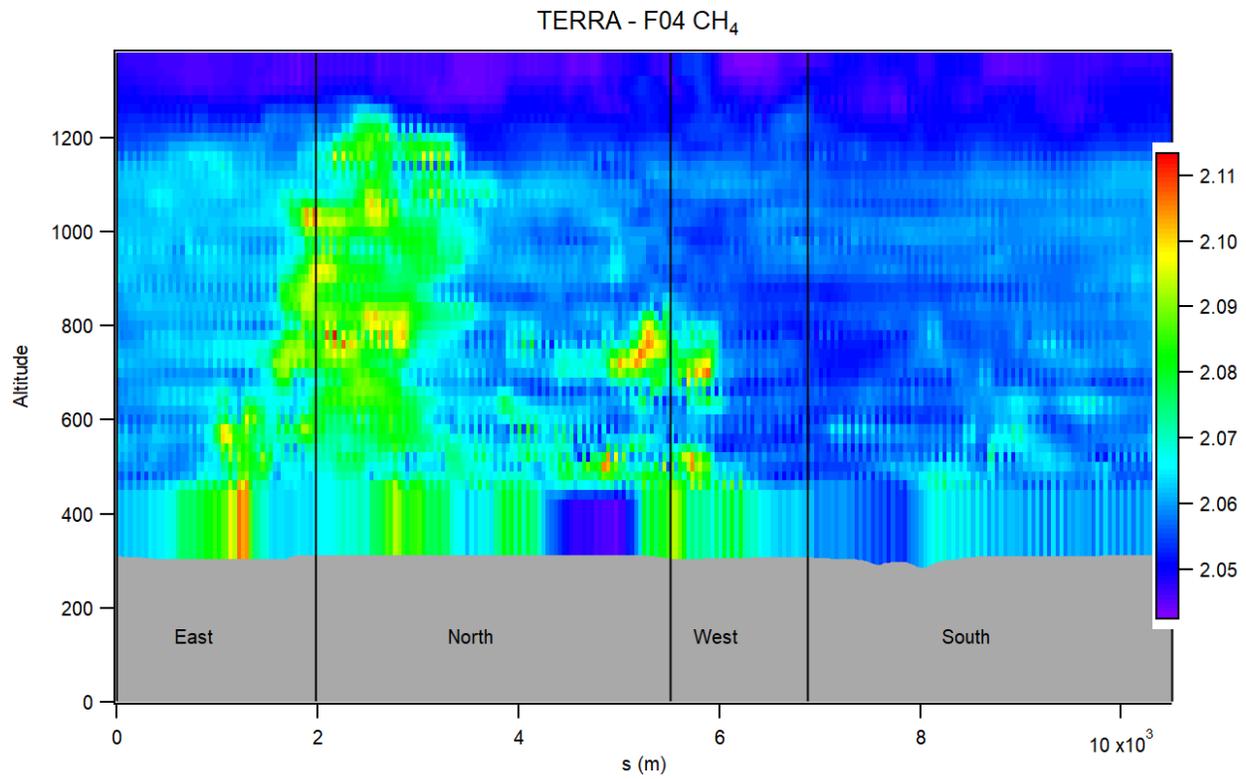


Figure 30. Same as Figure 26 for flight F04 CH₄ (ppm).

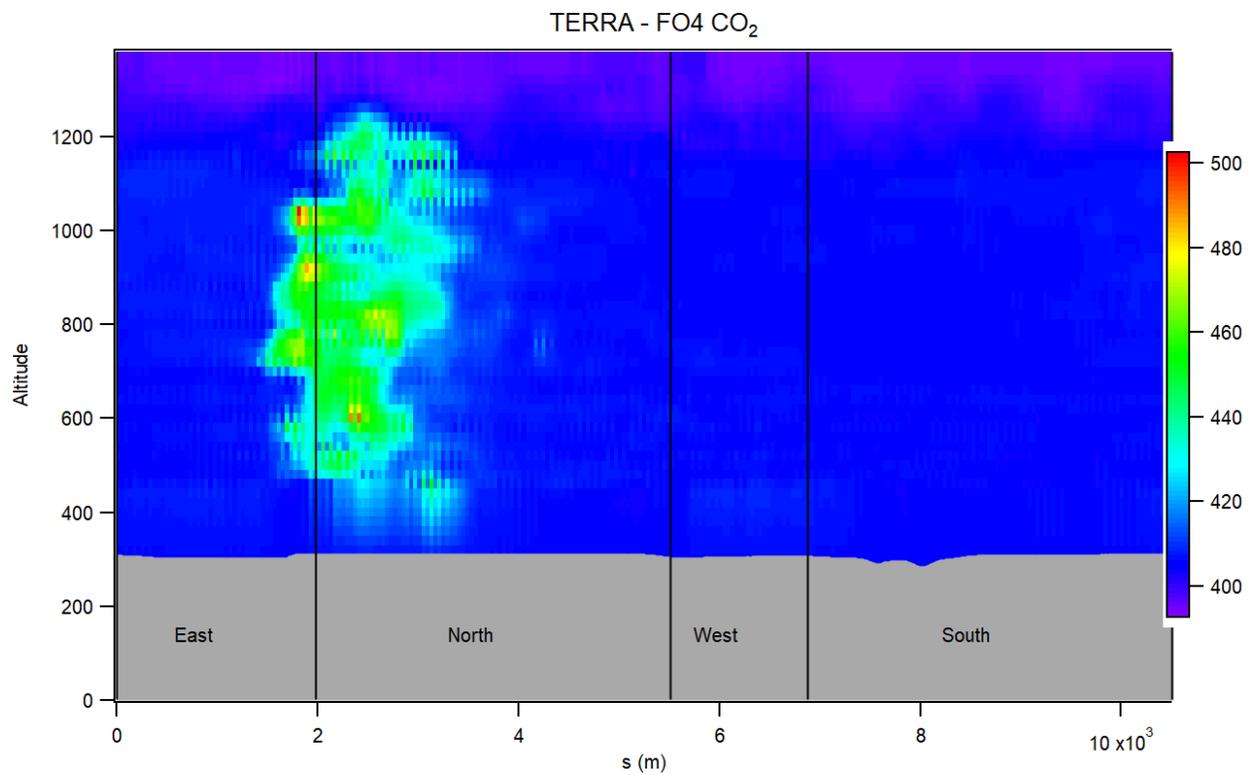


Figure 31. Same as Figure 26 for flight F04 CO₂ (ppm).

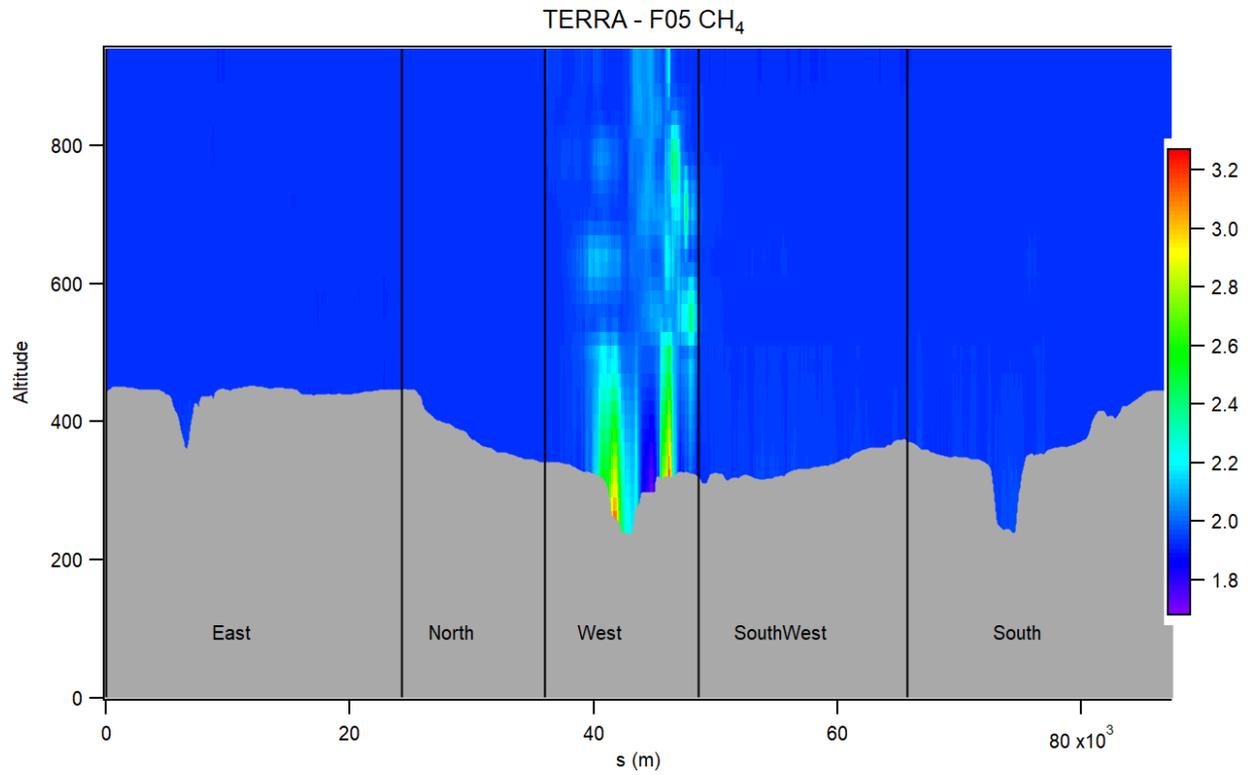


Figure 32. Same as Figure 26 for flight F05 CH₄ (ppm).

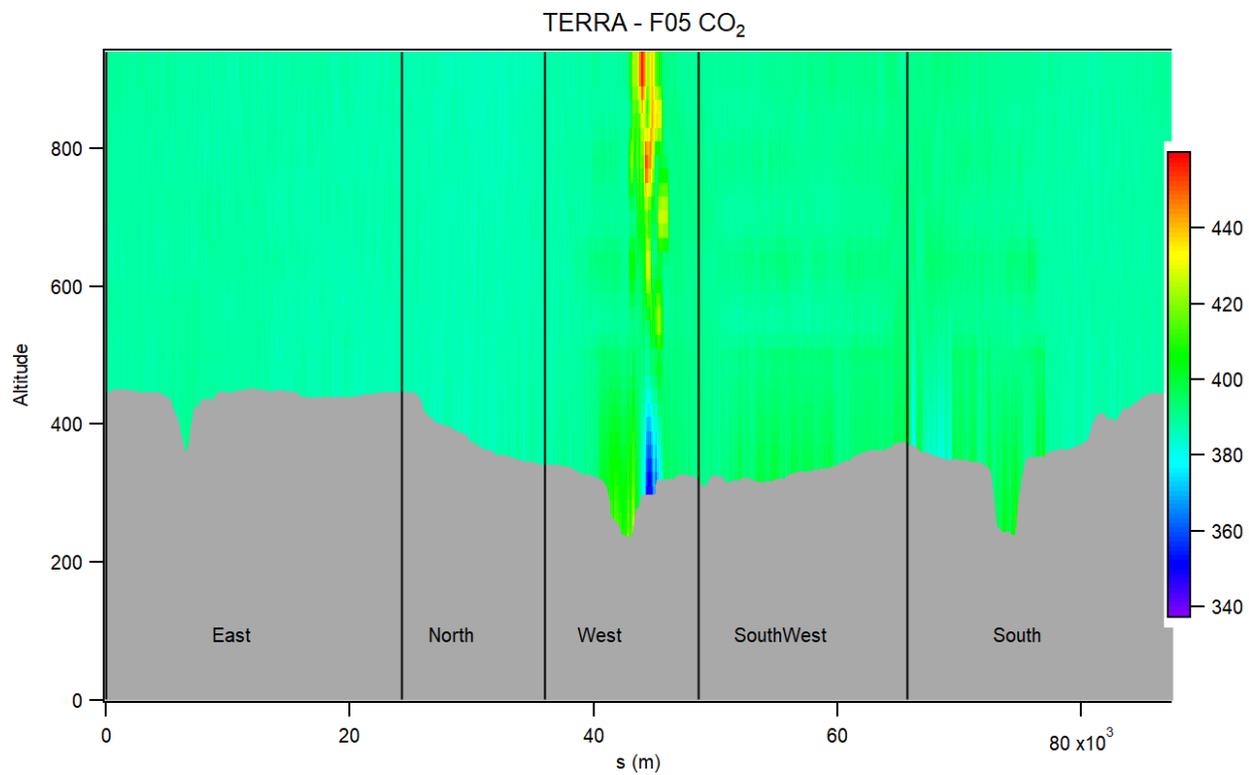


Figure 33. Same as Figure 26 for flight F05 CO₂ (ppm).

Section 1.2: TERRA Error Terms

The calculated TERRA error terms are provided in Table 12 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 $\pm 2\sigma/\sqrt{n}$, 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 E_{max}/E to E_{min}/E .

The box top height error was calculated in IGOR 9 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 12. 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

Section 1.3: Choosing the Background Mixing Ratio Values

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 was estimated by assessing the distribution of the raw data and the values used are given in Table 13. The detailed process of choosing these values is described in this section. The “background” and “linear interpolate to background” fits were avoided for F04 CH₄ due to larger uncertainty in the choice of background value.

Table 13. Background mixing ratio values used in the surface extrapolation of each 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

The mixing ratios of each emitted gas are enhanced when there is an emission plume and therefore the enhancements had to be removed from the raw data to approximate the mean background value for each flight. Summary statistics of the raw data before removing the enhancements are presented in Table 14. To calculate an approximate average background value, mixing ratios that could be assumed to be plume enhancements were excluded and a curve was fit to the remaining distribution of values. The limit for each fit was chosen by assessing the distribution of the raw data and conservatively set to exclude the distribution outliers of the large enhancements. The mean value was estimated from the new distribution and the values given in Table 15 were used for the background and linear interpolation to background surface extrapolations for TERRA. Values were also considered for continuity of the emission screen. Figure 36 illustrates the sensitivity to the different mixing ratio as it shows the different shade of blue for the surface extrapolation when using 1.90 ppm versus 1.88481 ppm for the background values for F02. The left screen using 1.90 ppm shows a lighter background value that is mismatched with the rest of the screen, whereas the right screen using 1.88481 ppm has continuity and the location of where the surface extrapolation starts is not noticeable (Figure 36). Figure 37 shows the more variable F04 CH₄ TERRA screens for a background surface extrapolation fit of 2.07 ppm (left) and 2.05365 ppm (right) illustrating the better cohesive fit of a lower background value. As a result, the CH₄ emission enhancement exclusion cut off was dropped from 2.075 to 2.06 ppm. Due to the uncertainty in choosing the F04 background mixing values, a constant surface extrapolation was used for the final emission estimate as it does not require an estimate of the background value.

Table 14. Summary statistics of the raw flight data in ppm.

Flight	CH ₄ Mean	CH ₄ Standard Deviation	CH ₄ Min	CH ₄ Max	CO ₂ Mean	CO ₂ Standard Deviation	CO ₂ Min	CO ₂ Max
F01	1.91613	0.01414	1.9109	2.4218	409.365	2.108	408.686	475.404
F02	1.88705	0.010898	1.8772	2.0283	383.262	4.878	378.229	445.840
F03	1.93751	0.024372	1.9190	2.1470	411.398	5.970	408.692	528.732
F04	2.06189	0.012393	2.0426	2.1255	409.705	13.350	394.518	522.505
F05	1.94174	0.049252	1.9154	2.5649	389.923	5.584	382.474	463.807

Table 15. Summary statistics of the raw flight data excluding emission enhancements in ppm.

Flight	CH ₄ Exclusion Limit	CH ₄ Mean	CH ₄ Standard Deviation	CO ₂ Exclusion Limit	CO ₂ Mean	CO ₂ Standard Deviation
F01	1.95	1.91613	0.003095	411	409.190	0.253334
F02	1.90	1.88481	0.002692	390	382.322	1.46831
F03	1.955	1.93007	0.006702	415	410.267	0.937985
F04	2.06	2.05365	0.004063	420	405.872	3.96142
F05	1.97	1.93103	0.006499	420	389.470	2.9545

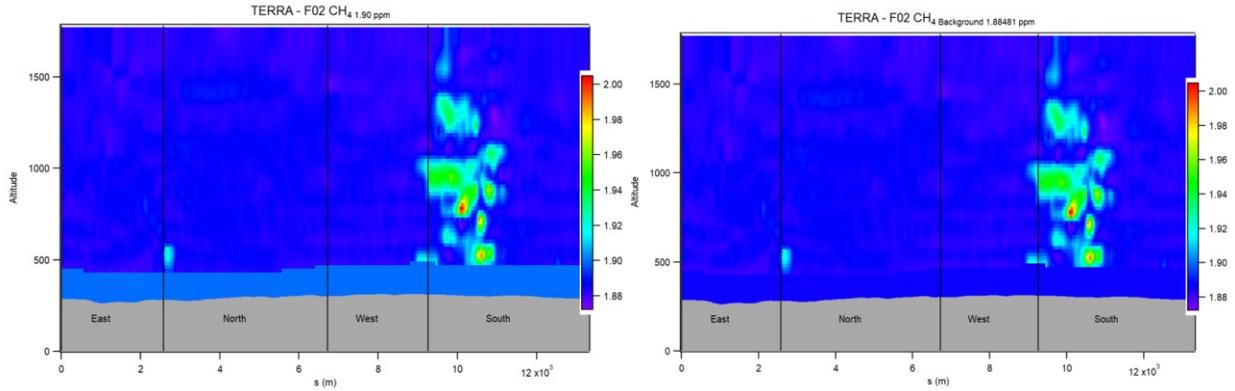


Figure 34. TERRA produced screens of the measured emission plume for F02 CH₄. The gap between the lowest flight lap and surface is filled using an estimated mean background value. The left screen shows the mismatch in the screen when a background concentration of 1.90 ppm is used in comparison to the value of 1.88481 ppm that was in analysis to calculate emissions.

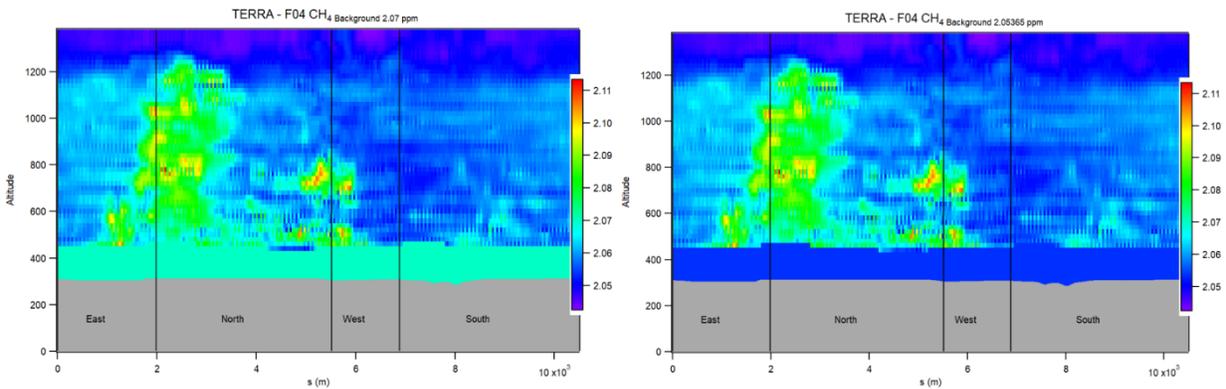


Figure 35. TERRA screens for F04 CH₄ when using 2.07 ppm (left) and 2.05365 ppm (right) as the background value.

Section 1.4: Sensitivity Analysis of the Background Value

A sensitivity analysis was run on two flights to see how the chosen background value affects the emission estimate for the “background”, and “interpolate to background” fits. F02 and F04 were analyzed as they are the two flights that use an interpolate to background surface extrapolation in the comparison analysis. F02 is an example of the average distribution of the flight raw data (Figure 38). It has a small variation in the mixing ratio and enhancements can easily be excluded. It illustrates the average shape of the distribution of the measured mixing ratio for each gas and the effect of fitting a new distribution that excludes the mixing ratio enhancements at a given limit (Table 15). F04 had a noticeably large dispersion of mixing ratio values compared to the other flights. The background value is more ambiguous for this flight and is described as the ‘worst case’ for estimating the value without an independent sample to determine the background. Assessing the screen plot (Figure 37) as well as the distribution of values was imperative for choosing a reasonable estimate as the distribution alone could not be relied on to estimate the average background. Figure 39 plots the F04 raw CH₄ values. Due to the large dispersion of mixing ratio values and the change in emission estimates given this range, fitting a surface extrapolation using a background value was discouraged. The sensitivity testing range for both flights included values chosen to span approximately the 1st percentile of the raw mixing ratio measurements to 90th, to assess sensitivity a large range of background values considered for the final analysis. For flights with a typical background value distribution, the emissions estimates were not sensitive to the choice in background value.

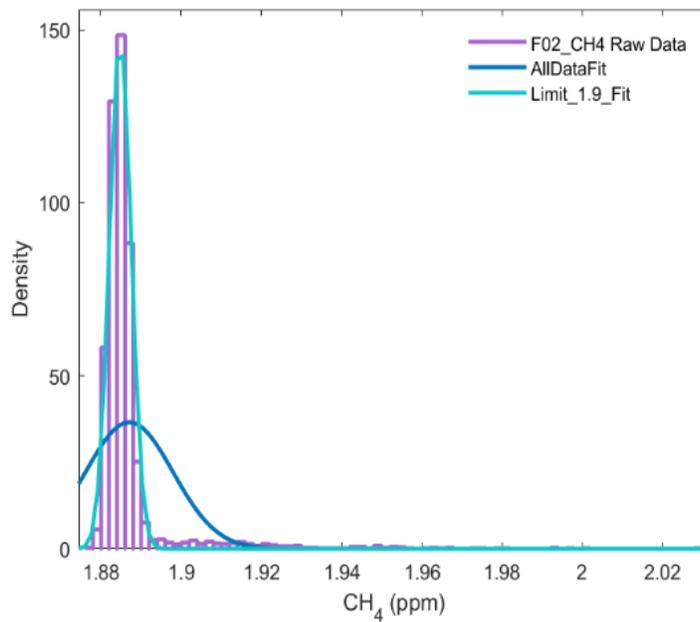


Figure 36. The histogram of F02 CH₄ raw data collected every second during the flight (purple). Normal curves that were calculated using the whole set of values (dark blue) and a subset of the data excluding values larger than 1.9 (light blue) are overlaid.

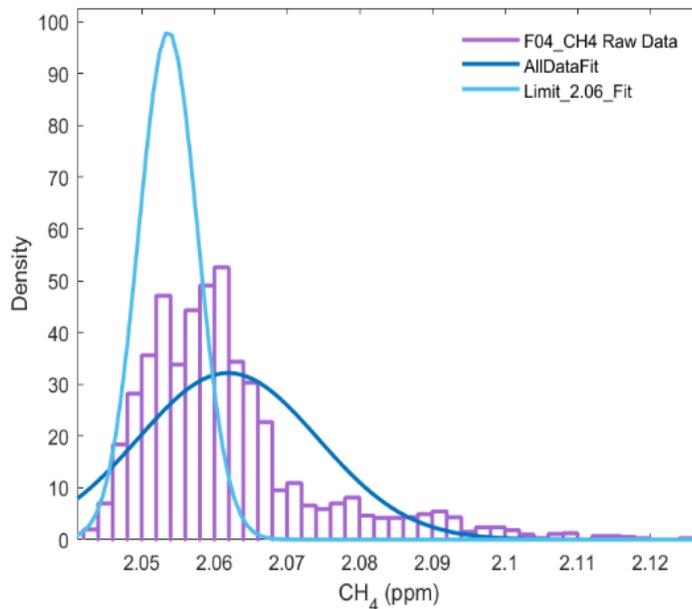


Figure 38. The histogram of F04 CH₄ raw data collected every second during the flight (purple). Normal curves that were calculated using the whole set of values (dark blue) and a subset of the data excluding values larger than 2.06 (light blue) are overlaid.

The percentage sensitivity for flights F02 and F04 for both CH₄ and CO₂ are given in Table 16. The percentage change was calculated between the minimum and maximum emission estimate calculated using the span of the 1st mixing ratio percentile to the 90th to assess the maximum error that could occur. The error when choosing a background value would be less than the min to max calculated values. F04 CH₄ was assessed three times due to its unusual distribution of values (Figure 39). The first percentage change was calculated for the span in percentiles, the second for a set excluding enhancements that were still evident (2.045 to 2.06 ppm), and the third given a more sensible set (2.049 to 2.055 ppm), that would approximate the error when choosing the value rather than an approximate of the maximum. The emission estimates produced by using varying background values that were used to calculate the percentage change for F02 and F04 are given in Tables 17-20. Having to choose the background value for the TERRA model surface extrapolation without independent samples was not a significant source of error for the comparison analysis.

Table 16. Percentage change in the emission estimate when using the min and max background values for the two surface extrapolations that require setting a background value in TERRA.

Flight and Gas	Percentage change: “Background and “Background to Linear”
F02 CH ₄	1.40 % and 0.67 %
F02 CO ₂	2.40 % and 1.17 %
F04 CH ₄ : fit 1	40.3 % and 19.1 %
F04 CH ₄ : fit 2	16.0 % and 7.6 %
F04 CH ₄ : fit 3	4.22 % and 1.89 %
F04 CO ₂	10% and 4.6 %

Table 17. F02 CH₄ (kg hr⁻¹) emissions estimates for each surface extrapolation fit given varying background mixing ratio values (ppm).

	1.881	1.883	1.884	1.885	1.886	1.887
Background	365.710	367.434	368.296	369.158	370.021	370.883
B. to Linear	393.585	394.468	394.909	395.351	395.792	396.234

Table 18. F02 CO₂ (kg hr⁻¹) emissions estimates for each surface extrapolation fit given varying background mixing ratio values (ppm).

	379	380	382	383	384
Background	477132	479497	484228	486593	488958
B. to Linear	511313	51524	514947	516158	517369

Table 19. F04 CH₄ (kg hr⁻¹) emissions estimates for each surface extrapolation fit given varying background mixing ratio values (ppm).

	2.045	2.049	2.05	2.05365	2.0537	2.054	2.055
Background	137.66	128.79	126.57	118.48	118.37	117.70	115.48
B. to Linear	128.88	124.93	123.95	120.35	120.30	120.00	119.02

	2.05796	2.0595	2.06	2.0605	2.061	2.065	2.07
Background	108.92	105.50	104.40	103.29	102.18	93.31	82.22
B. to Linear	116.10	114.58	114.09	113.60	113.10	109.16	104.23

Table 20. F02 CO₂ (kg hr⁻¹) emissions estimates for each surface extrapolation fit given varying background mixing ratio values (ppm).

	395	400	405	406	406.6	406.65	408	410
Background	602807	576204	545777	535704	536040	535736	527520	515349
B. to Linear	594083	584490	570964	564270	566636	566501	562849	557438

Appendix C: 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 21. The lap bin bounds also provided by Scientific Aviation are given in Table 22. Together these data can be used to recreate the SciAv profiles and emission calculations for each flight.

Table 21. The average divergence and uncertainty for each flight lap.

Flight	Lap Number	Altitude (m)	CH ₄ Average Divergence	CH ₄ Uncertainty	CO ₂ Average 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 22. 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

Section 1.1: Fitting the SciAv Surface Extrapolation

The following section provides the Scientific Aviation (SciAv) model surface extrapolation for all five different fits: Constant, Linear, Interpolate, Background, and Linear. The model'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. For each surface extrapolation the lowest bin was extrapolated to the surface using the given fit. The final estimate was determined by combining the emissions calculated from the binned profile added to the trapezoidal integration of area between the extrapolation fit, the surface, and a flux enhancement of zero (Chapter 2, Figure 10). The estimates produced by fitting the different surface extrapolations to the SciAv model that are shown in Chapter 2, Figure 13A - 13E are given in Table 21.

Table 23. SciAv method estimates.

Flight	Model	Gas	Method	Extrapolation	Emission (kg hr ⁻¹)
F01	SciAv	CH ₄	Binned	Constant	3842.1
F01	SciAv	CH ₄	Binned	Linear	5268.7
F01	SciAv	CH ₄	Binned	Interpolate	3899.6
F01	SciAv	CH ₄	Binned	Background	3046.7
F01	SciAv	CH ₄	Binned	Linear Weighted	4686.3
F01	SciAv	CH ₄	Binned	Averaged	4195.3
F01	SciAv	CO ₂	Binned	Constant	1036685
F01	SciAv	CO ₂	Binned	Linear	1248005
F01	SciAv	CO ₂	Binned	Interpolate	1043900
F01	SciAv	CO ₂	Binned	Background	783997
F01	SciAv	CO ₂	Binned	Linear Weighted	1304495
F01	SciAv	CO ₂	Binned	Averaged	1151884
F02	SciAv	CH ₄	Binned	Constant	361.8
F02	SciAv	CH ₄	Binned	Background	306.3
F02	SciAv	CH ₄	Binned	Linear	359.6
F02	SciAv	CH ₄	Binned	Interpolate	331.4
F02	SciAv	CH ₄	Binned	Linear Weighted	358.7
F02	SciAv	CH ₄	Binned	Averaged	360.9
F02	SciAv	CO ₂	Binned	Constant	563238
F02	SciAv	CO ₂	Binned	Background	488868
F02	SciAv	CO ₂	Binned	Linear	556080
F02	SciAv	CO ₂	Binned	Interpolate	522777
F02	SciAv	CO ₂	Binned	Linear Weighted	563078
F02	SciAv	CO ₂	Binned	Averaged	561828
F03	SciAv	CH ₄	Binned	Constant	496.9
F03	SciAv	CH ₄	Binned	Background	399.9
F03	SciAv	CH ₄	Binned	Linear	505.5
F03	SciAv	CH ₄	Binned	Interpolate	452.0

F03	SciAv	CH ₄	Binned	Linear Weighted	503.6
F03	SciAv	CH ₄	Binned	Averaged	503.2
F03	SciAv	CO ₂	Binned	Constant	525773
F03	SciAv	CO ₂	Binned	Background	401449
F03	SciAv	CO ₂	Binned	Linear	553281
F03	SciAv	CO ₂	Binned	Interpolate	479194
F03	SciAv	CO ₂	Binned	Linear Weighted	571394
F03	SciAv	CO ₂	Binned	Averaged	547970
F04	SciAv	CH ₄	Binned	Constant	348.8
F04	SciAv	CH ₄	Binned	Linear	332.4
F04	SciAv	CH ₄	Binned	Interpolate	312.9
F04	SciAv	CH ₄	Binned	Background	294.5
F04	SciAv	CH ₄	Binned	Linear Weighted	332.1
F04	SciAv	CH ₄	Binned	Averaged	340.7
F04	SciAv	CO ₂	Binned	Constant	1171101
F04	SciAv	CO ₂	Binned	Background	1017928
F04	SciAv	CO ₂	Binned	Linear	1089963
F04	SciAv	CO ₂	Binned	Interpolate	1048606
F04	SciAv	CO ₂	Binned	Linear Weighted	1110325
F04	SciAv	CO ₂	Binned	Averaged	1139170
F05	SciAv	CH ₄	Binned	Constant	3473.7
F05	SciAv	CH ₄	Binned	Linear	3711.4
F05	SciAv	CH ₄	Binned	Interpolate	3033.4
F05	SciAv	CH ₄	Binned	Background	2413.0
F05	SciAv	CH ₄	Binned	Linear Weighted	3672.1
F05	SciAv	CH ₄	Binned	Averaged	3589.1
F05	SciAv	CO ₂	Binned	Constant	849712
F05	SciAv	CO ₂	Binned	Linear	782514
F05	SciAv	CO ₂	Binned	Interpolate	750608
F05	SciAv	CO ₂	Binned	Background	661413
F05	SciAv	CO ₂	Binned	Linear Weighted	781836
F05	SciAv	CO ₂	Binned	Averaged	806397

Section 1.2: Change in SciAv Fitting Height

To estimate different surface extrapolations from the SciAv profiles, a height must be selected to fit the extrapolation from the divergence points to the surface. 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. Type I and II profiles can be fit nearer to TERRA's 300m profile with the aim to avoid fitting the enhancement whilst retaining the largest number of profile points to fit from. The type III profile is better suited to fitting over the whole sampling height to obtain a more robust estimate. The chosen fitting heights were: 1100m for F01, 400m for F02, 250m for

F03, 400m for F04, and 600m for F05. The extrapolation height was chosen by the density of points, (at least three points were required to fit), the profile shape (enhancements were avoided in the fittings), and the minimum flight height. To assess the stability of the chosen height for fitting the SciAv profile extrapolation a minimum height, and maximum height were fit to compare with the height chosen (Table 24). Estimates for the background, constant, and interpolate to background fits were not affected by a differing extrapolation height. Results are given in Table 25.

Table 24. Fitting heights used to test the stability of the surface extrapolation of each flight.

Flight	Height Used	Min Height	Max Height
F01	1100	400	1100
F02	400	300	1300
F03	250	175	800
F04	400	225	1100
F05	600	300	600

Table 25. SciAv model emissions estimates in kg hr⁻¹ given the heights in Table 24.

Flight	Gas	Method	Extrap	Height Used	Min Height	Max Height
F01	CH ₄	Binned	Linear	5269	5353	5269
F01	CH ₄	Binned	LinearW	4686	5391	4686
F01	CH ₄	Binned	Averaged	4195	5193	4195
F01	CH ₄	Trapz	Linear	4596	5353	4596
F01	CH ₄	Trapz	LinearW	4770	5391	4770
F01	CH ₄	Trapz	Averaged	4734	5193	4734
F01	CO ₂	Binned	Linear	1248005	1469447	1248005
F01	CO ₂	Binned	LinearW	1304495	1461100	1304495
F01	CO ₂	Binned	Averaged	1151885	1277900	1151885
F01	CO ₂	Trapz	Linear	1294681	1516122	1294681
F01	CO ₂	Trapz	LinearW	1351171	1507776	1351171
F01	CO ₂	Trapz	Averaged	1332110	1458125	1332110
F02	CH ₄	Binned	Linear	360	330	381
F02	CH ₄	Binned	LinearW	359	330	379
F02	CH ₄	Binned	Averaged	361	341	375
F02	CH ₄	Trapz	Linear	355	325	376
F02	CH ₄	Trapz	LinearW	354	325	375
F02	CH ₄	Trapz	Averaged	354	334	368
F02	CO ₂	Binned	Linear	556080	489620	589286

F02	CO ₂	Binned	LinearW	563078	489620	602214
F02	CO ₂	Binned	Averaged	561828	515188	585942
F02	CO ₂	Trapz	Linear	549825	483365	583031
F02	CO ₂	Trapz	LinearW	556824	483365	595960
F02	CO ₂	Trapz	Averaged	552360	505721	576474
F03	CH ₄	Binned	Linear	506	592	511
F03	CH ₄	Binned	LinearW	504	592	488
F03	CH ₄	Binned	Averaged	503	562	500
F03	CH ₄	Trapz	Linear	532	619	538
F03	CH ₄	Trapz	LinearW	530	619	514
F03	CH ₄	Trapz	Averaged	531	589	527
F03	CO ₂	Binned	Linear	553281	903011	574069
F03	CO ₂	Binned	LinearW	571394	903011	571686
F03	CO ₂	Binned	Averaged	547970	775085	554997
F03	CO ₂	Trapz	Linear	526782	876511	547569
F03	CO ₂	Trapz	LinearW	544895	876511	545187
F03	CO ₂	Trapz	Averaged	534039	761154	541065
F04	CH ₄	Binned	Linear	332	305	348
F04	CH ₄	Binned	LinearW	332	304	343
F04	CH ₄	Binned	Averaged	341	322	349
F04	CH ₄	Trapz	Linear	339	312	355
F04	CH ₄	Trapz	LinearW	339	311	350
F04	CH ₄	Trapz	Averaged	339	320	347
F04	CO ₂	Binned	Linear	1089963	1083590	1151001
F04	CO ₂	Binned	LinearW	1110325	1095482	1147692
F04	CO ₂	Binned	Averaged	1139170	1132098	1171972
F04	CO ₂	Trapz	Linear	1130430	1124057	1191468
F04	CO ₂	Trapz	LinearW	1157880	1135949	1188159
F04	CO ₂	Trapz	Averaged	1133658	1126586	1166460
F05	CH ₄	Binned	Linear	3711	4300	3601
F05	CH ₄	Binned	LinearW	3672	4300	3628
F05	CH ₄	Binned	Averaged	3589	3995	3538
F05	CH ₄	Trapz	Linear	3833	4422	3723
F05	CH ₄	Trapz	LinearW	3794	4422	3750
F05	CH ₄	Trapz	Averaged	3801	4207	3750
F05	CO ₂	Binned	Linear	782514	958552	795135
F05	CO ₂	Binned	LinearW	781836	958552	799407
F05	CO ₂	Binned	Averaged	806397	923982	816461
F05	CO ₂	Trapz	Linear	739880	915918	752501
F05	CO ₂	Trapz	LinearW	739202	915918	756773
F05	CO ₂	Trapz	Averaged	758750	876335	768815

Flight profiles for F01 through F05 and shown in Chapter 2, Figure 13 A-E and the profile shapes discussed and illustrated in Chapter 2 Figure 12. Flight profiles are described as stable if the emission estimates have low variation when the extrapolation fit type, or extrapolation heights are changed.

The profile shapes and best fit were investigated. The F01 type III profile shape is hypothesized to benefit from not following SciAv's methods of applying a constant extrapolation. By applying a linear weighted extrapolation and fitting over the whole profile the behaviour of the plume can be captured. Due to the small number of laps flown, fitting over the whole profile avoids the instability introduced when just the lowest profile points are used for extrapolation. F02 extrapolations were stable with very little difference in estimates. Due to type II shape of F02 fits that followed the behaviour of the plume may have been more appropriate, however the lower profile had some emission spikes requiring more information about the source to assume the divergence is approaching zero near the surface. As such a profile height was fit to maximize fitting along the points below the enhancement rather than tailor the height to profile behaviour. F03 had differing profile behaviour for each gas. The emission estimates from the profile for CH₄ was stable for both the height below the enhancement and extrapolation chosen. The profile of F03 CO₂ is unstable at the lower flight paths due to the highly variable divergence points. Both F03 samples are well suited to the SciAv method of fitting a constant extrapolation. By fitting a constant extrapolation F03 CH₄ avoids an arbitrary choice in the extrapolation type and F03 CO₂ avoids assuming an appropriate extrapolation height. F04 provides a unique profile type due to the differing profiles for the flight up compared to down. Averaging over a larger number of points was necessary to avoid fitting the profile spikes between the divergence points. F04 had very little difference in the extrapolation type chosen and the extrapolation height would have likely also been stable had there not been a dichotomy in profile points. F05 is a sample defined by the lack of plume capture and ambiguous plume behaviour. The extrapolation height was set to fit over the whole profile to avoid the instability of the highly variable divergence point profile shapes (Chapter 2, Figure 13D).

This sensitivity analysis of profile fit, and height indicates that the SciAv method of fitting a constant extrapolation is on average the best approach. When a profile has a prominent type III profile shape where emission enhancements are largely at the bottom of the profile and

are increasing towards the surface a differing extrapolation type might be more appropriate. A linear weighted extrapolation fit at an extrapolation height over the whole profile would capture the behaviour of the emission plume, account for the uncertainty in the laps, but provide a conservative assumption about the extent of the increasing surface emissions.

Section 1.3 Scientific Aviation Trapezoidal Integration Estimates

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.

Emissions estimates using the trapezoidal integration are given in Table 26. Figures 40A - 40E 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 40A). 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.

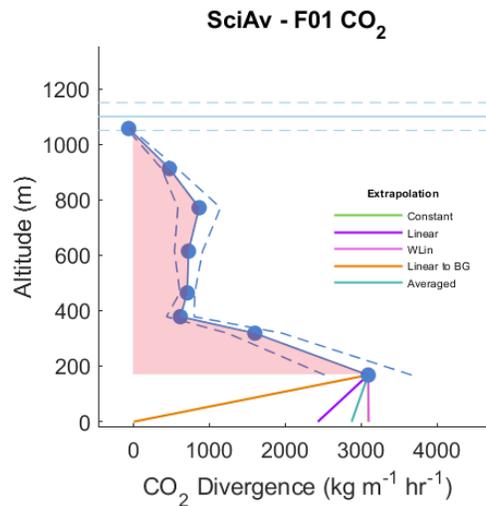
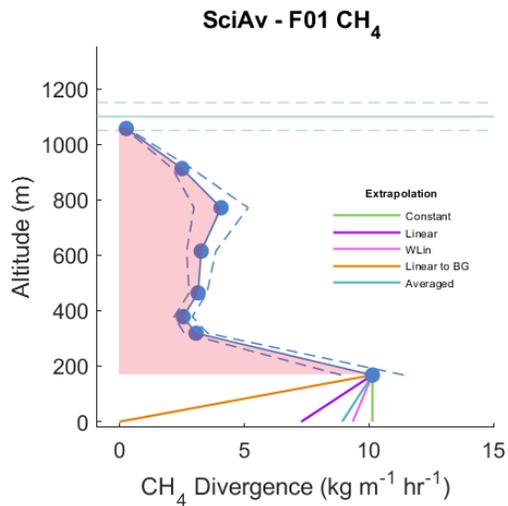


Figure 40A.

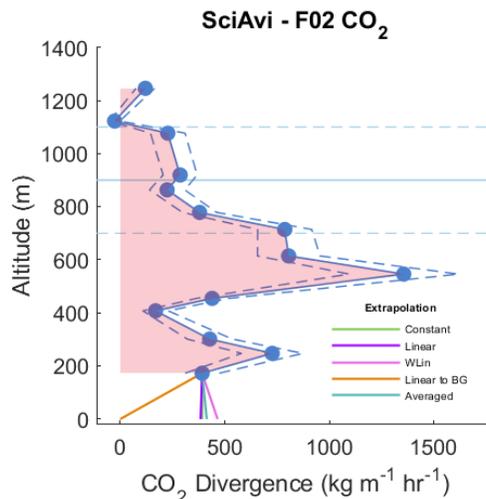
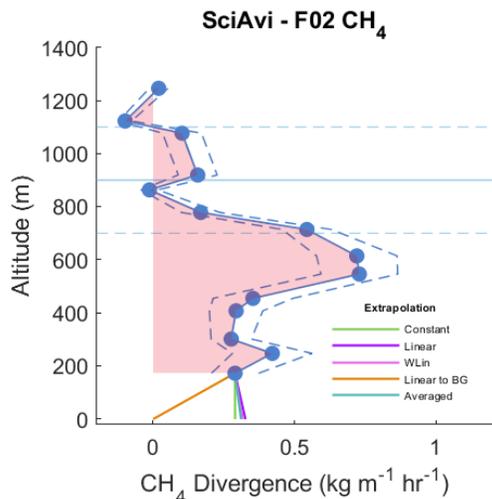


Figure 40B.

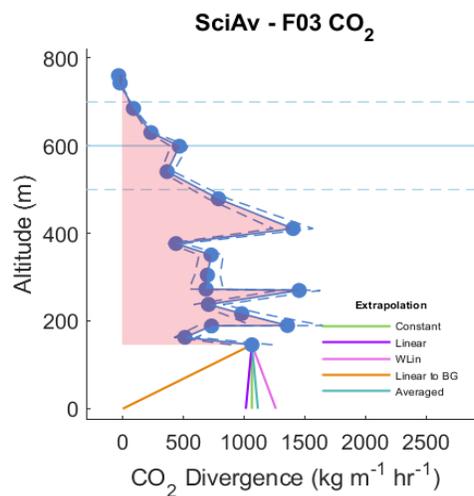
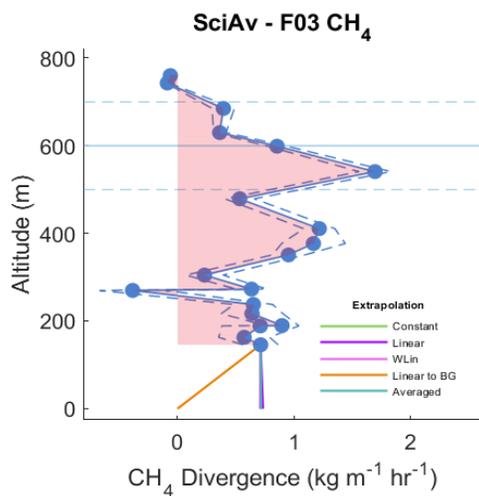


Figure 40C.

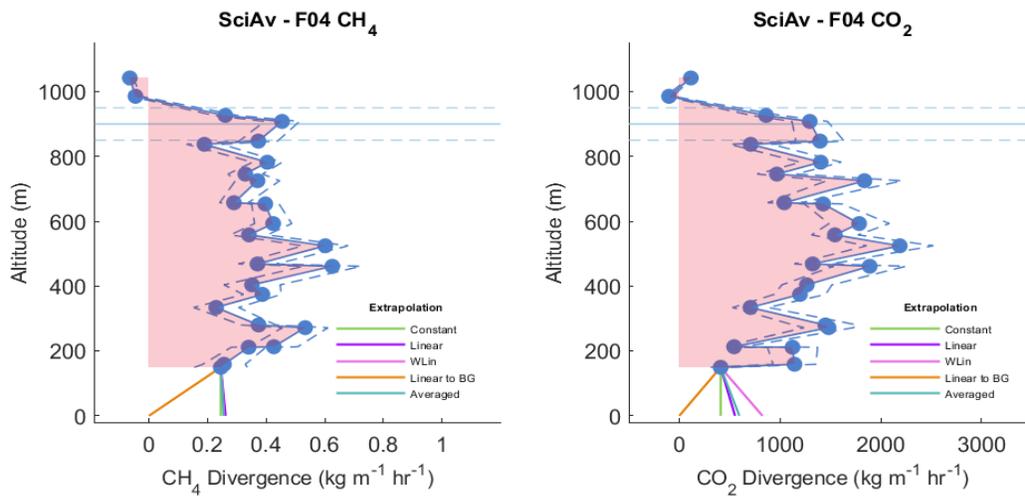


Figure 40D.

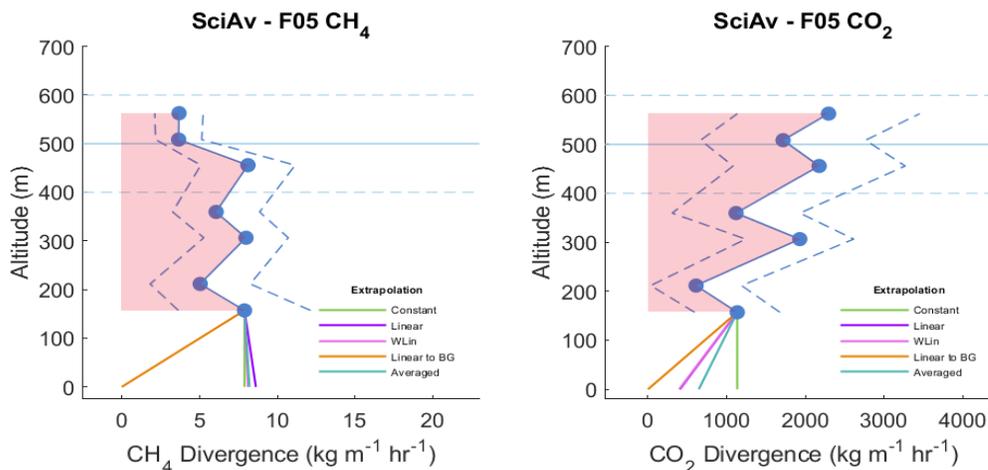


Figure 40E.

Figure 40. A - E of the SciAv profiles for flights F01 through F05. The blue points are the estimated 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.

Table 26. Emission estimates from an adapted SciAv method. Trapezoidal integration is used to estimate emissions over the profile rather than the binning method.

Flight	Model	Gas	Method	Extrapolation	Emission (kg/hr)
F01	SciAv	CH ₄	Trapezoid	Constant	4836.4
F01	SciAv	CH ₄	Trapezoid	Background	3130.6
F01	SciAv	CH ₄	Trapezoid	Linear	4596.4
F01	SciAv	CH ₄	Trapezoid	Interpolate	3983.5
F01	SciAv	CH ₄	Trapezoid	Linear Weighted	4770.2
F01	SciAv	CH ₄	Trapezoid	Averaged	4734.3
F01	SciAv	CO ₂	Trapezoid	Constant	1350478
F01	SciAv	CO ₂	Trapezoid	Background	830672
F01	SciAv	CO ₂	Trapezoid	Linear	1294681
F01	SciAv	CO ₂	Trapezoid	Interpolate	1090575
F01	SciAv	CO ₂	Trapezoid	Linear Weighted	1351171
F01	SciAv	CO ₂	Trapezoid	Averaged	1332110
F02	SciAv	CH ₄	Trapezoid	Constant	351.9
F02	SciAv	CH ₄	Trapezoid	Background	301.8
F02	SciAv	CH ₄	Trapezoid	Linear	355.0
F02	SciAv	CH ₄	Trapezoid	Interpolate	326.8
F02	SciAv	CH ₄	Trapezoid	Linear Weighted	354.2
F02	SciAv	CH ₄	Trapezoid	Averaged	353.7
F02	SciAv	CO ₂	Trapezoid	Constant	550432
F02	SciAv	CO ₂	Trapezoid	Background	482614
F02	SciAv	CO ₂	Trapezoid	Linear	549825
F02	SciAv	CO ₂	Trapezoid	Interpolate	516523
F02	SciAv	CO ₂	Trapezoid	Linear Weighted	556824
F02	SciAv	CO ₂	Trapezoid	Averaged	552360
F03	SciAv	CH ₄	Trapezoid	Constant	530.4
F03	SciAv	CH ₄	Trapezoid	Background	426.3
F03	SciAv	CH ₄	Trapezoid	Linear	531.8
F03	SciAv	CH ₄	Trapezoid	Interpolate	478.3
F03	SciAv	CH ₄	Trapezoid	Linear Weighted	530.0
F03	SciAv	CH ₄	Trapezoid	Averaged	530.7
F03	SciAv	CO ₂	Trapezoid	Constant	530440
F03	SciAv	CO ₂	Trapezoid	Background	374950
F03	SciAv	CO ₂	Trapezoid	Linear	526782
F03	SciAv	CO ₂	Trapezoid	Interpolate	452695
F03	SciAv	CO ₂	Trapezoid	Linear Weighted	544895
F03	SciAv	CO ₂	Trapezoid	Averaged	534039
F04	SciAv	CH ₄	Trapezoid	Constant	338.1
F04	SciAv	CH ₄	Trapezoid	Background	301.4

F04	SciAv	CH ₄	Trapezoid	Linear	339.3
F04	SciAv	CH ₄	Trapezoid	Interpolate	319.7
F04	SciAv	CH ₄	Trapezoid	Linear Weighted	338.9
F04	SciAv	CH ₄	Trapezoid	Averaged	338.8
F04	SciAv	CO ₂	Trapezoid	Constant	1119751
F04	SciAv	CO ₂	Trapezoid	Background	1058396
F04	SciAv	CO ₂	Trapezoid	Linear	1130430
F04	SciAv	CO ₂	Trapezoid	Interpolate	1089074
F04	SciAv	CO ₂	Trapezoid	Linear Weighted	1150792
F04	SciAv	CO ₂	Trapezoid	Averaged	1133658
F05	SciAv	CH ₄	Trapezoid	Constant	3775.9
F05	SciAv	CH ₄	Trapezoid	Background	2534.9
F05	SciAv	CH ₄	Trapezoid	Linear	3833.4
F05	SciAv	CH ₄	Trapezoid	Interpolate	3155.4
F05	SciAv	CH ₄	Trapezoid	Linear Weighted	3794.1
F05	SciAv	CH ₄	Trapezoid	Averaged	3801.1
F05	SciAv	CO ₂	Trapezoid	Constant	797169.3
F05	SciAv	CO ₂	Trapezoid	Background	618778.9
F05	SciAv	CO ₂	Trapezoid	Linear	739880
F05	SciAv	CO ₂	Trapezoid	Interpolate	707974
F05	SciAv	CO ₂	Trapezoid	Linear Weighted	739202
F05	SciAv	CO ₂	Trapezoid	Averaged	758750

Table 27 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 28 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.

Table 27. 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)	1735.372	1812.275	581.476
CO ₂ (kg hr ⁻¹)	274004.9	314857.1	138748.7

Table 28. 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.

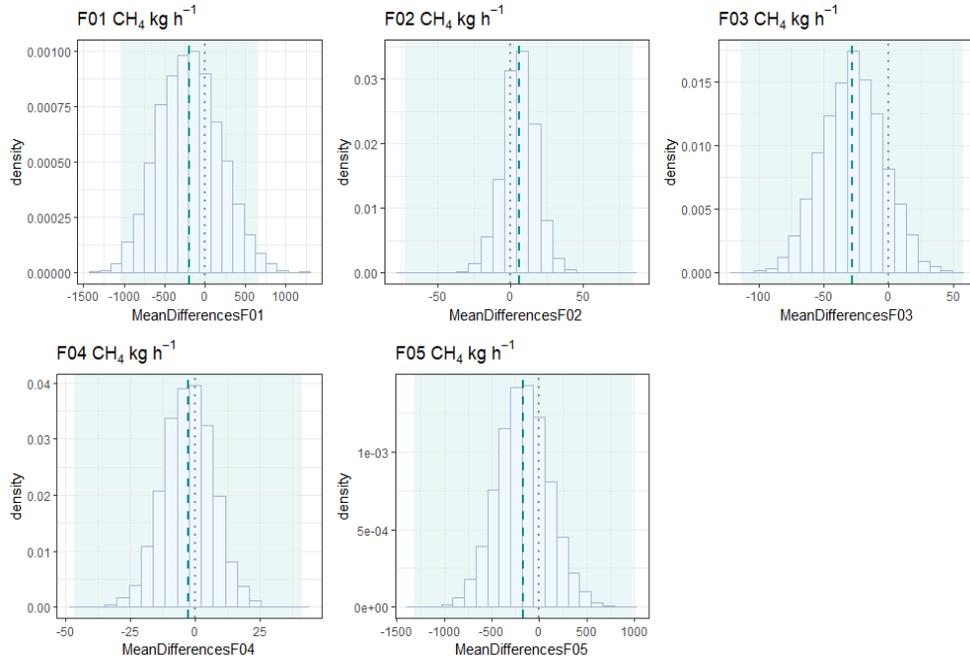
	Weighted t-test: p-value	Weighted t-test: t-value	Wilcoxon signed- rank test: p-value	Wilcoxon signed- rank test: V
CH ₄	0.9517	-0.062	0.1875	2
CO ₂	0.9524	-0.062	1	8

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:

$$\text{difference} = \text{mean}(\text{SciAvEstimate}_{\text{Binned}}) - \text{mean}(\text{SciAvEstimate}_{\text{Trapz}}) \quad (8)$$

As shown in Figure 41, 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.

Bootstrap Distribution of Mean Differences Between Scientific Aviation Integration Methods



Bootstrap Distribution of Mean Differences Between Scientific Aviation Integration Methods

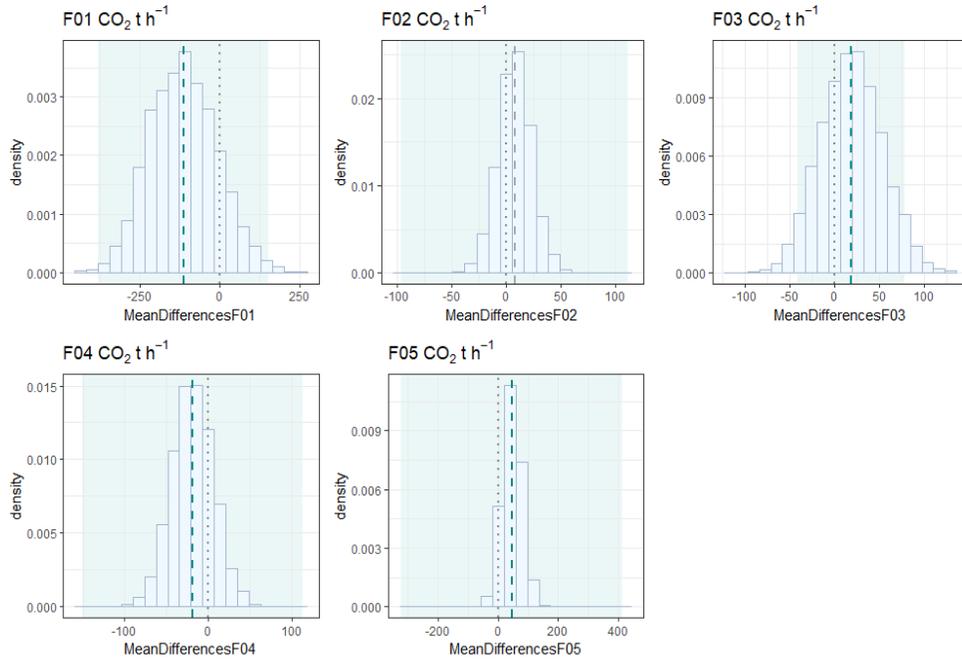


Figure 42. Distributions of the mean difference between all fits of CH₄ (top panel) and CO₂ (bottom panel) given the SciAv model for the binning and trapezoidal integration methods are drawn in light blue. The mean difference between the standard estimates of the models is plotted as a teal dashed lined 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 models.

Appendix D: Comparison of SciAv and TERRA Methods Using Background Surface Extrapolation Fits

To remove the effect of the surface extrapolation the two models were compared when using each model's "background" surface extrapolation fit. This section details the analysis and results. 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 model 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 42. The estimate uncertainties were only calculated for each model'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 model estimates is given in Table 29. Results of parametric (weighted t-test) and non-parametric (Wilcoxon signed-rank test) significance testing of differences between the two box-flight models using the set of all five flights given each the standard fit, background fit, and average of the fits estimates is given in Table 30.

Emission estimates using a background fit are systematically lower than the standard estimates (Figure 42), but no clear pattern is identifiable that would indicate systematic model 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 model agreement. The models 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 model assumptions are met, variation between the model estimates can largely be attributed to different surface extrapolation methods.

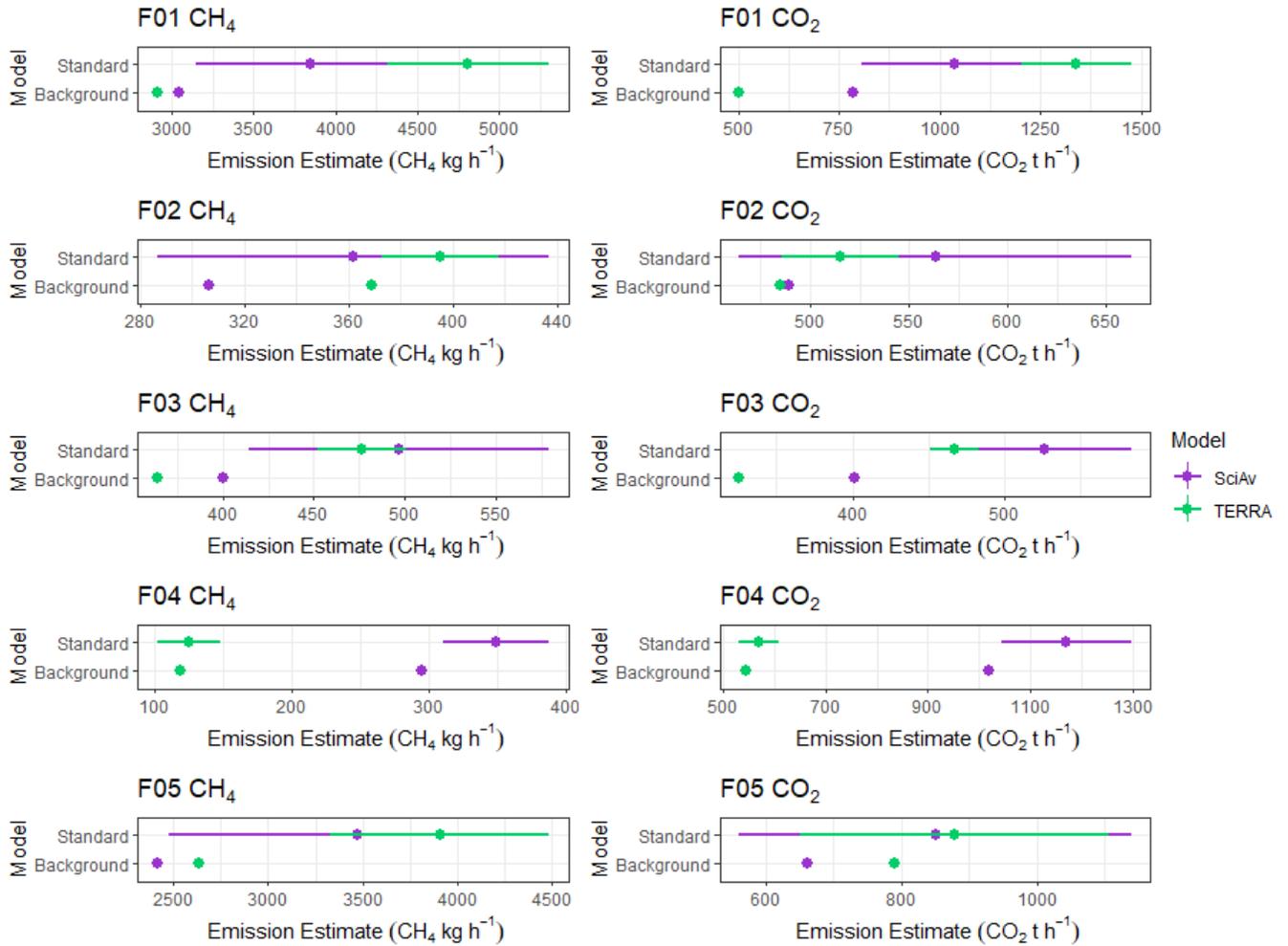


Figure 45. Model estimates for CH₄ and CO₂ given the background mixing ratio fit of no emission plume extrapolation to the surface compared to the standard estimates.

Table 29. Mean estimates and the standard deviation of each model’s surface extrapolation emission estimate. The difference between the mean model estimates is calculated as SciAv estimate – TERRA estimate.

	SciAv’s Six Surface Extrapolations Mean Estimate	SciAv’s Six Surface Extrapolations Standard Deviation	TERRA’s Five Surface Extrapolations Mean Estimate	TERRA’s Five Surface Extrapolations Standard Deviation	Difference Between Mean Model Estimates
F01 CH ₄ (kg hr ⁻¹)	4156.5	763.6	4042.5	745.7	114.0
F02 CH ₄ (kg hr ⁻¹)	346.5	22.8	402.9	21.6	-56.4
F03 CH ₄ (kg hr ⁻¹)	476.9	42.8	437.5	47.7	39.4
F04 CH ₄ (kg hr ⁻¹)	326.9	19.9	118.5	8.5	208.4
F05 CH ₄ (kg hr ⁻¹)	3315.5	505.6	3381.7	501.9	-66.2
F01 CO ₂ (kg hr ⁻¹)	1094828	186156	1061833	329776	32995
F02 CO ₂ (kg hr ⁻¹)	542665	30583	524562	24944	18103
F03 CO ₂ (kg hr ⁻¹)	513177	63253	424136	61742	89041
F04 CO ₂ (kg hr ⁻¹)	1096182	56721	572560	18916	523622
F05 CO ₂ (kg hr ⁻¹)	772080	63429	858699	44441	-86619

Table 30. The mean differences between the two model outputs amongst the five flights and results from significance testing were calculated for comparison between the standard, background and average of the models fit.

	CH ₄ Standard	CH ₄ Background	CH ₄ Average	CO ₂ Standard	CO ₂ Background	CO ₂ Average
Mean Difference (kg hr ⁻¹)	-238.5	12.4	47.8	75803	141445	115425
Pairwise t-test: p-value	0.8569	0.9888	0.9687	0.7237	0.3209	0.5248
Pairwise t-test: t-value	-0.1865	0.0145	0.0404	0.3669	1.0673	0.6651
Wilcoxon signed-rank test: p-value	0.4375	1	0.625	0.625	0.3125	0.3125
Wilcoxon signed-rank test: V	4	8	10	10	12	12

Appendix E: QA/QC and Meteorological Conditions Analysis 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 QA/QC 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 F04 non-stationarity.

Prior to this study, Scientific Aviation and Alberta Airshed Stewardship performed a quality assurance and quality control (QA/QC) analysis and inspection 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. A summary of their findings and the decision to accept F04 is shown in Figure 43.

Figure 44 shows the CH₄ and CO₂ mixing ratios values over the length of the flight by altitude. Plotting the raw values did suggest potential non-stationarity of the emission plume. Larger mixing ratio values can be seen on the left side of the figures during the flight up. The cause of the non-stationarity was investigated. Time was measured by Scientific Aviation in EPOCH Time, also called Unix time, as a count of the total seconds since January 1, 1970, at midnight.

Syncrude Plant – 2017/08/14

Syncrude - Plant-2017-08-14/N617DH

- Good winds? **Yes**
- Good Surface Extrapolation? **Maybe (re-do both)**
- Entire plume captured? **Yes (both)**
 - Upgrader plume sampled? **Yes**
 - Highest loop, $E_{CO_2} = 0.1 \text{ t/m/hr}$
 - Highest loop, $E_{CH_4} = 0.00 \text{ kg/m/hr}$
 - Max Altitude/BL = *TBD*
- Negligible upwind flux? **Yes (both)**
- Current Flux Estimates
 - $CO_2 = 1,139 \text{ t/hr}$
 - $CH_4 = 338 \text{ kg/hr}$

ACCEPT FLIGHT

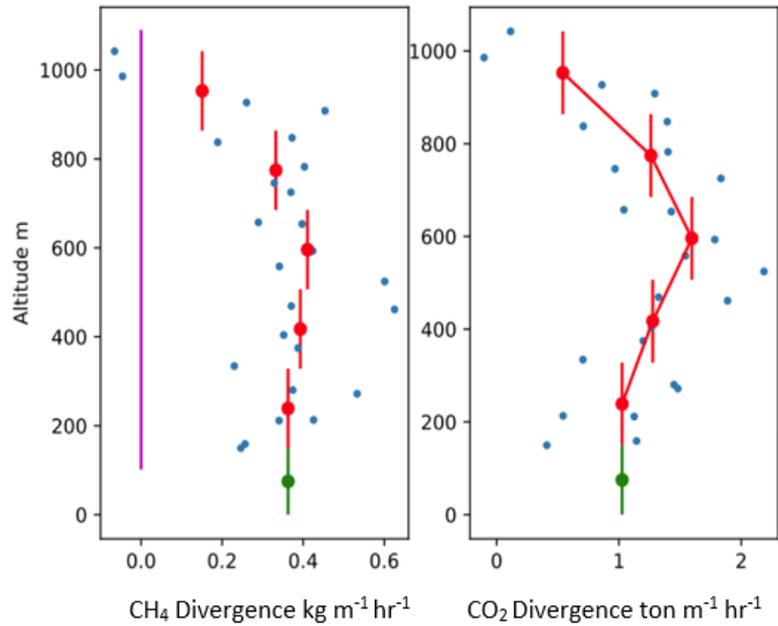


Figure 48. Summary of the preliminary QA/QC of flight F04 provided by the Alberta Environment and Parks and Scientific Aviation.

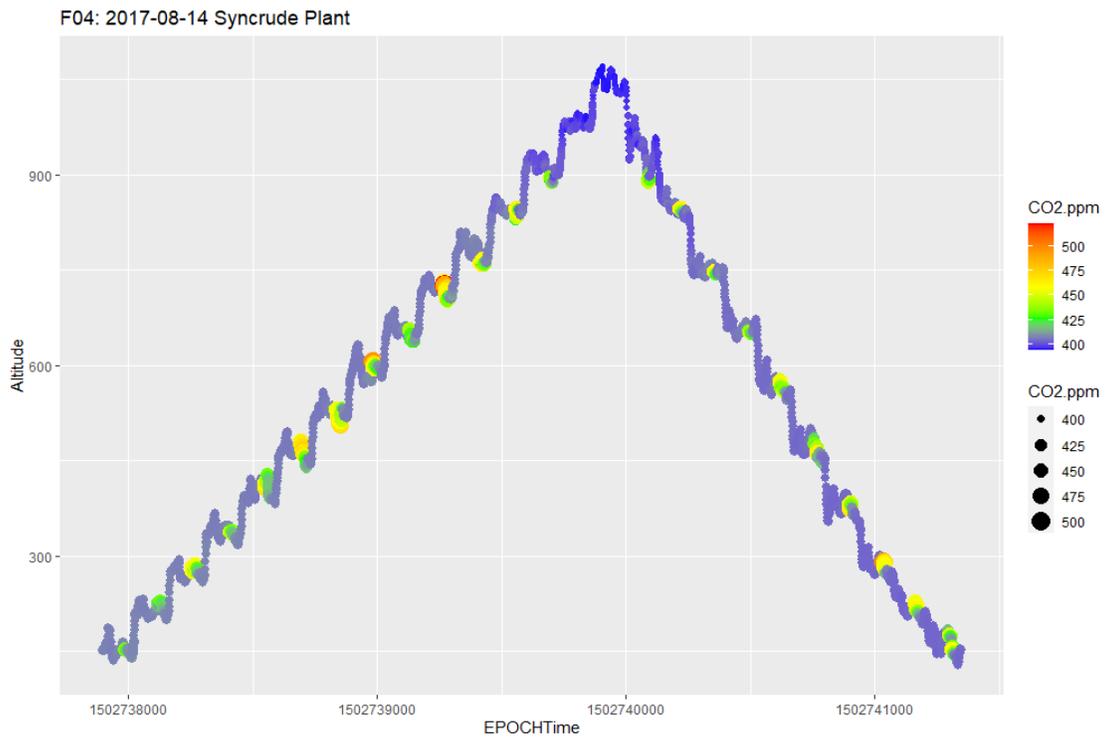
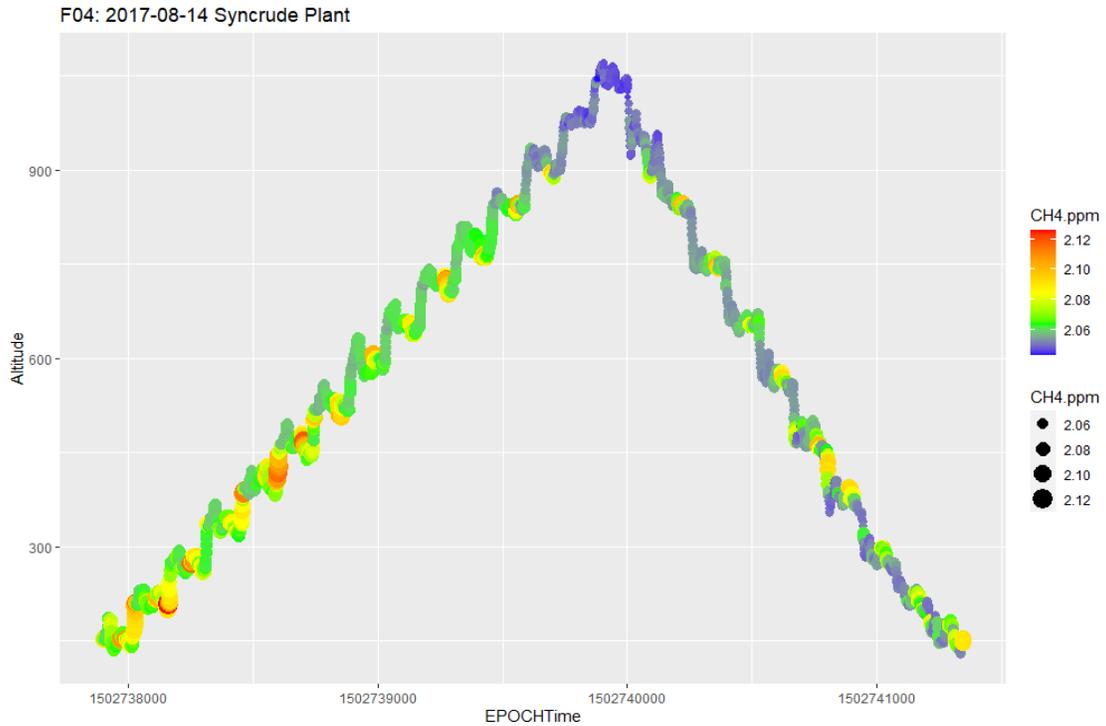


Figure 44. Mixing ratio values of CH₄ (top panel) and CO₂ (bottom panel) for F04 in ppm, plotted by altitude (in m above ground level) over time (EPOCHTime). Values were scaled by size to make enhancements more noticeable.

Meteorological conditions tend to be the most common source of non-stationarity of an emission plume and thus these were thoroughly examined. The wind profiles in the West to East (u - component) and South to North (v – component) wind directions were assessed along with the changes in wind components. Profiles were plotted following practices by Dai et al 2014. The profiles were plotted for the whole flight (Figure 45) and then separated into the flight up and the flight down (Figure 46) to assess any significant changes in their spread, shape, or direction. While there are fewer sample points for the flight down (1358) than the flight up (2092) the shape of the profiles does not noticeably change. This analysis along with the prior QA/QC work indicates that the winds remain stationary during the flight for F04.

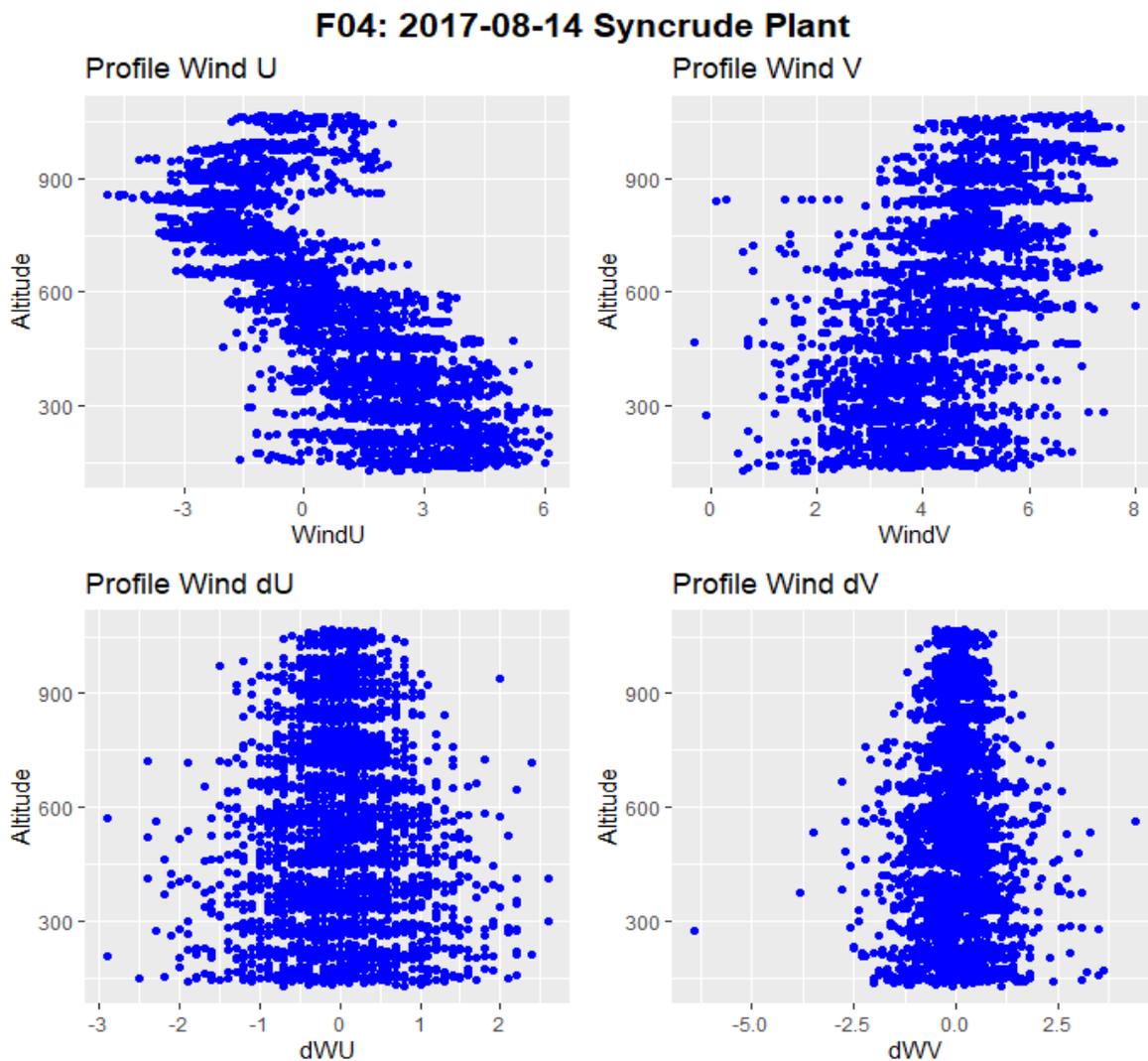
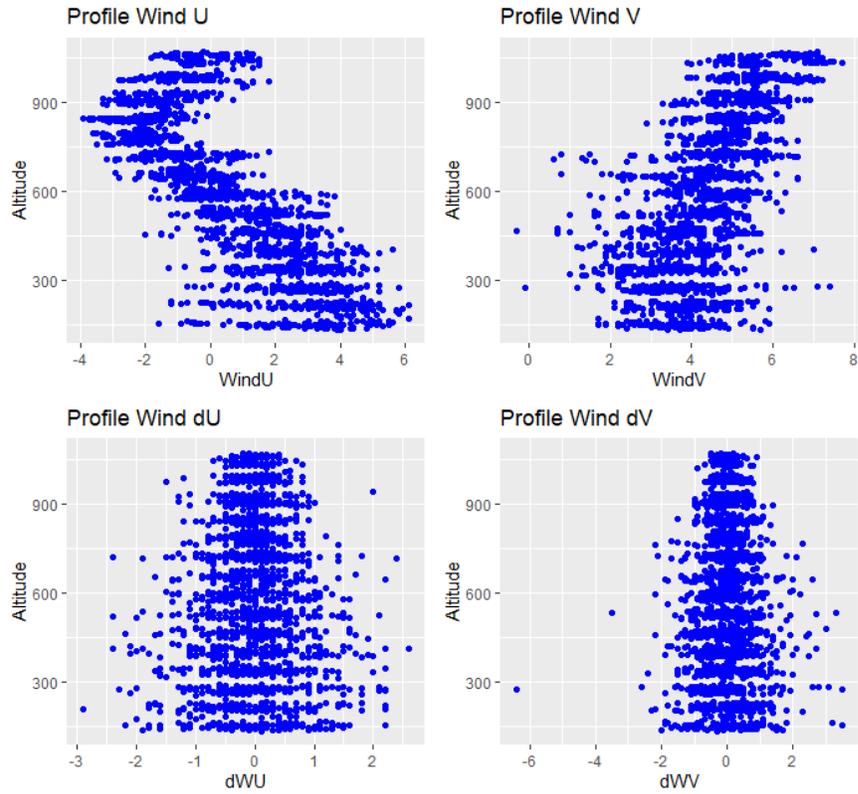


Figure 45. Wind profiles for the u and v components and change in the wind components (d) for the whole F04 flight.

F04: 2017-08-14 Syncrude Plant Up



F04: 2017-08-14 Syncrude Plant Down

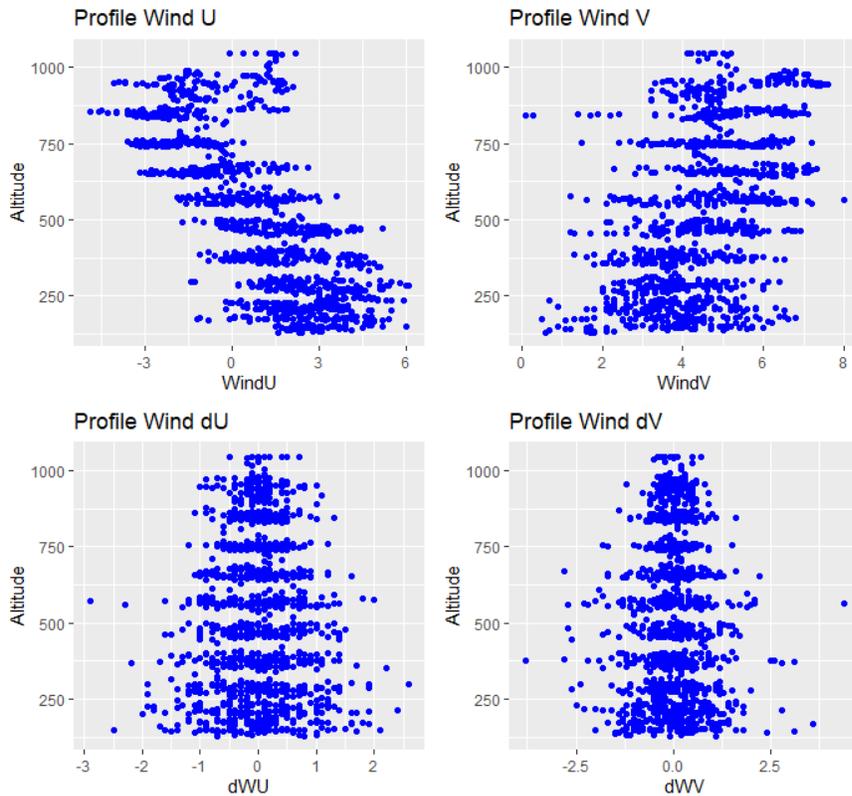


Figure 46. Wind profiles for the u and v components and change in the wind components (d) for the flight up (top panel) and the flight down (bottom panel). A few outliers occur that change the axis of the dV profile the shape.

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 analyzes uses the lap data to produce spatially gridded emission mixing ratios. Figure 47 shows the middle step of gridded screen for F04 CH₄. Some boxes contained a high (red) emission and a low (blue) emission enhancement (Figure 47) 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 the TERRA model, there is a noticeable difference in the CH₄ mixing ratios in the upward flight compared to the downward, with higher mixing ratios (red) detected in the ascending portion of the flight (Figure 48).

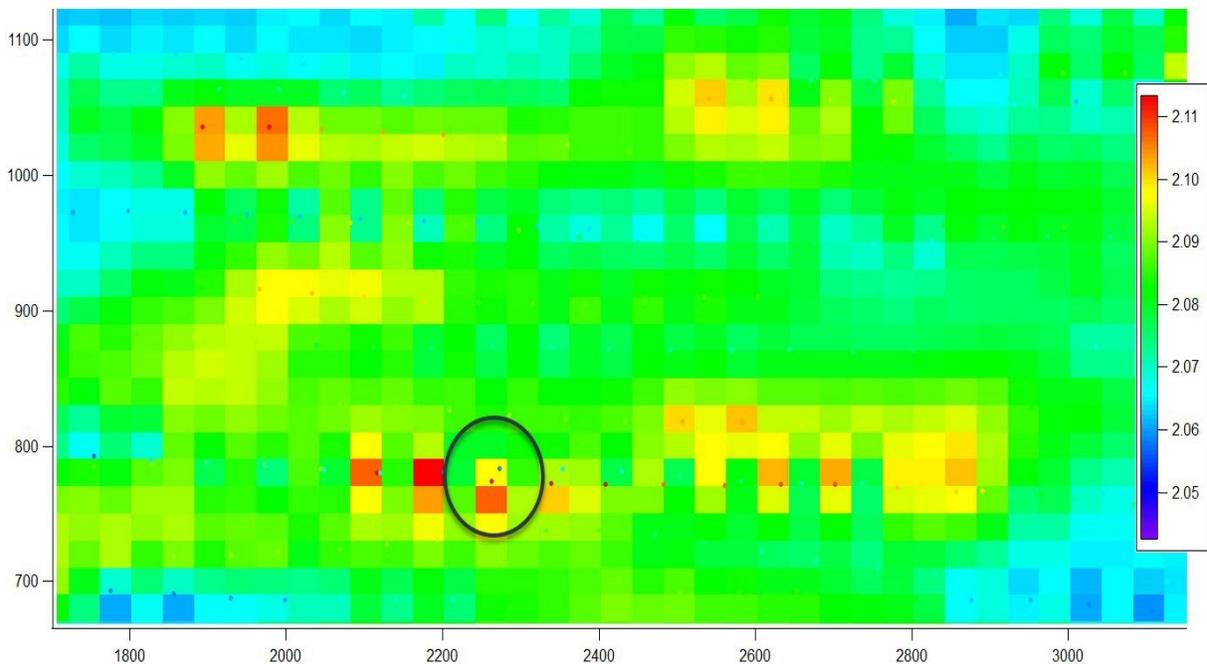


Figure 47. A zoomed in look at the final F04 CH₄ box of kriging-interpolated data. A black circle is drawn around a gridded square that contained a large red enhancement at the same location as a low blue mixing ratio. The square was averaged to a yellow CH₄ value. Imagery provided by Environment and Climate Change Canada.

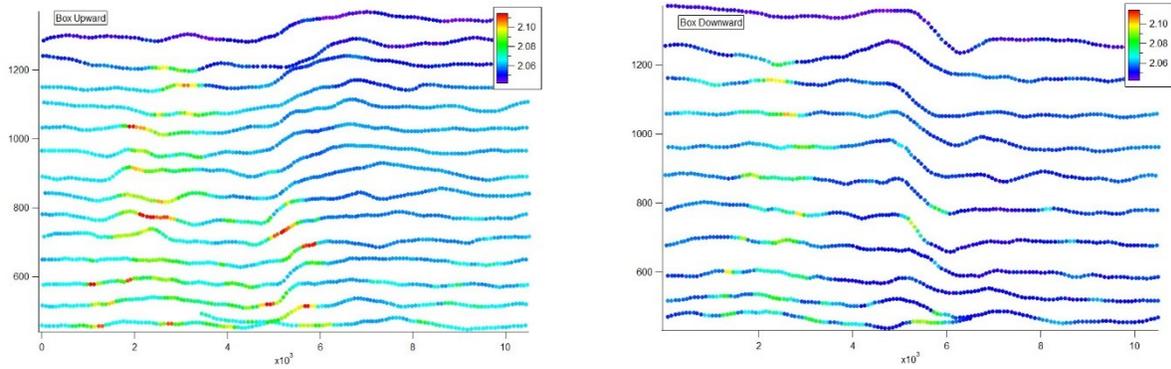


Figure 48. Per lap CH₄ mixing ratio screen data for the box upward (left) and downward (right) prior to kriging of the data. Imagery provided by Environment and Climate Change Canada.

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) six variables were computed: gradient Richardson number, bulk Richardson number, change in virtual temperature, change in potential temperature, and the change in virtual potential temperature. Figure 49 shows the calculated values for each measurement by time and altitude. The profiles for these values over the whole flight (Figure 50) and for the flight up and down (Figure 51) were also inspected. There was no obvious change in the variables used to estimates the boundary layer during the flight.

F04: 2017-08-14 Syncrude Plant

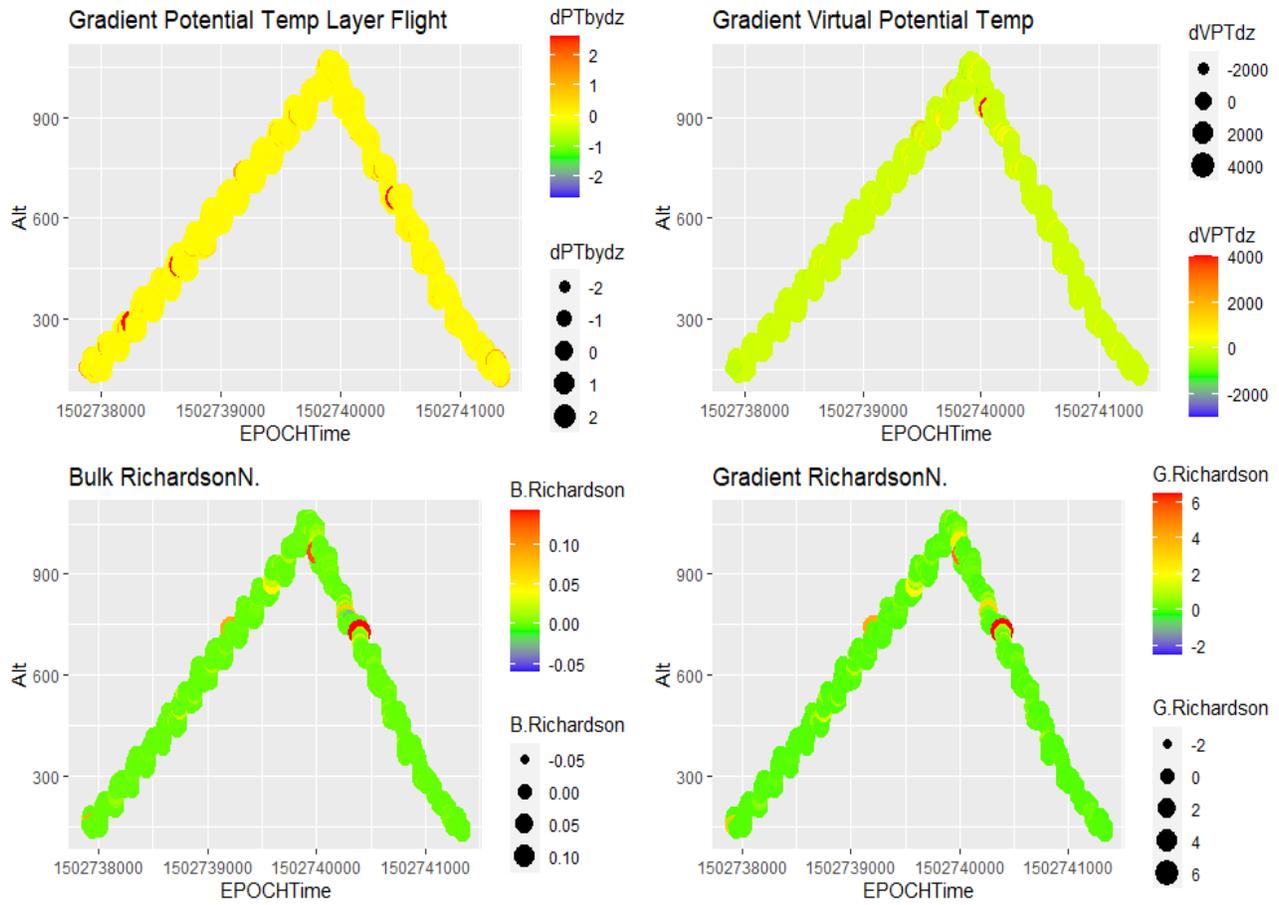


Figure 49. The gradient potential temperature, gradient virtual potential temperature, bulk Richardson number, and gradient Richardson numbers by altitude over time. Values were scaled by size to make enhancements more noticeable.

F04: 2017-08-14 Syncrude Plant

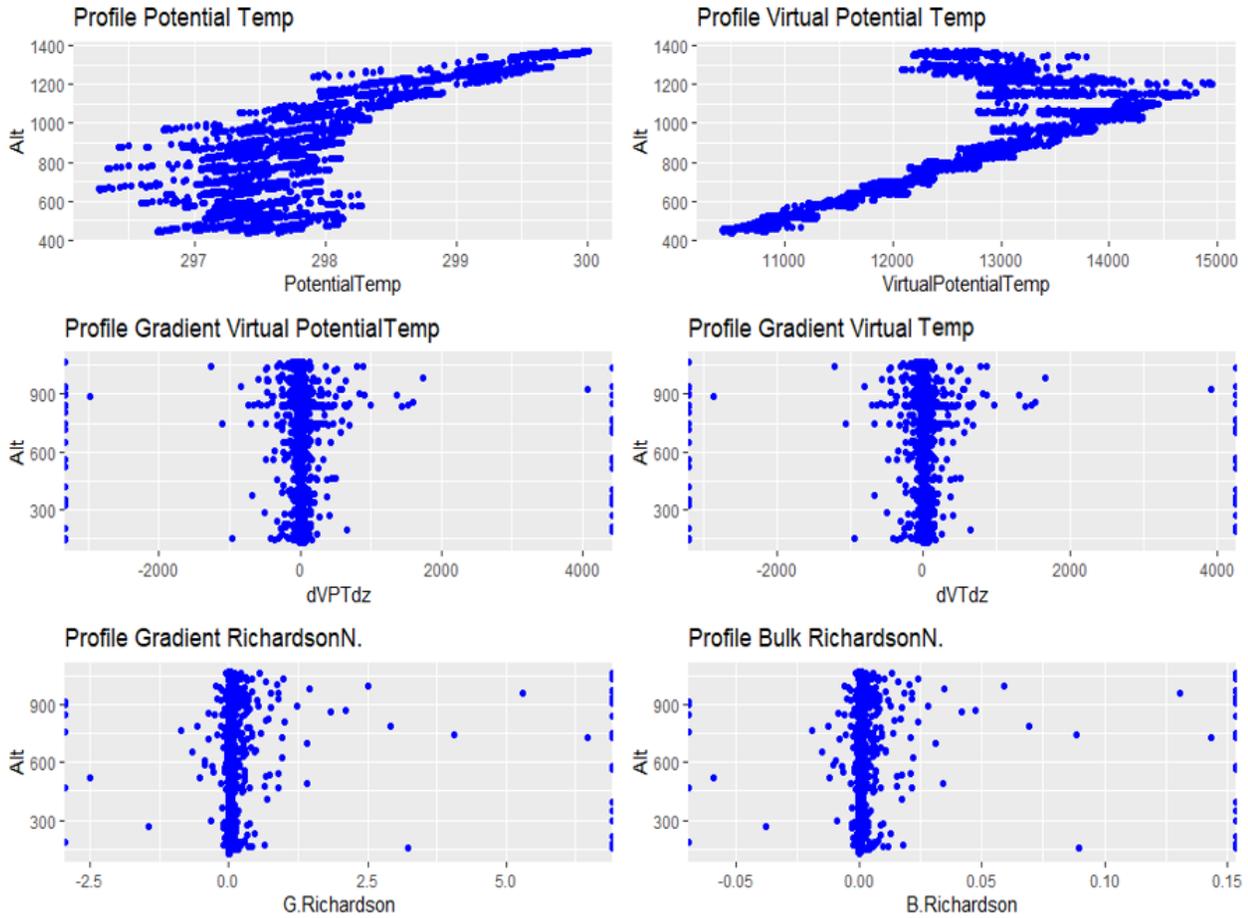
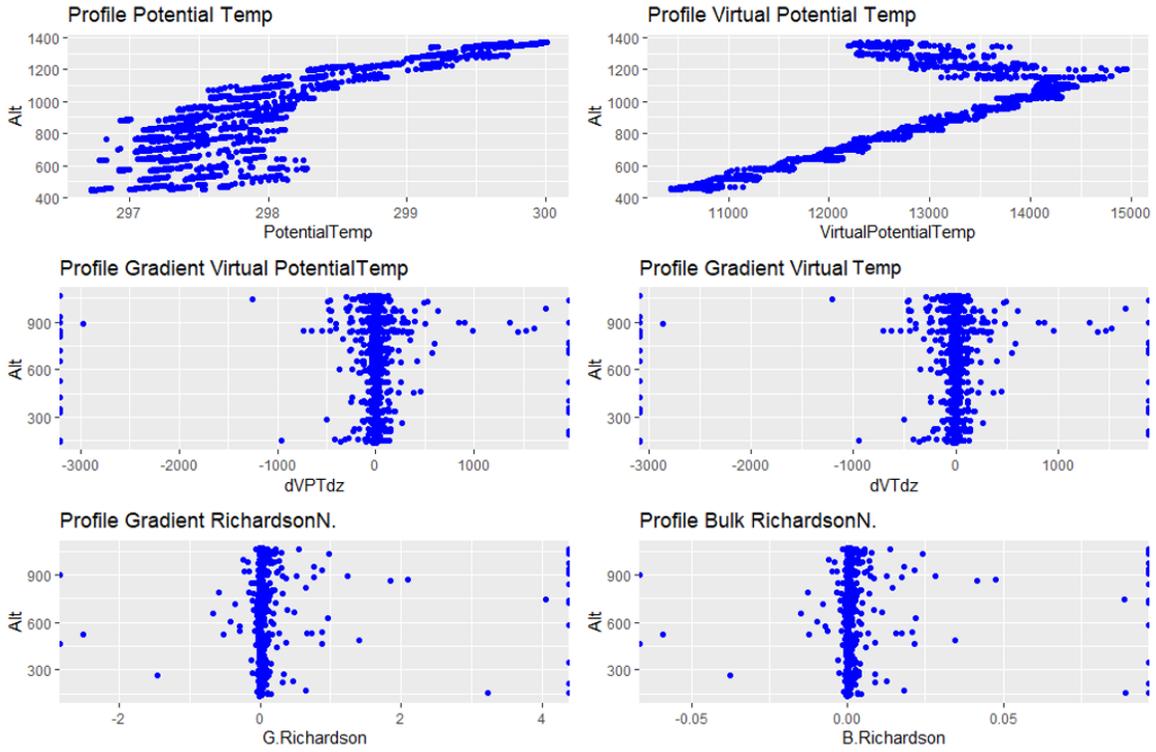


Figure 50. F04 profiles for the calculated potential temperature, virtual potential temperature, gradient (d) virtual potential temperature, gradient virtual temperature, gradient Richardson number, and bulk Richardson number.

F04: 2017-08-14 Syncrude Plant Up



F04: 2017-08-14 Syncrude Plant Down

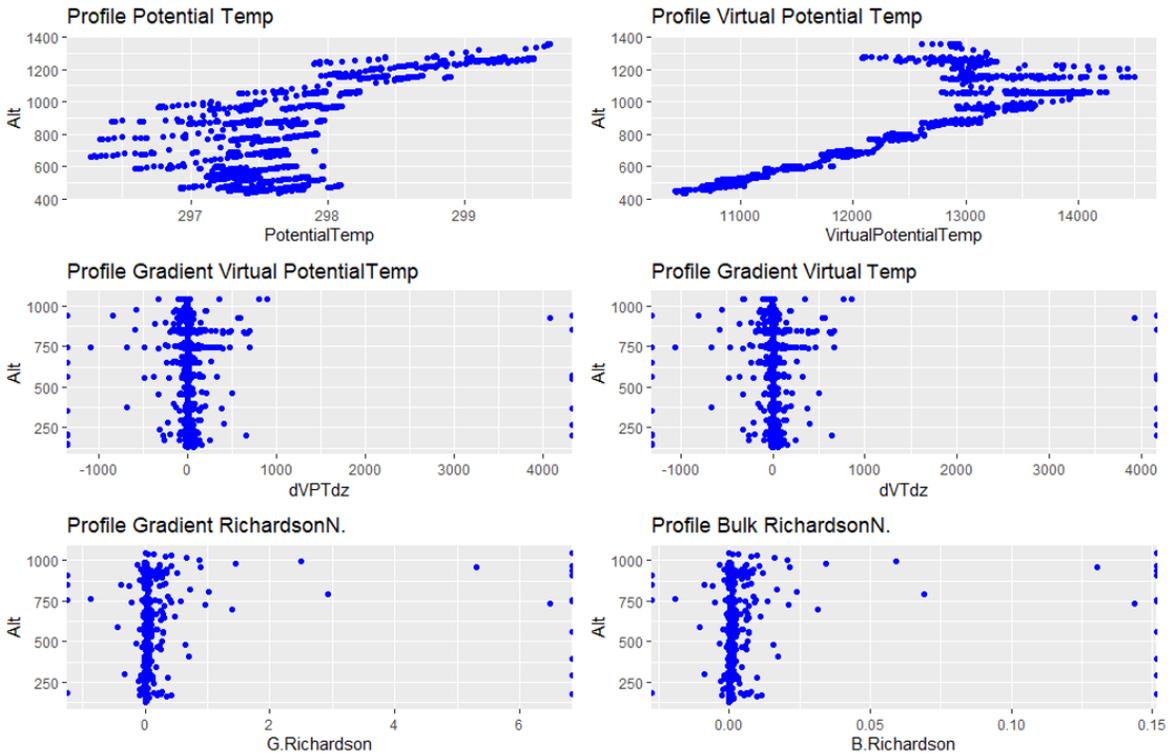


Figure 51. Same as for Figure 50, but for the F04 flight up (top panel) and flight down (bottom panel).

Code was developed as part of this thesis to estimate the boundary layer height 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 52 for F04. Analysis by Dai et al 2014 suggest the best practice is to estimate the boundary layer at maximum change in the virtual potential temperature. For this flight, the boundary layer estimate is very near SciAv's estimate and appears stable. This method was then applied to F04 split into approximate upwards and downwards flights (Figure 53 and 54). If the meteorological conditions caused a shift in the boundary layer this would likely show up during the reanalysis. While the max change in virtual temperature changed (which would not be used), the rest of the estimates remain the same. 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.

Boundary Layer Estimates F04: 2017-08-14

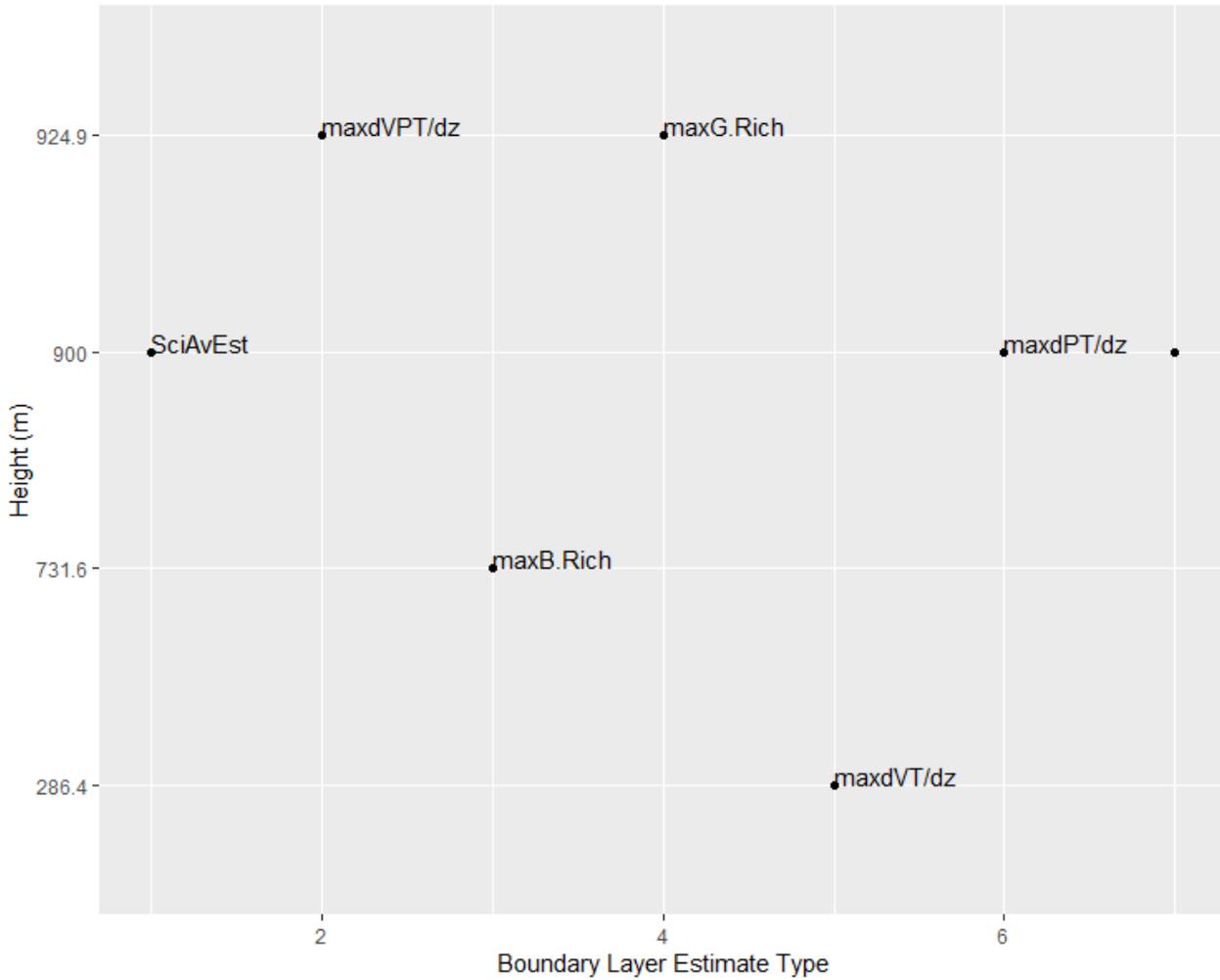


Figure 52. The five boundary layer estimates for the full F04 flight are plotted as the maximum: gradient Richardson number (maxG.Rich), bulk Richardson number (maxB.Rich), change in virtual temperature (dVT/dz), change in potential temperature (dPT/dz), and change in virtual potential temperature ($dVPT/dz$). The acronym dz is the change in altitude above ground level. The y-axis gives the estimated boundary layer in meters above ground level.

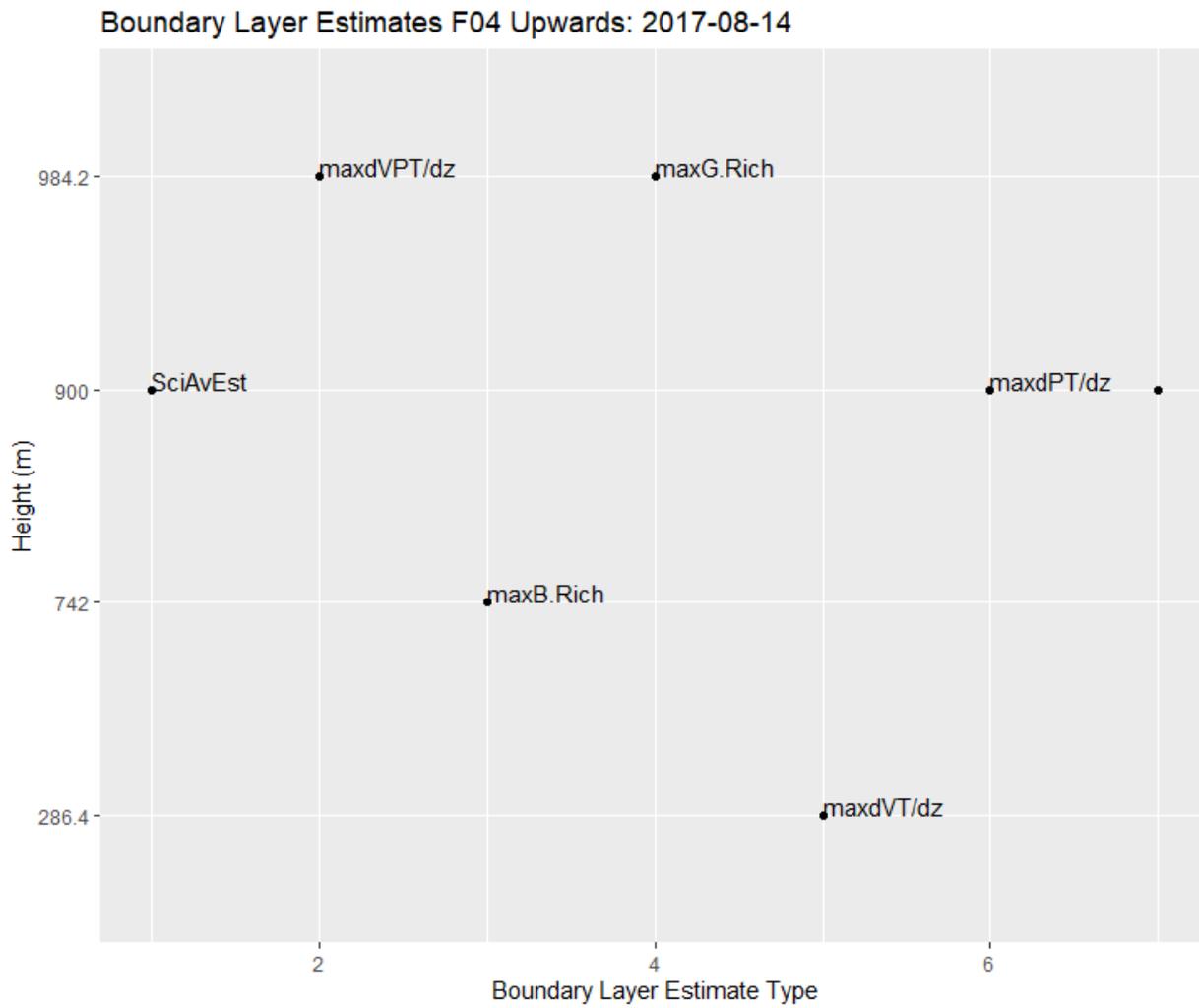


Figure 53. Same as for Figure 52 given the F04 split into the upward lap data for the flight.

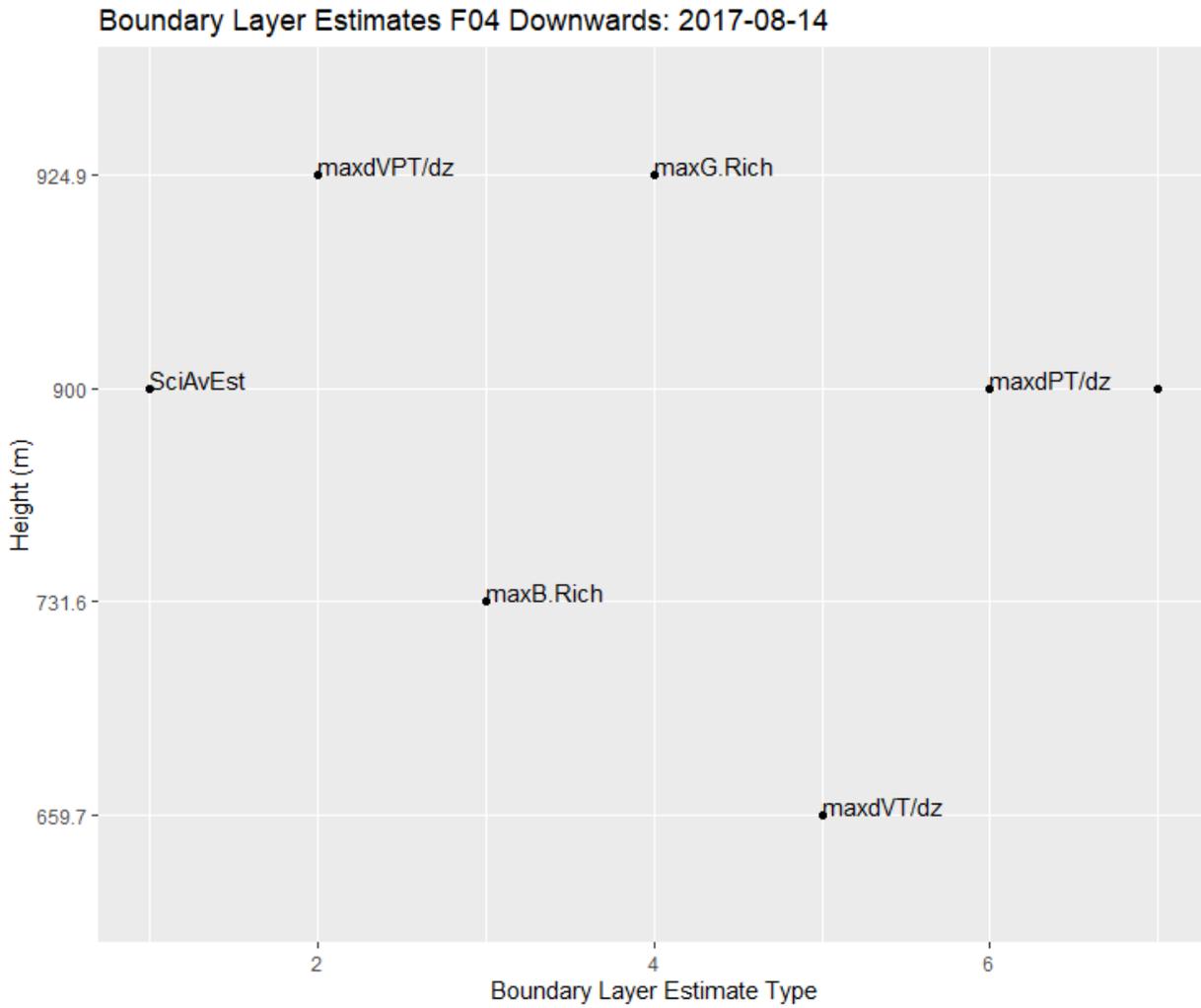


Figure 54. Same as for Figure 52 given the F04 split into the downwards lap data for the flight.