

Supplement of
**Comparing Airborne Algorithms for Greenhouse Gas Flux
Measurements over the Alberta Oil Sands**

5 Broghan M. Erland et al.

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

10 **1 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 section outlines the processes to evaluate those four requirements as instructed by Environment and Climate Change Canada (ECCC).

15 **1.1 TERRA Flight Screens**

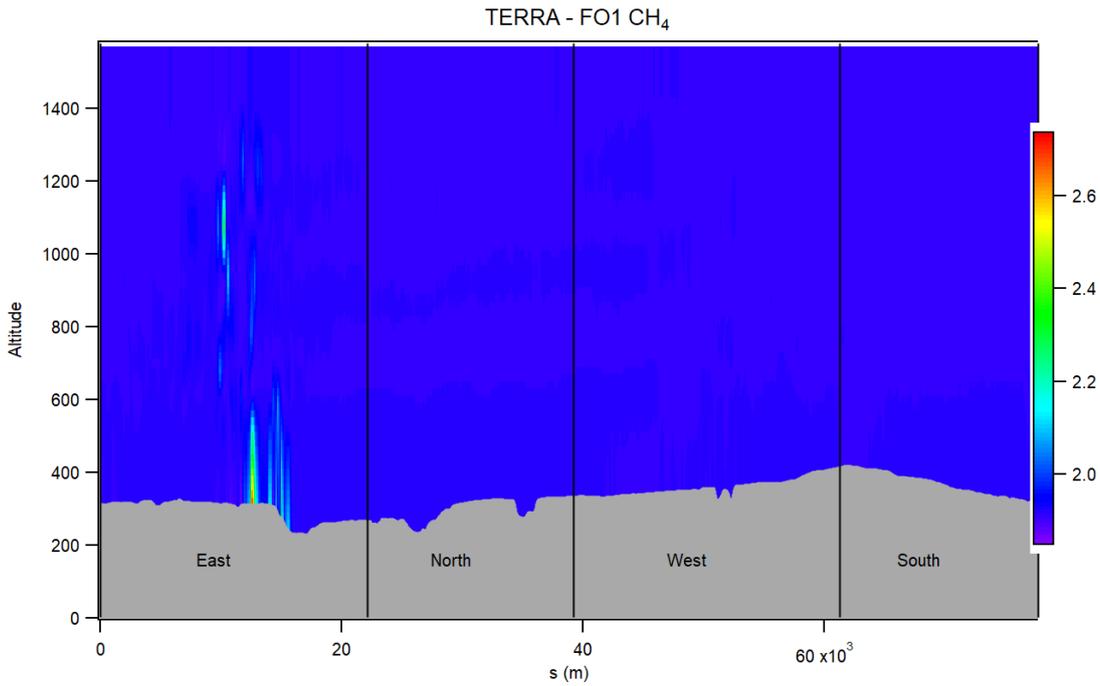
The emission screens produced by TERRA after fitting the chosen surface extrapolation are shown in Figure S 1 - 10. The left y-axis gives the altitude above ground level in meters, the right hand the colour scale of CH₄, or CO₂ in ppm. The length along each lap s(m) is plotted along the x-axis with the 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 S 7. The surface extrapolation sometimes estimates a decreasing emission plume towards the surface as per Figure S 2 and S 10. This leads to a larger range in the mixing ratio and a change from the typical royal blue background colour to a lighter shade such as cyan, or even light green to adjust for the lower scale. The background values are not affected by the change in colour. Aside from F04, the flight data used in the comparison analysis represent standard emissions screens for the TERRA method.

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

35

Table S 1: TERRA emission rate estimates using the method's five surface extrapolation fits.

Flight	Surface Extrapolation	CH ₄ Emission (kg h ⁻¹)	CO ₂ Emission (t h ⁻¹)
F01	Constant	4320	1220
F01	Linear	4810	1340
F01	Interpolate	3710	1040
F01	Background	2910	503
F01	Exponential	4460	1200
F02	Constant	412	535
F02	Linear	424	544
F02	Interpolate	395	515
F02	Background	369	485
F02	Exponential	414	543
F03	Constant	457	459
F03	Linear	476	467
F03	Interpolate	416	401
F03	Background	364	326
F03	Exponential	475	468
F04	Constant	125	589
F04	Linear	104	572
F04	Interpolate	123	569
F04	Background	119	543
F04	Exponential	122	589
F05	Constant	3540	882
F05	Linear	3910	877
F05	Interpolate	3150	8410
F05	Background	2630	790
F05	Exponential	3680	1220



40 **Figure S 1: 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.**

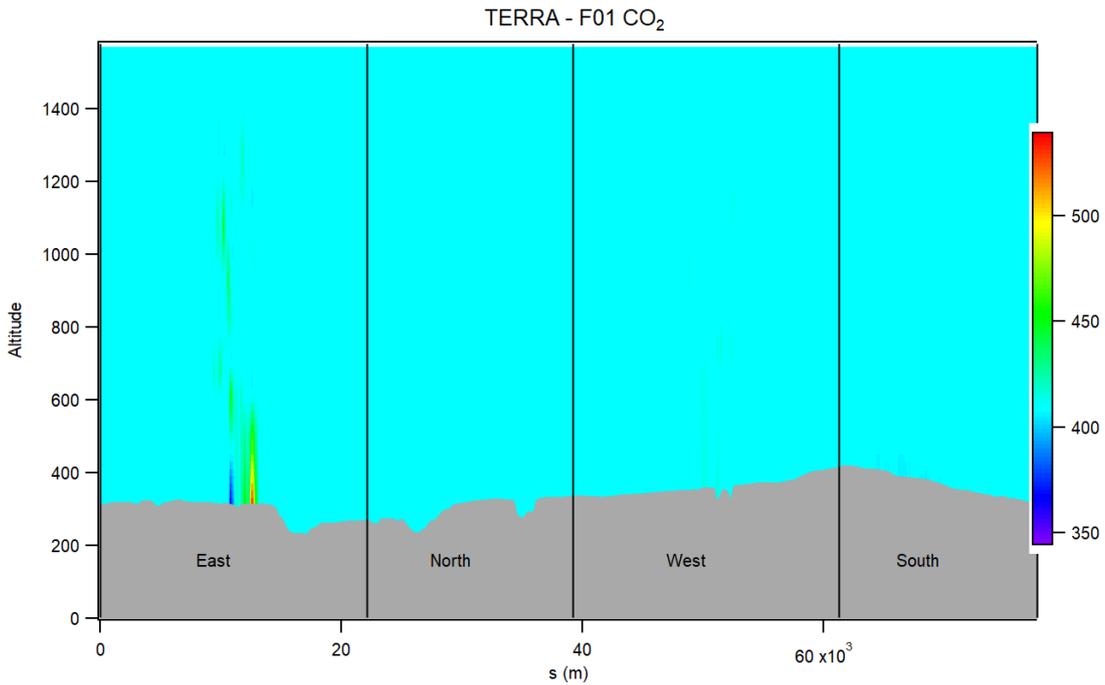
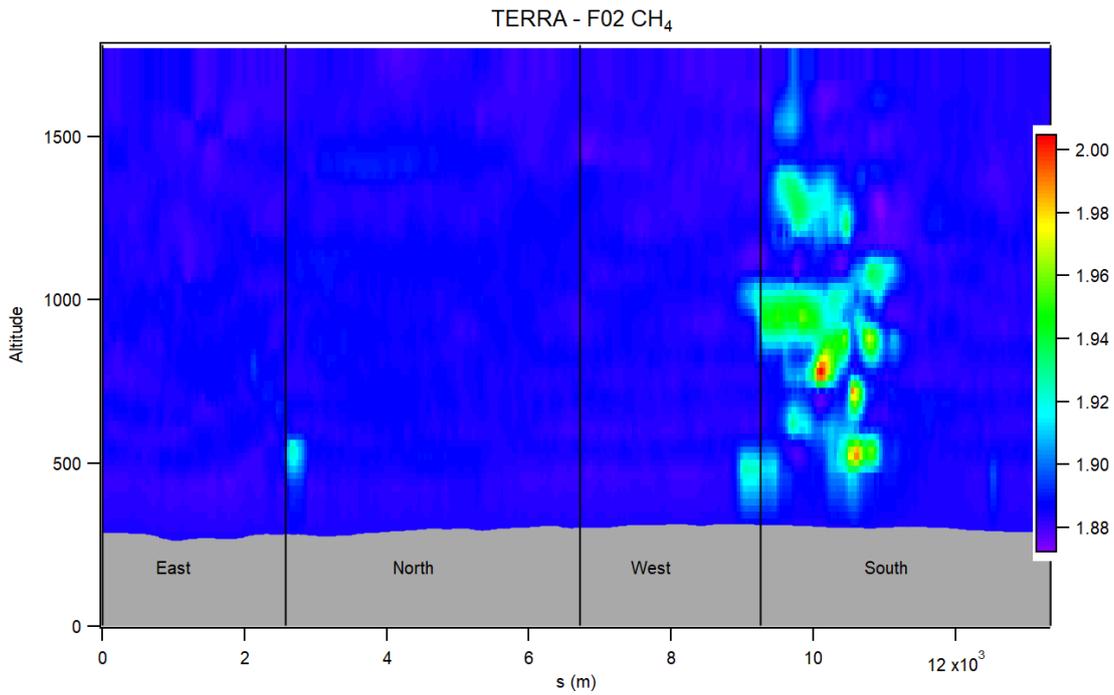


Figure S 2: Same as Figure S 1 for flight F01 CO₂ (ppm).



45 **Figure S 3:** Same as Figure S 1 for flight F02 CH₄ (ppm).

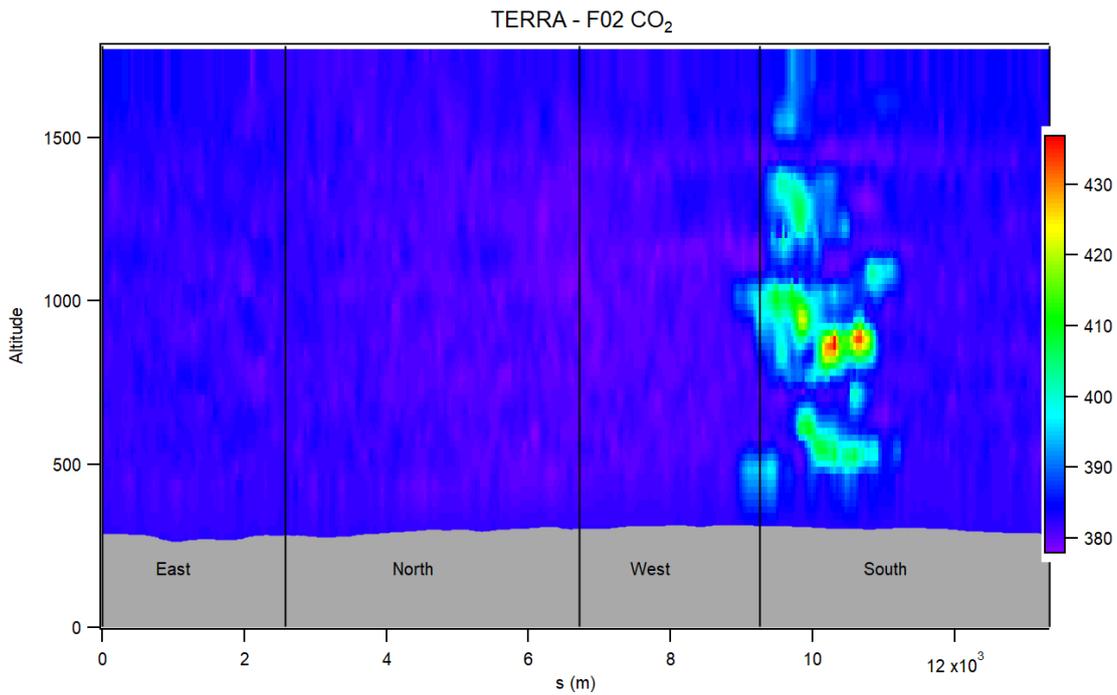


Figure S 4: Same as Figure S 1 for flight F02 CO₂ (ppm).

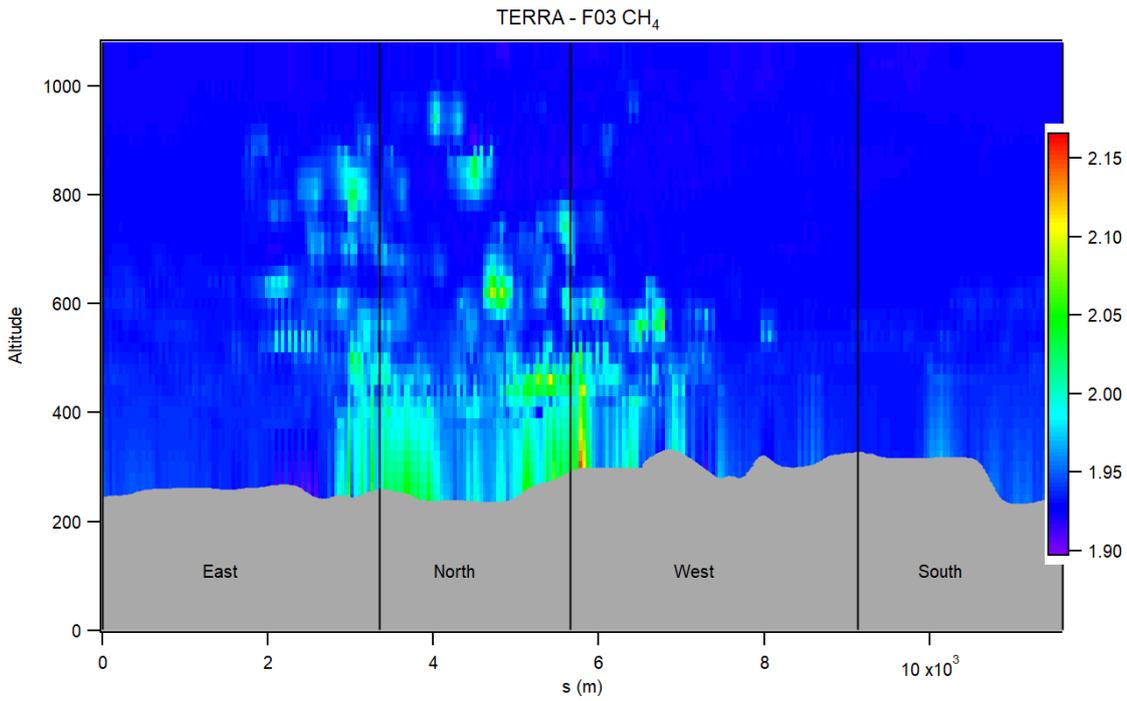


Figure S 5: Same as Figure S 1 for flight F03 CH₄ (ppm).

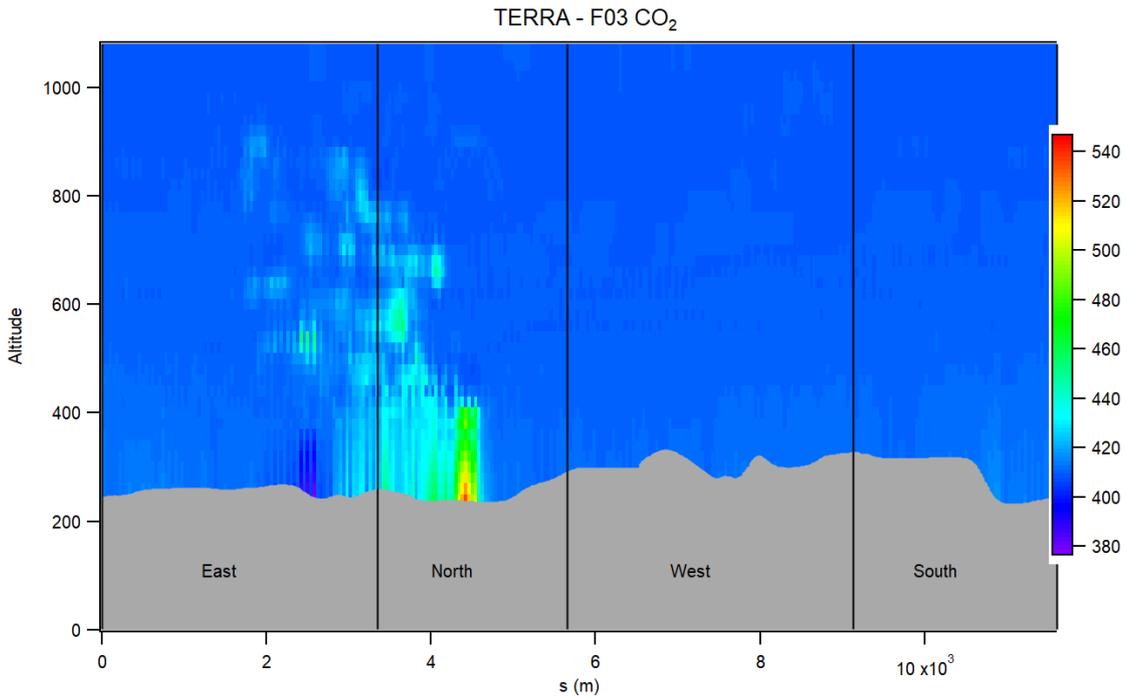


Figure S 6: Same as Figure S 1 for flight F03 CO₂ (ppm).

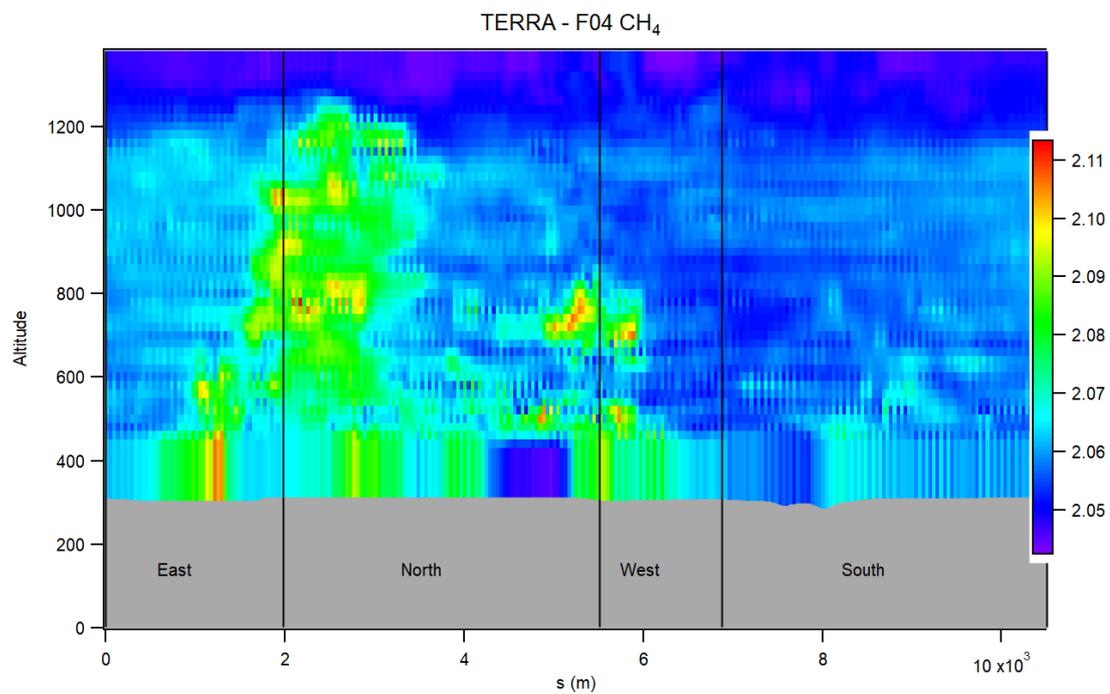
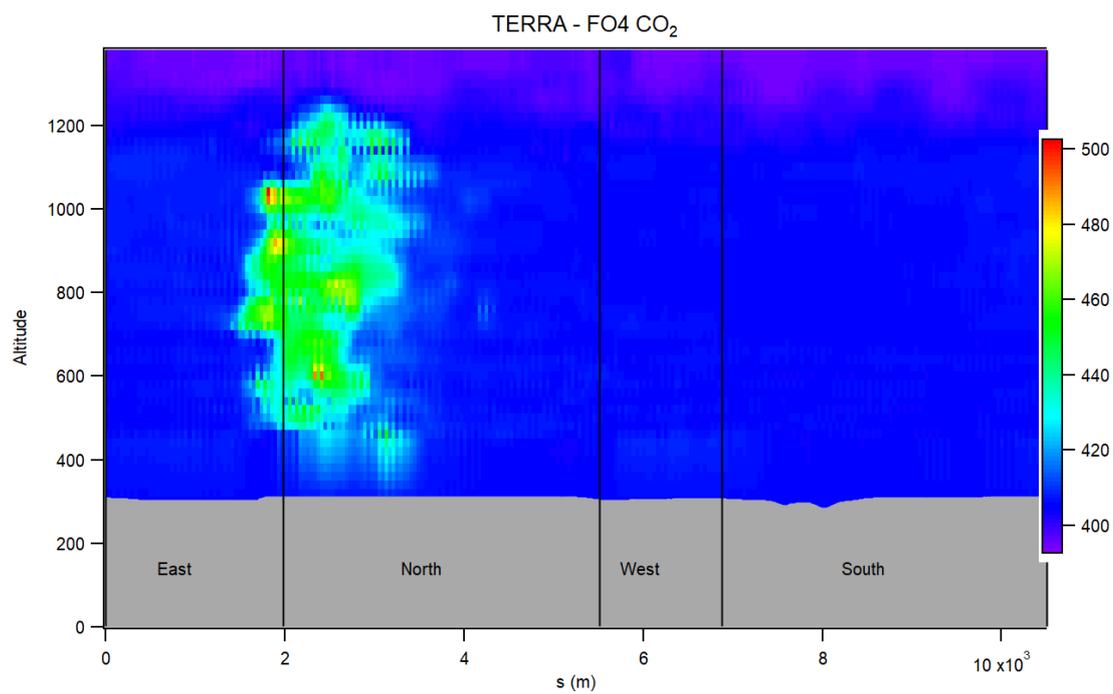


Figure S 7: Same as Figure S 1 for flight F04 CH₄ (ppm).



55 Figure S 8: Same as Figure S 1 for flight F04 CO₂ (ppm).

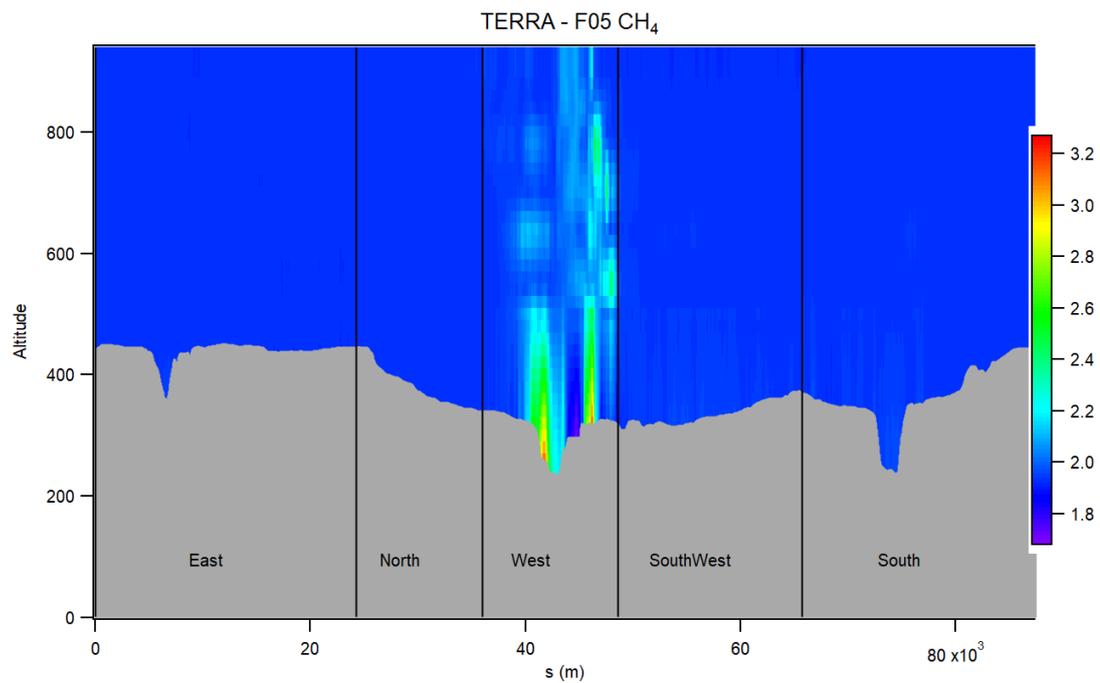


Figure S 9: Same as Figure S 1 for flight F05 CH₄ (ppm).

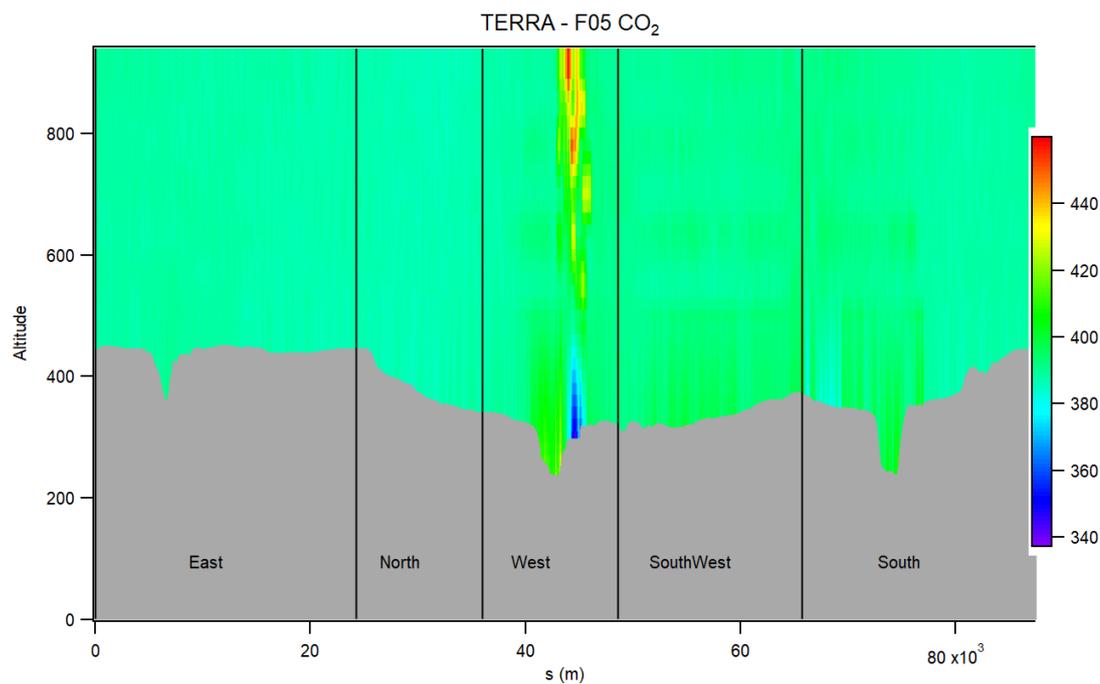


Figure S 10: Same as Figure 1 for flight F05 CO₂ (ppm).

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

Flight	CH ₄ (ppm)	CO ₂ (ppm)
F01	1.91613	409.1896
F02	1.88481	382.3219
F03	1.93007	410.2672
F04	2.05365	405.8721
F05	1.93103	389.4703

1.2 TERRA Error Terms

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

65

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).

70

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 ($\Delta p/p_i - \Delta T_i/T_i$) 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 .

75

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

85 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.

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

Flight	Measurement Error δ_M	Mixing ratio Extrapolation δ_{Ex}	Wind Extrapolation δ_{Wind}	Box-top mixing ratio δ_{Top}	Density Change δ_{dens}	Vertical Turbulence δ_{VT}	Box-top height δ_{BH}	Total Uncertainty δ
F01 - CH ₄	<1%	10.13	<1%	0.89	0.03	N/A	0.56	10.28
F01 - CO ₂	<1%	10.04	<1%	0.89	0.02	N/A	0.35	10.18
F02 - CH ₄	<1%	4.80	<1%	2.15	1.06	N/A	1	5.64
F02 - CO ₂	<1%	5.43	<1%	0.38	0.89	N/A	1	5.78
F03 - CH ₄	<1%	4.16	<1%	2.30	0.48	N/A	1	5.08
F03 - CO ₂	<1%	1.71	<1%	2.30	0.83	N/A	1	3.45
F04 - CH ₄	<1%	17.08	<1%	2.42	5.79	N/A	1	18.28
F04 - CO ₂	<1%	3.53	<1%	2.43	4.93	N/A	1	6.76
F05 - CH ₄	<1%	9.54	<1%	1.00	0.04	N/A	11.17	14.79
F05 - CO ₂	<1%	7.33	<1%	1.11	0.10	N/A	24.82	25.94

Adaptation of the SciAv Method in Further Detail

The raw data files for the five flights can be accessed through the Government of Alberta Portal:

95 <http://ckandata01.canadacentral.cloudapp.azure.com/dataset/aep-noaa-greenhouse-gas-measurement-flights>

The calculated average divergence for each lap, and associated errors provided by Scientific Aviation are provided in Table S 4. The lap bin bounds also provided by Scientific Aviation are given in Table S 5. Together these data can be used to recreate the SciAv profiles and emission calculations for each flight.

Table S 4: The average 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

100

105

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

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

115 1.3 Fitting the SciAv Surface Extrapolation

The following section provides the Scientific Aviation (SciAv) algorithm surface extrapolation for all five different fits: Constant, Linear, Interpolate, Background, and Linear Weighted. The algorithm's integration method of binning the estimates and adding in summation the average of each bin was applied for the measured part of the profile (Conley et al. 2017). For each surface extrapolation the lowest bin was extrapolated to the surface using the given fit. Certain profile shapes can be used to determine the fitting height depending on the number of points, minimum height, and shape of the bottom of the profile. The chosen fitting heights were: 1100m for F01, 400m for F02, 250m for F03, 400m for F04, and 600m for F05. The final estimate was determined by combining the emissions calculated from binned profile added to the trapezoidal integration of area between the extrapolation fit, the surface, and a flux enhancement of zero (main text, Figure S 4). The estimates produced by fitting the different surface extrapolations to SciAv are given in Table S 6. Figure S 11A shows that both emission profiles for F01 follows a type III shape with fits for linear, weighted linear, and averaged increasing towards the surface. The F01 surface extrapolation and resulting emission estimate would be larger and more closely follow plume behaviour if the lowest divergence point, rather than the lowest bin estimate (the method standard), was used (Figure S 11 A). The F02 SciAv profile follows a Type II shape (Figure S 11 B) with very little variation in the extrapolation fits, which are clustered around the standard constant fit. The F03 SciAv profile has a lot of between-lap variation with a profile shape that largely follows a Type I shape (Figure S 11 C). The F03 surface extrapolation was the most unstable fit due to the larger variation between laps at the bottom of the profiles. F05 was the only flight with varying profile shapes for and CH₄ and CO₂. The shape of the emission profile in Figure S 11 E 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 S 11 E shows incomplete sampling for CO₂ as the divergence points are still increasing at the highest altitude indicating that the plume was not fully captured.

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

Flight	Surface Extrapolation	CH ₄ Emission (kg h ⁻¹)	CO ₂ Emission (t h ⁻¹)
F01	Constant	3840	1040
F01	Linear	5270	1250
F01	Interpolate	3900	1040
F01	Background	3050	784
F01	Linear Weighted	4690	1300
F01	Averaged	4200	1150
F02	Constant	362	5630
F02	Linear	360	556
F02	Interpolate	331	523

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

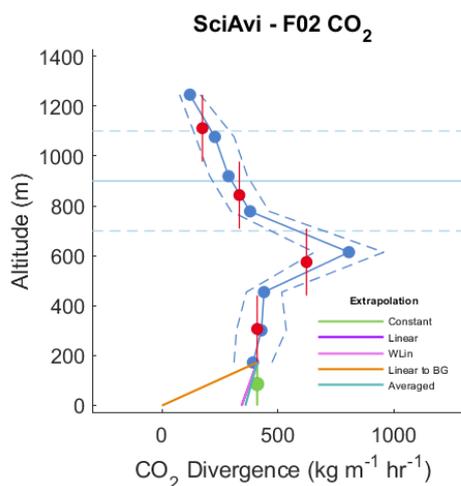
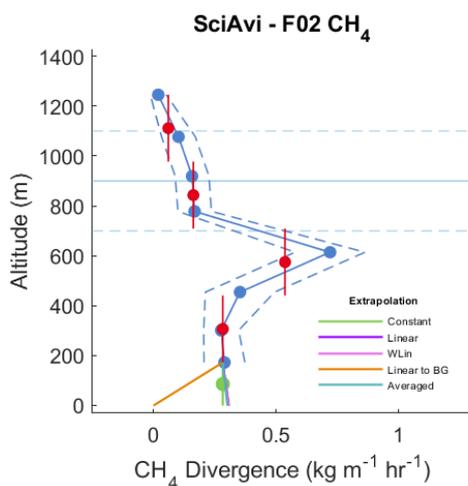
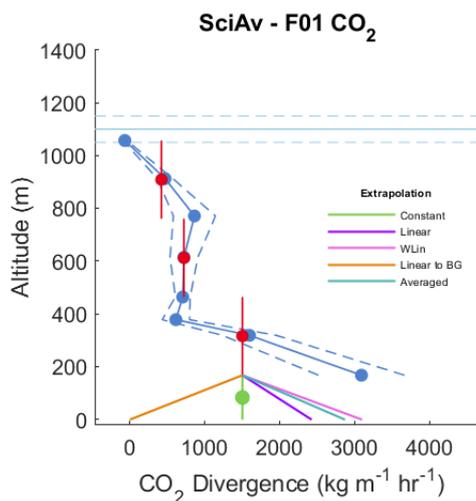
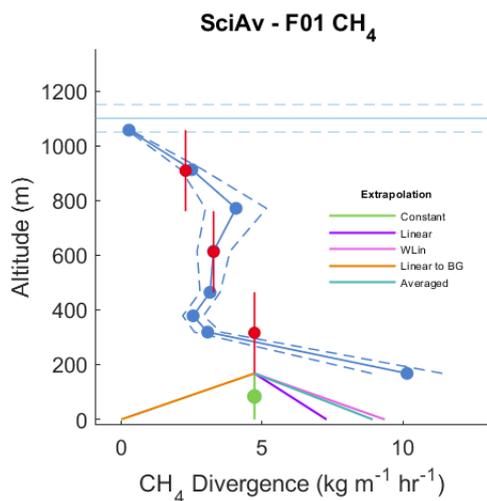


Figure S 11 B

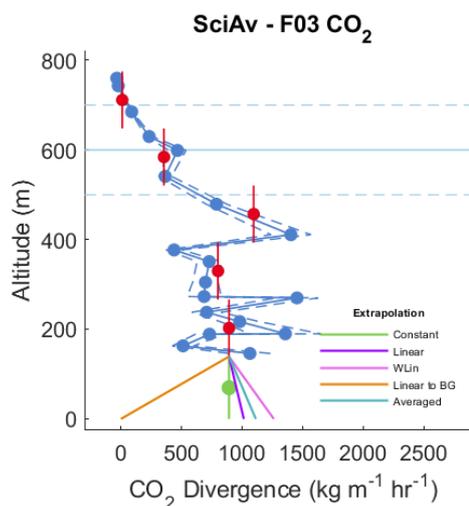
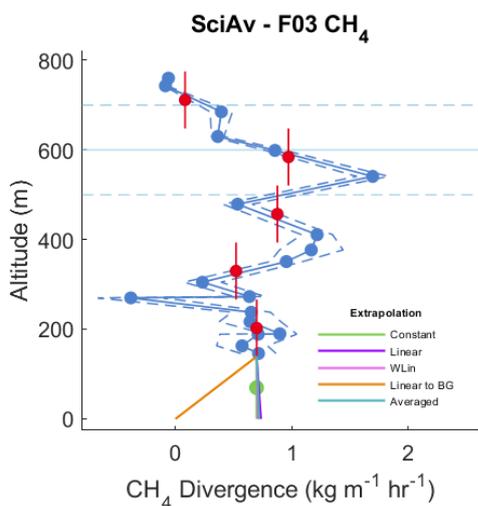


Figure S 11 C

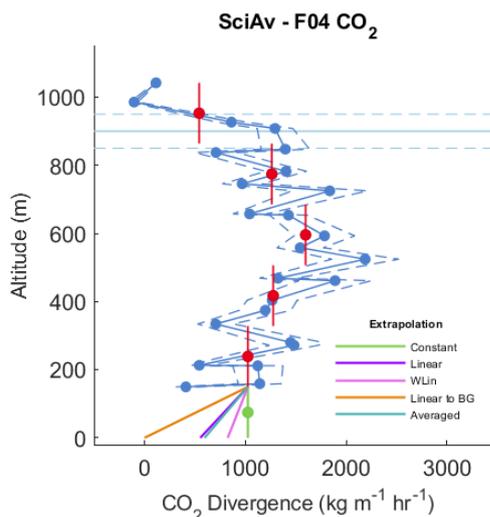
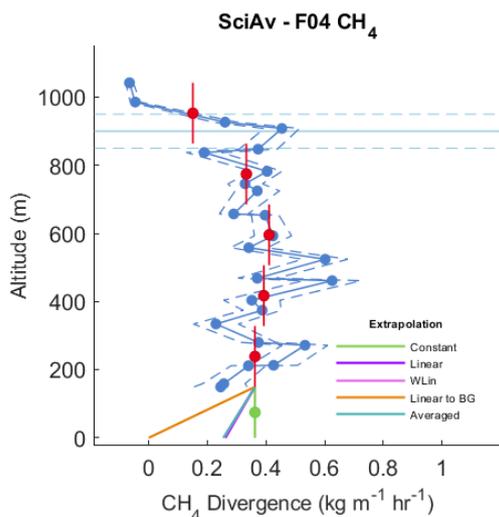


Figure S 11 D.

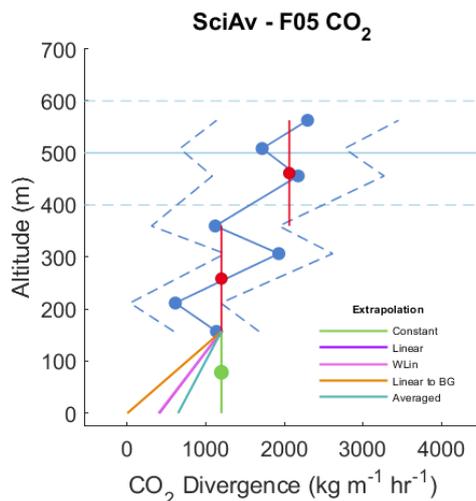
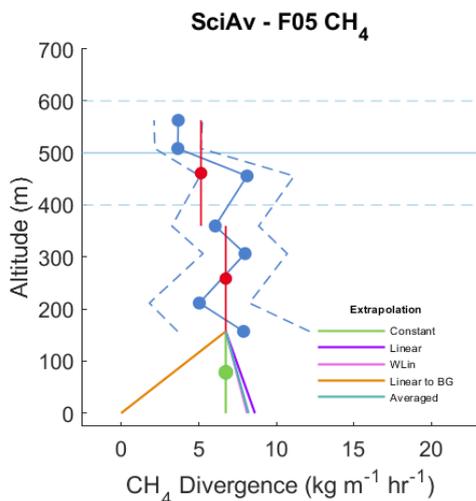


Figure S 11 E.

Figure S 11: The SciAv profile is shown for F01-F05 in Figures 8A – 8E. 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.

Comparison of SciAv and TERRA Methods Extended

1.4 Significance Testing and Further Comparison

The difference between estimates, relative mean percentage difference, and the propagated error in percentage for the average estimate of the two algorithms are shown in Table S 7. The whole set of results from five flights were formally tested for differences between the SciAv and TERRA estimates using a weighted t-test and Wilcoxon signed rank test. As a collective, the differences between the estimates were found to be insignificant for both CH₄ and CO₂ (Table S 8).

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

	F01	F02	F03	F04	F05
CH ₄ : Difference (kg h ⁻¹)	-967	-33.5	20.5	224	-436
CH ₄ : Relative Percentage Difference	22%	8%	4%	94%	12%
CH ₄ : Propagated Percentage Error	21%	22%	17%	18%	29%
CO ₂ : Difference (kg h ⁻¹)	-302000	47900	58700	602000	-27700
CO ₂ : Relative Percentage Difference	25%	9%	12%	69%	3%
CO ₂ : Propagated Percentage Error	24%	19%	12%	13%	43%

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

	Weighted t-test: p-value	Weighted t-test: t-value	Wilcoxon signed-rank test: p-value	Wilcoxon signed-rank test: V
CH ₄	0.306	-1.09	0.438	4
CO ₂	0.366	-0.96	0.625	10

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

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

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

Emission estimates using a background fit are systematically lower than the standard estimates (Figure S 12), but no clear pattern is identifiable that would indicate systematic disagreement. There is a considerably larger TERRA estimate for F01 between the standard fit compared to the background. The choice of surface extrapolation influences the extent of agreement. The algorithms tend to agree more when the effect of a different surface extrapolation is removed as much as possible. This indicates good agreement between the methods. When fundamental assumptions are met, variation between the estimates can largely be attributed to different surface extrapolation methods.

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

	CH ₄ Standard	CH ₄ Background	CH ₄ Average	CO ₂ Standard	CO ₂ Background	CO ₂ Average
Mean Difference (kg h ⁻¹)	-239	12.4	47.8	75800	141000	115000
Pairwise t-test: p-value	0.857	0.989	0.969	0.724	0.321	0.525
Pairwise t-test: t-value	-0.187	0.0145	0.0404	0.367	1.07	0.665
Wilcoxon signed-rank test: p-value	0.438	1	0.625	0.625	0.313	0.313
Wilcoxon signed-rank test: V	4	8	10	10	12	12

180

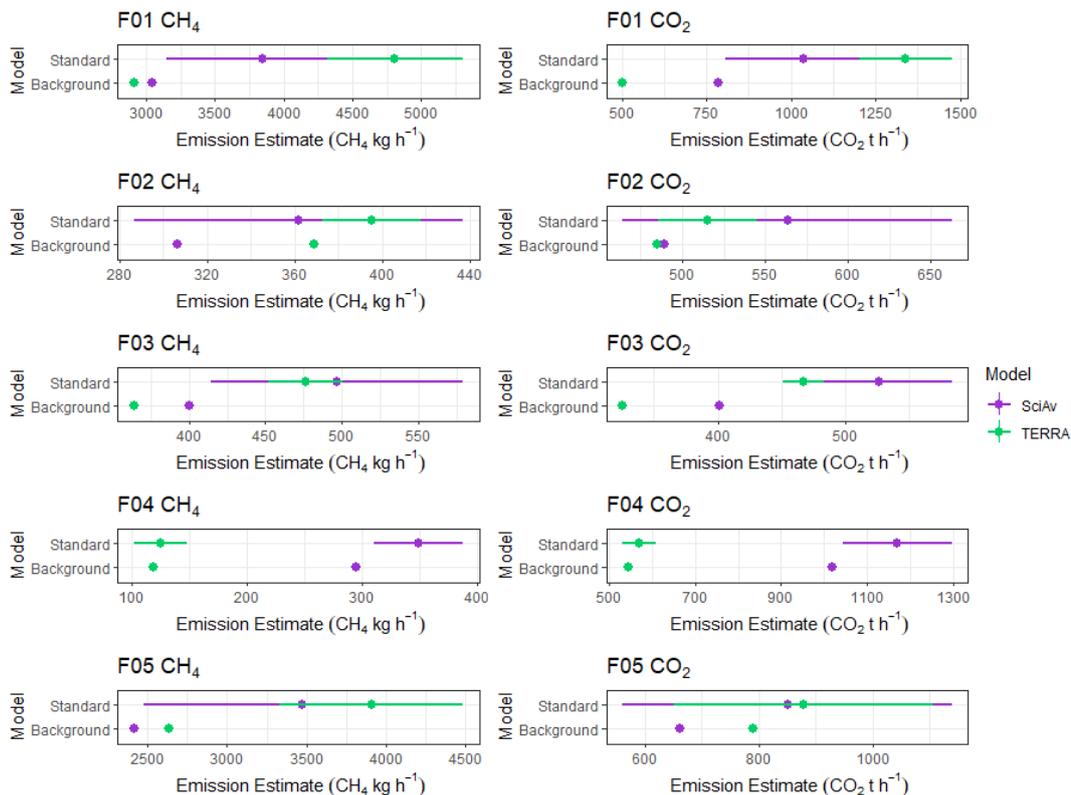


Figure S 12: Algorithm estimates for CH₄ and CO₂ given the background mixing ratio fit of no emission plume extrapolation to the surface compared to the standard estimates.

Table S 10 gives the percentage relative differences between each algorithm, and the standard deviation of the percentages for each surface extrapolation, calculated for the four surface extrapolations used for each method. In general, the choice in surface extrapolation varied less than the relative percentage difference between the algorithms. Fitting extrapolations with increasing emissions towards the surface increased agreement for F01 CO₂ from a percentage relative difference of 16% for the standard constant fit, to 7% using the linear fit.

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

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

195

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

$$200 \quad \text{difference} = \text{mean}(\text{Estimate}_{\text{SciAv}}) - \text{mean}(\text{Estimate}_{\text{TERRA}}) \quad (1)$$

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

205

Scientific Aviation Trapezoidal Integration Estimations

1.5 Comparison Results and Significance Testing

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

Table S 11 gives the SciAv trapezoidal integration method estimates. Figures S 13 A – 13 E show the SciAv profiles given the trapezoidal integration method. Surface extrapolations are fit from the lowest mean divergence lap estimate. The estimated end point for the surface extrapolation does not change between the two integration methods as it is determined by fitting curves by the position of the points and fitting height. As the surface extrapolation points were the same for both profiles which caused the F01 trapezoidal profiles to decrease despite their noticeable increasing to surface trend (Figure S 13 A). A less conservative fitting height procedure would likely produce a large range in emission estimates, and should be explored if the trapezoidal method is adapted, as it could produce an extrapolation truer to the nature of the divergence profile.

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

Flight	Surface Extrapolation	CH ₄ Emission (kg h ⁻¹)	CO ₂ Emission (t h ⁻¹)
F01	Constant	4840	1350
F01	Background	3130	831
F01	Linear	4600	1300
F01	Interpolate	3980	1090
F01	Linear Weighted	4770	1350
F01	Averaged	4730	1330
F02	Constant	352	550
F02	Background	302	483
F02	Linear	355	550
F02	Interpolate	327	517
F02	Linear Weighted	354	557

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

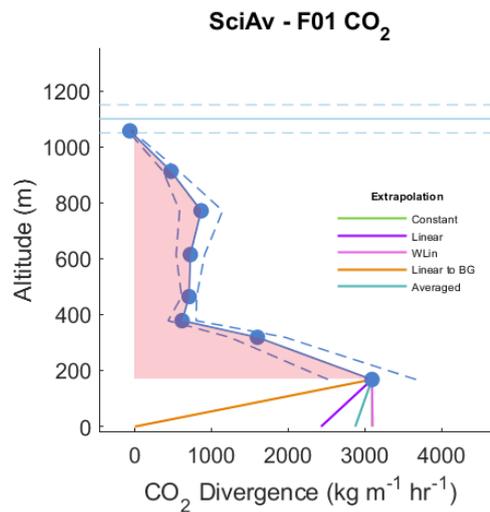
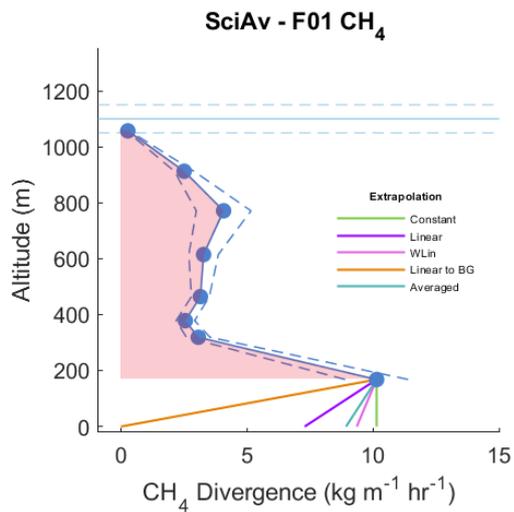


Figure S 13 A.

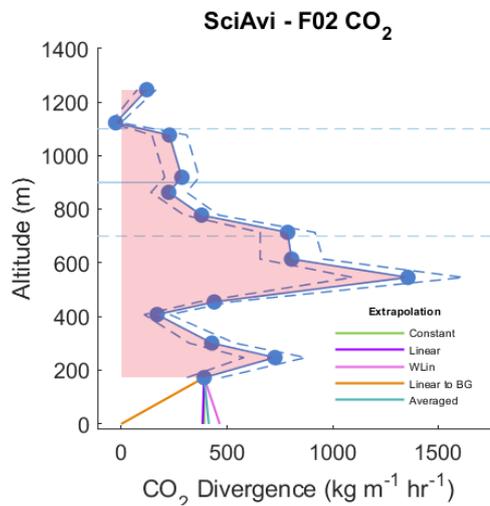
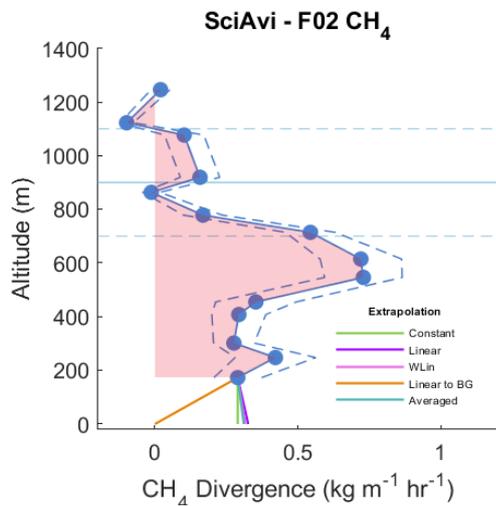


Figure S 13 B.

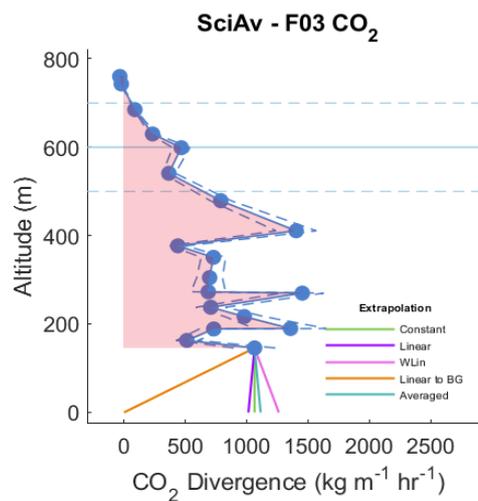
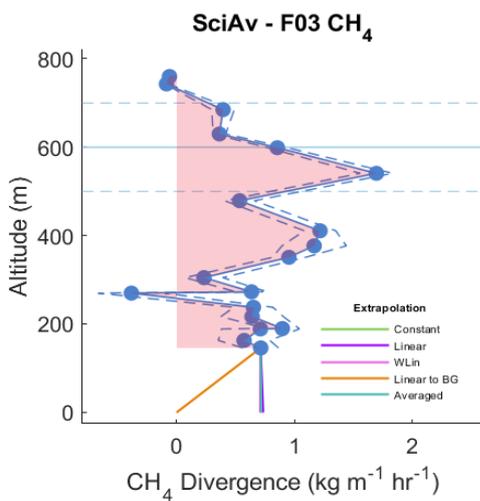


Figure S 13 C.

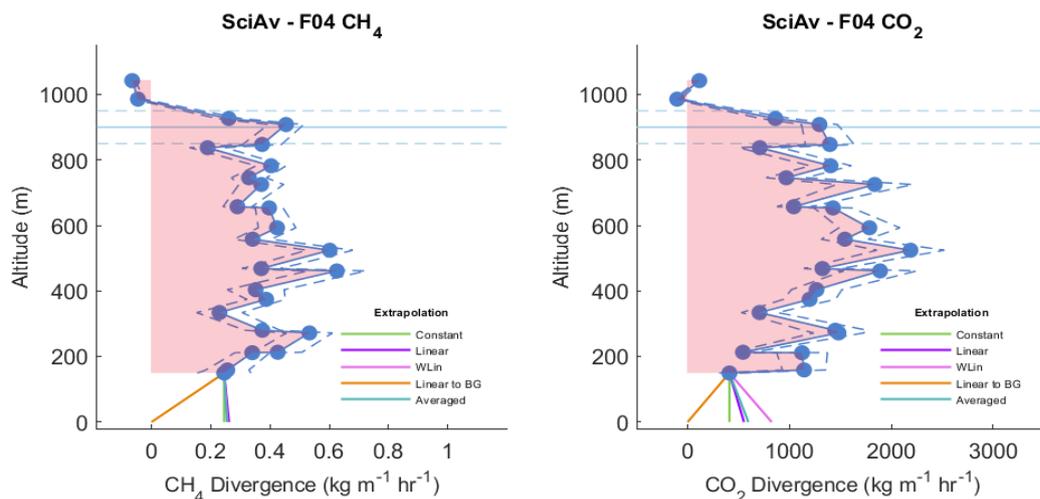


Figure S 13 D.

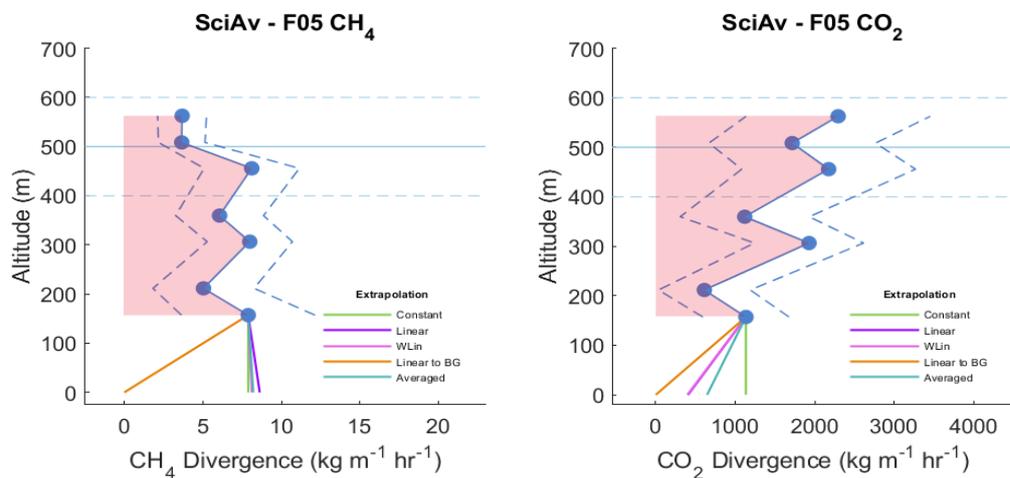


Figure S 13 E.

Figure S 13: 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.

Figure S 14 show the binned estimates compared to the trapezoidal (Trapz). No systematic pattern is evident that would indicate that one integration method produces consistently larger or smaller emissions estimates. There is a systematic gap between estimates that could be attributed to the use of different extrapolation points. The binned method uses the lowest bin for extrapolation, whereas the trapezoidal method uses the lowest divergence point.

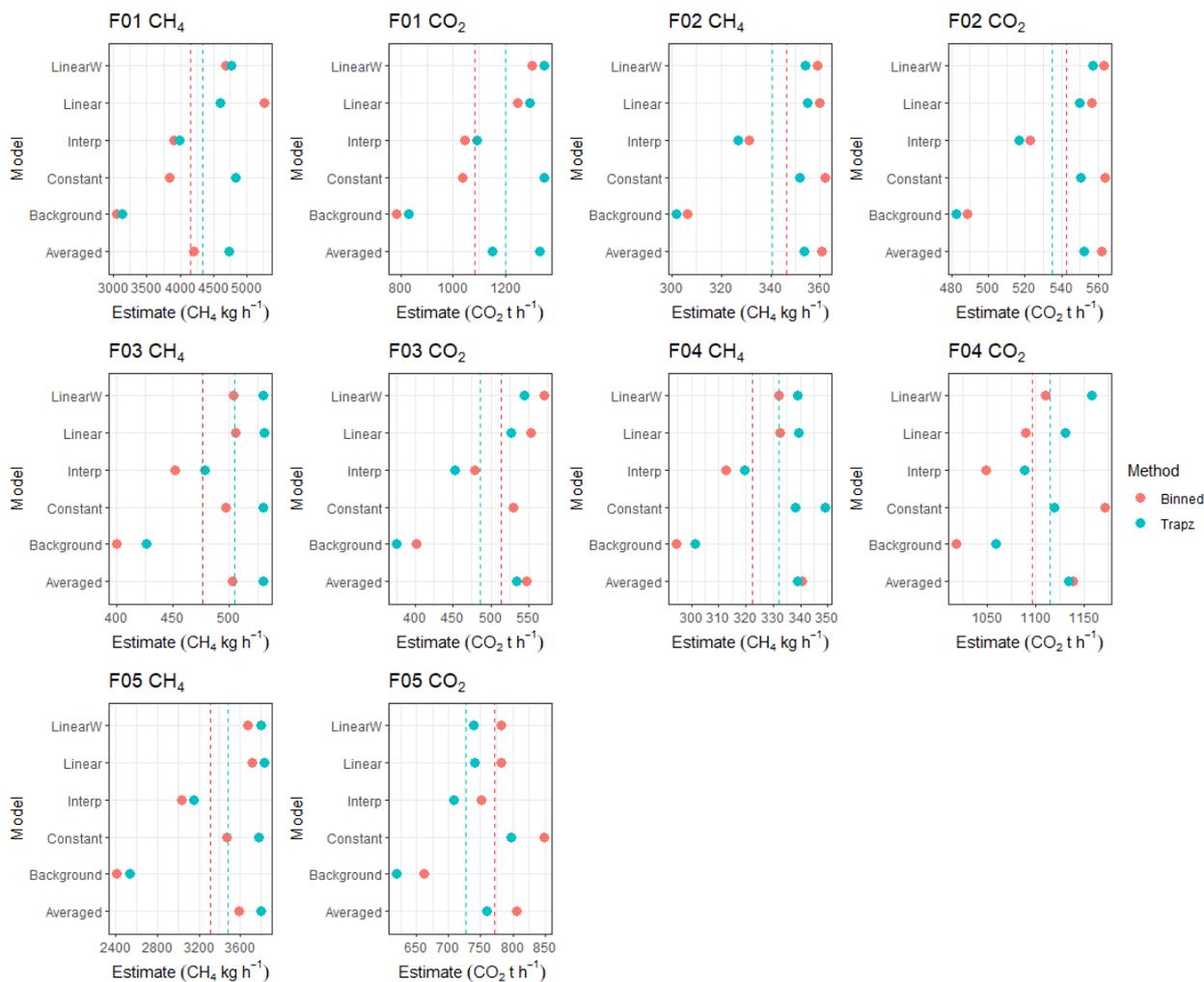


Figure S 14: SciAv emission estimates of CH_4 (kg h^{-1}) and CO_2 (t h^{-1}) derived using both a binning (red) and trapezoidal (blue) integration method. Mean values for each method are shown as a dotted line.

Inspecting Figure S 14 there is a larger variation between estimates due to the chosen surface extrapolation than integration method. There was one exception, there is a larger difference between the integration types when a constant extrapolation is chosen for F01. This flight has an increasing to surface profile point, and because the lowest divergence point differs greatly from the location of the lowest bin estimate, the extrapolation fits differ. Applying an appropriate surface extrapolation that follows the shape of the profile remedied the difference between integration methods.

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

255

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

	F01 CH ₄	F02 CH ₄	F03 CH ₄	F04 CH ₄	F05 CH ₄	F01 CO ₂	F02 CO ₂	F03 CO ₂	F04 CO ₂	F05 CO ₂	Mean
Binned %RSD	18	7	9	6	15	17	6	12	5	8	10
Trapezoid % RSD	15	7	9	5	15	17	5	14	3	8	10
Difference % RSD	12	1	1	2	2	9	1	3	3	0	3

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

	SD Binned Extrapolation Estimates	SD Trapezoidal Extrapolation Estimates	SD Difference in Methods Extrapolation Estimates
CH ₄ (kg hr ⁻¹)	1740	1810	581
CO ₂ (t hr ⁻¹)	274	315	139

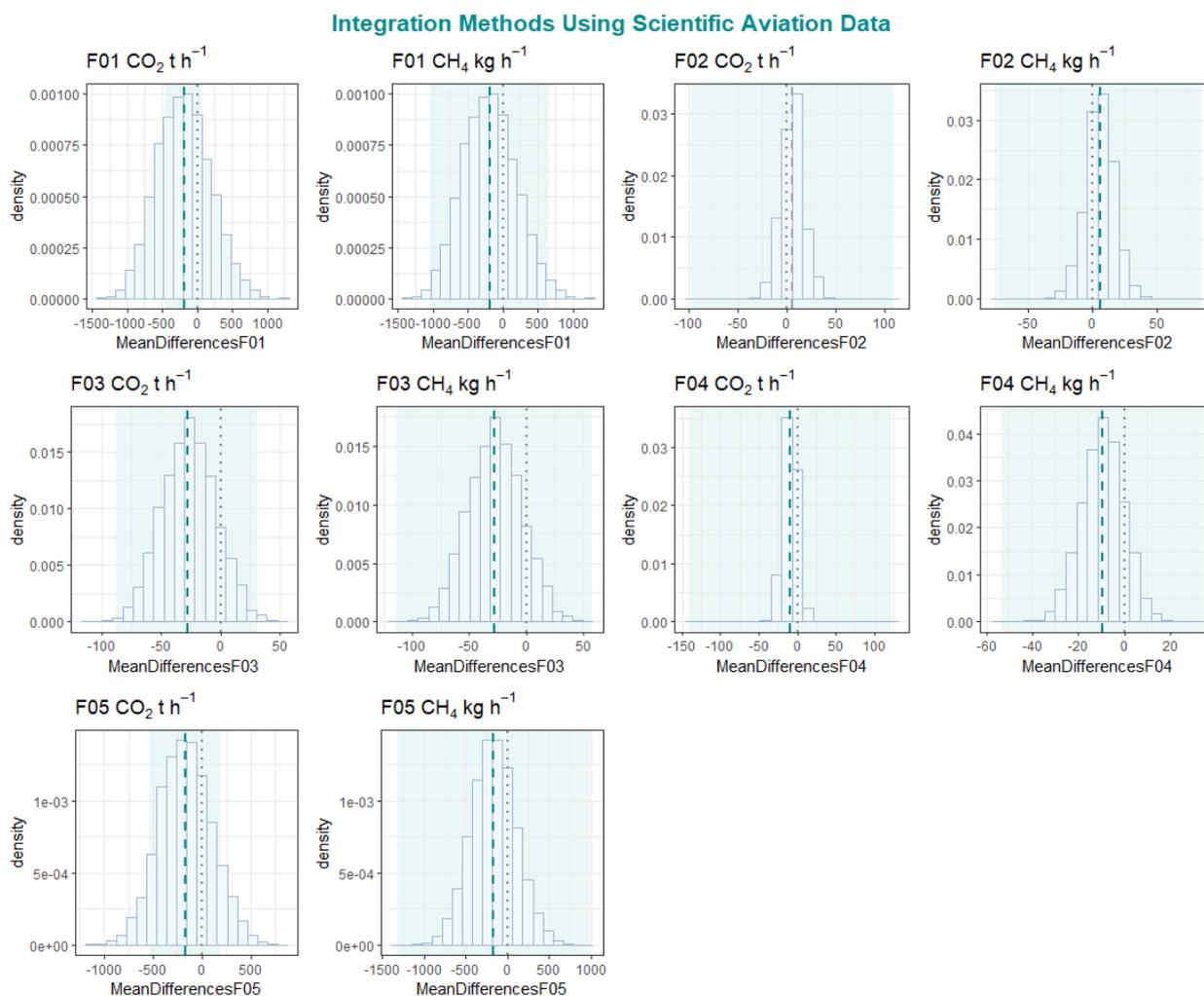
265 **Table S 14: Results of parametric (weighted t-test) and non-parametric (Wilcoxon signed-rank test) significance testing of differences between the two SciAv integration method, by calculating the mean of the estimates for each flight and tested over the whole set of five flights.**

	Weighted t-test: p-value	Weighted t-test: t-value	Wilcoxon signed-rank test: p-value	Wilcoxon signed-rank test: V
CH ₄	0.952	-0.062	0.188	2
CO ₂	0.952	-0.062	1	8

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

$$difference = mean(SciAvEstimate_{Binned}) - mean(SciAvEstimate_{Trapez}) \quad (2)$$

270 As shown in Figure S 15, 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
 275 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.



280 **Figure S 15: Distributions of the mean difference between all SciAv fits of CH₄ and CO₂, given the binning compared to the trapezoidal integration methods, in light blue. The mean difference between the standard estimates is plotted as a teal dashed lined and the range in the estimate as the light teal box behind the distribution. A grey dot dashed line is drawn at zero as a reference point for the location of exact agreement between the methods.**

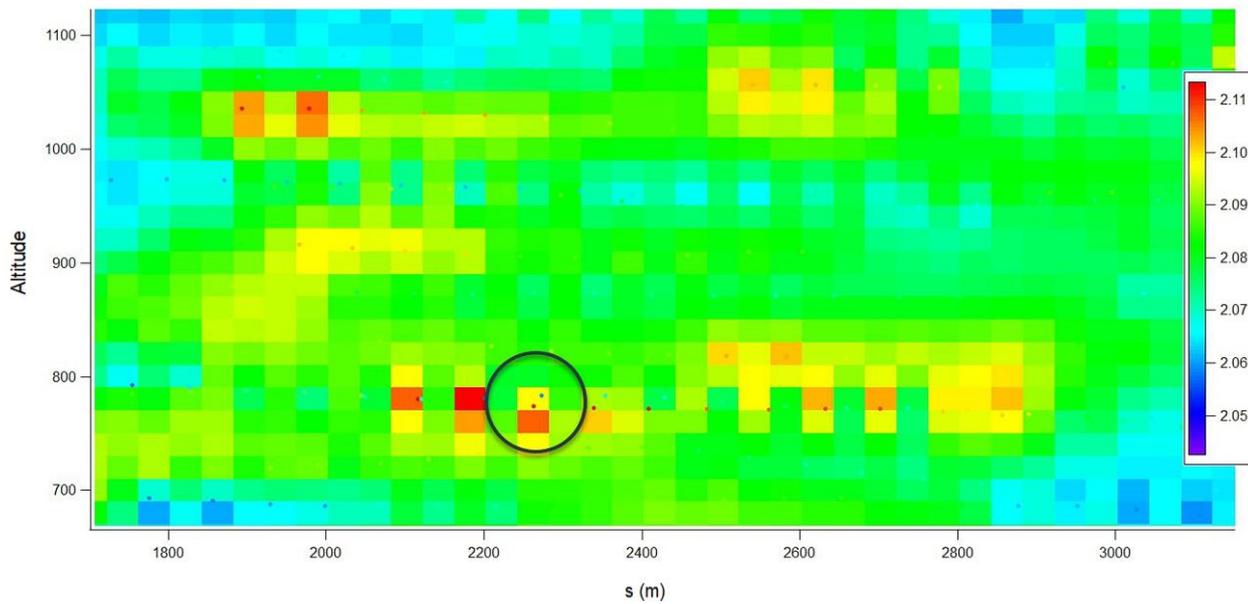
Box-flight Algorithm Assumptions Investigated

1.6 Flight Screening and Analysis of Meteorological Conditions During Flight F04

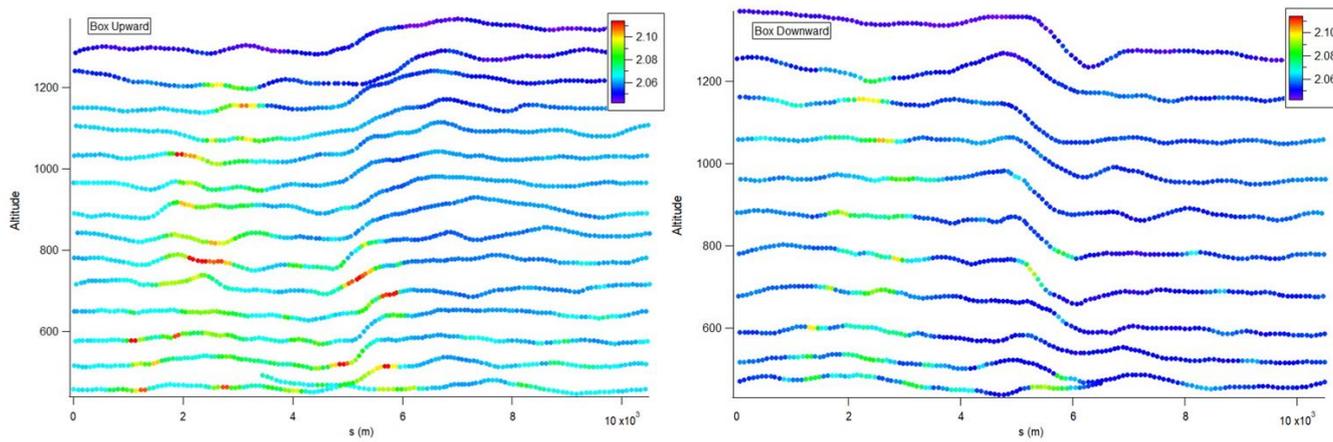
285 The potential cause of the non-stationarity of F04 was thoroughly investigated. Analysis of the raw data and emission profiles indicates that the emission plume itself was non-stationary as the mixing ratios enhancements noticeably differed between the upwards and downwards portion of the flight,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

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

300 The non-stationarity of F04 was not confirmed until post-hoc analysis was applied by Environment and Climate Change Canada. The TERRA step of kriging the data uses the lap data to produce spatially gridded emission mixing ratios. Figure S 20 shows the middle step of gridded screen for F04 CH₄. Some boxes contained a high (red) emission and a low (blue) emission enhancement (Figure S 16) 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
305 conditions were not shared by the facility. The non-stationarity of the emission plume biased the TERRA estimates as the large mixing ratio differences in each square led to lower estimates from the kriging. For TERRA, there is a noticeable difference in the 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 S 17).



310 **Figure S 16:** 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.

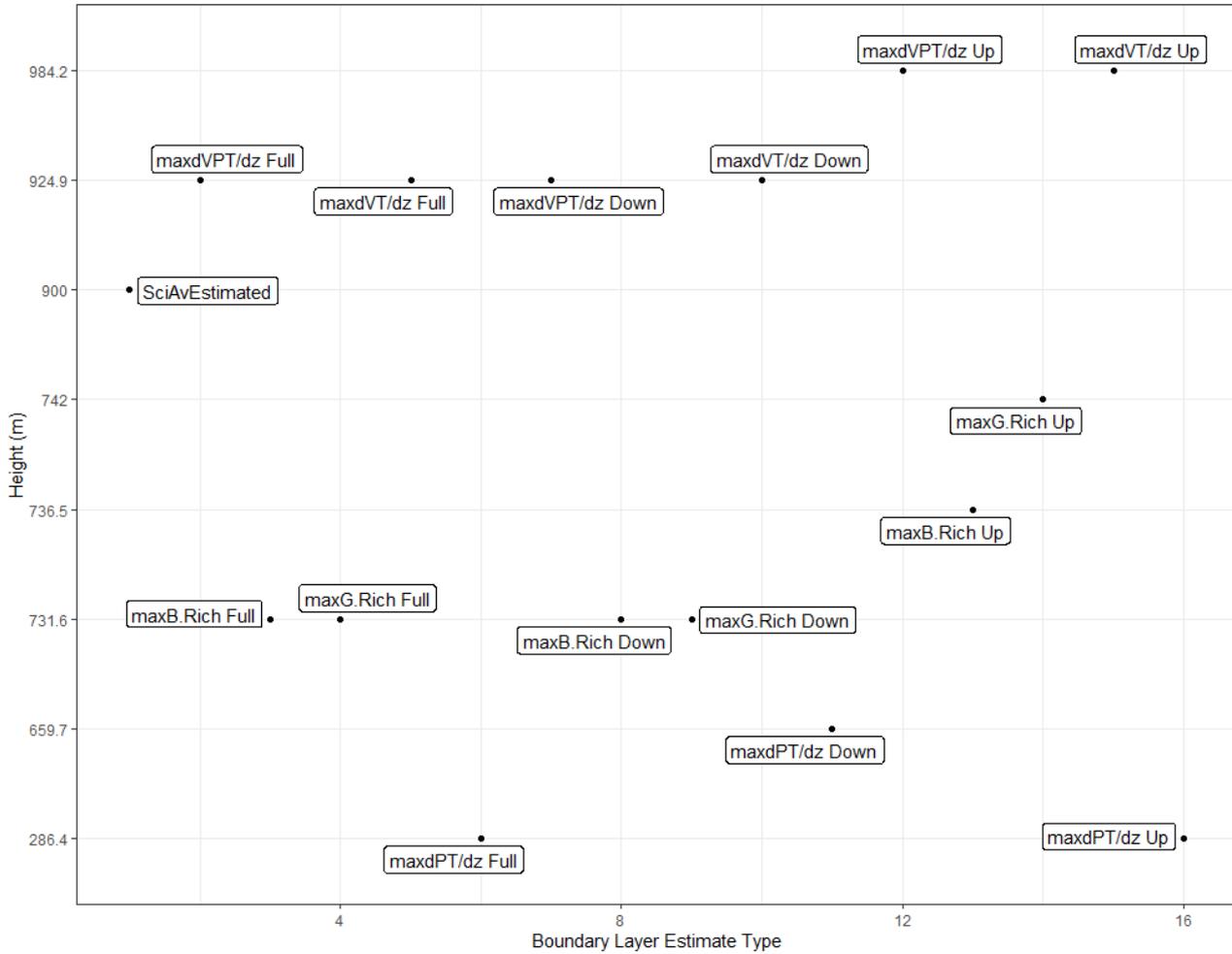


315 **Figure S 17:** Per lap CH₄ mixing ratio screen data for the box upward (left) and downward (right) prior to kriging of the data. Imagery provided by Environment and Climate Change Canada.

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) variables were computed: gradient Richardson number, bulk Richardson number, change in virtual temperature, change in potential temperature, and the change in virtual potential temperature. The boundary layer height was estimated from the maximum of each of these variables to assess any changes during the flight up versus down. These values were also compared to the SciAv estimate that was provided which was estimated the day of the flight. The calculated maximum values used to estimate the boundary layer are plotted in Figure S 18 for the full, upwards, and downwards F04 flight segments. Analysis by Dai et al. 2014 suggest the best practice is to estimate the boundary layer at maximum change in the virtual potential temperature ($\max VPT/dz$). For this flight, the boundary layer estimate is near SciAv's estimate and appears stable. If the meteorological conditions caused a significant shift in the boundary layer this would likely show up during the reanalysis. The consistency in boundary layer estimates for the flight up versus down further indicates that the boundary layer was stationary during the flight.

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

Boundary Layer Estimates F04 : 2017-08-14



340 **Figure S 18: The five boundary layer estimates for the full, up and down F04 flight segments plotted as the maximum: gradient Richardson number (maxG.Rich), bulk Richardson number (maxB.Rich), change in virtual temperature (dVT/dz), change in potential temperature (dPT/dz), and change in virtual potential temperature (dVPT/dz). The acronym dz is the change in altitude above ground level. The y-axis gives the estimated boundary layer in meters above ground level.**

345 **References**

- Conley, Stephen, Ian Faloona, 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>.
- 350 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>.
- 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>.
- Peng, Roger D. 2019. “Simpleboot: Simple Bootstrap Routines.” <https://github.com/rdpeng/simpleboot>.

355