

Optimizing Top-down Airborne Emission Retrievals through
High and Super-Resolution Numerical Modelling

SEPEHR FATHI

A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

Doctor of Philosophy (PhD)

GRADUATE PROGRAM IN PHYSICS AND ASTRONOMY
YORK UNIVERSITY
TORONTO, ONTARIO

June 2022

© 2022 Sepehr Fathi

Abstract

A multi-scale-modelling study of conventional top-down source emission-rate estimation methodologies was conducted. Two modelling systems were employed: Environment and Climate Change Canada's regional air quality model GEM-MACH at 2.5km resolution (high-resolution), and Weather Research and Forecasting (WRF) with ARW dynamical core at 50m resolution (super-resolution).

Using GEM-MACH, high-resolution air-quality model simulations were conducted for the period of an airborne campaign in 2013 over the Canadian oil sands facilities. Modelling products from these simulations were analyzed to investigate the application of the mass-balance technique in aircraft-based retrievals. The focus was on exploring the theoretical aspects and the underlying assumptions of the mass-balance technique. An extensive range of realistic meteorological and source emission conditions were considered. It was demonstrated how temporal variability in meteorology/emission conditions can give rise to storage-and-release events, where mass-balancing using only aircraft measurements can result in significant under-/over-estimates.

Using WRF-ARW, super-resolution (< 100 m) model simulations with Large-Eddy-Simulation (LES) subgrid-parameterization were developed/implemented. The objective was to resolve smaller dynamical processes at the spatio-temporal scales of the airborne measurements. This was achieved by multi-domain model nesting in the horizontal, grid-refining in the vertical, and down-scaling of reanalysis data from 31.25 km to 50 m. Further, WRF dynamical-solver source code was modified to simulate passive-tracer emissions within the finest resolution domain. Different meteorological case studies and several tracer emission sources were considered. Model-generated fields were evaluated against observational data and also in terms of tracer mass-conservation, results indicated high model performance.

Using the model output from the WRF super-resolution simulations, conventional aircraft-based retrievals were simulated/evaluated. It was shown that conventional methods can result in estimates with 30-50% uncertainty/error. Two major sources of uncertainty were identified: (a) the spatio-temporal variability in the sampled fields, and (b) the gap of information below the flight level. Optimal flight-time around one hour and sampling-distance between 10-15 km, were shown to minimize the uncertainty arising from (a)-(b). Finally, a new sampling/retrieval strategy is introduced where aircraft-based in-situ and remote measurements can be combined to improve the accuracy of top-down estimates by up to 30%. This method utilizes remote sensing to fill the information gap below flight level, characterize temporal trends in the environmental fields during flight-time (to estimate storage-rate), while reducing the required flight-time for more accurate source emission rate estimates.

Dedication

I dedicate this work to my family and friends, for their constant
love and support.

Acknowledgements

I would like to first express my deepest gratitude to Professor Mark Gordon of the Earth & Space Science Department at York University, for his invaluable supervision and continuous support during my PhD work. Professor Gordon's conduct, knowledge, and expertise have guided me throughout my career as a researcher, for which I am forever grateful.

Many thanks to my supervisory committee Professors James Whiteway and Yongsheng Chen of the Earth & Space Science Department at York University, whose insight and knowledge into the subject matter steered me through this research work. I wish to extend my appreciation to the other members of my dissertation committee for providing valuable feedback and helpful suggestions, and nominating my work for York University Faculty of Graduate Studies Dissertation prize: Professor Amir Aliabadi of School of Engineering at University of Guelph, Professor Robert McLaren of the Centre for Atmospheric Chemistry at York University, and Professor Matthew Johnson of Physics & Astronomy Department at York University.

I am extremely grateful to Dr. Paul Makar for providing invaluable insight, expertise and support throughout my research work in collaboration with the Air Quality Research Division (AQRD) at Environment and Climate Change Canada.

Thanks to Stefan Miller, Mohammad Koushifar, Xuanyi Zhang, Eric Ward, and Rhythm Reet for their support and friendship in the Air Pollution Lab at York University. I look forward to having you as colleagues and collaborators in the years ahead.

Finally, I must express my very profound gratitude to my family and friends for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing my dissertation. This accomplishment would not have been possible without your loving support and I dedicate my dissertation to you: Masoud and Fariba, Aida, Lachin and Reza, Sina and Niloofar. Thank you!

Table of Contents

Abstract	ii
Dedication	iii
Acknowledgments	iv
Table of Contents	v
List of Tables	vii
List of Figures	xi
List of Abbreviations	xxiii
1 Introduction	1
2 Evaluating the impact of storage-and-release on aircraft-based mass-balance methodology using a regional air-quality model (GEM-MACH)	7
2.1 Introduction	7
2.2 Methods	9
2.2.1 Model description and simulations	10
2.2.2 Divergence theorem, the mass-balance technique and emission rate retrievals	12
2.3 Results and discussion	21
2.3.1 Direct retrieval (DR)	22
2.3.2 TERRA retrieval (TR) approximation	25
2.3.3 Revised TERRA retrieval (TR*)	26
2.4 Three quantitative predictors of storage-and-release events	29
2.4.1 Gradient Richardson number (Ri) and changing atmospheric stability	29
2.4.2 Plume vertical motion and variations in the direction of the transport ($\Delta_t\alpha$)	32
2.4.3 Weighted upwind to downwind concentration ratio (ϕ)	34
2.4.4 Summary of predictive methods	38
2.5 Conclusions	40
3 Tracer Transport Modelling at Super-resolution with WRF-ARW	45
3.1 Introduction	45
3.2 Methods	47
3.2.1 Case Studies	47
3.2.2 Model and Technical Setup	48

3.2.3	Divergence Theorem and the Mass-balance Technique	53
3.3	Results and Discussions	55
3.3.1	Meteorological Evaluation	55
3.3.2	Plume Characteristics	62
3.3.3	Tracer Mass Conservation and Emission Rate Calculation	68
3.4	Conclusions	78
4	Optimizing Top-down Emission Rate Retrievals	81
4.1	Introduction	81
4.2	Methods	83
4.2.1	Mass-Balance technique: Control Volumes and Downwind Screens	84
4.2.2	Flying Point (FP) - aircraft-based in-situ measurements	86
4.2.3	Rolling Column (RC) - aircraft-based remote measurements	88
4.3	Results and Discussions	91
4.3.1	Error as a function of sampling time and downwind distance	96
4.3.2	Flying Point (FP): conventional airborne approach	101
4.3.3	Optimal flight time/radius: recommendations for FP airborne retrievals . . .	107
4.3.4	Rolling Colum (RC): combined in-situ and remote retrievals	109
4.4	Conclusions	115
5	Conclusions	117
	Bibliography	121
	Appendices	129
	Appendix A: GEM-MACH configuration (Table A.1) & Mass-balance terms	129
A.1	Derivations for mass-balance terms	129
A.2	Discrete integral expressions (Tables A.2–A.4)	130
	Appendix B: Storage-release as a function of downwind distance	133
	Appendix C: Prognostic TKE Closure	134
	Appendix D: Mass-balance turbulent flux terms	134
	Appendix E: WRF model meteorological and tracer profiles	135
E.1	Case 1: August 20, 2013	135
E.2	Case 2: August 26, 2013	136
E.3	Case 3: September 2, 2013	141
	Appendix F: Links to online time-lapse videos of model output (Case 1)	146
	Appendix G: Flying Point & Rolling Column analysis	146

List of Tables

2.1	Summary of the three retrieval methods: direct retrieval (DR), TERRA retrieval (TR) approximation and revised TERRA retrieval (TR*). See Eqs. (2.4)–(2.15) and equations in Appendix A for descriptions of each term.	21
2.2	Summary of correlations between NRMS error in TR estimates and the three forecast parameters for storage-release events, in terms of the Pearson correlation coefficient (P_r), and fit line (least squares) slopes (m) in Figs. 2.8, 2.10 and 2.12. The last row indicates the amount of change in forecast parameters that would result in $\geq 30\%$ error in estimates. $\phi < 10\%$ corresponds to $< 30\%$ error.	39
2.3	A summary of the meteorological conditions and source emission scenarios by the three oil sands facilities for the period of the nine JOSM 2013 box flight cases in our GEM-MACH model simulations. Flight date-time and duration is provided for each case, LT stands for local time (UT–6 h). The flight time average $\overline{R_i}$, wind speed and direction at plume height are shown. The three forecast parameters are also provided: $\Delta_t R_i$, $\Delta_t \alpha$ and ϕ . DR, TR and TR* method estimates are compared to model input emissions (MIEs). Storage (S) and/or release (R) events, their order of occurrence (for the cases where both were present) and their relative magnitude (the bigger contribution is represented by boldface letters) are given. The flight time average storage rate as a fraction (%) of the total retrieved emission rates are also provided for each case. The performance of the three retrieval methods against MIE, are shown in terms of NM and NRMS errors.	43
3.1	Three case studies during late August and early September of 2013 over oil sands facility CNRL. Times and locations for the case studies were chosen from three JOSM 2013 box flights.	48
3.2	Eleven tracer emission scenarios including seven point sources representing release from stack tops at various heights, three area sources including a large area tailing pond towards the western side of the facility and two smaller (in area) surface mines, and a line source approximately spanning the extent of the Horizon Highway south of the facility. Note that height/locations for sources with superscript † are hypothetical.	49

3.3	Case study model setup for model simulations with 5 domains with increasing resolution. The first 3 domains have the same coarse vertical grid. Domains d04 and d05 have increasing resolution via vertical grid refinement. The finest domain (d05) has $\Delta z = 11.62$ m for the first 40 levels near the surface. Δx and Δt show the horizontal grid size and the model simulation time-step for each domain, respectively. X and Y indicate domain dimensions. nx , ny and nz are the number of computational grid points in each direction. With model Z_{top} at 15.623 km (15.350 km agl) and P_{top} (pressure at model top level) at 10 kPa.	51
3.4	Evaluation of domain d05 simulations against d04 output fields at every 100 seconds at the geographical location of the CNRL oil sands facility (Lon=-111.738 and Lat=57.339). Note that positive/negative sings mean over/under-estimates by d05 relative to d04.	59
3.5	Normalized (to MIE) rates of change in tracer mass for all emission scenarios are shown. Mean rates are given for net, positive (+) and negative (-) mass. Standard deviations are also provided. Rates agree with model input emissions within 2% for 10 out 11 tracers (conserved).	74
3.6	Source emission rates (E_C) determined by performing mass-balance calculations on model output data, evaluated against model input emissions (MIE) for the three cases. Results are shown as normalized mean bias (nmb) in %. Contribution of three main terms in the mass-balance equation are shown, normalized to MIE (%). Values less than 0.1% are indicated with ϵ	76
4.1	Case 1 FP estimates of flux out are evaluated against the true $F_{C,out}$ and results are presented as the normalized (to true $F_{C,out}$) mean bias. Extrapolation methods zero p_0 , constant p_c , linear p_l , Gaussian p_g are compared for the the rectangular box, cylindrical box and the downwind screen estimates. Note that the large values for CNRLw and CNRLs for the two boxes are the results of division by very small $F_{C,out}$ values (normalization). Bold text is used for indicating the best extrapolation result in each category. Absolute values less than 0.01 are indicated with ϵ	106
4.2	Case 1 FP+RC ₀ and full RC (with 10-12 laps) generated estimates evaluated against the true model $F_{C,out}$ for the two flux boxes and the downwind screen (at 5 km). Results are shown in terms of normalized mean bias (NMB). Sampling was conducted at horizontal flight speed of 100 m/s around and downwind the emission sources for all sampling volumes/screen.	113
4.3	Case 1 normalized RMS error of the RC generated $F_{C,out}$ and E_C estimates, with S_C estimated according to Eq. 4.3 with 10-12 laps, evaluated against model Truth. Note that values in [brackets] for CNRLw and CNRLs are rms values (not normalized) due to E_C being very small for the upwind cases (e.g., CNRLs $E_C = 0.002$).	114
A.1	GEM-MACH Configuration details and references	131

A.2	Discrete integral expressions of Eqs. (2.4), (2.5), (A.1) and (A.3) to be substituted in Eq. (2.7). Constants Δx , Δy and Δs are equal to model grid resolution in the horizontal, 2.5 km. t is the time index; i , j and k are the 3D-space indices and s is the path index around the box.	132
A.3	Discrete integral expressions for Eqs. (A.4) and (2.8)–(2.10) to be substituted in Eq. (2.12) for TR method estimations. Terms and variables are listed in the order needed for calculating the dependent terms. Δs is constant and equal to model horizontal grid resolution, 2.5 km. t and s are time and path indices, respectively. A is the box base area. Point (i_o, j_o) is at the box horizontal center.	132
A.4	Discrete integral expressions for Eqs. (2.13)–(2.15) to be substituted in Eq. (2.16) for TR* method estimations. t is the time index; s and k are the horizontal and vertical space indices along the screens, respectively. A is the box base area.	132
G.1	Case 1 breakdown of FP sampled and unsampled portions of tracer plumes for the three sampling screen/volumes. Results are presented in three groups: CNRL stack plumes, surface plumes, and the upwind plumes. From top to bottom of the table, the downwind distance from source to sampling screen increases, except for upwind emissions from CNRLw and CNRLs. These two plumes were first sampled at the upwind wall (short distance) and then a second time at the downwind wall (long distance) of the two boxes. Release heights for CNRLs and CNRLw were below the lowest flight level (similar to other sources except CNRL0). Consequently >50% of plumes fell below the lowest flight level on the upwind wall, hence large unsampled % for CNRLs and CNRLw for the two boxes.	147

List of Figures

- 2.1 **(a)** GEM-MACH model horizontal grid structure with nested domains: the dark blue region shows the parent domain over North America at 10 km horizontal grid resolution. Output from GEM-MACH 10 km simulations provided initial and boundary conditions for the nested high-resolution GEM-MACH 2.5 km domain (red). The high-resolution domain was configured over the Canadian provinces of Alberta and Saskatchewan including the Alberta Athabasca oil sands region, shown as a dark shaded area. **(b)** © Google Earth map of the region with three oil sands facilities labelled. JOSM 2013 flight tracks on 14, 20 and 28 August are drawn with white curves. Model 4D fields were extracted for nine JOSM 2013 box flight cases over these three facilities, examples are shown as overlay SO₂ concentration maps. 10
- 2.2 SO₂ mixing ratio screen around CNRL extracted from the GEM-MACH model output at the start of case 4 on 20 August 2013 at 16:30 UT (10:30 LT). **(a)** Mixing ratio data (2D) on the lateral walls of the flux box, along with the flight tracks plotted as dark dots. The additional graphics (arrows and plumes) on **(a)** represent processes contributing to change in tracer mass within the box volume: source emission rate (E_C), horizontal flux ($E_{C,H}$), vertical flux ($E_{C,V}$), surface deposition ($E_{C,VD}$) and storage rate ($E_{C,S}$). The extracted fields around the box are unwrapped onto a 2D screen comprised of box lateral walls (east, north, west and south) and shown in **(b)** the normal air flux (positive outwards), **(c)** SO₂ mixing ratio and **(d)** SO₂ mass flux through the screen. Note that the net flux on box corners is the vector sum of the fluxes in west–east (U) and south–north (V) directions. 16
- 2.3 Emission rates determined by the three methods (DR, TR and TR*) are plotted against model input emissions (MIEs). The dark solid curve is the 1 : 1 MIE line; estimates below and above this line represent under- and over-estimation by the three methods. Case IDs are noted for each set. The vertical lines represent the error associated with storage-release events ($E_{C,S}$) in TR estimates relative to MIEs. Also shown in the figure are the $\pm 50\%$, 30% and 20% error limits (dashed lines) for reference. 22

2.4	Contribution (%) of different terms in Eq. (2.7) to the total retrieved emissions (E_C), for the nine JOSM 2013 cases. The 100 % contribution level is marked with a (purple) dot-dashed line, $\pm 30\%$ and 50% levels are also marked with (red) dashed lines on the figure. The main contributing terms were $E_{C,H}$ and $E_{C,S}$ ($> 97\%$). The other terms contributed $< 3\%$ combined.	23
2.5	Time series of the estimates by DR, TR and TR* methods, plotted along with MIEs. Normalized-root-mean-square (NRMS) error for the three methods are noted on the plot for each of the nine JOSM 2013 cases. JOSM 2013 flights on 24 August (case 2) and 2 September (case 6) included two consecutive box flights around the same facility, shown as shaded areas on their respective time-series plots. Note that the DR and TR* methods, unlike the TR method, account for the storage rate ($E_{C,S}$ and $E_{C,S}^{tr*}$).	24
2.6	This three-panel graph demonstrates the correlation between estimates of $\partial\chi_C/\partial t$ by analyzing 4D (box) and 3D (consecutive screens) data sets. The vertical side plot shows the range and distribution of $\partial\chi_C/\partial t$ estimates for the period of nine studied flights by 4D analysis, the horizontal side plot shows the same for 3D analysis. The main panel demonstrates the correlation between the two methods of estimation. Pearson correlation coefficient ($P_r = 0.74$) shows the strength of this relation. The least-squares fit line has slope $m = 1.36$ and intercept $b = 0.09$. Note that case 8 estimates are in low correlation due to relatively high upwind emissions during the time of that particular flight case (28 August), rendering the 3D estimates less representative of the 4D estimates. Excluding case 8 from this analysis, results in $P_r = 0.88$, $m = 1.07$ and $b = 0.03$. $E_{C,S}^{tr*}$ can be estimated from multiple consecutive 2D screens, by assuming 3D estimations of $\partial\chi_C/\partial t$ as representative of the rate of change in species mass within the flux box. The same can also be estimated from downwind vertical profiling of tracer concentrations over the sampling time.	27
2.7	This schematic demonstrates (qualitatively) the correlation between change in atmospheric stability represented by the gradient Richardson number (Ri), plume vertical motion and storage-release events for the period of case 4. (a) Ri vertical profiles at stack location over the flight time, with the color-scale symmetric around the critical gradient Richardson number $Ri_C = 0.25$. (b) SO_2 mixing ratio vertical profiles at stack location. (c) Time series of net horizontal flux ($E_{C,H}$) and (d) the storage rate ($E_{C,S}$), over sampling time. Note the level of anti-correlation between $E_{C,H}$ and $E_{C,S}$ ($P_r = -0.95$); oscillations in storage and release are occurring: changes in $E_{C,S}$ are matched by opposite changes in $E_{C,H}$. Wind speed and direction are shown as barb plots with units $m s^{-1}$ (half = 5, full = 10). For every incremental rise in plume height, a corresponding release event occurs. As the plume ascends into regions with faster wind speeds, more mass is released ($E_{C,S} < 0$) through the lateral walls resulting in an increase in $E_{C,H}$ as can be seen in (c)	28

2.8	Correlation between the NRMS error and the change in atmospheric stability represented by the absolute change in the gradient Richardson number $ \Delta_t R_i $. The shaded areas indicate the 90 % confidence region. Case ID is indicated for each data point on the graph along with the calculated Pearson correlation coefficients (P_r) for the three methods. Results indicate a strong positive correlation between the NRMS error and $ \Delta_t R_i $	31
2.9	SO ₂ mixing ratio horizontal cross sections at two (consecutive) plume center (max. conc.) heights (meters above sea level – m.a.s.l.) and four time steps, during the period of case 5 over CNRL. The strength and direction of the airflow at each cross section is represented by grey arrows. Corresponding time and average height are noted for each panel. As the plume rises into layers with different wind direction over time, it re-circulates within the volume of the box which in turn results in a storage event.	34
2.10	Correlation between the NRMS error and the change in wind direction at plume height ($\Delta_t \alpha$). The shaded areas show the 90 % confidence region. Case ID is noted for each data point on the graph along with the calculated Pearson correlation coefficients (P_r) for the three methods. Results indicate a positive correlation between the NRMS error and $\Delta_t \alpha$	35
2.11	Case 8 screens around Suncor: (a) air flux screen, (b) SO ₂ mixing ratio screen and (c) SO ₂ mass flux screen. Note the strong negative flux (inflow) on the western wall and the weaker positive flux (outflow) on the eastern wall resulting in (temporal avg.) $\phi = 13\%$	37
2.12	Correlations between the NRMS error and the forecast parameter ϕ , for DR, TR and TR* estimates. The shaded areas indicate the 90 % confidence region. Case ID is indicated for each data point. Pearson correlation coefficients (P_r) for the three methods are also noted on the graph. Results indicate a strong correlation between NRMS error and the ϕ parameter.	38
3.1	©Google map of the study region: The oil sands facility CNRL with case study emission sources marked on the map including seven point sources, two area sources, a large area surface source (tailing pond) and a multi-segment line source (road/Hwy). Direction of north (N) is shown with a compass arrow.	50
3.2	(a) WRF model grid (horizontal) configuration with five nests d01 - d05 with increasing resolution (decreasing size) at a ratio of 1 to 5. The finest domain, d05, is centred over the oil sands region. (b) ©Google map of the oil sands region with overlay of model domains d04 and d05. The CNRL facility is marked with a red star. Locations for two WBEA monitoring stations are also shown with red triangles, Bertha Ganter – Fort McKay and Barge Landing monitoring stations.	52

3.3	Topography of the modelling region for domain d05 (finest domain) is shown, lower elevations such as the Athabasca river valley is shown in dark blue and higher elevations are shown with colours green and brown. Our WRF model simulations use terrain-following model layers to ensure continuity in the modelling layers and proper treatment of the interaction between the atmospheric flow and the topography. Two model layers at levels 42 and 55 are also shown (mesh).	53
3.4	Model output sea level pressure (top left), 2-m relative humidity (top right), 2-m temperature (bottom left), and 10-m wind (bottom right) for domains d03, d04 and d05 on 20 August 2013 are shown. Boundaries of domains d04 (red rectangle) and d05 (blue rectangle) are also shown. Contours show terrain height in meters above seal level (masl). The geographical location of the CNRL facility is indicated with a star in the middle of domain d05. Output from the three domains agree to a good extent, with small overestimation in wind speed for d05 compared to the other two. Note that in the bottom-right panel, the arrow sizes are not representative of wind speeds, but model grid spacing (resolution).	56
3.5	Wind-rose diagrams comparing observational data from WBEA monitoring station Bertha Ganter–Fort McKay to model output winds at the location of the WBEA station from domains d03, d04 and d05 for 20 August 2013. Wind directions are consistent within 20 degrees, blowing from w-s-w. Wind speeds are biased high at various degrees for the three domains.	57
3.6	Model output for 2-m temperature, 2-m relative humidity and 10-m wind speed from domains d04 and d05 are compared to observational data from WBEA monitoring stations.	58
3.7	Model-generated wind fields (orange) compared to aircraft observations (blue dots) at several altitude levels during the 2013 JOSM airborne campaign, and WBEA observational data at 10 m agl (red diamonds). Horizontal bars show one standard deviation in model wind fields. Aircraft data and model fields agree within one standard deviation. WBEA wind speeds are lower than both aircraft and model wind speeds, red bars show one standard deviation in WBEA data.	60
3.8	Model output meteorological fields from the two fine resolution domains d04 and d05 evaluated against d03 output for the location of CNRL facility. d04 and d05 output fields show good agreement to d03 for sea level pressure, humidity, temperature and wind. d05 wind speeds were biased high by 1.57 m/s and 3.47 m/s.	61

3.9	Top view of the modelling domain (d05) and the location of tracer emission sources are shown. Sources include a big area source (Pond), a multi-segment line source (Highway), two small area sources (Mines 1, 2), five CNRL stacks 0-4 and two upwind point sources (west and south). Emission plumes from Pond are shown in blue, Highway in gray, Mines in green and stack/point sources in red. Colour darkness is proportional to log of column total tracer mass. Two control volumes (dark dashed boxes) for 4D mass-balance calculations are shown, one enclosing all the emission sources and the other marking the boundaries of the CNRL facility (the smaller box) and only including within facility sources.	63
3.10	Case 1: top view of the 50 m resolution modelling domain with tracer plumes from our eleven emission scenarios at 16:30UT on 20 August 2013. Tracer plumes are dispersed and advected well beyond the boundaries of the CNRL facility, shown with the dashed curve (same as the small box in Fig. 3.9). 10-m wind data is shown with wind barbs over the entire domain (flag: half 5 m/s, full 10 m/s). The shown data is the tracer column total normalized to the maximum value for each panel. (top left) point/stack sources CNRL0-4, CNRLs and CNRLw. (top right) surface/area sources MINE1 and MINE2. (bottom left) the large area (surface) sources POND. (bottom right) the multi-segment line (surface) source HWY. Refer to Table 3.2 for source specifications.	64
3.11	Domain d05 west-east vertical cross-section for case 1 on 20 August 2013. Vertical cross-section of tracer plumes from stack/point emission sources are shown. Release height above ground is indicated for each source. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for west-east distance in km is at domain centre.	65
3.12	Domain d05 south-north vertical cross-section for case 1 on 20 August 2013. Vertical cross-section of tracer plumes from stack/point emission sources are shown with release heights indicated for each source. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for south-north distance in km is at domain centre.	67
3.13	Domain d05 west-east vertical cross-section for case 1 on 20 August 2013. Vertical cross-section of tracer plumes from surface emission sources (area, line) are shown. Release height above ground is 6 m agl for all sources. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for west-east distance in km is at domain centre.	68
3.14	Domain d05 south-north vertical cross-section for case 1 on 20 August 2013. Vertical cross-section of tracer plumes from surface emission sources (area, line) are shown with release heights at 6 m agl for all sources. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for south-north distance in km is at domain centre.	69

- 3.15 Case 2: top view of the 50 m resolution modelling domain with tracer plumes from our eleven emission scenarios at 19:43UT on 26 August 2013. 10-m wind data is shown with wind barbs over the entire domain (flag: half 5 m/s, full 10 m/s). The shown data is the tracer column total normalized to the maximum value for each panel. (top left) point/stack sources CNRL0-4, CNRLs and CNRLw. (top right) surface/area sources MINE1 and MINE2. (bottom left) the large area (surface) sources POND. (bottom right) the multi-segment line (surface) source HWY. For this case tracer plumes were not advected considerably, due to low and spatially variable winds. Refer to Table 3.2 for source specifications. 70
- 3.16 Case 3: top view of the 50 m resolution modelling domain with tracer plumes from our eleven emission scenarios at 17:43UT on 2 September 2013. Tracer plumes are dispersed and advected well beyond the boundaries of the CNRL facility (dashed curve) as in case 1. Wind data at 10 m agl is shown with wind barbs over the entire domain (flag: half 5 m/s, full 10 m/s). The shown data is the tracer column total normalized to the maximum value for each panel. (top left) point/stack sources CNRL0-4, CNRLs and CNRLw. (top right) surface/area sources MINE1 and MINE2. (bottom left) the large area (surface) sources POND. (bottom right) the multi-segment line (surface) source HWY. Refer to Table 3.2 for source specifications. 71
- 3.17 The ratio of tracer mass present within the modelling domain to mass emitted over simulation time after the start of tracer release. Time series are separated into positive and negative mass in the bottom two panels. Tracer mass remains conserved over time, with the exception of CNRL 0. Note that the drop in present-to-emitted mass ratio beyond 16:30UT, marked by the vertical dashed line, is due to tracer mass exiting through domain boundaries and not a violation of mass conservation. . . . 72
- 3.18 Temporal rate of change in tracer mass for the entire domain $S_{C,d05}$ is shown. Time series are normalized to model input emission (MIE) rates. The net rate of change (top panel) is separated into rates for positive and negative mass in the bottom two panels. The creation of negative mass is offset by creation of excess positive mass. Rate of change in tracer mass match MIE (conserved) for all except one emission source (CNRL0). 73
- 3.19 Normalized (to model input emissions - MIE) retrieved emission rates for tracer emission scenarios of case 1. Emission rates (E_C) were calculated using the mass-balance equation Eq. 3.7 and the control volume shown in Fig. 3.9. The breakdown (%) of contributing terms are shown as absolute values, see Table 3.6 for the actual values. The vertical axis is in log scale. Refer to Table 3.2 for source specifications. . 75

3.20	Three panel snapshot of MINE 1 plume during case 1 at 5 min intervals. Note that the erroneous negative mass creation (blue contours) is concentrated upwind of the emission source and on plume edges. Consequently, 98% of the domain total negative mass was located within the control volume along with only 43% of total positive mass (red). The vertical dashed line shows the location of control volume downwind wall.	77
4.1	Top view of the modelling domain with tracer plumes from point/stack emission sources within the CNRL facility and two upwind sources (top left). Sampling volumes/surface for rectangular box, cylindrical box and downwind screen (at 5 km distance) are also superimposed on the map with dashed lines. The 3D projections of the two boxes (control volumes) and the downwind screen are also shown in different panels of the figure, viewed from north-east. Note that the vertical scale is exaggerated, sampling volumes/surfaces extend tens of kilometres in the horizontal and about 2 km in the vertical.	84
4.2	JOSM box flight on 2 September 2013 over CNRL. Topography of the region is shown as a 3D surface. Flight path is shown with dark dots. Flight started on the east side of CNRL near Athabasca river basin (shown in blue) in counterclockwise direction in a rectangular path, while gradually climbing in altitude. For our purposes the aircraft can be considered as a dimensionless point, and therefore herein we refer to this sampling strategy as Flying Point (FP).	86
4.3	Rolling column sampling strategy includes extracting/sampling vertical columns of environmental data around (downwind of) the emission source. During field campaigns this can be achieved by using aircraft-based remote measurement instruments such as LIDAR. By making column measurements, environmental fields will be sampled from the ground surface to the box top, eliminating the need for data extrapolations. This method can generate complete control volumes/screen at shorter time periods of about 10-15 min.	88
4.4	Transport and dispersion of tracer plume released from the CNRL4 source is demonstrated in three panels. Tracer is released at 17:00UT, it reaches the downwind screen at 17:05UT, the cylindrical box at 17:15UT and the boundaries of the rectangular box at 17:20UT. For reference, the rectangular-box dimensions are 20 km $\times$ 15 km. The spatial heterogeneity of the 10-m wind field is shown with wind barbs (flag: half=5m/s, full=10m/s). Note how the plume spreads as it is advected downwind and how the advection path (plume centre line) is altered along the way, as it interacts with the topography (passing over the Athabasca river basin at about 240 meter above sea level) and the heterogeneous wind field.	91

- 4.5 Time-series signals for the components of the mass-balance equation (Eq. 4.1) for tracer CNRL4 in Fig. 4.4. Note how S_C and $F_{C,out}$ are anti-correlated and their dependence on the downwind distance of the point of interception from the emission source. The oscillatory behaviour in the signals (period, amplitude) shows correlation to the downwind distance (box size), the bigger the distance the longer the period. . 93
- 4.6 Contribution breakdown of the terms in the mass-balance equation (Eq. 4.1) for case 1, where the mean retrieved emission rate E_C is shown with green bars, mean flux out $F_{C,out}$ with blue and the mean storage rate S_C with orange bars. Estimates for the rectangular box, cylindrical box, and the downwind screen are shown in their respective panels. Results show small contribution of S_C for stack/point sources (CNRL1-4) and relatively larger contribution for surface sources and upwind sources (CNRLs and CNRLw). This case represent desirable conditions for top-down retrievals based on $F_{C,out}$ alone as the contribution of S_C is small for most sources. . 94
- 4.7 Contribution breakdown of the terms in the mass-balance equation (Eq. 4.1) for case 2, where the mean retrieved emission rate E_C is shown with green bars, mean flux out $F_{C,out}$ with blue and the mean storage rate S_C with orange bars. Results for the Rectangular box, cylindrical box, and the downwind screen are shown in their respective panels. Results show the significant contribution of S_C for all emission and sampling scenarios and the relatively small contribution of $F_{C,out}$. This case of severe storage-release event represent undesirable conditions for top-down retrievals based on $F_{C,out}$ alone as the contribution of S_C is dominant in mass-balance equation. 95
- 4.8 Contribution breakdown of the terms in the mass-balance equation (Eq. 4.1) for case 3, where the mean retrieved emission rate E_C is shown with green bars, mean flux out $F_{C,out}$ with blue and the mean storage rate S_C with orange bars. Results for the rectangular-box, cylindrical-box, and the downwind-screen are shown in their respective panels. Results show moderate contribution of S_C for most of the emission scenarios and relatively larger contribution for POND, MINE 1 and CNRLw. This case represent conditions where top-down retrievals based on $F_{C,out}$ would be more significantly improved (relative to case 1) with estimates of S_C 96
- 4.9 Case 1 mass-balance estimates based on flux out ($F_{C,out}$) alone. (top-panel) $F_{C,out}$ time-series for select emission sources, normalized to model Truth. (bottom-panel) normalized mean bias (NMB) in estimates as a function of averaging period. Different averaging intervals are marked in the top-panel with vertical dashed lines. Error decreases as the averaging period increases for most sources, as shown in the bottom-panel. Error increases beyond 40 min for source emissions with relatively bigger contribution from the neglected storage rate term S_C (compare to Fig. 4.6). . 97

4.10	Box-plots of case 1 $F_{C,out}$ (normalized to input emissions rates) for several point and area sources during the sampling time of 2 hr and 10 min (at 1 Hz sampling frequency). S stands for the downwind-screen, C for the cylindrical, and R for the rectangular-box estimates. Note how the spread (and outliers) are larger for point sources CNRL1 and CNRL3. This spread is smaller for surface sources POND, MINE1 and MINE2. Surface sources are further from the sampling screens than the point sources.	99
4.11	Temporal standard deviation (σ) as a function of downwind distance for different emission scenarios. The locations of the sampling surfaces downwind screen (S), cylindrical (C) and rectangular box (R) are shown with dashed lines, and their relative distances (to source) are indicated. The temporal variability in tracer vertical mixing decreases as a function of downwind distance. The rate of this decrease is faster for the surface emissions (POND, MINE1, MINE2) compared to stack emission (CNRL1, CNRL3).	100
4.12	Flux estimations with multiple screens at 5 km consecutive downwind distances from the emission source. The location of the downwind screens are shown with dashed curves in the main and the top panels. The top panel shows the decrease in vertical variability (σ) as a function distance. The right panel shows the same along the horizontal extent of each consecutive screen (normalized to the max σ on the first screen). At larger downwind distances, σ drops to low levels as the plume mixes more in the 3D space.	101
4.13	The sampling strategy of flying point (FP) is demonstrated for the cylindrical and rectangular boxes. For the shown case, relevant fields are sampled around the facility and along the flight path that starts at about 150 m agl and climbs in altitude in counterclockwise direction. Tracer amount from the surface emission source MINE 2 along the flight path are shown.	102
4.14	Case 1 FP sampled rectangular-box for CNRL1 is shown. The FP sampled region was interpolated using the kriging method (top left). FP sampled/interpolated data are then extrapolated below the flight level for calculating the net flux out $F_{C,out}$ (bottom left). Model vertical profiles during the first and last pass/lap at the horizontal location of FP-sampled maximum are shown in the right panel (dashed lines). FP sampled points are shown with squares, and the kriging interpolated profile with diamonds. Extrapolation methods (constant p_c , linear p_l , Gaussian p_g) applied to sampled/interpolated vertical profile at the same horizontal location are also shown.	104

4.15	Vertical profiles of tracer amounts (normalized to max.) at the horizontal location of the plume maximum (centre) for the eleven emission sources. Note that these profiles are for different timestamps (indicated on top of each panel) and are at different horizontal locations (for different plume centrelines) along the rectangular-box walls. Vertical mixing is diverse and variable in time and space for different emission sources.	105
4.16	Tracer vertical profiles at consecutive 5 km downwind distances. Profiles for stack emission CNRL0 and CNRL1, remain non-uniform as a function of height up to 10 km. These approach constant profiles at about 15 km. The same is true for surface emissions, however they approach near constant profiles at shorter distances. The 15 km distance seems to be the optimal sampling radius, where data extrapolation below flight level can be done with more confidence.	108
4.17	Application of Rolling Column (RC) method for top-down retrievals. POND emissions on the cylindrical box are shown. (a) FP measurements combined with column measurements during the first/lowest pass around the facility RC_0 (first 10 min of the sampling time). (b) RC column measurements extended for the entire extent of the box (ground surface to the box top) with up and down facing remote column measurements. (c) RC sampling linearly interpolated in the horizontal to generate complete flux screens for every 10-12 min of the flight time.	110
4.18	RC generated time-series of $F_{C,out}$ for select sources using the cylindrical box are shown (orange step-plots). RC estimates follow the true model $F_{C,out}$ signal and capture the temporal variability in the exit flux. Normalized mean bias (nmb) scores are also indicated for each source.	112
4.19	Rolling Column (RC) estimates for case 1, source CNRL1 and rectangular-box with S_C estimated based on Eq.4.3. Normalized root mean square (nrms) error scores are mentioned for RC generated E_C and $F_{C,out}$ estimates evaluated against model true values for the same terms.	113
B.1	The contribution of the storage rate term ($E_{C,S}$) to the net retrieved emission rates (E_C) as a function of downwind distance for case 8. As shown in the wind-rose diagram, the wind direction for this case was towards south-east (S-E), and therefore the downwind dependence is shown in two directions west-east (green squares) and north-south (red circles). The contribution of storage increases as a function of downwind distance. This suggests that the impact of storage release may increase as the distance between the downwind screen (sampling location) and the emission sources increase. A similar dependence of the magnitude of the storage release with downwind distance was observed for the other cases.	133

E.1	Case 1 mean meteorological profiles on 20 August 2013: potential temperature θ , absolute temperature T , virtual temperature T_v , west-east wind component U , south-north wind component V , and vertical wind W . Bars show standard deviations during the simulation time.	135
E.2	Case 2 mean meteorological profiles on 26 August 2013: potential temperature θ , absolute temperature T , virtual temperature T_v , west-east wind component U , south-north wind component V , and vertical wind W . Bars show standard deviations during the simulation time.	136
E.3	Domain d05 west-east vertical cross-section for case 2 on 26 August 2013. Vertical cross-section of tracer plumes from stack/point emission sources are shown. Release height above ground is indicated for each source. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for west-east distance in km is at domain centre.	137
E.4	Domain d05 south-north vertical cross-section for case 2 on 26 August 2013. Vertical cross-section of tracer plumes from stack/point emission sources are shown with release heights indicated for each source. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for south-north distance in km is at domain centre.	138
E.5	Domain d05 west-east vertical cross-section for case 2 on 26 August 2013. Vertical cross-section of tracer plumes from surface emission sources (area, line) are shown. Release height above ground is 6 m agl for all sources. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for west-east distance in km is at domain centre.	139
E.6	Domain d05 south-north vertical cross-section for case 1 on 20 August 2013. Vertical cross-section of tracer plumes from surface emission sources (area, line) are shown with release heights at 6 m agl for all sources. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for south-north distance in km is at domain centre.	140
E.7	Case 3 mean meteorological profiles on 2 September 2013: potential temperature θ , absolute temperature T , virtual temperature T_v , west-east wind component U , south-north wind component V , and vertical wind W . Bars show standard deviations during the simulation time.	141
E.8	Domain d05 west-east vertical cross-section for case 3 on 2 September 2013. Vertical cross-section of tracer plumes from stack/point emission sources are shown. Release height above ground is indicated for each source. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for west-east distance in km is at domain centre.	142

E.9	Domain d05 south-north vertical cross-section for case 3 on 2 September 2013. Vertical cross-section of tracer plumes from stack/point emission sources are shown with release heights indicated for each source. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for south-north distance in km is at domain centre.	143
E.10	Domain d05 west-east vertical cross-section for case 3 on 2 September 2013. Vertical cross-section of tracer plumes from surface emission sources (area, line) are shown. Release height above ground is 6m agl for all sources. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for west-east distance in km is at domain centre.	144
E.11	Domain d05 south-north vertical cross-section for case 1 on 20 August 2013. Vertical cross-section of tracer plumes from surface emission sources (area, line) are shown with release heights at 6 m agl for all sources. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for south-north distance in km is at domain centre.	145
G.1	Rolling Column (RC) estimates for case 3, source CNRL1 and cylindrical-box with S_C estimated based on Eq. 4.3. Normalized root mean square (nrms) error scores are mentioned for RC generated E_C and $F_{C,out}$ estimates, evaluated against model true values for the same terms.	147

List of Abbreviations

agl	above ground level
amsl	above mean sea level
AVIRIS	Airborne Visible/Infrared Imaging Spectrometer
AQ	Air Quality
AQRD	Air Quality Research Division
ARW	Advanced Research WRF
ACP	Atmospheric Chemistry and Physics
CNRL	Canadian Natural Resources Ltd.
DIAL	Differential Absorption Lidar
ECCC	Environment and Climate Change Canada
FP	Flying Point
GEM-MACH	Global Environmental Multiscale-Modeling Air-quality and CHemistry
HYSPLIT	Hybrid Single Particle Lagrangian Integrated Trajectory
ILMC	Iterative Locally Mass-Conserving
JOSM	Joint Canada-Alberta Implementation Plan on Oil Sands Monitoring
LES	Large Eddy Simulation
LIDAR	Light Detection and Ranging
LT	Local Time
masl	metres above sea level
ME	Mean Error
NARR	North American Regional Reanalysis
NMB	Normalized Mean Bias
NME	Normalized Mean Error
NOAA	National Oceanic and Atmospheric Administration
NRMSE	Normalized Root Mean Square Error
MIE	Model Input Emission
RC	Rolling Column
RMS	Root Mean Square
SR	Super-Resolution
TERRA	Top-down Emission Rate Retrieval Algorithm
UAV	Unmanned Air Vehicle
UT	Universal Time
UTC	Coordinated Universal Time
WBEA	Wood Buffalo Environmental Association
WRF	Weather Research and Forecasting

Chapter 1

Introduction

Aircraft-based measurements of air pollutants provide an effective top-down approach for estimating total integrated emissions of sources ranging from individual stacks to large industrial complexes and cities. These top-down estimates are commonly employed for the evaluation of bottom-up emissions reported to inventories and can provide insights on air quality and health impacts of anthropogenic emissions. Top-down estimates can also be used to provide highly resolved emissions data for input into air quality models. Several studies have utilized aircraft-based top-down emission rate retrievals in the past (Cambaliza et al., 2014; Conley et al., 2017; Gordon et al., 2015; Kalthoff et al., 2002; Karion et al., 2015, 2013; Mays et al., 2009; Nathan et al., 2015; Ryoo et al., 2019; Tadić et al., 2017; Turnbull et al., 2011). Flight patterns from these studies include single-height transects, single-screen flights and box flights. The box flight, which improves upon single-leg (mainly downwind) flights, refers to flight patterns with enclosed shapes (polygonal, cylindrical) and is accomplished by flying in closed paths (loops) at multiple altitudes around the emission source while making measurements of meteorological fields and pollutant concentrations. The mass-balance technique and Gauss’s divergence theorem — where the net mass flux exiting a control volume is equated to the within-volume source emission rates — can be employed to infer point and area source emission rates from aircraft measured data. Kalthoff et al. (2002) used aircraft measured data to estimate CO and NO_x emissions of the city of Augsburg in southern Germany, using data from two research aircraft at several altitudes (polygonal box) around the boundaries of the city during an airborne measurement campaign (Slemr et al., 2002). Emission rate estimations were made using a mass-balance approach applied to aircraft measurements, and these estimates were used to improve emissions input data for an air quality model (KAMM/DRAIS). Through air-quality model simulations, Panitz et al. (2002) analyzed the validity of the assumptions made by Kalthoff et al. (2002) and also studied the contributions of all relevant processes (advective fluxes, deposition, turbulent transports, and chemical transformations) to the mass budget of chemical species (e.g. CO and NO_x). They concluded that, in an aircraft-based mass-balance approach, 85% and 95% of (CO and NO_x) emissions can be determined from the advective fluxes. Gordon et al. (2015) conducted top-down aircraft-based emission rate retrievals for SO₂ and CH₄ by analyzing polygonal (e.g. rectangular) box flights around individual oil sands facilities in Alberta, Canada. This study

also involved development of a methodology for extrapolation of aircraft measured data below the lowest flight level based on various assumptions and the analysis of the associated uncertainties. Gordon et al. (2015) estimated uncertainty in aircraft-based estimates as 2%-30%, with the source of this uncertainty attributed mainly to extrapolation of aircraft measurements below the lowest flight track. Conley et al. (2017) and Ryoo et al. (2019) made aircraft-based emission estimations by flying cylindrical box flights around different point/area emission sources. The study by Conley et al. (2017) also involved evaluation and optimization of the sampling strategy (e.g. optimized circling radius) in order to minimize the uncertainties due to data extrapolation below the lowest flight level, through tracer controlled release and large eddy simulations (LES).

These and similar studies, employed either mass transfer — e.g., the single-screen flight approach where the background-subtracted horizontal mass flux through a downwind vertical plane is equated to the upwind source emission rates — or a mass-balance approach (box flights) to infer point and area source emission rates from aircraft measured data. Aircraft-based estimates are generally (and where possible) compared to bottom-up inventory reported emission data (e.g., Baray et al., 2018; Cheng et al., 2020; Gordon et al., 2015; Li et al., 2017; Liggio et al., 2019), and may also be evaluated using Continuous Emissions Monitoring System (CEMS) data (when these data are available). These comparisons and their conclusions are dependent in part on the magnitude of the uncertainty levels associated with the estimates. The algorithms for emission rate retrieval require steady-state meteorological and emissions data. Cases where these conditions are not met have been avoided through the application of *a priori* criteria in flight planning, and post-flight decision making. For example, pre-flight decision criteria such as relatively invariant forecast wind speed, wind direction, and planetary boundary layer height, were used to determine “fly/no fly” decisions in past studies (e.g., Li et al., 2017; Liggio et al., 2019). We refine these concepts here to create quantitative metrics for assessing retrieval success which may also be used in forecasting flight conditions or during flights. We also investigate the potential causes of emissions retrieval uncertainty, using a variety of cases, including unsuitable cases which would usually not be used for retrievals, as tests of our criteria. Therefore, it is important to further refine the methodology used to predict uncertainties in retrievals, as well as suggest revised methodologies which may reduce emissions retrieval uncertainty.

In Chapter 2 we focus on investigating the underlying assumption of time-invariant meteorological conditions (during observation time), which is shared among most (if not all) top-down retrieval methodologies, in particular the impact of localized variations in meteorology (e.g. change in atmospheric stability or wind direction) on the application of the mass-balance technique in box flight emission rate retrievals. Under time-invariant and spatially uniform conditions, a mass-balance approach based on the Gauss’s divergence theorem equates the rate of emissions from sources within a control volume (box flight) to the net mass flux through the boundaries (box walls) by assuming the net rate of accumulation of mass within the volume to be negligible. When conditions (e.g. meteorology, emissions) deviate from steady state and/or localized inhomogeneity is developed in tracer concentration and meteorological fields, the rate of mass accumulation/release within the

control volume may become significant enough to affect the accuracy of the estimates based on the mass balance approach. Herein, we use the term **“Storage-and-Release”** to refer to events where emitted mass is stored within the control volume and then subsequently released at a later time.

Numerical dispersion modelling has been proven useful in assessing the uncertainties in top-down retrievals in the past (e.g., Panitz et al., 2002) and more recently (e.g., Angevine et al., 2020; Karion et al., 2019). In Chapter 2, we use Environment and Climate Change Canada’s (ECCC) air-quality model, Global Environmental Multiscale-Modeling Air-Quality and CHemistry (GEM-MACH), which is a regional chemical transport model. Through the use of GEM-MACH model 4D fields (known meteorology and input emissions), we have eliminated all other sources of uncertainty in the input information used to retrieve emissions (e.g. measurement error, interpolation) and have extensively explored the impact of storage-and-release events on aircraft-based top-down emission rate retrievals. We investigate the potential magnitude and the causes of emissions over- and underestimates, using GEM-MACH model’s input emissions and output concentrations as a proxy for aircraft observations. This approach allows us to more precisely examine the root causes of potential emissions retrieval over/underestimates, with ancillary information provided by the model, which may not be available from past aircraft studies. Throughout Chapter 2, as an example method, we refer to (and adopt schemes from) the aircraft-based retrieval methodology developed by ECCC and described in Gordon et al. (2015), the Top-down Emission Rate Retrieval Algorithm (TERRA). The concepts explored here are not specific to TERRA and are applicable to all top-down retrieval methodologies that use the mass-balance approach.

Having conducted an in-depth investigation in the principles (theoretical and practical) of employing the mass-balance technique for top-down source emission rate retrievals in Chapter 2, in the following chapters we focus on other potential sources of uncertainty. Uncertainties in aircraft-based top-down retrievals may arise from various sources (e.g. measurements, data interpolation and extrapolation), some of which have been studied in the past (e.g., Angevine et al., 2020; Cambaliza et al., 2014; Gordon et al., 2015). In order to conduct a thorough model-based study of the atmospheric dynamical processes that can affect aircraft-based retrievals, model simulations at resolutions on the scale of measurement frequency ($\Delta x < 100$ m, $\Delta t < 1$ s) are required. Herein we refer to such resolutions as “super-resolution” (as in recent real-case LES-modelling studies, e.g. Onishi et al., 2019; Watson et al., 2020; Wu et al., 2021), as opposed to “high resolution” which commonly refers to mesoscale model simulations with horizontal grid spacing of a few kilometres (e.g. GEM-MACH at 10 km and 2.5 km, see Chapter 2). The resolution of model simulations need to be sufficient enough to resolve the dynamical and turbulent processes at the scale at which aircraft measurements are conducted. During airborne campaigns, tracer plumes are sampled at a few kilometres downwind distance from sources of emission (e.g., JOSM, 2013). At such downwind distances, plumes are spread laterally (horizontal and vertical) for 5 to 10 kilometres with high spatial variability with length scales on the order of metres. For a moderate sampling frequency of one measurement every second spaced by a few tens of metres (e.g., flight speed of about 100 m/s), model resolutions in meters and seconds are required in order to be able to mimic airborne

measurements and explore the best practices for data post-processing and emission rate estimates. The GEM-MACH model simulations used in Chapter 2, have resolutions of 2.5 km and 2 min. While these resolution are suitable for investigating the concepts of mass-balancing and divergence theorem, they lack the required resolution for resolving atmospheric processes at the scale of airborne measurements. In Chapter 3 we explore super-resolution modelling of the atmosphere and tracer transport using the Weather Research and Forecasting (WRF) modelling platform with the ARW dynamical core (Skamarock, 2008). In developing our WRF model simulations we used North American Regional Reanalysis (NARR) data as our initial and boundary conditions, employed multi-nest model simulations for dynamical down-scaling of NARR data from about 31 km resolution to our desired super-resolutions (50 m), and used Large Eddy Simulation (LES) subgrid physical and dynamical parameterization in order to simulate the complex and turbulent atmospheric processes at the scale of aircraft-based measurements. We also modified WRF dynamical solver source code to model the continuous release of passive tracers and simulated various emission scenarios under different meteorological conditions. The process of preparing our WRF model simulations and the resulting modelling products are discussed in Chapter 3.

In Chapter 4, we use the modelling products from the super-resolution WRF model simulations of atmospheric tracer transport over the region of Canadian oil sands to evaluate the conventional methods for airborne campaigns where top-down source emission rate retrievals are performed. Using atmospheric modelling simulations at resolutions that resolve dynamical process at the scale of aircraft-based measurements, we investigate the common operational parameters (e.g., flight time, path, downwind distance) under different meteorological conditions and for various source emission scenarios (e.g. surface, stack, road). We also evaluate the conventional methods of data post-processing and analysis for top-down source emission rate retrievals and discuss the range and sources of uncertainty for such estimates. In our investigations we consider top-down retrievals based on the mass-transfer approach using single-screen downwind sampling and also we consider the mass-balancing for box flights with rectangular and cylindrical shapes. We investigate the impact of sampling radius (downwind distance) and temporal variability of atmospheric/environmental fields on airborne measurements and emission rate estimates based on these measurements.

Aircraft observations at few elevations between 150 m to 2500 m above ground level (agl) leave a gap of information below and above flight tracks. It has been shown that the unsampled regions particularly below the lowest flight level can result in up to 40% uncertainty in aircraft-based estimates (Angevine et al., 2020; Conley et al., 2017; Gordon et al., 2015). This range of uncertainty differs case by case, depending on emission and meteorological conditions, and is considered as a major source of error/uncertainty for airborne retrievals. We note that aircraft-based measurements compared to other sampling methods — e.g., tower (Andreae et al., 2015; Heintzenberg et al., 2011), radiosondes (Nygård et al., 2017), land vehicle (Davis et al., 2019; de Boer et al., 2021) — provide higher (spatio-temporal) sampling frequencies and range (several kilometres in the vertical and tens of kilometres in the horizontal). While all the different sampling methods have their own limitations

in data collection (range and resolution), the gap of information in the sampled data can be filled by combining data from different measurement approaches. For instance, aircraft measured data can be complimented by ground-base measurements (Bell et al., 2021; Brus et al., 2021b; Islam et al., 2021). Noting that synchronizing data from multiple observation platforms (e.g., frequency, timestamps, range) can be challenging and create additional sources of uncertainty, it can be argued that conducting in-situ and remote measurements from the same platform (e.g., aircraft) can be beneficial. For instance, in addition to conventional in-situ measurements of pollutants and meteorological fields, aircraft-based remote (lidar) measurements of wind fields (e.g., Bilbro et al., 1984; Serafin et al., 1986; Wekker et al., 2012) and tracer amounts (e.g., Aggarwal et al., 2018; Howell et al., 2014) can be conducted. These remote observations (on the same platform) would provide ancillary data to fill the information gap in in-situ measurements towards improved top-down estimates. They also have the potential to reduce the operational costs of airborne retrievals by shortening the required flight time.

In Chapter 4, we introduce a new sampling strategy where in-situ and remote aircraft-based measurements are combined to produce top-down estimates at higher confidence levels. Our proposed sampling strategy addresses the shortcomings of conventional methods and provides a novel practical approach for direct estimation of the storage rate term to be included in mass-balance calculations. This is something that is not possible with conventional methods and, as we show in Chapter 4, lack of an estimate for the storage rate is one of the two main sources of uncertainty in top-down retrievals. Our proposed approach also addresses the other major source of uncertainty for aircraft-based retrievals, the extrapolation of data to the unsampled region below the flight level and the associated range of uncertainty in the estimates. Our modified retrieval methodology shows improvement over the conventional methods.

In this work we conduct a model-based study of airborne mass-balance methodologies for various source emission scenarios and under different meteorological conditions, utilizing numerical model simulations at different scales — from mesoscale to micro-LES scales. Using the regional air-quality model GEM-MACH at the horizontal resolution of 2.5 km, we re-evaluate the principles of applying the mass-balance technique and the divergence theorem in source emission rate estimations. Using the WRF modelling system, we employ super-resolution model simulations at 50 m horizontal grid-spacing to simulate sampling at length/time scales of aircraft measurements (virtual airborne sampling). We evaluate the impact of small-scale atmospheric dynamical process on aircraft-based retrievals and provide recommendations for improved sampling and data analysis methods.

Chapter 2

Evaluating the impact of storage-and-release on aircraft-based mass-balance methodology using a regional air-quality model (GEM-MACH)

2.1 Introduction

In this chapter we investigate the potential for aircraft-based top-down emission rate retrieval over- and under-estimation using Environment and Climate Change Canada’s (ECCC) air-quality Model, Global Environmental Multiscale – Modeling Air-Quality and CHemistry (GEM-MACH). The results and discussions in this chapter have been previously published in the journal of Atmospheric Chemistry and Physics (ACP) under the same title, (Fathi et al., 2021). For this work, the PhD candidate Sepehr Fathi ran the GEM-MACH model simulations, implemented the mass-balance retrieval algorithm using GEM-MACH modelling products, conducted the full analysis, and wrote the paper under the guidance and supervision of professor Mark Gordon of York University and Dr. Paul A. Makar of Air Quality Research Division (AQRD) at ECCC. Dr. Ayodeji Akingunola (ECCC/AQRD), provided GEM-MACH model simulation support throughout the research project. The other authors provided advice during the analysis and contributed to paper revisions.

Aircraft-based source emissions characterization involves airborne measurements by flying downwind and/or around point and area sources. Observational data from single-leg downwind screen (or single-height) measurements have been previously analyzed to estimate emission rates based on mass flux calculations (Cambaliza et al., 2014; Peischl et al., 2010). To account for upwind and background emissions and utilize a complete mass-budget analysis, a control volume can be

defined by conducting airborne measurements following a closed shape (rectangular, cylindrical) at several altitudes. This control volume (box) is then considered for application of the mass-balance technique to estimate within-box source emission rates (Kalthoff et al., 2002; Ryoo et al., 2019). The mass-balance equation, equates the accumulation rate of pollutants (e.g., aerosols and tracers such as SO_2 , CO , CH_4) within the control volume to source emission rates, the difference between mass inflow and outflow rates, surface deposition rate, and the chemical production/consumption rate of the compounds (Gordon et al., 2015). Under the assumption of steady state conditions (e.g., constant emissions, constant height of the mixing layer, and constant wind fields) the within box accumulation rate can be assumed negligible, and by neglecting the contribution of chemical production/consumption, the within-box source emission rate can be equated to the integrated mass-flux through the surfaces of the box (Alferi et al., 2010). The study by Panitz et al. (2002) using air-quality modelling, concluded that between 85% to 95% of tracer emissions (e.g., CO , NO_x) can be obtained from the advective fluxes alone (with no estimate of the accumulation/storage rate). Previous model-based and/or observational (aircraft-based) studies estimated the uncertainty in airborne top-down retrievals between 20% to 40% (e.g., Angevine et al., 2020; Cambaliza et al., 2014; Gordon et al., 2015; Karion et al., 2019; Turnbull et al., 2011). These studies attributed the large uncertainties mainly to error in estimating the mixing layer height (e.g., Angevine et al., 2020; Karion et al., 2019) and the gap of information below the lowest flight level (Conley et al., 2017; Gordon et al., 2015). Our focus in this chapter is on investigating the uncertainty associated with the underlying assumption of steady state conditions and circumstances that lead to significant accumulation rates (storage-and-release events), which has not been thoroughly explored in previous studies.

Using air-quality model simulations with the GEM-MACH model, we focus on the uncertainty in mass-balance analysis arising from the temporal evolution of the sampled fields by considering an extensive range of realistic meteorological and emission conditions. Air-quality model simulations have been previously employed to characterize the error in airborne retrievals (e.g., Angevine et al., 2020; Conley et al., 2017; Karion et al., 2019). These studies were mainly concerned with tracer/pollutant transport and mixing conditions downwind the emission source and their impact on top-down estimations, rather than the tracer mass-budget. The use of air-quality modelling for mass-balance analysis applications requires careful considerations. Air-quality models commonly use post-advection mass conservation schemes to correct for the creation of spurious negative concentrations (de Grandpré et al., 2016). Depending on the specific applications of the study, one might choose a global or local mass conservation scheme. Fathi (2017) demonstrated that local mass conservation is essential for mass-balance analysis using model-generated fields, and concluded that the use of local schemes are recommended over global schemes. We use these results in our model-based study using the GEM-MACH model. In our investigations we consider in particular the application of the mass-balance approach in the Top-down Emission Rate Retrieval Algorithm (TERRA) described in Gordon et al. (2015). Aircraft-based mass-balance retrieval methodologies such as TERRA require specific conditions in order to reliably estimate emission rates from aircraft

observations – essentially, that meteorological conditions and source emission rates are relatively constant during the time of observations. Avoiding cases where the time-invariant requirement may not be met is one means of reducing emissions retrieval uncertainty, and quantitative metrics that may be used for retrieval accuracy estimation are therefore desirable - using these metrics has the potential to greatly improve emission rate retrieval accuracy. Here, we investigate the impact of meteorological variability on the accuracy of mass-balance emission rate retrievals, by using model simulated fields as a proxy for real world chemical and meteorological fields, in which virtual aircraft sampling of the GEM-MACH output was used for top-down mass balance estimates. We also explore the impact of upwind emissions from nearby sources on the accuracy of the retrieved emission rates. This approach allows the state of the atmosphere used for top-down estimates to be characterized in time and 3D space; the input meteorology and chemistry is “known”, and thus potential means for improving emission rate retrievals and determining the factors affecting retrieval accuracy may be investigated.

2.2 Methods

During an aircraft campaign conducted in August and September of 2013 as part of the Joint Canada Alberta implementation plan for Oil Sands Monitoring (JOSM), the National Research Council of Canada Flight Research Laboratory (NRC-FRL) Convair 580 research aircraft was used to make air-quality measurements over the Athabasca oil sands region (Alberta, Canada). Air-quality forecasts created using ECCC’s air-quality model GEM-MACH, provided advice on flight planning for the JOSM 2013 campaign (Makar et al., 2016). For this purpose, and for the post-study simulations conducted here, the GEM-MACH system was configured to create high-resolution nested air-quality forecasts with a fine-resolution domain (2.5 km) over the Canadian provinces of Alberta and Saskatchewan, including the Athabasca oil sands region (Fig. 2.1a). From 10 August to 10 September 2013 a total of 22 flights were conducted around and downwind of oil sands facilities, 13 of which were emission estimation (box) flights where the aircraft was flown around individual facilities at multiple altitudes, approximating a rectangular/polygonal shaped box (control volume), while making meteorological and air-quality measurements (JOSM, 2013). We chose our model input test cases from the same times and locations as JOSM 2013 box flights, to sample conditions (meteorology and emissions) similar to the range existing within the real atmosphere which might be considered for retrievals. Nine box flight cases over three oil sands facilities with significant SO₂ emissions (Syncrude, CNRL and Suncor; see Fig. 2.1b) were chosen because (1) SO₂ is emitted from stack sources but has low background levels (Gordon et al., 2015), reducing uncertainties in emission rate retrievals based on the mass-balance approach and (2) hourly SO₂ emissions and stack parameter data are available from the CEMS reports (Zhang et al., 2018) thereby ensuring the GEM-MACH inputs are realistic: CEMS-derived input emissions in GEM-MACH resulted in improved model plume rise and transport representation (model-simulated plume heights were evaluated in Akingunola et al., 2018). Output data from GEM-MACH model simulations for the

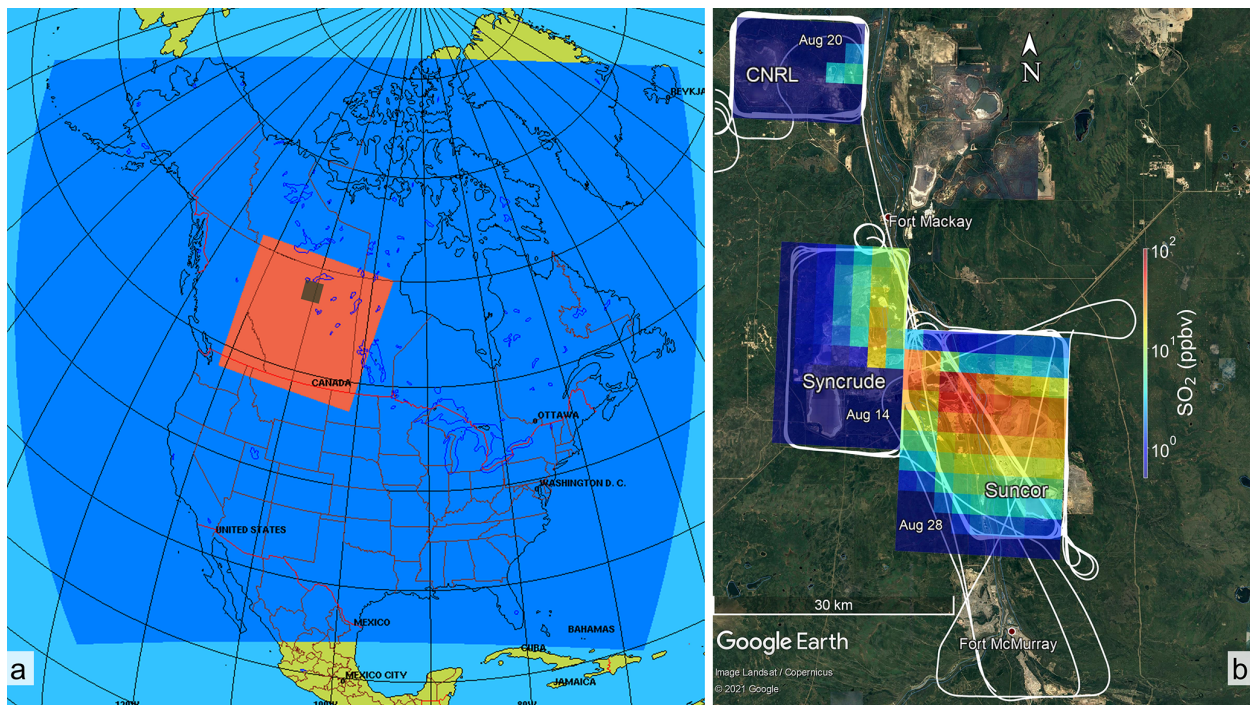


Figure 2.1: (a) GEM-MACH model horizontal grid structure with nested domains: the dark blue region shows the parent domain over North America at 10 km horizontal grid resolution. Output from GEM-MACH 10 km simulations provided initial and boundary conditions for the nested high-resolution GEM-MACH 2.5 km domain (red). The high-resolution domain was configured over the Canadian provinces of Alberta and Saskatchewan including the Alberta Athabasca oil sands region, shown as a dark shaded area. (b) © Google Earth map of the region with three oil sands facilities labelled. JOSM 2013 flight tracks on 14, 20 and 28 August are drawn with white curves. Model 4D fields were extracted for nine JOSM 2013 box flight cases over these three facilities, examples are shown as overlay SO_2 concentration maps.

period of the nine cases were analyzed and mass-balance estimates of model SO_2 input emissions were made. Model configuration and set-up and simulation scenarios for this work are described in Sect. 2.2.1. Section 2.2.2 describes the methodologies that were developed here for direct and virtual aircraft-based retrievals of model input emissions using the GEM-MACH model output fields as well as suggested improvements for flight planning.

2.2.1 Model description and simulations

ECCC’s air-quality model, GEM-MACH, is a fully coupled online air-quality and chemical transport model. The GEM-MACH modelling system resides within the Global Environmental Multi-scale (GEM) numerical weather prediction model (Côté et al., 1998a,b; Girard et al., 2014) and includes an atmospheric chemistry module (Moran et al., 2010) with gas and particle process representation. The details of the GEM-MACH configuration used here appear in Table A.1. We note that this work is not intended as an evaluation of the GEM meteorological model and its components, which have been extensively evaluated elsewhere in the literature (e.g., Bélair et al.,

2003a,b; Côté et al., 1998b; Fillion et al., 2010; Girard et al., 2014; Li and Barker, 2005; Milbrandt and Morrison, 2016; Milbrandt and Yau, 2005a,b). A recent detailed evaluation of GEM-MACH’s meteorological performance is also provided in Makar et al. (2021). Rather, we use the model as a proxy for observations since it can provide instantaneous 3D values for chemical concentrations and meteorological variables, which in turn allow us to construct parameters that can be used to predict storage-and-release events in observation-based applications.

For this work, a series of retrospective high-resolution nested air-quality simulations were made for the period of the JOSM summer 2013 aircraft intensive campaign (JOSM, 2013). The GEM-MACH model grid configuration consisted of a coarse parent domain with a horizontal resolution of 10 km covering North America and a nested 2.5 km high-resolution domain including the Canadian provinces of Alberta and Saskatchewan (see Fig. 2.1a). To prevent the model high-resolution meteorology from drifting chaotically from observed meteorology, the GEM-MACH 10 km simulations were updated every 24 h at 06:00 UT (local mid-night). Most of the JOSM 2013 cases examined here take place during the late mornings and early afternoons in local time (LT); by choosing to update the driving meteorology in the model at 06:00 UT, sudden shifts in the model meteorological fields were avoided during the local daytime, and more than sufficient spin-up time (6 h time frame) was provided for model meteorological fields (in particular the cloud microphysics fields) to reach equilibrium before local noon.

Flux box data, including meteorological fields and SO_2 concentrations, were extracted from the model output database for the period of nine JOSM 2013 flights over three different facilities (Fig. 2.1b). The mass-balance approach was employed to retrieve model input emissions using the extracted flux box data. Earlier work (Fathi, 2017) suggests that there are three main considerations which should be taken into account when using regional air-quality models as proxies for observations.

1. Air-quality models usually require the use of some form of “mass conservation” correction for their advection algorithms (de Grandpré et al., 2016), which can otherwise create spurious negative concentrations. However, mass conservation algorithms which conserve mass over the entire model domain, but do not conserve mass locally, effectively include spurious mass transport – in turn affecting emissions retrievals using model output. A local mass correction algorithm (three-shell ILMC approach, Sørensen et al., 2013) was therefore used in the GEM-MACH simulations carried out here. Model performance with the ILMC scheme in terms of local and global mass-conservation was investigated and insured through mass-balance analysis and comparison to other available methods in GEM-MACH (e.g., Bermejo and Conde, 2002). Model resolution also impacts mass conservation and needs to be sufficiently high that the sources may be resolved.
2. Regional air-quality models provide instantaneous output at every grid cell on a fixed (2 min) time step – however, if these are interpolated to a finer time resolution (e.g., along an aircraft flight track), errors in interpolation may affect the use of the model output in retrieval algorithms. Values of model variables at 3D grid points, in the vertical columns making up the

emissions “box”, were therefore used in the analysis which follows to avoid these (temporal and spatial) interpolation errors in the use of model output as a proxy for observations.

3. In aircraft-based sampling (both in real-world aircraft measurements and in extraction of emissions retrieval algorithm inputs from 4D air-quality model-predicted output), temporal lag between data points collected along flight tracks may represent an additional source of error. This may introduce additional errors if the “static atmosphere” assumption is not met. Here, instantaneous model output at each time step (2 min) was used for mass-balance estimates, to avoid the potential for interpolation from the model resolution (2.5 km) and time step to influence the retrievals generated here.

For this purpose, GEM-MACH model four-dimensional data (3D space, 1D time) were extracted at every 2 min model chemistry time step for the flux boxes over individual oil sands facilities for our nine case studies. Aircraft-based mass-balance retrievals were simulated by employing the divergence theorem in analyzing the model extracted data. The estimated emission rate time series and flight-time (1.5–2.5 h) averages were then compared to model input emissions (MIEs) to evaluate top-down mass balance retrievals.

2.2.2 Divergence theorem, the mass-balance technique and emission rate retrievals

During an actual aircraft campaign, measurements are made along a flight trajectory on the lateral walls of a flux box (control volume); measured data on a single 2D screen (closed surface) comprised of box lateral walls are all that is available for emission rate retrievals based on mass flux and mass balance calculations. Here we use instantaneous model values along lateral box walls as our 2D screen (“TR” approach, Sect. 2.2.2).

The temporal evolution of the system can be studied by increasing the number of flights around the same facility. Here, this approach is explored with GEM-MACH simulated fields by analyzing multiple 2D screens at consecutive model time steps (3D fields) and generating mass balance estimates (“TR*” approach, Sect. 2.2.2).

A truly comprehensive analysis of the tracer mass budget within the flux box is only possible with 3D-volumetric data over the entire extent of the box and over time (4D fields). However, collection of 4D (volumetric time series) data is not practically feasible during aircraft measurements as the aircraft cannot sample the entire volume of the box even with increased flights during a time window of only a few (e.g., 2–3) hours. One advantage of the use of a model such as GEM-MACH as a proxy for observations is that 4D forecast fields can be generated, to provide complementary information for mass balance retrieval investigation. Here we use GEM-MACH model 4D chemical and meteorological fields to conduct a comprehensive analysis of the transport of the emitted mass from sources within the control volume and to further investigate uncertainties associated with storage-and-release events in top-down mass balance retrievals (“DR” approach, Sect. 2.2.2).

Gauss’s divergence theorem states that the total integrated divergence of a vector field (e.g., fluid

flow) F in a region of space (control volume) V equals the total flux of F through the enclosing boundary (surface) S ,

$$\iiint_V \nabla \cdot F dV = \iint_S F \cdot dS. \quad (2.1)$$

A mass-balance approach (conservation of mass) based on the divergence theorem can be employed to estimate chemical (pollutant) emission rates by sources (e.g., stack emissions) within a control volume V where atmospheric dispersion and transport takes place. In order for mass to be conserved, the inflow of mass of species C into V ($E_{C,\text{in}}$) plus the rate of generation/emission of mass (E_C) within V must equal the rate of mass removal from the control volume ($E_{C,\text{out}}$) by the outgoing flux and processes such as surface deposition and the rate at which mass is accumulated/stored within the volume of the box ($E_{C,S}$), written in a format linked to the divergence theorem (Eq. 2.1),

$$E_C = (E_{C,\text{out}} - E_{C,\text{in}}) + E_{C,S}. \quad (2.2)$$

Note that the left-hand side and the terms in brackets on the right-hand side of Eq. (2.2), comprise the implementation of the divergence theorem in the mass-balance equation. In steady-state conditions, i.e., under the assumption of time-invariant meteorology and input emissions, the net accumulation (rate) of mass can be assumed zero and therefore the divergence theorem can be utilized to equate E_C to the net flux through the boundaries of the box ($E_{C,\text{out}} - E_{C,\text{in}}$). The divergence theorem is commonly utilized in aircraft-based studies with flight paths approximating flux boxes of polygonal (e.g., Gordon et al., 2015; Liggio et al., 2019) and/or cylindrical (e.g., Ryoo et al., 2019) shapes, where aircraft observations along flight tracks are projected (and interpolated) onto a 2D screen comprised of box lateral walls surrounding the emitting facility. A mass-balance technique is then employed to estimate point/area source (e.g., stack, tailing ponds) emission rates by calculating the net mass flux through the screen (e.g., TERRA).

We start with a comprehensive analysis of GEM-MACH 4D data for direct retrieval of model input emissions, where we explore all the processes (including storage-and-release) contributing to change in tracer mass within the flux box and over time (Sect. 2.2.2). We then simulate aircraft-based retrievals by limiting our analysis to 2D data on box lateral walls (Sect. 2.2.2). Finally, we combine the information from time-consecutive screens (3D) to explore the temporal trends in chemical and meteorological fields and introduce potential means for improving top-down estimates (Sect. 2.2.2).

Direct retrieval (DR) using GEM-MACH 4D fields

Here we apply the divergence theorem to the GEM-MACH model four-dimensional (volumetric time series) output data to estimate model input emissions. Data from volumes corresponding to the nine JOSM 2013 cases were extracted from the GEM-MACH model output at every 2 min model time step (no interpolation), creating nine sets of 4D metadata with variables (including wind speed and direction, air density, temperature, species mixing ratio, ground surface deposition). Model-extracted 4D flux box metadata, spanning model vertical levels from the surface up to

~ 2000 m above ground (average JOSM 2013 box flight top height), were analyzed using the mass balance approach to estimate the model input emissions. We note that this approach incorporates information not available for aircraft observation-based retrievals (i.e., volumetric time-series data) but may be used to analyze the importance of storage release and as a standard against which the other methods which follow may be evaluated and compared.

The total mass of compound C within the control volume V ($B_C(t, V)$) is calculated at each model time step (Δt) by integrating over the entire volume of the box. The rate of change in mass can then be calculated by taking the time derivative of $B_C(t, V)$ and following the chain rule,

$$\begin{aligned}\frac{\partial}{\partial t} B_C(t, V) &= \frac{\partial}{\partial t} \left[M_R \iiint \chi_C(t, V) \rho_{\text{air}}(t, V) dV \right] \\ &= M_R \iiint \left[\frac{\partial \chi_C}{\partial t} \rho_{\text{air}} + \chi_C \frac{\partial \rho_{\text{air}}}{\partial t} \right] dV,\end{aligned}\tag{2.3}$$

where M_R is the ratio of the compound molar mass (e.g., 64.07 g mol^{-1} for SO_2) to the molar mass of air (28.97 g mol^{-1}), $\chi_C(t, V)$ and $\rho_{\text{air}}(t, V)$ are compound mixing ratio and air density at each model grid point and time step, respectively, with the 3D integrand volume element $dV = dx dy dz$. The first integral on the right-hand side of Eq. (2.3) represents the rate of change in compound mass within V characterized as variations in compound mixing ratio $\chi_C(t, V)$ independent of the changing air density,

$$E_{C,S}(t) = M_R \iiint \frac{\partial \chi_C}{\partial t} \rho_{\text{air}} dV.\tag{2.4}$$

The second integral represents the apparent change in compound mass (which is calculated as a product of air density and compound mixing ratio) due to temporal variations in air density $\rho_{\text{air}}(t, V)$, caused by meteorological processes, and does not represent an actual change in compound mass within the control volume,

$$E_{C,M}(t) = -M_R \iiint \chi_C \frac{\partial \rho_{\text{air}}}{\partial t} dV.\tag{2.5}$$

The minus sign signifies that this is an apparent change in compound mass, which is accounted for in the mass balance equation. In applying the mass balance technique, we set the net rate of change of Eq. (2.3) as equal to the sum of all flux and chemical terms changing the compound mass (emissions, horizontal and vertical advection, vertical diffusion and chemical production). Solving for $E_{C,S}$, the processes contributing to the net change in compound mass within the control volume V can be described with the following expression:

$$E_{C,S} = E_C - E_{C,H} - E_{C,V} - E_{C,VD} + E_{C,M} + E_{C,X},\tag{2.6}$$

where E_C represents the emission rate of compound C from sources within the control volume. $E_{C,H}$ and $E_{C,V}$ represent the net horizontal and vertical fluxes through the lateral and top walls of the flux box, respectively (derivations appear in Appendix A, Eqs. A.1–A.3). $E_{C,VD}$ represents

the surface deposition rate, which is available as an output variable from GEM-MACH. Change in compound mixing ratio due to variations in air density can be accounted for by calculating $E_{C,M}$ from Eq. (2.5). $E_{C,X}$ represents the rate of change (i.e., the formation rate of C) due to chemical processes, and can be assumed negligible for a tracer compound such as SO_2 relative to the emission rates for these facilities since it has a lifetime of the order of days. In the context of emissions estimation using observed concentrations and winds along an aircraft trajectory, Eq. (2.4) cannot be estimated directly, since it requires a volume integral over the box enclosing the emission source, data that are not available along the flight path. However, when a regional air-quality model is used as a proxy for observations, and/or a strategy of repeat flights are used (see Sect. 2.2.2), $E_{C,M}$ and $E_{C,S}$ can be estimated directly. When model values are used these terms may be generated from Eqs. (2.4) and (2.5). For example $E_{C,S}$ is estimated using the mixing ratio rate of change ($\partial\chi_C/\partial t$) multiplied by the air density (ρ_{air}) at each model grid point and integrating over the volume of the emissions enclosure box, with a similar approach used for $E_{C,M}$. $E_{C,S}$ represents the net rate of increase/decrease in tracer mass within the box, a residual rate of change associated with the difference given in Eq. (2.6). We therefore refer to it as a “storage” term for tracer mass. Observation-based emission rate estimate studies (e.g., TERRA) assume $E_{C,S} = 0$, while in reality a non-zero $E_{C,S}$ term is incorporated in the E_C term; here we examine the potential magnitude of $E_{C,S}$ using the GEM-MACH model’s concentration predictions as a proxy for aircraft observations, and use the combination of mass-balance approach and GEM-MACH to fully investigate the potential impacts of storage of tracer mass on emission rates retrieval over-/under-estimates. We also use this combination of retrieval algorithm and regional air-quality model to generate meteorological forecast model variables which predict conditions in which the storage term may be of concern, reducing its influence on future emissions analysis (Sect. 2.4), and to demonstrate flight procedures which may be adopted a priori to reduce its influence on emissions retrievals (Sect. 2.2.2). Rearranging Eq. (2.6) and neglecting $E_{C,X}$ as being insignificant over the short distances (5–20 km) between source and emissions box wall, compound emission rates can be estimated based on the other terms as

$$E_C = E_{C,H} + E_{C,V} + E_{C,VD} - E_{C,M} + E_{C,S}. \quad (2.7)$$

Equation (2.7) equates the total integrated emission rate, originating from sources (point and/or area) within the box (control volume), to the flux leaving the box through the top and lateral walls $E_{C,V}$ and $E_{C,H}$, the surface deposition rate $E_{C,VD}$, and the volumetric storage rate $E_{C,S}$, while accounting for the effects of changing air density ($E_{C,M}$). Note that in our GEM-MACH driven analysis, $E_{C,H}$, $E_{C,V}$ and $E_{C,VD}$ are derived from GEM-MACH 2D fields (screens, box top, ground surface deposition) at each time step, while $E_{C,M}$ and $E_{C,S}$ are temporal gradient terms estimated from model 4D fields (volumetric time series) and employing a second-order central finite-difference (FD) scheme (i.e., $\partial\chi_C/\partial t \cong (\chi_C(t+\Delta t) - \chi_C(t-\Delta t))/2\Delta t$); the choice of FD scheme was made based on optimized performance and accuracy in our computations, and it was determined that higher-order schemes did not provide further advantages in terms of accuracy. A similar set of equations, lacking sources, sinks and a separate storage term, can be written for the air mass

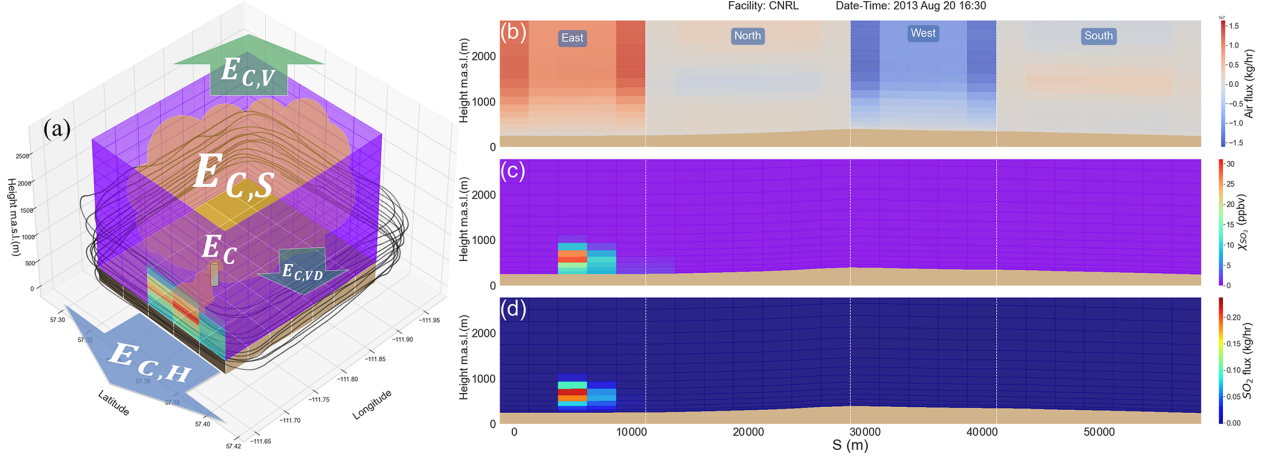


Figure 2.2: SO_2 mixing ratio screen around CNRL extracted from the GEM-MACH model output at the start of case 4 on 20 August 2013 at 16:30 UT (10:30 LT). (a) Mixing ratio data (2D) on the lateral walls of the flux box, along with the flight tracks plotted as dark dots. The additional graphics (arrows and plumes) on (a) represent processes contributing to change in tracer mass within the box volume: source emission rate (E_C), horizontal flux ($E_{C,H}$), vertical flux ($E_{C,V}$), surface deposition ($E_{C,VD}$) and storage rate ($E_{C,S}$). The extracted fields around the box are unwrapped onto a 2D screen comprised of box lateral walls (east, north, west and south) and shown in (b) the normal air flux (positive outwards), (c) SO_2 mixing ratio and (d) SO_2 mass flux through the screen. Note that the net flux on box corners is the vector sum of the fluxes in west–east (U) and south–north (V) directions.

changes within the emissions box (see Appendix A), which lead to the mass balance expression for air

$$E_{\text{air},M} = -E_{\text{air},H} - E_{\text{air},V}. \quad (2.8)$$

Equations (2.4), (2.5) and (2.7) (with other terms in Appendix A) were utilized in analyzing GEM-MACH model 4D flux box data for the nine cases, and DR method estimates were compared to the MIEs. Results are discussed in Sect. 2.3.1.

TERRA retrieval (TR) approximation using GEM-MACH 2D fields

Aircraft-based retrieval methods such as TERRA rely on measured data along the lateral walls of the box. In observation-based retrievals, the horizontal flux $E_{C,H}$ can be calculated (directly) from aircraft collected data and the other terms need to be estimated based on several assumptions about the atmospheric conditions and source properties (Gordon et al., 2015). Here, we explore the use of regional (GEM-MACH) model output in a similar fashion as aircraft observations, restricting our analysis to model output sampled along the lateral walls of a box enclosing an emitting facility within the model domain (GEM-MACH 2D fields). Figure 2.2a shows model simulated SO_2 concentrations (normalized to the maximum concentration) on the lateral walls of the flux box surrounding the CNRL facility (case 4). Flight tracks on 20 August 2013 (JOSM) are also plotted as dark dots around the facility, for reference. For emissions retrieval applications using aircraft

observations, the sampled observations must be interpolated (e.g., kriging) between successive circuits around the box and extrapolated to the ground surface in order for emissions to be estimated. Surface deposition rate $E_{C,VD}$ is also estimated using the extrapolated species mixing ratio data at the ground level. Gordon et al. (2015) estimated the uncertainty due to extrapolation of aircraft data to the ground surface as 0.4 and 12 % in TERRA estimates for SO_2 plumes. Here we avoid the uncertainties arising from such assumptions by extracting model data along box lateral walls from the ground surface up to the box top, and by using GEM-MACH model surface deposition ($E_{C,VD}$) values in our calculations.

For this portion of our analysis, in which we limit the inputs to mass-balance estimations to the information available on the box lateral walls in analogy to aircraft observations, instantaneous 2D screens of relevant fields (e.g., SO_2 concentrations) covering the entire area comprised of box lateral walls (as shown in Fig. 2.2) were extracted/generated from the GEM-MACH model output at every model time step. Such screens were also generated from GEM-MACH output for air flux, based on the extracted air density ρ_{air} , the calculated normal wind $U_{\perp}$ (positive outwards), and SO_2 mass flux, based on concentration and air flux screens (Fig. 2.2b–d). The integrated horizontal flux through the screen $E_{C,H}$ was then calculated using Eq. (A.1). In observation-based applications, TERRA estimates the changes in mass due to variations in air density over flight time, $E_{air,M}$ and $E_{C,M}$, based on observed meteorological trends in air density ($\frac{\partial}{\partial t}\rho_{air}$) from nearby weather stations (towers) and average species mixing ratios (χ_C) around the box. Here, we substituted the weather data with a single column (GEM-MACH-predicted) air density temporal rate of change $\frac{\partial}{\partial t}\rho_{air}(t, z)$ at point (x_o, y_o) located at the horizontal center of the box for each of the nine studied cases. Here, single column $\frac{\partial}{\partial t}\rho_{air}(t, x_o, y_o, z)$ is assumed to represent the spatial average rate of change in air density over the region of interest through the assumption of horizontal homogeneity, as in observational TERRA applications. Our comparisons between the GEM-MACH box average and the single column rate of change in air density indicate < 5 % difference between the two trends for the time periods of our studied cases. This gives

$$E_{air,M}^{tr}(t) = A \int \frac{\partial}{\partial t}\rho_{air}(t, x_o, y_o, z) dz, \quad (2.9)$$

$$E_{C,M}^{tr}(t) = -AM_R \int \overline{\chi_C}(t, z) \frac{\partial}{\partial t}\rho_{air}(t, x_o, y_o, z) dz, \quad (2.10)$$

where A is the base area of the box. $\overline{\chi_C}(t, z)$ is the average GEM-MACH-predicted species mixing ratio around the box along the flight path (s) at height z , which is the same as in aircraft-based applications of TERRA.

Furthermore, due to lack of measurements at the box top, the vertical flux $E_{C,V}$ in TERRA in observation-based applications is estimated in two steps: (a) the vertical air mass flux ($E_{C,V}$) is estimated based on the other terms by rearranging the mass balance expression for air (Eq. 2.8), and (b) the species mixing ratio (e.g., SO_2) at the box top is assumed to be uniform and equal to the average (along flight path s) screen top mixing ratio ($\overline{\chi_{C,top}}$) (Gordon et al., 2015). Here, the same procedure was employed in deriving the vertical flux through box top using GEM-MACH

output fields,

$$E_{C,V}^{\text{tr}}(t) = AM_R \overline{\chi_{C,\text{top}}}(t) E_{\text{air},V}(t). \quad (2.11)$$

As there are no volumetric data available in aircraft measurement applications of TERRA, the storage rate within the box cannot be inferred directly. Emission rate estimations in aircraft-based methods such as TERRA are made with the underlying assumption of steady-state conditions and therefore $E_{C,S}$ is assumed to be negligible, effectively incorporating it into E_C^{tr} (the superscript “tr” refers to TR method estimates). Thus, for aircraft-observation TERRA applications, mimicked here, Eq. (2.7) is reduced to

$$E_C^{\text{tr}} = E_{C,H} + E_{C,V}^{\text{tr}} + E_{C,VD} - E_{C,M}^{\text{tr}}. \quad (2.12)$$

Estimations of emission rates were made for each model time step (using instantaneous 2D screens) during the time periods of the nine studied JOSM cases by using GEM-MACH output to approximate the input data for TRs as described above. Note that the TR method, as described in this section, makes use of GEM-MACH 2D data at each time t to generate an estimate (one estimate at each time step t during the entire sampling period) while making approximate estimations of the temporal gradient terms (Eqs. 2.9 and 2.10) at time t by following procedures similar to aircraft-based applications of TERRA. Flight-time-averaged estimates are then compared to the flight-time-averaged MIEs (Sect. 2.3.2).

Revised TERRA retrieval (TR*) using GEM-MACH 3D fields

Utilizing time-consecutive instantaneous screens around an emission source, the temporal evolution of the system and its impact on mass-balance estimates can be studied in more detail, as discussed at the beginning of this section (Sect. 2.2.2). By using air density screens at every 2 min model time step (2D space, 1D time), as opposed to using a single column (vertical) profile in the TR method, estimations of $E_{\text{air},M}$ and $E_{C,M}$ can be improved (by 1%–3%, determined through comparisons with GEM-MACH box averaged temporal trends):

$$E_{\text{air},M}^{\text{tr}*}(t) = A \int \frac{\partial \overline{\rho_{\text{air}}}(t, z)}{\partial t} dz, \quad (2.13)$$

$$E_{C,M}^{\text{tr}*}(t) = -AM_R \int \overline{\chi_C}(t, z) \frac{\partial \overline{\rho_{\text{air}}}(t, z)}{\partial t} dz, \quad (2.14)$$

where $\overline{\rho_{\text{air}}}(t, z)$ and $\overline{\chi_C}(t, z)$ are the average GEM-MACH-predicted air density and species mixing ratio around the box (along s) at height z . A similar procedure (using time-consecutive screens) can be used for estimating the storage term $E_{C,S}$. The main difference between DR (Eq. 2.7) and TR (Eq. 2.12) retrievals is that the former explicitly accounts for and calculates the storage term $E_{C,S}$. As discussed later in Sect. 2.3.3, the results from the direct retrieval (DR) method demonstrate the potentially significant contribution of the storage term $E_{C,S}$ to the total integrated estimated emission rates E_C (Eq. 2.7). Ignoring this term, as in Eq. (2.12), could result in substantial

over-/under-estimates, if retrievals are attempted for conditions where the underlying assumption of time-invariant meteorology is not met and/or when significant upwind emissions (from other sources) enter the flux box. As noted in the Introduction, a priori metrics may be used to avoid these conditions. The key information for calculating $E_{C,S}$, is the estimate of the temporal gradient in species mixing ratio ($\partial\chi_C/\partial t$) as in Eq. (2.4). This is achieved in the DR method by utilizing GEM-MACH model 4D data (volumetric time series) and integration over the entire volume of the box at each model time step. These 4D data are not available for aircraft applications of retrieval algorithms, since instantaneous datasets cannot be created from flight information, and collection of volumetric data may not be operationally feasible in aircraft campaigns, where flight durations may be limited to a few hours. However, data from time-consecutive 2D screens may be used to approximate $E_{C,S}$, leading to improved estimates of emission rates – by providing ancillary information for the study of temporal trends in chemical and meteorological fields. As we show later in Sect. 2.3.3, estimates of the storage term from time-consecutive 2D screens ($E_{C,S}^{\text{tr}*}$) are strongly correlated ($P_r = 0.9$) with the corresponding DR estimates of $E_{C,S}$ using volumetric (3D) time series. This high correlation suggests that $E_{C,S}$ can be approximated using $\partial\chi_C/\partial t$ generated from time-consecutive 2D screens (3D fields) as representative of the rate of change in mass within the volume of the box. The approximation of $E_{C,S}$ using screen data alone is given by

$$E_{C,S}^{\text{tr}*}(t) = AM_R \int \overline{\rho_{\text{air}}}(t, z) \frac{\partial \overline{\chi_C}(t, z)}{\partial t} dz, \quad (2.15)$$

where, same as in Eq. (2.14), $\overline{\rho_{\text{air}}}(t, z)$ and $\overline{\chi_C}(t, z)$ are the average GEM-MACH-predicted air density and species mixing ratio along s (around the box) at height z . Note that the key difference between Eqs. (2.4) and (2.15) is that the former integrates over the volume, using consecutive time steps of chemical transport model output, while the latter integrates over average values along s at each height level z on the screen walls using consecutive time steps, and multiplies the result by the box horizontal area. However, Eq. (2.15) may also be generated from observations, something which is not possible for Eq. (2.4). For the case of aircraft observations, a repeat flight would be required: rather than just a single flight path starting near the surface and working upwards along the box boundary, the aircraft would execute (at least) two consecutive box flights around the same facility. A better approach would be the use of two (multiple) aircrafts, or UAVs as they are becoming more popular for air-quality measurements (e.g., Barchyn et al., 2018; Elston et al., 2015; Li et al., 2018; Mölders et al., 2015), flying in tandem following the same flight path in order to reduce time lag between samplings (in turn reducing the chance for meteorological variability to be a concern). Differencing between the fields obtained from consecutive flights would be used to estimate the value of $E_{C,S}^{\text{tr}*}$. As we discuss below, the use of either ground-based or airborne remote sensing is another potential means to estimate $E_{C,S}^{\text{tr}*}$. The correlation between $E_{C,S}$ and $E_{C,S}^{\text{tr}*}$ also indicates that the latter may be used to estimate, from repeat flight aircraft observations and/or remote sensing, the extent to which within-box mass storage may influence emission rate retrievals.

Considering the potentially prohibitive operational cost of conducting repeat flights and the fact

that few (e.g., two to three) time-consecutive measurements may not provide enough information for the estimate of temporal trends in within-box tracer mass budget, the following alternative is proposed. As we discuss later in Sect. 2.3.3, observed temporal trends of tracer concentrations ($\partial\chi_C/\partial t$) downwind of the emission source are also in similar strong correlation ($P_r = 0.9$) with the corresponding DR estimates (using volumetric time series). Furthermore, our studied cases suggest that downwind vertical profiling of tracer concentrations can provide similar estimates of the temporal trends required for the estimate of the storage rate $E_{C,S}^{\text{tr}*}$ and can be substituted in Eq. (2.15) for $\partial\overline{\chi_C}(t,z)/\partial t$. In observational applications, this can be achieved through ground-based or, preferably, aircraft-based vertical profiling remote measurements (e.g., lidar measurements, Aggarwal et al., 2018), where the sampling aircraft with a remote measurement setup would collect column data while flying around or downwind of the emission source. The use of an aircraft-based vertical profiling instrument for a small number of chemical fields, along with a repeated, single-loop flight path around a facility (10–15 min) would generate fields similar to those generated by the air-quality model used here, in turn allowing highly time-resolved estimates of the storage term. This alternative approach can be more cost (operational) efficient compared to the strategy of repeat flights (which would require a second aircraft traveling in tandem behind the first to achieve the same time resolution) while providing more time-consecutive data points for the study of the temporal trends in the tracer mass budget and further reducing the relatively small emissions biases associated with extrapolation to the ground of observed fields.

We note that the correlation between $E_{C,S}$ and $E_{C,S}^{\text{tr}*}$ was deduced here through GEM-MACH model simulations with model grid point resolution (in the horizontal) of 2.5 km. This correlation (and the corresponding time/length scales) can be further investigated through model simulations at higher resolutions (e.g., 50–250 m) and/or field observations (e.g., controlled tracer release). Equations (2.14) and (2.15) along with the other terms in Eq. (2.12), give the TR* method,

$$E_C^{\text{tr}*} = E_{C,H} + E_{C,V}^{\text{tr}} + E_{C,VD} - E_{C,M}^{\text{tr}*} + E_{C,S}^{\text{tr}*}. \quad (2.16)$$

Using the consecutive screens at every model time step (extracted from GEM-MACH model output fields), estimations of $E_{C,S}^{\text{tr}*}$ and $E_{C,M}^{\text{tr}*}$ were made for the period of the nine JOSM 2013 cases around Syncrude, CNRL and Suncor facilities. Retrieved emissions by the TR* method $E_C^{\text{tr}*}$ (where tr* refers to TR* and tr to TR method estimates) were compared to the MIEs as well as to the estimations made with the DR and TR methods (E_C and E_C^{tr} , respectively).

Table 2.1 summarizes the main features of the three retrieval methods (DR, TR and TR*), along with possible practical/observational applications for each. Meta-data-set types, terms for estimating the source emission rates, descriptions for each method, and their relevant equation numbers are also noted on the table. The horizontal flux and deposition terms $E_{C,H}$, $E_{C,VD}$ are shared among all three methods, with the other terms approximated for the TR and TR* methods. Note that here $E_{C,VD}$ is “directly” extracted from the model output, but in observational methods this term is also approximated based on the other measured fields (Table 2.1). Results and discussions follow.

Table 2.1: Summary of the three retrieval methods: direct retrieval (DR), TERRA retrieval (TR) approximation and revised TERRA retrieval (TR*). See Eqs. (2.4)–(2.15) and equations in Appendix A for descriptions of each term.

Method	Meta data	Directly derived terms	Approx. terms	Equation numbers	Model-based applications, methods developed here	Potential observational applications
DR: E_C	4D	$E_{C,H}$, $E_{C,V}$ $E_{C,VD}$ $E_{C,M}$ $E_{C,S}$	—	(A.1) (A.3) (2.4) (2.5) (2.7)	Comprehensive mass-balance analysis using model output volumetric time-series data.	Aircraft retrievals with air-quality model estimates of temporal trends and the impact of storage-release.
TR: E_C^{tr}	2D	$E_{C,H}$ $E_{C,VD}$	$E_{C,V}^{\text{tr}}$ $E_{C,M}^{\text{tr}}$	(A.1) (2.10) (2.11) (2.12)	Simulating observation-based mass-balancing by limiting the analysis to data along box walls (screens), no storage estimate.	Aircraft-based measurements (standard methods).
TR*: $E_C^{\text{tr}*}$	3D	$E_{C,H}$ $E_{C,VD}$	$E_{C,V}^{\text{tr}}$ $E_{C,M}^{\text{tr}*}$ $E_{C,S}^{\text{tr}*}$	(A.1) (2.14) (2.15) (2.16)	Improving upon the TR method by estimating the storage term from time-consecutive screen data.	Aircraft-based measurements with estimate of storage via (a) repeat flights, (b) remote profiling (lidar).

2.3 Results and discussion

Figure 2.3 summarizes the retrieved emissions by the three methods (DR, TR and TR*) and their performance against the MIEs; vertical lines represent the over-/under-estimates by the TR method due to the exclusion of the storage term ($E_{C,S}$) in Eq. (2.12), for each case. Also shown in Fig. 2.3 for reference are the $\pm 50\%$, 30% and 20% error limits, with $\pm 30\%$ corresponding to the maximum estimated uncertainty for TERRA applications in Gordon et al. (2015). Estimates for most of the flight cases studied here using GEM-MACH output fell within $\pm 30\%$ of MIEs, exceptions being cases on 26 and 28 August (cases 5 and 8). The latter two flights were not analyzed in past work due to unfavorable conditions for reliable application of TERRA, namely, light variable winds (case 5) and high upwind emissions (case 8). However, they have been included here using GEM-MACH predictions in mass-balance estimates, as extreme cases in which storage-and-release events have a significant influence on resulting retrievals. We herein refer to cases 5 and 8 as the rejected cases. Using the DR method, we were able to determine the relative contribution of the different terms to the net estimate of E_C (Fig. 2.4).

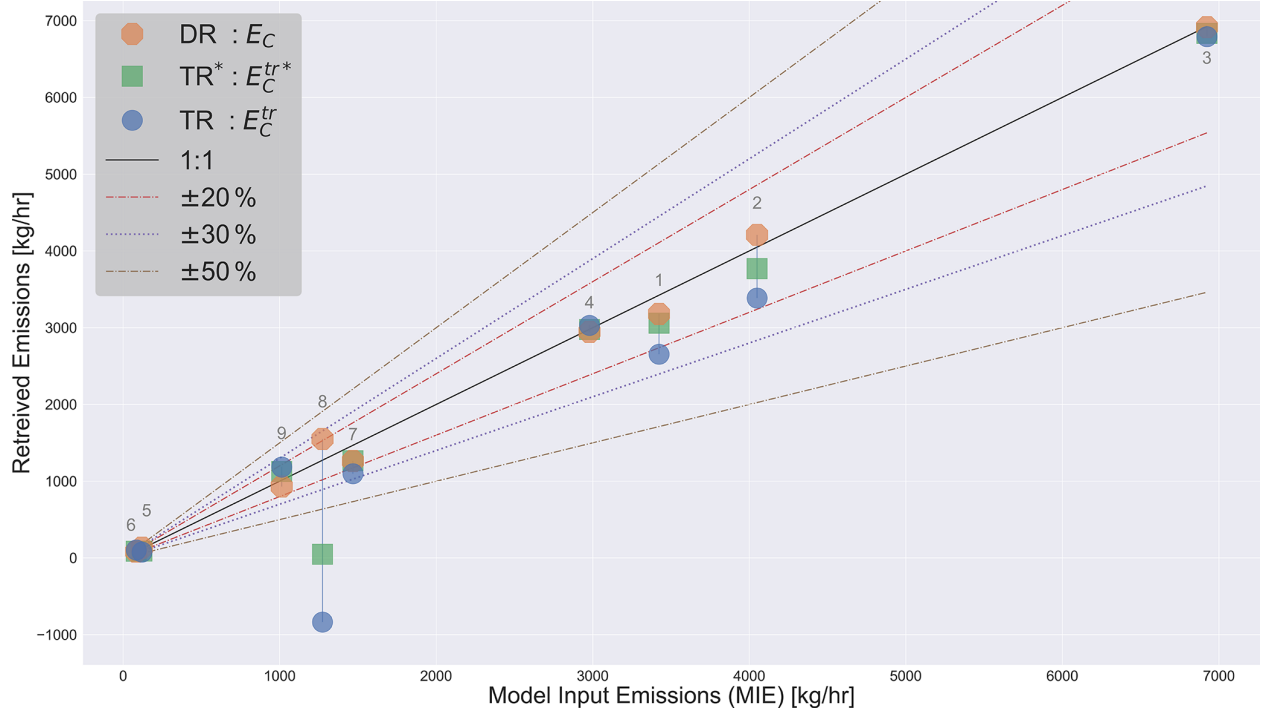


Figure 2.3: Emission rates determined by the three methods (DR, TR and TR*) are plotted against model input emissions (MIEs). The dark solid curve is the 1 : 1 MIE line; estimates below and above this line represent under- and over-estimation by the three methods. Case IDs are noted for each set. The vertical lines represent the error associated with storage-release events ($E_{C,S}$) in TR estimates relative to MIEs. Also shown in the figure are the $\pm 50\%$, 30% and 20% error limits (dashed lines) for reference.

2.3.1 Direct retrieval (DR)

DR estimations were made by analyzing model 4D data for flux boxes approximating the nine cases. The numerical integration expressions for calculating each term, using model data, are described in Table A.2. Estimates of E_C for SO_2 emission rates for the three oil sands facilities (Syncrude, CNRL and Suncor) were made by substituting the calculated terms from Table A.2 in Eq. (2.7). Figure 2.4 demonstrates the contribution of each of the terms in Eq. (2.7) to the total integrated emission rate E_C by DR from the GEM-MACH model 4D data for the nine flight cases. As shown in Fig. 2.4 the main contribution came from two terms, the net integrated horizontal flux $E_{C,H}$ and the storage term $E_{C,S}$, together contributing $> 97\%$ of the net change in mass. For seven out of nine studied cases, $E_{C,H}$ accounted for $> 80\%$ of the total emissions (E_C). This is in agreement with another model-based study by Panitz et al. (2002), where the contribution of advective fluxes ($E_{C,H}$ and $E_{C,V}$) were estimated as 85% and 95% with no estimation of the storage term. The combined contribution of the other terms ($E_{C,V}$, $E_{C,VD}$ and $E_{C,M}$) were $< 1\%$ for all cases except for cases 6 and 8 (2 September and 28 August) where it was 1.2% and 2.9% , respectively. For two out of nine studied cases, the contribution of the storage term ($E_{C,S}$) was $> 30\%$, the maximum estimated uncertainty for cases analyzed in Gordon et al. (2015). As discussed earlier, these flights were not

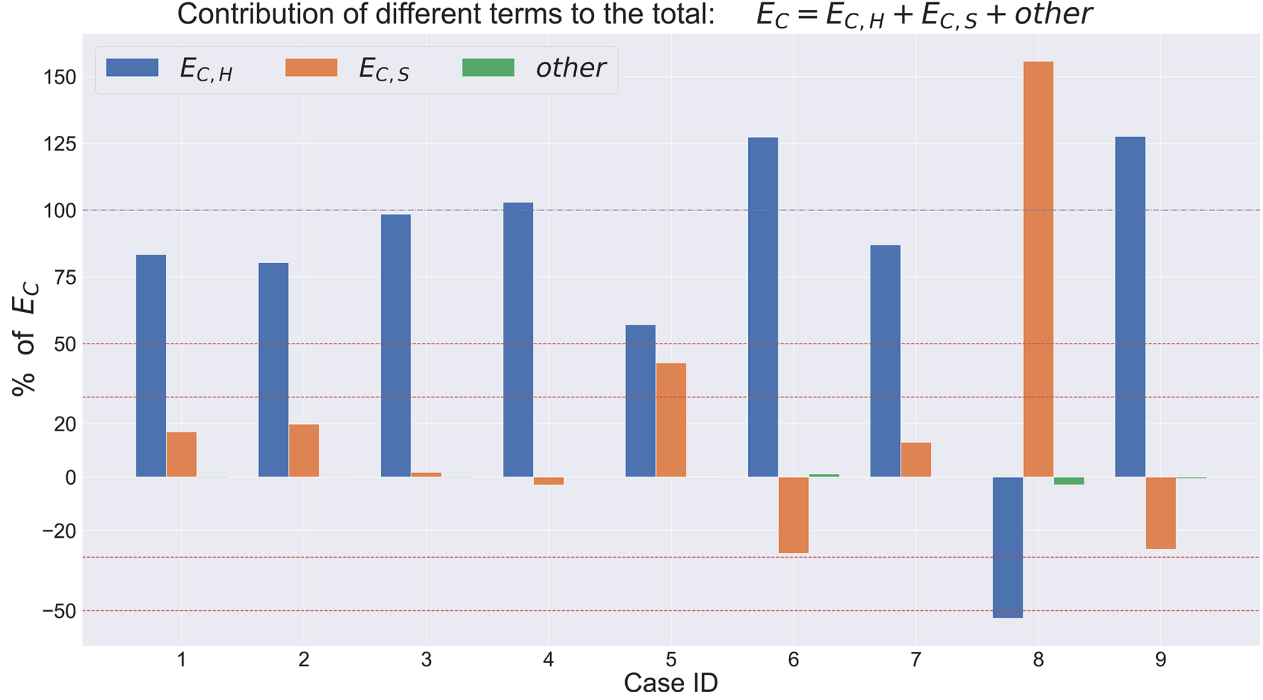


Figure 2.4: Contribution (%) of different terms in Eq. (2.7) to the total retrieved emissions (E_C), for the nine JOSM 2013 cases. The 100 % contribution level is marked with a (purple) dot-dashed line, ± 30 % and 50 % levels are also marked with (red) dashed lines on the figure. The main contributing terms were $E_{C,H}$ and $E_{C,S}$ (> 97 %). The other terms contributed < 3 % combined.

analyzed in past work — but they have been included here as good examples of cases in which time-variant meteorology and/or high upwind emissions result in significant over-/under-estimates due to storage-and-release events. We also show later in Sect. 2.3.3 how these estimates may be improved.

DR method estimates were compared to the MIEs, as shown in Figs. 2.3 and 2.5 and Table 2.3. Estimates were within 1 %–23 % of MIEs for the nine studied cases. We note that in some cases (4, 6 and 9), negative $E_{C,S}$ values reduced the net E_C estimate. Rejected cases 5 and 8 were the two cases with the largest storage term ($E_{C,S}$) contributions of 43 % and 156 % (averaged over flight time) of the total estimated emissions (E_C), respectively. Once corrected for storage (i.e., DR method), estimates for these two cases were within 12 % and 23 % of MIEs. Cases 3 and 4 had the lowest storage rates (2 % and -3 %), where storage-corrected estimates were within 1 % and 2 % of MIEs, respectively.

For the majority of our studied cases, our results indicate that the storage term (if not accounted for) contributes the bulk of emission rate over-/under-estimates for the variables investigated here, and hence methods to predict and/or reduce its influence are desirable for aircraft retrieval algorithms. We note that the impact of storage on emissions retrieval estimates will depend on the relative magnitude of the estimated emissions themselves in comparison to other sources of uncertainty. That is, an emissions estimate double that of previous estimates, which has a ± 30 % over-

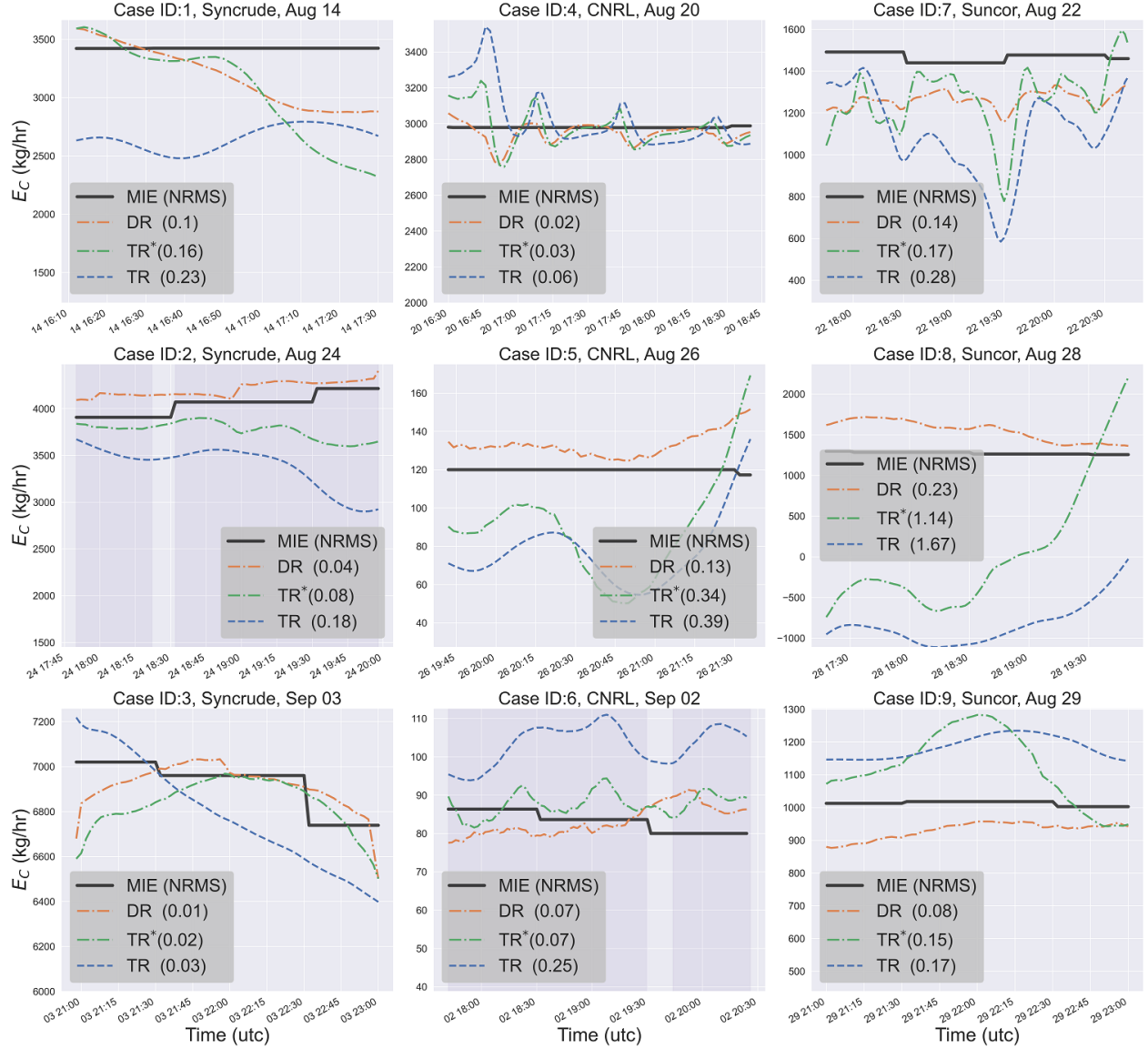


Figure 2.5: Time series of the estimates by DR, TR and TR* methods, plotted along with MIEs. Normalized-root-mean-square (NRMS) error for the three methods are noted on the plot for each of the nine JOSM 2013 cases. JOSM 2013 flights on 24 August (case 2) and 2 September (case 6) included two consecutive box flights around the same facility, shown as shaded areas on their respective time-series plots. Note that the DR and TR* methods, unlike the TR method, account for the storage rate ($E_{C,S}$ and $E_{C,S}^{tr}$).

or under-estimate associated with storage, may still be concluded as being significantly higher than the previous estimates (i.e., the emissions change is greater than the uncertainty associated with the retrieval itself). However, the storage term as defined here has been shown to be the main contributor to emissions over-/under-estimates, and will therefore be examined in more detail in the analysis which follows. We also note that the impact of $E_{C,S}$ is sometimes negative, a net release of mass from the box. Note that net releases of stored mass from the box would result in emissions

over-estimates; net storage of mass within the box would result in emissions under-estimates.

Storage-and-release events occur when mass is accumulated within the volume of the box and released at a later time through box lateral walls; see the detailed discussions in Sect. 2.4. Such events can be detected on the box walls by comparing the temporal variation in the net horizontal flux, $E_{C,H}$ with $E_{C,S}$. Figure 2.7c and d (Sect. 2.4.1) show time series of the horizontal flux $E_{C,H}$ and the storage rate $E_{C,S}$ for case 4 on 20 August 2013 (CNRL). Storage-and-release events are represented by peaks and troughs in the $E_{C,S}$ curve, which coincide with troughs and peaks in the $E_{C,H}$ curve. $E_{C,H}$ and $E_{C,S}$ are anti-correlated (with Pearson correlation coefficient of $P_r = -0.95$) for case 4. P_r values between $E_{C,H}$ and $E_{C,S}$ were in the range -0.68 to -0.98 for the nine cases, indicating strong anti-correlations for all the studied cases. Positive values of $E_{C,S}$ correspond to storage, which result in decreased net flux exiting through the box walls; negative values correspond to release, which results in increased net flux exiting through the box walls. Therefore, temporal variations in $E_{C,H}$ over the period of a flight around an emission source with relatively constant emission rates, can be attributed to mass storage/release. This provides a potential practical means for detecting storage-release events in aircraft-based retrieval methods as the horizontal flux ($E_{C,H}$) can be calculated directly using airborne measurements along the box lateral walls, if ancillary information is available, such as from repeated flights. Changes to the calculated $E_{C,H}$ can indicate the potential presence of storage-and-release events within emissions observation data collection time periods.

2.3.2 TERRA retrieval (TR) approximation

By limiting the analysis to the extracted model data along box lateral walls (Fig. 2.2), TR estimates of MIEs were made. Terms in Eq. (2.12) were solved numerically using model data, with $E_{C,H}$ calculated as described in Table A.2; numerical expressions for calculating the rest of the terms are described in Table A.3. Model surface deposition rates for the base area of the box (A) were used in TR estimations, similar to the DR method, the main difference between these two methods is that, unlike Eq. (2.7), Eq. (2.12) does not account for the storage term. TR estimates of SO_2 emission rates were within 4%–39% of MIEs for eight out of nine studied cases; see Table 2.3. For cases 3 and 4, the two cases with the lowest net storage, results were within 4%–6% of MIEs. For the rejected case 8 with the extreme storage event, the TR method under-estimated the model input emissions by -166% ; during the sampling period of this case, on average ~ 1.5 times more SO_2 mass entered the flux box through the upwind wall than left through the downwind wall.

Comparing the performance of the two methods (DR and TR) against MIEs reveals the significance of the storage term ($E_{C,S}$) in emission rate retrievals. Figure 2.5 shows the time series of retrieved emissions using the DR and TR methods for the nine cases, with model time-step temporal resolution ($\Delta t = 2$ min). MIE time series are also plotted for SO_2 emissions from all the sources within each facility. Note that in aircraft-based applications of TERRA, a single emission rate estimate (temporal average) is generated based on the observations during the entire flight-time (~ 2 h). Here, where we approximated the TERRA retrievals by the TR method (see Sect. 2.2.2),

instantaneous 2D screen fields (extracted from the GEM-MACH model output) were used to generate an emission rate estimate at each time step using the TR method. The flight-time average estimates by the DR and TR methods were compared to the flight-time average MIE; normalized root mean square (NRMS) error scores are noted for each method in Fig. 2.5.

By combining the information from Figs. 2.4 and 2.5, TR estimates for the nine cases can be categorized in three groups: (a) net positive storage (resulting in emission rates under-estimation), (b) net negative storage (i.e., release of stored mass from the box during the time period studied, resulting in emission rates over-estimation), and (c) low storage amounts, the best emission rate estimates. The first group (a) includes cases 1, 2, 5, 7 and 8. For these cases the net storage was positive, where 13%–156% of the emitted SO_2 mass was accumulated within the box volume and was not released through box walls during the sampling time. The stored mass was accumulated within the box either before or during the sampling time, which in turn resulted in decreased flux through box walls and the consequent under-estimation in TR estimations. Cases 6 and 9 fall into group (b) where the previously stored mass was released (−29% and −27%) through the box walls along with the SO_2 mass from emission sources during the flight time, resulting in TR method over-estimation. Group (c) include the ideal cases 3 and 4 where TR estimates were within 4% and 6% of MIE, respectively. These two cases captured consecutive storage-release pairs over the sampling period (see Sect. 2.4.1, Fig. 2.7c and d for case 4) that canceled each other out and therefore resulted in relatively low (sampling time average) storage rates ($E_{\text{C,S}}$) of 2% and −3%, respectively.

2.3.3 Revised TERRA retrieval (TR*)

Here we show that change in tracer mass within the flux box can also be observed on box lateral walls over time (using time-consecutive screens), as described in Sect. 2.2.2. Figure 2.6 demonstrates the correlation between the 4D (volume) and 3D (screens) estimates of the temporal change in species mixing ratio ($\partial\chi_{\text{C}}/\partial t$) for all the data points (2 min resolution) corresponding to the period of the nine studied cases. The side plots show the range (and distribution) of the estimated $\partial\chi_{\text{C}}/\partial t$ by 3D (horizontal axis) and 4D (vertical axis) calculations. The $\partial\chi_{\text{C}}/\partial t$ range is slightly higher for 4D estimates (2.52 ppbv h^{−1}) than for 3D estimates (2.17 ppbv h^{−1}), and the two variables are strongly correlated. The correlation coefficient $P_{\text{r}} = 0.7$ (with slope $m = 1.36$ and $b = 0.09$ intercept for the least-square fit line), indicates strong correlation between 4D and 3D estimates. An exception is rejected case 8, where the correlation is low due to relatively high upwind emissions during the time period of this case (on 28 August). Excluding case 8 from this analysis, increases the correlation between 3D and 4D estimates to $P_{\text{r}} = 0.9$ (with $m = 1.07$ and $b = 0.03$). As discussed earlier, the lack of information about the storage rate was the primary source of under-/over-estimates in TR method estimates for the majority of our studied cases. Here we demonstrate the use of an estimate of the storage term ($E_{\text{C,S}}^{\text{tr}*}$; see Sect. 2.2.2) in a revised retrieval approach. We note that this term may be estimated from consecutive screens (Eq. 2.15) or tandem aircraft samples for ambient atmosphere applications. As discussed in Sect. 2.2.2, an alternative to multiple

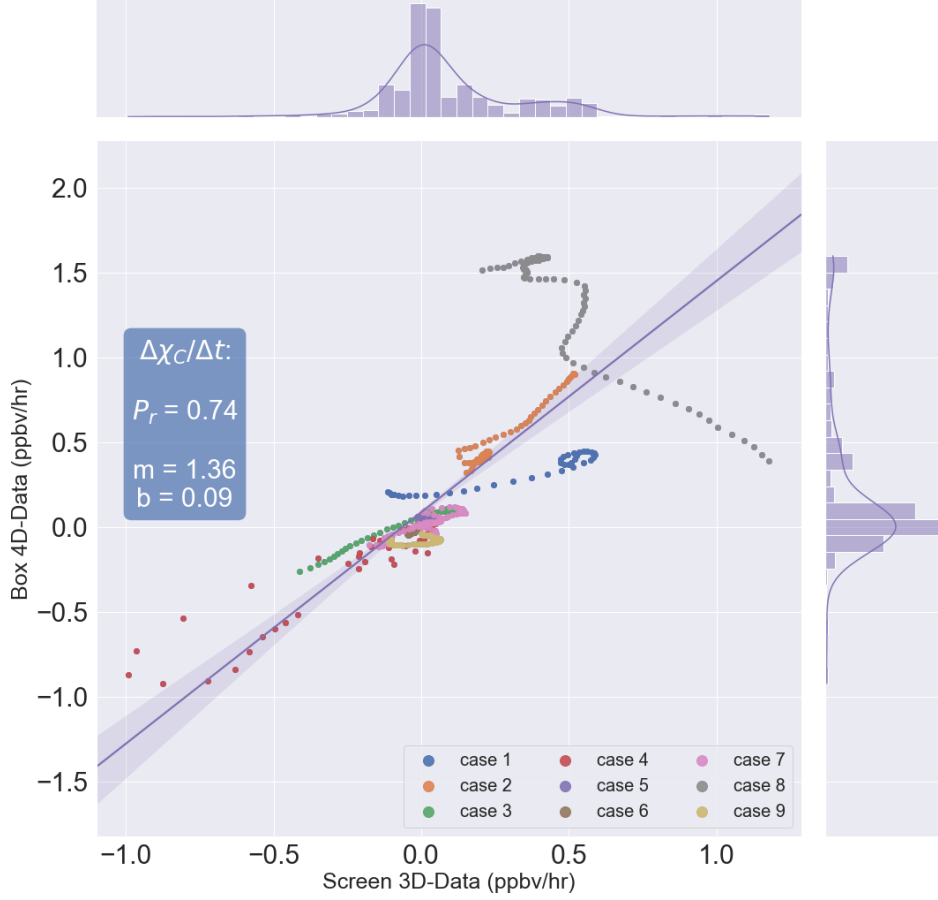


Figure 2.6: This three-panel graph demonstrates the correlation between estimates of $\partial\chi_C/\partial t$ by analyzing 4D (box) and 3D (consecutive screens) data sets. The vertical side plot shows the range and distribution of $\partial\chi_C/\partial t$ estimates for the period of nine studied flights by 4D analysis, the horizontal side plot shows the same for 3D analysis. The main panel demonstrates the correlation between the two methods of estimation. Pearson correlation coefficient ($P_r = 0.74$) shows the strength of this relation. The least-squares fit line has slope $m = 1.36$ and intercept $b = 0.09$. Note that case 8 estimates are in low correlation due to relatively high upwind emissions during the time of that particular flight case (28 August), rendering the 3D estimates less representative of the 4D estimates. Excluding case 8 from this analysis, results in $P_r = 0.88$, $m = 1.07$ and $b = 0.03$. $E_{C,S}^{\text{tr}*}$ can be estimated from multiple consecutive 2D screens, by assuming 3D estimations of $\partial\chi_C/\partial t$ as representative of the rate of change in species mass within the flux box. The same can also be estimated from downwind vertical profiling of tracer concentrations over the sampling time.

aircraft-based measurements is to estimate the temporal trends in tracer concentrations ($\partial\chi_C/\partial t$) through remote vertical profiling (e.g., lidar measurements). Our analysis of different flight cases show that downwind trends are also in strong correlation with DR method volumetric time-series estimates with the same correlation coefficient of $P_r = 0.9$ (with different $m = 0.57$ and $b = 0.03$). The advantage of remote (downwind and upwind) vertical profiling over multiple aircraft or UAV measurements, in addition to operational cost efficiency, would be the collection of more temporal measurements over the sampling time, which in turn can result in improved estimates

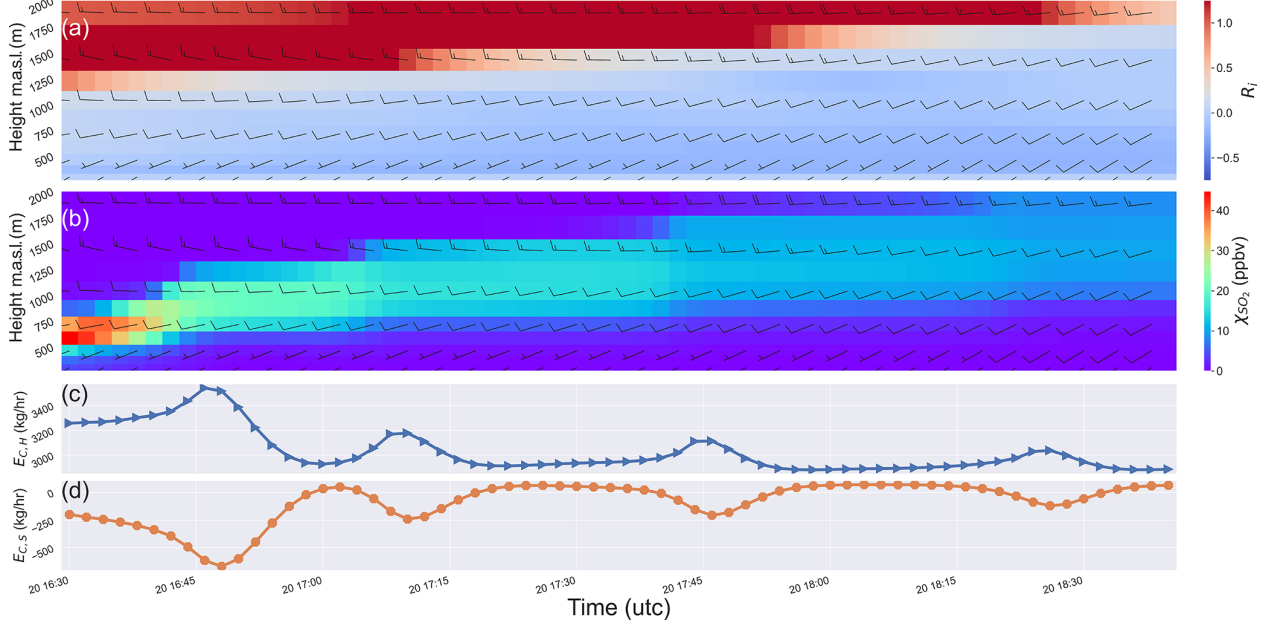


Figure 2.7: This schematic demonstrates (qualitatively) the correlation between change in atmospheric stability represented by the gradient Richardson number (Ri), plume vertical motion and storage-release events for the period of case 4. **(a)** Ri vertical profiles at stack location over the flight time, with the color-scale symmetric around the critical gradient Richardson number $Ri_C = 0.25$. **(b)** SO_2 mixing ratio vertical profiles at stack location. **(c)** Time series of net horizontal flux ($E_{C,H}$) and **(d)** the storage rate ($E_{C,S}$), over sampling time. Note the level of anti-correlation between $E_{C,H}$ and $E_{C,S}$ ($P_r = -0.95$); oscillations in storage and release are occurring: changes in $E_{C,S}$ are matched by opposite changes in $E_{C,H}$. Wind speed and direction are shown as barb plots with units $m s^{-1}$ (half = 5, full = 10). For every incremental rise in plume height, a corresponding release event occurs. As the plume ascends into regions with faster wind speeds, more mass is released ($E_{C,S} < 0$) through the lateral walls resulting in an increase in $E_{C,H}$ as can be seen in **(c)**.

of $\partial\chi_C/\partial t$ and the storage term ($E_{C,S}^{tr*}$), as opposed to estimates based on a few time-consecutive measurements (e.g., two to three) via multiple or in tandem aircrafts. Data from time-consecutive screens/measurements can then be used to numerically solve Eqs. (2.14) and (2.15) (see Table A.4 for discrete integral expressions). For GEM-MACH-driven retrievals as used here, these time-consecutive screens originate from consecutive model output times. Here, we show the effect of the $E_{C,S}^{tr*}$ estimate on the resulting TR^* emission rate estimates (E_C^{tr*}) using GEM-MACH fields for top-down retrievals. Improved/ revised TERRA retrievals (TR^*) can be made by plugging in $E_{C,M}^{tr*}$ and $E_{C,S}^{tr*}$ in Eq. (2.16). TR^* method estimates are compared to the MIEs in Figs. 2.3 and 2.5 and Table 2.3. For eight out of nine cases, TR^* estimates were within 2 % to 34 % of MIEs. TR^* estimates showed improvement over the TR method for all the cases (including the rejected case 8) by 2 %–52 % of MIEs; see Table 2.3. TR^* estimations rely on the availability of temporal data for estimating change in tracer concentrations ($\partial\chi_C/\partial t$) along the box lateral walls (or only on the downwind wall) for each emission rate retrieval case, which in our analysis with simulated fields were readily available with the GEM-MACH model data at each time step. Again,

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

2.4 Three quantitative predictors of storage-and-release events

The main processes contributing to the change in SO_2 mass within the volume of the box are illustrated in Fig. 2.2a (E_C , $E_{C,S}$, $E_{C,H}$, $E_{C,V}$ and $E_{C,VD}$). An SO_2 emission source continuously adds mass to the box at emission rate E_C . Most of the mass is advected out of the box through the top and lateral walls ($E_{C,V}$ and $E_{C,H}$), and some is deposited to the ground surface ($E_{C,VD}$). All the while, mass is continuously stored within the volume of the box and may later be released at net storage rate $E_{C,S}$ (positive to negative). Under time-invariant conditions (e.g., wind field, atmospheric stability, emission rate), the rate of generation of mass (E_C) is at net (mass) balance with the other mass-removing processes ($E_{C,H}$, $E_{C,V}$ and $E_{C,VD}$) and the total accumulated mass within the box volume (B_C) remains relatively constant over time, and the storage rate negligible ($E_{C,S} \simeq 0$). The mass-balance approach based on the divergence theorem (Eq. 2.12) can therefore accurately equate the source emission rate to the net integrated flux leaving the box through the enclosing surfaces. However, localized (spatial) variations in meteorological fields (e.g., atmospheric stability, wind) and chemical fields (e.g., temporal changes in source emission rate (E_C) and/or significant incoming upwind emissions) during the retrieval time, can shift the mass balance towards $E_{C,S} \neq 0$ (Eq. 2.7) – i.e., storage-and-release events. The scale and duration of the storage-and-release events contribute to over- and under-predictions in estimated emissions based on the mass-balance approach. Using GEM-MACH model 4D data we carried out extensive analysis of the storage-release events during our studied cases and devised three quantitative predictors for the severity of these events. In the following subsections, using results from DR, TR and TR* estimates, we discuss these parameters and their correlation with over- and under-estimates in top-down mass balance emission rate retrievals due to storage-and-release events.

2.4.1 Gradient Richardson number (Ri) and changing atmospheric stability

Aircraft-based emissions retrieval methods such as TERRA, utilize the mass-balance approach with the underlying assumption of mean steady-state conditions over the region of study ($\sim 100 \times 100 \text{ km}^2$) during the flight time ($\sim 2\text{--}3 \text{ h}$). However, localized and transient variations in atmospheric stability at the vicinity of the stack/source location can impact the pollutant plume rise and transport within the box and result in storage-release events. Horizontal wind velocity usually increases logarithmically with height. The height the plumes reach in the atmosphere depend on the stack parameters (exit velocity, exit temperature, stack height) and on the local stability conditions (temperature and wind profile near the stack location). Thus plume height

may change with varying stack parameters and stability conditions during the course of a retrieval. This in turn would move the mass to a different flow regime – e.g., higher horizontal velocities if the plume height has increased, and lower horizontal velocities if the plume height has decreased. These changes in plume height, in response to changes in atmospheric stability, may thus lead to a change in the rate at which emitted mass is transported out of the box. Stability changes are thus a good predictor of the possibility for mass storage and release. A measure of the change in dynamic stability is therefore desirable as a predictor for conditions under which this aspect of the “static meteorology” condition, required for reliable top-down emission rate retrievals, is not met.

Atmospheric dynamic stability can be described in terms of the gradient Richardson number (Ri), which is a dimensionless ratio, related to static stability and the shear induced turbulence (American Meteorological Society – AMS, 2021b),

$$Ri = \frac{\frac{g}{T_v} \frac{\partial \theta_v}{\partial z}}{\left(\frac{\partial U}{\partial z}\right)^2 + \left(\frac{\partial V}{\partial z}\right)^2} \quad (2.17)$$

where θ_v and T_v are virtual potential temperature and virtual absolute temperature, respectively, g is the gravitational constant and U and V are the horizontal wind components. The sign of Ri is determined by the static instability term ($\partial \theta_v / \partial z$) in Eq. (2.17); the denominator is always positive. In the presence of a strong wind shear Ri approaches 0. The critical gradient Richardson number (Ri_c) is about 0.25, below this value the atmosphere is dynamically unstable. The atmospheric Physics module within the GEM model includes the estimation of atmospheric stability and calculations of the gradient Richardson number Ri (Mailhot and Benoit, 1982), which is also used as a parameter (in addition to emission stack/source information such as temperature, exit velocity) to calculate plume rise height (Akingunola et al., 2018). Changes in Ri with respect to time are thus an indicator of the potential for the plume to change height during an aircraft emission rates retrieval; Ri may be predicted using a weather forecast model as part of pre-flight planning.

For our analysis, 4D fields of Ri were extracted from the model output dataset for each of the nine studied cases. Figure 2.7 shows the case for the JOSM flight on 20 August (case 4) with vertical profile time series at the location of the emitting stack, and compares Ri in (a), with the SO_2 mixing ratio in (b), the net horizontal flux $E_{C,H}$ in (c) and the storage rate $E_{C,S}$ in (d). As shown in Fig. 2.7a, the region of unstable air (blue) was gradually increasing in height during the sampling time, resulting in an increase in the vertical extent of the mixing layer. Note that the color-scale is symmetric around the critical gradient Richardson number $Ri_c = 0.25$ in Fig. 2.7a. The increase in the vertical extent of the mixing layer during the period of case 4, resulted in the SO_2 plume mixing into heights with higher wind speeds (Fig. 2.7b). The incremental plume rise resulted in a sequence of fast-release events, that is, the simulated plume was moved into a faster (higher level) horizontal flow, and SO_2 mass was carried to the boundaries of the box faster – the net rate of mass loss through the downwind walls thus increased during the sampling time period ($\frac{\partial}{\partial t}(E_{C,H}) > 0$, Fig. 2.7c; $\frac{\partial}{\partial t}(E_{C,S}) < 0$, Fig. 2.7d). Note the anti-correlation between $E_{C,S}$ and $E_{C,H}$ in Fig. 2.7c and d. We also note that although faster moving flow can result in more

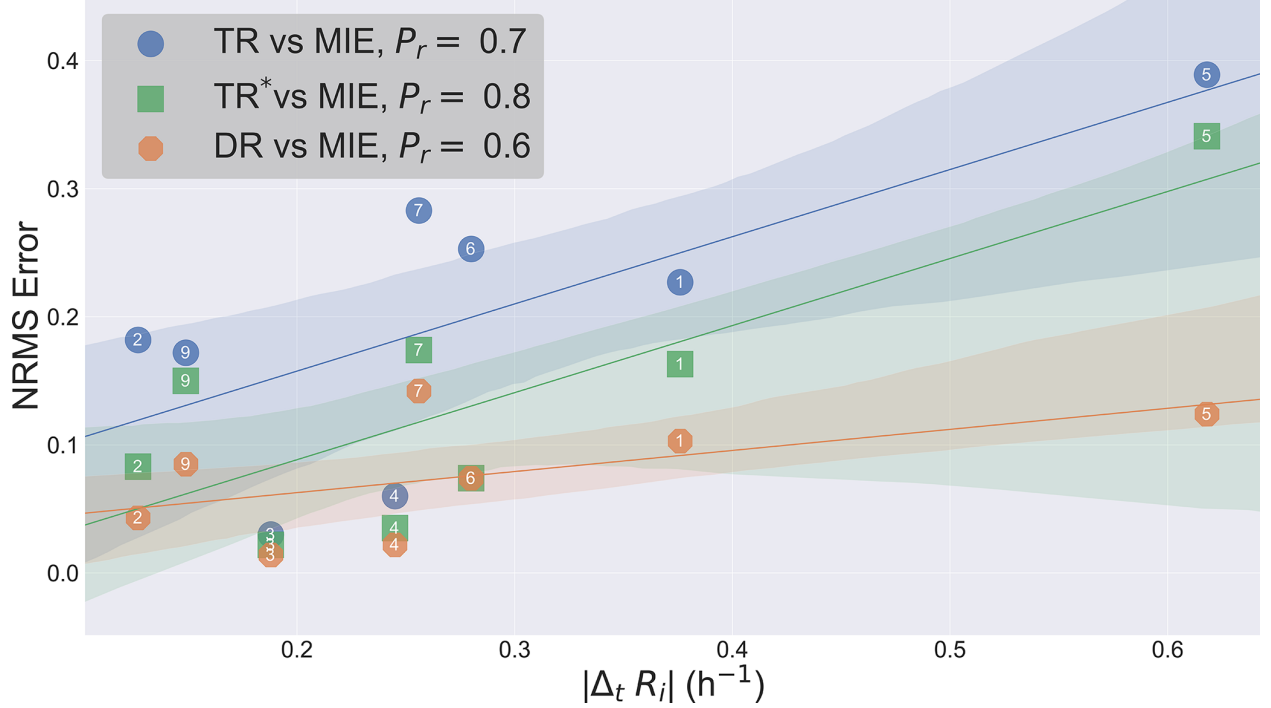


Figure 2.8: Correlation between the NRMS error and the change in atmospheric stability represented by the absolute change in the gradient Richardson number $|\Delta_t R_i|$. The shaded areas indicate the 90 % confidence region. Case ID is indicated for each data point on the graph along with the calculated Pearson correlation coefficients (P_r) for the three methods. Results indicate a strong positive correlation between the NRMS error and $|\Delta_t R_i|$.

mixing and the subsequent plume dilution, the decrease in compound mixing ratio is not necessarily enough to compensate for the increased exiting flux. If emissions retrievals are calculated under these conditions without accounting for the storage term (negative in this case), emissions estimates will be biased high (over-estimation). These fast-release events temporarily disturb the required balance between the addition of mass to the box volume by source emissions (E_C) and removal of the mass through vertical and horizontal fluxes and surface deposition, causing a negative change in the total stored mass (B_C) within the volume of the box ($E_{C,S} < 0$). These release events can be seen as troughs (peaks) in the $E_{C,S}$ ($E_{C,H}$) time-series plot (Fig. 2.7c and d). While the plume remained at a constant height, the balance was restored over time ($E_{C,S} \rightarrow 0$); this can be seen as peaks/plateaus in $E_{C,S}$ time series in Fig. 2.7d. After these periods of changing stability, the plume started to rise again, and the balance was shifted, again resulting in negative $E_{C,S}$ values. This process was repeated several times over the period of case 4, resulting in a sequence of release events. Note that the trough to peak duration (period) and scale (magnitude) of each event are inversely proportional to the wind speed at plume height, and they are reduced as the plume ascends into faster moving layers of air resulting in a damped oscillatory storage and release sequence. In this particular example, this periodic storage release during the sampling time, with subsequent equilibrium intervals, resulted in relatively small net storage of $E_{C,S} = -3\%$ over the sampling

time and TR estimates within 6 % of MIEs (slight over-estimation) for case 4. When consecutive storage-release pairs are captured during the flight/sampling time as in case 4, the average error due to storage term ($E_{C,S}$) is reduced. However, if the storage or release is of longer duration, so that an unpaired event occurs during observations, the net error increases, as was the case for flight case 9 where an unpaired release event (one storage-release pair and one additional release event) resulted in net storage of $E_{C,S} = -27\%$ and a consequent TR over-estimate. Monotonic storage or release events occur when mass is continuously stored (e.g., when a plume descends into slower moving air throughout a sampling interval; $\frac{\partial}{\partial t}(E_{C,S}) > 0$ and $\frac{\partial}{\partial t}(E_{C,H}) < 0$) or released (e.g., release of previously stored mass during a sampling interval; $\frac{\partial}{\partial t}(E_{C,S}) < 0$ and $\frac{\partial}{\partial t}(E_{C,H}) > 0$); such events result in TR under-/over-estimates as in case 2 where net storage of $E_{C,S} = 20\%$ (i.e., storage event: $E_{C,S} > 0$ and $\frac{\partial}{\partial t}(E_{C,H}) < 0$) resulted in TR under-estimations.

The average value $Ri = -0.14$ at plume height, indicates unstable conditions ($Ri < Ri_C$) during the period of case 4 – the average value was calculated by spatially averaging Ri at the vertical layer containing the plume center (maximum concentration in the vertical) over the horizontal extent of the box at each time step and then averaging over the sampling time. Atmospheric stability was further decreased at an average rate of $\Delta_t Ri = -0.25 \text{ h}^{-1}$ over the period of this case. Change in atmospheric stability (represented by $\Delta_t Ri$) is thus related to plume vertical motion and mixing, which in turn contributes to storage-release events and a corresponding over- or under-estimate in TR retrievals; $\Delta_t Ri$ is thus a useful forecast metric to use as an a priori indicator of the potential for changing atmospheric stability affecting emission rates retrieval accuracy, and may also be calculated from retrieved aircraft meteorological data. Table 2.3 lists (flight time) average Ri and $\Delta_t Ri$ values for each of the nine studied cases. Our analysis indicates a strong positive correlation between the absolute value $|\Delta_t Ri|$ and the NRMS error for the DR, TR and TR* methods with the Pearson correlation coefficients (P_r) of 0.6, 0.7 and 0.8, respectively. Figure 2.8 shows the correlation between $|\Delta_t Ri|$ and NRMS error for each method. Note that the rejected case 8 with $\Delta_t Ri = -0.34 \text{ h}^{-1}$ (see Table 2.3) was excluded from correlation analysis as an outlier, due to the extreme impact of high upwind emissions (and not meteorological variability) on emission rate estimates (further discussed in Sect. 2.4.3). $\Delta_t Ri$ can be calculated from aircraft-measured data along the lateral walls of the box and the measured data from additional profiling spiral flights upwind and downwind the source, such as those conducted during JOSM2013 campaign. Analysis of $\Delta_t Ri$ can provide insights regarding the change in atmospheric stability and the probability of plume vertical motion and the resulting storage-release events during flight time; the uncertainty levels associated with the storage rate ($E_{C,S}$) in the retrieved emissions can therefore be quantified.

2.4.2 Plume vertical motion and variations in the direction of the transport ($\Delta_t \alpha$)

The direction of the wind experienced by the plume may change when the plume remains at one altitude, or may change as the plume rises or falls in the atmosphere (e.g., via the well-known “Ekman spiral” of wind direction changes with increasing height, AMS, 2021a). Changes in wind direction for a fixed plume height would result in the rejection (for top-down retrievals) of the

flight – here we note that changes in wind direction due to plume vertical motion may also be a potential reason for flight rejection. Using GEM-MACH-predicted wind fields for the nine studied cases (chosen a priori for relatively constant wind speeds at each height), we here investigated the extent of variations in wind direction at the time-varying vertical level containing the plume center,

$$\Delta_t \alpha = \frac{\sigma_\alpha}{\Delta t^{\text{SP}}}, \quad (2.18)$$

where σ_α is the standard deviation in wind direction α_{wind} time series (spatially averaged over the horizontal extent of the box including lateral walls at each time step) and Δt^{SP} is the *sampling period* time duration. In aircraft-based retrievals (e.g., TERRA), $\Delta_t \alpha$ can be estimated from aircraft measurements around the box at plume center-point height. Our analysis of the wind fields for the nine studied cases indicates a positive correlation ($P_r = 0.6$) between the rate of variation in wind direction ($\Delta_t \alpha$) and the rate of change in atmospheric stability ($\Delta_t Ri$). Variations in the direction of the flow at plume height ($\Delta_t \alpha$) can result in the re-circulation of the plume, accumulation of the emitted mass within the volume of the box and its release at a later time (storage-release event). Such an event occurred during our GEM-MACH model simulations for the period of the JOSM flight on 26 August around CNRL (case 5). Figure 2.9 shows horizontal cross sections of the SO_2 mixing ratio data at two (consecutive) plume-maximum heights and over time for case 5. Starting from top left at 20:20 UT and height $H_1 = 909$ m a.s.l. (meters above sea level) to the bottom right at 21:30 UT and height $H_2 = 1079$ m a.s.l., horizontal cross sections are plotted at plume center-point (maximum concentration in the vertical) height at each time. Air flow direction and strength are shown as grey arrows on each panel. The model shows strong vertical wind shear between these two heights, with a $\sim 180^\circ$ shift in the wind direction between the upper and lower rows of panels which remains relatively constant with time. This is a subtle example of how wind direction changes may interact with changes in stability: while the wind direction is relatively static at each height, changes in stability may cause the plume to rise to different heights, where the wind direction may change or even reverse relative to the level below. The changes in turn will alter (slow) the rate of mass release from the walls of the box ($\frac{\partial}{\partial t}(E_{\text{C,H}}) < 0$), a storage event, with $\frac{\partial}{\partial t}(E_{\text{C,S}}) > 0$. For this particular case (case 5), between 20:20 and 20:50 UT, the plume center was at H_1 and was advected in a weak south-easterly wind; between 20:50 and 21:00 UT, the plume mixed upwards and rose to H_2 , where wind direction was north-north-westerly and remained there until 21:30 UT (which afterwards descended back to H_1). This change in wind direction at plume height ($\Delta_t \alpha$) as a consequence of plume vertical motion due to change in atmospheric stability ($\Delta_t Ri \neq 0$), resulted in a net storage of $E_{\text{C,S}} = 43\%$ during case 5 and consequently an under-estimation in emission rate using the TR method.

Cases 6 and 9 (both release events) along with rejected case 5 (storage event) were the three cases most impacted by variations in wind direction at plume-center height. Flight time average $\Delta_t \alpha$ values for all nine cases are given in Table 2.3. Our results indicate a positive correlation between $\Delta_t \alpha$ and NRMS error in DR, TR and TR* estimates, similar to $\Delta_t Ri$; Fig. 2.10 demonstrates this correlation for the three methods, with P_r scores 0.4, 0.7 and 0.6, respectively. As mentioned earlier,

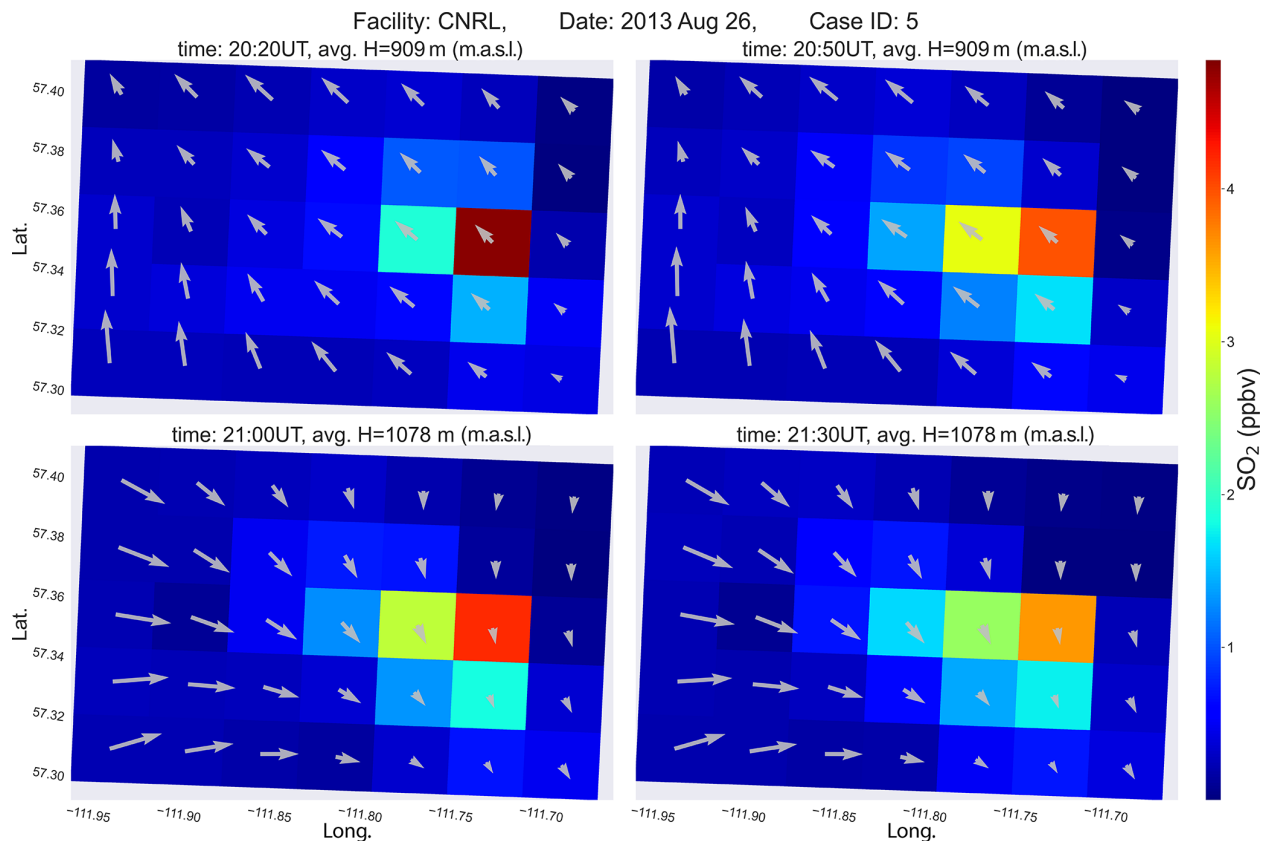


Figure 2.9: SO₂ mixing ratio horizontal cross sections at two (consecutive) plume center (max. conc.) heights (meters above sea level – m.a.s.l.) and four time steps, during the period of case 5 over CNRL. The strength and direction of the airflow at each cross section is represented by grey arrows. Corresponding time and average height are noted for each panel. As the plume rises into layers with different wind direction over time, it re-circulates within the volume of the box which in turn results in a storage event.

there is also a positive correlation between $\Delta_t \alpha$ and $\Delta_t Ri$ ($P_r = 0.6$); the combined effect of change in the direction of the transport and changing stability increases the uncertainty in the retrieved emission rates. In aircraft-based emission rate retrievals, wind shear strength in terms of $\Delta_t \alpha$ can be included in uncertainty analysis associated with plume vertical motion due to change in atmospheric stability ($\Delta_t Ri$). Aircraft-measured data along the box walls combined with measurements during additional spiral flights can be used to evaluate these metrics.

2.4.3 Weighted upwind to downwind concentration ratio (ϕ)

In applying the mass-balance technique, the mass flux associated with upwind emissions (and regional background levels) entering the box volume (inflow) is subtracted from the downwind outgoing flux containing the emissions within the box (outflow) to estimate the net emission rate from sources within the box volume. The mass inflow associated with regional background levels (e.g., for CH₄ and CO with relatively high background concentrations) is balanced out with the

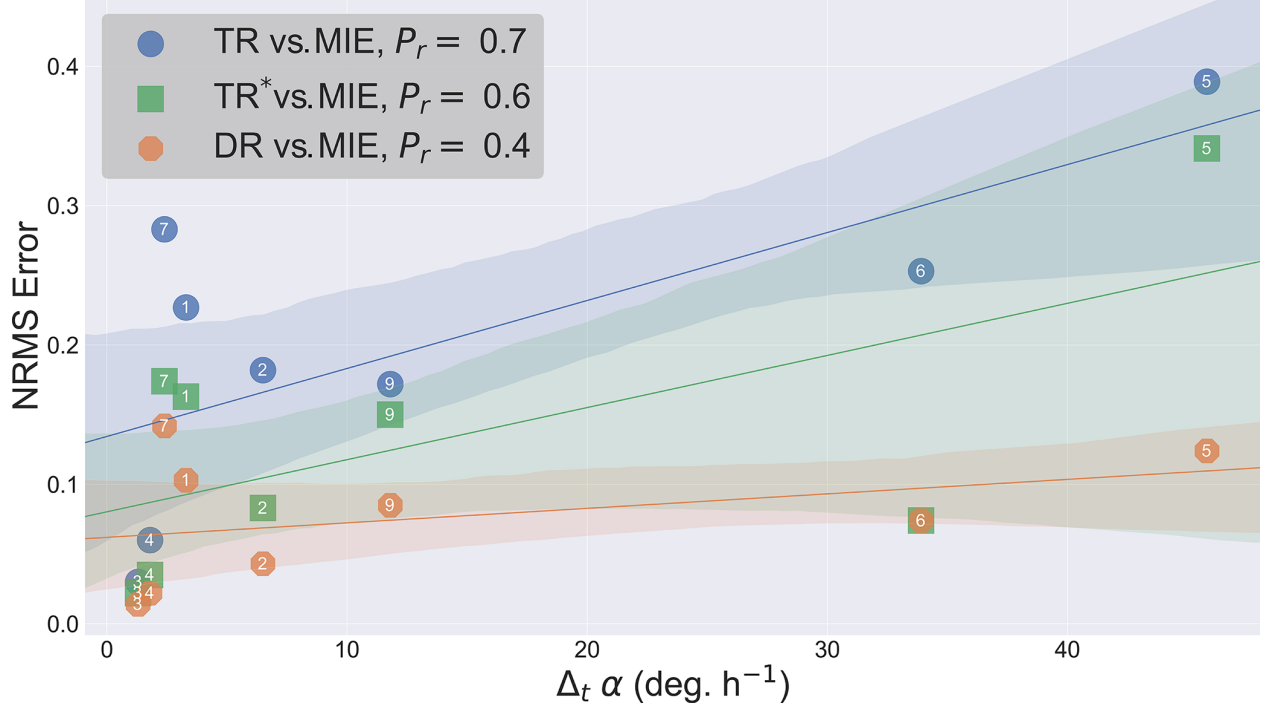


Figure 2.10: Correlation between the NRMS error and the change in wind direction at plume height ($\Delta_t \alpha$). The shaded areas show the 90 % confidence region. Case ID is noted for each data point on the graph along with the calculated Pearson correlation coefficients (P_r) for the three methods. Results indicate a positive correlation between the NRMS error and $\Delta_t \alpha$.

mass flux out of the box due to spatial (horizontal) homogeneity of background concentrations. However, emissions from upwind sources carried through the sampling region (box volume) are spatially heterogeneous (plumes) and therefore their inflow and outflow through a control volume is not necessarily equal and canceling. For instance, converging horizontal wind fields and/or plume vertical motion and mixing can result in temporary accumulation of the mass of these upwind plumes within the box (control volume). If the emissions from upwind sources are large relative to the within-box source(s), the emission rates retrieval may therefore become less accurate. That is to say, all the processes (including the ones described above in terms of $\Delta_t Ri$ and $\Delta_t \alpha$) that contribute to storage-release of the mass emitted by the source(s) within the flux box may also affect the mass emitted from upwind sources which are entering the box through the upwind wall(s) – leading to further increased uncertainty in the retrieved emission rates for sources within the box. The box upwind-to-downwind concentration ratio is in strong correlation with the relative magnitude of upwind emissions entering the box to the emissions from a given source within the box, with $P_r \geq 0.8$ as determined for our nine studied cases. To distinguish upwind emissions from regional background levels, this ratio can be weighted by temporal standard deviations in upwind and downwind concentrations. Hence we define

$$\phi = \left(\frac{\overline{\chi_{C,u}}}{\overline{\chi_{C,p}}} \right) \left(\frac{\sigma_{\chi_{C,u}}}{\sigma_{\chi_{C,d}}} \right), \quad (2.19)$$

where $\overline{\chi_{C,u}}$ is the screen average upwind concentration and $\overline{\chi_{C,p}}$ is the plume average concentration; here the plume is defined as downwind concentrations that are one standard deviation above the mean. $\sigma_{\chi_{C,u}}$ and $\sigma_{\chi_{C,d}}$ are upwind and downwind concentration standard deviations. Note that while the first ratio in Eq. (2.19) represents the magnitude of upwind to source emissions, the second ratio distinguishes upwind plumes from background concentrations: in the absence of upwind plumes, $\sigma_{\chi_{C,u}}$ approaches zero due to the uniform nature of background concentrations. Parameter ϕ can therefore provide a quantitative indication of the level of confidence in the retrieved emissions based on the mass-balance approach, due to this factor. Table 2.3 lists ϕ values in % for the nine studied cases, determined from our GEM-MACH-simulated fields.

For an isolated source with relatively low upwind emissions (which in turn depends on the mean wind direction and the location of the nearby sources) ϕ would be small and the application of the mass-balance technique would be expected to yield more accurate results. This was true for cases 3 and 4 with $\phi < 0.2\%$ (see Fig. 2.2d for case 4). JOSM 2013 cases 3 and 4 were conducted over CNRL and Syncrude oil sands facilities on 20 August and 3 September 2013. For the period of these two flight cases, wind directions were also consistently from the south-west, a condition that was accurately simulated in our GEM-MACH model runs for the same period. CNRL and Syncrude are the two facilities located on the western side of the Athabasca oil sands complex, with no facilities to their west and therefore no upwind sources of emissions when wind direction is from the west (Fig. 2.1b). These conditions resulted in low $E_{C,S}$ rates of 2% and -3% for cases 3 and 4. For these two cases, TR estimates were within 4%–6% of the MIEs. Our TR estimates of SO₂ emissions for CNRL were in strong agreement with TERRA SO₂ emission rate estimates by Gordon et al. (2015) using aircraft-measured data during JOSM 2013 campaign, where estimates were 4.8% higher than minute-by-minute emissions reported by CNRL. Our analysis indicates that about 3% of this over-estimation could be due to release of the stored mass during the flight on 20 August (case 4).

Small ϕ values represent favourable conditions for applying the mass-balance technique for emission rate retrievals. For seven out of nine studied cases, ϕ was less than 10% with NRMS errors between 4% and 28% for TR estimates. The opposite was true for rejected cases 5 and 8 with ϕ scores 11% and 13%, with NRMS errors 39% and 166% for TR estimates. This under performance was a combined result of upwind emissions which were higher than the source emissions, consistent converging wind fields (slower wind speeds within the flux box relative to the upwind wall throughout the flight time), and plume vertical motion due to change in atmospheric stability ($\Delta_t Ri \neq 0$). For rejected case 5, there was an additional contribution from variations in the direction of plume transport ($\Delta_t \alpha = 46^\circ$, sampling time average). Rejected case 8 was conducted around the Suncor facility on 28 August 2013, during which the mean wind direction was from the west; as a result, large upwind SO₂ plumes originating from the Syncrude facility, located on the western side of Suncor (see Fig. 2.1b), continuously entered the flux box during the period of this case. SO₂ mass was accumulated within the volume of the box resulting in a flight/sampling time average storage rate of 156% of the total retrieved emission rates (see Fig. 2.4). Note that

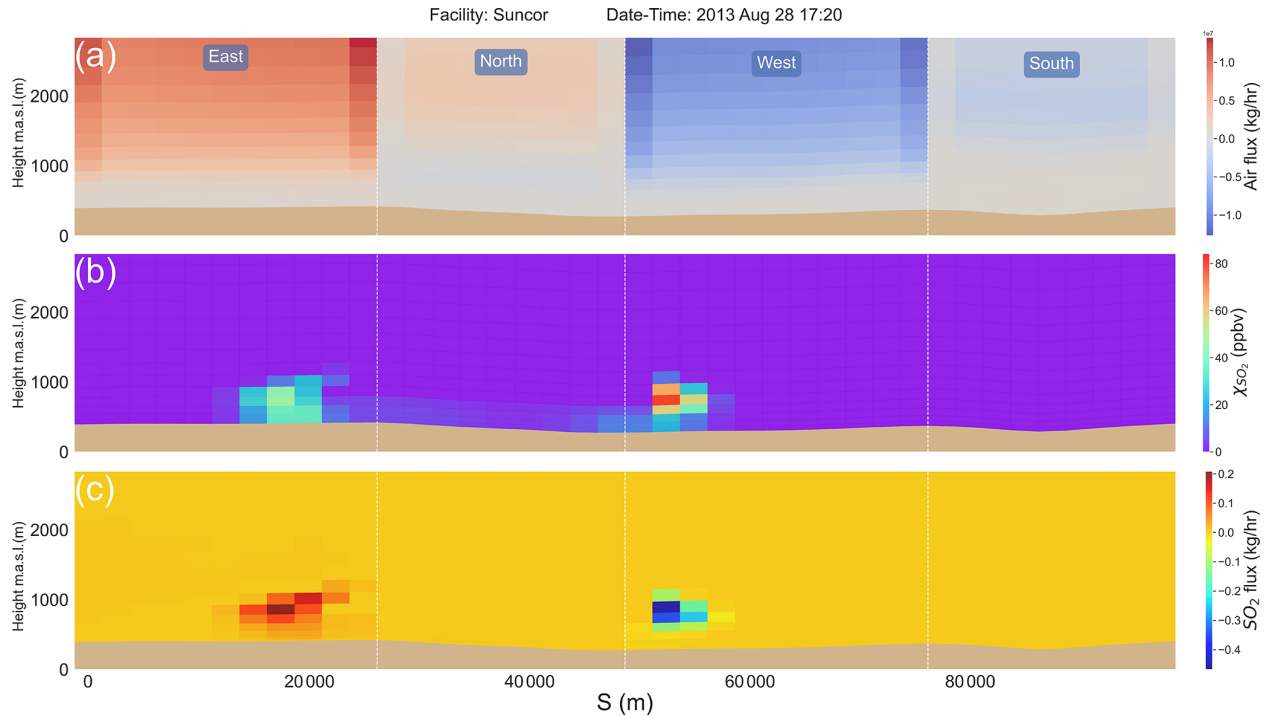


Figure 2.11: Case 8 screens around Suncor: (a) air flux screen, (b) SO_2 mixing ratio screen and (c) SO_2 mass flux screen. Note the strong negative flux (inflow) on the western wall and the weaker positive flux (outflow) on the eastern wall resulting in (temporal avg.) $\phi = 13\%$.

a large upwind source was a condition previously identified for rejecting this flight from emissions calculations (which is why emissions from this flight were not estimated in previous work). The ϕ predictor introduced here allows this effect to be quantified (from model forecasts used to plan aircraft flights, and the analysis of aircraft measured pollutant data time series). Figure 2.11 shows (a) air flux, (b) SO_2 mixing ratio (χ_{SO_2}) and (c) SO_2 mass flux screens around Suncor at the start of rejected case 8; note the strong negative (inwards) flux on the western wall (originating from upwind emissions from the Syncrude facility) and the weaker positive (outwards) flux on the eastern wall (Fig. 2.11). By accounting for the storage rate ($E_{\text{C},\text{S}}^{\text{tr}*}$), using model 3D data from time-consecutive screens, NRMS error was reduced by 52 % of MIEs in TR^* estimates for this case. Direct retrieval (DR) of the Suncor emission rates for the period of rejected case 8, where storage was estimated directly from the model 4D data (volumetric time series), was within 23 % of MIEs. $\phi = 13\%$ for rejected case 8 implies unsuitable conditions ($> 30\%$ uncertainty) for the application of the mass-balance technique, as the TR method failed to retrieve facility emissions ($E_{\text{C}} < 0$); the ϕ parameter allows a quantification of the conditions for rejection of flight data for emission rates retrieval purposes.

Our results indicate a strong positive correlation between ϕ and the NRMS error (strongest among the three forecast parameters), independent of the method. Figure 2.12 shows this correlation for the three retrieval methods, note that rejected case 8 was excluded from correlation analysis as an outlier (extreme storage event) since its inclusion would dominate the correlation and

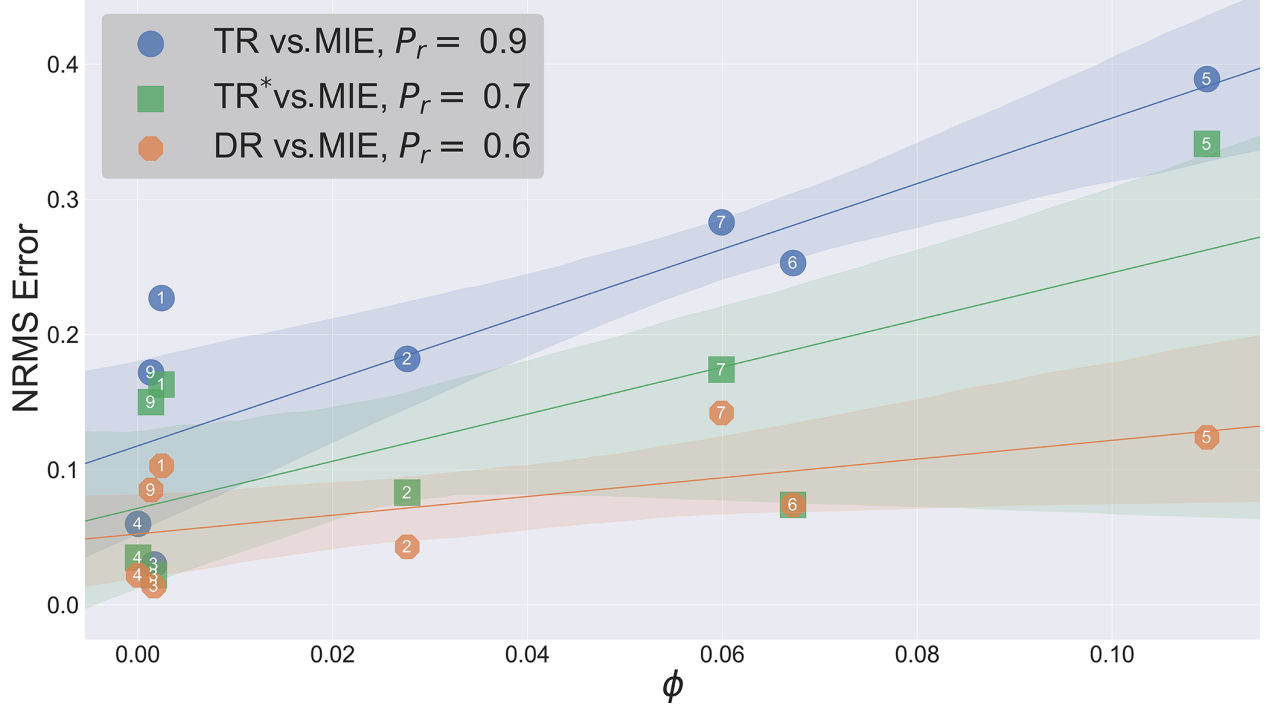


Figure 2.12: Correlations between the NRMS error and the forecast parameter ϕ , for DR, TR and TR* estimates. The shaded areas indicate the 90 % confidence region. Case ID is indicated for each data point. Pearson correlation coefficients (P_r) for the three methods are also noted on the graph. Results indicate a strong correlation between NRMS error and the ϕ parameter.

obscure the information from the other data points. Pearson correlation coefficients (P_r) for DR, TR and TR* were calculated as 0.6, 0.9 and 0.7, respectively. That is, a significant proportion of the variability in the emissions retrieval NRMS error can be accounted for by each of the metrics proposed here. In aircraft-based retrievals, analysis of ϕ can provide insights into the expected reliability of the retrieved emissions and the probability of storage-release events. Pre-flight model forecasts of ϕ can be used for the same purpose. For ϕ scores greater than 10 %, the uncertainty in the retrieved emission rates due to storage release is over 30 %. For such conditions, it is advised that the storage rate ($E_{C,S}$) be estimated from time-consecutive screens and/or vertical profiles of relevant fields. This can be achieved by conducting multiple box flights and/or downwind vertical profiling over the sampling time in order to estimate $E_{C,S}^{tr*}$ for improved top-down emission rate retrievals (TR*; see Sects. 2.2.2 and 2.3.3).

2.4.4 Summary of predictive methods

The three forecast parameters $\Delta_t Ri$, $\Delta_t \alpha$ and ϕ as discussed above provide a new approach for analyzing the uncertainty levels associated with potential storage-release events during aircraft-based emission rate retrieval flight time, either using a priori meteorological forecasts to provide guidance for emissions retrieval flight decision making or in post-processing of already completed flights. Table 2.2 summarizes the correlations between these parameters and the NRMS error in

estimates by the TR method in terms of P_r correlation coefficients and the slopes (m) of the best fit lines (least squares) in Figs. 2.8, 2.10 and 2.12. The last row in Table 2.2, demonstrates the correlation between the changes in the two meteorological forecast parameters $\Delta_t Ri$ and $\Delta_t \alpha$ and change in NRMS error levels; accordingly, net changes of 0.57 h^{-1} in $|\Delta_t Ri|$ and 60° h^{-1} in $\Delta_t \alpha$ during the flight/sampling time may result in 30 % (or more when both contribute) uncertainty in the estimates, the maximum estimated uncertainty (30 %) for TERRA SO_2 emission rate retrievals in Gordon et al. (2015). The combined effects of $|\Delta_t Ri| = 0.62 \text{ h}^{-1}$ and $\Delta_t \alpha = 46^\circ \text{ h}^{-1}$ during the $\sim 2 \text{ h}$ flight time of case 5, contributed to a net error of 39 % in TR estimates for this case. Further, for any forecasted/estimated rate of change in meteorology in terms of $\Delta_t Ri$ and $\Delta_t \alpha$, shorter sampling intervals (e.g., two flights in tandem, vertical profiling information) may result in more confidence in the estimated emission rates; small values of $\Delta_t Ri$ and $\Delta_t \alpha$ correspond to steady-state conditions, whereas higher values correspond to increased uncertainty in the estimates due to storage-release events. The NRMS errors in TR estimates for cases with $\phi < 10 \%$ (seven out of nine) were less than 30 %. $\phi > 10 \%$ correspond to over 30 % uncertainty in the estimates, as for rejected cases 5 and 8. Note that cases 5 and 8 were excluded (rejected) from TERRA retrievals in previous work due to unfavorable conditions for the application of TERRA. Here, using GEM-MACH-predicted fields, we were able to quantify the impact of changing meteorology and/or high upwind emissions in mass-balance emission rate retrieval for these two cases in terms of the three forecast parameters $\Delta_t Ri$, $\Delta_t \alpha$ and ϕ as described above.

Table 2.2: Summary of correlations between NRMS error in TR estimates and the three forecast parameters for storage-release events, in terms of the Pearson correlation coefficient (P_r), and fit line (least squares) slopes (m) in Figs. 2.8, 2.10 and 2.12. The last row indicates the amount of change in forecast parameters that would result in $\geq 30 \%$ error in estimates. $\phi < 10 \%$ corresponds to $< 30 \%$ error.

Parameter	$ \Delta_t Ri $	$\Delta_t \alpha_i$	ϕ
P_r correlation	0.7	0.7	0.9
$m = \frac{\Delta(\text{NRMS})}{\Delta(\text{Param.})}$	0.525 [h]	0.005 [$^\circ \text{ h}^{-1}$]	2.4
$\frac{30\%}{m}$	0.571 [h^{-1}]	60° [h^{-1}]	0.125

We have examined storage and release events in the specific context of “box” flights around an emitting facility, however note that other strategies have been put forward in the literature (e.g., near-field downwind transects Peischl et al., 2010). We note: (1) the heterogeneous nature of meteorology even at the scales employed for the box flights here suggests that the impact of storage and release will likely be greater as the distance between screen(s) and emission source(s) increases (for example, see Appendix B, Fig. B.1, for case 8, where the contribution of the storage term increases as a function of downwind distance): short distance “box” flights such as examined here

would reduce the storage and release impact; (2) our $\Delta_t Ri$ and $\Delta_t \alpha$ parameters may nevertheless be applied to “single-leg” downwind transects retrievals, and repeat flights (e.g., by a second aircraft following behind the first aircraft) may be one way of reducing storage and release uncertainties in these flights; (3) our ϕ parameter relies on the availability of upwind information and hence is not applicable for “single-leg” (e.g., downwind transects) retrievals. (4) Our ϕ parameter provided the strongest correlation between storage and release and retrieval outcomes and therefore, where possible, should be estimated from additional upwind observations. These findings suggest that, while some aspects of storage and release can be identified with “single-leg” downwind flights, the impacts of storage and release for these flight patterns may be higher than for “box” flight patterns such as studied here.

Table 2.3 provides a summary of all the meteorological conditions and emission scenarios in our GEM-MACH model simulations for the period of the nine studied JOSM 2013 cases. The three forecast parameters are also provided in the table: the average rate of change in atmospheric stability represented by $\Delta_t Ri$, the average rate of change in wind direction at plume height (change in direction of the transport) represented by $\Delta_t \alpha$ and the weighted upwind-to-downwind concentration ratio parameter (ϕ). Estimates by the three described methods (DR, TR and TR*) are compared to MIEs, and the corresponding normalized (to MIE) mean (NM) and root-mean-square (NRMS) errors are listed for each of the studied cases.

2.5 Conclusions

We have carried out a series of retrospective air-quality simulations employing Environment and Climate Change Canada’s (ECCC) air-quality model, Global Environmental Multiscale – Modelling Air-quality and Chemistry (GEM-MACH), for the period of the JOSM 2013 campaign over the Athabasca oil sands region, with the primary objective of evaluating aircraft-based mass balance emission rate retrievals such as applied in TERRA (i.e., the Top-down Emission Rate Retrieval Algorithm). We considered the simulations of nine JOSM 2013 emission estimation flight cases, over three oil sands facilities with high SO_2 emissions (CNRL, Syncrude and Suncor), and used GEM-MACH model 4D output data as a surrogate for aircraft measurements in emission rate estimations to evaluate the application of the mass-balance approach in top-down methodologies. Although the emphasis was on SO_2 emissions, the explored concepts and the three forecast parameters developed here are also applicable to species with high background levels (e.g., CH_4 , CO). Emissions retrievals were made by analyzing GEM-MACH model output data through three different approaches: (a) direct retrieval (DR) of the model input emissions (MIEs) by analyzing GEM-MACH model 4D data (volumetric time series) through a comprehensive analysis of all of the processes contributing to the change in mass within the control volume (flux box) including vertical and horizontal fluxes through box top and lateral walls, surface deposition and rate of storage of mass within the volume of the box, (b) TERRA retrieval (TR) approximations by limiting our analysis to the model data on 2D screens comprised of lateral walls of the flux box surrounding

the emission source, mimicking aircraft-based retrievals during an emission estimation (box) flight, and (c) improved TERRA retrievals (TR*) through estimation of the rate of change in the stored mass within the volume of the flux box by analyzing the data on time-consecutive 2D screens at each model time step (3D). See Table 2.1 for a summary of the three methods.

Results were compared to the MIEs and the performance of the three methods was analyzed in terms of normalized (to MIE) root-mean-square (NRMS) error. The DR method estimates (E_C) were within 1 %–23 % of MIE for all nine studied cases. TR and TR* (E_C^{tr} and $E_C^{\text{tr}*}$) estimates for seven out of nine cases were within 4 %–28 % and 2 %–17 % of MIEs, respectively. This range of emissions retrieval error was similar to the range (± 30 %) estimated from aircraft observations in previous work (Gordon et al., 2015), with the remaining 2 flights (rejected cases 5 and 8) excluded in previous work due to unsuitable meteorological conditions and knowledge of upwind emissions. Estimates for the JOSM case on 26 August 2013 around the CNRL facility (rejected case 5) were impacted by change in atmospheric stability at rate $|\Delta_t Ri| = 0.62 \text{ h}^{-1}$ and shift in wind direction at plume center-point (maximum concentration) height at rate $\Delta_t \alpha = 46^\circ \text{ h}^{-1}$ during the $\sim 2 \text{ h}$ flight time; NRMS errors for this case in estimates by the TR and TR* methods were 39 % and 34 %, respectively. Estimates for the case on 28 August around the Suncor facility (rejected case 8) were impacted by high emissions (relative to source emission rates) from the upwind oil sands facility Syncrude (avg. $\phi = 13$ %) and change in atmospheric stability at rate $|\Delta_t Ri| = 0.34 \text{ h}^{-1}$ during the $\sim 2.5 \text{ h}$ flight/sampling time; NRMS errors for this case were 166 % and 114 % by the TR and TR* methods, respectively. The under-performance in rejected cases 5 and 8 estimates were the consequence of several different factors and conditions during the flight time; we refer to the collective impact of these conditions as storage-and-release events. The nine studied cases were affected by storage-release events at various degrees, ranging from -29 % to 156 % with a median of 13 %. This range for the seven cases that are included in past work with TERRA applications is -29 % to 20 % with a 2 % median.

Storage-release events occur when temporal and spatial variations in meteorological conditions and/or source emission rate result in a temporary imbalance between the addition of mass to the volume of the box through source emissions and removal of mass by vertical and horizontal fluxes through box top and lateral walls and the deposition to ground surface. The transient storage of the emitted mass within box volume and its later release contribute to emission rate over- and under-estimations based on the mass-balance technique. We introduced a correction term, $E_{C,S}$, in the mass-balance equation to represent the rate of storage-release. Utilizing GEM-MACH model output data, we investigated the contribution of this term to the total integrated retrieved emission rates. The storage rate term ($E_{C,S}$) contributed over 20 % of the total retrieved emissions for five out of nine cases, signifying the impact of storage-release events in emission rate retrievals based on the mass-balance approach. The corresponding uncertainties in estimated emission rates were quantified for each of the nine studied cases, and conditions giving rise to storage-release events were extensively explored.

We have introduced an approach for estimating the storage term as the TR* method, based

on estimations of the rate of change in the stored mass within the box volume from data on time-consecutive 2D screens and/or vertical profiles around and downwind of the same emission source (3D dataset). We have shown that by estimating the storage term ($E_{C,S}^*$) in this way, estimates can be improved by 2%–52% of MIE. Three forecast parameters were introduced for predicting/identifying the probability of storage-release events: variations in atmospheric dynamic stability represented by change in gradient Richardson number ($\Delta_t Ri$), shift in the direction of transport represented by variations in wind direction at plume height ($\Delta_t \alpha$), and the weighted upwind to downwind concentration ratio (ϕ). We have demonstrated the strong correlation between these three parameters and the NRMS error in mass-balance estimates and discussed the potential of utilizing the three forecast parameters in analysis of the uncertainty levels in aircraft-based top-down emission rate retrieval methodologies. Model-based forecasts of these parameters can also provide practical advice for the planning of future airborne measurement campaigns.

The investigation carried out here has resulted in the creation of quantitative measures for the extent to which storage and release events may impact aircraft emissions retrieval accuracy. Retrieval uncertainties, using regional air-quality model output, were shown to be of similar magnitude to previously published values, with the exception of cases where the underlying assumption of time-invariant meteorological conditions was not valid and/or where significant upwind emissions impacted the estimates. We have also devised a methodology to reduce the impact of this form of emissions retrieval error by estimating the storage rate ($E_{C,S}^{tr*}$) through (a) the use of repeat flights, around the same facility, either with a single or multiple aircraft(s), or (b) aircraft-based remote vertical profiling (e.g., lidar measurements) of relevant fields during the sampling period.

Table 2.3: A summary of the meteorological conditions and source emission scenarios by the three oil sands facilities for the period of the nine JOSM 2013 box flight cases in our GEM-MACH model simulations. Flight date-time and duration is provided for each case, LT stands for local time (UT−6 h). The flight time average $\overline{R_i}$, wind speed and direction at plume height are shown. The three forecast parameters are also provided: $\Delta_t R_i$, $\Delta_t \alpha$ and ϕ . DR, TR and TR* method estimates are compared to model input emissions (MIEs). Storage (S) and/or release (R) events, their order of occurrence (for the cases where both were present) and their relative magnitude (the bigger contribution is represented by boldface letters) are given. The flight time average storage rate as a fraction (%) of the total retrieved emission rates are also provided for each case. The performance of the three retrieval methods against MIE, are shown in terms of NM and NRMS errors.

Facility	Syncrude			CNRL			Suncor		
Case ID	1	2	3	4	5	6	7	8	9
Date and start time	08/14 10:12	08/24 11:50	09/3 14:58	08/20 10:30	08/26 13:43	09/2 11:43	08/22 11:45	08/28 11:18	08/29 15:01
Duration	01:18	02:06	02:07	02:10	01:52	01:45	02:58	02:30	02:13
Atm. cond.	st	unst	unst	unst	st	unst	unst	unst	unst
$\overline{R_i}$	0.26	−0.03	−0.66	−0.14	1.01	−0.26	−0.19	−0.44	−0.58
wind dir. & spd. (m/s)	200° 4.3	264° 2.0	215° 4.9	259° 10.9	243° 0.8	100° 6.9	246° 11.4	297° 3.2	55° 2.0
$\Delta_t R_i$ (h ^{−1})	−0.38	−0.13	0.19	−0.25	0.62	−0.28	−0.26	−0.34	0.15
$\Delta_t \alpha$ (h ^{−1})	3.3°	6.5°	1.3°	1.8°	45.8°	33.9°	2.4°	3.3°	11.8°
ϕ (%)	0.244	2.766	0.167	0.004	10.97	6.725	5.992	13.04	0.137
Model input emiss., MIE	3423 kg h ^{−1}	4050 kg h ^{−1}	6923 kg h ^{−1}	2979 kg h ^{−1}	120 kg h ^{−1}	83 kg h ^{−1}	1467 kg h ^{−1}	1273 kg h ^{−1}	1013 kg h ^{−1}
Storage and /or release contribution	S − R 17%	S 20%	R − S 2%	R − S −3%	S 43%	R −29%	R − S 13%	S 156%	R −27%
NM error:									
DR (E_C)	−0.07	0.04	−0.00	−0.01	0.11	−0.00	−0.14	0.22	−0.08
TR (E_C^{tr})	−0.22	−0.16	−0.02	0.02	−0.36	0.24	−0.25	−1.66	0.17
TR* ($E_C^{\text{tr}*}$)	−0.11	−0.07	−0.01	−0.00	−0.26	0.05	−0.14	−0.96	0.11
NRMS error:									
DR (E_C)	0.10	0.04	0.01	0.02	0.12	0.07	0.14	0.23	0.08
TR (E_C^{tr})	0.23	0.18	0.04	0.06	0.39	0.25	0.28	1.66	0.17
TR* ($E_C^{\text{tr}*}$)	0.16	0.08	0.02	0.03	0.34	0.07	0.17	1.14	0.15

Chapter 3

Tracer Transport Modelling at Super-resolution with WRF-ARW

3.1 Introduction

In the context of airborne top-down source emission rate estimations, numerical model simulations can be employed in various approaches. These include off-line applications, where a meteorological model (e.g., Weather Research Forecasting - WRF) is used to replicate conditions during airborne and/or ground-based observations. The model generated meteorological fields are often used to drive a separate Lagrangian tracer dispersion model (e.g., HYSPLIT) either forward in time to simulate tracer concentrations at observation times and locations, or for inverse method analysis (Cui et al., 2015; Kia et al., 2022; Lauvaux et al., 2016). Previous airborne studies have also used model generated wind fields and aircraft measured concentrations for flux calculations and mass-balance analysis (Karion et al., 2015). The emission and transport of the passive tracers can be simulated in-line with meteorological fields within the same modelling platform such as the Eulerian WRF model, for source emission characterizations at the scale of observations (Ahmadov et al., 2015; Barkley et al., 2017; Nahian et al., 2020). For these applications, model generated fields are analyzed as complementary information for characterizing emissions based on airborne observational data. For instance, Ražnjević et al. (2022) have employed large-eddy-simulation (LES) modelling driven by reanalysis data for interpreting field observations of CH₄. Further, model simulations of tracer transport and dispersion have been previously used for assessing the uncertainties/errors in top-down retrievals and optimizing the observational approach (Angevine et al., 2020; Conley et al., 2017; Fathi et al., 2021). Numerical model simulations can also be used for simulating ground-based and/or airborne observations, where model generated fields are used as a proxy for measurement data (virtual sampling). For a robust model-based study of observational methods, model resolutions must be chosen to resolve the time and length scales of the measurements. For example, Gasch et al. (2020) simulated aircraft-based Doppler Lidar measurements of wind fields through LES modelling at 10 m resolution to investigate airborne

lidar measurements for a lidar range length of 72 m.

The study in Chapter 2, using a regional chemical transport model with physical and chemical process representations (GEM-MACH), was successful in evaluating the application of the mass-balance technique in top-down retrievals using model simulated fields as a proxy for the real world environmental fields (Fathi et al., 2021). However, the (relatively) coarse resolution (2.5 km, 2 min) of the model was insufficient for the investigation of aircraft-based retrievals through virtual airborne samplings within the model simulated 4D fields. In airborne campaigns, environmental observations (e.g. wind, temperature, tracer concentrations) are made while flying downwind or around emission sources. This data are then processed through various retrieval algorithms to estimate source emission rates (Gordon et al., 2015; Peischl et al., 2010; Ryoo et al., 2019). An underlying assumption common among retrieval algorithms is the steady-state conditions during the sampling time of several hours (Alfieri et al., 2010). Data collection during aircraft-based in-situ measurements are made through 3D space and over time, thus: (a) any point in space along the flight path is visited only once, and (b) spatially adjacent data points are collected at different (consecutive) times. By assuming stationarity (e.g., wind, emissions), the observational data are assumed to be representative of the average conditions during the sampling time. However, time-varying conditions (whether due to turbulence or weather trends) can reduce the representativeness of the sparsely collected environmental data. To study these effects through model simulations, the model resolutions should be chosen to resolve dynamical processes (turbulence) at the spatio-temporal scales at which aircraft in-situ measurements are made. For instance, to simulate (and evaluate) in-situ measurements at a flying/sampling speed of about 100 m/s (e.g., Conley et al., 2017; Gordon et al., 2015), model should be able to simulate (and output) atmospheric fields at length and time scales of $\Delta x \leq 100$ m and $\Delta t \leq 1$ s. Recent real-case LES-modelling studies have commonly referred to such resolutions ($\Delta x \leq 250$ m) as "super-resolution" (e.g., Onishi et al., 2019; Watson et al., 2020; Wu et al., 2021), herein we use the same terminology to describe our WRF model simulations.

The modelling requirements described above, motivated the development of super-resolution micro and LES scale atmospheric tracer transport model simulations, fine enough to resolve smaller-scale flow details and the effects of turbulence and changing stability in atmospheric mixing of tracer concentrations downwind of point and area sources of emission, enabling:

1. thorough dynamical evaluation of the application of the divergence theorem and the mass-balance technique in inferring source emission rates,
2. investigating the effects of flight pattern in aircraft-based top-down retrievals, utilizing model 4D output database,
3. exploring improved sampling approach through optimized flight design and multi-platform (in-situ, remote) measurements,
4. exploring improved data analysis, post-processing, and interpolation/extrapolation methods needed for flux calculations based on airborne observations.

In this chapter, we present a proof of concept for performing super-resolution model simulations of atmospheric tracer transport and dispersion with WRF-ARW. The concepts that are explored here include (a) the realistic modelling of the atmospheric boundary layer at large-eddy-simulation scale over complex terrain, (b) the mass-conserved modelling of atmospheric dispersion and transport of passive tracers under the conditions described in (a), and (c) generating modelling products at spatio-temporal scale of airborne observations (aircraft-based in-situ and remote measurements), useful for evaluating the observational methods and providing recommendations for future studies. We evaluate the performance of our model simulations against historical observational data from ground-based monitoring stations and aircraft-based observations from the 2013 JOSM airborne campaign. We further assess the performance of our simulations in terms of global (over the entire modelling domain) and local (sub-domain) mass-conservation, by conducting four-dimensional mass-balance analysis.

We explore three different cases (dates and times) during August and September of 2013 over Canadian oil sands (Athabasca, Alberta). We use reanalysis data as initial and boundary conditions for our case studies. To achieve the desired micro and LES scale resolutions, we perform multi-domain nested simulations with LES parameterization for the finest domains. Further, we modify the WRF source code (dynamical solver) to simulate the release of passive tracers from points and area sources within the finest model domain. The novel modelling approach with WRF presented in this chapter is comprised of: (1) dynamical down-scaling of reanalysis data from synoptic to LES resolution, (2) super-resolution model simulations through horizontal nesting and vertical grid refining, (3) LES sub-grid parameterization, and (4) passive tracer transport and dispersion simulations. The model output data from the super-resolution simulations in this chapter are used for evaluating the accuracy of aircraft-based emission rate retrieval methods in Chapter 4.

3.2 Methods

3.2.1 Case Studies

For this work we chose our case studies from the times and locations of box flights during the JOSM 2013 campaign over the Athabasca oil sand region (Alberta, Canada). This choice was made to enable qualitative comparisons to observations. We considered the geographical location of an oil sands facility, Canadian Natural Resources Ltd. (CNRL). We configured our WRF model domain centred over the CNRL facility. For our WRF simulations, we considered model simulation times overlapping those of three JOSM 2013 box flights over CNRL (see Table 3.1). Box flight refers to closed shape (e.g. rectangular, cylindrical) flight paths around the target emission source, where aircraft-based measurements are used to estimate source emission rates using the mass-balance technique. Note that case 2 on 26 August 2013 was a “rejected” case in the actual campaign analysis due to unsuitable atmospheric conditions for aircraft-based retrievals. Such conditions are discussed here by using the example of case 2, utilizing model generated atmospheric fields.

For each of these three cases, eleven tracer emission scenarios/sources were considered. Table

Table 3.1: Three case studies during late August and early September of 2013 over oil sands facility CNRL. Times and locations for the case studies were chosen from three JOSM 2013 box flights.

	Case 1	Case 2	Case 3
Date	20 Aug	26 Aug	2 Sep
Start Time (Local Time)	10:30	13:43	11:43
Start Time (UT)	16:30	19:43	17:43
Duration (hh:mm)	02:10	01:52	01:45
Model Simulation Time	15UT - 19UT	18UT - 21UT	15UT - 19UT

3.2 provides the spatial details for the different emission sources, including seven elevated point sources (representing stack emissions), two small area sources (representing surface mines), a large area source (representing the tailing pond west of CNRL) and a long multi-section line source (over the approximate extent of the Horizon Highway south of CNRL). Table 3.2 lists geographical coordinates and tracer release heights (stack top height) for all of the sources. Horizontal dimensions are also provided (in brackets) for the line and area sources. The horizontal dimensions for each of the point sources are equal to those of one grid cell in the finest model domain. Coordinates and heights for CNRL 1-4 exactly correspond to actual (real world) stacks in the CNRL facility. The hypothetical source CNRL 0 is co-located with CNRL 1 and 4, with stack-top/release height at over 4 times higher than the tallest facility stack (CNRL 1), simulating the initial (assumed instantaneous) plume rise due to buoyancy. Figure 3.1 shows a map of the region with the location and spatial extent of case study emission sources marked/labelled in red. The large area rectangular surface source represents the tailing pond on the west of the CNRL complex. The multi-section line surface source represents the Horizon Highway south of CNRL. Two small area sources labelled as Mine 1 and Mine 2 represent emissions from surface mine excavation sites within the CNRL complex. Figure 3.1 also shows the locations for two hypothetical point (stack) sources CNRL s (south) and CNRL w (west). During two of our case studies on 20 August 2013 (case 1) and 2 September 2013 (case 3), mean wind was from west and south-west, placing these two hypothetical stacks upwind of the CNRL facility.

3.2.2 Model and Technical Setup

For this work, the Weather Research and Forecasting (WRF - Skamarock, 2008) model with the ARW dynamical core was utilized. WRF-ARW provides a multi-scale simulation framework suitable for efficient parallel computing with a vast range of physical parameterizations adaptable for different scales and dynamical processes. For this project, the High Performance Computing (HPC) resources of Compute Canada (CC) were used. WRF model simulation were run on 20 CPUs with 20 nodes each (20×20) and 128GB RAM memory on CC’s super computers Graham and Cedar. To further speed up the model runs and model output, multiple nodes were used for parallel I/O using the PNETCDF module embedded in the WRF modelling platform, which sped up the output part of the modelling process by 6-10 times compared to the conventional serial I/O. Model simulations

Table 3.2: Eleven tracer emission scenarios including seven point sources representing release from stack tops at various heights, three area sources including a large area tailing pond towards the western side of the facility and two smaller (in area) surface mines, and a line source approximately spanning the extent of the Horizon Highway south of the facility. Note that height/locations for sources with superscript [‡] are hypothetical.

Source ID	Type	Lat.	Lon.	Height agl (m)	Description
CNRL 0 [‡]	Point	57.339	-111.738	483	Stack
CNRL 1	Point	57.339	-111.738	114	Stack
CNRL 2	Point	57.337	-111.740	54	Stack
CNRL 3	Point	57.336	-111.732	30	Stack
CNRL 4	Point	57.339	-111.738	54	Stack
CNRL w [‡]	Point	57.327	-112.014	102	Stack (upwind west)
CNRL s [‡]	Point	57.250	-111.867	102	Stack (upwind south)
HWY	Line	57.258	-111.765 (~ 20 km)	6	Horizon Highway
POND	Area	57.348	-111.918 (~ 50 km ²)	6	Tailing Pond
MINE 1	Area	57.337	-111.834 (550 m $\times$ 550 m)	6	Surface Mine 1
MINE 1	Area	57.325	-111.820 (350 m $\times$ 550 m)	6	Surface Mine 2

generated 12GB of output data for each model second. Super-resolution WRF simulations for our case studies generated over 120TB of model output data, which were central to the purposes of this research work and are analyzed in chapters 3 and 4.

Mesh refinement and increased resolution can be achieved in WRF through series or concurrent grid nesting in the horizontal dimensions. With concurrent grid nesting, multiple computational domains with increasing resolution can be integrated simultaneously; where the coarse “parent” domain’s output is interpolated to provide initial and lateral boundary conditions for the fine “child” domain, a process referred to as one-way nesting. Two-way nesting is achieved when information from the “child” domain is aggregated to write the overlapping regions of the “parent” domain. At high resolutions (< 3 km), mesh refining in WRF via grid nesting only in the horizontal dimensions limits the control over the grid aspect ratio which can lead to poor grid quality and numerical errors. It has been shown that grid quality affects the accuracy of numerical solutions (Lee and Tsuei, 1992; You et al., 2006). A procedure permitting vertical nesting for one-way concurrent simulation is developed and described in Daniels et al. (2016), which allows high resolutions in the order of meters while grid quality is maintained. This procedure permits one-way concurrent grid nesting in both the horizontal and vertical and this is herein utilized for WRF simulations with 5 domains (d01 - d05) with increasing resolutions from ~ 31 km to 50 m in the horizontal dimensions and up to near surface vertical resolution of ~ 10 m in the finest domain d05.

The North American Regional Reanalysis (NARR) GRIB data (at 3 hour intervals) from NOAA archives were used, for August and September of 2013 (period of 2013 JOSM campaign) over the Athabasca oil sands region (Alberta, Canada). Realistic WRF-ARW simulations were carried out with concurrent one-way grid nesting in the horizontal dimensions with 5 domains (d01-d05) at a ratio of 1:5 and with mesh grid refinement in the vertical for the two smallest and finest domains

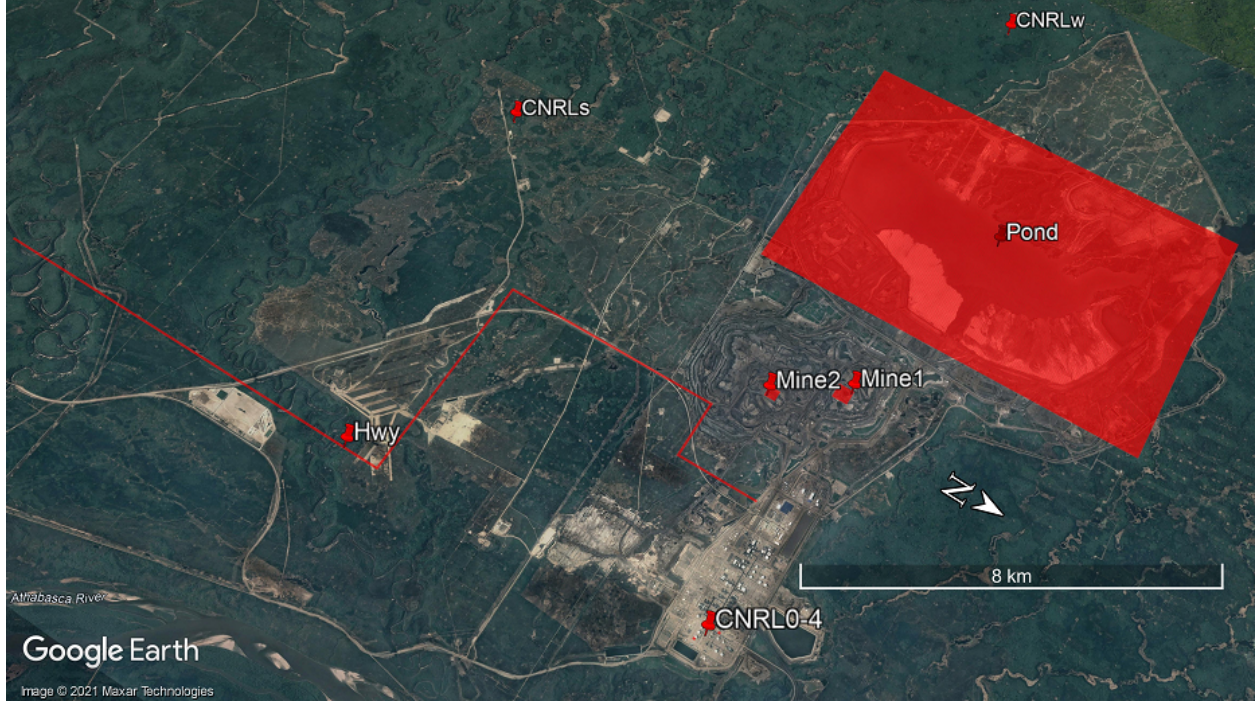


Figure 3.1: ©Google map of the study region: The oil sands facility CNRL with case study emission sources marked on the map including seven point sources, two area sources, a large area surface source (tailing pond) and a multi-segment line source (road/Hwy). Direction of north (N) is shown with a compass arrow.

(d04 and d05) by consecutively increasing the number of vertical levels near the surface. The finest domain (d05) has a horizontal grid size of $\Delta x = 50$ m over the entire domain, $\Delta t = 0.16$ s model simulation time-step and $\Delta z = 11.62$ m for the first 40 full grid levels near the surface; Δx and Δt configurations are set as such to ensure Courant-Friedrichs-Lewy (CFL) stability criterion, $\Delta t < \Delta x / |u_{max}|$ where $|u_{max}|$ is the maximum wind speed in the model (Jacobson, 2005). Table 3.3 provides the details of model grid configurations for the five domains.

Figure 3.2a shows the 5 domains of the model and their relative size. Model domains are centred on the region of Athabasca oil sands with the two finest domains (d04 and d05) centred on the CNRL Horizon facility on the north west quarter of the complex, west of the Athabasca river. Figure 3.2b shows a map of the region with the CNRL facility marked on the map (red star). Boundaries of the two finest domains, d04 and d05, are also overlaid on the map in Figure 3.2b. The relative location of the other oil sands facilities can be seen in the figure. CNRL is at the north west corner of the complex with no facilities to its north and west. The oil sands region is located on the Athabasca river valley with 400-500 m vertical relief within a few tens of kilometres of the facilities (mainly in the west-east direction). This may give rise to complex flow and frequent vertical wind shear in the valley (Gordon et al., 2018). In the following sections we discuss the performance of our super-resolution model simulations against observed meteorology for the same locations and time periods.

Table 3.3: Case study model setup for model simulations with 5 domains with increasing resolution. The first 3 domains have the same coarse vertical grid. Domains d04 and d05 have increasing resolution via vertical grid refinement. The finest domain (d05) has $\Delta z = 11.62$ m for the first 40 levels near the surface. Δx and Δt show the horizontal grid size and the model simulation time-step for each domain, respectively. X and Y indicate domain dimensions. nx , ny and nz are the number of computational grid points in each direction. With model Z_{top} at 15.623 km (15.350 km agl) and P_{top} (pressure at model top level) at 10 kPa.

Domain	Vertical Grid	$\Delta x(\text{m})$	$\Delta t(\text{s})$	$X(\text{m})$	$Y(\text{m})$	nx	ny	nz
d01	Coarse	31250	100	6281250	6281250	201	201	30
d02	Coarse	6250	20	3131250	3131250	501	501	30
d03	Coarse	1250	4	751250	751250	601	601	30
d04	Fine	250	0.8	175250	175250	701	701	48
d05	Fine	50	0.16	50050	50050	1001	1001	82

Figure 3.3 shows the topography and two topography-following model vertical levels (model levels 42 and 55) of domain d05. Domain d05 has a vertical resolution of about $\Delta z = 12$ m for the first 40 levels near the surface and the vertical resolution decreases gradually with altitude up to about 1000 metres above sea level (masl), beyond which it matches the vertical grid resolution of domain d04. Vertical resolution of about 12 m for the bottom 500 m agl (above ground level) is sufficient for investigating and evaluating different methods for extrapolating sampled data below the lowest flight level (typically ~ 150 m) where no aircraft-measurements are made, which is required for flux estimations in top-down emission rate retrieval methods (Gordon et al., 2015).

In order to simulate small-scale atmospheric dynamical processes, the finest two model domains (d04 and d05) were configured with the following Large Eddy Simulation (LES) sub-grid parameterization available in the WRF model (Skamarock, 2008),

1. Diffusion option was set to "Full Diffusion" mode (*diff_opt*= 2) to accurately compute horizontal gradients using full metric terms.
2. "K Option" was set to *km_opt*= 2 to solve a prognostic equation for turbulent kinetic energy (TKE) where the diffusion coefficient K is calculated based on TKE (see Appendix C).
3. 6th Order Horizontal Diffusion set to default option (*diff_6th_opt*= 1).

WRF 3.7+ includes a basic framework for initiating and creating continuous (and variable) emission tracer plumes. To create tracer transport simulations, WRF dynamical solver FORTRAN source codes were modified following Blaylock et al. (2017). Tracer amounts were initiated at several different horizontal locations within the finest domain (d05) after 30 minute model simulation. A meteorological spin-up time of 1 hour was considered through testing with the modelling setup for different initial and boundary conditions according to the following criteria: for the cases we considered, within the first hour of simulation time (a) the model boundary conditions propagated over the entire span of domain d05, (b) model winds (in west-east and south-north directions) assumed continuous profiles both horizontally and between model vertical layers, (c) water vapour

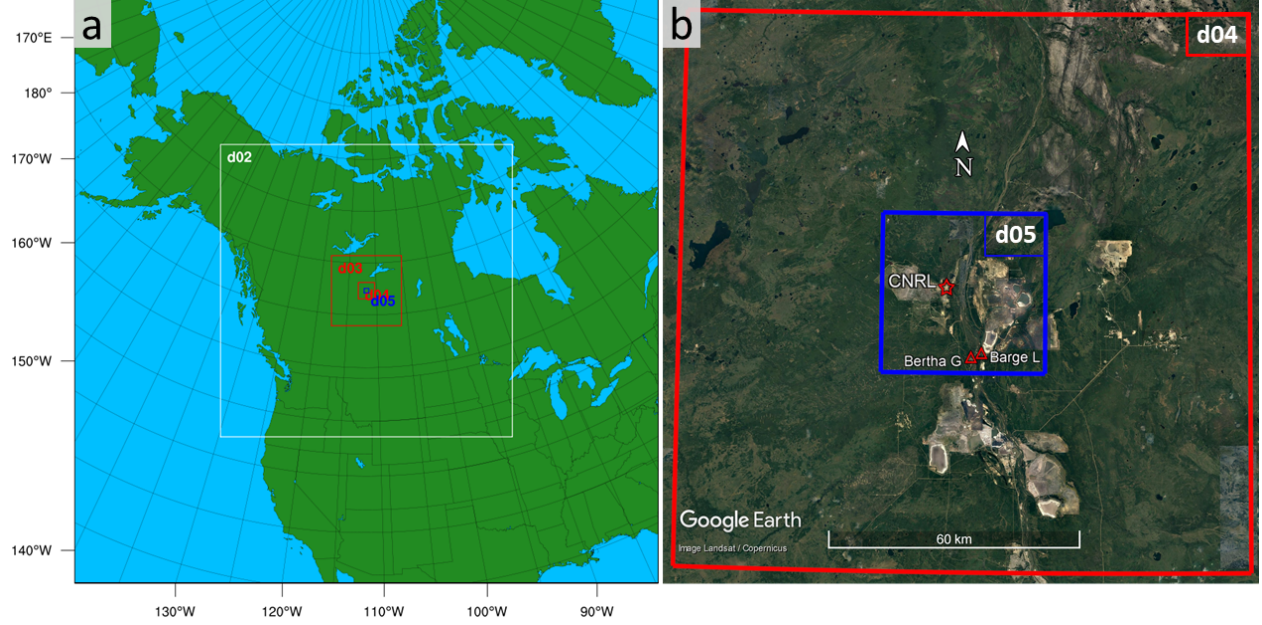


Figure 3.2: (a) WRF model grid (horizontal) configuration with five nests d01 - d05 with increasing resolution (decreasing size) at a ratio of 1 to 5. The finest domain, d05, is centred over the oil sands region. (b) ©Google map of the oil sands region with overlay of model domains d04 and d05. The CNRL facility is marked with a red star. Locations for two WBEA monitoring stations are also shown with red triangles, Bertha Ganter – Fort McKay and Barge Landing monitoring stations.

on model mass-layers (horizontal and vertical mass grid points) assumed continuous profiles within this time period. Tracer release for our considered emission scenarios (see Table 3.2) started after a 30 min spin-up time (half the meteorological spin-up time) and continued for the rest of the simulation period. These included surface emissions on several model grid points at the lowest level (i.e., level 1 at ~ 6 m agl) at various locations, and stack emissions at levels 3, 5, 9, 10 and 40 according to the stack top heights and horizontal locations described in Table 3.2. Note that in WRF-ARW vertical levels (Fig. 3.3) are configured using a terrain-following hydrostatic-pressure coordinate system (Skamarock, 2008) and therefore stack-top heights for our simulations are assigned to pressure levels with heights (in meters) closest to source heights. Depending on pressure changes, the height of pressure levels can vary over time. This variation was determined to be smaller than vertical level thickness for our simulations and therefore its impact on tracer release simulation is considered negligible in this work. All of the tracer emissions were implemented within the boundaries of the CNRL facility and the surrounding region, on the eastern half of the modelling domain. All the emission scenarios involved the emission of passive and non-buoyant tracers with no interactions with meteorology and no defined surface deposition rates. Note that the topography input information (land use indices) used in our WRF simulations were not modified to represent oil sands operations (e.g., tailing pond, excavation sites) and only represent natural features (e.g., river basin, hills). The release, dispersion and transport of tracers from our emission scenarios under different meteorological conditions are discussed in the Results section of this chapter.

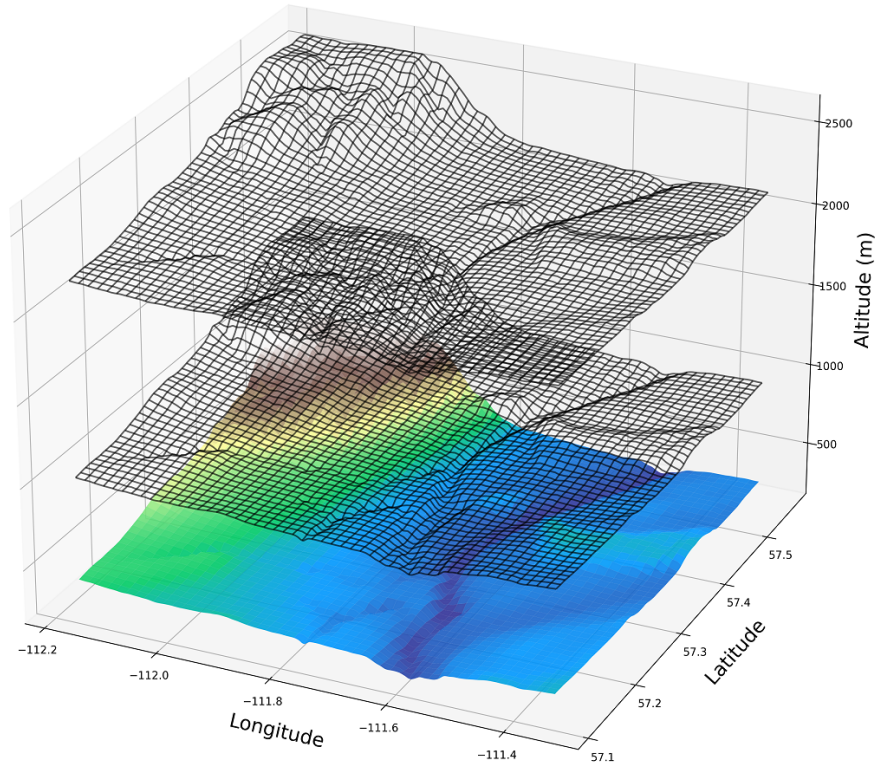


Figure 3.3: Topography of the modelling region for domain d05 (finest domain) is shown, lower elevations such as the Athabasca river valley is shown in dark blue and higher elevations are shown with colours green and brown. Our WRF model simulations use terrain-following model layers to ensure continuity in the modelling layers and proper treatment of the interaction between the atmospheric flow and the topography. Two model layers at levels 42 and 55 are also shown (mesh).

3.2.3 Divergence Theorem and the Mass-balance Technique

For this work, the mass-balance technique is utilized for calculating the net integrated flux out of virtual control volumes (emission box) enclosing the emission sources of interest. The calculation steps in this section follow those in Chapter 2, with slight modifications for use with WRF model output data. For a detailed discussion on the application of the mass-balance and divergence theorem in estimating source emission rates see Chapter 2, Section 2.2.2.

In applying the mass-balance technique to estimate the rate of emissions from sources within a flux box (control volume), the mass flux exiting the box through box lateral and top walls are equated to the emission rate of the tracer within the box. The processes contributing to the change of mass, for a passive tracer, within the control volume can be described with the following expression,

$$S_C = E_C - F_{C,H} - F_{C,V} - F_{C,HT} - F_{C,VT} \quad (3.1)$$

where the storage term S_C represents the change in mass of tracer C within the control volume, E_C represents the tracer emission rate, $F_{C,H}$ and $F_{C,V}$ represent the net horizontal and vertical advective fluxes exiting through lateral and top walls of the flux box, respectively. $F_{C,HT}$ and $F_{C,VT}$ represent horizontal and vertical turbulent fluxes across the box walls, respectively.

The total mass of the tracer within the control volume can be calculated by integrating the mass over the entire volume of the box at each model output timestamp ($\Delta t = 1$ sec),

$$B_{C,Tot}(t) = \iiint \chi_C(t, x, y, z) dx dy dz \quad (3.2)$$

where $\chi_C(t, x, y, z)$ is the tracer concentration at each model grid point. Further the storage term can be calculated by taking the time derivative of $B_{C,Tot}(t)$,

$$S_C(t) = \frac{\partial}{\partial t} B_{C,Tot}(t) \quad (3.3)$$

Horizontal advective flux through the lateral walls of the box can be calculated by extracting tracer concentration and normal wind (positive outwards) along the lateral walls of the box from model output,

$$F_{C,H}(t) = \iint \chi_C(t, s, z) U_{\perp}(t, s, z) ds(x, y) dz \quad (3.4)$$

where $ds(x, y)$ is the path $s(x, y)$ increment along the walls. The normal wind along the box walls is calculated as,

$$U_{\perp}(t, s, z) = \frac{U \frac{ds(x, y)}{dy} - V \frac{ds(x, y)}{dx}}{\sqrt{\frac{ds(x, y)}{dx}^2 + \frac{ds(x, y)}{dy}^2}} \quad (3.5)$$

where U is the west-east wind and V is the south-north wind. $U_{\perp}(t, s, z)$ as calculated above is positive outward for $s(x, y)$ increasing in the counterclockwise orientation. Similarly, the vertical flux through the box top can be calculated as,

$$F_{C,V}(t) = \iint \chi_{C,top}(t, x, y) W_{top}(t, x, y) dx dy \quad (3.6)$$

where $\chi_{C,top}(t, x, y)$ and $W_{top}(t, x, y)$ are tracer concentration and vertical wind speed at box top, respectively. See Appendix D for turbulent flux terms. As we show later, the vertical advective ($F_{C,V}$) and the turbulent fluxes ($F_{C,HT}$, $F_{C,VT}$) have negligible relative contributions to the mass-balance equation, with the horizontal advective flux $F_{C,H}$ being the dominant term removing mass from the box. We collect all the flux terms (advective and turbulent) contributing to the removal of tracer mass from the box into a single flux out term as $F_{C,out} = F_{C,H} + F_{C,V} + F_{C,HT} + F_{C,VT}$.

By rearranging Eq. 3.1, the tracer emission rate can be estimated based on the other terms as,

$$E_C = S_C + F_{C,out} \quad (3.7)$$

By extracting the required fields from the model output 4D database, Equation 3.7 can be utilized to determine the source emission rate based on the mass-balance equation which can then be compared to the known input emission rate. Following the above described calculation process, source emission rates can be estimated at each model output time-step and compared to the model input emissions to evaluate model performance in terms of local mass-conservation and mass-flux consistency. At the same time, the application of the mass-balance technique for retrieving source emission rates can be evaluated as all the terms in Eqs. 3.1 or 3.7 can be independently inferred from the model output data.

3.3 Results and Discussions

Model simulations were carried out for the period between 15UT to 19UT for cases 1 and 3 on 20 August and 2 September 2013, respectively. The simulation period for case 2 on 26 August 2013 was between 18UT to 21UT (see Table 3.1). NARR reanalysis data at 31.25 km resolution was used as initial and boundary conditions (IC and BC) for the coarsest domain d01 with the same resolution. Through model nesting at an increasing resolution (and decreasing size) ratio of 1:5, the input reanalysis data were down-scaled to consecutively higher resolutions all the way to 50 m in the fines domain d05. Each parent domain provided initial and boundary conditions for their nested child domain: $d01 \Rightarrow d02 \Rightarrow d03 \Rightarrow d04 \Rightarrow d05$. Note that feedback (two-way nesting) between parent and nested domain was turned off to allow for vertical grid refining for domains d04 and d05. Output frequency was set to 3 hours for domains d01 – d03, 100 seconds for d04, and 1 second for d05.

3.3.1 Meteorological Evaluation

Figure 3.4 compares case 1 meteorological output fields from the finest three domains d03, d04 and d05. Note that output from d04 and d05 are overlaid on top of d03 output with domain boundaries for d04 shown in red and d05 in blue. The geographical location of the CNRL facility is indicated with a star. Output from these three domains agree to a great extent for sea level pressure with $< 0.1\%$ difference, relative humidity (at 2 m agl) with 2% difference, and temperature (at 2 m agl) with 0.1% difference. Wind directions are also consistent with 2 to 7 degrees difference for a mean wind direction from about 240 degrees (west-south-west). Wind speeds are biased high with 0.7 m/s to 1.5 m/s for domain d04 and by 3 m/s to 4 m/s for domain d05 relative to domain d03 with mean wind speed of about 6 m/s.

Output from the finest three domains were compared to the historical observational data from the Wood Buffalo Environmental Association (WBEA) continuous monitoring stations Bertha Gan-

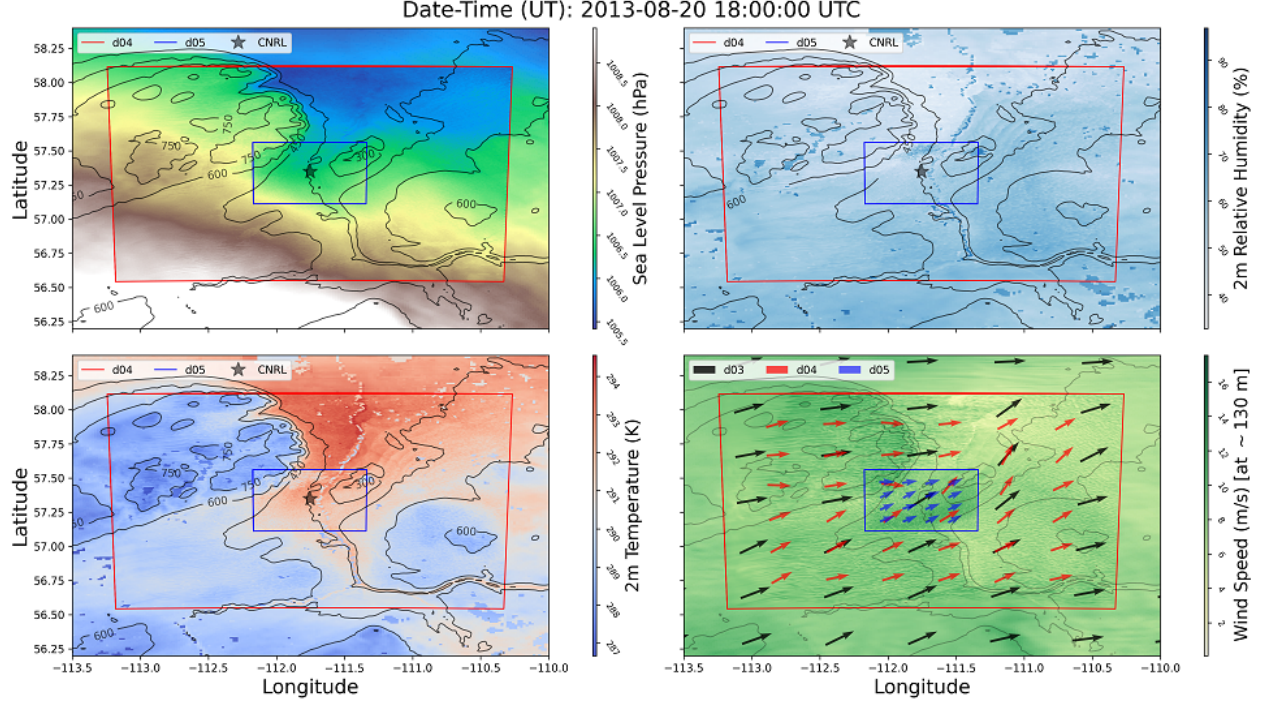


Figure 3.4: Model output sea level pressure (top left), 2-m relative humidity (top right), 2-m temperature (bottom left), and 10-m wind (bottom right) for domains d03, d04 and d05 on 20 August 2013 are shown. Boundaries of domains d04 (red rectangle) and d05 (blue rectangle) are also shown. Contours show terrain height in meters above seal level (masl). The geographical location of the CNRL facility is indicated with a star in the middle of domain d05. Output from the three domains agree to a good extent, with small overestimation in wind speed for d05 compared to the other two. Note that in the bottom-right panel, the arrow sizes are not representative of wind speeds, but model grid spacing (resolution).

ter – Fort McKay and Barge Landing, which have been monitoring since October 1997 and July 2001 respectively (<https://wbea.org/historical-monitoring-data/>). Figure 3.2b shows the locations for the two WBEA stations on the map of the region (red triangles). Both stations provide complete observational historical data for our periods of interest in August and September 2013. Documentations for Bertha Ganter – Fort McKay and Barge Landing monitoring stations indicate the height of meteorological sensors as 10 m above ground level (agl). We compared model sea level pressure, 2-m relative humidity (Rh), 2-m temperature and 10-m wind to WBEA observational data. Figure 3.5 shows wind-rose diagrams for case 1 on August 20 where output from model domains d03, d04 and d05 are compared to WBEA data at the location of Bertha Ganter–Fort McKay monitoring station. Wind directions for this case were from west and west-south-west during the simulation time, which is consistent with the observed WBEA wind directions. Model wind directions are within 20 to 30 degrees of the WBEA observational data. Model wind speeds are higher for all three domains compared to WBEA observational data for the locations of the two monitoring stations (with mean wind speed of 3 m/s): by 2 – 3 m/s for domain d03, by 3 – 4 m/s for domain d04, and by 4 – 7 m/s for domain d05. Note that the Large Eddy Simulation

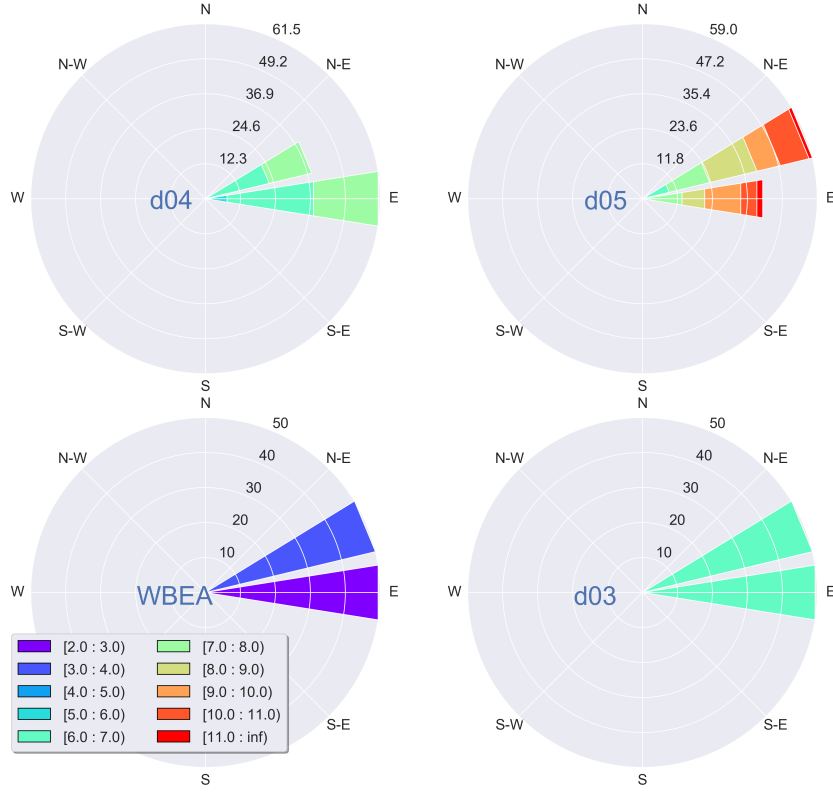


Figure 3.5: Wind-rose diagrams comparing observational data from WBEA monitoring station Bertha Ganter–Fort McKay to model output winds at the location of the WBEA station from domains d03, d04 and d05 for 20 August 2013. Wind directions are consistent within 20 degrees, blowing from w-s-w. Wind speeds are biased high at various degrees for the three domains.

(LES) subgrid parameterization was used for d04 and d05 simulations. Model winds for these two domains, especially for d05, are highly variable over time and space. Figure 3.6 shows output from d04 and d05 compared to observational data from the two WBEA monitoring stations for wind, 2-m temperature and 2-m relative humidity for case 1. Comparisons for the other cases show similar results, with agreements within 1% to 5% for 2-m temperature, 2-m relative humidity, sea level pressure and wind direction. Similar to case 1, wind speeds were higher than WBEA winds by between 2 m/s to 8 m/s for cases 2 and 3.

We note the following considerations when comparing model generated fields (e.g., wind) to WBEA observational data:

1. The lack of observational data from continuous monitoring stations for more spatial locations, especially closer to the centre of domain d05 (where CNRL is located) is a source of uncertainty. Note that the two available WBEA stations are located close to the southern boundary of domain d05, less than 200 model grid points from the boundary, as shown in Figure 3.2b. Model fields close to domain boundaries are highly impacted by the boundary

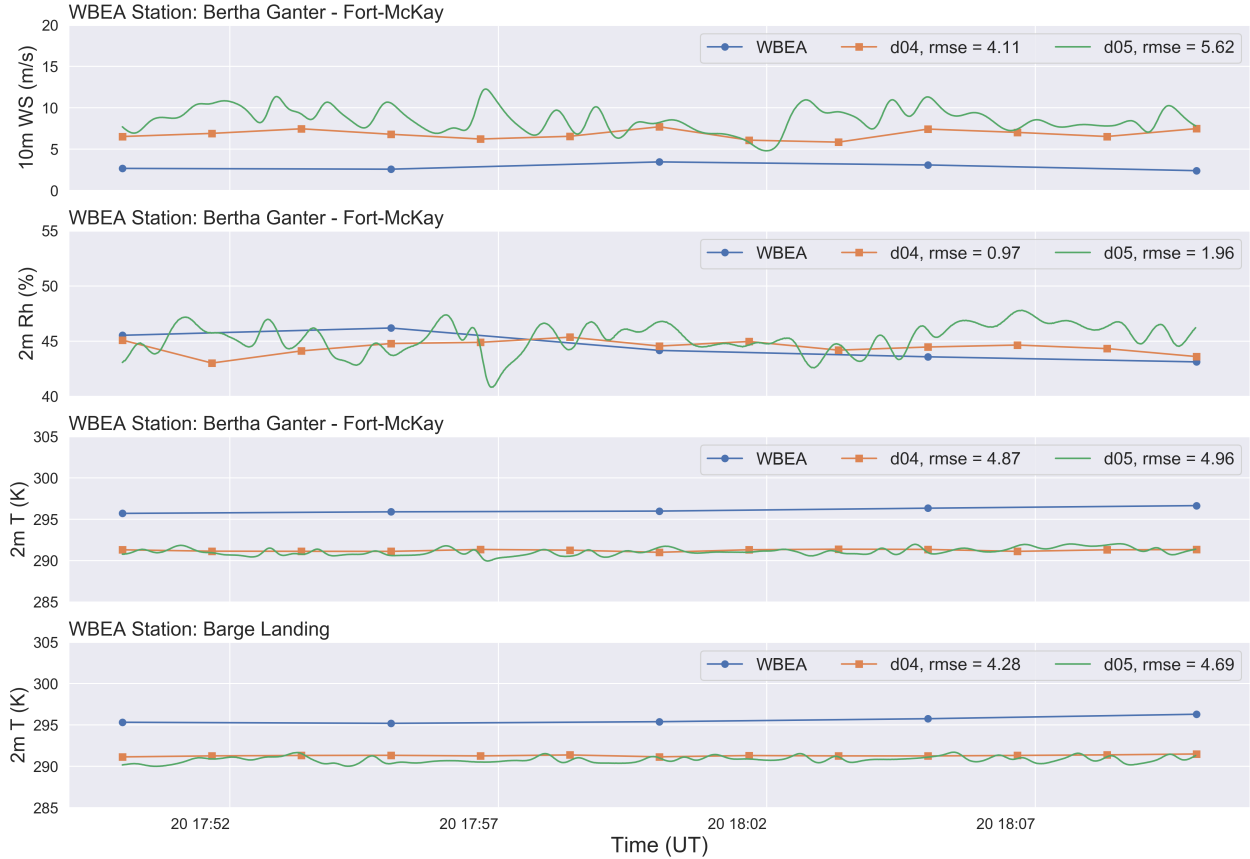


Figure 3.6: Model output for 2-m temperature, 2-m relative humidity and 10-m wind speed from domains d04 and d05 are compared to observational data from WBEA monitoring stations.

conditions from the parent domain, and are usually not included in model output analysis. For this work, we considered a buffer zone of 100 grid points on each side and excluded data from this zone in our analysis. We note that discrepancies between model fields and WBEA observational data are smaller for domains d03 and d04 compared to d05, where WBEA locations are well within the interior of the modelling domains (far from lateral boundaries).

2. The wind speeds in the NARR reanalysis data (at 31 km resolution) used as input for our simulations, were higher than WBEA observed values by 2-3 m/s for the region and the periods of interest. These are very similar to domain d03 output fields, which indicate the high fidelity in d03 domain simulations. As a result we can use d03 output fields at 1250 m resolution as a base case for evaluating d04 and d05 simulations. Note that wind speeds were biased high by only about 1 m/s for d04 and by about 3 m/s for d05 simulations (see Fig 3.8 for evaluations against d03).
3. Dynamical down-scaling of NARR reanalysis data from 31.25 km resolution to 50 m resolution with five nested domains and vertical grid refining, is another source of uncertainty. In concurrent grid nesting as used in this work, output from parent domain is interpolated

to provide initial and boundary conditions for each respective nested domain. Horizontal, vertical and temporal interpolation errors are therefore compounded with each nesting (five in this case). Daniels et al. (2016) demonstrated how horizontal grid nesting and vertical grid refining can result in 1-2 m/s bias in wind speeds for each nesting. This is consistent with our results where d04 wind speeds are higher than d03 by about 1 m/s, and d05 winds are higher than d04 by about 1-2 m/s.

4. The accuracy of subgrid scale LES parameterization in domains d04 and d05, is also a source of uncertainty. Liu et al. (2011) discussed concurrent nested modelling from synoptic scale to the LES scale with 4 domains. They demonstrate how simulated wind speeds differ by 2-5 m/s for the 4 domains, with weakest winds in the coarsest domain and stronger winds in the finest domain. Which is consistent with our results where wind speeds are biased high for the finest domain d05 compared to d04 and d03 by 1-4 m/s.

We also compared wind fields from domain d05 to aircraft observations during the 2013 JOSM campaign over the oil sands region for the same time periods as our model simulations. Figure 3.7 compares model wind speeds and directions for our three cases to aircraft observations for altitude levels of airborne measurements. WBEA data at 10 m agl are also included on the figure for comparison. Horizontal bars on Fig. 3.7 show one standard deviation in model-generated fields over domain d05. Model wind fields (vertical profiles) overlap with aircraft observation within one standard deviation for the three cases. As can be seen from Fig. 3.7, WBEA wind speeds are lower than both model and aircraft wind speeds. Note the high spatial heterogeneity in wind fields captured by both aircraft observations (blue dots) and model simulations (orange bars). Spatial (horizontal and vertical) variability in wind fields were more severe for case 2 compared to the other two cases. We discuss later how the conditions of case 2 resulted in weak advection of tracer mass and rendered this case unsuitable for top-down mass-balance retrievals.

Table 3.4: Evaluation of domain d05 simulations against d04 output fields at every 100 seconds at the geographical location of the CNRL oil sands facility (Lon=-111.738 and Lat=57.339). Note that positive/negative sings mean over/under-estimates by d05 relative to d04.

	Score	SLP (hPa)	2m RH (%)	2m T (°C)	10m U (m/s)	10m V (m/s)
Case 1	d04 mean	1006.42	51.18	19.32	5.05	3.20
	rms error	0.11	1.89	0.25	3.16	1.37
	mean error	-0.04	-1.20	-0.17	3.11	1.14
Case 2	d04 mean	1013.44	52.59	20.30	-0.766	0.799
	rms error	0.56	7.59	1.84	0.59	2.20
	mean error	0.55	6.86	-1.81	-0.55	-2.16
Case 3	d04 mean	1006.14	62.40	21.95	3.57	3.90
	rms error	0.09	2.44	0.38	2.77	1.01
	mean error	0.06	-2.24	0.36	2.73	0.75

For all cases, output from domain d03 showed the best agreement to observational data. Since observational data were not available at the location of CNRL oil sands facility and due to the fact

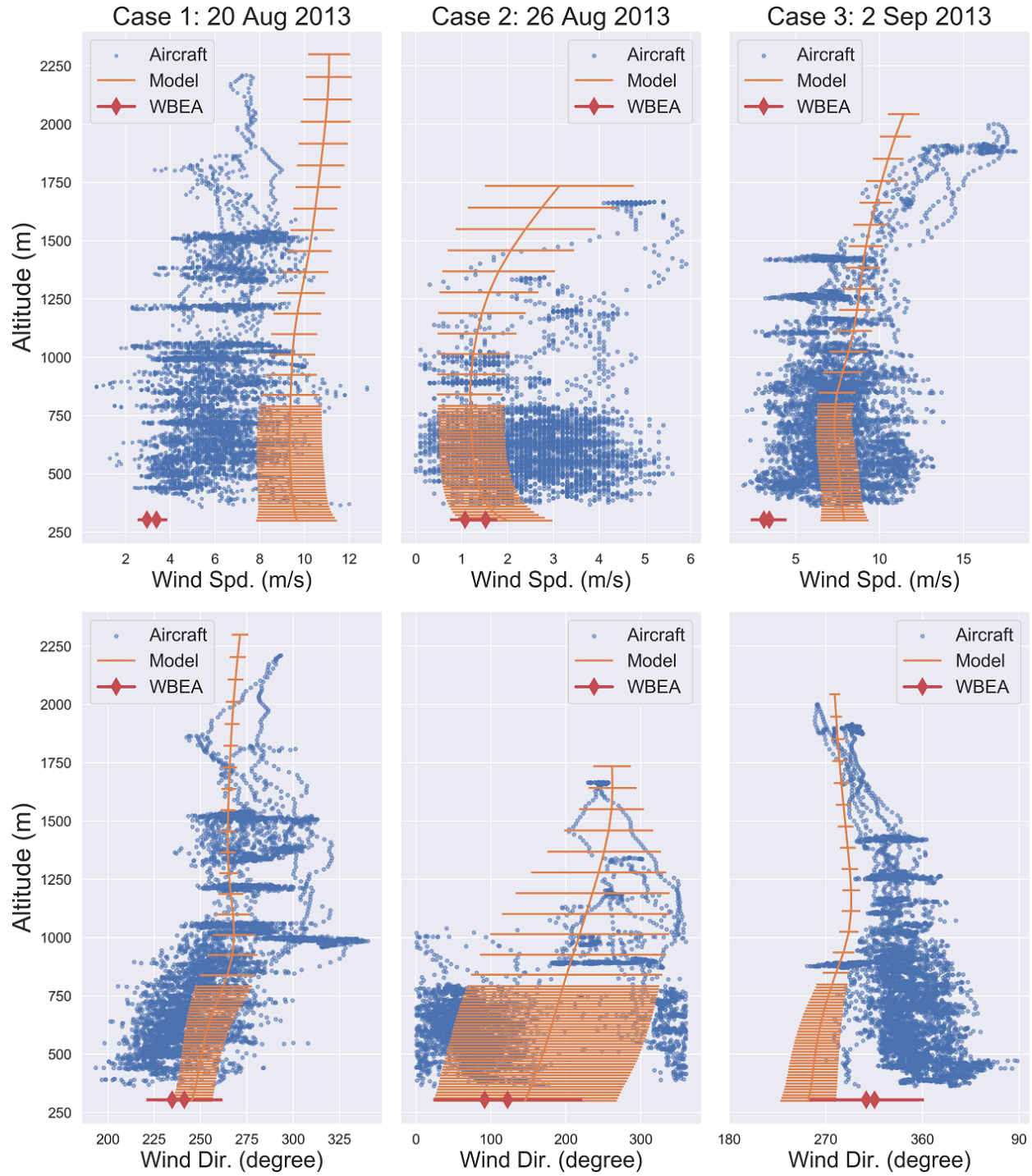


Figure 3.7: Model-generated wind fields (orange) compared to aircraft observations (blue dots) at several altitude levels during the 2013 JOSM airborne campaign, and WBEA observational data at 10 m agl (red diamonds). Horizontal bars show one standard deviation in model wind fields. Aircraft data and model fields agree within one standard deviation. WBEA wind speeds are lower than both aircraft and model wind speeds, red bars show one standard deviation in WBEA data.

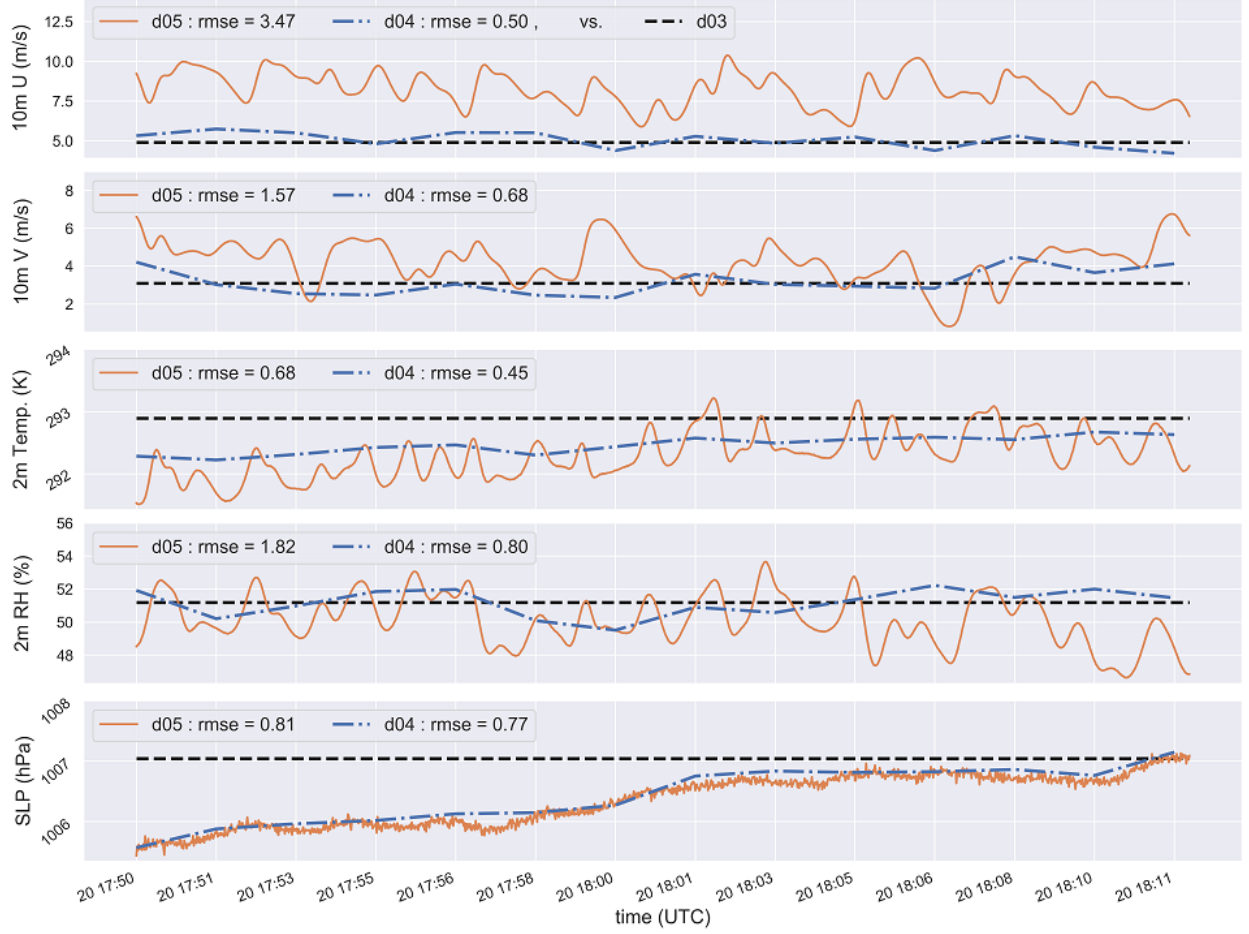


Figure 3.8: Model output meteorological fields from the two fine resolution domains d04 and d05 evaluated against d03 output for the location of CNRL facility. d04 and d05 output fields show good agreement to d03 for sea level pressure, humidity, temperature and wind. d05 wind speeds were biased high by 1.57 m/s and 3.47 m/s.

d03 output fields showed good agreement (see above) to observational data for the two monitoring stations, we evaluated the performance of d04 and d05 simulations to d03 as a proxy for observational data at the geographical location of CNRL. Figure 3.8 shows this comparison for case 1 for east (U) and north (V) wind components, 2-m temperature, 2-m relative humidity and sea level pressure. RMS error scores were small (e.g., 0.5 m/s for U , 0.68 m/s for V , 0.45 °C for temperature) for d04 simulations at 250 m horizontal resolution and with LES parameterization. RMS errors for d05 simulations were also similarly low for sea level pressure, temperature and humidity, with wind speed biased high by 1.57 m/s and 3.47 m/s for V and U respectively (at 10 m agl).

As the model output data from domain d03 were only for every 3 hours, to evaluate the performance of domain d05 simulations at CNRL at a higher temporal resolutions we performed evaluations against domain d04 with model output every 100 seconds. Evaluation results for all three cases are summarized in Table 3.4. Root mean square (rms) error ranges from 0.09 hPa to 0.56 hPa for sea level pressure (SLP), 1.89 % to 7.59 % for 2-m relative humidity (RH), 0.25 °C to

1.84 °C for 2-m temperature, 0.59 m/s to 3.16 m/s for west-east wind component U , and 0.75 m/s to 2.20 m/s for south-north wind component.

For case 1 on 20 August 2013, 2-m temperatures ranged from 17 °C to 20 °C and 10-m wind speed from 6.6 m/s to 11.4 m/s at the geographical location of main CNRL stack sources. For case 2 on 26 August 2013, the 2-m temperature range was similar to case 1 but the 10-m wind speeds were much lower ranging from 0.8 m/s to 3.5 m/s with mean 2.8 m/s. For case 3 on 2 September 2013, 10-m wind speed ranged from 5.3 m/s to 9.4 m/s with 2-m temperatures slightly higher than the other cases ranging from 21 °C to 23 °C. The temperature and wind speed ranges mentioned for the three cases are for periods between 18:00UT to 20:00UT (local noon to 2pm).

3.3.2 Plume Characteristics

Continuous passive tracer emissions for surface and stack sources were initiated simultaneously at different locations within the finest resolution modelling domain (d05), after a 30 min initial model spin-up time. Source locations, spatial extent, and release heights are provided in Table 3.2. Figure 3.9 shows tracer plumes for case 1 on 20 Aug 2013 at 17:35 UT, 2 hours and 5 minutes after the initial release. For the example shown, tracer plumes have propagated the downwind span of the modelling domain and reached the opposite lateral boundary. The top-down retrieval study in Chapter 4 makes use of tracer plume simulations in this chapter. As we discuss later in Chapter 4, plume characteristics have the potential to impact the accuracy of top-down retrievals using aircraft-based sampling of environmental fields. Therefore, in this section we discuss plume behaviour for our three case studies with emphasis on case 1 as our main case. In the next chapter, we refer to plume characteristics presented here in the context of top-down retrievals.

Super-resolution model simulations for case 1 for a period during 20 August 2013 started at 15:00UT. At 15:30UT tracer release for our 11 emission scenarios were initiated. For point/stack sources, continuous tracer release was assigned to select model grid cells geographically co-located with CNRL facility major point sources. Further point emissions were considered for two hypothetical upwind sources. The mean direction of wind for the first 1.5 hour of the simulation for this case was from south west, which transitioned to winds from west and west-south-west in the following hours (Fig. 3.9). Figure 3.10 shows tracer plumes at 16:30UT for case 1. By this point, tracer plumes are dispersed considerably and have been advected towards the east and north-east well beyond the boundaries of the CNRL facility (shown with the dashed rectangle in Fig. 3.10).

As mentioned before, we have discarded 100 grid points on either side and considered output data for the inner sub-domain in our analysis. In Chapter 4, where we simulate aircraft-based measurements of the environmental fields, the start of the sampling period is 16:30UT (similar to a JOSM 2013 box flight time conducted on the same date). Sampling starts at this point in the simulation time, as the addition (through source emissions) and removal of tracer mass (via advective and turbulent fluxes) through the boundaries of the control volume (box) have reached a mass-balance and a relative steady-state. Steady-state conditions are essential for simulating top-down retrievals based on sampling along the lateral boundaries of the control volume.

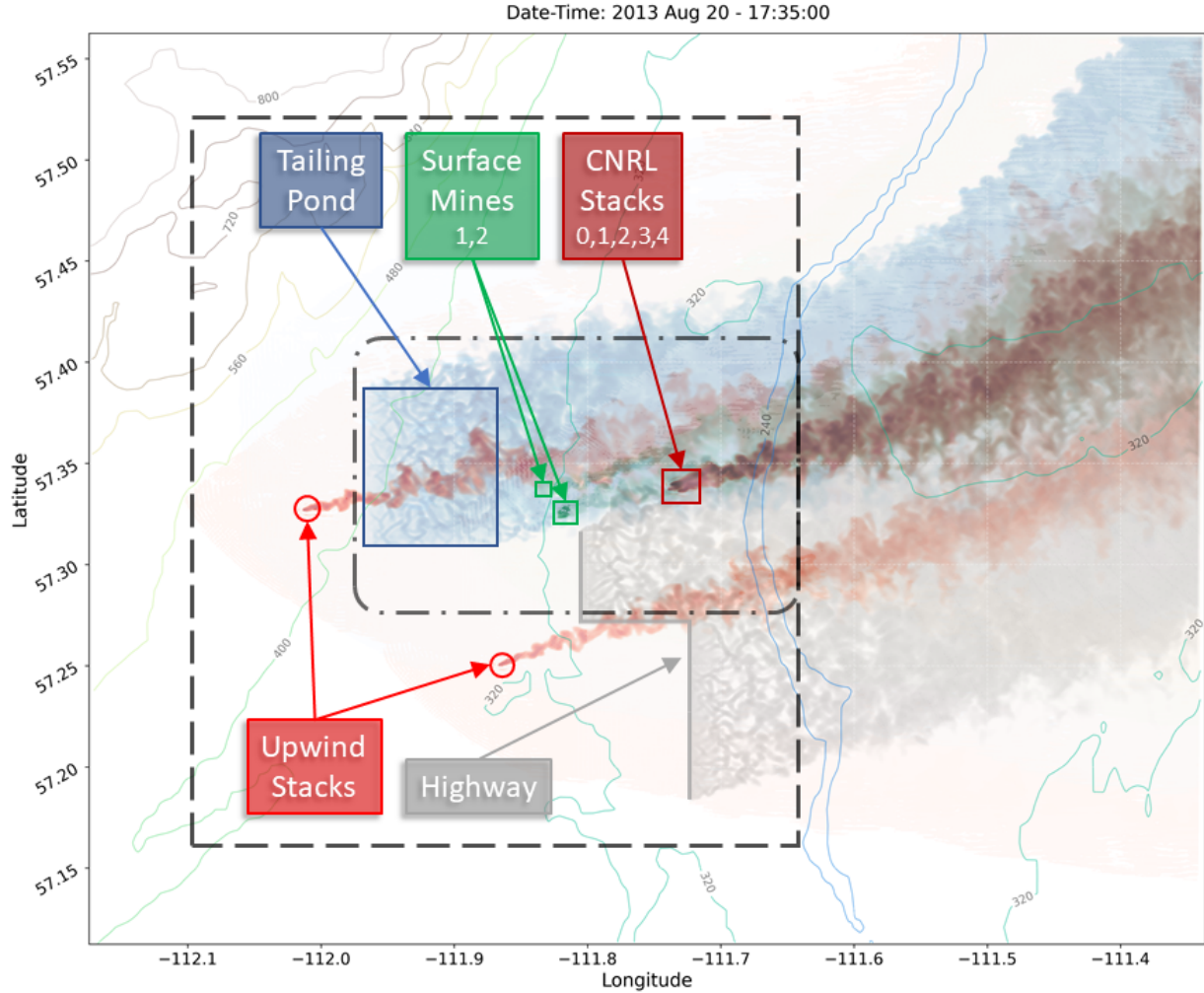


Figure 3.9: Top view of the modelling domain (d05) and the location of tracer emission sources are shown. Sources include a big area source (Pond), a multi-segment line source (Highway), two small area sources (Mines 1, 2), five CNRL stacks 0-4 and two upwind point sources (west and south). Emission plumes from Pond are shown in blue, Highway in gray, Mines in green and stack/point sources in red. Colour darkness is proportional to log of column total tracer mass. Two control volumes (dark dashed boxes) for 4D mass-balance calculations are shown, one enclosing all the emission sources and the other marking the boundaries of the CNRL facility (the smaller box) and only including within facility sources.

As we discuss in Chapter 4, the spatio-temporal variability in plume/flow characteristics has the potential to impact the accuracy of top-down estimates. The spatial heterogeneity of wind fields for case 1 can be seen over the span of the modelling domain by considering the wind barbs and also the tracer plume horizontal profiles (e.g., compare upwind plumes to main CNRL stack plumes in Fig. 3.10 top left panel). Temporal variability in mean conditions can also be seen from the snapshot in Fig. 3.10 by considering plume streaklines.¹ For all tracer plumes, an elbow at

¹“Streaklines are lines deposited in the flow during a time interval by continuous emission of a tracer from a fixed point. Examples can be seen in aerial photographs of smoke plumes emitted from smokestacks, or volcanic ash

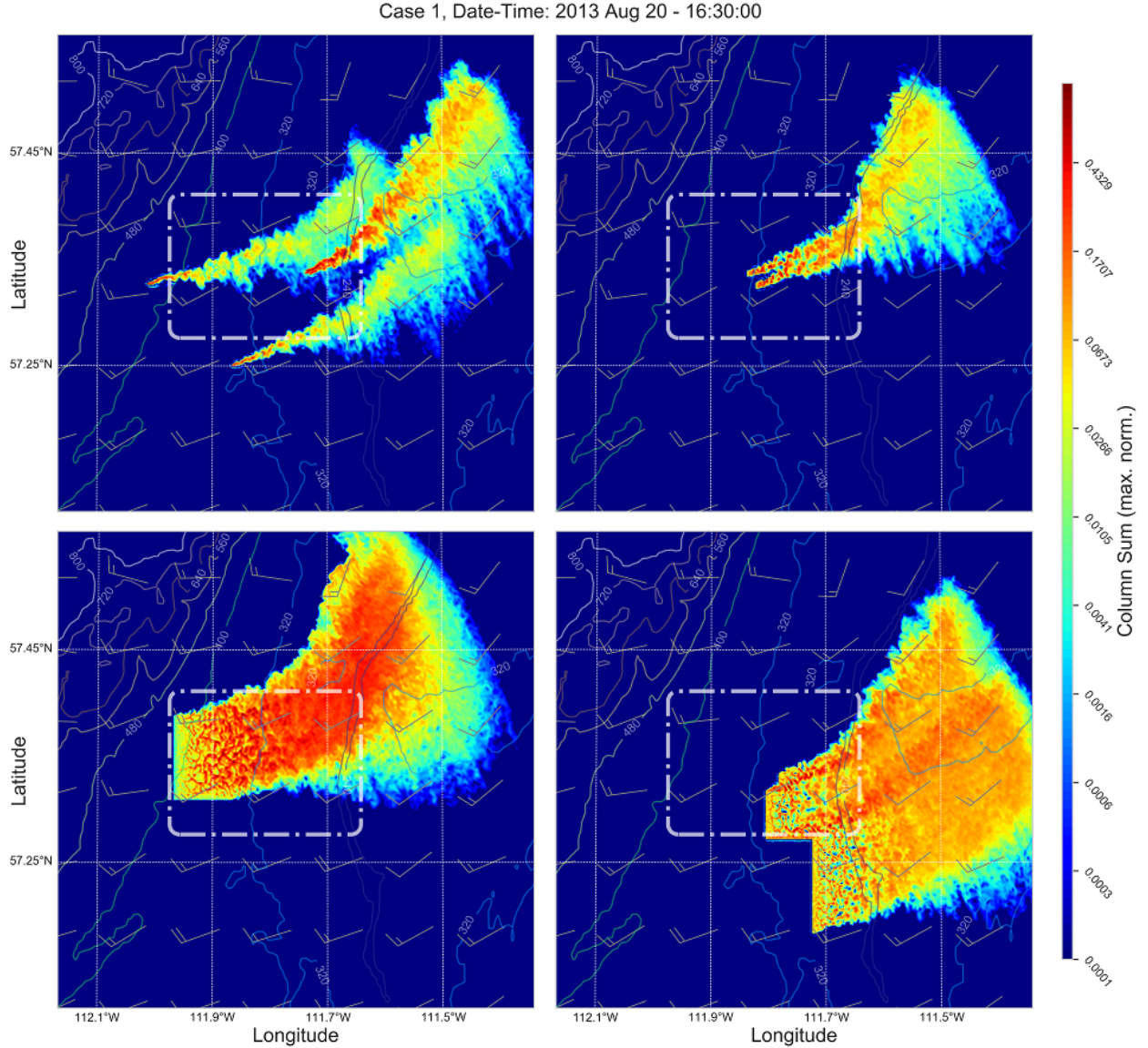


Figure 3.10: Case 1: top view of the 50 m resolution modelling domain with tracer plumes from our eleven emission scenarios at 16:30UT on 20 August 2013. Tracer plumes are dispersed and advected well beyond the boundaries of the CNRL facility, shown with the dashed curve (same as the small box in Fig. 3.9). 10-m wind data is shown with wind barbs over the entire domain (flag: half 5 m/s, full 10 m/s). The shown data is the tracer column total normalized to the maximum value for each panel. (top left) point/stack sources CNRL0-4, CNRLs and CNRLw. (top right) surface/area sources MINE1 and MINE2. (bottom left) the large area (surface) sources POND. (bottom right) the multi-segment line (surface) source HWY. Refer to Table 3.2 for source specifications.

<45 degrees angle is visible in the plume streakline over the eastern boundary of CNRL facility (dashed curve). This directional change is partially due to the spatial variability in the flow and partially due to interactions with topography (Fig. 3.10 contours). The Athabasca river is located

clouds” (Stull, 2017).

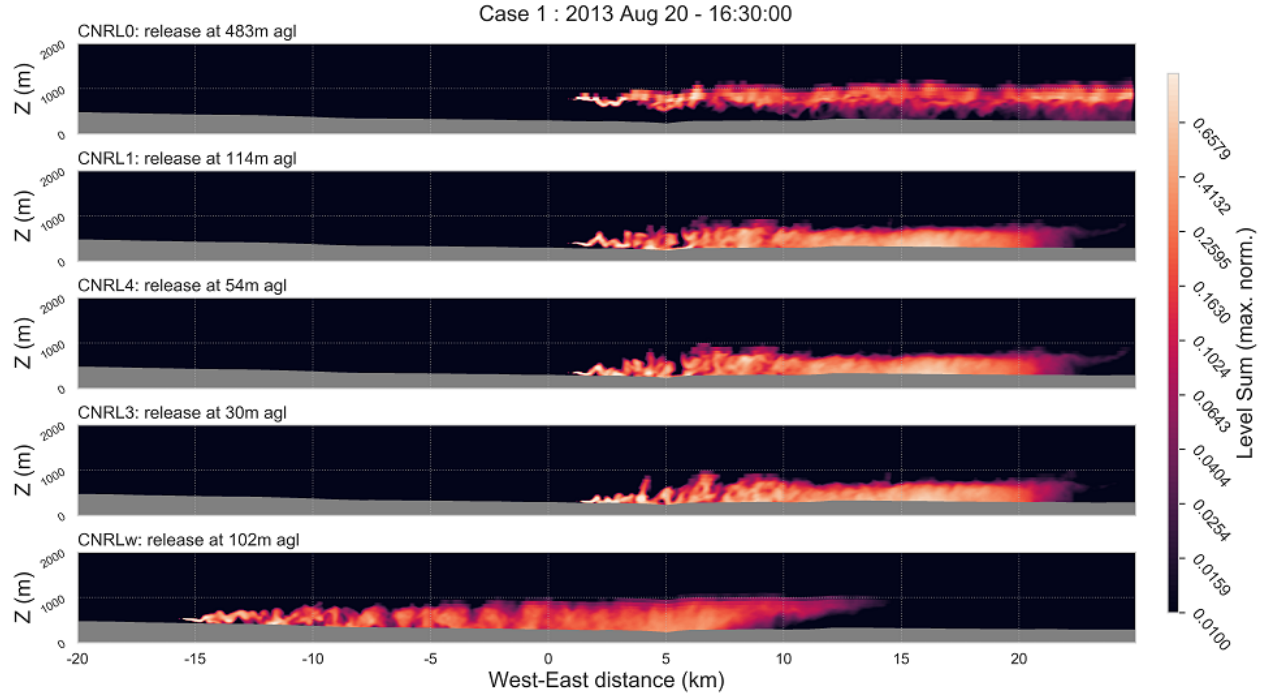


Figure 3.11: Domain d05 west-east vertical cross-section for case 1 on 20 August 2013. Vertical cross-section of tracer plumes from stack/point emission sources are shown. Release height above ground is indicated for each source. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for west-east distance in km is at domain centre.

to the east of the CNRL facility. There is also a temporal change in the direction of the flow from north to east during the previous 30 minutes. Fig. 3.9 shows the same case about an hour later at 17:35 UT, where flow is mostly towards east. See Appendix F for links to model time-lapse videos.

Tracer amounts were advected at different vertical levels and flow regimes. Fig. 3.10 (wind barbs) show wind speeds at 10 m above ground level (agl), which were between 10-20 m/s for this case. Flow was different at other vertical levels. For instance, surface emissions from MINE 1 and MINE 2 (Fig. 3.10 top right panel) were advected in slower air flows and covered less downwind range during the same time period compared to stack (elevated) emissions CNRL 0-4 (Fig. 3.10 top left panel). Figure 3.11 shows a west-east vertical cross-section of the modelling domain, with tracer plumes from select stack/point emission sources. CNRL 1-4 tracer release heights correspond to real-world CNRL stacks, CNRL 0 is released at 483 m agl representing a (hypothetical) buoyant plume rise, and CNRLw, which represents (hypothetical) upwind emissions, is released at a height similar to CNRL 1. Advection in west-east orientation is unidirectional at all vertical layers as can be seen in Fig. 3.11. The plume centre-line for CNRL 0 emissions remains at about the initial release height, while mixing in the vertical as it is advected downwind. CNRL 1-4 emissions, although released at different heights above ground (by a few tens of meters), show similar vertical mixing profiles along the downwind advection path. Emissions from CNRL 1-4, mixed to the ground surface within the 5 km range and assumed a near uniform vertical mixing beyond 15 km

downwind distance. Note how plumes from these sources interact with the Athabasca river basin at the 5 km distance. There is an apparent discontinuity in tracer concentrations beyond which mixing in the vertical is intensified. This is likely due to different surface fluxes over the river and the ground on either side with stronger updrafts. This interaction is less visible in CNRLw plume which is relatively well mixed by 5 km distance, but it can be seen that the vertical mixing becomes more uniform on the other side of the river for this plume. Vertical mixing as a function of downwind distance and its impact on aircraft-based retrievals is discussed in Chapter 4.

Change in plume buoyancy or atmospheric stability can place the plume at different flow regimes at different vertical heights which in turn can change the advection time/path and the downwind horizontal flux. Figure 3.12 shows the south-north vertical cross-section for the same case, with CNRLw replaced by CNRLs (the southern upwind stack). Wind and therefore transport in the south-north direction were weaker than the transport in the west-east direction. There was also a strong vertical shear in south-north wind (V). See Appendix Figure E.1 for meteorological vertical profiles for this case. While all tracer plums for release heights of 30 m to 114 m agl were advected north, CNRL 0 at 483 m agl was advected south (Fig. 3.12 top panel). On the other hand, CNRL 0 tracer amounts that were mixed towards ground surface were advected north. This was also true for tracer amounts from other sources that were mixed in the vertical up to over 500 m agl, which are not visible for the data/colour scale in Fig. 3.12 due to small concentrations. This type of vertical wind shear can result in tracer amounts being dispersed and advected in different directions, which in turn results in the temporary accumulation/storage of mass in the volume/environment at the vicinity of the emission source. Similar atmospheric processes governed the dispersion and transport of tracer plumes from the surface emission sources. Figures 3.13 and 3.14 show west-east and south-north vertical cross-sections for surface emissions.

WRF model simulations for case 2 for the period during 26 August 2013 started at 18UT. As mentioned before, the JOSM emission flight for this day was rejected for emission rate calculations in relevant publications due to low and variable wind speeds (Fathi et al., 2021). This case is referred to as "rejected" throughout this work. We recreated the meteorological and tracer transport conditions on this day with our super-resolution simulations, which are presented here as an example of unsuitable conditions for top-down retrieval. For this case, tracer emissions from various sources were initiated at 18:30UT. During the period between 19UT to 21UT, wind fields over the region of interest were highly variable (spatially) with very low wind speeds. Figure 3.15 is a top view of the modelling domain for this case, showing tracer plumes from the various emission sources and the wind data as wind barbs. Wind speeds at a height of 10 m agl were less than 5 m/s over the modelling domain. The mean wind direction was towards the south and south-west near the ground, transitioning towards the south-east up to 1000 m agl, and towards east and north-east above 1000 m agl. The south-north wind component V ranged from -1 m/s to 1.5 m/s in the vertical, with the west-east wind component U ranging from -0.5 m/s near the ground surface to 7.5 m/s up to 2500 m agl. See Appendix Figure E.2 for meteorological vertical profiles for this case. The weak advection and the strong vertical wind shear resulted in tracer plumes being transported

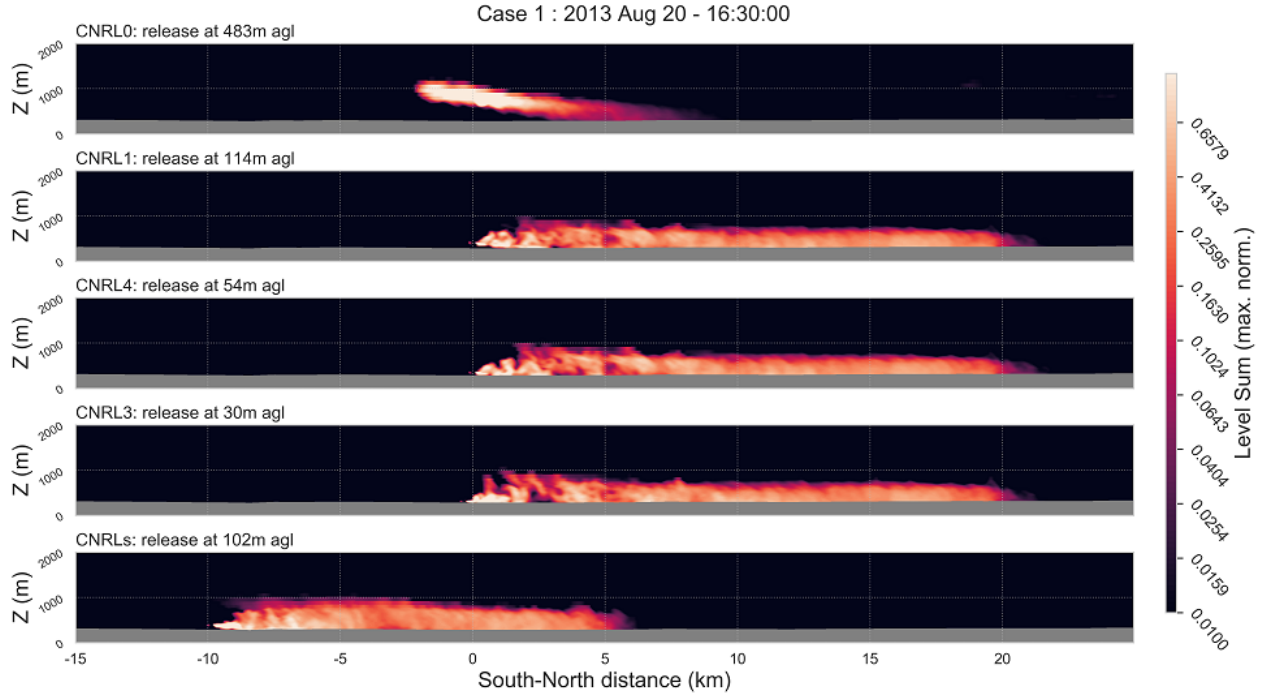


Figure 3.12: Domain d05 south-north vertical cross-section for case 1 on 20 August 2013. Vertical cross-section of tracer plumes from stack/point emission sources are shown with release heights indicated for each source. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for south-north distance in km is at domain centre.

for only a few kilometres in the horizontal and mainly staying within the boundaries of the CNRL facility, as shown in Fig. 3.15.

Unstable atmospheric conditions persisted during the simulation period over the region of interest for case 2. Gradient Richardson number (Ri) values, which is a measure for atmospheric dynamic stability (see Chapter 2 Eq. 2.17), were below the critical value of $Ri_c = 0.25$ up to 400 m agl. Ri values below 0.25 correspond to unstable conditions in the atmosphere (AMS, 2021b). Consequently, tracer plumes from emission sources mixed in the vertical up to 2000 m during the simulation time. See Figures E.3 to E.6 for west-east and south-north vertical cross-sections for the different emission scenarios of this case. Similar meteorological and atmospheric conditions were observed during the the JOSM 2013 field campaign during the same period. These conditions were also simulated using Environment and Climate Change Canada’s (ECCC) air-quality model GEM-MACH at 2.5 km resolution. Our super-resolution (50 m) WRF model simulations were also successful in recreating the conditions on 26 August 2013 over CNRL. See Chapter 2 (section 2.4.2) for detailed discussions on how low and variable wind and unstable conditions affect the atmospheric transport of tracer plumes for this case.

Case 3 model simulations at super-resolutions for the period during 2 September 2013 started at 15UT. Tracer emissions were initiated at 15:30UT. Atmospheric conditions during this case were fairly stable, with Gradient Richardson number Ri exceeding the critical values $Ri_c = 0.25$

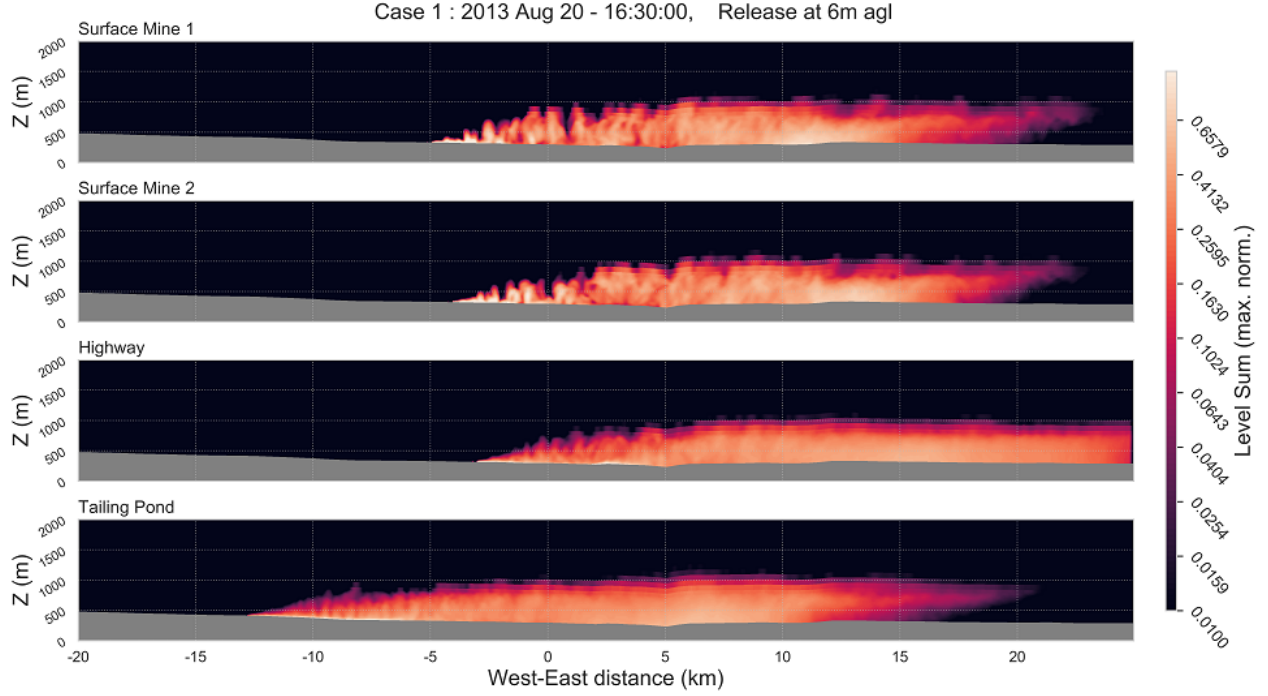


Figure 3.13: Domain d05 west-east vertical cross-section for case 1 on 20 August 2013. Vertical cross-section of tracer plumes from surface emission sources (area, line) are shown. Release height above ground is 6 m agl for all sources. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for west-east distance in km is at domain centre.

(indicating atmospheric dynamic stability) below 100 m above ground level (agl). Wind speeds were higher for this case compared to the other cases with west-east wind component U ranging from 5 m/s near ground surface to 15 m/s up to 2000 m agl. The south-north component of the wind V was about 5 m/s near the surface, with a strong shear in the vertical transitioning towards south at about 500 m agl. See Figure E.7 for meteorological vertical profiles during the time period of this case. Figure 3.16 shows the top view of the modelling domain with tracer plumes and wind data at 10 m agl for case 3. Similar to case 1, the mean direction of transport was towards east and north-east, but with a stronger northward component compared to case 1.

3.3.3 Tracer Mass Conservation and Emission Rate Calculation

In this section we evaluate the conservation of tracer mass within domain d05. The Runge-Kutta transport scheme in our WRF modelling setup is conservative, however it does not guarantee positive definiteness on its own. Negative mass creation is offset by positive mass such that tracer mass is conserved over the modelling domain (Skamarock, 2008). Negative mass can be set to zero, but this will result in erroneous increase of tracer mass within the modelling domain. By choosing a positive-definite scalar advection option in WRF, as we did in our simulations, a flux re-normalization is applied to the transport step to remove the nonphysical effects such as creation of the negative mass and the excess positive mass (Skamarock and Weisman, 2009). The turbulent

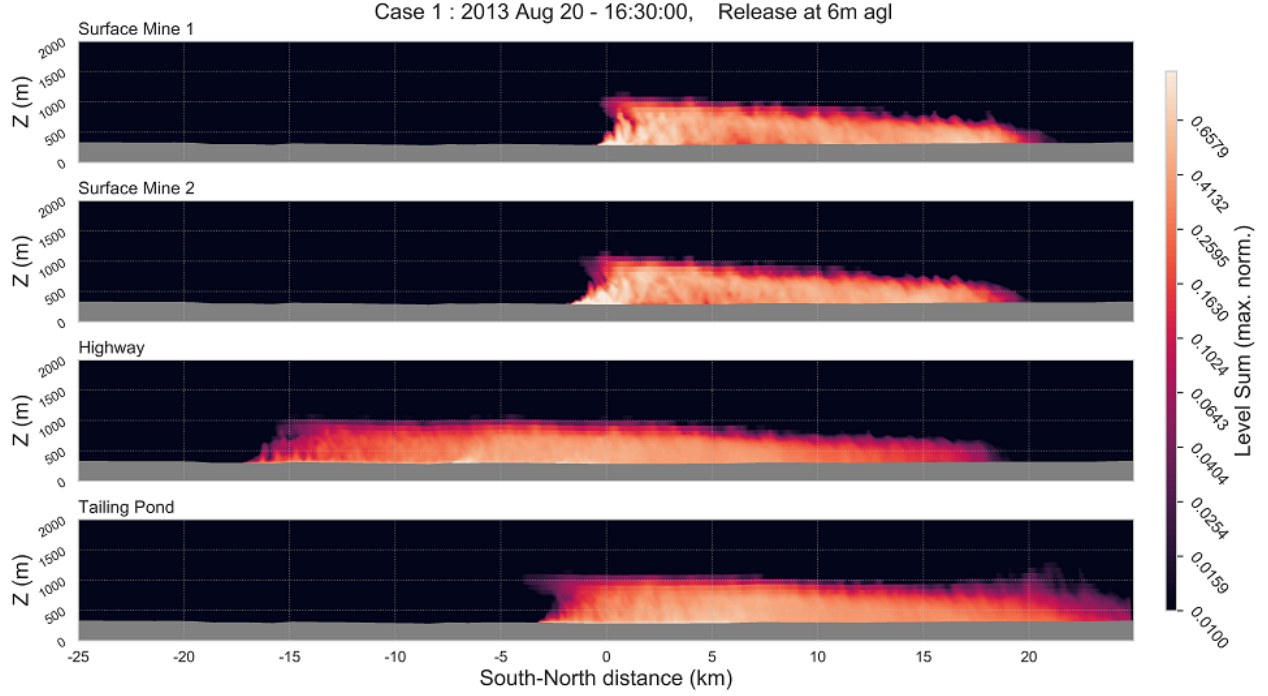


Figure 3.14: Domain d05 south-north vertical cross-section for case 1 on 20 August 2013. Vertical cross-section of tracer plumes from surface emission sources (area, line) are shown with release heights at 6 m agl for all sources. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for south-north distance in km is at domain centre.

diffusion step in the model is also prone to creating negative mass, but to a lesser degree. As we show in the following, the positive-definite re-normalization applied to the advective transport is not always able to filter out the negative mass (and the excess positive mass) created by the diffusion transport. This can result in local mass deficit/surplus.

We investigated the tracer mass budget within domain d05 by integrating over the entire domain at each model output time-step (Eq. 3.2). Figure 3.17 shows the resulting time-series for the eleven emission scenarios for the first hour after tracer release was initiated. In Fig. 3.17, the net mass present within the modelling domain is separated into negative and positive mass. Time-series are shown as the ratio of mass present in the domain to mass emitted (normalized) for each emission source. As can be seen from Fig. 3.17, initially up to 14% negative mass (figure bottom panel) is created within the domain which is offset by about the same amount of excess positive mass creation (figure middle panel). The net mass (figure top panel) is conserved as the negative and the excess positive mass cancel out. The creation of negative mass is reduced over simulation time until it falls below 2% after about 30 min simulation time. The creation of negative mass is likely due to sharp gradients in tracer mass immediately after the initial release, which are concentrated mostly upwind of the emission sources. As tracer mass is advected and dispersed over several grid points, the gradient is smoothed out and negative mass creation becomes less pronounced. Fig. 3.17 top panel shows that present-to-emitted mass ratio for 10 out of 11 sources are conserved and remain

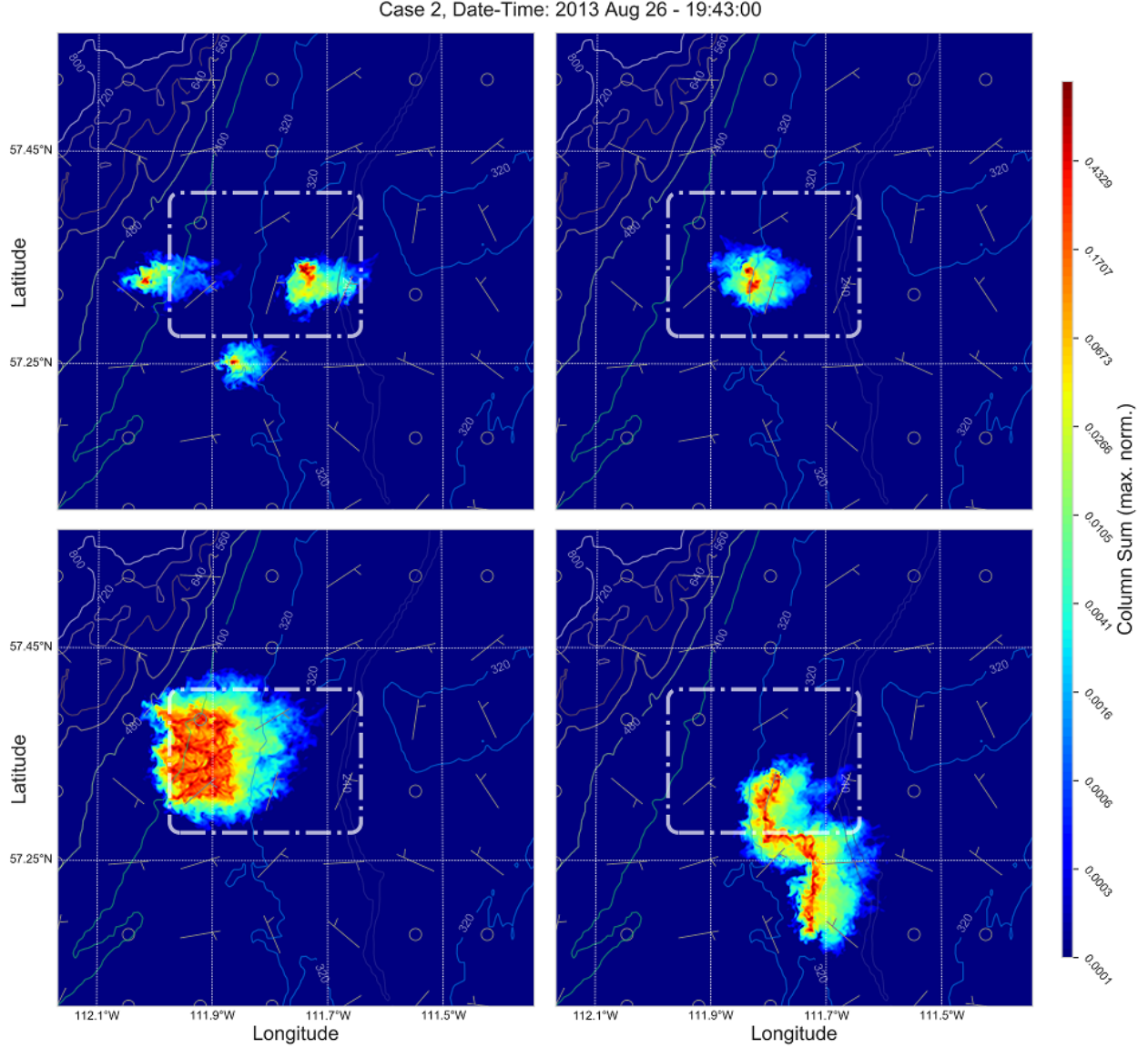


Figure 3.15: Case 2: top view of the 50 m resolution modelling domain with tracer plumes from our eleven emission scenarios at 19:43UT on 26 August 2013. 10-m wind data is shown with wind barbs over the entire domain (flag: half 5 m/s, full 10 m/s). The shown data is the tracer column total normalized to the maximum value for each panel. (top left) point/stack sources CNRL0-4, CNRLs and CNRLw. (top right) surface/area sources MINE1 and MINE2. (bottom left) the large area (surface) sources POND. (bottom right) the multi-segment line (surface) source HWY. For this case tracer plumes were not advected considerably, due to low and spatially variable winds. Refer to Table 3.2 for source specifications.

equal to unity up to about 16:30UT when they reach the domain boundaries and are removed from the modelling domain. The exceptions is CNRL 0, which does not keep up with the emissions and, except for a few minutes after the initial release, is always less than one. Note that the time-series in Fig. 3.17 are the ratio of mass present in the modelling domain at each timestamp to the total

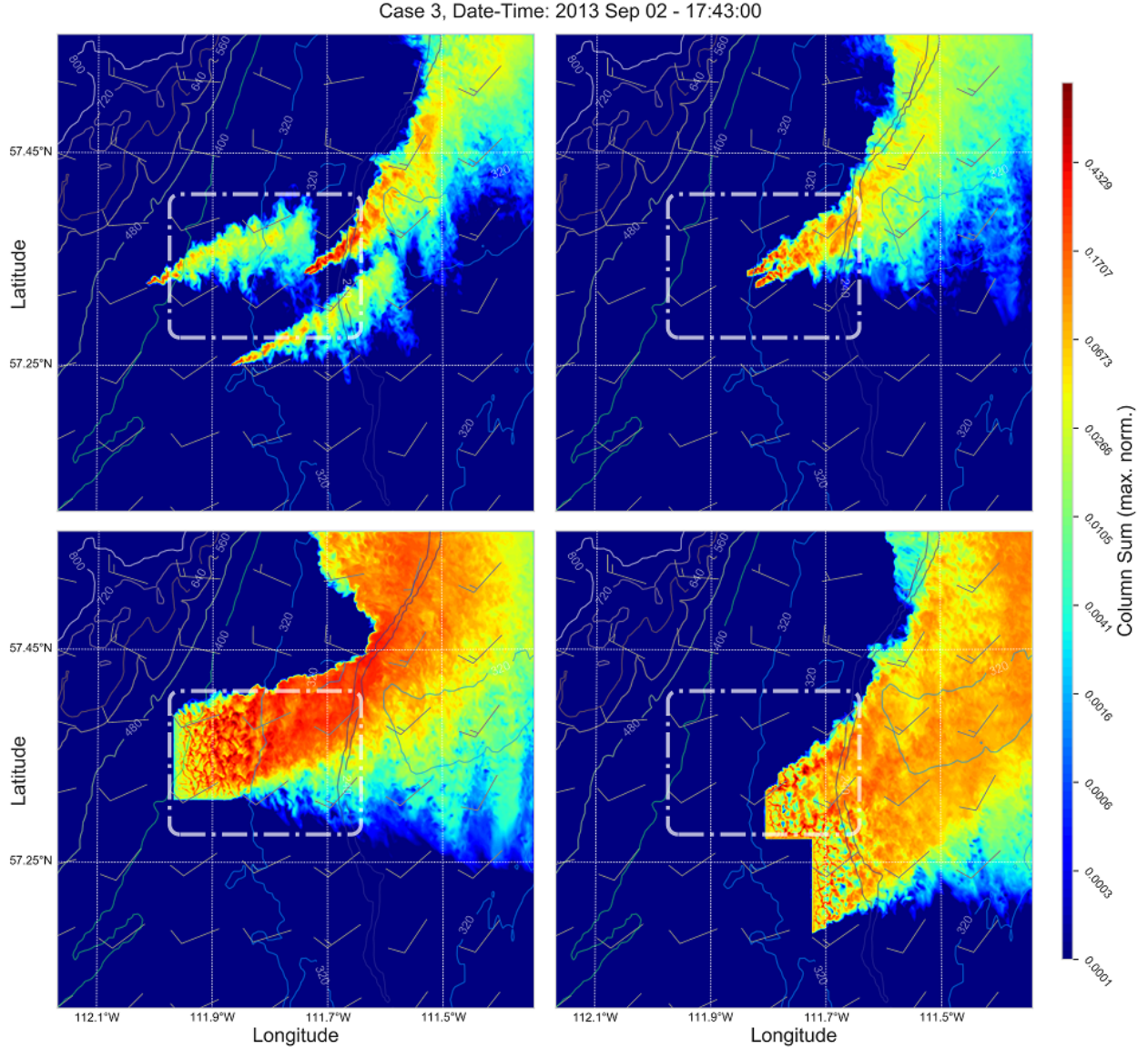


Figure 3.16: Case 3: top view of the 50 m resolution modelling domain with tracer plumes from our eleven emission scenarios at 17:43UT on 2 September 2013. Tracer plumes are dispersed and advected well beyond the boundaries of the CNRL facility (dashed curve) as in case 1. Wind data at 10 m agl is shown with wind barbs over the entire domain (flag: half 5 m/s, full 10 m/s). The shown data is the tracer column total normalized to the maximum value for each panel. (top left) point/stack sources CNRL0-4, CNRLs and CNRLw. (top right) surface/area sources MINE1 and MINE2. (bottom left) the large area (surface) sources POND. (bottom right) the multi-segment line (surface) source HWY. Refer to Table 3.2 for source specifications.

mass emitted up to that point. The mass present in the domain (not normalized) increases initially and later plateaus (approaching an asymptote) as plumes start exiting through domain boundaries at rates less than or equal to source emission rates.

Figure 3.18 shows temporal rate of change in tracer mass for each emission scenario, normalized

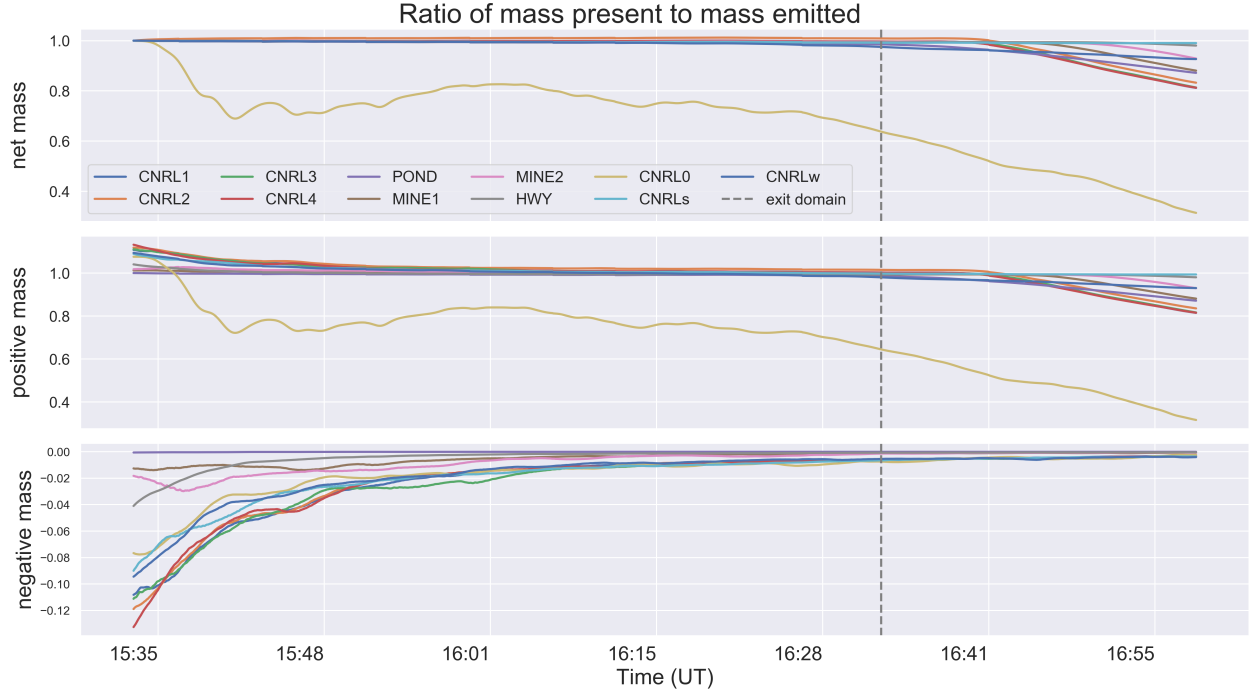


Figure 3.17: The ratio of tracer mass present within the modelling domain to mass emitted over simulation time after the start of tracer release. Time series are separated into positive and negative mass in the bottom two panels. Tracer mass remains conserved over time, with the exception of CNRL 0. Note that the drop in present-to-emitted mass ratio beyond 16:30UT, marked by the vertical dashed line, is due to tracer mass exiting through domain boundaries and not a violation of mass conservation.

to model input emission rate (MIE). This rate of change was calculated by differentiating the domain mass time-series shown in Fig. 3.17 according to Eq. 3.3, which is equivalent to the storage rate term for the entire domain ($S_{C,d05}$). The net rate of change (top panel) is separated into rates for positive and negative mass in the bottom two panels, for each tracer case. The rate of creation of negative mass oscillates between -10% and 5% and is damped over time. This is offset by the creation of excess positive mass, resulting in net mass conservation over the modelling domain (an artifact of the model transport scheme). For tracer mass to be conserved, the net rate of change over the entire domain must be equal to model input emission rate for the time period before tracer plumes start exiting through domain boundaries (before about 16:30UT for most tracers as shown in Fig. 3.18). Tracer mass at domain boundaries was set to zero (boundary condition). As plumes reached the boundaries the tracer mass was removed from the domain (at different times for each tracer depending on the transport speed and source proximity to the boundaries), hence the drop in $S_{C,d05}$ as shown in Fig. 3.18. This was true for all emission scenarios as can be seen from Fig. 3.18 top panel. The $S_{C,d05}$ time-series for ten out of eleven sources were at MIE level, indicating mass conservation for these cases. The only exceptions was CNRL 0 for which $S_{C,d05}$ oscillated around MIE without converging. It is evident from Figures 3.17 and 3.18 that CNRL 0 mass was

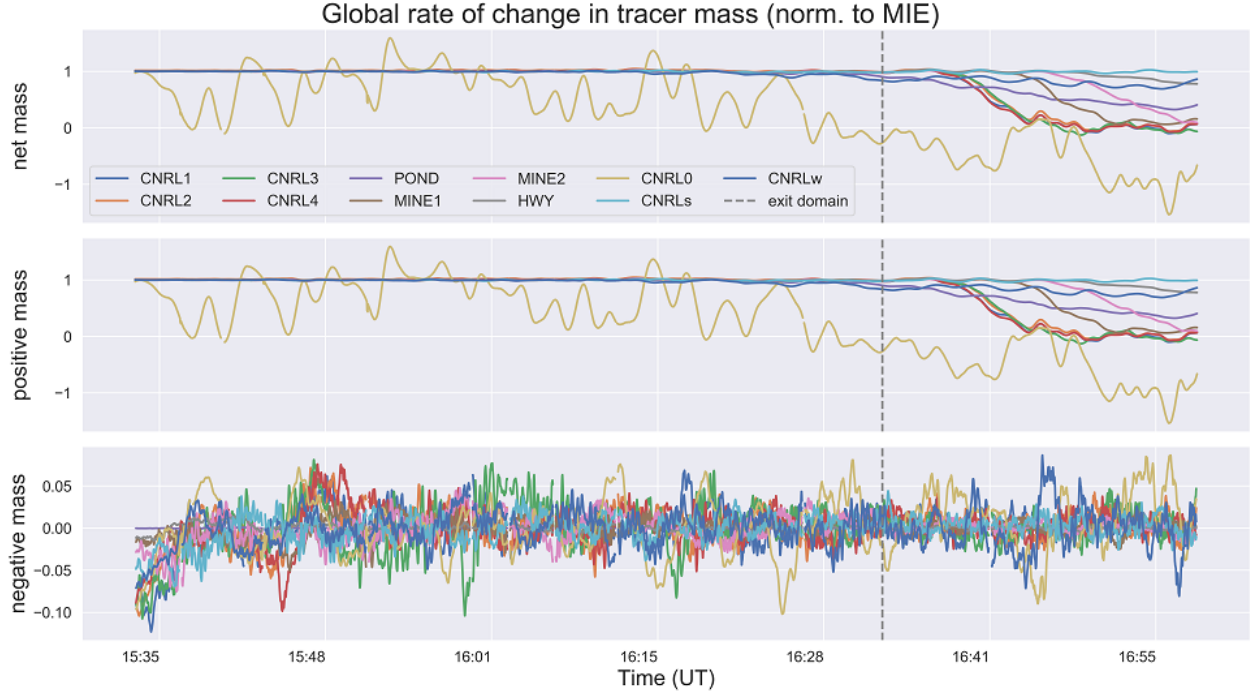


Figure 3.18: Temporal rate of change in tracer mass for the entire domain $S_{C,d05}$ is shown. Time series are normalized to model input emission (MIE) rates. The net rate of change (top panel) is separated into rates for positive and negative mass in the bottom two panels. The creation of negative mass is offset by creation of excess positive mass. Rate of change in tracer mass match MIE (conserved) for all except one emission source (CNRL0).

not conserved.

For the case of CNRL 0, tracer release height was at the model vertical level where model resolution (vertical) transitions from model layer thickness of about 12 m to progressively increasing thicknesses (and decreasing resolutions). Transport (advection and diffusion) of tracer amounts between model vertical layers of varying resolutions, resulted in erroneous and unbalanced (by excess positive mass) creation of negative mass (both in magnitude and spatial location). The positive-definite transport scheme in our WRF setup failed to conserve the tracer mass for this case, which resulted in about 31% loss. Table 3.5 lists statistics for the first hour of tracer emissions (before exiting through domain boundaries) for all emission scenarios. The temporal means (μ) and standard deviations (σ) are indicated for the normalized rates ($S_{C,d05}$) shown in Fig. 3.18. As mentioned before, CNRL 0 is the only non-conservative case with a mean value of 0.69 and standard deviation of 0.39. Therefore, in this chapter (and Chapter 4) we discuss CNRL 0 mainly in terms of plume behaviour rather than a conserved quantity in mass-balance calculations. The mean normalized rates for CNRL 2 and CNRL w were within 2% and 1% of MIE, respectively. For the remaining 8 out of 11 cases, the mean normalized rate were equal to 1.00 (perfect agreement with model input emission rates). These results apply to all three case studies within 2-4%.

Results in Table 3.5 indicate global (over the entire modelling domain) mass conservation for 10

Table 3.5: Normalized (to MIE) rates of change in tracer mass for all emission scenarios are shown. Mean rates are given for net, positive (+) and negative (−) mass. Standard deviations are also provided. Rates agree with model input emissions within 2% for 10 out 11 tracers (conserved).

Tracer ID	Temporal Mean (μ)			Standard Deviation (σ)		
	net	+	−	net	+	−
CNRL 1	1.00	1.00	0.00	0.07	0.07	0.03
CNRL 2	1.02	1.00	0.00	0.07	0.07	0.03
CNRL 3	1.00	1.00	0.00	0.07	0.07	0.04
CNRL 4	1.00	1.00	0.00	0.07	0.07	0.03
POND	1.00	1.00	0.00	0.07	0.07	0.00
MINE 1	1.00	1.00	0.00	0.07	0.07	0.02
MINE 2	1.00	1.00	0.00	0.07	0.07	0.03
HWY	1.00	1.00	0.00	0.07	0.07	0.01
CNRL s	1.00	1.00	0.00	0.07	0.07	0.03
CNRL w	0.99	0.99	0.00	0.07	0.07	0.03
CNRL 0	0.69	0.69	0.00	0.39	0.39	0.03

out 11 emission cases. We further evaluated model local mass conservation through mass-balance and flux calculations. A control volume enclosing all the emission sources, with the downwind wall at a 5 km distance from the main CNRL stacks, was considered (large rectangle in Fig. 3.9). We conducted 4D mass-balance calculations using this control volume for all case studies, to evaluate the performance of model simulations in the context of local mass conservation and transport of tracer amounts within a sub-domain (control volume). The mass-balance calculations for this portion of our analysis were conducted for the period between 1 to 2 hours after the tracer release started, well after plumes crossed the box (control volume) walls.

Mass-balance calculations were done according to Eq. 3.7. The net flux out term $F_{C,out}$ was calculated using the instantaneous fields (e.g., wind speed, tracer concentrations) along top and lateral boundaries of the control volume. $F_{C,out}$ includes horizontal ($F_{C,H}$, $F_{C,HT}$) and vertical ($F_{C,V}$, $F_{C,VT}$) advective and turbulent fluxes across box walls. The rate of mass storage/release S_C within the control volume (box) at each model output time-step (1 second) was calculated by integrating over the entire volume of the box and differentiating with respect to time. Source emission rates E_C were calculated as the sum of $F_{C,out}$ and S_C and were compared to model input emission rates (MIE). Contributions (absolute value) of different terms are shown in Figure 3.19 for case 1. The main contribution came from the horizontal advective flux, along with significant contributions from the storage rate term S_C . The horizontal turbulent flux contributed between −0.3% to 2.3% for the three cases. Contributions from vertical fluxes were negligible as the box top was chosen at a height of over 4 km agl, well above the mixing layer height at about 2-3 km agl. The breakdown of contributing terms for the three cases are summarized in Tables 3.6.

For case 1 (Table 3.6), mass-balance estimates were within 4% of MIE for 10 out of 11 tracers. Estimates for the (globally) non-conservative tracer CNRL 0 were biased low by 51%. Estimated emission rates for all the other emission scenarios were in over 96% agreement with model input

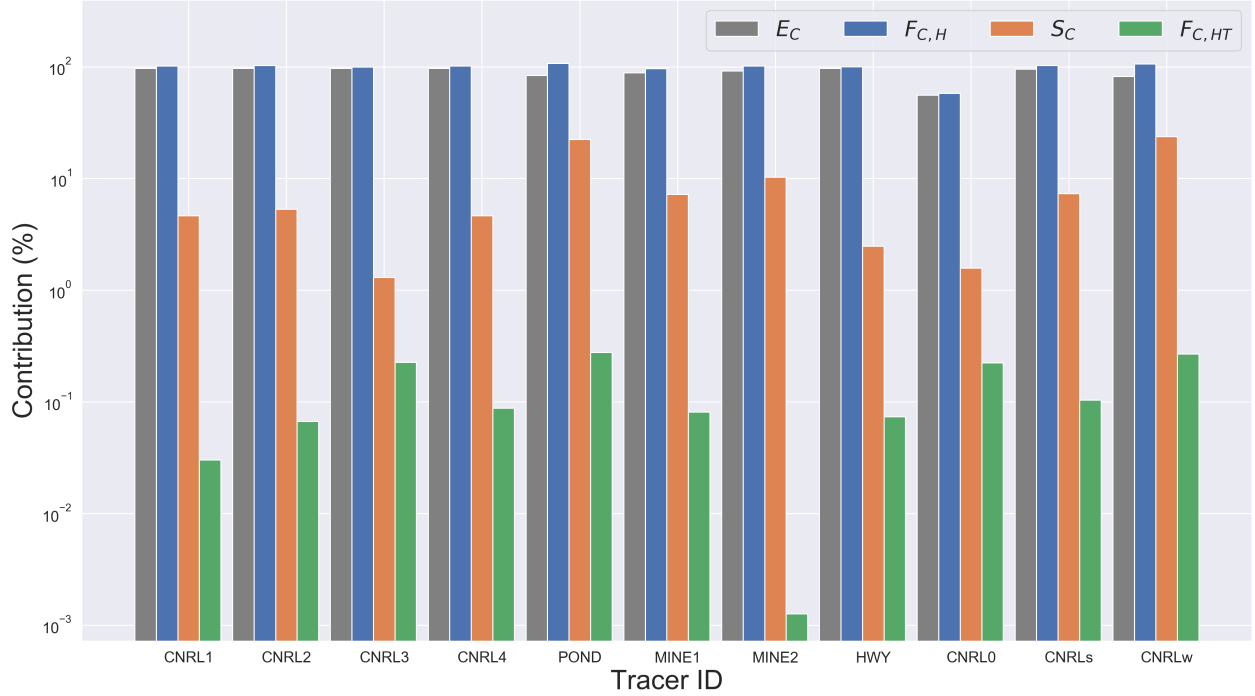


Figure 3.19: Normalized (to model input emissions - MIE) retrieved emission rates for tracer emission scenarios of case 1. Emission rates (E_C) were calculated using the mass-balance equation Eq. 3.7 and the control volume shown in Fig. 3.9. The breakdown (%) of contributing terms are shown as absolute values, see Table 3.6 for the actual values. The vertical axis is in log scale. Refer to Table 3.2 for source specifications.

emission rates (MIE). Contributions from the storage rate term S_C were negative (release of mass from the control volume) for ten emission scenarios and ranged between 1% to 30%. The contribution of S_C for MINE 2 was positive (storage of mass in the control volume) and about 4%. $F_{C,H}$ contributions were positive and ranged between 51% to 130%, which were offset by negative S_C contributions. Our results with GEM-MACH model simulations in Chapter 2 for SO_2 emissions (stack/point source) for the same time period, were very similar (97% $F_{C,H}$ and 3% S_C) to stack emissions simulated here with WRF at super-resolution. The model based study by Panitz et al. (2002) attributes 85% to 95% of the emissions to advective fluxes (for SO_2 and CO), which is also consistent with our estimates. Our results here with WRF (LES) simulations are within the range of previous model studies.

Mass-balance estimates for case 2 were within 5% of MIE for the ten emission scenarios (Tables 3.6). Estimates for CNRL 0 were biased low, similar to case 1, by about 50% (not conserved). For this case, where wind speeds were very low and spatially heterogeneous, the tracer amounts were mainly stalled within the control volume over the simulation time. Consequently, the storage rate term S_C was the dominant term with contributions (positive) ranging from 88% to 103%. The horizontal advective flux $F_{C,H}$ contributed to the mass-balance equation between 0.6% to 12%. The contributions from vertical and turbulent fluxes were less than 1%, except for CNRLw with

Table 3.6: Source emission rates (E_C) determined by performing mass-balance calculations on model output data, evaluated against model input emissions (MIE) for the three cases. Results are shown as normalized mean bias (nmb) in %. Contribution of three main terms in the mass-balance equation are shown, normalized to MIE (%). Values less than 0.1% are indicated with ϵ .

Tracer ID	Case 1				Case 2				Case 3			
	E_C	Contributions (%)			E_C	Contributions (%)			E_C	Contributions (%)		
	nmb (%)	S_C	$F_{C,H}$	$F_{C,HT}$	nmb (%)	S_C	$F_{C,H}$	$F_{C,HT}$	nmb (%)	S_C	$F_{C,H}$	$F_{C,HT}$
CNRL 1	-4.0	-2.9	99.2	-0.3	-1.7	97.7	0.7	ϵ	1.5	-27.5	127.3	1.7
CNRL 2	-3.2	-4.4	101.2	-0.1	3.3	102.7	0.6	ϵ	1.3	-26.6	126.3	1.5
CNRL 3	-3.7	-3.7	100.0	-0.1	-2.6	94.6	2.8	ϵ	1.0	-27.5	127.1	1.4
CNRL 4	-3.8	-3.9	100.2	-0.1	-3.9	92.4	3.6	ϵ	1.4	-27.2	127.0	1.6
POND	-4.4	-10.9	106.2	0.3	-4.2	92.0	3.8	ϵ	-3.8	15.4	80.6	0.2
MINE 1	-0.4	-8.6	108.4	-0.2	-1.9	96.1	2.0	ϵ	-1.2	5.1	93.3	0.3
MINE 2	3.9	-3.0	107.3	-0.3	1.6	97.4	4.2	ϵ	-0.9	5.6	92.7	0.7
HWY	-1.2	-3.0	101.8	ϵ	-4.8	90.3	4.7	0.2	-0.6	-19.2	118.3	0.3
CNRL s	-3.8	-8.7	105.0	ϵ	-4.9	90.7	4.3	ϵ	-1.4	10.0	88.0	0.6
CNRL w	-3.8	-29.8	126.2	-0.1	3.1	88.9	11.9	2.3	-3.1	72.1	24.6	0.1
CNRL 0	-50.7	-1.3	50.9	-0.3	-49.5	49.7	0.8	ϵ	-54.1	5.7	40.6	-0.4

2.3% contribution. Our estimates for SO_2 emission rates for CNRL with GEM-MACH simulations (Chapter 2) showed similar results. We note that conditions of weak advection were also observed during JOSM 2013 airborne campaign for the period of case 2 on 26 August 2013 (see Fig. 3.7). As a result of such conditions, the emission estimation flights for this time period were not included (rejected) for top-down retrievals in the relevant published work.

Mass-balance estimates for case 3 were similar to estimates for case 1, with over 96% agreement with model input emissions (MIE) for the ten emission scenarios. Similar to case 1 and case 2, CNRL 0 estimates were biased low (at 54%). For case 3, the contribution of the storage rate term S_C ranged between -27.5% to 72.1%. The horizontal advective flux $F_{C,H}$ contributed between 25% to 127% to the mass-balance equation. The horizontal turbulent flux contributed between -0.4% to 1.7%. Similar to case 1 and case 2, contribution of vertical fluxes were negligible.

As can be seen from Table 3.6, our mass-balance estimates for the three cases were within 5% of model input emissions (MIE). Estimates were partially affected by local mass deficit/surplus for these cases. As discussed at the beginning of this section, the turbulent diffusion step in our WRF model setup created erroneous negative mass (locally) within the modelling domain. The negative mass was mainly created near the emission source (upwind) and at plume edges, as shown in Figure 3.20. The advection of positive/negative mass passed the box-downwind-wall (vertical dashed line in Fig. 3.20) was not always balanced, due to different spatial distributions for positive and negative amounts. Fig. 3.20 shows three snapshots at 5 min intervals for tracer MINE 1 during the final hour of case 1. As indicated on the figure, during this time period 98% of the domain total negative mass remained inside the control volume along with only about 43% of the domain total positive mass. This resulted in a mismatch between the negative mass and the excess positive

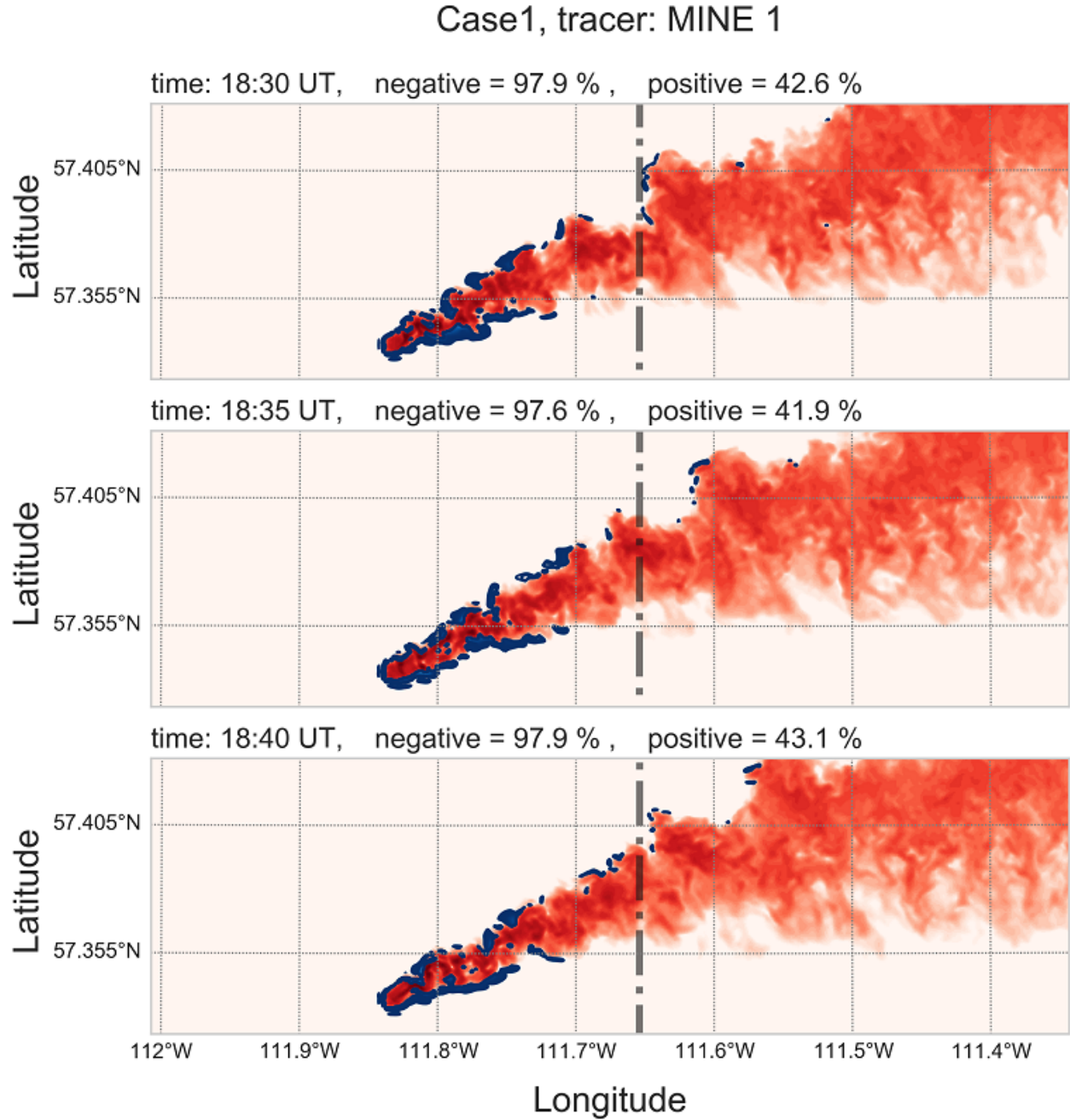


Figure 3.20: Three panel snapshot of MINE 1 plume during case 1 at 5 min intervals. Note that the erroneous negative mass creation (blue contours) is concentrated upwind of the emission source and on plume edges. Consequently, 98% of the domain total negative mass was located within the control volume along with only 43% of total positive mass (red). The vertical dashed line shows the location of control volume downwind wall.

mass and the consequent mass emission rate underestimations for this and similar cases. Estimates were also partially affected by the changing vertical grid spacing for upper model layers, for tracer amounts mixing to higher altitudes (> 2.5 km). This is similar to mass loss for CNRL 0 emissions,

but to a much lesser extent. We conclude that all three cases (with ten emission scenarios) were globally (on the entire modelling domain) and locally (control volume) mass conserved within 4% and 5% of MIE, respectively.

3.4 Conclusions

We developed and implemented super-resolution (< 100 m) model simulations employing the WRF-ARW atmospheric modelling system. We used NARR reanalysis data at 31 km resolution as Initial and Boundary conditions for our coarsest domain at the same resolution. Dynamical down-scaling of reanalysis data was done through numerical model nesting with five domains from 31 km to 50 m at the finest resolution domain over the oil sands facility CNRL. We chose our model simulation times and locations from three emission rate retrieval flights during the JOSM field campaign in August and September 2013. We performed model simulations for three days in August and September, representing different meteorological conditions. Model simulations for the finest resolution domain were conducted using Large Eddy Simulation (LES) sub-grid parameterization. The main objective was to model the state of the atmosphere at high enough resolution to simulate atmospheric dynamical processes at spatial and temporal scales of airborne measurements. Model output data from our simulation cases were evaluated against historical observational data from two WBEA monitoring stations, as well as aircraft observations during 2013 JOSM campaign for the same locations and time periods. Model output fields showed good agreement with observational data within 5% in terms of 2-m temperature, 2-m relative humidity, and 10-m winds. General wind directions generated by the super-resolution model were within 20 to 30 degrees of the observational data. Model wind speeds were in agreement with aircraft observations within one standard deviation.

We modified WRF model dynamical solver source code to simulate passive tracer emissions within the finest modelling domain, centred over the CNRL facility. Our case simulations included eleven different emission scenarios with stack/point and surface/area sources. Continuous emissions were simulated for time periods of 2-3 hours during three different days. We evaluated the atmospheric dispersion and transport of tracer plumes from the eleven emission scenarios under different meteorological conditions. Unstable atmospheric conditions, low wind speeds, and strong vertical wind shear during our case study on 26 August 2013 resulted in weak advection of tracer plumes during the simulation time. During case studies on 20 August and 2 September, atmospheric conditions were relatively more stable with higher wind speeds of about 5-15 m/s. Tracer plumes from emission sources were advected mainly towards east and north-east for these two cases. Similar conditions have been observed during the JOSM 2013 field campaign for the same time periods and locations as our three case studies.

We evaluated the performance of our model simulations in terms of global (over the entire domain) mass conservation. For one case out of eleven, creation of erroneous mass by the model transport step resulted in loss of tracer mass of about 30%. For this emission case, tracer release

was placed at the model vertical level where model vertical resolution transitions from the super-resolution of about 10 m to progressively coarser resolutions. The unbalanced creation of erroneous mass at sharp concentration gradients, such as at the vicinity of point sources, is intensified on an irregular grid (an artifact of model transport scheme). Therefore, it is recommended that tracer emissions to be assigned to model grid points with regular grid spacing for several adjacent model cells along the x , y , z directions. The positive-definite re-normalization scheme in our modelling setup was successful in preventing such nonphysical effects for 10 out of 11 emission sources (assigned on regular grid). The rate of change for tracer mass was calculated by integrating over the entire modelling domain and differentiating with respect to time. Results were within 2-4% of model input emissions (MIE) for 10 out of 11 emission scenarios, indicating global mass conservation in our simulation cases.

We further investigated local mass conservation by mass-balance calculations over a sub-domain (control volume). Mass-balance calculations were conducted by considering a control volume (box) enclosing all emission sources. We estimated within-box-source emission rates by calculating the net exiting flux through box top and lateral walls, and the temporal rate of change in tracer mass within the control volume. Under normal advective conditions (> 5 m/s wind speeds), the horizontal advective flux was equal to over 90% of the emission rate while turbulent fluxes were less than 2%. The remaining contribution came from the storage rate term, the release/accumulation rate of tracer mass within the control volume. Our mass-balance estimates were within 5% of MIE for tracer release heights below 200 m agl for our three case studies (30 emission scenarios), indicating local mass conservation for our super-resolution WRF simulations.

Except for one hypothetical emission source, our results for various tracer emission scenarios under different meteorological conditions were globally and locally mass conserved within 4% and 5% of model input emission rates, respectively. In Chapter 4 we use model output from this super-resolution WRF simulation for the model-based study of airborne top-down source emission rate retrievals. We evaluate the conventional methods and propose improved approaches for aircraft-based retrievals.

Chapter 4

Optimizing Top-down Emission Rate Retrievals

4.1 Introduction

During airborne field studies for top-down retrieval of source emission rates, the environmental fields (e.g., meteorology, air quality) are sampled around and/or downwind emission sources. Measurements are made over time and through the 3D space using aircraft-based remote and/or in-situ observations (e.g., Aggarwal et al., 2018; Peischl et al., 2010). The time and spatial resolutions of such measurements are determined by the sampling frequency and range of the measuring instrument, the speed of the sampling platform, the sampling path and geographical locations (Conley et al., 2017; Gordon et al., 2015). The sparse spatial measurements that are made over time are processed and analyzed according to various assumptions regarding how representative they are of the mean and real-time environmental conditions (e.g., wind field, emissions: Alfieri et al., 2010). The post-processed data are then used for estimating emission rates from sources of pollution (Gordon et al., 2015; Karion et al., 2015; Ryoo et al., 2019).

Regardless of the measurement approach, the spatial heterogeneity and temporal variability of meteorology and concentration fields can result in large uncertainties in top-down estimates. Previous studies have attributed large uncertainties between 20% to 40% in estimates to the gap of information (spatial and temporal) in the sampled data (e.g., Angevine et al., 2020). For instance, airborne measurements are commonly made at elevations above 150 m agl for safety considerations. Gordon et al. (2015) identified the unsampled region below the lowest flight level as a large source of uncertainty in mass-balance analysis (e.g., up to 26% for CH₄ plumes). The gap of information in the sampled data, due to limitations on spatio-temporal resolution and range of the sampling method, can be partially filled by combining data from different measurement platforms. For instance, airborne samplings (aircraft, UAV) can be complimented by ground-base measurements (Bell et al., 2021; Brus et al., 2021b; Islam et al., 2021). Fixed location in-situ measurement techniques of meteorology and tracer concentrations include tower measurements at heights of up

to 350 m agl (Andreae et al., 2015; Heintzenberg et al., 2011), and radiosonde (tethered/balloon) measurements include heights up to 2 km agl (Nambiar et al., 2020; Nygård et al., 2017), which improve on measurement range over tower measurements. Ground-based remote measurements (e.g. Doppler Lidar - DL) can provide profiles at higher vertical range and temporal resolution, compared to tower and radiosondes (Bell et al., 2021; Islam et al., 2021). Ground-based remote-sensing can also be conducted from mobile surface land vehicles (Davis et al., 2019; de Boer et al., 2021), generating column measurements at higher spatial (horizontal) resolution. Remote-sensing datasets can be analyzed in conjunction with airborne measurements for both validation (Davis et al., 2020) and as complementary information (Brus et al., 2021a) in air quality studies.

Taking into account the logistical/operational challenges of synchronizing multi-platform measurements (ground-based and airborne, stationary and mobile), the benefits of combined in-situ and remote measurements on the same mobile platform (e.g., aircraft) can be considered. For example, Aggarwal et al. (2018) conducted airborne lidar measurements over Canadian oil sands during the time period of in-situ airborne measurements on board another (Convair) aircraft (JOSM, 2013) where laser eye safety for pilots in the Convair aircraft was a limiting factor for collocated measurements. Airborne remote measurements of pollutants (e.g., Aggarwal et al., 2018; Haarig et al., 2019) and wind fields (e.g., Bilbro et al., 1984; Serafin et al., 1986; Wekker et al., 2012) have been successfully conducted in the past. Therefore, combining in-situ measurements with remote-sensing on the same mobile platform (aircraft) is tested, feasible, and can: (a) fill the information gap during airborne in-situ sampling, and (b) potentially reduce the operational cost of top-down retrievals. As we discuss later, while traditional in-situ airborne measurements require several hours of flight time depending on the size of the area under study (which could be urban or industrial sources), combined in-situ/remote measurements have the potential to reduce the flight time to several minutes. Here we emphasize complementing in-situ observations by remote measurements, not replacing them, noting that the same level of accuracy and capacity for measuring different aerosol and tracer concentrations using aircraft in-situ sampling may not be available with remote measurements (Aggarwal et al., 2018).

The anthropogenic sources of pollution include large emitting stacks in industrial complexes such as the Canadian oil sands (Li et al., 2017; Liggio et al., 2019). These kind of operations (in addition to stack emissions) include roads that connect different operations, surface sources of pollution that span over a large area such as surface mine excavation sites, and larger area sources such as tailing ponds (Baray et al., 2018; Davis et al., 2020). Various emission scenarios (e.g., point, area sources) and tracer dispersion and transport under different meteorological conditions were considered in the WRF model simulations described in Chapter 3. Here, we use the output data from super-resolution (with LES parameterization) WRF modelling as described in Chapter 3 as a proxy for real-world environmental fields (virtual sampling). The range of spatial and temporal variability in fields sampled by a mobile platform for top-down retrievals can impact the accuracy of the estimates. For example, the spatio-temporal variability in sampled fields is dependent on the downwind distance and hence investigating the optimized sampling radius using model data

is desirable and can provide valuable advice in terms of observational flight planning and data processing. By studying the output fields from several different WRF simulation scenarios, we investigate the impact of different sampling strategies (e.g. the 3D flight paths, combined in-situ and remote measurements) on the accuracy of top-down estimates and provide operational recommendations for general and specific cases.

4.2 Methods

The super-resolution WRF model output data from Chapter 3 span a geographical location over the northwest portion of Athabasca oil sands region, Alberta, Canada (see Figure 3.2). The 4D (3D space, 1D time) output data from the WRF simulations are on three different days in August and September of 2013, which overlap the period of the JOSM 2013 field campaign over the oil sands region (JOSM, 2013). More specifically, our chosen time periods overlap with three aircraft emission retrieval flights over the CNRL facility in the northwest area of the oil sands region. In this study, we simulate aircraft measurements around and downwind of the CNRL facility using the model output data as proxy for real-world environmental fields (virtual sampling). To achieve this, we extract data from the model 4D fields for time periods and along flight paths similar to those conducted during airborne field campaigns. The super-resolution ($\Delta x = 50$ m, $\Delta z \simeq 10$ m, $\Delta t < 1$ s) of our model-generated atmospheric fields allow us to sample data at temporal and spatial scales of airborne measurements without the need for interpolation of model generated data. Our super-resolution model simulations were also capable of resolving atmospheric dynamical processes (turbulence) at airborne sampling scales, and therefore provide us with ample information for investigating the impact of spatio-temporal variability in sampled fields on the accuracy of top-down estimates. We analyze the model-generated data as if they were observational data to retrieve emission rates from various sources within the facility. We then evaluate these estimates against the known modelling conditions (i.e. emission rates, meteorology) to investigate the sources of error using the data that were sparsely sampled in space and over time. By quantifying the contribution of different sources of uncertainty, we provide recommendations for improved data post processing and analysis. We also, introduce recommendations for operational methods during field campaigns that combine in-situ and remote measurements and have the potential to increase the accuracy of estimates at acceptable operational costs. All of this is done by taking advantage of model output data from super-resolution simulations that can resolve the scale of turbulent dynamical processes (atmospheric transport and dispersion) which affect aircraft-based estimates. We start with a short discussion on the application of the mass-balance technique using various sampling control volumes/surfaces. Next we discuss the conventional observational methods where in-situ aircraft-based measurements are used for top-down retrievals. Finally we discuss a new measurement methodology where aircraft-based in-situ and remote measurements are used in a complementary manner for high accuracy top-down retrievals.

4.2.1 Mass-Balance technique: Control Volumes and Downwind Screens

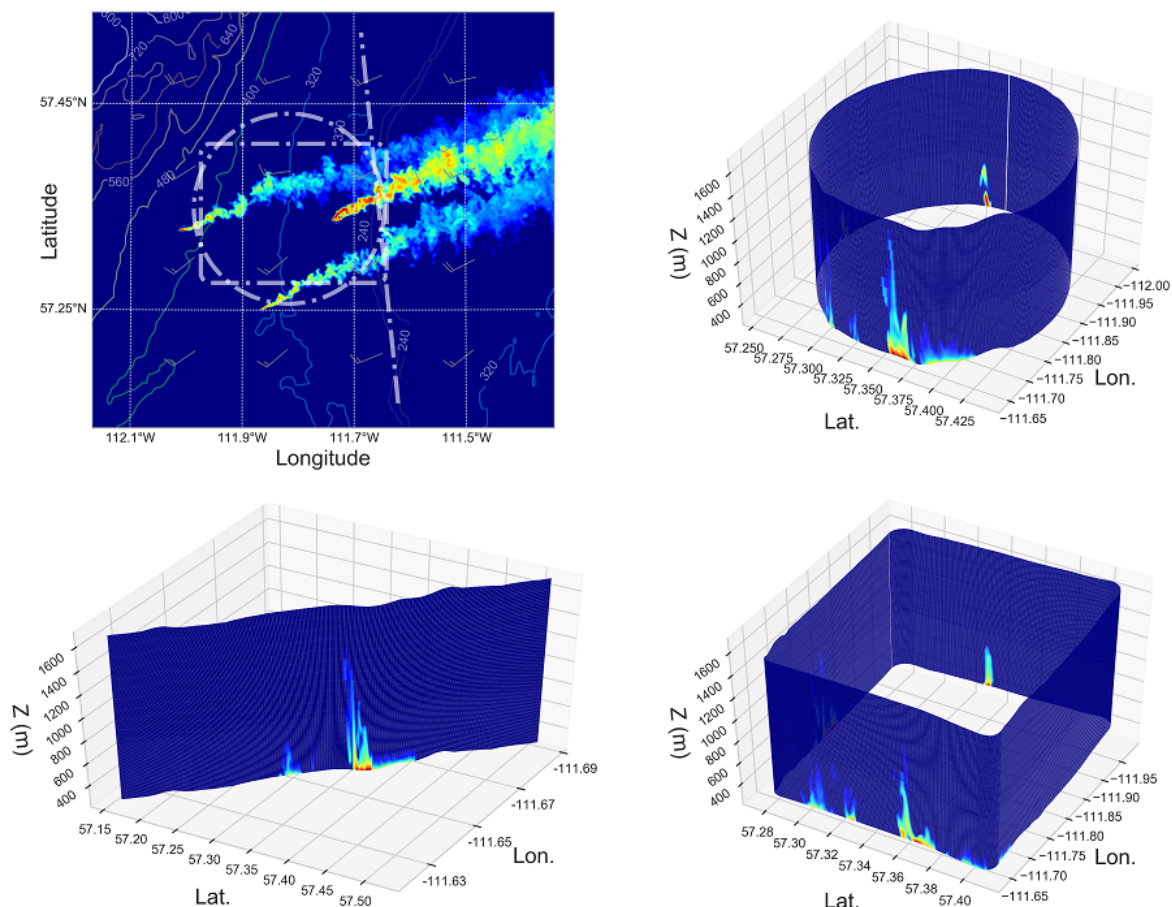


Figure 4.1: Top view of the modelling domain with tracer plumes from point/stack emission sources within the CNRL facility and two upwind sources (top left). Sampling volumes/surface for rectangular box, cylindrical box and downwind screen (at 5 km distance) are also superimposed on the map with dashed lines. The 3D projections of the two boxes (control volumes) and the downwind screen are also shown in different panels of the figure, viewed from north-east. Note that the vertical scale is exaggerated, sampling volumes/surfaces extend tens of kilometres in the horizontal and about 2 km in the vertical.

Airborne top-down emission rate retrievals are usually accomplished by flying at a distance downwind of the emission source and at various altitude levels. The measured data are then converted into source emission rates using various algorithms. One of the most commonly used methods is to calculate the downwind mass flux of desired species by combining the tracer/particle concentrations and the wind speed data (e.g., Karion et al., 2015; Peischl et al., 2010). The rate of advective mass transport in this calculation is then equated to the emission rates of the source of interest. In the presence of upwind sources of emission, upwind fluxes must be subtracted from the mass flux of the source of interest. In this situation it is common to combine the downwind measurements with lateral and upwind measurements (a closed shape around the emission source).

In the case of airborne measurements, the aircraft would execute what is referred to as an emission flight or box flight (see Fathi et al., 2021; Gordon et al., 2015). During a box flight, the aircraft would fly in horizontal closed loops (e.g., rectangular, cylindrical) at several altitude levels, to create a control volume (box) around the emission source. The net flux exiting the lateral walls of this control volume is then factored in estimating emission rates. See Chapter 2 for detailed discussion on the application of the mass-balance technique and the divergence theorem in aircraft-based top down retrievals. In this chapter we consider three distinct sampling volumes/surfaces: rectangular box, cylindrical box, and downwind-screen (at various distances). Note that throughout this work we use “box” and control-volume interchangeably. Figure 4.1 shows the top view of the modelling domain with the spatial extent of the flux boxes and the downwind screen at 5 km distance superimposed with dashed lines (top left). The figure also shows the 3D projections of the flux boxes/screens that extend in the vertical direction. Plumes from stack emission sources from our WRF simulations are also shown in Figure 4.1, which includes major stacks within the CNRL facility (CRNL 1-4) and two upwind sources (outside the control volumes) on the west and south of the facility. We extracted data from 4D model output fields of meteorology and tracer concentrations for control volumes and along downwind screens. In order to calculate source emission rates from the extracted fields, we apply the mass-balance technique and the divergence theorem as described with the following equation,

$$E_C = S_C + F_{C,out} \quad (4.1)$$

where the integrated emission rate E_C (of tracer C) from sources within the control volume (or upwind the sampling screen) are equated to net flux out ($F_{C,out} = F_{C,H} + F_{C,V} + F_{C,HT} + F_{C,VT}$, see Chapter 3), and the storage/accumulation rate S_C of tracer mass within the control volume (or upwind of the screen). As the emission source emits tracer mass into the control volume at rate E_C , some of it will accumulate within the control volume (box) at storage rate S_C and some will exit the box through lateral walls at mass flux rate $F_{C,out}$. The balance between these three terms constitutes the mass-balance equation. Knowledge of two (e.g., through direct measurements), can be used for determining the third. For our purposes, we assume E_C is unknown. We investigate how well $F_{C,out}$ and S_C can be determined using observational methods. Then we use the observationally determined $F_{C,out}$ and S_C contributions to estimate E_C , which is then evaluated against the known model input emission rates and fluxes. Here we use “observational” to mean virtually sampled from the model output fields.

We extract instantaneous data for flux boxes/screens from model output data and perform mass balancing to retrieve source emission rates at 1 second frequency. The performance of the model in terms of mass conservation and transport are discussed in Chapter 3. Throughout this chapter, we consider the model simulated environmental fields and tracer mass emission rates that are retrieved from the instantaneous control volumes/surfaces as the “Truth”, and evaluate the estimates based on the other sampling strategies (e.g. Flying point, Rolling column) against this Truth.

4.2.2 Flying Point (FP) - aircraft-based in-situ measurements

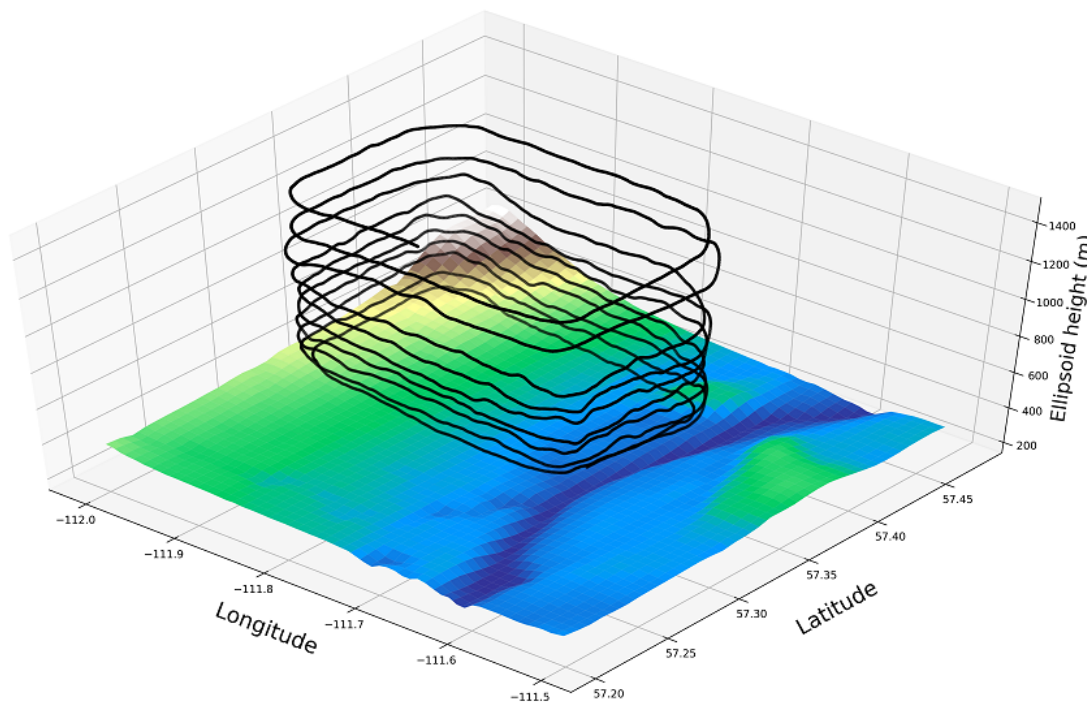


Figure 4.2: JOSM box flight on 2 September 2013 over CNRL. Topography of the region is shown as a 3D surface. Flight path is shown with dark dots. Flight started on the east side of CNRL near Athabasca river basin (shown in blue) in counterclockwise direction in a rectangular path, while gradually climbing in altitude. For our purposes the aircraft can be considered as a dimensionless point, and therefore herein we refer to this sampling strategy as Flying Point (FP).

As discussed earlier, during a typical box flight the aircraft would fly in closed loops at multiple altitude levels to trace a control volume enclosing the emission source. Figure 4.2 shows the JOSM 2013 flight path on 2 September 2013 over CNRL (corresponding to our case 3). For this box flight the aircraft started close to the ground on the east side of CNRL (close to Athabasca river basin), following a rectangular path counterclockwise while climbing to higher altitudes after each circuit up to about 1500 m. The aircraft sampled meteorological fields and tracer concentrations along the flight path. At any point in time the aircraft made measurements at a single point in space and never returned to the same point. Therefore, this airborne measurement approach is considered as a flying (sampling) point in our analysis. Flight paths can start near the surface or at the top, and progress in clockwise or counterclockwise direction. The environmental fields that are sampled along the flight path constantly change over time, wind direction can vary during the sampling time and the tracer plume can rise or descend to different altitude levels. The sampled points along the flight path each represent conditions at a single space-time point. These points are separated in space-time according to flight speed and pattern. In this context, the flight path determines the degree at which point measurements represent the mean and/or real-time conditions at other points in time and space. For example, for a flight path starting near the surface and climbing in

altitude while sampling a tracer plume that is also rising during the flight time, oversampling of the plume can result in overestimation of source emission rates. While the opposite can result in underestimation.

In this strategy, environmental fields are sampled somewhat sparsely in space and time (4D) along the lateral walls of the box. In conventional aircraft-based retrievals, to calculate the net flux out of the control volume $F_{C,out}$, the data collected along the flight path is projected on a 2D screen comprised of box lateral walls and interpolated (e.g., Gordon et al., 2015; Liggió et al., 2019; Ryoo et al., 2019). One of the most commonly used interpolation methods for airborne measurements is the Kriging method (e.g., DiBiase and Wentz, 2006; Goovaerts, 2019; Hohn, 1991). Sampled data are interpolated between consecutive points and flight tracks at different altitude levels, the interpolated data on the 2D screen are then extrapolated to the ground surface at unsampled altitudes below the lowest flight track. The method of data extrapolation is chosen according to assumptions regarding vertical mixing of tracer concentrations (e.g., uniform/constant, exponential, Gaussian) and meteorological profiles (e.g., linear, logarithmic — see Gordon et al., 2015). Vertical mixing of tracer concentrations depend on various factors including emission height, downwind distance and atmospheric stability conditions. The interpolation/extrapolation is commonly done in space only, under the assumption of steady state condition. However, atmospheric fields can be highly variable in time periods of several hours (typical emission estimation flight time), which in turn can effect aircraft-based top-down retrievals.

In this chapter we explore these effects and the possible sources of uncertainty in top-down retrievals, by making use of output fields from our super-resolution WRF model simulations capable of simulating similar spatio-temporal variability in environmental fields. We simulate aircraft-based measurements by extracting data along flight paths in 3D space and over time, similar to that shown in Figure 4.2. Herein we refer to this form of extracting model data as the sampling strategy of Flying Point (FP). We then interpolate the sampled data on a 2D screen using Kriging and examine different extrapolation methods. We use the 2D data to calculate the tracer mass flux out of the control volume (or through the downwind screen) $F_{C,out}$, mimicking aircraft-based retrievals. We note that for estimates using a downwind screen, where no upwind sampling is conducted, tracer concentrations are background subtracted. Background levels are usually estimated from measurements on screen edges, far from plume concentrations (e.g., Peischl et al., 2010). Note that in the mass-balance equation Eq. 4.1 there is an additional term, the storage rate S_C , which is not possible to estimate in conventional aircraft-based retrievals. As can be seen from Fig. 4.2, airborne measurements are performed on the lateral walls of the control volume (2D). While for estimating S_C volumetric data over time is required (4D). Therefore the lack of an estimate of S_C emerges as another major source of error in top-down retrievals when the contribution of this term in the mass-balance equation (Eq. 4.1) is significant. This concept was discussed extensively in Chapter 2 in the context of mass-balance applications, and will be considered in the following discussions in the context of observation-based estimates.

4.2.3 Rolling Column (RC) - aircraft-based remote measurements

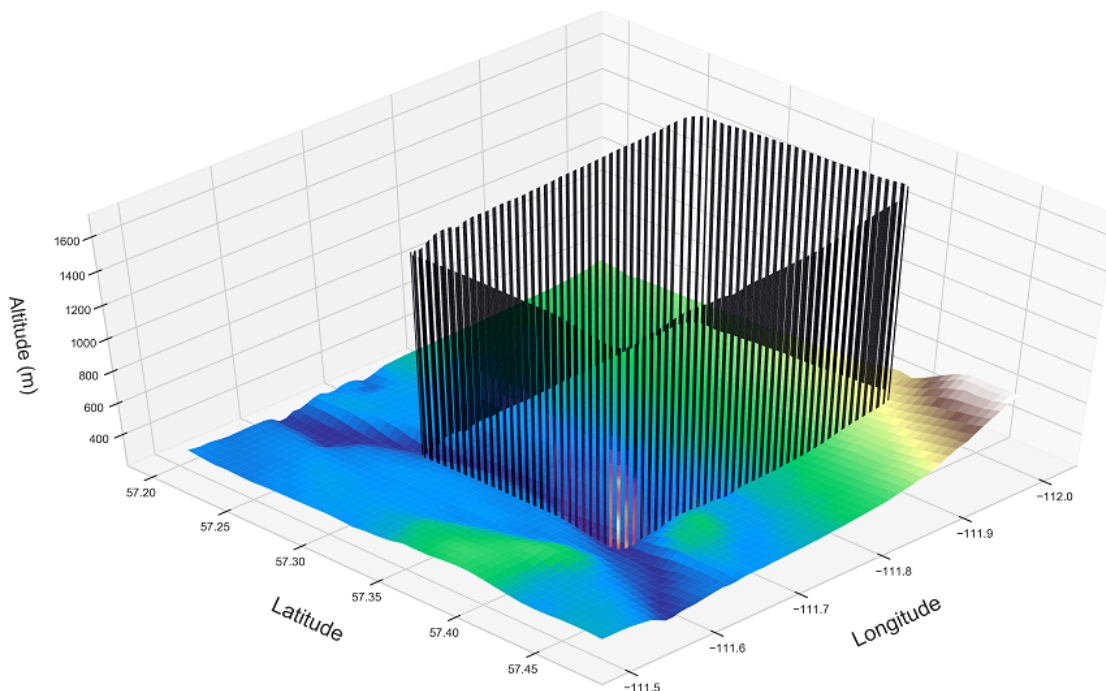


Figure 4.3: Rolling column sampling strategy includes extracting/sampling vertical columns of environmental data around (downwind of) the emission source. During field campaigns this can be achieved by using aircraft-based remote measurement instruments such as LIDAR. By making column measurements, environmental fields will be sampled from the ground surface to the box top, eliminating the need for data extrapolations. This method can generate complete control volumes/screen at shorter time periods of about 10-15 min.

As discussed earlier, the lack of data below the lowest flight level is a major source of uncertainty in aircraft-based top-down retrievals. Remote vertical profiling for heights below the flight track is a potential solution for filling this gap of information. Additionally, aircraft-based remote vertical profiling at all flight levels would provide ancillary data for retrievals that are not available from in-situ aircraft-based measurements alone — i.e. via Flying Point (FP) sampling. An aircraft equipped with remote vertical profiling instruments (down and/or up facing instruments) such as lidar, can produce column measurements at horizontal locations along the flight path. For example, Howell et al. (2014) and Aggarwal et al. (2018) conducted airborne lidar measurements of tracer gas and aerosols over Canadian oil sands region. Further, Doppler lidars are capable of making wind vector measurements based on the optical Doppler effect (e.g., Bilbro et al., 1984; Serafin et al., 1986; Wekker et al., 2012). Side-door-mounted lidar scanner beams can be oriented in different scanning directions to profile environmental fields above, on, and below the flight level (Wekker et al., 2012). While among the most effective remote-sensing techniques and provide high resolution observations, lidars have several limitations in sampling environmental fields. For instance:

1. Depending on the scanning pattern (e.g., velocity azimuth display - VAD, NADIR-stare),

airborne lidar measurements might be at coarser horizontal resolutions compared to in-situ sampling (e.g., at ~ 1 km: Weissmann et al., 2005; Wekker et al., 2012),

2. The back scatter of the emitted lidar pulse from the ground surface, can obscure the aerosol signal near the surface (e.g., below 50 m agl) and also make near surface wind field retrievals challenging and error prone (Godwin et al., 2012),
3. Although differential absorption lidar (DIAL) instruments are capable of sampling mass concentration profiles of several tracers and aerosols, they lack the capacity and confidence level of aircraft-based in-situ instruments in sampling concentrations (Haarig et al., 2019).

Despite the shortcomings, lidars are effective in profiling concentrations and trends. Lidar vertical profiles can be calibrated with co-located in-situ measurements to characterize emissions (e.g., Aggarwal et al., 2018; Howell et al., 2014). The corrected/calibrated column data of tracer concentrations and wind vectors can then be used for mass-balance analysis.

Full box flights around large industrial complexes take 1-3 hours to complete. During such time period the environment under study can evolve dramatically, which can result in increased uncertainty in retrievals based on measurements along the flight path. By making column measurements along the horizontal path, the entire vertical extent of the flux box can be sampled at every circuit/lap around the source which takes only 10-15 min to complete. Herein we refer to this sampling strategy as the Rolling Column (RC). RC can be achieved by an aircraft flying at a single height around the emission source while making column measurements. This method can be combined with in-situ measurements for a flight path that also climbs to higher altitudes. This approach would provide complementary in-situ and remote measurements that would address some of the shortcomings of the FP approach by filling the information gap created by lack of sampling between and outside (above and below) the flight tracks, and by generating complete flux boxes at shorter time periods (10-15 min as opposed to 1-3 hours), which is desirable when investigating a system that is evolving over time. In this chapter, this concept is presented for the first time in the context of aircraft-based emission rate retrievals. To explore the application of RC, we extract vertical profiles along box lateral boundaries (and along the downwind screen) from the ground surface to the box top (well above the altitude at which any considerable tracer concentrations are detectable) as shown in Figure 4.3. Note that the operational limitations of remote column measurements (see above) are not simulated here. Vertical profiles/columns are extracted over time at typical aircraft speeds (e.g., 100 m/s) around the source of interest, completing a circuit every 10-15 min, generating a complete sampling screen for the same period. We denote the time period for one full circuit/lap around and/or downwind the source as t_i where $i = 1, 2, 3, \dots, n$ indicates the lap number. Extracted/sampled columns of data are then interpolated in space (simple linear interpolation) and used for calculating the net flux out $F_{C,out}$. This strategy has the potential to generate several 10-15 min (t_i) estimates over a 1-3 hour flight time with the following benefits over the FP strategy:

1. estimates with the assumption of steady state condition are made for shorter periods of 10-15 min, reducing the uncertainty arising from temporal evolution of the system,
2. columns of data are sampled/extracted for the entire vertical extent of the box, no need for interpolating/extrapolating in the vertical, only interpolating in the horizontal between tightly spaced columns of data, further reducing the uncertainty arising from spatial interpolation of data collected over time,
3. mass-balance emission rate estimations can be achieved via shorter flights (under an hour of flight time), potentially reducing the operational cost of airborne top-down retrievals.
4. ancillary data is generated for studying the temporal evolution of the system (e.g., concentrations, meteorology) which would enable the estimation of the storage rate term S_C using observational data for the first time.

As discussed in Chapter 2 (Fathi et al., 2021), storage-and-release events can occur due to variations in source emission rates, localized changes in atmospheric stability, and change in the direction of transport (wind direction). During such events, the contribution of the storage rate term S_C becomes significant enough to impact the accuracy of mass-balance estimates. Hence it is desirable to estimate S_C using observational data, as we demonstrate in the following. S_C can be estimated based on the temporal changes in sampled tracer concentration levels. As we discussed in Chapter 2, Section 2.3.3 (see Fig. 2.6), in the absence of volumetric time-series data S_C can be estimated from time-consecutive screens of concentration (Fathi et al., 2021). Here we follow the same approach using RC generated time-consecutive screens. Full box/screen of tracer concentration $\chi_C(t, z, s)$ can be generated from RC sampling for each lap time period t_i of 10-15 min, which we will denote as ψ_i ($i = 1, 2, 3, \dots, n$ indicates the lap number). $\psi_i(z, s)$ is then averaged along the horizontal path s and integrated for the control volume/screen as

$$\mu_i = A \int \overline{\psi_i}(z) dz, \quad (4.2)$$

where A is the base area of the control volume (flux box). Differentiating μ_i with respect to time t_i using consecutive RC screens data yield,

$$S_{C,i} = \frac{\mu_i - \mu_{i-1}}{t_i} \quad (4.3)$$

$S_{C,i}$ provides an estimate for the rate of storage between time-consecutive screens i and $i - 1$.

Box (control volume) dimensions are readily available for the box flights (rectangular and cylindrical in this work). However when a single-leg (one-sided) downwind screen is used, A is not clearly defined. To address this problem we consider an effective volume for each emission source, which is defined as a triangle with base equal to the horizontal extent of the downwind screen and height equal to the downwind distance from the emission source to the downwind screen where sampling

takes place. Note that in this approach, tracer concentrations are assumed to be at background levels on the upwind walls of this virtual/effective control-volume.

In the following, we first derive S_C , $F_{C,out}$ and E_C directly from model 4D output fields for each sampling box/screen. We will refer to these estimates, determined through 4D mass-balance calculations, as model “Truth” for each set (emission scenario, meteorology, sampling box/screen). We then evaluate Flying Point (FP) and Rolling Column (RC) estimates against this Truth.

4.3 Results and Discussions

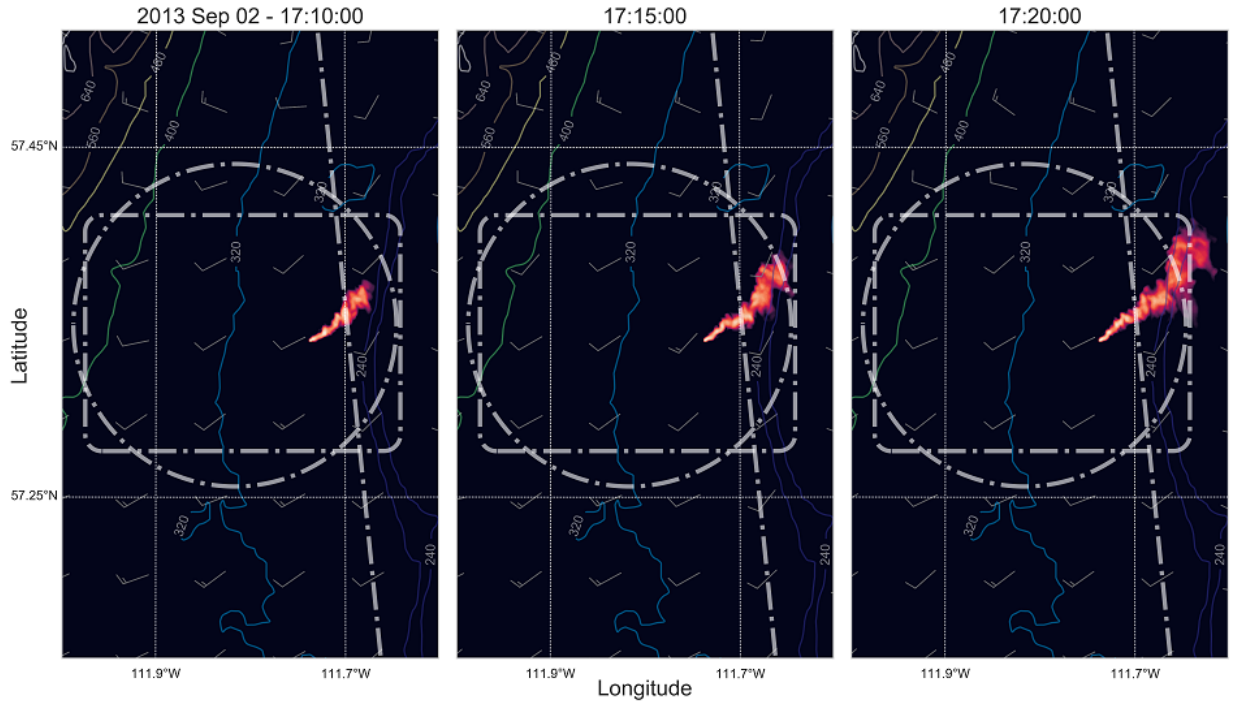


Figure 4.4: Transport and dispersion of tracer plume released from the CNRL4 source is demonstrated in three panels. Tracer is released at 17:00UT, it reaches the downwind screen at 17:05UT, the cylindrical box at 17:15UT and the boundaries of the rectangular box at 17:20UT. For reference, the rectangular-box dimensions are 20 km $\times$ 15 km. The spatial heterogeneity of the 10-m wind field is shown with wind barbs (flag: half=5m/s, full=10m/s). Note how the plume spreads as it is advected downwind and how the advection path (plume centre line) is altered along the way, as it interacts with the topography (passing over the Athabasca river basin at about 240 meter above sea level) and the heterogeneous wind field.

Figure 4.4 demonstrates the transport and dispersion of a test tracer (CNRL 4) during our third simulation case on 2 September 2013. We considered two flux boxes: the rectangular box and the cylindrical box, and several downwind screens: at 2.5 km to 20 km downwind distances. The two boxes and the screen at 2.5 km are shown with dashed curves in Fig. 4.4. These control volumes/screens are used for mass-balance calculations as described by Eq. 4.1. For the case shown in Fig. 4.4, the tracer is released within the control volume at about 17:00UT and is

advected downwind until it reaches the first sampling point at about 3 km distance, the downwind screen, around 17:05UT. Up to this point the emission rate is equal to the mass buildup upwind the screen ($E_C = S_C$) as shown in Figure 4.5 top panel; after this point as the mass starts to pass through the screen, the flux out $F_{C,out}$ starts increasing while S_C decreases. The sum of S_C and $F_{C,out}$ amount to the net emission rate of E_C . Note that as discussed in Chapter 3, mass-balance determined E_C for our simulated cases were within 5% of model input emission rates (MIE). As shown in Fig. 4.4, the plume reaches the cylindrical box at about 17:15UT, and the rectangular box at about 17:20UT. The time scale for the initial storage and the eventual release of tracer mass depends on the box size and the distance of the downwind leg/wall from the emission point and the speed of the flow of air carrying the tracer amounts (i.e. wind speed). In this case the exit point on the downwind-screen is at 3 km distance, on the cylindrical box at 6 km, and on the rectangular box at 7 km distance as can be seen from the top view map in Fig. 4.4 and the time-series (signal) graph in Fig. 4.5. Note that the advection path is diagonal with both west-east and south-north components. Therefore the downwind distance to the point of interception by the sampling screen/box is determined by the wind direction. With a spatially fixed screen/box, the downwind distance can vary with change in the direction of advection. Further, the wind speed along the advection path will determine the advection time from the point of emission to the point of sampling. The advection path can be a straight line, from source to the point of detection, at the altitude level at which the tracer was released, or at the final plume rise height for buoyant plumes. However, the vertical motion of the air flow along the advection path due to the presence of updrafts and/or local variation in atmospheric stability can increase the distance the tracer will travel before arriving at the point of sampling. The vertical motion of the flow can displace the path of advection between different flow regimes (different wind speeds) at different altitude levels, which in turn contribute to change in advection time. All of these processes contribute to turbulent transport of emitted tracers and the oscillatory behaviour in $F_{C,out}$ and S_C signals as can be seen in Figure 4.5. Our WRF simulations at super-resolution ($\Delta x = 50$ m, $\Delta z = 10$ m, $\Delta t < 1$ s) were successful in simulating this turbulent behaviour. This concept is integral to the topics that will be explored in this chapter, as we explore the uncertainty in emission rate retrievals arising from the time varying environmental fields and the dynamically turbulent tracer transport.

Note the oscillatory trends in the time-series shown in Fig. 4.5. The emitted mass from source(s) within the control volume, constantly accumulates in the box (upwind of the screen) and then is released through the downwind wall. This results in the observed oscillations in the $F_{C,out}$ and S_C signals. Figure 4.5 shows that the period (peak to peak) of the oscillation can depend on the downwind distance of the interception point (box dimensions). For instance, the signals for the downwind screen at 3 km distance (Fig 4.5 top panel) have shorter oscillation periods compared to cylindrical and rectangular boxes at longer downwind distances. Due to the fact that S_C cannot be estimated directly in real-life measurement campaigns (i.e., Flying Point sampling - FP), top-down estimates mainly rely on calculating the flux out $F_{C,out}$ directly from sampled fields on box lateral walls. Given that the signal for $F_{C,out}$ is oscillatory over time, FP sampling can

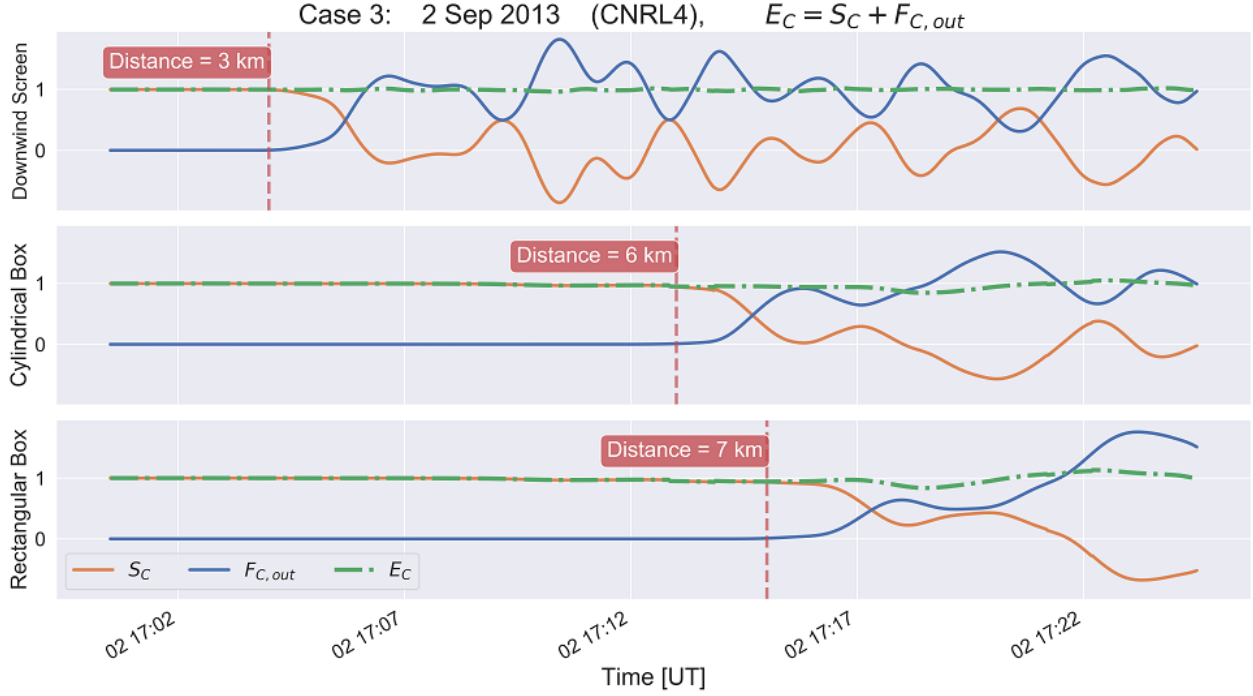


Figure 4.5: Time-series signals for the components of the mass-balance equation (Eq. 4.1) for tracer CNRL4 in Fig. 4.4. Note how S_C and $F_{C,out}$ are anti-correlated and their dependence on the downwind distance of the point of interception from the emission source. The oscillatory behaviour in the signals (period, amplitude) shows correlation to the downwind distance (box size), the bigger the distance the longer the period.

yield over-/under-estimates. Averaging can only result in an accurate estimate if it is performed over the right time period. Firstly, this period should be after the initial purely storage period where $F_{C,out} < S_C$. The storage period can restart during the sampling time if source emission rates change. It can also happen if the transport path/time changes, either due to variations in wind direction or the vertical movement of the plume due to varying buoyancy and/or atmospheric stability. Such temporal changes would result in transient accumulation of tracer mass within the control-volume ($F_{C,out} < S_C$), as changes in transport/emission conditions at the source location would be detected at the sampling point with a time delay (depending on the downwind distance and the advection speed). Secondly, the averaging period must cover complete oscillation periods — i.e., if a peak/trough is left out of the averaging period (e.g. 1.5 period) the estimate will be biased low/high.

As the oscillation period for a shorter downwind distance is shorter, the averaging time needed for an accurate estimate will be shorter too. At longer downwind distances (e.g., rectangular box in Fig. 4.5), a longer averaging time will be needed. If the oscillation period can be detected/estimated from multiple observations over time at the same downwind distance and height, the averaging time can be chosen accordingly to reduce the uncertainty in the estimate. While this provides a potential approach to correct for the storage rate S_C and improve the estimates of E_C , due to the fact that

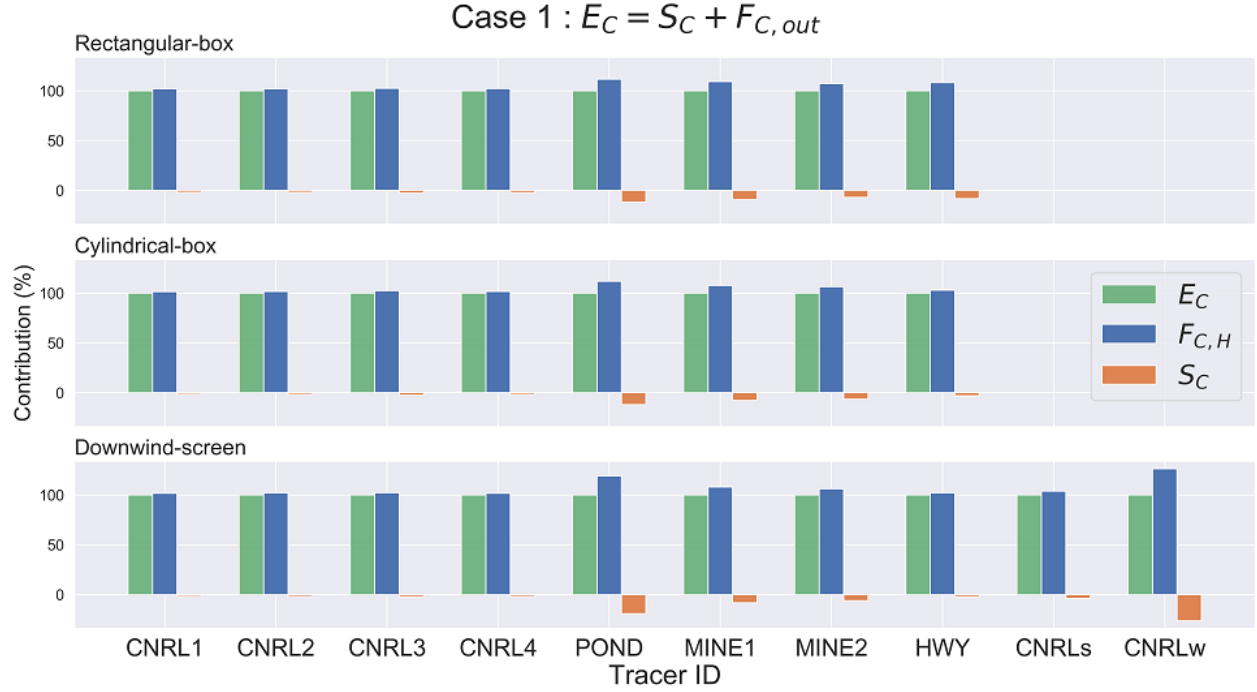


Figure 4.6: Contribution breakdown of the terms in the mass-balance equation (Eq. 4.1) for case 1, where the mean retrieved emission rate E_C is shown with green bars, mean flux out $F_{C,out}$ with blue and the mean storage rate S_C with orange bars. Estimates for the rectangular box, cylindrical box, and the downwind screen are shown in their respective panels. Results show small contribution of S_C for stack/point sources (CNRL1-4) and relatively larger contribution for surface sources and upwind sources (CNRLs and CNRLw). This case represent desirable conditions for top-down retrievals based on $F_{C,out}$ alone as the contribution of S_C is small for most sources.

the oscillation amplitude is also variable over time, S_C remains as a major source of uncertainty in top-down estimates based on an estimate of the net flux out $F_{C,out}$. It is therefore desirable to investigate the contribution of each term in the mass-balance equation (Eq. 4.1) under different meteorology and emission conditions.

Figure 4.6 shows the breakdown for case 1 for the ten tracer emission scenarios. Estimates for the two flux boxes and the downwind screen are shown, all with their downwind wall at 5 km distance from the main CNRL stacks. Note that estimates for the upwind sources CNRLs and CNRLw (sources located outside the box) resulted in negligible E_C estimates for the two flux boxes (flux-in/-out cancel), and therefore are only shown for the downwind-screen. $F_{C,out}$ contributions were between 102% to 122% with mean 106% and median 105% for the rectangular-box, 102% to 112% with mean 105% and median 103% for the cylindrical-box, and 102% to 126% with mean 107% and median 104% for the downwind-screen. S_C contributed the remainder ($100\% - F_{C,out}$).

During the time period of this case the mean wind conditions were consistent, blowing from west with wind speeds between 5 m/s to 15 m/s at 10 m agl. Overall, the transport conditions during case 1 represent ideal conditions for top-down retrievals based on the estimated flux out $F_{C,out}$ due to small contribution of the storage rate term for stack/point emissions ($S_C < 2\%$). For

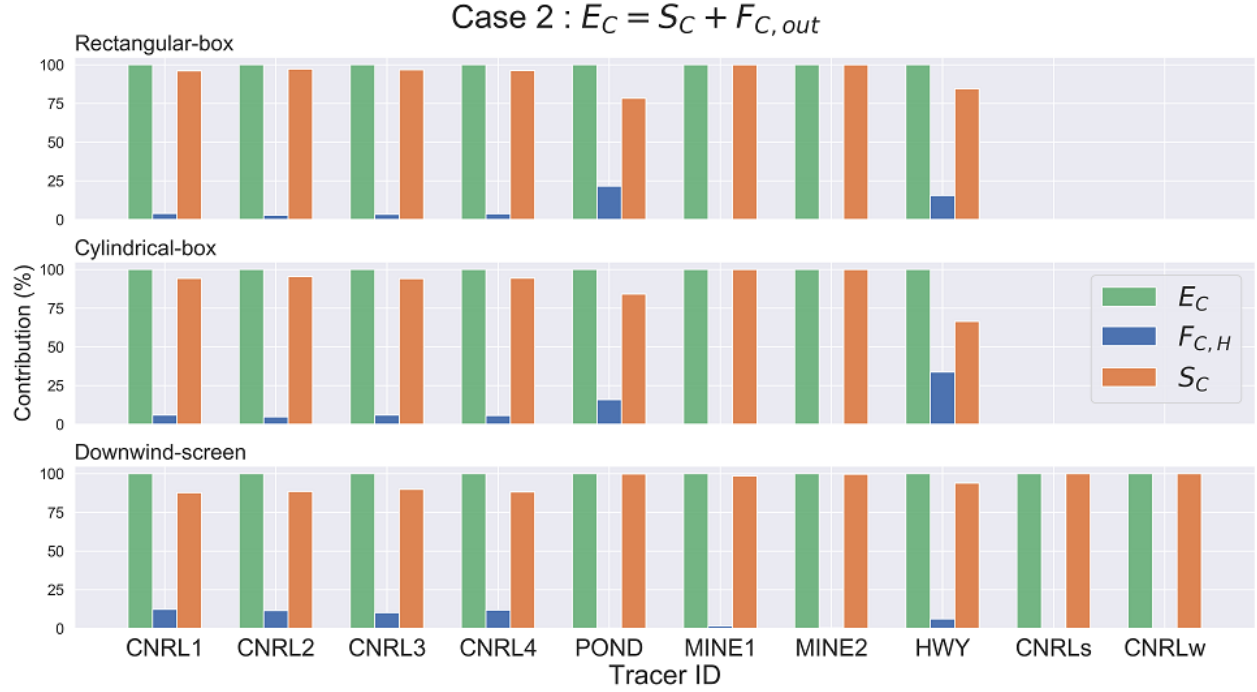


Figure 4.7: Contribution breakdown of the terms in the mass-balance equation (Eq. 4.1) for case 2, where the mean retrieved emission rate E_C is shown with green bars, mean flux out $F_{C,out}$ with blue and the mean storage rate S_C with orange bars. Results for the Rectangular box, cylindrical box, and the downwind screen are shown in their respective panels. Results show the significant contribution of S_C for all emission and sampling scenarios and the relatively small contribution of $F_{C,out}$. This case of severe storage-release event represent undesirable conditions for top-down retrievals based on $F_{C,out}$ alone as the contribution of S_C is dominant in mass-balance equation.

surface sources (or upwind sources at larger distances), the longer transport times to the point of interception (sampling wall) are correlated with higher S_C rates. Sampling at shorter distances could yield more accurate estimates based on $F_{C,out}$ alone. For instance, for the surface source HWY (which is at less than 5 km distance to the sampling wall) S_C contributed only about 2%.

Figure 4.7 shows the breakdown for case 2. During the period of this case (26 August 2013), wind speeds at 10 m agl were < 3 m/s with a strong wind shear in the vertical (see Chapter 3). Tracer amounts released from sources within the control volumes, were transported only 5-15 km in the horizontal in different directions during the sampling time and dispersed mostly in the vertical direction up to over 2 km (see Figures E.3 to E.6). This was due to slow horizontal wind speeds and strong wind shear in the vertical. Similar meteorological conditions were observed for this time period (26 August 2013) during JOSM 2013 field campaign and were also recreated with GEM-MACH model simulations for the same period. See the detailed discussions in Chapter 2. These conditions result in large storage of mass within control volumes (S_C) and relatively small exit flux $F_{C,out}$ during the sampling time for all emission scenarios (Fig. 4.7). Average $F_{C,out}$ contributions were less than 50% for this case. Case 2 provides an example of worse-case conditions where mass-balance estimations should be avoided, due to potentially large under/over-estimates

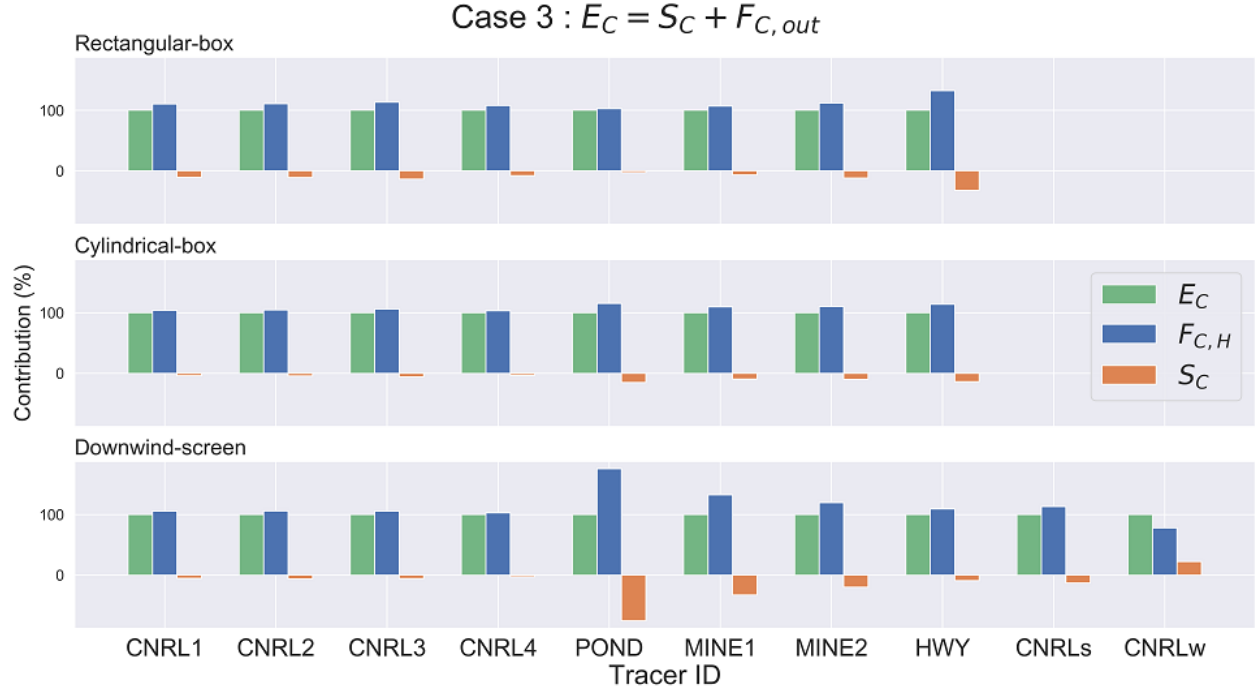


Figure 4.8: Contribution breakdown of the terms in the mass-balance equation (Eq. 4.1) for case 3, where the mean retrieved emission rate E_C is shown with green bars, mean flux out $F_{C,out}$ with blue and the mean storage rate S_C with orange bars. Results for the rectangular-box, cylindrical-box, and the downwind-screen are shown in their respective panels. Results show moderate contribution of S_C for most of the emission scenarios and relatively larger contribution for POND, MINE 1 and CNRLw. This case represent conditions where top-down retrievals based on $F_{C,out}$ would be more significantly improved (relative to case 1) with estimates of S_C .

arising from the storage-release events.

Figure 4.8 shows the same breakdown for case 3 on 2 September 2013. Conditions during the time period of this case were somewhat similar to case 1, but with higher contribution from the storage rate term (S_C) resulting in both over and underestimates based on flux $F_{C,out}$ calculations. S_C contributions were between -13% to -4% for CNRL 1-4, resulting in overestimates by $F_{C,out}$ of 4-13%. S_C contributions for surface sources were also negative resulting in overestimates by $F_{C,out}$ with 3% to 75% for POND, 7% to 32% for MINE 1, 10% to 20% for MINE 2, and 9% to 32% for HWY. For upwind sources CNRLs and CNRLw S_C values were -13% and 22%, respectively. Case 3 represent conditions where top-down retrievals based on flux calculations may be attempted provided that the contribution of S_C to E_C is accounted for.

4.3.1 Error as a function of sampling time and downwind distance

As discussed before, airborne top-down emission estimations rely on measurements along the 2D sampling screens and calculations of the flux out $F_{C,out}$. This is because direct calculations of the storage rate S_C , the other main term in the mass-balance equation, is impossible due to the

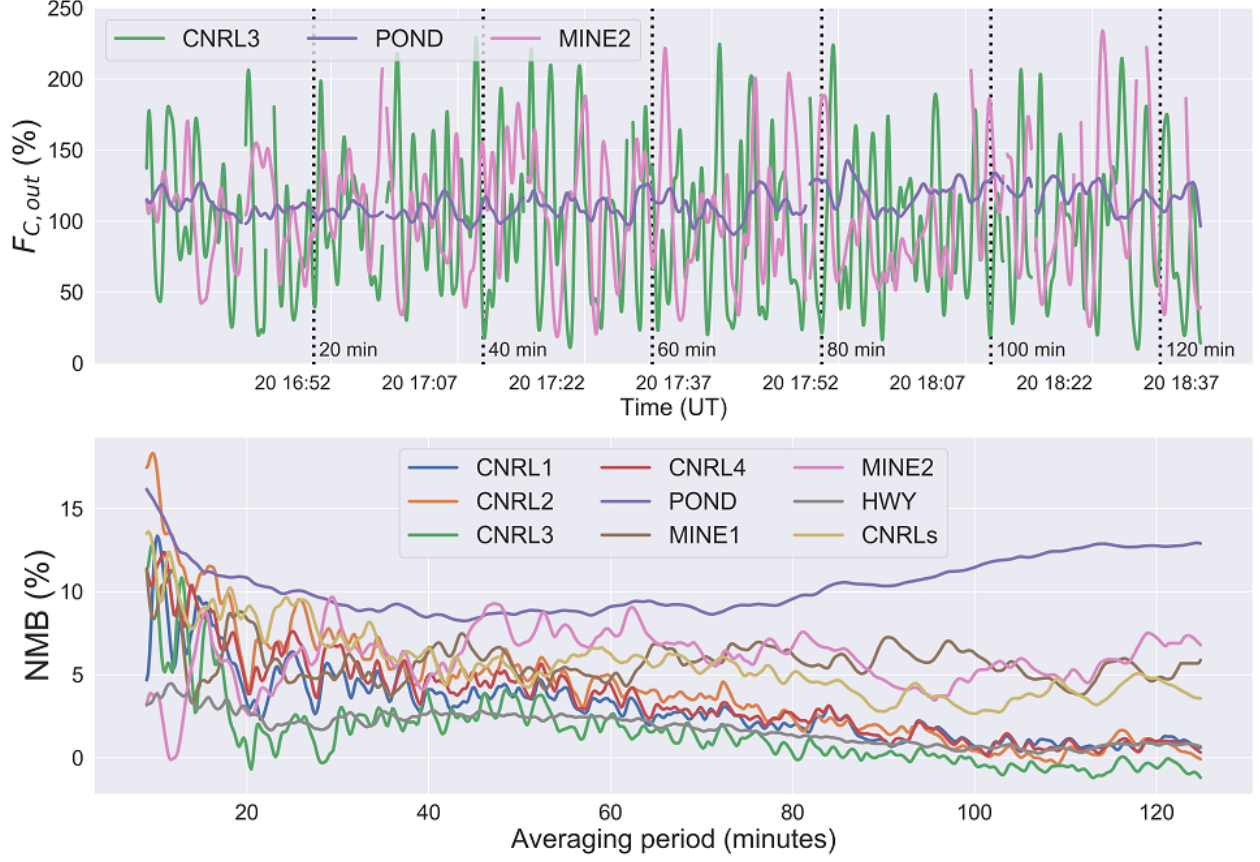


Figure 4.9: Case 1 mass-balance estimates based on flux out ($F_{C,out}$) alone. (top-panel) $F_{C,out}$ time-series for select emission sources, normalized to model Truth. (bottom-panel) normalized mean bias (NMB) in estimates as a function of averaging period. Different averaging intervals are marked in the top-panel with vertical dashed lines. Error decreases as the averaging period increases for most sources, as shown in the bottom-panel. Error increases beyond 40 min for source emissions with relatively bigger contribution from the neglected storage rate term S_C (compare to Fig. 4.6).

lack of volumetric time-series data (4D). The accuracy of airborne estimates are determined by how closely they can estimate the sampling-time-average $F_{C,out}$. However, the correlation of the sampling time average $F_{C,out}$ with source emission rate depends on the averaging period (due to the oscillatory nature of the $F_{C,out}$ signal, see Figure 4.9 top panel). Note that under the assumption of steady state conditions the averaging period for airborne retrievals is effectively the duration of the flight, and the estimated emission rate is considered as the mean source emission rate during the flight time. Here we investigate the assumption of steady state condition by considering flux $F_{C,out}$ through instantaneous complete screens around/downwind the source (idealization). This is achieved by extracting model output fields along the 2D screen comprised of box walls (downwind-screen) and calculating the total (instantaneous) flux out $F_{C,out}$ passing through this screen at each timestamp. We investigate the temporal variability in $F_{C,out}$ for different screens (control volumes) at different downwind distances from the emission sources. Note that the instantaneous sampling of the entire screen is an idealization. In a realistic airborne measurement, discussed in the following

sections, different parts of the screen are sampled at different times (i.e., along the flight path).

The temporal variations in meteorology and transport of tracer amounts during the sampling time, result in corresponding variations in flux out ($F_{C,out}$) through the sampling screen. We investigated the impact of variations in $F_{C,out}$ using WRF model output fields for different averaging periods (flight times). $F_{C,out}$ estimates, using instantaneous screens/boxes, were compared to known emission rates (model Truth). The fluctuating trend in flux out $F_{C,out}$ can be seen in Fig. 4.9 top panel. This fluctuating behaviour is the sum of the turbulent transport and the periodic storage-release events. The short period fluctuations correspond to the turbulent transport, and the longer period trends correspond to storage-release events (Fig. 4.9 top-panel). Figure 4.9 bottom panel shows the normalized mean bias (NMB) error for case 1 (20 August 2013) estimates, as a function of averaging period. For sources with small storage rate S_C values (see Fig. 4.6), as the averaging time increases the NMB error decreases (Fig. 4.9 bottom panel). For the sources with relatively larger contribution from the storage rate S_C (e.g., POND), the error increases with averaging periods larger than about 40 min. For such sources (mainly surface sources), the storage-release period is between 30 min to 120 min due to longer transport times from source to the sampling screen(s). For the shown sources, error remains under 10% between 20 min to 80 min averaging time. The optimal time window of sampling depends on whether a complete storage-release period is captured in the averaging interval. The storage-release time is determined by the transport time which depends on the strength of the advection and the downwind distance of sampling. This demonstrates how model simulation of transport conditions can be utilized to determine the optimal sampling/flight time during airborne retrievals.

The spatio-temporal variability of the tracer amounts are functions of the downwind distance. Figure 4.10 shows box-plots of $F_{C,out}$ during the 2 hr and 10 min sampling time (case 1) for a select number of emission scenarios (stack/point and surface/area sources), for the three sampling volumes/screen. The temporal variations in $F_{C,out}$ can be seen as the large spread (and presence of several outliers) in the data. Note that the spread is larger for point/stack sources than for surface sources. As we discuss later, the extent of the variation in $F_{C,out}$ has higher correlation with the downwind distance (from source to the sampling screen) than the release elevation. In Fig. 4.10 flux estimates for different tracers are arranged by increasing downwind distance from bottom to top, with CNRL 3 being the closest and POND the farthest source to the sampling screen. It is apparent from the box-plots in Fig. 4.10 that the spread in the data is reduced as the downwind distance increases. This is due to the fact that at shorter ranges plumes from emission sources are not yet well mixed especially in the vertical, whereas at larger downwind distances plumes are spread and mixed more uniformly in the 3D space with less temporal variability in the mixing profiles.

Figure 4.11 shows the temporal standard deviation (σ) in tracer vertical profiles as a function of downwind distance. POND is at the largest downwind distance from the three sampling surfaces at 15.8 km, 16.4 km and 16.6 km. For this surface source the variability decreases rapidly as a function of distance and drops to small values by the time it reaches the three sampling screens.

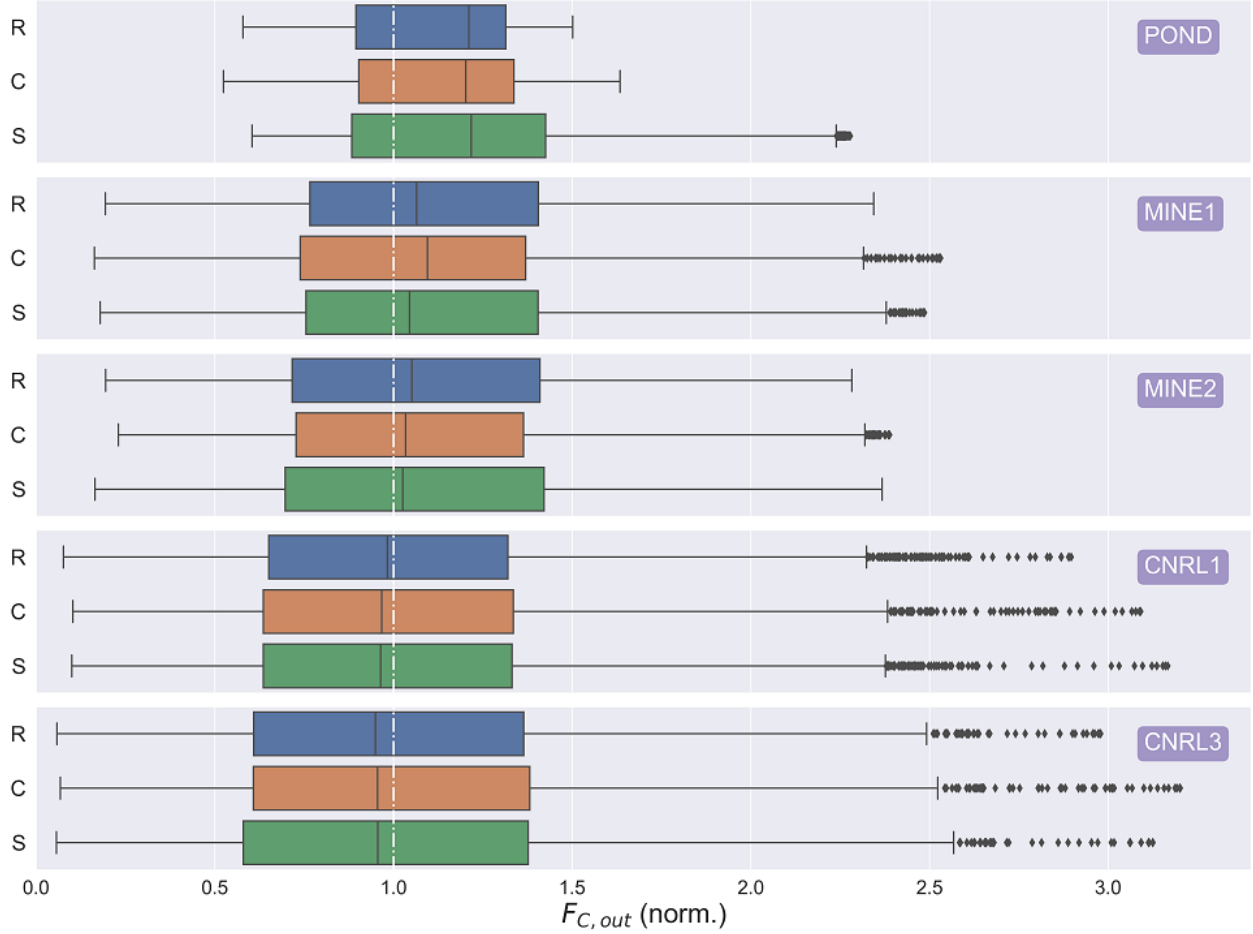


Figure 4.10: Box-plots of case 1 $F_{C,out}$ (normalized to input emissions rates) for several point and area sources during the sampling time of 2 hr and 10 min (at 1 Hz sampling frequency). S stands for the downwind-screen, C for the cylindrical, and R for the rectangular-box estimates. Note how the spread (and outliers) are larger for point sources CNRL1 and CNRL3. This spread is smaller for surface sources POND, MINE1 and MINE2. Surface sources are further from the sampling screens than the point sources.

This results in the small spread in $F_{C,out}$ data as shown in Fig. 4.10. Similarly for the other two surface sources, MINE1 and MINE2, the standard deviation rapidly decreases as a function of downwind distance before reaching the sampling screens. For the two stack emissions CNRL 1 and CNRL 3 shown in Fig. 4.11, the variability drops at a lower rate compared to surface emissions. Which in concert with the fact that they are at shorter distances to the sampling screens, results in larger spread in $F_{C,out}$ as depicted in Fig. 4.10.

We further investigated the reduction in $F_{C,out}$ standard deviation (σ) at larger downwind distances with downwind-screens at 5, 10, 15, and 20 km, as shown in Figure 4.12. The right side panel shows the total σ on the screens (normalized to max. σ on the first screen) for CNRL 1. It can be seen from the figure that the plume is less spread on the first screen (blue curve at 5 km), and highly variable. As we move further downwind, plume is more mixed and spread on

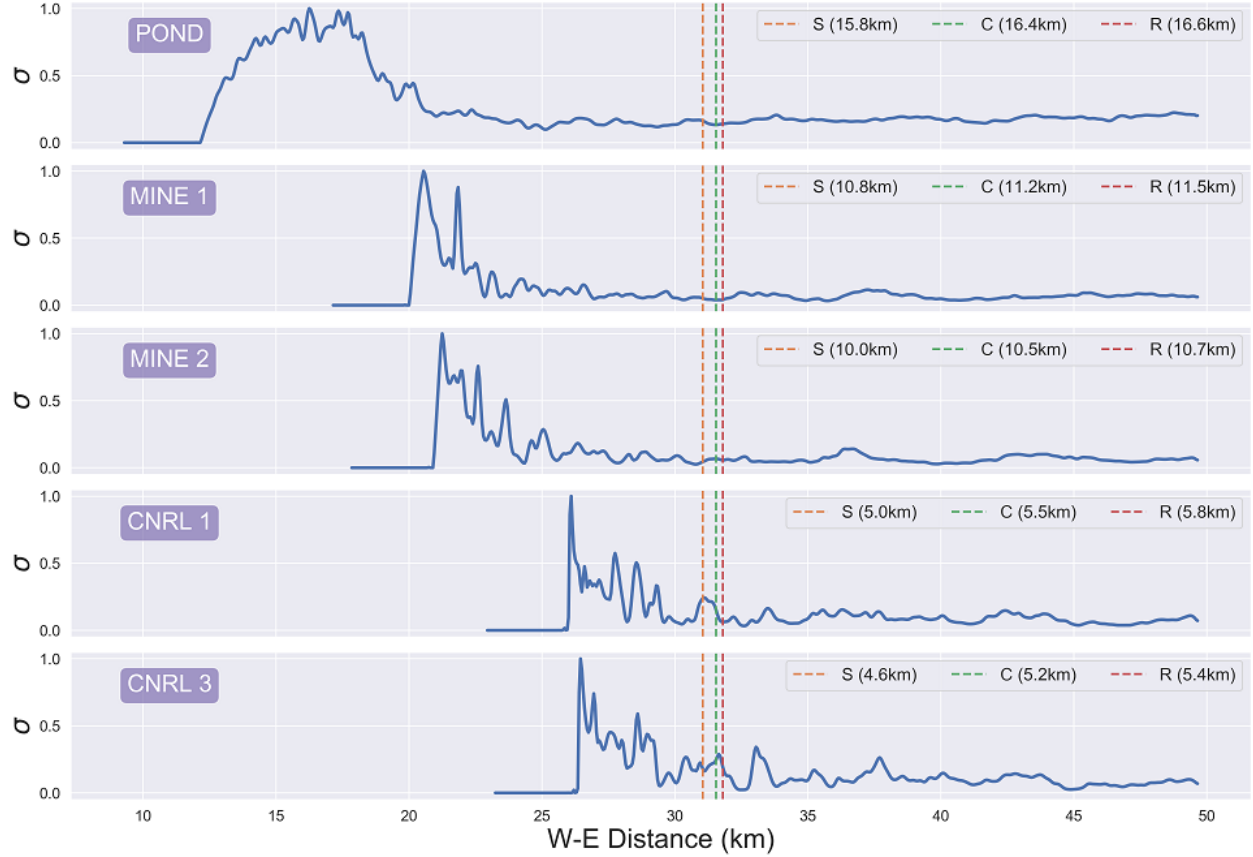


Figure 4.11: Temporal standard deviation (σ) as a function of downwind distance for different emission scenarios. The locations of the sampling surfaces downwind screen (S), cylindrical (C) and rectangular box (R) are shown with dashed lines, and their relative distances (to source) are indicated. The temporal variability in tracer vertical mixing decreases as a function of downwind distance. The rate of this decrease is faster for the surface emissions (POND, MINE1, MINE2) compared to stack emission (CNRL1, CNRL3).

the screen and the variability in terms of σ is reduced (e.g. red curve at 20 km). The top panel shows σ (normalized to max.) as a function of downwind distance as in Fig. 4.11. Similar analysis was done for the other emission sources with similar results. For all sources the variability in flux $F_{C,out}$ is reduced as a function of distance, both in the vertical and horizontal, on the consecutive sampling screens. Sampling beyond the 10 km downwind distance yields more confidence in flux estimates for the stack sources. This distance for surface emissions is about 6 km, as plumes from surface sources tend to mix and spread more rapidly downwind of the sources. This is due to lower flow speeds and increased turbulence near the surface and interaction with the topography (see Chapter 3). Whereas for stack emissions where tracers are released at 30 m to 114 m agl levels, plumes travel further downwind before spreading to the surface and approach uniform mixing in the vertical.

For this portion of our analysis, we used estimates from instantaneous screens (idealized). As we show in the following sections, the realistic airborne FP and RC estimates are impacted more

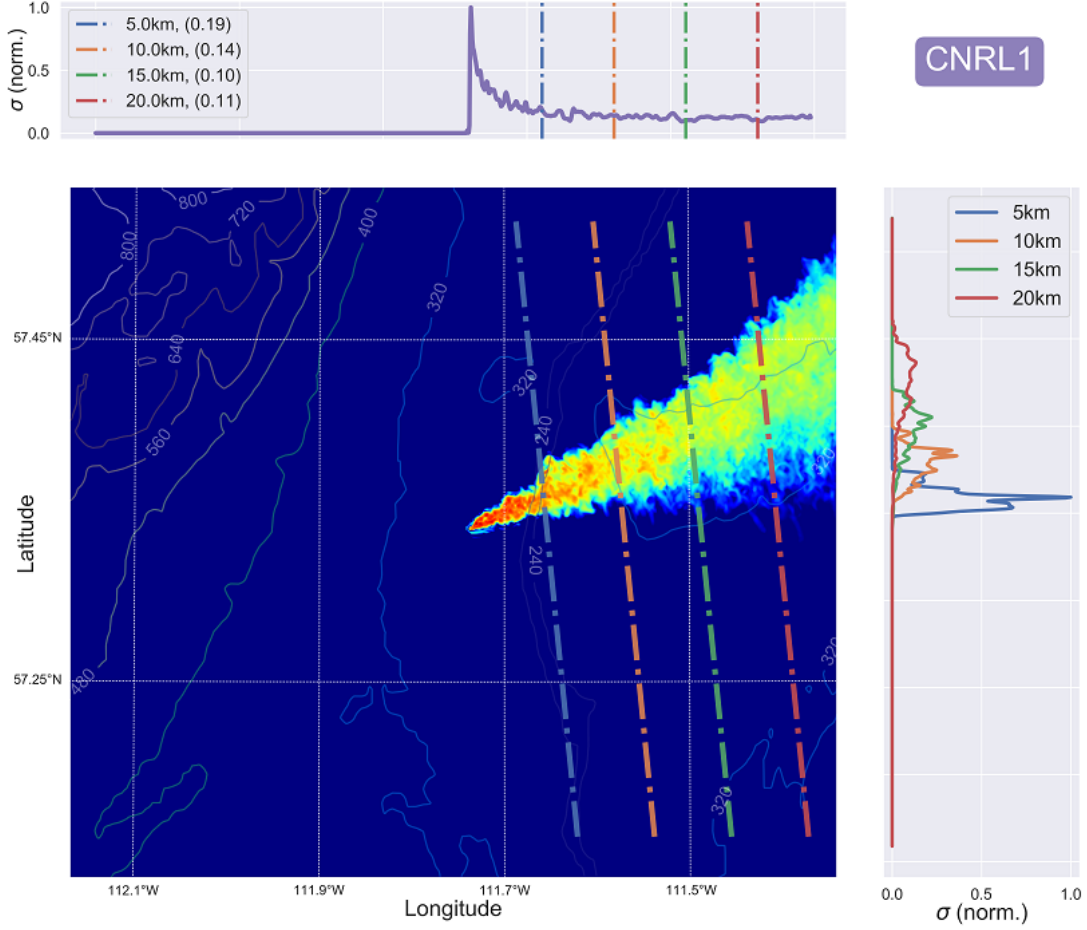


Figure 4.12: Flux estimations with multiple screens at 5 km consecutive downwind distances from the emission source. The location of the downwind screens are shown with dashed curves in the main and the top panels. The top panel shows the decrease in vertical variability (σ) as a function distance. The right panel shows the same along the horizontal extent of each consecutive screen (normalized to the max σ on the first screen). At larger downwind distances, σ drops to low levels as the plume mixes more in the 3D space.

by the spatio-temporal variability in $F_{C,out}$. The extent of this variability affects both estimates from the sampled/interpolated part of the screen (FP and RC) and the extrapolated portion below the flight level (FP). If the variability is high at the sampling screen, the measured/interpolated data become less representative of the mean conditions. Whereas if the variability is low and tracer amounts are uniformly mixed in 3D space by the time they arrive at the sampling screen, data interpolation/extrapolation can be done with more confidence.

4.3.2 Flying Point (FP): conventional airborne approach

We simulated aircraft-based measurements using WRF generated meteorology and tracer concentration fields as proxy for real-world environmental fields. The method of flying point was employed

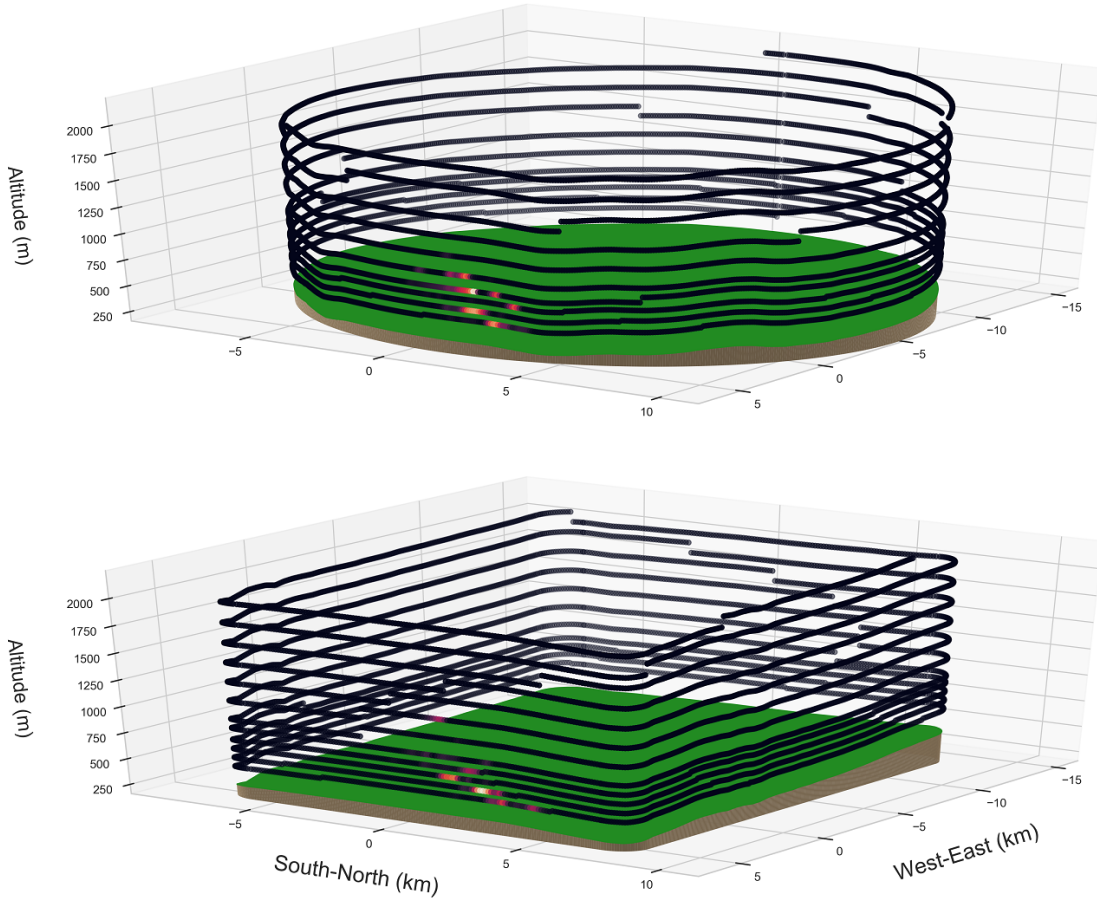


Figure 4.13: The sampling strategy of flying point (FP) is demonstrated for the cylindrical and rectangular boxes. For the shown case, relevant fields are sampled around the facility and along the flight path that starts at about 150 m agl and climbs in altitude in counterclockwise direction. Tracer amount from the surface emission source MINE 2 along the flight path are shown.

to extract/sample relevant information in 3D space and time in a fashion similar to aircraft measurements (see Appendix F, second link). We considered three sampling strategies: rectangular-box, cylindrical-box, and downwind-screen. Figure 4.13 shows two of these sampling strategies, the rectangular box and the cylindrical box. We considered a horizontal flight speed of 100 m/s, similar to box flight speeds of the JOSM field campaign in 2013. For the illustrated case, the flight path starts near the surface (about 150 m above ground level) in the counterclockwise direction and climbs in altitude around the emission source at rates of 0.5-1 m/s for the first 500-600 m agl and at 1-2 m/s aloft. The climbing rate was optimized according to changing model resolution in the vertical, with minimal impact on our analysis. The flight radius for the cylindrical box is about 10 km, centred about 5 km west and 1 km north of the main CNRL stacks (see Figure 4.4 for the top view). The horizontal dimensions for the rectangular box are 20 km $\times$ 15 km (long side along the west-east direction) and it is centred about 4.25 km west and 0.5 km north of main CNRL stacks. One complete revolution around the facility takes 10 min to 14 min of flight time. Both control volumes

(flux boxes) completely enclose all the emission sources except for the upwind sources CNRLw and CNRLs, and for HWY which is partially captured. For both cases the box top is at an altitude of about 2.5 km, well above the altitude at which considerable tracer amounts are detectable.

Following the common practices for aircraft-based retrievals (Gordon et al., 2015; Ryoo et al., 2019), the sampled data were interpolated onto the 2D surface comprised of box lateral walls using the kriging interpolation method. Kriging calculates the weights of each sampled point based on the spatial distance between the sampled locations and points on the 2D screen, and then data are interpolated according to the calculated weights (Goovaerts, 2019). Figure 4.14 shows the resulting (Kriging) interpolated rectangular-box for CNRL 1 emissions during case 1. Interpolation is done in 2D space and only between flight tracks (i.e. heights below/above lowest/highest flight tracks were excluded). For the case shown in Fig. 4.14, the tracer was released at a height of 114 m above ground level (agl), and advected downwind for about 5 km before being intercepted by the flying point (virtual aircraft) on the east wall of the rectangular box. The maximum detected flux on the interpolated screen is at 187 m agl (470 m altitude), which indicates a 73 m rise relative to the ground level. Note that the topography height declines in the west-east direction from the location of the main CNRL stacks towards the Athabasca river basin where the box east wall is located.

During the period of case 1, CNRL 1 plume ascended to higher altitudes over time as the mixing layer grew in the vertical towards the local noon (18:00UT). During the same period, the direction of advection also shifted more towards the east from the initial north-east direction at 16:30UT. FP sampled the relevant information for this case, while flying in counterclockwise direction and climbing in altitude. Consequently, the same plume was sampled multiple times at different locations, as can be seen in Fig. 4.14. The maximum tracer amounts sampled at each altitude level progressively shifted towards south on the east wall, indicating the shift in the advection direction. The maximum tracer amount was sampled during the first pass on the east wall around the facility, which corresponds to the plume centre on the interpolated screen. Figure 4.14 (right panel) shows tracer ($F_{C,out}$) vertical profiles (dashed lines) at the horizontal location of plume maximum during the first and last pass (first and last 10-12 min of flight time). Note how the first FP data point (lowest square) intersects with the vertical profile for the first pass (blue line). At this point in space and time, FP sampled the maximum tracer amount. Next time around (about 11 min later), the FP sampled value at the same horizontal location dropped to a negligible level (second lowest square). FP data points (at every 10-12 min) for the same horizontal location remain relatively small for the rest of the flight, which is indicative of plume moving to another region on the 2D screen and not due to turbulent transport/dispersion of tracer amounts. The vertical profile during the last pass over the same location (orange line), also shows relatively small amounts.

On average, 51% of the plume mass shown in Fig. 4.14, fell below the lowest flight level and therefore was not sampled by FP. For this sampling strategy, the 2D screen data need to be extrapolated for the unsampled region below the lowest flight level to the ground surface. Figure 4.14 right panel, shows extrapolated profiles using three different methods:

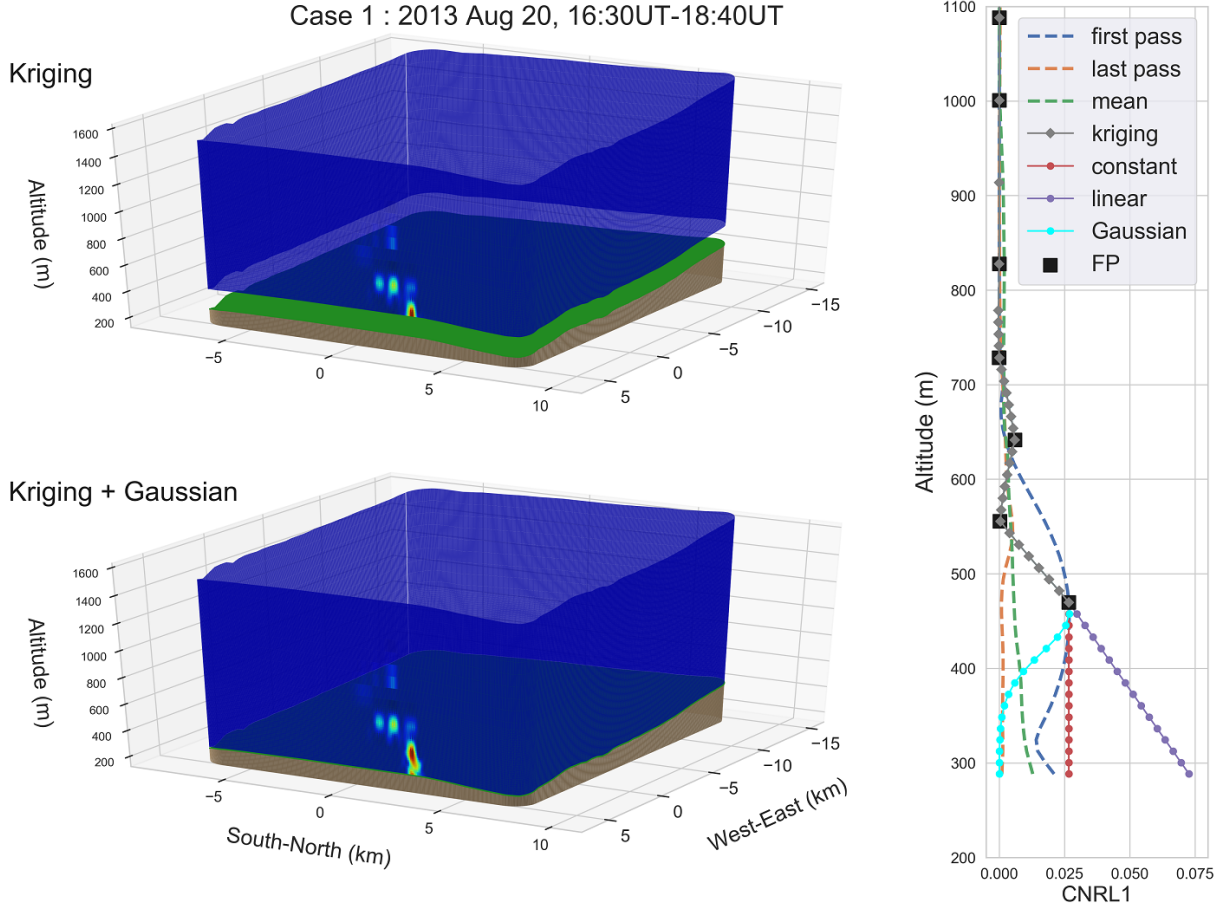


Figure 4.14: Case 1 FP sampled rectangular-box for CNRL1 is shown. The FP sampled region was interpolated using the kriging method (top left). FP sampled/interpolated data are then extrapolated below the flight level for calculating the net flux out $F_{C,out}$ (bottom left). Model vertical profiles during the first and last pass/lap at the horizontal location of FP-sampled maximum are shown in the right panel (dashed lines). FP sampled points are shown with squares, and the kriging interpolated profile with diamonds. Extrapolation methods (constant p_c , linear p_l , Gaussian p_g) applied to sampled/interpolated vertical profile at the same horizontal location are also shown.

- Constant p_c , where the value of the lowest sampled point is assigned to the unsampled heights,
- Linear p_l , where data is extrapolated to the ground surface according to a linear fit to the interpolated data up to an altitude of about 600 m,
- Gaussian p_g , where a Gaussian fit is used for data extrapolation. See Eq. G.1 for more details.

We also considered a “zero method” p_0 , where concentrations below the lowest flight level are assumed to be zero. The extrapolated profiles agree only at one point (lowest sampled point) and diverge below. For the case shown in 4.14, compared to vertical profile during the first pass, the extrapolated profiles are biased low at higher altitudes and biased high at lower altitudes for p_c and p_l , and alternate for p_g . The same doesn’t apply to other emission cases. Figure 4.15 shows

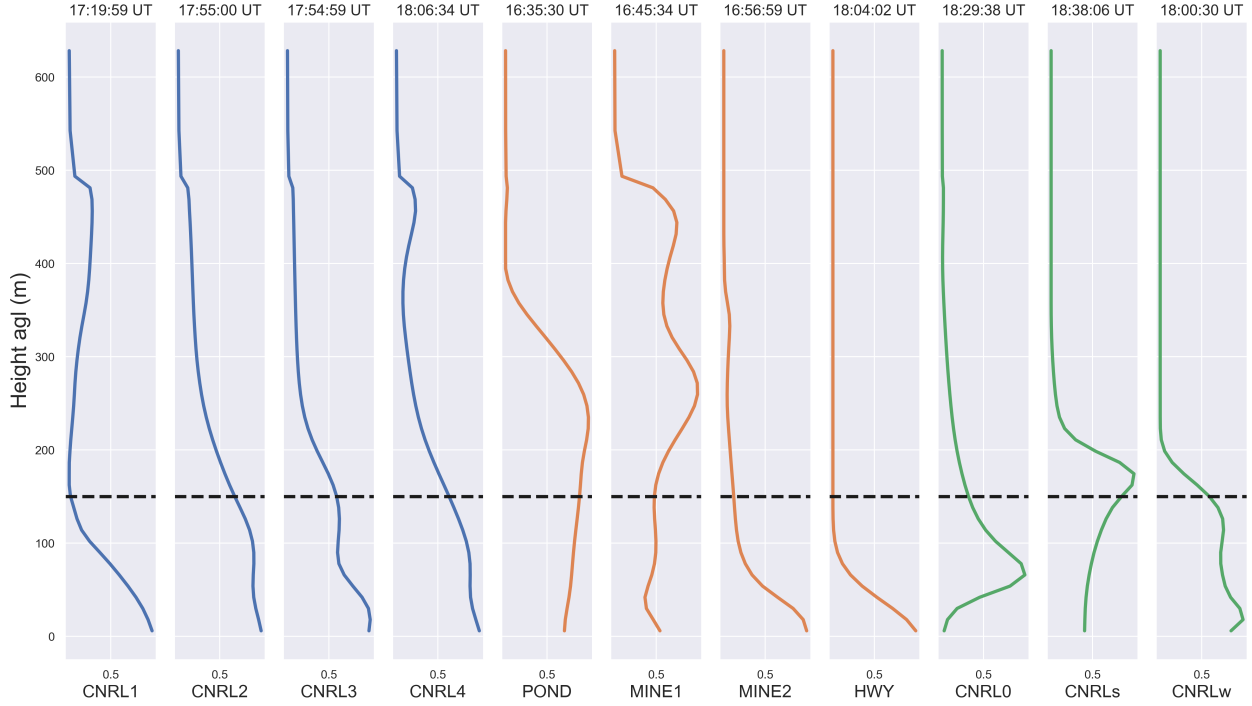


Figure 4.15: Vertical profiles of tracer amounts (normalized to max.) at the horizontal location of the plume maximum (centre) for the eleven emission sources. Note that these profiles are for different timestamps (indicated on top of each panel) and are at different horizontal locations (for different plume centrelines) along the rectangular-box walls. Vertical mixing is diverse and variable in time and space for different emission sources.

the vertical mixing profiles for eleven tracer plumes at different times and horizontal locations. It can be seen that the profiles are very diverse, which is a combined result of temporal variability and different downwind distances. For some of the tracers, profile maximum is above the lowest sampling level, for the others it is below (within the unsampled region). For some (e.g. POND, MINE1), mixing of tracer amounts below the lowest level is relatively uniform and consistent with the lowest sampled point, for the others tracer amounts increase (e.g. CNRL4) or decrease (CNRLs) towards the ground surface. Stack plumes (blue curves) have maximums at higher altitudes, while large portions of them fall below the lowest level. These profiles are variable over time and across the sampling surfaces. This type discrepancy makes the choice of extrapolation method challenging and error prone, which in turn creates a significant level of uncertainty for top-down estimates of the flux $F_{C,out}$ based on the interpolated/extrapolated (sampling) screens.

The portion of plume falling below the lowest flight level (unsampled region) ranged from 16% to over 50% of the total plume mass for different emission scenarios (Table G.1). Therefore, the range and level of uncertainty in flux estimates due data extrapolation was different for each emission scenario. For the case shown in Fig. 4.14, estimates by the three extrapolation methods differ by about 18%. The constant method p_c performs best by generating an estimate with less than 1% bias. For this case the p_0 estimate, where only the sampled/interpolated region is considered, is about

Table 4.1: Case 1 FP estimates of flux out are evaluated against the true $F_{C,out}$ and results are presented as the normalized (to true $F_{C,out}$) mean bias. Extrapolation methods zero p_0 , constant p_c , linear p_l , Gaussian p_g are compared for the the rectangular box, cylindrical box and the downwind screen estimates. Note that the large values for CNRLw and CNRLs for the two boxes are the results of division by very small $F_{C,out}$ values (normalization). Bold text is used for indicating the best extrapolation result in each category. Absolute values less than 0.01 are indicated with ϵ .

	Rectangular Box				Cylindrical Box				Downwind Screen			
	p_0	p_c	p_l	p_g	p_0	p_c	p_l	p_g	p_0	p_c	p_l	p_g
CNRL1	-0.46	ϵ	0.13	-0.18	-0.42	0.28	0.42	-0.14	-0.14	0.19	-0.27	-0.07
CNRL2	-0.35	-0.11	-0.32	-0.25	-0.58	-0.40	-0.70	-0.53	-0.16	0.50	0.36	0.05
CNRL4	-0.42	-0.16	-0.30	-0.25	-0.40	0.28	0.35	-0.22	0.02	0.47	0.20	0.09
CNRL3	-0.48	-0.19	-0.28	-0.39	-0.26	-0.10	-0.84	-0.24	-0.46	0.05	ϵ	-0.38
POND	-0.33	0.19	0.29	0.08	-0.35	0.19	0.30	0.02	-0.20	0.42	0.53	0.22
MINE1	-0.16	0.34	0.32	-0.02	-0.41	-0.27	-0.63	-0.39	-0.26	0.34	0.35	0.03
MINE2	-0.28	0.11	0.06	-0.11	-0.27	0.34	0.49	-0.06	0.07	0.60	0.39	0.15
HWY	0.54	1.48	1.39	1.21	-0.35	0.24	0.41	0.06	-0.26	0.22	0.27	0.01
CNRLs	1.01	3.97	5.29	1.67	-2.63	-5.63	-8.21	-4.54	-0.38	-0.01	0.02	-0.12
CNRLw	-3.33	-3.41	-2.47	-2.20	-1.08	1.25	4.03	-1.05	-0.22	0.48	0.62	0.30

54% of the true (mean) flux $F_{C,out}$ (-46% bias) while the model average indicates that the sampled region should account for only 49% of $F_{C,out}$. Therefore the sampled/interpolated extent of the screen overestimates the true flux for the same portion of the screen by 5%. Model average is 51% for the unsampled region below the flight level, while p_c predicts 46% contribution (underestimating by 5%). Therefore while the mean absolute error for this case is about 10%, the cancelling out of the overestimation for the sampled region and underestimation for the unsampled region results in less than 1% bias. On the other hand, the same constant method (p_c) overestimates the true $F_{C,out}$ using data from the cylindrical-box and the downwind-screen for the same tracer and during the same sampling time (at different downwind distances) by 28% and 19%, respectively. Similar discrepancies occur for the other sources and methods. For example, for the surface emission source POND the Gaussian method p_g overestimates the true flux by 8%, 2% and 22%, while p_c overestimates the same by 19% and 42%. Estimates for the different emission sources were compared to true $F_{C,out}$ and results are summarized in Table 4.1 as normalized mean bias (nmb) scores. Comparisons against true E_C showed similar bias levels.

The best FP estimate for each category (source and sampling screen) is indicated with bold text in Table 4.1. For stack sources CNRL1-4 with the rectangular-box, the constant method p_c generated the best estimates with bias ranging between -1% to -19% with median -13.5%. Our results are comparable to the uncertainty ranges of 2% and 27% for SO_2 and CH_4 , estimated by Gordon et al. (2015) for aircraft-derived (observational) emission rates for the same case (CNRL, 20 August 2013, constant p_c). For the cylindrical-box with the downwind wall closer to CNRL sources by a few hundred meters relative to the rectangular-box, the best estimates were split between the

constant method (CNRL2 and CNRL3) and the Gaussian method (CNRL1 and CNRL4) as shown in Table 4.1. For the downwind-screen at the shortest distance of 5 km (500 m closer than the cylindrical-box) to CNRL stacks, the Gaussian method gave the best estimates (except for CNRL3) with bias ranging between -7% to 9% with median 5%. This suggests that at shorter distances, the Gaussian method may perform better than the constant and linear methods. The Gaussian method generally performed better than the other two methods for the surface sources and upwind sources with the two boxes (with two exceptions - see Table 4.1). These results are consistent with conclusions by Gordon et al. (2015) regarding CH₄ profiles on the downwind wall versus the upwind wall; where a constant profile was chosen for extrapolation on the downwind wall, and an exponential or (half-) Gaussian profile for the upwind wall.

Evaluation of FP estimates against the model Truth, shows that data extrapolation below the lowest flight level introduces the largest source of uncertainty for top-down estimates based on flux ($F_{C,out}$) calculations. While up to 50% of uncertainty in FP estimates can be attributed to the information gap below the lowest flight level, the temporal variability of the sampled fields is another major source of uncertainty. Note that the performance of the extrapolation methods are also impacted by the discontinuity in the sampled data, due to plume displacement over time. As can be seen in Fig. 4.14, while the instantaneous vertical profiles (first and last pass) can be closely approximated by at least one of the extrapolation methods, temporal variations in sampled profiles reduce the accuracy of data extrapolation. As we discussed in the previous section, this temporal variability is correlated with the downwind distance of sampling.

Previous aircraft-/model-based studies support substantial uncertainties in top-down estimates due to temporal variability in the sampled fields and the gap of information below the flight level (e.g., Conley et al., 2017; Karion et al., 2019; Peischl et al., 2010). For instance, Angevine et al. (2020) estimated the error in airborne top-down estimates against a known source (four flights), and also by using a Lagrangian dispersion model (HYSPLIT). Their results indicate an uncertainty range of 30% to 40% for top-down estimates under ideal conditions, and up to 100% for less than ideal conditions. Our evaluations of FP estimates indicate a wide range of uncertainty under different conditions, from a few percent to over 50%. As we discuss next, model simulations of atmospheric tracer transport and dispersion can be employed for optimizing the top-down retrievals.

4.3.3 Optimal flight time/radius: recommendations for FP airborne retrievals

Our analysis with instantaneous sampling screens/volumes, indicated that error due to temporal variability of the sampled fields in estimates based on flux $F_{C,out}$ calculations, remain under 10% of the true emission rates for sampling/averaging times between 20 min to 80 min (Fig. 4.9). One revolution at about 100 m/s flight speed around the flux boxes that we considered (rectangular and cylindrical), which are similar in dimensions to box flights during the JOSM 2013 airborne campaign, takes about 10-12 minutes to complete. A flight time of 1.5 hour is enough for 5 to 7 revolutions at multiple altitude levels around the emission source of interest. Assuming well mixed conditions at the sampling plane (box downwind wall), observations at 5 to 7 altitude levels spaced

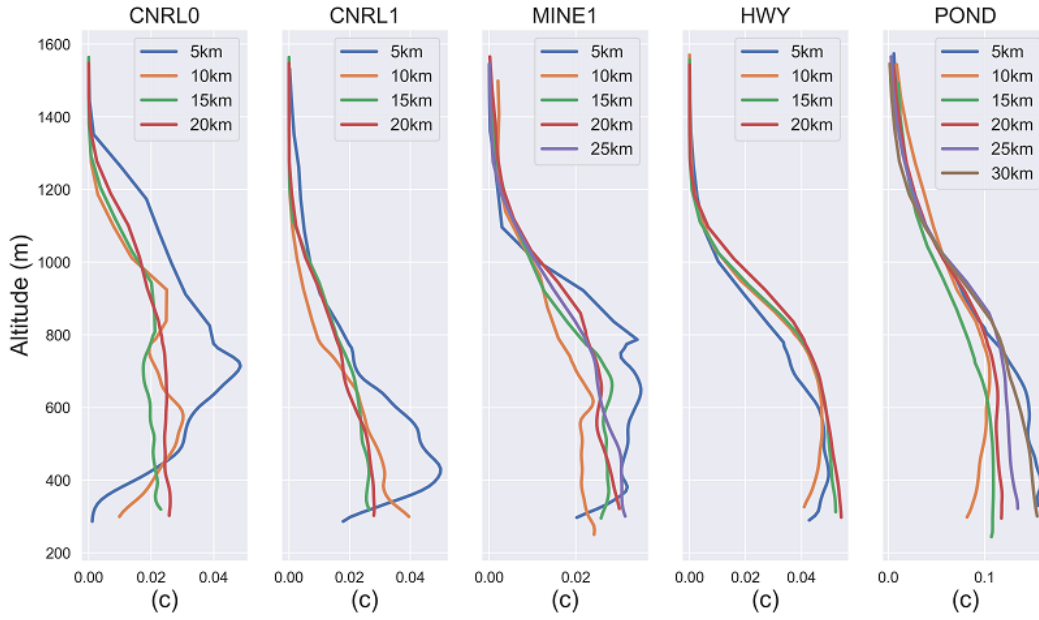


Figure 4.16: Tracer vertical profiles at consecutive 5 km downwind distances. Profiles for stack emission CNRL0 and CNRL1, remain non-uniform as a function of height up to 10 km. These approach constant profiles at about 15 km. The same is true for surface emissions, however they approach near constant profiles at shorter distances. The 15 km distance seems to be the optimal sampling radius, where data extrapolation below flight level can be done with more confidence.

(vertically) by 100-200 meters can provide sufficient information for estimating the mixing profiles of sampled tracer plumes. This then can be used, along with observed wind profiles, for mass-balance estimates of source emission rates. The validity of the assumption of well mixed plumes depends on the spatio-temporal variability of concentration fields at the sampling distance.

For stack emissions, normalized spatio-temporal variability (σ) in the sampled data drop to 10% at downwind distances between 10-15 km. For surface emissions, σ drops to the 10% level between 6-10 km downwind (Fig. 4.11). σ values below 10% correspond to nearly uniform mixing in the vertical especially for low plumes. Figure 4.16 shows vertical profiles for select tracers at downwind distances of 5 km, 10 km, 15 km and 20 km. It can be seen that profiles are highly variable as a function of altitude for stack emissions (e.g., CNRL 1) up to 10 km, and they approach constant profiles beyond the 15 km distance. Profiles for surface emissions (MINE1, HWY and POND) are already well mixed at 5-10 km, but they too approach constant profiles at around 15 km. Analysis for the other emission sources and sampling volumes/screens show similar results. Our results are very similar to other LES studies (e.g., Conley et al., 2017; Vinuesa and Galmarini, 2009). For instance, LES simulations of tracer transport by Vinuesa and Galmarini (2009) show near uniform mixing in tracer vertical profiles at 12-24 km downwind distance. According to our results for various emission scenarios, the shortest downwind distance where vertical profiles approach

uniform mixing is between 10-15 km. At such distances, the spatio-temporal variability (σ) in tracer concentration is low and less than 10%. FP sampling of uniformly mixed tracer profiles would result in more accurate estimates based on flux out $F_{C,out}$ calculations. Furthermore, under well mixed conditions, the sampled data can be extrapolated below the flight level at higher confidence levels.

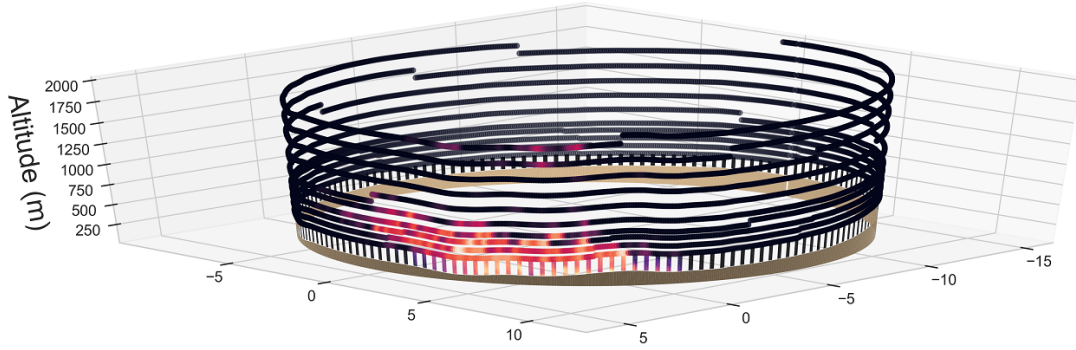
Determining the optimal sampling radius depends on several factors. Sampling needs to be done at far enough distances where tracer plumes are well mixed and less variable over time. It also needs to be close enough to the emission source so the concentration signal is strong and well characterized for accurate estimates. Tracer mass can be lost due to surface deposition or chemical transformation (not simulated here) at large downwind distances of 15-20 km (Angevine et al., 2020; Conley et al., 2017). Flight time is another factor in determining the optimal sampling radius. Flight time for single-leg downwind sampling screens, such as the ones shown in Figure 4.12, is independent of the chosen sampling distance (unless the plume gets significantly wider downwind). Flight paths for downwind screens are usually shorter than box flights and can sample more altitude levels in one hour of flight time. However, top-down retrievals of source emission rates with single-leg screens are only accurate when the contribution of the storage term S_C (which is being neglected in this approach) is small and upwind emissions are negligible and do not interfere with top-down retrievals for the source(s) of interest. In the presence of high upwind concentrations (background, upwind sources), the recommended approach is sampling with flux boxes (control volumes). Flux boxes are certainly feasible for flight times of under two hours, and with downwind walls at 10-15 km distance from the emission source. Our estimates with the rectangular and cylindrical boxes show no significant differences, and so the estimates are not affected by the choice of flight path.

We note that the results discussed here are for a specific case with consistent winds and under strong advection conditions. Meteorological conditions are very important in determining the optimal sampling strategy. We demonstrated here that real-case LES simulations can be utilized for campaign planning, where the optimal flight/sampling approach can be determined for various emission sources and atmospheric transport conditions.

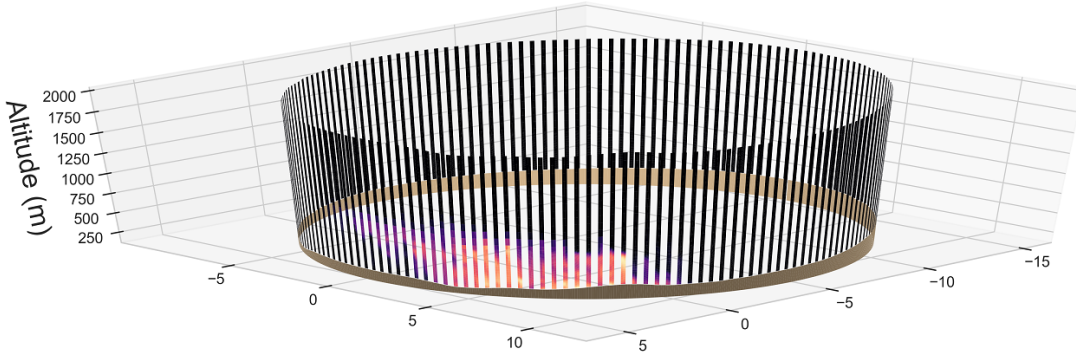
4.3.4 Rolling Colum (RC): combined in-situ and remote retrievals

In the previous section we discussed the uncertainties in $F_{C,out}$ estimates arising from the lack of spatial and temporal information based on the sampling strategy of Flying Point (FP). Namely, the gap of information below the lowest flight level to the ground surface which is responsible for 16% to 52% of uncertainty in the estimates. Filling this gap by conducting column sampling can improve the confidence level in the top-down retrievals. This can be achieved through ground-based (e.g., land-vehicle lidar: de Boer et al., 2021) and/or aircraft-based (e.g., airborne lidar: Aggarwal et al., 2018) remote column measurements, combined with in-situ sampling. The vertical profile of tracer amounts can be detected by remote column measurements. Column sampled trends can then be fitted to aircraft-based in-situ measurements (i.e. FP) at the same horizontal locations and timestamps to generate estimates for the actual vertical profiles. The culmination of these steps

(a) FP + RC



(b) RC



(c) RC interpolated

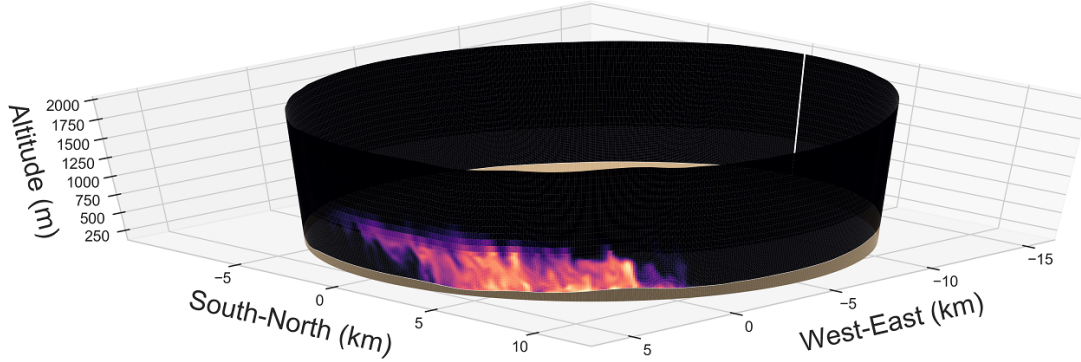


Figure 4.17: Application of Rolling Column (RC) method for top-down retrievals. POND emissions on the cylindrical box are shown. (a) FP measurements combined with column measurements during the first/lowest pass around the facility RC_0 (first 10 min of the sampling time). (b) RC column measurements extended for the entire extent of the box (ground surface to the box top) with up and down facing remote column measurements. (c) RC sampling linearly interpolated in the horizontal to generate complete flux screens for every 10-12 min of the flight time.

give the method of Rolling Column (RC). For practical applications the uncertainty in column measurements based on the chosen remote sensing approach (e.g., LIDAR, AVIRIS: Aggarwal et al., 2018; Thorpe et al., 2021) and the fitting process to the in-situ measurements need to be accounted for in the final product which is an estimate of flux out $F_{C,out}$. For our purposes we

focus on the theoretical aspects of using the ancillary data provided through column measurements in top-down retrievals rather than the instrumental/operational uncertainties.

We start with the gap of information below the lowest flight level. Figure 4.17a shows the FP sampled data for the cylindrical box and emissions from the surface source POND. The gap below the lowest flight level is filled with column measurements during the lowest flight pass/lap (the first 10 min of the sampling time) around the facility. This provides an alternative to extrapolating the sampled data for heights below the flight level, which provides a relatively higher degree of confidence in estimates for the region that is not sampled by the FP method. We refer to such estimates as FP+RC₀, where we combine FP sampling with the column measurements during the first/lowest circuit (RC₀). For the surface source POND, FP+RC₀ results in up to 25% improvement over FP estimates. FP+RC₀ estimates for all sources and sampling screens were evaluated against the true $F_{C,out}$ in terms of normalized mean bias (NMB) and are summarized in Table 4.2. FP+RC₀ performs best when applied to low plumes where the main portion of the plume is sampled during the first few circuits around the facility. Note that FP+RC₀ estimates are discussed here to further highlight the uncertainty introduced by lack of sampling below the flight level with the FP approach, rather than a proposed airborne sampling method.

By extending the RC column measurements for the entire vertical extent of the box/screen, from ground surface to the box top, complete flux screens can be generated for each 10-15 min segment of the total sampling time (Figures 4.17b and c). This approach eliminates two major sources of uncertainty: (1) interpolation of sparsely sampled (spatially and temporally) FP data points collected over the entire sampling period (>2 hr) and along the flight path, (2) data extrapolation below the lowest flight level. Further, generating flux $F_{C,out}$ estimates using time-consecutive RC screens, provides temporal data for investigating the evolution of the system over time and provides a means for estimating the storage rate for each source. Figure 4.18 shows RC generated time-series of estimated flux out $F_{C,out}$ for select sources using the cylindrical box shown in Fig. 4.17. Compared to the true model $F_{C,out}$ for the cylindrical box (blue curves), RC estimates (orange step-plot with markers) perform well with normalized mean bias (nmb) scores of -10% to -2%. Results for all emission scenarios and sampling methods are summarized in Table 4.2. Note that with the RC method, only one lap around (or downwind) the facility is enough to generate a complete screen and an estimate of $F_{C,out}$. This reduces the required flight time from several hours to several minutes. In Fig. 4.18 we show multiple RC estimates (12 laps) during the full 2 hour and 10 min flight time for case 1. As can be seen from the figure, the individual RC estimates agree well with model $F_{C,out}$ values for the corresponding time periods. This suggests that one lap with RC sampling can generate enough information for a reliable estimate of the flux out $F_{C,out}$.

RC results show overall improvements over FP+RC₀ for $F_{C,out}$ estimates. However, the true value of the RC method lies in the generation of temporal data that can be utilized to estimate the storage rate term S_C , something that is not possible with the single-value estimates (one estimate for the entire sampling period) by FP and FP+RC₀ methods. RC sampling generates time-consecutive screens of meteorological data as well as tracer concentrations. Temporal concentration data on RC



Figure 4.18: RC generated time-series of $F_{C,out}$ for select sources using the cylindrical box are shown (orange step-plots). RC estimates follow the true model $F_{C,out}$ signal and capture the temporal variability in the exit flux. Normalized mean bias (nmb) scores are also indicated for each source.

screens can be used to estimate the rate of change in tracer mass within the box (effective) volume. As mentioned in the methods section, for downwind-screen sampling where a control volume is not defined by the flight path, the effective volume is considered as the volume of a virtual box that has a triangular base area with height equal to the distance from source to the downwind screen and base equal to the horizontal extent of the downwind screen. Such calculations can be made according to equations 4.2 and 4.3. Note that estimating S_C from RC measurements, requires a minimum of two laps around the emission source. Conducting additional laps will generate more temporal data and can potentially increase the confidence in S_C estimates.

Figure 4.19 shows RC estimates for case 1 CNRL1 with the rectangular box, where storage rate S_C is calculated according to Eq. 4.3 with 12 laps. The source emission rate is calculated as $E_C = F_{C,out} + S_C$. RC generated S_C improves the E_C estimate for case 1 CNRL1 by 17% relative to true model E_C . See Figure G.1 for RC estimates for case 3 CNRL1 with a cylindrical box, where the RC generated S_C for this case improves E_C estimates by 14%. RC retrievals with estimates of the storage rate term S_C were carried out for all emission sources and sampling surfaces and were evaluated against model $E_{C,S}$ for each set. Results for case 1 are summarized in Table 4.3.

Estimates based on Eq. 4.3 where S_C is calculated based on difference in tracer concentration on the consecutive sampling screen, result in improvements between 2-17% with median of 12% for rectangular-box, 3-31% with median of 6% for cylindrical-box, and 1-10% with median of 5% for the downwind-screen. Case 3 estimates show similar results to case 1, where estimates of S_C

Table 4.2: Case 1 FP+RC₀ and full RC (with 10-12 laps) generated estimates evaluated against the true model $F_{C,out}$ for the two flux boxes and the downwind screen (at 5 km). Results are shown in terms of normalized mean bias (NMB). Sampling was conducted at horizontal flight speed of 100 m/s around and downwind the emission sources for all sampling volumes/screen.

	Rectangular Box		Cylindrical Box		Downwind Screen	
	FP+RC ₀	RC	FP+RC ₀	RC	FP+RC ₀	RC
CNRL1	-0.01	0.18	0.54	-0.07	0.34	0.07
CNRL2	0.14	0.26	-0.13	-0.06	0.48	0.19
CNRL4	-0.005	0.16	0.53	-0.13	0.58	0.15
CNRL3	-0.31	-0.08	0.05	0.03	-0.09	0.23
POND	0.20	0.01>	0.20	-0.02	0.44	-0.02
MINE1	0.21	-0.07	-0.26	-0.10	0.35	-0.04
MINE2	0.14	0.24	0.09	-0.09	0.28	0.01>
HWY	0.47	0.13	0.26	0.03	0.16	0.01>
CNRLs	4.28	1.23	-1.16	-0.77	-0.04	-0.03
CNRLw	-5.29	0.61	-0.64	-0.07	0.46	-0.16

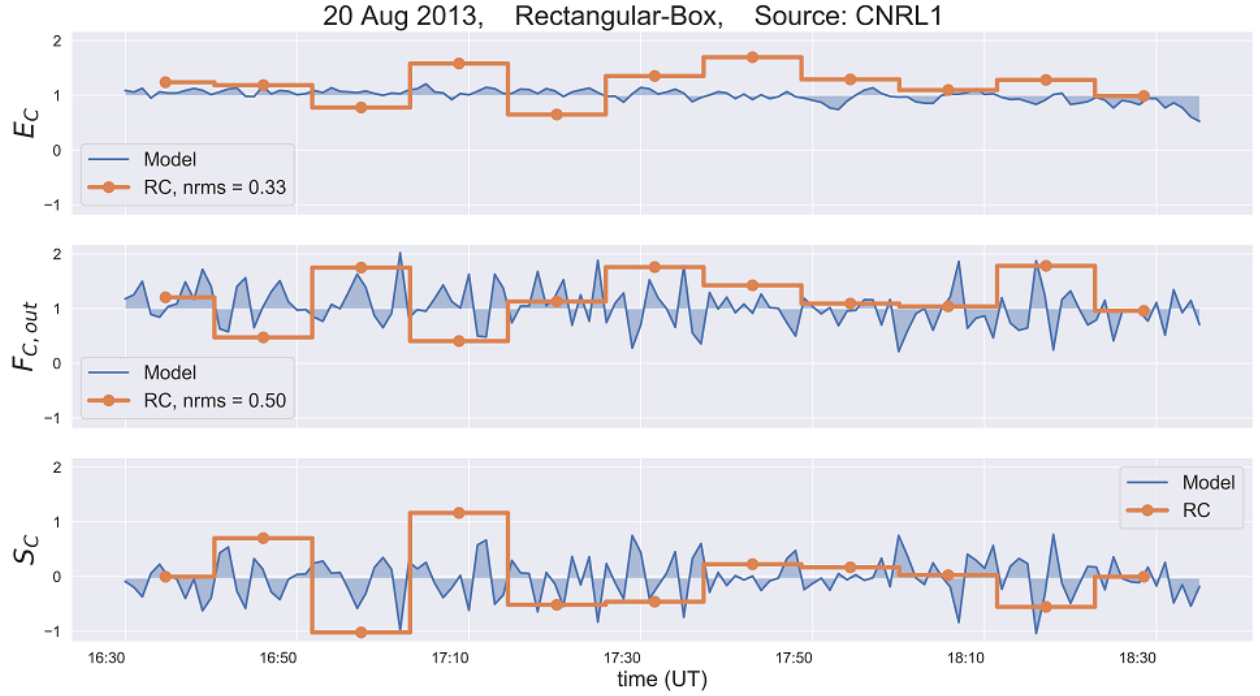


Figure 4.19: Rolling Column (RC) estimates for case 1, source CNRL1 and rectangular-box with S_C estimated based on Eq.4.3. Normalized root mean square (nrms) error scores are mentioned for RC generated E_C and $F_{C,out}$ estimates evaluated against model true values for the same terms.

improve the retrievals by 1-23% with median 12% for rectangular-box, 1-17% with median 9% for cylindrical-box, and 4-10% with median 8% for the downwind-screen. As we mentioned at the beginning of the Results and Discussion section (also see Chapter 3), the weak advection conditions

within the control volume during case 2 (26 August) resulted in weak tracer concentration signal on the sampling screens and substantial storage levels (up to 90% of E_C). The RC method, although providing a means for estimating S_C , relies on tracer concentration measurements on the sampling screens. This makes RC estimates of S_C less reliable during weak advection conditions like case 2, and therefore estimates for this case are not included here.

Table 4.3: Case 1 normalized RMS error of the RC generated $F_{C,out}$ and E_C estimates, with S_C estimated according to Eq. 4.3 with 10-12 laps, evaluated against model Truth. Note that values in [brackets] for CNRLw and CNRLs are rms values (not normalized) due to E_C being very small for the upwind cases (e.g., CNRLs $E_C = 0.002$).

	Rectangular Box		Cylindrical Box		Downwind Screen	
	$F_{C,out}$	E_C	$F_{C,out}$	E_C	$F_{C,out}$	E_C
CNRL1	0.50	0.33	0.33	0.29	0.58	0.55
CNRL2	0.56	0.43	0.30	0.20	0.61	0.55
CNRL4	0.52	0.39	0.40	0.34	0.60	0.57
CNRL3	0.53	0.37	0.63	0.55	0.75	0.68
POND	0.13	0.13	0.15	0.12	0.19	0.18
MINE1	0.30	0.28	0.26	0.20	0.30	0.23
MINE2	0.42	0.37	0.40	0.37	0.40	0.35
HWY	0.46	0.43	0.13	0.11	0.16	0.08
CNRLs	[2.70]	[2.24]	[2.95]	[2.64]	0.24	0.14
CNRLw	[3.01]	[1.08]	[3.67]	[0.88]	0.31	0.25

4.4 Conclusions

We considered model output fields from the super-resolution WRF model simulations in Chapter 3 as a proxy for real-world environmental fields for top-down retrievals. We simulated aircraft-based sampling by extracting model data in 3D space and over time along predetermined flight paths similar to those conducted in real-world airborne campaigns. We refer to this sampling strategy as Flying Point (FP). We analyzed FP sampled data following conventional data analysis methods for top-down emission rate retrieval campaigns using aircraft measurements. We evaluated the estimates against known model information (meteorology, emission rates) and determined a range of uncertainty for such estimates. Our results indicate an uncertainty range of about 30% to 50% under different meteorological conditions and for various tracer emission scenarios. Our estimated uncertainty range is comparable but somewhat larger than the estimates by similar previous studies (e.g., Angevine et al., 2020; Gordon et al., 2015) at 30% to 40%. Our results show that the temporal variability of the sampled fields over the sampling time, and the extrapolation of data below the lowest flight level represent the biggest sources of uncertainty. We also demonstrated how lack of information regarding the storage rate term in the mass-balance calculation can lead to higher uncertainty levels in top-down estimates based on FP sampling. We show that downwind distances between 10 km to 15 km represent optimal sampling radii where tracer mixing vertical profiles can be assumed as well-mixed, and the sampled data can be extrapolated below the flight level with low uncertainty. This is consistent with the 12-24 km range for well mixed conditions estimated by Vinuesa and Galmarini (2009). At 10-15 km distances, plume concentration profiles are also less temporally variable for the sampled heights, further reducing the uncertainty in flux calculations based on the sparsely sampled data points.

We discussed a new sampling strategy where aircraft-based in-situ and remote measurements can be combined for improved top-down retrievals. Note that remote measurements can also be achieved by ground-based remote-sensing instruments (e.g., land vehicles) and analyzed in conjunction with airborne measurements. However, synchronizing ground-based and airborne observations (e.g., sampling frequency, geographical range, collocation, road access) may pose additional challenges. We have shown that using remote column measurements to fill the information gap below the lowest flight level has the potential to improve the estimates by up to 33%. We also discussed extending remote column measurements for the entire vertical extent of the sampling volume/screen. This column information in combination with the in-situ measurements has the potential to generate flux box/screens covering heights from ground surface to the box top for shorter 10-15 min time intervals. We refer to this sampling strategy as the Rolling Column (RC). RC sampling has the potential to improve top-down estimates by generating flux box/screen data at higher temporal resolutions compared to conventional airborne retrievals. Using this temporal data, the rate of storage can be estimated based on aircraft measurements. This method provides the first practical means for estimating the storage rate, which can improve the estimates by up to 31% relative to known model emission rates. The RC sampling has the potential to reduce the required flight time for reliable source emission rate estimates, as enough information can be generated by flying only

a few laps (e.g., 20-60 min flight time) around/downwind the emission source. This can potentially reduce the operational costs of airborne observations.

We note that there are several limitations associated with airborne remote measurements (e.g., lidar). For instance, wind field measurements using a Doppler lidar have lower horizontal resolution compared to airborne in-situ measurements (Weissmann et al., 2005; Wekker et al., 2012) and retrieving near ground wind fields and pollutant concentrations can be challenging and error prone due to the reflection of the lidar pulse from the ground surface (Godwin et al., 2012). Therefore, airborne remote measurements need to be analyzed in conjunction with ground-based and aircraft-based in-situ measurements for reliable mass-balance estimates. The practical implementation of the RC method needs to be investigated further and instrumental/observational uncertainties associated with combining in-situ and remote measurements for top-down retrievals need to be explored. This method has the potential to provide a practical means for generating flux estimates at higher temporal resolution during an airborne measurement campaign and therefore a means (for the first time) for estimating the storage rate term for improved mass-balance calculations; while at the same time eliminating the main source of error which is the extrapolation of the sampled/interpolated data below the flight level.

Chapter 5

Conclusions

For this study, we conducted a multi-scale (from mesoscale to microscale) model-based study of aircraft-based emission rate retrieval methodologies using two different modelling platforms with realistic input information. Environment and Climate Change Canada’s (ECCC) air-quality model GEM-MACH was employed for air-quality model simulations at high resolution with 2.5 km model horizontal grid-spacing over the region of Canadian oil sands in Athabasca, Alberta. Realistic meteorological and source emission input information were used to generate environmental fields for the period of the JOSM 2013 airborne campaign over the oil sands region. We investigated the impact of meteorological variability on mass-balance emission rate retrieval accuracy, by using model simulated fields as a proxy for real-world chemical and meteorological fields, in which virtual aircraft sampling of the GEM-MACH output was used for top-down mass balance estimates. We also explored the impact of upwind emissions from nearby sources on the accuracy of the retrieved emission rates. This approach allowed the state of the atmosphere used for top-down estimates to be characterized in time and 3D space; the input meteorology and chemistry was “known”, and thus potential means for improving emission rate retrievals and determining the factors affecting retrieval accuracy was investigated. We demonstrated that emissions retrieval accuracy can be predicted via three key quantitative criteria, evaluated a priori from forecasts and/or from observations during the sampling period: (1) changes to the atmospheric stability (described as the change in gradient Richardson number), (2) variations in the direction of transport, as a result of plume vertical motion and in the presence of vertical wind shear, and (3) the upwind to downwind concentration ratio, when weighted by the ratio of upwind to downwind concentration standard deviations. We showed that cases where these criteria indicate high temporal variability and/or high upwind emissions can result in “storage-and-release” events within the sampled region (control volume), which decrease emission rate retrieval accuracy. Storage-and-release events may contribute the bulk of mass-balance emission rate retrieval under- and over-estimates, ranging in the tests carried out here from -166% to 24% of the known (input) emissions, with a median of -16%. Previous studies have estimated the uncertainty in airborne mass-balance estimates (under the assumption of steady-state conditions) between 30% to 40% (e.g., Angevine et al., 2020; Gordon et al., 2015). Our results suggest that on average up to 20% error in mass-balance estimates can result from storage-

and-release events (see Fathi et al., 2021).

The comprehensive study on mass-balance analysis applications using the regional air-quality model GEM-MACH at 2.5 km resolution, with physical and chemical process representations in on-line (feedback) mode, was successful in characterizing the mass-budget of pollutants under atmospheric transport and dispersion. This study identified the storage rate term as a major contributing term in the mass-balance equation and discussed the corresponding uncertainties in top-down estimates. In order to explore the other sources of uncertainty/error in airborne retrievals we needed to replicate the tracer transport and dispersion conditions at a much higher resolution, the time and length scales of airborne observations. We used the Weather Research and Forecasting (WRF) model to develop and implement super-resolution model simulations at 50 m horizontal resolution. We used North American Regional Reanalysis (NARR) data at 31 km resolution as input and boundary conditions. Using the WRF model, we performed dynamical down-scaling of the reanalysis data from 31 km to 50 m through numerical model nesting with 5 domains. For the two inner most domains at 250 m and 50 m resolutions, we used LES subgrid parameterization. The finest resolution domain was centred on the CNRL facility within the oil sands region. The WRF model dynamical solver source code was modified to simulate emission of passive-tracers from stack and surfaces sources within the CNRL facility, as well as a number of hypothetical sources. Simulation times were chosen for two days in August and one day in September 2013, corresponding to three airborne measurement flights during the JOSM campaign. Model-generated meteorological fields from the finest resolution domain were evaluated against aircraft observations during the corresponding JOSM 2013 flights, and also against historical observational data from two WBEA monitoring stations within the region of study. Model wind speeds and directions were in agreement with observational data within one standard deviation. Model temperature and humidity were within 5% of observational values. We evaluated the model performance in terms of global (over the entire modelling domain) and local (sub-domain) mass conservations for the tracer amounts emitted from several surface/area and stack/point sources. Results were within 2% to 5% of model input emissions for 10 out of 11 sources, indicating mass-conservation for our model simulations. For one hypothetical source on a non-regular (vertical) grid, the creation of erroneous concentrations resulted in mass imbalance and rejection of this source for further mass-balance analysis. We recommend source emission model simulations on a regular grid (both horizontal and vertical), especially for point sources with steep concentration gradients close to the point of emission. For our simulated sources, the WRF model transport step with the positive-definite filter was successful in correcting for spurious mass creation on a regular grid (fixed grid spacing). Our simulated tracer mass budget using WRF was consistent with the GEM-MACH generated fields at 2.5 km resolution for the same locations and time periods (Fathi et al., 2021). The model output from our super-resolution simulations are consistent with previous LES studies, both in terms of meteorology and generated wind fields (e.g., Daniels et al., 2016; Liu et al., 2011) and downwind tracer mixing conditions (e.g., Conley et al., 2017; Vinuesa and Galmarini, 2009).

Using the model output from WRF super-resolution simulations (at spatio-temporal scale

of airborne measurements) as a proxy for real-world environmental fields, we evaluated conventional aircraft-based emission rate retrievals. We simulated airborne in-situ measurements by sampling/extracting model generated fields along predetermined flight paths through 3D space and over time, at sampling resolutions and frequencies similar to real-world aircraft campaigns (through sampling strategy of flying point - FP). Following the conventional mass-balance methods (e.g., Ryoo et al., 2019), we analyzed the sampled wind and concentration data and generated emission rate estimates for various emission sources under different meteorological conditions. Our results indicate an uncertainty range of between 30% to 50% in top-down estimates. This range is comparable but larger than the uncertainty estimated by previous studies between 20% to 40% (e.g., Angevine et al., 2020; Gordon et al., 2015). We identified three main sources of uncertainty: (1) temporal variability of the sampled fields, (2) the gap of information below the flight level, and (3) lack of an estimate for the storage rate term in mass-balance analysis. We demonstrated that the first two sources of uncertainty can be addressed by optimizing the flight time and radius (sampling period and downwind distance). Aircraft-based estimates rely on calculation of the tracer mass flux downwind the emission source, using spatially sparse measurements of wind and concentrations over a flight time of several hours (i.e., FP). We demonstrated that the extent of uncertainty/error with respect to extended flight time can be determined through super-resolution modelling. This can be utilized for optimizing the flight time during airborne campaigns (e.g., 20 min to 80 min for the studied cases). Further we showed that at downwind distances between 10-15 km for stack and surface emissions, concentrations are well mixed in the 3D space, assume near constant profiles in the vertical, and are less variable over time. This is consistent with the 10-20 km estimated range from previous LES studies (e.g., Conley et al., 2017; Vinuesa and Galmarini, 2009). Sampled fields at such distances can be interpolated/extrapolated in vertical and horizontal direction at higher confidence levels. This in turn has the potential to minimize the large uncertainties associated with the extrapolation of sampled data below the flight level in mass-balance methodologies (Gordon et al., 2015).

We discussed a new sampling/retrieval strategy, the Rolling Column (RC), to address the major sources of uncertainty in estimates based on FP sampling. The RC method relies on combining in-situ and remote (e.g., lidar) measurements to fill the information gap in the sampled data below and/or between the flight tracks, and has the potential to generate flux estimates at shorter times (period of one lap/circuit around the emission source). This can reduce the required flight time for reliable mass-balance estimates (from several hours with FP sampling to several minutes). Further, the temporal data (concentrations, wind field) generated through RC sampling can be used to estimate the storage rate term for improved mass-balance analysis. We demonstrated that this approach can potentially improve the top-down estimates by over 30%. We note that there are limitations to the remote sensing technology compared to in-situ measurements, such as lower horizontal sampling resolution for airborne Doppler lidar measurements of wind fields (Weissmann et al., 2005; Wekker et al., 2012) and challenge in retrieving near surface values due to ground return of the lidar pulse that can obscure the back-scatter signal from aerosols (Godwin et al.,

2012). We recommend the analysis of remote observations as complementary information to in-situ measurements. While remote sensing might not provide the same capacity and accuracy for measuring tracer and aerosol concentrations, the characterization of emissions for a select number of species can be used to model the storage rate for the other sampled compounds. The RC method provides the first practical means for estimating the storage rate from aircraft observations. Further, it has the potential to address all three main sources of uncertainty in aircraft-based retrievals, while reducing the required flight time for reliable mass-balance estimates.

We conducted a multi-scale study of airborne mass-balance emission rate estimation methodologies using two different modelling platforms, and provided recommendations for improved sampling and data analysis approaches. As a follow up to the concepts explored here, the following is planned in terms of future research work: (a) model-based re-evaluation of two previous (JOSM) airborne campaigns in 2013 and 2018 over the Canadian oil sands region, (b) technical survey of aircraft-based remote measurement technologies and their potential field applications (future campaigns) with the RC method, and (c) developing machine learning applications such as utilizing clustering in plume discrimination on downwind sampling screens in order to substitute the need for upwind measurements. In this work we have demonstrated that transport conditions and different sampling approaches, and their impact on the accuracy of top-down estimates can be replicated through model simulations. This suggests that machine learning applications such as model-driven neural network training can potentially be utilized for improved top-down estimates, which we plan to undertake as part of a future research work.

Bibliography

- (2021a). American meteorological society, 2021: Ekman spiral. glossary of meteorology.
- (2021b). American meteorological society, 2021: Gradient richardson number. glossary of meteorology.
- Abdul-Razzak, H. and Ghan, S. J. (2002). A parameterization of aerosol activation 3. sectional representation. *Journal of Geophysical Research: Atmospheres*, 107(D3):AAC 1–1–AAC 1–6.
- Aggarwal, M., Whiteway, J., Seabrook, J., Gray, L., Strawbridge, K., Liu, P., O’Brien, J., Li, S.-M., and McLaren, R. (2018). Airborne lidar measurements of aerosol and ozone above the canadian oil sands region. *Atmospheric Measurement Techniques*, 11(6):3829–3849.
- Ahmadov, R., McKeen, S., Trainer, M., Banta, R., Brewer, A., Brown, S., Edwards, P. M., de Gouw, J. A., Frost, G. J., Gilman, J., Helmig, D., Johnson, B., Karion, A., Koss, A., Langford, A., Lerner, B., Olson, J., Oltmans, S., Peischl, J., Pétron, G., Pichugina, Y., Roberts, J. M., Ryerson, T., Schnell, R., Senff, C., Sweeney, C., Thompson, C., Veres, P. R., Warneke, C., Wild, R., Williams, E. J., Yuan, B., and Zamora, R. (2015). Understanding high wintertime ozone pollution events in an oil- and natural gas-producing region of the western us. *Atmospheric Chemistry and Physics*, 15(1):411–429.
- Akingunola, A., Makar, P. A., Zhang, J., Darlington, A., Li, S.-M., , M., Moran, M. D., and Zheng, Q. (2018). A chemical transport model study of plume-rise and particle size distribution for the athabasca oil sands. *Atmospheric Chemistry and Physics*, 18(12):8667–8688.
- Alfieri, S., Amato, U., Carfora, M., Esposito, M., and Magliulo, V. (2010). Quantifying trace gas emissions from composite landscapes: A mass-budget approach with aircraft measurements. *Atmospheric Environment*, 44(15):1866–1876.
- Andreae, M. O., Acevedo, O. C., Araùjo, A., Artaxo, P., Barbosa, C. G. G., Barbosa, H. M. J., Brito, J., Carbone, S., Chi, X., Cintra, B. B. L., da Silva, N. F., Dias, N. L., Dias-Júnior, C. Q., Ditas, F., Ditz, R., Godoi, A. F. L., Godoi, R. H. M., Heimann, M., Hoffmann, T., Kesselmeier, J., Könemann, T., Krüger, M. L., Lavric, J. V., Manzi, A. O., Lopes, A. P., Martins, D. L., Mikhailov, E. F., Moran-Zuloaga, D., Nelson, B. W., Nölscher, A. C., Santos Nogueira, D., Piedade, M. T. F., Pöhlker, C., Pöschl, U., Quesada, C. A., Rizzo, L. V., Ro, C.-U., Ruckteschler, N., Sá, L. D. A., de Oliveira Sá, M., Sales, C. B., dos Santos, R. M. N., Saturno, J., Schöngart, J., Sörgel, M., de Souza, C. M., de Souza, R. A. F., Su, H., Targhetta, N., Tóta, J., Trebs, I., Trumbore, S., van Eijck, A., Walter, D., Wang, Z., Weber, B., Williams, J., Winderlich, J., Wittmann, F., Wolff, S., and Yáñez Serrano, A. M. (2015). The amazon tall tower observatory (atto): overview of pilot measurements on ecosystem ecology, meteorology, trace gases, and aerosols. *Atmospheric Chemistry and Physics*, 15(18):10723–10776.
- Angevine, W. M., Peischl, J., Crawford, A., Loughner, C. P., Pollack, I. B., and Thompson, C. R. (2020). Errors in top-down estimates of emissions using a known source. *Atmospheric Chemistry and Physics*, 20(20):11855–11868.

- Baray, S., Darlington, A., Gordon, M., Hayden, K. L., Leithead, A., Li, S.-M., Liu, P. S. K., Mittermeier, R. L., Moussa, S. G., O'Brien, J., Staebler, R., Wolde, M., Worthy, D., and McLaren, R. (2018). Quantification of methane sources in the athabasca oil sands region of alberta by aircraft mass balance. *Atmospheric Chemistry and Physics*, 18(10):7361–7378.
- Barchyn, T. E., Hugenholtz, C. H., Myshak, S., and Bauer, J. (2018). A uav-based system for detecting natural gas leaks. *Journal of Unmanned Vehicle Systems*, 6(1):18–30.
- Barkley, Z. R., Lauvaux, T., Davis, K. J., Deng, A., Miles, N. L., Richardson, S. J., Cao, Y., Sweeney, C., Karion, A., Smith, M., Kort, E. A., Schwietzke, S., Murphy, T., Cervone, G., Martins, D., and Maasackers, J. D. (2017). Quantifying methane emissions from natural gas production in north-eastern pennsylvania. *Atmospheric Chemistry and Physics*, 17(22):13941–13966.
- Bell, T. M., Klein, P. M., Lundquist, J. K., and Waugh, S. (2021). Remote-sensing and radiosonde datasets collected in the san luis valley during the lapse-rate campaign. *Earth System Science Data*, 13(3):1041–1051.
- Bermejo, R. and Conde, J. (2002). A conservative quasi-monotone semi-lagrangian scheme. *Monthly Weather Review*, 130(2):423 – 430.
- Bilbro, J., Fichtl, G., Fitzjarrald, D., Krause, M., and Lee, R. (1984). Airborne doppler lidar wind field measurements. *Bulletin of the American Meteorological Society*, 65(4):348 – 359.
- Bélair, S., Brown, R., Mailhot, J., Bilodeau, B., and Crevier, L.-P. (2003a). Operational implementation of the isba land surface scheme in the canadian regional weather forecast model. part ii: Cold season results. *Journal of Hydrometeorology*, 4(2):371 – 386.
- Bélair, S., Crevier, L.-P., Mailhot, J., Bilodeau, B., and Delage, Y. (2003b). Operational implementation of the isba land surface scheme in the canadian regional weather forecast model. part i: Warm season results. *Journal of Hydrometeorology*, 4(2):352 – 370.
- Blaylock, B. K., Horel, J. D., and Crosman, E. T. (2017). Impact of Lake Breezes on Summer Ozone Concentrations in the Salt Lake Valley. *Journal of Applied Meteorology and Climatology*, 56(2):353–370.
- Briggs, G. A. (1984). Plume rise and buoyancy effects. *Atmospheric Science and Power Production, D. Randerson, Ed., U.S. Dept. of Energy DOE/TIC-27601, available from NTIS as DE84005177*, page 327–366.
- Brus, D., Gustafsson, J., Kemppinen, O., de Boer, G., and Hirsikko, A. (2021a). Atmospheric aerosol, gases, and meteorological parameters measured during the lapse-rate campaign by the finnish meteorological institute and kansas state university. *Earth System Science Data*, 13(6):2909–2922.
- Brus, D., Gustafsson, J., Vakkari, V., Kemppinen, O., de Boer, G., and Hirsikko, A. (2021b). Measurement report: Properties of aerosol and gases in the vertical profile during the lapse-rate campaign. *Atmospheric Chemistry and Physics*, 21(1):517–533.
- Cambaliza, M. O. L., Shepson, P. B., Caulton, D. R., Stirn, B., Samarov, D., Gurney, K. R., Turnbull, J., Davis, K. J., Possolo, A., Karion, A., Sweeney, C., Moser, B., Hendricks, A., Lauvaux, T., Mays, K., Whetstone, J., Huang, J., Razlivanov, I., Miles, N. L., and Richardson, S. J. (2014). Assessment of uncertainties of an aircraft-based mass balance approach for quantifying urban greenhouse gas emissions. *Atmospheric Chemistry and Physics*, 14(17):9029–9050.
- Cheng, Y., Li, S.-M., Liggio, J., Gordon, M., Darlington, A., Zheng, Q., Moran, M., Liu, P., and Wolde, M. (2020). Top-down determination of black carbon emissions from oil sand facilities in alberta, canada using aircraft measurements. *Environmental Science & Technology*, 54(1):412–418.

- Coats, Jr, C. J. (1996). High performance algorithms in the sparse matrix operator kernel emissions (smoke) modeling system.
- Conley, S., Faloona, I., Mehrotra, S., Suard, M., Lenschow, D. H., Sweeney, C., Herndon, S., Schwietzke, S., Pétron, G., Pifer, J., Kort, E. A., and Schnell, R. (2017). Application of gauss’s theorem to quantify localized surface emissions from airborne measurements of wind and trace gases. *Atmospheric Measurement Techniques*, 10(9):3345–3358.
- Côté, J., Desmarais, J.-G., Gravel, S., Méthot, A., Patoine, A., Roch, M., and Staniforth, A. (1998a). The Operational CMC-MRB Global Environmental Multiscale (GEM) Model. Part II: Results. *Monthly Weather Review*, 126(6):1397–1418.
- Côté, J., Gravel, S., Méthot, A., Patoine, A., Roch, M., and Staniforth, A. (1998b). The Operational CMC-MRB Global Environmental Multiscale (GEM) Model. Part I: Design Considerations and Formulation. *Monthly Weather Review*, 126(6):1373–1395.
- Cui, Y. Y., Brioude, J., McKeen, S. A., Angevine, W. M., Kim, S.-W., Frost, G. J., Ahmadov, R., Peischl, J., Bousseres, N., Liu, Z., Ryerson, T. B., Wofsy, S. C., Santoni, G. W., Kort, E. A., Fischer, M. L., and Trainer, M. (2015). Top-down estimate of methane emissions in california using a mesoscale inverse modeling technique: The south coast air basin. *Journal of Geophysical Research: Atmospheres*, 120(13):6698–6711.
- Daniels, M. H., Lundquist, K. A., Mirocha, J. D., Wiersema, D. J., and Chow, F. K. (2016). A New Vertical Grid Nesting Capability in the Weather Research and Forecasting (WRF) Model. *Monthly Weather Review*, 144(10):3725–3747.
- Davis, Z. Y. W., Baray, S., McLinden, C. A., Khanbabakhani, A., Fujs, W., Csukat, C., Debosz, J., and McLaren, R. (2019). Estimation of NO_x and SO₂ emissions from sarnia, ontario, using a mobile max-doas (multi-axis differential optical absorption spectroscopy) and a NO_x analyzer. *Atmospheric Chemistry and Physics*, 19(22):13871–13889.
- Davis, Z. Y. W., Frieß, U., Strawbridge, K. B., Aggarwal, M., Baray, S., Schnitzler, E. G., Lobo, A., Fioletov, V. E., Abboud, I., McLinden, C. A., Whiteway, J., Willis, M. D., Lee, A. K. Y., Brook, J., Olfert, J., O’Brien, J., Staebler, R., Osthoff, H. D., Mihele, C., and McLaren, R. (2020). Validation of max-doas retrievals of aerosol extinction, SO₂, and NO₂ through comparison with lidar, sun photometer, active doas, and aircraft measurements in the athabasca oil sands region. *Atmospheric Measurement Techniques*, 13(3):1129–1155.
- de Boer, G., Waugh, S., Erwin, A., Borenstein, S., Dixon, C., Shanti, W., Houston, A., and Argrow, B. (2021). Measurements from mobile surface vehicles during the lower atmospheric profiling studies at elevation – a remotely piloted aircraft team experiment (lapse-rate). *Earth System Science Data*, 13(1):155–169.
- de Grandpré, J., Tanguay, M., Qaddouri, A., Zerroukat, M., and McLinden, C. A. (2016). Semi-Lagrangian Advection of Stratospheric Ozone on a Yin–Yang Grid System. *Monthly Weather Review*, 144(3):1035–1050.
- DiBiase, D., D. M. J. A. K. K. L. A. T. P. B. and Wentz, E. (2006). Principles of kriging. the geographic information science & technology body of knowledge.
- Elston, J., Argrow, B., Stachura, M., Weibel, D., Lawrence, D., and Pope, D. (2015). Overview of small fixed-wing unmanned aircraft for meteorological sampling. *Journal of Atmospheric and Oceanic Technology*, 32(1):97–115.
- Fathi, S. (2017). Evaluating the Top-down Emission Rate Retrieval Algorithm (TERRA) Using Virtual Aircraft-based Sampling Within the GEM-MACH Model. Master’s thesis, York University.
- Fathi, S., Gordon, M., Makar, P. A., Akingunola, A., Darlington, A., Liggio, J., Hayden, K., and Li, S.-M. (2021). Evaluating the impact of storage-and-release on aircraft-based mass-balance methodology using a regional air-quality model. *Atmospheric Chemistry and Physics*, 21(20):15461–15491.

- Fillion, L., Tanguay, M., Lapalme, E., Denis, B., Desgagne, M., Lee, V., Ek, N., Liu, Z., Lajoie, M., Caron, J.-F., and Pag?, C. (2010). The canadian regional data assimilation and forecasting system. *Weather and Forecasting*, 25(6):1645 – 1669.
- Gasch, P., Wieser, A., Lundquist, J. K., and Kalthoff, N. (2020). An les-based airborne doppler lidar simulator and its application to wind profiling in inhomogeneous flow conditions. *Atmospheric Measurement Techniques*, 13(3):1609–1631.
- Girard, C., Plante, A., Desgagné, M., McTaggart-Cowan, R., Côté, J., Charron, M., Gravel, S., Lee, V., Patoine, A., Qaddouri, A., Roch, M., Spacek, L., Tanguay, M., Vaillancourt, P. A., and Zadra, A. (2014). Staggered Vertical Discretization of the Canadian Environmental Multiscale (GEM) Model Using a Coordinate of the Log-Hydrostatic-Pressure Type. *Monthly Weather Review*, 142(3):1183–1196.
- Godwin, K. S., Wekker, S. F. J. D., and Emmitt, G. D. (2012). Retrieving winds in the surface layer over land using an airborne doppler lidar. *Journal of Atmospheric and Oceanic Technology*, 29(4):487 – 499.
- Gong, S. L., Barrie, L. A., Blanchet, J.-P., von Salzen, K., Lohmann, U., Lesins, G., Spacek, L., Zhang, L. M., Girard, E., Lin, H., Leaitch, R., Leighton, H., Chylek, P., and Huang, P. (2003). Canadian aerosol module: A size-segregated simulation of atmospheric aerosol processes for climate and air quality models 1. module development. *Journal of Geophysical Research: Atmospheres*, 108(D1):AAC 3–1–AAC 3–16.
- Gong, S. L., Barrie, L. A., and Lazare, M. (2002). Canadian aerosol module (cam): A size-segregated simulation of atmospheric aerosol processes for climate and air quality models 2. global sea-salt aerosol and its budgets. *Journal of Geophysical Research: Atmospheres*, 107(D24):AAC 13–1–AAC 13–14.
- Gong, W., Makar, P., Zhang, J., Milbrandt, J., Gravel, S., Hayden, K., Macdonald, A., and Leaitch, W. (2015). Modelling aerosol–cloud–meteorology interaction: A case study with a fully coupled air quality model (gem-mach). *Atmospheric Environment*, 115:695 – 715.
- Goovaerts, P. (2019). Kriging interpolation. the geographic information science & technology body of knowledge.
- Gordon, M., Li, S.-M., Staebler, R., Darlington, A., Hayden, K., O’Brien, J., and Wolde, M. (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):3745–3765.
- Gordon, M., Makar, P. A., Staebler, R. M., Zhang, J., Akingunola, A., Gong, W., and Li, S.-M. (2018). A comparison of plume rise algorithms to stack plume measurements in the athabasca oil sands. *Atmospheric Chemistry and Physics*, 18(19):14695–14714.
- Haarig, M., Walser, A., Ansmann, A., Dollner, M., Althausen, D., Sauer, D., Farrell, D., and Weinzierl, B. (2019). Profiles of cloud condensation nuclei, dust mass concentration, and ice-nucleating-particle-relevant aerosol properties in the saharan air layer over barbados from polarization lidar and airborne in situ measurements. *Atmospheric Chemistry and Physics*, 19(22):13773–13788.
- Heintzenberg, J., Birmili, W., Otto, R., Andreae, M. O., Mayer, J.-C., Chi, X., and Panov, A. (2011). Aerosol particle number size distributions and particulate light absorption at the zotto tall tower (siberia), 2006–2009. *Atmospheric Chemistry and Physics*, 11(16):8703–8719.
- Hohn, M. E. (1991). An introduction to applied geostatistics: by edward h. isaaks and r. mohan srivastava, 1989, oxford university press, new york, 561 p., isbn 0-19-505012-6, isbn 0-19-505013-4 (paperback), 55.00cloth,35.00 paper (us). *Computers & Geosciences*, 17(3):471–473.

- Howell, S. G., Clarke, A. D., Freitag, S., McNaughton, C. S., Kapustin, V., Brekovskikh, V., Jimenez, J.-L., and Cubison, M. J. (2014). An airborne assessment of atmospheric particulate emissions from the processing of athabasca oil sands. *Atmospheric Chemistry and Physics*, 14(10):5073–5087.
- Islam, A., Shankar, A., Houston, A., and Detweiler, C. (2021). University of nebraska unmanned aerial system (uas) profiling during the lapse-rate field campaign. *Earth System Science Data*, 13(6):2457–2470.
- Jacobson, M. Z. (2005). *Fundamentals of Atmospheric Modeling*. Cambridge University Press, 2 edition.
- JOSM (2013). Joint oil sands monitoring plan, integrated monitoring plan for the oil sands, air quality component. page 72.
- Kalthoff, N., Corsmeier, U., Schmidt, K., Kottmeier, C., Fiedler, F., Habram, M., and Slemr, F. (2002). Emissions of the city of augsburg determined using the mass balance method. *Atmospheric Environment*, 36, Supplement 1:19 – 31. Evaluation of Modeled Emission Inventories of Ozone Precursors. A Case Study for an Urban Area (Augsburg, Germany).
- Karion, A., Lauvaux, T., Lopez Coto, I., Sweeney, C., Mueller, K., Gourdji, S., Angevine, W., Barkley, Z., Deng, A., Andrews, A., Stein, A., and Whetstone, J. (2019). Intercomparison of atmospheric trace gas dispersion models: Barnett shale case study. *Atmospheric Chemistry and Physics*, 19(4):2561–2576.
- Karion, A., Sweeney, C., Kort, E. A., Shepson, P. B., Brewer, A., Cambaliza, M., Conley, S. A., Davis, K., Deng, A., Hardesty, M., Herndon, S. C., Lauvaux, T., Lavoie, T., Lyon, D., Newberger, T., Pétron, G., Rella, C., Smith, M., Wolter, S., Yacovitch, T. I., and Tans, P. (2015). Aircraft-based estimate of total methane emissions from the barnett shale region. *Environmental Science & Technology*, 49(13):8124–8131. PMID: 26148550.
- Karion, A., Sweeney, C., Pétron, G., Frost, G., Michael Hardesty, R., Kofler, J., Miller, B. R., Newberger, T., Wolter, S., Banta, R., Brewer, A., Dlugokencky, E., Lang, P., Montzka, S. A., Schnell, R., Tans, P., Trainer, M., Zamora, R., and Conley, S. (2013). Methane emissions estimate from airborne measurements over a western united states natural gas field. *Geophysical Research Letters*, 40(16):4393–4397.
- Kia, S., Flesch, T. K., Freeman, B. S., and Aliabadi, A. A. (2022). Calculating gas emissions from open-pit mines using inverse dispersion modelling: A numerical evaluation using calpuff and cfd-ls. *Journal of Wind Engineering and Industrial Aerodynamics*, 226:105046.
- Lauvaux, T., Miles, N. L., Deng, A., Richardson, S. J., Cambaliza, M. O., Davis, K. J., Gaudet, B., Gurney, K. R., Huang, J., O’Keefe, D., Song, Y., Karion, A., Oda, T., Patarasuk, R., Razlivanov, I., Sarmiento, D., Shepson, P., Sweeney, C., Turnbull, J., and Wu, K. (2016). High-resolution atmospheric inversion of urban co2 emissions during the dormant season of the indianapolis flux experiment (influx). *Journal of Geophysical Research: Atmospheres*, 121(10):5213–5236.
- Lee, D. and Tsuei, Y. (1992). A formula for estimation of truncation errors of convection terms in a curvilinear coordinate system. *Journal of Computational Physics*, 98(1):90 – 100.
- Li, J. and Barker, H. W. (2005). A radiation algorithm with correlated-k distribution. part i: Local thermal equilibrium. *Journal of the Atmospheric Sciences*, 62(2):286 – 309.
- Li, S.-M., Leithead, A., Moussa, S. G., Liggio, J., Moran, M. D., Wang, D., Hayden, K., Darlington, A., Gordon, M., Staebler, R., Makar, P. A., Stroud, C. A., McLaren, R., Liu, P. S. K., O’Brien, J., Mittermeier, R. L., Zhang, J., Marson, G., Cober, S. G., Wolde, M., and Wentzell, J. J. B. (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–E3765.

- Li, X.-B., Wang, D., Lu, Q.-C., Peng, Z.-R., Fu, Q., Hu, X.-M., Huo, J., Xiu, G., Li, B., Li, C., Wang, D.-S., and Wang, H. (2018). Three-dimensional analysis of ozone and pm2.5 distributions obtained by observations of tethered balloon and unmanned aerial vehicle in shanghai, china. *Stochastic Environmental Research and Risk Assessment*, 32(5):1189–1203.
- Liggio, J., Li, S.-M., Staebler, R. M., Hayden, K., Darlington, A., Mittermeier, R. L., O’Brien, J., McLaren, R., Wolde, M., Worthy, D., and Vogel, F. (2019). Measured canadian oil sands co2 emissions are higher than estimates made using internationally recommended methods. *Nature Communications*, 10(1):1863.
- Liu, Y., Warner, T., Liu, Y., Vincent, C., Wu, W., Mahoney, B., Swerdlin, S., Parks, K., and Boehnert, J. (2011). Simultaneous nested modeling from the synoptic scale to the les scale for wind energy applications. *Journal of Wind Engineering and Industrial Aerodynamics*, 99(4):308–319. The Fifth International Symposium on Computational Wind Engineering.
- Mailhot, J. and Benoit, R. (1982). A finite-element model of the atmospheric boundary layer suitable for use with numerical weather prediction models. *Journal of Atmospheric Sciences*, 39(10):2249 – 2266.
- Makar, P., Bouchet, V., and Nenes, A. (2003). Inorganic chemistry calculations using hetv—a vectorized solver for the $\text{SO}_4\text{--NO}_3\text{--NH}_4^+$ system based on the isorropia algorithms. *Atmospheric Environment*, 37(16):2279–2294.
- Makar, P., Gong, W., Hogrefe, C., Zhang, Y., Curci, G., Žabkar, R., Milbrandt, J., Im, U., Balzarini, A., Baró, R., Bianconi, R., Cheung, P., Forkel, R., Gravel, S., Hirtl, M., Honzak, L., Hou, A., Jiménez-Guerrero, P., Langer, M., Moran, M., Pabla, B., Pérez, J., Pirovano, G., San José, R., Tuccella, P., Werhahn, J., Zhang, J., and Galmarini, S. (2015a). Feedbacks between air pollution and weather, part 2: Effects on chemistry. *Atmospheric Environment*, 115:499 – 526.
- Makar, P., Gong, W., Milbrandt, J., Hogrefe, C., Zhang, Y., Curci, G., Žabkar, R., Im, U., Balzarini, A., Baró, R., Bianconi, R., Cheung, P., Forkel, R., Gravel, S., Hirtl, M., Honzak, L., Hou, A., Jiménez-Guerrero, P., Langer, M., Moran, M., Pabla, B., Pérez, J., Pirovano, G., San José, R., Tuccella, P., Werhahn, J., Zhang, J., and Galmarini, S. (2015b). Feedbacks between air pollution and weather, part 1: Effects on weather. *Atmospheric Environment*, 115:442 – 469.
- Makar, P. A., Akingunola, A., Aherne, J., Cole, A. S., Aklilu, Y.-A., Zhang, J., Wong, I., Hayden, K., Li, S.-M., Kirk, J., Scott, K., Moran, M. D., Robichaud, A., Cathcart, H., Baratzedah, P., Pabla, B., Cheung, P., Zheng, Q., and Jeffries, D. S. (2018). Estimates of exceedances of critical loads for acidifying deposition in alberta and saskatchewan. *Atmospheric Chemistry and Physics*, 18(13):9897–9927.
- Makar, P. A., Akingunola, A., Chen, J., Pabla, B., Gong, W., Stroud, C., Sioris, C., Anderson, K., Cheung, P., Zhang, J., and Milbrandt, J. (2021). Forest-fire aerosol–weather feedbacks over western north america using a high-resolution, online coupled air-quality model. *Atmospheric Chemistry and Physics*, 21(13):10557–10587.
- Makar, P. A., Stroud, C., Zhang, J., Moran, M., Akingunola, A., Gong, W., Gravel, S., Pabla, B., Cheung, P., Zheng, Q., Marson, G., Li, S. M., Brook, J., Hayden, K., Liggio, J., Staebler, R., and Darlington, A. (2016). *High Resolution Model Simulations of the Canadian Oil Sands with Comparisons to Field Study Observations*, pages 503–508. Springer International Publishing, Cham.
- Mays, K. L., Shepson, P. B., Stirm, B. H., Karion, A., Sweeney, C., and Gurney, K. R. (2009). Aircraft-based measurements of the carbon footprint of indianapolis. *Environmental Science & Technology*, 43(20):7816–7823. PMID: 19921899.
- Milbrandt, J. A. and Morrison, H. (2016). Parameterization of cloud microphysics based on the prediction of bulk ice particle properties. part iii: Introduction of multiple free categories. *Journal of the Atmospheric Sciences*, 73(3):975 – 995.

- Milbrandt, J. A. and Yau, M. K. (2005a). A multimoment bulk microphysics parameterization. part i: Analysis of the role of the spectral shape parameter. *Journal of the Atmospheric Sciences*, 62(9):3051 – 3064.
- Milbrandt, J. A. and Yau, M. K. (2005b). A multimoment bulk microphysics parameterization. part ii: A proposed three-moment closure and scheme description. *Journal of the Atmospheric Sciences*, 62(9):3065 – 3081.
- Mölders, N., Butwin, M., Madden, J., Tran, H., Sassen, K., and Kramm, G. (2015). Theoretical investigations on mapping mean distributions of particulate matter, inert, reactive, and secondary pollutants from wildfires by unmanned air vehicles (uavs). *Open Journal of Air Pollution*, 04(03):149—174.
- Moran, M., Ménard, S., Talbot, D., Huang, P., Makar, P., Gong, w., Landry, H., Gravel, S., Gong, S., Crevier, L.-P., Kallaur, A., and Sassi, M. (2010). Particulate-matter forecasting with gem-mach15, a new canadian air-quality forecast model. *Air Pollution Modelling and Its Application XX*, pages 289–292M.
- Nahian, M. R., Nazem, A., Nambiar, M. K., Byerlay, R., Mahmud, S., Seguin, A. M., Robe, F. R., Ravenhill, J., and Aliabadi, A. A. (2020). Complex meteorology over a complex mining facility: Assessment of topography, land use, and grid spacing modifications in wrf. *Journal of Applied Meteorology and Climatology*, 59(4):769 – 789.
- Nambiar, M. K., Byerlay, R. A. E., Nazem, A., Nahian, M. R., Moradi, M., and Aliabadi, A. A. (2020). A tethered air blimp (tab) for observing the microclimate over a complex terrain. *Geoscientific Instrumentation, Methods and Data Systems*, 9(1):193–211.
- Nathan, B. J., Golston, L. M., O’Brien, A. S., Ross, K., Harrison, W. A., Tao, L., Lary, D. J., Johnson, D. R., Covington, A. N., Clark, N. N., and Zondlo, M. A. (2015). Near-field characterization of methane emission variability from a compressor station using a model aircraft. *Environmental Science & Technology*, 49(13):7896–7903. PMID: 26011292.
- Nygård, T., Tisler, P., Vihma, T., Pirazzini, R., Palo, T., and Kouznetsov, R. (2017). Properties and temporal variability of summertime temperature inversions over dronning maud land, antarctica. *Quarterly Journal of the Royal Meteorological Society*, 143(702):582–595.
- Onishi, R., Sugiyama, D., and Matsuda, K. (2019). Super-resolution simulation for real-time prediction of urban micrometeorology. *SOLA*, 15:178–182.
- Panitz, H.-J., Nester, K., and Fiedler, F. (2002). Mass budget simulation of nox and co for the evaluation of calculated emissions for the city of augsburg (germany). *Atmospheric Environment*, 36, Supplement 1:33 – 51. Evaluation of Modeled Emission Inventories of Ozone Precursors. A Case Study for an Urban Area (Augsburg, Germany).
- Peischl, J., Ryerson, T. B., Holloway, J. S., Parrish, D. D., Trainer, M., Frost, G. J., Aikin, K. C., Brown, S. S., Dubé, W. P., Stark, H., and Fehsenfeld, F. C. (2010). A top-down analysis of emissions from selected texas power plants during texaq 2000 and 2006. *Journal of Geophysical Research: Atmospheres*, 115(D16).
- Ražnjević, A., van Heerwaarden, C., van Stratum, B., Hensen, A., Velzeboer, I., van den Bulk, P., and Krol, M. (2022). Technical note: Interpretation of field observations of point-source methane plume using observation-driven large-eddy simulations. *Atmospheric Chemistry and Physics*, 22(10):6489–6505.
- Ryoo, J.-M., Iraci, L. T., Tanaka, T., Marrero, J. E., Yates, E. L., Fung, I., Michalak, A. M., Tadić, J., Gore, W., Bui, T. P., Dean-Day, J. M., and Chang, C. S. (2019). Quantification of CO₂ and CH₄ emissions over sacramento, california, based on divergence theorem using aircraft measurements. *Atmospheric Measurement Techniques*, 12(5):2949–2966.
- Serafin, R. J., Szejwach, G., and Phillips, B. B. (1986). Applications of airborne remote sensing in atmospheric sciences research. *Journal of Geophysical Research: Oceans*, 91(C2):2510–2516.

- Skamarock, W. C., K. J. B. D. J. G. D. O. B. D. D. M. G. . P. J. G. (2008). A description of the advanced research wrf version 3 (no. ncar/tn-475+str). *University Corporation for Atmospheric Research*.
- Skamarock, W. C. and Weisman, M. L. (2009). The impact of positive-definite moisture transport on nwp precipitation forecasts. *Monthly Weather Review*, 137(1):488–494.
- Slemr, F., Friedrich, R., and Seiler, W. (2002). The research project eva - general objectives and main results. *Atmospheric Environment*, 36(SUPPL. 1):1–6. Cited By :4.
- Sørensen, B., Kaas, E., and Korsholm, U. S. (2013). A mass-conserving and multi-tracer efficient transport scheme in the online integrated enviro-hirlam model. *Geoscientific Model Development*, 6(4):1029–1042.
- Stockwell, W. and Lurmann, F. (1989). Intercomparison of the adom and radm gas-phase chemical mechanisms.
- Stull, R. (2017). *Practical Meteorology: An Algebra-based Survey of Atmospheric Science*. Univ. of British Columbia, 3 edition.
- Tadić, J. M., Michalak, A. M., Iraci, L., Ilić, V., Biraud, S. C., Feldman, D. R., Bui, T., Johnson, M. S., Loewenstein, M., Jeong, S., Fischer, M. L., Yates, E. L., and o, J.-M. (2017). Elliptic cylinder airborne sampling and geostatistical mass balance approach for quantifying local greenhouse gas emissions. *Environmental Science & Technology*, 51(17):10012–10021. PMID: 28727429.
- Thorpe, A. K., O’Handley, C., Emmitt, G. D., DeCola, P. L., Hopkins, F. M., Yadav, V., Guha, A., Newman, S., Herner, J. D., Falk, M., and Duren, R. M. (2021). Improved methane emission estimates using aviris-ng and an airborne doppler wind lidar. *Remote Sensing of Environment*, 266:112681.
- Turnbull, J. C., Karion, A., Fischer, M. L., Faloona, I., Guilderson, T., Lehman, S. J., Miller, B. R., Miller, J. B., Montzka, S., Sherwood, T., Saripalli, S., Sweeney, C., and Tans, P. P. (2011). Assessment of fossil fuel carbon dioxide and other anthropogenic trace gas emissions from airborne measurements over sacramento, california in spring 2009. *Atmospheric Chemistry and Physics*, 11(2):705–721.
- Vinuesa, J.-F. and Galmarini, S. (2009). Turbulent dispersion of non-uniformly emitted passive tracers in the convective boundary layer. *Boundary-Layer Meteorology*, 133(1):1–16.
- Watson, C. D., Wang, C., Lynar, T., and Weldemariam, K. (2020). Investigating two super-resolution methods for downscaling precipitation: Esrgan and car.
- Weissmann, M., Braun, F. J., Gantner, L., Mayr, G. J., Rahm, S., and Reitebuch, O. (2005). The alpine mountain–plain circulation: Airborne doppler lidar measurements and numerical simulations. *Monthly Weather Review*, 133(11):3095 – 3109.
- Wekker, S. F. J. D., Godwin, K. S., Emmitt, G. D., and Greco, S. (2012). Airborne doppler lidar measurements of valley flows in complex coastal terrain. *Journal of Applied Meteorology and Climatology*.
- Wu, Y., Teufel, B., Sushama, L., Belair, S., and Sun, L. (2021). Deep learning-based super-resolution climate simulator-emulator framework for urban heat studies. *Geophysical Research Letters*, 48(19):e2021GL094737. e2021GL094737 2021GL094737.
- You, D., Mittal, R., Wang, M., and Moin, P. (2006). Analysis of stability and accuracy of finite-difference schemes on a skewed mesh. *Journal of Computational Physics*, 213(1):184 – 204.
- Zhang, J., Moran, M. D., Zheng, Q., Makar, P. A., Baratzadeh, P., Marson, G., Liu, P., and Li, S.-M. (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):10459–10481.

Appendices

Appendix A GEM-MACH configuration & Mass-balance terms

A.1 Derivations for mass-balance terms

The net horizontal flux exiting through lateral walls of the box ($E_{C,H}$) is calculated by extracting compound mixing ratio (χ_C), air density (ρ_{air}) and the wind normal to the lateral walls of the box from the model output, and integrating over the surface (2D screen) comprised of box lateral walls at each time step,

$$E_{C,H}(t) = M_R \iint \chi_C(t, s, z) \rho_{\text{air}} U_{\perp} ds dz, \quad (\text{A.1})$$

where $s(x, y)$ is the screen path coordinate, which follows the horizontal projection of the box flight path around the facility in a counter-clockwise direction.

The normal wind (positive outwards) along the screen is calculated as

$$U_{\perp}(t, s, z) = \frac{U \frac{ds(x,y)}{dy} - V \frac{ds(x,y)}{dx}}{\sqrt{\left(\frac{ds(x,y)}{dx}\right)^2 + \left(\frac{ds(x,y)}{dy}\right)^2}}, \quad (\text{A.2})$$

where U and V are wind components towards east (x) and north (y), respectively.

The vertical flux through the box top can be calculated as

$$E_{C,V}(t) = M_R \iint \chi_{C,\text{top}} \rho_{\text{air},\text{top}} W_{\text{top}} dx dy, \quad (\text{A.3})$$

where $\chi_{C,\text{top}}(t, x, y)$ and $W_{\text{top}}(t, x, y)$ are compound mixing ratio and the vertical wind component (positive upwards) at the box top. In the observational application of the retrieval methodology, the top of the box corresponds to an aircraft measurement altitude which is above the plume.

The horizontal air flux through the screen (positive outwards) is calculated from extracted air density and wind fields along the screen at time t as

$$E_{\text{air},H}(t) = \iint \rho_{\text{air}}(t, s, z) U_{\perp} ds dz. \quad (\text{A.4})$$

The rate of change in the air mass (accumulation) within the box due to changing air density

(compressible fluid) can be estimated as

$$E_{\text{air,M}}(t) = \iiint \frac{\partial \rho_{\text{air}}(t, V)}{\partial t} dV. \quad (\text{A.5})$$

The vertical air flux leaving the box (positive upwards) is calculated similar to Eq. (A.3),

$$E_{\text{air,V}}(t) = \iint \rho_{\text{air,top}}(t, x, y) W_{\text{top}}(t, x, y) dx dy. \quad (\text{A.6})$$

A.2 Discrete integral expressions (Tables A.2–A.4)

The numerical integration expressions for calculating the terms in Eqs. (2.7), (2.12) and (2.16), using model data, are described in Tables A.2–A.4. Throughout, the second-order central finite-difference scheme Jacobson (2005) was used to numerically solve the time derivatives ($\Delta/\Delta t$) with the residual error of order $O(\Delta t^2)$, where Δt is equal to the 2 min (1/30 h) model time step. Surface deposition rates ($E_{\text{C,VD}}^t$) were extracted directly from the model output for the area covered by the base of the flux box for each case, to eliminate deposition estimates as a potential source of error in retrievals. Note that in Table A.3, the species vertical mass flux ($E_{\text{C,V}}^t$) is calculated using the estimated vertical air mass flux ($E_{\text{air,V}}^t$) from the mass-balance equation for air (Eq. 2.8) and the box top average mixing ratio ($\bar{\chi}_{\text{C,top}}^t$).

Table A.1: GEM-MACH Configuration details and references

Model Component	Description	Reference
Driving numerical weather prediction model	Global Environmental Multiscale (GEM), v5.0	Côté et al. (1998a,b); Girard et al. (2014)
Driving air-quality model	Global Environmental Multiscale – Modelling Air-quality and Chemistry (GEM-MACH) v2	Moran et al. (2010); Makar et al. (2015a,b)
Model grid and nesting, time stepping	North American 10km res. parent domain provides boundary conditions for 2.5km high-res., 64 vertical staggered hybrid level (terrain-following coordinate) 2.5km domain (Figure 2.1a). Time stepping: Parent domain: 5 min physics, 15 min chemistry. 2.5km domain: 1 min physics and 2 min chemistry.	Girard et al. (2014); Makar et al. (2015a,b)
Weather/Aerosol feedbacks	Direct effect via binary water-dry aerosol mix with <i>Mie</i> algorithm optical property calculations. Indirect effect via aerosols providing cloud condensation nuclei via the Abdul-Razzak and Ghan scheme.	Abdul-Razzak and Ghan (2002); Gong et al. (2015); Makar et al. (2015a,b)
Gas-phase chemistry	Acid Deposition and Oxidant Mechanism, version 2 (ADOM-II), gas-phase chemistry for 42 gas species, integrated using a Young and Boris solver.	Stockwell and Lurmann (1989)
Particle microphysics	Sectional approach – 8 particle species (sulphate, nitrate, ammonium, primary organic carbon, secondary organic carbon, black carbon, sea salt, crustal material), and 12 particle bins.	Gong et al. (2003, 2002)
Aqueous chemistry and gas and aerosols scavenging	Cloud scavenging of gases and aerosols, aqueous phase chemistry using a Young and Boris solver (combined time-resolved and steady-state chem.).	Gong et al. (2015)
Deposition	Gas (Robichaud scheme) and particle dry deposition (Zhang scheme)	Makar et al. (2018)
Particle inorganic thermodynamics	Sulphate-nitrate-ammonium non-ideal (high concentration) thermodynamic equilibrium system solved using a nested iterative approach.	Makar et al. (2003)
Anthropogenic plume rise	Modified Briggs (1984) eqs. for calculating plume rise. CEMS stack parameters data (e.g. stack height and diameter, exit temperature, exit velocity) were used for plume rise calculations in the model simulations.	Akingunola et al. (2018); Gordon et al. (2018)
Advection and mass conservation	Chemical transport in GEM-MACH is solved utilizing an implicit semi-Lagrangian (SL) advection space-time integration scheme. SL is not inherently mass conserving, and thus requires the use of a post advection mass conservation step (3D ILMC used here).	Bermejo & Conde 2002; Sørensen et al. (2013); de Grandpré et al. (2016)
Emissions data	Emissions are processed based upon the Sparse Matrix Operator Kernel Emissions – SMOKE; emissions data from hybrid Oil Sands Database. Large stack data derived from Continuous Emissions Monitoring (CEMS).	Coats (1996); Zhang et al. (2018)

Table A.2: Discrete integral expressions of Eqs. (2.4), (2.5), (A.1) and (A.3) to be substituted in Eq. (2.7). Constants Δx , Δy and Δs are equal to model grid resolution in the horizontal, 2.5 km. t is the time index; i , j and k are the 3D-space indices and s is the path index around the box.

Term	Numerical integration
$E_{C,S}^t =$	$M_R \Delta x \Delta y \sum_{i,j,k}^{n_x, n_y, n_z} \frac{\Delta \chi_{C,ijk}^t}{\Delta t} \rho_{air,ijk}^t \Delta z_k$
$E_{C,M}^t =$	$-M_R \Delta x \Delta y \sum_{i,j,k}^{n_x, n_y, n_z} \chi_{C,ijk}^t \frac{\Delta \rho_{air,ijk}^t}{\Delta t} \Delta z_k$
$E_{C,H}^t =$	$M_R \Delta s \sum_{s,k}^{n_s, n_z} \chi_{C,sk}^t \rho_{air,sk}^t U_{\perp,sk}^t \Delta z_k$
$E_{C,V}^t =$	$M_R \Delta x \Delta y \sum_{i,j}^{n_x, n_y} \chi_{C,ij}^{t,top} \rho_{air,ij}^{t,top} W_{ij}^{t,top}$

Table A.3: Discrete integral expressions for Eqs. (A.4) and (2.8)–(2.10) to be substituted in Eq. (2.12) for TR method estimations. Terms and variables are listed in the order needed for calculating the dependent terms. Δs is constant and equal to model horizontal grid resolution, 2.5 km. t and s are time and path indices, respectively. A is the box base area. Point (i_o, j_o) is at the box horizontal center.

Term	Numerical integration
$E_{air,H}^t =$	$\Delta s \sum_{s,k}^{n_s, n_z} \rho_{air,sk}^t U_{\perp,sk}^t \Delta z_k$
$E_{air,M}^t =$	$A \sum_k^{n_z} \frac{\Delta}{\Delta t} \rho_{air,i_o j_o k}^t \Delta z_k$
$\bar{\chi}_{C,k}^t =$	$\frac{1}{n_s} \sum_s^{n_s} \chi_{C,sk}^t$
$E_{C,M}^t =$	$-AM_R \sum_k^{n_z} \bar{\chi}_{C,k}^t \frac{\Delta}{\Delta t} \rho_{air,i_o j_o k}^t \Delta z_k$
$E_{air,V}^t =$	$-E_{air,M}^t - E_{air,H}^t$
$\bar{\chi}_{C,top}^t =$	$\frac{1}{n_s} \sum_s^{n_s} \chi_{C,s}^{t,top}$
$E_{C,V}^t =$	$AM_R \bar{\chi}_{C,top}^t E_{air,V}^t$

Table A.4: Discrete integral expressions for Eqs. (2.13)–(2.15) to be substituted in Eq. (2.16) for TR* method estimations. t is the time index; s and k are the horizontal and vertical space indices along the screens, respectively. A is the box base area.

Term	Numerical integration
$\bar{\rho}_{air,k}^t =$	$\frac{1}{n_s} \sum_s^{n_s} \rho_{air,sk}^t$
$E_{air,M}^t =$	$A \sum_k^{n_z} \frac{\Delta \bar{\rho}_{air,k}^t}{\Delta t} \Delta z_k$
$E_{C,M}^{t*} =$	$-AM_R \sum_{sk}^{n_s, n_z} \bar{\chi}_{C,sk}^t \frac{\Delta \bar{\rho}_{air,sk}^t}{\Delta t} \Delta z_k$
$E_{C,S}^{t*} =$	$AM_R \sum_{sk}^{n_s, n_z} \bar{\rho}_{air,sk}^t \frac{\Delta \bar{\chi}_{C,sk}^t}{\Delta t} \Delta z_k$

Appendix B Storage-release as a function of downwind distance

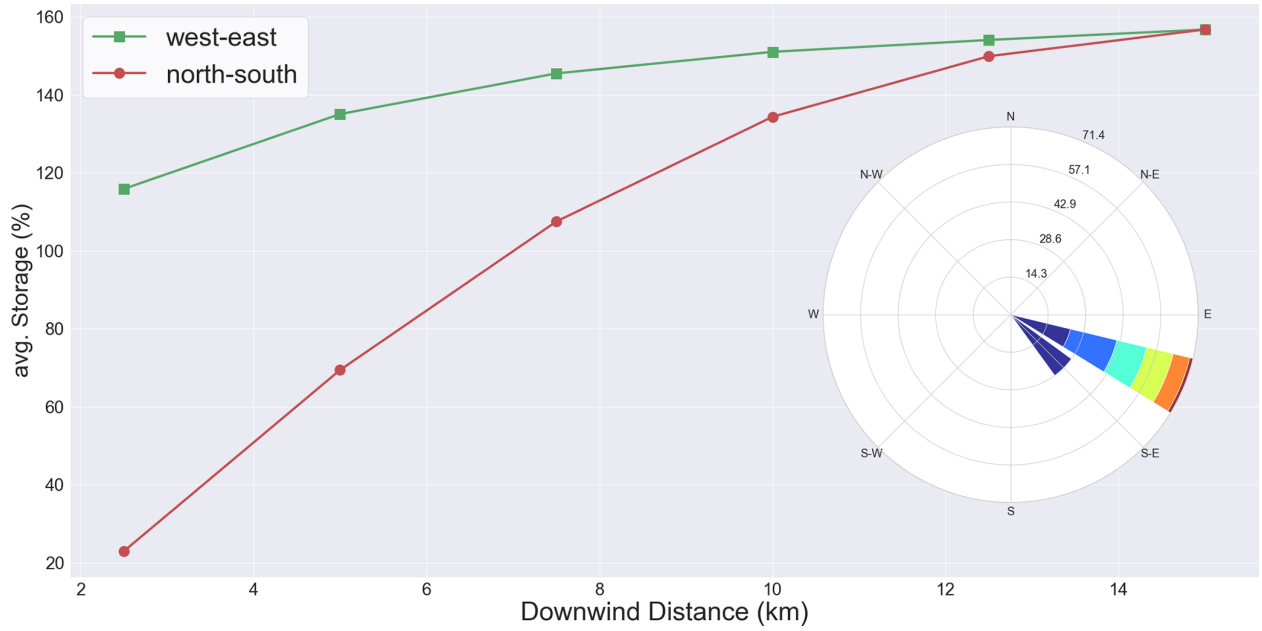


Figure B.1: The contribution of the storage rate term ($E_{C,S}$) to the net retrieved emission rates (E_C) as a function of downwind distance for case 8. As shown in the wind-rose diagram, the wind direction for this case was towards south-east (S-E), and therefore the downwind dependence is shown in two directions west-east (green squares) and north-south (red circles). The contribution of storage increases as a function of downwind distance. This suggests that the impact of storage release may increase as the distance between the downwind screen (sampling location) and the emission sources increase. A similar dependence of the magnitude of the storage release with downwind distance was observed for the other cases.

Appendix C Prognostic TKE Closure

The WRF model eddy viscosity (diffusivity) for the predicted turbulent kinetic energy option (K option $km_opt=2$), are computed using (Skamarock, 2008),

$$K_{h,v} = C_k l_{h,v} \sqrt{e} \quad (\text{C.1})$$

where e is the turbulent kinetic energy (prognostic in this scheme), C_k is a constant (typically $0.15 < C_k < 0.25$), and l is a length scale given as follows for the anisotropic option,

$$l_h = \sqrt{\Delta x \Delta y} \quad (\text{C.2})$$

and,

$$l_v = \min [\Delta z, 0.76\sqrt{e}/N] \quad \text{for } N^2 > 0 \quad (\text{C.3})$$

$$l_v = \Delta z \quad \text{for } N^2 \leq 0. \quad (\text{C.4})$$

where N is Brunt-Väisälä frequency, see Skamarock (2008) for derivations. The eddy viscosity used for mixing scalars is divided by a turbulent Prandtl number P_r . The Prandtl number is 1/3 for the horizontal eddy viscosity K_h , and $P^{-1} = 1 + 2l/\Delta z$ for the vertical eddy viscosity K_v . Note that the above are for the anisotropic mixing option (mix_isotropic = 0, default) that was used in our WRF simulations.

Appendix D Mass-balance turbulent flux terms

Derivations for turbulent flux terms in mass-balance equation Eq. 3.1,

$$F_{C,HT}(t) = - \iint K_h \frac{d\chi_C}{dx_\perp}(t, s, z) ds dz \quad (\text{D.1})$$

$$F_{C,VT}(t) = - \iint K_v \frac{d\chi_C}{dz_\perp}(t, x, y) dx dy \quad (\text{D.2})$$

where K_h and K_v are horizontal and vertical eddy diffusivity coefficients, respectively (see Appendix C). $d\chi_C/dx_\perp$ and $d\chi_C/dz_\perp$ are tracer concentration gradients across box lateral and top walls.

Appendix E WRF model output meteorology and tracer profiles

E.1 Case 1: August 20, 2013

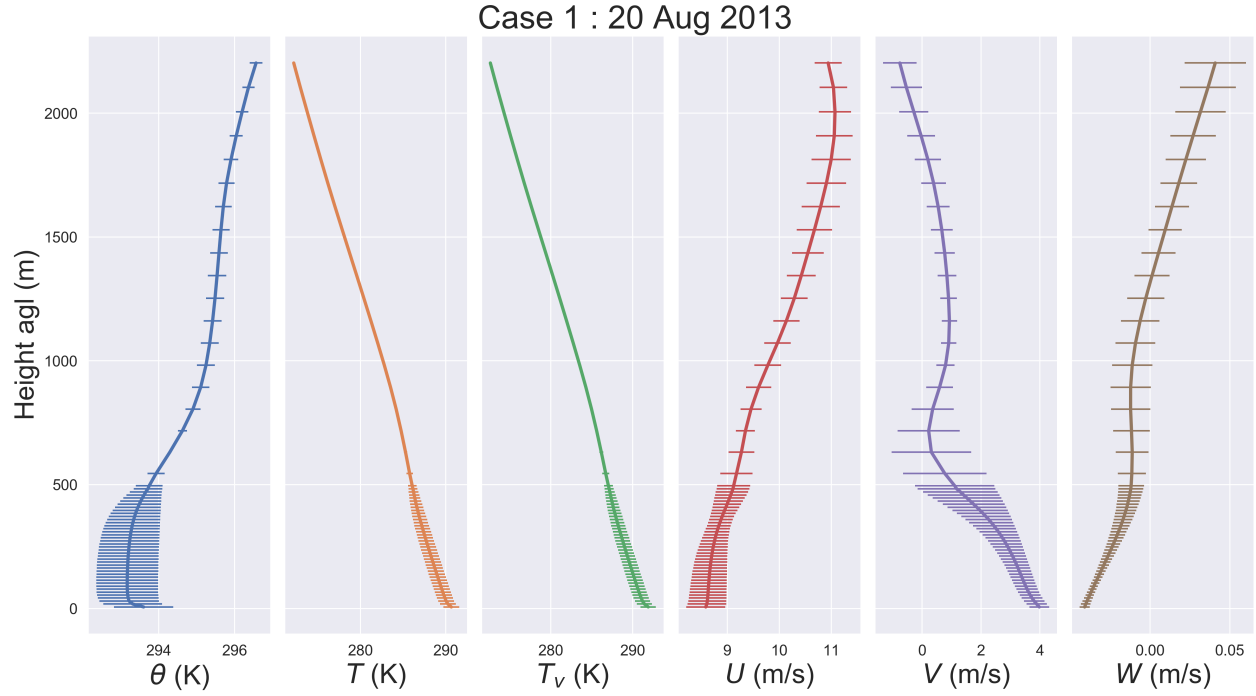


Figure E.1: Case 1 mean meteorological profiles on 20 August 2013: potential temperature θ , absolute temperature T , virtual temperature T_v , west-east wind component U , south-north wind component V , and vertical wind W . Bars show standard deviations during the simulation time.

E.2 Case 2: August 26, 2013

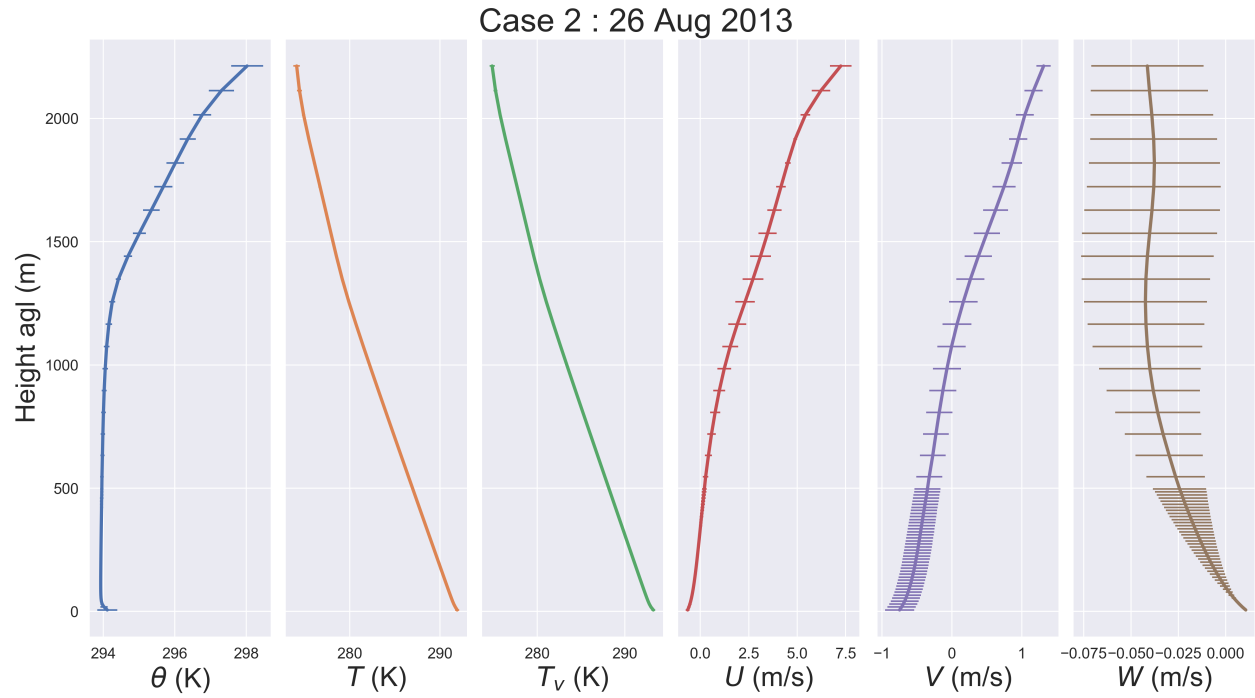


Figure E.2: Case 2 mean meteorological profiles on 26 August 2013: potential temperature θ , absolute temperature T , virtual temperature T_v , west-east wind component U , south-north wind component V , and vertical wind W . Bars show standard deviations during the simulation time.

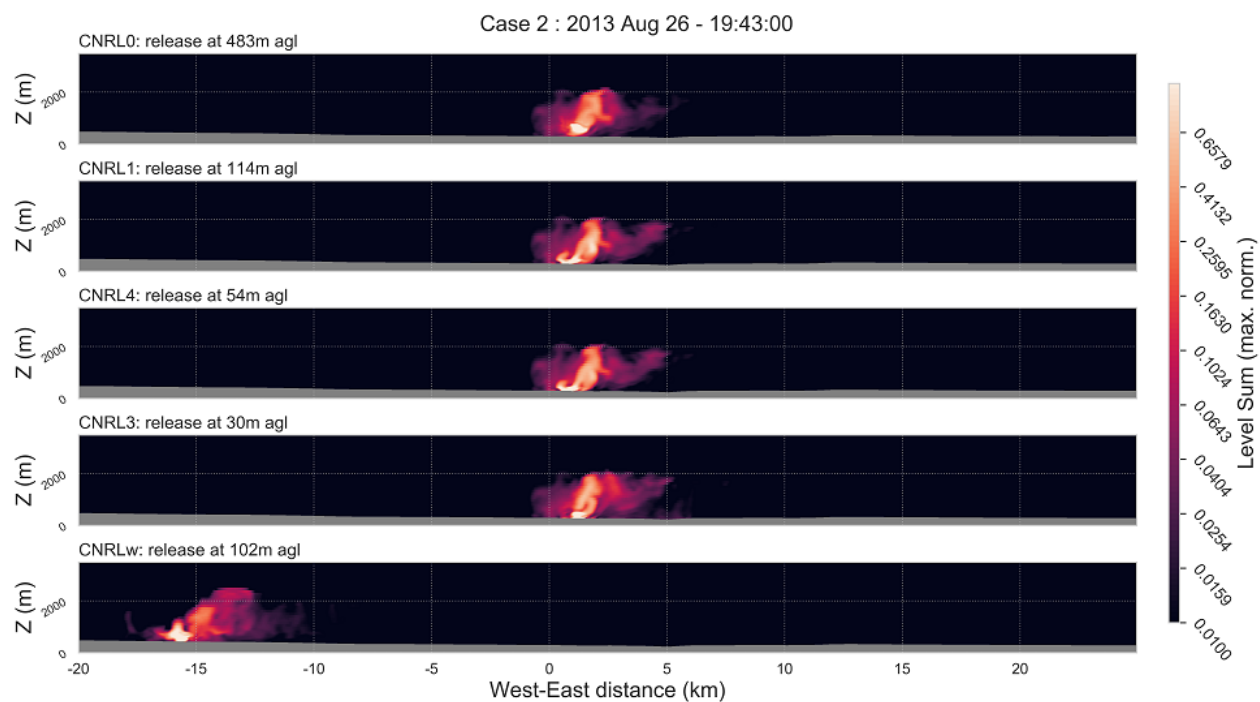


Figure E.3: Domain d05 west-east vertical cross-section for case 2 on 26 August 2013. Vertical cross-section of tracer plumes from stack/point emission sources are shown. Release height above ground is indicated for each source. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for west-east distance in km is at domain centre.

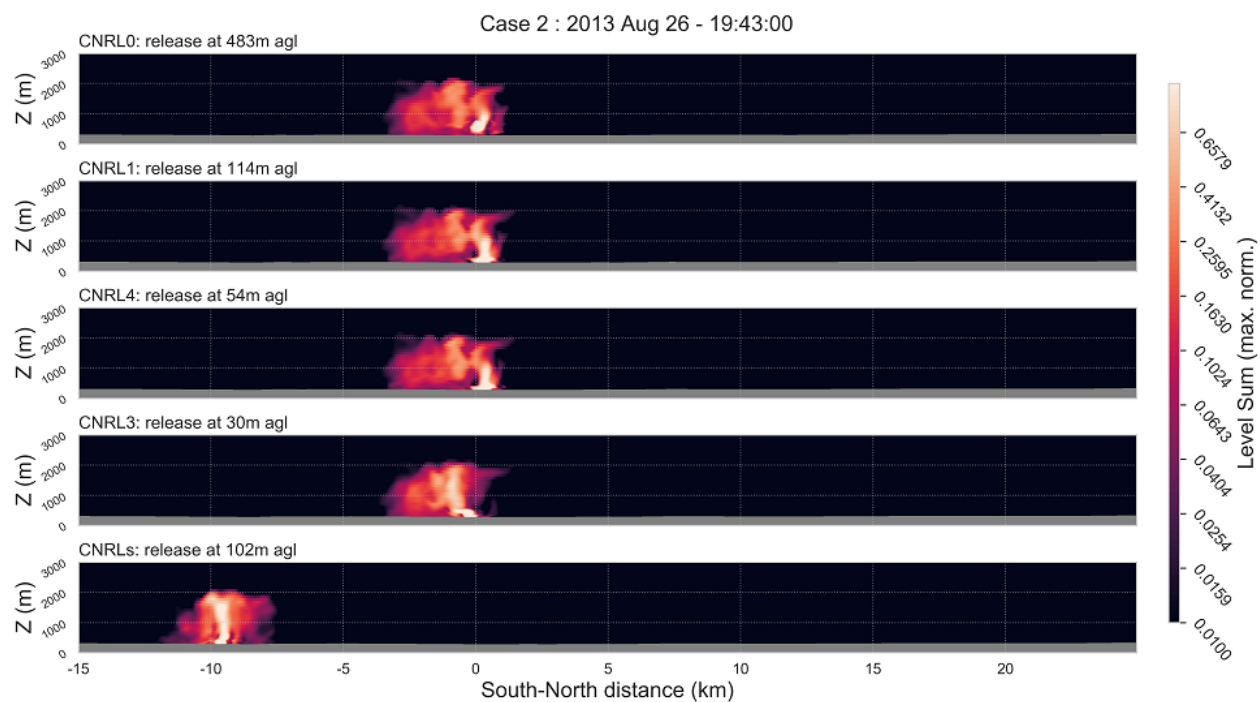


Figure E.4: Domain d05 south-north vertical cross-section for case 2 on 26 August 2013. Vertical cross-section of tracer plumes from stack/point emission sources are shown with release heights indicated for each source. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for south-north distance in km is at domain centre.

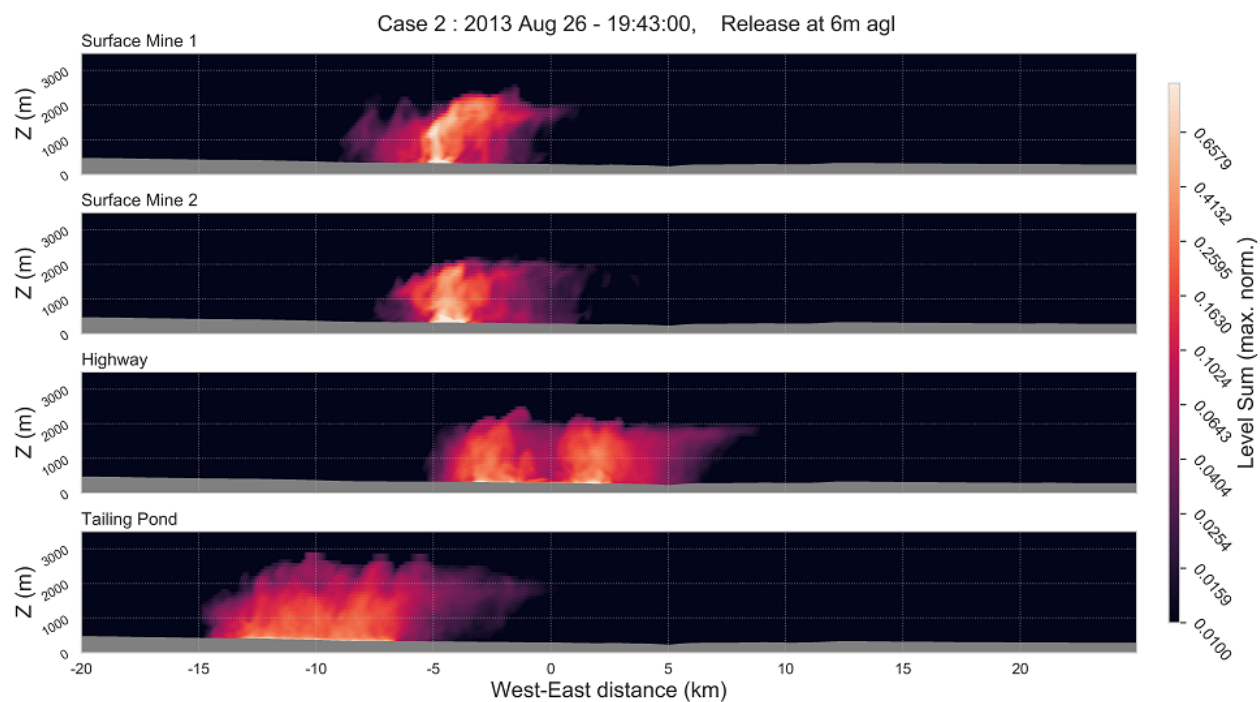


Figure E.5: Domain d05 west-east vertical cross-section for case 2 on 26 August 2013. Vertical cross-section of tracer plumes from surface emission sources (area, line) are shown. Release height above ground is 6 m agl for all sources. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for west-east distance in km is at domain centre.

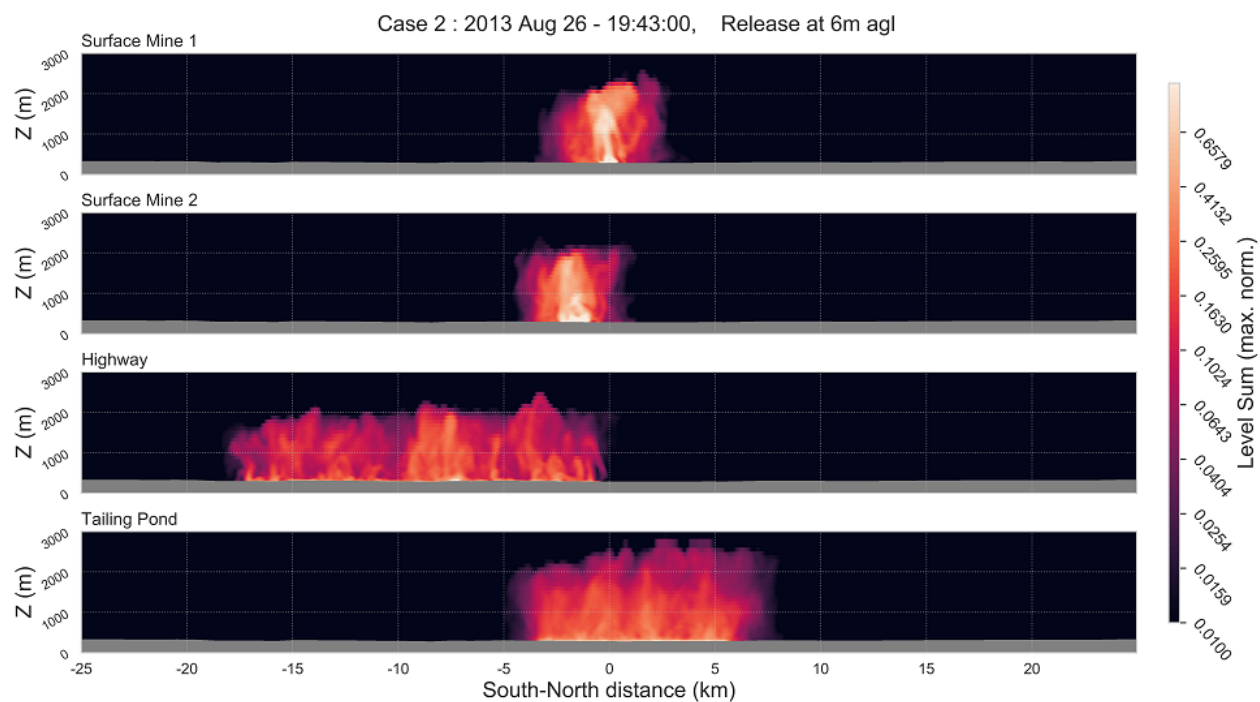


Figure E.6: Domain d05 south-north vertical cross-section for case 1 on 20 August 2013. Vertical cross-section of tracer plumes from surface emission sources (area, line) are shown with release heights at 6 m agl for all sources. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for south-north distance in km is at domain centre.

E.3 Case 3: September 2, 2013

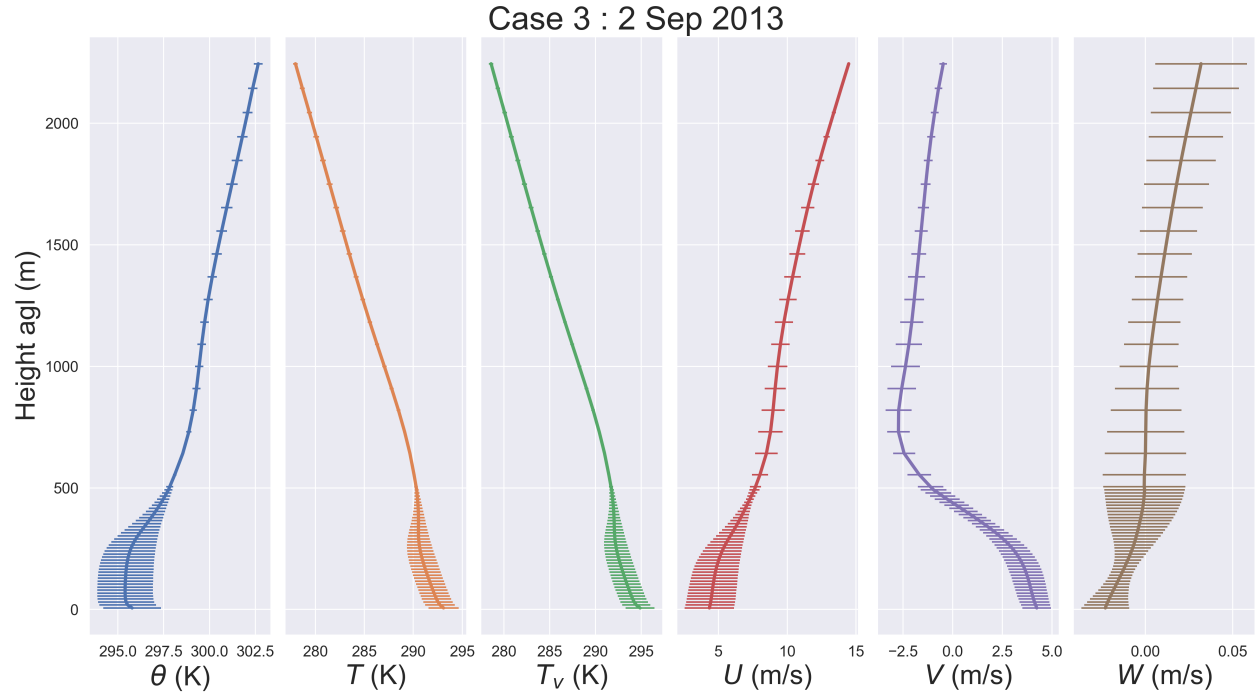


Figure E.7: Case 3 mean meteorological profiles on 2 September 2013: potential temperature θ , absolute temperature T , virtual temperature T_v , west-east wind component U , south-north wind component V , and vertical wind W . Bars show standard deviations during the simulation time.

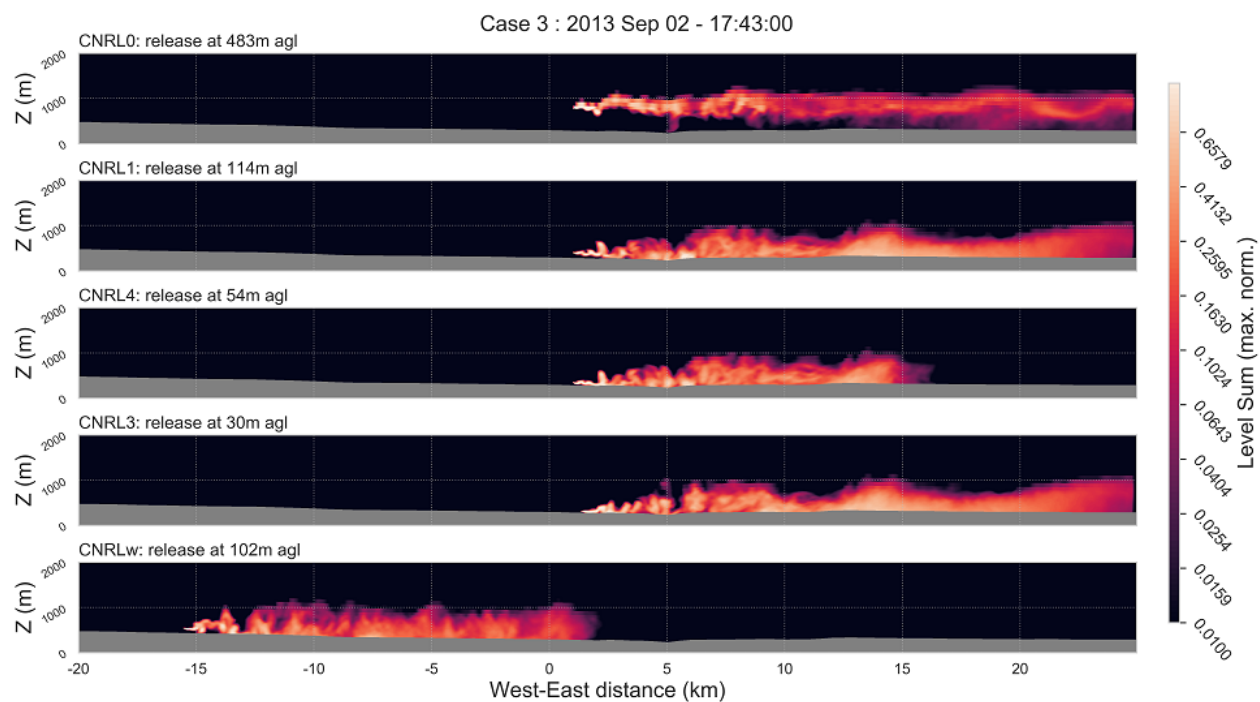


Figure E.8: Domain d05 west-east vertical cross-section for case 3 on 2 September 2013. Vertical cross-section of tracer plumes from stack/point emission sources are shown. Release height above ground is indicated for each source. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for west-east distance in km is at domain centre.

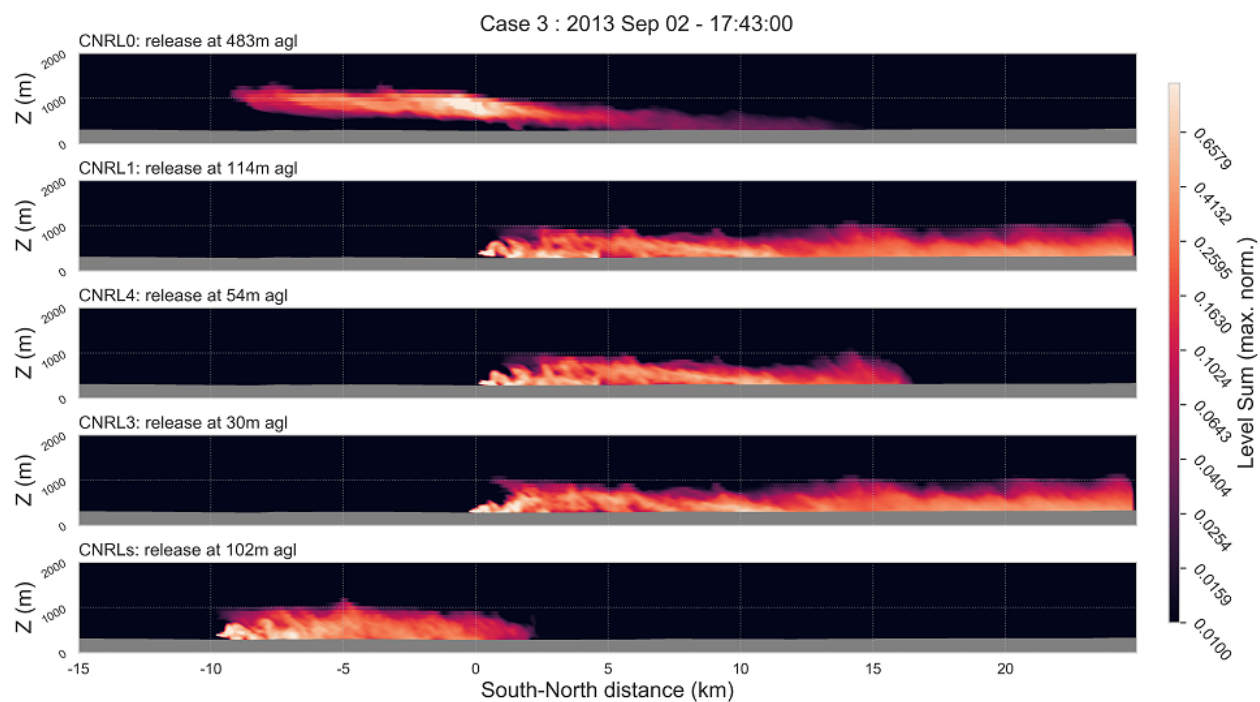


Figure E.9: Domain d05 south-north vertical cross-section for case 3 on 2 September 2013. Vertical cross-section of tracer plumes from stack/point emission sources are shown with release heights indicated for each source. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for south-north distance in km is at domain centre.

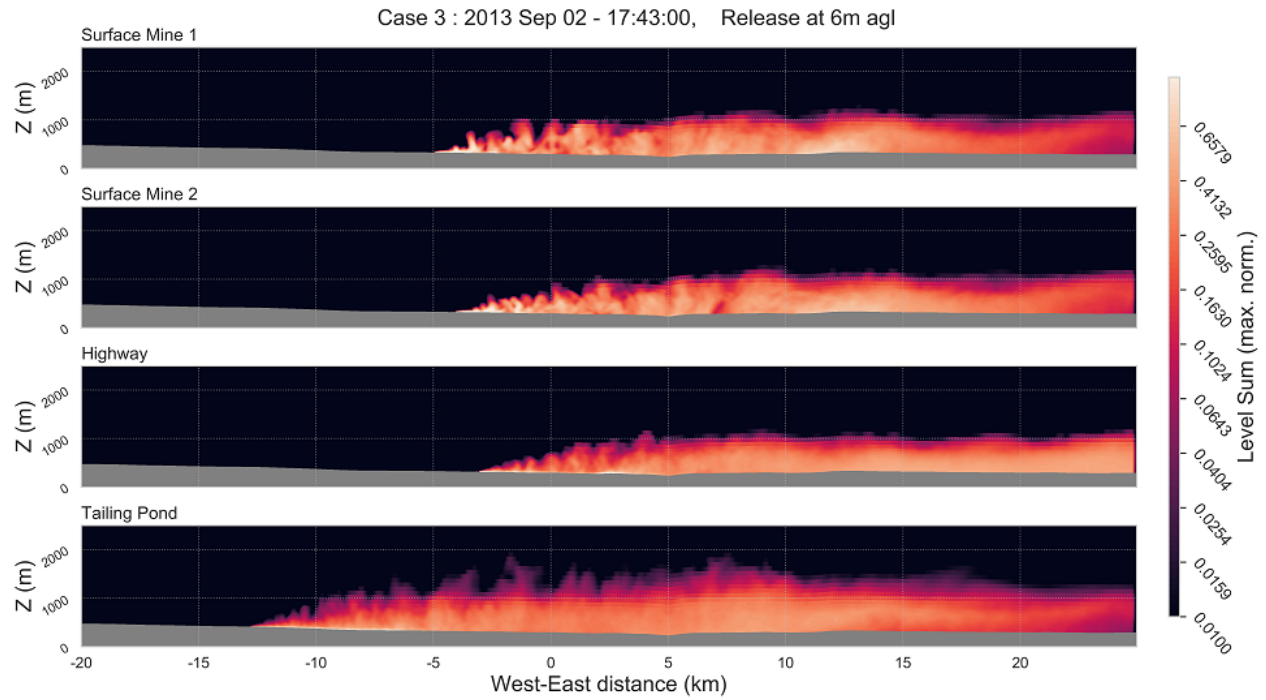


Figure E.10: Domain d05 west-east vertical cross-section for case 3 on 2 September 2013. Vertical cross-section of tracer plumes from surface emission sources (area, line) are shown. Release height above ground is 6m agl for all sources. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for west-east distance in km is at domain centre.

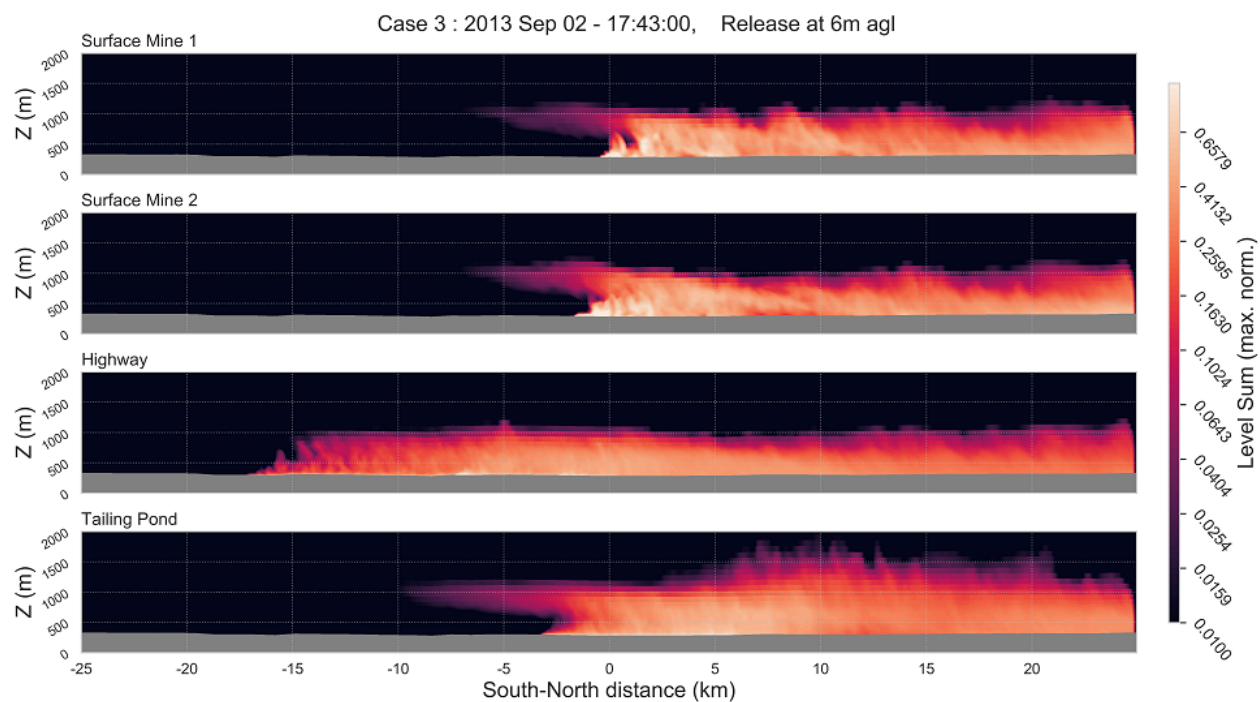


Figure E.11: Domain d05 south-north vertical cross-section for case 1 on 20 August 2013. Vertical cross-section of tracer plumes from surface emission sources (area, line) are shown with release heights at 6 m agl for all sources. Data shown is the level tracer count sum normalized to maximum values for each source. The origin for south-north distance in km is at domain centre.

Appendix F Links to online videos of model output (Case 1)

- **Sources POND and CNRL1:** “Simulating tracer emission and transport under atmospheric dispersion with WRF-ARW” <https://youtu.be/1hi6hss5l7U>
- **Virtual airborne sampling, CNRL1:** “Simulating aircraft-based pollution measurements at super-resolution (LES-scale) using WRF-ARW model” <https://youtu.be/oNfy3yphD9s>

Appendix G Flying Point & Rolling Column analysis

Gaussian Extrapolation below the lowest flight level: extrapolating tracer concentrations $\chi(s, z)$ for heights below the flight level ($z < z_l$) by least-squares fitting up to 1000 m agl at each horizontal point along each s ,

$$\chi(s, z) = \chi_p(s) \exp\left(-(z - z_p)^2 / (2D^2)\right) \quad (\text{G.1})$$

where $\chi_p(s)$ is the peak concentration in the vertical, z_p is the peak height, and D is the vertical-profile standard deviation (Gaussian function scaling distance).

Table G.1: Case 1 breakdown of FP sampled and unsampled portions of tracer plumes for the three sampling screen/volumes. Results are presented in three groups: CNRL stack plumes, surface plumes, and the upwind plumes. From top to bottom of the table, the downwind distance from source to sampling screen increases, except for upwind emissions from CNRLw and CNRLs. These two plumes were first sampled at the upwind wall (short distance) and then a second time at the downwind wall (long distance) of the two boxes. Release heights for CNRLs and CNRLw were below the lowest flight level (similar to other sources except CNRL0). Consequently >50% of plumes fell below the lowest flight level on the upwind wall, hence large unsampled % for CNRLs and CNRLw for the two boxes.

	Rectangular-box		Cylindrical-box		Downwind-screen	
	sampled	unsampled	sampled	unsampled	sampled	unsampled
CNRL0	85 %	15 %	72 %	28 %	88	12 %
CNRL1	49 %	51 %	41 %	59 %	54	46 %
CNRL2	60 %	40 %	48 %	52 %	55	45 %
CNRL4	58 %	42 %	48 %	52 %	60	40 %
CNRL3	61 %	39 %	57 %	43 %	52	48 %
HWY	84 %	16 %	58 %	42 %	65 %	35 %
MINE1	66 %	34 %	63 %	37 %	65 %	35 %
MINE2	70 %	30 %	61 %	39 %	72 %	28 %
POND	59 %	41 %	59 %	41 %	62 %	38 %
CNRLs	15 %	85 %	47 %	53 %	63 %	37 %
CNRLw	5 %	95 %	17 %	83 %	63 %	37 %

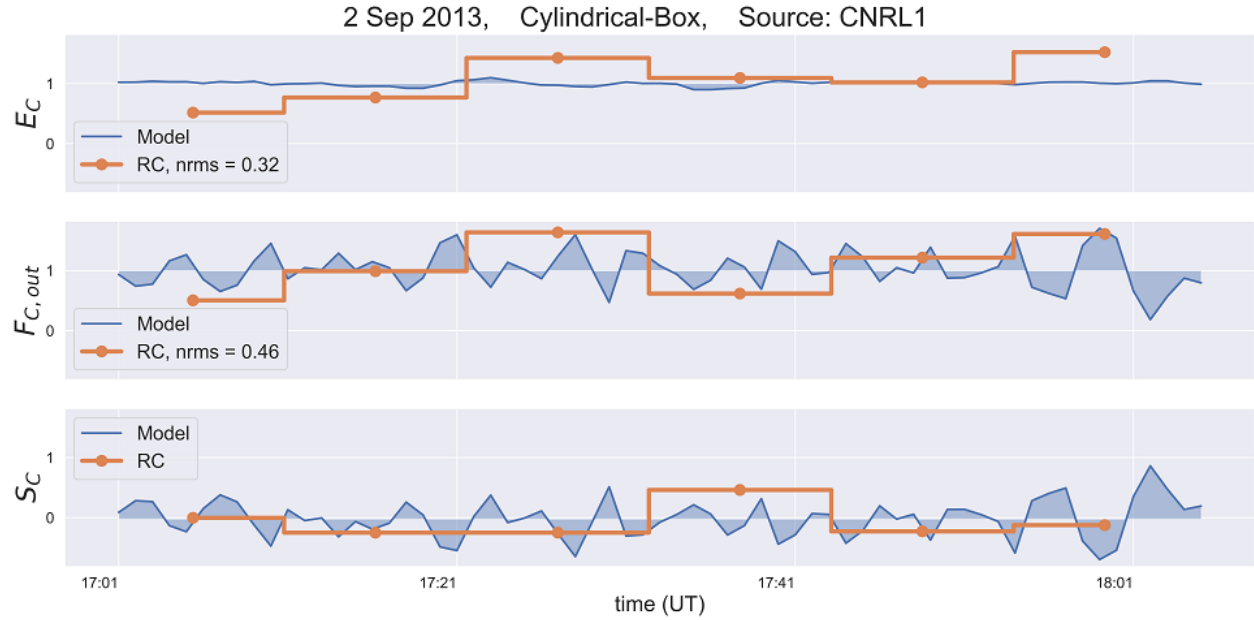


Figure G.1: Rolling Column (RC) estimates for case 3, source CNRL1 and cylindrical-box with S_C estimated based on Eq. 4.3. Normalized root mean square (nrms) error scores are mentioned for RC generated E_C and $F_{C,out}$ estimates, evaluated against model true values for the same terms.