

Master's Thesis

Evaluation and Optimization of Airborne In-Situ Measurement Strategies for Emission Estimation Using High-Resolution Large-Eddy Simulations of Greenhouse Gas Plumes

by

Zhoutong Ma

at the

Chair of Atmospheric Physics

Faculty of Physics

Ludwig-Maximilians-University Munich

in cooperation with the

Institute of Atmospheric Physics,

German Aerospace Center (DLR)

Supervisor: Dr. Friedemann Reum

Supervisor: Prof. Dr. Julia Marshall

Reviewer: Prof. Dr. Markus Rapp

September 15th, 2025



Masterarbeit

**Evaluation und Optimierung
luftgestützter In-situ-Messstrategien
zur Emissionsabschätzung mittels
hochaufgelöster Large-Eddy-Simulationen
von Treibhausgas-Fahnen**

von

Zhoutong Ma

am

Lehrstuhl für Physik der Atmosphäre
Fakultät für Physik
Ludwig-Maximilians-Universität München

in Zusammenarbeit mit dem
Institut für Physik der Atmosphäre
Deutsches Zentrum für Luft- und Raumfahrt (DLR)

Betreuer: Dr. Friedemann Reum

Betreuer: Prof. Dr. Julia Marshall

Gutachter: Prof. Dr. Markus Rapp

15. September 2025



Abstract

Greenhouse gas (GHG) emission estimates from atmospheric observations are often limited by variability in the atmospheric boundary layer. In this study, the Weather Research and Forecasting (WRF) model is used to simulate passive tracer plumes from a continuous point source. The simulations provide a controlled environment for testing airborne in-situ measurement strategies.

The focus is on the widely used mass-balance method, which is used to estimate source strength by integrating tracer concentrations and wind speed across vertical cross-sections downwind of the source. Different flight strategies are systematically tested for accuracy and precision of emission estimates. The results show that both meteorological variability and flight design can introduce significant bias into emission estimates.

The findings highlight the importance of flight path configuration and the influence of weather variability. The results provide initial practical guidance for planning airborne GHG campaigns and contribute to a better understanding of the limitations of current emission estimation methods.

Contents

Abstract

1	Introduction	1
1.1	Motivation	2
1.2	Current Research	3
1.3	Research Design	4
1.4	Research Questions	4
2	Methods	5
2.1	Area and Period of Interest	5
2.2	Plume Simulations	9
2.3	Flight Path and Observation Location Generation	11
2.4	Sampling and Interpolation	13
2.5	Emission Estimation Using the Mass-Balance Approach	14
2.6	Ensemble Generation and Evaluation	15
2.7	Flight Path Cases	16
2.8	Analysis of Ensembles of Turbulent Realizations	20
3	Results and Discussion	22
3.1	Model Verification	22
3.2	The Base Case	26
3.3	Modification A: Influence of Measurement Frequency	28
3.4	Modification B: Number of Transects	32
3.5	Modification C: Minimum Flight Height	35
3.6	Modification D: Distance to Source	43
3.7	Impact of Meteorological Events	47
4	Conclusion	53
4.1	Impact of flight design is smaller than impact of meteorological variation	53
4.2	Distance matters	54
4.3	Minimum height and ground measurements are important (at short distance)	54
4.4	Frequency and number of transects just need to be sufficient	55
4.5	Suggestions	55

5	Outlook and Future Work	56
5.1	Limitations of This Study	56
5.2	Broader Applications	57
5.3	Final Thoughts	57
6	Appendix	59
6.1	WRF Configuration and Namelist Options	59
6.2	Table of Experiments	60

1 Introduction

The mass-balance method is widely used to quantify greenhouse gas (GHG) emissions from point sources in both research and operational contexts (Pühl et al. 2024; Förster et al. 2025). Conceptually derived from Gauss’s divergence theorem, it relates the flux through a control surface to the source strength within. In airborne applications, emissions are estimated by integrating the mass flux of a tracer across a vertical plane downwind of the source, assuming steady-state conditions with negligible storage and end effects.

The mass-balance method is subject to several sources of uncertainty, including measurement errors in concentration, imperfect knowledge of variable wind speed, limited spatial or temporal coverage, and restricted vertical sampling. Additional challenges arise from limited available flight time, atmospheric turbulence, vertical mixing, and diurnal variability.

Previous work has highlighted these limitations. Conley et al. (2017) showed that optimized flight trajectories significantly affect estimation accuracy. Wolff et al. (2021) demonstrated that airborne CO₂ lidar measurements are strongly influenced by wind speed uncertainties and turbulence, with precision degrading under turbulent conditions. Averaging multiple overflights can reduce errors, but turbulence still introduces large variations. More recently, Gałkowski et al. (2024) reported that turbulent wind fields near emission sources cause uncertainties of about 10% in mass-flux methods, which cannot be fully corrected. Together, these studies highlight the importance of accurate wind data, strategic flight planning, and appropriate measurement distances.

Nevertheless, important gaps remain. Previous studies have not systematically addressed how plume turbulence, flight-path design, flight time, and limited vertical coverage contribute to uncertainty under realistic atmospheric conditions. Experimental campaigns are further constrained by flight time, safety concerns, and variable weather, limiting opportunities for systematic testing.

This study addresses these gaps by using the Weather Research and Forecasting (WRF) model in large-eddy simulation (LES) mode to generate synthetic turbulent plume data. This controlled environment allows systematic evaluation of the mass-balance method, with a focus on the combined effects of turbulence and flight design choices. The objective is to identify critical sources of error, quantify their impact on emission retrievals, and provide practical guidance for future airborne campaigns.

1.1 Motivation

Climate change and GHGs Climate change represents one of the most pressing global challenges of the 21st century, driven primarily by the accumulation of GHGs such as carbon dioxide (CO₂) and methane (CH₄) in the atmosphere. Reducing anthropogenic emissions of these gases is essential to meet international climate targets, including the goal of limiting global warming to well below 2°C and pursuing efforts to limit it to 1.5°C above pre-industrial levels, as outlined in the Paris Agreement (UNFCCC 2015).

Anthropogenic CO₂ and methane emissions from point sources A share of global GHG emissions originates from discrete point sources, including power plants, industrial facilities, and fossil fuel extraction sites. Quantifying emissions from these sources with high spatial and temporal resolution is critical for assessing the efficacy of climate change mitigation measures, regulatory compliance and scientific assessment. However, accurately estimating point-source emissions remains technically challenging due to complex atmospheric transport processes, variability in emission strength (Knapp et al. 2023), and measurement limitations. Improved methods for detecting, quantifying, and attributing emissions, particularly using aircraft-based platforms, are therefore essential for advancing atmospheric research and informing climate policy.

Emission monitoring and verification support (MVS) Reliable estimation of GHG emissions is essential not only for scientific understanding but also for supporting climate policy, regulatory frameworks, and mitigation strategies. To this end, the concept of Monitoring and Verification Support (MVS) has emerged as a key element in ensuring transparency and accountability in emission reporting and assessment. Two complementary approaches, “bottom-up” and “top-down”, are commonly used in emission estimation:

- **Bottom-up emission inventories** are estimates based on activity data and emission factors, typically reported by countries or sectors such as power plants and coal mines. These inventories are developed using statistical, engineering, or economic data, often following IPCC guidelines (UNFCCC 2015). The term “bottom-up” describes the aggregation of local or sectoral data to estimate total emissions (Meijer et al. 2019).
- **Top-down emission estimation** uses atmospheric measurements and inverse

modeling techniques to infer emissions from observed concentrations. This approach provides an independent means of verifying reported emissions and identifying discrepancies or unreported sources, offering insights into emission mitigation potential (Ciais et al. 2017). The term “top-down” refers to deriving emissions by interpreting atmospheric signals rather than activity data (Meijer et al. 2019).

Combining top-down and bottom-up approaches under a consistent framework enhances the robustness of emission monitoring systems and supports international efforts toward emission reductions. The work presented here contributes specifically to improving top-down estimates by testing and refining airborne mass-balance strategies, focusing on the influence of flight design and variability on emission quantification.

1.2 Current Research

Recent advances in airborne in-situ measurement techniques have significantly contributed to the quantitative assessment of GHG emissions from point sources. In particular, aircraft-based platforms have proven to be effective tools for high-resolution, direct sampling of plumes. Within this context, we operate the HELiPOD system, a helicopter-borne measurement platform originally developed at TU Braunschweig (Reum et al. 2025; Förster et al. 2025; Pätzold et al. 2023). Our group contributes expertise in greenhouse gas monitoring, specifically in the instrumentation, calibration, operation, and scientific analysis. The system is equipped with a Picarro GHG analyzer (Crosson 2008), which provides high-precision measurements of methane and carbon dioxide and forms the core of the airborne GHG payload.

1.2.1 Methane-to-Go campaign and HELiPOD

A recent application of the HELiPOD is the *Methane-to-Go* (MTG) campaign in Poland, which targeted methane emissions from coal-mining operations. The HELiPOD was used to survey the air near major point sources (Förster et al. 2025). Although results from this campaign are not analyzed in this study, the campaign provides valuable context and motivation for evaluating emission estimation techniques under realistic turbulent conditions. The HELiPOD platform is suspended 25 m beneath a helicopter flying at an average speed of 40 m/s (Pätzold et al. 2023). It is equipped with fast-response sensors for GHG concentrations, meteorological

instruments, and GPS systems for precise navigation and positioning. These capabilities make it particularly well-suited for performing mass-balance calculations of trace gas fluxes in the lower atmosphere.

1.3 Research Design

This study investigates the uncertainty in emission estimates obtained using the mass-balance method under turbulent atmospheric conditions. The analysis uses simulations of atmospheric tracer transport, following an approach similar to that of Wolff et al. (2021), who demonstrated the suitability of synthetic datasets for evaluating the robustness of airborne emission quantification methods.

Simulations offer a cost-effective and flexible alternative to real-world flight campaigns, which would require substantial time and resources to repeatedly measure the same emission source under varying conditions. In contrast, the simulated environment allows for arbitrary adjustments to emission scenarios and flight paths with minimal effort.

Flight paths are generated automatically and used to sample the simulated tracer fields. Sampling, interpolation, and mass-balance calculations are applied to each path to produce an ensemble of emission estimates. This enables a systematic analysis of the effects of turbulence on the accuracy and uncertainty of airborne mass-balance methods.

1.4 Research Questions

This study aims to improve understanding of how atmospheric turbulence affects the accuracy and uncertainty of airborne GHG emission estimates derived using the mass-balance method. To this end, the following research questions are addressed:

1. To what extent does atmospheric turbulence influence the accuracy of emission estimates obtained using the airborne mass-balance approach?
2. How do different flight path geometries affect the robustness of the mass-balance method under varying turbulence conditions?

By addressing these questions using synthetic tracer fields generated with WRF, this study provides guidance for optimizing airborne measurement strategies and improving the reliability of mass-balance-based emission assessments.

2 Methods

2.1 Area and Period of Interest

The simulation domain used in this study is in the region investigated by Fiehn et al. (2020) and Förster et al. (2025), covering the Upper Silesian Coal Basin, a coal mining area in Poland known for significant point-source emissions of CH_4 . Here, one emission source targeted during MTG-Poland is chosen for analysis. The nested modeling domains as shown in Fig. 1, denoted as d01, d02, d03 and d04, have resolutions of 3750 m, 1250 m, 250 m and 50 m, respectively. Domain d04 is centered on the primary emission source and provides high spatial resolution for resolving plume dynamics and turbulent structures. ERA5 reanalysis data were used to drive the WRF simulation. Figure 2 provides an zoomed overview of the domains d03 and the innermost high-resolution domain d04. The WRF simulations were conducted for a summer day under convective boundary layer conditions to ensure realistic turbulence development and plume dispersion. 16 June, 2022 was chosen, as it was a day on which measurements were carried out during the MTG Poland campaign, as shown in Fig. 3.

Recent airborne campaigns such as the *Methane-to-Go* project in Poland have demonstrated the potential of in-situ measurements for assessing emissions from coal-mining operations. These efforts highlight the need for improved methodologies that can account for atmospheric turbulence and optimize flight strategies for emission quantification.

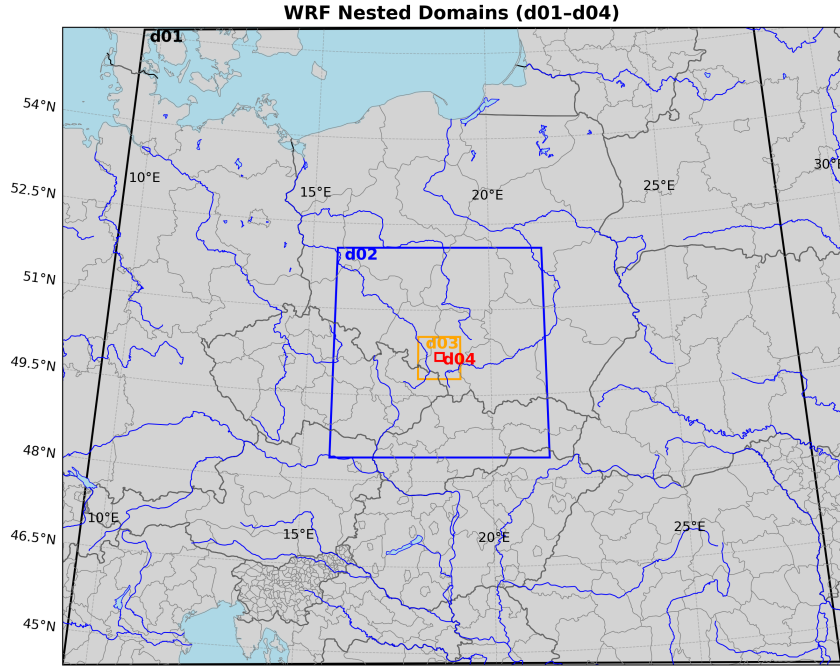


Figure 1: **Overview of the nested WRF-LES modeling domains used in this study.** The four nested domains (d01–d04) are outlined in black, blue, orange, and red, respectively. The domains have horizontal grid resolutions of 3750 m, 1250 m, 250 m, and 50 m. The domains are centered over southern Poland.

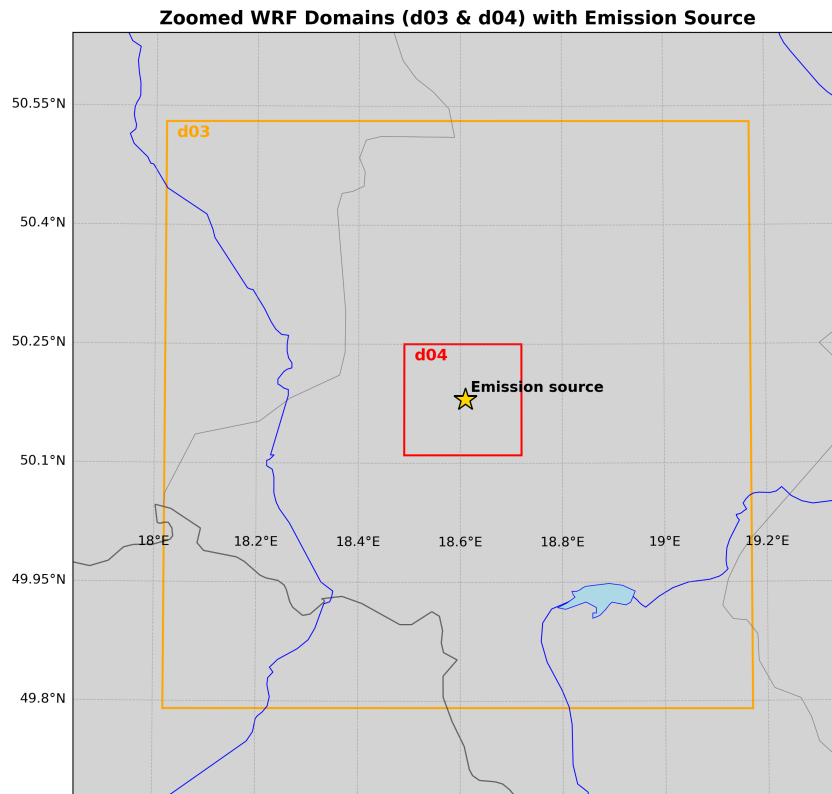


Figure 2: **Zoomed view of domains d03 and d04.** The domains are centered on the emission source (yellow star) in southern Poland. Domain d03 (orange) has a horizontal resolution of 250 m, while domain d04 (red) has a horizontal resolution of 50 m.

In-situ methane measurements during MTG Poland campaign

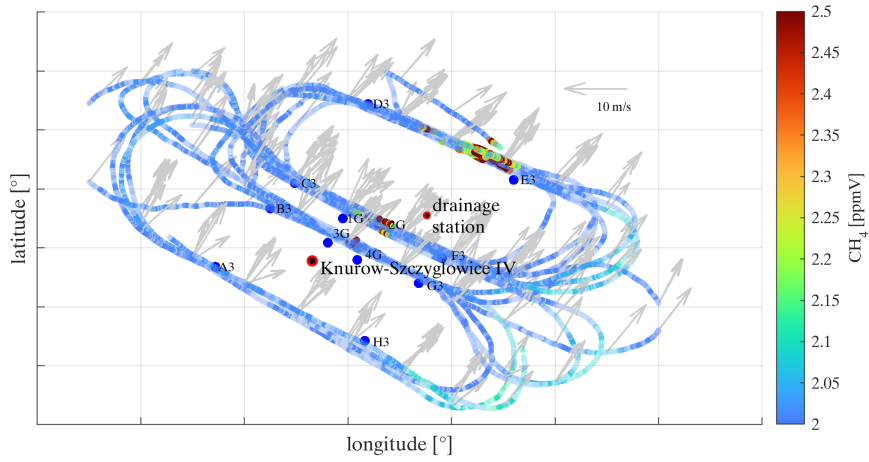


Figure 3: **In-situ methane measurements during MTG Poland campaign, on 16 June 2022.** Blue lines indicate the HELiPOD flight path, with color coding representing measured CH_4 concentrations (ppmV). Wind vectors (grey arrows) illustrate the local wind field during the flight. Measurement sites, including the Knurów-Szczygłowie IV mine and drainage stations, are marked in red. Latitude and longitude values are omitted. Courtesy of Eric Förster, DLR. Further details on the HELiPOD instrument are provided in Sect. 1.2.1.

2.2 Plume Simulations

2.2.1 WRF-Chem

The Weather Research and Forecasting (WRF) model is a non-hydrostatic, mesoscale numerical weather prediction system widely used in operational forecasting and atmospheric research (Skamarock et al. 2019). For this study, the model is used in its chemistry-enabled configuration (WRF-Chem) (Grell et al. 2005), which includes a greenhouse gas (GHG) module (WRF-GHG) (Beck, Koch, Thomas, and Kretschmer, Roberto 2011) for passive tracer transport. This configuration generates synthetic GHG concentration fields from an idealized point source.

The simulation domain is configured with nested grids (Sect. 2.1). The plumes simulated in the innermost domain (d04) are analyzed here. The d04 domain consists of 320×320 grid cells (16 km x 16 km) with 50 m horizontal spacing. Output from d04 is saved every minute to resolve turbulent plume dynamics.

A full list of relevant physical parameters and namelist options, including boundary-layer schemes and advection settings, is provided in Appendix 6.1.

2.2.2 Large-Eddy Simulations

Large-Eddy Simulations (LES) are used to explicitly resolve turbulent structures within the planetary boundary layer (PBL). This approach enables detailed modeling of atmospheric transport and mixing processes, which are critical for simulating realistic dispersion of GHG plumes. In this study, LES is applied to domain d03 and to the innermost domain d04 with high spatial and temporal resolution to ensure that the influence of turbulence on plume evolution is well captured. The LES framework in WRF follows the standard approach of resolving large-scale motions while parameterizing unresolved turbulence through sub-grid-scale closure schemes (Nieuwstadt et al. 1993).

2.2.3 Cell Perturbation Method

To efficiently generate realistic turbulence in the LES domain, the Cell Perturbation Method (CPM) is used (Muñoz-Esparza and Kosović 2018). CPM introduces stochastic perturbations at the inflow boundary to accelerate turbulent spin-up while preserving physical realism. Specifically, virtual potential temperature is perturbed in inflow cells according to the CPM formulation. The perturbation amplitude is controlled by the Eckert number (E_c), a non-dimensional parameter governing energy

input to the flow. The CPM perturbation amplitude is controlled by the Eckert number, with lower values (e.g., $E_c = 0.2$) producing stronger perturbations and therefore more rapid turbulence development. Kilroy et al. (2024) found an Eckert number of $E_c = 0.2$ yields realistic turbulence. In the case analyzed here, it was suspected early on that $E_c = 0.2$ may lead to exaggerated turbulence. Though not analyzed systematically, a slightly weaker turbulence generation with $E_c = 0.4$ was chosen. A systematic evaluation of how well the simulated turbulence characteristics (e.g. Turbulence Kinetic Energy) correspond to actual conditions on this specific day was not conducted. The reason is that the representativeness of the results is limited to the simulated conditions regardless of fidelity to actual conditions, and in case of deviations, they may be viewed as representative for a day with slightly different meteorological conditions than observed. A follow-up study should broaden the scope of the present study by investigating the sensitivity of the results to a range of different turbulence conditions, including for meteorological conditions and domains.

The CPM configuration follows the setup developed and validated by Mazzaro et al. (2019), who showed that CPM-accelerated turbulence generation compares well with observation-based datasets. Kilroy et al. (2024) adapted the CPM implementation and tested simulated WRF output against measurements from the Krummendeich research wind farm at 5 m horizontal resolution.

Using CPM in this context reduces computational cost by allowing turbulence to develop in a smaller domain than without applying CPM.

2.2.4 Example plume

The GHG source emits continuously at a rate of $1.0 \times 10^6 \text{ t yr}^{-1}$ (1 megatonne per year, 1 Mt yr^{-1}), which is approximately 31.71 kg/s . This emission rate serves as the reference value for the emission estimation. The tracer simulation is, however, generic and applicable to all passive tracers. For scaling purposes, a CO_2 source is assumed as reference. For other gases the source strength can be converted accordingly. Since measurement errors are not considered, the absolute choice of the reference gas and emission rate is not essential for the interpretation of results and serves to illustrate the dynamics. Therefore, the approach presented here is equally applicable to other passive tracers such as methane.

Figure 4 shows a horizontal cross-section of the passive tracer concentration at the height of emission. The plume displays spatial variability and turbulent features

similar to those found in other state-of-the-art LES plume studies at comparable scales. Because its shape changes over the course of the day, different structures appear that resemble the simulated plumes reported in both Wolff et al. (2021) and Brunner et al. (2023). While we cannot verify that the simulated plume matches the actual meteorological conditions of the case, the similarity in plume morphology and turbulence patterns to published simulations and observations supports the suitability of the LES configuration for the study objectives.

2.3 Flight Path and Observation Location Generation

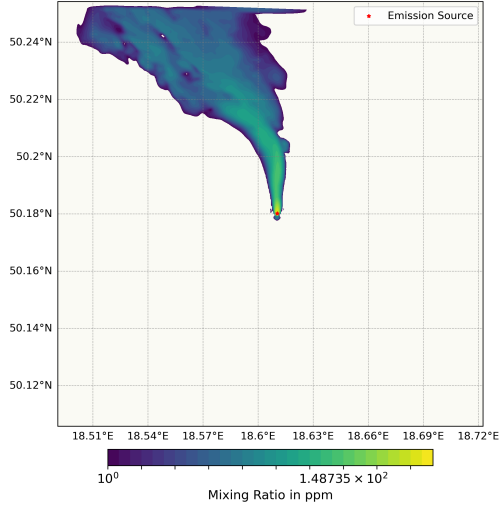
To simulate realistic airborne sampling of the plume, a software tool was developed to automatically generate virtual flight paths through the LES domain. The design mimics strategies used in real airborne campaigns, where aircraft perform multiple transects perpendicular to the wind to intersect the plume at varying altitudes.

A representative flight pattern is shown in Fig. 5 with its cross-section in Fig. 6. Each flight path consists of a sequence of horizontal legs forming virtual “walls” that cross the plume. Flights are aligned perpendicular to the mean wind direction, the geometry of the sampling wall is defined by configurable parameters.

The key parameters that define each flight path include:

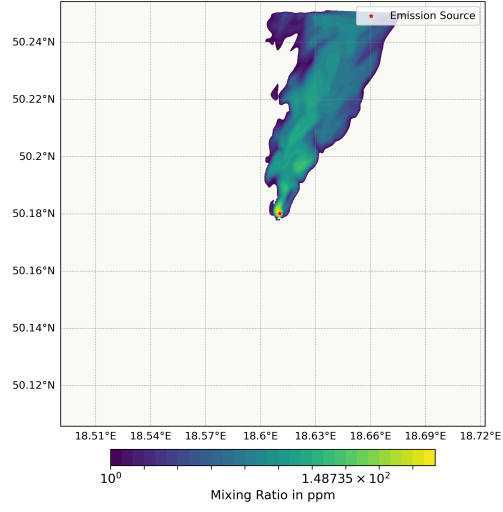
- **Wall distance from source:** Horizontal distance between the emission source and the sampling wall.
- **Minimum and maximum flight height:** Vertical extent of the sampling wall. The **minimum flight height** sets the lowest altitude for a transect, the **maximum height** is set dynamically from the simulated plume-top height. (See below)
- **Number of transects (nTS):** Count of horizontal legs within one wall, distributed evenly between the minimum and maximum height.
- **Ground leg:** A low-altitude transect placed 1 m above the surface, representing measurements from a ground platform (i.e. from a vehicle). For simplicity, the horizontal location of the ground leg is aligned with the transects.
- **Measurement frequency:** Temporal resolution of the virtual sampling, i.e., the time interval between two successive sampling points ("timesteps") along the flight path. This parameter affects the resolution of the sampled data but does not alter the geometry of the flight path itself.

Horizontal cross-section of simulated plume at 04:00 UTC



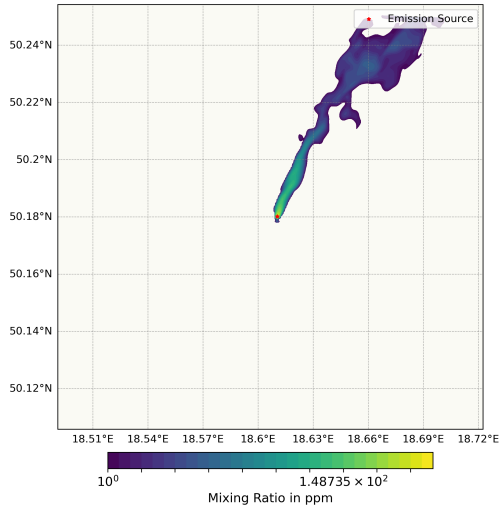
(a) 04:00 UTC

Horizontal cross-section of simulated plume at 06:00 UTC



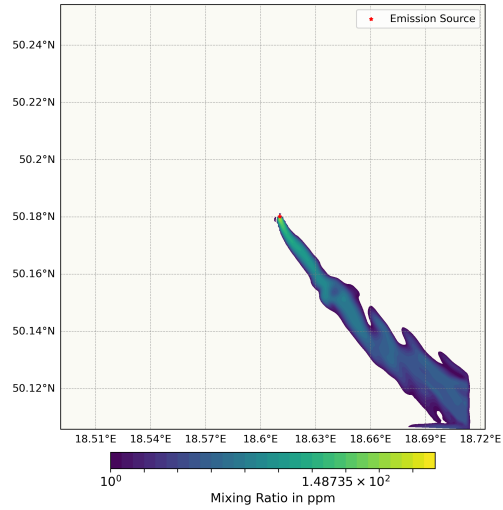
(b) 06:00 UTC

Horizontal cross-section of simulated plume at 08:00 UTC



(c) 08:00 UTC

Horizontal cross-section of simulated plume at 13:00 UTC



(d) 13:00 UTC

Figure 4: **Horizontal cross-sections of simulated passive tracer plumes.** Cross-sections at 50 m above ground level are shown for four times: (a) 04:00 UTC, (b) 06:00 UTC, (c) 08:00 UTC, and (d) 13:00 UTC. The plume originates from a continuous point source at 50 m and shows dispersion and spatial variability due to turbulent mixing within the LES domain. Concentrations are expressed as volume mixing ratios (ppmV) and displayed on a logarithmic color scale.

To ensure sufficient plume coverage, the maximum flight height is adjusted dynamically using the plume-top height (PTH) derived from LES tracer fields. The plume top is identified by analyzing vertical cross-sections of tracer concentration and setting the upper limit where the integrated concentration is more than 99.9% of the total concentration on that cross-section. For each flight, the maximum plume-top height over all sampled WRF output frames was calculated and used as the representative PTH. An analogous maximum is also determined for the PBLH, which is directly provided by the WRF output. This adaptive approach ensures that no tracer mass is passing above the wall.

The generated flight paths are then used to extract concentration and wind data from the LES output at virtual observation locations, forming the basis for emission estimates with the mass-balance method.

To mimic real flight dynamics, the simulation includes turnarounds between transects and ferry segments to and from the airport. These portions are excluded from the emission estimation but are accounted for in timing and sampling continuity. This yields more realistic campaign duration and spatial coverage.

The script generates loops with two transects to simulate the real flight path, always resulting in two walls.

2.4 Sampling and Interpolation

After generating the virtual flight paths, the next step is to sample and interpolate relevant model output along each measurement point of the trajectory. At each sampling location, atmospheric variables pressure, temperature, wind speed, wind direction, and GHG tracer concentration are required. Because values are not available at arbitrary locations, they need to be calculated from the discrete WRF grid using spatial and temporal interpolation. These parameters are critical for the emission estimation, as explained in Eq. 3.

This process is implemented in Python using the `numpy` and `scipy` libraries. Each point along the flight path is defined by latitude, longitude, and altitude. Spatial interpolation is carried out in three dimensions: (i) the two WRF levels that bracket the altitude of the measurement point are located; (ii) for each bracketing level, a 2D bilinear interpolation is performed in the horizontal (latitude–longitude); and (iii) a 1D linear interpolation between those two levels yields the value at the exact altitude. The lowest WRF level was about 9 m above ground. The values at 9 m are therefore extrapolated downward to the ground. This procedure constitutes a full 3D spatial

interpolation.

Temporal interpolation is also required because WRF output is typically stored at coarse time intervals (e.g. every 60 s), whereas the virtual flight path may require sub-second data. For each measurement point, spatially interpolated values are first obtained at the two nearest model times (preceding and following), and then a linear interpolation in time is applied between them.

This sampling procedure mimics the behavior of real airborne instruments, which measure at high frequency as the aircraft traverses the plume.

To manage the computational time and cost of processing such large 4D (x, y, z, t) model files at high sampling rates, these operations are parallelized on the high-performance computing system Levante, operated by the Deutsches Klimarechenzentrum (DKRZ).

The interpolated data from all measurement points are stored in structured `numpy` arrays and serve as input to the subsequent mass-balance emission estimation.

Measurement error and background variability are neglected in this study to focus on turbulence. A follow-up study should address them once the variability introduced by turbulence is better understood, enabling a more complete assessment of emission estimation uncertainty.

2.5 Emission Estimation Using the Mass-Balance Approach

After spatial and temporal interpolation is completed, the resulting dataset resembles real in-situ airborne measurements, except that measurement errors are not included. The next step is to estimate the emission strength of the point source using the mass-balance approach, which is commonly applied in airborne measurement campaigns. This setup follows established practice in airborne studies, e.g. Pühl et al. (2024).

The total tracer flux is calculated using the mass-balance approach as

$$Flux_{\text{total}} = \sum_i Flux_i \quad (1)$$

where $Flux_i$ of each measurement point i is,

$$Flux_i = \Delta C_i \cdot \frac{p_i \cdot M}{R \cdot T_i} \cdot V_{\perp,i} \cdot \Delta x_i \cdot D_i \quad (2)$$

Expanding the $Flux_i$ term in Eq. 1 we get

$$Flux_{\text{total}} = \sum_i^{\text{transects}} \left(\Delta C_i \cdot \frac{p_i \cdot M}{R \cdot T_i} \cdot V_{\perp,i} \cdot \Delta x_i \cdot D_i \right), \quad (3)$$

where the variables are defined as:

- $\Delta C_i = C_i - C_0$ is the tracer concentration enhancement in segment i relative to the background C_0 . Here, unlike with real in-situ measurements, $C_0 = 0$.
- M is the molar mass of dry air, a scalar quantity
- R is the universal gas constant, scalar quantity
- p_i is the ambient pressure,
- T_i is the ambient temperature,
- $V_{\perp,i}$ is the wind speed perpendicular to the flight transect,
- Δx_i is the horizontal width of the segment, and
- D_i is the vertical distance between transects.

Two common approaches are used to account for the layer between the lowest transect and the ground. One integrates only halfway to the ground, assuming a decreasing plume gradient. The other assumes equal distribution and integrates the mixing ratio of the lowest transect all the way to the ground. For well-mixed plume this is a good approach where homogeneity within the PBL justifies this assumption. However, for plumes that have not reached it's well-mixed state, the potential bias of this approach must be considered in a separate study. Due to limited time and resources, it can not be concluded here, which approach is closest to the true value. To remain consistent with the method used for MTG-Poland from Förster et al. (2025), the latter approach was adopted.

2.6 Ensemble Generation and Evaluation

Due to the stochastic nature of atmospheric turbulence, individual emission estimates can vary significantly. To quantify this variability, an ensemble of virtual flights was generated for each flight path, with one flight every 3 min between 04:00 and 20:00 UTC (06:00–22:00 local time in Poland). **Unless otherwise specified, all times**

are given in UTC. Earlier flight times were excluded from the analysis for two reasons: (1) plume top heights are often too low during the night to allow reliable observation under real-world minimum flight altitude restrictions, and (2) calm winds at night can cause tracer accumulation, hindering accurate emission estimation. A follow-up study could address this limitation by rigorously assessing night-time flights and quantifying the effect of these constraints. Each ensemble therefore consists of 321 realizations under different turbulent flow structures, reflecting realistic variability in operational campaigns. The emission estimate from each flight is stored, and the resulting ensembles are analyzed using four statistical indicators:

- **Mean:** Represents the central tendency of the emission estimates and provides a first-order approximation of the true emission rate.
- **Median:** Represents the central tendency of the emission estimates while being less sensitive to outliers, and is therefore a robust indicator of the typical emission rate.
- **Standard deviation (SD):** Quantifies the spread of the ensemble at each time step, with larger values indicating greater variability and higher uncertainty.
- **Standard error (SE):** Defined as $SE = SD/\sqrt{N}$, where N is the number of ensemble members. SE quantifies how precisely the ensemble mean estimates the underlying population mean.

All four statistical indicators are computed using the Python `numpy` library. These statistical indicators are a core part of the final results and are essential for assessing the robustness and reliability of emission estimates derived from airborne observations. In addition to their temporal evolution, they provide the basis for discussing and interpreting uncertainties in the results.

2.7 Flight Path Cases

To investigate the influence of flight path configuration on emission estimation uncertainty, a set of flight cases was defined by modifying selected parameters of a base case configuration derived from the 2022 MTG Poland mission. The modifications were applied independently to isolate their influence. A verification case was run first to validate the emission estimation method.

Table 1 provides an overview of the varied parameters across the cases.

Table 1: Overview of flight path cases and modified parameters compared to the base case.

—	Model Verification	Quasi-instantaneous flights for verifying agreement between estimated and true emissions.
—	Base case, no modification	Replicates 2022 MTG Poland mission
A	Frequency	Modified measurement frequency
B	Number of transects	Increased/decreased to assess vertical resolution effect
C	Altitude range	Modified min/max flight altitude and ground measurement
D	Distance to source	Modified distance to emission source

2.7.1 Experiments

Each parameter was modified multiple times, with each modification treated as a separate experiment, resulting in approximately 40 experiments in total. The detailed setups and results are provided in Appendix 6.2.

2.7.2 Model verification

To check the model setup, a special verification case was created. The flight speed and measurement frequency were both increased by a factor of 10000 so that the aircraft could cover the full flight path instantly while keeping the same sampling density. This way, the whole measurement was finished within one WRF output (1 min), which excludes the effect of meteorological variability over time. The specific parameters of the base case configuration are:

- Wall distances: 1150 m (Wall 1), 1200 m (Wall 2)
- MFH: 1 m
- Measurement frequency: 4.5 Hz
- nTS: 60
- Ground measurements: unnecessary (1 m MFH)
- Wall width: 10000 m

The wall width was set to 10000 m to ensure the full plume was captured. Emission rates were then calculated with the same mass-balance method that is also used in the other experiments. Details and results are discussed in the subsequent Sect. 3.1.

2.7.3 Base case configuration

The base case replicates the configuration of the *Methane-to-Go* campaign flight in Poland on 16 June 2022 (Fig. 3). This date was selected due to favorable meteorological conditions and the availability of flight-planning data. The flight path includes two vertical sampling walls placed 600 m and 1200 m downwind of the emission source, respectively. The wall width is 2500 m for base case configuration. 5000 m for walls beyond 4000 m downwind, as used in other cases. For 6000 m and greater distances, the width was increased to 7500 m to allow research on further distance. There is little guidance from HELiPOD campaigns at such ranges, as measurements at this distance are rare due to limited flight time. The minimum flight height (MFH) is set to 100 m, and the maximum flight height is determined dynamically from the plume-top height, as described in Sect. 2.3.

In the real mission, the minimum flight height was constrained by vegetation and regulatory limits, typically between 100–150 m. On this day, the minimum flight height was 100 m.

In real campaigns, the planetary-boundary-layer height (PBLH) is estimated using forecasts and initial transects. In this study, the plume-top height derived from WRF simulations was used instead, to avoid bias from inaccurate PBL estimates. This ensures full vertical plume coverage in each simulated case.

Flight paths are aligned with the dominant wind direction to ensure perpendicular plume intersections. The wind direction is calculated by averaging wind vectors near the source over the entire flight time.

The specific parameters of the base case configuration are:

- Wall distances: 600 m (Wall 1), 1200 m (Wall 2)
- MFH: 100 m
- Measurement frequency: 1.5 Hz
- nTS: 10
- Ground measurements: included

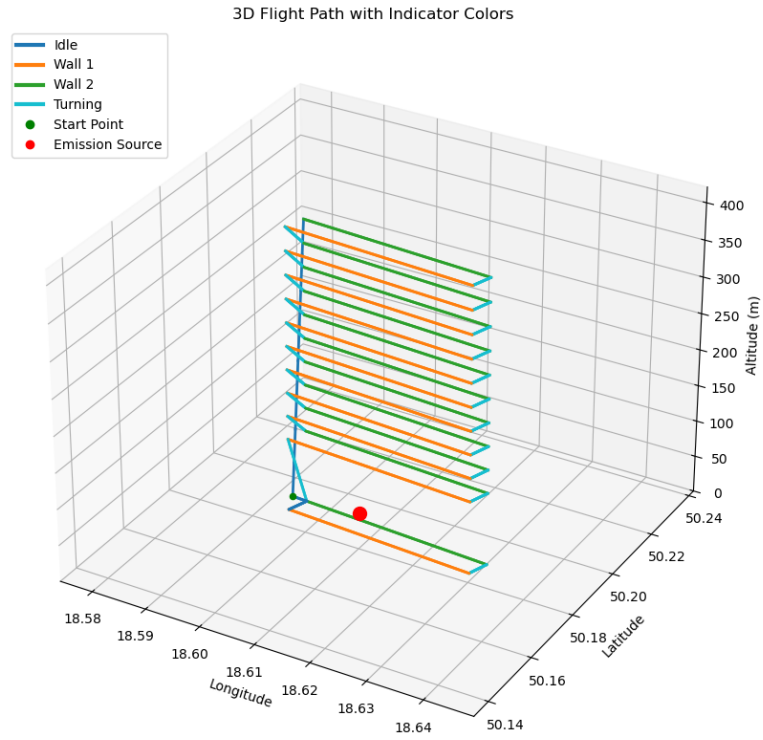


Figure 5: **Three-dimensional representation of the base case flight path.** Colors indicate different flight phases: transporting (blue), Wall 1 (orange), Wall 2 (green), and turning maneuvers (cyan). The red dot marks the emission source location, and the green dot indicates the airport location. Longitude and latitude are shown on the horizontal axes, and altitude (m) is shown on the vertical axis.

- Wall width: 2500 m

The minimum flight height (MFH) is the altitude of the HELiPOD, which is 25 m below the helicopter. Because of meteorological variability, most analysis focuses between 04:00 and 12:00, the reason is discussed in Sect. 3.1.

2.7.4 Flight path variations

The base case is modified in each variant to test how measurement design affects emission retrievals. Table 1 summarizes the modifications applied to these parameters. Variant of modifications:

- **Modification A: Measurement frequency:** Varied to mimic upgrades or degradations in sensor performance. Picarro instruments on HELiPOD typically sample CH_4 at 1.5–2.2 Hz. In these tests, the frequency is increased by $3\times$ and reduced to $1/3$ of nominal to assess impacts on plume resolution and integration accuracy.
- **Modification B: Number of Transects:** Directly affects vertical sampling resolution and thus could affect integration accuracy in the mass-balance calculation. Lower counts risk missing key plume structures.
- **Modification C: Minimum Flight Height:** The altitude of the lowest measurement, which could potentially lead to an incomplete capture of the plume gradient and influence the emission estimate
- **Modification D: Wall distance:** Influences spread and dilution of the sampled cross section of the plume. Shorter distances may not capture full dispersion, longer distances increases width and height of the plume, potentially requiring more flight time for sampling.

2.8 Analysis of Ensembles of Turbulent Realizations

To assess the influence of various flight path modifications, ensemble-based emission statistics for each case are compared with the base case. The comparison focuses on changes in the ensemble mean, standard deviation, and standard error to quantify how specific parameters affect the robustness and uncertainty of emission estimates. By systematically varying one parameter per case (e.g., transect count, distance from the source, measurement frequency), the sensitivity of the mass-balance method to

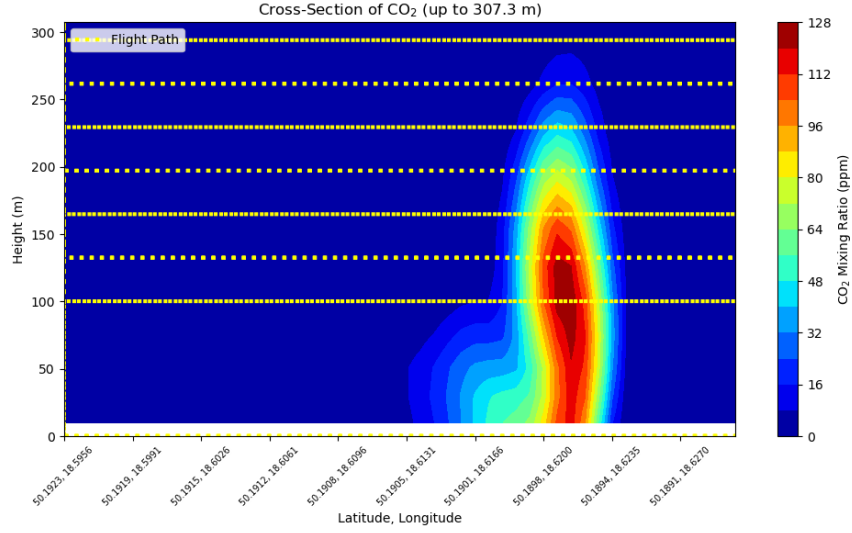


Figure 6: **Cross-section of the simulated plume with the base case flight path.** The color scale indicates the mixing ratio (ppmV). Yellow dashed lines denote the flight path. Latitude and longitude are shown on the horizontal axis, and height (m) on the vertical axis.

flight configuration is explored. Cases that show large changes in the mean suggest a systematic bias introduced by the modification, whereas increases in standard deviation or standard error indicate greater measurement uncertainty.

Each modification's impact is interpreted in terms of:

- change in the retrieved emission rate (mean),
- change in the spread of retrievals (standard deviation),
- change in the confidence of the estimate (standard error).

The discussion highlights the parameters most critical for reliable emission estimation under turbulent atmospheric conditions. A summary table ranks each modification by its influence on the ensemble statistics and forms the basis for practical recommendations for designing future airborne mass-balance missions for accuracy and reliability.

3 Results and Discussion

3.1 Model Verification

Before the other experiments, the method is validated. The validation setup includes simulated airborne measurements conducted effectively instantaneously by increasing flight speed by a factor of 10000, making every flight path a quasi-instantaneous measurement. The measurement frequency is also boosted with the same factor to keep the sampling density unchanged. This high-speed setup allows the entire measurement to be sampled within one minute, minimizing the influence of transient meteorological variability during the measurement.

Two vertical measurement walls are placed downwind of the source, at distances of 1150 m and 1200 m. Each wall contains 60 transects, six times more than in the base case. The walls are oriented perpendicular to the mean wind direction near the plume. The wall width is set to 10000 m, four times larger than the base case, to ensure full plume coverage. The measurement frequency is 4.5 Hz, for multiple measurement points per WRF grid cell. Emissions are estimated using the same mass-balance algorithm as in the other cases. Figure 7 shows the path in 3D.

Results:

- Wall 1: mean = 34.97 kg/s, standard deviation = 24.92 kg/s, standard error = 1.39 kg/s, median = 31.98 kg/s
- Wall 2: mean = 35.14 kg/s, standard deviation = 26.13 kg/s, standard error = 1.46 kg/s, median = 32.02 kg/s

Both walls indicate an overestimation. Wall 1 shows an overestimation of 10.3%, while Wall 2 shows a slightly larger overestimation of 10.81%. When comparing medians instead of means, Wall 1 is only 0.85% above the reference value (31.71 kg/s), and Wall 2 is 0.98% above. The medians therefore agree well with the reference value.

The discrepancy between the mean and median reflects both the influence of high variability and the presence of a positively skewed distribution, with occasional high outliers pulling the mean upward relative to the median. Figure 8 shows that the highest emission variability occurs in the afternoon, between 12:00 and 18:00. Figure 10 further illustrates a clear weather shift beginning around 12:00. From this time onward, the estimated emissions fluctuate between 0 kg/s and 270 kg/s, nine times the reference value. This explains the extremely high standard deviation and standard error.

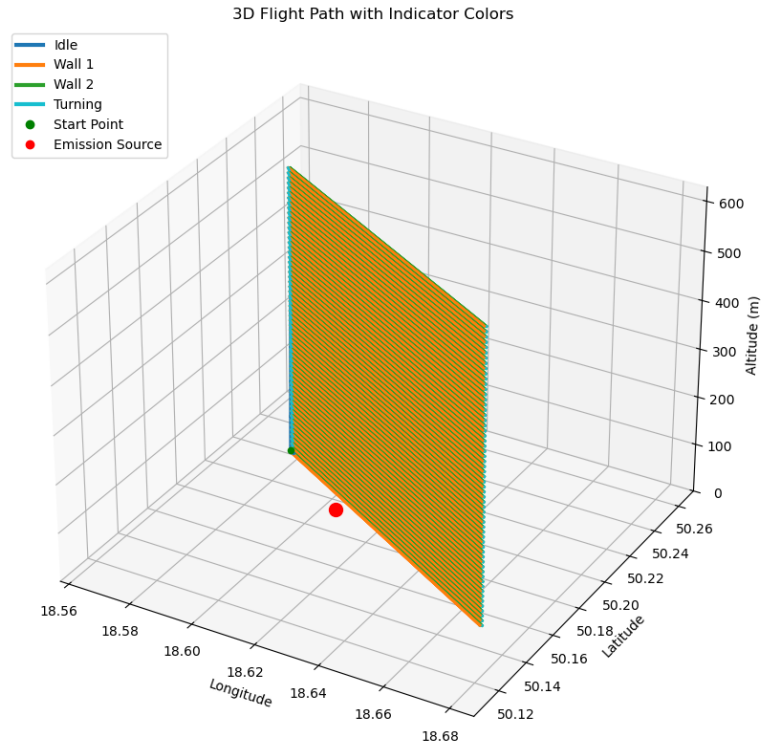


Figure 7: **Three-dimensional representation of the verification flight path.** Colors indicate different flight phases: transporting (blue), Wall 1 (orange), Wall 2 (green), and turning maneuvers (cyan). The red dot marks the emission source location, and the green dot indicates the airport location. Longitude and latitude are shown on the horizontal axes, and altitude (m) on the vertical axis.

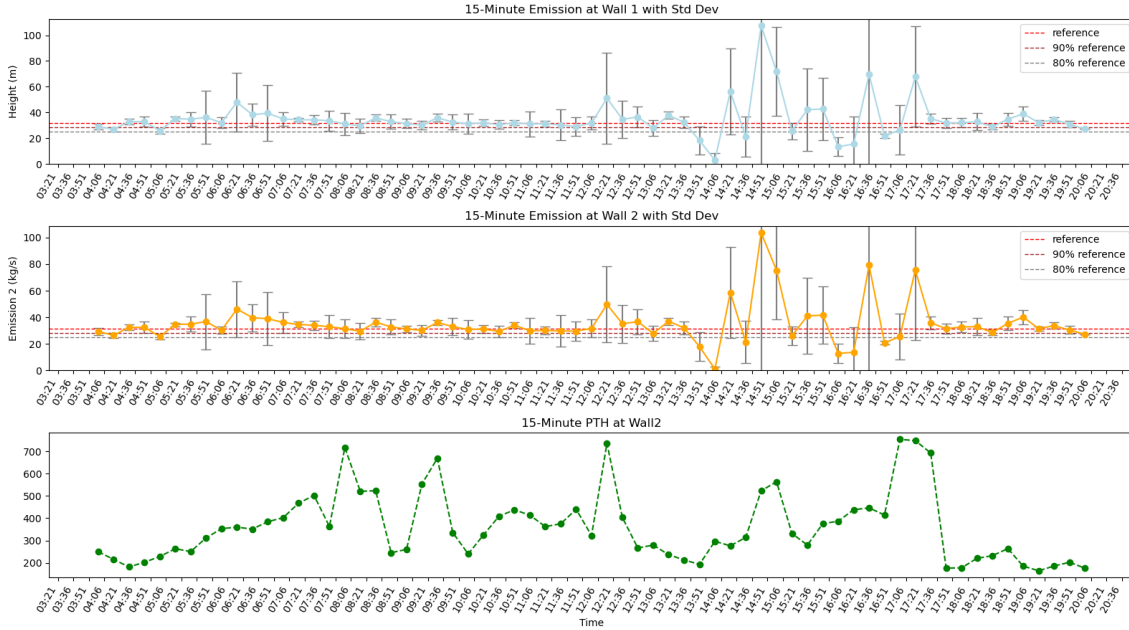


Figure 8: **Mean and standard deviation of verification flights.** Results are shown in 15-minute intervals, corresponding to five flights per interval. (a) Emission estimates at Wall 1 (1150 m) in kg/s with standard deviation. (b) Emission estimates at Wall 2 (1200 m) in kg/s with standard deviation. (c) Plume top height (PTH) at Wall 2. Horizontal dashed lines indicate the reference emission (red), 90% of the reference (dark grey), and 80% of the reference (light grey).

Nevertheless, the emission tracer cannot be lost, and the daily median remains close to the reference value. The observed overestimation by the mean probably originates from rapid wind-direction changes, which cause the walls to register the tracer multiple times. An example of this is shown for the flight started at 12:21 UTC in Fig. 9.

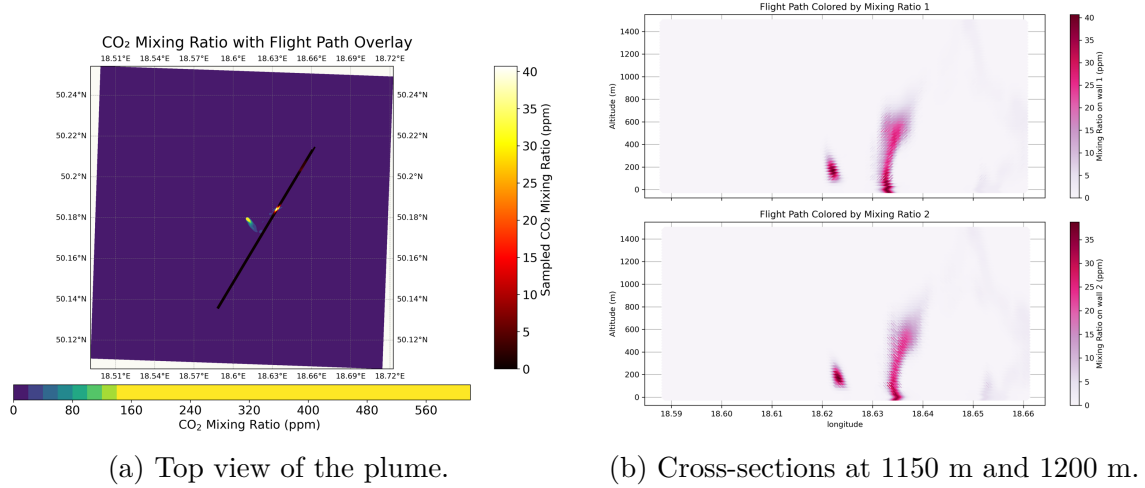


Figure 9: Impact of wind direction change on plume measurements. (a) Top view of the plume, upper part of emission already measured by the frame before, now measuring old tracer and new tracer due to sudden wind change. Sampled mixing ratio in ppmV. (b) Cross-sections at both walls. Due to a rapid wind direction change, tracer emitted at different times is intersected, resulting in two apparent plumes and causing overestimation of emissions. On the left is tracer mass from an earlier time, on the right is tracer mass emitted later.

These results illustrate the importance of less variable meteorological conditions for accurate mass-balance estimation. The window between 12:00 and 18:00 shows significantly fluctuating emission estimate and meteorological conditions. For this reason, the remainder of this study focuses on the time window between 04:00 and 12:00.

A new verification was conducted between 04:00 and 12:00. The first wall was positioned at a distance of 600 m from the emission source, and the second wall at 1200 m.

Results:

- Wall 1: mean = 33.21 kg/s, standard deviation = 5.91 kg/s, standard error = 0.47 kg/s, median = 32.90 kg/s

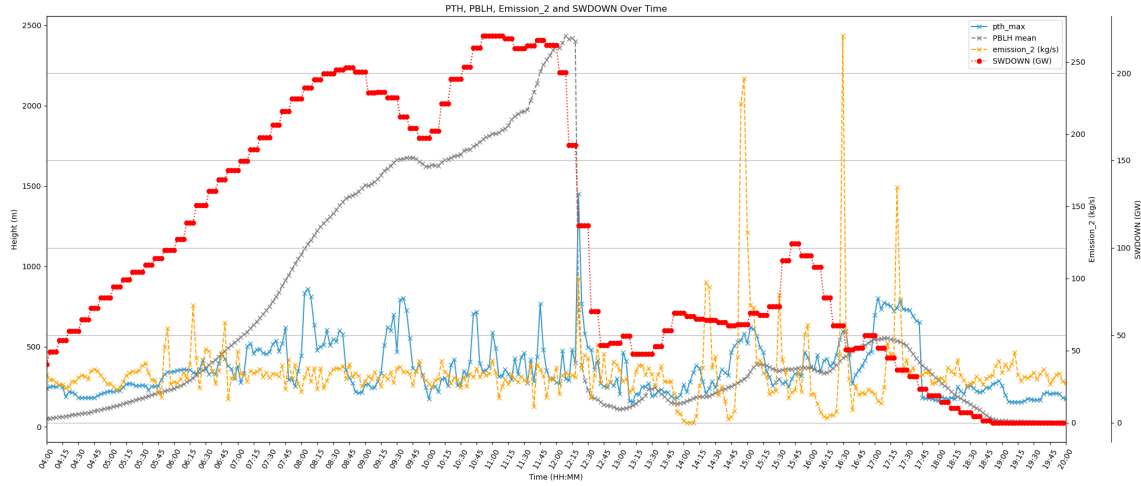


Figure 10: **Time series of plume-top height and related meteorological variables.** Plume-top height (PTH, blue), planetary boundary layer height (PBLH, grey), emission rate (yellow), and short-wave downward radiation (red, integrated in GW) are shown at 1200 m downwind of the emission source. The left axis denotes height (m), while the right axis denotes emission rate (kg/s) and short-wave radiation (GW).

- Wall 2: mean = 32.90 kg/s, standard deviation = 8.41 kg/s, standard error = 0.66 kg/s, median = 32.13 kg/s

Here, mean and median were close. Hence, only the mean is reported in most further analyses. The verification within the selected time window yielded mean values closer to the reference value. Both the standard deviation and the standard error were markedly smaller compared with the full-day data. When measurements from the afternoon were excluded, the resulting uncertainty was reduced to approximately 20–35% of the value obtained for the entire day.

The mean emission estimate of the verification case approaches the true emissions. Thus, the case demonstrates that method works as intended.

3.2 The Base Case

Base Case Results:

- Wall 1: mean = 24.61 kg/s, standard deviation = 10.09 kg/s, standard error = 0.80 kg/s, median = 22.53 kg/s
- Wall 2: mean = 27.46 kg/s, standard deviation = 11.41 kg/s, standard error = 0.91 kg/s, median = 26.42 kg/s

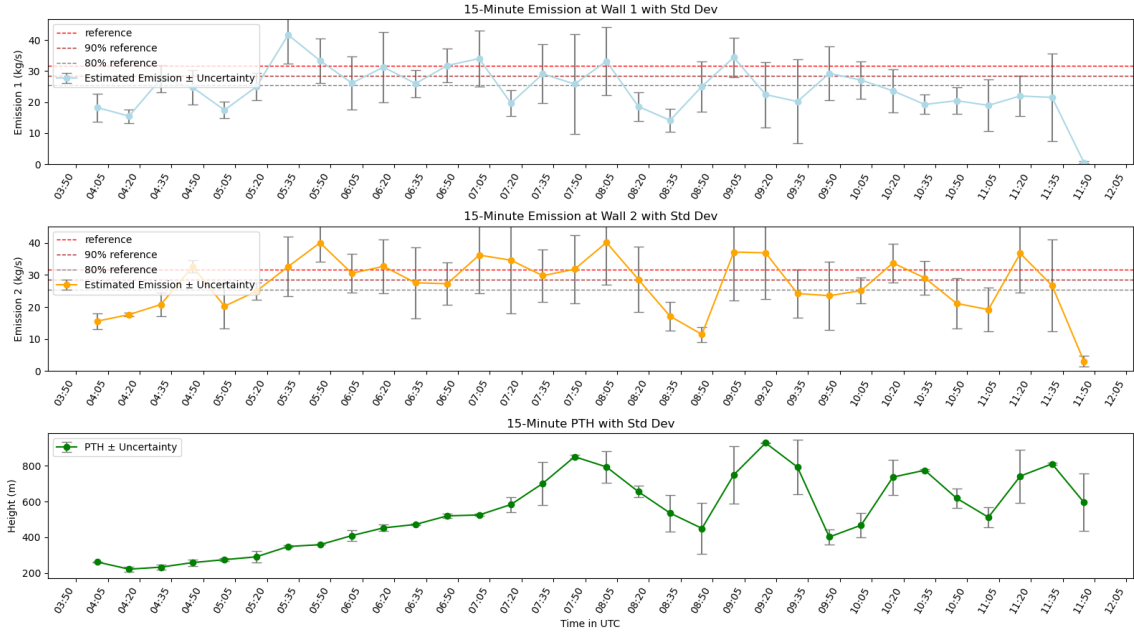


Figure 11: **Base case emission estimates and plume-top height.** Results are shown in 15-minute intervals of virtual flights, corresponding to five flights per interval. (a) Emission estimates at Wall 1 (600 m) in kg/s with standard deviation. (b) Emission estimates at Wall 2 (1200 m) in kg/s with standard deviation. (c) Plume-top height at Wall 2 with uncertainty. Horizontal dashed lines in panels (a) and (b) indicate the reference emission (red), 90% of the reference (dark grey), and 80% of the reference (light grey).

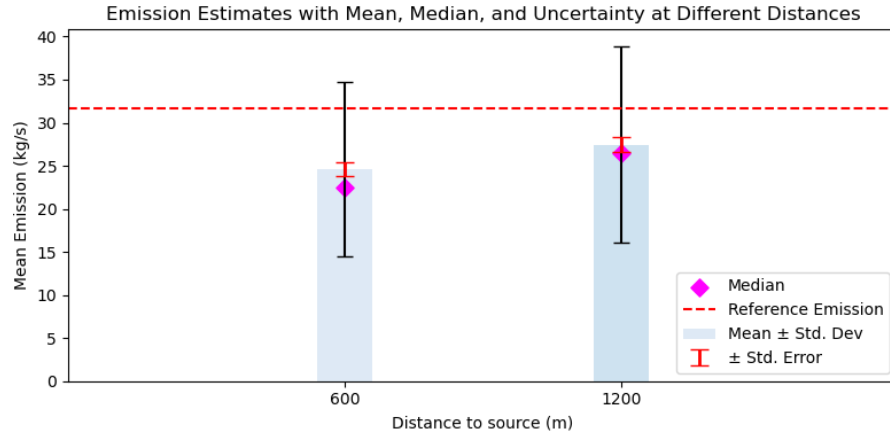


Figure 12: **Base case emission estimates at different distances from the source.** Results are shown for 600 m and 1200 m downwind in kg/s, using a measurement frequency of 1.5 Hz, ground measurements, and a minimum flight height (MFH) of 100 m. The wall width was 2500 m and the total flight time was 43 minutes.

These values show a clear underestimation compared to the known true emission of 31.71kg/s (Fig. 12 and Fig. 11). Compare to the verification settings, the base case has 1/6 the number of transects, 1/3 the measurement frequency, and no measurement coverage between the ground leg and the lowest transect due to the minimum height limit. It also has only 1/4 of the wall width of the verification case, which was made extremely long to ensure horizontal coverage. The flight time is 43 minutes. This significant underestimation could be due to a superposition of all these differences to the verification case. The modification cases will quantitatively compare with the base case to assess and quantify the improvement and identify the causes of the underestimation.

3.3 Modification A: Influence of Measurement Frequency

In this case, the influence of measurement frequency on the accuracy of the emission estimate was investigated. Three configurations were tested:

- Low frequency: 0.5 Hz
- Base case: 1.5 Hz
- High frequency: 4.5 Hz

Results: Wall 1 (600 m):

- Frequency 0.5 Hz: mean = 24.12 kg/s, standard deviation = 10.33 kg/s, standard error = 0.82 kg/s
- Frequency 1.5 Hz: mean = 24.61 kg/s, standard deviation = 10.09 kg/s, standard error = 0.80 kg/s
- Frequency 4.5 Hz: mean = 24.57 kg/s, standard deviation = 10.14 kg/s, standard error = 0.81 kg/s

Wall 2 (1200 m):

- Frequency 0.5 Hz: mean = 27.33 kg/s, standard deviation = 11.67 kg/s, standard error = 0.93 kg/s
- Frequency 1.5 Hz: mean = 27.46 kg/s, standard deviation = 11.41 kg/s, standard error = 0.91 kg/s

- Frequency 4.5 Hz: mean = 27.36 kg/s, standard deviation = 11.38 kg/s, standard error = 0.91 kg/s

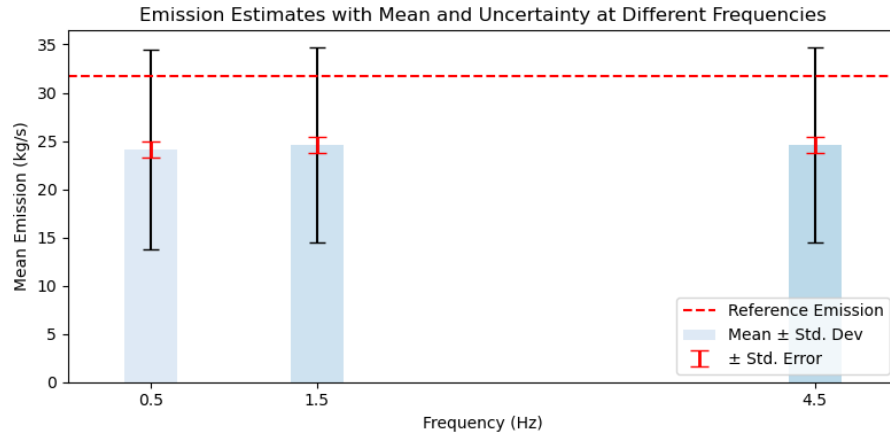


Figure 13: **Emission estimates at Wall 1 (600 m) for different measurement frequencies.** Results are shown in kg/s for 0.5 Hz, 1.5 Hz, and 4.5 Hz. Bars indicate the mean with standard deviation, black error bars represent the standard deviation, and red error bars denote the standard error. The dashed red line marks the reference emission.

Uncertainty: Across all frequency settings, the standard deviations remained within a 1% variation range, and the standard errors differed by only 0.01 kg/s. The ensemble mean varied by less than 2%. This stability indicated that the effects of measurement frequency on the ensemble mean, standard deviation, and standard error were negligible.

Discussion: Increasing the frequency from 1.5 to 4.5 Hz did not improve the emission estimate, suggesting that the measurement frequency used in the base case setup was sufficient.

In contrast, decreasing the frequency to 0.5 Hz resulted in a reduction of the emission estimate by 2% at Wall 1 and 0.5% at Wall 2 as shown in Fig. 13 and Fig. 14. These decreases were small and did not provide evidence that lowering the frequency to 0.5 Hz substantially affected the ensemble mean.

A modified verification case was also performed in which only the frequency parameter was reduced from 4.5 Hz to 1.5 Hz.

Results at 1200 m Wall:

- Frequency 1.5 Hz: mean = 35.14 kg/s, standard deviation = 26.13 kg/s, standard error = 1.46 kg/s

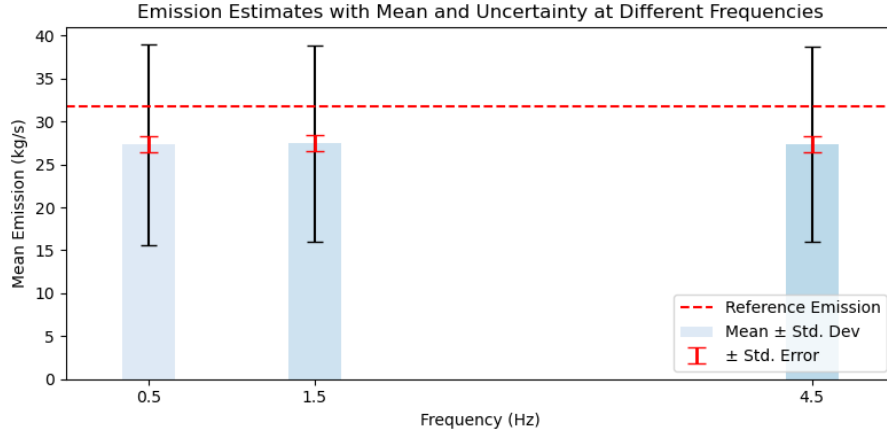


Figure 14: **Emission estimates at Wall 2 (1200 m) for different measurement frequencies.** Results are shown in kg/s for 0.5 Hz, 1.5 Hz, and 4.5 Hz. Bars indicate the mean with standard deviation, black error bars represent the standard deviation, and red error bars denote the standard error. The dashed red line marks the reference emission.

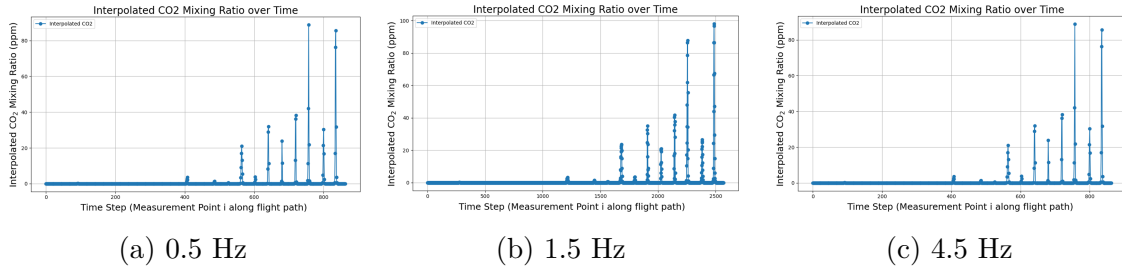


Figure 15: Chronological plots of the sampled mixing ratio (ppmV) along the flight path for increasing measurement frequencies (0.5, 1.5, and 4.5 Hz, left to right). The apparent increase in mixing ratio in each plot reflects the descending altitude in the measurement order, with stronger tracer signals observed at lower altitudes.

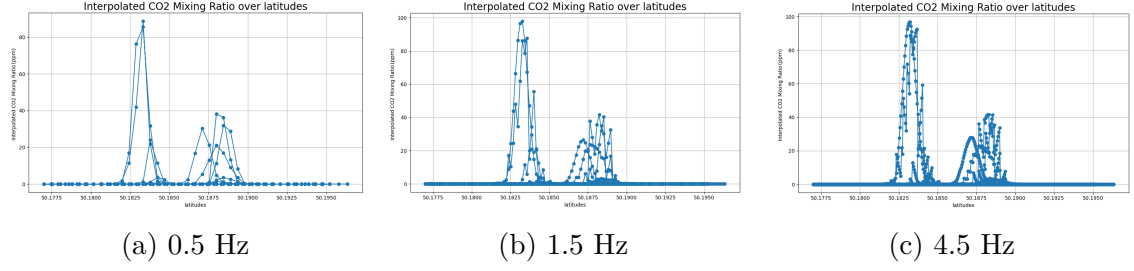


Figure 16: Latitudinal variation of the interpolated mixing ratio (ppmV) from flights with increasing measurement frequencies (0.5, 1.5, and 4.5 Hz, left to right). With increasing frequency, the plume pattern becomes smoother and better resolved, interestingly the main peaks are already well captured at 0.5 Hz.

- Frequency 4.5 Hz: mean = 35.14 kg/s, standard deviation = 26.20 kg/s, standard error = 1.46 kg/s

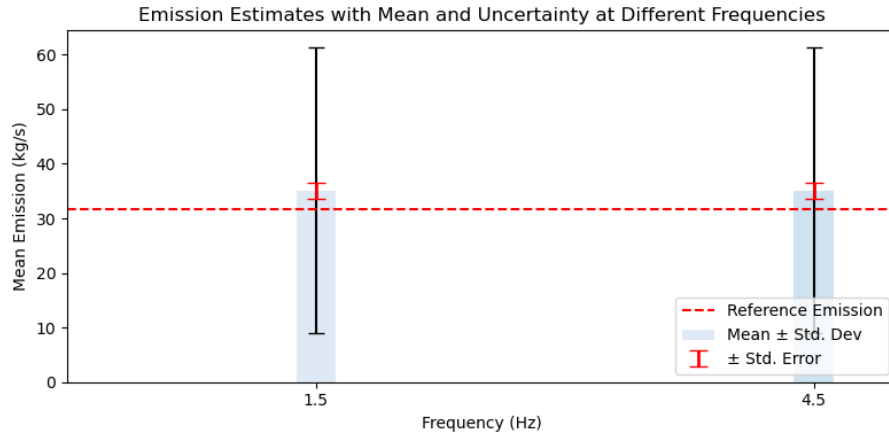


Figure 17: **Emission estimates at Wall 2 (1200 m) from the verification setup.** Results are shown in kg/s for 1.5 Hz, and 4.5 Hz. Bars indicate the mean with standard deviation, black error bars represent the standard deviation, and red errors bar denote the standard error. The dashed red line marks the reference emission.

The results at 1.5 Hz and 4.5 Hz were nearly identical as shown in Fig. 17. This can be explained by the fact that at 1.5 Hz, the WRF-simulated grid cells were already sampled sufficiently, and higher frequencies led only to oversampling without improving the estimate. This trend is shown in Fig. 15 and Fig. 16. To be specific, at a flight speed of 40 m/s, this corresponds to sampling intervals of 26.7 m and 8.9 m respectively. Since the WRF grid resolution was 50 m, both frequencies already

oversampled the grid cells, and oversampling at higher frequencies did not improve the estimate. In real-world applications, however, oversampling may not occur in the same way, potentially allowing higher frequencies to yield slightly more precise estimates. Nevertheless, based on both instantaneous cases, the improvement from higher measurement frequencies in real-world scenarios is expected to remain small. In summary, a sampling frequency of 1.5 Hz is already sufficient for this plume structure and model resolution. Increasing the frequency beyond this value did not improve accuracy, while decreasing the frequency introduced a small but noticeable reduction in accuracy. Therefore, measurement frequencies substantially below 0.5 Hz may introduce larger errors.

3.4 Modification B: Number of Transects

This section examines how the number of vertical transects per wall affected the accuracy and stability of the emission estimate. The base case used 11 transects (10 airborne transects and one ground measurement), while additional cases tested 20 and 30 airborne transects to assess the impact of vertical resolution.

Experimental setups

The following configurations were compared:

Results: Wall 1 (600 m):

- nTS = 10: mean = 24.61 kg/s, standard deviation = 10.09 kg/s, standard error = 0.80 kg/s
- nTS = 20: mean = 23.19 kg/s, standard deviation = 11.08 kg/s, standard error = 0.88 kg/s
- nTS = 30: mean = 20.61 kg/s, standard deviation = 11.78 kg/s, standard error = 0.94 kg/s

The results (Fig. 18) showed a decreasing and thus less accurate ensemble mean with increasing nTS, which was unexpected. Further analysis indicated that this decrease was caused by a weather shift beginning at 12:00 (see Fig. 10). Flights with more transects required longer measurement durations. For example, flights with nTS = 30, starting at 12:00 ended around 13:30, when the meteorological changes led to lower emission estimates.

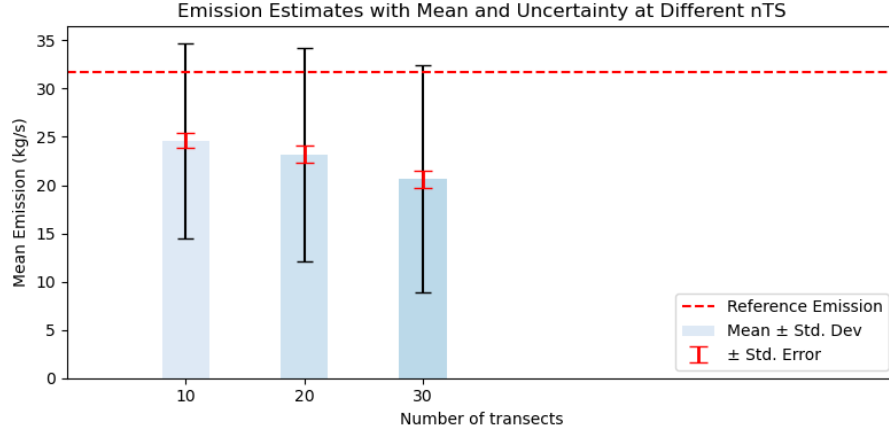


Figure 18: **Emission estimates at Wall 1 (600 m) for different nTS.** Results are shown in kg/s for 10, 20 and 30 transects. Base case flights with nTS modified. Bars indicate the mean with standard deviation, black error bars represent the standard deviation, and red errors bar denote the standard error. The dashed red line marks the reference emission.

To ensure comparability, the aircraft speed was increased while maintaining the spatial distance the same, so that the total flight time remained similar across cases.

New results with boosted aircraft speed: Wall 1 (600 m):

- nTS = 10, with ground measurement, speed $\times 1$: mean = 24.61 kg/s, standard deviation = 10.09 kg/s, standard error = 0.80 kg/s
- nTS = 20, with ground measurement, speed $\times 2$: mean = 23.69 kg/s, standard deviation = 9.46 kg/s, standard error = 0.75 kg/s
- nTS = 30, with ground measurement, speed $\times 3$: mean = 23.47 kg/s, standard deviation = 8.71 kg/s, standard error = 0.69 kg/s
- nTS = 60, with ground measurement, speed $\times 6$: mean = 23.22 kg/s, standard deviation = 8.18 kg/s, standard error = 0.64 kg/s

These results showed more stable ensemble means at 600 m with decreasing standard deviations as nTS increased (Fig. 19). While the means still underestimated the true emissions by 22%, the uncertainty decreased with larger sample sizes. This behavior is consistent with the expectation that averaging over the same time interval reduces uncertainty as the number of samples increases. The reason why the higher number of transects still results in slightly (6%) smaller and thus less accurate emission estimates

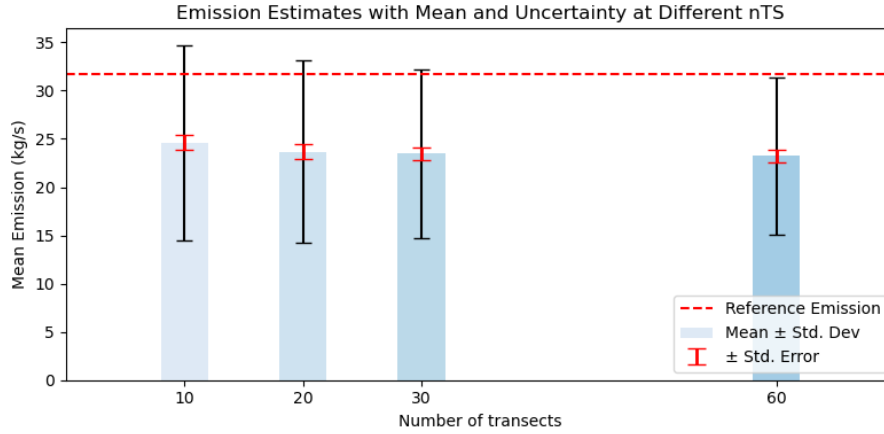


Figure 19: **Emission estimates at Wall 1 (600 m) for different nTS.** Results are shown in kg/s for 10, 20, 30 and 60 transects. Measurement flights with ground measurement included and boosted aircraft speed. Bars indicate the mean with standard deviation, black error bars represent the standard deviation, and red errors bar denote the standard error. The dashed red line marks the reference emission.

was not determined, as addressing the overall underestimation was the more urgent issue.

To isolate the effect of nTS independent of flight time, the verification case with walls at 600 m and 1200 m was repeated, reducing nTS from 60 to 11 while holding all other parameters constant.

Results: Wall 1 (600 m):

- nTS = 60, with ground measurement: mean = 33.21 kg/s, standard deviation = 5.91 kg/s, standard error = 0.47 kg/s
- nTS = 10, with ground measurement: mean = 33.09 kg/s, standard deviation = 6.05 kg/s, standard error = 0.48 kg/s

Wall 2 (1200 m):

- nTS = 60, with ground measurement: mean = 32.90 kg/s, standard deviation = 8.41 kg/s, standard error = 0.66 kg/s
- nTS = 10, with ground measurement: mean = 32.89 kg/s, standard deviation = 8.35 kg/s, standard error = 0.66 kg/s

Figure 20 shows the difference between nTS = 60 and nTS = 11 was less than 0.5% for both walls when emissions were estimated instantaneously. Standard deviations and

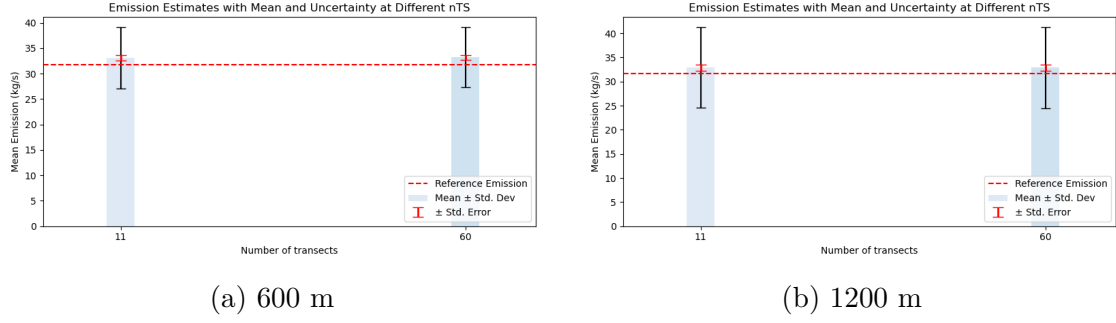


Figure 20: **Emission estimates at Wall 1 (600 m) and Wall 2 (1200 m) for different nTS.** Results are shown in kg/s for 11 and 60 transects at 600 m and 1200 m. Flight from verification case with modified nTS, with ground measurement. Bars indicate the mean with standard deviation, black error bars represent the standard deviation, and red errors bar denote the standard error. The dashed red line marks the reference emission.

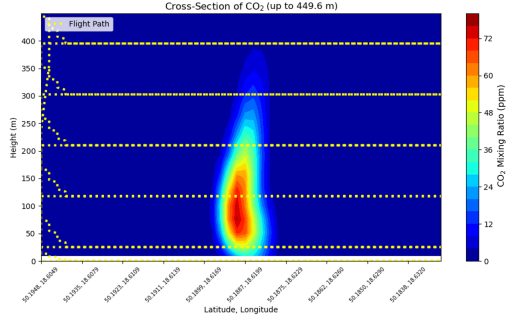
standard errors were also considerably smaller than in the base case. Notably, even with $nTS = 11$, the estimates were already 99.5% accurate relative to the verification. The higher ensemble means observed in the verification case were not caused by the number of transects, since reducing nTS from 60 to 11 did not lower the mean. This shows that undersampling in the vertical direction was not responsible for the bias. The experiments with adjusted flight speed also led to a conclusion that longer flight time led to less accurate and less precise emission estimation, presumably due to higher risk of influence from wind direction changes. If the plume shifts during a long sampling period, the transects may no more fully capture all tracers. Other factors, such as the minimum flight altitude and the distance from the source, need to be examined instead.

3.5 Modification C: Minimum Flight Height

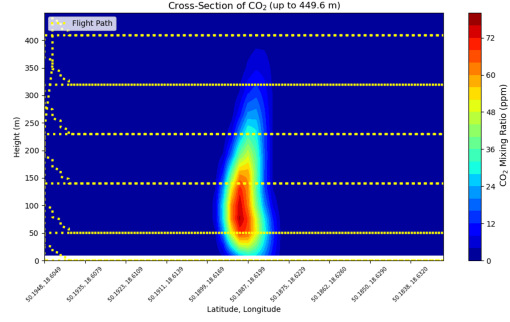
This section investigates the sensitivity of emission estimates to the minimum altitude of flight transects. The base case used a minimum altitude of 100 m, while alternative configurations tested 150 m, 75 m, 50 m, 25 m, and 1 m. These variations are important because, under shallow boundary-layer conditions, a significant portion of the plume can remain near the surface. If this lower region is not sampled, flux estimates may become biased.

Results:

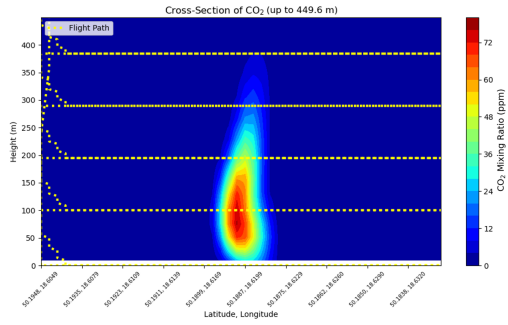
Wall 1 (600 m):



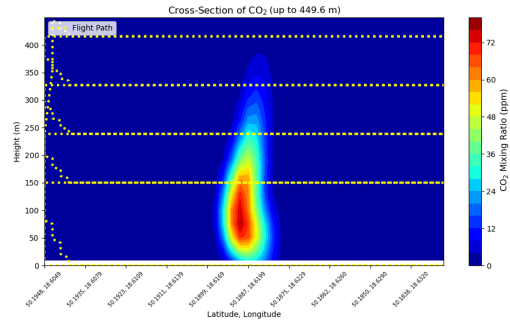
(a) Plume cross-section at 08:32 with MFH = 25 m.



(b) Plume cross-section at 08:32 with MFH = 50 m.



(c) Plume cross-section at 08:32 with MFH = 100 m.



(d) Plume cross-section at 08:32 with MFH = 150 m.

Figure 21: **Plume cross-sections at different minimum flight heights (MFH).** Results are shown for MFH = 25, 50, 100, and 150 m with nTS = 10, with ground-based measurements, at 1200 m downwind from the source. Some transects are not visible at this time because the plume-top height was low, but the highest transect is defined using the maximum PTH over the entire flight.

- MFH = 1 m: mean = 29.73 kg/s, standard deviation = 10.57 kg/s, standard error = 0.84 kg/s
- MFH = 25 m: mean = 30.62 kg/s, standard deviation = 10.28 kg/s, standard error = 0.82 kg/s
- MFH = 50 m: mean = 32.97 kg/s, standard deviation = 11.03 kg/s, standard error = 0.88 kg/s
- MFH = 75 m: mean = 29.01 kg/s, standard deviation = 10.04 kg/s, standard error = 0.80 kg/s
- MFH = 100 m: mean = 24.61 kg/s, standard deviation = 10.09 kg/s, standard error = 0.80 kg/s
- MFH = 150 m: mean = 22.31 kg/s, standard deviation = 11.33 kg/s, standard error = 0.90 kg/s

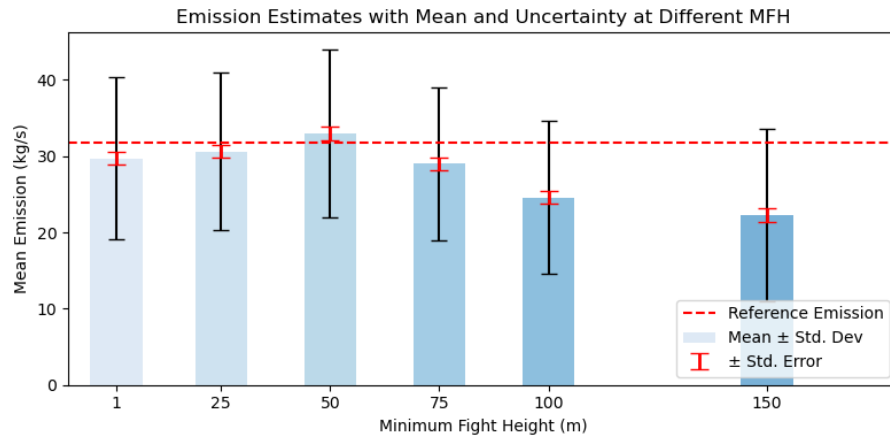


Figure 22: **Emission estimates at Wall 1 (600 m) for varying minimum flight heights.** Results are shown in kg/s for MFH values of 1, 25, 50, 75, 100, and 150 meter. Bars indicate the mean with standard deviation, black error bars represent the standard deviation, and red error bars denote the standard error. The dashed red line marks the reference emission.

Wall 2 (1200 m):

- MFH = 1 m: mean = 30.48 kg/s, standard deviation = 11.39 kg/s, standard error = 0.91 kg/s

- MFH = 25 m: mean = 32.16 kg/s, standard deviation = 11.69 kg/s, standard error = 0.93 kg/s
- MFH = 50 m: mean = 32.22 kg/s, standard deviation = 12.04 kg/s, standard error = 0.96 kg/s
- MFH = 75 m: mean = 30.05 kg/s, standard deviation = 11.48 kg/s, standard error = 0.91 kg/s
- MFH = 100 m: mean = 27.46 kg/s, standard deviation = 11.41 kg/s, standard error = 0.91 kg/s
- MFH = 150 m: mean = 25.54 kg/s, standard deviation = 13.03 kg/s, standard error = 1.04 kg/s

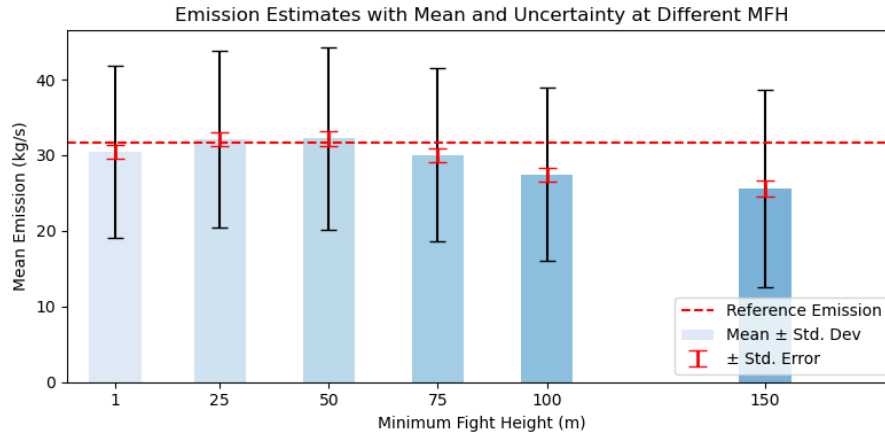


Figure 23: **Emission estimates at Wall 2 (1200 m) for varying minimum flight heights.** Results are shown in kg/s for MFH values of 1, 25, 50, 75, 100, and 150 meter. Bars indicate the mean with standard deviation, black error bars represent the standard deviation, and red error bars denote the standard error. The dashed red line marks the reference emission.

Discussion

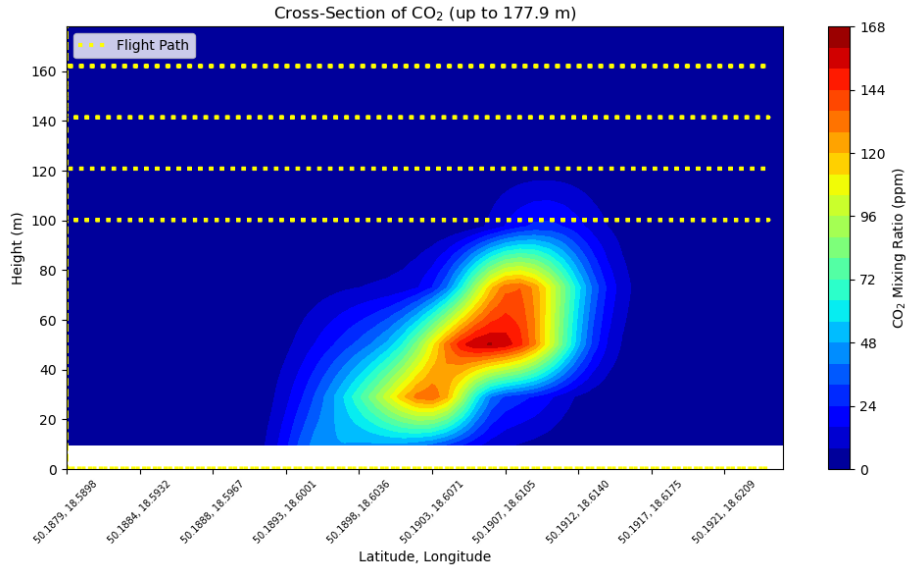
Lowering the minimum flight height improved the emission estimates, particularly at Wall 1 (Fig. 22). When MFH was reduced from 100 m to 50 m, the mean increased by 33.9% (from 24.61 to 32.97 kg/s). At Wall 2 (Fig. 23), the increase was smaller (about 17%). Both walls showed a peak in the mean emission estimate at MFH = 50 m, where the plume maximum was located near the lowest transect (Fig. 21).

Outside this peak at 50 m, the results declined again toward both lower and higher MFH values.

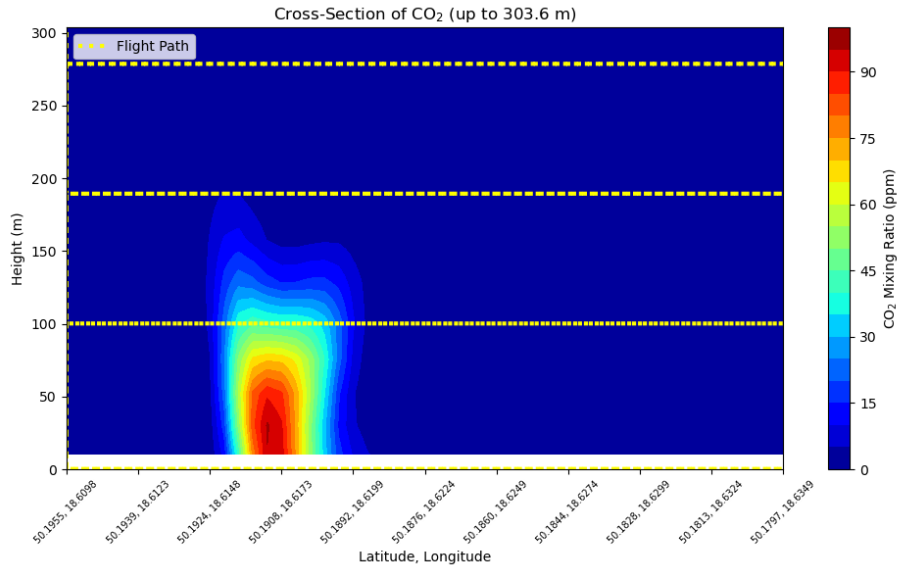
This pattern is consistent with physical expectations. Under shallow boundary-layer conditions, much of the plume can remain below 100 m. If flights are restricted to higher altitudes, transects may miss the plume core, as illustrated at 04:00 (Fig. 24a) and during a weather change around 11:30 (Fig. 24b).

The highest accuracy (Wall 1 = 32.97 kg/s) was achieved at MFH = 50 m rather than at 1 m. This can be explained by the effective emission height, which in this case was about 50 m. As shown in Fig. 21b, transects at 50 m therefore had a high probability of capturing the plume maximum. The results indicate that simply lowering MFH does not always improve accuracy. Instead, optimal estimates occur when the transects intersect the plume core. Note that the maximum flight height (maximum of PTH) was often very high because it was the maximum over the entire flight time. This spaced out vertical distance a lot, as can be seen in Fig. 24a and Fig. 24b. This is the reason why the MFH of 1 m even with 10 transects yields a good result in the verification case (quasi-instantaneous flight path) but not in this case. When **no ground measurements are available** and the plume maximum is captured at the **lowest transect**, the mixing ratio from that transect is integrated down to the ground using the measured value. This leads to three possible outcomes:

1. **Correct estimation:** This occurs only when the plume is either well mixed throughout the boundary layer or stagnant down to the ground with approximately uniform concentrations, so that integrating the lowest transect value down to the ground yields a correct column integral.
2. **Underestimation:** This occurs when the plume lies below the lowest transect, as shown in Fig. 24, which then only captures the weakened upper portion of the plume. In this case, the flux is integrated down to the ground without ever sampling the plume core, leading to an underestimated value for that flight.
3. **Overestimation:** This occurs when the lowest transect coincides with the plume maximum, so the value is integrated throughout the column down to the ground. Such overestimation can affect individual flights but does not necessarily bias the ensemble mean, since it only occurs when the plume lies very low (near the MFH). Figure 21b illustrates this: without the ground measurement, the flux would be overestimated, whereas including the ground measurement allows



(a) At 04:00 plume core missed.



(b) At 11:30 plume core missed.

Figure 24: **Examples of extremely low plumes that cannot be properly captured with a minimum flight height (MFH) of 100 m.** (a) At 04:00, the plume core lies below the MFH, and (b) at 11:30, the plume is similarly missed. Both cases demonstrate how low plumes lead to incomplete sampling. Flight path is from base case setup, mixing ratio is in ppmV and time in UTC.

the lowest transect to be integrated only halfway to the ground, yielding a correct emission estimate.

This effect depends on which part of the plume is sampled by the lowest transect and is distinct from the random over- and underestimations introduced by turbulent plume puffs. Notably, at greater distances this effect becomes much weaker due to larger distance for vertical mixing.

This makes ground measurements especially important. To assess the impact of their absence, additional tests without ground measurement were performed.

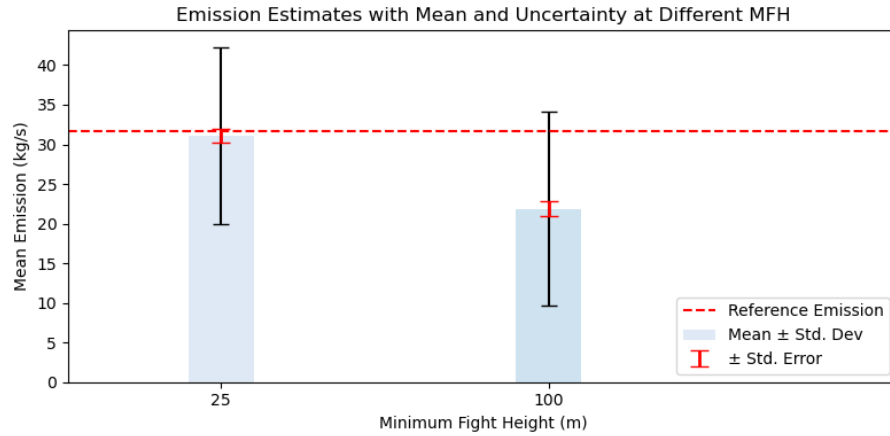


Figure 25: **Emission estimates at Wall 1 (600 m) without ground measurements.** Results are shown in kg/s for minimum flight heights of 25 m and 100 m. Bars indicate the mean with standard deviation, black error bars represent the standard deviation, and red error bars denote the standard error. The dashed red line marks the reference emission.

3.5.1 Results without ground measurement (MFH = 25 and 100 m)

Wall 1 (600 m):

- MFH = 100: mean = 21.88 kg/s, standard deviation = 12.28 kg/s, standard error = 0.98 kg/s
- MFH = 25: mean = 31.06 kg/s, standard deviation = 11.19 kg/s, standard error = 0.89 kg/s

Wall 2 (1200 m):

- MFH = 100: mean = 25.24 kg/s, standard deviation = 11.58 kg/s, standard error = 0.92 kg/s

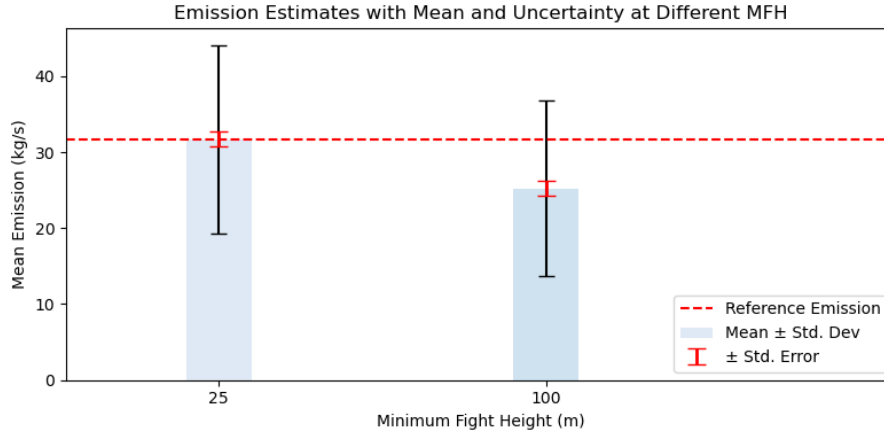


Figure 26: **Emission estimates at Wall 2 (1200 m) without ground measurements.** Results are shown in kg/s for minimum flight heights of 25 m and 100 m. Bars indicate the mean with standard deviation, black error bars represent the standard deviation, and red error bars denote the standard error. The dashed red line marks the reference emission.

- MFH = 25: mean = 31.70 kg/s, standard deviation = 12.36 kg/s, standard error = 0.98 kg/s

Figure 25 and Fig. 26 show results at 600 m and 1200 m without ground measurement. Measurement flight with MFH = 100 m further underestimated the emission, flight while MFH = 25 m estimated values similar to those with ground measurements. This supports the above interpretation that the treatment of the lowest transect strongly influences results.

It is important to note that there are two main approaches to calculate the emission flux at the lowest level: (1) integrating the full vertical distance to the ground, assuming the plume is well mixed, or (2) integrating only half the distance, assuming reduced concentrations near the ground. In this study, the full path was used. Future work should evaluate which approach is more appropriate under different meteorological conditions and source distances (cf. Conley et al. (2017)).

In summary, lower MFH did not guarantee better estimates. The key factor was whether transects intersected the plume core, which depends on the effective emission height. For buoyant plumes with higher effective release heights (e.g., 400 m), an MFH of 100 m would not affect accuracy. In this case, the optimal MFH was about 50 m, because it ensured a flight leg close to the plume maximum at the studied distance.

3.6 Modification D: Distance to Source

The results from the MFH experiments indicated that the wall at 1200 m almost always provided more accurate emission estimates than the wall at 600 m. This raised the question of how the distance from the source influences the accuracy of flux estimates.

The verification experiments, however, did not provide strong evidence that increasing distance alone systematically improved results. However, from the nTS modification we know that longer flight time from more nTS would lead to less emission estimate (Compare with Sect.3.4). To investigate this effect further, six distances were tested. The verification setup with speed boost was modified to use 11 transects (lowered from 60), MFH = 100 m (originally 1 m) with a ground measurement, a sampling frequency of 4.5 Hz, and quasi-instantaneous aircraft speed so that results were independent of temporal variability of the meteorological conditions.

Results:

- Distance = 600 m: mean = 25.74 kg/s, median = 24.34 kg/s, standard deviation = 8.39 kg/s, standard error = 0.66 kg/s
- Distance = 1200 m: mean = 28.39 kg/s, median = 27.70 kg/s, standard deviation = 10.22 kg/s, standard error = 0.81 kg/s
- Distance = 2400 m: mean = 31.14 kg/s, median = 29.89 kg/s, standard deviation = 10.47 kg/s, standard error = 0.82 kg/s
- Distance = 3600 m: mean = 32.31 kg/s, median = 31.75 kg/s, standard deviation = 9.50 kg/s, standard error = 0.75 kg/s
- Distance = 4800 m: mean = 33.49 kg/s, median = 31.82 kg/s, standard deviation = 10.79 kg/s, standard error = 1.03 kg/s
- Distance = 6000 m: mean = 34.38 kg/s, median = 32.35 kg/s, standard deviation = 12.26 kg/s, standard error = 1.17 kg/s

These results showed clear underestimation at 600 m and 1200 m. At greater distances, mean estimates increased by at least 10% relative to 1200 m. However, both standard deviation and standard error also increased with distance, reflecting greater variability. The mean values from walls further away tend to overestimate the true emission, whereas the medians are closer to the true emission rate.

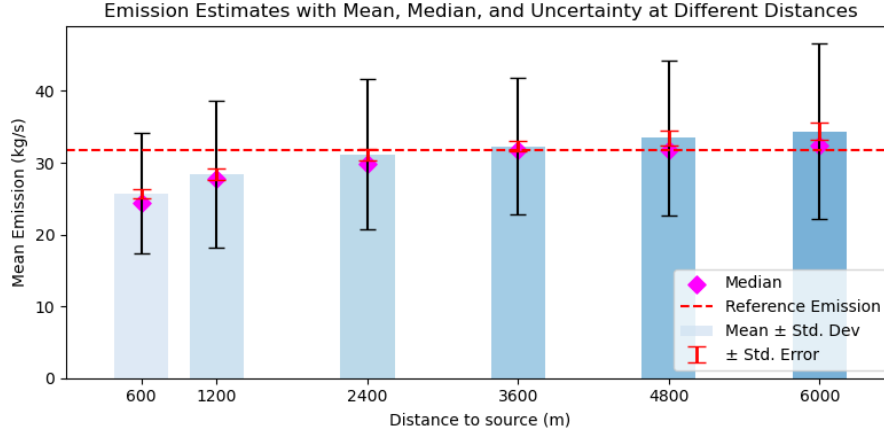


Figure 27: **Emission estimates at varying distances from the source.** Results from are shown in kg/s for distances of 600, 1200, 2400, 3600, 4800, and 6000 m downwind. Measurement flights are verification cases with modified distances, MFH = 100 m, with ground measurement. Bars indicate the mean, black error bars represent the standard deviation, and red error bars denote the standard error. Magenta diamonds show the median, and the dashed red line marks the reference emission.

A possible reason could be that the double-counting effect discussed in Sect. 3.1 is more likely to occur at greater distances, though the times when this phenomenon was most likely (12:00–18:00) were excluded. In this set of quasi-instantaneous experiments, slow meteorological variability was excluded (increasing PBLH, slow wind direction change), while rapid fluctuations can still shape the plume (Fig. 9). This suggests that the increasing overestimation beyond 3600 m may be linked to the combined effects of distance, wall width, turbulence (puffs) and rapid change of wind direction. However, no dedicated tests were performed to confirm this, so a firm conclusion cannot yet be drawn.

Compared to the mean, the median values vary more slowly after 3600 m. The median of 32.35 kg/s at 6000 m clearly overestimates the true value of 31.71 kg/s, but is still smaller than the verification case (Sect. 3.1) at 2400 m (32.90 kg/s). To confirm the more accurate median observed at greater wall distances for real flights, the base case setup was evaluated at the same distances. Instead of accelerating the flight speed, flights extending beyond 12:00 were excluded from the ensemble to minimize the influence of weather changes around noon. The wall width was also adjusted. In the instantaneous flights it was fixed at 10,000 m (made extremely long to ensure capture of all emission), whereas in the real flights it was chosen according to experimental

guidance (Sect. 2.7.3).

Results:

Wall 1 (600 m) and Wall 2 (1200 m), **wall width = 2500 m:**

- Wall 1: mean = 25.27 kg/s, median = 23.12 kg/s, standard deviation = 9.52 kg/s, standard error = 0.79 kg/s
- Wall 2: mean = 27.73 kg/s, median = 26.65 kg/s, standard deviation = 10.79 kg/s, standard error = 0.89 kg/s
- **flight time: 43 minutes**

Wall 1 (2400 m) and Wall 2 (3600 m), **wall width = 5000 m:**

- Wall 1: mean = 30.08 kg/s, median = 27.12 kg/s, standard deviation = 13.62 kg/s, standard error = 1.17 kg/s
- Wall 2: mean = 31.05 kg/s, median = 30.03 kg/s, standard deviation = 14.58 kg/s, standard error = 1.25 kg/s
- **flight time: 72 minutes**

Wall 1 (4800 m) and Wall 2 (6000 m), **wall width = 7500 m:**

- Wall 1: mean = 30.08 kg/s, median = 29.65 kg/s, standard deviation = 12.71 kg/s, standard error = 1.12 kg/s
- Wall 2: mean = 29.46 kg/s, median = 28.77 kg/s, standard deviation = 10.80 kg/s, standard error = 0.95 kg/s
- **flight time: 95 minutes**

These results are more realistic than the instantaneous experiments shown in Fig. 27 because they account for the effects of variability in wind direction and PTH (meteorology variability). The walls at 600 m and 1200 m still underestimated the emissions, while distances of 2400 m and greater produced estimates close to the true value. The increase from 600 m to 2400 m is 19%, showing too close wall placement would cause strong underestimation. At greater distances from the source, the mean emission estimates decreased. A possible explanation is that small parts of the plume extended beyond the lateral boundaries of the measurement wall, so the wall width did not fully capture the plume throughout the 95-minute flight. Increasing the wall width would

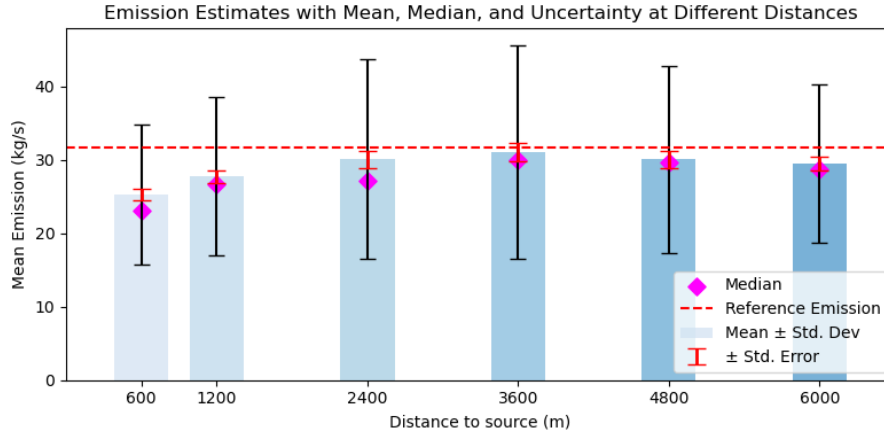


Figure 28: **Emission estimates at varying distances from the source.** Results are shown in kg/s for distances of 600, 1200, 2400, 3600, 4800, and 6000 m downwind. Base case flights with modified distances and wall widths. Bars indicate the mean, black error bars represent the standard deviation, and red error bars denote the standard error. Magenta diamonds show the median, and the dashed red line marks the reference emission. Flights ending after 12:00 UTC were excluded to minimize the influence of weather changes around noon.

improve coverage, but this would also lengthen the required flight time. Longer flights raise the likelihood of wind direction changes during the measurement, which can lead to partial plume loss. Thus, there may be an optimum distance from the source that balances the benefit of a shorter flight time with sufficient plume coverage and vertical mixing. Between 3600 m and 4800 m, while the mean values decreased, the median values showed only a smaller reduction. This suggests that most individual estimates remained relatively stable, but a few strongly underestimated cases pulled the mean downward. One could debate which measure is the more representative average, but in real campaigns, with only 6–10 walls per day, such an ensemble is too small for a robust conclusion. Overall, these results show that distance influences the accuracy of emission estimates, particularly at short ranges where the plume is less vertically developed. At greater distances, the plume has more time to mix vertically, reducing the sensitivity of the estimates to MFH limitations. This finding is consistent with Conley et al. (2017). In principle, the distance required for a well-mixed PBL can be calculated. A systematic analysis of the optimal wall width at greater distances for a fixed flight time remains to be done, so no definitive conclusions can yet be drawn for walls at 4800 m and 6000 m. However, if emission estimates

do not improve at these ranges, then allocating additional flight time and resources to such long-range measurements may not be worthwhile. The present comparisons correspond to flight times of 43, 72, and 95 minutes, respectively. When resources are limited, the trade-off between improved accuracy and longer measurement duration must be carefully considered.

3.7 Impact of Meteorological Events

During the simulation period, several meteorological transitions strongly influenced the emission estimates. These events highlight how atmospheric stability, solar radiation, wind variability, and mixing depth govern plume behavior and affect the accuracy of mass-balance calculations. The variation in wind direction shown in Fig. 30 also indicates that a narrow wall combined with a long flight time may fail to capture the entire plume. For instance, between 04:00 and 06:00 UTC the wind direction shifted by about 50° . At a distance of 6000 m from the source, this corresponds to a plume center displacement of roughly 5110 m, calculated using the chord-length formula Eq. 4, where r is the distance from the source to the wall and θ is the wind-direction change in radians.

$$\text{Chord} = 2r \cdot \sin\left(\frac{\theta}{2}\right), \quad (4)$$

With $r = 6000$ m and $\theta = 50^\circ$. Even a horizontal variation of only 5° would require a wall width of at least 6000 m to ensure full plume coverage.

3.7.1 Plume accumulation due to wind-direction change

A dramatic event occurred around 13:50 when cloud cover became complete and the PBLH collapsed below 200 m. This was accompanied by reduced wind speed and, most importantly, a rapid change in wind direction as shown in Fig. 30. As a result, emission tracers accumulated near the source, forming a stagnant plume that failed to propagate downwind. The whole accumulation event as shown in Fig. 32 lasted for more than an hour. At 600 m downwind, some plume signatures were still detected (Fig. 29), but at 3600 m almost no signal was observed (0–5 kg/s, Fig. 30), confirming that the plume never reached that distance. After about 30 minutes, the accumulated gases were suddenly released, producing a sharp peak at 15:06 in emission estimates (Fig. 30).

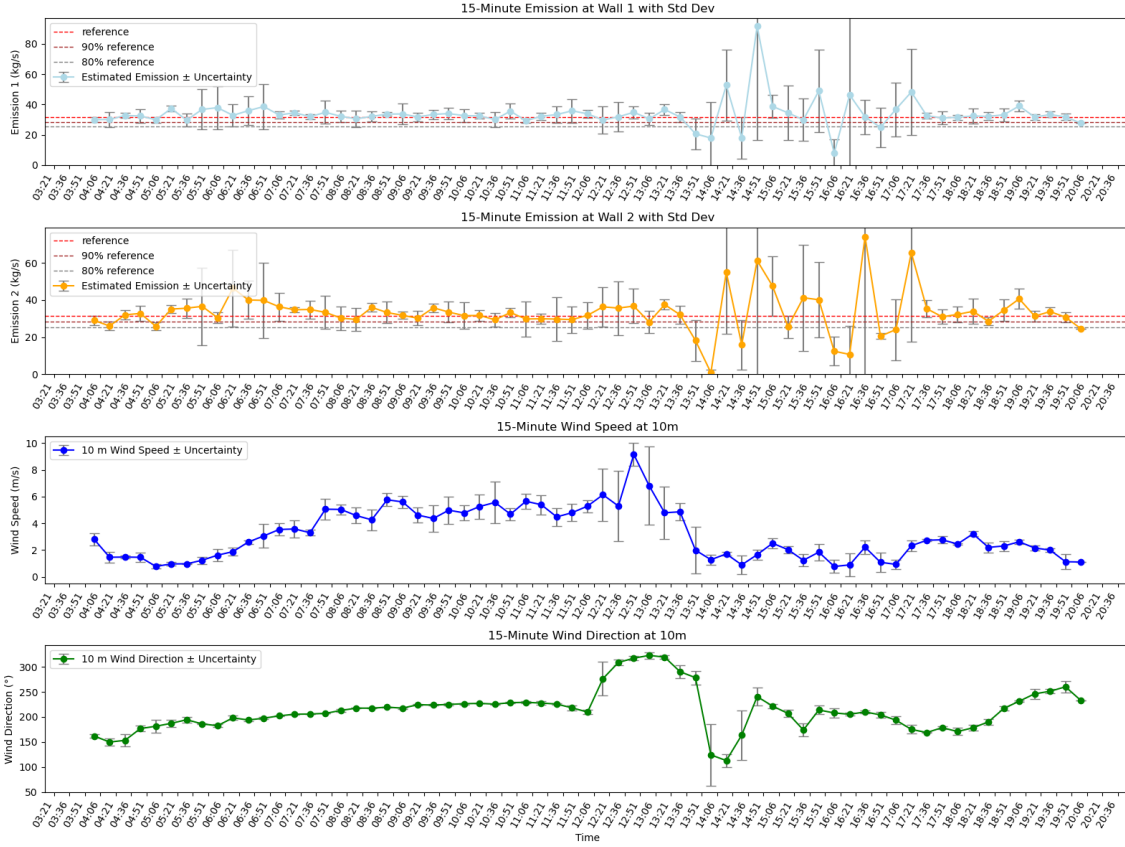


Figure 29: Time series of 15-minute averages of (a) estimated emissions at Wall 1 (600 m downwind) and (b) Wall 2 (1200 m downwind), each shown with standard deviations as error bars. Horizontal dashed lines indicate the reference emission rate and its 90% and 80% values. Panels (c) and (d) show the corresponding 15-minute mean wind speed and wind direction at 10 m above ground, measured 1200 m from the source. Base case configuration. Time in UTC.

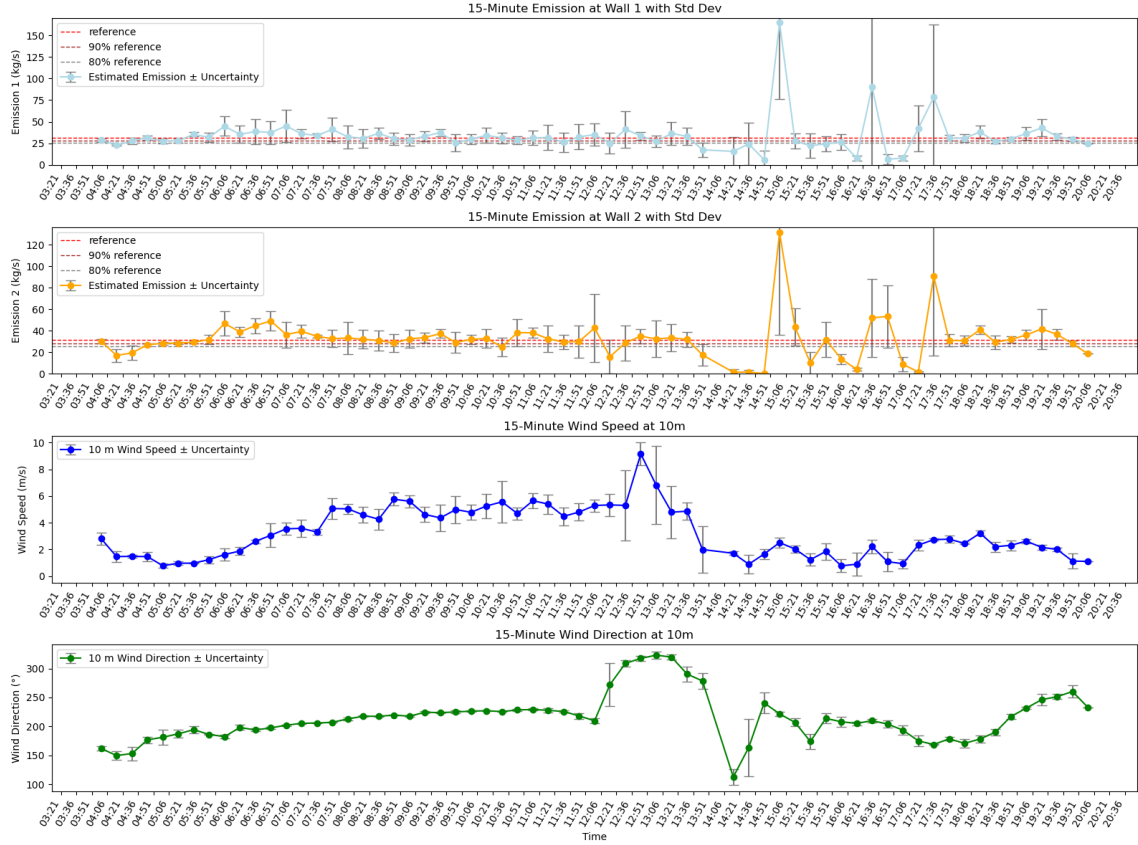


Figure 30: Time series of 15-minute averages of (a) estimated emissions at Wall 1 (2400 m downwind) and (b) Wall 2 (3600 m downwind), each shown with standard deviations as error bars. Horizontal dashed lines indicate the reference emission rate and its 90% and 80% values. Panels (c) and (d) show the corresponding 15-minute mean wind speed and wind direction at 10 m above ground, measured 3600 m from the source. Time in UTC.

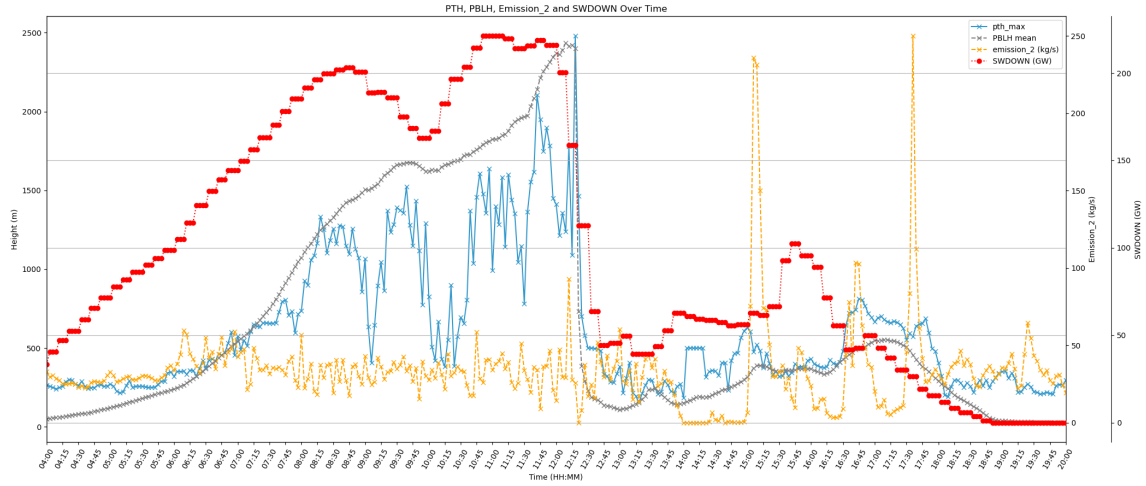


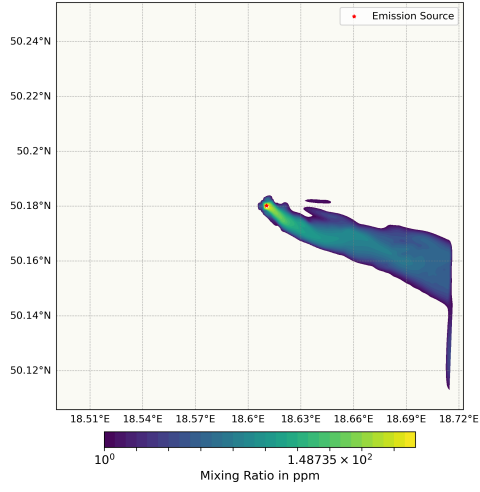
Figure 31: Plume-top height (PTH, blue), planetary boundary layer height (PBLH, grey), emission rate (yellow), and short-wave downward radiation (red, integrated in GW) at 3600 m downwind of the emission source.

This case demonstrates the scale of meteorological influence on emission estimates. Under such variable conditions, even well-designed flight plans may fail to capture the plume, leading to both under- and overestimation. Rapid wind-direction shifts also complicated flight planning, sometimes pushing the plume outside the sampling walls and causing further underestimation.

3.7.2 Cloud impact and sudden wind-speed change at 12:00 UTC

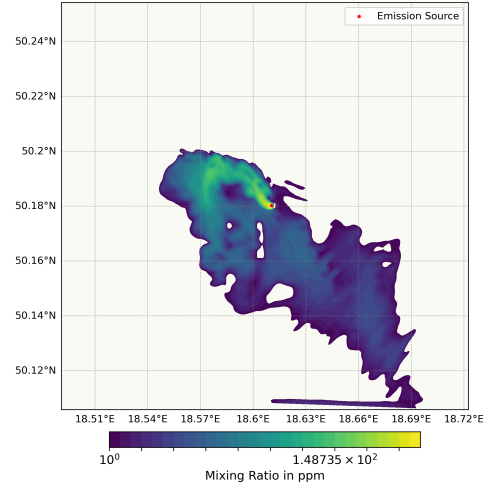
As shown in Fig. 31, during the morning, clear skies led to steadily increasing SWDOWN and a growing PBLH. In the early afternoon (12:00 UTC), both parameters abruptly dropped to about one-fifth of their prior values within only 30 minutes. At 3600 m, the plume-top height followed the collapse of the PBLH (Fig. 31), while at 1200 m the PTH change was less visible because vertical mixing distance was insufficient. Notably, the PBLH collapse lagged behind the drop in SWDOWN (an indicator of cloud cover). This delay might reflect the time needed for turbulence in the boundary layer to decay once surface heating was cut off, as the air column possibly lost the buoyancy support provided by convection, ultimately leading to the collapse of convection. Compared to the stable morning hours, this early afternoon variability was strong. The fluctuations in SWDOWN likely altered convection-driven turbulence, directly affecting plume shape and dispersion. This hypothesis remains to be confirmed in further studies.

Horizontal cross-section of simulated plume at 14:00 UTC



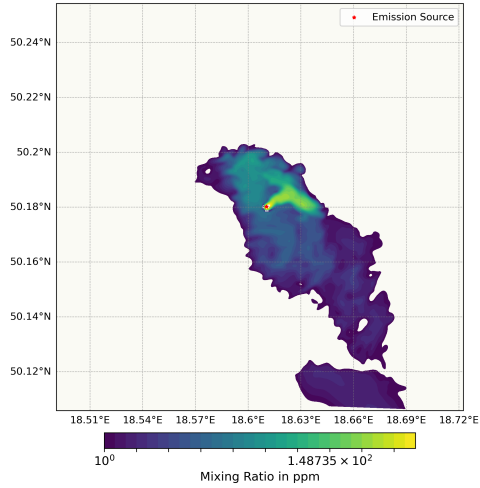
(a) Before accumulation at 14:00 UTC.

Horizontal cross-section of simulated plume at 14:35 UTC



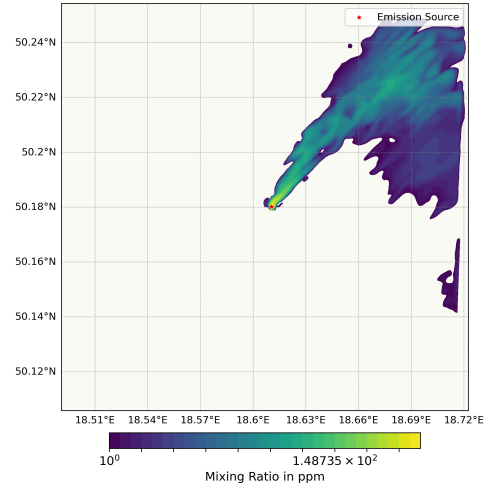
(b) Accumulation at 14:35 UTC.

Horizontal cross-section of simulated plume at 14:55 UTC



(c) Downwash at 14:55 UTC.

Horizontal cross-section of simulated plume at 15:25 UTC



(d) After accumulation at 15:25 UTC.

Figure 32: Emission tracer accumulated at the source due to rapid change of wind direction and released after 30 minutes. Entire process lasted over an hour. Zero estimate and peak estimate clearly visible in Fig. 31 around 13:51 and 15:15 UTC. With wind direction change shown in Fig. 30 at the same interval. This event demonstrates the strong impact of wind variability on emission estimation.

Such a sudden collapse of the PBLH not only modifies plume dispersion but also has direct consequences for flux measurements. In particular, it rapidly shift the plume core to lower altitudes. Flights with longer durations are often configured with walls covering a wide vertical range. In such cases, a sudden shift of the plume core can have two possible outcomes:

1. If the plume previously had a large vertical extent and was already partially sampled above, then sampling the shifted plume core at lower transects may lead to a higher estimate for that measurement flight.
2. If the plume did not extend far upward, the shifted core would be captured only at the lowest transects, which could still yield a reasonable estimate but may also result in underestimation.

In either case, the effect is to increase variability in the estimates and add uncertainty. However, if the flight time is shorter, the smaller variation range of the PTH allows for a more appropriate flight top height, which can lead to more reliable emission estimates. Notably, a reversed order of measurement, for example sampling from the ground upward to the PTH, could completely change this effect. Further research needs to account for this and quantify the potential improvements for different flight patterns, orders, and durations.

3.7.3 Impact of low wind conditions in afternoon and night-time on mass-balance estimates

Between 14:00 and 18:00 UTC, wind speeds vary around 2 m/s (Fig. 30). Varon et al. (2018) specifically note that mass-balance method should not be used under calm-wind conditions, and identified 10 m winds slower than 2 m/s as a threshold where performance degrades. The calm-wind condition causes large variability in plume shape and dispersion, because turbulent motions dominate over mean advection. Therefore it also violates mass-balance assumption, which require a quasi-steady, advected plume crossing the flux plane. The variation in wind direction, wind speed and emission estimate in afternoon align with the conclusion from Varon et al. (2018). It is also worth recalling the statement from experimental colleagues regarding night-time flights in Sect. 2.6:

Earlier flight times were not considered for the analysis for two reasons:

- *First, the plume-top height may be too low during the night to ensure reliable observation of the plume given real-world minimum flight height restrictions.*
- *Second, calm winds during the night may result in accumulations of the tracer, hindering accurate emission estimation.*

Both the cloudy afternoon period analyzed here and the night-time experience reported by colleagues were characterized by low wind speeds. A follow-up study could therefore benefit from a systematic assessment of night-time flights and a quantification of the effects of these potential restrictions.

These findings highlight that mass-balance emission estimates are highly sensitive to meteorological conditions. In particular, shallow PBLH, sudden cloud impacts, and wind-direction changes can produce large biases, while sustained wind speeds below 2 m/s violate a core assumption of the method. For practical campaigns, this underlines the need to avoid flights during low wind speeds or rapidly changing conditions, and to rely on accurate meteorological forecasts for planning.

4 Conclusion

4.1 Impact of flight design is smaller than impact of meteorological variation

Differences caused by flight design choices (measurement frequency, number of transects, minimum flight height, and wall distances) were small compared to the effects of weather. Afternoon changes in wind direction led to very large variations in estimated emissions and dominated the overall uncertainty. The wind speed and direction and the PBLH mainly control how the plume moves and spreads. Depending on the width of the wall, fast wind shifts can cause double counting of the plume, or completely missing it. A sudden collapse of the PBLH (weather change) can rapidly shift the plume core to lower altitudes, resulting in large variability in the estimates and greater uncertainty.

The findings are robust, though certain effects are still speculative. In particular, under strongly variable meteorological conditions, flight time and sampling order may also affect the result. Shorter flights may reduce PTH variability and improve reliability, while a reversed sampling order could significantly alter outcomes. These hypotheses should be tested and quantified in future research.

For practical applications, measurements should be avoided when wind speeds fall below about 2 m/s, since under such calm conditions the mass-balance approach becomes unreliable (Varon et al. 2018). Overall, meteorological effects are much stronger than those from flight design, but many of them can be anticipated in advance with the help of weather forecasts. In particular, cloud cover, wind speed, and wind direction provide important guidance.

4.2 Distance matters

Placing the walls farther downwind (e.g., from 600–1200 m to 2400–3600 m) reduced bias because the plume had more time to mix vertically. However, longer distances also mean longer flights and a greater risk of changing weather. A balance is therefore needed between enhanced mixing and practical limitations. At even greater distances, an apparent increase in bias was observed, which remains unexplained and highlights the uncertainty of applying the method too far downwind. In addition, placing the wall too far downwind requires a much larger wall width, which consumes considerably more flight time while providing only limited improvement in the estimates. The same flight time could be used more effectively to perform additional measurements, thereby enlarging the ensemble and reducing the standard deviation, which is equivalent to lowering the measurement uncertainty. Thus, there may be an optimum distance from the source that balances the benefit of a shorter flight time with sufficient plume coverage and vertical mixing, and this should be investigated further.

4.3 Minimum height and ground measurements are important (at short distance)

Close to the source (600–1200 m), estimates improve when the lowest flight transect crosses the plume core (around 50–60 m in this study). Ground measurements help capture near-surface mass and reduce underestimation when the PBLH is shallow. However, simply flying as low as possible is not always optimal, the key is to align with the plume maximum. At greater distances, the MFH and ground measurements become less critical, as vertical mixing drives the plume core toward a more homogeneous distribution within the PBL. In this study, the method was applied consistently with the MTG-Poland (Förster et al. 2025), integrating mixing ratio to the ground. However, the demonstrated influence of the minimum flight height implies that alternative extrapolation approaches could contribute to better

emission estimate and should be investigated in future work.

4.4 Frequency and number of transects just need to be sufficient

For the tested plume and model, about 1.5 Hz sampling and 10 transects were enough. Higher frequencies mainly oversampled without improving results. More transects reduced variability only when flight time was kept the same. Very low frequencies or too few transects should be avoided.

4.5 Suggestions

4.5.1 Weather forecasting is most important

Accurate and reliable forecasts of wind direction, wind speed, PBLH, and cloud cover are essential for planning flights. Periods with rapid changes, such as sudden cloud formation or the arrival of weather fronts, should be avoided. When wind speed drop below 2 m/s, the mass-balance approach loses validity and emission estimate becomes unreliable (Varon et al. 2018). The increasing PBLH is not a problem, as long as meteorological parameters don't vary rapidly. In this study, the most favorable conditions occurred in the early morning, between 07:00 and 11:00 local time with clear sky and consistent wind speed and direction. The actual flights took place around 08:00 local time, well within this window.

4.5.2 Keep measurement windows short and consistent

Flight walls should be flown during periods with consistent wind conditions, when wind speed and direction remain relatively constant (e.g., in the morning in this study), in order to reduce errors from changing meteorological conditions. The flights themselves do not necessarily have to be short, but the winds should remain consistent throughout the entire measurement window.

4.5.3 Choosing a sweet spot for the wall distance

Wall distances should be chosen to allow sufficient plume development, but not so far that flights become excessively long (e.g., requiring very wide walls) or overly sensitive to changing meteorological conditions. A few kilometers downwind is often a good compromise. The well-mixed distance can be calculated following Conley et al. (2017), as it depends on wind speed and PBLH. In our case, with wind speeds varying

between 2 m s^{-1} and 6 m s^{-1} and a changing PBLH between 100 m and 1500 m, no single fixed distance can be defined. Moreover, the ideal well-mixed distance, assuming 5 m/s wind speed and 1000 m PBLH, would only be reached at 7500 km downwind, which would require very wide sampling walls and long flight times. Based on experimental practice, measurements at shorter distances are preferred, as they already provide sufficiently accurate results without requiring excessive flight time. This is consistent with the conclusions of this study, namely that selecting an optimal wall distance is a compromise between precision and flight time.

4.5.4 Use ground measurement and minimum height to improve short-distance flights

At short distances, ground measurements and well-chosen minimum heights can help capture the plume fully and reduce underestimation. It is important not to simply lower the minimum flight height, but to ensure measuring the plume core.

4.5.5 Ensure enough sampling (preferably ensembles)

Because variability is high, sampling should be sufficient but not excessive. Combining multiple flights into ensembles provides more robust results and reduces the effect of short-term outliers.

5 Outlook and Future Work

5.1 Limitations of This Study

While this study provides detailed insights into the performance of mass-balance flux estimation under a range of flight path configurations and meteorological conditions, several limitations must be acknowledged. First, the simulation framework relies on synthetic meteorology derived from high-resolution LES output. Although realistic in structure, it may not capture the full complexity and stochasticity of atmospheric turbulence in real-world environments.

Second, the emission source in this study was assumed to be constant over time, which simplifies both transport and emission estimation but may overlook important temporal variability in real emissions (e.g., diurnal cycles, operational cycles, or maintenance events).

Third, the placement and orientation of the flux integration walls were optimized in post-processing, and assumed to be perfectly perpendicular to the mean wind. This does not reflect operational limitations of aircraft navigation, or real-time wind variability. The measurement uncertainty was also neglected.

5.2 Broader Applications

The modeling framework developed in this study, while tailored to methane point source emission quantification, has broader applicability. It can be extended to other GHGs such as CO₂ or N₂O.

Furthermore, the same mass-balance principles can be applied in different environments. In urban contexts, the technique may help quantify distributed emissions from traffic or industrial sectors. However, added complexity such as building-induced turbulence and heterogeneous sources would require refined wall placement strategies and higher spatial resolution.

In contrast, in countryside or remote areas, the framework could support verification of landfill emissions, with relatively fewer limitations to wall setup.

Importantly, the simulation setup can also support mission planning for real aircraft campaigns. By running pre-mission synthetic tests, researchers can assess how different flight geometries and atmospheric conditions influence retrieval performance and optimize accordingly.

5.3 Final Thoughts

This study presents an investigation into the performance and sensitivity of airborne mass-balance emission estimation using synthetic flight simulations. The results highlight the technique’s strengths in capturing overall source strength under favorable conditions, while also revealing its vulnerabilities, especially to meteorological variability and the importance of plume sampling geometry.

Despite its limitations, the method remains an essential method in atmospheric scientist’s toolbox, particularly when supported by high-resolution weather forecast and flexible flight planning. By identifying key factors that govern retrieval accuracy, such as minimum flight altitude, meteorological condition, time of the day, and sampling frequency, this work contributes to a more rigorous foundation for future campaign design and data interpretation.

Ultimately, airborne mass-balance methods will continue to play a vital role in validating emission inventories, detecting unreported sources, and supporting mitigation policies. However, their full precision can only be reached when the limitations are considered.

6 Appendix

6.1 WRF Configuration and Namelist Options

The following excerpts from the `namelist.input` provide the key settings used in this study. Only physics, dynamics, CPM, and tracer-related options are included here. The full namelist file is available upon request or in the supplementary material.

```
&domains
  max_dom      = 4,
  e_we         = 321, 319, 321, 321,
  e_sn         = 321, 319, 321, 321,
  e_vert       = 36, 57, 70, 82,
  dx           = 3750, 1250, 250, 50,
  dy           = 3750, 1250, 250, 50,
  parent_grid_ratio = 1, 3, 5, 5,
  parent_time_step_ratio = 1, 3, 5, 5,
  num_metgrid_levels = 92,
/

&physics
  mp_physics    = 6, 6, 6, 6,
  ra_lw_physics = 4, 4, 4, 4,
  ra_sw_physics = 4, 4, 4, 4,
  bl_pbl_physics = 1, 1, 0, 0,
  cu_physics    = 3, 0, 0, 0,
  sf_sfclay_physics = 1, 1, 1, 1,
  sf_surface_physics = 2, 2, 2, 2,
/

&dynamics
  rk_ord        = 3,
  diff_opt      = 1, 1, 2, 2,
  km_opt        = 4, 4, 5, 5,
  diff_6th_opt  = 2, 2, 2, 2,
  diff_6th_factor = 0.12, 0.12, 0.12, 0.12,
  damp_opt      = 3,
  zdamp         = 3000., 3000., 3000., 3000.,
  h_mom_adv_order = 5, 5, 5, 5,
  v_mom_adv_order = 3, 3, 3, 3,
  h_sca_adv_order = 5, 5, 5, 5,
  v_sca_adv_order = 3, 3, 3, 3,
  scalar_adv_opt  = 2, 2, 2, 2,
  chem_adv_opt    = 2, 2, 2, 2,
  tke_adv_opt     = 2, 2, 2, 2,
/

&dynamics ! --- CPM settings for LES
  cpm_opt        = 0, 0, 2, 2,
  cpm_meso_pblh  = 0, 1, 1, 1,
  cpm_ngc_h      = 0, 0, 8, 8,
  cpm_ncells_h   = 0, 0, 3, 3,
  cpm_ngc_v      = 0, 0, 1, 1,
  cpm_lambda     = 0.0, 0.0, 0.875, 0.875,
  cpm_pec        = 0.0, 0.0, 0.4, 0.4,
/

&chem
  chem_opt       = 16, 16, 16, 16,
  emiss_opt      = 16, 16, 16, 16,
  tracer_opt     = 1,
  vertmix_onoff  = 1, 1, 1, 1,
  have_bcs_tracer = .true., .true., .true., .true.,
/
```

6.2 Table of Experiments

Table 2: Experiment setup and result

Group	A01	A02	A03	A04	A05	B01	B02	B03	B04	B05
Modified Parameter	validation	validation	validation	validation	validation	basecase	basecase+full day	basecase+noG	basecase+far	basecase+far+noG
Distance 1/m	1150	600	600	3600	1150	600	600	600	2400	2400
Distance 2/m	1200	1200	1200	4000	1200	1200	1200	1200	3600	3600
nTS	60	60	60	60	60	11	11	10	11	10
MFH/m	1	1	1	1	1	100	100	100	100	100
Frequency/Hz	4.5	4.5	4.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Speedup	x10000	x10000	x10000	x10000	x10000					
Start	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00
End	20:00:00	20:00:00	12:00:00	20:00:00	20:00:00	12:00:00	20:00:00	12:00:00	12:00:00	12:00:00
Ground	noG	noG	noG	noG	noG	yesG	yesG	noG	yesG	noG
Wall Width/m	10000	10000	10000	10000	10000	2500	2500	2500	5000	5000
Mean1/kg/s	34.96	34.96	33.21	33.26	34.97	24.61	20.72	21.88	30.02	31.15
Mean2/kg/s	35.14	35.14	32.90	33.26	35.14	27.46	24.79	25.24	30.73	32.24
Median1/kg/s	31.95	31.95	32.22	30.53	31.98	22.53	19.55	21.10	27.53	29.62
Median2/kg/s	32.11	32.11	32.13	30.36	32.03	26.42	22.37	24.19	29.52	30.86
Std. dev 1/kg/s	24.96	24.96	5.91	26.89	24.92	10.09	13.20	12.28	13.53	14.32
Std. dev2/kg/s	26.20	26.20	8.41	27.85	26.13	11.41	14.37	11.58	14.46	14.19
Std. err1/kg/s	1.39	1.39	0.47	1.50	1.39	0.80	0.77	0.98	1.11	1.17
Std. err2/kg/s	1.46	1.46	0.66	1.55	1.46	0.91	0.84	0.92	1.18	1.16

Table 3: Experiment setup and result

Group	C01	C02	C03	C04	D01	D02	D03	D04	D05
Modified Parameter	Frequency	Frequency	Frequency	Frequency+far	nTS	nTS	nTS + Speed	nTS + Speed	nTS + Speed
Distance 1/m	600	600	600	2400	600	600	600	600	600
Distance 2/m	1200	1200	1200	3600	1200	1200	1200	1200	1200
nTS	11	11	11	11	21	31	21	31	61
MFH/m	100	100	100	100	100	100	100	100	100
Frequency/Hz	0.3	0.5	4.5	0.5	1.5	1.5	1.5	1.5	1.5
Speedup							x2	x3	x6
Start	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00
End	12:00:00	12:00:00	12:00:00	12:00:00	12:00:00	12:00:00	12:00:00	12:00:00	12:00:00
Ground	yesG	yesG	yesG	yesG	yesG	yesG	yesG	yesG	yesG
Wall Width/m	2500	2500	2500	5000	2500	2500	2500	2500	2500
Mean1/kg/s	24.75	24.12	24.57	29.83	23.19	20.61	23.69	23.47	23.22
Mean2/kg/s	28.03	27.33	27.36	30.42	24.73	22.80	26.63	26.00	25.48
Median1/kg/s	22.53	22.22	22.63	26.86	24.55	22.48	23.49	23.62	23.18
Median2/kg/s	26.56	25.40	26.17	29.65	25.08	25.24	26.50	25.95	25.79
Std. dev 1/kg/s	10.46	10.33	10.14	13.96	11.08	11.78	9.46	8.71	8.18
Std. dev2/kg/s	11.94	11.67	11.38	14.04	11.75	13.31	11.44	11.18	10.85
Std. err1/kg/s	0.83	0.82	0.81	1.14	0.88	0.94	0.75	0.69	0.64
Std. err2/kg/s	0.95	0.93	0.91	1.15	0.93	1.06	0.90	0.88	0.86

Table 4: Experiment setup and result

Group	D06	D07	D08	E01	E02	E03	E04	E05	E06	E07	E08
Modified Parameter	nTS	+ nTS	+ nTS	+ Min Height	Min Height	Min Height	Min Height	Min Height	Min Height + far	Min Height + noG	Min Height + noG
	Speed+1m MFH	Speed+1m MFH	Speed+1m MFH								
Distance 1/m	600	2400	2400	600	600	600	600	600	2400	600	600
Distance 2/m	1200	3600	3600	1200	1200	1200	1200	1200	3600	1200	1200
nTS	61	61	61	11	11	11	11	11	11	10	10
MFH/m	1	1	1	1	25	50	75	150	50	25	75
Frequency/Hz	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Speedup	x6	x6	x6								
Start	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00
End	12:00:00	12:00:00	12:00:00	12:00:00	12:00:00	12:00:00	12:00:00	12:00:00	12:00:00	12:00:00	12:00:00
Ground	noG	noG	noG	yesG	yesG	yesG	yesG	yesG	yesG	noG	noG
Wall Width/m	2500	5000	5000	2500	2500	2500	2500	2500	5000	2500	2500
Mean1/kg/s	30.98	27.25	27.25	29.73	30.62	32.97	29.01	22.31	32.09	31.06	29.82
Mean2/kg/s	30.05	25.58	25.58	30.48	32.16	32.22	30.05	25.54	32.18	31.70	30.06
Median1/kg/s	30.61	26.47	26.47	28.92	29.60	33.44	28.40	20.96	29.10	29.75	29.97
Median2/kg/s	29.15	25.92	25.92	29.18	31.25	31.55	29.29	24.11	29.99	29.57	28.98
Std. dev 1/kg/s	10.60	13.13	13.13	10.57	10.28	11.03	10.04	11.33	15.35	11.19	10.88
Std. dev2/kg/s	11.91	12.55	12.55	11.39	11.69	12.04	11.48	13.03	15.52	12.36	12.09
Std. err1/kg/s	0.84	1.04	1.04	0.84	0.82	0.88	0.80	0.90	1.26	0.89	0.87
Std. err2/kg/s	0.94	0.99	0.99	0.91	0.93	0.96	0.91	1.04	1.27	0.98	0.96

Table 5: Experiment setup and result

Group	F01	F02	F03	F04	F05	B06	B07	B08	B09
Modified Parameter	Instant+1m MFH	Instant+100m MFH	Instant+100m MFH+far	Instant+100m MFH+far	Instant+100m MFH+far	basecase+far	basecase+cut	basecase+cut+far	basecase+cut+far
Distance 1/m	600	600	2400	4800	4800	4800	600	2400	4800
Distance 2/m	1200	1200	3600	6000	6000	6000	1200	3600	6000
nTS	11	11	11	11	11	11	11	11	11
MFH/m	1	100	100	100	100	100	100	100	100
Frequency/Hz	4.5	4.5	4.5	4.5	1.5	1.5	1.5	1.5	1.5
Speedup	x10000	x10000	x10000	x10000	x10000				
Start	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00	04:00:00
End	12:00:00	12:00:00	12:00:00	12:00:00	12:00:00	12:00:00	12:00:00	12:00:00	12:00:00
Ground	noG	yesG	yesG	yesG	yesG	yesG	yesG	yesG	yesG
Wall Width/m	10000	10000	10000	10000	10000	7500	2500	5000	7500
Mean1/kg/s	33.09	25.74	31.14	33.49	32.89	29.62	25.27	30.08	30.08
Mean2/kg/s	32.89	28.39	32.31	34.38	33.02	29.29	27.73	31.05	29.46
Median1/kg/s	31.89	24.34	29.89	31.82	31.12	29.06	23.12	27.12	29.65
Median2/kg/s	32.01	27.70	31.75	32.35	31.12	28.12	26.65	30.03	28.77
Std. dev 1/kg/s	6.05	8.39	10.47	10.79	11.32	12.62	9.52	13.62	12.71
Std. dev2/kg/s	8.35	10.22	9.50	12.26	12.08	10.66	10.79	14.58	10.80
Std. err1/kg/s	0.48	0.66	0.82	1.03	0.89	1.06	0.79	1.17	1.12
Std. err2/kg/s	0.66	0.81	0.75	1.17	0.95	1.07	0.89	1.25	0.95

References

- Beck, Veronika, Koch, Thomas, and Kretschmer, Roberto (2011). *The WRF Greenhouse Gas Model (WRF-GHG)*. TECHNICAL REPORTS 25.
- Brunner, Dominik et al. (Feb. 27, 2023). “Evaluation of simulated CO₂ power plant plumes from six high-resolution atmospheric transport models”. In: *Atmospheric Chemistry and Physics* 23.4. Publisher: Copernicus GmbH, pp. 2699–2728. ISSN: 1680-7316. DOI: 10.5194/acp-23-2699-2023. URL: <https://acp.copernicus.org/articles/23/2699/2023/>.
- Ciais, Philippe et al. (2017). *An operational anthropogenic CO₂ emissions monitoring & verification system: baseline requirements, model components and functional architecture*. Publications Office of the European Union. ISBN: 978-92-79-72101-4. URL: <https://data.europa.eu/doi/10.2760/08644>.
- Conley, Stephen et al. (Sept. 13, 2017). “Application of Gauss’s theorem to quantify localized surface emissions from airborne measurements of wind and trace gases”. In: *Atmospheric Measurement Techniques* 10.9. Publisher: Copernicus GmbH, pp. 3345–3358. ISSN: 1867-1381. DOI: 10.5194/amt-10-3345-2017. URL: <https://amt.copernicus.org/articles/10/3345/2017/>.
- Crosson, E.R. (Sept. 1, 2008). “A cavity ring-down analyzer for measuring atmospheric levels of methane, carbon dioxide, and water vapor”. In: *Applied Physics B* 92.3, pp. 403–408. ISSN: 1432-0649. DOI: 10.1007/s00340-008-3135-y. URL: <https://doi.org/10.1007/s00340-008-3135-y>.
- Fiehn, Alina et al. (Nov. 3, 2020). “Estimating CH₄, CO₂ and CO emissions from coal mining and industrial activities in the Upper Silesian Coal Basin using an aircraft-based mass balance approach”. In: *Atmospheric Chemistry and Physics* 20.21. Publisher: Copernicus GmbH, pp. 12675–12695. ISSN: 1680-7316. DOI: 10.5194/acp-20-12675-2020. URL: <https://acp.copernicus.org/articles/20/12675/2020/>.
- Förster, Eric et al. (Mar. 12, 2025). “A helicopter-based mass balance approach for quantifying methane emissions from industrial activities, applied for coal mine ventilation shafts in Poland”. In: *EGUsphere*. Publisher: Copernicus GmbH, pp. 1–26. DOI: 10.5194/egusphere-2025-1010. URL: <https://egusphere.copernicus.org/preprints/2025/egusphere-2025-1010/>.
- Gałkowski, Michał et al. (Nov. 12, 2024). *Impact of atmospheric turbulence on the accuracy of point source emission estimates using satellite imagery*. DOI: 10.5194/

- egusphere-2024-2792. URL: <https://egusphere.copernicus.org/preprints/2024/egusphere-2024-2792/>.
- Grell, Georg A. et al. (Dec. 1, 2005). “Fully coupled “online” chemistry within the WRF model”. In: *Atmospheric Environment* 39.37, pp. 6957–6975. ISSN: 1352-2310. DOI: 10.1016/j.atmosenv.2005.04.027. URL: <https://www.sciencedirect.com/science/article/pii/S1352231005003560>.
- Kilroy, Gerard et al. (June 1, 2024). “Evaluation of turbulence characteristics in WRF simulations at WiValdi wind park”. In: *Journal of Physics: Conference Series* 2767.5, p. 052063. ISSN: 1742-6588, 1742-6596. DOI: 10.1088/1742-6596/2767/5/052063. URL: <https://iopscience.iop.org/article/10.1088/1742-6596/2767/5/052063>.
- Knapp, M et al. (Apr. 2023). “Spectrometric imaging of sub-hourly methane emission dynamics from coal mine ventilation”. In: *Environmental Research Letters* 18.4. Publisher: IOP Publishing, p. 044030. ISSN: 1748-9326. DOI: 10.1088/1748-9326/acc346. URL: <https://dx.doi.org/10.1088/1748-9326/acc346>.
- Mazzaro, L. J. et al. (2019). “Random Force Perturbations: A New Extension of the Cell Perturbation Method for Turbulence Generation in Multiscale Atmospheric Boundary Layer Simulations”. In: *Journal of Advances in Modeling Earth Systems* 11.7. _eprint: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019MS001608>, pp. 2311–2329. ISSN: 1942-2466. DOI: 10.1029/2019MS001608. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001608>.
- Meijer, Yasjka et al. (2019). *An operational anthropogenic CO₂ emissions monitoring & verification support capacity: needs and high level requirements for in situ measurements : report from the CO₂ monitoring task force*. Publications Office of the European Union. ISBN: 978-92-76-09004-5. URL: <https://data.europa.eu/doi/10.2760/182790>.
- Muñoz-Esparza, Domingo and Branko Kosović (June 2018). “Generation of Inflow Turbulence in Large-Eddy Simulations of Nonneutral Atmospheric Boundary Layers with the Cell Perturbation Method”. In: *Monthly Weather Review* 146.6, pp. 1889–1909. ISSN: 0027-0644, 1520-0493. DOI: 10.1175/MWR-D-18-0077.1. URL: <https://journals.ametsoc.org/doi/10.1175/MWR-D-18-0077.1>.
- Nieuwstadt, F. T. M. et al. (1993). “Large-Eddy Simulation of the Convective Boundary Layer: A Comparison of Four Computer Codes”. In: *Turbulent Shear Flows 8*. Ed. by Franz Durst et al. Berlin, Heidelberg: Springer, pp. 343–367. ISBN: 978-3-642-77674-8. DOI: 10.1007/978-3-642-77674-8_24.

- Pätzold, Falk et al. (Oct. 18, 2023). “HELiPOD—Revolution and evolution of a helicopter-borne measurement system for multidisciplinary research in demanding environments”. In: *Elementa: Science of the Anthropocene* 11.1, p. 00031. ISSN: 2325-1026. DOI: 10.1525/elementa.2023.00031. URL: <https://doi.org/10.1525/elementa.2023.00031>.
- Pühl, Magdalena et al. (Jan. 24, 2024). “Aircraft-based mass balance estimate of methane emissions from offshore gas facilities in the southern North Sea”. In: *Atmospheric Chemistry and Physics* 24.2, pp. 1005–1024. ISSN: 1680-7324. DOI: 10.5194/acp-24-1005-2024. URL: <https://acp.copernicus.org/articles/24/1005/2024/>.
- Reum, Friedemann et al. (Jan. 15, 2025). “Airborne observations reveal the fate of the methane from the Nord Stream pipelines”. In: *Nature Communications* 16.1. Publisher: Nature Publishing Group, p. 351. ISSN: 2041-1723. DOI: 10.1038/s41467-024-53780-7. URL: <https://www.nature.com/articles/s41467-024-53780-7>.
- Skamarock, William C. et al. (Mar. 29, 2019). “A Description of the Advanced Research WRF Model Version 4”. In: UCAR/NCAR. DOI: 10.5065/1DFH-6P97. URL: <https://opensky.ucar.edu/islandora/object/opensky:2898>.
- UNFCCC (2015). *Paris Agreement*. United Nations Framework Convention on Climate Change. URL: https://unfccc.int/sites/default/files/english_paris_agreement.pdf.
- Varon, Daniel J. et al. (Oct. 18, 2018). “Quantifying methane point sources from fine-scale satellite observations of atmospheric methane plumes”. In: *Atmospheric Measurement Techniques* 11.10. Publisher: Copernicus GmbH, pp. 5673–5686. ISSN: 1867-1381. DOI: 10.5194/amt-11-5673-2018. URL: <https://amt.copernicus.org/articles/11/5673/2018/>.
- Wolff, Sebastian et al. (Apr. 8, 2021). “Determination of the emission rates of CO₂ point sources with airborne lidar”. In: *Atmospheric Measurement Techniques* 14.4. Publisher: Copernicus GmbH, pp. 2717–2736. ISSN: 1867-1381. DOI: 10.5194/amt-14-2717-2021. URL: <https://amt.copernicus.org/articles/14/2717/2021/>.

Acknowledgments

I would like to express my sincere gratitude to Prof. Dr. Markus Rapp, Dr. Anke Roiger, Prof. Dr. Julia Marshall, and Dr. Friedemann Reum for giving me the opportunity to work on this project. It has been a pleasure to work with you, and I have learned extensively in a short time thanks to your outstanding expertise.

I am also grateful to Eric Förster, Heidi Huntrieser, Magdalena Pühl, Klaus-Dirk Gottschaldt, and all other colleagues for their support regarding the MTG Poland campaign and the HELiPOD measurements.

I would like to thank Sebastian Wolff and Gerard Kilroy for their patient explanations of the Cell-Perturbation method and WRF.

I would like to sincerely thank my family and relatives for their constant encouragement and support during this journey. I also thank my friends from the university for their encouragement and for cheering me up during the course of this thesis.

Finally, I acknowledge the Levante team at DKRZ for providing computing resources and technical support.

Declaration:

I hereby declare that this thesis is my own work, and that I have not used any sources and aids other than those stated in the thesis.

München, 15 September 2025

ZHOUTONG MA