THESIS

GAUSSIAN PLUME AND BACKWARD LAGRANGIAN STOCHASTIC MODELING OF METHANE EMISSIONS: AN EXPERIMENT AT METEC

Submitted by

Aashish Upreti

Department of Mechanical Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2026

Master's Committee:

Advisor: Daniel B. Olsen
Co-Advisor: Daniel J. Zimmerle

Kira Shonkwiler
Ellison Carter

ABSTRACT

GAUSSIAN PLUME AND BACKWARD LAGRANGIAN STOCHASTIC MODELING OF METHANE EMISSIONS: AN EXPERIMENT AT METEC

Methane (CH_4) is a potent greenhouse gas, and emissions from the oil and gas (O&G) sector are a crucial contributor to it. Accurate quantification of CH_4 emission from the O&G sector is important. This study investigates the performance of Gaussian plume (GP) and backward Lagrangian stochastic (bLS) dispersion models through controlled release experiments conducted at the METEC facility at Colorado State University for release rates of 0.5 to 6 kg h^{-1}. Experiments were performed in near-field conditions (<100 m) with a 5-minute sampling period across varying release heights and atmospheric conditions. Results reveal that performance from the bLS model was better than GP, as indicated by the factor of 2 (FAC2;35% for GP and 52% for bLS), geometric mean bias (MG; 0.08 for GP and 0.63 for bLS), and mean factor of error (MFoE; 3.5 for GP and 1.9 for bLS). Models' performance slightly improved when the release height was closer to the sampling height and in moderately unstable atmospheric conditions. This study suggests that the bLS model is more suitable than GP for mock near-field O&G facilities like METEC, when experiments are conducted under a single controlled release.

ACKNOWLEDGEMENTS

I am grateful to Dr. Kira Shonkwiler for her mentorship and advice throughout this project. Her dedication to coaching and teaching me has been a valuable insight for my learning journey. I appreciate the constructive feedback and advice to improve some of my weaknesses during the project. My sincere gratitude towards advisor Dr. Daniel Olsen and co-advisor Daniel Zimmerle for allowing me to the part of the research I conducted. I am thankful to Dr. Ellison Carter for agreeing to be the committee member for this thesis and for managing time for reviewing, feedback, and defense.

I would like to thank Ryan Brouwer and Daniel Fleishmann for their support during my field measurements at the METEC site. Without them, I could not have engaged in safety and instrumentation effectively.

I am thankful to all the members of the Zimmerle Research Group for their insightful feedback on this project and for supporting me throughout the whole period.

DEDICATION

To my parents Samjhana Upreti and Deepak Upreti

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

kg	Kilogram
m	Meter
cm	Centimeter
h	Hour
Hz	Hertz
$\mu g\ m^{-3}$	Micrograms per cubic meter
ppm	Parts per million
$kg\ s^{-1}$	Kilogram per second
$kg\ h^{-1}$	Kilogram per hour
$m\ s^{-1}$	Meter per second

Chapter 1: INTRODUCTION

Methane (CH_4) is a major component of natural gas (NG) and a potent greenhouse gas (GHG) with a global warming potential (GWP_{100}) of approximately 27.9 relative to CO_2 on a 100-year horizon [1]. Anthropogenic activities like oil and gas (O&G) production, agriculture, biomass burning, and waste disposal account for 60 to 65% of the total CH_4 emission globally, while natural emitters like termites, freshwater ecosystems, and wetlands contribute to the remaining fraction [2]. Atmospheric CH_4 concentration has increased by approximately 11 ppb yr^{-1} as reported in 2023 [3,4]. This rise has played a significant role in driving shifts in the global climate over the past decade. Therefore, mitigating CH_4 emissions from O&G production sites is crucial for reducing overall greenhouse gas emissions. Achieving effective mitigation, however, depends on accurate quantification of facility-level emissions, which form the basis for mitigation planning and regulatory compliance. Yet, obtaining reliable emission estimation remains a challenge under varying atmospheric conditions in near-field O&G settings [5,6,7]. O&G operations are recognized as a major anthropogenic source, yet emission estimates at the facility level remain highly uncertain [8,9]. Large-scale approaches like aerial surveys, satellite observations, and regional inversions are essential in constraining basin-wide CH_4 fluxes, but their limited temporal coverage makes it difficult to capture short-duration or intermittent emissions from individual facilities [10]. To address this gap, near-field techniques that integrate dispersion modeling and direct concentration measurements are being increasingly used for facility-level emissions, but their accuracy is limited due to meteorological variability and violated modeling assumptions such as flat terrain, unobstructed wind fields, and steady state flow, which are rarely valid at complex O&G sites [8,11,12].

Among dispersion models, Gaussian plume (GP) and backward Lagrangian stochastic (bLS) are commonly applied in CH_4 emission estimations across various source types [13,14]. The GP method assumes the spread of the plume from a point source follows a Gaussian profile with a steady-state condition in which the emission occurs with constant meteorological conditions [15,16,17]. Similarly, bLS requires greater computational efforts and is sensitive to boundary conditions, sampling geometry, and configuration [12,13,18]. Although bLS in industrial settings has often focused primarily on area sources like livestock enclosures [12,13], it is also being increasingly applied to near-field point-source emission estimations [14]. Despite their limitations, GP and bLS models are widely used for rapid and near-field estimations at O&G facilities. GP-based approaches are widely used in regulatory and engineering tools for quantification in O&G facilities [14,16], while bLS is more embedded in research communities and specialized monitoring programs like area and point source, such as agricultural and industrial operations, where complex source-sensor configurations are important [7,12,13]. Hence, this study directly compares and evaluates GP and bLS models for near-field single source release conditions at a proxy O&G production facility. However, the comparative performance of these models for single-source controlled release under realistic meteorological variability and plume-sensor geometry in near-field conditions remains poorly constrained, especially for short averaging times.

Determining the conditions under which GP and bLS yield reliable estimates is essential for optimizing the CH_4 monitoring approach at O&G sites [11,14,16,17]. Service providers and facility operators use dispersion models to select monitoring technologies, interpret short-term measurement campaigns, and design effective sampling strategies for leak detection and quantification [13]. Regulators also depend on model performance when establishing acceptable methods for controlled release tests and compliance assessments [15]. By comparing GP and bLS

models under realistic near-field conditions at METEC, this study may support the decisions about model choice, controlled release test design, and the geometric and meteorological conditions that allow reliable quantification.

This study examines single controlled point-source releases in the near-field (<100 m) in relatively level terrain. The analysis evaluates the reliability and performance of GP and bLS models in quantifying emission estimates under controlled release conditions and highlights the role of plume geometry, meteorological variability, and simulation parameters on models' performance. The study provides an understanding of the strengths and limitations of the GP and bLS model in near-field applications at a proxy oil and gas production facility. The bLS model is expected to infer better emission estimations under unstable and turbulent atmospheric conditions due to its stochastic particle-based formulation [12,13,18], more accurately representing the plume dispersion in varying wind conditions, whereas the GP model is more reliable under steady and uniform wind conditions [15,16,17].

This study is guided by three research questions: (1) How accurately can the GP and bLS models predict the true CH_4 emission rates? How is the performance for predicting emissions influenced by meteorological variability (atmospheric stability, wind meander) and plume geometry factors (height match score, plume centeredness, and centerline distance)? (3) What is the total particle count used in bLS simulations? To address these questions, emission estimates are inferred from GP and bLS models and compared against known controlled-release values. It additionally examines the model performance across meteorological variability and plume geometry factors to identify conditions that affect model performance. Also, the total particle counts required to infer estimates from bLS model across various meteorological variability and plume geometry factors are determined.

This study does not intend to optimize the model algorithm to introduce new formulations for model implication nor to encompass basin or regional-scale atmospheric modeling. Thus, the results are most relevant to near-field facility monitoring applications. By outlining the specific assumptions and methodological boundaries and comparing the performance of GP and bLS, this study highlights the practical limitations of the models for CH_4 emission estimates in O&G facilities.

Findings from this study may contribute to further research and CH_4 monitoring operations in the O&G field by understanding the practical constraints of GP and bLS models across varying atmospheric near-field conditions. The outcomes of this study may be helpful for regulators in improving the controlled release testing protocols and in identifying the limits of the quantification approach. The role of varying near-field meteorological factors, plume geometry, and particle requirements on the performance of the respective GP and bLS models will be the major takeaway from this work.

The subsequent chapters present methods with experimental design and modeling approaches, and results from this study in detail. Chapter 2 includes methodology that encompasses the experimental setup and the models' description. Chapter 3 contains results and evaluations across wind conditions, true release rate, signal strength, plume geometry (centeredness, height match, plume centerline distance) for both GP and bLS models. Similarly, Chapter 4 discusses the conclusion with recommendations and major takeaways for future research and scope.

Chapter 2: METHODS

2.1 Site Description and Experimental Design

The experiment was performed at the Methane Emission Technology Evaluation Center (METEC) test facility, operated by Colorado State University's Energy Institute and located in Fort Collins, Colorado (69 miles north of Denver). METEC is a prototypical or experimental site that simulates an O&G production facility by simulating emission points from above-ground point sources such as tanks, separators, and well-heads (Figure 2.1). NG is stored in large cylinders across the site, and the emission rate is controlled using orifices, pressure regulators, and solenoid valves.

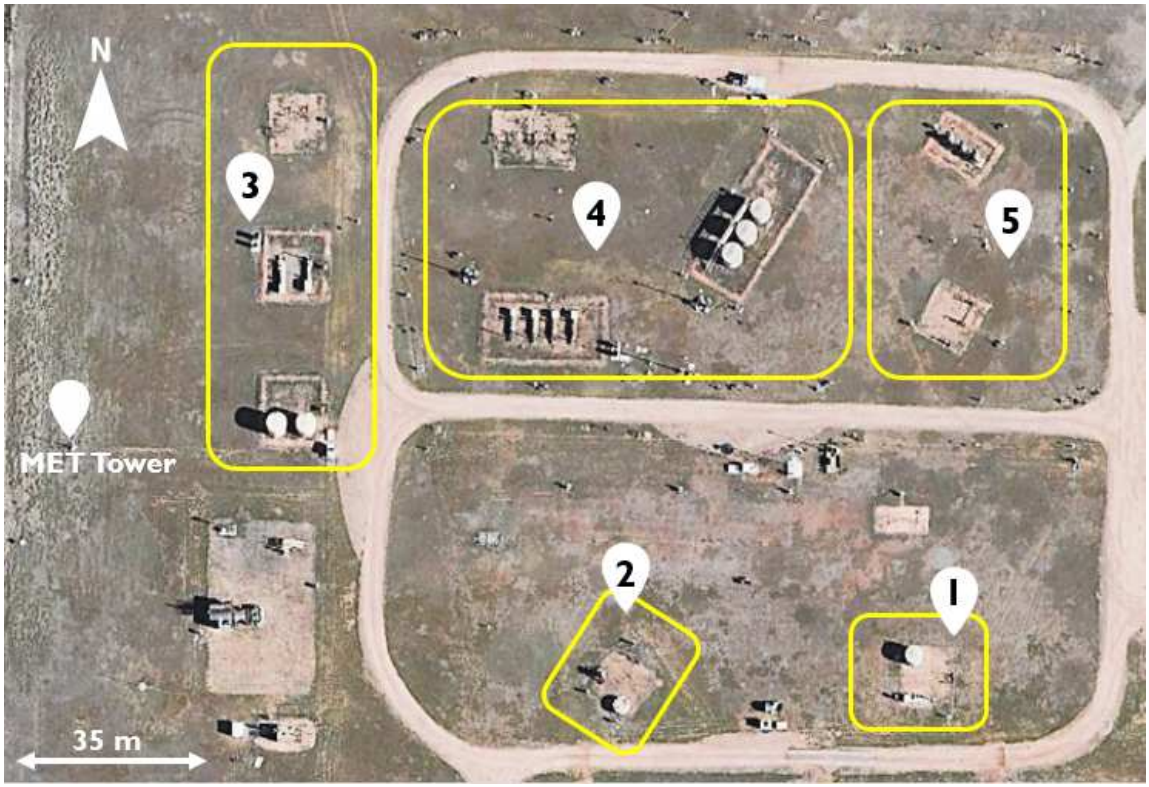


Figure 2.1. Aerial view of METEC's experimental layout in Ft. Collins, Colorado. Each labeled "pad" consists of varying types of decommissioned oil and gas equipment fitted with controlled release points at dozens of places (per equipment type) in locations where real-world leaks are known to occur. In this study, releases occurred from pads 1–

5. Equipment types for releases in these experiments were wellheads (pad 4), separators (pads 1–5), and tanks (pad 4) at varying rates and release heights/locations. METEC has expanded its capabilities in the time since experiments were performed and since this photo was taken.

The experimental procedure entailed placement of a portable methane concentration sensor (see section 2.3 for details) downwind of a release point on one of the pads, noting the location with a GPS unit for later determination of plume centerline and crosswind distances. Note that release locations and rates changed from experiment to experiment. Upon placement, and after ensuring continued optimal operation and wind alignment, a 5-minute measurement would occur for that downwind position. After a "successful" measurement was finished, the sensor would be moved downwind again—usually at least 10 meters—and a new measurement would be attempted until around 6 measurements were taken per experiment (in a few cases, it was 5, and a few others had 7). In this case, "success" is defined as the sensor operator viewing its instantaneous output throughout the 5-minute sampling timeframe and making a subjective determination that the concentration signal was sufficiently elevated throughout the measurement to continue to the next position.

Table 2.1. Date and time of experiments. The summary under #Samples includes total measurements taken on a particular experiment, total measurements under a single release, and multiple releases. WS Criterion represents the total measurement in the corresponding experiment in which wind speed > 1m s^{-1}. (*Wind speed <1 m s^{-1}; note this only occurred for a single release in experiment 21.)

			#Samples			
Exp#	Date	Time Range (UTC)	Total	Single	Multiple	WS Criterion*
1	3/19/2024	16:40:59.98-17:17:00.02	6	6	0	6
2	3/21/2024	16:33:00.00-17:17:59.97	6	6	0	6
3	4/23/2024	19:09:59.93-19:47:59.99	5	5	0	5
4	6/11/2024	17:36:00.25-18:06:10.17	5	4	1	4
5	6/17/2024	21:59:59.76-22:37:00.46	5	2	3	2
6	6/18/2024	17:40:59.96-18:17:10.26	6	2	4	2
7	6/19/2024	18:55:59.45-19:28:10.25	6	4	2	4
8	6/19/2024	21:09:59.84-21:41:39.64	6	1	5	1
9	7/9/2024	17:46:00.16-18:40:09.57	6	6	0	6
10	7/10/2024	18:20:59.05-18:52:30.04	6	6	0	6
11	7/11/2024	20:50:59.96-21:26:00.27	6	5	1	5
12	7/17/2024	21:29:22.27-22:15:21.76	6	5	1	5

13	7/18/2024	17:49:22.15-18:23:22.06	6	6	0	6
14	7/18/2024	19:51:22.16-20:23:32.05	6	5	1	5
15	8/26/2024	17:18:46.45-17:58:06.85	7	7	0	7
16	9/16/2024	17:37:24.27-18:40:55.76	6	6	0	6
17	9/16/2024	19:43:55.16-20:25:55.06	6	6	0	6
18	9/16/2024	21:48:54.96-22:41:54.96	6	6	0	6
19	9/17/2024	18:38:54.66-19:20:55.36	6	6	0	6
20	9/17/2024	20:20:54.76-20:55:04.96	6	6	0	6
21	9/18/2024	19:18:54.76-23:01:54.65	5	4	1	3*
22	9/19/2024	18:05:54.76-18:38:05.36	6	6	0	6
23	9/19/2024	19:04:55.27-19:49:55.26	6	6	0	6
24	9/20/2024	19:19:54.76-19:54:55.26	6	6	0	6
		Total	141	122	19	121

A total of 121 measurements were used in the study that represent single point source releases conducted between March and September 2024 across 18 different days (Table 2.1). Measurements taken during the multiple point source release were excluded from the analysis because the study was only focused on single point-controlled releases for simplification. After the data were collected and merged/time-synched, 5-minute averages were determined for the respective 5-minute downwind measurements. Samples were started when the instantaneous concentration recording and wind aligned with being downwind of a release (not necessarily occurring at rounded time intervals or on easily divisible increments).

Datasets will be discussed in the following sections and include 1 Hz release data, 1 Hz meteorological data, and downwind concentration data.

2.2 Meteorological Measurements

At METEC, meteorological data were recorded at 1 Hz and matched in time to the start/end of the associated concentration measurements. The wind data were measured by an R.M. Young 81000 3D sonic anemometer (R.M. Young, Traverse City, MI, USA) permanently mounted at METEC, 6.7 meters above ground (Figure 2.2), and ranged between 46 to 147 meters from the release locations used for the experiments. The sonic anemometer provides wind speed (accuracy

of ± 1%), wind direction (accuracy of ± 2°), and sonic temperature, while a Met One 597 (Met One Instruments, Grant Pass, OR, USA) provided ambient temperature (accuracy of ±2 °C), RH (accuracy of ±2%), and barometric pressure in hPa (P; at 6.2 m). Solar radiation measurements were used to quantify Pasquill-Gifford stability class (PGSC) and were acquired from 5-minute data reported via a nearby weather station maintained by the Colorado Climate Center (weather station named FTC01), within 175 m of METEC with similar surface characteristics (i.e., roughness length, displacement height, topography). A LI-200R-SMV model pyranometer was used for the solar radiation measurement. Although 5-minute data from FTC01 varied in its offset from most of the 5-minute downwind measurements, the nearest 5-minute mean was matched based on rounding to the nearest 5-minute sample for solar radiation to be as reflective as possible of the corresponding measurement conditions.

Figure 2.2. Meteorological tower in the METEC site. The anemometer is kept at 6.7 m above ground level, and the temperature and barometric pressure probe at 6.2 m.

In the atmospheric surface layer, wind speed is zero at the ground because of surface friction and increases with height as turbulence transports momentum downwards. The logarithmic or power-law profiles represent this vertical structure, which reflects the effect of stability at the rate of wind speed increase. The turbulence rather than air density differences governs this behaviour. Stable condition weakens the mixing, and unstable conditions intensify the mixing.

The Pasquill-Gifford stability classification (PGSC) explains vertical motion and related turbulence characteristics of the atmosphere, categorized using the strength of solar radiation and wind speed [19,20]. Two different approaches to solar radiation classification were included for the determination of PGSC: the solar radiation delta-T (SRDT) method [19] and a simplified approach based on Turner's framework (Table 2.2) [20]. The simplified subjective approach is based on visual meteorological indicators like cloud cover and ceiling height and then paired with wind speed.

Table 2.2. Pasquill-Gifford stability class (PGSC) categorized by wind speed and sky conditions and summarized for both a simplified subjective approach to defining solar radiation (see Appendix Table A.1) and more specific SRDT standards (Appendix C1). Classes represent unstable: A (Extreme), B (Moderate), C (Slight); neither unstable or stable: D (Neutral); and stable: E (Slight), F (Moderate). In this study, filtered data coincided with PGSCs of A, B, and C only.

Wind Speed (m s^{-1})	Simplified (Solar Radiation W m^{-2})	SRDT (Solar Radiation W m^{-2})
<2	>700 → A 350–700 → B <350 → B	≥925 → A 925–675 → A 675–175 → B <175 → D
2-3	>700 → A 350–700 → B <350 → C	≥925 → A 925–675 → B 675–175 → C <175 → D

3-5	>700 → B 350–700 → C <350 → C	≥925 → B 925–675 → B 675–175 → C <175 → D
5-6	>700 → C 350–700 → D <350 → D	≥925 → C 925–675 → C 675–175 → D <175 → D
≥6	>700 → C 350–700 → D <350 → D	≥925 → C 925–675 → D 675–175 → D <175 → D

2.3 Methane Measurement

Downwind methane concentration data were obtained using an ABB Micro-portable Greenhouse Gas Analyzer (MGGA; ABB, Zurich, Switzerland) [21]. The analyzer was connected to 3.175 mm diameter tubing whose inlet was affixed at 1 m above the ground using a tripod (Figure 2.3). Sensor height of 1 m above the ground for concentration measurement was consistent with the near-field surface sampling guideline in the U.S. Environmental Protection Agency (EPA) Other Test Method (OTM)-33A [22]. The sample cell flow rate of the MGGA was 2.62 L min^{-1}. Similarly, the measurement accuracy was ≤ 1% of the reading. As for the resolution, MGGA achieves 0.9 ppb (at 1 second) and 0.3 ppb (at 3 seconds). The methane concentration in this study was measured in ppm, and a small variation between 1 s and 3 s intervals (≤ 0.0009 ppm) was considered negligible. Also, the sensor was placed within a 100 m distance from the source, which follows OTM-33A guidelines [22]. Before the concentration measurement from a single controlled release was made, the background concentration was measured for 15 minutes and averaged. The concentration enhancement (difference of sampling concentration during controlled release and background concentration) was later used in dispersion models to infer emission estimates.

Figure 2.3. Example of a field deployment of methane concentration sampling. The tubing inlet connected to the gas analyzer (MGGA) and mounted to the tripod at 1 m above the ground level.

There were 2 different recording intervals used for concentration measurements: the first covered experiment 2 from 0.201 seconds or 4.967 Hz to 0.152 seconds or 6.597 Hz, and the second covered the remaining experiments from 0.999 seconds or 1.001 Hz to 0.996 seconds or 1.004 Hz. The minor variation of sampling rate in both recording intervals was due to the built-in clock and data buffering behavior of MGGA, which operates independently from the controlled-release systems. For the concentration measurements, 5-minute averages consisted of anywhere from 282 to 302 records for the near-1 Hz data, and from 1492 to 1965 records for the higher-speed data.

2.4 Controlled Release System

The controlled release experiment was performed between March 19th and September 20th, 2024. At METEC, NG was stored in large cylinders across the site, and the emission rate was controlled using orifices, pressure regulators, and solenoid valves. The gas composition of the controlled release NG was determined using a gas chromatograph (7890 GC, Agilent Technologies, Santa Clara, CA, USA) with nominally 85 mol% CH_4. The NG release rate ranged

between 0.53 to 6.03 kg h^{-1} (0.45 to 5.13 kg CH_4 h^{-1}), based on the 85 mol% CH_4 composition. It is comparable to moderate and fugitive leaks in upstream O&G operations [23,24]. Field studies at natural gas production and transmission facilities have shown that the most equipment-level methane leaks are generally under 1 kg CH_4 h^{-1} [23], while component-level leaks in the storage and transmission sector typically range between 0.05 to 5 kg CH_4 h^{-1} [24]. The selected release rates of NG fall in the range that represent emissions comparable to moderate or fugitive leak events in upstream O&G operations. The releases were from separators, tanks, and wellheads from heights between 0.4 to 6.7 m above ground.

2.5 Modeling

Transport dispersion models help with understanding how pollutants are emitted and dispersed into the atmosphere from sources. In downwind measurement applications, like specified by the U.S. EPA OTM-33A method [22], the dispersion model is applied to infer emission rates from sources and concentrations measured at specific locations where the plume is dispersed in the prevailing wind direction, along the plume centerline distance [25,26]. The selection of the dispersion model for the downwind method is determined by the study context, depending on site conditions, spatial scale, and emission type–point source [25,26]. A model like the Gaussian plume (GP) is recognized under the OTM-33A framework developed for near-field conditions [22]. Similarly, the backward Lagrangian stochastic (bLS) model is also widely applied for emission quantification in similar conditions [14].

2.5.1 Gaussian Plume (GP) Method

The GP model is a widely employed analytical tool in atmospheric studies for estimating emissions of pollutants from continuous point sources. The model works based on the assumption of steady-state dispersion of the plume in the downwind direction [15,16,25,27]. The GP model

assumes the pollutant concentration follows a normal (Gaussian) distribution in the vertical (z) and crosswind/lateral (y) directions from the centerline of the plume (x). The horizontal (σ_y) and vertical dispersion coefficients (σ_z) were calculated using Tables A1 to A3 (see appendix). The general expression of concentration C at any given point x, y, z is given as:

$$C\ (x,y,z) = \frac{Q}{2\pi u U \sigma_z} e^{-\frac{y^2}{2\sigma_y^2}} \left[\left(e^{-\frac{(z-H)^2}{2\sigma_z^2}}\right) + \left(e^{-\frac{(z+H)^2}{2\sigma_z^2}}\right)\right] \qquad \text{(Equation 1)}$$

where:

C: CH_4 Concentration (μ g m^{-3})

Q: Emission rate from a point source (g s^{-1})

U: Mean wind speed (m s^{-1})

H: Height of emission (m)

σ_y: Horizontal dispersion coefficient (standard deviation) (m)

σ_z: Vertical dispersion coefficient (standard deviation) (m)

x, y, z: Plume centerline distance, crosswind/lateral distance, and vertical sampling height from release (m), respectively.

The second exponential term in equation 1 is defined as ground-reflection term. This component is also interpreted as a 'mirror-source' in physical term, which ensures the plume reflects upwards after it strikes the ground rather than penetrating inside.

The Gaussian plume inverse method was applied in this study, in which measured concentration was used to infer emission rates, solving equation 1 for Q [in SI.xlsx]. Assumptions made for the implementation of the inverse method are [28]:

a. The point source is emitting pollutants at a constant rate.
b. Flat terrain between the source and the sensor.
c. The sensor should be kept >100 m from the source and be able to detect the emissions.

This method always infers emission results even if the assumptions are violated [11,28]. This model is mathematically simple and efficient in terms of computation, which makes it easy to use. In contrast, the GP model is constrained by several limitations, as the model is fundamentally based on the steady-state assumption of emission with constant meteorological conditions over homogenous terrain, which does not apply to the real-world conditions at oil and gas sites (which include flow obstructions and are often located in nonhomogeneous surroundings) [26,29]. In this study, the model assumptions were violated due to the distance between the source and downwind sensor being <100 m, the presence of equipment structures that obstructed flow, and nonhomogeneous surface characteristics.

2.5.2 *Backward Lagrangian Stochastic (bLS) Model*

To contrast with the GP results, a bLS modeling is also used for emission estimation, wherein it tracks the trajectories of released particles in the atmosphere [12,29,30,31]. Here, the 'particle' is defined as a numerical/computational tracer representing the stochastic/random motion of air parcels that simulate the turbulent flow of the gas. The bLS model is based on stochastic processes that simulate the turbulent nature of the dispersion or "particle flow" [12]. It models the trajectories of particles solving stochastic differential equations, enabling the model to capture variability in atmospheric turbulence, wind speed, and direction [12,29,30]. However, bLS can be computationally demanding, in its particle count requirement, and has difficulty resolving low wind conditions [12,29,30].

The bLS model used in this study was WindTrax (version 2.0.9.7, developed by Thunder Beach Scientific, Nova Scotia, Canada) [31]. The data used to feed the model consists of wind speed, wind direction, atmospheric temperature, barometric pressure, PGSC, measured CH_4 concentration downwind, background concentration, roughness length, meteorological

measurement location/heights, source location/height, and measurement location/height [see Appendix B]. WindTrax accounts for the differences between the concentration inlet and the meteorological collection heights by applying a surface-layer flow model to compute height height-dependent wind field, ensuring consistency between concentration measurement and wind observations.

Meteorological and concentration data were used to infer estimates from GP and bLS models, and the results were compared against the known controlled-release values. This comparison has provided the foundation to evaluate models' performance on varying meteorological conditions and plume geometries. The next section presents the results and interpretation of the models' outcome and their performance.

For this study, the initial number of particles used for the first run of each 5-minute sample/measurement for all 121 samples was 500,000. Although the default particle count used in WindTrax is 50,000, due to the restricted plume volume collection (collected using a tube with an internal diameter of 3.175 mm) and short sampling period (5 minutes), a higher particle count was used (500,000) to achieve stable quantification. If a quantification error occurred, particle count was increased to 1M (where M denotes million), then 2M, and subsequently increased in steps of 2M (i.e., 4M, 6M, and so on) up to a maximum of 20M, unless the error was not resolved. Despite the adjustments of increasing particle count in the model, 18 samples did not resolve errors after 20M particles, resulting in 103 modeled samples among 121. The average particle count used for the simulation was 6.4±4.8M.

2.6 Performance Analysis

The results were examined using statistical metrics common in atmospheric dispersion studies to evaluate the model's performance [32,33]. The metrics used to assess model

performance for this study are: predictions within a factor of two (FAC2), geometric mean (MG) bias, and median factor of error (MFoE) [32,33].

The equations used to calculate these metrics are mentioned below.

If 'i' is the number of observations, Q_i is the estimated emission rate, and $\hat{Q}_i$ is the known emission rate, the ratio r_i is defined as:

$$r_i = \frac{Q_i}{\hat{Q}_i}, \qquad i = 1, 2, \ldots, N \qquad \text{Equation 2}$$

The ratio r_i (Equation 2) is the foundational variable for evaluating model performance. Model evaluation metrics were determined using the ratio applied in MG and MFoE (Equations 3 and 4), which are derived from the ratio r_i.

$$MG = exp\left(\frac{1}{N}\sum_{i=1}^{N} \ln r_i\right) \qquad \text{Equation 3}$$

$$MFoE = median\,[(\exp|(\ln r_i)|] \qquad \text{Equation 4}$$

FAC2 measures the proportion of model predictions that fall within a factor of two, also defined as the metric in which the ratio of emission estimated by GP to the known emission rate is between 0.5 and 2 [32,33]. MG explains the over- or underestimation of model prediction. If the MG value is 1, the model is unbiased, <1 indicates underestimation, and >1 indicates overestimation [32,33]. MFoE captures the multiplicative spread of error, which is always ≥1, and values close to 1 mean that the model has a small deviation from the prediction [32,33]. These three indicators, when taken together, provide an aspect of model behavior in varying conditions.

Some other metrics used for the analysis of the performance of the models are mean absolute error (MAE), root mean square error (RMSE) and mean absolute bias (MAB).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|Q_i - \hat{Q}_i| \quad \text{Equation 5}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Q_i - \hat{Q}_i)^2} \quad \text{Equation 6}$$

$$MAB = \frac{1}{N}\sum_{i=1}^{N}(Q_i - \hat{Q}_i) \quad \text{Equation 7}$$

The MAE measures the average magnitude of error between the estimated and true emission rates, regardless of direction. The RMSE predicts the square root of the average of the squared errors, which is highly sensitive to the larger errors. Unlike MAE and RMSE, MAB shows the bias direction of the error. The positive error means that the model is overestimating, and the negative means that it is underestimating.

Chapter 3 RESULTS

3.1 Meteorological Analysis

The 5-minute wind speeds during the measurements ranged between 0.9 and 7.9 m s^{-1}, with a median of 3.1 m s^{-1} and an average of 3.3±1.4 m s^{-1} (Appendix C: Table C1). The data was filtered for wind speeds <1 m s^{-1}, as low speeds are not as reliable for adequate transport conditions for well-defined plume analysis, often defined as calm wind conditions [29,33,35] (n=1; Table 2.1). Also, 5-minute wind directions observed during the sampling periods encompassed winds from nearly every direction (between 18 to 347°), ensuring good coverage around METEC's equipment.

The meteorological data presented here are representative of the number of samples shown in the far-right column (Appendix C: Table C1). The measurements were conducted at air temperatures between 3.7 and 33.4 °C, reflecting the seasonal variability during which the experiments were conducted. Also, relative humidity was between 10% to 59% across all measurements, reflective of the dry atmospheric conditions of Northern Colorado. Wind direction sectors and corresponding wind speeds during the measurement (Figure 3.1) show most experiments were performed under southeasterly flow, often at moderate wind conditions.

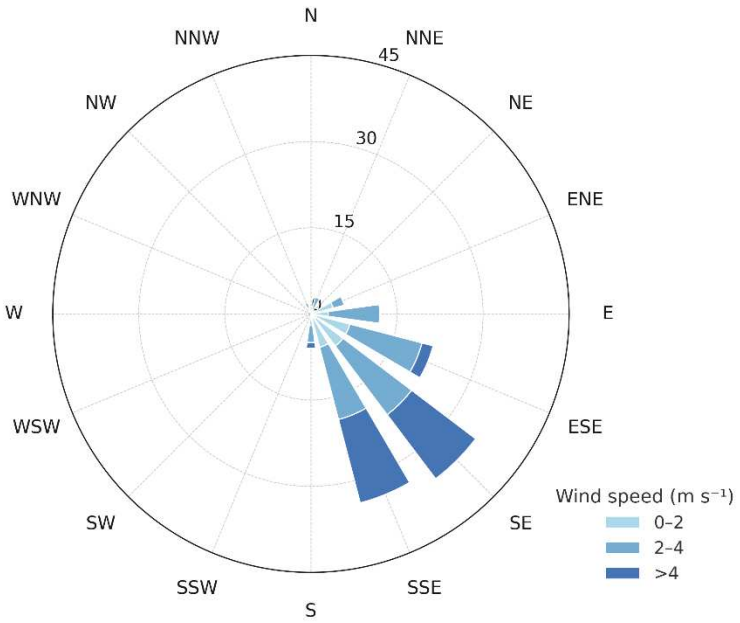


Figure 3.1. Wind rose diagram representing only the sampling periods. The rose is divided into 16 sectors with an interval of 22.5°, labeled by compass points (N, NNE, NE, and so on). Concentric circles are the sample counts in increments of 15.

Pasquill-Gifford divided stability into A, B, C, D, E, and F classifications, where each represents turbulent conditions that are extremely unstable, moderately unstable, slightly unstable, neutral, slightly stable, and moderately stable, respectively [18,26,27]. Analysis revealed that filtered measurements coincided with only three of these for this study: extremely unstable (A), moderately unstable (B), and slightly unstable (C) (see section 3.2.2).

3.2 Gaussian Plume Analysis

The general performance of the GP model across all measurements is summarized in Table 3.2. For the GP analysis under the simplified solar radiation method, the overall FAC2 value of 34.7% (FAC2 percentage out of 121 samples) highlights the limited predictive performance of the GP model across the dataset. The geometric mean bias of 0.17 indicates that the model has underpredicted the emissions by 83% across all samples. Similarly, the MFoE value of 3.4 reveals that the predictive error of the model was nearly three and a half times the true emissions.

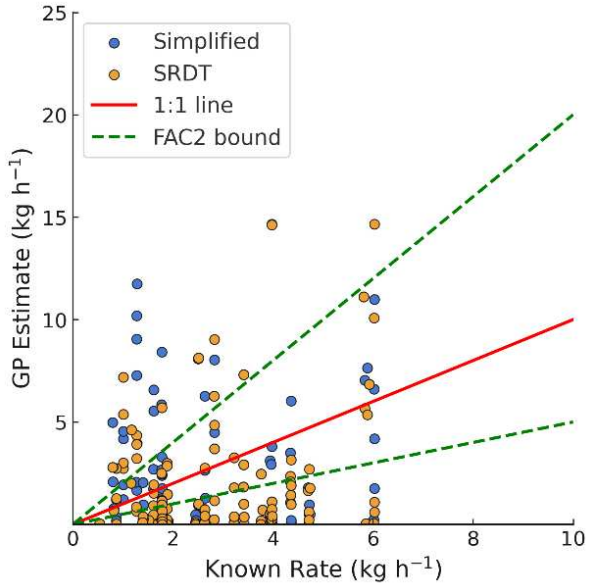
By comparison, results using the simplified classification method produced slightly better performance than SRDT, as the overall FAC2 of SRDT was 24.8%. The overall FAC2 is based on the ratio of total FAC2 hits to the number of samples. Similarly, MFoE from the SRDT approach (MFoE = 3.5) was almost the same as that of the simplified approach (MFoE = 3.4). However, the geometric mean bias decreased to 0.08, showing slightly stronger underestimation of emissions compared to the simplified approach (summary in Table 3.1).

In addition to FAC2, MG, and MFoE, the error metrics MAE, RMSE, and MAB provide a more detailed evaluation of the magnitude and directional bias of the model deviations. The MAE of the simplified approach of 2.5 kg h^{-1} indicates that the overall model predictions differ from the known emission rate by 2.5 kg h^{-1}, and the SRDT approach also showed a similar tendency with an MAE of 2.4 kg h^{-1}. The RMSE values of 3.3 kg h^{-1} and 3 kg h^{-1} for simplified and SRDT approaches, respectively, indicate that both approaches reflect occasional larger errors. The MAB value of -0.3 (simplified) and -0.8 (SRDT) confirmed systematic underestimation in both approaches. Together, these metrics complement FAC2, MG, and MFoE by evaluating not just the frequency of the model's accurate predictions but also the directional bias and average error magnitude, which helps comprehensive evaluation of performance.

Table 3.1: Overall performance metrics of the GP model using two stability classification techniques (by simplified and SRDT). The metrics include factors of two (FAC2), geometric mean bias (MG), median factor of error (MFOE), mean absolute error (MAE), root mean square error (RMSE), and mean absolute bias (MAB).

Metric	Derived from a simplified approach	Derived from SRDT approach
FAC2 (%)	34.7	24.8
MG	0.17	0.08
MFoE	3.4	3.5
MAE	2.5 kg h^{-1}	2.4 kg h^{-1}
RMSE	3.3 kg h^{-1}	3.0 kg h^{-1}
MAB	-0.3	-0.8

The estimated emission rate from GP is plotted against known emission rates (Figure 3.2) to show a direct comparison between true and estimated emission rates. The FAC2 bound is the benchmark for moderately acceptable predictions, and the 1:1 line is a benchmark for "perfect" predictability. The scatter around this line reflects the non-linear variability of model prediction because the coefficients of determination (R^2) for the simplified and SRDT approaches were 0.05 and 0.001, respectively. These low R^2 values reflect the w



eak linear relationship between estimated and known emissions and emphasize the necessity of assessing quantitative performance metrics like FAC2, MG, MFoE, MAE, RMSE, and MAB.

Figure 3.2: Scatterplot of emission estimated by GP versus the known emission rate from controlled experiments (n=121), using simplified stability classification (blue scatter points), and SRDT (orange scatter points).

Overall, the quantitative performance metrics indicate that the predictive accuracy of models varies under stability classification approaches. However, numerical performance metrics alone are insufficient to explain the variation in the model's predictive capacity and emphasize physical interpretation [32,33]. Therefore, to better understand the model's performance, the

following sections evaluate the effect of various other factors like release rates, signal strength, wind meander, and plume geometry (like height match score along with crosswind and plume centerline distance).

3.2.1 Signal Strength and Release Rate

Two groupings related to methane data were considered during the analysis: signal strength and release rate bins. Signal strength is defined as concentration scaled by measurement detection limits, which helps differentiate between reliable and weak detection. Signal strength is a crucial factor to analyze because weak signals can represent unstable wind conditions or diluted plumes, which can significantly impact emission estimation. In contrast, strong concentration signals generally indicate a greater likelihood of plume detection.

Signal strength bins were grouped into evenly populated quartiles: ≤2.3, 2.3–2.9, 2.9–4.0, and >4.0 ppm (Table 3.2).

For the simplified classification of stability, the range of FAC2 hit percentage across quartiles was 22.6-53.4%, and under the SRDT classification, it was 19.4-30%, showing no trend of systematic improvement with stronger signals within the scope of this dataset (see Table 3.2). The results suggest that model performance is not very dependent on signal strength. Factors like localized turbulence at METEC, limited sampling length, and transient plume fluctuations may have masked potential improvements in model performance at higher signal levels.

Table 3.2: Quartiles classification for filtered data based on signal strength (in ppm) for the GP performance analysis acquired from both simplified and the SRDT method of stability classification. FAC2 hits are the number of samples within a factor of two, and FAC2 hit rate (%) is the percentage of FAC2 hits on specific bins.

Quartile Concentration (ppm)	n_{bin}	Simplified		SRDT	
		FAC2 hits (n)	FAC2 hit rate (%)	FAC2 hits (n)	FAC2 hit rate (%)
≤2.3	31	7	22.6	6	19.4
2.3-2.9	30	11	36.7	6	20
2.9-4.0	30	18	53.4	9	30

>4.0	30	8	26.7	9	30
Overall	121	42	34.7	30	24.8

Release rate bins were created from known emission rates to group sources into categories representing different emission magnitudes. Evaluating model estimates over these emission bins helps to determine whether the model performs consistently across varying source strengths or not.

Release rate bins (kg h^{-1}) are grouped as tertiles to capture low, medium, and high emission scenarios. Cutoff values were defined as low when the emission rate was ≤1.8 kg h^{-1}, medium between 1.8-3.4 kg h^{-1}, and high between 1.8-3.4 kg h^{-1} (Table 3.3). This classification is relative to controlled release rates rather than fixed thresholds of emission strength.

The low, medium, and high release rate bins had FAC2 hit rate (percentage of FAC2 hit under each bin) of 40% & 27.5%, 27.9% & 20.9, and 36.8% & 26.3% for simplified and SRDT approach, respectively. Although the simplified method yielded slightly better FAC2 agreement, the trend does not show any improvement with increasing release rate in both approaches.

The lack of a clear trend across release rate bins suggests that the GP model was not highly influenced by emission magnitude within the tested range. It was expected that with higher release rates, the CH_4 plume carries more mass, allowing the sensor to more easily detect elevated concentrations, resulting in better performance [20,28]. However, other factors such as short sampling duration, atmospheric stability, and wind variability likely influence the dispersion rather than the release strength only.

Table 3.3: Tertiles classification based on low, medium, and high release rates (kg CH_4 h^{-1}) used in this study for the GP results acquired from both the simplified and the SRDT methods of stability classification. FAC2 hits are the number of samples within a factor of two, and FAC2 hit rate (%) is the percentage of FAC2 hits on specific bins.

Release rate bin tertiles (kg h^{-1})	n_{bin}	Simplified		SRDT	
		FAC2 hits (n)	FAC2 hit rate (%)	FAC2 hits (n)	FAC2 hit rate (%)
Low: ≤1.8	40	16	40	11	27.5
Medium: 1.8-3.4	43	12	27.9	9	20.9
High: 3.4-6.0	38	14	36.8	10	26.3

Overall	121	42	34.7	30	24.8

3.2.2 Atmospheric Stability

Atmospheric stability strongly influences the accuracy of the GP model, as it is incorporated into the model's formulation (equation 1) [18,25,29]. Stability was categorized using PGSC, derived from solar radiation and wind speed characterization (Table 2.2). All measurements at METEC fell under unstable daylight conditions, specifically classes A (extremely unstable), B (moderately unstable), and C (slightly unstable).

Under the simplified approach, the FAC2 hit rate for class A was the highest (41.7%), followed by class B (32.8%) and class C (20%). Like simplified, the SRDT approach also had the highest FAC2 hit rate (33.3%) in class A, and classes B and C showed almost equal FAC2 (B with 23% and C with 23.1%) (Table 3.4).

Additionally, the number of samples varied considerably depending on whether the simplified or SRDT method was implemented. Using the simplified method, the number of samples was distributed as 48 in A, 58 in B, and 15 in C. But using SRDT's method, samples were distributed as 21 for A, 74 for B, and 26 for C. It highlights the difference in sample distribution based on the choice of what stability classification method was applied.

It was expected to achieve a higher FAC2 hit rate in class C compared to classes A and B, because the dispersion of the plume in class C has generally steadier horizontal and vertical mixing without excessive disruption due to turbulence [18,29]. But class A exhibited relatively higher FAC2 agreement, while classes B and C performed similarly. Both the approaches (simplified and SRDT) showed a decline of FAC2 from class A to B. It is because the result may be influenced by the considerable variation in sample distribution across stability classes in both approaches.

Table 3.4: Number of samples under various PGSC and corresponding FAC2. Sampled data had the PGSC of classes A, B, and C, which were characterized using the simplified method and the SRDT method. FAC2 hits are the number of samples within a factor of two, and FAC2 hit rate (%) is the percentage of FAC2 hits on specific bins.

PGSC	Simplified			SRDT		
	n_{bin}	FAC2 hits (n)	FAC2 hit rate (%)	n_{bin}	FAC2 hits (n)	FAC2 hit rate (%)
A	48	20	41.7	21	7	33.3
B	58	19	32.8	74	17	23
C	15	3	20	26	6	23.1
Overall	121	42	34.7	121	30	24.8

3.2.3 Wind Direction Meander

Wind direction variability (or meander) captures the shifting nature of wind fields [18,30]. In this study, we examined wind meander, defined as the standard deviation of wind direction during the 5-minute averaging interval, to quantify the potential influence of wind direction on the performance of GP. Wind meander score indicates whether dispersion is governed by steadier airflow or by turbulence (possibly from rapidly shifting weather or from nearby equipment obstructions, or a combination of both). Strong variation in wind direction can disrupt the steady state assumption of GP, thereby reducing the reliability of model estimates [18,33].

Wind meander was categorized into ±5° bins (±0°-5°, ±5°-10°, ±10°-15°, ±15°-20°, ±20°-25°, ±25°-30°) and > ±30°. This metric allows the evaluation of how increasing wind direction variability might influence GP performance. For each bin, the sample count and the number of FAC2 falling under those bins (defined as FAC2 hits) were calculated. The two classification techniques (simplified and SRDT) were compared (Table 3.5). Under both methods, none of the measurements agreed with FAC2 when the meander was ±0-5°. Similarly, there was no FAC2 agreement hit under meander ±5-10° using the SRDT method, while a 33.3% FAC2 hit rate was observed using the simplified method. However, the ±5-10° meander with the simplified method

consists of only 6 samples, which is just 5% of the total samples (n = 121). Thus, under near steady wind conditions (meander = ±0-10°), a limited number of samples could not adequately allow for assessing the model's performance.

Table 3.5: Wind meander classification based on its standard deviation for the GP performance analysis acquired from both simplified and the SRDT's method of stability classification. FAC2 hits are the number of samples within a factor of two, and FAC2 hit rate (%) is the percentage of FAC2 hits on specific bins.

Wind meander bins (°)		Simplified		SRDT	
	n_{bin}	FAC2 hits (n)	FAC2 hit rate (%)	FAC2 hits (n)	FAC2 hit rate (%)
±0°-5°	0	0	0	0	0
±5°-10°	6	2	33.3	0	0
±10°-15°	35	16	45.7	11	31.4
±15°-20°	30	8	26.7	7	23.3
±20°-25°	25	7	28	6	24
±25°-30°	15	5	33.3	5	33.3
> ±30°	10	4	40	1	0.8
Overall	121	42	34.7	30	24.8

The highest relative agreement of the GP model originates from the meander range 10°-15° for simplified and ±20°-25° for SRDT. However, because of the uneven distribution of samples across meander bins, the interpretation of these results should be treated cautiously. Although the low variation in the wind direction might reflect stable plume alignment, it does not necessarily enhance the model's performance, as the trajectory of the plume may still bypass the sensor array due to smaller turbulence.

3.2.4 Plume Centeredness (Crosswind Distance) and Height Match Score

The plume centeredness score is the ratio of crosswind distance (y) to the horizontal dispersion coefficient (σ_y), a measure of lateral plume spread [25]. This score indicates the closeness of a measurement point to the centerline of a plume. Larger values indicate the measurement position was near the plume edge, where concentrations are typically lowest, and

smaller values correspond to the plume axis, where concentrations are generally highest [18]. The physical significance of plume centeredness is that the GP model assumes the highest concentration is along the plume's centerline, and it decreases laterally with distance following a Gaussian distribution [18,25,29]. If the sensor is farther away from the plume centerline, errors in plume spread and wind direction can lead to a decrease in the predictive performance of the GP model, whereas measurements closer to the plume centerline axis are theoretically less affected by directional uncertainty.

Centeredness score was categorized into four bins: 0-0.5 (n=27), 0.5-1 (n=28), 1-2 (n=33), 1.5->2 (n=33) (Table 3.6), and the analysis was conducted based on the GP model acquired using simplified and SRDT stability classification methods (Table 2.1). Each bin of centeredness represents how far the sensor was positioned laterally from the plume centerline, expressed as multiples of the plume width standard deviation (σ_y). Low values (0-0.5) mean that the sensor was located very close to the plume centerline, intermediate or mid-range values (0.5-1 and 1-2) indicate sensors were positioned farther from the plume centerline, one to two standard deviations of the plume axis. Similarly, the larger values (>2) indicate that the sensor was positioned far from the plume centerline within the outer plume margins, where concentration signals are weakest under the Gaussian distribution centered on the plume axis.

Table 3.6: Plume centeredness score based on the ratio of crosswind offset (y) to lateral plume spread (σ_y) for the GP performance analysis acquired from both simplified and the SRDT method of stability classification. FAC2 hits are the number of samples within a factor of two, and FAC2 hit rate (%) is the percentage of FAC2 hits on specific bins.

Centeredness Score	n_{bin}	Simplified		SRDT	
		FAC2 hits (n)	FAC2 hit rate (%)	FAC2 hits (n)	FAC2 hit rate (%)
0-0.5	27	11	40.7	8	29.6
0.5-1	28	16	57.1	12	42.9
1-2	33	15	45.5	9	27.3
>2	33	0	0	1	3
Overall	121	42	34.7	30	24.8

FAC2 hit rate was limited in the highest score > 2 (simplified = 0, SRDT = 3%). The highest agreement was observed in the 0.5-1 score (simplified = 57.1% and SRDT = 42.9%). The second-highest agreement for simplified (40.7%) and SRDT (27.3%) approaches was in the score of 0-0.5 and 1-2, respectively. This suggests slightly improved performance of GP emission estimations was observed within a half to one standard deviation laterally from the plume axis.

It was expected that the highest FAC2 hit rate would have occurred in the 0-0.5 bin, because the concentration carried near the centerline of the plume is at a maximum. This indicates that centeredness alone is not decisive for the performance of the GP model. When the plume centeredness score is small (0-0.5), a small deviation in wind direction may cause large swings in GP estimations [18], especially for shorter averaging intervals that may miss short-term deviations in the signal, much like the 5-minute measurements used in this study.

The height match score is determined by using the ratio of the absolute difference between source and sensor height to the plume's vertical spread (or vertical dispersion coefficient, σ_z), which is used to determine the extent of vertical overlap between the sensor level and plume. Each bin of height match score represents how far the sensor is positioned vertically from the plume centerline, expressed as multiples of plume height standard deviation (σ_z). The score 0-0.5 indicates that the sensor height lies within $0.5\sigma_z$ (within a half of the plume height standard deviation). In contrast, a height match score of one indicates that sensors are vertically one standard deviation away from the plume's height. The values greater than one indicate that the sensor lies outside the effective vertical reach of the plume. With the evaluation of the height match score, the role of vertical alignment of the sensor and the source can be examined.

Height match score is binned to 0-0.5 (n=98), 0.5-1 (n=19), and 1-2 (n=4) (Table 3.7).

The highest FAC2 hit rate for the simplified method was observed in 0.5-1 (42.1%), and for SRDT, it was in 0-0.5 (26.5%). Under both classifications, the lowest agreement was observed in the highest height-match score (1-2), as there was only 1 FAC2 hit for simplified and none in SRDT.

The results indicate that the model's performance is higher when the height match score < 1. When the height match score increases, there is greater vertical variability between the plume axis and the sensor; thus, the model performance decreases. This is reflected in the SRDT case but not with the simplified one. Due to the considerable variation in sample distributions across height match score bins (samples in the 0-0.5 bin are the maximum; n=98 compared to the remaining bins; n=23), the trend of decrement of FAC2 hit rate with increasing height match score is not exhibited by the simplified method. The largest FAC2 hits were observed in the 0-0.5 score, which also had the highest number of samples among the bins.

Table 3.7: Height match score classification based on the ratio of vertical plume spread (σ_z) to source height (z) for the GP performance analysis acquired from both simplified and the SRDT method of stability classification. FAC2 hits are the number of samples within a factor of two, and FAC2 hit rate (%) is the percentage of FAC2 hits on specific bins.

Height match score	n_{bin}	Simplified		SRDT	
		FAC2 hits (n)	FAC2 hit rate (%)	FAC2 hits (n)	FAC2 hit rate (%)
0-0.5	98	33	33.7	26	26.5
0.5-1	19	8	42.1	4	21.1
1-2	4	1	25	0	0
Overall	121	42	34.7	30	24.8

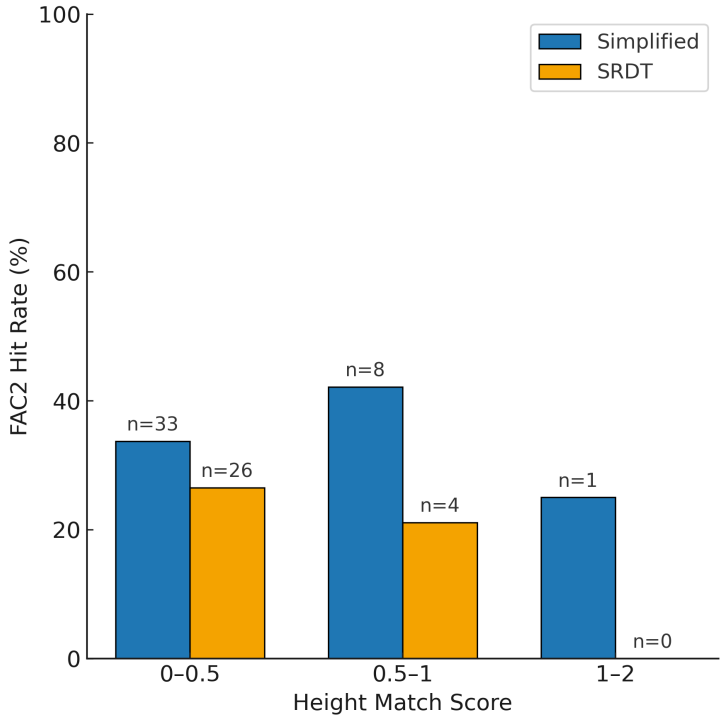


Figure 3.3: Percentage of FAC2 on each height match score bin. Blue bars are the FAC2 acquired from GP modeling using a simplified way of classifying stability. Orange bars are the FAC2 acquired from GP modeling using the SRDT way of classifying stability.

3.2.5 Plume Centerline Distance

The plume centerline distance represents the distance between the source and the sensor measured along the plume's trajectory. Dispersion and dilution increase with plume centerline distance, which can influence the GP model's predictability. At a shorter distance, the plume is more concentrated, while with increasing centerline distance, mixing is enhanced, and signal strength is also reduced [18,25], which may lower the performance of GP. To evaluate this, the centerline distances were binned into three categories: 0-40, 40-80, and >80 m.

Under both approaches, the highest FAC2 hit rate was observed in the plume centerline distance of >80 m, but it is important to note that there were only 7 samples and 4 FAC2 hits from the total samples of 121. Therefore, the FAC2% in this range may deceive that beyond 80m the performance of GP is the best. After >80 m distance, the highest hit rate for the simplified (37.2%) and SRDT (23.8%) method was in 0-40 m and 40-80 m distances, respectively.

It was anticipated that the predictive performance of the model would have been highest in 40-80 m centerline distance [18], but limited samples beyond 80m (n = 7), and higher FAC2 hit rate in 0-40 m (37.2%) than in 40-80 m (30.2%) in the simplified method prevent a robust conclusion.

Table 3.8: Plume centerline distance classification based on the downwind distance between source and sensor along plume axis for the GP performance analysis acquired from both simplified and the SRDT method of stability classification. FAC2 hits are the number of samples within a factor of two, and FAC2 hit rate (%) is the percentage of FAC2 hits on specific bins.

Plume Centerline Distance (m)	n_{bin}	Simplified		SRDT	
		FAC2 hits (n)	FAC2 hit rate (%)	FAC2 hits (n)	FAC2 hit rate (%)
0-40	51	19	37.2	11	21.6
40-80	63	19	30.2	15	23.8
>80	7	4	57.1	4	57.1
Overall	121	42	34.7	30	24.8

In summary, the Gaussian plume analysis highlighted how meteorological variability and plume properties can influence the emission estimations. Following this, a backward Lagrangian stochastic (bLS) model was used to evaluate emissions estimates under the same release conditions.

3.3 Backward Lagrangian Stochastic Analysis

The backward Lagrangian stochastic (bLS) model was implemented using WindTrax, as discussed in Section 2.5.2. Unlike the GP model, which yielded results for all the applied 121 single-point release measurements, WindTrax produced results for just 103 samples. The remaining 18 simulations did not yield valid outputs because the sensor did not detect sufficient particles to infer estimation.

The meteorological inputs used for WindTrax are the same as those applied in the GP model because the meteorological conditions and sampling periods were identical. As there were two analyses for GP based on simplified and SRDT methods of PGSC classification, the WindTrax

analysis was run using PGSC via the SRDT method. This is because SRDT needed solar radiation data, which brings the analysis to common quantifiable values, while the simplified approach in this study is an approximation of what is considered the level of cloud cover, which is subjective and is outlined in [36]. This ensured consistency when comparing model performance under the same classification basis.

Another reason for running bLS using the SRDT method, but not with both approaches, was that WindTrax simulations are computationally intensive with increasing particle number applied to a simulation if it didn't produce results at a smaller particle number, making it longer to generate results. The effect became more pronounced under simplified classification, with a higher number of extremely unstable conditions (class A; Table 3.5 and section 3.3.2) that required increased particle counts for a result. In contrast, higher proportions of class B and C occurred using the SRDT approach compared to class A. Thus, using the SRDT method for the WindTrax application reduced overall computational time and maintained an adequate representation of the dataset for consistent model comparison.

The overall performance of the bLS model is summarized in Table 3.9. FAC2 using WindTrax was 51.5%, suggesting almost half the modeled samples were within a factor of two. Similarly, the geometric mean bias was 0.63 (equation 3), indicating a systematic underprediction of emissions by the model is 37%. The MFoE value of 1.9 (equation 4) implies that the predictive error of the model is roughly two times the true emission value. The MAE (equation 5) shows that the model deviated from the known emissions by 1.5 kg h^{-1} on average, and the RMSE (equation 6) value of 2.6 kg h^{-1} indicates the influence of larger individual errors on some predictions by the model. The very low value of MAE (equation 7; 0.04) reveals almost no systematic directional bias.

Table 3.9: General performance metrics of the bLS model. The metrics include factor of two (FAC2), geometric mean bias (MG), median factor of error (MFOE), mean absolute error (MAE), root mean square error (RMSE), and mean absolute bias (MAB).

Metric	Value
FAC2 (%)	51.5
MG	0.63
MFoE	1.9
MAE	1.5 kg h^{-1}
RMSE	2.6 kg h^{-1}
MAB	0.04

To observe the error structure of the bLS results, the estimated emissions and true emissions were compared using a scatter plot with the FAC2 bounds and a 1:1 reference line (Figure 3.4). The bLS model underpredicts emissions, as many points (66 out of 103) are below the 1:1 line. Furthermore, as 51.5% of modeled emissions fall within the FAC2 bound, the plot also shows that roughly half of the points are within it.

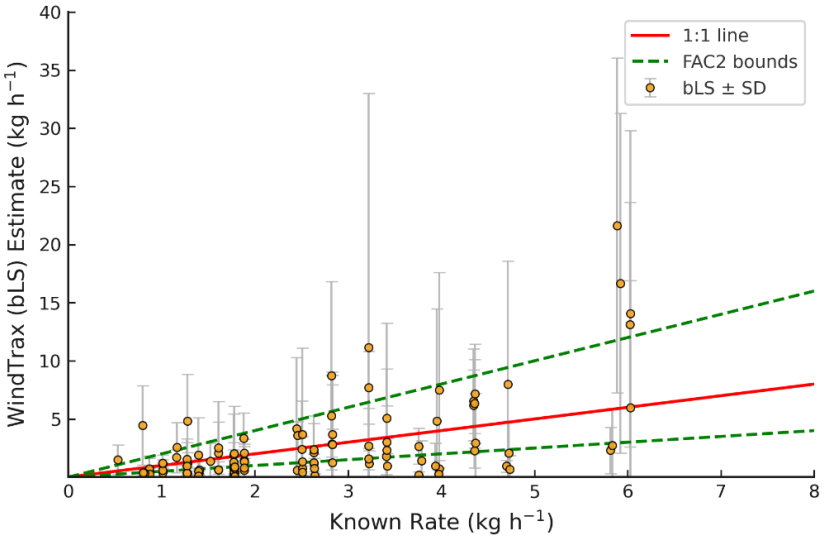


Figure 3.4: Scatter plot of emission estimated by bLS versus the known emission rate from controlled experiments (n=103). Vertical bars represent ± 1 standard deviation (s.d) around each estimate. Standard deviation values in SI.xlsx sheet 'WindTrax'.

Similar to the GP analysis, the subsequent sections evaluate the effect of meteorological conditions and plume geometry on bLS performance. The bLS demonstrated better predictions

compared to GP, so the following subsections aim to investigate the factors contributing to the bLS model's improved predictive ability.

3.3.1 Signal Strength and Release Rate

The samples were divided into the same signal strength quartiles used for the GP analysis: ≤2.3, 2.3–2.9, 2.85–4.0, and >4.0 ppm (Table 3.10). In each bin, 26.9% fell into the FAC2 range in <2.3 ppm, 61.5% into 2.3–2.9 ppm, 48% into 2.9–4 ppm, and 69.2% above 4 ppm. The highest FAC2 hit rate occurred in the fourth quartile signal, followed by the second quartile, while the third and first quartiles had lower agreement.

The highest particle counts (9.5M ± 5.6M) were needed when the signal strength was the lowest (<2.3 ppm), and the particle requirement decreased when the signal strength increased, which suggests that if the concentration is lower, then the particle numbers needed for simulation increase. This pattern likely reflects that when the signal is low, limited released particles reach the sensor due to other factors like atmospheric stability and plume alignment with the sensor.

We would expect a higher FAC2 hit rate with increasing signal strength, but the pattern of the results is not entirely consistent. Weaker signals typically reflect that the plume overlaps the sensor partially, limiting the detection signal, leading to inaccuracy in emission estimation [29,30]. In contrast, stronger signals provide clear and more stable detection, leading to improved estimates.

Table 3.10: Quartiles classification based on signal strength (in ppm) for the bLS performance analysis. FAC2 hits are the number of samples within a factor of two. n_{bin} is the total number of estimates within the specified range/bin, and n_{no_result} is the total number of failed estimates by the model, and FAC2 hit rate (%) is the percentage of FAC2 hits on specific bins.

Strength Quartile (ppm)	n_{bin}	FAC2 hits (n)	FAC2 hit rate (%)	$n_{no\_results}$	Particle# ± s.d.
<2.3	26	7	26.9	8	9.5 M ± 5.6 M
2.3-2.9	26	16	61.5	4	7.1 M ± 3.9 M
2.9-4.0	25	12	48	1	6.4 M ± 5.4 M
>4.0	26	18	69.2	5	4.0 M ± 3.5 M

Overall	103	53	51.5	18

The release rate bins for bLS results were categorized in the same way as for GP (section 3.2.1). The FAC2 hit rate in low (≤1.8 kg h^{-1}), medium (1.8-3.4 kg h^{-1}), and high (1.8-3.4 kg h^{-1}) release rate tertiles was 44.7%, 40%, and 80% respectively (Table 3.11), indicating a narrow variation across low and medium bin (4.7%), but the increased FAC2 agreement in high release rate bin.

The narrow FAC2% band between low and medium bin reflects only slight differences in bLS performance within <1.8 kg h^{-1} and 3.2 kg h^{-1}, but increased FAC2% in the high release rate suggests that the performance of bLS has increased comparatively in > 3.2 kg h^{-1}. The result suggests that with a high release rate (>3.2 kg h^{-1} in this study), the predictive performance of bLS may increase. The bLS performance is also dependent on plume geometry and atmospheric turbulence (section 3.3.2-3.3.5), not entirely on release rate magnitude, because there is no trend of increasing FAC2% when the release rate jumped from low to medium range.

The required particle counts increased with an increase in the controlled release rate. As the release rate increased from <1.8 kg h^{-1} to 1.8 – 3.2 kg h^{-1} to 3.2 – 6.0 kg h^{-1}, the particle number requirement rose from 5.3 M to 9.1 M (Table 3.11). Although the result highlights that more particle counts are required for the high release rate bin than the low, the medium bin required fewer particles compared to the low bin. Also, there is a consistent increase in particle variability (s.d.) with increasing release rate bins, suggesting that the relationship between particle requirement and FAC2 agreement may not be strictly physical but potentially influenced by plume offset and unequal sample distribution.

Table 3.11: Tertiles classification based on release rates (kg CH_4 h^{-1}) for the bLS performance analysis. FAC2 hits are the number of samples within a factor of two. n_{bin} is the total number of estimates within the specified range/bin, n_{no_result} is the total number of failed estimates by the model, and FAC2 hit rate (%) is the percentage of FAC2 hits on specific bins.

Release rate bin tertiles (kg h^{-1})	n_{bin}	FAC2 hits (n)	FAC2 hit rate (%)	$n_{no\_results}$	Particle# ± s.d.
Low: <1.8	38	17	44.7	2	5.3 M ± 4.0 M
Medium: 1.8-3.2	40	16	40	3	5.2 M ± 4.2 M
High: 3.2-6.0	25	20	80	13	9.1 M ± 5.6 M
Overall	103	53	51.5	18	

3.3.2 Atmospheric Stability

As stated prior (section 3.3), the dataset used in bLS contained PGSC classification derived from the SRDT method, including the unstable PGSC classes A, B, and C, with A being the most unstable and C being the least. Their corresponding sample sizes are mentioned in Table 3.13. The highest amount of FAC2 hit rate was observed in PGSC category C (56.3%), compared to A and B (A = 50% and B = 50.7%). Slightly unstable (class C) conditions provide enhanced horizontal and vertical mixing with fewer disturbances in plume structure, leading to improved model predictive performance, whereas more unstable conditions (class A and B) enhance concentration variability due to strong turbulence and erratic plume behavior, leading to less reliable estimations [18,29].

The decrease in particle number requirement from A (7.8 M) to B (6.4 M) to C (5.9 M) (Table 3.12) suggests that extremely unstable conditions need the most particles compared to moderately unstable conditions, followed by slightly unstable conditions. It means under stronger turbulence, trajectories of random particles become more erratic, thereby reducing the likelihood of successful particle tracking to the sensor's location in the bLS model, and vice versa under more stable conditions.

Table 3.12: Number of samples under various PGSC and corresponding FAC2. Sampled data had the PGSC of classes A, B, and C, which were characterized using the SRDT's method. FAC2 hits are the number of samples within a factor of two. n_{bin} is the total number of estimates within the specified range, n_{no_result} is the total number of failed estimates by the model, and FAC2 hit rate (%) is the percentage of FAC2 hits on specific bins.

PGSC	n_{bin}	FAC2 hits (n)	FAC2 hit rate (%)	$n_{no\_results}$	Particle# ± s.d.
A	18	9	50	3	7.8 M ± 5.9 M
B	69	35	50.7	5	6.4 M ± 4.7 M
C	16	9	56.3	10	5.9 M ± 5.0 M
Total	103	53	51.5	18	

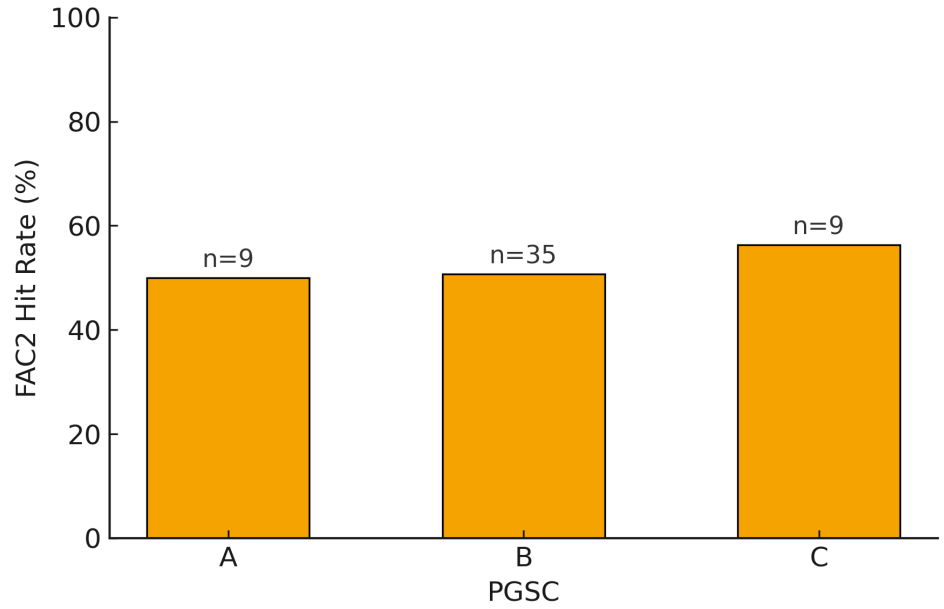


Figure 3.5: Percentage of FAC2 on each PGSC classification. n is the number of samples that fall under FAC2 among 103 samples.

3.3.3 Wind Direction Meander

Wind meander was grouped into 5° bins (0°-±5°, ±5°-±10°, ±10°-±15°, ±15°-±20°, ±20°-±25°, ±25°-±30°) and >±30° to <±60° (Table 3.13). Wind meander represents the standard deviation of wind direction during each 5-minute sampling period. Thus, by definition, an increase in the meander value indicates a wider lateral plume around its centerline.

The highest FAC2 hit rate was observed in ±5°-10° (80%), but it should be noted that only 5 out of 103 samples were in this category with 4 FAC2 hits. After ±5°-10° meander, the highest

FAC2% was in ±15°-20° (57.1%), followed by 55.6% in the ±10°-15° bin. FAC2 peaked between ±5° to ±20° and then gradually fell for all other meander bins. Overall, 34% of FAC2 shares out of the overall 51.5% were observed between ±5° to ±20°. However, this trend should be carefully interpreted, considering the uneven sample distribution. It is important to note that the number of samples in ±0°-5° was none, and in ±5°-10° was only 5. Similarly, only 22 out of 103 samples were beyond ±25°. Most samples (76 out of 103) were between ±10° to ±25° meander about the plume centerline. At a very low meander (<±10°), the plume is narrowly constrained so that a small change in wind direction may cause the sensor to miss signals. This is because of a 5-minute sampling period; the sudden variation/fluctuation in wind direction may not have been captured, but between ±0°-5° of meander, bLS has shown its capability of enhanced performance, reflecting its ability to consider the low meander (but should be cautiously interpreted due to fewer samples in this range). Conversely, when the meander is higher (>±25°), overall lateral coverage of the plume increases, which reduces the chances of capturing clear signals, leading to poor performance. Additionally, the wind meander bins where higher FAC2 agreement was observed (above ±10°) needed more particle counts compared to the remaining meander bins. It is because when the variability of wind direction increases, the plume may miss the sensor array and the requirement of particles increases.

Table 3.13: Wind meander classification based on its standard deviation for the bLS performance analysis. FAC2 hits are the number of samples within a factor of two. n_{bin} is the total number of estimates within the specified range/bin, n_{no_result} is the total number of failed estimates by the model, and FAC2 hit rate (%) is the percentage of FAC2 hits on specific bins.

Wind meander bins (°)	n_{bin}	FAC2 hits (n)	FAC2 hit rate (%)	$n_{no\_results}$	Particle# ± s.d.
0°-5°	0	0	0	0	
5°-10°	5	4	80	1	3.0 M ± 3.1 M
10°-15°	27	15	55.6	8	5.5 M ± 3.3 M
15°-20°	28	16	57.1	2	7.0 M ± 5.5 M
20°-25°	21	11	52.4	4	7.2 M ± 5.6 M
25°-30°	14	6	42.9	1	7.0 M ± 4.9 M

>30°	8	1	12.5	2	8.1 M ± 6.3 M
Total	103	53	51.5	18	

3.3.4 Plume Centeredness (Crosswind Distance) and Height Match Score

The plume centeredness scores were grouped into 0-0.5, 0.5-1, 1-2, and >2 (Table 3.14). This is a normalized metric derived from the ratio of crosswind offset (y) to lateral plume spread/horizontal dispersion coefficient (σ_y), for which lower values indicate greater closeness (or decreased crosswind distance) of the measurement point from the plume centerline. Concentration decreases with an increase in centeredness and vice versa. The lowest and highest hit rate was observed in 0.5-1 (37%) and >2 (72.2%) bins, respectively. However, no clear systematic trend or relationship between centeredness and FAC2 hit rate was observed, suggesting that within the dataset, centeredness did not contribute to the predictability of bLS in this study.

The number of particles needed for the model to converge for estimation increased as centeredness increased (Table 3.14). This is because, as centeredness increases, the concentration decreases, and a higher particle count is required for the model to produce estimates.

Table 3.14: Plume centeredness score based on the ratio of crosswind offset (y) to lateral plume spread (σ_y) for the bLS performance analysis. FAC2 hits are the number of samples within a factor of two. n_{bin} is the total number of estimates within the specified range/bin, n_{no_result} is the total number of failed estimates by the model, and FAC2 hit rate (%) is the percentage of FAC2 hits on specific bins.

Centeredness (m)	n_{bin}	FAC2 hit (n)	FAC2 hit rate (%)	$n_{no\_results}$	Particle# ± s.d.
0-0.5	27	13	48.1	0	4.7 M ± 4.0 M
0.5-1	27	10	37	1	6.7 M ± 4.4 M
1-2	31	17	54.8	2	7.2 M ± 5.8 M
>2	18	13	72.2	15	7.8 M ± 4.9 M
Total	103	53	51.5	18	

The height match score was divided into three bins: 0-0.5, 0.5-1, and 1-2, among which the largest samples occurred in the smallest bin (0-0.5) (Table 3.15). The height match score is the absolute vertical offset between release and sensor heights normalized by vertical plume spread

(σ_z). As a result, a low value, like <0.5, indicates the sensor was positioned at less than half the plume's vertical standard deviation, where concentration peaks and chances of detection were high. In contrast, the samples above 0.5 represent that the plume axis was more than $0.5\sigma_z$ (half of the vertical dispersion coefficient) above the sensor level, which likely reduces the detection and thus lowers the model's performance. Ninety out of 103 samples were within the 0-0.5 bin, which means that, in most of the measurements, the sensor height was generally aligned with the plume centerline.

The highest FAC2 hit rate occurred in the 1-2 bin (100%), but it was due to only one sample, and that sample got the FAC2 hit (Figure 3.6). So, it does not contribute to the finding that in a 1-2 height match score, the model performs well. The result reflects that with increasing height match score FAC2 hit rate also increases, but it is deceptive as it is important to note that the samples beyond a height match score of 0-0.5 share only ~13% of the total sample, while 0-0.5 has 90 out of 103 samples (87% of the samples).

The particles needed when the height match score was 0-0.5 averaged 6.3M with a wide standard deviation (±5M), followed by 8.8M in the 0.5-1 bin. The height match score between 1-2 had only one valid sample. The lower particle requirement in a 0-0.5 bin suggests that if the plume centerline and sensor are vertically well aligned, fewer particles are required to obtain the emission estimates from WindTrax.

Table 3.15: Height match score classification based on the ratio of vertical plume spread (σ_z) to source height (z) for the bLS performance analysis. FAC2 hits are the number of samples within a factor of two. n_{bin} is the total number of estimates within the specified range/bin, n_{no_result} is the total number of failed estimates by the model, and FAC2 hit rate (%) is the percentage of FAC2 hits on specific bins.

Height match score	n_{bin}	FAC2 hits (n)	FAC2 hit rate (%)	$n_{no\_results}$	Particle# ± s.d.
0-0.5	90	45	50	8	6.3 M ± 5.0 M
0.5-1	12	7	58.3	7	8.8 M ± 4.0 M
1-2	1	1	100	3	6.0 M ± 5.8 M
Total	103	53	51.5	18	

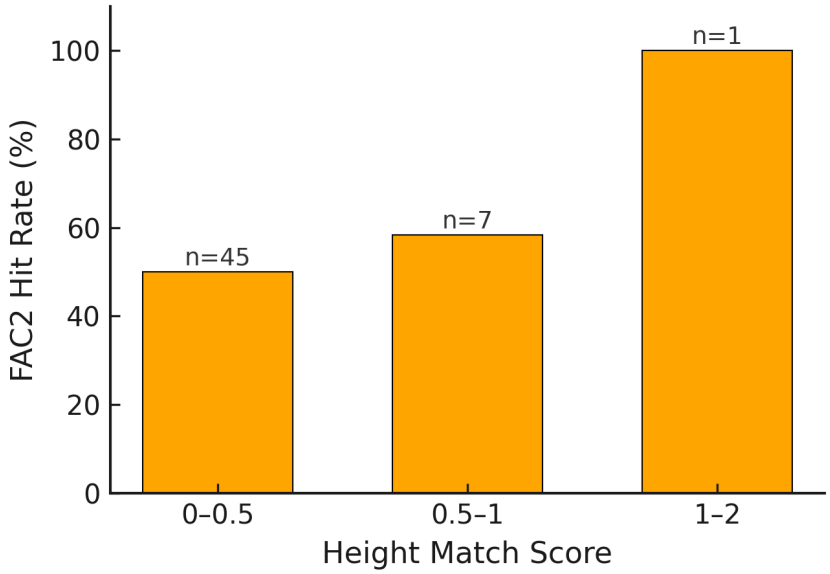


Figure 3.6: Percentage of FAC2 on each height match score bin. n is the number of samples that fall under FAC2 among 103 samples.

3.3.5 Plume Centerline Distance

Plume centerline distance is the distance between the measurement point and source along the plume axis and is categorized into three different bins for this dataset: 0-40, 40-80, and >80 m (Table 3.16). Many of the measurements (n = 52) were conducted between centerline distances of 40-80 m, followed by 0-40 m (n = 44) and >80 m (n = 7). The highest FAC2 hit rate was observed between 40 to 80 meters (57.7%), and the least was >80 m (28.6%), likely due to the limited sample size of only 7 for >80 m distance. Although the FAC2 hit rate in 40-80 m bins was the highest, this should not be interpreted as evidence that the bLS model always performs better in this range, due to the inconsistency in sample size. It is because 40-80 m is the distance at which the plume sufficiently disperses and concentration stabilizes, leading to better detection, whereas at shorter distances (<40 m), the plume is still developing and sensitive to small turbulence (especially in unstable conditions). Conversely, in greater distances (>80 m), atmospheric mixing enhances plume dispersion which weakens the signal detection.

The requirement of particle numbers increased with an increase in plume centerline distance (Table 3.16). It is because as the distance between the sensor and source increases along the plume axis, the released particles are diluted, reducing the concentration and limiting the sensor from achieving a strong signal [18].

Table 3.16: Plume centerline distance classification based on the downwind distance between source and sensor along the plume axis for the bLS performance analysis. FAC2 hits are the number of samples within a factor of two. n_{bin} is the total number of estimates within the specified range/bin, n_{no_result} is the total number of failed estimates by the model, and FAC2 hit rate (%) is the percentage of FAC2 hits on specific bins.

Plume Centerline Distance (m)	n_{bin}	FAC2 hits (n)	FAC2 hit rate (%)	$n_{no\_results}$	Particle# ± s.d.
0-40	44	21	47.7	7	4.3 M ± 4.2 M
40-80	52	30	57.7	11	8.0 M ± 4.7 M
>80	7	2	28.6	0	9.9 M ± 5.9 M
Total	103	53	51.5	18	

3.3.6 Particle Number Distribution

The particle numbers used in WindTrax for the stochastic simulation influence the quality of emissions estimations [29,37]. As explained in section 2.5.2, the initial particle count used for each sample was 500,000, and the count was increased to 1M, then 2M, and subsequently by 2M after that (4M, 6M, 8M…) up to 20M, unless the result was achieved. However, increasing the particle count also significantly increases computational demand.

It is important to recall that concentration measurements were conducted with a 5-minute averaging period, and the bLS simulation was run with a sampling volume of 0.05 m x 0.05 m (Appendix B: Table B1). Larger sampling volume, especially, would have required fewer particles per run and thus, less computational effort. A longer averaging period (>5 minutes) would likely better represent wind variability with higher statistical reliability, and this could reduce uncertainty arising from short-term fluctuation of signals. Therefore, using longer sampling times and larger physical inlet sampling volumes should be prioritized for future experiments.

The estimated emission from WindTrax was evaluated across the particle count bins (0.5-5M, 5-10M, 10-15M, and 15 -20M, where M denotes million). FAC2 among these bins was analyzed to observe the impact of particle numbers on the model's performance (Table 3.17). The results showed that when 0.5 to 5 M particles were used, 48 out of 103 samples produced estimations. Similarly, when 5 to 10 M particles were required, 38 estimates were obtained. The number dropped to 9 and 8 when 10-15 M and 15-20 M particles were used, respectively. The number of samples beyond 10 M has decreased, and with this sample distribution, the FAC2 hit rate has not produced any trend with particle count requirement (Figure 3.17). When 121 equivalent samples from the GP analysis were applied to bLS, 18 did not generate estimates when 20 M particles were used, which limits our samples for bLS to 103 (Table 3.17). Although the FAC2 agreement is higher in 10-15 M and 15-20 M, it is important to note that samples requiring higher particle counts were considerably smaller (10-15M; n=5 and 15-20M; n=6).

Table 3.17: Particle count distribution bin for the bLS performance analysis. FAC2 hits are the number of samples falling under the FAC2 agreement, and FAC2 hit rate (%) is the percentage of FAC2 hits on specific bins.

Particle Count Bin (in Million)	n_{bin}	FAC2 hits	FAC2 hit rate (%)
1-5M	48	26	51.2
5-10M	38	16	42.1
10-15M	8	5	62.5
15-20M	9	6	66.7
Overall	103	53	51.5
20M; no result	18		

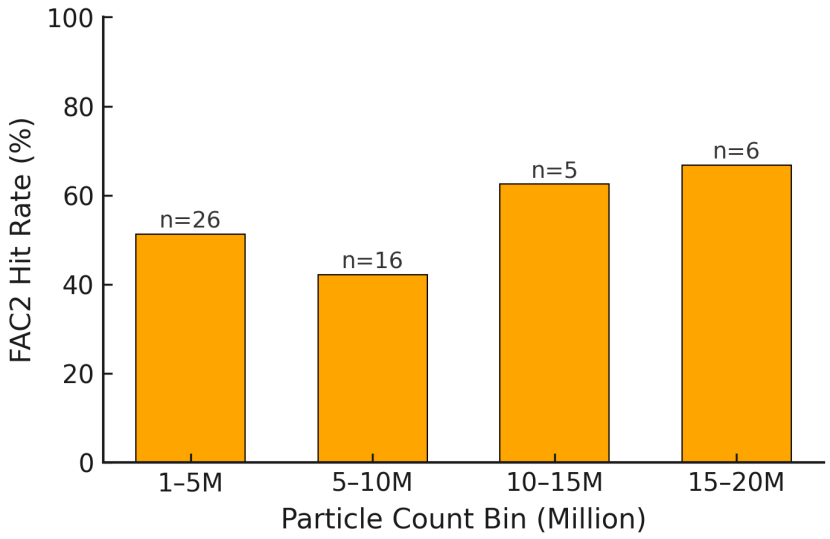


Figure 3.7: FAC2 hit rate on each particle count bin used in WindTrax run. n is the number of samples that fall under FAC2 among 103 samples.

The evaluation to assess the sensitivity of various parameters (i.e., known release rate, vertical height match score, plume centerline distance, plume centeredness score, and wind meander) on required particle counts used for the simulation, a multivariate regression analysis (Table 3.18) was performed, with particle counts as the independent variable and the remaining parameters as dependent.

Table 3.18: Regression result showing the influence of experimental parameters on the particle counts required for WindTrax runs. R^2 is the coefficient of determination, *m* is the regression gradient (slope), p-value for statistical significance, and the intercept of the regression line (*b*).

Variable	R^2	Slope (*m*)	p-value	Intercept (*b*)
Known release rate (kg CH_4 h^{-1})	0.18	0.02	0.98	2
Height match score	0.09	0.002	0.03	0
Plume centerline (meters)	0.21	0.32	1.29×10^{-5}	29
Plume centeredness score	0.11	0.01	3.25×10^{-4}	1
Wind Meander (°)	0.002	0.03	0.1	17

The coefficient of determination (R^2) measures the variability of particle count explained by each independent variable. R^2 ranges between 0 to 1; values near 1 indicate strong explanatory and dependence on a particular variable, and vice versa when the value is near 0. Among the

analyzed variables, the plume centerline distance showed the highest coefficient of determination ($R^2 = 0.21$), followed by known release rate ($R^2 = 0.18$). It means that only 21% of the variation in particle count can be explained by plume centerline distance, while 18% is attributed to the known release rate. Although the R^2 value is small, the plume centerline distance has the strongest influence on particle counts among the dependent variables.

The considerable difference in statistical significance can be observed with the plume centerline distance and known release rate. P-value tests whether the variable is statistically significant or not, and a value <0.05 indicates that it is significant. P-values closer to 0 increase the statistical significance, and the observed results are unlikely due to random chance. The plume centerline distance was highly significant with a p-value 1.29×10^{-5}, whereas the known release rate was not (p-value = 0.98). Another parameter showing measurable influence was the centeredness score, with p-value 3.25×10^{-4} and R^2 of 0.11.

Overall, the results suggest that the particle requirement for WindTrax simulation depends more on plume alignment and geometry than on the emission itself.

3.4 Comparison of GP and bLS Model Performance

The differences in the results produced by both models can be directly compared because the same performance metrics were used to analyze the GP and bLS models. A comparative evaluation of the two models under the SRDT-based stability classification and identical sampling and meteorological conditions was performed (Table 3.19).

Comparing the result with GP, bLS demonstrated stronger overall performance. Overall FAC2 hit rate increased from 24.8% for GP to 51.5% using bLS, and the multiplicative predictive error decreased from 3.5 to 1.9. Similarly, the geometric mean bias improved with bLS as it systematically underpredicts by 37%, which is better than that for GP (92% underprediction).

Both models have underpredicted emissions (GP by 92% and bLS by 37%); estimation by bLS was notably closer to true emissions (Table 3.19). The reason is that the bLS simulates random particle trajectories, enabling the model to capture turbulent dispersion more realistically than GP.

Error-based measures like MAE, RMSE, and MAB further highlight the stronger comparative performance of bLS. The MAE and RMSE were lower for bLS predictions (1.5 and 2.6 kg h^{-1}, respectively) compared to GP predictions (2.4 and 3 kg h^{-1}, respectively). The MAE of bLS is improved as the bLS has produced only a minor residual bias (0.04), while GP has a more pronounced underprediction (-0.8).

Table 3.19: Comparison of GP and bLS models based on their performance metrics.

Metric	GP	bLS	Absolute difference
FAC2 (%)	24.8	51.5	26.7
MG	0.08	0.63	0.55
MFoE	3.5	1.9	reduced by a factor of about 2
MAE	2.4 kg h^{-1}	1.5 kg h^{-1}	0.9
RMSE	3.0 kg h^{-1}	2.6 kg h^{-1}	0.4
MAB	-0.8	0.04	0.84 (bias reduction)

The differences in how the two models responded to atmospheric stability were also examined. The highest FAC2 hit rate for GP was in stability class A (33%), and for bLS it was class C (56.3%). It is a contrasting result and indicates the influence of the atmospheric stability class on GP and bLS performance. However, it must be taken into consideration that the range of FAC2 hit rate within the classes analyzed for GP and bLS is minimal (10.3% and 6.3%, respectively), and the stability class B had the highest samples compared to A and C, which shows the variation in sample distribution.

When compared with respect to plume geometry parameters, both models showed distinct behavior. The agreement across plume centerline bins was consistently higher in bLS, with the strongest agreement between 40-80 m range (section 3.3.5; 57.7% for bLS). The highest agreement

for GP was observed in >80 m plume centerline distance (section 3.2.5; 57.1% for GP). But it is important to consider that the total sample >80 m distance is only 7. So, if we undermine >80 m samples, the FAC2 hit rate at 40-80 m distance (23.8%) for GP is greater than in 0-40 m (21.6%). The bLS model has consistently performed better in each centerline bin because it represents particle transport and turbulence, which allows sufficient distance for the plume to disperse.

For height match score in 1-2 bin, the number of samples in GP and bLS were 4 and 1, respectively. This may not represent the analytical conclusion of models performing better at this height score. However, in 0-0.5 and 0.5-1 match score ranges, the bLS had a better FAC2 hit rate in each of these bins. The FAC2 hit rate in 0-0.5 and 0.5-1 for GP and bLS were 26.5% & 21.1% and 50% & 58.3%, respectively. So, the result reflects that the bLS model is more capable of considering the vertical overlap between sensor level and plume compared to the GP model.

While comparing the models with respect to plume-centeredness score, the highest FAC2 hit rate for GP (42.9%) and bLS (72.2%) was observed in the 0.5-1 and >2 score, respectively. In >2 bin, where bLS had the highest FAC2 hit rate, the GP had a FAC2 hit rate of only 3%. It shows that the performance of bLS model has not been significantly influenced by how far the sensors are kept laterally from the plume centerline for the measurement. It suggest that bLS model was able to capture and represent the plume dispersion better than the GP model in terms of the centeredness score.

For signal strength bins, the highest FAC2 hit rate was under the same category (signal strength >4 ppm), although the bLS had a greater agreement (30% for GP and 69.2% for bLS). A contrasting result was observed with performance analysis based on release rate, as the highest agreement for GP was in the ≤ 1.8 bin and bLS in the 3.2 to 6 kg h^{-1} bin. Along with this contrasting result for release rate, it should be noted that the range of FAC2 hit rate for GP and bLS across all

the bins was 6.6% and 40% respectively. The GP showed a narrow variation in hit rates, whereas the bLS showed a much wider spread. It means that, unlike the GP model, bLS is more sensitive to release rate and tends to perform better, particularly in high release rate conditions (3.4-6 kg h^{-1}).

The plume geometry is better represented by WindTrax as random particle trajectories from the source to the sensor are captured in a dispersion pattern more effectively.

In case of wind meander, the largest portion of FAC2 hit for bLS was between ±10° to ±20° (31 out of 103; ~30%), and in GP, it was 18 out of 121 (~15%). This suggests that under wind variability between ±10°- ±20° in this study, the WindTrax simulation was more effective than the GP. The reason for this is associated with the bLS approach for accounting for varying wind flow patterns in the simulation as random particles, whereas the GP model assumes uniform dispersion occurred without considering meander.

The estimated emission rate (blue and orange) and true emission rate (green dashed) for both GP and bLS models are shown in Figure 3.8. The plot helps to visualize how far the emission estimates are from true rates.

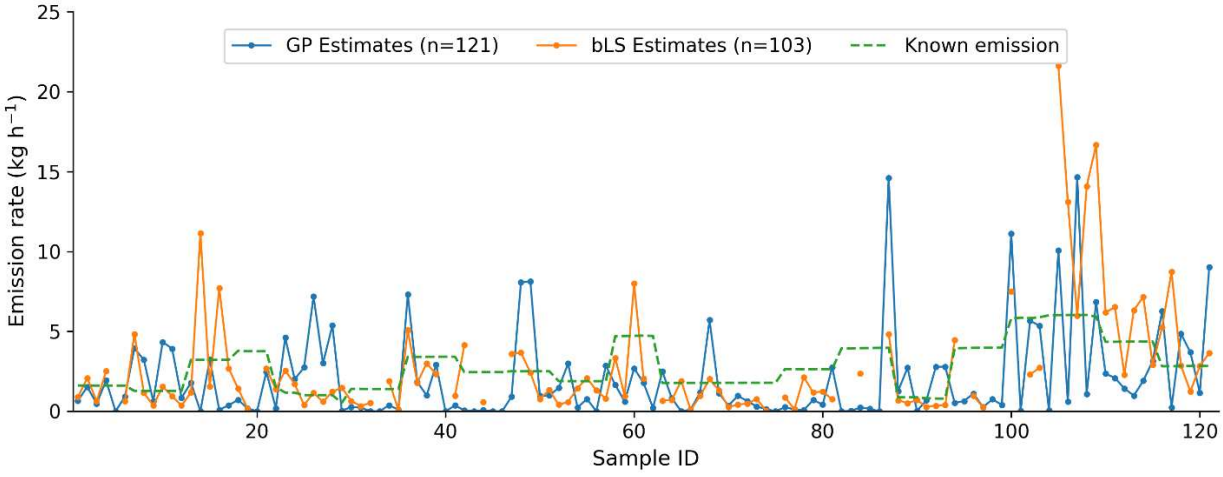


Figure 3.8: Comparison of emissions estimates from GP (blue) and bLS (orange), and corresponding known emission rates (green dash) by sample ID. Gaps in the bLS series reflect samples for which WindTrax estimates were unavailable; all estimates are aligned by sample ID.

The differences in performance of GP and bLS models can be explained by how these models represent the dispersion process and atmospheric turbulence. The bLS simulates individual random particles driven by atmospheric turbulent variation, allowing it to capture the varying nature of particle trajectories within a plume. This stochastic nature of the bLS model adapts to variation in wind conditions that leads to better near-field estimations. Unlike bLS, the GP model depends on parameterized dispersion coefficients to represent average plume spread, restricting the ability to capture and simulate turbulence variations directly. As a result, bLS achieves better agreement with true emission rates and higher FAC2 hits due to its ability to simulate the behavior of the plume more realistically than the GP approach.

Overall, the bLS model had greater predictive reliability compared to the GP model, with approximately two times more overall FAC2. While the GP model was faster and more convenient for estimations, its accuracy declined due to wind variability and complexity in plume geometry. Conversely, the bLS model, though computationally more intensive, more accurately estimated the emissions under unstable near-field O&G conditions. But it is to note that the bLS estimate through WindTrax produces the results with a standard deviation on emission estimates, indicating variability in model outputs.

Chapter 4 CONCLUSION AND RECOMMENDATIONS

This study evaluated the performance of two commonly used atmospheric dispersion models – the Gaussian plume (GP) and the backward Lagrangian stochastic (bLS) models for methane emissions estimation under controlled near-field releases conducted at the METEC facility at Colorado State University. Using controlled release conditions, model estimations were compared to the true emissions rates. A short sampling period of a 5-minute interval was used during the concentration measurement to test estimates of emission for shorter time frames. The analysis focused on how varying wind conditions, stability class, source-sensor alignment, and plume geometry affect the performance of the GP and bLS models.

The atmospheric stability, plume geometry, and wind conditions are described separately in the results; they are not independent in the dispersion modeling. Plume geometry (centeredness, height match score, and centerline distance) governs the performance of dispersion models, as poor alignment of the sensor to the source or plume degrades emission estimates. Atmospheric stability affects the model performance through its influence on turbulence and wind meander, as unstable conditions enhance the sensitivity of models. Therefore, atmospheric stability and plume geometry are combined drivers of models' performance and should not be treated as independent factors in dispersion modeling applications.

The meteorological conditions during the experiment were all conducted during the daytime, with light to moderate wind speeds. The atmospheric turbulence conditions defined with the Pasquil-Gifford stability classification were all in unstable regimes (classes A, B, and C).

The simplified and solar radiation delta-T (SRDT) approach for PGSC classification was applied to estimate emissions from the GP model [20,29,36]. The simplified method is based on

wind speed and subjective assessment of solar radiation from Turner's approach, whereas SRDT is based on an explicit range of solar radiation and wind speed that objectively defines Pasquill-Gifford stability classes. Results demonstrated that the GP model using the simplified method of PGSC classification yielded slightly better results compared to using the SRDT approach, as the overall FAC2 hit rate for SRDT was 24.8% and that for the simplified approach was 34.7%. The systematic underprediction using the simplified approach (83%) was less than the SRDT approach (92%). However, the multiplicative spread error from SRDT and the simplified approach were almost the same (simplified = 3.4, SRDT = 3.5). Similarly, the mean absolute error and root mean square error of bLS was slightly less than GP (by 0.9 and 0.4 kg h^{-1}, respectively).

The bLS model was implemented using the SRDT method of classifying stability, and the results were compared with GP analysis using the SRDT method under the same meteorological and experiment inputs. The bLS model was applied using the WindTrax software package [31]. The resultant FAC2 hits from bLS quantification demonstrated better agreement of the model (53 FAC2 hits out of 103 samples; 51.5%) with known emissions over GP (30 FAC2 hits out of 121 samples; 24.8%). The systematic mean underestimation by bLS, shows better performance compared to the GP results. The stochastic method of transporting the particles from the source to the sensor allowed WindTrax to better capture varying atmospheric turbulence [12]. As a result, the analysis of plume geometry and wind variations on the performance of both the models reveals that bLS can approximate the estimates closer to the true emission values better than the GP model. The performance of bLS was better than the GP model, under various analyzed parameters like release rate, signal magnitude, wind meander stability, plume centeredness, centerline distance, and height match score.

The study helps to understand the behavior of two commonly used dispersion models, GP and bLS, however, both models are constrained to certain limitations. The GP model is based on the steady-state assumption of emission and wind flow over homogeneous terrain [26,29], which restricts its applicability to complex O&G fields' conditions. The bLS model also assumes similar homogeneity, but it is also computationally demanding due to its large particle requirement [12,29,30].

The 5-minute sampling period has shown a rapid wind meander, which violates the assumptions of steady state flow. This averaging period may not be enough to capture the wind variability present in the near-field conditions, especially when the atmospheric scope is limited to daytime unstable classes (PGSC: A, B, and C) (see Tables 3.5, 3.6, 3.13, 3.14). Measurements performed within 100 m distant from the source also undermine the GP assumption of well-developed plume dispersion. Due to the time constraints in bLS application, the higher particle simulations beyond 20M were not performed, adding another limitation associated with WindTrax.

For near-field, single-source O&G conditions similar to METEC, findings suggest that bLS is generally a more suitable choice than GP because bLS model achieved higher agreement with true release rates (smaller bias, lower MFoE, and higher FAC2) and was less sensitive to specific combinations of plume centeredness score, height match score, centerline distance, and stability class. In comparison, GP often underestimated emissions and had larger errors when compared to known release rates. The results suggest that GP is appropriate in settings with uncomplicated geometries and conditions that closely align with its underlying assumptions, like moderate winds, better lateral and vertical alignment of plume centerline and sensor. Under more turbulent and complex atmospheric conditions, bLS generally offers more reliable estimates, provided that sufficient particle counts and accurate meteorological inputs are used. However, bLS model is

time-consuming due to its computational efforts compared to GP. While GP achieved higher FAC2 hit rates in some subsets, such cases were primarily associated with favorable plume geometry and small sample sizes, indicating the improved performance is due to geometric alignment and sample distribution rather than robust stability conditions.

For practitioners relying on GP-based tools for near-field quantification, results suggest that this model should be applied under the condition of a steady wind field, uncomplicated site geometry, and well-aligned plume-sensor configurations. Sampling strategies should prioritize strong concentration enhancements and avoid highly unstable turbulence conditions. Also, the longer average sampling period may improve the representativeness of wind conditions that may enhance the model's performance.

Future research should focus on overcoming the limitations outlined in this study. Longer averaging duration should be considered to reduce the influence of short-term wind variability on emission estimates and improve model predictive performance [38,39]. The reason for using a 5-minute interval for this study was due to the limited availability of a testing facility for single control release, because shutting down all the operations for the single project was impractical. Extending the averaging time would likely improve wind representativeness, which is an important factor driving the plume.

Experiments should incorporate multiple measurement heights. The height match score analysis (sections 3.2.4 and 3.3.4) revealed that samples collected from closely aligned source heights have higher FAC2 hits (for height match scores of 0 to 0.5: FAC2 hits in GP = 26 out of 30 total hits; in bLS = 45 out of 53 total hits). Multi-level sensors for the measurement collection would help better represent and characterize plume dispersion. The measurements incorporating multiple releases should also be considered because they can replicate real O&G conditions where

multiple sources are emitting at once. Also, model performance on overlapping plumes can be studied.

Beyond modification of the experimental setup, upcoming work should expand the range of atmospheric stability because this study evaluated the models' performance under classes A, B, and C. The evaluation under neutral and stable conditions is therefore necessary to assess model performance under a wide range of atmospheric turbulence. This study shows the performance of models dependent on PGSC category, as the GP model had the highest FAC2 hit rate at class A (33.3%; see section 3.2.2), and bLS showed an improving FAC2 hit rate with increasing stability from class A to C (section 3.3.2).

Overall, the findings highlight that better emission estimations from dispersion models depend on model parameters and configuration on meteorological and geometric conditions.

REFERENCES

[1] Intergovernmental Panel on Climate Change (IPCC). Chapter 7: The Earth's energy budget, climate feedbacks, and climate sensitivity. In Climate Change 2021 – The Physical Science Basis: Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, 2021. Accessed: Oct. 24, 2025. Available at: https://www.ipcc.ch/report/ar6/wg1/chapter/chapter-7/

[2] M. Saunois, A. Martinez, B. Poulter, Z. Zhang, P. A. Raymond, P. Regnier, J. G. Canadell, R. B. Jackson, P. K. Patra, P. Bousquet, P. Ciais, E. J. Dlugokencky, X. Lan, G. H. Allen, D. Bastviken, D. J. Beerling, D. A. Belikov, D. R. Blake, S. Castaldi, M. Crippa, B. R. Deemer, F. Dennison, G. Etiope, N. Gedney, L. Höglund-Isaksson, M. A. Holgerson, P. O. Hopcroft, G. Hugelius, A. Ito, A. K. Jain, R. Janardanan, M. S. Johnson, T. Kleinen, P. B. Krummel, R. Lauerwald, T. Li, X. Liu, K. C. McDonald, J. R. Melton, J. Mühle, J. Müller, F. Murguia-Flores, Y. Niwa, S. Noce, S. Pan, R. J. Parker, C. Peng, M. Ramonet, W. J. Riley, G. Rocher-Ros, J. A. Rosentreter, M. Sasakawa, A. Segers, S. J. Smith, E. H. Stanley, J. Thanwerdas, H. Tian, A. Tsuruta, F. N. Tubiello, T. S. Weber, G. R. van der Werf, D. E. J. Worthy, Y. Xi, Y. Yoshida, W. Zhang, B. Zheng, Q. Zhu, Q. Zhu and Q. Zhuang. Global methane budget 2000–2020. *Earth System Science Data*, vol. 17, no. 5, pp. 1873–1958, May 2025. Available at: https://doi.org/10.5194/essd-17-1873-2025

[3] T. Stein. No sign of greenhouse gases increases slowing in 2023. NOAA Research, 2023. Accessed: Oct. 24, 2025. Available at: https://research.noaa.gov/no-sign-of-greenhouse-gases-increases-slowing-in-2023/

[4] National Oceanic and Atmospheric Administration (NOAA). Current GHG levels. NOAA Climate.gov, 2025. Accessed: Oct. 24, 2025. Available at: https://www.climate.gov/ghg/current-levels

[5] National Academies of Sciences, Engineering, and Medicine. Methane emission measurement and monitoring methods. In Improving Characterization of Anthropogenic Methane Emissions in the United States. *National Academies Press (US)*, 2018. Accessed: Oct. 03, 2025. Available at: https://www.ncbi.nlm.nih.gov/books/NBK519293/

[6] S. Peng. Challenges and opportunities in the global methane cycle. *iScience*, vol. 26, no. 6, p. 106878, Jun. 2023. Available at: https://doi.org/10.1016/j.isci.2023.106878

[7] I. M. L. D. Storm, A. L. F. Hellwing, N. I. Nielsen and J. Madsen. Methods for measuring and estimating methane emission from ruminants. *Animals (Basel*), vol. 2, no. 2, pp. 160–183, Apr. 2012. Available at: https://doi.org/10.3390/ani2020160

[8] R. A. Alvarez, D. Zavala-Araiza, D. R. Lyon, D. T. Allen, Z. R. Barkley, A. R. Brandt, K. J. Davis, S. C. Herndon, D. J. Jacob, A. Karion, E. A. Kort, B. K. Lamb, T. Lauvaux, J. D. Maasakkers, A. J. Marchese, M. Omara, S. W. Pacala, J. Peischl, A. L. Robinson, P. B. Shepson, C. Sweeney, A. Townsend-Small, S. C. Wofsy and S. P. Hamburg. Assessment of methane emissions from the U.S. oil and gas supply chain. *Science*, vol. 361, no. 6398, pp. 186–188, Jul. 2018. Available at: https://doi.org/10.1126/science.aar7204

[9] D. R. Lyon, R. A. Alvarez, D. Zavala-Araiza, A. R. Brandt, R. B. Jackson and S. P. Hamburg. Aerial surveys of elevated hydrocarbon emissions from oil and gas production sites. *Environmental Science & Technology*, vol. 50, no. 9, pp. 4877–4886, May 2016. Available at: https://doi.org/10.1021/acs.est.6b00705

[10] D. H. Cusworth, R. M. Duren, A. K. Thorpe, W. Olson-Duvall, J. Heckler, J. W. Chapman, M. L. Eastwood, M. C. Helmlinger, R. O. Green, G. P. Asner, P. E. Dennison and C. E. Miller. Intermittency of large methane emitters in the Permian Basin. *Environmental Science & Technology Letters,* vol. 8, no. 7, pp. 567–573, Jul. 2021. Available at: https://doi.org/10.1021/acs.estlett.1c00173

[11] A. Anand, S. Riddick, K. B. Shonkwiler, A. ti, M. Moy, E. Kiplimo, M. Mbua and D. J. Zimmerle. Computational fluid dynamics-based modeling of methane flows around oil and gas equipment. *Atmosphere*, vol. 16, no. 7, p. 811, Jul. 2025. Available at: https://doi.org/10.3390/atmos16070811

[12] T. K. Flesch, J. D. Wilson and E. Yee. Backward-time Lagrangian stochastic dispersion models and their application to estimate gaseous emissions. *Journal of Applied Meteorology*, vol. 34, no. 6, pp. 1320–1332, Jun. 1995. Available at: https://doi.org/10.1175/1520-0450(1995)034<1320:BTLSDM>2.0.CO;2

[13] S. M. McGinn, K. A. Beauchemin, T. K. Flesch and T. Coates. Performance of a dispersion model to estimate methane loss from cattle in pens. *Journal of Environmental Quality*, vol. 38, no. 5, pp. 1796–1802, Sep. 2009. Available at: https://doi.org/10.2134/jeq2008.0531

[14] S. N. Riddick, R. Ancona, M. Mbua, C. S. Bell, A. Duggan, T. L. Vaughn and K. Bennett. A quantitative comparison of methods used to measure smaller methane emissions typically observed from superannuated oil and gas infrastructure. *Atmospheric Measurement Techniques*, vol. 15, no. 21, pp. 6285–6296, Nov. 2022. Available at: https://doi.org/10.5194/amt-15-6285-2022

[15] D. Ball, U. Ismail, N. Eichenlaub, N. Metzger and A. Lashgari. Performance evaluation of multi-source methane emission quantification models using fixed-point continuous

monitoring systems. *EGUsphere* [preprint], 2025. Available at: https://doi.org/10.5194/egusphere-2025-1266

[16] D. C. Blackmore, J.-P. Hickey, A. Wigle, K. Osadetz and K. J. Daun. A Bayesian technique for quantifying methane emissions using vehicle-mounted sensors with a Gaussian plume model. *Atmospheric Environment*, vol. 344, p. 121002, Mar. 2025. Available at: https://doi.org/10.1016/j.atmosenv.2024.121002

[17] M. Jia, W. S. Daniels and D. M. Hammerling. Comparison of the Gaussian plume and puff atmospheric dispersion models for methane modeling on oil and gas sites. *ChemRxiv* preprint, May 03, 2023. Available at: https://doi.org/10.26434/chemrxiv-2023-hc95q-v2

[18] S. R. Hanna, G. A. Briggs and R. P. Hosker Jr. *Handbook on atmospheric diffusion*. DOE/TIC-11223, 5591108, Jan. 1982. Available at: https://doi.org/10.2172/5591108

[19] D. B. Turner. *Workbook of atmospheric dispersion estimates*. Public Health Service Publication No. 999-AP-26, United States Department of Health, Education, and Welfare, Public Health Service, National Air Pollution Control Administration, Cincinnati, Ohio, revised 1970. Available through the U.S. Environmental Protection Agency National Environmental Publications Information System (NEPIS) and the National Technical Information Service (NTIS), Accession No. PB191482. Available at: https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=9100JEIO.TXT:

[20] United States Environmental Protection Agency (US EPA). Meteorological monitoring guidance for regulatory modeling applications. EPA-454/R-99-005. Research Triangle Park, North Carolina: Office of Air Quality Planning and Standards, Office of Air and Radiation, February 2000. Available at: https://www.epa.gov/sites/default/files/2020-10/documents/mmgrma_0.pdf

[21] ABB Inc. Micro-portable greenhouse gas analyzer (GLA131-GGA) datasheet – Rev. 20J. ABB Measurement & Analytics, 2024. Available at: https://library.e.abb.com/public/0fa167abde3c42f6b88668387e25c5e9/3KXG167001R1001_J_GLA131-GGA%20Datasheet-EN%20Rev.%20J.pdf

[22] United States Environmental Protection Agency (US EPA). OTM 33A – Point Source Gaussian (PSG) analysis standard operating procedure. U.S. EPA Office of Research and Development, National Risk Management Research Laboratory, September 2013 (Revision 0). Available at: https://www.epa.gov/sites/default/files/2020-08/documents/otm_33a_appendix_f1_psg_analysis_sop.pdf

[23] D. T. Allen, V. M. Torres, J. Thomas, D. W. Sullivan, M. Harrison, A. Hendler, S. C. Herndon, C. E. Kolb, M. P. Fraser, A. D. Hill, B. K. Lamb, J. M. Mohan, R. F. Robinson, J. E. Stone, E. T. Szwec and R. L. Yacovitch. Measurements of methane emissions at natural gas production sites in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 44, pp. 17768–17773, Oct. 2013. Available at: https://doi.org/10.1073/pnas.1304880110

[24] D. J. Zimmerle, T. L. Vaughn, C. M. Bell, M. L. Bennett, G. A. Duggan, B. K. Harrison, J. Schneider, A. L. Robertson, M. K. Heath, D. R. Lyon, S. P. Hamburg, A. R. Brandt and R. A. Alvarez. Methane emissions from the natural gas transmission and storage system in the United States. *Environmental Science & Technology*, vol. 49, no. 15, pp. 9374–9383, Aug. 2015. Available at: https://doi.org/10.1021/acs.est.5b01669

[25] J. M. Stockie. The mathematics of atmospheric dispersion modeling. *SIAM Review*, vol. 53, no. 2, pp. 349–372, Jan. 2011. Available at: https://doi.org/10.1137/10080991X

[26] J. H. Seinfeld and S. N. Pandis. *Atmospheric chemistry and physics: from air pollution to climate change*. Third edition. Hoboken, New Jersey: John Wiley & Sons, 2016.

[27] United States Environmental Protection Agency (US EPA). Air quality dispersion modeling – alternative models. Environmental Protection Agency, 2025. Accessed: Oct. 07, 2025. Available at: https://www.epa.gov/scram/air-quality-dispersion-modeling-alternative-models

[28] S. N. Riddick, M. Mbua, C. Laughery and D. J. Zimmerle. A review of offshore methane quantification methodologies. *Atmospher*e, vol. 16, no. 5, p. 626, May 2025. Available at: https://doi.org/10.3390/atmos16050626

[29] T. K. Flesch, J. D. Wilson, L. A. Harper, B. P. Crenna and R. R. Sharpe. Deducing ground-to-air emissions from observed trace gas concentrations: a field trial. *Journal of Applied Meteorology*, vol. 43, no. 3, pp. 487–502, Apr. 2004. Available at: https://doi.org/10.1175/1520-0450(2004)043<0487:DGEFOT>2.0.CO;2

[30] T. K. Flesch, S. M. McGinn, D. Chen, J. D. Wilson and R. L. Desjardins. Data filtering for inverse dispersion emission calculations. *Agricultural and Forest Meteorology*, vol. 198–199, pp. 1–6, Nov. 2014. Available at: https://doi.org/10.1016/j.agrformet.2014.07.010

[31] B. Crenna. An introduction to WindTrax. Thunder Beach Scientific, Ontario, Canada, 2006. Available at: http://thunderbeachscientific.com/

[32] J. C. Chang and S. R. Hanna. Air quality model performance evaluation. *Meteorology and Atmospheric Physics*, vol. 87, no. 1, pp. 167–196, Sep. 2004. Available at: https://doi.org/10.1007/s00703-003-0070-7

[33] S. R. Hanna, J. C. Chang and D. G. Strimaitis. Hazardous gas model evaluation with field observations. *Atmospheric Environment Part A: General Topics*, vol. 27, no. 15, pp. 2265–2285, Oct. 1993. Available at: https://doi.org/10.1016/0960-1686(93)90397-H

[34] Atmospheric Dispersion Modelling Liaison Committee (ADMLC). Annual report 1996/97. NRPB-R302. Chilton, United Kingdom: National Radiological Protection Board, 1999. Available at: https://admlc.com/wp-content/uploads/2014/09/r302.pdf

[35] South Coast Air Quality Management District (SCAQMD). Meteorological data for AERMOD. *South Coast AQMD*, 2025. Accessed: Oct. 07, 2025. Available at: https://www.aqmd.gov/home/air-quality/meteorological-data/data-for-aermod

[36] A. J. Prussin II, L. C. Marr, D. G. Schmale III, R. Stoll, and S. D. Ross. Experimental validation of a long-distance transport model for plant pathogens: Application to Fusarium graminearum. *Agricultural and Forest Meteorology*, vol. 203, pp. 118–130, Jan. 2015. Available at: https://doi.org/10.1016/j.agrformet.2014.12.009

[37] D. E. Aylor and T. K. Flesch. Estimating spore release rates using a Lagrangian stochastic simulation model. *Journal of Applied Meteorology*, vol. 40, no. 7, pp. 1196–1208, Jul. 2001. Available at: https://doi.org/10.1175/1520-0450(2001)040<1196:ESRRUA>2.0.CO;2

[38] A. Ražnjević, D. Argyriou, T. K. Flesch, and C. Feigenwinter. Evaluation of two common source estimation measurement strategies using large-eddy simulation of plume dispersion under neutral atmospheric conditions. *Atmospheric Measurement Techniques*, vol. 15, no. 7, pp. 3611–3628, Jul. 2022. Available at: https://doi.org/10.5194/amt-15-3611-2022

[39] S R. Hanna. Air Quality Model Evaluation and Uncertainty. *Journal of Air Pollution Control Association*, vol. 38, issue 4 pp. 406-412, 1988. Available at: https://doi.org/10.1080/08940630.1988.10466390

[40] U.S. Environmental Protection Agency. User's Guide for the Industrial Source Complex (ISC3) Dispersion Models, Volume II: Description of Model Algorithms. EPA-454/B-95-003. Washington, DC: U.S. Environmental Protection Agency, 1995. Available at: https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=0000324F.PDF

APPENDIX

Appendix A PGSC Definition, σ_y and σ_z Calculations

Table A1: Estimation of stability class using wind speed and incoming solar strength (for daytime values from Table 3-1 in [19] and Table 6-3 in [20])

Stability Class	Day		
WS (m s^{-1})	Strong	Moderate	Light
<2	A	A	B
2-3	B	B	C
3-5	B	C	C
5-6	C	C	D
>6	C	D	D

Table A2: Parameters used to calculate σ_y, where x is in kilometers, Table 1-1 in [40].

$$\sigma_y = 465.11628\ (x) \tan (TH)$$
$$TH = 0.017453293\ [c - d \ln (x)]$$

PGSC Category	c	d
A	24.1670	2.5334
B	18.3330	1.8096
C	12.5000	1.0857
D	8.3330	0.72382
E	6.2500	0.54287
F	4.1667	0.36191

Table A3: Parameters used to calculate σ_z, where x is in kilometers.
* If the calculated value of σ_z exceeds 5000, σ_z is set to 5000 m, Table 1-2 in [40]

PGSC Category	σ_z (meters) = ax^b (x in km)		
	x (km)	a	b
A	<0.10	122.80	0.9447
	0.10-0.15	158.08	1.0542
	0.16-0.20	170.22	1.0932
	0.21-0.25	179.52	1.1262
	0.26-0.30	217.41	1.2644
	0.31-0.40	258.89	1.4094
	0.41-0.50	346.75	1.7283

	0.51-3.11	453.85	2.1166
	>3.11	*	*
B	<0.20	90.673	0.93198
	0.21-0.40	98.483	0.98332
	>0.40	109.300	1.09710
C	All	61.141	0.91465
D	<0.30	34.459	0.86974
	0.31-1.00	32.093	0.81066
	1.01-3.00	32.093	0.64403
	3.01-10.00	33.504	0.60486
	10.01-30.00	36.650	0.56589
	>30.00	44.053	0.51179
E	<0.10	24.260	0.83660
	0.10-0.30	23.331	0.81956
	0.31-1.00	21.628	0.75660
	1.01-2.00	21.628	0.63077
	2.01-4.00	22.534	0.57154
	4.01-10.00	24.703	0l.50527
	10.01-20.00	26.970	0.46713
	>40.00	35.420	0.37615
F	<0.20	15.209	0.81558
	0.21-0.70	14.457	0.78407
	0.71-1.00	13.953	0.68465
	1.01-2.00	13.953	0.6327
	2.01-3.00	14.823	0.54503
	3.01-7.00	16.187	0.46490
	7.01-15.00	17.836	0.4150
	15.01-30.00	22.651	0.32681
	30.01-60.00	27.027	0.2436
	>60.00	43.219	0.2176

Appendix B WindTrax Configuration

This section describes the WindTrax (Thunder Beach Scientific, Nova Scotia, Canada) configuration parameters used in this study for the bLS modeling. The configuration described below includes input variables, particle release conditions, and model integration parameters.

Table B1: Model domain and setup

Parameter	Description
Model type	LS model
Source type	Point source
Source coordinates	(0, 0, H) where H is a release height
Sensor coordinates	(x, y, z) where x = plume centerline distance, y = crosswind distance, and z = inlet height of sensor (1 m)
Surface roughness	2.3 cm
Terrain	Uniform
Particle count	500,000 initially (increased up to 20M if error resulted)
Release species	CH_4 in Kg h^{-1}
Measured species	CH_4 in ppmv
Collection volume type	Fixed
Collection box height	0.05 m
Collection box width	0.05 m
Random number generation	Precalculated
Stability	Directly specified (Pisquill-Gifford class input)
Tower coordinates	(x, y) meter distance from the source at 10 m height
Anemometer height	6.7 m
T/P/RH sensor height	6.2 m
Surface elevation	4982 ft

Appendix C Meteorological Data During Experiments

Table C1: Summary of meteorological conditions for single release measurements. Variables listed: mean and standard deviation of wind speed (U), air temperature (T_{air}), relative humidity (RH), mean of barometric pressure P (kPa), and sector (prevailing wind direction). The standard deviation of P is not mentioned because P never deviated more than 0.3 hPa. *Number of samples (n = 121) reflects the final total filtered for single releases (low winds: U < 1 ms^{-1} excluded).

Exp #	Date (2024)	*U* (m/s)	Sector	T_{air} (°C)	P (hPa)	RH (%)	#Samples*
1	3/19	2.0 ± 0.4	SE	10.1 ± 0.8	816.5 ± 0.0	30.6 ± 1.1	6
2	3/21	2.1 ± 0.3	SE	11.0 ± 0.9	812.1 ± 0.1	30.3 ± 3.7	6
3	4/23	3.8 ± 0.6	SE	13.3 ± 0.3	844.7 ± 0.2	31.1 ± 0.5	5
4	6/11	1.6 ± 0.2	SE	21.9 ± 0.5	836.0 ± 0.1	21.9 ± 0.5	4
5	6/17	4.7 ± 0.5	SE	26.1 ± 0.4	822.6 ± 0.2	36.3 ± 0.4	2
6	6/18	4.6 ± 0.7	SE	16.6 ± 0.5	832.8 ± 0.1	28.0 ± 0.4	2
7	6/19	4.5 ± 0.4	SE	16.0 ± 0.2	841.8 ± 0.1	46.7 ± 2.8	4
8	6/19	6.2 ± 0.3	SE	18.7 ± 0.2	839.5 ± 0.1	37.0 ± 2.7	1
9	7/9	2.2 ± 0.7	NE	27.4 ± 0.4	835.6 ± 0.1	15.1 ± 0.6	6
10	7/10	3.1 ± 0.4	SE	28.2 ± 0.0	836.9 ± 0.1	15.5 ± 0.9	6
11	7/11	3.0 ± 0.5	NE	32.9 ± 0.3	833.6 ± 0.1	10.3 ± 0.2	5
12	7/17	4.2 ± 0.6	SE	28.5 ± 0.4	842.5 ± 0.2	22.3 ± 0.9	5
13	7/18	2.5 ± 0.2	SE	24.1 ± 0.4	842.9 ± 0.1	36.5 ± 2.1	6
14	7/18	3.0 ± 0.4	SE	27.3 ± 0.5	842.0 ± 0.1	31.2 ± 1.3	5
15	8/26	2.1 ± 0.7	SE	20.0 ± 0.2	839.9 ± 0.1	58.5 ± 0.9	7
16	9/16	1.5 ± 0.3	NE	23.0 ± 0.6	825.9 ± 0.0	25.7 ± 0.9	6
17	9/16	1.8 ± 0.4	SE	25.0 ± 0.3	825.1 ± 0.1	22.9 ± 1.4	6
18	9/16	3.7 ± 0.3	SE	25.6 ± 0.3	823.9 ± 0.1	25.6 ± 0.1	6
19	9/17	3.3 ± 0.8	SE	23.5 ± 0.5	820.7 ± 0.1	28.0 ± 0.9	6
20	9/17	2.0 ± 0.4	SE	24.5 ± 0.6	818.9 ± 0.3	27.6 ± 2.0	6
21	9/18	2.4 ± 1.1	SE	22.1 ± 2.8	821.3 ± 1.0	26.9 ±12.0	5*
22	9/19	4.4 ± 0.2	SE	19.6 ± 0.3	827.6 ± 0.1	21.7 ± 0.3	6
23	9/19	4.1 ± 0.3	SE	21.0 ± 0.4	827.1 ± 0.1	21.1 ± 1.0	6
24	9/20	4.8 ± 0.5	SE	24.7 ± 0.2	823.5 ± 0.1	17.2 ± 1.0	6

Appendix D PGSC Definition By Mean Solar Radiation and Wind Speed

Table D1: Solar Radiation Delta-T (SRDT) approach for Pasquill-Gifford (P-G stability) from table 6-7 in [20]. The experiment was conducted in daytime, so nighttime information was not taken into account for analysis.

Daytime				
	Solar Radiation (W/m^2)			
Wind Speed ($m\ s^{-1}$)	≥925	925-675	675-175	<175
<2	A	A	B	D
2-3	A	B	C	D
3-5	B	B	C	D
5-6	C	C	D	D
≥6	C	D	D	D
Nighttime				
	Vertical Temperature Gradient			
Wind Speed ($m\ s^{-1}$)	<0		≥0	
<2	E		F	
2-2.5	D		E	
≥2.5	D		D	

Appendix E: List of Abbreviations And Acronyms

CH_4	Methane
NG	Natural Gas
GHG	Greenhouse Gas
O&G	Oil and Gas
IPCC	Intergovernmental Panel on Climate Change
US EPA	United States Environmental Protection Agency
GP	Gaussian Plume
BLS	Backward Lagrangian Stochastic Model
METEC	Methane Evaluation Technology Evaluation Center
MGGA	Micro-portable Greenhouse Gas Analyzer