



# Mobile sensing of point-source gas emissions using Bayesian inference: An empirical examination of the likelihood function

Xiaochi Zhou<sup>a,1</sup>, Amir Montazeri<sup>a,b</sup>, John D. Albertson<sup>a,\*</sup>

<sup>a</sup> School of Civil and Environmental Engineering, Cornell University, Ithaca, NY, 14853, USA

<sup>b</sup> Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY, 14853, USA

## ARTICLE INFO

### Keywords:

Bayesian inference  
Environmental sensing  
Mobile monitoring  
Methane

## ABSTRACT

This paper evaluates likelihood function forms for Bayesian inference of point-source gas emissions using a mobile sensor. Whereas Bayesian inference has been successfully used to estimate emission rates from time-averaged concentration data measured by stationary sensors, data collected by mobile sensors do not represent ensemble or time-averaged conditions. To examine the potential impact of this contrast, controlled release experiments were conducted with a mobile sensor measuring concentrations repeatedly along transverse cross sections of the downwind plumes. Experiments were conducted with measurements made at different downwind distances, different sensor heights, and with different obstacle states. An examination is made between two commonly-used likelihood functions, the Gaussian and the log-normal. For experiments conducted in the absence of obstacles, the Bayesian estimates using the log-normal likelihood function yield a much smaller bias than those based on the Gaussian likelihood function. This finding is consistent with the non-Gaussian nature of concentration fluctuations near a point-source. For experiments conducted in the presence of obstacles, the Bayesian inference based on the Gaussian likelihood function exhibits a better performance. This can be explained by the enhanced turbulent mixing due to the obstacle-introduced wake eddies. Overall, we find that the selection of the likelihood function can be physically related to the underlying conditions, and the proper selection is critical to ensure the performance of the Bayesian inference for source characterization using mobile sensing data.

## 1. Introduction

Mobile sampling of air quality is becoming popular as it efficiently offers coverage of an extended area of interest. Sensors have been mounted on ground vehicles to map air quality variables such as particulate matter, carbon dioxide, and carbon monoxide (Devarakonda et al., 2013; Marjovi et al., 2015; Apte et al., 2017; Lee et al., 2017) in urban areas, and mounted on aerial platforms to characterize combustion from open area sources (Aurell et al., 2011; Zhou et al., 2017). Due to the potentially high spatial resolution of the mobile sensing data, researchers are able to explore the link between human exposure to traffic-related air pollution and adverse health effects at community level (Alexeeff et al., 2018).

With the recent expansion of natural gas production, there is heightened interest in estimating fugitive methane emission rates from point sources. A good example is the recent effort to develop innovative technologies and algorithms to cost-effectively detect methane

emissions from oil and gas well pads (with a typical size of 10 m by 10 m), so as to target mitigation efforts (ARPA-E, 2015). In fact, fugitive natural gas emissions can jeopardize potential advantage of natural gas over coal with respect to greenhouse gas forcing per unit energy produced. (Alvarez et al., 2012). Motivated by concern for potential leaks in natural gas distribution systems, mobile sensing has been used along roadways to measure elevated methane levels (Phillips et al., 2013; Jackson et al., 2014; von Fischer et al., 2017), however, these efforts have primarily concluded with concentration mapping with no rigorous source characterization or leak localization. Several studies have used mobile sensing to estimate fugitive methane emissions from natural gas production sites (Yacovitch et al., 2015; Lan et al., 2015; Rella et al., 2015; Albertson et al., 2016). However, the estimated uncertainty reported by some studies has not been rigorously tied to the prevailing experimental conditions, but rather referenced to accuracy found in prior controlled release experiments conducted under different experimental conditions. Ideally, a mobile sensing method should provide

\* Corresponding author.

E-mail addresses: [xiaochi.zhou@arb.ca.gov](mailto:xiaochi.zhou@arb.ca.gov) (X. Zhou), [am2774@cornell.edu](mailto:am2774@cornell.edu) (A. Montazeri), [albertson@cornell.edu](mailto:albertson@cornell.edu) (J.D. Albertson).

<sup>1</sup> Current address: California Air Resources Board, Sacramento, CA, 95814, USA.

emission estimation with a reported uncertainty defined in terms of the actual sampling conditions (e.g. wind speed, atmospheric stability, and site complexity) as well as the instrumental characteristics (e.g. accuracy, precision, detection limit, and time-resolution), as needed to understand the value of potential additional measurements to reduce estimation uncertainty. Similarly, measurement efforts could then be designed based on relevant variables, so as to achieve the desired level of accuracy (Cai and Ferrari, 2009; Zhang et al., 2009). A recent effort proposed an information-driven algorithm that optimizes sensor paths to minimize the overall source estimation uncertainty (Albertson et al., 2016).

Source characterization using downwind measurements (either by stationary or mobile sensors) is essentially an inverse problem, with uncertainties stemming from both the plume dispersion model and the sensor data (Rao, 2005). Bayesian tools provide a logical way to incorporate both model and data uncertainties into source inference, thus allowing a rigorous quantification of the estimation uncertainty. Such an approach has been applied to estimate point source emission rates using downwind fixed (stationary) sensors (Keats et al., 2007; Yee and Flesch, 2010; Yee, 2012; Humphries et al., 2012), typically using time-averaged gas concentrations (e.g. averaged over 20–30 min). In contrast, a mobile sensor can traverse a plume in less than several seconds, thus providing more of an instantaneous picture of the plume. Albertson et al. (2016) applied a recursive Bayesian method to estimate point-source emissions using mobile sensor data. They used a Gaussian likelihood function, assuming the errors from both model and data are normally distributed, similar to those studies using ensemble-averaged data (e.g. Yee, 2012). The applicability of the Gaussian assumption (i.e. under what conditions it will hold/fail) remains to be examined experimentally or supported theoretically, for that matter. Logically, the form of the likelihood function is related to the concentration distribution, and the Gaussian likelihood function may fail close to the source where the concentrations often exhibit a marked non-Gaussian distribution (Csanady, 1973; Fackrell and Robins, 1982; Yee et al., 1993).

In this paper, we evaluate and compare the performance of Bayesian inference with two mostly used likelihood functions (the Gaussian and the log-normal), using data collected by a mobile sensor downwind from a controlled point-source emission. First, an instantaneous plume dispersion formulation is developed to guide the analysis, and then data collected from a series of controlled release experiments are used to evaluate the candidate approaches.

## 2. Technical background

Most point-source characterization approaches are based on time-averaged measurements. However, here it is necessary to introduce a formulation for the instantaneous plume to support the analysis of the mobile sensor data and the associated uncertainty. This formulation is valid for passive scalars, which are diffusive contaminants with low concentrations such that they pose no dynamical effects (i.e. via buoyancy) on the fluid flow (Warhaft, 2000).

### 2.1. An instantaneous view of plume transport

Consider a local coordinate system with a steady state, point-source located at the origin  $O$  (Fig. 1).  $u$ ,  $v$ , and  $w$  [m/s] are defined as the wind velocity components in  $x$ ,  $y$ , and  $z$  directions, respectively. A control volume is defined, from  $O$  to a downwind vertical plane  $x_m$ ,  $y_{min}$  to  $y_{max}$  laterally, and  $z_{min}$  to  $z_{max}$  vertically, to encompass the entire plume upwind of the mobile sensor (Fig. 1). From conservation of mass, the emission rate (mass per time),  $Q_0$ , can be expressed as:

$$Q_0 = F(x_m, t) + \frac{dS(t)}{dt}, \quad (1)$$

where  $S(t)$  is total mass of the emitted gas in the control volume,  $t$  is

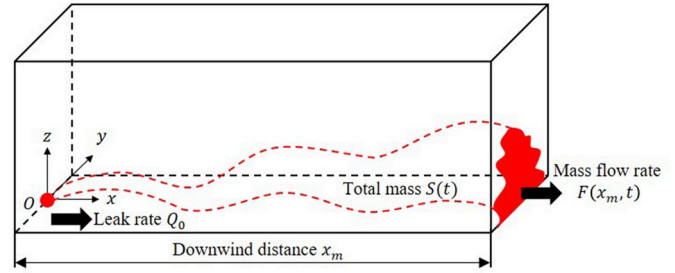


Fig. 1. A control volume containing an emission source (located at  $O$ ) with a mass flow rate of  $Q_0$ , and a cross-plane view of the plume mass flow rate at downwind distance  $x_m$ .

time, and  $F(x_m, t)$  is the mass flow rate exiting of the control volume. Provided that the control volume is defined such that no mass exits anywhere other than the downstream face, then

$$F(x_m, t) = \int_{z_{min}}^{z_{max}} \int_{y_{min}}^{y_{max}} c(x_m, y, z, t) u(x_m, y, z, t) dy dz, \quad (2)$$

where  $c$  is the mass concentration of the passive scalar [ $\text{g}/\text{m}^3$ ]. Note that molecular diffusion is ignored given the large Reynolds numbers typically observed in the atmospheric boundary layer (Stull, 1988).

It is instructive to define a normalized distribution of the mass concentrations ( $D$ , [ $\text{m}^{-2}$ ]) and a plume-weighted advection velocity ( $u_e$ , [ $\text{m}/\text{s}$ ]) at the exit face ( $x_m, y, z$ ) as

$$D(x_m, y, z, t) = \frac{c(x_m, y, z, t)}{\int_{z_{min}}^{z_{max}} \int_{y_{min}}^{y_{max}} c(x_m, y, z, t) dy dz} \quad (3a)$$

$$u_e(x_m, t) = \int_{z_{min}}^{z_{max}} \int_{y_{min}}^{y_{max}} D(x_m, y, z, t) u(x_m, y, z, t) dy dz. \quad (3b)$$

By substituting Equation (3a) into (3b) and using Equation (2), we can rewrite  $u_e(x_m, t)$  as:

$$u_e(x_m, t) = \frac{\int_{z_{min}}^{z_{max}} \int_{y_{min}}^{y_{max}} c(x_m, y, z, t) u(x_m, y, z, t) dy dz}{\int_{z_{min}}^{z_{max}} \int_{y_{min}}^{y_{max}} c(x_m, y, z, t) dy dz} = \frac{F(x_m, t)}{c(x_m, y, z, t)/D(x_m, y, z, t)}. \quad (4)$$

With Equation (4) and Equation (1), the mass concentrations trajectory when traversing the plume ( $c(x_m, y, z, t)$ ) can be related to the other important variables as:

$$c(x_m, y, z, t) = \frac{Q_0 - dS(t)/dt}{u_e(x_m, t)} D(x_m, y, z, t) \quad (5)$$

In practice,  $u_e(x_m, t)$  can be approximated with nearby meteorological measurements, adjusted as needed by height difference through the vertical scaling of the wind profile based on the Monin Obukhov Similarity Theory (MOST) (Obukhov, 1971), as detailed in Appendix A. Accordingly, we replace  $u_e(x_m, t)$  by  $u_e^M(x_m, t) \delta_u(x_m, t)$ , so that  $\delta_u(x_m, t)$  accounts for the ratio between the actual  $u_e(x_m, t)$  and the approximated  $u_e^M(x_m, t)$ . The superscript  $M$  stands for model estimated quantities. Similarly, we introduce  $\delta_S(x_m, t) = 1 - \frac{1}{Q_0} \frac{dS(t)}{dt}$  to represent the non-steadiness in the mass stored in the control volume, normalized by  $Q_0$ . With these non-stationary terms, Equation (5) becomes:

$$c(x_m, y, z, t) = \frac{Q_0}{u_e^M(x_m, t)} \left( \frac{\delta_S(x_m, t)}{\delta_u(x_m, t)} \right) D(x_m, y, z, t), \quad (6)$$

which describes the instantaneous plume transport while sharing an underlying form with other commonly used models based on the ensemble-averaged view (Horst and Weil, 1992). The main difference is the time dependence of the function  $D$  to represent the stochastic nature of the turbulent plume, and the presence of  $\delta_u(x_m, t)$  and  $\delta_S(x_m, t)$  to account for non-stationarity in wind speed and storage in the control

volume.

$D(x_m, y, z, t)$  is a random variable capturing plume movement in time as well as in the  $y$  and  $z$  dimensions, as it responds to the instantaneous turbulent velocity components in the  $y$  and  $z$  directions, which would scale with the respective standard deviations of velocity components ( $\sigma_y$  and  $\sigma_z$ ). It is well-understood that  $\sigma_w$  can be described well by local scaling based on MOST (Kaimal and Finnigan, 1994). However,  $\sigma_y$  is affected by random large scale motions in the atmosphere, which cannot be described accurately by local scaling laws (Lumley and Panofsky, 1964). Consequently, we can deduce that a greater degree of randomness exists in  $y$  than in  $z$  direction for  $D(x_m, y, z, t)$ .

This observation motivates the idea of integrating both sides of Equation (6) over  $y$  (i.e. across the plume), such that the uncertainty associated with lateral plume dispersion can be effectively removed (Albertson et al., 2016).

$$c_y(x_m, z, t) = \frac{Q_0}{u_e^M(x_m, t)} \left( \frac{\delta_S(x_m, t)}{\delta_u(x_m, t)} \right) D_y(x_m, z, t), \quad (7)$$

where  $c_y(x_m, z, t) = \int_{y_{min}}^{y_{max}} c(x_m, y, z, t) dy$  is the cross-plume integrated mass concentrations [ $g/m^3$ ]. In practice, the sensor path is typically constrained by roadways and is not always perpendicular to the wind direction. When the road segments and the wind direction are not perpendicular (i.e. the acute angle between them is  $\theta_r$ ),  $c_y(x_m, z, t)$  can be estimated by numerical integration of the mass concentrations along the path based on trigonometry ( $c_y(x_m, z, t) = \sum_{y_{min}}^{y_{max}} c(x_m, y, z, t) \Delta t V \sin(\theta_r)$ , where  $\Delta t$  is the sensor acquisition time step and  $V$  is the vehicle speed).  $D_y(x_m, z, t) = \int_{y_{min}}^{y_{max}} D(x_m, y, z, t) dy$  is a reflection of the vertical profile of mass concentrations at  $x_m$ , which is a random variable mainly driven by the stochastic nature of the vertical transport dynamics in the turbulent flow.

Given the stochastic nature of  $\delta_S(x_m, t)$ ,  $\delta_u(x_m, t)$ , and  $D_y(x_m, z, t)$ , it is useful to introduce a fluctuating  $D_{y,e}(x_m, z, t) = \frac{\delta_S(x_m, t)}{\delta_u(x_m, t)} D_y(x_m, z, t)$  to account for all stochasticity in  $c_y(x_m, z, t)$  and to submit itself to empirical analysis. Thus we rewrite Equation (7) as:

$$c_y(x_m, z, t) = \frac{Q_0}{u_e^M(x_m, t)} D_{y,e}(x_m, z, t). \quad (8)$$

Equation (8) becomes the descriptive equation for instantaneous plume transport after cross-plume integration. It can be used in a forward manner to estimate the downwind cross-plume integrated mass concentrations  $c_y(x_m, z, t)$  given emission rate  $Q_0$ , or in a inverse manner to infer  $Q_0$  based on measured  $c_y(x_m, z, t)$ . Since most models can only approximate the ensemble-averaged  $D_{y,e}(x_m, z, t)$ , we apply Bayesian inference to account for the fluctuation of  $D_{y,e}(x_m, z, t)$  from its ensemble mean (detailed in the following section). For simplicity of notation,  $x_m$ ,  $z$ , and  $t$  will be dropped hereafter.

## 2.2. Bayesian inference

Following Bayes's rule, the posterior probability distribution ( $p(Q|c_y)$ ) of the emission rate  $Q$  given the observation of  $c_y$  is (Yee, 2008; Albertson et al., 2016):

$$p(Q|c_y) = \frac{p(Q)p(c_y|Q)}{p(c_y)}, \quad (9)$$

where  $p(Q|c_y)$ ,  $p(Q)$ ,  $p(c_y|Q)$ , and  $p(c_y)$  are probability density functions (PDFs).  $p(Q)$  is the prior, is the likelihood function, and the  $p(c_y)$  is the evidence term that ensures  $p(Q|c_y)$  integrates to unity.

Prior to the proposed sampling activities, past measurements of similar facilities (e.g. Brantley et al., 2014) may be used to formulate  $p(Q)$ . Assuming that the prior knowledge of  $Q$  is limited to its lower and upper bounds (i.e.  $Q_{min}$  and  $Q_{max}$ , respectively), a uniform prior is adopted (Yee, 2007, 2008), which is usually considered as sufficiently uninformative based on the principle of maximum entropy (Jaynes,

1968). After the first sensor pass, we follow a recursive approach that takes the posterior PDF ( $p(Q|c_y)$ ) derived after the previous pass to be the prior PDF of the next pass (Albertson et al., 2016).

$$\begin{cases} p(Q) = 1/(Q_{max} - Q_{min}) & j = 1 \\ p(Q) = p(Q|c_y)_{j-1} & j > 1 \end{cases} \quad (10)$$

where  $j$  is a counter for successive sensor passes.

All the information provided by the measurement  $c_y$  about the unknown  $Q$  should be contained in the likelihood function ( $p(c_y|Q)$ ), which describes the probability of observing  $c_y$  given  $Q$  (Jaynes, 2003). Most studies adopted a Gaussian form of likelihood function (Keats et al., 2007; Yee, 2008; Yee and Flesch, 2010; Albertson et al., 2016), and the selection is based on the argument that the underlying distribution of the measurement is unknown and the principle of maximum entropy supports the application of the Gaussian distribution (Jaynes, 2003). When moving close to the source, however, plumes are mainly driven by meandering of the wind, thus gas concentrations often exhibit a non-Gaussian distribution with non-zero skewness and strong intermittency. For example, the exponential distribution was found to fit data collected in wind tunnel (Fackrell and Robins, 1982), as well as outdoor tracer release experiments with a downwind distance from 50 to 100 m (Mylyne and Mason, 1991). Meanwhile, Yee et al. (1993) suggested a log-normal distribution for gas concentrations collected in their short-range experiments (i.e. downwind distance at 25 and 50 m), and a gamma distribution for long-range experiments (i.e. downwind distance at 80 and 100 m). With the presence of an obstacle array, Davidson et al. (1995) recommended a clipped normal distribution for gas concentrations measured within the array.

Note that the above-mentioned distributions were observed with stationary sensors; the distribution of path-integrated mass concentrations ( $c_y$ ) is less explored. We propose to consider the widely-used Gaussian distribution as a candidate likelihood function, and as an alternative to allow non-zero skewness of  $c_y$ , we consider the log-normal distribution inspired by Yee et al. (1993).

$$p(c_y|Q) = \begin{cases} \frac{1}{\sigma_e^G \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{c_y - c_y^M(Q)}{\sigma_e^G}\right)^2\right) \\ \frac{1}{\sigma_e^{LN} c_y \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{\ln(c_y) - \ln(c_y^M(Q))}{\sigma_e^{LN}}\right)^2\right) \end{cases}, \quad (11)$$

where  $c_y^M(Q) = \frac{Q}{u_e^M} D_y^M$  is a model estimated value at some considered candidate  $Q$ .  $D_y^M$  is the estimated value of  $D_{y,e}$  based on a Lagrangian Stochastic Model (LSM). The LSM describes plume dispersion in a turbulent flow by modeling the paths of fluid particles driven by the random velocity field, which can be modeled by the generalized Langevin equation (Wilson and Sawford, 1996). By imposing the so-called well-mixed condition, we adopt Thomson's simplest solution for statistically stationary and horizontally homogeneous turbulence (Thomson, 1987). The inputs of the LSM are the meteorological measurements (i.e. friction velocity, surface roughness, standard deviation of  $u$  and  $w$ , and Obukhov length) and the estimated distance between the emission source and the sensor. Details of the LSM are reported in Appendix B.

$\sigma_e^G$  and  $\sigma_e^{LN}$  are the error (i.e. uncertainty) scale parameters, which can be estimated for the observed data:

$$\begin{cases} \sigma_e^G = \sqrt{\frac{1}{N-1} \sum_{j=1}^N (c_{y,j} - c_y^M(Q_0))^2} \\ \sigma_e^{LN} = \sqrt{\frac{1}{N-1} \sum_{j=1}^N (\ln(c_{y,j}) - \ln(c_y^M(Q_0)))^2} \end{cases}, \quad (12)$$

where  $N$  is the number of passes per experiment.

Both  $\sigma_e^G$  and  $\sigma_e^{LN}$  are estimated from the controlled release experiments and are known as prior for the Bayesian inference. When the error scale parameters are not known, they can be estimated using the error propagation method (Rao, 2005; Zhou et al., 2019). More

specifically, the error scale parameters are:  $\sigma_S^2 + \sigma_M^2 + \sigma_D^2$ , where  $\sigma_S^2$  represents error due to the stochastic nature of atmospheric plume dispersion,  $\sigma_M^2$  is related to the simplification of the plume dispersion model, and  $\sigma_D^2$  includes errors from the model input data. Here we assume that these errors are uncorrelated and random (Rao, 2005). The parameterization of  $\sigma_S$ ,  $\sigma_M$ , and  $\sigma_D$  will depend on the local meteorological conditions, the dispersion model, as well as the measurement data quality. A good example is shown in the Supplemental Materials of Zhou et al. (2019).

The selection of the likelihood function (Equation (11)) should depend on the distribution of  $c_y - c_y^M(Q_0)$  or  $\ln(c_y) - \ln(c_y^M(Q_0))$ , which includes uncertainties from both the measurement and the model. Data used to inform the analysis are reported in the Results and Discussion section (Fig. 3).

Equation (9) has no analytical solution therefore is solved numerically. We first discretize  $Q$  from  $Q_{min}$  to  $Q_{max}$  with a uniform  $\Delta Q$ , resulting in a series of candidate  $Q$  values to be considered. A small  $\Delta Q$  is used, i.e.  $\Delta Q = (Q_{max} - Q_{min})/10000$ , to reduce the discretization error. With a measurement of  $c_y$ , the prior ( $p(Q)$ ) and the likelihood function ( $\text{}$ ) can be evaluated for each candidate  $Q$  using Equations (10) and (11), respectively. After evaluating all candidate  $Q$  values, the evidence term can be calculated using numerical integration:  $p(c_y) = \int_{Q_{min}}^{Q_{max}} p(Q)p(c_y|Q)dQ$ . After estimating all these terms (i.e. prior, likelihood, and the evidence), we can calculate the posterior using Equation (9), evaluated at each candidate  $Q$  value. After making an additional mobile pass, the posterior is updated by re-applying the above-mentioned numerical procedure. A new prior (i.e. the posterior derived after the previous mobile pass, as shown in Equation (10)) will be used, and the likelihood function will be evaluated using a newly measured  $c_y$ .

After the  $j^{\text{th}}$  pass, the emission rate and the associated uncertainty can be estimated from the posterior PDF  $p(Q|c_y)_j$ . For example, the emission rate can be calculated as the mean, median, or mode of the posterior PDF  $p(Q|c_y)_j$ . Considering that a uninformative was adopted, and the median and mean are heavily affected by the prior in the early stage of analysis, we use the mode of the posterior PDF  $p(Q|c_y)_j$  to represent the estimated emission rate in this study:

$$E_j^Q = \arg \max_Q \left( p(Q|c_y)_j \right). \quad (13)$$

After performing additional mobile passes, the effects of the prior will be gradually reduced and the difference among the mean, median, and mode of  $p(Q|c_y)_j$  will become small. When an informative prior can be derived from the past experiments, the mean or median of the posterior PDF  $p(Q|c_y)_j$  may be better incorporate the prior information.

Meanwhile, the associated uncertainty of the emission rate estimation is often calculated as the standard deviation ( $\sigma_j^Q$ ) of  $p(Q|c_y)$ :

$$(\sigma_j^Q)^2 = \int (Q - \bar{Q}_j)^2 \times p(Q|c_y)_j dQ, \quad (14)$$

where  $\bar{Q}_j = \int Q \times p(Q|c_y)_j dQ$  is the expectation of the posterior PDF.

### 2.3. Model performance assessment

After the final pass ( $N^{\text{th}}$ ), the accuracy of the estimated emission rate ( $E_N^Q$ ) can be assessed against the controlled release rate ( $Q_0$ ), e.g. the relative error  $R = (E_N^Q - Q_0)/Q_0$ . However, quantitative assessment of the accuracy of the modeled uncertainty ( $\sigma_N^Q$ ) is not well defined in the literature. It is most common to simply check whether  $Q_0$  is within the range of  $E_N^Q \pm \sigma_N^Q$  by visual inspection (Yee, 2008, 2012; Humphries et al., 2012; Keats et al., 2007). Among other reasons, this may mistakenly favor methods that tend to overestimate uncertainty, such that  $E_N^Q \pm \sigma_N^Q$  is wide enough to include  $Q_0$ .

To overcome this problem, we propose a method to infer the accuracy of  $\sigma_N^Q$  by assessing the posterior PDF,  $p(Q|c_y)$ . We first estimate the empirical distribution of the relative error,  $p(R_e) = (Q_0 - E_N^Q)/Q_0$ ,

obtained from experiments conducted under similar experimental conditions (e.g. similar obstacle condition). The distribution of modeled relative error,  $p(R_m)$ , can be readily derived from the posterior PDF  $p(Q|c_y)$  by substituting  $Q$  with  $R_m = (Q - E_N^Q)/E_N^Q$ .  $R_m$  becomes the relative error of  $Q$  from the estimated emission rate  $E_N^Q$ , and the newly constructed  $p(R_m)$  represents the PDF of modeled relative error. The distribution of the averaged  $R_m$  from experiments under similar conditions, denoted as  $p(\bar{R}_m)$ , was derived to represent the modeled relative error under those conditions.

Ideally,  $p(\bar{R}_m)$  should follow  $p(R_e)$  closely, and the agreement between them can be assessed by the p-value from the Kolmogorov-Smirnov (KS) test (Massey, 1951). Since  $p(R_m)$  is directly derived from the posterior PDF,  $p(Q|c_y)$ , its shape is mostly controlled by  $\sigma_N^Q$ . Therefore, we consider the p-value from the KS test to be a good indicator of the accuracy of  $\sigma_N^Q$ .

### 3. Experiments

Controlled methane release experiments were conducted at the McGovern soccer training field of Cornell University (Game Farm Rd, Ithaca, NY, USA) during the middle of June and early August 2016. The site was covered with short grass (~5 cm) and located in a relatively open field: ~500 m from a residential area in the west, ~150 m from a small forest in the north, and around ~400 m (500 m) from roads on the east (south) side. A point-source emission of methane (99.9% pure gas) was controlled by a mass flow controller (SmartTrak 100 from Sierra Instruments Inc., Monterey, CA, USA), with a mass-flow rate accuracy of  $\pm 1\%$ . The height of the release was similar to the height of grass (~5 cm). Two small towers were set-up on the east and west side of the field to measure local meteorological conditions. On each tower, a 3D sonic anemometer (CSAT-3, Campbell Scientific Inc., Logan, UT, USA) was installed to measure the three components of the wind velocity and air temperature at 10 Hz. The heights of the 3D sonic anemometers were 2.24 m and 2.21 m in the experiments conducted in June, and 2.49 m and 2.31 m in the experiments conducted in August. For experiments conducted in June, no local obstacles were present between emission source and sensor, and the field was quite open. This is designed to mimic emissions from an isolated wellhead (e.g. from the lower-level valves that are close to the ground). For experiments conducted in August, a 1.4 m high barrier (windbreak) was established in a circle around the methane source. The radius of the circle was ~3 m, and the source was placed in the center of the circle. This is to approximate a emission source surrounded by other low-level structures, for example, a wellhead located in a densely organized well pad or a pipeline within a small natural gas metering station.

A mobile measurement platform (MMP) was configured with a precise GPS unit (Trimble Geo 7X handheld from Trimble Inc., Sunnyvale, CA, USA) to track its position at a sampling frequency of 1 Hz. The GPS unit has an accuracy of 5 – 15 cm for > 97% data points after post-processing. The MMP was equipped with two methane analyzers: a Picarro G2204 cavity ring-down spectrometer (Picarro, Inc., Santa Clara, CA, USA) and a LI-COR LI-7700 open-path methane analyzer (LI-COR Biosciences, Lincoln, NE, USA). Both analyzers output methane mixing ratios in the unit of parts per million (ppm). The Picarro and the LI-COR analyzers were operated at a frequency of 1 Hz and 10 Hz, respectively. The inlet of the Picarro analyzer was attached to the MMP front at a height of 0.3 m, and the LI-COR analyzer was positioned at a height of 1.3 m with its path oriented horizontally. A multi-point calibration was performed before the experiment for the Picarro analyzer, and the calibration curve (slope 1.03 and intercept -0.04 ppm) was applied to process the raw data. The LI-COR analyzer was calibrated by the manufacturer less than a month before the experiment, and is designed with an open-path configuration for long-term monitoring without regular re-calibration. The time delay due to flow in the sampling line of the Picarro analyzer was measured, and the Picarro data were adjusted accordingly to be compatible with the LI-

**Table 1**

A summary of sampling day of year (DOY), and the time-averaged ambient methane mixing ratios measured prior to the experiments ( $\bar{c}_p$ ) by both Picarro and Licor analyzers.

DOY	Licor $\bar{c}_p$ [ppm]	Picarro $\bar{c}_p$ [ppm]
169	1.99	1.97
171	1.97	1.94
214	2.00	2.00
215	2.03	2.02
216	1.99	1.99
217	2.02	2.03
218	1.97	1.98
219	2.00	2.03

COR data. An inter-comparison between the Picarro and Licor analyzers is performed by concurrently measuring the ambient methane mixing ratios prior to the controlled release experiments ( $c_p$ , in ppm). In Table 1, we report the time-averaged  $c_p$  over a span of about 5 min, denoted as  $\bar{c}_p$ . Despite small daily variability,  $\bar{c}_p$  reported by the two analyzers are in good agreement.

A conversion factor is applied to translate the above-ambient mixing ratios ( $c_a$ , in ppm) into mass concentration  $c$  (in  $\text{g}/\text{m}^3$ ) in Equation (2), considering the molecular weight of methane (16.04  $\text{g}/\text{mol}$ ) and the air temperature that affects the molar volume of gas. The above-ambient mixing ratios ( $c_a$ ) is calculated as:  $c_a = c_r - c_b$ , where  $c_r$  is the raw methane mixing ratios measured by the analyzers and  $c_b$  is the ambient methane mixing ratios.  $c_b$  was estimated as the 5th percentile of the ranked time series of  $c_r$  (Brantley et al., 2014; Foster-Wittig et al., 2015). The estimated  $c_b$  is compared to the measured methane mixing ratios prior to the experiments ( $c_p$ ) and the difference is minimal (< 2%), suggesting that the determination of  $c_b$  is robust.

Stake flags are placed in three circles that are centered at the source with radius of 10, 20, and 30 m, respectively. Repeated passes were conducted along these circles such that sensor paths are almost perpendicular to the wind direction. The averaged travel speed is very low (around 2 m/s) in order to better capture the plume structure. Data was aggregated within 30 min periods, during which the meteorological parameters, such as the Obukhov length ( $L$ ) and the friction velocity ( $u_*$ ) were derived from data collected by the two met towers. We consider the set of passes during a single 30-min period to be a single experiment.

### 3.1. Data quality control

A total of 36 experiments were conducted, half of which with the obstacle present. Twelve experiments were conducted at each sensor-to-source distance ( $x_m$  of 10, 20, and 30 m). Two data quality requirements were established to filter out unacceptable experimental conditions. The first one admitted only experiments conducted under neutral or unstable atmospheric condition, thus excluding four experiments conducted under stable atmospheric condition. The second one excluded two experiments conducted under low wind ( $\bar{u} < 1.0$  m/s) and high turbulent intensity ( $I_u > 0.5$ ) conditions. Additionally, one experiment was excluded due to battery failure and data loss. Consequently, 29 experiments are available for further analysis, and their experimental conditions are summarized in Table 2. Meteorological conditions measured by the two towers were very similar due to their close proximity. Therefore, only the meteorological conditions reported by the west side of the tower, which is closer to the emission source, are reported in Table 2 and used in the analysis.

## 4. Results and Discussion

Before looking across all the experiments, we explore one experiment (ID # is 8 in Table 2) as an example. The above-ambient methane

mixing ratio ( $c_a$ ) measured along two sensor passes are plotted in Fig. 2. The plume position differs between passes due to the meandering of the wind. Also, mixing ratios measured by sensors located at different heights show different shapes. This is mainly caused by the vertical structure of the instantaneous plume, as well as the difference in sampling frequencies for the two analyzers (i.e. 10 Hz and 1 Hz for the Licor and Picarro analyzers that are installed at 1.3 m and 0.3 m, respectively).

The cross-plume integrated mass concentrations,  $c_y$ , is calculated by numerically integrating the geo-referenced mass concentrations along each pass. Then, we plot the distribution of  $c_y - c_y^M(Q)$ , represented as the probability mass function (PMF), for experiments conducted under similar conditions (e.g., similar sensor-to-source distance and obstacles) in Fig. 3. The PMF is fitted using both Gaussian and log-normal functions, and the p-value based on the KS test is calculated between the PMF and the best-fit curves. The results obtained at  $z_m = 0.3$  m are similar thus not shown here.

Without the presence of obstacles, the distributions of  $c_y - c_y^M(Q_0)$  exhibit a non-Gaussian shape with a long tail, and the skewness becoming less pronounced with a greater sensor-to-source distances ( $x_m$ ). This is supported by the better fitting capability using a non-Gaussian curve (e.g. log-normal) than a Gaussian, with a larger p-value obtained from applying the KS test (Fig. 3a, c, and e). This is attributed to the plume becoming better mixed by turbulence over the longer travel distance, which effectively reduces the skewness of  $c_y - c_y^M(Q_0)$ . On the other hand, the presence of obstacles continuously injects wake eddies that strongly mixed the plume and reduced the variability of  $c_y - c_y^M(Q_0)$  (Davidson et al., 1995, 1996). The distributions of  $c_y - c_y^M(Q_0)$  in the presence of obstacles more closely resemble a Gaussian distribution than a non-Gaussian, except for the case with  $x_m = 20$  m (Fig. 3b, d, and f). As shown in Equation (11), the distributions of ( $c_y - c_y^M(Q)$ ) should determine the form of the likelihood function and the error scale parameter ( $\sigma_e$ ). These observations motivate and intuitively inform the selection of different likelihood functions for Bayesian inference, which will be supported by further analysis detailed in the following sections.

### 4.1. Emission rate estimation using Bayesian inference

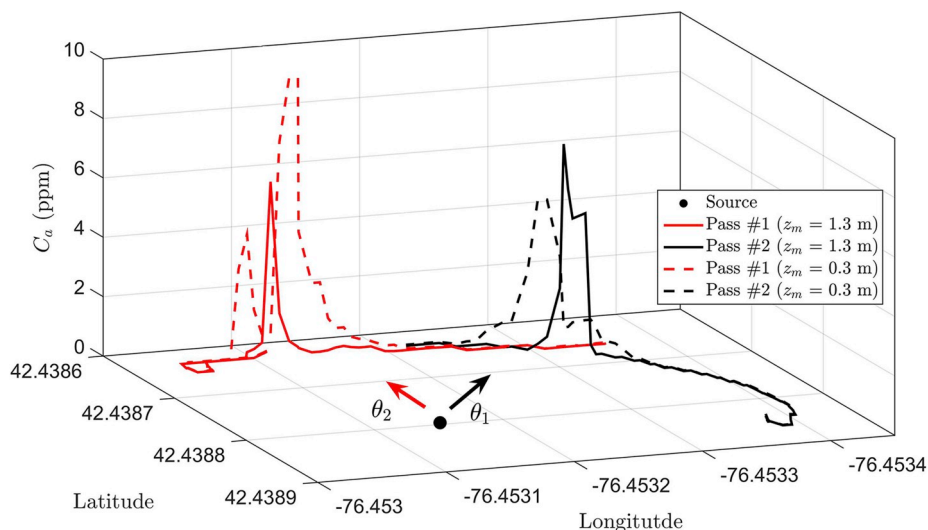
We first show emission rate estimation for one experiment (Experiment ID is 8 in Table 2) using both log-normal and Gaussian likelihood functions, as an example (Fig. 4). The lower and upper bounds of  $Q$ , denoted as  $Q_{min}$  and  $Q_{max}$  in Equation (10), are specified as 0 and 2.0 g/s.  $Q_{min}$  is indeed the lower bound since  $Q \geq 0$ .  $Q_{max}$  is determined by trail-and-error such that the tail of the derived posterior PDF ( $p(Q|c_y)$ ) is close to zero. Using a larger  $Q_{max}$  will not affect the accuracy of the Bayesian inference, but it is unnecessary and will increase the computational demand for the Bayesian inference. As shown in Fig. 4, the  $p(Q|c_y)$  is fairly small at  $Q = 0.3$  g/s, suggesting the determination of  $Q_{max} = 2.0$  g/s is effective. For both log-normal and Gaussian likelihood functions, the  $p(Q|c_y)$  tends to “sharpen” with additional sensor passes, from a relatively broad PDF (representing a large uncertainty of  $Q$ ) to a more narrow PDF (representing a low uncertainty of  $Q$ ). Fig. 4a shows that the posterior PDF with a log-normal likelihood function becomes fairly symmetrical (skewness < 0.5) after the final pass, suggesting that standard deviation is a good measure for the estimation uncertainty. Comparing against the controlled emission rate ( $Q_0$ ), the utilization of the log-normal likelihood function successfully enables the mode of  $p(Q|c_y)$  to approach  $Q_0$ . In contrast, the mode of the PDF using Gaussian likelihood function overestimates  $Q_0$ .

After examining the example case, we evaluate across all the experiments. For each experiment, two results obtained using the sensor located at  $z_m$  of 1.3 m and 0.3 m are color coded in red and black, respectively (Fig. 5). After the final sensor pass (the  $N^{th}$  pass), we plot  $E_N^Q \pm \sigma_N^Q$  against  $Q_0$  for all experiments (Fig. 5), with  $E_N^Q$  and  $\sigma_N^Q$  estimated using Equation (13) and (14) respectively. Overall, the

**Table 2**

A summary of experimental conditions, including the identification number (ID), the presence of obstacles, approximate sensor-to-source distance ( $x_m$ ), the number of passes per experiment ( $N$ ), and sampling day of year (DOY), and meteorological conditions measured by a nearby met tower, including the mean and standard deviation of streamwise velocity ( $\bar{u}$  and  $\sigma_u$ ), turbulent intensity ( $I_u$ ), friction velocity ( $u^*$ ), mean wind direction ( $\theta_m$ ) clockwise from the north, sensible heat flux ( $H$ ), and atmospheric stability ( $z/L$ ). All meteorological variables are derived from data collected during each experiment (~30 min).

ID	obstacle	$x_m$ [m]	$N$	DOY	$\bar{u}$ [m/s]	$\sigma_u$ [m/s]	$I_u$ [-]	$u^*$ [m/s]	$\theta_m$ [deg]	$H$ [W/m <sup>2</sup> ]	$z/L$ [-]
1	N	30	22	169	2.65	0.97	0.30	0.24	300	211.54	-0.34
2	N	20	17	169	2.51	0.92	0.32	0.21	293	196.86	-0.51
3	N	10	23	169	1.71	1.26	0.43	0.24	38	205.50	-0.34
4	N	30	18	169	2.55	1.12	0.33	0.23	72	167.32	-0.32
5	N	20	15	169	1.51	0.86	0.45	0.21	22	85.67	-0.21
6	N	10	24	169	1.98	1.13	0.40	0.13	48	129.82	-1.38
7	N	10	19	171	1.92	1.01	0.40	0.20	304	212.38	-0.57
8	N	20	14	214	1.15	0.63	0.53	0.16	42	38.28	-0.24
9	N	10	16	214	1.29	0.91	0.52	0.24	349	13.08	-0.06
10	N	30	16	215	2.36	0.78	0.32	0.20	309	66.12	-0.22
11	N	20	22	215	1.44	1.01	0.48	0.23	324	141.93	-0.31
12	N	30	16	215	1.80	1.09	0.43	0.23	73	17.23	-0.04
13	Y	20	20	216	2.19	0.53	0.30	0.15	151	10.50	-0.08
14	Y	10	20	216	2.82	0.84	0.28	0.22	170	37.55	-0.09
15	Y	30	18	217	2.94	0.98	0.28	0.18	147	77.67	-0.32
16	Y	20	20	217	2.72	1.06	0.31	0.23	149	148.97	-0.31
17	Y	10	18	217	2.49	0.98	0.33	0.29	172	138.13	-0.14
18	Y	30	14	217	2.72	1.15	0.31	0.24	152	161.21	-0.30
19	Y	20	19	217	2.95	1.08	0.29	0.22	146	171.01	-0.41
20	Y	10	19	217	2.48	0.97	0.33	0.21	159	159.46	-0.46
21	Y	30	15	218	3.41	1.34	0.28	0.37	204	219.77	-0.11
22	Y	20	16	218	3.70	1.27	0.26	0.36	211	223.92	-0.12
23	Y	10	15	218	3.82	1.33	0.26	0.38	212	207.86	-0.10
24	Y	30	20	218	4.18	1.31	0.24	0.36	194	171.36	-0.09
25	Y	20	16	218	4.31	1.31	0.24	0.37	184	173.08	-0.08
26	Y	10	14	218	3.99	1.23	0.25	0.40	179	129.53	-0.05
27	Y	30	16	219	2.25	1.01	0.35	0.27	320	118.53	-0.16
28	Y	20	15	219	2.75	1.04	0.30	0.16	318	125.83	-0.77
29	Y	10	14	219	2.29	1.04	0.35	0.20	315	140.29	-0.47



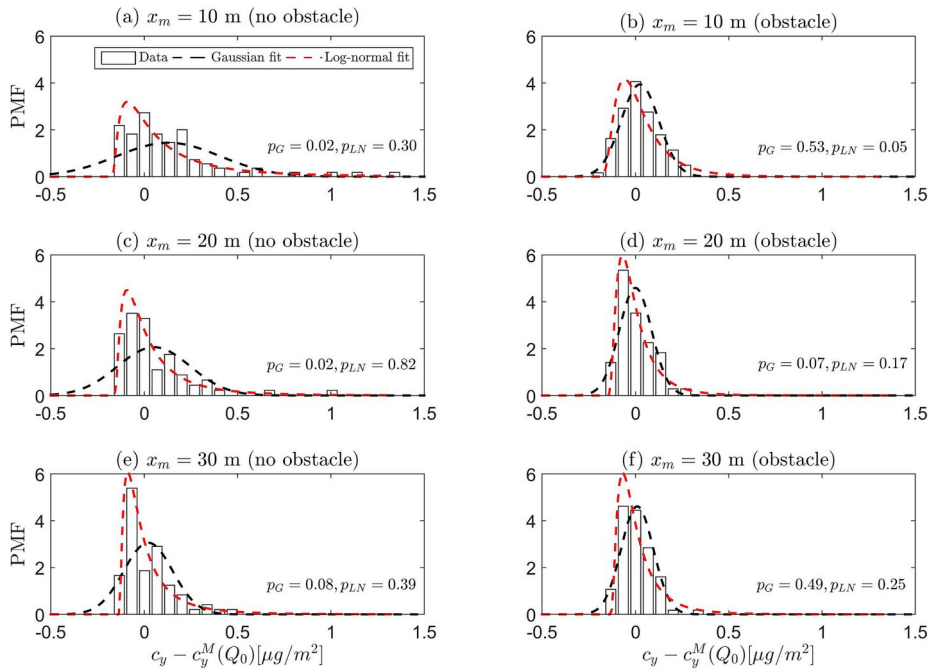
**Fig. 2.** An example (ID # is 8) of above-ambient methane mixing ratios measured at two different sensor heights ( $z_m$ ) over two passes. The corresponding wind directions for the first ( $\theta_1$ ) and the second pass ( $\theta_2$ ) are shown as a red and a black arrow, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

estimation uncertainty ( $\sigma_N^Q$ ) is smaller for experiments conducted in the presence of obstacles than under unobstructed conditions for the same inference method. This is mainly caused by the reduced spread of  $c_y - c_y^M(Q_0)$  (Fig. 3), which leads to a reduced  $\sigma_N^Q$  for experiments conducted in the presence of obstacles.

For experiments conducted under unobstructed conditions,  $Q_0$  is within  $E_N^Q \pm \sigma_N^Q$  for most experiments with the implementation of the log-normal likelihood function (Fig. 5a). The averaged  $E_N^Q$  across all experiments is 0.084 g/s, which is only 1% higher than  $Q_0$  (0.083 g/s) and suggests a good performance of the Bayesian inference using the

log-normal likelihood function. The application of the Bayesian inference with the Gaussian likelihood function resulted in an averaged  $E_N^Q$  of 0.12 g/s, which is more than 40% higher than  $Q_0$  (0.083 g/s) as shown in Fig. 5c.

For experiments conducted in the presence of obstacles,  $E_N^Q$  derived from the two methods closely match  $Q_0$ . Using Bayesian inference with the log-normal likelihood function, the averaged  $E_N^Q$  across all experiments is 0.076 g/s, which is 8% lower than  $Q_0$  (0.083 g/s). In contrast,  $E_N^Q$  estimated by the Gaussian-based Bayesian inference slightly overestimate  $Q_0$  at 0.086 g/s, which is 4% higher than  $Q_0$  (0.083 g/s).

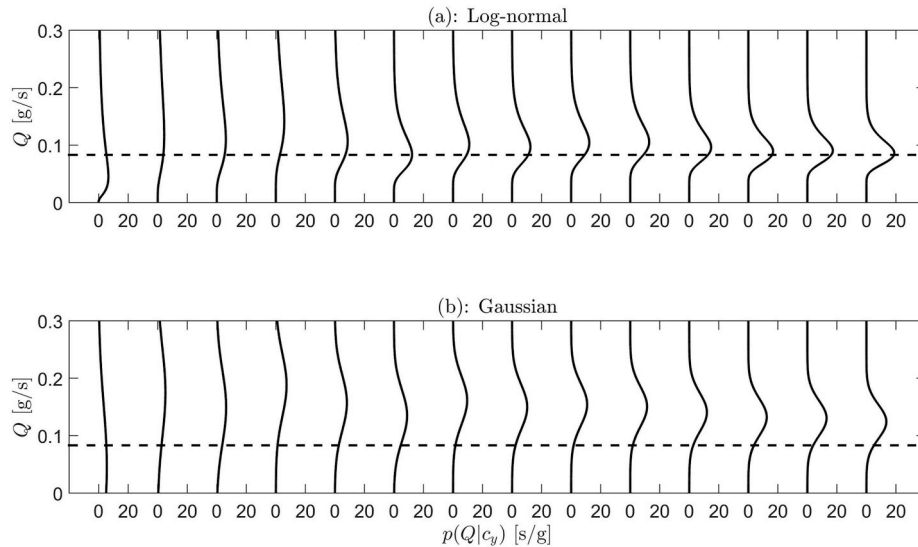


**Fig. 3.** The probability mass function (PMF) of the difference between the measured ( $c_y$ ) and modeled ( $c_y^M(Q_0)$ ) cross-plume integrated mass concentrations with source-to-sensor distance ( $x_m$ ) of 10 m (a and b), 20 m (c and d), and 30 m (e and f). The red and black dash lines are Gaussian and log-normal fit curves for the PMF, respectively. A p-value based on the Kolmogorov-Smirnov test is calculated between the PMF and the best-fit Gaussian ( $p_G$ ) and log-normal ( $p_{LN}$ ) curves. Experiments conducted in the absence of obstacles are showing on the left column (a, c, and e), and experiments conducted in the presence of obstacles are showing on the right column (b, d, and f). Note that all the results are collected at a sensor height ( $z_m$ ) of 1.3 m. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

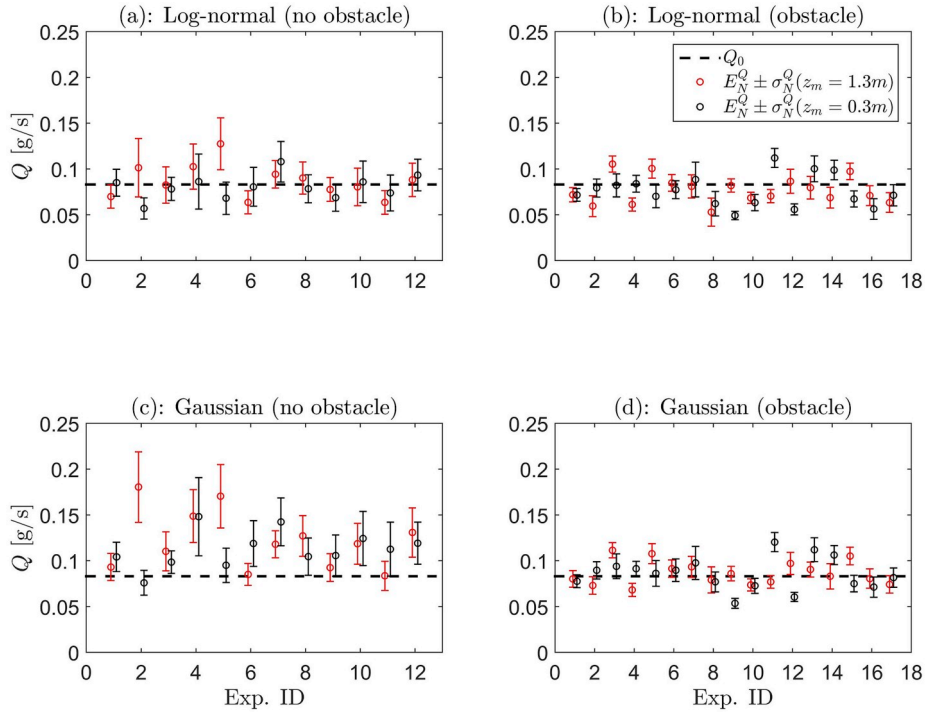
To evaluate the accuracy of model uncertainty ( $p(R_m)$ ) derived from different models, we estimate the cumulative density function (CDF) of  $p(R_c)$  and  $p(\bar{R}_m)$  based on the log-normal and Gaussian likelihood functions (Fig. 6). A p-value based on the Kolmogorov-Smirnov test is calculated between the empirical and the modeled relative errors. For experiments conducted under unobstructed conditions (Fig. 6a and c), Bayesian inference using the log-normal likelihood function shows a much better performance than that using the Gaussian likelihood function ( $p = 0.43$  versus  $p < 0.01$ ). For experiments conducted in the presence of obstacles, the performance of Bayesian inference using the

Gaussian likelihood function ( $p = 0.13$ ) clearly outperforms the log-normal likelihood function ( $p < 0.01$ ), as shown in Fig. 6b and d.

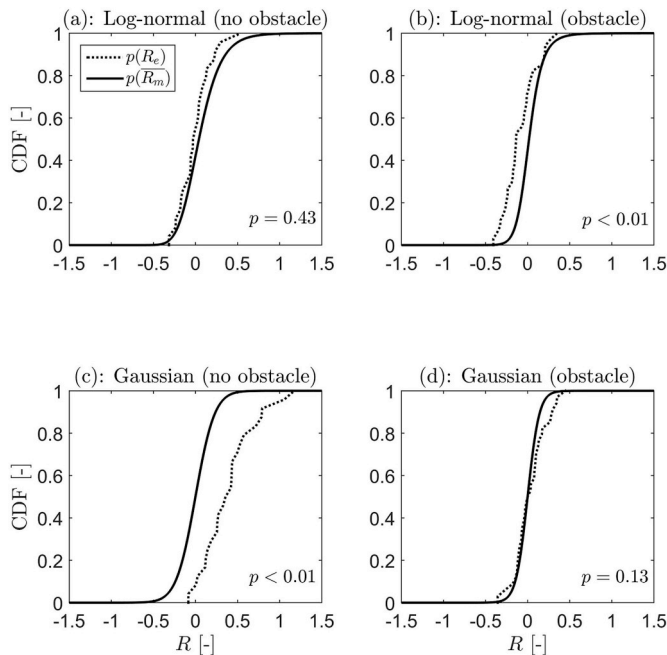
Overall, it is clear that the appropriate likelihood function ensure the performance of the Bayesian inference. More specifically, these results suggest that the implementation of a log-normal likelihood function is suitable for near-source unobstructed conditions, while a Gaussian likelihood function is desirable in the presence of obstacles, for the Bayesian inference of characterizing point-source emission in the near-field using a mobile sensor.



**Fig. 4.** The posterior PDF ( $p(Q|c_y)$ ) derived from the Bayesian inference using (a): log-normal and (b): Gaussian likelihood function. From left to right, results are shown after each sensor pass (a total of 14). The controlled release rate ( $Q_0$ ) is shown with a black dash line. This corresponds to Experiment ID of 8 (no obstacles) and  $z_m$  of 0.3 m.



**Fig. 5.** After the final pass (the  $N^{\text{th}}$  pass), the estimated emission rates  $\pm$  standard deviation ( $E_N^Q \pm \sigma_N^Q$ ) derived from Bayesian inference using the log-normal (a and b) and Gaussian (c and d) likelihood functions. For each experiment, two results obtained using the sensor located at  $z_m$  of 1.3 m and 0.3 m are color coded in red and black, respectively. The black dash line is the controlled release rate ( $Q_0$ ). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



**Fig. 6.** The cumulative density function (CDF) of the empirical relative error ( $p(R_e)$ ) and the modeled relative error  $p(R_m)$  when applying Bayesian inference using log-normal likelihood function (a and b) and Gaussian likelihood function (c and d). A p-value based on the Kolmogorov-Smirnov test is calculated between the empirical and the modeled relative errors.

**4.2. Assessment of the effects of sensor-to-source distance, sensor height, and wind speed on source characterization**

Here, we examine the effects of different experimental conditions, such as downwind distance ( $x_m$ ), sensor height ( $z_m$ ), and wind speed

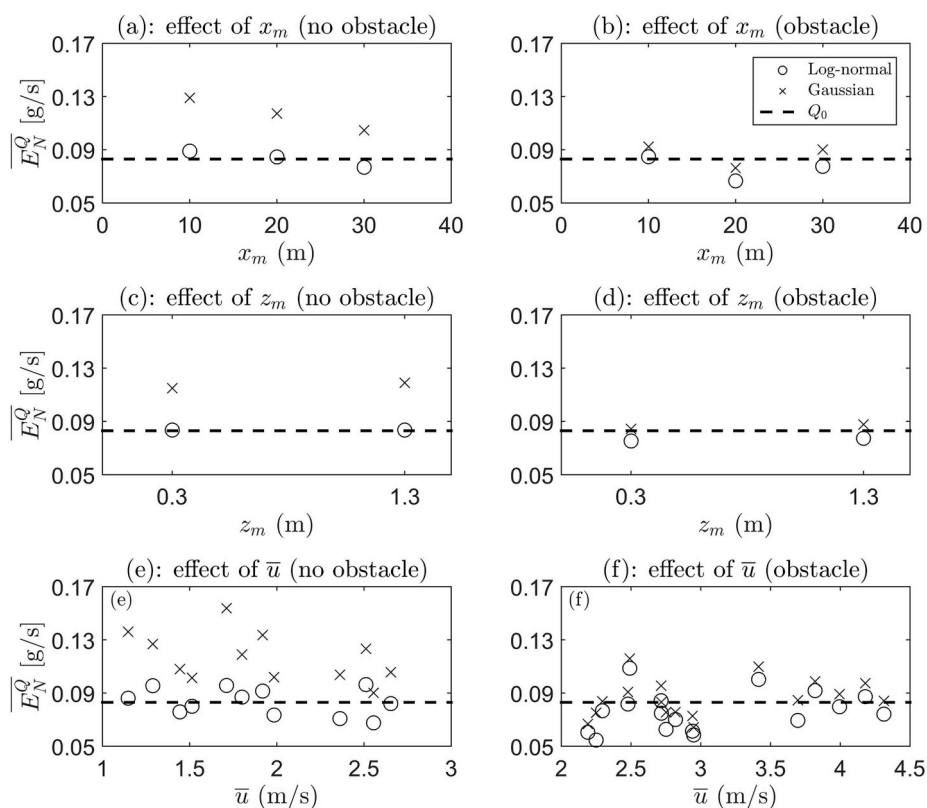
( $\bar{u}$ ), on the accuracy of source characterization. The effect of obstacles has been shown, therefore all results will be presented separately, with and without the presence of obstacles. For example, to highlight the effect of downwind distance, we averaged  $E_N^Q$  (denoted as  $\bar{E}_N^Q$ ) over experiments sharing the same  $x_m$ .

As shown in Fig. 7a, we find that the sensor-to-source distance ( $x_m$ ) is an important factor in determining source characterization performance for experiments conducted under unobstructed conditions. Current literature often favors the application of Gaussian likelihood function for far-field applications (Albertson et al., 2016). However, Bayesian inference based on the log-normal likelihood function shows the best performance when  $x_m \leq 30$  m, consistent with our results of Figs. 5 and 6. For experiments conducted with the presence of obstacles, however, little dependence of model performance on  $x_m$  is observed.

When comparing model performance using sensors of different heights, we find that  $\bar{E}_N^Q$  is insensitive to  $z_m$ . This is true for both the log-normal and the Gaussian likelihood functions, with and without the presence of obstacle (Fig. 7b and d). This demonstrates the flexibility of the inference methods with respect to the two sensor heights.

Finally, we evaluate the effects of wind speed ( $\bar{u}$ ) on the accuracy of source characterization. As shown in Fig. 7e, we note that a stronger wind speed can help reduce estimation bias when applying the Gaussian likelihood function under unobstructed conditions. A higher wind speed is usually coupled with a stronger plume mixing, which favors the application of symmetric likelihood function (e.g. Gaussian). This is similar to the effect of  $x_m$  on  $\bar{E}_N^Q$  as shown in Fig. 7a. At low wind speed, however, plumes are often transported by wind gust with a strong intermittency, and the log-normal likelihood function demonstrates a better performance. With the presence of obstacles, we find little correlation between  $\bar{E}_N^Q$  and  $\bar{u}$  for both the log-normal and the Gaussian likelihood functions (Fig. 7f).





**Figure 7.** Assessment of the effect of (a and b) sensor-to-source distance ( $x_m$ ), (c and d) sensor height ( $z_m$ ), and (e and f) mean wind speed ( $\bar{u}$ ) on the accuracy of source characterization using the Bayesian inference method. The left column (a, c, and e) are experiments conducted under unobstructed condition, and the right column (b, d, and f) are experiments conducted with the presence of obstacle. The black dash line is the controlled release rate ( $Q_0$ ).

## 5. Conclusions and future work

This paper addressed the application of Bayesian inference to estimating gas emission rates from point sources in the context of mobile sensor data, which provide more of an instantaneous view of the plume than the classic ensemble or time-averaged approaches. A series of controlled release experiments were conducted, using mobile sensors traversing cross-sections of the plume at different downwind distances, both with and without the presence of obstacles. The use of Gaussian and log-normal likelihood functions were evaluated across a range of experimental conditions, with the motivation being that instantaneous plume concentration distributions are known to exhibit a marked asymmetry, especially near the source. The results showed that the selection of the appropriate likelihood function can be guided by the underlying plume transport and mixing conditions. The plumes measured in the near-field, without obstructions, exhibited a strong skewness in the concentration fluctuations that followed the log-normal likelihood function better than the commonly-employed Gaussian form. On the contrary, at longer downwind distances the plume concentration distributions showed a greater symmetry (due to the greater extent of mixing over the travel history) and the Gaussian likelihood distribution performed better than the log-normal form. The presence of significant obstacles enhanced the plume mixing so as to favor the Gaussian distribution, even at shorter travel distances. Considering the high scalability of mobile sensing, these results are encouraging for characterizing dispersed point emissions over extensive regions (e.g. oil and gas well pads in a shale basin). Future experiments under more of real-world emission scenarios (e.g. unsteady emission rates, multiple leaks, and more complex obstacle set-up) will be useful to expand the applicability of this method.

The current study used error scale parameters directly derived from the controlled experiments, and provided an error prorogation method to estimate them when the experimental data are absent. Future work will be focused on building a look-up table for the error scale parameters under varying experimental conditions with controlled release experiments, and assessing the applicability of the proposed error prorogation method for estimating the error scale parameters. Under some conditions, we found that the performance of the inference can be improved using data simultaneously collected by vertically separated sensors. Studies of the optimal sampling strategies will be beneficial for future applications of emission sources characterization using mobile sensing data.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgment

We thank Iain Wright from University of Waterloo and Soren Przygocki from Ithaca High school, NY for their help with the experiment, Joe Rudek from the Environmental Defense Fund who lent us the Picarro G2204 methane analyzer, and Matthew Coats from the Cornell athletic department for granting access to the experimental field. This project was supported by David R. Atkinson Center for a Sustainable Future (ACSF) at Cornell University, and DOE ARPA-E's Methane Observation Networks with Innovative Technology to Obtain Reductions (MONITOR) program under grant DE-AR0000749.

## Appendix A. Determination of $u_e^M(x_m, t)$ based on the logarithmic wind profile

Here we introduce the estimation of  $u_e^M(x_m, t)$ . In a horizontally homogeneous Atmospheric Boundary Layer, the mean streamwise velocity  $\bar{u}$  follows the logarithmic profile based on the Monin Obukhov Similarity Theory (MOST) (Kaimal and Finnigan, 1994):

$$\bar{u} = \frac{u_*}{k_v} \left( \ln \left( \frac{z}{z_0} \right) - \psi \left( \frac{z}{L} \right) \right), \quad (\text{A.1})$$

where  $u_*$  is the friction velocity,  $k_v$  is the von Karman constant (0.4),  $z_0$  is the surface roughness (1.0 cm for short grassland (Brutsaert, 2013)),  $L = -\frac{u_*^2 \bar{T}}{k_v g w' T'}$  is the Obukhov length (Kaimal and Finnigan, 1994), where  $g$  is the gravitational acceleration (9.81 m/s<sup>2</sup>),  $\bar{T}$  is the mean air temperature (in Kelvin), and  $\overline{w'T'}$  is the mean covariance of the instantaneous  $w$  and  $T$ .  $\psi \left( \frac{z}{L} \right)$  is a dimensionless stability correction function (Stull, 1988):

$$\psi \left( \frac{z}{L} \right) = \begin{cases} -4.7 \frac{z}{L}, & L \geq 0 \\ 2 \ln \left( \frac{1+\varphi}{2} \right) + 2 \ln \left( \frac{1+\varphi^2}{2} \right) - 2 \tan^{-1}(\varphi) + \frac{\pi}{2}, & L < 0 \end{cases} \quad (\text{A.2})$$

where  $\varphi = (1 - 16z/L)^{1/4}$ .

According to Equation (3b) in the main text,  $u_e(x_m, t)$  is defined as the weighted plume advection velocity at the exit surface  $(x_m, y, z)$ :

$$u_e(x_m, t) = \int_{z_{\min}}^{z_{\max}} \int_{y_{\min}}^{y_{\max}} D(x_m, y, z, t) u(x_m, y, z, t) dy dz. \quad (\text{A.3})$$

To calculate  $u_e^M(x_m, t)$ , we first replace  $u(x_m, y, z, t)$  by  $\bar{u}$ , which is a function of  $z$  and  $t$ . Then,  $\int_{y_{\min}}^{y_{\max}} D(x_m, y, z, t) dy$  can be modeled using the Lagrangian Stochastic Model (LSM) as  $D_y^M(x_m, z, t)$ , as detailed in the following section. With these modifications,  $u_e^M(x_m, t)$  can be calculated as:

$$u_e^M(x_m, t) = \int_{z_{\min}}^{z_{\max}} D_y^M(x_m, z, t) \bar{u} dz. \quad (\text{A.4})$$

## Appendix B. Lagrangian Stochastic Model (LSM) for estimation of $D_y^M(x_m, z, t)$

The LSM describes the plume dispersion by calculating the trajectories of marked fluid particles in 2D (longitudinal and vertical directions). The particle positions ( $x_p$  and  $z_p$ ) in the downwind and vertical directions are calculated as:

$$dx_p = (u_p + \bar{u}) dt, \quad (\text{B.1a})$$

$$dz_p = w_p dt, \quad (\text{B.1b})$$

where  $dt$  is the time step.  $u_p$  and  $w_p$  are the particle's Lagrangian velocity in the  $x$  and  $z$  directions, which follows the generalized Langevin equation (Thomson, 1987):

$$du_p = a_u dt + b_u dW, \quad (\text{B.2a})$$

$$dw_p = a_w dt + b_w dW, \quad (\text{B.2b})$$

where  $dW$  is an incremental Wiener process with a zero mean and a variance of  $dt$ .  $a$  and  $b$  are parameters that need to satisfy Kolmogorov's hypothesis of local isotropy in a Lagrangian frame of reference (Kolmogorov, 1941) and the well-mixed condition (Thomson, 1987). According to the simplest solution (Thomson, 1987) and considering a flat homogeneous surface layer,  $a$  and  $b$  can be formulated as:

$$a_u = -\frac{b_u^2}{A} (\sigma_w^2 u_p - \overline{u'w'} w_p) + \frac{1}{A} \left( \sigma_w^2 \frac{\partial \sigma_u^2}{\partial z} u_p w_p - \overline{u'w'} \frac{\partial \sigma_u^2}{\partial z} w_p^2 \right), \quad (\text{B.3a})$$

$$a_w = -\frac{b_w^2}{A} (\sigma_u^2 w_p - \overline{u'w'} u_p) + \frac{1}{A} \left( -\overline{u'w'} \frac{\partial \sigma_w^2}{\partial z} u_p w_p + \sigma_u^2 \frac{\partial \sigma_w^2}{\partial z} w_p^2 \right) + \frac{1}{2} \frac{\partial \sigma_w^2}{\partial z}, \quad (\text{B.3b})$$

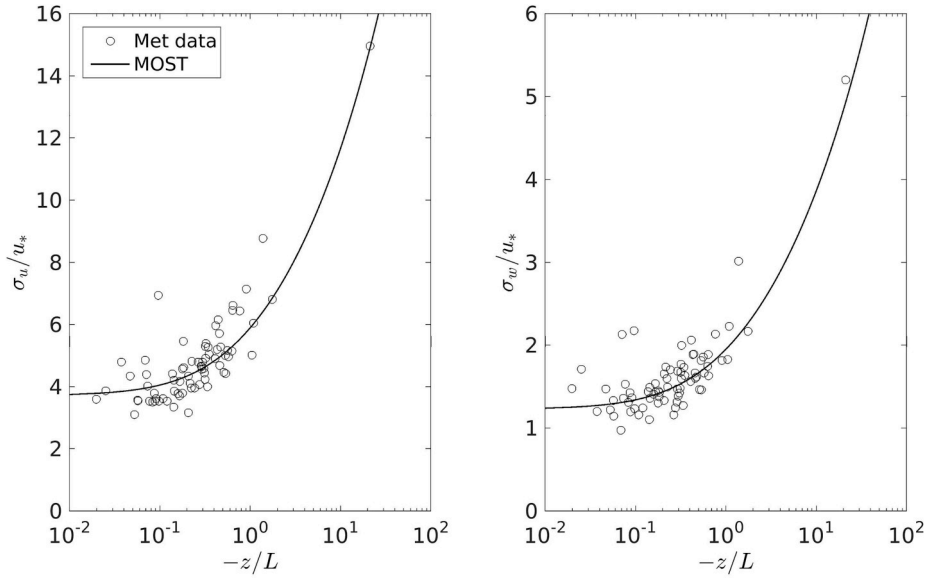
$$b_u = b_w = \sigma_w \sqrt{2/T_L}, \quad (\text{B.3c})$$

where  $A = 2(\sigma_u^2 \sigma_w^2 - \overline{u'w'}^2)$ ,  $\sigma_u$  and  $\sigma_w$  are the standard deviations of the Eulerian velocity in the longitudinal and vertical directions,  $\overline{u'w'}$  is the Reynolds stress, and  $T_L$  is the Lagrangian velocity time scale.

To solve Equations (B.1)–(B.3), vertical profiles of  $\bar{u}$ ,  $\sigma_u$ ,  $\sigma_w$ ,  $\overline{u'w'}$  and  $T_L$  are required. The Reynolds stress  $\overline{u'w'}$  =  $-u_*^2$  is assumed to be a constant in the surface layer (Stull, 1988). Since measurements were available only at a single height during each experiment, the MOST was applied to describe the vertical profiles of all wind statistics required by the LSM ( $\bar{u}$ ,  $\sigma_u$ ,  $\sigma_w$ ). More specifically,  $\bar{u}$  is described by logarithmic wind profile (Equation (A.1)).  $\sigma_u$  and  $\sigma_w$  can be described as (Hsieh and Katul, 1997):

$$\sigma_u = 3.7 \times u_* \left( 1 - 3 \frac{z}{L} \right)^{1/3}, \quad \sigma_w = 1.26 \times u_* \left( 1 - 3 \frac{z}{L} \right)^{1/3}. \quad (\text{B.4})$$

The empirical parameters 3.7 and 1.26 are estimated by fitting the measured  $\sigma_u$  and  $\sigma_w$  by the 3-D sonic anemometer against  $\left( 1 - 3 \frac{z}{L} \right)^{1/3}$  (Figure B.1).



**Fig. B.1.** Normalized velocity standard deviations as a function of atmospheric stability. The experimental data are shown with open circles from the 3D sonic data. The solid lines are the MOST conventional 1/3 relation (Equation (B.4)) with best-fit empirical coefficients.

In order to solve the Langevin equation, we should also estimate the Lagrangian time scale.  $T_L$  can be estimated using the K theory (Duman et al., 2014):

$$T_L = \frac{K}{\sigma_w^2}, \quad (\text{B.5})$$

where  $K$  is the diffusion coefficient (Rodean, 1996)

$$K = \frac{\kappa u_* z}{\psi_h}, \quad (\text{B.6})$$

and  $\psi_h$  is the stability correction (Hsieh et al., 2000)

$$\psi_h = \begin{cases} \left(1 - 3\frac{z}{L}\right), & z/L \geq 0 \\ 0.32\left(0.037 - \frac{z}{L}\right)^{-1/3}, & z/L < 0 \end{cases}. \quad (\text{B.7})$$

During a LSM run, particles' velocity and location are stored in the  $x - z$  domain. Given a point source with a unit emission rate, the predicted mean concentration after integrating over  $y$ , denoted as  $C_y^{LSM}$ , can be described as (Flesch et al., 1995; Rannik et al., 2012):

$$C_y^{LSM}(x, z, t) = \frac{1}{N_p \Delta x} \sum \frac{1}{|w_p(x, z)|}, \quad (\text{B.8})$$

where  $N_p$  is the number of fluid particles.  $\Delta x$  and  $\Delta z$  are the grid spacing of the LSM in  $x$  and  $z$  directions, respectively.

The LSM predicted  $D_y^M(x_m, z)$  can be calculated as:

$$D_y^M(x_m, z, t) = \frac{C_y^{LSM}(x_m, z, t)}{\sum_{z_{min}}^{z_{max}} C_y^{LSM}(x_m, z, t) \Delta z}, \quad (\text{B.9})$$

where  $x_{max}$ ,  $z_{max}$  are the size of the computational domain in the  $x$ ,  $z$  directions, respectively.

A computational domain of  $(x_{max}, z_{max}) = (100 \text{ m}, 20 \text{ m})$  is used with the grid cell size of  $(\Delta x, \Delta z) = (0.5 \text{ m}, 0.01 \text{ m})$ .  $dt$  is determined dynamically as  $dt = \min(0.02T_L, \Delta z/w(t - dt))$  to satisfy the necessary condition of  $dt \ll T_L$  and to prevent a large jump in the vertical direction (Duman et al., 2014). For each run, a total of  $N_p = 10^6$  particle are released from the source and their trajectories are calculated using the LSM. The ground and the boundary-layer top (set to be 1 km for neutral and unstable conditions (Duman et al., 2014)) are considered as perfect reflectors, such that a particle will be perfectly reflected back in the vertical direction and the sign of both  $u_p$  and  $w_p$  being reversed (Duman et al., 2014). Particles can only exit the end of the domain when  $x(t) > x_{max}$ . A summary of parameters used in the LSM is shown in Table B.1.

**Table B.1**  
A summary of parameters used in the LSM

Name	Value
Number of fluid particles ( $N_p$ )	$10^6$
Computational domain size ( $x_{max}, z_{max}$ )	100, 20 (m)
Computational grid size ( $\Delta x, \Delta z$ )	0.5, 0.01 (m)

Table B.1 (continued)

Name	Value
Surface roughness ( $z_0$ )	0.01 (m)
von Karman constant ( $\kappa$ )	0.4
Height of sensor ( $z_m$ )	0.3, 1.3 (m)
Height of source ( $z_s$ )	0.02 (m)
Friction velocity ( $u^*$ )	case dependent
Standard deviation of $u$ and $w$ ( $\sigma_u, \sigma_w$ )	case dependent
Obukhov length ( $L$ )	case dependent

$D_y^M(x_m, z, t)$  is calculated without considering obstacle, which may be acceptable at  $x_m = 20$  m and 30 m but will introduce errors at  $x_m = 10$  m. For simplicity, we assume that  $D_y^M(x_m, z, t)$  is vertically well-mixed by obstacle-injected wake eddies from ground to  $1.5 \times z_f$  at  $x_m = 10$  m, where  $z_f$  is the height of the obstacle. Therefore,  $D_y^M(10, z < 1.5z_f, t) = \frac{1}{1.5z_f} \sum_0^{1.5z_f} D_y^M(10, z, t) \Delta z$ .

## References

- Albertson, J.D., Harvey, T., Foderaro, G., Zhu, P., Zhou, X., Ferrari, S., Amin, M.S., Modrak, M., Brantley, H., Thoma, E.D., 2016. A mobile sensing approach for regional surveillance of fugitive methane emissions in oil and gas production. *Environ. Sci. Technol.* 50 (5), 2487–2497.
- Alexeeff, S.E., Roy, A., Shan, J., Liu, X., Messier, K., Apte, J.S., Portier, C., Sidney, S., Van Den Eeden, S.K., 2018. High-resolution mapping of traffic related air pollution with Google street view cars and incidence of cardiovascular events within neighborhoods in Oakland, CA. *Environ. Health* 17 (1), 38.
- Alvarez, R., Pacala, S., Winebrake, J., Chameides, W., Hamburg, S., 2012. Greater focus needed on methane leakage from natural gas infrastructure. *Proc. Natl. Acad. Sci.* 109 (17), 6435–6440.
- Apte, J.S., Messier, K.P., Gani, S., Brauer, M., Kirchstetter, T.W., Lunden, M.M., Marshall, J.D., Portier, C.J., Vermeulen, R.C., Hamburg, S.P., 2017. High-resolution air pollution mapping with Google street view cars: exploiting big data. *Environ. Sci. Technol.* 51 (12), 6999–7008.
- ARPA-E, 2015. Methane observation networks with innovative technology to obtain reductions (monitor). <https://arpa-e.energy.gov/?q=arpa-e-programs/monitor>.
- Aurell, J., Gullett, B.K., Pressley, C., Tabor, D.G., Gribble, R.D., 2011. Aerostat-lofted instrument and sampling method for determination of emissions from open area sources. *Chemosphere* 85 (5), 806–811.
- Brantley, H., Thoma, E., Squier, W., Guven, B., Lyon, D., 2014. Assessment of methane emissions from oil and gas production using mobile measurements. *Environ. Sci. Technol.* 48, 14508–14515.
- Brutsaert, W., 2013. *Evaporation into the Atmosphere: Theory, History and Applications*, vol. 1 Springer Science & Business Media.
- Cai, C., Ferrari, S., 2009. Information-driven sensor path planning by approximate cell decomposition. *IEEE Trans. Syst. Man Cybern. B Cybern.* 39 (3), 672–689.
- Csanady, G.T., 1973. *Turbulent Diffusion in the Environment*. Reidel.
- Davidson, M.J., Mylne, K.R., Jones, C.D., Phillips, J.C., Perkins, R.J., Fung, J.C.H., Hunt, J.C.R., 1995. Plume dispersion through large groups of obstacles - a field investigation. *Atmos. Environ.* 29 (22), 3245–3256.
- Davidson, M.J., Snyder, W.H., Lawson, R.E., Hunt, J.C.R., 1996. Wind tunnel simulations of plume dispersion through groups of obstacles. *Atmos. Environ.* 30, 3715–3731.
- Devarakonda, S., Sevusu, P., Liu, H., Liu, R., Iftoed, L., Nath, B., 2013. Real-time air quality monitoring through mobile sensing in metropolitan areas. In: *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*. ACM, pp. 15.
- Duman, T., Katul, G.G., Siqueira, M.B., Cassiani, M., 2014. A velocity-dissipation Lagrangian stochastic model for turbulent dispersion in atmospheric boundary-layer and canopy flows. *Boundary-Layer Meteorol.* 152 (1), 1–18.
- Fackrell, J.E., Robins, A.G., 1982. Concentration fluctuations and fluxes in plumes from point sources in a turbulent boundary layer. *J. Fluid Mech.* 117, 1–26.
- Flesch, T.K., Wilson, J.D., Yee, E., 1995. Backward-time Lagrangian stochastic dispersion models and their application to estimate gaseous emissions. *J. Appl. Meteorol.* 34 (6), 1320–1332.
- Foster-Wittig, T.A., Thoma, E.D., Albertson, J.D., 2015. Estimation of point source fugitive emission rates from a single sensor time series: a conditionally-sampled Gaussian plume reconstruction. *Atmos. Environ.* 115, 101–109.
- Horst, T.W., Weil, J.C., 1992. Footprint estimation for scalar flux measurements in the atmospheric surface layer. *Boundary-Layer Meteorol.* 59, 279–296.
- Hsieh, C.I., Katul, G., Chi, T.W., 2000. An approximate analytical model for footprint estimation of scalar fluxes in thermally stratified atmospheric flows. *Adv. Water Resour.* 23 (7), 765–772.
- Hsieh, C.-I., Katul, G.G., 1997. Dissipation methods, Taylor's hypothesis, and stability correction functions in the atmospheric surface layer. *J. Geophys. Res.: Atmos.* 102 (D14), 16391–16405.
- Humphries, R., Jenkins, C., Leuning, R., Zegelin, S., Griffith, D., Caldwell, C., Berko, H., Feitz, A., 2012. Atmospheric tomography: a bayesian inversion technique for determining the rate and location of fugitive emissions. *Environ. Sci. Technol.* 46 (3), 1739–1746.
- Jackson, R.B., Down, A., Phillips, N.G., Ackley, R.C., Cook, C.W., Plata, D.L., Zhao, K., 2014. Natural gas pipeline leaks across Washington, dc. *Environ. Sci. Technol.* 48 (3), 2051–2058.
- Jaynes, E.T., 1968. Prior probabilities. *IEEE Trans. Syst. Sci. Cybern.* 4 (3), 227–241.
- Jaynes, E.T., 2003. *Probability Theory: the Logic of Science*. Cambridge university press.
- Kaimal, J.C., Finnigan, J.J., 1994. *Atmospheric Boundary Layer Flows: Their Structure and Measurement*. Oxford University Press.
- Keats, A., Yee, E., Lien, F.-S., 2007. Bayesian inference for source determination with applications to a complex urban environment. *Atmos. Environ.* 41 (3), 465–479.
- Kolmogorov, A.N., 1941. Dissipation of energy in locally isotropic turbulence. In: *Doklady Akademiia Nauk SSSR*, vol. 32. pp. 16–18.
- Lan, X., Talbot, R., Laine, P., Torres, A., 2015. Characterizing fugitive methane emissions in the barnett shale area using a mobile laboratory. *Environ. Sci. Technol.* 49 (13), 8139–8146.
- Lee, J.K., Christen, A., Ketler, R., Nestic, Z., 2017. A mobile sensor network to map carbon dioxide emissions in urban environments. *Atmos. Meas. Tech.* 10 (2), 645–665.
- Lumley, J.L., Panofsky, H.A., 1964. *The Structure of Atmospheric Turbulence*. Interscience Monographs and Texts in Physics and Astronomy. Wiley, New York.
- Marjovi, A., Arfire, A., Martinoli, A., 2015. High resolution air pollution maps in urban environments using mobile sensor networks. In: *Distributed Computing in Sensor Systems (DCOSS)*, 2015 International Conference on. IEEE, pp. 11–20.
- Massey, F.J.J., 1951. The Kolmogorov-smirnov test for goodness of fit. *J. Am. Stat. Assoc.* 46 (253), 68–78.
- Mylne, K.R., Mason, P.J., 1991. Concentration fluctuation measurements in a dispersing plume at a range of up to 1000 m. *Q. J. R. Meteorol. Soc.* 117 (497), 177–206.
- Obukhov, A., 1971. Turbulence in an atmosphere with a non-uniform temperature. *Boundary-Layer Meteorol.* 2 (1), 7–29.
- Phillips, N.G., Ackley, R., Crosson, E.R., Down, A., Hutya, L.R., Bronfield, M., Karr, J.D., Zhao, K., Jackson, R.B., 2013. Mapping urban pipeline leaks: methane leaks across boston. *Environ. Pollut.* 173, 1–4.
- Rannik, Ü., Sogachev, A., Foken, T., Göckede, M., Kljun, N., Leclerc, M.Y., Vesala, T., 2012. Footprint analysis. In: *Eddy Covariance: A Practical Guide to Measurement and Data Analysis*. Springer, pp. 211–261.
- Rao, K.S., 2005. Uncertainty analysis in atmospheric dispersion modeling. *Pure Appl. Geophys.* 162 (10), 1893–1917.
- Rella, C.W., Tsai, T.R., Botkin, C.G., Crosson, E.R., Steele, D., 2015. Measuring emissions from oil and natural gas well pads using the mobile flux plane technique. *Environ. Sci. Technol.* 49 (7), 4742–4748.
- Rodean, H.C., 1996. *Stochastic Lagrangian Models of Turbulent Diffusion*, vol. 45 Springer.
- Stull, R.B., 1988. *An Introduction to Boundary Layer Meteorology*. Kluwer Academic Publishers.
- Thomson, D.J., 1987. Criteria for the selection of stochastic models of particle trajectories in turbulent flows. *J. Fluid Mech.* 180, 529–556.
- von Fischer, J.C., Cooley, D., Chamberlain, S., Gaylord, A., Griebenow, C.J., Hamburg, S.P., Salo, J., Schumacher, R., Theobald, D., Ham, J., 2017. Rapid, vehicle-based identification of location and magnitude of urban natural gas pipeline leaks. *Environ. Sci. Technol.* 51 (7), 4091–4099. <https://doi.org/10.1021/acs.est.6b06095>.
- Warhaft, Z., 2000. Passive scalars in turbulent flows. *Annu. Rev. Fluid Mech.* 32 (1), 203–240.
- Wilson, J.D., Sawford, B.L., 1996. Review of Lagrangian stochastic models for trajectories in the turbulent atmosphere. *Boundary-Layer Meteorol.* 78 (1–2), 191–210.
- Yacovitch, T.I., Herndon, S.C., Pétron, G., Kofler, J., Lyon, D., Zahniser, M.S., Kolb, C.E., 2015. Mobile laboratory observations of methane emissions in the barnett shale region. *Environ. Sci. Technol.* 49 (13), 7889–7895. <https://doi.org/10.1021/es506352j>.
- Yee, E., 2007. Bayesian probabilistic approach for inverse source determination from limited and noisy chemical or biological sensor concentration measurements. In: *Chemical and Biological Sensing VIII*, vol. 6554. International Society for Optics and Photonics, pp. 65540W.
- Yee, E., 2008. Theory for reconstruction of an unknown number of contaminant sources using probabilistic inference. *Boundary-Layer Meteorol.* 127 (3), 359–394.
- Yee, E., 2012. Probability theory as logic: data assimilation for multiple source reconstruction. *Pure Appl. Geophys.* 169 (3), 499–517.
- Yee, E., Flesch, T., 2010. Inference of emission rates from multiple sources using bayesian probability theory. *J. Environ. Monit.* 12 (3), 622–634.

- Yee, E., Kosteniuk, P.R., Chandler, G.M., Biltoft, C.A., Bowers, J.F., 1993. Statistical characteristics of concentration fluctuations in dispersing plumes in the atmospheric surface layer. *Boundary-Layer Meteorol.* 65 (1), 69–109.
- Zhang, G., Ferrari, S., Qian, M., 2009. An information roadmap method for robotic sensor path planning. *J. Intell. Robot. Syst.* 56 (1), 69–98.
- Zhou, X., Aurell, J., Mitchell, W., Tabor, D., Gullett, B., 2017. A small, lightweight multipollutant sensor system for ground-mobile and aerial emission sampling from open area sources. *Atmos. Environ.* 154, 31–41.
- Zhou, X., Passow, F.H., Rudek, J., von Fisher, J.C., Hamburg, S.P., Albertson, J.D., 2019. Estimation of methane emissions from the us ammonia fertilizer industry using a mobile sensing approach. *Elem Sci Anth* 7 (1).