Research Article

Inverse Dispersion for an Unknown Number of Sources: Model Selection and Uncertainty Analysis

Eugene Yee

Defence R&D Canada – Suffield, P.O. Box 4000 Stn Main, Medicine Hat, AB, Canada T1A 8K6

Correspondence should be addressed to Eugene Yee, eyee0309@gmail.com

A simple recursive method is presented for performing the inverse dispersion modeling of an unknown number of (localized) sources, given a finite number of noisy concentration data acquired by an array of detectors. Bayesian probability theory is used to address the problem of selecting the source model which is most plausible in view of the given concentration dataset and all the available prior information. The recursive algorithm involves subtracting a predicted concentration signal arising from a source model consisting of N localized sources from the measured concentration data for increasing values of N and examining the resulting residual data to determine if the residuals are consistent with the estimated noise level in the concentration data. The method is illustrated by application to a real concentration dataset obtained from an atmospheric dispersion experiment involving the simultaneous release of a tracer from four sources.

1. Introduction

Significant advances in sensing technology for concentration measurements of contaminants (e.g., toxic industrial materials; chemical, biological, and radiological agents) released into the atmosphere, either accidentally or deliberately, have fostered interest in exploiting this information for the reconstruction of the contaminant source distribution responsible for the observed concentration. Indeed, innovative methods for measuring and collecting concentration data in formal observation (or monitoring) networks of ever more compact (nanotechnology-enabled) sensors, including the adoption of new wireless sensor network technologies for acquisition of data in situ and for rapid delivery of this information through wireless transmission [1], have placed the emphasis on data assimilation and inverse dispersion modeling.

This should not be too surprising owing to the fact that inverse dispersion modeling will enable the full realization of the benefits in this new data context, allowing the innovative fusion of sensor data with predictive models for atmospheric dispersion for determination of the unknown source characteristics. In turn, this will lead to a greatly improved situational awareness and result in a significantly enhanced common operational picture required for making more informed decisions for mitigation of the consequences of a (toxic) contaminant release. In this context, inverse dispersion modeling for source reconstruction has important implications for public safety and security.

In the past, two different methodologies have been used to address the inverse dispersion modeling problem, namely, deterministic optimization and stochastic Bayesian approaches. In the optimization method, the parameters θ describing the source model are obtained as the solution of a nonlinear least-squares optimization problem:

$$\hat{\theta} = \arg \min_{\theta \in \mathbb{R}^n} \{ f_O(\theta) + f_R(\theta) \}, \tag{1.1}$$

where n is the number of parameters (unknowns) used to parameterize the source model, $f_O(\theta)$ is an objective functional that measures the total mismatch (usually taken as the sum-of-squared differences) between the measured concentration data and the synthetic (predicted) concentration data associated with the current source model θ, and $f_R(\theta)$ is a regularization functional that is used either to impose an additional constraint on the solution or to incorporate *a priori* information (which is required to produce a mathematically unique solution). Most efforts in application of the optimization approach to inverse dispersion modeling have been focused on the development of numerical methods for solution of the optimization problem or on the prescription of the regularization functional for incorporation of various forms of *a priori* information. The optimization method for inverse dispersion modeling has been used by Robertson and Langner [2], Bocquet [3], Thomson et al. [4], Allen et al. [5], and Issartel et al. [6] (among others) based on various numerical methods for solution of the optimization problem (e.g., variational data assimilation, genetic algorithm, conjugate gradient algorithm with line search, augmented Lagrangian method) and a number of different forms of regularization (e.g., energy, entropy, smoothness, or flatness).

While the optimization approach for inverse dispersion modeling seeks to provide a single optimal solution for the problem, the stochastic Bayesian approach seeks to generate multiple solutions to the problem with the evaluation of the degree of plausibility for each solution. In contrast to the optimization approach, the stochastic Bayesian approach allows the quantification of the uncertainty in the source reconstruction (arising as such from the use of incomplete and noisy concentration data and imperfect models of atmospheric dispersion for the inverse dispersion modeling). This approach has been developed and used by a number of researchers (e.g., [7–10]).

In all the previous studies cited above, the inverse dispersion modeling involved the problem of identification of source parameters for a *single* localized source. The determination of the characteristics of multiple localized sources was briefly addressed by Yee [11] and by Sharan et al. [12] using, respectively, a stochastic Bayesian approach and an optimization (least-squares) approach that were similar to those applied previously for the recovery of a single source. In both these cases, this was possible because it was assumed in these investigations that the number of sources was known *a priori*. The problem of reconstruction of an unknown number of localized sources using a finite number of noisy concentration measurements is a significantly more difficult problem. A solution for this problem was

proposed by Yee [13, 14] who approached the problem as a *generalized* parameter estimation problem in which the number of localized sources, N, in the (unknown) source distribution was included explicitly in the parameter vector θ, in addition to the usual parameters that characterize each localized source (e.g., location, emission rate, source-on time, source-off time).

Solving the reconstruction of an unknown number of localized sources as a generalized parameter estimation problem posed a number of conceptual and technical difficulties. The primary conceptual difficulty resided in the fact that when the number of sources is unknown *a priori*, the dimension of the parameter space (or, equivalently, the dimension or length of the associated parameter vector θ) is an unknown that needs to be estimated using the noisy concentration data. Yee [13, 14] overcame this conceptual difficulty by using Bayesian probability theory to formulate the full joint probability density function (PDF) for the number of sources N and the parameters of the N localized sources and then demonstrated how a reversible-jump Markov chain Monte Carlo (RJMCMC) algorithm can be designed to draw samples of candidate source models from this joint PDF, allowing for changes in the dimensionality of the source model (or equivalently, changes in the number of localized sources in the source distribution). Furthermore, it was found that the RJMCMC algorithm sampled the parameter space of the unknown source distribution rather inefficiently. To overcome this technical difficulty, Yee [13, 14] showed how the RJMCMC algorithm can be combined either with a form of parallel tempering based on a Metropolis-coupled MCMC algorithm [13] or with a simpler and computationally more efficient simulated annealing scheme [14] to improve significantly the sampling efficiency (or, mixing rate) of the Markov chain in the variable dimension parameter space.

In this paper, we address the inverse dispersion modeling problem for the difficult case of multiple sources when the number of sources is unknown *a priori* as a model selection problem, rather than as a generalized parameter estimation problem as described by Yee [13, 14]. The model selection problem here is formulated in the Bayesian framework which involves the evaluation of model selection statistics that gauges rigorously the tradeoff between the goodness-of-fit of the source model structure to the concentration data and the complexity of the source model structure. The model selection approach proposed herein for reconstruction of an unknown number of localized sources is conceptually and algorithmically simpler than that based on a generalized parameter estimation approach [13, 14], leading as such to a simpler and more efficient computer implementation. More specifically, the model selection approach is simpler than the generalized parameter estimation approach in that it does not need to deal with the complexity of the variable dimensionality of the parameter space, which requires generally the development of sophisticated and computationally intensive algorithms.

2. Bayesian Model Selection

Let m_N denote a source distribution (model) which consists of a *known* number, N, of localized sources. The model m_N is characterized by a set of parameters θ. Bayesian inference for θ, when given the concentration data $\mathfrak{D}$, is based on Bayes' theorem (rule):

$$p(\theta \mid m_N, \mathfrak{D}, I) = \frac{p(\theta \mid m_N, I) p(\mathfrak{D} \mid m_N, \theta, I)}{p(\mathfrak{D} \mid m_N, I)}, \tag{2.1}$$

where I denotes the background (contextual) information that defines the multiple source reconstruction problem (e.g., background meteorology, atmospheric dispersion model used to determine the source-receptor relationship, etc.) and the vertical bar "|" denotes "conditional upon" or "given." Furthermore, in (2.1), $p(\theta \mid m_N, \mathfrak{D}, I) \equiv p(\theta)$ is the posterior probability distribution of the parameters θ, $p(\theta \mid m_N, I) \equiv \pi(\theta)$ is the prior probability distribution of the parameters θ before the data $\mathfrak{D}$ was made available, and $p(\mathfrak{D} \mid m_N, \theta, I) \equiv \mathcal{L}(\theta)$ is the likelihood function and defines the probabilistic model of how the data were generated. Finally, $p(\mathfrak{D} \mid m_N, I)$ is the evidence and, in the context of parameter estimation, ensures proper normalization of the posterior probability distribution (assuming $\theta \in \mathbb{R}^n$):

$$p(\mathfrak{D} \mid m_N, I) = \int_{\mathbb{R}^n} \mathcal{L}(\theta)\pi(\theta)d\theta \equiv Z(m_N). \tag{2.2}$$

However, in our problem, the number of sources (N) is *unknown a priori*, so the relevant question that needs to be addressed is as follows: given a set of possible source models $S \equiv \{m_0, m_1, \ldots, m_{N_{\max}}\}$ (m_N denotes the source model consisting of N localized sources), which source model is the most plausible (probable) given the concentration data $\mathfrak{D}$? This question is addressed rigorously in the Bayesian framework by using Bayes' theorem to compute the posterior probability for source model m_N ($N = 0, 1, \ldots, N_{\max}$) to give

$$p(m_N \mid \mathfrak{D}, I) = \frac{p(m_N \mid I)p(\mathfrak{D} \mid m_N, I)}{p(\mathfrak{D} \mid I)}. \tag{2.3}$$

Here, $p(m_N \mid I)$ is the prior probability of source model m_N given only the information I, and $p(\mathfrak{D} \mid m_N I)$ is the global likelihood of the concentration data $\mathfrak{D}$ given the source model m_N and I and represents the goodness-of-fit of the model to the data. Note that the global likelihood of (2.3) is identical to the evidence $Z(m_N)$ defined in (2.2). Finally, the marginal probability of the data given the background information, $p(\mathfrak{D} \mid I)$, is a normalization constant over all source models:

$$p(\mathfrak{D} \mid I) = \sum_{N=0}^{N_{\max}} p(m_N \mid I)p(\mathfrak{D} \mid m_N, I). \tag{2.4}$$

Given any two source models m_{N_0} and m_{N_1} from the set S, the "odds" in favor of model m_{N_1} relative to model m_{N_0} are given by the ratio of the posterior probabilities of the two models, which on using (2.3) and (2.2), reduces to

$$K_{10} \equiv \frac{p(m_{N_1} \mid \mathfrak{D}, I)}{p(m_{N_0} \mid \mathfrak{D}, I)} = \frac{p(\mathfrak{D} \mid m_{N_1}, I)}{p(\mathfrak{D} \mid m_{N_0}, I)} \cdot \frac{p(m_{N_1} \mid I)}{p(m_{N_0} \mid I)} = \frac{Z(m_{N_1})}{Z(m_{N_0})} \cdot \frac{p(m_{N_1} \mid I)}{p(m_{N_0} \mid I)}. \tag{2.5}$$

Note that the posterior odds ratio K_{10} is the product of the evidence ratio $Z(m_{N_1})/Z(m_{N_0})$ and the prior odds ratio $p(m_{N_1} \mid I)/p(m_{N_0} \mid I)$. If every source model in the set S is equally probable (namely, $p(m_N \mid I) = 1/(N_{\max} + 1)$ for $N = 0, 1, \ldots, N_{\max}$), then the prior odds ratio is unity, and the posterior odds ratio is *identically equal* to the evidence ratio. When viewed as a model selection problem, the inverse dispersion modeling of an unknown number of sources is conceptually simple: compute the evidence $Z(m_N)$ for each source model m_N

in S given the input concentration data $\mathfrak{D}$ and the background information I, select the model with the largest value of the evidence as the most probable model, and recover the parameters θ corresponding to this most probable model.

The key to the model selection problem reduces to the computation of the evidence $Z(m_N)$ for an arbitrary value of N. A perusal of (2.2) shows that the evidence for the source model structure m_N (namely, a source model consisting of exactly N localized sources) involves a computation of the overlap integral (or inner product) of the likelihood $\mathcal{L}(\theta)$ and the prior distribution $\pi(\theta)$ over the entire parameter space for model m_N. This inner product can be interpreted as a metric (statistic) for model selection and, as such, is an intrinsic element of the source model structure m_N in the sense that it depends on both the set of parameters to be varied and the prior ranges for those parameters but is independent of the most probable (preferred) values for the parameters.

Note that the model selection metric selects preferentially a source model structure with the largest overlap of the likelihood and prior distribution, implying that in order for a more complex source model (with a necessarily higher-dimensional parameter space) to be selected, this requires the prior information to be smeared over a larger hypervolume in the parameter space in order to overlap significantly the likelihood. This mechanism embodies automatically an Occam's razor in that a simpler source model structure is preferentially selected unless the likelihood of the data for a more complex source model increases significantly more than what would be expected from simply fitting the noise (residual between the measured concentration and the predicted concentration provided by the model).

The solution of the model selection problem is conceptually simple involving, as such, the computation of $Z(m_N)$ ($m_N \in S$). Unfortunately, this computation is a technically difficult problem for two reasons. Firstly, as N (or, equivalently, as the number of localized sources) increases, the overlap integral in (2.2) needs to be evaluated over a parameter space of increasing dimensionality n. More specifically as N increases, the dimensionality of the parameter space n increases linearly with N. Secondly, the integrand of the overlap integral of (2.2) becomes ever more complex with increasing N, owing to a counting degeneracy in the problem. In particular, it is noted that $\mathcal{L}(\theta)$ is invariant under a reordering (relabelling) of the identifiers used to label the individual localized sources in the source model. This degeneracy simply corresponds to different (but physically equivalent) identifications of what is meant by a particular localized source in the source distribution, implying that the number of modes in $\mathcal{L}(\theta)$ increases with m_N as $N!$. The factor $N!$ here corresponds simply to the number of possible permutations of the labels for the localized sources in the source model m_N. In summary, the increase in the dimensionality of the parameter space and in the complexity of $\mathcal{L}(\theta)$ with increasing N makes the problem of computation of $Z(m_N)$ technically difficult.

In view of the difficulty arising from the numerical evaluation of $Z(m_N)$ for increasing values of N, we consider an alternative (but closely related) methodology for addressing the model selection problem in the context of inverse dispersion modeling of an *a priori* unknown number of localized sources. To this purpose, the question that needs to be posed is after the removal of a predicted concentration for a source model m_N from the measured concentration data, are the residuals that remain consistent with the estimated noise level or is there further evidence for the existence of an additional localized source in the residuals? This question is easily addressed using Bayes' theorem as follows.

Towards this purpose, the following model is assumed for the measured concentration data d_i ($i = 1, 2, \ldots, N_m$):

$$d_i = \overline{C}_i + e_i^{(1)} = \tilde{C}_i + e_i^{(1)} + e_i^{(2)} \equiv \tilde{C}_i + e_i, \qquad (2.6)$$

where $\overline{C}_i$ is the true (albeit unknown) mean concentration, $e_i^{(1)}$ is the measurement error, $\tilde{C}_i$ is the predicted mean concentration obtained from an atmospheric dispersion model, $e_i^{(2)}$ is the model error incurred by using the (necessarily) imperfect atmospheric model to predict $\overline{C}_i$, and e_i is the composite error that includes both the measurement and model error. Furthermore, it is assumed that the expectation value $\langle e_i \rangle$ of the composite error is zero, and the variance $\langle e_i^2 \rangle$ is assumed (for now) to be known and given by σ_i^2. Hence, it is assumed that we are given the concentration data $\mathfrak{D} \equiv \{d_1, d_2, \ldots, d_{N_m}\}$ and corresponding uncertainties $\{\sigma_1, \sigma_2, \ldots, \sigma_{N_m}\}$ where σ_i is the standard deviation (square root of the variance) for the i-th concentration datum d_i. Normally, exact values for the uncertainties σ_i are unknown so that all we would have available would be some estimated values for these uncertainties s_i, rather than the true values σ_i.

If we assume *a priori* that there are N localized sources in the domain, we can analyse the concentration data $\mathfrak{D}$ given that m_N is the correct model structure and draw samples of source models with exactly N localized sources (encoded by θ) from the posterior distribution given by (2.1). Assume that N_s samples of source models with N localized sources have been drawn from the posterior distribution; namely, we have available the samples $\theta_k(m_N)$ ($k = 1, 2, \ldots, N_s$) where $\theta_k(m_N)$ is kth source model sample consisting of exactly N localized sources. To determine if N is the correct number of localized sources, we can compute the residual data $\mathcal{E} \equiv \{e_1, e_2, \ldots, e_{N_m}\}$ where e_i is the mean residual for the ith concentration datum (with realizations of the residuals obtained by subtracting the predicted concentration $\tilde{C}_i(m_N)$ for a sample of the source model m_N encoded in $\theta_k(m_N)$ from the measured concentration d_i). The mean residual datum e_i is estimated from the ensemble of source model samples $\theta_k(m_N)$ ($k = 1, 2, \ldots, N_s$) as follows:

$$\hat{e}_i = \frac{1}{N_s} \sum_{k=1}^{N_s} \left(d_i - \tilde{C}_i(\theta_k(m_N)) \right). \qquad (2.7)$$

Similarly, the uncertainty s_i^e (error) in the residual datum e_i is estimated as $s_i^e = \left[\hat{e_i^2} - (\hat{e}_i)^2 \right]^{1/2}$, where

$$\hat{e_i^2} = \frac{1}{N_s} \sum_{k=1}^{N_s} \left(d_i - \tilde{C}_i(\theta_k(m_N)) \right)^2. \qquad (2.8)$$

Now, given the residual data $\mathcal{E}$, we evaluate the evidence for the presence of an additional source by considering two models (hypotheses) for $\mathcal{E}$: (1) H_0: there is no additional source and the residual data $e_i = \tilde{C}_i(m_0) + e_i = e_i$ (where $\tilde{C}_i(m_0)$ is the predicted concentration for a source model with no localized sources and hence has zero value), and (2) H_1: there is an additional source and the residual data $e_i = \tilde{C}_i(m_1) + e_i$ (where $\tilde{C}_i(m_1)$ is the predicted concentration for a source model consisting of one localized source).

Under hypothesis H_0 (null hypothesis), the residuals are consistent with the estimated noise level e_i. To decide which hypothesis is favoured by the data, we calculate the posterior odds ratio for H_1 relative to H_0. From (2.5) and assuming that hypotheses H_0 and H_1 are equally probable so $p(H_0 \mid I) = p(H_1 \mid I) = 1/2$, we get

$$K_{10}^\epsilon \equiv \frac{p(H_1 \mid \mathcal{E}, I)}{p(H_0 \mid \mathcal{E}, I)} \equiv \frac{p(m_1 \mid \mathcal{E}, I)}{p(m_0 \mid \mathcal{E}, I)} = \frac{p(\mathcal{E} \mid m_1, I)}{p(\mathcal{E} \mid m_0, I)} \equiv \frac{Z^\epsilon(m_1)}{Z^\epsilon(m_0)}. \tag{2.9}$$

If K_{10}^ϵ exceeds some preassigned threshold, then there is evidence for an additional source (namely, hypothesis H_1 is favoured with respect to hypothesis H_0). If this is the case, then the number of sources N is increased by one, and we repeat the analysis of the concentration data $\mathfrak{D}$ given that m_{N+1} is the correct model structure. This analysis would involve drawing samples from the posterior distribution of (2.1) with m_N replaced by m_{N+1}, computing the residual data $\mathcal{E}$ (obtained now by subtracting the predicted concentration $\tilde{C}_i(m_{N+1})$ from the measured concentration d_i), and then determining once more the evidence in favour of H_1 relative to that of H_0. When this test fails (namely, K_{10}^ϵ is less than or equal to the preassigned threshold value), the algorithm terminates. At this point, the number of sources N_* has been determined, and the samples drawn from the posterior distribution of (2.1) for m_{N_*} can be used to estimate any posterior statistic of interest for the source parameters θ.

To summarize, the algorithm for the inverse dispersion modeling of an unknown number of sources consists of the following steps.

(1) Input concentration data $\mathfrak{D} = \{d_1, d_2, \dots, d_{N_m}\}$ and associated estimated uncertainties $\{s_1, s_2, \dots, s_{N_m}\}$.

(2) Set $N = 1$ and specify a threshold K_* for the posterior odds ratio. Draw N_s samples of source models (encoded by the parameter vectors $\theta_k(m_N)$, $k = 1, 2, \dots, N_s$) from the posterior distribution $p(\theta \mid m_N, \mathfrak{D}, I)$ given by (2.1).

(3) Compute the evidences $Z(m_N)$ and $Z(m_{N-1})$, and use these values to determine K_{10} in accordance to (2.5). If $K_{10} \leq K_*$, stop (no sources are present in the domain, so $N_* = 0$).

(4) Use samples $\theta_k(m_N)$ ($k = 1, 2, \dots, N_s$) to estimate the residual data $\mathcal{E} = \{\hat{e}_1, \hat{e}_2, \dots, \hat{e}_{N_s}\}$ and the associated uncertainty in this data $\{s_1^\epsilon, s_2^\epsilon, \dots, s_{N_m}^\epsilon\}$ in accordance to (2.7) and (2.8), respectively.

(5) Using the information from the previous step, compute K_{10}^ϵ in accordance to (2.9). If $K_{10}^\epsilon \leq K_*$, set $N_* = N$, output $\theta_k(m_N)$ ($k = 1, 2, \dots, N_s$), and stop.

(6) Increase $N \rightarrow N + 1$. Draw N_s samples $\theta_k(m_N)$ ($k = 1, 2, \dots, N_s$) of source models from $p(\theta \mid m_N, \mathfrak{D}, I)$ given by (2.1). Continue from Step 4.

The algorithm summarized previously determines the minimum number of localized sources needed to represent the concentration data $\mathfrak{D}$ down to the estimated noise level, without the requirement to compute the evidence $Z(m_N)$ for $N \geq 2$ (a technically difficult problem computationally). Steps 2 and 3 in the algorithm correspond to the signal detection phase which addresses the specific question: is there evidence in the concentration data $\mathfrak{D}$ for the existence of one or more localized sources in the domain? On the other hand, Steps 4 to 6 in the algorithm address the estimation phase and answer the question given that there is evidence for one or more localized sources in the domain (detection has been confirmed), how many localized sources are present, and what are the values of the source parameters that best characterize each of these localized sources?

3. Source-Receptor Relationship (Dispersion Modeling)

In this paper, we are interested in the reconstruction of an unknown number N of localized sources. In view of this, the model for the source density function has the following explicit form (for source model m_N):

$$S(\mathbf{x}, t) = \sum_{j=1}^{N} Q_j \delta(\mathbf{x} - \mathbf{x}_{s,j}) [U(t - T_{b,j}) - U(t - T_{e,j})], \tag{3.1}$$

where $\delta(\cdot)$ and $U(\cdot)$ are the Dirac delta and Heaviside unit step functions, respectively, and, Q_j, $\mathbf{x}_{s,j}$, $T_{b,j}$, and $T_{e,j}$ are the emission rate, vector position (location), source-on time, and source-off time, respectively, of the jth source ($j = 1, 2, \ldots, N$). For a source model (or molecule) composed of N localized sources (or source atoms), it is convenient to define the parameter vector as follows: $\theta \equiv (\mathbf{x}_{s,1}, T_{b,1}, T_{e,1}, Q_1, \ldots, \mathbf{x}_{s,N}, T_{b,N}, T_{e,N}, Q_N) \in \mathbb{R}^{6N}$.

To apply the Bayesian probability theory, it is necessary to relate the hypotheses of interest about the unknown source model S to the modeled (predicted) concentration $\tilde{C}$ [cf. (2.6)]. The predicted concentration $\tilde{C}$ can be compared directly to the measured concentration d "seen" by a sensor at receptor location $\mathbf{x}_r$ and time t_r, averaged over the sensor volume and sampling time. The comparison can be effected by averaging the mean concentration $C(\mathbf{x}, t)$ over this sensor volume and sampling time to give

$$\tilde{C}(\mathbf{x}_r, t_r) \equiv \int_0^T dt \int_{\mathcal{R}} d\mathbf{x} \, C(\mathbf{x}, t) h(\mathbf{x}, t \mid \mathbf{x}_r, t_r) \equiv \langle C \mid h \rangle (\mathbf{x}_r, t_r), \tag{3.2}$$

where $h(\mathbf{x}, t \mid \mathbf{x}_r, t_r)$ is the spatial-temporal filtering function for the sensor concentration measurement at $(\mathbf{x}_r, t_r)$ and $\mathcal{R} \times [0, T]$ corresponds to the space-time volume that contains the source distribution and the receptors (sensors). Note that the mean concentration $\tilde{C}(\mathbf{x}_r, t_r)$ "seen" by a sensor can be expressed succinctly as the inner or scalar product $\langle C \mid h \rangle$ of the mean concentration C and the sensor response function h.

A source-oriented forward-time Lagrangian stochastic (LS) model can be used to predict $C(\mathbf{x}, t)$. In this approach, C is estimated from the statistical characteristics of "marked" particle trajectories described by the following stochastic differential equation [15]:

$$d\mathbf{X}(t) = \mathbf{V}(t) \, dt,$$

$$d\mathbf{V}(t) = \mathbf{a}(\mathbf{X}(t), \mathbf{V}(t), t) \, dt + (C_0 \epsilon(\mathbf{X}(t), t))^{1/2} \, d\mathbf{W}(t), \tag{3.3}$$

where $\mathbf{X}(t) \equiv (X_i(t)) = (X_1(t), X_2(t), X_3(t))$ is the Lagrangian position and $\mathbf{V}(t) \equiv (V_i(t)) = (V_1(t), V_2(t), V_3(t))$ is the Lagrangian velocity of a "marked" fluid element at time t (marked by the source as the fluid element passes through it at some earlier time t'). The state of the fluid particle at any time t after its initial release from the source density function S is defined by $(\mathbf{X}, \mathbf{V})$. In (3.3), C_0 is the Kolmogorov "universal" constant that is associated with the Kolmogorov similarity hypothesis for the form of the second-order Lagrangian velocity structure function in the inertial subrange; ϵ is the mean dissipation rate of turbulence kinetic energy; $d\mathbf{W}(t) \equiv (dW_i(t)) = (dW_1(t), dW_2(t), dW_3(t))$ are the increments of a vector-valued (three-dimensional) Wiener process; $\mathbf{a} \equiv (a_i) = (a_1, a_2, a_3)$ is the drift coefficient vector (or, more precisely, the conditional mean acceleration vector).

Unfortunately, for the current application, the computational demands of the source-oriented approach make it impractical as a direct tool for sampling from the posterior distribution because this involves necessarily exploring a large number of source parameter hypotheses. This is highly computer intensive, as the simulation-based Bayesian inference procedure requires a large number of forward calculations of the mean concentration to be performed, each of which involves the numerical solution of (3.2) and (3.3). In this particular case, it may be useful to construct an emulator for the simulation model and use this emulator as an inexpensive surrogate to replace the forward model [16–18]. Fortunately, applying this type of approximation method for the forward model is not required for the current problem. It was shown by Keats et al. [7] and Yee et al. [10] that an *exact* computationally efficient procedure (appropriate for use in a Bayesian inference scheme) exists in the form of a receptor-oriented strategy for the representation of the source-receptor relationship.

Towards this purpose, the modeled sensor concentration $\widetilde{C}(\mathbf{x}_r, t_r)$ can be computed alternatively using the following dual representation:

$$
\begin{aligned}
\widetilde{C}(\mathbf{x}_r, t_r) &\equiv \int_0^T dt \int_{\mathcal{R}} d\mathbf{x}\, C(\mathbf{x}, t) h(\mathbf{x}, t \mid \mathbf{x}_r, t_r) \\
&= \int_0^{t_r} dt_0 \int_{\mathcal{R}} d\mathbf{x}_0\, C^*(\mathbf{x}_0, t_0 \mid \mathbf{x}_r, t_r) \mathcal{S}(\mathbf{x}_0, t_0) \equiv \langle C^* \mid \mathcal{S} \rangle (\mathbf{x}_r, t_r),
\end{aligned}
\tag{3.4}
$$

where $C^*(\mathbf{x}_0, t_0 \mid \mathbf{x}_r, t_r)$ is the adjunct (or dual) concentration at space-time point $(\mathbf{x}_0, t_0)$ associated with the concentration datum at location $\mathbf{x}_r$ at time t_r (with $t_0 \leq t_r$). A comparison of (3.2) and (3.4) implies the duality relationship $\langle C \mid h \rangle = \langle C^* \mid \mathcal{S} \rangle$ between C and C^* through the source functions h and $\mathcal{S}$. Moreover, C^* is uniquely defined in the sense that it is explicitly constructed so that it verifies this duality relationship.

The adjunct field C^* in the dual representation can be estimated from the statistical characteristics of "marked" particle trajectories corresponding to a receptor-oriented backward-time LS trajectory model, defined as the solution to the following stochastic differential equation (with $dt' > 0$):

$$
d\mathbf{X}^b(t') = \mathbf{V}^b(t') dt',
\tag{3.5}
$$

$$
d\mathbf{V}^b(t') = \mathbf{a}^b\left(\mathbf{X}^b(t'), \mathbf{V}^b(t'), t'\right) dt' + \left(C_0 \epsilon\left(\mathbf{X}^b(t'), t'\right)\right)^{1/2} d\mathbf{W}(t'),
\tag{3.6}
$$

where at any given time t', $(\mathbf{X}^b, \mathbf{V}^b)$ is a point in the phase space along the backward trajectory of the "marked" fluid element (here assumed to be marked or tagged as a fluid particle which at time t_r passed through the spatial volume of the detector at location $\mathbf{x}_r$). Here, $(\mathbf{X}^b, \mathbf{V}^b)$ defines the state of a fluid particle at any time $t' < t_r$ before its "final" release from the sensor space-time volume h at time t_r. It can be shown [15, 19] that C^* obtained from (3.5) for a detector with the filtering function h and C obtained from (3.2) for a release from the source density $\mathcal{S}$, are exactly consistent with the duality relationship $\langle C \mid h \rangle = \langle C^* \mid \mathcal{S} \rangle$ provided: (1) $\mathbf{V}^b$ in (3.5) is related to $\mathbf{V}$ in (3.3) as $\mathbf{V}^b(t) = \mathbf{V}(t)$, and (2) $\mathbf{a}^b$ in (3.5) is related to $\mathbf{a}$ in (3.3) as

$$
a_i^b(\mathbf{x}, \mathbf{u}, t) = a_i(\mathbf{x}, \mathbf{u}, t) - C_0 \epsilon(\mathbf{x}, t) \frac{\partial}{\partial u_i} \ln P_E(\mathbf{u}),
\tag{3.7}
$$

where $P_E(\mathbf{u})$ is the background Eulerian velocity PDF. Thomson [15] provides one particular solution for the drift coefficient vector $\mathbf{a}$ that is consistent with the well-mixed criterion for the case where $P_E(\mathbf{u})$ has a Gaussian functional form (namely, for Gaussian turbulence).

Finally, if we substitute (3.1) into (3.4), the predicted concentration $\widetilde{C}(\theta) \equiv \widetilde{C}(\mathbf{x}_r, t_r; \theta)$ "seen" by the sensor at $(\mathbf{x}_r, t_r)$ is given explicitly by the following expression:

$$\widetilde{C}(\theta) = \sum_{j=1}^{N} Q_j \int_{T_{b,j}}^{\min(T, T_{e,j})} C^*(\mathbf{x}_{s,j}, t_s \mid \mathbf{x}_r, t_r) dt_s. \tag{3.8}$$

4. Prior and Likelihood

4.1. Prior

Assuming that each parameter in θ (for source model m_N) is logically independent, the prior $\pi(\theta)$ can be factored as follows:

$$\pi(\theta) = \prod_{j=1}^{N} \pi(Q_j) \pi(\mathbf{x}_{s,j}) \pi(T_{b,j}) \pi(T_{e,j}). \tag{4.1}$$

In this paper, the component prior distributions are assigned uniform distributions over an appropriate range for each source parameter. Furthermore, it is noted that the parameter ranges specified in the prior define the region in $\mathbb{R}^n$ ($n \equiv 6N$) over which the evidence integral is carried out [cf. (2.2)]. More specifically, for $j = 1, 2, \ldots, N$, $Q_j \sim \mathcal{U}([0, Q_{\max}])$, where $Q_{\max}$ is an *a priori* upper bound on the emission rate; $\mathbf{x}_{s,j} \sim \mathcal{U}(\mathcal{R})$ where $\mathcal{R} \subset \mathbb{R}^3$ is the *a priori* specified spatial domain that is assumed to contain the source $\mathcal{S}$; and, $T_{b,j} \sim \mathcal{U}([0, T_{\max}])$ and $T_{e,j} \sim \mathcal{U}([T_{b,j}, T_{\max}])$, where $T_{\max}$ is the upper bound on the time at which the source was turned on or turned off. Here, the symbol $\sim$ denotes "is distributed as" and $\mathcal{U}([a, b])$ is the uniform distribution defined on the closed interval $[a, b]$. Finally, note in the specification of the prior for $T_{e,j}$ that the lower bound of the parameter range is $T_{b,j}$, encoding the fact that the time the source is turned off must necessarily occur after it has been turned on.

4.2. Likelihood

In the absence of a detailed knowledge of the noise distribution e_i in (2.6), other than the fact that it has a finite variance σ_i^2, the application of the principle of maximum entropy [20] informs us that a Gaussian distribution is the most conservative choice for the direct probability (likelihood) of the concentration data $\mathfrak{D}$. In consequence, the likelihood function $\mathcal{L}(\theta)$ has the following form:

$$\mathcal{L}(\theta) \equiv \mathcal{L}(\theta \mid \mathfrak{D}, \varsigma) = \frac{1}{\prod_{i=1}^{N_m} \sqrt{2\pi} \sigma_i} \exp\left(-\frac{1}{2} \chi^2(\theta)\right), \tag{4.2}$$

where

$$\chi^2(\theta) \equiv \sum_{i=1}^{N_m} \left(\frac{d_i - \tilde{C}_i(\theta)}{\sigma_i} \right)^2.$$ (4.3)

In (4.2), the notation for the likelihood function was adjusted to include the standard deviation for the noise $\varsigma \equiv (\sigma_1, \sigma_2, \ldots, \sigma_{N_m})$ in the conditioning to emphasize the fact that $\{\sigma_i\}_{i=1}^{N_m}$ are assumed to be known quantities.

Unfortunately, as mentioned previously, the noise term e_i in (2.6) is rather complicated, arising as such from a superposition of both model and measurement errors. In consequence, reliable estimates for σ_i ($i = 1, 2, \ldots, N_m$) are difficult to obtain in practical applications. In view of this, let us denote by s_i the quoted estimate of the standard deviation for the noise term e_i for which the true (but unknown) standard deviation is σ_i. Now, let us characterize the uncertainty in the specification of the standard deviation of e_i with an inverse gamma distribution of the following form (or, equivalently, prior distribution for σ_i):

$$\pi(\sigma_i \mid s_i, \alpha, \beta) = 2 \frac{\alpha^\beta}{\Gamma(\beta)} \left(\frac{s_i}{\sigma_i} \right)^{2\beta} \exp\left(-\alpha \frac{s_i^2}{\sigma_i^2} \right) \frac{1}{\sigma_i}, \quad i = 1, 2, \ldots, N_m,$$ (4.4)

where $\Gamma(\mathbf{x})$ denotes the gamma function and α and β are scale and shape parameters, respectively, that define the inverse gamma distribution. The inverse gamma distribution for σ_i has mean $\langle \sigma_i \rangle = s_i \alpha^{1/2} \Gamma(\beta - 1/2) / \Gamma(\beta)$ and variance $\mathrm{Var}[\sigma_i] = s_i^2 \alpha [1/(\beta - 1) - \Gamma^2(\beta - 1/2)/\Gamma^2(\beta)]$. Again, the parameters α and β have been added to the PDF of the noise uncertainty in (4.4) to indicate that the values for these parameters are assumed to be known.

In view of (4.4), the true but unknown standard deviations σ_i of e_i that appear in (4.2) can be treated as nuisance parameters and eliminated by considering the integrated likelihood

$$\mathcal{L}(\theta \mid \mathfrak{D}, \mathbf{s}, \alpha, \beta) = \int \mathcal{L}(\theta \mid \mathfrak{D}, \varsigma) \pi(\varsigma \mid \mathbf{s}, \alpha, \beta) d\varsigma$$

$$= \int \mathcal{L}(\theta \mid \mathfrak{D}, \varsigma) \prod_{i=1}^{N_m} 2 \frac{\alpha^\beta}{\Gamma(\beta)} \left(\frac{s_i}{\sigma_i} \right)^{2\beta} \exp\left(-\alpha \frac{s_i^2}{\sigma_i^2} \right) \frac{1}{\sigma_i} d\varsigma,$$ (4.5)

where $\mathbf{s} \equiv (s_1, s_2, \ldots, s_{N_m})$ are the estimated standard deviations for the noise $(e_1, e_2, \ldots, e_{N_m})$ and $d\varsigma \equiv d\sigma_1 d\sigma_2 \cdots d\sigma_{N_m}$. Now, substituting the form for $\mathcal{L}(\theta \mid \mathfrak{D}, \varsigma)$ from (4.2) and (4.3) into (4.5) and performing the integration with respect to ς, we obtain the integrated likelihood function given by

$$\mathcal{L}(\theta \mid \mathfrak{D}, \mathbf{s}, \alpha, \beta) = \prod_{i=1}^{N_m} \frac{\alpha^\beta \Gamma(\beta + 1/2)}{\sqrt{2\pi} s_i \Gamma(\beta)} \frac{1}{\left[\alpha + \left(d_i - \tilde{C}_i(\theta) \right)^2 / (2s_i^2) \right]^{\beta + 1/2}}.$$ (4.6)

The integrated likelihood of (4.6) can be interpreted simply as an average over all conditional likelihoods given ς, weighted by their prior probabilities. In so doing, the integrated

likelihood incorporates the uncertainties regarding the standard deviations for e_i (i = $1, 2, \ldots, N_m$).

The integrated likelihood function given by (4.6) depends explicitly on the hyperparameters α and β for which values need to be assigned. In this paper, the values of α and β are assigned as $\alpha = \pi^{-1}$ and $\beta = 1$. The assignment $\alpha = \pi^{-1}$ ensures that $\langle \sigma_i \rangle = s_i$ encoding our belief that our estimates s_i of the standard deviation of e_i are unbiased. Furthermore, the assignment $\beta = 1$ results in a very heavy-tailed distribution for $\pi(\sigma_i \mid s_i, \alpha, \beta)$ which allows significant deviations of the noise uncertainty from the quoted value of s_i (provided by the user). Indeed, with the choice $\beta = 1$, the variance associated with $p(\sigma_i \mid s_i, \alpha, \beta)$ in (4.4) becomes infinite. The heavy tail of the distribution is chosen to account for possibly significant under-estimations of the actual uncertainty (namely, the quoted uncertainty $s_i \ll \sigma_i$). This could arise from inconsistencies in the model concentration predictions owing to structural model error or to "outliers" in the measured concentration data owing to either measurement error or perhaps distortion of the measured concentration data due to some unrecognized spurious (background) source.

In order to compute $Z^e(m_1)$ required in the evaluation of K_{10}^e in (2.9), we are required to specify a functional form for the likelihood $\mathcal{L}(\theta(m_1) \mid \mathcal{E}, \mathbf{s}^e)$ of the estimated mean residual data $\mathcal{E} \equiv \{\widehat{e}_1, \widehat{e}_2, \ldots, \widehat{e}_{N_s}\}$ and the associated uncertainty in this data $\mathbf{s}^e \equiv \{s_1^e, s_2^e, \ldots, s_{N_m}^e\}$. To this end, the integrated likelihood function of (4.6) is also used for the specification of $\mathcal{L}(\theta(m_1) \mid \mathcal{E}, \mathbf{s}^e)$:

$$\mathcal{L}(\theta(m_1) \mid \mathcal{E}, \mathbf{s}^e) \equiv \mathcal{L}(\theta(m_1) \mid \mathcal{E}, \mathbf{s}^e, \alpha, \beta)$$

$$= \prod_{i=1}^{N_m} \frac{\alpha^\beta \Gamma(\beta + 1/2)}{\sqrt{2\pi} s_i^e \Gamma(\beta)} \frac{1}{\left[\alpha + \left(\widehat{e}_i - \widetilde{C}_i(\theta(m_1))\right)^2 / (2s_i^{e2})\right]^{\beta + 1/2}}. \tag{4.7}$$

Here, $\theta(m_1)$ are the parameters of a source model consisting of a single localized source, used in the test to determine if the mean residual data $\mathcal{E}$ contains evidence for the presence of an additional source. Towards this objective, the likelihood function of (4.7) is used to calculate the evidence $Z^e(m_1)$ for this test as

$$Z^e(m_1) = \int_{\mathbb{R}^n} \mathcal{L}(\theta(m_1) \mid \mathcal{E}, \mathbf{s}^e, \alpha, \beta) \pi(\theta(m_1)) d\theta(m_1), \tag{4.8}$$

where $n = 6$ (corresponding to the dimension of the parameter space for one source).

5. Computational Methods

There are two major computational problems that need to be addressed in order to fully specify the algorithm summarized in Section 2, namely, (1) specification of a method to draw samples from the posterior distribution $p(\theta \mid m_N, \mathfrak{D}, I)$ (Steps 2 and 6) and (2) specification of a method for computation of the evidence $Z(m_1)$ (Step 3, cf. (2.2)) or $Z^e(m_1)$ (Step 5, cf. (4.8)).

The most popular method that can be used for sampling from $p(\theta \mid m_N, \mathfrak{D}, I)$ (see (2.1)) is a stochastic algorithm referred to as Markov chain Monte Carlo (MCMC). There has been considerable effort expended to improve the efficiency and the chain mixing efficacy

of MCMC sampling procedures in order to allow rapid chain convergence to and efficient sampling from the posterior distribution of θ. Improvements to MCMC sampling methods include: introduction of an adaptive Metropolis algorithm by Haario et al. [21], formulation of the differential evolution Monte Carlo algorithm by Ter Braak [22], development of a differential evolution adaptive Metropolis algorithm (DREAM) by Vrugt et al. [23], and design of multiple-try Metropolis algorithms by Liu et al. [24].

The are a number of options that are available for the computation of the evidences $Z(m_1)$ and $Z^e(m_1)$. Owing to the fact that the determination of $Z(m_1)$ and $Z^e(m_1)$ involves only an evaluation of a multidimensional integral in a low-dimensional space ($\mathbb{R}^6$ in this case), it is in principle possible to apply a brute force numerical integration [25] to address this problem. An alternative would be to apply a methodology that is applicable to the evaluation of the evidence in the general case, involving the evaluation of an overlap integral in a high-dimensional parameter space $\mathbb{R}^n$ ($n \gg 1$). These methodologies include importance sampling estimators [26] and thermodynamic integration [27].

Although there are various alternatives (leading to many different combinations) that could be used potentially to draw samples from $p(\theta \mid m_N, \mathfrak{D}, I)$ and to evaluate $Z(m_1)$ and $Z^e(m_1)$, we have used instead a single methodology to do both. The methodology that is used in this paper to address both these problems is nested sampling developed by Skilling [28] for the efficient evaluation of the evidence Z for the general case. The nested sampling algorithm transforms the multidimensional evidence integral of (2.2) into a simple one-dimensional representation:

$$Z \equiv Z(m_N) = \int_0^1 \mathcal{L}(\chi)\,d\chi, \tag{5.1}$$

where

$$\chi(\mathcal{L}^*) = \int_{\mathcal{L}(\theta) > \mathcal{L}^*} \pi(\theta)\,d\theta \tag{5.2}$$

is the prior mass in the parameter (hypothesis) space enclosed with likelihood greater than $\mathcal{L}^*$ and $\mathcal{L}(\chi)$ is the inverse which labels the likelihood contour that encloses a prior mass χ. If we evaluate the likelihood $\mathcal{L}(\chi)$ at a sequence of m points χ_i ($i = 1, 2, \ldots, m$) with $0 < \chi_m < \chi_{m-1} < \cdots < \chi_2 < \chi_1 < 1$ and $\mathcal{L}(\chi_i) > \mathcal{L}(\chi_j)$ ($i > j$), the evidence Z can be approximated from the likelihood-ordered samples as the following weighted sum:

$$\hat{Z} = \sum_{i=1}^m \mathcal{L}(\chi_i)\delta\chi_i, \quad \delta\chi_i = \chi_{i-1} - \chi_i. \tag{5.3}$$

If the prior mass points χ_i are sampled in a logarithmic manner as $\chi_i = \prod_{j=1}^i t_j$ where $t_j \in (0,1)$ is a shrinkage ratio, then the nested algorithm consists of the following steps. The reader is referred to Skilling [28] for further details of the algorithm.

(1) Set $i = 0$, $\chi_0 = 1$, $Z_0 = 0$ and $f = 0.5$ (preset fraction used in the stopping criterion). Randomly draw M samples θ from the prior $\pi(\theta)$ to give an ensemble of samples. Evaluate the likelihood $\mathcal{L}(\theta)$ for each of the samples in the ensemble.

(2) Increase $i \rightarrow i+1$. Select the sample having the lowest likelihood (which we label as $\mathcal{L}_i$) in the ensemble and remove (discard) it. Shrink the prior mass to $\chi_i = \chi_{i-1}e^{-1/M}$.

(3) Draw a new sample θ from the prior $\pi(\theta)$ subject to the hard likelihood constraint $\mathcal{L}(\theta) > \mathcal{L}_i$, and add this sample to the ensemble of samples.

(4) Increment the evidence: $Z_i = Z_{i-1} + \mathcal{L}_i(\chi_{i-1} - \chi_i)$.

(5) If $\mathcal{L}_{\max}\chi_i < fZ_i$ ($\mathcal{L}_{\max}$ is the largest value of the likelihood in the ensemble of samples), add in contributions to the evidence Z_i from the ensemble of samples (remaining M samples that have not been discarded), and stop; otherwise, continue from Step 2.

In Step 5 of the algorithm, if the stopping criterion is satisfied, the estimate for the evidence $\widehat{Z}$ is completed by adding the contribution of the remaining M samples in the ensemble to Z_i; namely, $\widehat{Z} \rightarrow Z_i + M^{-1}(\mathcal{L}(\theta_1) + \mathcal{L}(\theta_2) + \cdots + \mathcal{L}(\theta_M))\chi_i$ where $\mathcal{L}(\theta_k)$ ($k = 1, 2, \ldots, M$) are the likelihood values evaluated at the remaining M samples. It is noted that the algorithm summarized above for estimation of Z (overlap integral of $\mathcal{L}(\theta)$ and $\pi(\theta)$) automatically provides weighted samples $\theta_k(m_N)$ drawn from the posterior distribution $p(\theta \mid m_N, \mathfrak{D}, I) \propto \mathcal{L}(\theta)\pi(\theta)$ (cf. (2.1) and (2.2)). More specifically, the ith discarded sample in Step 2 of the algorithm can be interpreted as a sample drawn from the posterior distribution of θ with weight given by $p_i = \mathcal{L}_i(\chi_{i-1} - \chi_i)/\widehat{Z}$ where $\widehat{Z}$ is the estimate of the evidence obtained in Step 5 on termination of the algorithm.

The key part of the algorithm is Step 3 involving drawing a sample from the prior $\pi(\theta)$ within a prescribed hard likelihood constraint $\mathcal{L}(\theta) > \mathcal{L}^*$. To this purpose, Feroz et al. [29] developed a very efficient algorithm (which the authors refer to as MULTINEST) for sampling from a prior within a hard likelihood constraint involving a very sophisticated procedure for decomposition of the support of the likelihood above a given bound $\mathcal{L}^*$ into a set of overlapping ellipsoids and then sampling from the resulting ellipsoids. The algorithm is appropriate for sampling from posterior distributions with multiple modes and with pronounced curving degeneracies in a high-dimensional parameter space.

Finally, the posterior odds ratio K_{10}^ϵ in (2.9) is a summary of the evidence for H_1 (additional source is present in the residuals) against H_0 (no source is present in the residuals). To interpret how strong (or weak) is the evidence for H_1 against H_0 provided by the residual data $\mathcal{E}$, we use a reference scale suggested by Jeffreys [30]. In this scale, $\log(K_{10}^\epsilon) < 1$ corresponds to inconclusive evidence for H_1 against H_0 (or the evidence for H_1 is not worth more than a bare mention and corresponds to a posterior odds ratio of less than about 3 to 1 in favor of H_1). On this same scale, $\log(K_{10}^\epsilon) > 5$ corresponds to very strong evidence for H_1 against H_0 (associated, as such, with a posterior odds ratio greater than about 150 to 1). In view of this scale for interpreting K_{10}^ϵ, we choose the threshold K_* for the posterior odds ratio (required to be specified in Step 2 of the algorithm summarized in Section 2) to be $\log(K_*) = 1$ (or, equivalently, $K_* \approx 2.7$).

6. Example: An Application of the Methodology

In this section, we apply the proposed methodology for inverse dispersion modeling of an unknown number of sources to a real dispersion data set obtained from a specific experiment conducted under the FUsing Sensor Information from Observing Networks (FUSION) Field Trial 2007 (FFT-07). The scientific objective of this field campaign was to acquire a comprehensive meteorological and dispersion dataset that can be used to validate

methodologies developed for source reconstruction. Details of the instrumentation deployed and the experiments conducted in FFT-07 are given in Storwold [31], so only a brief summary of FFT-07 will be presented here. In particular, only the relevant details of the particular experiment that are required for the interpretation of the results in this paper are emphasized.

The experiments in FFT-07 were carried out in September 2007 at Tower Grid on US Army Dugway Proving Ground, Utah about 2 km west of Camel Back Ridge on the Great Salt Lake Desert. The easterly through southerly drainage flows that predominate during the early morning hours at this site originate on the higher terrain to the southeast and are channeled by Camel Back Ridge. Generally, the terrain was flat, uniform, and homogeneous consisting primarily of short grass interspersed with low shrubs with a height between about 0.25 and 0.75 m. The momentum roughness length, z_0, was estimated to be $z_0 = 1.3 \pm 0.2$ cm.

In all the experiments, propylene (C_3H_6) was used as the tracer gas. The concentration detectors used were fast-response digital photoionization (dPID) detectors. In the experiments, a plume was formed in the atmospheric surface layer by releasing propylene from one or more (up to a maximum of four) purpose-designed gas dissemination systems. The network (or array) of concentration detectors consisted of up to 100 dPIDs, arranged in a staggered configuration of 10 rows of 10 detectors, with the rows of detectors separated by 50 m and the detectors along each row spaced 50 m apart. The concentration detectors along the ten sampling lines in the array were placed at a height, z_d, of 2.0 m.

In the experiment used to test the inverse dispersion modeling methodology proposed herein, tracer gas was released continuously over a period of 10 min from four source locations at a height, z_s, of 2.0 m: (1) source 1 is at $(x_s, y_s) = (33.0, 171.0)$ m; (2) source 2 is at $(x_s, y_s) = (33.8, 240.7)$ m; (3) source 3 is at $(x_s, y_s) = (30.0, 312.9)$ m; (4) source 4 is at $(x_s, y_s) = (26.0, 384.4)$ m. The coordinates reported here for the source locations are referenced with respect to a local Cartesian coordinate system. Unfortunately, in this experiment, only the mass flow controller for source 3 functioned properly. The mass flow controllers for sources 1, 2, and 4 failed to properly regulate the flow, so the emission rates from these sources were unknown.

A three-dimensional sonic anemometer, placed at the 2 m level of a lattice tower located upwind of the array of concentration detectors, was used to characterize the background micrometeorological state of the atmospheric surface layer. For this experiment, the horizontal mean wind speed S_2 at the 2 m level, the friction velocity (u_*), the atmospheric stability (Obukhov length L), and the standard deviations of the velocity fluctuations in the alongwind (σ_u), crosswind (σ_v), and vertical (σ_w) directions were $S_2 = 3.61$ m s^{-1}, $u_* = 0.282$ m s^{-1}, $L = -27.3$ m, $\sigma_u/u_* = 2.33$, $\sigma_v/u_* = 1.86$, and $\sigma_w/u_* = 1.10$. The mean wind direction in the experiment was normally incident on the detector array (namely, the mean wind direction corresponded to a wind from the $+x$-direction and, hence, is perpendicular to the sampling lines of detectors along the y-axis).

The wind velocity and turbulence statistics were used in conjunction with Monin-Obukhov similarity theory relationships [32] as input to the atmospheric dispersion model (namely, the backward-time LS particle trajectory model described briefly in Section 3) for provision of the predicted concentration $\tilde{C}$ (cf. (3.4)). More specifically, the backward-time LS model of (3.5) was applied with the Kolmogorov constant $C_0 = 4.8$, a value which was recommended by Wilson et al. [33] from a calibration of the model against concentration data obtained from the Project Prairie Grass atmospheric dispersion experiments.

The example used here to illustrate the inverse dispersion methodology involve continuously emitting sources, so the relevant source parameters for source model m_N are $\theta \equiv \theta(m_N) = (\mathbf{x}_{s,1}, Q_1, \ldots, \mathbf{x}_{s,N}, Q_N)$. Furthermore, it is assumed that the height of the sources

above ground level (z_s = 2.0 m) is known *a priori*, so the only unknown location parameters are the (x_s, y_s) coordinates of the sources. Owing to the horizontal homogeneity of the mean flow and turbulence statistics in the current example, the adjunct concentration $C^*(\mathbf{x}_s \mid \mathbf{x}_d)$ in (3.4) can be precalculated for one detector location $\mathbf{x}_d$ (for the known source height z_s), with the adjunct concentration C^* (considered as a function of $\mathbf{x}_s$) at all other detector locations obtained simply by a linear translation of $C^*(\mathbf{x}_s \mid \mathbf{x}_d)$ in the horizontal (x, y)-plane.

In order to calculate the likelihood function given by (4.6), we need to provide values for the estimated standard deviations s_i for the noise e_i. The noise error variance σ_i^2 includes the sensor sampling error variance, $\sigma_{d,i}^2$, in the measurement of the concentration datum d_i and the model error variance, $\sigma_{m,i}^2$, in the prediction of $\tilde{C}_i$. The two contributions are combined in quadrature to give the noise error variance $\sigma_i^2 = \sigma_{d,i}^2 + \sigma_{m,i}^2$. The measurement error standard deviation is estimated as $\hat{\sigma}_{d,i} \equiv s_{d,i} = \max(0.05, 0.02d_i)$ ppm (parts per million by volume) where the lower limit of 0.05 ppm represents the precision in the concentration measurements using the dPIDs. The model error standard deviation is estimated to be 20% of the predicted value of the concentration $\tilde{C}_i(\theta)$ (namely, $\hat{\sigma}_{m,i} \equiv s_{m,i} = 0.20\tilde{C}_i(\theta)$). In consequence, the estimated noise error standard deviation is given by $\hat{\sigma}_i \equiv s_i = (s_{d,i}^2 + s_{m,i}^2)^{1/2}$.

The algorithm described in Section 2 was used for the source reconstruction, applied to the concentration data $\mathfrak{D} \equiv \{d_1, d_2, \ldots, d_{N_m}\}$ obtained from 62 detectors in the array (N_m = 62) (see Figure 1 which shows the locations of the 62 detectors as filled squares). The hyperparameters defining the prior $\pi(\theta)$ have been chosen as follows: $Q_j \sim \mathcal{U}([0, Q_{max}])$ with Q_{max} = 100 g s^{-1} and $(x_{s,j}, y_{s,j}) \sim \mathcal{U}(\mathcal{R})$ with $\mathcal{R}$ = [0, 100]m × [0, 500] m for $j = 1, 2, \ldots, N$.

After Steps 2 and 3 of the algorithm were completed, we found $\log(K_{10}) = \log(Z(m_1)/Z(m_0))$ = 64.0 ± 0.2 implying that the concentration data strongly support the source model m_1 (one localized source or "signal") against the alternative source model m_0 (no localized source or "no signal"). In other words, at this point in the algorithm, a source for the concentration has been detected in the data $\mathfrak{D}$. Owing to the fact that $\log(K_{10}) > \log(K_*)$ = 1, the algorithm continues to the iterative loop involving Steps 4 to 6. This iterative loop executes four times before it is terminated. The results of the execution of each of these four iterations are exhibited in Figure 1(a) (N = 1), Figure 1(b) (N = 2), Figure 1(c) (N = 3), and Figure 1(d) (N = 4). The caption in Figure 1 summarizes the values for $\log(K_{10}^e) = \log(Z^e(m_1)/Z^e(m_0))$ obtained from testing the alternative hypothesis H_1 (additional source present in the residual data) against the "null" hypothesis H_0 (no additional source present in the residual data). Each panel in Figure 1 also displays a density plot of samples of the source model m_N drawn from the posterior distribution $p(\theta \mid m_N, \mathfrak{D}, I)$ and projected onto the (x, y) horizontal plane (namely, each point in the plot corresponds to the (x, y) location of a sample of the source model m_N drawn from the posterior distribution). Note that the algorithm terminates with N_* = N = 4 with $\log(K_{10}^e)$ = 0.92 ≤ $\log(K_*)$ = 1, providing the correct number of localized sources in the source distribution (N_* = 4 in this case).

The samples of source models m_N (N = 4) drawn from the posterior distribution $p(\theta \mid m_N, \mathfrak{D}, I)$ in the last iteration of the algorithm (see Figure 1(d)) can be used to determine the characteristics (location, emission rate) of the four localized sources. Towards this objective, Figure 2 displays the univariate (diagonal) and bivariate (off-diagonal) marginal posterior distributions for the parameters of each source. For each univariate marginal posterior distribution of a parameter for a given source, the solid vertical line indicates the true value of the parameter (for this source), and the dashed vertical line corresponds to the best estimate of the parameter (for this source) obtained as the posterior mean. For the bivariate
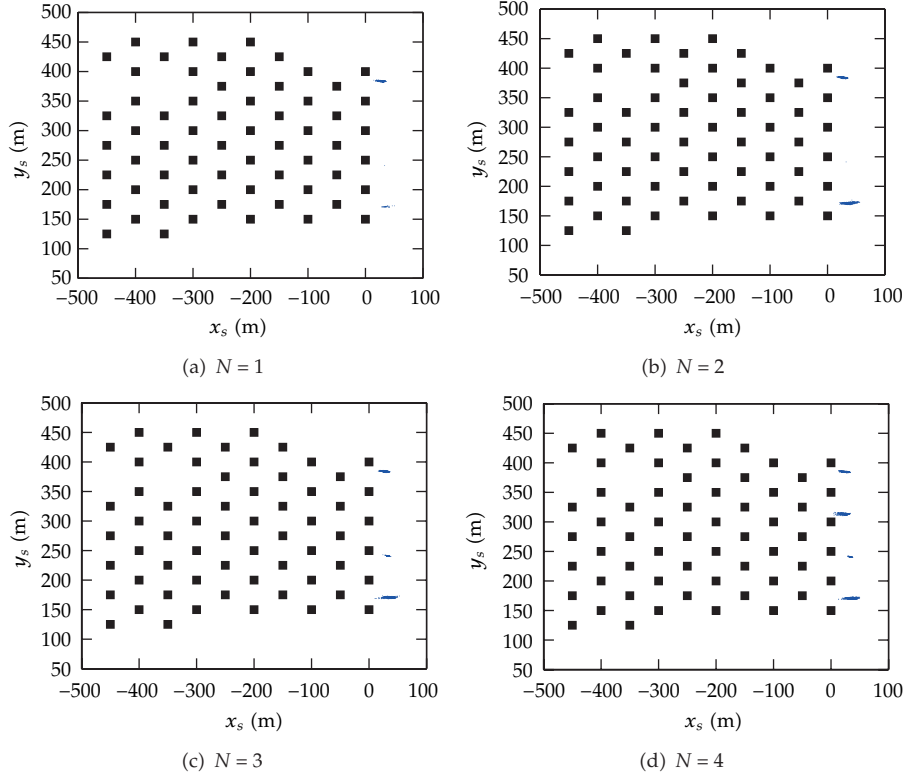
Figure 1: Four iterations of the algorithm of Section 2 showing the samples of source model m_N drawn from the posterior distribution $p(\theta \mid m_N, \mathfrak{D}, I)$ and the computation of K_{10}^e for (a) $N = 1$, $\log(K_{10}^e) = 62.4 \pm 0.2$; (b) $N = 2$, $\log(K_{10}^e) = 57.3 \pm 0.3$; (c) $N = 3$, $\log(K_{10}^e) = 39.1 \pm 0.3$; (d) $N = 4$, $\log(K_{10}^e) = 0.92 \pm 0.33$. A solid square denotes a detector.

Table 1: The posterior mean, posterior standard deviation, and lower and upper bounds of the 95% HPD interval of the parameters $x_{s,j}$ (m), $y_{s,j}$ (m), and $Q_j \equiv q_{s,j}$ (g s^{-1}) for $j = 1$, 2, 3, and 4 calculated from samples of source model m_N with $N = 4$ drawn from the posterior distribution $p(\theta|m_N, \mathfrak{D}, I)$ [cf. Figure 1(d)].

	Parameter	Mean	Standard deviation	95% HPD	Actual
$j = 1$	x_s (m)	34.9	4.3	(27.6, 44.4)	33.0
	y_s (m)	170.8	0.6	(169.5, 171.8)	171.0
	q_s (g s^{-1})	8.9	0.4	(8.0, 9.3)	—
$j = 2$	x_s (m)	33.7	1.1	(31.6, 35.9)	33.8
	y_s (m)	240.7	0.3	(240.2, 241.4)	240.7
	q_s (g s^{-1})	7.5	0.4	(6.7, 8.1)	—
$j = 3$	x_s (m)	25.4	3.9	(17.1, 31.2)	30.0
	y_s (m)	313.2	0.6	(311.9, 314.6)	312.9
	q_s (g s^{-1})	4.2	0.2	(3.7, 4.6)	3.8
$j = 4$	x_s (m)	25.1	3.0	(18.9, 30.3)	26.0
	y_s (m)	384.5	0.5	(383.5, 385.5)	384.4
	q_s (g s^{-1})	6.3	0.3	(5.8, 6.8)	—

(a) Source 1
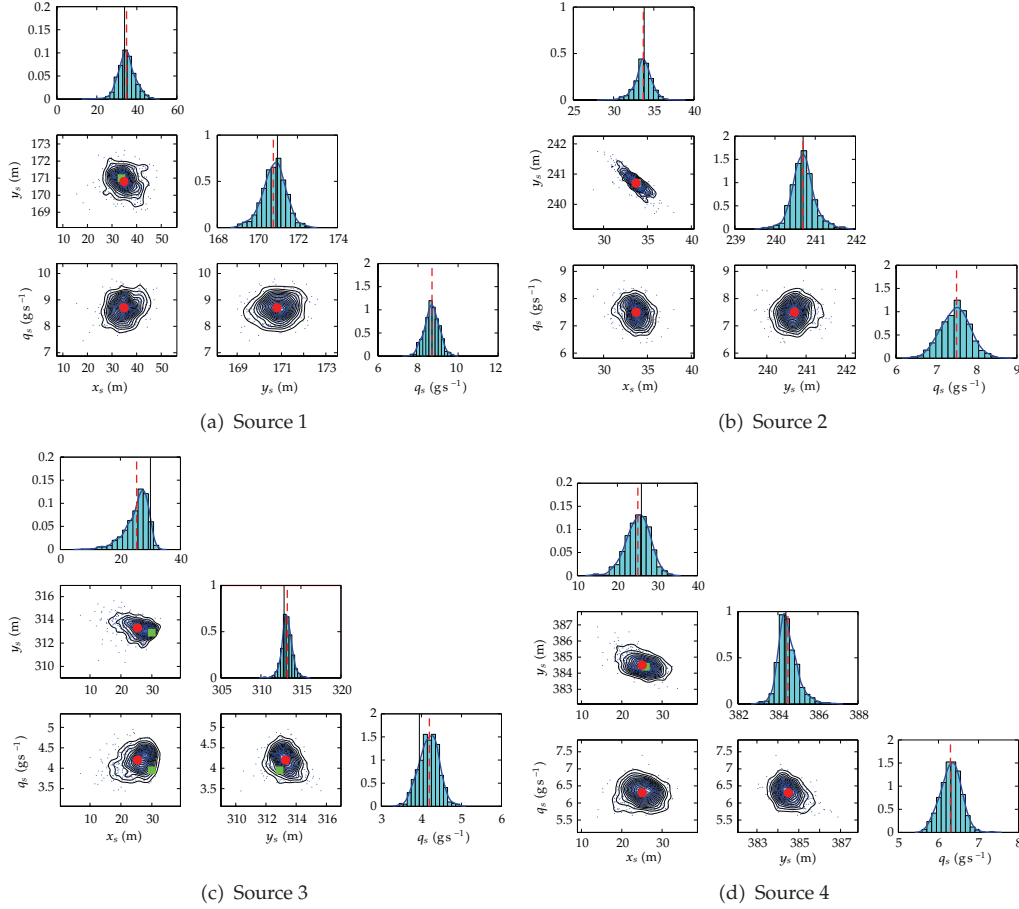
(b) Source 2

(c) Source 3

(d) Source 4

Figure 2: Univariate (diagonal) and bivariate (off-diagonal) marginal posterior distributions of the source parameters, namely, alongwind location x_s, crosswind location y_s, and emission rate q_s that characterize sources 1 (a), 2 (b), 3 (c), and 4 (d) obtained from samples of source model m_N ($N = 4$) drawn from the posterior distribution $p(\theta \mid m_N, \mathfrak{D}, I)$ ([cf. Figure 1(d)).

marginal posterior distribution of parameters of a given source, the solid square represents the position of the true source parameter values, and the solid circle indicates the best estimate of the true source parameter values obtained as the posterior means. The posterior mean, posterior standard deviation, and the lower and upper bounds for the 95% highest posterior distribution (HPD) interval of the source parameters for each identified source are summarized in Table 1. Comparing these estimated values of the source parameters, it can be seen that the proposed algorithm has adequately recovered the true parameters for each source (if these are known) to within the stated uncertainties.

7. Conclusions

In this paper, we have proposed a Bayesian inference approach for addressing the inverse dispersion of an unknown number of sources using model comparison (selection). The necessary integrations to compute the evidence $Z(m_N)$ for $N > 1$ can be very

computationally demanding (as well as technically difficult), and, as a consequence, we developed an efficient and robust algorithm for model comparison that recursively removes the influence of a source model m_N from the measured concentration data $\mathfrak{D}$ and tests the resulting residual data $\mathcal{E}$ to determine if the residual data are consistent with the estimated noise level. This test requires nothing more than the computation of the evidence $Z^\epsilon(m_N)$ for $N = 1$ which is usually computationally simple. The procedure finds the minimum number of localized sources necessary to represent the concentration signal in the data $\mathfrak{D}$ down to the estimated noise level. Furthermore, the uncertainty in the estimated noise level (which includes contributions from both measurement and model errors) is treated by using an integrated likelihood, which was implemented using a specific prior (inverse gamma distribution) to represent the uncertainty in the estimated noise level. Nested sampling is used for the evidence computation, as well as for sampling from the posterior distribution $p(\theta \mid m_N, \mathfrak{D}, I)$ in the proposed algorithm for inverse dispersion modeling.

The new algorithm has been applied successfully to a real concentration dataset obtained from an atmospheric dispersion experiment conducted in the FFT-07 field campaign consisting of the simultaneous release of a tracer from four sources. It is shown that the proposed algorithm performed well for this example: the number of sources was determined correctly ($N_* = 4$), and for each of the identified sources the corresponding parameters (e.g., location, emission rate) were estimated reliably along with the determination of the uncertainty in the parameter (in the form of either a standard deviation or a credible interval that encloses a prespecified posterior probability mass). The methodology proposed herein offers a simpler alternative for the inverse dispersion modeling of an unknown number of sources that was addressed previously [13, 14] as a generalized parameter estimation problem, the latter of which involves necessarily the complexities in the design of an appropriate reversible-jump MCMC sampling algorithm to allow transdimensional jumps in the generalized parameter space (namely, jumps between the parameter space of source models m_N of different dimensions involving different numbers of sources N).

References

[1] L. Porta, T. H. Illangasekare, P. Loden, Q. Han, and A. P. Jayasumana, "Continuous plume monitoring using wireless sensors: proof of concept in intermediate scale tank," *Journal of Environmental Engineering*, vol. 135, no. 9, pp. 831–838, 2009.

[2] L. Robertson and J. Langner, "Source function estimate by means of variational data assimilation applied to the ETEX-I tracer experiment," *Atmospheric Environment*, vol. 32, no. 24, pp. 4219–4225, 1998.

[3] M. Bocquet, "Reconstruction of an atmospheric tracer source using the principle of maximum entropy. I: theory," *Quarterly Journal of the Royal Meteorological Society*, vol. 131, no. 610 B, pp. 2191–2208, 2005.

[4] L. C. Thomson, B. Hirst, G. Gibson et al., "An improved algorithm for locating a gas source using inverse methods," *Atmospheric Environment*, vol. 41, no. 6, pp. 1128–1134, 2007.

[5] C. T. Allen, G. S. Young, and S. E. Haupt, "Improving pollutant source characterization by better estimating wind direction with a genetic algorithm," *Atmospheric Environment*, vol. 41, no. 11, pp. 2283–2289, 2007.

[6] J. P. Issartel, M. Sharan, and S. K. Singh, "Identification of a point source by use of optimal weighted least squares," *Pure and Applied Geophysics*, vol. 169, pp. 467–482, 2012.

[7] A. Keats, E. Yee, and F. S. Lien, "Bayesian inference for source determination with applications to a complex urban environment," *Atmospheric Environment*, vol. 41, no. 3, pp. 465–479, 2007.

[8] A. Keats, E. Yee, and F. S. Lien, "Efficiently characterizing the origin and decay rate of a nonconservative scalar using probability theory," *Ecological Modelling*, vol. 205, no. 3-4, pp. 437–452, 2007.

[9] F. K. Chow, B. Kosović, and S. Chan, "Source inversion for contaminant plume dispersion in urban environments using building-resolving simulations," *Journal of Applied Meteorology and Climatology*, vol. 47, no. 6, pp. 1533–1572, 2008.

[10] E. Yee, F. S. Lien, A. Keats, and R. D'Amours, "Bayesian inversion of concentration data: source reconstruction in the adjoint representation of atmospheric diffusion," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 96, no. 10-11, pp. 1805–1816, 2008.

[11] E. Yee, "Bayesian probabilistic approach for inverse source determination from limited and noisy chemical or biological sensor concentration measurements," in *Chemical and Biological Sensing VIII*, A. W. Fountain III, Ed., vol. 6554, 65540W, of *Proceedings of the SPIE*, p. 12, April 2007.

[12] M. Sharan, S. K. Singh, and J. P. Issartel, "Least square data assimilation for identification of the point source emissions," *Pure and Applied Geophysics*, vol. 169, pp. 483–497, 2012.

[13] E. Yee, "Theory for reconstruction of an unknown number of contaminant sources using probabilistic inference," *Boundary-Layer Meteorology*, vol. 127, no. 3, pp. 359–394, 2008.

[14] E. Yee, "Probability theory as logic: data assimilation for multiple source reconstruction," *Pure and Applied Geophysics*, vol. 169, pp. 499–517, 2012.

[15] D. J. Thomson, "Criteria for the selection of stochastic models for particle trajectories in turbulent flows," *Journal of Fluid Mechanics*, vol. 180, pp. 529–556, 1987.

[16] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *Journal of the Royal Statistical Society B*, vol. 63, no. 3, pp. 425–464, 2001.

[17] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and analysis of computer experiments," *Statistical Science*, vol. 4, no. 4, pp. 409–435, 1989.

[18] D. Higdon, J. Gattiker, B. Williams, and M. Rightley, "Computer model calibration using high-dimensional output," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 570–583, 2008.

[19] 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, 1995.

[20] E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, Mass, USA, 2003.

[21] H. Haario, E. Saksman, and J. Tamminen, "An adaptive Metropolis algorithm," *Bernoulli*, vol. 7, no. 2, pp. 223–242, 2001.

[22] C. J. F. Ter Braak, "A Markov chain Monte Carlo version of the genetic algorithm differential evolution: easy Bayesian computing for real parameter spaces," *Statistics and Computing*, vol. 16, no. 3, pp. 239–249, 2006.

[23] J. A. Vrugt, C. J. F. Ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon, "Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling," *International Journal of Nonlinear Sciences and Numerical Simulation*, vol. 10, no. 3, pp. 273–290, 2009.

[24] J. S. Liu, F. Liang, and W. H. Wong, "The multiple-try method and local optimization in Metropolis sampling," *Journal of the American Statistical Association*, vol. 95, no. 449, pp. 121–134, 2000.

[25] J. Berntsen, T. O. Espelid, and A. Genz, "An adaptive algorithm for the approximate calculation of multiple integrals," *ACM Transactions on Mathematical Software*, vol. 17, no. 4, pp. 437–451, 1991.

[26] M. A. Newton and A. E Rafferty, "Approximate Bayesian inference by the weighted likelihood bootstrap," *Journal of the Royal Statistical Society B*, vol. 3, pp. 3–48, 1994.

[27] A. Gelman and X.-L. Meng, "Simulating normalizing constants: from importance sampling to bridge sampling to path sampling," *Statistical Science*, vol. 13, no. 2, pp. 163–185, 1998.

[28] J. Skilling, "Nested sampling for general Bayesian computation," *Bayesian Analysis*, vol. 1, no. 4, pp. 833–859, 2006.

[29] F. Feroz, M. P. Hobson, and M. Bridges, "MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics," *Monthly Notices of the Royal Astronomical Society*, vol. 398, no. 4, pp. 1601–1614, 2009.

[30] H. Jeffreys, *Theory of Probability*, Oxford University Press, Oxford, UK, 3rd edition, 1961.

[31] D. P. Storwold, *Detailed Test Plan for Fusing Sensor Information from Observing Networks (FUSION) Field Trial 2007 (FFT-07)*, WDTC-TP-07-078, West Desert Test Center, US Army Dugway Proving Ground, 2007.

[32] R. B. Stull, *An Introduction to Boundary-Layer Meteorology*, Kluwer Academic Publishers, 1988.

[33] J. D. Wilson, T. K. Flesch, and L. A. Harper, "Micro-meteorological methods for estimating surface exchange with a disturbed windflow," *Agricultural and Forest Meteorology*, vol. 107, no. 3, pp. 207–225, 2001.