Efficient MCMC Inference for Remote Sensing of Emission Sources

by

Christopher L. Dean

B.S. Computer Science and Engineering
The Ohio State University, 2013

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science
in Electrical Engineering and Computer Science
at the Massachusetts Institute of Technology

June 2015

Signature of Author: _____

Department of Electrical Engineering and Computer Science
May 20, 2015

Certified by: _____

John W. Fisher III
Senior Research Scientist
Thesis Supervisor

Accepted by: _____

Professor Leslie A. Kolodziejski
Chair, Department Committee on Graduate Students

Efficient MCMC Inference for Remote Sensing of Emission Sources

by

Christopher L. Dean

Submitted to the Department of Electrical Engineering and Computer Science
on May 20, 2015 in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

A common challenge in environmental impact studies and frontier exploration is identifying the properties of some emitters from remotely obtained concentration data. For example, consider estimating the volume of some pollutant that a chemical refinery releases into the atmosphere from measurements of pollutant concentrations. Previous methods assume a known number of emitters, low ambient concentrations, or measurements from a group of stationary sensors. In contrast, we use measurements from a mobile sensor and detect source contributions that are several orders of magnitude smaller than ambient concentrations.

Here, we develop and analyze a method for inferring the location, emission rate, and number of emitters from measurements taken by an aircraft. We use Reversible-jump Markov chain Monte Carlo sampling to jointly infer the posterior distribution of the number of emitters and emitter properties. Additionally, we develop performance metrics that can be efficiently computed using the sample-based representation of the posterior distribution.

We investigate the expected performance of the inference algorithm with respect to certain model parameters in a series of synthetic experiments and use these performance metrics for evaluation. These experiments provide insight into subtleties of the model, including the identifiability of source configurations, the effect of various path geometries, and the effects of incorporating data from multiple flights. We also provide intuition for best-case performance when running on real-world data using a synthetic experiment. Finally, we demonstrate our ability to process and analyze real-world data for which the true source configuration is unknown.

Thesis Supervisor: John W. Fisher III
Title: Senior Research Scientist

Acknowledgments

First, I would like to thank my advisor, John Fisher. His guidance over the past two years has been truly invaluable and I am very lucky to have had the opportunity to work with him and the other members of the Sensing, Learning, and Inference group at MIT. Among my labmates, I would especially like to thank Randi Cabezas, Julian Straub, and Sue Zheng for all of their help what feels like all of the time. Guy Rosman has also provided considerable insights and a wealth of new ideas as I have been working on this project. Thanks are also due to Hossein Mobahi for his help with the optimization procedure.

I would also like to thank the Office of Naval Research and the Army Research Office Multidisciplinary Research Initiative programs and Royal Dutch Shell through the MIT Energy Initiative for supporting aspects of the research presented in this thesis. Special thanks also go out to Bill Hirst, David Randell, Philip Jonathan, and Matthew Jones for their assistance in gaining a detailed understanding of their work, on which this is largely based.

Finally, none of this would have been possible without my wife Laura. Her constant support, encouragement, and enthusiasm was truly invaluable.

Contents

List of Figures

List of Algorithms

List of Tables

Chapter 1

Introduction

Remote sensing is the science of obtaining information about objects or areas from a distance, e.g. from an aircraft or spacecraft. A variety of data may be sensed remotely, including satellite imagery, radar, and airborne light detection and ranging (LIDAR) measurements. These types of data are then processed and used for a wide range of applications, including oceanography, disaster response, meteorology, and military operations. In particular, one type of object that may be sensed remotely is an emission source. These come in many forms, including chimneys, landfills, explosions, and waste-water treatment exit areas. We can sense these objects remotely by recording local concentrations of some pollutant and inferring emitter properties from these measurements.

This corresponds to solving an *inverse problem* in the atmospheric sciences literature. In contrast, the *forward* problem pertains to mapping the flow of particles of some pollutant after they are ejected by an emitter. Specifically, the forward problem can be stated as follows: given the location and other properties of an emitter, what is the concentration that would be measured by a sensor at some location downwind of the emitter? A variety of methods and models exist to describe this mapping. The inverse problem, and the one to which we will restrict our attention, asks the following question: given a set of concentration measurements obtained at a set of locations, where is the emitter and what are its properties? This problem provides a rich setting for probabilistic inference.

Probabilistic inference is concerned with inferring the values of some unobservable random variables given the observed values of some other variables. One commonly-used approach to probabilistic inference is Markov chain Monte Carlo (MCMC) methods, which allows us to generate a set of samples from the posterior distribution and compute statistics about these variables using this set of samples. One key issue in such approaches is determining the convergence rate of the set of samples to the correct distribution. We apply an MCMC method to solve the aforementioned inference problem efficiently. Critically, the approach described here is not restricted to the specific application of atmospheric emissions. Instead, the results and modeling considerations may be extended to any type of emissions propagating through any type of fluid, provided an appropriate transport model is available.

To enable in-depth discussion of the atmospheric dispersion application, Chapter 2

provides a primer on atmospheric dispersion models, an introduction to MCMC methods, and an overview of relevant prior work in solving the inverse problem. The probabilistic model is formulated in Chapter 3 and inference on the model is discussed in Chapter 4. Finally, Chapter 5 develops a set of performance metrics for analyzing system performance. Chapter 6 uses these metrics and presents a series of synthetic experiments that reveal certain properties of the model and inference procedure. We also present the results from processing a real-world data set.

Chapter 2

Background

In this chapter we provide a brief primer on atmospheric dispersion models and previous formulations of the emission source detection problem. Although the topics described here are restricted to problems in the atmospheric sciences, and particle transport in particular, we stress that the inference algorithms discussed can be applied to other problem domains.

2.1 Atmospheric dispersion models

Atmospheric dispersion models attempt to capture the physical processes associated with particle transport through the atmosphere. Gaseous particles, which are emitted from some source, rise into the atmosphere due to both particle momentum, related to the ejection velocity, and heat-related particle buoyancy. Particles eventually level off due to the effects of gravity and cooling of the gasses. Transport and dispersion through the atmosphere is caused by the wind, which consists of both large and small-scale eddies, or currents. The large scale eddies are referred to as "wind" and transport the particles in the downwind direction as a group. The smaller-scale eddies, called "turbulence", cause the particles to disperse randomly in the cross-wind and vertical directions. This transport and dispersion process is affected by a variety of factors, including the height of the atmospheric boundary layer—the layer of the atmosphere in which most gasses are contained—and the roughness of the ground [8].

We first consider a plume emitted by some emission source. The emission source may be located at altitude, in which case it is called a stack, or located at ground level. As mentioned earlier, the plume rises and stabilizes at a certain height, fans out in the horizontal and vertical directions, and has a shape that fluctuates randomly. This fluctuation is due to the turbulence acting on the plume and is well-represented by a random process. However, we can compute an ensemble average pollutant concentration at some location averaged over some short period of time. This is the high-level goal of atmospheric dispersion models, which predict the concentration of a pollutant at some measurement time and location given an emitter with certain properties (e.g. location and emission schedule) and information about the atmosphere through which the diffusion occurs.

In this section we will introduce three different dispersion models of varying levels of

complexity. We first discuss the Gaussian plume dispersion model, which represents the concentration profile at a point downwind from the source as a Gaussian distribution. Despite its simplicity, the Gaussian plume model can capture all the factors affecting plume transport and diffusion with the exception of obstacles and changing landscapes. Next we provide a review Lagrangian simulation, which samples particle trajectories as a stochastic process and computes ensemble-averaged statistics over the simulated trajectories. Finally, we will introduce a Computational Fluid Dynamics model which implements the Navier-Stokes equations and includes some approximations of the small-scale turbulence eddies. In subsequent chapters we utilize the Gaussian plume model; the other two dispersion models are presented for context.

■ 2.1.1 Gaussian plume model

In general, the release and transport of some pollutant through the atmosphere is described by the *advection-diffusion equation* (ADE):

$$\frac{\partial C}{\partial t} + \boldsymbol{u} \cdot \nabla C = \nabla \cdot (\mathbf{K} \nabla C) + S \tag{2.1}$$

where $C(\boldsymbol{x}, t)$ gives the concentration observed at location $\boldsymbol{x}$ at time t, $\boldsymbol{u}$ is the wind velocity, $\mathbf{K}$ is a diagonal matrix of turbulent eddy diffusivities, and $S(\boldsymbol{x}, t)$ gives the source emission rate. A closed-form solution can be computed in many cases, depending on boundary constraints and turbulence diffusivities. Stockie [31] steps through the derivation of the Gaussian plume model from the ADE, Equation 2.1, by applying a particular set of assumptions. In particular, the Gaussian plume model is only valid under these assumptions:

- The source is emitting at a constant rate.
- The source is a mathematical point (i.e. has no area).
- Wind speed and direction are constant across space and time.
- Statistics of turbulence are constant across space and time.
- The wind speed is sufficiently strong such that dispersion in the downwind direction is negligible compared to the advection (plume rise).
- Mass is conserved within the plume, or the pollutant neither deposits onto the ground nor undergoes chemical reaction within the atmosphere.

Any practical application will see nearly all of these assumptions violated. However, they are often *approximately* satisfied, in which case the Gaussian plume model is expected to be accurate to within a factor of two [8].

The Gaussian plume model, shown in Figure 2.1, predicts the expected concentration at a measurement location given some emitter location. Since the plume shape

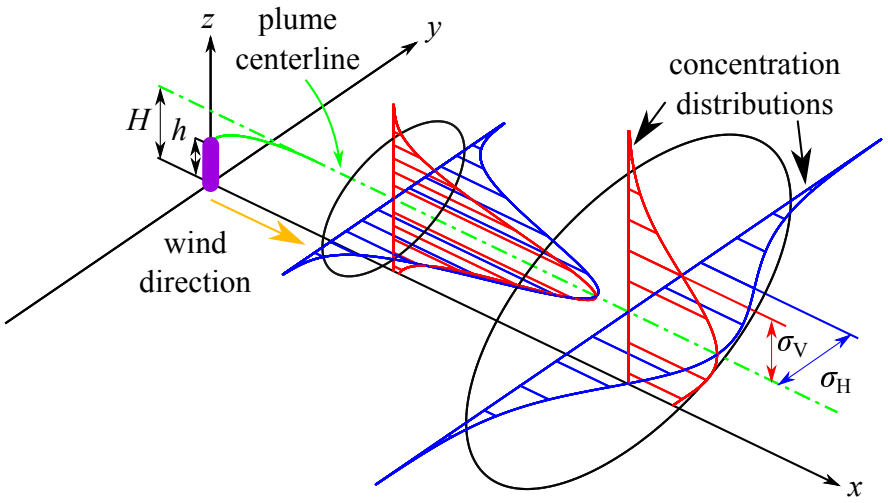

Figure 2.1: An illustration of the Gaussian plume model. The source has height h and the plume ascends to a stable height H. At some distance x downwind of the source, the horizontal and vertical concentration profiles are modeled as independent Gaussians with standard deviations $\sigma_{\mathrm{H}}(x)$ and $\sigma_{\mathrm{V}}(x)$, respectively. We also refer to these as σ_y and σ_z when the y-axis is in the cross-wind direction.

fluctuates randomly due to turbulence, the Gaussian plume model computes an ensemble average over some sample time. Given a long enough sample time, the ensemble average of the concentration at some downwind distance x is represented by a Gaussian distribution in both the crosswind (y) and vertical (z) direction. We cannot obtain an instantaneous concentration measurement profile from the Gaussian plume model as it cannot consider temporally-varying emission rates; more complicated models will be described for this task in later sections. By convention we place the source at $(0, 0, 0)$ and assume the relative measurement location (x, y, z). The expected concentration at this location, $C(x, y, z)$ factors into the downwind, cross-wind, and vertical coordinate directions as follows:

$$C(x, y, z) = \phi_x \phi_y \phi_z, \tag{2.2}$$

where the $C(\cdot, \cdot, \cdot)$ is unitless, ϕ_x has units of m^3/m, and ϕ_y and ϕ_z both have units of $1/\mathrm{m}$.

In the x direction the plume is deterministically diluted by the wind. The source is assumed to be emitting at some constant rate Q in m^3/s. This concentration is diluted downwind by a wind of speed u in m/s; thus, we model the amount of pollutant per meter in the downwind direction by

$$\phi_x = \frac{Q}{u}. \tag{2.3}$$

The volume of pollutant ϕ_x is then randomly and independently distributed in both the crosswind and vertical directions. This physically corresponds to the turbulence that

causes the atoms to disperse. The extent of the diffusion is dictated by the parameters σ_y and σ_z, the standard deviation of the ensemble average in both the crosswind and vertical directions. A variety of methods exist for computing the values of these parameters. We will consider several common methods later in this section. In the crosswind direction, at some distance y meters from the plume center-line, the ensemble-averaged concentration profile is modeled by

$$\phi_y = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left\{-\frac{y^2}{2\sigma_y^2}\right\}. \tag{2.4}$$

In the vertical direction the observed concentration is also distributed normally when vertical diffusion is unbounded. In such unbounded cases the vertical concentration profile is

$$\phi_z = \frac{1}{\sqrt{2\pi}\sigma_z} \exp\left\{-\frac{(z-H)^2}{2\sigma_z^2}\right\} \tag{2.5}$$

where $H = h + \Delta h$ is the effective source height, h is the physical source height, and Δh is the plume rise. In reality, the plume does not experience unbounded dispersion in the vertical direction. The ground provides one such bound, and since most pollutants deposit onto the ground very slowly, we can conservatively treat the pollutant as being reflected off the ground. This manifests as a virtual source located at height $-H$, such that the concentrations from this virtual source intersect the ground at exactly the same downwind distance as the actual source. Equation (2.5) then becomes

$$\phi_z = \frac{1}{\sqrt{2\pi}\sigma_z} \left(\exp\left\{-\frac{(z-H)^2}{2\sigma_z^2}\right\} + \exp\left\{-\frac{(z+H)^2}{2\sigma_z^2}\right\}\right). \tag{2.6}$$

Using Equation (2.6) instead of (2.5) in (2.2) yields the *classical Gaussian plume equation*

$$a(x,y,z) = \frac{Q}{2\pi u \sigma_y \sigma_z} \exp\left\{-\frac{y^2}{2\sigma_y^2}\right\} \left(\exp\left\{-\frac{(z-H)^2}{2\sigma_z^2}\right\} + \exp\left\{-\frac{(z+H)^2}{2\sigma_z^2}\right\}\right). \tag{2.7}$$

The vertical diffusion is also bounded from above by a physical phenomenon called temperature inversion. The region between the ground and the temperature inversion layer is often called the mixing layer. The depth D of this mixing layer is often greater in the daytime than in the evening primarily due to atmospheric heating, but other factors also play a role. When the plume is confined to this mixing layer, an infinite number of reflections may occur. As such, we can further extend Equation (2.6) to

$$\begin{aligned}\phi_z =& \frac{1}{\sqrt{2\pi}\sigma_z} \\ &\times \sum_{j=-\infty}^{\infty} \left(\exp\left\{-\frac{(z-(H+2jD))^2}{2\sigma_z^2}\right\} + \exp\left\{-\frac{(z+(H+2jD))^2}{2\sigma_z^2}\right\}\right).\end{aligned} \tag{2.8}$$

	Day			Night	
u (m/s)	Strong	Moderate	Slight	Cloudy ($\geq 4/8$)	Clear ($\leq 3/8$)
< 2	A	A–B	B	E	F
2–3	A–B	B	C	E	F
3–5	B	B–C	C	D	E
5–6	C	C–D	D	D	D
> 6	C	D	D	D	D

Table 2.1: Criteria for Pasquill-Gifford stability classes. The above division into "Cloudy" and "Clear" is based on the fractional cloud cover, ($\geq$ 4/8) and ($\leq$ 3/8) respectively.

In practice only a few of the infinitely many terms in Equation (2.8) need be calculated. In general, computing the terms for j from -2 to 2 is sufficient [8].

Concentrations obtained by the Gaussian plume equation are highly dependent on accurate estimations of the dispersion parameters σ_y and σ_z. We will now present two different approaches for computing these dispersion parameters.

Parameterizations based on stability classes

The first of the two methods classifies the environment into "stability classes" which have different empirical estimates for σ_y and σ_z. Although stability-classes are no longer the preferred method for computing dispersion parameters they are still commonly used in many pollution screening applications. The most commonly-used notion of stability classes was developed by Pasquill [24] and Gifford [11]. A table for computing the stability class is shown in Table 2.1. Each stability class corresponds to curves of σ_y and σ_z versus the downwind distance x. The dispersion parameter values were originally read from graphs presented by Turner [32]; the graphs are reproduced here in Figure 2.2.

Many empirical equations have been set forth to describe these curves. Those proposed by Briggs [2] have seen the greatest popularity. These equations take the form

$$\sigma_y = \frac{ax}{(1+bx)^c} \tag{2.9}$$

$$\sigma_z = \frac{dx}{(1+ex)^f} \tag{2.10}$$

where $a,\ldots,f$ are empirical constants given in Table 2.2 and x is the downwind distance. The measurement data used to derive the Briggs equations were based on 10-minute averages, but they are usually treated as hour-long averages according to the standards established by the U.S. EPA.

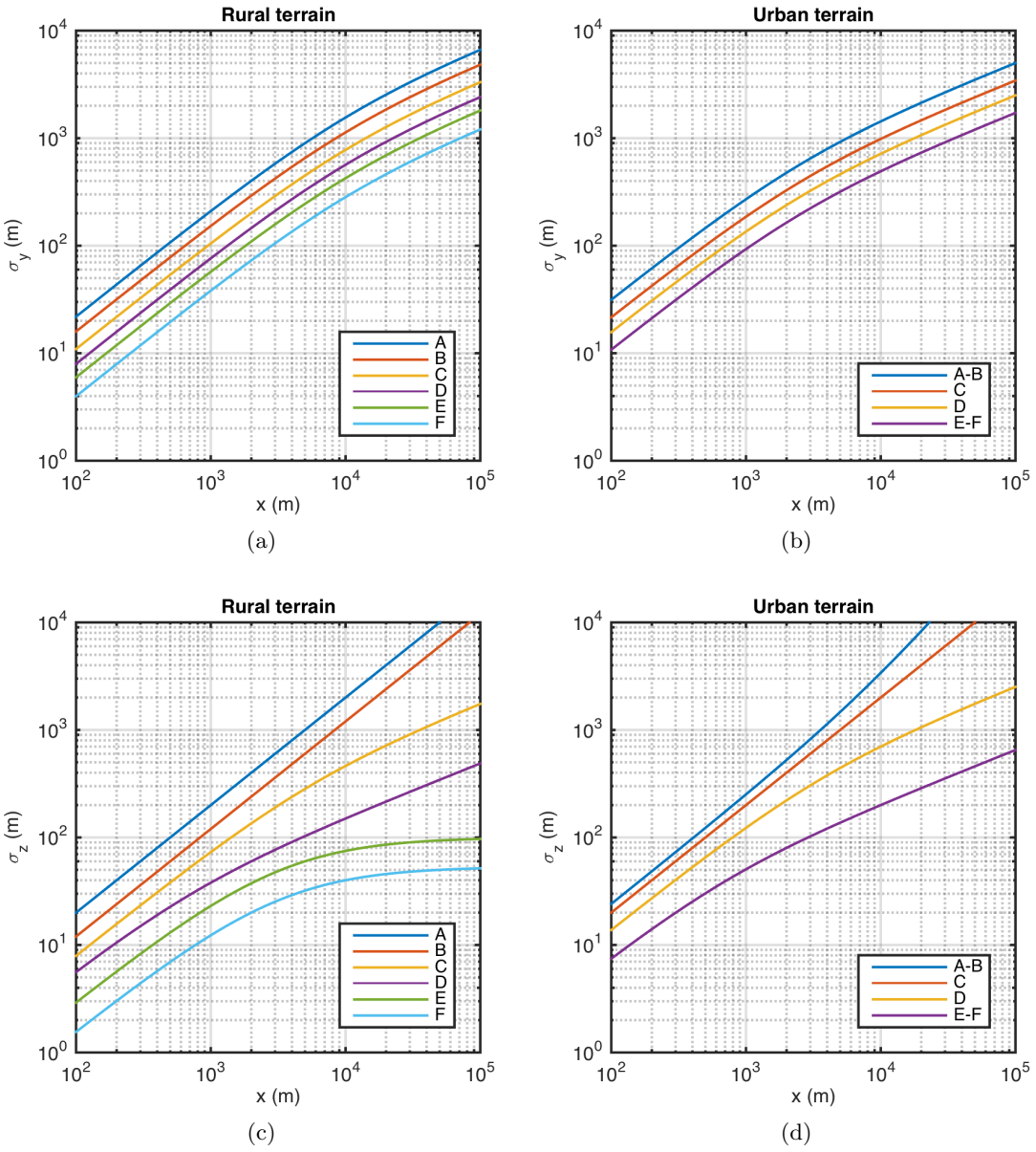

Figure 2.2: Horizontal and vertical dispersion parameters calculated with the Briggs equation [2] based on the graphs by Turner [32].

	Rural					
	σ_y			σ_y		
Stability class	a	b	c	d	e	f
A	0.22	0.0001	0.5	0.2	0	—
B	0.16	0.0001	0.5	0.12	0	—
C	0.11	0.0001	0.5	0.08	0.0002	0.5
D	0.06	0.0001	0.5	0.03	0.0003	1
F	0.04	0.0001	0.5	0.016	0.0003	1
	Urban					
	σ_y			σ_y		
Stability class	a	b	c	d	e	f
A–B	0.32	0.0004	0.5	0.24	0.0001	-0.5
C	0.22	0.0004	0.5	0.2	0	—
D	0.16	0.0004	0.5	0.14	0.0003	0.5
E–F	0.11	0.0004	0.5	0.08	0.0015	0.5

Table 2.2: Coefficients for the Briggs equations for computing σ_y and σ_z. For notes on application see [8].

Continuous parameterization of dispersion parameters

The stability-class method for computing the dispersion parameters is tractable when only the general atmospheric characteristics are known. However, in practice it is common to collect additional data about the wind field. This allows for the computation of much more accurate dispersion parameter values through incorporating statistical knowledge of the turbulence.

Consider particles being ejected from a source at location $(x, y, z) = (0, 0, 0)$ at time $t = 0$. Particles are dispersed by a random wind vector with downwind component u and crosswind component v, where v is distributed as a zero-mean Gaussian with variance σ_v^2. Neglecting wind speed fluctuation in the x direction, we have that all particles will be at $x = \bar{u}t$ after a very short period of time t where $\bar{u}$ is the average wind speed. Over the same period of time t, each of the particles is subject to a *different* wind speed in the cross-wind direction leading to new cross-wind direction is $y = vt$, assuming without loss of generality that the mean wind speed is along the x axis. We can compute the standard deviation of the y position of the particles to be $\sigma_y = \sigma_v t$. Due to the stochastic nature of the wind speed and direction, this is the maximum value of σ_y. Over time, as the wind speed is less likely to be the same, this σ_y is some fraction of $\sigma_v t$. This depends on how long the wind speed stays the same. The same analysis applies in the vertical direction.

Many works attempt to describe the influence of cross-wind and vertical eddies,

or turbulence, on the extent of the plume. Pasquill [25] proposed the following parameterization that includes computing the two non-dimensional functions $f_y(\cdot)$ and $f_z(\cdot)$,

$$\sigma_y = \sigma_v t f_y(t/T_L) \tag{2.11}$$

$$\sigma_z = \sigma_w t f_z(t/T_L) \tag{2.12}$$

where t is the diffusion time and σ_v and σ_w are the standard deviations of the horizontal and vertical wind components, which can be either measured or estimated using a formula. The functions $f_y(\cdot)$ and $f_z(\cdot)$ slowing the rate of change of σ_y and σ_z as the downwind distance grows very large. The scale at which this occurs is governed by the Lagrangian time scale T_L, a convenient measure of the average duration of autocorrelation; T_L is different in all three directions x, y, and z.

In practice we will usually want to estimate the dispersion parameters at some downwind distance x instead of some diffusion time t. Draxler [9] reparameterizes (2.11) and (2.12) as

$$\sigma_y = \sigma_\theta x f_y(t/T_\mathrm{i}) \tag{2.13}$$

$$\sigma_z = \sigma_\phi x f_z(t/T_\mathrm{i}) \tag{2.14}$$

We note that the diffusion time $t = x/u$, where x is the downwind distance and u is the wind speed. Similarly $T_i = 1.64 T_L$ is a normalization factor proportional to the Lagrangian time scale in each direction that indicates the time required for $f_y(\cdot)$ or $f_z(\cdot)$ to become equal to 0.5. Finally, we note that the standard deviations of wind vector azimuth and elevation angles as $\sigma_\theta = \arctan(\sigma_v/u)$ and $\sigma_\phi = \arctan(\sigma_w/u)$. For small angles we have $\sigma_\theta \simeq \sigma_v/u$ and $\sigma_\phi \simeq \sigma_w/u$. Making the appropriate substitutions in Equations (2.11) and (2.12) yield Equations (2.13) and (2.14).

We now turn our attention to the non-dimensional functions f_y and f_z, as well as the Lagrangian integral time scales T_L. For short distances the wind speed fluctuations do not change much, so the dispersion is the maximum possible value. In other words, for $t \ll T_L$ we have $f_y = f_z = 1$. As the diffusion time increases, the wind speed changes along the path from the emitter to the measurement location and can be assumed to follow a random walk in the crosswind and vertical directions. As such, for $t \gg T_L$ we would expect $f_y < 1$ and $f_z < 1$. This results in a tapering-off of the standard deviation as a function of the downwind distance. Draxler [9] notes that for downwind distance of about 10 km, the dependence on the time scale parameter T_L can be dropped, resulting in an expression for σ_y and σ_z that are proportional to the downwind distance x. Others have suggested additional modifications in the years since. Seinfeld and Pandis [29] summarized the approaches, offering the following forms for unstable atmosphere:

$$f_y(t/T_L) = \frac{1}{1 + (t/T_L)^{0.5}}, \qquad f_z(t/T_L) = \frac{1}{1 + 0.9(t/T_L)^{0.5}}. \tag{2.15}$$

Under stable conditions in the y direction the Lagrangian time scale is often taken to be $T_L = 1000$ s. For unstable conditions in the z direction the value $T_L = 500$ s is suggested; stable conditions in the z direction suggest using $T_L = 100$ s.

It is noted that once the functions f_y and f_z have been determined, σ_y and σ_z are directly given by observations of σ_θ and σ_ϕ. These standard deviations of the wind direction and wind elevation angle components are generally computed over 10 minutes of observations. A comparison of methods for obtaining these values are given in [7].

Extensions of the Gaussian plume model

One of the primary deficiencies of the standard Gaussian plume model is an inability to deal with changing wind directions. Several extensions to the Gaussian plume model have been proposed to take a changing wind field into account.

The segmented Gaussian plume model accounts for a temporally-varying wind model [4]. The wind field is assumed spatially constant so that at a given time all of the plume is being acted on by the same field. The Gaussian plume is broken up into segments, parameterized by time. Each segment is itself Gaussian and concentrations are computed by the equations previously given.

The Gaussian puff model abandons the plume representation altogether, allowing for nonstationary emissions in temporally and spatially varying wind conditions. The emitter's continuous emission at the rate Q is assumed to be broken up into segments of length Δt. After Δt seconds the emitter injects a plume of mass $\Delta M = Q\Delta t$ into the atmosphere. Each puff may be treated independently by any spatially or temporally-varying wind conditions. Further information may be found in Zanetti [37]. This model has been used for source estimation by Neumann et al. [23].

■ 2.1.2 Lagrangian stochastic model

Gaussian plume models are deterministic in nature and represent an ensemble average of the particles emitted from a plume. They work well in situations where the wind speed is constant, but break down quickly as the particle dispersion becomes more complicated. Lagrangian stochastic models manage to avoid such problems by simulating the trajectories of emitted particles in a turbulent flow given statistical knowledge of the velocity field. These particles represent a unit of "fluid" that is propagated through the air by some wind field and is subject to random small-scale effects of turbulent flows. We can then compute statistics over a large collection of simulated trajectories to approximate the ensemble-average concentration measured at some point in space-time.

Since they simulate actual particle motion through the wind field, Lagrangian stochastic models tend to be more accurate than Gaussian plume models and even Gaussian puff models. Plume model accuracy tends to degrade at about 30-50 km downwind of the source and puff models tend to break down around 200 km. Lagrangian particle models are accurate to much further distances provided that a sufficient number of particle trajectories are simulated [8]. One additional advantage of simulating actual trajectories is that, unlike Gaussian plume models, Lagrangian stochastic models do not assume that the system is at steady-state conditions. Hence, these models can incorporate the time dependency in the expected concentration and sources that turn

on and off. Unfortunately they also carry a much greater computational burden that makes them less commonly used in iterative source estimation methods.

However, there still remain some situations in which Lagrangian stochastic models fail to accurately estimate the concentration field. This is common in urban environments, where flow may be blocked by buildings and other obstacles. In such environments, one may resort to using even more complicated model that simulate motion from first principles and predicts expected concentrations with a high degree of accuracy.

■ 2.1.3 Computational fluid dynamics model

The Gaussian plume model and the Lagrangian stochastic model are both based on a series of assumptions about the physical processes governing transport and diffusion. For instance, they both assume that the wind speed and statistics of turbulence profiles are known or can be classified as a stability class. It is also assumed that the terrain is relatively smooth and that there are no large obstacles that would impact the flow of fluids from the source to the sensor. When either of these assumptions are severely violated, one may resort to simulating directly from the physical model of atmospheric turbulence.

The Navier-Stokes equation and the mass conservation equation govern the physics of turbulent flow; the energy balance equation is occasionally required as well. Numerical simulation from these equations is called *computational fluid dynamics* (CFDs). CFDs model effects on a huge range of scales, ranging from the Kolmogorov length scale (millimeters) to the size of the planetary boundary layer (kilometers). This wide range of scales precludes a naive implementation of the equations. Instead, to make computation tractable, approximations are generally used for the smallest time scales. Two common approaches are called Reynolds-averaged Navier-Stokes (RANS) and large eddy simulation (LES). De Visscher [8] provides an overview of these two methods and an introduction to the aforementioned equations governing particle motion.

CFD models are so computationally demanding that they are usually only applied when more efficient methods would fail to correctly model the dispersion. For example, Chow et al. [6] uses CFDs to estimate source locations within an urban environment containing tall buildings. These tall buildings funnel the wind between them, causing horizontal gusts that could neither be captured in a Gaussian plume model nor a Lagrangian stochastic dispersion model. Most applications in the source estimation literature consider dispersion across a rural environment for which CFD models are unnecessary, so these computationally intensive methods are rarely employed.

■ 2.2 Markov chain Monte Carlo methods

In this section we shift topics from atmospheric dispersion to Bayesian inference. Specifically, here we consider a particular method for approximate Bayesian inference called Markov chain Monte Carlo methods. Suppose we are interested in the posterior distribution $p(x \mid y)$ of some scalar parameter x given some measurement y. We call this the

target distribution and write it as $p(x)$ for brevity. MCMC methods generate a Markov chain of samples $x^{(1)}$, $x^{(2)}$, ..., $x^{(k)}$ that converges (under mild regularity conditions) to the target distribution $p(x)$. Given a collection of samples from the posterior, we can compute any desired statistic of $p(x \mid y)$, including marginal distributions and *maximum a posteriori* (MAP) estimates.

We can still sample from the target distribution even if we can only evaluate the target up to some multiplicative constant. This is common in Bayesian inference where we want to evaluate the posterior

$$p(x \mid y) = \frac{p(y \mid x)p(x)}{\int p(y \mid x)p(x)\,\mathrm{d}x} = \frac{\tilde{p}(x)}{Z} \tag{2.16}$$

and $p(y \mid x)$ and $p(x)$ are the data likelihood and parameter prior distributions. From the generative model we can easily evaluate the numerator $\tilde{p}(x) = p(y \mid x)p(x)$ but computing the denominator $Z = \int p(y \mid x)p(x)\,\mathrm{d}x$, known as the *partition function* or *evidence*, is generally intractable.

Markov chain Monte Carlo methods are an iterative approach to approximating any probability distribution by a set of samples. Excellent treatments of MCMC methods are provided in [10], [12], and [27].

Most MCMC methods rely on the Metropolis-Hastings (MH) algorithm to construct the chain of samples from the target distribution [14]. The MH algorithm is a method for constructing the k-th sample from the distribution given the $(k-1)$-th. It relies on a *proposal distribution* $q(x^* \mid x^{(k-1)})$ from which we can easily sample. Generating the k-th sample involves the following steps:

1. Sample a new state x^* from the proposal distribution, i.e.

$$x^* \sim q\Big(x \mid x^{(k-1)}\Big). \tag{2.17}$$

2. Compute the *Hastings ratio*

$$r = \frac{p(x^*)q\big(x^{(k-1)} \mid x^*\big)}{p\big(x^{(k-1)}\big)q\big(x^* \mid x^{(k-1)}\big)}. \tag{2.18}$$

3. Set

$$x^{(k)} = \begin{cases} x^* & \text{with } \textit{acceptance probability } \alpha = \min(1, r) \\ x^{(k-1)} & \text{otherwise.} \end{cases} \tag{2.19}$$

The MH algorithm, Algorithm 2.1, begins at at some arbitrary starting point $x^{(0)}$ for which $p(x^{(0)}) > 0$. If $p(x^{(0)}) = 0$ the initial Hastings ratio is undefined. The above steps are repeated as many times as desired. We note that the form of the Hastings ratio means we can use $p(x)$ or the unnormalized distribution $\tilde{p}(x)$ when computing α. The same holds for the proposal distribution $q(\cdot \mid \cdot)$.

procedure MetropolisHastings(x^0, K)
  for $k = 1, 2, \ldots, K$ **do**
    $x \leftarrow x^{k-1}$
    $x' \sim q(x' \mid x)$ ▷ sample from proposal distribution
    $\alpha \leftarrow \dfrac{\tilde{p}(x')q(x \mid x')}{\tilde{p}(x)q(x' \mid x)}$ ▷ compute acceptance probability
    $r \leftarrow \min(1, \alpha)$
    $u \sim \mathcal{U}(u; 0, 1)$
    if $u < r$ **then**
      $x^k \leftarrow x'$ ▷ accept the proposed sample
    else
      $x^k \leftarrow x^{k-1}$ ▷ reject the proposed sample
    end if
  end for
  return $\{x^k\}_{k=1}^K$
end procedure

Algorithm 2.1: Metropolis-Hastings algorithm

It is known that samples obtained by the MH algorithm are guaranteed to converge to the stationary distribution *regardless* of the choice of proposal distribution under suitable regularity conditions [28]. As such, proposals are generally chosen to be convenient or easy to evaluate. One common class of proposal distributions are called *symmetric*, i.e.

$$q(x' \mid x) = q(x \mid x'). \tag{2.20}$$

For symmetric proposal distributions the Hastings ratio becomes the ratio of the target distributions

$$r = \frac{p(x^*)q(x^{(k-1)} \mid x^*)}{p(x^{(k-1)})q(x^* \mid x^{(k-1)})} = \frac{p(x^*)}{p(x^{(k-1)})}.$$

A subclass of symmetric proposal distributions are *symmetric random walk* proposals, where

$$q(x' \mid x) = \mathcal{N}(x'; x, \tau^2) \tag{2.21}$$

for some variance τ^2. The rate at which our chain of samples converges to the stationary distribution is usually tied to the *acceptance rate* $\mathbb{E}[\alpha]$, which is generally computed over some number of the most recent samples. An acceptance rate $\mathbb{E}[\alpha]$ that is large corresponds to not adequately exploring the target distribution and making very small steps around the current state of the Markov chain due to a small proposal variance τ^2. Conversely, if $\mathbb{E}[\alpha]$ is very low then most of the proposals are being rejected and the variance τ^2 is too large. The same intuition also applies to proposals that are not symmetric random walks.

The MH algorithm extends easily to a random vector of parameters $\boldsymbol{x} = (x_1, \ldots, x_N)$. To sample from the target distribution $p(\boldsymbol{x})$ we sample each component conditioned on the other components being constant. To draw the (k)-th sample of $\boldsymbol{x}$ we perform component-wise MH updates with the proposal distributions $q_i(x_i^* \,|\, x_i, \boldsymbol{x}_{-i}^{(k-1)})$ where

$$\boldsymbol{x}_{-i}^{(k-1)} = \left(x_1^{(k)}, \ldots, x_{i-1}^{(k)}, x_{i+1}^{(k-1)}, \ldots, x_N^{(k-1)}\right).$$

The Hastings ratio for the ith component is

$$r = \frac{p\left(x_i^*, \boldsymbol{x}_{-i}^{(k-1)}\right) q_i\left(x_i^{(k-1)} \,|\, x_i^*, \boldsymbol{x}_{-i}^{(k-1)}\right)}{p\left(x_i^{(k-1)}, \boldsymbol{x}_{-i}^{(k-1)}\right) q_i\left(x_i^* \,|\, x_i^{(k-1)}, \boldsymbol{x}_{-i}^{(k-1)}\right)}. \tag{2.22}$$

As long as we update each component x_i at every iteration we are still guaranteed to converge to the target distribution $p(\boldsymbol{x})$.

Alternative methods exist for sampling the target distribution for multivariate parameter vectors. In one common method called Gibbs sampling, we sample each component x_i from its full conditional distribution $p(x_i \,|\, \boldsymbol{x}_{-i})$. That is, we draw the (k)-th sample from the $(k-1)$-th using

$$x_i^{(k)} \sim p\left(x_i \,|\, \boldsymbol{x}_{-i}^{(k-1)}\right) \tag{2.23}$$

and iterating from $i = 1, \ldots, N$. This is commonly used when the full conditional distributions can be written in closed form and sampled from.

We can show that Gibbs sampling is just a specific case of the MH algorithm for which the acceptance probability $\alpha = 1$ since the proposals are always accepted. Sampling from the full conditional for some component x_i corresponds to using the proposal distribution

$$q_i\left(x_i' \,|\, x_i, \boldsymbol{x}_{-i}\right) = p\left(x_i' \,|\, \boldsymbol{x}_{-i}\right). \tag{2.24}$$

Plugging into Equation 2.22 we get

$$\begin{aligned} r &= \frac{p\left(x_i^*, \boldsymbol{x}_{-i}^{(k-1)}\right) q_i\left(x_i^{(k-1)} \,|\, x_i^*, \boldsymbol{x}_{-i}^{(k-1)}\right)}{p\left(x_i^{(k-1)}, \boldsymbol{x}_{-i}^{(k-1)}\right) q_i\left(x_i^* \,|\, x_i^{(k-1)}, \boldsymbol{x}_{-i}^{(k-1)}\right)} \\ &= \frac{\left(p\left(x_i^* \,|\, \boldsymbol{x}_{-i}^{(k-1)}\right) p\left(\boldsymbol{x}_{-i}^{(k-1)}\right)\right) p\left(x_i^{(k-1)} \,|\, \boldsymbol{x}_{-i}^{(k-1)}\right)}{\left(p\left(x_i^{(k-1)} \,|\, \boldsymbol{x}_{-i}^{(k-1)}\right) p\left(\boldsymbol{x}_{-i}^{(k-1)}\right)\right) p\left(x_i^* \,|\, \boldsymbol{x}_{-i}^{(k-1)}\right)} \\ &= 1, \end{aligned}$$

which proves Gibbs sampling is a special case of the MH algorithm. As with component-wise MH, Gibbs sampling is guaranteed to converge when all components are sampled at each iteration. Gibbs sampling can also be used within component-wise MH by using

the full conditional as the proposal distribution for some particular component or set of components. Convergence to the target distribution can be accelerated by sampling variables simultaneously, as in block Gibbs sampling.

Reversible-jump MCMC (RJ-MCMC) is a technique used when the size of the parameter set changes with the model order and the model order is one of the parameters being inferred [13]. Its formulation is usually tightly coupled with the model of interest, so we will postpone our introduction to this method until we apply it in Chapter 4.

2.3 Source estimation methods

In this section we will review existing approaches for source estimation. Source estimation refers to solving an *inverse problem* with the following general structure. There exist some number of static emission sources releasing a pollutant. We make concentration measurements using one or more possibly mobile sensors; each measurement has an associated location and time. Based on this set of measurements we would like to infer some properties of the emission sources, potentially including the location, rate of emission, activation or deactivation time, and number of sources. Additionally, we may know some of these properties *a priori*, or in advance. Another name for these inverse problems is atmospheric dispersion event reconstruction problems.

There are many approaches to solving these inverse problems. Rao [26] divides them into two main categories: *forward* modeling methods and *backward* modeling methods. Forward modeling methods are generally iterative and use a forward-simulating dispersion model, i.e. concentrations are transported from source to receptor. Backward modeling methods simulate transport backwards from the measurement location to possible emitter locations. They generally require only one run. We now provide further description and recent work for each type of method.

2.3.1 Methods based on forward simulation models

Forward modeling methods are iterative models which generate candidate source configurations upon each iteration. In each iteration, a forward atmospheric dispersion model uses the candidate source configurations to obtain a set of predicted concentration measurements. The differences between the predicted and actual concentration measurements are used to propose new source configurations, and the approach terminates when some criteria is met. By far the most common methods for forward source estimation are MCMC methods.

MCMC methods for a known number of emitters

Atmospheric dispersion event reconstruction problems attempt to quantify source properties which include a time component. For a chemical, biological, or radiological (CBR) release time would be critical, since responders would want to know not just the location and rate of emissions but also when the emissions began. In most works dealing with this problem a single source is assumed and a sensor network monitors the survey

area taking measurements at some time intervals. An MCMC formulation of this was given in Johannesson et al. [19]. A variation on the Gaussian plume model was used and the terrain was assumed to be smooth over the large survey area. These methods were extended to smaller survey areas containing urban environments, which require more complicated dispersion models. Neumann et al. [23] employed a Gaussian puff model to take buildings into account. Similar work by Chow et al. [6] used a full CFD model to estimate the concentrations observed by the sensor network. In both cases, reasonable solutions to the event reconstruction problem were obtained from the MCMC procedure.

When the emissions of a source are time-invariant, the forward dispersion model can be restricted to the steady-state case. In Senocak [30] MCMC methods based on [19] are used to infer the location and emission rate of a single source in a large rural survey area. Since emissions are time-invariant, the standard Gaussian plume model is used along with dispersion parameters given by Briggs curves. This work uses a more complicated sensor model in which the sensors only issue a non-zero noisy concentration measurement when the observed concentration is above some detectable threshold. MCMC efficiently provided reasonable results in the experiments shown, including one that accommodated shifting wind fields using a segmented Gaussian plume mode.

It is often the case that there are multiple emission sources. Both Yee [34] and Keats et al. [20] consider this case for CBR emission sources. Both assume that the number of sources is known *a priori* and that each source has its own activation and deactivation times. These properties are inferred along with the source locations and emission rates are using efficient MCMC inference. In both cases the dispersion model is solution to the adjoint ADE equation. Keats et al. additionally accounts for "highly disturbed" flow fields, such as those seen in an urban environment.

In certain applications with time-varying emissions it is not sufficient to know just the times when the emissions begin and end. Instead, it may be crucial to know how the source is emitting over time. Hazart et al. [16] applies MCMC methods to estimate the location, emission rate, and emission profile of a single source. The paper considers seven different emission profiles, including a Dirac delta, a step function, a window function, and a Gaussian signal shape. A variation on the Gaussian plume model is derived and several different prior parametrizations are considered. Hazart et al. also examines conditions under which some temporal release profiles closely approximate others.

MCMC methods for an unknown number of emitters

When we don't know the number of emitters, the standard MCMC approach breaks down. We cannot simply append the number of sources to our parameter vector $\boldsymbol{\theta}$, since now the dimensionality of $\boldsymbol{\theta}$, or the number of parameters we are estimating, changes with the number of sources. Two methods currently exist for handling this: run MCMC with a very large number of sources and then select a small subset at the very end, or use a variation of MCMC that adaptively estimates the number of sources along with

all of the parameters.

Following the first approach, Huang et al. [18] infers the location, emission rate, and activation time for a number of sources using MCMC sampling. During the sampling procedure the number of sources is assumed known. Following sampling the deviance information criterion (DIC) is applied to select the number of sources. An alternative approach by Wade and Senocak [33] extends the model developed in [30] by assuming multiple sources. Here, Wade and Senocak use a composite ranking approach to estimate the number of sources by combining multiple score functions.

An alternative approach embeds estimation of the number of sources into the MCMC sampling procedure. Reversible-jump MCMC (RJ-MCMC) adds the number of sources to the parameter vector and augments the sampling procedure with a trans-dimensional proposal step to allow sampling models of different orders [13]. It was originally used for the event reconstruction problem by Yee [35] which extends [34] to estimate the number of sources. Here the trans-dimensional proposal step is uniformly selected from a source creation proposal or a source annihilation proposal. Yee et al. compares RJ-MCMC to importance sampling with progressive correction is given in [36].

Hirst et al. [17] also infers an unknown number of sources in the case of time-invariant emissions. Unlike the previously-described applications with static sensors, Hirst et al. collect data from an aircraft; consequently they only observe the concentration at one instant of time per location, instead of how the concentration at that location changes with time. To compensate they impose additional structure on the ambient background concentrations that are also part of the measurement. RJ-MCMC is implemented using the two trans-dimensional proposals from [35] along with source splitting and merging proposals. Reasonable results are obtained for both real and synthetic data.

Other forward-modeling methods

Genetic algorithms (GA) have also been used to solve the inverse problem in various forms. GAs maintain populations of candidate source configurations that change over time. Traditionally the populations change by the standard genetic operators of mutation and crossover from Darwinian evolutionary theory and the suitability or fitness of a candidate configuration is given by some score function. Allen et al. [1] first proposed using GAs to solve the optimization problem associated with source estimation. Haupt et al. [15] applied a continuous parameter GA to obtain an estimate of the source location, emission rate, and wind direction using a Gaussian plume model for forward modeling. This was extended by Long et al. [22] to estimate the source location, height, strength, and release time based on the observed concentration measurements. Another related method, called evolutionary strategies (ES), has recently gained popularity. Cervone et al. [5] develops an algorithm called Adaptive ES (AES) in the context of source estimation. In AES the genetic operators of mutation and crossover have been dispatched in favor of an optimization procedure with random perturbations. This approach is used to identify the location, size, and emission rate of a single source

over multiple iterations, where dispersion is accomplished through a single-reflection Gaussian plume model.

Another forward-modeling method is a variant of the Expectation-Maximization (EM) algorithm, applied to the inverse problem by Khemka et al. [21]. Here it is assumed that a known number of sources is distributed over an survey area as a 2D homogeneous Poisson process. Noisy measurements are taken at ground level by sensors that only activate at a certain threshold. The applied Gaussian plume equation is obtained as a solution to the ADE for ground-level observations. They formulate the probability distribution of source locations over a grid where each cell may be a source with some emission rate and obtain the MAP estimate using the Generalized EM algorithm.

■ 2.3.2 Methods based on backward simulation models

Backwards algorithms simulate transport and dispersion in the reverse direction. Backwards modeling methods often only use one run in the reverse direction from the receptors to find the unknown upwind source(s), hence they are computationally more efficient and are often preferred when there are more sources than receptors or when the number of receptors is limited and the number of sources is unknown. These methods include Kalman filtering, adjoint and tangent linear models, and variational data assimilation. Rao [26] provides further details and references for each of these methods.

Chapter 3

Model Description

In this chapter we present a mathematical model for the source estimation problem. It is designed to make inference tractable for remotely-obtained measurements, but is also flexible enough to fit any data acquisition process where measurements have a known time and location. First we develop the observation model, and then we examine the background model. We then examine the generative graphical model corresponding to this problem. The model largely follows that of Hirst et al. [17].

Consider a collection of m ground-level emission sources (where m is unknown) that emit some material at a constant rate. These emissions are then subject to atmospheric transport and diffusion due to wind and turbulence. Our goal is to use some noisy concentration measurements in order to infer the number, location, and emission rate of the sources. For our purposes, we will assume that concentration measurements are collected by an aircraft which introduces constraints on how the measurements can be collected. Furthermore, it is generally the case that *ambient* background sources introduce significant interference into the measurement process. Finally, uncertainty in the wind direction and variability further complicates both modeling and inference.

The problem and approach specified here differ from much of the literature on source estimation and event reconstruction. We recognize that we are modeling a source estimation problem and not the more general event reconstruction problem. This is due to an assumption that the sources are emitting at a constant rate, hence we are not concerned with the activation and deactivation times inferred for event reconstruction problems. Another significant difference between this approach and much of the literature pertains to the types of measurements we consider. In most approaches, static sensors are placed in set locations and take many measurements in time at their corresponding locations. This is possible when we have a small *survey area* in which we "look" for sources; in our cases the survey area is usually very large, e.g. 40 km by 40 km. To survey this large area we collect measurements remotely via aircraft. In contrast with the case of static sensors, our mobile sensor platform takes only a single measurement at many locations through time and space. Since we only obtain one measurement per location, we need to impose some additional structure on the latent, or background, concentration levels to help our inference procedure differentiate background concentrations from concentrations due to the sources.

We describe the emissions from our m sources as having a spatial distribution given

by the two-dimensional Gaussian with mean locations $\boldsymbol{x} = \{\boldsymbol{x}_j\}_{j=1}^m$ and isotropic covariance matrices $w_j^2 \cdot \mathbf{I}^{2\times 2}$, where $\mathbf{I}^{d\times d}$ is the $d \times d$ identity matrix. Lt $\boldsymbol{w} = \{w_j\}_{j=1}^m$ denote the widths of the sources and $\boldsymbol{s} = \{s_j\}_{j=1}^m$ denote the source emission rates (i.e. strengths), which are modeled as constant over time and are in units of m^3/s.

By conceptualizing the emissions as particles, we can make an immediate physical interpretation of this model. Consider the release of discrete particles of volume δv by the j-th emission source. Every second the source releases $N_p = s_j/\delta v$ particles and the exact release location for each particle is drawn from a 2D Gaussian distribution $\mathcal{N}(\boldsymbol{x}_j, w_j^2\, \mathbf{I}^{2\times 2})$. This intuition generalizes to the continuous-emission case by letting $\delta v \to 0$, hence $N_p \to \infty$.

Observations are made remotely from an aircraft flying a known trajectory and carrying a sensor with a known sampling rate f_s. This produces n scalar concentration measurements $\boldsymbol{z} = \{z_i\}_{i=1}^n$ at measurement locations $\boldsymbol{y} = \{\boldsymbol{y}_i\}_{i=1}^n$ and times $\boldsymbol{t} = \{t_i\}_{i=1}^n$, also known. These n measurements contain contributions from the sources defined above, plus the background contribution and measurement noise. The vector of background contributions is $\boldsymbol{b} = \{b_i\}_{i=1}^n$, where b_i is the ambient concentration at measurement location $\boldsymbol{y}_i$ and time t_i. We model measurement noise as the uncorrelated white Gaussian noise vector $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$ where $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. In general, the source contributions are within an order of magnitude of 10^{-8}; the measurement noise is on the order of 10^{-9}, however, the background contributions are on the order of 10^{-6}. As such, an accurate background model is essential for reliable estimation of source parameters. We note that this particular formulation can be easily applied to non-remote settings where the sensors are stationary and take many measurements through time by appropriately defining the measurement time and location vectors, $\boldsymbol{t}$ and $\boldsymbol{y}$.

Our analysis uses wind and turbulence data defined at each measurement time and location; we denote the set as $\mathcal{W} = \{\{\boldsymbol{u}_i\}_{i=1}^n, \gamma_\mathrm{H}, \gamma_\mathrm{V}\}$. These parameters are used by the dispersion model and as a part of the background model, which we will describe later this chapter. We have a two-dimensional wind vector $\boldsymbol{u}_i$ defined at each measurement location, in addition to the wind direction standard deviation γ_H, also known as the horizontal plume opening angle, and the vertical wind direction standard deviation γ_V, or the vertical plume opening angle. We stress that although our model is in 3D, the wind model is currently limited to 2D. When working with real data the wind field measurements $\mathcal{W}$ are provided by the U.K. Meteorological Organization (UKMO) and are treated as parameters; otherwise we generate a synthetic wind field.

■ 3.1 Observation model

At each measurement location we make noisy observations according to

$$z_i = \sum_{j=1}^{m} a_{ij} s_j + b_i + \varepsilon_i \tag{3.1}$$

where b_i is the background contribution and measurement error is $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, as mentioned above. The value a_{ij} is called the "coupling strength" between measurement location i with each source j. This strength is given by $a_{ij} = a(\boldsymbol{y}_i, \boldsymbol{x}_j, w_j; \mathcal{W})$, where $a(\cdot)$ is the "coupling function" that gives the expected concentration measured at location $\boldsymbol{y}_i$ from a source with unit emission rate, location $\boldsymbol{x}_j$, and width w_j subject to wind field $\mathcal{W}$. For conciseness, $\mathcal{W}$ is often omitted. We will often write Equation 3.1 in the equivalent matrix form

$$\boldsymbol{z} = \mathbf{A}\boldsymbol{s} + \boldsymbol{b} + \boldsymbol{\varepsilon} \tag{3.2}$$

where $\mathbf{A} = \{a_{ij}\}_{i=1,j=1}^{n,m}$ is referred to as the "coupling matrix".

■ 3.1.1 Application of the Gaussian plume dispersion model

For simplicity we use the Gaussian plume model from Section 2.1.1 to model atmospheric dispersion. This provides a closed-form expression for the expected concentration measured at location $\boldsymbol{y}_i$ due to the source with a unit emission rate located at $\boldsymbol{x}_j$ with width w_j. Transport is assumed to follow the wind vector at measurement location i, $\boldsymbol{u}_i = \begin{pmatrix} u_1^i & u_2^i \end{pmatrix}$, which has wind speed $|\boldsymbol{u}_i|$ and direction $\theta_{\boldsymbol{u}_i}$. Additional wind parameters are contained in $\mathcal{W}$.

In our original development of the Gaussian plume equation our coordinate axes were determined by the wind: we had the downwind distance x, the cross-wind distance y, and the vertical distance z. In order to compute these wind-aligned quantities for the distance between source $\boldsymbol{x}_j$ and measurement $\boldsymbol{y}_i$ we must rotate the difference into coordinate axes defined by the wind $\boldsymbol{u}_i$:

$$\begin{pmatrix} \delta_\mathrm{R} \\ \delta_\mathrm{H} \\ \delta_\mathrm{V} \end{pmatrix} = \begin{pmatrix} \cos(\theta_{\boldsymbol{u}_i}) & \sin(\theta_{\boldsymbol{u}_i}) & 0 \\ -\sin(\theta_{\boldsymbol{u}_i}) & \cos(\theta_{\boldsymbol{u}_i}) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} | & | \\ \boldsymbol{y}_i - \boldsymbol{x}_j & \\ | & | \end{pmatrix} \tag{3.3}$$

where to avoid confusion we have renamed the downwind distance δ_R, the cross-wind distance δ_H, and vertical distance to δ_V. As an approximation we model the wind direction as constant between each measurement location and all source locations.

Recall that the original plume equation involved a point source, i.e. a source for which $w_j = 0$. The expected concentration profile in the cross-wind direction is Gaussian-distributed about the emission location with some variance σ_H^2 that is a function of the downwind distance δ_R. In our case we do not know the emission location exactly; instead the emission location is modeled as a Gaussian random variable with mean $\boldsymbol{x}_j$ and variance $w_j^2\,\mathbf{I}^{2\times 2}$. As such, the expected concentration in the cross-wind direction is the sum of two independent normally-distributed random variables, for which the means and the variances add. Note that this only applies in the cross-wind direction, as the Gaussian plume model assumes negligible diffusion in the downwind direction. With regard to the dispersion parameters, or the standard deviations σ_H and σ_V of the horizontal and vertical concentration distributions, we have

$$\sigma_\mathrm{H}^2 = \delta_\mathrm{R}^2\sigma_\theta^2 + w_j^2 \qquad \text{and} \qquad \sigma_\mathrm{V}^2 = \delta_\mathrm{R}^2\sigma_\phi^2, \tag{3.4}$$

where the wind direction and elevation angle standard deviations $\sigma_\theta = \tan(\gamma_\mathrm{H})$ and $\sigma_\phi = \tan(\gamma_\mathrm{V})$ are written in terms of the horizontal and vertical plume opening angles γ_H and γ_V. This also corresponds to Equations 2.13 and 2.14, where the terms for the nondimensional functions $f_y(\cdot)$ and $f_z(\cdot)$ have been dropped. As mentioned before, these factors play little role for downwind distances less than 10 km. Although we are generally interested in areas of up to 50 km by 50 km in size, we make multiple passes through the survey area; thus, we almost always have at least one observation within 10 km of any given source and this should not degrade performance. Further, [17] showed that including these factors in the forward model made a negligible difference for an inference procedure and application very similar to ours.

After decomposing the distance into its axis-aligned components, we can now compute the expected concentration at some measurement distance $\boldsymbol{\delta} = (\delta_R, \delta_H, \delta_V)$ away from a source with width w_j and unit emission rate:

$$a(\boldsymbol{\delta}, w_j; \mathcal{W}) = \frac{1}{2\pi|\boldsymbol{u}_i|\sigma_\mathrm{H}\sigma_\mathrm{V}} \exp\left\{-\frac{\delta_\mathrm{H}^2}{2\sigma_\mathrm{H}^2}\right\} \times \sum_{j=-N_\mathrm{rfl}}^{N_\mathrm{rfl}} \exp\left\{-\frac{(\delta_\mathrm{V} - (H + 2jD))^2}{2\sigma_\mathrm{V}^2}\right\} + \exp\left\{-\frac{(\delta_\mathrm{V} + (H + 2jD))^2}{2\sigma_\mathrm{V}^2}\right\}, \quad (3.5)$$

As before, H represents the source height above ground level and the ABL begins at height D above ground level. For our application we can assume $H = 0$; different values can easily be incorporated into the model. Further, for our purposes $N_\mathrm{rfl} = 2$ is sufficient. [17] discusses sensitivity of this choice.

We note that the coupling strength given by this model is an ensemble average, or the average contribution over many different realizations of the turbulence and flow of the wind field. Since it is efficiently computable in closed-form it is especially well-suited for sampling-based inference methods that may need to compute the coupling matrix $\mathbf{A}$ tens of thousands of times. Additionally, the form of this coupling equation is reminiscent of the Gaussian distribution: the horizontal and vertical strengths f_H and f_V are akin to the exponential terms in a multivariate Gaussian distribution; the horizontal and vertical plume standard deviations σ_H and σ_V correspond to the standard deviations of a Gaussian that factors completely.

■ 3.2 Background model

The background contributions b_i represent the ambient, or background, concentration that is captured in measurement i. They are constrained to be positive and both spatially and temporally smooth. Gaussian Markov random fields (GMRFs) are commonly used to model spatio-temporally smooth fields. Before collecting measurements, we cannot know the exact values of background measured at each measurement location; however, we can characterize the relationships *between* the background contributions at each point. A certain type of GMRF, a first-order intrinsic GMRF, is commonly used

as a prior distribution when we lack strong prior information about the actual value. A detailed treatment of these models is provided in [28].

■ 3.2.1 Intrinsic Gaussian Markov random fields

Gaussian MRFs model a set of jointly Gaussian random variables. This model is often applied to random variables that are related through some spatio-temporal process. In our case these random variables are the background contributions b_i measured at each point; we use this model to encode spatio-temporal smoothness of the background. Formally, we consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where each the background contribution from each measurement is represented by a node in the graph's vertex set $\mathcal{V}$. The edge set $\mathcal{E}$ exactly defines the structure of a precision matrix $\mathbf{J}$, which specifies the dependencies between background contributions. Although in truth each background values may be dependent on many neighboring background values, here we restrict the number of dependencies to ensure a sparse $\mathbf{J}$, which in turn allows efficient inference. Since the background contributions are Gaussian distributed about some mean $\boldsymbol{b}_0$ we can write the general form

$$\boldsymbol{b} \sim \mathcal{N}\Big(\boldsymbol{b}; \boldsymbol{b}_0, (\mu\mathbf{J})^{-1}\Big) \tag{3.6}$$

where μ is a parameter that tunes how heavily we weight the smoothness of the background relative to the noise in the data. In our model we set the $\mu = 1$.

In the absence of strong prior information about the background mean, we can write this as an intrinsic GMRF of the first order. This implies that the precision matrix $\mathbf{J}$ has rank $n - 1$, or exactly one zero eigenvalue, and that $\mathbf{J1} = \mathbf{0}$. We can equivalently write this condition as

$$\sum_{j=1}^{n} J_{ij} = 0 \text{ for all } i. \tag{3.7}$$

Importantly, note that in an intrinsic GMRF the parameters $\boldsymbol{b}_0$ and $\mathbf{J}$ *do not* represent the mean and precision, which formally no longer exist (although we will continue to refer to them as such). As discussed in [28] the Markov properties of an IGMRF should be interpreted as those obtained from the limit of a proper density.

We now elucidate why these models are commonly used in absence of prior information about the mean of the distribution. The density of an intrinsic GMRF of the first order with mean $\boldsymbol{b}_0$ and precision $\mathbf{J} \in \mathbb{R}^{n\times n}$ of rank $n - 1$ is defined as

$$p(\boldsymbol{b}) = (2\pi)^{-\frac{n-1}{2}} (|\mathbf{J}|^*)^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{b} - \boldsymbol{b}_0)^\top \mathbf{J}(\boldsymbol{b} - \boldsymbol{b}_0)\right) \tag{3.8}$$

where $|\cdot|^*$ denotes the generalized determinant. Recall that the generalized determinant of a matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ of rank $n - k > 0$ is denoted $|\mathbf{A}|^*$ and is the product of the $n - k$ non-zero eigenvalues of $\mathbf{A}$. We note that the mean appears only in the term inside the

exponent. If $\boldsymbol{b}_0 = \mathbf{1}c$ for some constant c, this simplifies to

$$\begin{aligned}(\boldsymbol{b}-\boldsymbol{b}_0)^\top \mathbf{J}(\boldsymbol{b}-\boldsymbol{b}_0) &= \boldsymbol{b}^\top\mathbf{J}\boldsymbol{b} - \boldsymbol{b}_0^\top\mathbf{J}\boldsymbol{b} - \boldsymbol{b}^\top\mathbf{J}\boldsymbol{b}_0 + \boldsymbol{b}_0^\top\mathbf{J}\boldsymbol{b}_0 \\ &= \boldsymbol{b}^\top\mathbf{J}\boldsymbol{b} - \mathbf{0}^\top\boldsymbol{b} - \boldsymbol{b}^\top\mathbf{0} + \boldsymbol{b}_0^\top\mathbf{0} \\ &= \boldsymbol{b}^\top\mathbf{J}\boldsymbol{b}\end{aligned}$$

where $\mathbf{J}\boldsymbol{b}_0 = \mathbf{J}\mathbf{1}c = \mathbf{0}c = \mathbf{0}$ by the condition $\mathbf{J}\mathbf{1} = \mathbf{0}$. The effect of this particular precision structure is that the distribution of $\boldsymbol{b}$ is translation invariant; the precision matrix only constrains the relationships between the different elements of the vector $\boldsymbol{b}$. This leads to the common use of intrinsic GMRFs in applications where the mean of the distribution is unknown *a priori*.

In addition to being translation invariant, this model also allows us to encode exactly the spatio-temporal smoothness constraint we wish to place on the background contributions, distributed by

$$\boldsymbol{b} \sim \mathcal{N}\big(\boldsymbol{b}; \mathbf{0}, \mathbf{J}^{-1}\big) \tag{3.9}$$

where the mean $\boldsymbol{b}_0 = \mathbf{0}$ without loss of generality. We now discuss how this smoothness constraint is encoded through the precision matrix.

■ 3.2.2 Defining the background precision matrix

In reality, the background concentration at every node is dependent with the background concentration at every other node; fully modeling this would result in a matrix that was dense and the corresponding inference algorithms would not be efficient. Instead, we approximate the dense precision matrix by a sparse matrix that only includes a small number of significant edges per node.

The precision matrix $\mathbf{J}$ is completely specified by the edges in $\mathcal{E}$ combined with the wind field $\mathcal{W}$ and the measurement locations $\boldsymbol{y}$ and times $\boldsymbol{t}$. As with all MRFs, the edges in $\mathcal{E}$ correspond to nonzero elements of the precision matrix $\mathbf{J}$. For two random variables b_i and b_j, $i \neq j$,

$$\{i,j\} \in \mathcal{E} \Leftrightarrow \alpha_{ij} = J_{ij} = J_{ji} \neq 0 \tag{3.10}$$

by the global Markov property. The edge strength α_{ij} is based on how correlated the edges are—more correlated elements have higher values of α_{ij}. To impose the sum-to-one constraint that results in the precision matrix having rank $n-1$, the diagonal elements are set as

$$J_{ii} = -\sum_{j:j\sim i} J_{ij} \quad \forall i \in \mathcal{V}, \tag{3.11}$$

where $\{j : j \sim i\}$ is the set of nodes j which share an edge with node i. Hence, each row sums to 0.

The correlation is approximated using heuristic expectations for how the background concentration varies in space and time. We define the constant $c_\mathrm{T} = 3 \times 10^{-15}$, which is the expected change in concentration per unit time at a single measurement location,

and $c_{\mathrm{D}} = 10^{-12}$, the expected change in concentration per unit distance for simultaneous measurements. For two background contributions b_i and b_j, the edge strength between them is defined as

$$\alpha_{ij} = \frac{1}{\left(c_{\mathrm{T}}|\Delta T| + c_{\mathrm{D}}\Delta D\right)^2} \tag{3.12}$$

where ΔT and ΔD are the distance in time and space, respectively. The difference in time is simply $\Delta T = t_j - t_i$, where t_i and t_j are the times when the measurements were collected. In order to properly define the distance ΔD, we must consider the physical processes through which background contributions are smooth across space and time.

Physically, similarity between background measurements arises in two ways. We encapsulate these two ways by defining the full edge set $\mathcal{E} = \mathcal{E}_{\mathrm{s}} \cup \mathcal{E}_{\mathrm{t}}$ as the union of two edge sets, each of which captures one of the mechanisms through which background contributions can be correlated. The spatial distance between the two measurements, ΔD, is computed differently depending on which edge set $\{i, j\}$ comes from.

Sequential measurement links

The first way in which highly-correlated background contributions will arise is if the measurements are taken sequentially. Such measurements are both temporally and spatially very close. We capture this effect through the set of edges between sequential measurements, called $\mathcal{E}_{\mathrm{s}}$.

$$\mathcal{E}_{\mathrm{s}} = \bigcup_{i=1}^{n-1} \{i, i+1\}. \tag{3.13}$$

In this case, the distance used in computing the edge strength is simply given by $\Delta D = \|\boldsymbol{y}_j - \boldsymbol{y}_i\|_2$, the Euclidean distance between the two measurement locations.

Wind-based measurement links

The second way in which background measurements can be similar stems from transport by the wind field $\mathcal{W}$. The background concentration present at one measurement location can be transported downwind by the wind field to reappear at some subsequent measurement location. Due to diffusion, some background component may be a mixture of the background components from multiple upwind measurement locations. The set of edges between measurements due to wind transport is called $\mathcal{E}_{\mathrm{t}}$. To keep $|\mathcal{E}_{\mathrm{t}}|$ small, we only impose one transport-based edge per measurement:

$$\mathcal{E}_{\mathrm{t}} = \bigcup_{i=1}^{n-1} \{i, j^*\} \tag{3.14}$$

where j^* is the measurement that maximizes the edge strength between nodes i and j subject to j being a distance of at least τ away from node i. This is meant to model the strongest correlated measurement directly upwind or downwind of the measurement i.

We have two cases to consider when computing the distance ΔD between some node i and a candidate node j, although the equations used are exactly the same: (1) $t_i < t_j$ and (2) $t_i > t_j$.

The first case corresponds to $t_i < t_j$. In this case, we consider the future location of a packet of air at location $\boldsymbol{y}_i$ at time t_i. The air packet would be acted on by the wind vector $\boldsymbol{u}_i$ for a time of $\Delta T = t_j - t_i > 0$ seconds, giving the packet a location at future time t_j of

$$\tilde{\boldsymbol{y}}_i = \boldsymbol{y}_i + \boldsymbol{u}_i \Delta T. \tag{3.15}$$

We then compute the distance as

$$\Delta D = \|\boldsymbol{y}_j - \tilde{\boldsymbol{y}}_i\|_2. \tag{3.16}$$

The second case corresponds to $t_i > t_j$. In this case, we consider the previous location of a packet of air at location $\boldsymbol{y}_i$ at time t_i, again transported by the wind vector $\boldsymbol{u}_i$ for $\Delta T = t_j - t_i < 0$. This gives the packet a location at previous time t_j of $\tilde{\boldsymbol{y}}_i$ by Equation 3.15 since $\Delta T < 0$. Again the distance is computed via Equation 3.16. It doesn't matter whether the packet is blown upwind or downwind, since emissions are considered to be at steady state and the background will be spatio-temporally smooth. The node j^* is chosen according to

$$j^* = \underset{j \in \mathcal{V} \setminus \mathrm{ne}(i)}{\arg\max}(\alpha_{ij}) \tag{3.17}$$

which maximizes the link strength α_{ij}. The set $\mathcal{V} \setminus \mathrm{ne}(i)$ restricts us to nodes sufficiently far away from node i so that we are capturing the correct effect. Often this will correspond to selecting nodes at least τ meters upwind or downwind from i. Several other methods exist for choosing j^*, including that of [17].

Figure 3.1 shows an example of the structure of $\mathbf{J}$ plotted over a sample flight trajectory. The trajectory is plotted; it is easily recognizable by the dark blue lines indicating strong correlation between successive measurements, i.e. from the edge set $\mathcal{E}_\mathrm{s}$. The lighter, or weaker, edges shown come from the transport-based edge set, $\mathcal{E}_\mathrm{t}$.

Several different interpretations exist for the background concentrations. One such interpretation is that they are well-mixed contributions due to sources far removed from the survey area. We note that the links imposed by the background are not exact. For any given node i, it is assumed that the wind field $\boldsymbol{u}_i$ holds over the entire survey area. This is obviously not true for non-constant wind fields. More complex wind models would allow us to pair nodes more accurately relative to the actual correlation of concentration measurements.

■ 3.3 Probabilistic model

Until now we have refrained from applying probability distributions to many of the quantities involved in the source estimation model. We now describe the generative model, which details the prior and conditional distributions over all of the random

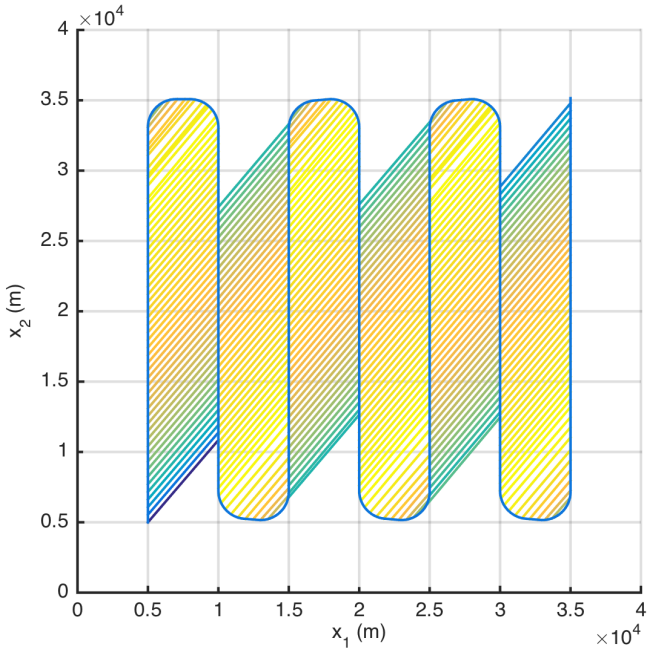

Figure 3.1: Wind links for a sample aircraft trajectory. The wind field is uniformly pointed north-east (top-right). Lighter edges correspond to weaker correlations between measurements.

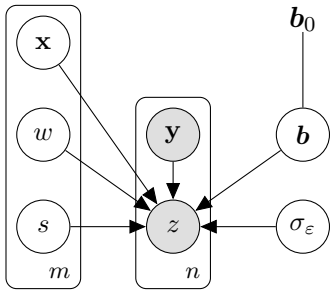

Figure 3.2: A graphical representation of the source estimation model. Shaded nodes correspond to observed random variables.

variables previously mentioned. This generative model corresponds to the graphical model in Figure 3.2.

First, note that we are omitting many parameters that are $\boldsymbol{z}$ depends on, including measurement times $\boldsymbol{t}$ and wind parameters $\mathcal{W}$, the wind field and the horizontal and vertical opening angles at each measurement location. In reality the measurement

locations are accurate up to several cm and the relative times between different elements of $\boldsymbol{t}$ are accurate to a very small fraction of a second, so we can model each of these as parameter values that are known exactly. The plate notation indicates that there are multiple independent sets of the variables positioned on the plate, e.g. there are m different $\boldsymbol{x}$, w, and s values [3].

The number of sources is selected from a uniform prior,

$$p(m) = \mathcal{U}(m;\, [0, m_{\max}]), \tag{3.18}$$

indicating no real prior information over the number of sources. Our model for source distribution could be converted to a Poisson point process by simply modeling the number of sources as a Poisson distribution with some rate parameter λ, which is the average number of sources per unit area.

For the remainder of this section we assume that the number of sources m is fixed. Developments in Chapter 4 will introduce the modeling assumptions used in Reversible-jump MCMC, in which the model order m is treated as a parameter to be estimated. That section will explicitly use the distribution $p(m)$ over the number of sources.

■ 3.3.1 Prior information of the target parameters

Each source j has a location $\boldsymbol{x}_j$, width w_j, and emission rate s_j. The sources are uniformly distributed within the survey area $I_{\boldsymbol{x}}$ according to

$$p(\boldsymbol{x}_j) = \mathcal{U}(\boldsymbol{x}_j;\, \mathrm{C}^\star(I_{\boldsymbol{x}}, \boldsymbol{y}, \mathcal{W})) \quad \text{and} \quad p(\boldsymbol{x} \,|\, m) = \prod_{j=1}^{m} p(\boldsymbol{x}_j) \tag{3.19}$$

where $\mathrm{C}^\star(I_{\boldsymbol{x}}, \boldsymbol{y}, \mathcal{W})$ pertains to the region of $I_{\boldsymbol{x}}$ where sources are actually observable, i.e.

$$\mathrm{C}^\star(I_{\boldsymbol{x}}, \boldsymbol{y}, \mathcal{W}) = \left\{ \boldsymbol{x}_j : \boldsymbol{x}_j \in \mathcal{I}_{\boldsymbol{x}} \wedge \max_i a(\boldsymbol{y}_i, \boldsymbol{x}_j, w_j; \mathcal{W}) \geq \tau \right\}, \tag{3.20}$$

and the minimum observable source contribution is $\tau = 3 \times 10^{-10}$. This allows us to restrict our estimated sources to the region where we would actually be able to observe concentrations from those sources. In the absence of such a condition there is no probabilistic penalty for producing a source that is neither supported nor contradicted by the observed measurements. The source widths and emission rates are constrained loosely by uniform priors. The prior distribution on the source widths is

$$p(w_j) = \mathcal{U}(w_j;\, [w_{\min}, w_{\max}]) \quad \text{and} \quad p(\boldsymbol{w} \,|\, m) = \prod_{j=1}^{m} p(w_j) \tag{3.21}$$

where usually $w_{\min} = 0$ and $w_{\max}$ can range from 50 m to 100 m. Similarly we assume that source emission rates are distributed as

$$p(s_j) = \mathcal{U}(s_j;\, [s_{\min}, s_{\max}]) \quad \text{and} \quad p(\boldsymbol{s} \,|\, m) = \prod_{j=1}^{m} p(s_j) \tag{3.22}$$

where $s_{\min} = 0$ and $s_{\max} = 1\ \mathrm{m}^3/\mathrm{s}$. We can write the full prior for the source parameters as

$$p(\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{w} \,|\, m) = \prod_{j=1}^{m} p(\boldsymbol{x}_j) p(s_j) p(w_j). \tag{3.23}$$

The density of the background given by Equation 3.9 gives the prior distribution

$$p(\boldsymbol{b}) = (2\pi)^{-\frac{n-1}{2}} (|\mathbf{J}|^*)^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{b}^\top \mathbf{J} \boldsymbol{b}\right) \tag{3.24}$$

where $|\cdot|^*$ is the generalized determinant. A procedure for sampling from an intrinsic GMRF is given in [28].

From Equation 3.2 we recall that our measurements are observations of the quantity $\mathbf{A}\boldsymbol{s} + \boldsymbol{b}$, where the noise is independently Gaussian with variance σ_ε^2. This corresponds to the likelihood distribution

$$p(\boldsymbol{z} \,|\, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{s}, \boldsymbol{b}, \sigma_\varepsilon) = \mathcal{N}\big(\boldsymbol{z};\ \mathbf{A}\boldsymbol{s} + \boldsymbol{b}, \mathbf{I}\sigma_\varepsilon^2\big). \tag{3.25}$$

For the sake of brevity we will sometimes group the parameters into two sets: the true parameters $\boldsymbol{\theta} = \{\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{w}, \boldsymbol{b}\}$, which we infer, and the auxiliary parameters $\boldsymbol{\phi} = \{\sigma_\varepsilon, \theta_d, \alpha_{\mathrm{H}}, \alpha_{\mathrm{V}}\}$.

We can hence write the posterior distribution as a function of the likelihood and the prior distributions by Bayes' rule:

$$p(\boldsymbol{\theta}, \boldsymbol{\phi} \,|\, \boldsymbol{z}) = \frac{p(\boldsymbol{z} \,|\, \boldsymbol{\theta}, \boldsymbol{\phi}) p(\boldsymbol{\theta}, \boldsymbol{\phi})}{\iint p(\boldsymbol{z} \,|\, \boldsymbol{\theta}, \boldsymbol{\phi}) p(\boldsymbol{\theta}, \boldsymbol{\phi}) \,\mathrm{d}\boldsymbol{\phi}\,\mathrm{d}\boldsymbol{\theta}} \propto p(\boldsymbol{z} \,|\, \boldsymbol{\theta}, \boldsymbol{\phi}) p(\boldsymbol{\theta}, \boldsymbol{\phi}). \tag{3.26}$$

We can expand this notation to include the distributions described in the generative model

$$p(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{s}, \boldsymbol{b}, \boldsymbol{\phi} \,|\, \boldsymbol{z}) \propto p(\boldsymbol{z} \,|\, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{s}, \boldsymbol{b}, \boldsymbol{\phi}) p(\boldsymbol{x}) p(\boldsymbol{w}) p(\boldsymbol{s}) p(\boldsymbol{b}) p(\boldsymbol{\phi}). \tag{3.27}$$

■ 3.3.2 Modeling the auxiliary parameters

Auxiliary parameters include the measurement error standard deviation σ_ε, a constant additive wind direction bias θ_d and α_{H} and α_{V}. Although some of these quantities were omitted from the original development of the model for brevity, we now describe them here. For all of these parameters there was originally no prior distribution assumed for them. We can, however, estimate them using the stochastic inference procedure outlined in the following chapter. In many cases this allows us to better fit the data compared to other methods that do not estimate these as parameters.

The measurement error standard deviation is currently uniformly distributed by

$$p(\sigma_\varepsilon) = \mathcal{U}(\sigma_\varepsilon;\ [0, \varepsilon_{\max}]) \tag{3.28}$$

where $\varepsilon_{\max} = 1 \times 10^{-6}$. Usually $\sigma_\varepsilon = 1 \times 10^{-9}$, as dictated by the sensor specifications.

We also stochastically estimate three parameters related to the wind measurements. The first is an constant additive bias in our measurements of the wind direction; the second and third are multipliers that act as coefficients of the horizontal and vertical plume standard deviations σ_H and σ_V. As originally described, we receive wind measurements $\boldsymbol{u}_i$ corresponding the wind at each measurement location. The measured wind vector $\boldsymbol{u}_i$ could be derived from the true wind state $\boldsymbol{u}_i'$ by

$$|\boldsymbol{u}_i| = |\boldsymbol{u}_i'| \quad \text{and} \quad \theta_{\boldsymbol{u}_i} = \theta_{\boldsymbol{u}_i'} + \theta_d. \tag{3.29}$$

The default value for the measurement error bias is $\theta_d = 0$. The dispersion parameter multipliers modify the horizontal and vertical dispersion parameters from the "true" values σ_{H}' and σ_{V}' as

$$\sigma_{\mathrm{H}} = \alpha_{\mathrm{H}}\sigma_{\mathrm{H}}' \quad \text{and} \quad \sigma_{\mathrm{V}} = \alpha_{\mathrm{V}}\sigma_{\mathrm{V}}' \tag{3.30}$$

where the α_{H} and α_{V} are the horizontal and vertical plume dispersion multipliers, respectively. The default value is 1. All three of these auxiliary parameters have uniform prior distributions so that the inference procedure will select whichever best matches the data. Such stochastic approaches were used successfully in both [17] and [30].

Chapter 4

Inference Procedure

Inference involves reasoning about the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{z})$. In some limited cases, closed-form analysis is possible. However, for more complex models such as ours, one must resort to alternative methods. Here, we consider Markov chain Monte Carlo (MCMC) methods for inference. Such methods rely on drawing samples from the posterior that, when regarded as a set, exactly describe the distribution. MCMC methods generally require some start-up period, called "burn-in," during which they are converging to the true distribution.

In the case of our source estimation model, inference is non-trivial. First, we do not know the number of sources *a priori*. This requires us to use more complex methods to truly sample from the posterior. Another aspect of the source estimation problem that makes inference challenging is that there is some ambiguity inherent in the model. For instance, a source emitting at a low rate could produce measurement contributions arbitrarily close to a source further away but emitting at a higher rate. Similar ambiguity pertains to the width parameter. Generally, such choices are made by the prior, but our prior distributions are simply uniform over the range of valid parameter values. Additionally, we are using a dispersion model that provides an ensemble average. The measurements that are realized in practice will be highly correlated, whereas the Gaussian plume model assumes independent realizations in the forward predictions.

■ 4.1 Initialization by optimization

The initial starting point for the sampling algorithm is obtained by an optimization procedure over a reduced set of the variables. By using an optimization method to obtain an initial solution we dramatically reduce the length of time required for our MCMC procedure to converge to the stationary distribution. We impose an $M \times N$ grid over the full survey area. At the center of each grid cell we assume a point source, i.e. we have $\tilde{m} = MN$ sources with defined locations $\boldsymbol{x}$ and assign every source an initial width w_init. We also take as input the set of measurements $\boldsymbol{z}$, measurement locations $\boldsymbol{y}$, and auxiliary parameters $\boldsymbol{\phi}$. In our experiments, $M = N = 100$ and we initialize all sources to have a width of 20 meters. We set the grid dimensions so that it roughly corresponds to the maximum expected density of sources, which increases the likelihood that we can differentiate sources that are close together even at this early

stage.

The objective of this optimization procedure is to find the configuration $\boldsymbol{\theta}$ that maximizes the posterior distribution. Since optimizing over all the parameters is non-trivial and time-consuming, we instead optimize over a reduced parameter set containing the background $\boldsymbol{b}$ and the source emission rates $\boldsymbol{s}$, and condition on the observations $\boldsymbol{z}$ and the fixed parameters $\boldsymbol{x}$, $\boldsymbol{w}$, and $\boldsymbol{\phi}$:

$$\begin{aligned} \max_{\boldsymbol{s},\boldsymbol{b}} \quad & p(\boldsymbol{s},\boldsymbol{b} \,|\, \boldsymbol{z};\boldsymbol{x},\boldsymbol{w},\boldsymbol{\phi}) \\ \text{s.t.} \quad & 0 \leq s_j \leq s_{\max}, \quad j = 1,\dots,m \\ & 0 \leq b_i, \quad i = 1,\dots,n \end{aligned} \tag{4.1}$$

where the constraints will be denoted by ($\bigstar$). The above formulation is equivalent to

$$\begin{aligned} \max_{\boldsymbol{s},\boldsymbol{b}} \quad & p(\boldsymbol{z} \,|\, \boldsymbol{b},\boldsymbol{s};\boldsymbol{x},\boldsymbol{w},\boldsymbol{\phi})p(\boldsymbol{b})p(\boldsymbol{s}) \\ \text{s.t.} \quad & (\bigstar), \end{aligned} \tag{4.2}$$

where we refer to data likelihood distribution from Equation 3.25

$$p(\boldsymbol{z} \,|\, \boldsymbol{s},\boldsymbol{b};\boldsymbol{x},\boldsymbol{w},\boldsymbol{\phi}) \propto \exp\left(-\frac{1}{2\sigma_\varepsilon^2}\|\mathbf{A}\boldsymbol{s}+\boldsymbol{b}-\boldsymbol{z}\|_2^2\right) \tag{4.3}$$

where $\boldsymbol{z}$ is $n \times 1$, $\boldsymbol{s}$ is $\tilde{m} \times 1$, and $\mathbf{A}$ is $n \times \tilde{m}$. We also have the background prior distribution from Equation 3.24

$$p(\boldsymbol{b}) \propto \exp\left(-\frac{1}{2}\boldsymbol{b}^\top \mathbf{J}\boldsymbol{b}\right). \tag{4.4}$$

Following [17], we assume a sparse Laplace prior on the source emission rates only during the optimization phase:

$$p(\boldsymbol{s}) \propto \exp(-\lambda\|\mathbf{Q}\boldsymbol{s}\|_1) \tag{4.5}$$

where we set $\mathbf{Q} = \mathbf{I}$.

Since logarithms are strictly increasing we can rewrite Problem 4.2 as

$$\begin{aligned} \min_{\boldsymbol{s},\boldsymbol{b}} \quad & -\log p(\boldsymbol{z} \,|\, \boldsymbol{b},\boldsymbol{s};\boldsymbol{x},\boldsymbol{w},\boldsymbol{\phi}) - \log p(\boldsymbol{b}) - \log p(\boldsymbol{s}) \\ \text{s.t.} \quad & (\bigstar) \end{aligned} \tag{4.6}$$

where we have also converted the maximization to a minimization by negating the objective function. Plugging in Equations 4.3-4.5 and the constraints ($\bigstar$) we get

$$\begin{aligned} \min_{\boldsymbol{s},\boldsymbol{b}} \quad & Z(\boldsymbol{s},\boldsymbol{b}) = \frac{1}{2\sigma_\varepsilon^2}\|\mathbf{A}\boldsymbol{s}+\boldsymbol{b}-\boldsymbol{z}\|_2^2 + \frac{1}{2}\boldsymbol{b}^\top\mathbf{J}\boldsymbol{b} + \lambda\sum_i s_i \\ \text{s.t.} \quad & 0 \leq s_j \leq s_{\max}, \quad j = 1,\dots,m \\ & 0 \leq b_i, \quad i = 1,\dots,n. \end{aligned} \tag{4.7}$$

where we have used the nonnegativity constraint on $\boldsymbol{s}$ to rewrite the L1-norm as a sum:

$$\lambda \|\mathbf{Q}\boldsymbol{s}\|_1 = \lambda \sum_i s_i.$$

Note that our objective function $Z(\boldsymbol{s}, \boldsymbol{b})$ is a convex function. It is obvious that the likelihood term and the emission rate term are both convex functions. Since $\mathbf{J}$ is PSD by construction we know that $\boldsymbol{b}^\top \mathbf{J}\boldsymbol{b}$, hence $Z(\boldsymbol{s}, \boldsymbol{b})$ is the sum of three convex functions and is also convex. For minimization of a convex function, local optima are the global optima, so any optimization routine that converges is guaranteed to converge to the global optimum.

We first solve for the optimal background $\boldsymbol{b}$, then for the optimal emission rates $\boldsymbol{s}$, and repeat until convergence. This is known as *optimizing by alternation*, a procedure which is guaranteed to converge to a local optima. Further, since we are minimizing a convex $Z(\boldsymbol{s}, \boldsymbol{b})$, this local minimum is also the global minimum. Importantly, we do have several constraints on the values that $\boldsymbol{s}$ and $\boldsymbol{b}$ can take. These constraints define a feasible solution set. Since our objective function is convex, there are only two possible outcomes for our inference procedure: (1) the solution is unique and fully within the feasible set, or (2) the solution is on the boundary of the feasible set and may not be unique.

We begin by setting $\boldsymbol{s} = \mathbf{0}$ for the sources defined at the center of each grid cell. The optimization steps outlined above are repeatedly applied until we converge to a solution is reached or we exceed some maximum number of iterations. The optimization procedure moves towards convergence, at each iteration attributing a greater concentration to the sources and less to the ambient background. We then initialize by sampling without replacement m_{init} sources, or grid cells, from the optimization solution. Each cell has normalized weight

$$\tilde{s}_j = \frac{s_j \mathbb{1}(\boldsymbol{x}_j \in C^*)}{\sum_{i=1}^{\tilde{m}} s_i \mathbb{1}(\boldsymbol{x}_i \in C^*)} \tag{4.8}$$

where $\mathbb{1}(\boldsymbol{x}_j \in C^*)$ ensures that source j is inside the set of observable locations given the measurement locations $\boldsymbol{y}$ and the wind direction $\boldsymbol{u}$.

We now describe the two optimization steps in more detail.

■ 4.1.1 Optimizing over the background

When solving for the optimal background we are attempting to solve the following restricted version of Equation 4.7:

$$\begin{aligned} \min_{\boldsymbol{b}} \quad & Z(\boldsymbol{b}) = \frac{1}{2\sigma_\varepsilon^2} \|\mathbf{A}\boldsymbol{s} + \boldsymbol{b} - \boldsymbol{z}\|_2^2 + \frac{1}{2}\boldsymbol{b}^\top \mathbf{J}\boldsymbol{b} \\ \text{s.t.} \quad & b_i \geq 0, \quad i = 1, \ldots, n \end{aligned} \tag{4.9}$$

First, assume that the constraints $\boldsymbol{b} \geq 0$ will be implicitly satisfied by any critical point of our objective function $Z(\boldsymbol{b})$. We will argue this momentarily. Since $Z(\boldsymbol{b})$ is the

sum of two squared forms we can find its critical point by writing it as a single squared form:

$$\begin{aligned} Z(\boldsymbol{b}) &= \frac{1}{2\sigma_\varepsilon^2}\|\mathbf{A}\boldsymbol{s}+\boldsymbol{b}-\boldsymbol{z}\|_2^2+\frac{1}{2}\boldsymbol{b}^\top\mathbf{J}\boldsymbol{b} \\ &= \frac{1}{2\sigma_\varepsilon^2}(\boldsymbol{b}-(\boldsymbol{z}-\mathbf{A}\boldsymbol{s}))^\top(\boldsymbol{b}-(\boldsymbol{z}-\mathbf{A}\boldsymbol{s}))+\frac{1}{2}\boldsymbol{b}^\top\mathbf{J}\boldsymbol{b} \\ &= \frac{1}{2}(\boldsymbol{b}-\boldsymbol{m})^\top\mathbf{R}(\boldsymbol{b}-\boldsymbol{m})+C \end{aligned}$$

where

$$\begin{aligned} \mathbf{R} &= \mathbf{I}\sigma_\varepsilon^{-2}+\mathbf{J} \\ \boldsymbol{m} &= \mathbf{R}^{-1}\sigma_\varepsilon^{-2}(\boldsymbol{z}-\mathbf{A}\boldsymbol{s}) \\ &= \left(\mathbf{I}\sigma_\varepsilon^{-2}+\mathbf{J}\right)^{-1}\sigma_\varepsilon^{-2}(\boldsymbol{z}-\mathbf{A}\boldsymbol{s}) \\ C &= \frac{1}{2\sigma_\varepsilon^2}(\boldsymbol{z}-\mathbf{A}\boldsymbol{s})^\top(\boldsymbol{z}-\mathbf{A}\boldsymbol{s})-\frac{1}{2}\boldsymbol{m}^\top\mathbf{J}_c\boldsymbol{m} \\ &= \frac{1}{2\sigma_\varepsilon^2}(\boldsymbol{z}-\mathbf{A}\boldsymbol{s})^\top(\boldsymbol{z}-\mathbf{A}\boldsymbol{s})-\frac{1}{2\sigma_\varepsilon^4}(\boldsymbol{z}-\mathbf{A}\boldsymbol{s})^\top\left(\mathbf{I}\sigma_\varepsilon^{-2}+\mathbf{J}\right)^{-1}(\boldsymbol{z}-\mathbf{A}\boldsymbol{s}). \end{aligned}$$

We know that $\mathbf{R}$ is guaranteed to be positive definite (PD) since $\mathbf{J}$ is PSD but we add σ_ε^{-2} to the diagonal elements making it PD. It follows that

$$(\boldsymbol{b}-\boldsymbol{m})^\top\mathbf{R}(\boldsymbol{b}-\boldsymbol{m})\geq 0 \tag{4.10}$$

for all $\boldsymbol{b}$. Hence, the value of $\boldsymbol{b}$ that minimizes the quadratic term $Z(\boldsymbol{b})$ is

$$\boldsymbol{b}^*=\boldsymbol{m}=\left(\mathbf{I}\sigma_\varepsilon^{-2}+\mathbf{J}\right)^{-1}\sigma_\varepsilon^{-2}(\boldsymbol{z}-\mathbf{A}\boldsymbol{s}) \tag{4.11}$$

For this to be the optimal solution to the above problem the constraints $\boldsymbol{b}^*\geq\mathbf{0}$ must be satisfied. From above we know that $\mathbf{R}=\mathbf{I}\sigma_\varepsilon^{-2}+\mathbf{J}$ is PD, so we know that $\mathbf{R}^{-1}=\left(\mathbf{I}\sigma_\varepsilon^{-2}+\mathbf{J}\right)^{-1}$ is also PD. Therefore, $\boldsymbol{b}^*\geq\mathbf{0}$ if $\boldsymbol{z}-\mathbf{A}\boldsymbol{s}\geq\mathbf{0}$. We now show that this is true based on the orders of magnitude of the terms involved. We recall that the measurements are modeled as

$$z_i=\boldsymbol{a}_i^\top\boldsymbol{s}+b_i+\varepsilon_i \tag{4.12}$$

where $\boldsymbol{a}_i^\top\boldsymbol{s}\approx 10^{-8}$, $b_i\approx 10^{-6}$, and $\varepsilon_i\approx 10^{-9}$. These are the magnitudes commonly observed for these various quantities in practice. Hence, z_i is also on the order of 10^{-6} and is at least one order of magnitude greater than the source contribution $\boldsymbol{a}_i^\top\boldsymbol{s}$. Thus, we know that $\boldsymbol{z}-\mathbf{A}\boldsymbol{s}\geq\mathbf{0}$ and that the $\boldsymbol{b}^*$ given in Equation 4.11 satisfies the constraints $\boldsymbol{b}\geq\mathbf{0}$. Therefore, $\boldsymbol{b}^*$ is the optimal solution.

We note that although the matrix $\mathbf{R}$ is large (i.e. 1000×1000), it is also independent of $\boldsymbol{b}$ and $\boldsymbol{s}$ so it only needs to be inverted once. This allows for very rapid optimization of the background $\boldsymbol{b}$.

An alternative approach to optimizing over the background parameters is to minimize a quadratic form of the objective function for each individual component b_i. This allows us to explicitly verify that the background constraints are satisfied but it is much slower to converge.

■ 4.1.2 Optimizing over the source emission rates

The process of solving for $\boldsymbol{s}$ entails solving the following

$$\begin{aligned} \min_{\boldsymbol{s}} \quad & Z(\boldsymbol{s}) = \frac{1}{2\sigma_\varepsilon^2}\|\mathbf{A}\boldsymbol{s} + \boldsymbol{b} - \boldsymbol{z}\|_2^2 + \lambda \sum_i s_i \\ \text{s.t.} \quad & 0 \leq s_j \leq s_{\max}, \quad j = 1, \ldots, m \end{aligned} \tag{4.13}$$

which is a quadratic function of $\boldsymbol{s}$. When optimizing $Z(\boldsymbol{b})$ we recognized that the constraints would almost never be active based on the initialization. This allowed us to treat the problem as the unconstrained minimization of a convex function, and since we could write the objective function as a quadratic equation we could obtain its critical point in closed-form. Although we have a similar structure here and we can indeed write $Z(\boldsymbol{s})$ as a quadratic equation with a closed-form critical point, the constraints will not be implicitly satisfied as before. We thus apply optimization by alternation over the sources j to obtain the optimal $\boldsymbol{s}$.

Since $Z(\boldsymbol{s})$ has a quadratic form, $Z(s_j; \boldsymbol{s}_{-j})$ also has a quadratic form so we can solve for the critical point in closed form and manually enforce the constraints. For convenience of notation we will write $\mathbf{A} = \begin{bmatrix} \mathbf{A}_j & \mathbf{A}_{-j} \end{bmatrix}$ where $\mathbf{A}_j$ is the j-th column of $\mathbf{A}$ and $\mathbf{A}_{-j}$ are all except the j-th column of $\mathbf{A}$. Similarly we define $\boldsymbol{s} = \begin{bmatrix} s_j & \boldsymbol{s}_{-j}^\top \end{bmatrix}^\top$ where $\boldsymbol{s}_{-j}$ contains all except the j-th element of $\boldsymbol{s}$.

The objective $Z(s_j; \boldsymbol{s}_{-j})$ has a critical point at the value

$$s_j^* = \frac{-b}{2a} \tag{4.14}$$

where

$$a = \frac{1}{2\sigma_\varepsilon^2}\left(\mathbf{A}_j^\top \mathbf{A}_j\right) \qquad \text{and} \qquad b = \frac{1}{\sigma_\varepsilon^2}(\mathbf{A}_{-j}\boldsymbol{s}_{-j} + \boldsymbol{b} - \boldsymbol{z})^\top \mathbf{A}_j + \lambda.$$

This critical point is the solution to the unconstrained minimization of the objective function. In the constrained problem the minimum may either be at this critical point or a point on the boundary. Since $Z(s_j; \boldsymbol{s}_{-j})$ is a univariate function we can easily check the three possibilities for the constrained minimization solution s_j':

$$s_j' = \begin{cases} 0 & s_j^* \leq 0 \\ s_j^* & 0 \leq s_j^* \leq s_{\max} \\ s_{\max} & s_j^* \geq s_{\max}. \end{cases} \tag{4.15}$$

Derivation of the critical point

We now derive the expression for the critical point obtained for each source j. To begin we expand the objective function

$$\begin{aligned} Z(s_j; \boldsymbol{s}_{-j}) &= \frac{1}{2\sigma_\varepsilon^2}\|\mathbf{A}\boldsymbol{s} + \boldsymbol{b} - \boldsymbol{z}\|_2^2 + \lambda \sum_i s_i \\ &= \frac{1}{2\sigma_\varepsilon^2}(\mathbf{A}\boldsymbol{s} + \boldsymbol{b} - \boldsymbol{z})^\top(\mathbf{A}\boldsymbol{s} + \boldsymbol{b} - \boldsymbol{z}) + \lambda \sum_i s_i \\ &= \frac{1}{2\sigma_\varepsilon^2}(\boldsymbol{s}^\top\mathbf{A}^\top\mathbf{A}\boldsymbol{s} + 2(\boldsymbol{b} - \boldsymbol{z})^\top\mathbf{A}\boldsymbol{s} + (\boldsymbol{b} - \boldsymbol{z})^\top(\boldsymbol{b} - \boldsymbol{z})) + \lambda \sum_i s_i. \end{aligned}$$

We now write some of the terms in the above quadratic form as functions of s_j:

$$\begin{aligned} \mathbf{A}\boldsymbol{s} &= \mathbf{A}_j s_j + \mathbf{A}_{-j}\boldsymbol{s}_{-j} \\ \boldsymbol{s}^\top\mathbf{A}^\top\mathbf{A}\boldsymbol{s} &= \mathbf{A}_j^\top\mathbf{A}_j s_j^2 + 2\boldsymbol{s}_{-j}^\top\mathbf{A}_{-j}^\top\mathbf{A}_j s_j + \boldsymbol{s}_{-j}^\top\mathbf{A}_{-j}^\top\mathbf{A}_{-j}\boldsymbol{s}_{-j} \\ \lambda\sum_i s_i &= \lambda s_j + \lambda \sum_{k=-j} s_k. \end{aligned}$$

Plugging these into $Z(\boldsymbol{s})$ and grouping terms by dependence on s_j yields

$$\begin{aligned} Z(s_j; \boldsymbol{s}_{-j}) &= \frac{1}{2\sigma_\varepsilon^2}\Big(\boldsymbol{s}^\top\mathbf{A}^\top\mathbf{A}\boldsymbol{s} + 2(\boldsymbol{b} - \boldsymbol{z})^\top\mathbf{A}\boldsymbol{s} + (\boldsymbol{b} - \boldsymbol{z})^\top(\boldsymbol{b} - \boldsymbol{z})\Big) + \lambda \sum_i s_i \\ &= \frac{1}{2\sigma_\varepsilon^2}\Big(\mathbf{A}_j^\top\mathbf{A}_j s_j^2 + 2\boldsymbol{s}_{-j}^\top\mathbf{A}_{-j}^\top\mathbf{A}_j s_j + \boldsymbol{s}_{-j}^\top\mathbf{A}_{-j}^\top\mathbf{A}_{-j}\boldsymbol{s}_{-j} \\ &\qquad +2(\boldsymbol{b} - \boldsymbol{z})^\top(\mathbf{A}_j s_j + \mathbf{A}_{-j}\boldsymbol{s}_{-j}) + (\boldsymbol{b} - \boldsymbol{z})^\top(\boldsymbol{b} - \boldsymbol{z})\Big) + \lambda s_j + \lambda \sum_{k=-j} s_k \\ &= \frac{1}{2\sigma_\varepsilon^2}\Big(\mathbf{A}_j^\top\mathbf{A}_j s_j^2 + 2\boldsymbol{s}_{-j}^\top\mathbf{A}_{-j}^\top\mathbf{A}_j s_j + 2(\boldsymbol{b} - \boldsymbol{z})^\top\mathbf{A}_j s_j + \boldsymbol{s}_{-j}^\top\mathbf{A}_{-j}^\top\mathbf{A}_{-j}\boldsymbol{s}_{-j} \\ &\qquad +2(\boldsymbol{b} - \boldsymbol{z})^\top\mathbf{A}_{-j}\boldsymbol{s}_{-j} + (\boldsymbol{b} - \boldsymbol{z})^\top(\boldsymbol{b} - \boldsymbol{z})\Big) + \lambda s_j + \lambda \sum_{k=-j} s_k. \end{aligned}$$

The last equation is easily rewritten as the quadratic equation

$$Z(s_j; \boldsymbol{s}_{-j}) = as_j^2 + bs_j + c \tag{4.16}$$

where we have coefficients

$$a = \frac{1}{2\sigma_\varepsilon^2}\left(\mathbf{A}_j^\top \mathbf{A}_j\right) \tag{4.17}$$

$$b = \frac{1}{2\sigma_\varepsilon^2}\left(2\boldsymbol{s}_{-j}^\top \mathbf{A}_{-j}^\top \mathbf{A}_j + 2(\boldsymbol{b}-\boldsymbol{z})^\top \mathbf{A}_j\right) + \lambda \tag{4.18}$$

$$= \frac{1}{\sigma_\varepsilon^2}(\mathbf{A}_{-j}\boldsymbol{s}_{-j} + \boldsymbol{b} - \boldsymbol{z})^\top \mathbf{A}_j + \lambda \tag{4.19}$$

$$c = \frac{1}{2\sigma_\varepsilon^2}\left(\boldsymbol{s}_{-j}^\top \mathbf{A}_{-j}^\top \mathbf{A}_{-j}\boldsymbol{s}_{-j} + 2(\boldsymbol{b}-\boldsymbol{z})^\top \mathbf{A}_{-j}\boldsymbol{s}_{-j} + (\boldsymbol{b}-\boldsymbol{z})^\top(\boldsymbol{b}-\boldsymbol{z})\right) + \lambda \sum_{k=-j} s_k. \tag{4.20}$$

The objective $Z(s_j; \boldsymbol{s}_{-j})$ has a critical point at the value s_j^* at which the derivative goes to zero. From the above we have

$$Z'(s_j) = 2as_j + b, \tag{4.21}$$

and the critical point

$$s_j^* = \frac{-b}{2a} \tag{4.22}$$

where a and b are defined as above.

Implementation concerns

A naive implementation of the above equations for s_j^* will be computationally expensive, since for each j we compute $\mathbf{A}_{-j}\boldsymbol{s}_{-j}$, where the number of columns of $\mathbf{A}_{-j}$ is $\tilde{m} \approx 10^4$ and we would repeat this multiplication $\tilde{m}$ times per iteration. Dramatic gains can be obtained by noting that

$$\mathbf{A}_{-j}\boldsymbol{s}_{-j} = \mathbf{A}\boldsymbol{s} - \mathbf{A}_j s_j \tag{4.23}$$

and computing this instead. It is worth noting that at the end of each iteration we need to update the product

$$\mathbf{A}\boldsymbol{s} = \mathbf{A}_{-j}\boldsymbol{s}_{-j} + \mathbf{A}_j s_j'. \tag{4.24}$$

Although numerical instabilities have not been an issue in our implementations, one could easily guard against them by simply recomputing the full product $\mathbf{A}\boldsymbol{s}$ periodically. In our experiments we do this every ten iterations.

■ 4.2 Sampling from the posterior distribution

Our inference procedure falls into the category of Markov chain Monte Carlo (MCMC) methods, or sampling-based methods. In this section we discuss the application of the techniques in Section 2.2 to the model of Chapter 3. Since the number of sources in the model is unknown and the number of parameters (i.e. the dimensionality of $\boldsymbol{x}$, $\boldsymbol{w}$, and $\boldsymbol{s}$) change based on the number of sources, we apply RJ-MCMC to infer the

procedure NEWSAMPLE($\boldsymbol{x}$, $\boldsymbol{w}$, $\boldsymbol{s}$, $\boldsymbol{b}$, σ_ε, m)
  for $i = 1, \ldots, M$ **do** ▷ sample new source locations
    $x'_{i,1} \leftarrow$ SAMPLEMH($x_{i,1}$, S_x)
    $x'_{i,2} \leftarrow$ SAMPLEMH($x_{i,2}$, S_x)
  end for
  for $i = 1, \ldots, M$ **do** ▷ sample new source half-widths
    $w'_i \leftarrow$ SAMPLEMH(w_i, S_w)
  end for
  for $i = 1, \ldots, M$ **do** ▷ sample new source emission rates
    $s'_i \leftarrow$ SAMPLEMH(s_i, S_s)
  end for
  $\boldsymbol{b}' \leftarrow$ SAMPLEGIBBS($\boldsymbol{b}$) ▷ sample new background vector
  $\sigma'_\varepsilon \leftarrow$ SAMPLEMH(σ_ε, R_σ) ▷ sample new measurement error
  $\langle m', \boldsymbol{x}', \boldsymbol{w}', \boldsymbol{s}' \rangle \leftarrow$ SAMPLEORDER(m, $\boldsymbol{x}'$, $\boldsymbol{w}'$, $\boldsymbol{s}'$) ▷ sample new source count
  return $\langle \boldsymbol{x}', \boldsymbol{w}', \boldsymbol{s}', \boldsymbol{b}', \sigma'_\varepsilon, m' \rangle$
end procedure

Algorithm 4.1: RJMCMC inference procedure

posterior distribution of the source parameters given the data. Finally, we derive the trans-dimensional "jumping" steps that allow us to estimate the number of sources.

The general inference procedure is outlined in Algorithm 4.1, which obtains a new sample of parameter values from an existing sample. The SAMPLEMH routine entails sampling the specified parameter using one iteration of the Metropolis-Hastings algorithm as in Section 4.2.1. The function SAMPLEGIBBS refers to running an iteration of Gibbs sampling over the background contribution vector $\boldsymbol{b}$ as is described in Section 4.2.2. The SAMPLEORDER function allow the chain to jump between models of different orders, or number of sources. Section 4.2.3 discusses the nature of these jumps.

■ 4.2.1 Sampling with Metropolis-Hastings

The Metropolis-Hastings algorithm is used to sample source parameters $\boldsymbol{x}$, $\boldsymbol{w}$, and $\boldsymbol{s}$ and the measurement error σ_ε. All of these have Gaussian proposal distributions with the following standard deviations: $S_x = 50$ m, $S_w = 10$ m, $S_s = 0.0016$ m^{3}s^{-1}, and $S_\varepsilon = 0.05 \log_{10}(1 \times 10^{-9})$ ppb. Since Gaussian proposals are symmetric, the Hastings ratio reduces to the ratio of the likelihoods for each configuration.

For example, we show the Hastings ratio for a new source emission rate proposal. Let's assume that the proposed emission rate is valid according to the prior—if this were not the case the Hastings ratio would evaluate to zero. Let's also note that $\boldsymbol{\theta}^* = \{\{\boldsymbol{\theta}^{(k-1)} \backslash s_j^{(k-1)}\} \cup s_j^*\}$, so that the only difference between $\boldsymbol{\theta}^{(k-1)}$ and $\boldsymbol{\theta}^*$ is the

single value that has been proposed.

$$\begin{aligned} r &= \frac{p(\boldsymbol{\theta}^* \mid \boldsymbol{z}) q(\boldsymbol{\theta}^{(k-1)} \mid \boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^{(k-1)} \mid \boldsymbol{z}) q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(k-1)})} \\ &= \frac{p(\boldsymbol{z} \mid \boldsymbol{\theta}^*)}{p(\boldsymbol{z} \mid \boldsymbol{\theta}^{(k-1)})} \cdot \frac{p(\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^{(k-1)})} \cdot \frac{q(\boldsymbol{\theta}^{(k-1)} \mid \boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(k-1)})} \\ &= \frac{\mathcal{N}(\boldsymbol{z}; \mathbf{A}^* \boldsymbol{s}^* + \boldsymbol{b}^*, \mathbf{I}(\sigma_\varepsilon^*)^2)}{\mathcal{N}(\boldsymbol{z}; \mathbf{A}^{(k-1)} \boldsymbol{s}^{(k-1)} + \boldsymbol{b}^{(k-1)}, \mathbf{I}(\sigma_\varepsilon^{(k-1)})^2)} \cdot 1 \cdot \frac{\mathcal{N}(\boldsymbol{\theta}^{(k-1)}; \boldsymbol{\theta}^*, S_s^2)}{\mathcal{N}(\boldsymbol{\theta}^*; \boldsymbol{\theta}^{(k-1)}, S_s^2)} \\ &= \frac{\mathcal{N}(\boldsymbol{z}; \mathbf{A}^* \boldsymbol{s}^* + \boldsymbol{b}^*, \mathbf{I}(\sigma_\varepsilon^*)^2)}{\mathcal{N}(\boldsymbol{z}; \mathbf{A}^{(k-1)} \boldsymbol{s}^{(k-1)} + \boldsymbol{b}^{(k-1)}, \mathbf{I}(\sigma_\varepsilon^{(k-1)})^2)}, \end{aligned} \tag{4.25}$$

which is just the ratio of the likelihoods. The only thing that would change for any of the other variables listed is the proposal variance, which doesn't even appear in the likelihood ratio because the proposals are symmetric.

■ 4.2.2 Sampling via Gibbs sampling

We sample the background $\boldsymbol{b}$ from its full conditional using Gibbs sampling. Since the background is conditionally independent of the other parameters given the measurements the full conditional distribution is given by

$$p(\boldsymbol{b} \mid \boldsymbol{z}) \propto p(\boldsymbol{z} \mid \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{s}, \boldsymbol{b}, \sigma_\varepsilon) p(\boldsymbol{b}; \boldsymbol{b}_0) \tag{4.26}$$

$$\propto \mathcal{N}(\boldsymbol{z}; \mathbf{A}\boldsymbol{s} + \boldsymbol{b}, \mathbf{I}\sigma_\varepsilon^2) \mathcal{N}(\boldsymbol{b}; \boldsymbol{b}_0, \mathbf{J}_b^{-1}) \tag{4.27}$$

$$\propto \mathcal{N}(\boldsymbol{b}; \mathbf{J}^{-1}\boldsymbol{h}, \mathbf{J}^{-1}) \tag{4.28}$$

where

$$\begin{aligned} \boldsymbol{h} &= \mathbf{I}\sigma_\varepsilon^{-2}(\boldsymbol{y} - \mathbf{A}\boldsymbol{s}) + \mathbf{J}_b \boldsymbol{b}_0 \\ \mathbf{J} &= \mathbf{I}\sigma_\varepsilon^{-2} + \mathbf{J}_b. \end{aligned}$$

Gibbs sampling generates a new value for each element b_i by applying the following steps:

1. Compute the new information mean:

$$\boldsymbol{h}_i' = \boldsymbol{h}_i - \prod_{j \in \mathrm{nbr}(i)} \mathbf{J}_{ij} \boldsymbol{b}_j \tag{4.29}$$

2. Sample a new $\boldsymbol{b}_i$:

$$\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{J}_{ii}^{-1} \boldsymbol{h}_i', \mathbf{J}_{ii}^{-1}) \tag{4.30}$$

This is done for every element of $\boldsymbol{b}$ to sample a new state. It is worth noting that it can take many iterations for the sequence of states to converge, depending on the initial value of $\boldsymbol{b}$. We accelerate convergence by initializing with $\boldsymbol{b} = \mathbf{J}^{-1}\boldsymbol{h}$.

This step includes the inversion of a large matrix which is computationally intensive. We can reduce the frequency with which we perform inversion by only recomputing $\mathbf{J}_*^{-1}$ when it changes. Specifically, $\mathbf{J}^{-1}$ only needs to be recomputed when the proposed σ_ε was accepted on the previous step, or when one of the auxiliary parameters for the wind field was accepted, hence changing $\mathbf{J}$. When we do have to perform an inversion, we invert the lower triangular matrix obtained by the Cholesky decomposition instead of the full matrix.

■ 4.2.3 Sampling model order via RJ-MCMC

The fact that we do not know the number of sources poses a challenge to performing inference on our model. The standard MH algorithm could not be applied in a straightforward manner, since we could not explicitly define our parameter set. An extension of the MH algorithm, called *Reversible-jump Markov chain Monte Carlo* (RJ-MCMC) sampling, addresses this issue. This method was first developed in [13]. Further considerations for this approach are found in [10].

Suppose our parameter set $\boldsymbol{\theta}$ corresponds to some model M_k, where the set of all possible models is given by $\mathcal{M} = \{M_k\}_{k=0}^{m_{\max}}$. In our case, the model order k corresponds directly to the number of sources in the model. In effect, we add the model order k to the set of parameters to be estimated, so that our target (posterior) distribution is:

$$p(k, \boldsymbol{\theta}_k \mid \boldsymbol{z}) \propto p(\boldsymbol{z} \mid \boldsymbol{\theta}_k, k) p(\boldsymbol{\theta}_k \mid k) p(k). \tag{4.31}$$

Here, we use $\boldsymbol{\theta}_k$ to denote that this particular $\boldsymbol{\theta}$ is of model M_k, or is of model order k. Suppose that our current estimate is $\boldsymbol{\theta}_k$ of model M_k and we want to propose some estimate $\boldsymbol{\theta}'_{k'}$ of model $M_{k'}$. We can generate and accept a new proposal of the desired model order by the following procedure.

1. Propose a new model order $M_{k'}$ with jump probability $j(k' \mid k)$.
2. Generate an auxiliary random variable $\boldsymbol{\psi}$ from the proposal distribution $g_{k,k'}(\boldsymbol{\psi} \mid \boldsymbol{\theta}_k)$.
3. Compute the proposed model's parameters $\boldsymbol{\theta}'_{k'}$ using a deterministic function, $(\boldsymbol{\theta}'_{k'}, \boldsymbol{\psi}') = h_{k,k'}(\boldsymbol{\theta}_k, \boldsymbol{\psi})$.
4. Compute the new Hastings ratio:

$$r = \frac{p(k', \boldsymbol{\theta}'_{k'} \mid \boldsymbol{z})}{p(k, \boldsymbol{\theta}_k \mid \boldsymbol{z})} \frac{j(k \mid k')}{j(k' \mid k)} \frac{g_{k',k}(\boldsymbol{\psi}' \mid \boldsymbol{\theta}'_{k'})}{g_{k,k'}(\boldsymbol{\psi} \mid \boldsymbol{\theta}_k)} \left| \frac{\partial h_{k,k'}(\boldsymbol{\theta}_k, \boldsymbol{\psi})}{\partial(\boldsymbol{\theta}_k, \boldsymbol{\psi})} \right| \tag{4.32}$$

5. Accept the new model order with acceptance probability $\alpha = \min\{1, r\}$.

There are several things to note here. We are effectively forming a set $\{\boldsymbol{\theta}_k, \boldsymbol{\psi}\}$ such that $\dim\{\boldsymbol{\theta}_k, \boldsymbol{\psi}\} = \dim\{\boldsymbol{\theta}'_{k'}, \boldsymbol{\psi}'\}$. This allows us to "jump" across dimensions. We also define the function h^{-1} for the reverse step such that $(\boldsymbol{\theta}_k, \boldsymbol{\psi}) = h^{-1}_{k,k'}(\boldsymbol{\theta}'_{k'}, \boldsymbol{\psi}')$. These are defined for all valid combinations of k and k'.

As required by the RJMCMC framework, we augment the set of parameters to be estimated with the model order k. We define the model order as the number of sources estimated in the model, so model order k consists of k sources. The goal is to transition from parameter vector $\boldsymbol{\theta}_m$ to a derived parameter vector $\boldsymbol{\theta}_{k'}$ with k' sources.

We propose a parameter vector $\boldsymbol{\theta}_{k'}$ using one of the following four equally-likely jump types to change the model order. We note that the jump types are paired so that the Markov chain of samples is reversible. For example, the birth and death types are paired. Moreover, the product of the Hastings ratios for paired steps must equal 1.

1. Source birth ($k' = k + 1$): generate the $(k+1)$th source by drawing new random variables.

$$\begin{aligned} \boldsymbol{x}_{j*} &= \boldsymbol{r}_x & \boldsymbol{r}_{x1} &\sim \mathcal{U}(I_{x1}) \text{ and } \boldsymbol{r}_{x2} \sim \mathcal{U}(I_{x2}) \\ w_{j*} &= r_w & r_w &\sim \mathcal{U}(0, R_w) \\ s_{j*} &= r_s & r_s &\sim \mathcal{U}(0, R_s) \end{aligned}$$

   where I_{x1} and I_{x2} are the x_1 and x_2 domain of the survey area, $R_w = w_{\max} - w_{\min}$ is the range of source widths, and $R_s = s_{\max} - s_{\min}$ is the range of emission rates.

2. Source death ($k' = k - 1$): remove an existing source $j \sim \mathcal{U}(j; 1, k)$ from the parameter set.

3. Source split ($k' = k + 1$): split an existing source $j^* \sim \mathcal{U}(j^*; 1, k)$ into two new sources j^*_- and j^*_+

$$\begin{aligned} \boldsymbol{x}_{j^*_\pm} &= \boldsymbol{x}_{j^*} \pm \boldsymbol{r}_x & \boldsymbol{r}_x^{(1)} &\sim \mathcal{U}(-\tfrac{E_{x1}}{2}, \tfrac{E_{x1}}{2}) \text{ and } \boldsymbol{r}_x^{(2)} \sim \mathcal{U}(-\tfrac{E_{x2}}{2}, \tfrac{E_{x2}}{2}) \\ w_{j^*_\pm} &= w_{j^*} \pm r_w & r_w &\sim \mathcal{U}(-\tfrac{E_w}{2}, \tfrac{E_w}{2}) \\ s_{j^*_\pm} &= s_{j^*} \pm r_s & r_s &\sim \mathcal{U}(-\tfrac{E_s}{2}, \tfrac{E_s}{2}) \end{aligned}$$

   where $E_{x1} = 100$ m, $E_{x2} = 100$ m, $E_w = 10$ m, and $E_s = 0.1$ m^{3}s^{-1}.

4. Source merge ($k' = k - 1$): select two sources randomly and generate a new source whose parameters are obtained by averaging the parameters of the randomly selected sources.

After selecting the jump type and computing the new parameter vector $\boldsymbol{\theta}_{k'}$ as described above, we compute the Hastings ratio

$$r = \frac{p(k', \boldsymbol{\theta}'_{k'} \,|\, \boldsymbol{z})}{p(k, \boldsymbol{\theta}_k \,|\, \boldsymbol{z})} \frac{j(k \,|\, k')}{j(k' \,|\, k)} \frac{g_{k',k}(\boldsymbol{\psi}' \,|\, \boldsymbol{\theta}'_{k'})}{g_{k,k'}(\boldsymbol{\psi} \,|\, \boldsymbol{\theta}_k)} \left| \frac{\partial h_{k,k'}(\boldsymbol{\theta}_k, \boldsymbol{\psi})}{\partial(\boldsymbol{\theta}_k, \boldsymbol{\psi})} \right|. \tag{4.33}$$

We note that the $p(k, \boldsymbol{\theta}_k \,|\, \boldsymbol{z})$ is our posterior (target) distribution. The proposal is generated by j and g: $j(k' \,|\, k)$ is the likelihood of jumping from model order k to order k', and $g_{k,k'}(\cdot \,|\, \boldsymbol{\theta}_k)$ is the distribution of auxiliary variables required to jump from model order k to order k'. Finally, $h_{k,k'}(\boldsymbol{\theta}_k, \boldsymbol{\psi})$ is a deterministic function that augments the parameter vector of order k with the auxiliary variables $\boldsymbol{\psi}$ into a parameter vector of order k'. We now specialize Equation 4.33 for the four different jump types outlined above.

Source birth and death

For birth jumps, we have $k = m$ and $k' = k + 1$, which we insert into Equation 4.33 to obtain

$$\alpha = \frac{p(m+1, \boldsymbol{\theta}'_{m+1} \,|\, \boldsymbol{z})}{p(m, \boldsymbol{\theta}_m \,|\, \boldsymbol{z})} \frac{j(m \,|\, m+1)}{j(m+1 \,|\, m)} \frac{g_{m+1,m}(\boldsymbol{\psi}' \,|\, \boldsymbol{\theta}'_{m+1})}{g_{m,m+1}(\boldsymbol{\psi} \,|\, \boldsymbol{\theta}_m)} \left| \frac{\partial h_{m,m+1}(\boldsymbol{\theta}_m, \boldsymbol{\psi})}{\partial(\boldsymbol{\theta}_m, \boldsymbol{\psi})} \right| \tag{4.34}$$

For convenience, we'll label each j, g, and h as belonging to a birth step ($m \to m+1$) or a death step ($m+1 \to m$). For a birth step, the acceptance ratio is:

$$\alpha = \frac{p(m+1, \boldsymbol{\theta}'_{m+1} \,|\, \boldsymbol{z})}{p(m, \boldsymbol{\theta}_m \,|\, \boldsymbol{z})} \frac{j_{\mathrm{d}}(m+1)}{j_{\mathrm{b}}(m)} \frac{g_{\mathrm{d}}(\boldsymbol{\psi}' \,|\, \boldsymbol{\theta}'_{m+1})}{g_{\mathrm{b}}(\boldsymbol{\psi} \,|\, \boldsymbol{\theta}_m)} \left| \frac{\partial h_{\mathrm{b}}(\boldsymbol{\theta}_m, \boldsymbol{\psi})}{\partial(\boldsymbol{\theta}_m, \boldsymbol{\psi})} \right| \tag{4.35}$$

First we consider the birth step, where we have current parameter vector $\boldsymbol{\theta}_m$ that has m sources. The probability of the birth jump from model M_k is

$$j_{\mathrm{b}}(k) = 0.25 \cdot \mathbb{1}(k < m_{\max})$$

and the probability of the death jump from model M_k is

$$j_{\mathrm{d}}(k) = 0.25 \cdot \mathbb{1}(k > 0).$$

In a birth step we sample an auxiliary parameter vector $\boldsymbol{\psi} = [\, \boldsymbol{r}_x, r_w, r_s \,]$ from the proposal distribution

$$\begin{aligned} g_{\mathrm{b}}(\boldsymbol{\psi} \,|\, \boldsymbol{\theta}_k) &= \mathcal{U}(\boldsymbol{r}_{x1};\, I_{x1}) \mathcal{U}(\boldsymbol{r}_{x2};\, I_{x2}) \mathcal{U}(r_w;\, 0, R_w) \mathcal{U}(r_s;\, 0, R_s) \\ &= (R_{x1} R_{x2} R_w R_s)^{-1} \end{aligned} \tag{4.36}$$

to form the augmented state vector $\{\boldsymbol{\theta}_k, \boldsymbol{\psi}\}$. In a death step we draw the parameter vector $\boldsymbol{\psi}' = [\, j^* \,]$ from the following distribution, where j^* is the index of the source that dies:

$$g_{\mathrm{d}}(\boldsymbol{\psi}' \,|\, \boldsymbol{\theta}'_{k'}) = (k')^{-1}.$$

For the birth and death steps the deterministic function h is just an assignment operator, concatenating $\boldsymbol{\psi}$ into the old vector, so

$$\left| \frac{\partial h_{\mathrm{b}}(\boldsymbol{\theta}_k, \boldsymbol{\psi})}{\partial(\boldsymbol{\theta}_k, \boldsymbol{\psi})} \right| = \left| \frac{\partial h_{\mathrm{d}}(\boldsymbol{\theta}'_{k'}, \boldsymbol{\psi}')}{\partial(\boldsymbol{\theta}'_{k'}, \boldsymbol{\psi}')} \right| = 1.$$

We can now use these distributions and the generative model defined in Equations 3.25 through 3.28 to compute the Hastings ratios for the birth and death jumps by substituting into Equation 4.33. For a birth jump from $k = m$ to $k' = m + 1$ we have

$$\alpha(\boldsymbol{\theta}'_{m+1} \,|\, \boldsymbol{\theta}_m) = \frac{\mathcal{N}(\boldsymbol{z}; \mathbf{A}'\boldsymbol{s}' + \boldsymbol{b}, \mathbf{J}_\varepsilon^{-1})}{\mathcal{N}(\boldsymbol{z}; \mathbf{A}\boldsymbol{s} + \boldsymbol{b}, \mathbf{J}_\varepsilon^{-1})} \cdot \frac{1}{m+1} \tag{4.37}$$

The acceptance probability for the death jump from $k = m$ to $k' = m - 1$ we have

$$\alpha(\boldsymbol{\theta}'_{m-1} \,|\, \boldsymbol{\theta}_m) = \frac{p(m-1, \boldsymbol{\theta}'_{m-1} \,|\, \boldsymbol{z})}{p(m, \boldsymbol{\theta}_m \,|\, \boldsymbol{z})} \frac{j_\mathrm{b}(m-1)}{j_\mathrm{d}(m)} \frac{g_\mathrm{b}(\boldsymbol{\psi}' \,|\, \boldsymbol{\theta}'_{m-1})}{g_\mathrm{d}(\boldsymbol{\psi} \,|\, \boldsymbol{\theta}_m)} \left| \frac{\partial h_\mathrm{d}(\boldsymbol{\theta}_m, \boldsymbol{\psi})}{\partial(\boldsymbol{\theta}_m, \boldsymbol{\psi})} \right| \tag{4.38}$$

$$= \frac{\mathcal{N}(\boldsymbol{z}; \mathbf{A}'\boldsymbol{s}' + \boldsymbol{b}, \mathbf{J}_\varepsilon^{-1})}{\mathcal{N}(\boldsymbol{z}; \mathbf{A}\boldsymbol{s} + \boldsymbol{b}, \mathbf{J}_\varepsilon^{-1})} \cdot m \tag{4.39}$$

where Equation 4.38 is just the reciprocal of Equation 4.35 up to a change of index in the model order.

Source splitting and merging

Just as in the birth step derivation, we have two operators that are reversible: a split step ($m \to m+1$) or a merge step ($m+1 \to m$). For convenience, we'll label each j, g, and h as belonging to either merge (m) or split (s). For a split step, the Hastings ratio is:

$$\alpha = \frac{p(m+1, \boldsymbol{\theta}'_{m+1} \,|\, \boldsymbol{z})}{p(m, \boldsymbol{\theta}_m \,|\, \boldsymbol{z})} \frac{j_\mathrm{m}(m+1)}{j_\mathrm{s}(m)} \frac{g_\mathrm{m}(\boldsymbol{\psi}' \,|\, \boldsymbol{\theta}'_{m+1})}{g_\mathrm{s}(\boldsymbol{\psi} \,|\, \boldsymbol{\theta}_m)} \left| \frac{\partial h_\mathrm{s}(\boldsymbol{\theta}_m, \boldsymbol{\psi})}{\partial(\boldsymbol{\theta}_m, \boldsymbol{\psi})} \right| \tag{4.40}$$

First we consider the split step, where we have current parameter vector $\boldsymbol{\theta}_k$ that has k sources. The unnormalized probability of the split jump from model M_k is

$$j_\mathrm{s}(k) = 0.25 \cdot \mathbb{1}(k < m_\mathrm{max})$$

and the unnormalized probability of the merge jump from model M_k is

$$j_\mathrm{m}(k) = 0.25 \cdot \mathbb{1}(k > 0).$$

Next we randomly select a source j^* from a uniform distribution over all the sources. We draw a random variable $\boldsymbol{\psi} = \{j^*, \boldsymbol{r}_x, r_w, r_s\}$ from the distribution, where j^* is the source to split and $\boldsymbol{r}_x$, r_w, and r_s are the offsets used in each dimension of the source parameterization.

$$g_\mathrm{s}(\boldsymbol{\psi} \,|\, \boldsymbol{\theta}_k) = \mathcal{U}(j^*; 1, k)\mathcal{U}(\boldsymbol{r}_{x1}; -\tfrac{E_{x1}}{2}, \tfrac{E_{x1}}{2})\mathcal{U}(\boldsymbol{r}_{x2}; -\tfrac{E_{x2}}{2}, \tfrac{E_{x2}}{2})\mathcal{U}(r_w; -\tfrac{E_w}{2}, \tfrac{E_w}{2})\mathcal{U}(r_s; -\tfrac{E_s}{2}, \tfrac{E_s}{2})$$

or, equivalently,

$$g_\mathrm{s}(\boldsymbol{\psi} \,|\, \boldsymbol{\theta}_k) = (k \cdot E_{x1}E_{x2}E_wE_s)^{-1}$$

to form the augmented state vector $\{\boldsymbol{\theta}_k, \boldsymbol{\psi}\}$. In the reverse (merge) step, we randomly select two of the $k+1$ sources to merge. So, our set of auxiliary random variables is $\boldsymbol{\psi}' = \{j_-^*, j_+^*\}$.

$$g_{\mathrm{m}}(\boldsymbol{\psi}' \,|\, \boldsymbol{\theta}'_{k'}) = (k \cdot k')^{-1}$$

The deterministic functions h_{s} and h_{m} take more complicated forms now. $h_{\mathrm{s}}(\boldsymbol{\theta}_k, \boldsymbol{\psi}) = (\boldsymbol{\theta}'_{k'}, \boldsymbol{\psi}')$ represents $\boldsymbol{x}_{j_\pm^*} = \boldsymbol{x}_{j*} \pm \boldsymbol{r}_x$, $w_{j_\pm^*} = w_{j*} \pm r_w$, and $s_{j_\pm^*} = s_{j*} \pm r_s$, where source j^* in the ordering is now replaced with two sources $j_-^* \to j^*$ and $j_+^* \to j^* + 1$. The deterministic function h_{m} gives the average of the j^* and $j^* + 1$ sources as the new parameters. [17] suggests that

$$\left| \frac{\partial h_{\mathrm{s}}(\boldsymbol{\theta}_k, \boldsymbol{\psi})}{\partial(\boldsymbol{\theta}_k, \boldsymbol{\psi})} \right| = 8 \quad \text{and} \quad \left| \frac{\partial h_{\mathrm{m}}(\boldsymbol{\theta}'_{k'}, \boldsymbol{\psi}')}{\partial(\boldsymbol{\theta}'_{k'}, \boldsymbol{\psi}')} \right| = \frac{1}{8}.$$

We will now derive an expression for the split move acceptance probability in terms of the components of $\boldsymbol{\theta}_k$ for a model where $k = m$ and $k' = m + 1$.

$$\begin{aligned} \alpha(\boldsymbol{\theta}'_{m+1} \,|\, \boldsymbol{\theta}_m) &= \frac{p(m+1, \boldsymbol{\theta}'_{m+1} \,|\, \boldsymbol{z})}{p(m, \boldsymbol{\theta}_m \,|\, \boldsymbol{z})} \cdot 1 \cdot \frac{(m(m+1))^{-1}}{(m \cdot E_{x1} E_{x2} E_w E_s)^{-1}} \cdot 8 \\ &= \frac{\mathcal{N}(\boldsymbol{z}; \mathbf{A}'\boldsymbol{s}' + \boldsymbol{b}, \mathbf{J}_\varepsilon^{-1})}{\mathcal{N}(\boldsymbol{z}; \mathbf{A}\boldsymbol{s} + \boldsymbol{b}, \mathbf{J}_\varepsilon^{-1})} \cdot \frac{E_{x1} E_{x2} E_w E_s}{R_{x1} R_{x2} R_w R_s} \cdot \frac{1}{m+1} \cdot 8 \end{aligned} \tag{4.41}$$

The acceptance probability for a merge step from $k = m$ to $k' = m - 1$ is the reciprocal of the probability for a split step, up to some shift in the model order.

$$\alpha(\boldsymbol{\theta}'_{m-1} \,|\, \boldsymbol{\theta}_m) = \frac{\mathcal{N}(\boldsymbol{z}; \mathbf{A}'\boldsymbol{s}' + \boldsymbol{b}, \mathbf{J}_\varepsilon^{-1})}{\mathcal{N}(\boldsymbol{z}; \mathbf{A}\boldsymbol{s} + \boldsymbol{b}, \mathbf{J}_\varepsilon^{-1})} \cdot \frac{R_{x1} R_{x2} R_w R_s}{E_{x1} E_{x2} E_w E_s} \cdot m \cdot \frac{1}{8} \tag{4.42}$$

Chapter 5

Performance Metrics

In this chapter we establish several methods of examining the output of the inference algorithm described in Chapter 4. We restrict this discussion to summaries of the results from a single run of the algorithm and a single true configuration of sources. This "true" configuration may or may not be used in each of the summaries we will discuss. We will begin by examining commonly-sought statistics in general inference problems. We will then introduce several additional metrics that target specific features of the source estimation problem. All of the summaries discussed in this chapter can be easily and efficiently computed from the collection of samples drawn from the posterior.

5.1 General statistics of interest

Two of the most fundamental Bayesian inference tasks are MAP estimation and the acquisition of posterior distributions for the parameters. These can be computed exactly for probabilistic models of certain types, i.e. those with tree-structured graphical models. The model developed in Chapter 3 does not have a structure suitable for exact inference; instead, we use MCMC techniques to obtain samples from the posterior distribution. From this set of samples we can compute nearly any statistic of the posterior, including the common summaries mentioned above as well as far more complex event probabilities.

5.1.1 MAP and MLE estimation

The *maximum a posteriori* (MAP) estimator for some random variable $\boldsymbol{\theta}$ with posterior distribution $p(\boldsymbol{\theta} \,|\, \boldsymbol{z})$ is the value of $\boldsymbol{\theta}$ given by

$$\boldsymbol{\theta}_{\text{MAP}} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta} \,|\, \boldsymbol{z}). \tag{5.1}$$

This provides a "point estimate", in that it represents a single point for which the posterior distribution has the maximum value. A point estimate can be very useful if the posterior distribution is unimodal and in a low-dimensional space. Even for unimodal distributions, in high-dimensional spaces the probability can change very rapidly very close to the MAP estimate. Of course, if the our posterior distribution is multimodal

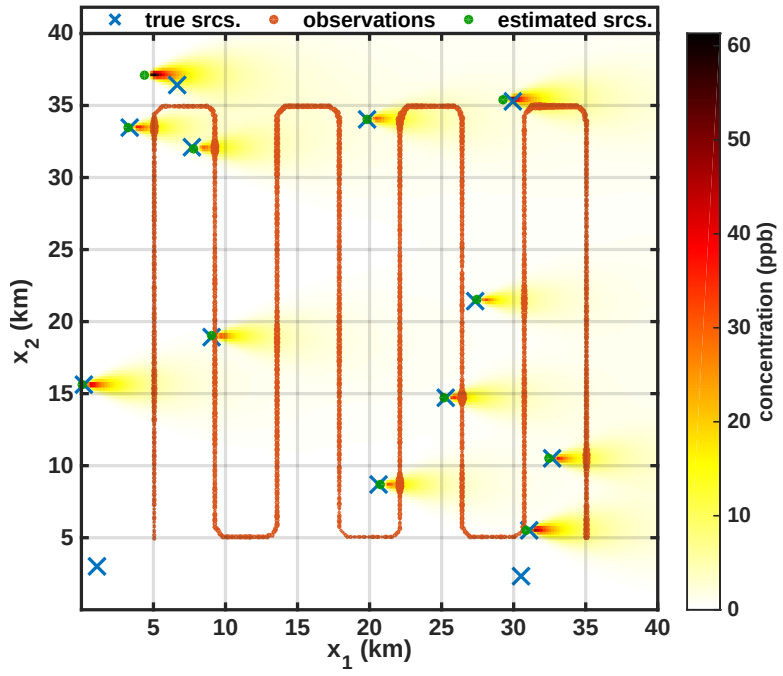

Figure 5.1: An approximate MAP estimate for the source configuration indicated by the blue x's. Ground truth sources were randomly placed and observations were generated according to the forward model.

then that MAP estimate only reveals one mode of the distribution by nature of being a single point, which makes it even more unreliable.

We can easily compute a quantity similar to the MAP estimate from the set of samples from the posterior. When using MCMC, the MAP estimate is computed as follows. We recall that the posterior distribution is proportional to the joint distribution, which we can compute directly. Hence, for the sample-based MAP estimate, we can simply take the sample $\boldsymbol{\theta}^{(k)}$ that maximizes the joint distribution instead of optimizing the posterior distribution over all possible values of $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}}_{\text{MAP}} = \arg\max_k p\Big(\boldsymbol{\theta}^{(k)} \,|\, \boldsymbol{z}\Big) = \arg\max_k p\Big(\boldsymbol{z} \,|\, \boldsymbol{\theta}^{(k)}\Big) p\Big(\boldsymbol{\theta}^{(k)}\Big). \tag{5.2}$$

Figure 5.1 is an example of the MAP estimate obtained by sampling from the posterior distribution. The forward model was used to randomly generate the true configuration and observed concentrations for a set aircraft trajectory. The underlying image shows the expected concentration at each pixel location as predicted by the Gaussian plume model. Nearly all of the sources are very closely estimated by a sampled source in the MAP solution. We can see that there seems to be some uncertainty about source location in the downwind direction—some sources have an estimate upwind of

the actual location and others have one downwind. The two southern-most true sources are not estimated at all, although considering the wind direction (due east) they likely have very little supporting evidence within the data.

Another common point estimate is the *maximum likelihood* (ML) estimate. Instead of maximizing the posterior distribution, the ML estimate maximizes the likelihood distribution $p(\boldsymbol{z} \mid \boldsymbol{\theta})$. This is useful in cases where we do not want our prior distribution $p(\boldsymbol{\theta})$ to influence the estimate. This estimate may be computed from the samples from the posterior in a similar manner as the MAP estimate.

■ 5.1.2 Event probabilities and marginal posterior distributions

One limitation of the point estimates described in the previous section is that they do not convey any information of how broad or narrow the posterior distribution is around the mode. Sampling-based approaches allow us to estimate not just this, but also to accurately compute expectations of *any* functions of the distribution. The function can be selected in such a way that reveals nearly any property of interest, including the commonly-sought posterior distributions of our model parameters. Note that we refer to the collection of marginal posterior distributions over each random variable as our "posteriors." The joint posterior distribution of all the variables is just called the "posterior." When we refer to the posterior distributions of several variables jointly, we will be explicit.

A numerical technique called Monte-Carlo integration is central to our ability to compute expectations of functions of the posterior distribution. Monte-Carlo integration refers to approximately computing some integral

$$\mathbb{E}[f(x)] = \int_x f(x)p(x)\,\mathrm{d}x$$

where $f(x)$ is an arbitrary function and $p(x)$ is the distribution on x. It is assumed that we can draw samples from $p(x)$. The Monte Carlo approximation for the above integral is

$$\mathbb{E}[f(x)] \approx \frac{1}{K}\sum_{k=1}^{K} f\left(x^{(k)}\right) \tag{5.3}$$

where we have drawn K i.i.d. samples $x^{(k)} \sim p(x)$.

In our case we have already drawn the samples from our posterior distribution $p(\boldsymbol{\theta} \mid \boldsymbol{z})$. Monte Carlo integration gives us a way of computing the expectation of any arbitrary function $f(\boldsymbol{\theta})$ of the posterior $p(\boldsymbol{\theta} \mid \boldsymbol{z})$:

$$\mathbb{E}_{\boldsymbol{\theta}\mid\boldsymbol{z}}[f(\boldsymbol{\theta}) \mid \boldsymbol{z}] = \int_{\boldsymbol{\theta}} f(\boldsymbol{\theta})p(\boldsymbol{\theta} \mid \boldsymbol{z})\,\mathrm{d}\boldsymbol{\theta} \approx \frac{1}{K}\sum_{k=1}^{K} f\left(\boldsymbol{\theta}^{(k)}\right). \tag{5.4}$$

In the context of probability, an event is some region of the state space of our variable $\boldsymbol{\theta}$. For some arbitrary event A we can easily specify a function that, in expectation,

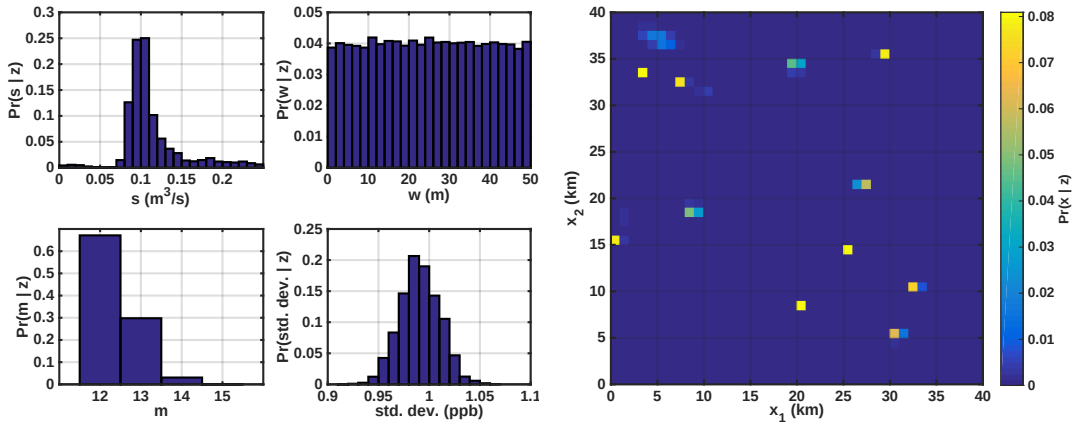

Figure 5.2: Posterior distributions approximated from the set of posterior samples. The same simulation was used to compute the MAP estimate shown in Figure 5.1.

gives the posterior *event probability* given the data:

$$f_A(\boldsymbol{\theta}) = \mathbb{1}(\boldsymbol{\theta} \in A). \tag{5.5}$$

The posterior marginal distributions are special cases of these event probabilities. Consider, for example, the posterior distribution over the number of sources m given the observed data $\boldsymbol{z}$. We denote the number of sources defined in the k-th sample of the posterior $\boldsymbol{\theta}^{(k)}$ by $m^{(k)}$. The posterior distribution $p(m \,|\, \boldsymbol{z})$ is estimated by computing

$$\Pr\{m = \mu \,|\, \boldsymbol{z}\} \approx \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}\left(m^{(k)} = \mu\right) \tag{5.6}$$

for each possible number of sources $\mu = 0, \ldots, m_{\max}$. The posteriors are similarly defined for other parameters, although posteriors for continuous parameters are specified in terms of ranges and not exact values.

Figure 5.2 shows the posterior distributions for the source parameters, generated from the same set of samples as the MAP estimate shown in Figure 5.1. These statistics provide several other insights that were missing from the MAP estimate. We can identify that the most likely number of sources is 12 by the marginal distribution $\Pr(m \,|\, \boldsymbol{z})$, recognize the source emission rates as being clustered around 0.1 m^3/s, and that the poorly-estimated source in the upper left corner of Figure 5.1 has great uncertainty in its position judging by the diffusivity of the corresponding location in the joint posterior distribution over x_1 and x_2.

■ 5.2 Source-based event probabilities

This section introduces several other event probabilities that are specially-tailored to our problem. We find that event probabilities provide a natural way of evaluating

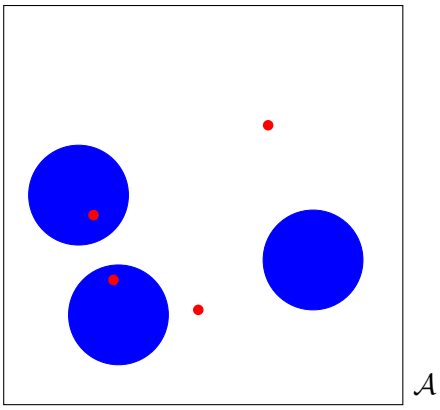

Figure 5.3: A visualization of the sources contained in the true configuration $\boldsymbol{\theta}^*$, shown as blue x's, and the sources from a single sample of the posterior, shown as red dots. The light-blue disk around each true source shows the detection region $\mathcal{D}_i^R$ for that source.

the performance of our system. While posterior distributions provide great insight into the results of inference, they also require some level of interpretation and manual examination. These event probabilities can be efficiently computed from the samples of the posterior using Equations 5.4 and 5.5 and often provide some immediate insights.

Our source estimation problem is, at a high level, an estimation problem where we have some true set of point sources $(\boldsymbol{x}, \boldsymbol{y})$ and we generate many "sampled" configurations of point sources $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$. A common concern is identifying how well our true set of point sources match our sampled sets of point sources.

Several factors are considered in designing our metrics. The first is that we do not *necessarily* know which estimated source corresponds to which true source, in fact, we have no guarantees that there will be a one-to-one mapping. Any metric that we devise should be efficient; hence, we would like to avoid solving the "assignment" problem or making any (hard or soft) assignment of estimated sources to true sources. Additionally, we want to capture two specific types of undesirable results that correspond to type I and type II errors. Type I errors occur when any true source is not matched by any sampled sources. Similarly, type II error refer to sampling spurious sources that are not matched by any true sources. Sampled sources that do not match true sources are not relevant results in that they do not tell us any further information about the true sources. Hence, we wish to maximize the relevance of sampled sources to minimize type II error. In light of these we will thus avoid referring to matches or mappings between true and sampled sources. Instead, we refer to true sources as having been "estimated" and sampled sources as being "relevant."

In this context we represent all sources, both true and sampled, as point sources. Consequently, we will never have one source directly atop another. This motivates our particular definition of "estimations" to be in terms of the distance between the true

and sampled source. Consider the set of sources in the true configuration $\boldsymbol{\theta}^*$, shown for some arbitrary example in Figure 5.3 as the set of blue x's. We also consider the set of sampled source locations from the k-th sample from the posterior, $\boldsymbol{\theta}^{(k)}$, which appear as the (arbitrarily placed) red dots in Figure 5.3.

Definition 5.2.1 (Detection region)**.** *For each of the m^* true sources, the* detection region of source i, $\mathcal{D}_i^R$, *is a disk of radius R centered at the source location $\boldsymbol{x}_i^*$:*

$$\mathcal{D}_i^R \triangleq \{\boldsymbol{x} \,:\, \|\boldsymbol{x} - \boldsymbol{x}_i^*\|_2 \leq R\}. \tag{5.7}$$

The (complete) detection region *for the true source configuration is the union of the detection regions for each source:*

$$\mathcal{D}^R \triangleq \bigcup_{i=1}^{m^*} \mathcal{D}_i^R.$$

Any sampled source that falls within the composite region $\mathcal{D}^R$ is an "estimate" of at least one true source (as the disks may overlap for closely-spaced sources). The radius R quantifies the notion of "how close is close enough" when estimating sources. It enables a physical interpretation where if we are standing at one of the sampled source locations and we are within R meters of a true source location then we are "close enough" to the true source.

In Figure 5.3 the individual light-blue disks show detection region $\mathcal{D}_i^R$ for some radius R. As pictured, two of the three true sources have at least one red dot in their detection regions. In other words, the proportion of our true sources that are "estimated" is $2/3$. Conversely, only two of the four sampled sources (red dots) fall into the light-blue detection region $\mathcal{D}^R$. We interpret this as only two of the four estimates actually being "relevant" to our set of true sources. We can then say that the proportion of sampled sources that are "relevant" is $2/4$. These proportions convey some useful information about the performance of our algorithm, which acts as a "source detector". We formally define these metrics below.

First we design an event that conveys how well we "match" the location for a particular true source $i \in \{1, \ldots, m^*\}$. For some sample from the posterior $\hat{\boldsymbol{\theta}}$ we define the following event.

Definition 5.2.2 (Event: *estimation* of a true source)**.** *The i-th true source is* estimated to within R *if and only if there exists a sampled source in $\hat{\boldsymbol{\theta}} \sim p(\boldsymbol{\theta} \,|\, \boldsymbol{z})$ that is within a distance R of the location of the i-th source. Formally, this event is given by the indicator function*

$$f_{EST_i}(\hat{\boldsymbol{\theta}}; R) \triangleq \mathbb{1}(\exists\, j \in \{1, \ldots, \hat{m}\} \text{ s.t. } \hat{\boldsymbol{x}}_j \in \mathcal{D}_i^R) \tag{5.8}$$

where $\hat{\boldsymbol{x}}_j$ and $\hat{m}$ are the location of the j-th sampled source in $\hat{\boldsymbol{\theta}}$ and the number of sampled sources in $\hat{\boldsymbol{\theta}}$, respectively.

Given a set of samples from the posterior we can apply Equations 5.4 and 5.5 to compute the *probability of estimation* for the i-th true source:

$$\Pr\{\text{EST}_i(R)\} = \mathbb{E}_{\boldsymbol{\theta}\,|\,\boldsymbol{z}}[f_{\text{EST}_i}(\boldsymbol{\theta}; R)\,|\,\boldsymbol{z}] \approx \frac{1}{K}\sum_{k=1}^{K} f_{\text{EST}_i}\left(\boldsymbol{\theta}^{(k)}; R\right). \tag{5.9}$$

Additionally, although we will not do so, one could also easily compute the *average* probability of estimation by averaging over the estimation probabilities for all m^* of the true sources.

The probability of estimation in Equation 5.9 is parameterized by the radius R of the detection region; by varying R we can produce a curve that shows how the estimation probability changes as we relax our detection threshold. These curves will feature prominently in our system-level analysis of the inference procedure. Specifically, $\Pr\{\text{EST}_i(R)\} = 0$ for $R = 0$, since the probability of two points overlapping is zero. As the radius $R \to \infty$ we get $\Pr\{\text{EST}_i(R)\} \to 1$ as eventually *any* sampled source in the survey area falls within the detection region of true source i. This limiting behavior assumes that we have at least one sampled source in the survey area, which we have found to be the case even though it may not correspond to the true source (if there is one).

The above event targeted the first type of undesirable sampler behavior: failing to estimate true sources. The second type of undesirable behavior, corresponding to type II error, is generating spurious sources that cannot be considered "estimates" of any true sources. For some sample from the posterior $\hat{\boldsymbol{\theta}}$ we define the following event.

Definition 5.2.3 (Event: *relevance* of a sampled source)**.** *The j-th sampled source in $\hat{\boldsymbol{\theta}}$ is* relevant to within R *if and only if there exists a true source in $\boldsymbol{\theta}^*$ that is within a distance R of the location of the j-th sampled source. Formally, this event is given by the indicator function*

$$f_{REL_j}(\hat{\boldsymbol{\theta}}; R) \triangleq \mathbb{1}(\exists\, i \in \{1, \ldots, m^*\}\ s.t.\ \hat{\boldsymbol{x}}_j \in \mathcal{D}_i^R) \qquad j = 1, \ldots, \hat{m} \tag{5.10}$$

where $\hat{\boldsymbol{x}}_j$ and $\hat{m}$ are the location of the j-th sampled source in $\hat{\boldsymbol{\theta}}$ and the number of sampled sources in $\hat{\boldsymbol{\theta}}$, respectively.

Since the indices, and even the number, of sampled sources are not consistent across different samples from the posterior, it is meaningless to consider the probability of relevance for the j-th sampled source. Instead, we can compute the *probability of relevance* for some sample from the posterior $\hat{\boldsymbol{\theta}}$:

$$f_{\text{REL}}(\hat{\boldsymbol{\theta}}; R) \triangleq \frac{1}{\hat{m}}\sum_{j=1}^{\hat{m}} f_{\text{REL}_j}(\hat{\boldsymbol{\theta}}; R) \tag{5.11}$$

This gives the percentage of sampled sources that are within R meters of at least one true source, or the probability that a randomly-chosen sampled source falls in the

detection region. Given a set of samples from the posterior we can apply Equations 5.4 and 5.5 to compute the *average* probability of relevance:

$$\Pr\{\mathrm{REL}(R)\} = \mathbb{E}_{\boldsymbol{\theta}\,|\,\boldsymbol{z}}[f_{\mathrm{REL}}(\boldsymbol{\theta}; R)\,|\,\boldsymbol{z}] \approx \frac{1}{K}\sum_{k=1}^{K} f_{\mathrm{REL}}\left(\boldsymbol{\theta}^{(k)}; R\right). \tag{5.12}$$

Sampling of spurious sources leads to values of $\Pr\{\mathrm{REL}(R)\}$ that are lower than 1.

As discussed above for the probability of estimation, we can compute a curve for the relevance probability by varying the detection region radius R. This has the same limiting behavior as $R \to 0$ and $R \to \infty$ as the other event. In both cases higher event probabilities indicate better system performance–given only the locations, not the emission rates or the widths of the sources, the best possible performance would have every true source estimated and every sampled source relevant. We also note that when the number of true sources and the number of sampled sources are both 1, the curves traced by $\Pr\{\mathrm{EST}(R)\}$ and $\Pr\{\mathrm{REL}(R)\}$ are identical.

■ 5.2.1 ROC-style analysis

Although the probability of relevance does provide some useful insights about sources where we have no sources *locally*, it does not truly provide us a way of examining type II error, or instances in which we sample a source where no source actually exists anywhere in the survey area. Instead, we could simulate a new measurement vector $\boldsymbol{z}'$ by setting the emission rate of the i-th true source to zero and generate samples from the posterior $p(\boldsymbol{\theta}\,|\,\boldsymbol{z}')$. We could then compute the probability of estimation for this new posterior distribution. We would know with certainty that all of the sampled sources from the new posterior are *not* due to the i-th true source, which is explicitly blocked from contributing to any of the measurements in $\boldsymbol{z}'$. If the number of true sources is 1 then this only requires one additional simulation and the analysis is straight-forward. For simplicity, and to ensure easily-interpreted results, we will only consider the case where $m = 1$ in this thesis.

In the simply-defined and easily-analyzed case where $m = 1$, we will run a second inference procedure where the data is using a source with emission rate zero and all other parameters and realizations are identical to the original. From the two resulting chains of samples we can then compute the estimation probabilities $\Pr\{\mathrm{EST}(R)\,|\,s > 0\}$ from the original chain of samples and $\Pr\{\mathrm{EST}(R)\,|\,s = 0\}$, where s is the emission rate of the single existing source in each run. This immediately suggests parallels between this approach and the well-studied problems of Bayesian hypothesis testing for signal detection.

Consider the problem of attempting to classify a received signal into one of two groups: $H = 0$, the null-hypothesis, for which the signal is just i.i.d. white Gaussian noise, or $H = 1$, for which the received signal is the signal we want to detect plus i.i.d. white Gaussian noise. The detector decides $\hat{H} = 1$ or $\hat{H} = 0$ based on some set of costs, which implicitly define a threshold τ. Bayesian hypothesis testing is evaluated based

on two quantities: the *probability of detection* $P_{\mathrm{D}} = \Pr\{\hat{H} = 1 \,|\, H = 1\}$, also called the *true-positive rate*, and the *probability of false alarm* $P_{\mathrm{FA}} = \Pr\{\hat{H} = 1 \,|\, H = 0\}$, also called the *false-positive rate*. The probability of detection is very similar to our original estimation probability $\Pr\{\mathrm{EST}(R) \,|\, s > 0\}$. Similarly, the probability of false alarm is closely related to the quantity $\Pr\{\mathrm{EST}(R) \,|\, s = 0\}$.

In Bayesian hypothesis testing a "receiver operating characteristic" (ROC) curve is generated by plotting the probability of detection versus the probability of false alarm and allowing varying the decision threshold τ from 0 to ∞. These curves go from the point $(0, 0)$ for $\tau = 0$ to $(1, 1)$ for $\tau \to \infty$. We can similarly plot $\Pr\{\mathrm{EST}(R) \,|\, s > 0\}$ versus $\Pr\{\mathrm{EST}(R) \,|\, s = 0\}$ by varying the radius R from 0 to ∞ and obtain similarly-shaped curves. ROC curves in hypothesis testing provide several useful insights that we attempt to use in our analogous case. First, all valid ROC curves are above the line of guessing, $P_{\mathrm{D}} = P_{\mathrm{FA}}$. We empirically observe that this holds true. Second, all ROC curves are concave down. In general, the curves that we obtain with our approximation exhibit this property. Notable exceptions are edge artifacts encountered from having a bounded survey area and lack of samples. Finally, if one ROC curve is completely above another ROC curve, it is said that the upper curve "dominates" the lower one, meaning that the sensor corresponding to the dominant curve is an all-around better sensor. In our case, we can vary our "emission source sensor" by changing the parameters of the sensor or of the true parameters used to generate the data. By analogy, then, we can say: that our inference procedure performs better under certain conditions than others; identify what these conditions are; and quantify the impact on system performance. As a side note, if one ROC curve is above another curve for some range of τ, it is said that that curve dominates the other for the set of cost ratios corresponding to that range of τ.

■ 5.3 Grid-based event probabilities

The previously-described event probabilities treated all of the sources as point sources, i.e. as if they had width $w = 0$. Given that the size of the sources is very small relative to the size of the survey area, this is not a terrible assumption; regardless, it still leaves some aspects of our model parameters unexamined. By considering event probabilities defined on a grid over the survey area we can easily take the source width and even the emission rate into account.

For both of the event probability definitions that follow, we use a discretized representation of the survey area $\mathcal{A}$ into a grid of cells as shown in Figure 5.4. In the figure we have chosen a grid with eight rows by eight columns; we can generally define a grid to have $R \times C$ cells. The East/West dimension of the survey area is divided evenly into $x_0, \ldots, x_C$ and the North/South dimension into $y_0, \ldots, y_R$. We denote grid cell (r, c) as the set of points in the 2D-interval given by $[x_{c-1}, x_c] \times [y_{r-1}, y_r]$. For example, Figure 5.4 shows the cell $(3, 5)$ highlighted in orange. We will generally wish to compute the probability of some event over each of the $R \times C$ grid cells. We now describe two

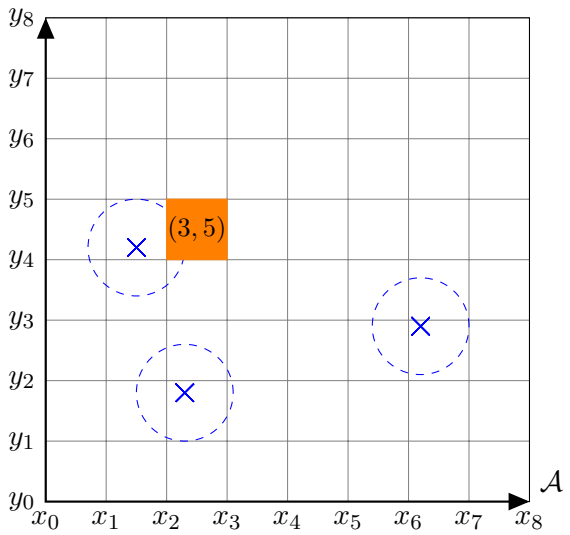

Figure 5.4: An 8×8 grid of cells fit to the survey area $\mathcal{A}$. The blue x's show several generic source locations and the blue dashed lines are the 1σ-contours for the corresponding multivariate normal distributions defined by the source locations and widths. We can define event probabilities for each cell in the grid, e.g., we may wish to compute the probability of some event occurring within the highlighted cell $(3, 5)$.

such events and the intuitions they provide.

■ 5.3.1 Source location histogram

In this method we visualize the 2D distribution of emission sources for some source configuration $\boldsymbol{\theta}$, which may be the true $\boldsymbol{\theta}^*$ or some sample $\boldsymbol{\theta}^{(k)}$ from the posterior distribution. Recall that the emission location for each of the diffuse sources $j = 1, \ldots, m$ in $\boldsymbol{\theta}$ is a multivariate Gaussian distribution with location $\boldsymbol{x}_j$ and an isotropic covariance matrix $w_j^2\, \mathbf{I}^{2\times 2}$. Hence

$$p_j(\boldsymbol{x}) = \mathcal{N}\big(\boldsymbol{x}; \boldsymbol{x}_j, w_j^2\, \mathbf{I}^{2\times 2}\big) \tag{5.13}$$

gives the pdf of emission locations for source j. Note that we can interchangeably write $\boldsymbol{x} = (x, y) = (x_1, x_2)$. We now define the following event:

Definition 5.3.1 (Event: source j emits from cell (r, c))**.** *The j-th sampled source in $\boldsymbol{\theta}$ emits within the cell (r, c) with probability given by the integral of the pdf $p_j(\boldsymbol{x})$ over*

the region $[x_{c-1}, x_c] \times [y_{r-1}, y_r]$. *We can efficiently compute this event probability as:*

$$\begin{aligned} P_j(r,c) &= \int_{y_{r-1}}^{y_r} \int_{x_{c-1}}^{x_c} p_j(\boldsymbol{x})\, dx\, dy \\ &= \int_{y_{r-1}}^{y_r} \int_{x_{c-1}}^{x_c} \mathcal{N}\left(\begin{pmatrix} x \\ y \end{pmatrix}; \begin{pmatrix} x_j \\ y_j \end{pmatrix}, \begin{pmatrix} w_j^2 & 0 \\ 0 & w_j^2 \end{pmatrix} \right) dx\, dy \\ &= \left(\int_{x_{c-1}}^{x_c} \mathcal{N}\big(x; x_j, w_j^2\big)\, dy \right) \left(\int_{y_{r-1}}^{y_r} \mathcal{N}\big(y; y_j, w_j^2\big)\, dy \right). \end{aligned} \tag{5.14}$$

The matrix $\mathbf{P}_j$ *is referred to as the* probability map *for source* j.

A natural extension of this involves computing the probability of *any* source emitting in the grid cell (r, c). This defines the *average* probability map, which we can easily compute by averaging over the probability maps for each source:

$$P(r,c) = \frac{1}{m} \sum_{j=1}^{m} P_j(r,c), \tag{5.15}$$

where we have assumed that an emission is equally likely to be generated by any of the sources. This effectively provides us a distribution over the possible emission locations assuming that all sources are equally likely.

We note that the true probability maps $\mathbf{P}^*$ are immediately accessible given the true configuration $\boldsymbol{\theta}^*$, which defines the m^* probability maps $\mathbf{P}_1^*, \ldots, \mathbf{P}_{m^*}^*$. We can also compute the expected probability map with respect to the posterior distribution $p(\boldsymbol{\theta} \,|\, \boldsymbol{z})$ by averaging over the K average probability maps generated by the sampled source configurations $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(K)}$. We can easily visualize this "sampled" probability map for comparison to the true probability map.

■ 5.3.2 Expected emission rate maps

The above probability-map based approach did not consider the emission rates of the sources. For example, it would be possible to have a very weakly emitting source with low width that appears very prominently on the probability map for source locations. In practice this would be a nuisance and we would far prefer to highlight the strongly-emitting sources. We encapsulate this notion by formulating a rate map $\mathbf{Q}$, in which each cell contains the expected emissions in the cell.

The above formulation combined the probability maps by weighting all sources equally. From a physical standpoint, we can say that the probability map gives the expected emission location for a particle emitted by a uniformly-chosen emission source. In reality, we would expect the sources with a higher emission rate to emit much more frequently. We use this intuition when combining the probability maps from each source:

$$Q(r,c) = \sum_{j=1}^{m} s_j P_j(r,c). \tag{5.16}$$

This is equivalent to replacing the pdf in Equation 5.13 by

$$q_j(\boldsymbol{x}) = s_j \mathcal{N}\big(\boldsymbol{x}; \boldsymbol{x}_j, w_j^2\, \mathbf{I}^{2\times 2}\big) \tag{5.17}$$

and integrating over the cells as described in the above section. We note that the function $q_j(\boldsymbol{x})$ is *not* a probability density function; instead, it gives the expected level of "emissions" arising in some area when integrated over the area. This leads to

$$\iint_{\mathcal{A}} q_j(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = s_j \tag{5.18}$$

which says that the sum of all emissions from source j over the entire survey area equal the emission strength s_j. As with the probability maps, we can easily compute and visualize these results for true configurations, any of the samples drawn from the posterior, and in expectation over the posterior distribution.

■ 5.4 Evaluating system performance

While the Monte-Carlo approximation discussed in Section 5.1.2 provides a tractable method for computing event probabilities given a single measurement realization $\boldsymbol{z}$, it does not inherently provide a way of performing high-level system analysis. In order to obtain these higher-level summaries we must *marginalize over* some of the parameters of our model that are used to generate the measurement vector $\boldsymbol{z}$. Luckily, Monte-Carlo approximation can also be applied here to make such marginalization tractable. In this section we outline the technique used.

Much of our previous discussion has dealt with the parameter vector $\boldsymbol{\theta}$, which contains information about the sources and the background measurements. We also have a set of auxiliary parameters $\boldsymbol{\phi}$, which includes information about the wind field, the measurement error standard deviation σ_ε, the measurement times and locations, and any other information. The union $\boldsymbol{\theta} \cup \boldsymbol{\phi}$ defines our full parameter set and completely specifies the distribution of the observations, $p(\boldsymbol{z} \,|\, \boldsymbol{\theta}, \boldsymbol{\phi})$. Now consider a different partition of our full parameter set into a set $\boldsymbol{\omega}$ of control variables—those that we will change and examine system performance with respect to—and a set $\boldsymbol{\nu}$ of nuisance variables that we do not care about and wish to marginalize over. For instance, suppose we are interested in the ability of our system to estimate a source with a given emission rate. The emission rate would be in the set of control parameters $\boldsymbol{\omega}$; the path, background realizations, and source widths would be in the set of nuisance variables $\boldsymbol{\nu}$.

In the previous sections we were interested in quantifying system performance through the expectation of some function $f(\boldsymbol{\theta})$ with respect to the distribution $p(\boldsymbol{\theta} \,|\, \boldsymbol{z})$. Here, instead, we are interested in the expectation of $f(\cdot)$ with respect to the distribution $p(\boldsymbol{\theta} \,|\, \boldsymbol{\omega})$. We can write this distribution in several terms of several others:

$$\begin{aligned} p(\boldsymbol{\theta} \,|\, \boldsymbol{\omega}) &= \int_{\boldsymbol{\nu}} \int_{\boldsymbol{z}} p(\boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{\nu} \,|\, \boldsymbol{\omega})\, \mathrm{d}\boldsymbol{z}\, \mathrm{d}\boldsymbol{\nu} \\ &= \int_{\boldsymbol{\nu}} \int_{\boldsymbol{z}} p(\boldsymbol{\theta} \,|\, \boldsymbol{z}) p(\boldsymbol{z} \,|\, \boldsymbol{\nu}, \boldsymbol{\omega}) p(\boldsymbol{\nu})\, \mathrm{d}\boldsymbol{z}\, \mathrm{d}\boldsymbol{\nu} \end{aligned}$$

where we know that $p(\boldsymbol{z} \mid \boldsymbol{\nu}, \boldsymbol{\omega})$ is Gaussian with mean given by a deterministic function of $\{\boldsymbol{\nu}, \boldsymbol{\omega}\}$ and a variance given by σ_ε, which is an element of either $\boldsymbol{\nu}$ or $\boldsymbol{\omega}$. We can assume a uniform prior distribution $p(\boldsymbol{\nu})$. We now write the expectation of some function $f(\boldsymbol{\theta})$ with respect to $p(\boldsymbol{\theta} \mid \boldsymbol{\omega})$ as a series of nested integrals, which will allow us to repeatedly apply Monte Carlo integration. Our desired quantity takes the form

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta} \mid \boldsymbol{\omega}}[f(\boldsymbol{\theta})] &= \int_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \boldsymbol{\omega}) \, \mathrm{d}\boldsymbol{\theta} \\
&= \int_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) \left(\int_{\boldsymbol{\nu}} \int_{\boldsymbol{z}} p(\boldsymbol{\theta} \mid \boldsymbol{z}) p(\boldsymbol{z} \mid \boldsymbol{\nu}, \boldsymbol{\omega}) p(\boldsymbol{\nu}) \, \mathrm{d}\boldsymbol{z} \, \mathrm{d}\boldsymbol{\nu} \right) \mathrm{d}\boldsymbol{\theta} \\
&= \int_{\boldsymbol{\nu}} \left(\int_{\boldsymbol{z}} \left(\int_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \boldsymbol{z}) \, \mathrm{d}\boldsymbol{\theta} \right) p(\boldsymbol{z} \mid \boldsymbol{\nu}, \boldsymbol{\omega}) \, \mathrm{d}\boldsymbol{z} \right) p(\boldsymbol{\nu}) \, \mathrm{d}\boldsymbol{\nu}
\end{aligned} \tag{5.19}$$

where we have really just applied the *law of iterated expectations* several times:

$$\mathbb{E}_{\boldsymbol{\theta} \mid \boldsymbol{\omega}}[f(\boldsymbol{\theta})] = \mathbb{E}_{\boldsymbol{\nu}}\left[\mathbb{E}_{\boldsymbol{z} \mid \boldsymbol{\nu}, \boldsymbol{\omega}}\left[\mathbb{E}_{\boldsymbol{\theta} \mid \boldsymbol{z}}[f(\boldsymbol{\theta})]\right]\right]. \tag{5.20}$$

The form of Equation 5.19 suggests the following sampling-based approximation for computing the desired expectation based on the Monte Carlo approximation for each integral.

1. Sample N_ν nuisance parameter values $\boldsymbol{\nu}^{(1)}, \ldots, \boldsymbol{\nu}^{(N_\nu)} \sim p(\boldsymbol{\nu})$
2. For each $\boldsymbol{\nu}^{(i)}$: sample N_z measurement vectors $\boldsymbol{z}_{(i)}^{(1)}, \ldots, \boldsymbol{z}_{(i)}^{(N_z)} \sim p\big(\boldsymbol{z} \mid \boldsymbol{\nu}^{(i)}, \boldsymbol{\omega}\big)$
   - This marginalizes over measurement realizations, background realizations, and any random elements in the model.
3. For each $\boldsymbol{z}_{(i)}^{(j)}$: sample N_s parameter vectors $\boldsymbol{\theta}_{(i,j)}^{(1)}, \ldots, \boldsymbol{\theta}_{(i,j)}^{(N_s)} \sim p\Big(\boldsymbol{\theta} \mid \boldsymbol{z}_{(i)}^{(j)}\Big)$
   - These samples are drawn via the sampling procedure in Chapter 4.

We then approximate the desired quantity by

$$\mathbb{E}_{\boldsymbol{\theta} \mid \boldsymbol{\omega}}[f(\boldsymbol{\theta})] \approx \frac{1}{N_\nu} \sum_{i=1}^{N_\nu} \frac{1}{N_z} \sum_{j=1}^{N_z} \frac{1}{N_s} \sum_{k=1}^{N_s} f\Big(\boldsymbol{\theta}_{(i,j)}^{(k)}\Big). \tag{5.21}$$

This technique will be widely used to assess system performance in Chapter 6.

Chapter 6

Experimental Results

In this chapter we employ the performance metrics and the procedure described in Chapter 5 to assess system performance on a variety of experiments. These experiments use synthetic data so that ground truth is known and available for comparison to the results. We additionally show results from running on a real-world data set.

The procedure for evaluating system performance given in Section 5.4 will be widely applied here; we will also adopt the technique of partitioning our parameter set into control parameters $\boldsymbol{\omega}$ and nuisance parameters $\boldsymbol{\nu}$. Generally the control variable set $\boldsymbol{\omega}$ include one or more variables that we are holding constant plus one or more variables for which we want to examine system performance. We then sample multiple realizations for a variety of marginalized parameter configurations. Each test that we ran included nearly 4,000 different source and path configurations. We ran the sampler for 20,000 iterations, the first half of which were discarded as burn-in.

For the set of synthetic experiments we are interested in, the path of the aircraft is usually an example of a parameter that we wish to marginalize out. Domain expert knowledge indicates that flight paths are usually designed as a series of passes perpendicular to the direction of the wind, starting with a pass furthest upwind and moving further downwind with each subsequent pass. Unless explicitly stated, in our examples we set the wind direction to due East (positive x_1 direction) and make multiple passes along the x_2 (North/South) dimension. The x_1 starting location of the paths and the density of the passes are included in the set of nuisance parameters $\boldsymbol{\nu}$ and marginalized over. Figure 6.1 shows a set of paths commonly-used throughout the experiments. Every setting of the remaining parameters (including any other nuisance parameters and all of the control parameters) are computed for each of the trajectories shown in Figure 6.1.

In order for our testing to be meaningful we have to assume that the model represents the real world "accurately enough". Specifically, we know that the real world violates several modeling assumptions, e.g. the wind speed and direction is not from source to sensor and the sources may not be point sources. Specifically, we assume that the survey area is very large compared to the size of the sources and that the sources are sparse in the survey area. As such, the sources are likely to be well-separated in space. We also need the background model to adequately model the variability that one would observe in real-world background concentrations. This allows us to marginalize

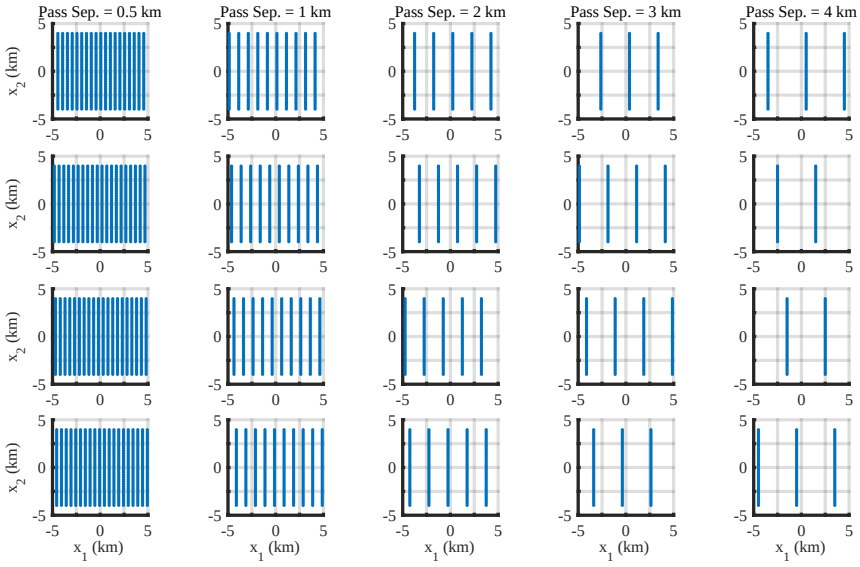

Figure 6.1: Path settings are included in the set of variables to marginalize over. The above path configurations are used to represent a uniform sampling from the set of possible paths. We assume that any path will begin far upwind and make passes perpendicular to the wind direction (towards positive x_1).

out the background by simply sampling from the background prior $p(\boldsymbol{b})$ for each of the different configurations. Finally, we assume that the current formulation of the wind model is also "close enough" to reality within some small region around each measurement location. This entails that the wind direction measured at one point is "relatively the same" between that point and the downwind measurement location with which it shares an edge in the background precision matrix $\mathbf{J}_b$. Additionally, violating this assumption prevents us from being able to realistically generate observations from the model. Accurate simulation would instead require one of the Lagrangian stochastic particle simulation methods introduced in Chapter 2.

The remainder of the section will discuss the results of a series of experiments. The first examines our ability to correctly estimate a known source location as the emission rate of the source varies. We attempt to find the emission rate at which our system can no longer reliably identify the source. Next, we examine how far apart a pair of sources must be before we identify them as two separate sources instead of a single larger source. The third experiment deals with equivalence classes imposed by the dispersion parameter equations and the inherent ambiguity between a close source that is large and a small source that is far away. This has potential implications for flight path selection. Another experiment examines several modifications to the data collection process, namely, the comparative benefits of increasing the measurement frequency (taking more measurements) versus reducing the measurement error noise

(taking "better" measurements). Finally, we explore model mismatch by performing combined inference on two separate flights where some subset of sources turned on or off in between the flights but the model assumes sources are either "always on" or "always off". The performance of the system in each of these cases is evaluated using the performance metrics from Chapter 5.

6.1 Source visibility

Our ability to estimate the location of a source, or even detect it at all, is very important to understand. A major factor in our ability to detect a source is its emission rate, which effectively controls the magnitude of the source contributions to the measurements relative to the other signal components, such as the background contribution and the measurement noise. In this experiment we investigate how our ability to detect a source changes as a function of its emission rate. We consider a single source in isolation and marginalize over a collection of flight paths shown in Figure 6.1. We also marginalize over the source width parameter.

Our system performance is evaluated through computing probability of estimation for the single ground-truth source, as well as the average probability of relevance for the sampled sources. We take the expectations of these probabilities with respect to the control parameters (emission rate), marginalizing out all of the other parameters via sampling. Figure 6.2 shows how the probabilities $\Pr\{\text{EST}(R) \mid s\}$ and $\Pr\{\text{REL}(R) \mid s\}$ change as a function of the radius R and conditioned on the emission rate s. We note that in this case the estimation and relevance probabilities are equal for all values of R, which occurs when the ground-truth and all sampled source configurations contain only a single source. We observe that even for the relatively low emission rate of $s = 0.02$ m^3/s we still obtain nearly 90% probability of detection for up to 500 m accuracy. Additionally, on both plots all five curves have a "bump" that occurs around $R = 5000$m. This is most likely an edge artifact arising as the disk radius grows just past the size where it is fully contained inside the survey area, which affects the rate at which additional sampled sources fall into the detection region.

Figure 6.3 shows $\Pr\{\text{EST}(R) \mid s = s_i\}$ versus $\Pr\{\text{EST}(R) \mid s\}$ and $\Pr\{\text{EST}(R) \mid s = 0\}$ for several values of s_i. This particular plot is analogous to an ROC curve, which generally plots the probability of detection P_{D} versus probability of false alarm P_{FA}. Each point on the curve corresponds to a particular radius R; as R increases from 0 to ∞ we trace the curve from $(0, 0)$ to $(1, 1)$. The intuition is that every point on each curve gives a different *operating point*, or radius, which corresponds to different probabilities of estimation and spurious estimations. This plot clearly indicates the level at which our performance begins to degrade rapidly. We perform well with $s = 0.02$ m^3/s, shown as the blue line. As far as realistic source emission rates go, this is considered fairly low. Halving the emission rate to $s = 0.01$ m^3/s sees a considerable drop in performance: in order to have an estimation probability of 80% we incur a spurious estimation rate of 15%. Halving the emission rate again results in an even greater drop in performance: an

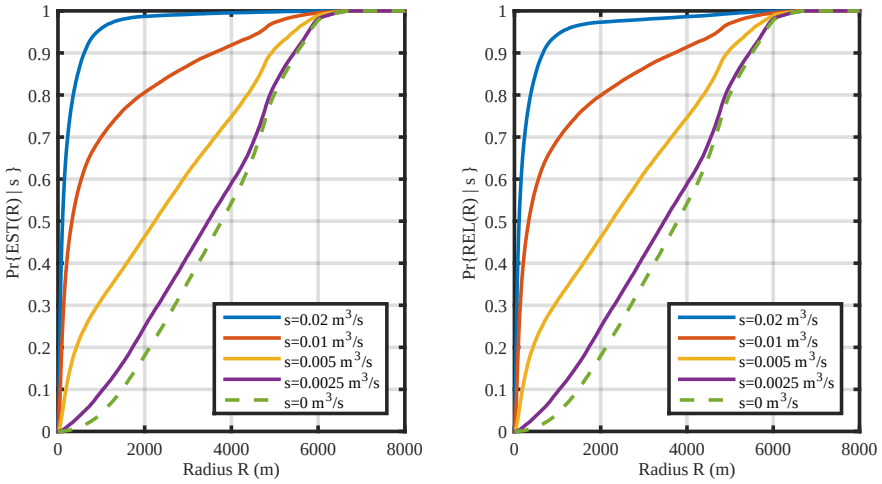

Figure 6.2: Plots of the estimation probability $\Pr\{\mathrm{EST}(R) \,|\, s\}$ and the relevance probability $\Pr\{\mathrm{REL}(R) \,|\, s\}$ as a function of the detection radius R and the emission rate s. The line corresponding to $s = 0$ m^3/s corresponds to chance.

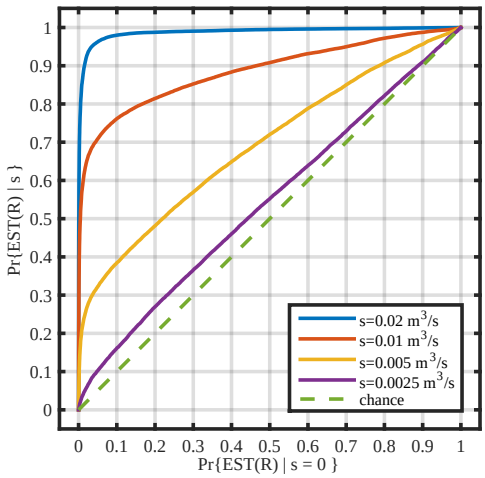

Figure 6.3: This plots our estimation probability conditioned on some source rate versus the spurious estimation probability $\Pr\{\mathrm{EST}(R) \,|\, s = 0\}$. System performance degrades sharply for low emission rates, as indicated by the curves for $s = 0.005$ m^3/s and $s = 0.0025$ m^3/s being very close to the curve of "chance". The curve is traced from $(0, 0)$ to $(1, 1)$ by varying the radius R from 0 to ∞.

estimation probability of 80% now incurs a spurious estimation probability of over 60%, which means that the detection radius is now large enough that it covers about 60% of the survey area. Similar arguments can be made by inspection of Figure 6.2. The main takeaways are that performance consistently degrades as the source rate diminishes, and that curves traced for higher emission rates dominate the curves for lower emission rates.

This analysis assumes that we have a single source, but we can interpret the result in the case where we have multiple sources. Given some radius of detection R we can approximate the expected estimation probability of a source of strength s' in a multiple-source setting by the corresponding probability $\Pr\{\text{EST}(R) \mid s = s'\}$ that is computed in the single-source setting. We now consider whether this approximation provides an upper bound or a lower bound on the true average probability of estimation given some new source configuration.

Suppose we have two sources. If they are far enough apart that their plumes do not intersect (i.e. that there are no measurements for which they both contribute more than some amount ϵ), then we can treat them as independent and each source's expected estimation probability will be given by the value obtained through our analysis, assuming we have estimation probabilities for an isolated source of emission rate s_j. We now consider how each source's expected estimation probability changes as the sources move close together. As the sources become increasingly nearer, at some point their detection regions, parameterized by R, will intersect. It is nearly unavoidable that we will have an estimate from source A (w.l.o.g.) land within the detection region of both sources, hence counting towards the estimation probability of the source A *and* source B. This "extra" sampled source will give source B a larger expected estimation probability than in the isolated case that we use for our analysis in this section. As a result, we see that the expected probability of estimation obtained in this analysis is a lower bound for the expected estimation probability for any given source in a multiple-emitter setting.

■ 6.2 Source resolvability

In this experiment we examine our system's ability to identify when we have two sources instead of just a single source. As a quick thought experiment, consider two sources of equal size and strength separated by a very small distance ϵ. The inference algorithm will estimate them as a single source of the same size with an emission rate equal to the sum of the two emission rates. Now move the sources apart gradually. At some distance Δ our inference algorithm will correctly *resolve* them into two separate sources. Our goal in this experiment is to identify the conditions under which a pair of sources can be resolved.

As before, we marginalize over the path variables via the configurations shown in Figure 6.1. In order to ensure that the effects we observe are completely related to the diffusion model and not a problem of observability, we set the source emission rates to

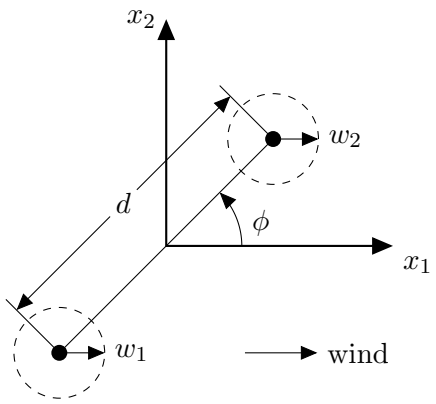

Figure 6.4: A diagram for the setup of the resolvability experiment.

be very high, $s_1 = s_2 = 0.1$ m$^3/s$. The two sources are arranged as shown in Figure 6.4. Each black dot in the figure represents a source; these sources have widths w_1 and w_2 meters, respectively. We consider two different width configurations: one where the sources are both of equal widths, $w_1 = w_2 = 10$ m, and one where $w_1 = 20$ m and $w_2 = 0$ m. Finally, the sources are separated by some distance d m and the separation is at some angle ϕ relative to the wind direction. The midpoint between the two sources has (x_1, x_2) location $(0, 0)$. When $\phi = 0°$ one source is directly downwind of the other; when $\phi = 90°$ the sources are separated exclusively in the cross-wind direction. In our experiment the separation distance d can take values of 50 m, 100 m, 200 m, 300 m, 400 m, or 500 m. The angle of separation ϕ can take values of 0°, 30°, 60°, or 90°.

Our intuition would indicate that sources would become resolvable once separated by several times their combined widths. Additionally, we would expect that sources are resolvable earlier when they are separated in the cross-wind direction than in the downwind direction. We evaluate our results by examining the posterior distribution over the number of sources and examining the probability maps associated with the posterior distribution.

Figure 6.5 shows the posterior distributions of the number of sources m given the control parameters ϕ and d. The two sources are better resolved in posterior distributions that are most sharply peaked around $m = 2$. From Figure 6.5 we see that sources separated in the downwind direction are *very* difficult to differentiate–even sources separated by up to 500 meters are not decisively identified as separate in 60% of cases. As the axis of separation becomes increasingly perpendicular to the wind the source separability improves dramatically. Comparing the second and third rows, for $\phi = 30°$ and $\phi = 60°$, we see that the latter row places more probability mass on $m = 2$ and adds greater probability mass to $m = 2$ for smaller values of d. For $\phi = 90°$ we do not observe a similar improvement over $\phi = 60°$. In fact, in certain cases there is actually *less* probability mass assigned to the correct number of sources $m = 2$. We hypothesize that some factor not visible in this experiment is influencing these results.

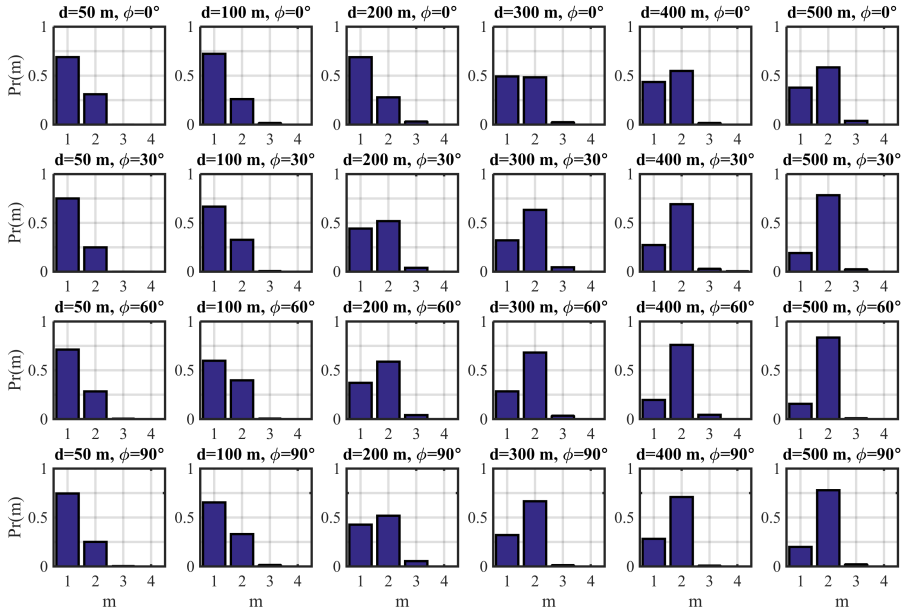

Figure 6.5: The posterior distribution over the number of sources m for the values of source separation distance d and angle of separation ϕ relative to the wind direction.

We investigate the posterior distribution of source locations using the *probability map* approach outlined in Chapter 5. Figure 6.6 shows the probability map corresponding to the ground-truth source configuration $\boldsymbol{\theta}^*$ for each of our possible control variable combinations. Figure 6.7 shows the expected probability map, where the expectation is taken over the distribution $p(\boldsymbol{\theta} \,|\, \boldsymbol{z})$. The latter figure is most of interest to us. The first two columns clearly show that the sources are not yet differentiated and are still being represented as one source. At some separation distance between 100 and 200 m the sources become resolvable. The third column shows that, especially for $\phi = 60°$ and $\phi = 90°$, the sources are often differentiated but sometimes not. The posterior distribution over source locations is bimodal in this case. We can either have two sources spread apart and further downwind or a single source further upwind with the combined emission rate. This trend continues as d continues to increase.

Curiously, as the separation between the two ground-truth sources increases, the position of the estimated third source decreases in the upwind direction. We suspect that this corresponds to samples from the posterior where the plume envelope of the single larger source further back completely encompasses the envelopes of the two smaller sources further downwind. This possibly highlights some ambiguity between how wide or diffuse a source is and its distance upwind of some measurement locations. This connection will be explored in a further experiment.

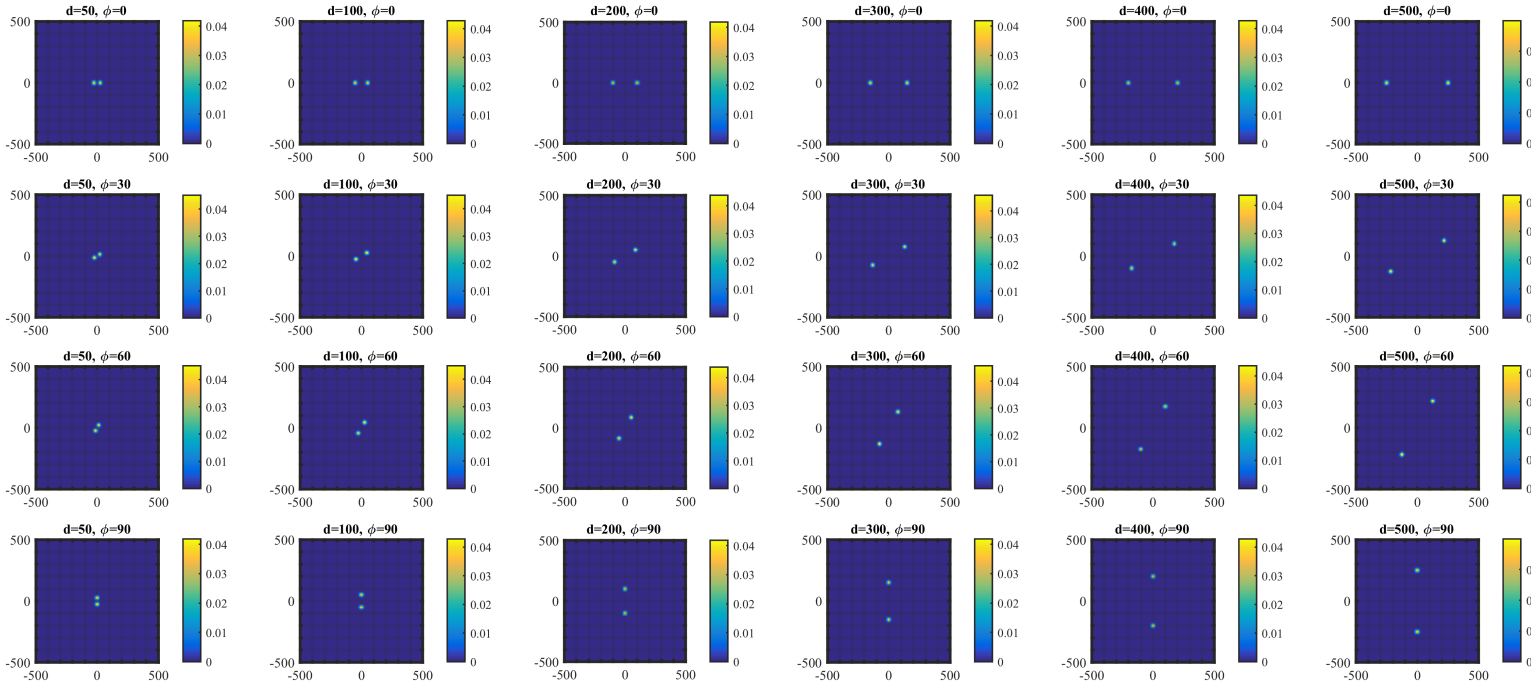

Figure 6.6: Probability maps for source locations, showing the ground truth distribution. The two sources we are trying to differentiate each have width of 10 m.

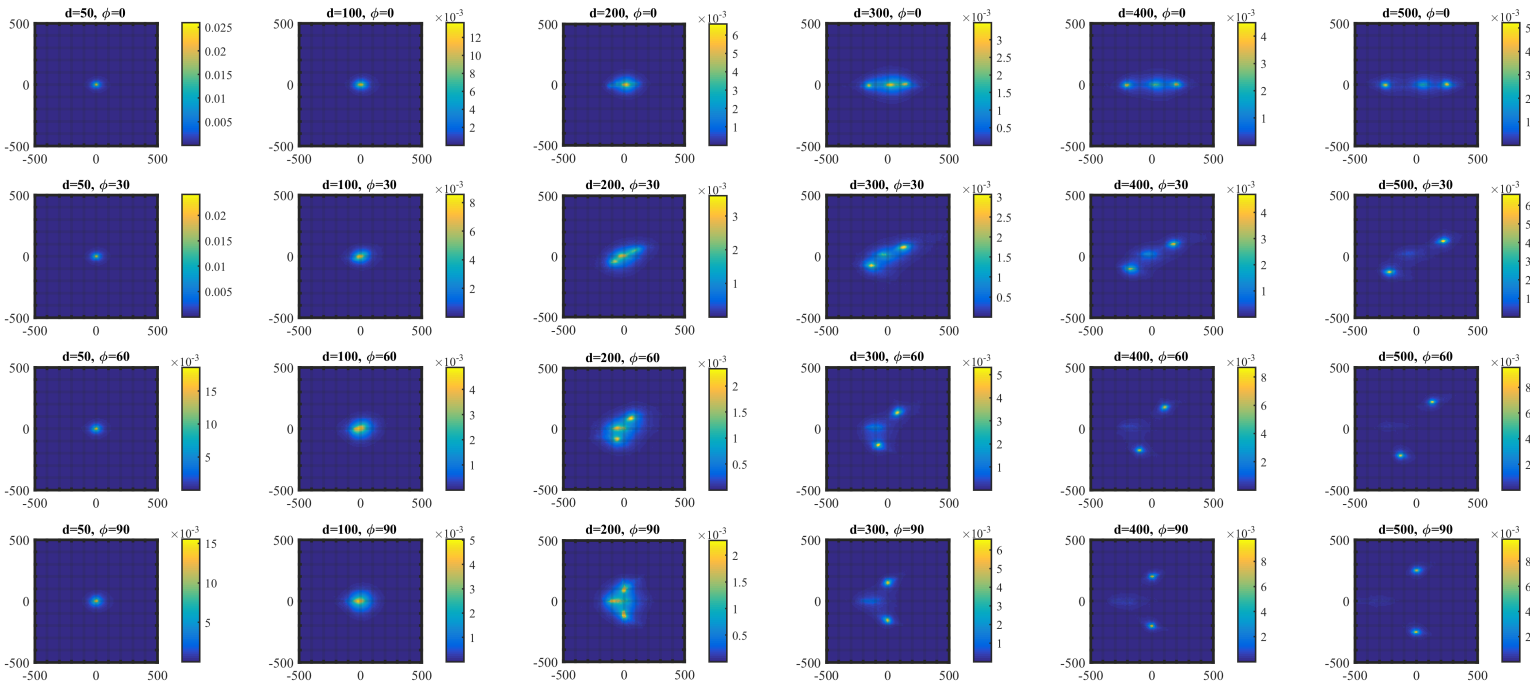

Figure 6.7: Probability maps over source locations, showing the mean of the estimated distributions. The two sources we are trying to differentiate each have width of 10 m.

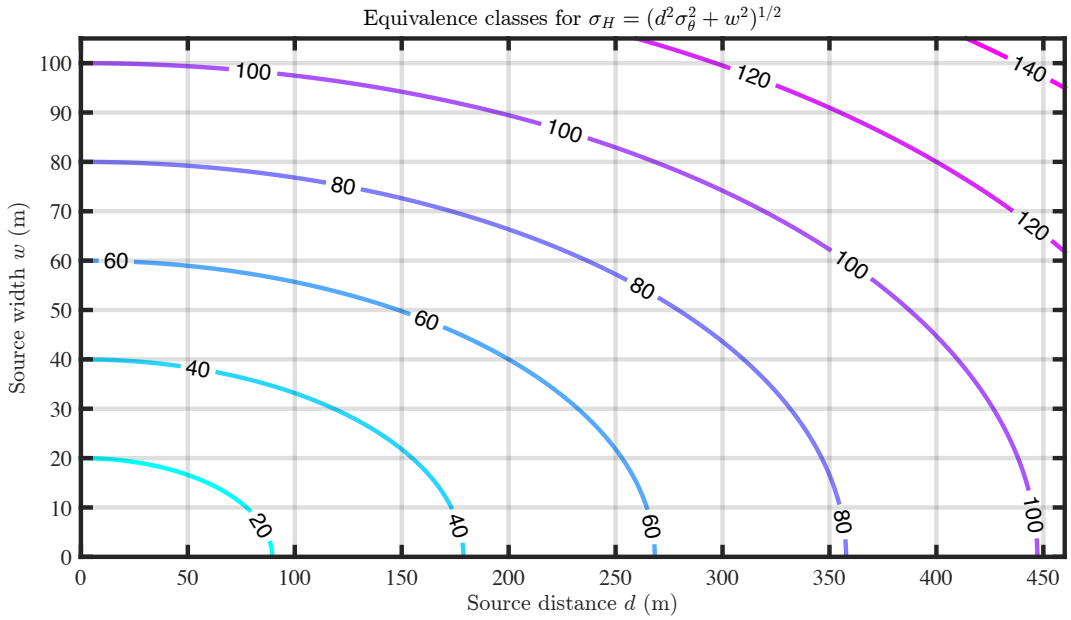

Figure 6.8: A plot showing sets of upwind source locations δ and the corresponding source widths w that result in identical plume horizontal dispersion parameter values σ_H.

■ 6.3 Source widths and flight direction

In the previous experiment we uncovered a potential ambiguity in the model regarding the source width parameter. Specifically, there is a tradeoff between the width of a source and its wind-aligned distance from the measurement location d, as indicated by the definition of the horizontal concentration plume standard deviation:

$$\sigma_H(d, w) = \left(d^2\sigma_\theta^2 + w^2\right)^{1/2} \tag{6.1}$$

where d is the distance of the measurement location, equivalent to the upwind distance from the measurement to the source location; w is the source width; and σ_θ is the wind angle standard deviation.

Recall that the horizontal plume standard deviation σ_H defines the width of the plume concentration profile, which along with the vertical plume standard deviation σ_V, completely determines the statistical properties the observed concentrations. Hence, if we temporarily ignore the effect of σ_V, having multiple source configurations that provide the same σ_H means that these configurations are completely indistinguishable for our model and inference procedure.

Equation 6.1 implicitly defines an equivalence class for each possible horizontal plume standard deviation σ_H, which can arise from many possible values of d and w. Figure 6.8 shows these combinations of d and w for several values of σ_H. For

instance, consider a plume width of $\sigma_H = 100$ m. This may come from a source of width $w = 90$ m located 200 m upwind or a source of width $w = 0$ located 450 m upwind—these two cases, and any of the configurations on the line between them, cannot be differentiated as they provide an identical concentration distribution. Hence, given some set of measurements all taken at the same downwind location, our sampling algorithm will generate sources along the curves on Figure 6.8 instead of converging to the exact point (d^*, w^*).

It is tempting to expect the vertical plume standard deviation σ_V, which is uniquely defined by the downwind distance d, to offer assistance in resolving this ambiguity. Unfortunately, the plume vertical standard deviation is not especially helpful due to the vertical plume reflections that occur within the model.

We anticipate that when we take measurements from multiple downwind distances from some source, we are effectively obtaining a set curves similar to those from Figure 6.8. The configuration that satisfies the measurements at all of the downwind distances will effectively be the point of intersection of the curves defined by each pass. Recall that each pass is perpendicular to the wind direction, so the downwind distance from every point along a pass to the source is identical. That is, so we only obtain one (d, w) curve for each pass.

We now consider angling the passes relative to the wind direction to help identify the true source location and width. We also wish to see if, and how, this ambiguity manifests and changes for sources of different sizes. Hence, our control variables for this experiment are the source size w, which we consider for values of 25 m, 50 m, 75 m, and 100 m, and flight angle relative to the wind θ_{path}, for which we consider values of 90°, 75°, 60°, and 45°. We marginalize over path separation, offset, background and measurement noise realizations. As before, we hold the emission rate constant at a high value to ensure that our single source at $(0, 0)$ is very easily observable and that any and all ambiguity is not related to our ability to "see" the source.

We would expect that the larger the source is the greater variation we would observe in its position along the direction of the wind. Specifically, we expect to see a "tail" behind the true source location where the source width is being traded off for downwind distance. We would also expect that flying at an angle to the wind provides measurements that allow the algorithm to better differentiate between members of the equivalence classes, so the "tails" described above would be shorter for values of θ_{path} that are closer to 45° than 90°.

Figure 6.9 shows the expected probability maps obtained after marginalization via sampling. The true probability map for each source size is shown in the top row; each column corresponds to a different source size. In the second through fifth row the path angle θ_{path} is adjusted from 90° to 45° in 15°-decrements. The wind is moving in the positive x_1 direction. We observe immediately that sources do have a "tail" that extends upwind, and that this is especially noticeable for larger sources. Interestingly, the area of highest probability is approximately mid-way back on the tail, as if our inference algorithm resorts to sampling uniformly from the members of the equivalence

classes and the area of the highest probability works out to be the median of this uniform distribution. This means that the most likely estimate of our source location will be inherently biased slightly upwind of the source, where the amount of the bias is somehow related to the width of the ground-truth source. Finally, we note that the path angle does not appear to have a significant effect on our ability to differentiate between members of the equivalence class.

These experimental results indicate that further work is needed to better understand the implications of the source dispersion model. It is unclear whether the elements of a given equivalence class are truly unidentifiable or identifiable in theory but not in practice due to noise. We may consider this problem in future work. In most practical scenarios the source widths are actually rather small ($w \approx 20$ m) so we may just set $w_{\max}$ to a similar value to ensure that the range of possible source locations is also small.

■ 6.4 Sensor-based improvements

In this section we examine two different modifications that could be made to the sensor used in the measurement process. Suppose we could design a sensor that had a higher sampling rate, which increases our number of measurements n. Suppose we could also design a sensor that had a higher precision, this would correspond in our model to having a lower variance for the additive Gaussian white measurement noise. In this experiment we consider the impact of changing these two parameters on our ability to detect a source.

Without loss of generality, here we investigate the detection of a single ground-truth source with location $(0, 0)$ and size $w = 20$ meters. Our control parameters are the sensor's sampling rate f_s and the sensor's measurement error standard deviation σ_ε, denoted here by σ. We allow the sampling rate f_s to take values of 1/4 Hz, 1/2 Hz, 1 Hz, and 2 Hz; the measurement error standard deviation σ takes values 10 ppb, 1 ppb, and 0.1 ppb, equivalently 10^{-8}, 10^{-9}, and 10^{-10}. Our current real-world system has $f_s = 1/3$ Hz since the alternatives are all within the realm of possibility; similarly our alternative values for σ represent an order of magnitude improvement or degradation. We marginalize over the paths shown in Figure 6.1, as well as source emission rates of $s = 0.1$ m^3/s and $s = 0.01$ m^3/s. Finally, we also marginalize over background realizations and measurement noise realizations.

Naturally we expect that the more measurements we obtain the better our ability to estimate a source; similarly, we expect that the better the measurements we obtain the better our ability to estimate a source. Figure 6.10 shows our estimation probability changes as a function of radius R for various settings of f_s and σ. We note that "bump" near $R = 5000$ m is due to the same edge artifact that was mentioned in the source visibility experiment. In agreement with our intuition, we see that both the estimation probability and the probability of relevance increase as the sampling rate increases, and also as the amount of noise in our measurements decreases. Interestingly, we also note

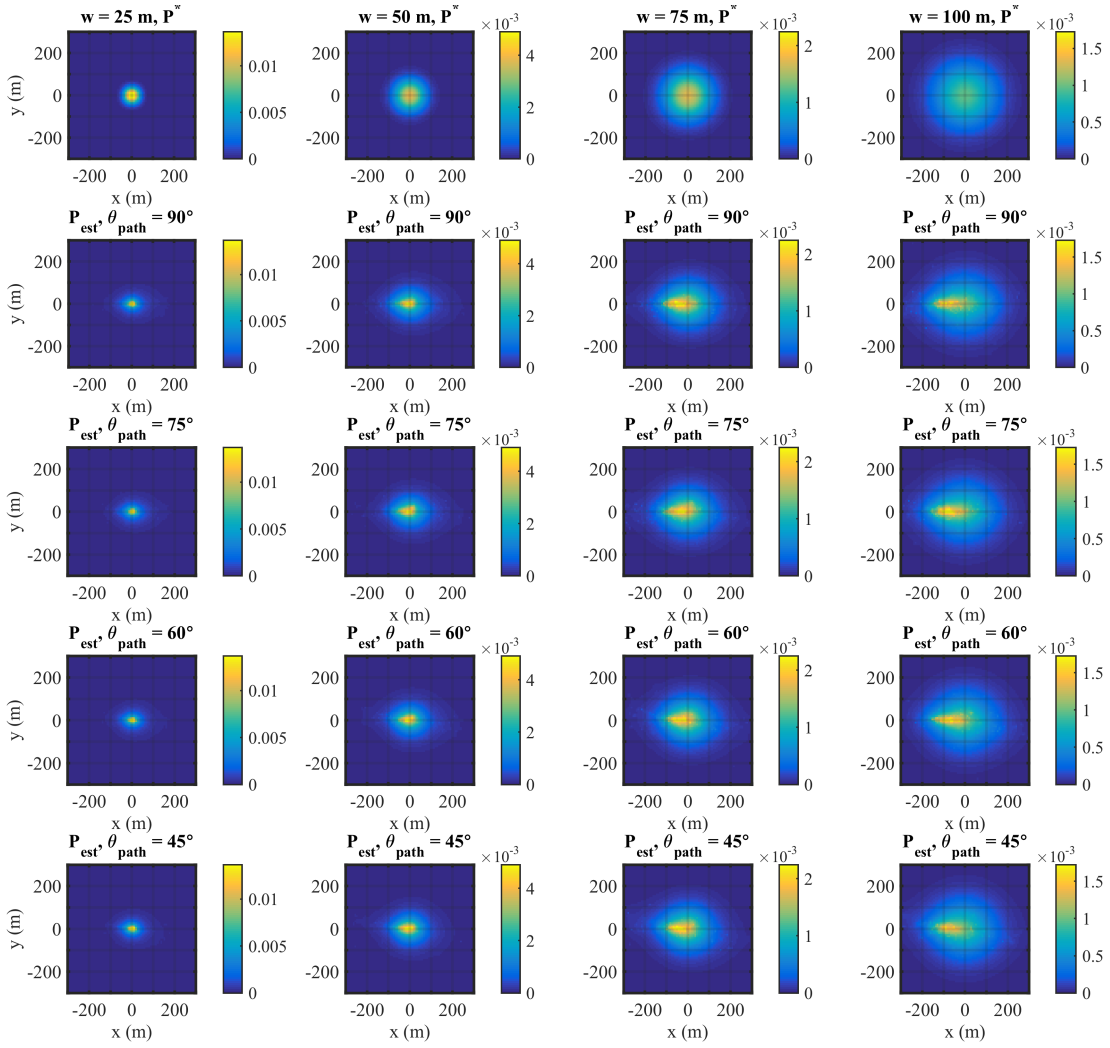

Figure 6.9: Probability maps for the true (top row) and estimated source location distributions (other rows). The ground-truth source location is at $(0, 0)$. The width of the source varies with each column. Each of the rows other than the first shows the average estimated source location distribution given a flight path at the designated angle to the wind. Note that larger sources have higher location ambiguity in the x direction and that the pass angle θ_{path} appears to have no discernible effect on the probability map for the source location.

that the dashed lines, which indicate the probability of spurious estimations, decrease as the amount of noise in our measurements decreases. This is consistent with what we would expect.

Figure 6.11 shows four plots, all containing curves that are analogous to ROC curves. The estimation probability for non-zero emissions is plotted versus the probability of estimation given no source contributions, also called our spurious estimation probability. We note that as the sampling rate increases the marginal benefit of increasing the sensor's accuracy by an order of magnitude decreases substantially.

Interestingly, in Figure 6.11, when the sampling rate is $f_s = 2$ Hz we perform better with $\sigma_\varepsilon = 1$ ppb than with $\sigma_\varepsilon = 0.1$ ppb since the red curve is *barely* above the green curve for some range of radii. In fact, if we would further reduce σ_ε without making any other modifications to the inference procedure we would see performance drop considerably. This is due to the way that we explore the parameter space via our Markov chain, which is affected by changing the number of measurements n (through f_s) and the measurement error standard deviation σ_ε. Consider the Hastings ratio in Equation 4.25. As n grows to be extremely large or σ_ε becomes extremely small our Hastings ratio approaches zero, so our acceptance probability drops dramatically. While our posterior distribution is much narrower our proposal variances do not change. Hence, what was previously a small (and relatively acceptable) step near the current point we are instead making very large steps that are very rarely accepted. This problem could be avoided by *a priori* setting the proposal variances for $\boldsymbol{x}$, $\boldsymbol{s}$, and $\boldsymbol{w}$ to be smaller when we have smaller σ_ε or larger n.

■ 6.5 Inference over multiple flights

The final synthetic experiment concerns a question of model mismatch. One ability of our system that has not yet been discussed is the capability to run inference simultaneously on data from multiple flights. That is, we can easily create a stacked measurement vector, measurement location matrix, and time vectors where times from the second flight are shifted relative to the beginning of the first flight. We can then run inference on this multi-flight survey. The measurement model becomes

$$\begin{bmatrix} \boldsymbol{z}^1 \\ \boldsymbol{z}^2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}^1 \\ \mathbf{A}^2 \end{bmatrix} \boldsymbol{s} + \begin{bmatrix} \boldsymbol{b}^1 \\ \boldsymbol{b}^2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}^1 \\ \boldsymbol{\varepsilon}^2 \end{bmatrix} \tag{6.2}$$

where the coupling matrices $\mathbf{A}^i$ could be generated from different wind fields $\boldsymbol{u}$. The prior on $\boldsymbol{b}$ can similarly be easily combined, since for large enough time differences between the flights it can be shown that the joint distribution of $\boldsymbol{b}^1$ and $\boldsymbol{b}^2$ has a block-diagonal precision matrix

$$\mathbf{J} \to \begin{bmatrix} \mathbf{J}^1 & \mathbf{0} \\ \mathbf{0} & \mathbf{J}^2 \end{bmatrix}$$

which we can invert far more efficiently than a non-block-diagonal precision. We have empirically found that the backgrounds become practically independent for temporal

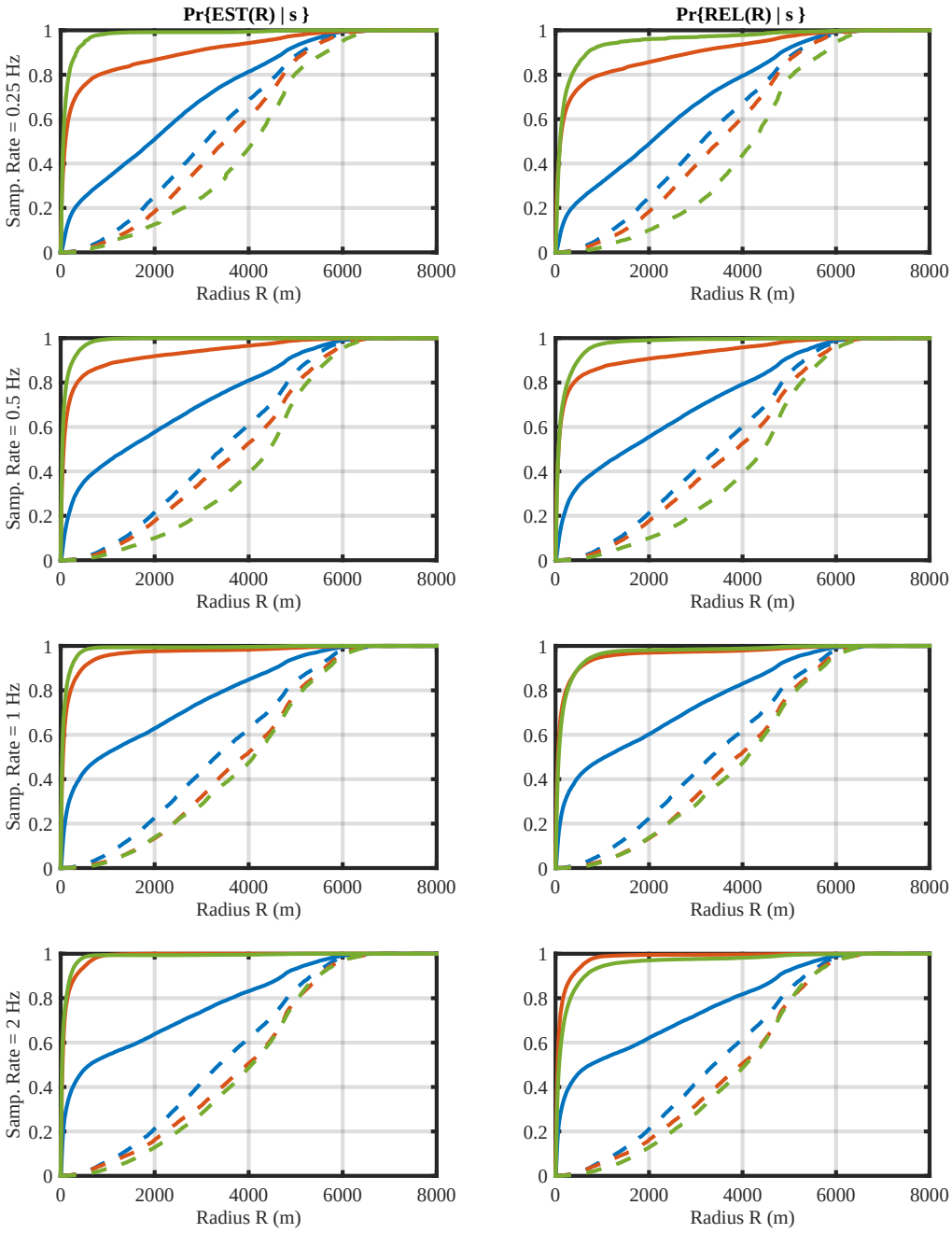

Figure 6.10: Plots of (left) the probability of having an estimated source within R meters of the ground-truth source at (0,0); (right) the probability of having a ground-truth source within R meters of a randomly-selected *estimated* source. Green lines correspond to $\sigma_\varepsilon = 0.1$ ppb, red lines to $\sigma_\varepsilon = 1$ ppb, and blue lines to $\sigma_\varepsilon = 10$ ppb. Solid lines are the average probability given data with a simulated ground-truth source at (0,0) and dashed lines are the average probability given data containing no source and all other parameters equal.

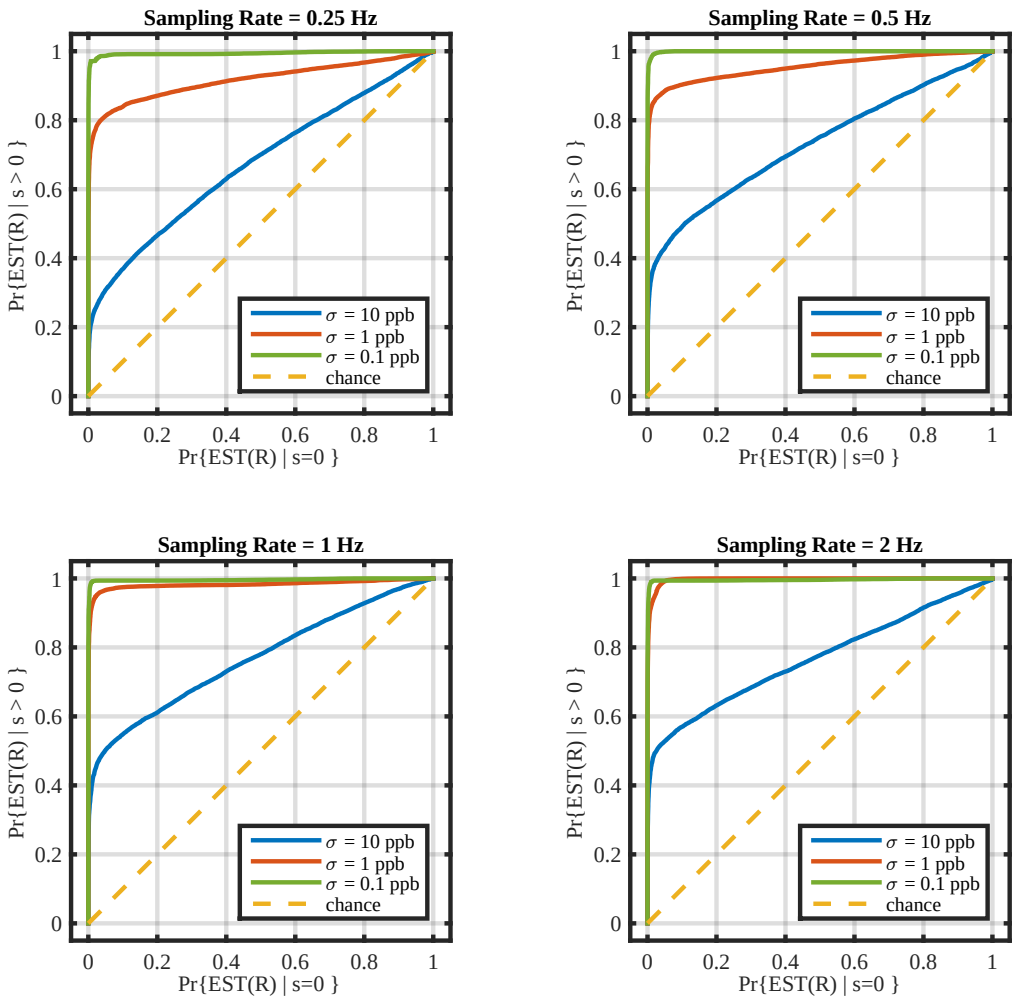

Figure 6.11: Curves analogous to ROC curves for the sensor-oriented experiment. We see that performance improves for higher sampling rates and lower measurement noise standard deviations, indicated by the curves for those configurations lying completely above the curves for lower sampling rates and higher measurement noise standard deviations.

flight separation on the order of 24 hours. For simultaneously-occurring flights a combined precision matrix may also be computed by the standard procedure outlined in Chapter 3.

We recognize that this model assumes that source parameters including locations $\boldsymbol{x}_j$, widths w_j, and emission rates s_j are constant between the two flights. This is *exactly* the assumption that we will violate in our experiment on model mismatch. In this experiment we consider our ability to detect a single source located at $(0, 0)$. In the first flight we set the emission rate $s_1 = 0.02$ m^3/s so that the source is just visible enough to be estimated most of the time while still leaving room for improvement. We then violate the constant-source assumption of the model by allowing the emission rate used in the second flight to take the value $s_2 = s_1/2 = 0.01$ m^3/s, $s_2 = s_1/4 = 0.005$ m^3/s, or $s_2 = 0$ m^3/s. Our key concern here is to quantify how much our system performance degrades by not accounting for possible changes in the emission rate of a source between flights, and whether or not this capacity needs to be incorporated into the model.

We are additionally interested in the impact of a changing wind direction between the two flights on our ability to do inference. It is suspected that our ability to accurately estimate sources improves because the sources and plumes will effectively be seen from multiple angles. The wind fields are assumed constant during each individual flight. Without loss of generality we set the wind direction in the first flight to be towards positive x_1 and let the wind direction in the second flight be ϕ, where ϕ can take values $0°$, $30°$, $60°$, and $90°$. In both flights we assume that we follow standard procedure and make passes perpendicular to the wind direction.

We marginalize over the source size w, as well as the flight paths for the two experiments. For each flight we select one of the flight paths shown in Figure 6.1, rotated such that the passes are perpendicular to the wind for that flight, in keeping with standard procedure. The paths are constrained to have the same pass separation in both flights.

Figure 6.12 shows the probability of estimating a source within R meters of the ground-truth source for varying values of R. Interestingly, as long as the wind directions are very close between the two flights there is relatively little degradation in the estimation probability as the emission rates differ. Even if the source is completely off in the second flight our probability of estimation only degrades by about 15% for $R \approx 1000$. We note a similar change in the probability of relevance for a small change ϕ in wind angle. As the difference in the wind angle grows, however, we generally see that the more consistent emission rates see an improvement in performance, whereas the performance reduction for a source deactivation becomes much larger–this is especially noticeable in the bottom-left plot ($\phi = 90$) where the blue and red lines move closer to the upper-left corner and the purple line drops from its position in the $\phi = 60°$ case. Similar patterns are observed in Figure 6.13.

This experiment would indicate that model mismatch, at least for sources with an emission rate of $s_1 = 0.02$ m^3/s, does not severely degrade performance. Further investigations may consider whether or not this degradation is acceptable and the results

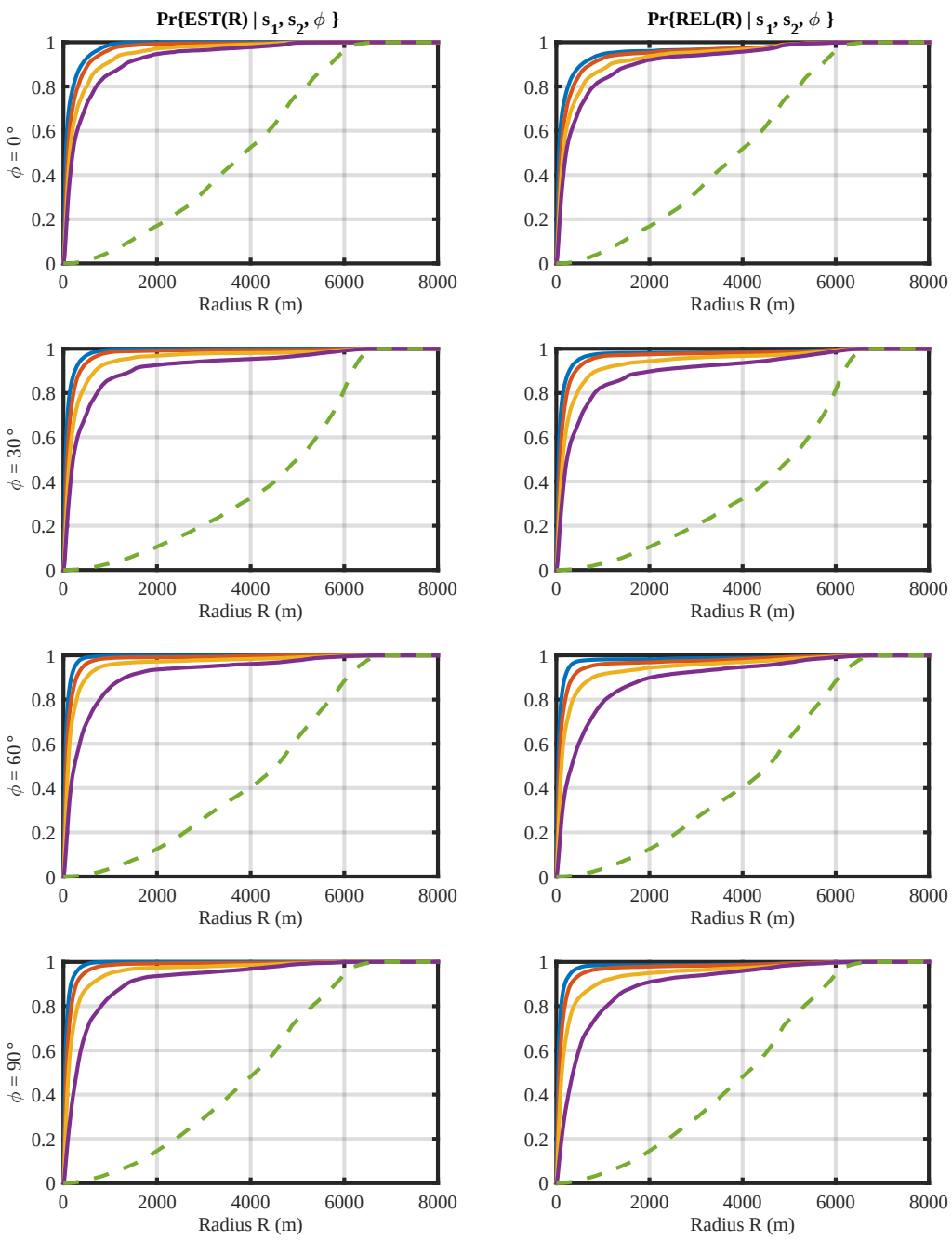

Figure 6.12: Plots of (left) the probability of having an estimated source within R meters of the ground-truth source at (0,0); (right) the probability of having a ground-truth source within R meters of a randomly-selected *estimated* source. The blue line indicates $s_2 = s_1$, the red line shows $s_2 = s_1/2$, the gold shows $s_2 = s_1/4$, and the purple shows $s_2 = 0$ The green line indicates the corresponding value for $s_1 = s_2 = 0$, corresponding with not having any source in either flight.

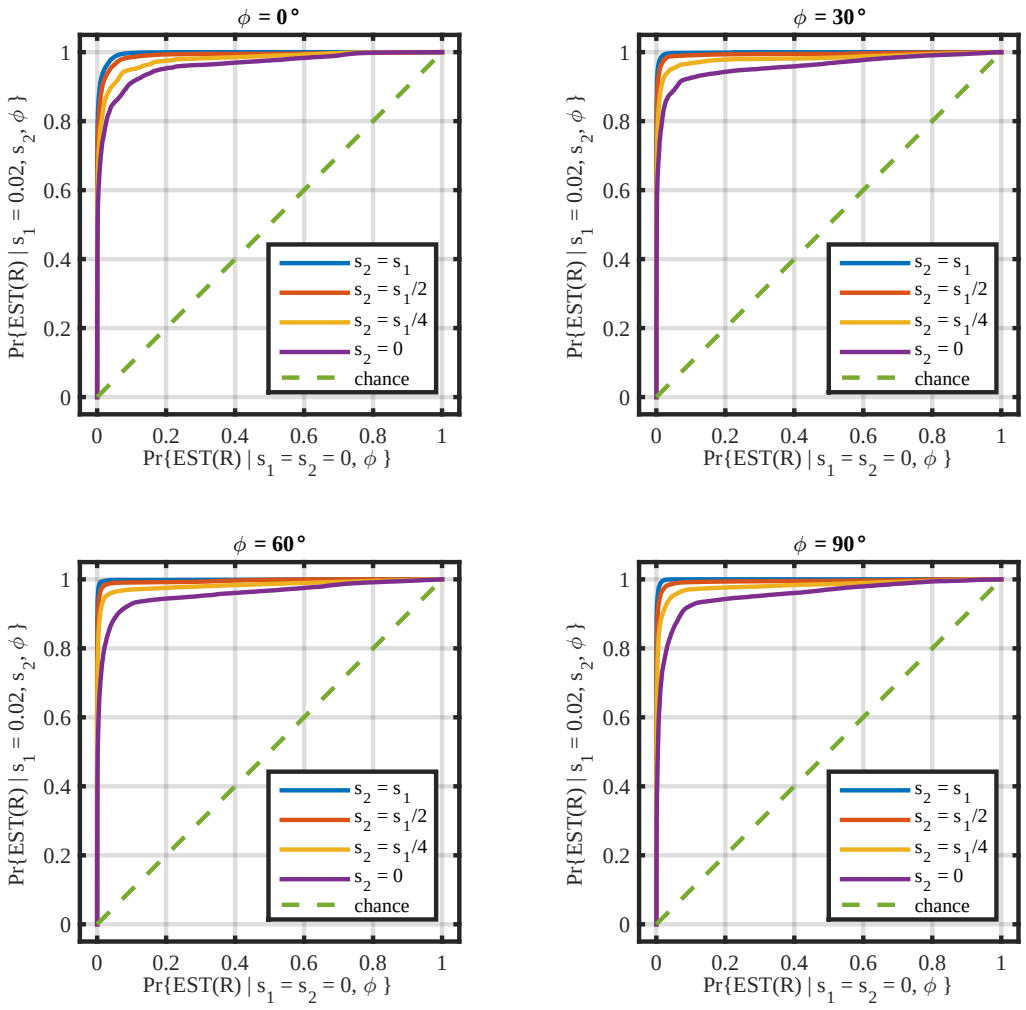

Figure 6.13: Curves analogous to ROC curves for the model mismatch experiment. We see that a more consistent emission rate s_2 in the second flight leads to better performance in all of the cases. As the difference ϕ in the wind angle between the two flights grows, strongly-performing configurations improve and poorly-performing configurations degrade, as shown by a widening of the gap between the blue/red/gold curves and the purple curve.

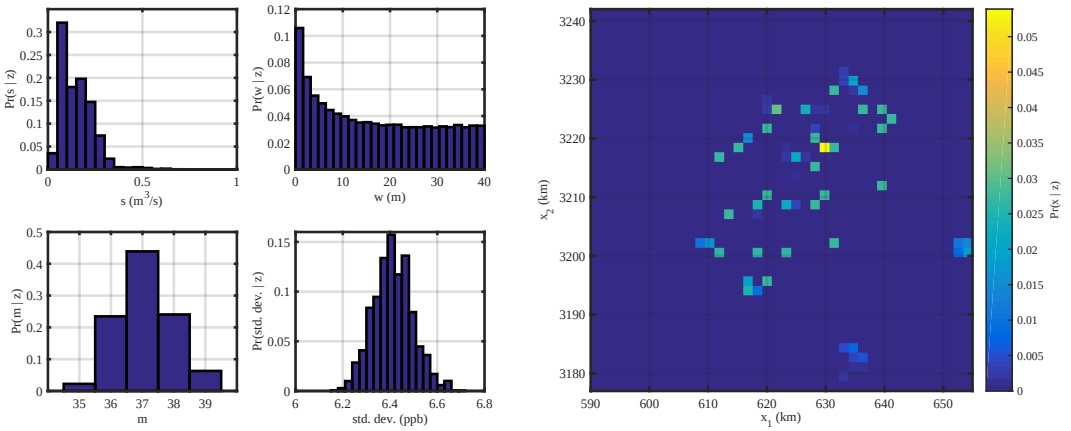

Figure 6.14: Marginal distributions for the various parameters.

of similar experiments for more highly- or lowly-emitting sources.

■ 6.6 Results from real data

In this section we show the results of performing inference on a set of real data. Unlike in synthetic cases, in real data we often do not know the ground-truth parameter values. When this is the case, many of the performance metrics developed in the previous chapter can no longer be applied. However, we can still compute posterior distributions, expected emission rate maps, and estimate the MAP solution from our sampler output.

For the data set used here, an aircraft flew over a rural area at an altitude of 150 meters and took measurements roughly every three seconds. The survey area is 65 km by 65 km. Unfortunately ground-truth is not known. The sampler was run for 20,000 iterations and the first 10,000 were discarded as burn-in.

The marginal distributions for the various parameters are shown in Figure 6.14. Our distribution over the source count m has a mode at $m = 37$ sources with locations scattered as shown by the joint marginal distribution over x_1 and x_2. The values in the distribution over emission rates are all well-within the expected range. Finally, our distribution over the width parameter looks nearly uniform, likely due to the ambiguity that we studied earlier in the chapter.

Figure 6.15 shows the flight path as a series of blue dots. The size of the dots corresponds to the magnitude of the measurement. Wind is blowing towards the north-west. The MAP solution is given by the green dots; the plumes generated by the MAP solution are shown as background color. We note that the measurements closer to the north-west corner of the survey area have much higher magnitude, as indicated by the larger size of the dots. Our inference results suggest that this is due to the background in that region. The sources are generally positioned upwind of a sequence of measurements that have a sharp increase and then decrease in magnitude, indicating

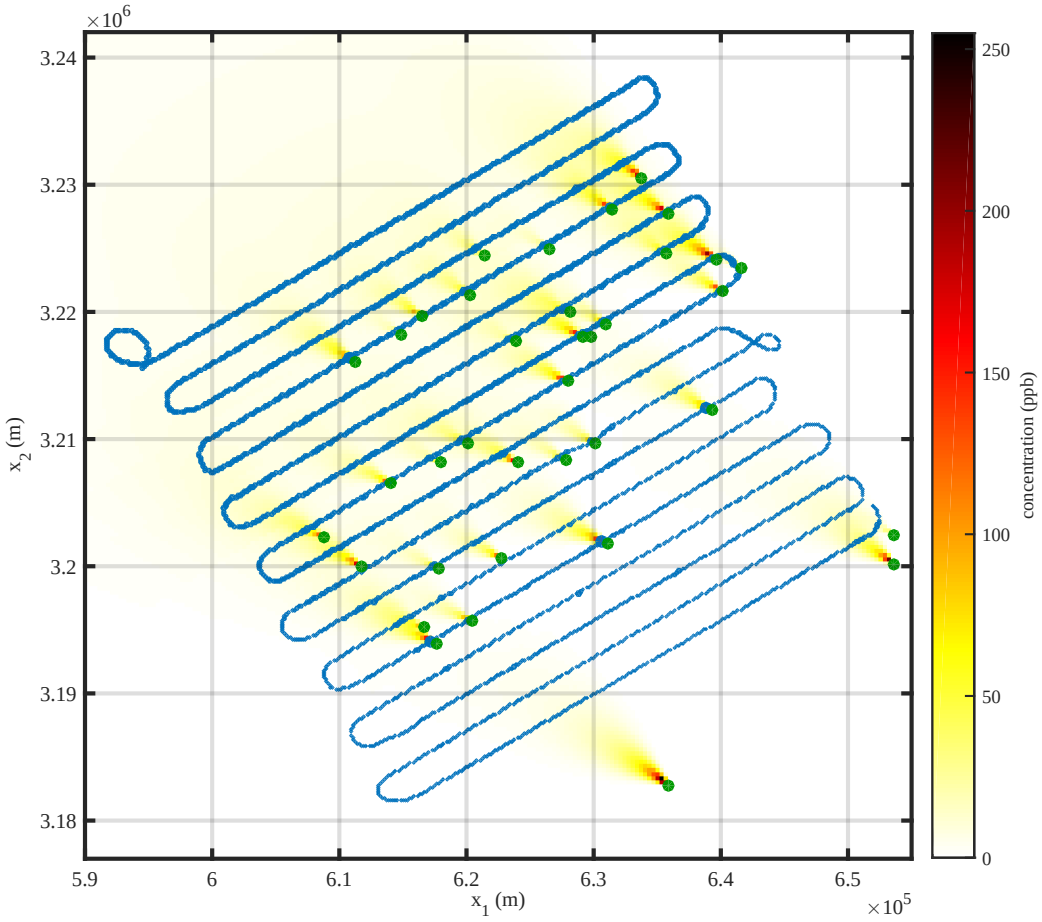

Figure 6.15: MAP solution for the inference on real data. Colormap shows expected concentration at each location according to the sources in the MAP estimate and dispersed according to the Gaussian plume model.

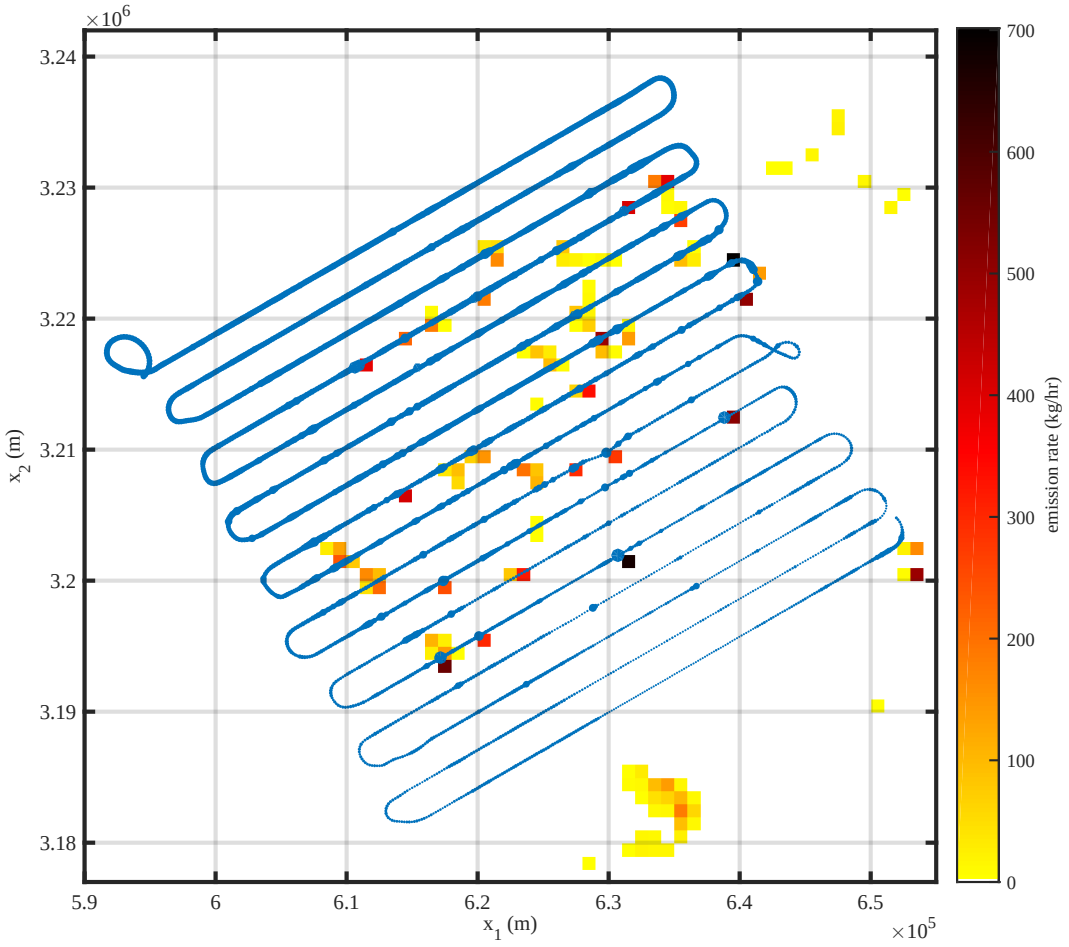

Figure 6.16: Expected emission rate map. This is computed as described in Chapter 5.

sensible placement. One exception to this is the source near the bottom right that is very far from the measurements. The MAP solution does not give any indication as to how "certain" we are that this source exists–for this intuition one should examine the expected emission rate map.

The expected emission rate map indicates, for each cell, the expected emission rate produced within this cell. The exact form was specified in Chapter 5. This is very similar to the joint marginal distribution over location, except it also provides some notion of source strength. For instance, the source that is near $(6.18 \times 10^5, 3.195 \times 10^5)$

appears to be emitting very strongly judging by the emission rate map. In the absence of real ground-truth data we cannot make further remarks.

Clearly inference on real data is a challenging task. Many assumptions used by our system, especially those for the dispersion model, are broken in real-world scenarios. Still, our algorithm generates plausible and sensible output. Several factors might improve our ability to perform inference, particularly on real-world data, including adding a non-uniform prior distributions over the number of sources, some of the source parameters, and the measurement error standard deviation. The incorporation of a more-robust wind model would also provide a large performance gain.

Chapter 7

Conclusions and future work

Identifying some emission source, and properties thereof, from remotely obtained concentration measurements is a problem of great interest in several intersecting areas of study. In this thesis we have focused on a single application—remote detection of gaseous emission sources—and developed a model and inference procedure suitable to the task. The selected problem differs from many in the existing literature. In particular, we do not assume the number of sources is known *a priori* and we sought source contributions that were embedded in a background that was several orders of magnitude larger. We formulated this problem as a probabilistic graphical model with associated prior, likelihood, and posterior distributions. Using a well-founded Bayesian inference technique, Reversible-jump MCMC, we are able to draw samples from the posterior distribution.

We developed a variety of performance metrics that leverage the sample-based representation of the posterior distribution. These metrics were then applied to a range of synthetic experiments for which we had ground-truth available. This allowed us to quantify the performance of the inference algorithm for some particular set of parameters. By effectively marginalizing over all but a small subset of the parameters, we were able to estimate the average system performance with respect to that small subset of parameters.

Future work could explore several different directions. Now that we have the mechanisms in place to evaluate different modeling choices, we would like to consider using non-uniform priors for many of the parameters including emission rate, source width, and standard deviation. We may also consider changing the model for the sources so that they are drawn from an inhomogeneous Poisson point process for which the rate function is unknown. An alternative direction would consider different wind models, including the incorporation of Lagrangian particle simulation methods.

Bibliography

[1] 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*, 41(11):2283–2289, 2007.

[2] G. A. Briggs. Diffusion estimation for small emissions. Technical report, NOAA Report ATDL-106, 1973.

[3] W. L. Buntine. Operations for Learning with Graphical Models. *Journal of Artificial Intelligence Research*, 2:159–225, 1994.

[4] L. W. Burger and M. V. Mulholland. Real-time prediction of point-source distributions using an anemometer-bivane and a microprocessor. *Atmospheric Environment (1967)*, 22(7):1309–1217, 1988.

[5] G. Cervone, P. Franzese, and A. Grajdeanu. Characterization of atmospheric contaminant sources using adaptive evolutionary algorithms. *Atmospheric Environment*, 44(31):3787–3796, 2010.

[6] F. K. Chow, B. Kosović, and S. T. Chan. Source inversion for contaminant plume dispersion in urban environments using building-resolving simulations. *Journal of Applied Meteorology and Climatology*, 47(6):1533–1572, 2008.

[7] B. M. Davies and D. J. Thomson. Comparisons of some parametrizations of wind direction variability with observations. *Atmospheric Environment*, 33:4909–4917, 1999.

[8] A. De Visscher. *Air Dispersion Modeling: Foundations and Applications*. John Wiley & Sons, Inc., first edition, 2014.

[9] R. R. Draxler. Determination of atmospheric diffusion parameters. *Atmospheric Environment*, 10:99–105, 1976.

[10] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, third edition, 2014.

[11] F. Gifford. Use of routine meteorological observations for estimating the atmospheric dispersion. *Nuclear Safety*, 2(4):47–57, 1961.

[12] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC, 1996.

[13] P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732, 1995.

[14] W. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[15] S. E. Haupt, G. S. Young, and C. T. Allen. A genetic algorithm method to assimilate sensor data for a toxic contaminant release. *Journal of Computers (Finland)*, 2(6):85–93, 2007.

[16] A. Hazart, J.-F. Giovannelli, S. Dubost, and L. Chatellier. Inverse transport problem of estimating point-like source using a Bayesian parametric method with MCMC. *Signal Processing*, 96:346–361, 2014.

[17] B. Hirst, P. Jonathan, F. González del Cueto, D. Randell, and O. Kosut. Locating and quantifying gas emission sources using remotely obtained concentration data. *Atmospheric Environment*, 74:141–158, Aug. 2013.

[18] C. Huang, T. Hsing, N. Cressie, A. R. Ganguly, V. A. Protopopescu, and N. S. Rao. Bayesian source detection and parameter estimation of a plume model based on sensor network measurements. *Applied Stochastic Models in Business and Industry*, 26(4):331–348, July 2010.

[19] G. Johannesson, B. Hanley, and J. Nitao. Dynamic Bayesian models via Monte Carlo - an introduction with examples. 2004.

[20] A. Keats, E. Yee, and F.-S. Lien. Bayesian inference for source determination with applications to a complex urban environment. *Atmospheric Environment*, 41:465–479, 2007.

[21] A. Khemka, C. A. Bouman, and M. R. Bell. Inverse problems in atmospheric dispersion with randomly scattered sensors. *Digital Signal Processing: A Review Journal*, 16(5):638–651, 2006.

[22] K. J. Long, S. E. Haupt, and G. S. Young. Assessing sensitivity of source term estimation. *Atmospheric Environment*, 44(12):1558–1567, 2010.

[23] S. Neumann, L. Glascoe, B. Kosovi, K. Dyer, W. Hanley, and J. Nitao. Event reconstruction for atmospheric releases employing urban puff model UDM with stochastic inversion methodology. *Sixth Symposium on the Urban Environment, American Meteorological Society*, 2006.

[24] F. Pasquill. Estimation of the dispersion of windborne material. *Meteorological Magazine*, 90:33–49, 1961.

[25] F. Pasquill. Atmospheric dispersion of pollution. *Quarterly Journal of the Royal Meteorological Society*, 97:369–395, 1971.

[26] K. S. Rao. Source estimation methods for atmospheric dispersion. *Atmospheric Environment*, 41(33):6964–6973, 2007.

[27] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer, New York, NY, 2004.

[28] H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications*. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, 2005.

[29] J. H. Seinfeld and S. N. Pandis. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. John Wiley & Sons, Inc., 2nd edition, 2006.

[30] I. Senocak, N. W. Hengartner, M. B. Short, and W. B. Daniel. Stochastic event reconstruction of atmospheric contaminant dispersion using Bayesian inference. *Atmospheric Environment*, 42:7718–7727, 2008.

[31] J. M. Stockie. The Mathematics of Atmospheric Dispersion Modeling. *SIAM Review*, 53:349–372, 2011.

[32] D. B. Turner. *Workbook of Atmospheric Dispersion Estimates*. CRC Press, 2 edition, 1994.

[33] D. Wade and I. Senocak. Stochastic reconstruction of multiple source atmospheric contaminant dispersion events. *Atmospheric Environment*, 74:45–51, Aug. 2013.

[34] E. Yee. Bayesian probabilistic approach for inverse source determination from limited and noisy chemical or biological sensor concentration measurements. In *Proceedings of SPIE*, volume 6554, page 65540W, 2007.

[35] E. Yee. Theory for Reconstruction of an Unknown Number of Contaminant Sources using Probabilistic Inference. *Boundary-Layer Meteorology*, 127(3):359–394, Mar. 2008.

[36] E. Yee, A. Gunatilaka, and B. Ristic. Comparison of Two Approaches for Detection and Estimation of Radioactive Sources, 2011.

[37] P. Zannetti. *Air Pollution Modeling: Theories, Computational Methods and Available Software*. Springer Science+Business Media, 1990.