Model-Data Fusion and Adaptive Sensing for Large Scale Systems: Applications to Atmospheric Release Incidents

by

Reza Madankan

A dissertation submitted to the Faculty of the Graduate School of the

State University of New York at Buffalo

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Mechanical and Aerospace Engineering

May 14, 2014

Acknowledgement

This dissertation wouldn't be complete without the help of many people and I would like to deeply appreciate all the ones who helped, inspired, and taught me during all the times I was working on my dissertation. I would love to express my deepest gratitude to my academic adviser, Dr. Puneet Singla, for his excellent guidance, caring, patience, and support during my stay in University at Buffalo. I am also greatly thankful to my committee member Dr. Tarunraj Singh who was always supportive and helpful when it was needed. It is also a great pleasure to thank Dr. John Crassidis, my committee member and Dr. Peter Scott for their invaluable feedback and suggestions.

Many thanks to the members of volcanology research group, Dr. Abani Patra, Dr. E. Bruce Pitman, Dr. Marcus Bursik, Dr. Peter Webley, Dr. Mat Jones, Dr. John Dehn, and Mike Pavolonis, Solene Pouget, E. Ramona Stefanescu, and Dr. Christopher Hughes for their great helps, constructive feedback, and cooperation.

I would also like to thank all the friends in Buffalo, especially Ehsan and Saba, who were always there when I needed and were a source of motivation and support during past few years. Special thanks to my colleagues, especially Nagavenkat Adurthi and Vishwajeet Kumar, in Laboratory of Autonomous and Intelligent Robotic Systems (LAIRS) who were more than just colleague during my stay in university at Buffalo.

Last but not least, I would love to thank my parents, my sister and my brother whom I haven't seen them for six years, but they have always been a source of inspiration and helped me through these years.

Reza Madankan

Spring 2014

This page is left blank intentionally

Contents

List of Figures

List of Tables

This page is left blank intentionally

Abstract

All across the world, toxic material clouds are emitted from sources, such as industrial plants, vehicular traffic, and volcanic eruptions can contain chemical, biological or radiological material. With the growing fear of natural, accidental or deliberate release of toxic agents, there is tremendous interest in precise source characterization and generating accurate hazard maps of toxic material dispersion for appropriate disaster management.

In this dissertation, an end-to-end framework has been developed for probabilistic source characterization and forecasting of atmospheric release incidents. The proposed methodology consists of three major components which are combined together to perform the task of source characterization and forecasting. These components include Uncertainty Quantification, Optimal Information Collection, and Data Assimilation.

Precise approximation of prior statistics is crucial to ensure performance of the source characterization process. In this work, an efficient quadrature based method has been utilized for quantification of uncertainty in plume dispersion models that are subject to uncertain source parameters. In addition, a fast and accurate approach is utilized for the approximation of probabilistic hazard maps, based on combination of polynomial chaos theory and recently developed Conjugate Unscented Transformation.

Besides precise quantification of uncertainty, having useful measurement data is also highly important to warranty accurate source parameter estimation. The performance of source characterization is highly affected by applied sensor orientation for data observation. Hence, a general framework has been developed for the optimal allocation of data observation sensors, to improve performance of the source characterization process. The key goal of this framework is to optimally locate a set of mobile sensors such that measurement of *better* data is guaranteed. This is achieved by maximizing the mutual information between model predictions and observed data, given a set of kinetic constraints on mobile sensors. Dynamic programming method has been utilized to solve the resulting optimal control problem.

To complete the loop of source characterization process, two different estimation techniques, minimum variance estimation framework and Bayesian Inference method has been developed to fuse model forecast with measurement data.

Incomplete information regarding the distribution of associated noise signal in measurement data, is another major challenge in the source characterization of plume dispersion incidents. This frequently happens in data assimilation of atmospheric data by using the satellite imagery. This occurs due to the fact that satellite imagery data can be polluted with noise, depending on weather conditions, clouds, humidity, etc. Unfortunately, there is no accurate procedure to quantify the error in recorded satellite data. Hence, using classical data assimilation methods in this situation is not straight forward. In this dissertation, the basic idea of a novel approach has been proposed to tackle these types of real world problems with more accuracy and robustness. A simple example demonstrating the real-world scenario is presented to validate the developed methodology.

Keywords: Data Assimilation, Inverse Problem, Optimal Sensor Placement, Dynamic Data Driven Application and System (DDDAS), Adaptive Sensing

Chapter 1

Introduction

1.1 Source Characterization of Large Scale Systems

Distributed parameter systems [1] have an important role in modeling different physical phenomena ranging from spatial modeling of infectious diseases to transport of amino acids within cells, diffusion of oxygen into blood, doping of semiconductors, air quality modeling, and diffusion/dispersion of a pollutant material through atmosphere. Most of proposed numerical approaches for solving distributed parameter systems, use discretization techniques to obtain approximate solution of the system. This often leads to a system of differential equation of large number of states which should be solved numerically. For instance, applied numerical solutions for the shallow water model [2] converts a set of partial differential equations into a set of differential equations, where the number of states depends on applied resolution for discretization. There often exists a trade off between the number of states for discretized system and desired accuracy for approximate numerical solution.

Plume dispersion phenomenon is one of the examples for distributed parameter systems, where its behavior is governed by a dispersion/advection partial differential equation. Emission of toxic material clouds from sources, such as industrial plants, vehicular traffic, deliberate toxic releases, and volcanic eruptions is one of the potential threats to environment and human society. With increasing number of instances of toxic material release, there is tremendous interest in precise source characterization and generating accurate hazard maps of toxic material dispersion for appropriate disaster management

Different algorithms have been proposed to solve the source characterization problem

for the plume dispersion phenomenon. These algorithms can be divided into three major categories: adjoint based techniques, optimization based methods, and probabilistic methods [3]. Adjoint based techniques are based on sensitivity analysis of cost function at final time with respect to changes of states/parameters in earlier time. One should note that backward solution is needed to perform this sensitivity analysis. The major drawback of adjoint methods is that they use linear approximations of the dynamics. Hence, adjoint based methods are useful when linearized approximations are valid. In addition, application of adjoint based methods is restricted to the cases when applied cost function is differentiable.

The basic idea of optimization based techniques is to find the source parameters in such a way that it minimizes the difference between the observed data and model prediction. Clearly, different cost function and optimization algorithms can be used to solve this problem. One of the major drawbacks of optimization based methods is their dependency on the applied initial value for optimization, which can specially restrict their application in non-convex optimization problems. In addition, optimization based techniques result in a point estimate for source parameters and no confidence bound is associated with these point estimates. One might be interested in statistical confidence bounds, based on possible presence of error in model prediction or presence of noise in measurement data.

Applied probabilistic approaches for source identification in hazardous chemical releases can be divided into three different categories: Kalman based methods [4,5], Bayesian Monte Carlo (BMC) method [6], and Bayesian Markov Chain Monte Carlo (BMCMC) technique [7]. Kalman filtering methods provide posterior mean and variance of source parameters, based on linearity assumption for dynamics of the system and Gaussian assumption for prior distribution of uncertain parameters and distribution of present noise in measurement data. These assumptions restrict applicability of Kalman filtering methods.

BMC is another statistical tool which is widely used for source characterization of plume dispersion incidents. The basic idea of this approach is to make use of Bayesian inference framework and Monte Carlo simulations. In this approach, numerical model is simulated for a large set of Monte Carlo samples, which are generated based on prior distribution of source parameters. Then, only those Monte Carlo samples whose outputs are consistent with observation data are preserved. As one can see, a large set of Monte Carlo samples is required to get a reasonable approximation for the posterior distribution of source parameters.

One can use BMCMC approach to alleviate computational complexities involved in BMC

method. In BMCMC, Monte Carlo samples are replaced with a Markov Chain Monte Carlo samples. Hence, less number of numerical simulations is needed to get the same approximation for posterior distribution of source parameters. However, still large number of MCMC samples is required to accurately approximate posterior distribution of source parameters. In addition, generating MCMC samples in high dimensions can still be computationally expensive.

1.2 Optimal Information Collection

Independent of applied approach for source estimation, locations of data observation sensors are crucially important while characterizing source parameters in plume dispersion incident. It is clear that poor data measurement, caused by misplacing the sensors over the domain of interest, results in poor estimate of source parameters. Hence finding optimal location of the sensor is very important for the source parameter estimation problem. Due to dynamic of plume dispersion phenomenon, it is much more efficient to apply mobile sensors, instead of static sensors for data monitoring purposes. Different strategies have been suggested to determine the optimal path of applied mobile sensors while source parameters characterization of plume dispersion phenomenon [8–12]. Earlier works to tackle this problem can be categorized as Chemotaxis [13, 14], Anemotaxis [15, 16], and Fluxotaxis [17]. These methods provide algorithms for moving mobile sensors based on concentration gradient, wind-field data, or dispersal material flux.

The major deficiency of aforementioned approaches lies in their restriction to continuous releases incidents. In other words, none of chemotaxis, anemotaxis, and fluxotaxis methods are applicable in presence of spontaneous releases. This is due to the presence of discontinuities in plume flow which hinders evaluation of concentration gradient or flux of the plume during the time. We will discuss about these methods in more detail in Chapter 5.

Besides aforementioned methods, recently multiple researches have been done on the field of optimal sensor placement, based on information theory [8–12]. The basic idea of these methods is to locate data observation sensors at spatial location where *mutual information* of model output and observed data is maximized. Based on dynamics of applied sensors, most of the works in this category can be divided into two different types. In the first scheme, a large network of *stationary* sensors is considered and few of these sensors at each time step

are selected based on calculated mutual information map [9] [10]. Hence, no dynamics is considered for applied sensors in this technique. The major drawback of this approach is that a large number of *stationary* sensors is needed (while a small portion of them is used at each time) to assure performance of proposed approach.

In the second scheme, a set of *mobile* sensors are used which move toward spatial locations where the value of mutual information increases [11] [12]. Hence, much lower number of sensors is needed in this technique. However, path control of applied mobile sensors needs to be considered. Different approaches are developed for path control of applied mobile sensors, ranging from optimal control [12] to moving along gradient of mutual information [11].

The major drawback of present methods in using mobile sensors is that they are susceptible to be trapped in local optima. In other words, none of the existing techniques provide a universal methodology to ensure global optimality of mobile sensor location.

1.3 Challenge of Noise Statistics

As mentioned in Section 1.1, there exist numerous researches [18–24] regarding probabilistic source estimation of dynamic systems, given some measurement data. In all these approaches, there is a crucial need for knowledge of statistics of associated noise in observed data. For instance, mean and variance of the noise are required in all minimum variance based methods. Similarly, in Bayesian inference framework, complete knowledge of noise signal distribution is needed to construct the likelihood function.

A major challenge in source parameter estimation of dynamic systems, especially plume dispersion incidents, is when there is no accurate information regarding the sensor error characteristics, i.e. there exists no accurate information about the statistics of the noise polluted in measurement data. This frequently happens in data assimilation of atmospheric incidents by using satellite imagery. This occurs due to the fact that satellite imagery data are often polluted with noise, depending on weather conditions, clouds, humidity, etc. Unfortunately, there is no precise procedure to quantify the error in the recorded satellite data. Hence, using classical data assimilation methods in this situation is not straight forward. One way to proceed in this situation is to assume some statistics for the associated noise in measurement data and perform the estimation. However, depending on assumed values for noise statistics, obtained estimation results may not be as *inclusive* and *accurate*

as it should be in reality.

1.4 Contribution and Outline of the Dissertation

Contribution

The overall objective of this dissertation is to develop an end-to-end computational framework for uncertainty quantification and data assimilation of large scale systems, especially plume dispersion incidents, with emphasis on addressing above discussed problems in literature. The main contributions of the dissertation is listed as:

1. Develop an end-to-end computational framework for uncertainty quantification and source characterization of large scale systems, with emphasis on plume dispersion incidents. A major advantage of the proposed methodology is its ability to characterize non-Gaussian uncertainty with limited computational effort.

2. Develop a source estimation method which provides a confident estimate of source parameters along their associated uncertainty bounds.

3. Develop a sensor placement algorithm which is implementable in *near real time* and improves performance of applied source estimation method by optimally locating mobile sensors over the spatial domain at each time step. The proposed methodology exploit the non-Gaussian uncertainty in model forecast to accurately compute information metrics.

4. Present a novel approach for source estimation of large scale systems in absence of sensor error characteristics.

Outline of Dissertation

The method presented in this dissertation consists of three different components which are combined together to perform the task of source parameter estimation. These components consist of i) Uncertainty Quantification (UQ), ii) Optimal Information Trajectory Design, and iii) Data Assimilation (DA). Schematic view of whole estimation process is shown in Fig. 1.1.

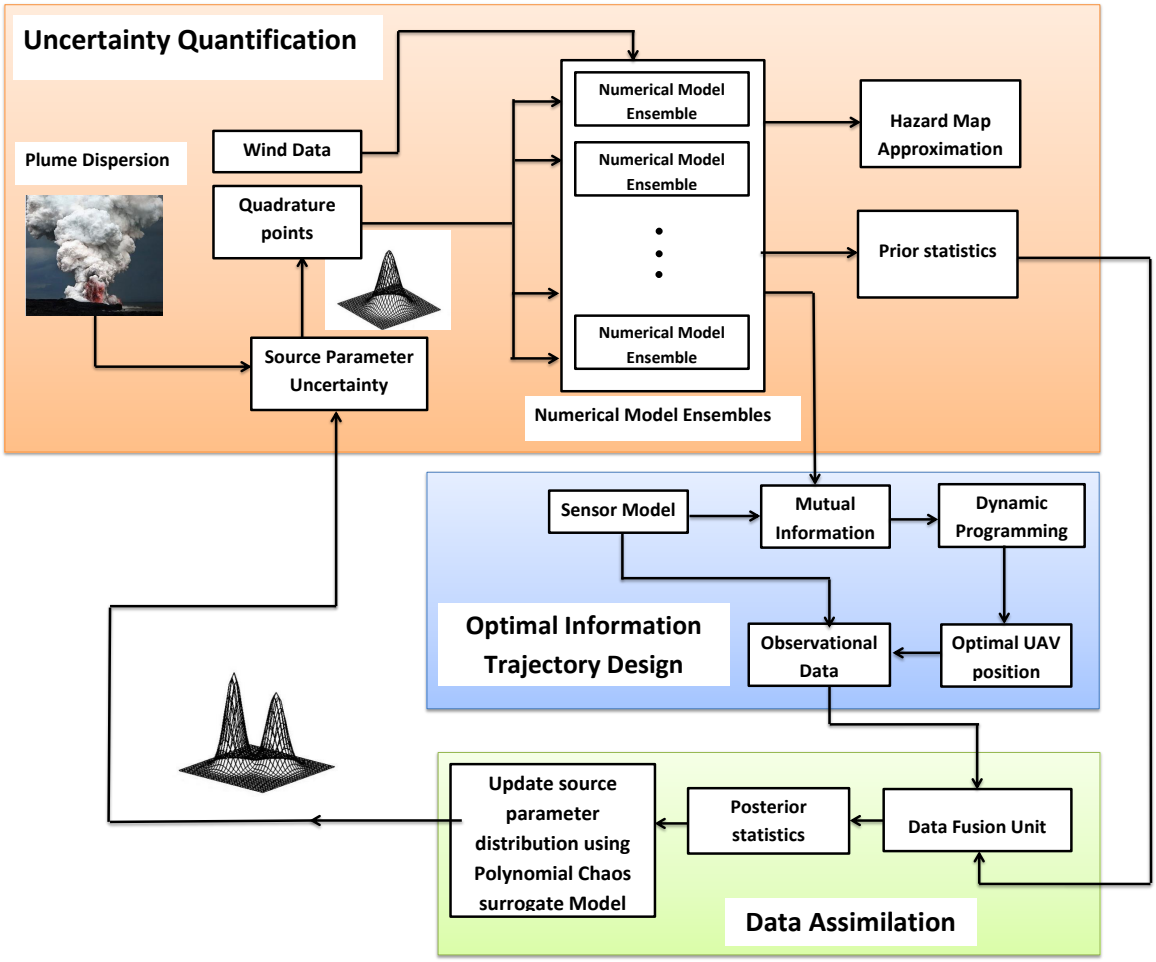


Figure 1.1: Schematic view of estimation process

As shown in Fig. 1.1, estimation process starts with a given uncertainty in source parameters of a dispersion phenomenon. The first step to perform the estimation is to quantify the effect of uncertain source parameters on spatial-temporal distribution of concentration of

dispersive material. This is performed by propagation of a set of quadrature points through the numerical model. Weighted average of propagated quadrature points are then used to determine statistics of dispersive material (mean and covariance) over the spatial domain at a given time. Obtained statistics by pure propagation of quadrature points through numerical model are called *prior* statistics. Accurate approximation of prior statistics is crucial to ensure the performance of estimation process. Hence, utilizing an *appropriate* and *efficient* set of quadrature points is of high importance. In addition to prior statistics of the plume, applied quadrature points can also be combined with a polynomial chaos framework to find probabilistic hazard map for presence of the plume over the spatial domain at each time.

Besides precise quantification of uncertainty, having *useful* measurement data is also highly important to warranty accurate source parameter estimation. Optimal Information Trajectory Design module is used to perform this task during estimation process. The key goal of Optimal Information Trajectory Design is to optimally locate a set of mobile sensors (Unmanned Aerial Vehicles) such that measurement of *better data* is guaranteed at each time step. This is achieved by maximizing the mutual information between model predictions and observed data, given a set of kinetic constraints on mobile sensors. Dynamic Programming [25] [26] is used to solve this optimization problem. This will result in a set of control signals which are applied to each mobile sensor at each time step t_k. The obtained control policy determines the optimal locations for data monitoring sensors during the time.

To complete the estimation process, obtained prior statistics from Uncertainty Quantification and observed data from Optimal Information Trajectory Design are combined together in a source estimation framework. This will result in *posterior* value for statistics (e.g. mean and covariance) of source parameters estimates.

The outline of this dissertation is as follows:

In Chapter 2, we present a quadrature based method for uncertainty quantification of large scale systems. In addition, an algorithm based on generalized Polynomial Chaos surrogate model [27] is presented to approximate probability of presence of dispersal material over spatial domain. Applied techniques for uncertainty quantification of plume dispersion incidents in this chapter provides us with model forecast which is needed for data assimilation in future chapters.

In Chapter 3, two different source estimation approaches are utilized for source parameter estimation of plume dispersion phenomenon. At first, a polynomial chaos based minimum

variance estimation [18] is introduced to perform data assimilation of nonlinear dynamic systems in presence of affine pollution of observational data with noise. In addition, a Bayesian inference data assimilation tool is introduced to tackle source estimation problem in presence of non-affine pollution of observation data with noise signal.

In Chapter 4, developed methods in Chapter 2 and Chapter 3 are then applied to the problem of source parameter estimation and forecasting of Eyjafjallajökull eruption which occurred in April 2010.

The focus of Chapter 5 is about developing a universal scheme for optimal sensor placement. In this chapter, we will develop an optimal sensor placement strategy based on information theory and dynamic programming method to ensure optimal sensor placement of mobile sensors over spatial domain at each time step.

In addition to provided methods in Chapter 3, we present the basic idea of a moment based approach in Chapter 6 to overcome the problem of source estimation of large scale systems in the absence of sensor error characteristics. In this chapter, we will describe a novel approach for source estimation of this type of real world problems which makes use of image comparison techniques and quadrature methods.

Chapter 7 is the final chapter which summarizes the major contributions of the dissertation and discusses the results and the future work in this area.

In the end, please note that even though emphasis of this dissertation is about source characterization of plume dispersion phenomena, all the mathematical developments in this document can be applied to other instances of large scale systems.

Chapter 2

Uncertainty Quantification

2.1 Introduction

The science of Uncertainty Quantification (UQ) describes the propagation of uncertainty in dynamical systems and its major goal is to determine the likelihood of certain outcomes of dynamical systems, given some aspects of dynamical systems to not be exactly known. In other words, UQ provides mathematical tools to characterize the behavior of outputs of dynamical systems, given some sources of uncertainties involved in their structure.

In general, there exist two different categories of uncertainty: Aleatoric uncertainty and Epistemic uncertainty [28]. Aleatoric uncertainty represents the randomness in physical phenomena, while epistemic uncertainty accounts for ignorance in accurate modeling of physical phenomena. The major difference between aleatory and epistemic uncertainty is that the aleatory uncertainty can not be reduced and it can only be better characterized, while epistemic uncertainty can be diminished by collection of information about the system under study.

The focus of this dissertation is on characterizing the uncertainty of dynamical systems in presence of epistemic uncertainty. The very first step to achieve this goal is to quantify the model output uncertainty in presence of epistemic uncertainty, e.g. presence of uncertain parameters or initial condition in model structure. This is also denoted as *uncertainty propagation* or *uncertainty quantification*.

There exist numerous techniques to perform the task of uncertainty quantification of dynamic systems in presence of uncertain parameters and/or initial conditions, ranging from

simple Monte Carlo (MC) method to spectral decomposition based methods, like stochastic finite element [29] and generalized Polynomial Chaos (gPC) [30] [31], and solution of Fokker-Planck-Kolmogorov Equation (FPKE) [19]. One should note that the choice of each method for uncertainty quantification depends on desired level of accuracy and available computational resources.

Ideally, Kolmogorov equation [32] provides exact solution for propagation of uncertainty through dynamical systems that are in presence of parametric and/or model input uncertainties. Exact analytical description of uncertainty, provided by Kolmogorov equation, makes this approach very elegant candidate for uncertainty quantification of dynamical system. However, solving Kolmogorov equation is not usually straightforward. Unfortunately, exact solution of Kolmogorov equation is not attainable other than very few cases. Hence, alternative methods are proposed to approximate the solution of Kolmogorov equation [33–35]. Unfortunately there exists significant computational cost involved in proposed approximate methods which restricts their applicability. Also, in higher dimensions, the computational cost involved in these methods becomes prohibitive. Hence, these methods can not be used to large scale systems.

On the other hand, Monte Carlo method is one of the most simplistic approaches that is widely used for uncertainty quantification of dynamical systems. In Monte Carlo method [36], a large set of points are generated based on distribution of uncertain parameter/initial conditions. Generated random points are then propagated through dynamic model to approximate statistics of model output in future time steps. However, ease of implementation and simplicity of MC approach makes it a very desirable candidate for uncertainty quantification of dynamical system, but its low rate of convergence hinders its applicability when a precise approximation of model output uncertainty is needed. In other words, depending on involved nonlinearities, a large number of random points need to be propagated through dynamic model to ensure convergence of obtained statistics. Clearly, larger number of MC samples results in more accurate statistics, but it results in more computational load at the same time which is undesirable. Schematic view of Monte Carlo method is shown in Fig. 2.1.

Spectral decomposition based techniques [29, 30, 37, 38] are another set of tools that are often used to avoid computational burden of Monte Carlo approach in uncertainty quantification of dynamic systems. The essence of spectral decomposition methods lies in decomposition of uncertain variable into a linear combination of a set of basis functions which

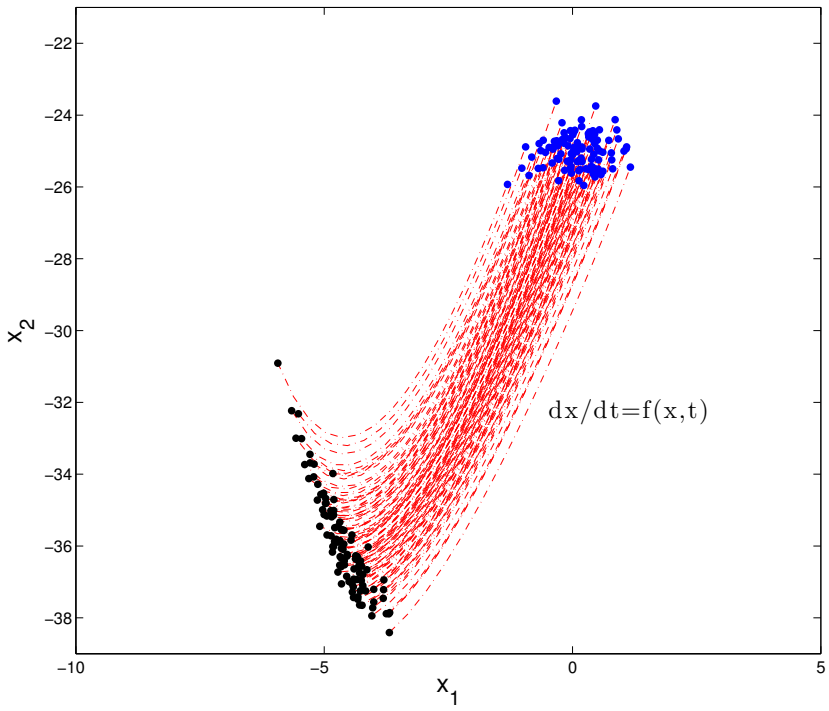


Figure 2.1: Schematic illustration of Monte Carlo approach

are function of standardized random variable $\boldsymbol{\xi} = [\xi_1, \xi_2, \cdots, \xi_m]^T \in \mathbb{R}^m$, where m is the number of existing independent uncertain variables in dynamic model. The major benefit of spectral decomposition based methods is their low computational cost and their ease of implementation. However, convergence issues in presence of severe nonlinearities, large number of independent uncertain variables, and long time uncertainty propagation restrict applicability of spectral decomposition methods.

An alternative approach to perform the task of uncertainty quantification is to make use of quadrature points. In quadrature points method, a set of *intelligently* selected points will be propagated through dynamical model (which can be given as a black-box model or a set of dynamical equations) and statistics of the output (e.g. mean and variance) are then determined by weighted average of model outputs. The major benefit of quadrature scheme is its significantly less computational cost with respect to Monte Carlo method. Also, the computational complexity involved in quadrature methods is far less than the computational complexities involved in solution of Kolmogorov equation. In addition, quadrature method can be combined with spectral decomposition based methods like gPC theory to improve

their performance [39] [40].

To summarize our discussion, a comparison between discussed methods is provided in Fig. 2.2. As one can see, method of quadrature points requires the minimal computational effort, while at the same time it provides reasonable degree of accuracy, comparing with other methods.

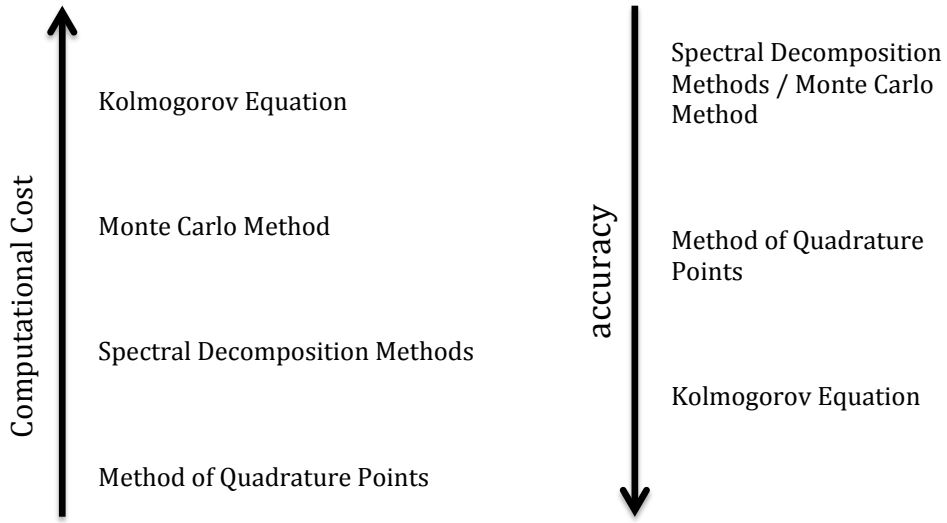


Figure 2.2: Comparison between different applied methods for uncertainty quantification

In this dissertation, due to mentioned benefits of the method of quadrature points with respect to other approaches like MC, spectral decomposition methods, and Kolmogorov equation, a quadrature based method is being used for quantification of uncertainty in plume dispersion models that are subject to uncertain source parameters. In the following, we first describe the method of quadrature points in Section 2.2. Mathematical developments for approximation of natural hazard maps are then explained in Section 2.3. In Section 2.4, we will present multiple numerical simulations to illustrate performance of the proposed approach. Finally, summary of the chapter is presented in Section 2.5.

2.2 Method of Quadrature Points

Propagation of uncertainty due to uncertain input parameters and initial conditions can be approximated by propagation of a set of quadrature points, expanded over a domain defined by uncertain parameters. The resulting method can be viewed as a MC-like evaluation of

system of equations, but with sample points selected by quadrature rules. To explain this in more detail, consider the following dynamical system:

$$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}, \boldsymbol{\Theta}) \tag{2.1}$$

where $\mathbf{x} \in \mathbb{R}^n$ represent states of the system at a given time t which is function of uncertain model parameter vector $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_m]^T \in \mathbb{R}^m$. Note that state $\mathbf{x}$ is a function of parameter $\boldsymbol{\Theta}$, i.e. $\mathbf{x} = \mathbf{x}(t, \boldsymbol{\Theta})$. The parameter $\boldsymbol{\Theta}$ is assumed to be time invariant and function of a random vector $\boldsymbol{\xi} = [\xi_1, \xi_2, \cdots, \xi_m]^T \in \mathbb{R}^m$ defined by a pdf $p(\boldsymbol{\xi})$ over the support Ω.

Based on the method of quadrature points, N^{th} order moment of state $\mathbf{x}$, at a given time t can be written as:

$$\mathcal{E}[\mathbf{x}^N] = \int_{\boldsymbol{\xi}} \mathbf{x}^N(\boldsymbol{\Theta}, t) p(\boldsymbol{\xi}) d\boldsymbol{\xi} \simeq \sum_{q}^{M} w_q \mathbf{x}^N(\boldsymbol{\Theta}(\boldsymbol{\xi}^q), t), \quad N = 1, 2, \cdots \tag{2.2}$$

where, M denotes total number of applied quadrature points and $\boldsymbol{\Theta}(\boldsymbol{\xi}^q) \in \mathbb{R}^{m \times 1}$ represents q^{th} quadrature point, generated based on applied quadrature scheme. Similarly, N^{th} order central moments of state $\mathbf{x}$ at each time t can be evaluated by shifting the quadrature points by the computed mean and then using Eq. (2.2).

Hence, the moments at any given time can be approximated as a weighted sum of the outputs of simulation runs initiated at the quadrature points generated from the initial uncertain parameter distribution. The natural choice for these quadrature points is the set of Gaussian quadrature points which is defined by choosing the points optimally in the sense of maximizing the degree of polynomial function that integrates exactly.

Different types of quadrature rules like classical Gaussian quadrature rule [41], Clenshaw-Curtis quadrature [42], and sparse grid [43] can be used to evaluate the mentioned integrals. In the following, we briefly describe few of these methods.

Gaussian Quadrature

The original $N-$point Gaussian quadrature rule is constructed such that it provides an exact result for integration of polynomials of degree $2N-1$ or less over the interval [-1,+1] by a suitable choice of the quadrature points ξ^i and weights w_i for $i = 1, ..., n$, i.e.

$$\int_{-1}^{1} f(\xi)d\xi \simeq \sum_{i=1}^{n} w_i f(\xi^i) \tag{2.3}$$

For integration over a general interval $[a, b]$, Gaussian quadrature can be written as:

$$\int_{a}^{b} f(y)dy \simeq (\frac{b-a}{2}) \sum_{i=1}^{n} w_i f((\frac{b-a}{2})\xi^i + (\frac{b+a}{2})) \tag{2.4}$$

where, ξ^is are corresponding quadrature points defined over interval $[-1, +1]$.

The integration problem can be reformulated in slightly more general form by introducing a positive weight function into the integral, which can be seen as a probability density function, i.e.

$$\int_{a}^{b} f(y)dy = \int_{a}^{b} \underbrace{\frac{f(y)}{p(y)}}_{g(y)} p(y)dy = \int_{a}^{b} g(y)p(y)dy \tag{2.5}$$

Different versions of Gaussian quadrature points are developed based on the type of probability density function $p(y)$ in Eq. (2.5). Table 2.1 illustrates different types of quadrature points which are being used based on the distribution $p(y)$.

Interval $[a, b]$	$p(y)$	Quadrature Points
$[-1, +1]$	Uniform	Gauss-Legendre quadrature
$[0, \infty)$	Exponential	GaussLaguerre quadrature
$[-\infty, +\infty)$	Normal	GaussHermite quadrature

Table 2.1: Applied Gaussian Quadrature points, based on distribution $p(y)$ [41]

Note that in an n-dimensional parameter space, the tensor product of 1-dimension quadrature points is used to generate quadrature points. As a consequence of this, the number of quadrature points increases exponentially as the number of input parameters increases.

Clenshaw-Curtis quadrature

The basic idea of Clenshaw-Curtis quadrature is to expand the integrand in terms of Chebyshev polynomials:

$$f(\xi) = \frac{a_0}{2} T_0(\xi) + \sum_{k=1}^{\infty} a_k T_k(\xi) \tag{2.6}$$

where, $T_k(\xi)$s are corresponding Chebyshev polynomials. Note that quadrature points in Clenshaw-Curtis scheme correspond to the exterma of the Chebyshev polynomials. Hence, given N basis function in Eq. (2.6) we will have N Clenshaw-Curtis quadrature points.

One should notice that given N quadrature points, Clenshaw-Curtis scheme exactly integrates polynomials upto degree $N-1$. Also, similar to Gaussian quadrature scheme, tensor product of 1-dimensional Clenshaw-Curtis quadrature points is used to generate corresponding quadrature points in higher dimensional.

Conjugate Unscented Transform

For a generic n-dimensional integral, the tensor product of 1-dimensional Gaussian quadrature points or Clenshaw-Curtis quadrature points results in an undesirable exponential growth of the number of points. Hence, using regular quadrature schemes will be computationally expensive in higher dimensions. In this dissertation, we have used Conjugate Unscented Transform (CUT) recently developed by Nagavenkat *et al.* [44–47], to overcome this drawback of regular quadrature points. The proposed CUT points are efficient in terms of accuracy while integrating polynomials and yet just employ a small fraction of the number of points used by the traditional Gaussian quadrature scheme.

Conjugate Unscented Transform(CUT) approach can be considered as an extension to the conventional Unscented Transformation method, see [48], by satisfying higher order moment constraint equations. The widely used Gaussian Quadratures can integrate all 1-Dimension polynomials of degree $2N-1$ with only N quadrature points. When extended to generic n-Dimensions, one would have to take a tensor product of 1-Dimension quadrature points totaling to N^n point, thus leading to an exponential growth of points with dimension. The main idea of the CUT approach is to judiciously select specific structures of *symmetric* points rather than taking the full tensor product of 1-D points. As a result, the obtained points can still exactly integrate polynomials of total degree $2N-1$ in n-dimensional space but with far lower number of points than N^n. To illustrate the Conjugate Unscented Transform approach in more detail, consider the problem of approximating the expected value of a function $f(\mathbf{x})$ in the following form:

$$\mathcal{E}[f(\mathbf{x})] = \int_{\Omega} f(\mathbf{x})p(\mathbf{x})d\mathbf{x} \simeq \sum_{i=1}^{N} w_i f(\mathbf{x}_{(i)}) \tag{2.7}$$

where, $\mathbf{x} = [x_1, x_2, \cdots, x_n]^T \in \mathbb{R}^n$ and $p(\mathbf{x})$ is a *uniform or Gaussian density function* defined over the domain $\Omega \subset \mathbb{R}^n$. Also, $\mathbf{x}_{(i)} \in \Omega$ are cubature points which are used to approximate the integral, and $w_i > 0$ are their corresponding scalar weights. Assuming that $f(\mathbf{x})$ has a valid Maclaurin series, given by:

$$f(\mathbf{x}) = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \cdots \sum_{n_n=0}^{\infty} \frac{x_1^{n_1} x_2^{n_2} \cdots x_m^{n_n}}{n_1! n_2! \cdots n_n!} \frac{\partial^{n_1+n_2+\cdots+n_n} f}{\partial x_1^{n_1} \partial x_2^{n_2} \cdots \partial x_n^{n_n}} \tag{2.8}$$

the expectation of $f(\mathbf{x})$ can be written as:

$$\mathcal{E}[f(\mathbf{x})] = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \cdots \sum_{n_m=0}^{\infty} \frac{\mathcal{E}[x_1^{n_1} x_2^{n_2} \cdots x_m^{n_m}]}{n_1! n_2! \cdots n_m!} \frac{\partial^{n_1+n_2+\cdots+n_m} f}{\partial x_1^{n_1} \partial x_2^{n_2} \cdots \partial x_n^{n_n}} \tag{2.9}$$

Hence, the problem of evaluating the expected value of $f(\mathbf{x})$ is reduced to computing higher order moments of random vector $\mathbf{x}$ according to the pdf $p(\mathbf{x})$. Now, substitution of Eq. (2.8) into Eq. (2.7) leads to the following expression:

$$\sum_{i=1}^{N} w_i f(\mathbf{x}_{(i)}) = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \cdots \sum_{n_n=0}^{\infty} \frac{\left(\sum_{i=1}^{N} w_i \prod_{k=1}^{n} x_{(i,k)}^{n_k}\right)}{n_1! n_2! \cdots n_n!} \frac{\partial^{n_1+n_2+\cdots+n_m} f}{\partial x_1^{n_1} \partial x_2^{n_2} \cdots \partial x_n^{n_n}}(\mathbf{0}) \tag{2.10}$$

Comparing Eq. (2.10) and Eq. (2.9) results in a set of algebraic equations known as *Moment Constraint Equations (MCE)*:

$$\mathcal{E}[x_1^{n_1} x_2^{n_2} \cdots x_N^{n_n}] = \sum_{i=1}^{N} w_i \prod_{k=1}^{n} x_{(i,k)}^{n_k} \tag{2.11}$$

where $x_{(i,k)}^{n_k}$ is the k^{th} coordinate of the i^{th} point $\mathbf{x}_{(i)}$. Notice that the left hand side of the aforementioned equation contain actual moments of input parameter density function while right hand side is function of unknown position of quadrature points. The CUT methodology involves finding quadrature points such that they satisfy Eq. (2.11) upto a desired order of moments. Assuming $p(\mathbf{x})$ to be symmetric, the cubature points are preferred to lie symmetrically on following appropriately defined directions:

- *Principal Axes:* Generally in a n-dimensional cartesian space, there exists n orthogonal coordinate axes centered at the origin which correspond to eigen-vectors of the covariance of input random variable. These axes are called *principal axes* which are denoted by σ. Also, the corresponding point on these axes are shown by $\sigma_j, j = 1, 2, \cdots, 2n$.

- *Conjugate Axes:* Conjugate axes denoted by c^P $(P \leq n)$, are the axes constructed from all the combinations and sign permutations of the set of principal axes taken P at a time. These points are shown by c_i^P, where $i = 1, 2, \cdots, 2^P \begin{pmatrix} n \\ P \end{pmatrix}$.

- *Scaled Conjugate Axes:* The remainder of the cubature points are found from n^{th}-*Scaled Conjugate axes* which are constructed from all the combinations including sign permutations of the set of principal axes such that in every combination exactly one principal axis is scaled by a scaling parameter h. These set of axes are labeled as $s^n(h)$, and the points are listed as $s_i^n(h)$ where $i = 1, 2, \cdots, n2^n$.

Table 2.2 illustrates a sample point for each of explained points in n dimension. As well, schematic view of the mentioned points in 3 dimensional space can be seen in Fig. 2.3.

Table 2.2: Different types of CUT points defined in n-dimensional space

Type	Sample Point	Number of Points
σ	$(1, 0, 0, \cdots, 0)$	$2n$
c^P	$(\underbrace{1, 1, \cdots, 1}_{n}, \underbrace{0, 0, \cdots, 0}_{n-P})$	$2^P \begin{pmatrix} n \\ P \end{pmatrix}$
$s^n(h)$	$(h, 1, 1, \cdots, 1)$	$n2^n$

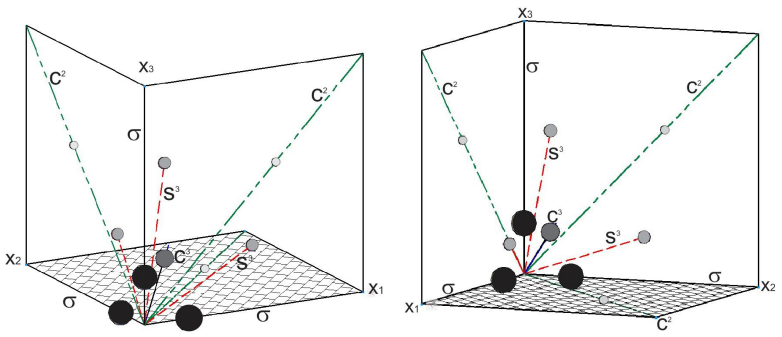


Figure 2.3: Cubature points in 3 dimensional cartesian space in different views

The next step is to select different combinations of mentioned points. It should be noted that all the selected points on the same set of symmetric axes should be equidistant from origin and should have equal weight. For each selected point, two unknown variables w_i

and r_i are assigned which are weight and scaling variable, respectively. Finally, the moment constraints equations for desired order of moments are derived in terms of unknown variables r_i's and w_i's. Solutions of this set of equations results in the values of these variables. Note that due to symmetrical properties of cubature points, the odd order moment constraints equations are automatically satisfied. Hence, variables w_i's and r_i's can be found by solving just the even order moment constraints equations. Note that different sets of cubature points can be found depending on the order of moment constraint equations and dimension n. In Refs. [44, 49], CUT scheme is employed to compute quadrature points of different order for Gaussian and uniform density functions.

Fig. 2.4 represents the number of 8^{th} order quadrature points required by different quadrature schemes (CUT, Gauss-Legendre, Clenshaw-Curtis and Sparse Grid) for a uniformly distributed random vector versus the dimensionality of the random vector. From this figure, it is clear that the growth of number of quadrature points with increase in dimensionality according to the CUT methodology is much lower as compared to the Gauss-Legendre and Clenshaw-Curtis. Furthermore, it is apparent that the CUT methodology requires less than one half of quadrature points as required by the sparse grid Smolyak approach. For example, 59 CUT quadrature points are required to satisfy 8^{th} order moments in 3-dimensional space as compared to 165 Sparse Grid quadrature points, 125 Gauss-Legendre quadrature points and 729 Clenshaw-Curtis quadrature points. Hence, the CUT methodology can be very useful in reducing the number of numerical model runs which needs to be performed for accurate computation of prior statistics.

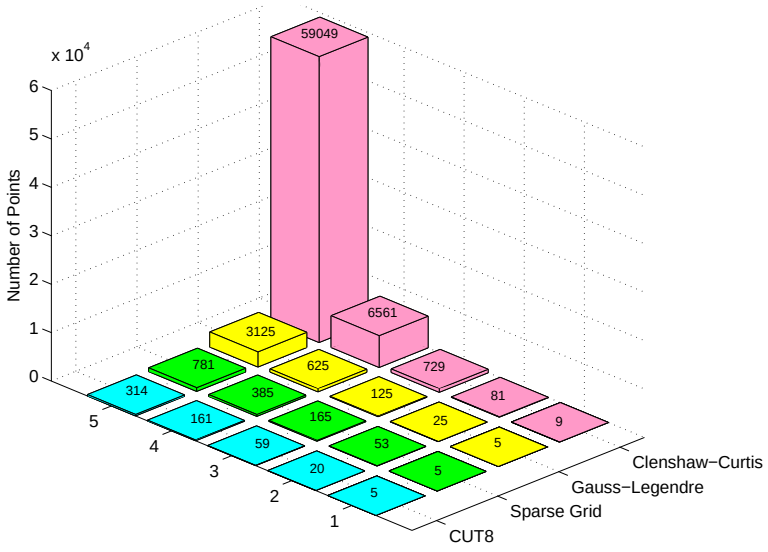


Figure 2.4: Comparison of number of 8^{th} order quadrature points required according to different quadrature scheme versus dimension of random variable.

2.3 Approximation of Probability Maps

The problem of generating hazard maps corresponds to computing the probability of Quantity of Interest (QOI), such as the amount of ash present in the atmosphere at a given geographical location, given the probability distribution for model input parameters. The accurate computation of probabilistic hazard map requires forward propagation of variability in model input parameters, in order to compute the probability of the QOI at a specified place and time. A simplistic approach to compute hazard maps entails running numerical simulations with a range of input values, and computing the relative frequency of a QOI. Unfortunately, a large number of realizations, depending on number of uncertain parameters/initial conditions and model nonlinearities, are generally required to get a good convergence in probability for the QOI using Monte Carlo simulations. This computational load renders this simplistic approach impractical for many dynamic models. Instead one needs a more judiciously chosen method for computing probabilities, recognizing the potential for a trade-off between computational efficiency and the accuracy of probability computations.

In this work, we follow an approach outlined in Dalbey et. al. [39] in which the generalized Polynomial Chaos (gPC) methodology was employed to create a fast, computationally cheap *polynomial surrogate model* , which is used to evaluate a large number of samples at

minimal computational cost. In the standard gPC methodology, Galerkin collocation is used to generate a system of deterministic differential equations for the expansion coefficients. The Galerkin collocation step fails when applied to problems with non-polynomial non-linearities, and can produce non-physical solutions when applied to hyperbolic equations. Non-intrusive spectral projection (NISP) or stochastic collocation methods can overcome these difficulties [50–52]. A different formulation of the NISP idea [39] known as polynomial chaos quadrature (PCQ) is used here. PCQ replaces the projection step of NISP with numerical quadrature. Thus our approach for computing the probability for a QOI involves (1) computing coefficients of the polynomial surrogate model according to the PCQ formulation; (2) sampling the surrogate at a large number of inputs at minimal computational cost. Let us decribe this approach in more detail.

2.3.1 Polynomial Chaos Quadrature

Let $\mathbf{x}(t, \boldsymbol{\Theta}) \in \mathbb{R}^n$ represent a vector of n quantities of interest which is a function of the uncertain model parameter vector $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_m]^T \in \mathbb{R}^m$. For example, in an ash transportation problem due to volcanic eruption, the vector $\mathbf{x}$ might represent the height at the top of an ash cloud and/or the ash concentration, at a specified geographical location, and the parameter vector $\boldsymbol{\Theta}$ might contain volcano source parameters like vent size, particle velocity at the vent, and grain size distribution. Parameter vector $\boldsymbol{\Theta}$ is assumed to be time invariant, and a function of a standardized random vector $\boldsymbol{\xi} = [\xi_1, \xi_2, \cdots, \xi_m]^T \in \mathbb{R}^m$ defined by a pdf $p(\boldsymbol{\xi})$ with support Ω. For instance, the uncertain model parameter vector $\boldsymbol{\Theta}$ can be assumed to be uniformly distributed random vector which lies in the range:

$$\mathbf{a} \leq \boldsymbol{\Theta} \leq \mathbf{b} \tag{2.12}$$

where, $a, b \in \mathbb{R}^m$ are constant vectors. Hence, $\boldsymbol{\Theta}$ can be written as a function of $\boldsymbol{\xi}$ consisting of m standardized uniform random variables between -1 and 1:

$$\theta_j = \frac{a_j + b_j}{2} + \frac{b_j - a_j}{2}\xi_j, \quad j = 1, 2, \cdots, m \tag{2.13}$$

If $\boldsymbol{\Theta}$ is assumed to be Gaussian random vector with prescribed mean and covariance matrix, then $\boldsymbol{\xi}$ can be a vector of Gaussian random variables with zero mean and identity covariance. Note that $\boldsymbol{\Theta}$ is not restricted to have uniform or Gaussian distribution. Ideally, it can have

any prescribed distribution. Now, the QOI (e.g. ash top-height at a geolocation) can be approximated as a linear combination of $N+1$ polynomial functions of $\boldsymbol{\xi}$:

$$x_i(t,\boldsymbol{\Theta}) = \sum_{k=0}^{N} x_{i_k}(t)\phi_k(\boldsymbol{\xi}) \tag{2.14}$$

where, $\phi_k(\boldsymbol{\xi})$ are orthogonal polynomial basis function set with respect to $p(\boldsymbol{\xi})$. One can use the Gram-Schmidt orthogonalization to compute these basis function. Note that the total number of terms in the expansion is $N+1$ and is determined by the chosen highest order (l) of basis polynomials $\phi_k(\boldsymbol{\xi})$ and the dimension of uncertain parameter vector $\boldsymbol{\Theta}$ and it is given as

$$N+1 = \frac{(l+m)!}{m!l!} \tag{2.15}$$

In general, according to the PCQ methodology, the uncertain QOI, $\mathbf{x}(t,\boldsymbol{\Theta})$ and model parameter $\boldsymbol{\Theta}$ can be written as a linear combination of orthogonal polynomial basis functions, $\phi_k(\boldsymbol{\xi})$, which span the space of random variables $\boldsymbol{\xi} = [\boldsymbol{\xi}_1, \cdots \boldsymbol{\xi}_m]^T$ and results in following polynomial surrogate model:

$$x_i(t,\boldsymbol{\Theta}) = \sum_{k=0}^{N} x_{i_k}(t)\phi_k(\boldsymbol{\xi}) = \mathbf{x}_i^T(t)\boldsymbol{\Phi}(\boldsymbol{\xi}) \Rightarrow \mathbf{x}(t,\boldsymbol{\xi}) = \mathbf{X}_{pc}(t)\Phi(\boldsymbol{\xi}), \quad i = 1,2,\cdots,n \tag{2.16}$$

$$\theta_j(\boldsymbol{\xi}) = \sum_{k=0}^{N} \theta_{j_k}\phi_k(\boldsymbol{\xi}) = \boldsymbol{\theta}_j^T\boldsymbol{\Phi}(\boldsymbol{\xi}) \Rightarrow \boldsymbol{\Theta}(t,\boldsymbol{\xi}) = \boldsymbol{\Theta}_{pc}\Phi(\boldsymbol{\xi}), \quad j = 1,2,\cdots,m \tag{2.17}$$

Here $\mathbf{X}_{pc}$ and $\boldsymbol{\Theta}_{pc}$ are matrices composed of coefficients of the PC expansion for $\mathbf{x}$ and $\boldsymbol{\Theta}$. The coefficients θ_{i_k} are obtained by making use of the *normal equation*:

$$\theta_{i_k} = \frac{\mathcal{E}[\theta_i(\boldsymbol{\xi})\phi_k(\boldsymbol{\xi})]}{\mathcal{E}[\phi_k(\boldsymbol{\xi})\phi_k(\boldsymbol{\xi})]} \tag{2.18}$$

In this expression, the expected value of a sufficiently smooth function $u(\boldsymbol{\xi})$ is defined as:

$$\mathcal{E}[u(\boldsymbol{\xi})] = \int u(\boldsymbol{\xi})p(\boldsymbol{\xi})d\boldsymbol{\xi} \tag{2.19}$$

Similarly, the coefficients of x_{i_k}'s can be found from:

$$x_{i_k} = \frac{\mathcal{E}[x_i(t,\boldsymbol{\theta}(\boldsymbol{\xi}))\phi_k(\boldsymbol{\xi})]}{\mathcal{E}[\phi_k(\boldsymbol{\xi})\phi_k(\boldsymbol{\xi})]} \tag{2.20}$$

In our calculations, numerical quadrature replaces exact integration. Specifically, the integrals in Eq. (3.29) and Eq. (2.20) can be written as:

$$\mathcal{E}[\phi_i(\boldsymbol{\xi})\phi_j(\boldsymbol{\xi})] = \int \phi_i(\boldsymbol{\xi})\phi_j(\boldsymbol{\xi})p(\boldsymbol{\xi})d\boldsymbol{\xi} \simeq \sum_{q=1}^{M} w_q \phi_i(\boldsymbol{\xi}^q)\phi_j(\boldsymbol{\xi}^q) \tag{2.21}$$

$$\mathcal{E}[x_i(t,\boldsymbol{\theta}(\boldsymbol{\xi}))\phi_j(\boldsymbol{\xi})] = \int x_i(t,\boldsymbol{\theta}(\boldsymbol{\xi}))\phi_j(\boldsymbol{\xi})p(\boldsymbol{\xi})d\boldsymbol{\xi} \simeq \sum_{q=1}^{M} w_q x_i(t,\boldsymbol{\xi}^q)\phi_j(\boldsymbol{\xi}^q) \tag{2.22}$$

$$\mathcal{E}[\theta_i(\boldsymbol{\xi})\phi_j(\boldsymbol{\xi})] = \int \theta_i(\boldsymbol{\xi})\phi_j(\boldsymbol{\xi})p(\boldsymbol{\xi})d\boldsymbol{\xi} \simeq \sum_{q=1}^{M} w_q \theta_i(\boldsymbol{\xi}^q)\phi_j(\boldsymbol{\xi}^q) \tag{2.23}$$

Notice that $x_i(t,\boldsymbol{\xi}^q)$ represent the quantity of interest at time t with model parameter vector being evaluated at $\boldsymbol{\xi}^q$, where $\boldsymbol{\xi}^q$ corresponds to quadrature value of parameter vector $\boldsymbol{\xi}$. That is, numerical model (expressed by Eq. (2.1)) is solved for each input parameter vector $\boldsymbol{\xi}^q$, and the QOI is then computed from these simulations. The resulting method can be viewed as a "smart" MC-like evaluation of the model equations, with sample points selected by quadrature rules. Now, instead of performing intensive simulations, the polynomial surrogate model Eq. (2.16) can be substituted in order to calculate the probability of the QOI at a given location.

To summarize, given a specific location, the following algorithm can be used to compute a hazard map for a QOI:

- Step 1: In the space of random variables, generate sampling points as combinations of input parameters, treated as random variables, corresponding to the selected quadrature scheme;
- Step 2: Perform a simulation at each sample point using the numerical model to generate a map of the QOI, as a function of position;
- Step 3: Use Eq. (2.20) to compute the PC expansion coefficients corresponding to the QOI, for each location;
- Step 4: Choose a large set of secondary sample points in the stochastic space, generated from to the probability density function $p(\xi)$.
- Step 5: Compute the QOI for each secondary sample point from the surrogate model.
- Step 6: Given the threshold $thresh$, compute the the relative frequency of QOI.

- Step 7: Computed frequency is an approximation for probability of QOI to be greater than or equal to $thresh$, denoted by $Pr(QOI \geq thresh)$.

2.4 Numerical Simulations

In this section, we provide two different examples to illustrate performance of the proposed methodology in this chapter. The first example is about simulation of atmospheric dispersion of Propane in neighborhood of New York city, while the second example illustrates dispersion of Chlorine in neighborhood of city of Denver, Co. A Second-order Closure Integrated PUFF (SCIPUFF) [53] is used to simulate atmospheric advection/dispersion for both examples. In the following, we first briefly describe applied SCIPUFF numerical model and then numerical simulations for each example are presented.

2.4.1 SCIPUFF Numerical Model

SCIPUFF [53] is a Lagrangian transport and diffusion model for atmospheric dispersion applications. The acronym SCIPUFF stands for Second-order Closure Integrated PUFF and describes two basic aspects of the model. SCIPUFF uses a Gaussian puff representation for the concentration field of a dispersing contaminant. A three dimensional Gaussian is completely described by its spatial integral moments up to second-order, and can be written in the form:

$$c(x) = \frac{Q}{2\pi^{3/2}[det(\sigma)]^{1/2}} exp[-\frac{1}{2}\sigma_{ij}^{-1}(x_i - \bar{x}_i)(x_j - \bar{x}_j)] \tag{2.24}$$

The specific Gaussian variation mentioned in Eq.2.24 applies to an individual puff, but in general the local concentration field will be composed of a sum of contributions from a number of such puffs.

The advection-diffusion equation for a scalar quantity in an incompressible flow field can be written as:

$$\frac{\partial c}{\partial t} + \frac{\partial}{\partial x_i}(u_i \bar{c}) = k\nabla^2 c + S \tag{2.25}$$

where $u_i(x,t)$ is the turbulent velocity field, k is the molecular diffusivity, and S represents the source terms.

The atmospheric velocity field is generally turbulent, so the Reynolds averaging technique is used to define a mean and a turbulent fluctuation value. Denoting the mean by an overbar and the fluctuation by a prime, we have $u = \bar{u} + u'$, and a similar decomposition can be applied to the concentration, c. The Reynolds averaged conservation equation for the mean scalar concentration is thus:

$$\frac{\partial \bar{c}}{\partial t} + \frac{\partial}{\partial x_i}(\bar{u}_i \bar{c}) = -\overline{u_i' c'} \frac{\partial \bar{c}}{\partial x_i} + k\nabla^2 \bar{c} + \bar{S} \tag{2.26}$$

The dispersion of any species in a turbulent velocity field is a random process since the turbulent fluctuations are effectively chaotic and cannot be measured or predicted in detail. The scalar concentration is therefore a stochastic quantity, with a probability distribution that depends on the distribution of velocity fluctuations. Traditional deterministic estimates of atmospheric dispersion only provide a single concentration value as a function of space and time, and this corresponds to the mean value, $\bar{c}$, for some definition of the statistical ensemble. The mean value is the first moment of the probability distribution, and contains no information about the statistical variability in the prediction. Higher moments are required to give a quantitative description of the variability. The probabilistic aspect of the SCIPUFF dispersion prediction is based on a transport equation for the statistical variance in the concentration value, that is the second moment of the probability distribution. The concentration fluctuation variance equation can be obtained from the scalar mass conservation equation in the form:

$$\frac{\partial \overline{c'^2}}{\partial t} + \frac{\partial}{\partial x_i}(\bar{u}_i \bar{c'^2}) = -2\overline{u_i' c'} \frac{\partial \bar{c}}{\partial x_i} - \frac{\partial}{\partial x_i}(\overline{u_i^2 c'^2}) - 2k(\frac{\partial \bar{c'}}{\partial x_i})^2 + k\nabla^2 \bar{c'^2} \tag{2.27}$$

We emphasize in here that through out this dissertation, we assume that we have perfect knowledge about the applied wind-field and all the turbulence in the wind-field in assumed to be negligible. For detailed information regarding SCIPUFF numerical model, please see [53].

2.4.2 Test Case 1

Consider dispersion/advection of propane over New York city area. The domain of interest and the applied wind-field (at one specific time) are shown in Fig. 2.5. Simulation time is

considered to be 24 $hrs.$, starting from 00 : 00 of September 1^{st}, 2013. North American Regional Reanalysis wind data at pressure level 100 kpa (height $\simeq$ 100 $m.$) is used as the corresponding wind-field for simulation. Three instantaneous mass releases are considered where their location is known and the only uncertain parameters are their amount of mass release. It is assumed that releases happen simultaneously at 00 : 00 of September 1^{st}. All mass releases are assumed to be uniformly distributed between 100 kg and 300 kg. Fig. 2.5 illustrates source locations and a snapshot of the wind-field (at $t = 0$ $hrs.$) over the two dimensional spatial domain.

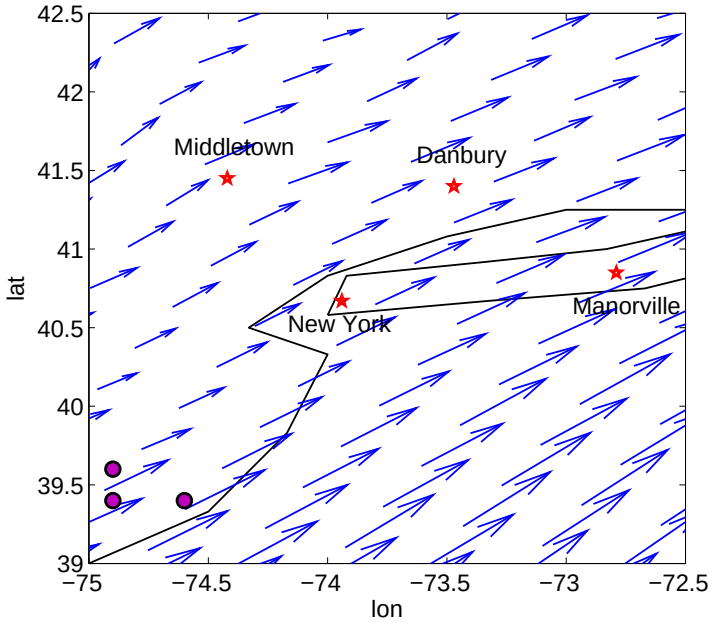
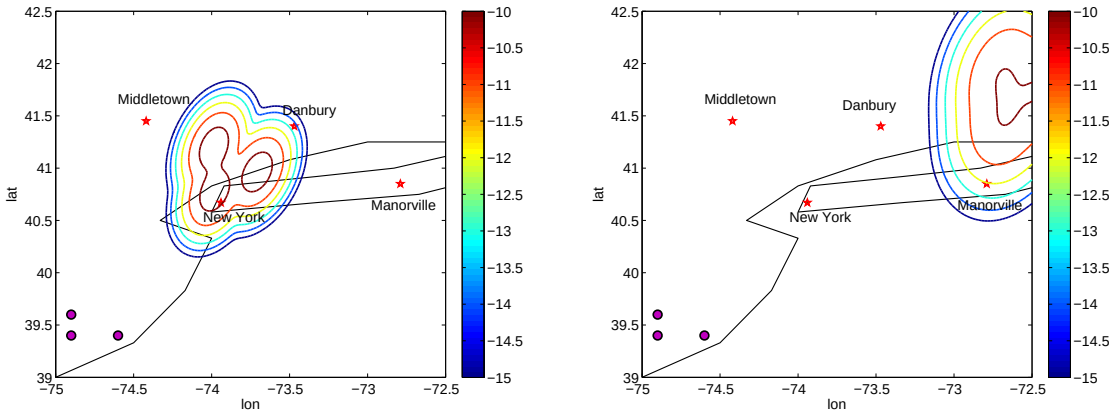


Figure 2.5: Schematic layout of Propane release over New York region, source locations are shown with purple circles, the wind-field (at $t = 00 : 00$ $hr.$ and pressure level = 100 kPa) is shown over the two dimensional domain with blue vector field. Geographical location of different cities can be seen on the background map

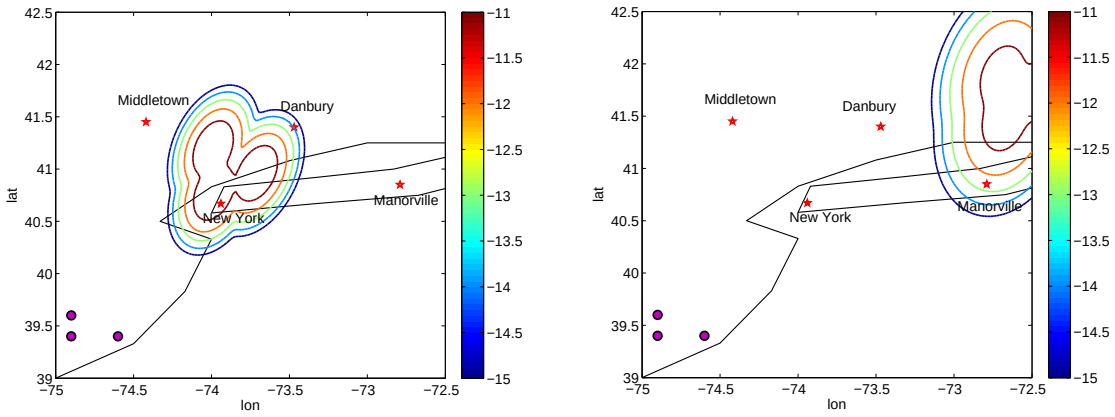
A set of 59 eighth order CUT quadrature points (CUT8) are used to quantify the uncertainty involved in concentration of dispersal material. Simulation of dispersion/advection has been performed using SCIPUFF numerical model, where concentration of propane is recorded every 10 $mins$. Applied CUT8 points are used to evaluate expected value and variance of concentration field (denoted by c) at each time step. Fig. 2.6 illustrates logarithm

of expected value of concentration field ($log(\mathcal{E}[c])$) at 12 $hrs.$ and 24 $hrs.$ after release time. Note that logarithm of expected value has been used to better illustrate variability of concentration field over spatial domain. Similarly, Fig. 2.7 illustrates logarithm of standard deviation of concentration field , denoted by $log(std(c))$, at 12 and 24 $hrs.$ after release time.



(a) $log(\mathcal{E}[c])$ at time $12:00$ $hr.$ (b) $log(\mathcal{E}[c])$ at time $24:00$ $hr.$

Figure 2.6: Logarithm of Expected value of concentration field at a) $t = 12:00$ $hr.$, b) $t = 24:00$ $hr.$ after release instant.

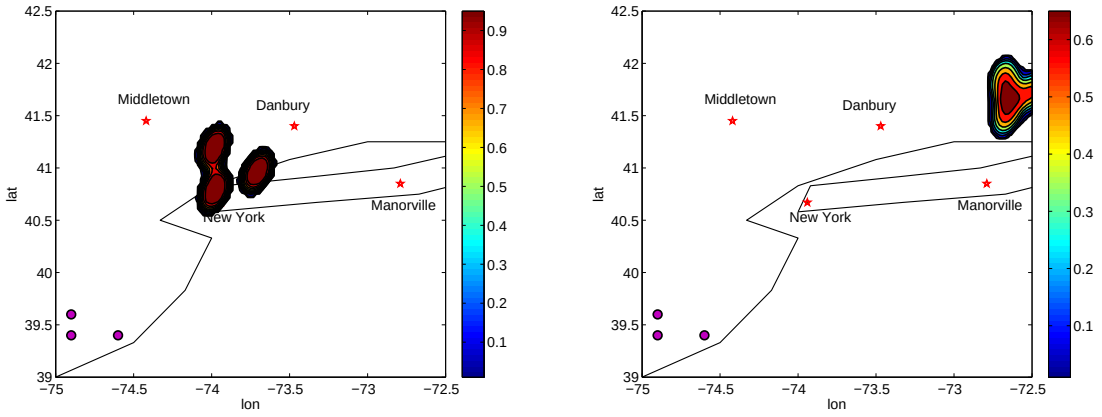


(a) $log(std(c))$ at time $12:00$ $hr.$ (b) $log(std(c))$ at time $24:00$ $hr.$

Figure 2.7: Logarithm of Standard deviation of concentration field at a) $t = 12:00$ $hr.$, b) $t = 24:00$ $hr.$ after release instant.

Probability of concentration to be greater than or equal to $thresh = 10^{-9} gr/m^3$ is shown in Fig. 2.8. We have used 5^{th} order polynomial chaos expansion for approximation of concentration field at each spatial location. Hence, there are totally $\frac{(5+3)!}{5!3!} = 56$ terms in polynomial

chaos expansion of concentration field whose coefficients have been found using Eq. (2.20). Also, a number of 5×10^4 secondary sample points of $\boldsymbol{\xi} \in \mathbb{R}^3$ are used for calculation of corresponding probability maps.



(a) $Pr(c > 10^{-9})$ at time $12:00\ hr.$ (b) $Pr(c > 10^{-9})$ at time $24:00\ hr.$

Figure 2.8: $Pr(c > 10^{-9})$ a) $t = 12:00\ hr.$, b) $t = 24:00\ hr.$ after release instant.

2.4.3 Test Case 2

A more complicated scenario is being simulated as the last example. For simulation purposes, we considered dispersion/advection of Chlorine over a *three dimensional* spatial domain in neighborhood of city of Denver, Co. Simulation time is considered to be 10 $hrs.$, starting from $12:00$ of June 6^{th}, 2012. Three dimensional North American Regional Reanalysis wind data and a $33 \times 46 \times 20$ grid points were used to perform simulations over the spatial domain. In our simulation, height variation is between 0 and 1000 m and a discretization of 50 m in vertical direction is performed to consider the effect of the wind-field at different heights. Fig. 2.4.3 illustrates projection of uncertain source location and a snapshot of the wind-field (at $height = 100\ m.$ and $t = 12:00$ of June 6^{th}) over the two dimensional spatial domain.

Uncertain source parameters are assumed to be source location $(lon_{src}, lat_{src}, z_{src})$ and mass release rate of chlorine (m_{src}). Source location is assumed to be *uniformly* distributed over the region $(lon, lat, z) \in ([-103.8, -103.5] \times [39.45, 39.65] \times [300, 400])$. The mass release rate of chlorine is considered to be uniformly distributed between 6 kg/min and 10 kg/min

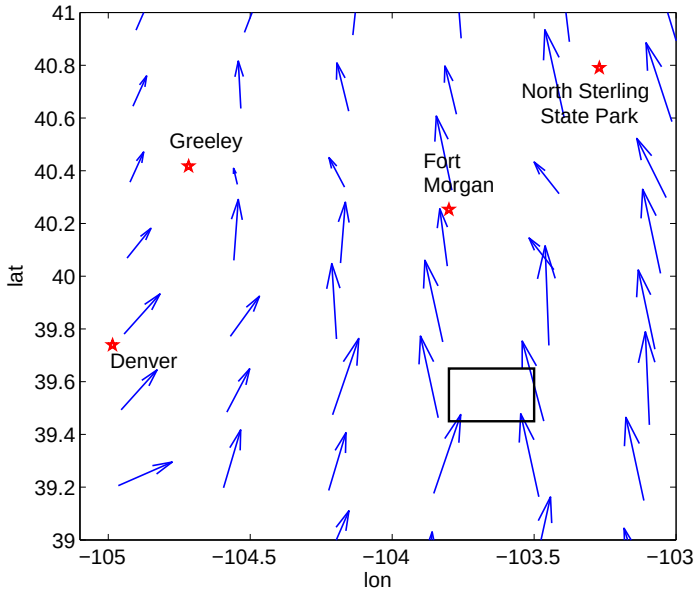


Figure 2.9: Schematic layout of Chlorine release over Denver region. Applied wind-field (at initial time and pressure level = 100 kPa) is shown over the two dimensional domain with blue vector field. Black rectangle illustrates projection of uncertain source location over lon-lat plane. Geographical location of different cities can be seen on the background map.

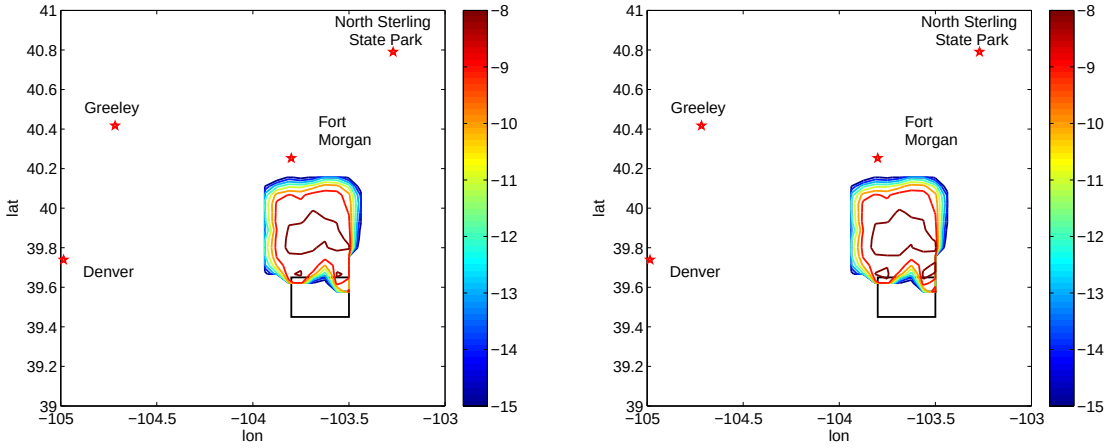
and total release time is 1 hr. Table 2.3 illustrates uncertain parameters for Test Case 2.

Table 2.3: Uncertain parameters considered for dispersion of Chlorine over city of Denver, Co.

Parameter	Distribution/Range
lon_{src}	Uniform $\in [-103.8, -103.5]$
lat_{src}	Uniform $\in [39.45, 39.65]$
z_{src}	Uniform $\in [300, 400]$ $m.$
m_{src}	Uniform $\in [4, 10]$ kg/min

A set of 161 CUT8 quadrature points quadrature points are used to quantify the uncertainty involved in concentration of pollutant material. Fig. 2.10 illustrates logarithm of expected value of concentration field, denoted by $log(\mathcal{E}[c])$, at $t = 1$ $hr.$ after release time

and at two different heights. We have used logarithm of expected value to better illustrate the variability of concentration field over spatial domain. Similarly, Fig. 2.11 illustrates logarithm of expected value of concentration field ($log(\mathcal{E}[c])$) at 10 $hr.$ after release time and at $z = 100\ m$ and $z = 500\ m$. Comparison of Fig. 2.10 and Fig. 2.11 shows that the difference between expected value of concentration field at $z = 100\ m.$ and $z = 500\ m$ is almost negligible.

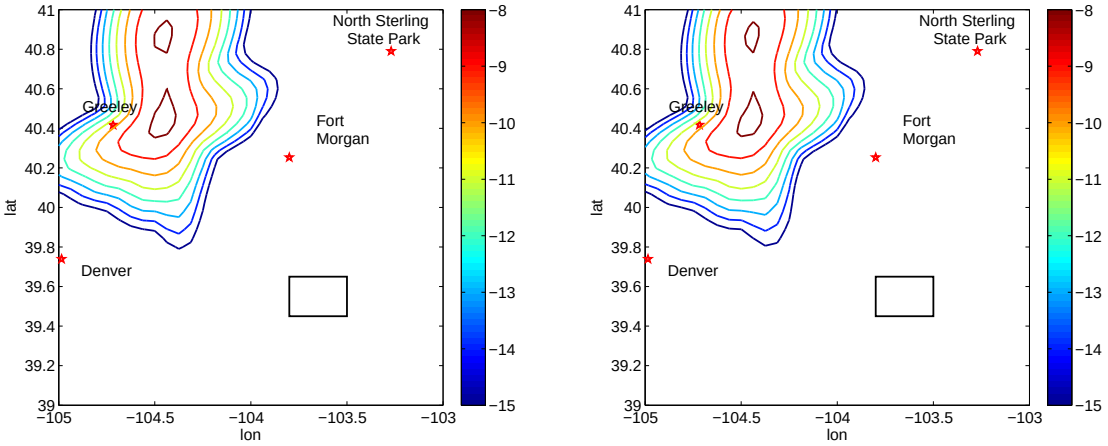


(a) $log(\mathcal{E}[c])$ at time $1:00\ hr.$ and $z = 100\ m.$ (b) $log(\mathcal{E}[c])$ at time $1:00\ hr.$ and $z = 500\ m.$

Figure 2.10: Logarithm of Expected value of concentration field at $t = 1:00\ hr.$ after release at different heights a) $z = 100\ m.$, b) $z = 500\ m.$

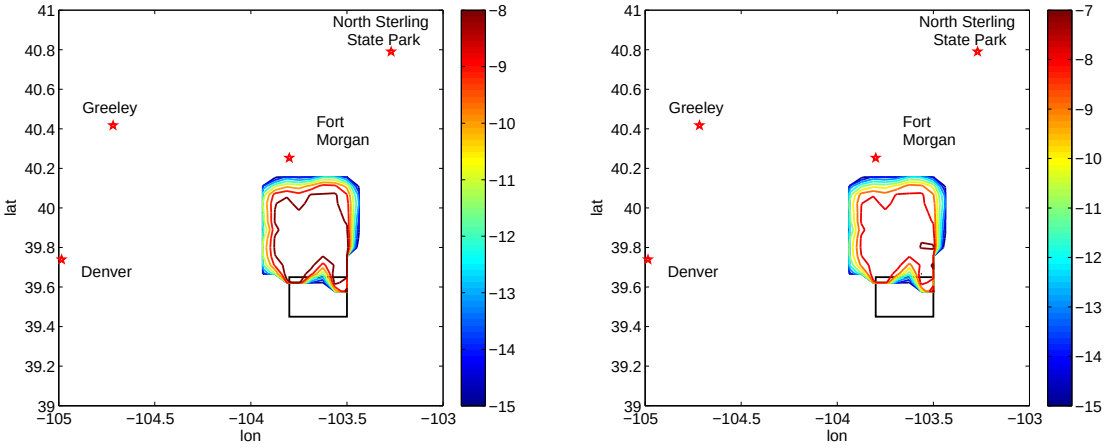
Fig. 2.12 and Fig. 2.13 represent logarithm of standard deviation of concentration field, denoted by $log(std(c))$, at 1 $hr.$ and 10 $hr.$ after release time. Similar to expected value of concentration field, comparison of Fig. 2.12 and Fig. 2.13 shows that the difference between standard deviation of concentration field at $z = 100\ m.$ and $z = 500\ m$ is almost negligible.

Fig. 2.14 and Fig. 2.15 show probability of concentration to be greater than or equal to $thresh = 10^{-10} gr/m^3$ over spatial domain and at different times. We have used 6^{th} order polynomial chaos expansion for approximation of concentration field at each spatial location. Hence, there are totally $\frac{(6+4)!}{6!4!} = 210$ terms in polynomial chaos expansion of concentration field whose coefficients have been found using Eq. (2.20). A number of 5×10^4 secondary sample points of $\boldsymbol{\xi} \in \mathbb{R}^4$ were used for calculation of corresponding probability maps.



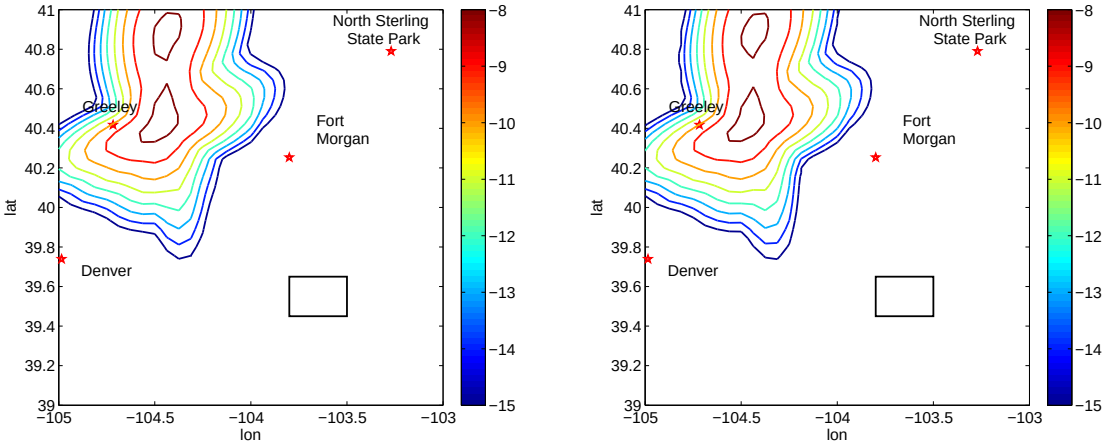
(a) $log(\mathcal{E}[c])$ at time $10:00$ $hr.$ and $z = 100$ $m.$ (b) $log(\mathcal{E}[c])$ at time $10:00$ $hr.$ and $z = 500$ $m.$

Figure 2.11: Logarithm of Expected value of concentration field at $t = 10:00$ $hr.$ after release at different heights a) $z = 100$ $m.$, b) $z = 500$ $m.$ Black rectangle represents projection of uncertain source location over lon-lat plane.



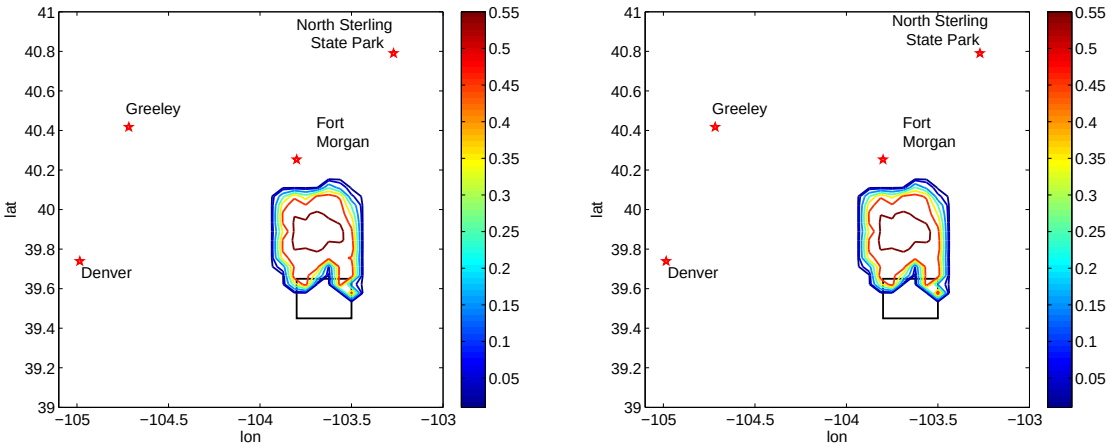
(a) $log(std(c))$ at time $1:00$ $hr.$ and $z = 100$ $m.$ (b) $log(std(c))$ at time $1:00$ $hr.$ and $z = 500$ $m.$

Figure 2.12: Logarithm of standard deviation of concentration field at $t = 1:00$ $hr.$ after release at different heights a) $z = 100$ $m.$, b) $z = 500$ $m.$ Black rectangle represents projection of uncertain source location over lon-lat plane.



(a) $log(std(c))$ at time $10:00\ hr.$ and $z = 100\ m.$ (b) $log(std(c))$ at time $10:00\ hr.$ and $z = 500\ m.$

Figure 2.13: Logarithm of standard deviation of concentration field at $t = 10:00\ hr.$ after release at different heights a) $z = 100\ m.$, b) $z = 500\ m.$ Black rectangle represents projection of uncertain source location over lon-lat plane.

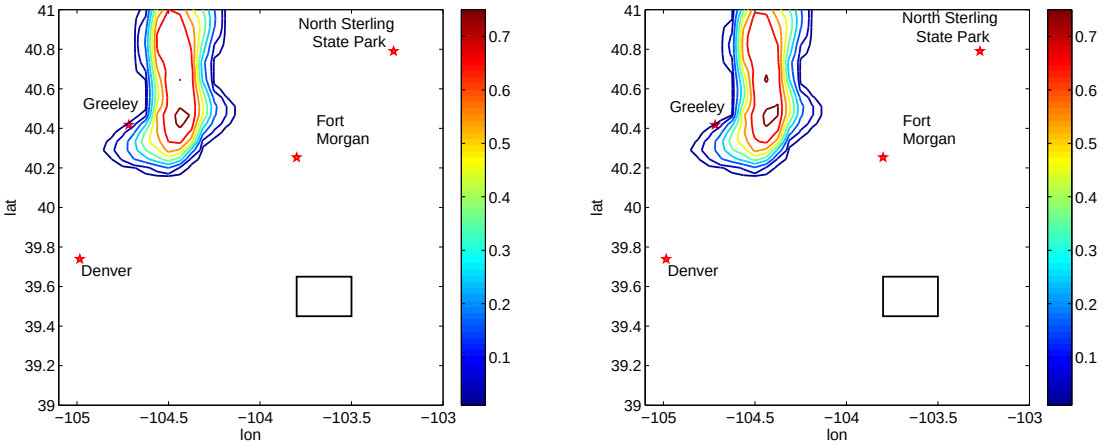


(a) $z = 100\ m.$ (b) $z = 500\ m.$

Figure 2.14: $Pr(c > 10^{-10})$ at $t = 1:00\ hr.$ after release at a) $z = 100\ m.$, b) $z = 500\ m.$ Black rectangle represents projection of uncertain source location over lon-lat plane.

2.5 Summary

In this chapter, we presented mathematical details for uncertainty quantification of dynamical systems in presence of parametric uncertainty. In Section 2.2 we discussed the method



(a) $z = 100\ m.$ (b) $z = 500\ m.$

Figure 2.15: $Pr(c > 10^{-10})$ at $t = 10:00\ hr.$ after release at a) $z = 100\ m.$, b) $z = 500\ m.$ Black rectangle represents projection of uncertain source location over lon-lat plane.

of quadrature points for evaluation of *prior* statistics of model output due to presence of uncertain parameters. As well, we utilized a polynomial chaos surrogate model to approximate hazard maps of QOI in Section 2.3. The major benefit of utilized technique for approximation of hazard maps lies in its considerable computational performance with respect to simple Monte Carlo approach. We verified performance of the method of quadrature points and introduced technique for approximation of hazard map by two numerical simulations. Simulation results for both examples demonstrates computational advantage and accuracy of discussed methods. Note that we used a recently developed Conjugate Unscented Transform technique to generate corresponding quadrature points for our numerical simulations. The major benefit of CUT methodology is that it employs small fraction of quadrature points used by traditional Gaussian quadrature scheme, while being efficient and accurate in evaluation of integrals.

In the next Chapter, we discuss about fusion of *prior* statistic, evaluated by the help of the method of quadrature points, with measurement data, which is obtained from data observation sensors. wind-field

Chapter 3

Source Parameter Estimation

3.1 Introduction

In the previous chapter, the method of quadrature points was discussed in detail as a tool to evaluate prior statistics of the model output at every time, given parametric or initial condition uncertainty. The use of observational data to refine the dynamical model forecast so as to reduce the associated uncertainty is a logical improvement over purely model-based predictions. Hence, one can fuse both model forecasts and observational data to improve the level of confidence in uncertain parameters and model forecasts in future time steps.

Different algorithms exist in literature that are used to fuse model forecast information and observational data for better estimation of uncertain parameters. In the context of plume dispersion phenomenon, these algorithms can be divided into two different categories: optimization based methods and probabilistic methods. The main idea of optimization based methods is to find the source location in such a way that it minimizes the difference between the observed data and model prediction [54–57]. Haupt *et al.* [58] presented a Genetic Algorithm based method for source characterization in plume dispersion phenomenon. In her method, source terms are obtained by minimizing the normalized difference between the measurement data and model predicted data. Allen *et al.* [59] applied the method developed by Haupt to the cases with noisy measurement. In another research, Allen *et al.* [54] extended his Genetic Algorithm based approach to characterize the source terms along estimation of wind direction. Similarly, Long *et al.* [60], presented a Genetic Algorithm based approach to identify the basic source term information by minimizing a slightly different cost function.

He also extended his method to find other information about the source like source height, source strength, surface wind direction, surface wind speed, and time of release that produces a concentration field [55]. Note that optimization based methods extremely depend on the applied initial value for optimization, specially in non-convex optimization. However, most of the presented methods apply Genetic Algorithm to guarantee global minimization. Another major drawback of optimization based methods is that all of them provide only a point estimate of uncertain parameters and there is no uncertainty bounds associated with these estimates.

On the other hand, probabilistic methods provides posterior estimate of source parameters along with their associated uncertainty bounds, depending on accuracy of the model and observational data. Different probabilistic methods are developed for source identification in hazardous chemical releases [5, 61–63]. These methods can be divided into three different categories: Kalman based methods [4, 5], Bayesian Monte Carlo (BMC) method [6], and Bayesian Markov Chain Monte Carlo (BMCMC) technique [7].

Kalman filtering methods provide posterior mean and variance of source parameters based on linear combination of prior statistics (mean and covariance) and observation data. Kalman Filtration method is used in many atmospheric release applications for source characterization purposes [64–69].

There are major assumptions in derivation of Kalman Filter which restrict its applicability for general problems. The first restriction of Kalman Filter is that it assumes the uncertainties involved in prior distribution and associated noise signal in observational data to be Gaussian, which is usually far from reality. For instance, amount of a release from a source is always a positive quantity and can not be assumed to be normally distributed. Similarly, associated noise signal in data observation can not be considered to be Gaussian always. The other restriction that is embedded in Kalman Filter is the linearity assumption for the structure of dynamic model and observation model. Most of the time, the observation and dynamic model are nonlinear function of model parameters $\boldsymbol{\Theta}$. Hence, the assumption of linearity for observation and dynamic model is really restrictive. One way to alleviate this restriction is to utilize Extended Kalman Filter [70], which linearizes dynamic and observation model and then uses Kalman update equations. However, the linearization in Extended Kalman Filter may result in convergence issues in case of highly nonlinear dynamic systems or observation models, or in case of sparse measurement updates.

The basic idea of Bayesian Monte Carlo is to make use of Bayesian inference framework and Monte Carlo simulations. In this approach, numerical model is simulated for a large set of Monte Carlo samples, which are generated based on prior distribution of source parameters. Then, only those Monte Carlo samples whose outputs are consistent with observation are preserved.

There have been multiple researches regarding application of BMC approach for source characterization of plume dispersion incidents. Sohn *et al.* [21] applied BMC method for source localization of atmospheric release in building environment. His work includes some *pre-event planning* phase in which he simulates the atmospheric release numerical model for many Monte Carlo samples of parameter realizations, given prior distribution of $\boldsymbol{\Theta}$ and store simulation data in a library of model simulations. Then in event of a release, he uses Bayesian Inference framework to rapidly find the realizations of parameter $\boldsymbol{\Theta}$ which are the most consistent with observational data. Even though Bayesian Inference can be implemented rapidly, but this approach requires *pre-event planning* which can be computationally expensive. In addition, there is no convergence criteria for deciding the number of required Monte Carlo realizations of parameter $\boldsymbol{\Theta}$. In a similar research by Sreedharan [22], Bayesian interpretation approach is utilized to characterize an indoor release using threshold sensor data. Similar strategy proposed by Sohn *et al.* [21] is being applied in this paper, while using different environment and different data observation sensors. Deguillaume *et al.* [71] applied BMC method for estimation of emission of NOx and volatile organic compounds over the Ile-de-France region. Similar to previous works, his method consists in performing a large number of successive simulations with the same model but with a distinct set of model input parameters at each time. Then a posteriori weights are attributed to individual Monte Carlo simulations by comparing them with observations from the sensor network.

Bayesian Markov Chain Monte Carlo approach is another alternative that can be used to alleviate computational complexities involved in BMC method. In BMCMC, Monte Carlo samples are replaced with Markov Chain Monte Carlo samples. Hence, less number of numerical simulations is needed to get the same approximation for posterior distribution of source parameters. However, still large number of simulations is required to accurately approximate posterior distribution of source parameters. In addition, generating MCMC samples in high dimensions can still be computationally expensive.

Due to mentioned privilege of BMCMC with respect to BMC, there have been more

concentration on application of BMCMC in source characterization of atmospheric release incidents. Keats *et al.* [23] and Yee *et al.* [72] applied BMCMC approach in conjunction with adjoint based source-receptor relationship for source characterization of atmospheric release incidents in different experimental setups. However, some restrictive assumptions were considered in these works. For instance, associated noise and model error were considered to be normally distributed, which is different from reality in atmospheric release phenomena.

Johanesson *et al.* [24, 73] have a comprehensive article regarding application of BMCMC for source characterization of plume dispersion incidents. This report provides an introduction to a Bayesian probabilistic approach to modeling a dynamic system, with emphasis on stochastic methods for posterior inference. In addition, an overview is given about different sampling algorithms like Markov chain Monte Carlo (MCMC) approach and sequential Monte Carlo (SMC) approach. Finally some numerical simulations are provided, to represent application of BMCMC to source parameter estimation of an atmospheric release.

BMCMC is used in similar researches for source characterization of plume dispersion incidents [74–77]. For more detail about application of Bayesian Inference methods (BMC and BMCMC), please see [3] and [78].

In the end, we emphasize that the associated noise of measurement data is considered to be normally distributed in most of mentioned applications of BMCMC, which is against realistic assumptions. In addition, one should note that performing Markov Chain Monte Carlo sampling for obtained posterior distribution can be computationally intractable in presence of large dimensions for parameter $\boldsymbol{\Theta}$.

To summarize, Fig. 3.1 represents pros and cons of applied methods for source parameter estimation. For a more comprehensive review of the optimization based and probabilistic based methods for source characterization problem in plume dispersion phenomenon, please refer to [78] and [3].

In this Chapter, we utilize two different estimation techniques, recently developed by Madankan *et al.* [18], for estimation purpose. First, we describe mathematical details for a minimum variance based estimation approach in Section 3.2. Then, theoretical basis for Bayesian estimation method is explained in Section 3.3. In Section 3.4, we describe a methodology for reconstruction of posterior distribution of estimated parameters, based on polynomial chaos surrogate model. Performance of introduced estimation methods is demonstrated by some numerical simulations in Section 3.5. In the end, we summarize this

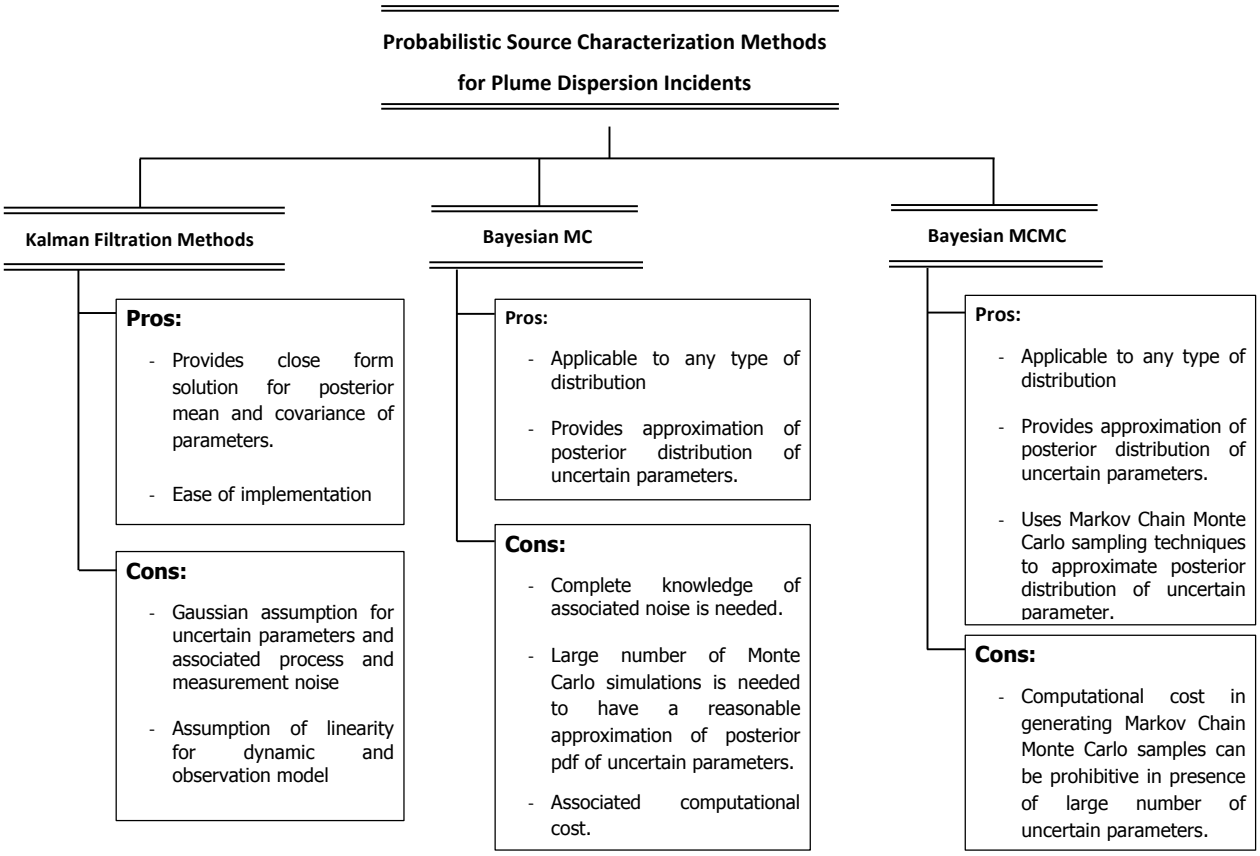


Figure 3.1: Classification of different source characterization methods for atmospheric release incidents

chapter by discussing the differences and benefits of each of the proposed methods in Section 3.6.

3.2 Minimum Variance Estimation

To describe the minimum variance estimation framework, once again consider a dynamical model given by Eq. (2.1), i.e.

$$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}, \boldsymbol{\Theta})$$

Now, let us assume the following sensor model for data observation:

$$\mathbf{y}_k \triangleq \mathbf{y}(t_k) = \mathbf{h}(\mathbf{x}_k, \boldsymbol{\Theta}) + \nu_k \tag{3.1}$$

where, $\mathbf{y}_k \in \mathbb{R}^b$ is observed sensor data and the nonlinear function $\mathbf{h}(\cdot)$ captures the sensor model and $\boldsymbol{\nu}_k$ is the measurement noise with zero mean and covariance matrix $\mathbf{R} \in \mathbb{R}^{b\times b}$. We are interested in a *linear* and *unbiased* minimum variance estimation framework which minimizes the trace of the posterior parameter covariance matrix:

$$J = \min_{\boldsymbol{\Theta}} \; Tr\left[\mathcal{E}[(\boldsymbol{\Theta} - \mathcal{E}[\boldsymbol{\Theta}])(\boldsymbol{\Theta} - \mathcal{E}[\boldsymbol{\Theta}])^T]\right] \tag{3.2}$$

It should be noted that the minimum variance formulation is valid for any pdf, although the formulation makes use of only the mean and covariance information. As well, it provides the maximum a-posteriori estimate when model dynamics and measurement model is linear and state uncertainty is Gaussian.

Minimizing the cost function J subject to the constraint of being an unbiased estimate, and using linear updating, allows us to compute the first two moments of the posterior distribution [18, 79]:

$$\hat{\boldsymbol{\Theta}}_k^+ = \hat{\boldsymbol{\Theta}}_k^- + \mathbf{K}_k[\mathbf{y}_k - \mathcal{E}^-[\mathbf{h}(\mathbf{x}_k, \boldsymbol{\Theta})]] \tag{3.3}$$

$$\boldsymbol{\Sigma}_k^+ = \boldsymbol{\Sigma}_k^- + \mathbf{K}_k \boldsymbol{\Sigma}_{\boldsymbol{\theta}y} \tag{3.4}$$

where, superscripts $-$ and $+$ denotes prior and posterior value of corresponding variable. In this update, the gain matrix $\mathbf{K}$ is given by:

$$\mathbf{K}_k = \boldsymbol{\Sigma}_{\boldsymbol{\theta}y}^T \left(\boldsymbol{\Sigma}_{hh}^- + \mathbf{R}_k + \mathbf{Q}_k\right)^{-1} \tag{3.5}$$

Note that $\hat{\boldsymbol{\Theta}}_k^-$ represents the prior mean for the parameter vector $\boldsymbol{\Theta}$ while incorporating measurements up to time interval t_{k-1} and $\hat{\boldsymbol{\Theta}}_k^+$ represents the posterior mean for parameter vector $\boldsymbol{\Theta}$ while incorporating measurements up to time interval t_k:

$$\hat{\boldsymbol{\Theta}}_k^- \triangleq \mathcal{E}^-[\boldsymbol{\Theta}_k] = \int_{\boldsymbol{\xi}} \boldsymbol{\Theta}_k^-(\boldsymbol{\xi}) p(\boldsymbol{\xi}) d\boldsymbol{\xi} \tag{3.6}$$

$$\hat{\boldsymbol{\Theta}}_k^+ \triangleq \mathcal{E}^+[\boldsymbol{\Theta}_k] = \int_{\boldsymbol{\xi}} \boldsymbol{\Theta}_k^+(\boldsymbol{\xi}) p(\boldsymbol{\xi}) d\boldsymbol{\xi} \tag{3.7}$$

Similarly, prior and posterior covariance matrices $\boldsymbol{\Sigma}_k^-$ and $\boldsymbol{\Sigma}_k^+$ can be written as:

$$\boldsymbol{\Sigma}_k{}^- \triangleq \mathcal{E}^-[(\boldsymbol{\Theta}_k - \hat{\boldsymbol{\Theta}}_k^-)(\boldsymbol{\Theta}_k - \hat{\boldsymbol{\Theta}}_k^-)^T] = \int_{\boldsymbol{\xi}} (\boldsymbol{\Theta}_k^-(\boldsymbol{\xi}) - \hat{\boldsymbol{\Theta}}_k^-)(\boldsymbol{\Theta}_k^-(\boldsymbol{\xi}) - \hat{\boldsymbol{\Theta}}_k^-)^T p(\boldsymbol{\xi}) d\boldsymbol{\xi} \in \mathbb{R}^{m\times m} \tag{3.8}$$

$$\boldsymbol{\Sigma}_k{}^+ \triangleq \mathcal{E}^+[(\boldsymbol{\Theta}_k - \hat{\boldsymbol{\Theta}}_k^+)(\boldsymbol{\Theta}_k - \hat{\boldsymbol{\Theta}}_k^+)^T] = \int_{\boldsymbol{\xi}} (\boldsymbol{\Theta}_k^+(\boldsymbol{\xi}) - \hat{\boldsymbol{\Theta}}_k^-)(\boldsymbol{\Theta}_k^+(\boldsymbol{\xi}) - \hat{\boldsymbol{\Theta}}_k^-)^T p(\boldsymbol{\xi}) d\boldsymbol{\xi} \in \mathbb{R}^{m\times m} \tag{3.9}$$

Also, $\mathbf{Q}_k$ denotes the model error covariance matrix in Eq. (3.5) which encapsulates the model's inaccuracies. The matrices $\boldsymbol{\Sigma}_{\boldsymbol{\theta}y}$ and $\boldsymbol{\Sigma}_{hh}$ are defined as:

$$\hat{\mathbf{h}}_k^- \triangleq \mathcal{E}^-[\mathbf{h}(\mathbf{x}_k, \boldsymbol{\Theta})] = \int_{\boldsymbol{\xi}} \underbrace{\mathbf{h}(\mathbf{x}_k^-(\boldsymbol{\xi}), \boldsymbol{\Theta}^-(\boldsymbol{\xi}))}_{\mathbf{h}_k} p(\boldsymbol{\xi}) d\boldsymbol{\xi} \tag{3.10}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}y} \triangleq \mathcal{E}^-[(\boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}}_k)(\mathbf{h}_k - \hat{\mathbf{h}}_k^-)^T] = \int_{\boldsymbol{\xi}} (\boldsymbol{\Theta}^-(\boldsymbol{\xi}) - \hat{\boldsymbol{\Theta}}_k^-)(\mathbf{h}_k - \hat{\mathbf{h}}_k^-)^T p(\boldsymbol{\xi}) d\boldsymbol{\xi} \tag{3.11}$$

$$\boldsymbol{\Sigma}_{hh}^- \triangleq \mathcal{E}^-[(\mathbf{h}_k - \hat{\mathbf{h}}_k^-)(\mathbf{h}_k - \hat{\mathbf{h}}_k^-)^T] = \int_{\boldsymbol{\xi}} (\mathbf{h}_k - \hat{\mathbf{h}}_k^-)(\mathbf{h}_k - \hat{\mathbf{h}}_k^-)^T p(\boldsymbol{\xi}) d\boldsymbol{\xi} \tag{3.12}$$

Here again, the expectation integrals in Eq. (3.10), Eq. (3.11), and Eq. (3.12) can be computed by suitable quadrature rules:

$$\hat{\mathbf{h}}_k^- \triangleq \mathcal{E}^-[\mathbf{h}(\mathbf{x}_k, \boldsymbol{\Theta})] \simeq \sum_{q=1}^{M} w_q \underbrace{\mathbf{h}(\mathbf{x}_k(\boldsymbol{\xi}^q), \boldsymbol{\Theta}(\boldsymbol{\xi}^q))}_{\mathbf{h}_q} \tag{3.13}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}y} \triangleq \mathcal{E}^-[(\boldsymbol{\Theta}_k - \hat{\boldsymbol{\Theta}}_k)(\mathbf{h}(\mathbf{x}_k, \boldsymbol{\Theta}) - \hat{\mathbf{h}}_k^-)^T] \simeq \sum_{q=1}^{M} w_q (\boldsymbol{\Theta}_k(\boldsymbol{\xi}^q) - \hat{\boldsymbol{\Theta}}_k^-)(\mathbf{h}_q - \hat{\mathbf{h}}_k^-)^T \tag{3.14}$$

$$\boldsymbol{\Sigma}_{hh}^- \triangleq \mathcal{E}^-[(\mathbf{h}(\mathbf{x}_k, \boldsymbol{\Theta}) - \hat{\mathbf{h}}_k^-)(\mathbf{h}(\mathbf{x}_k, \boldsymbol{\Theta}) - \hat{\mathbf{h}}_k^-)^T] \simeq \sum_{q=1}^{M} w_q (\mathbf{h}_q - \hat{\mathbf{h}}_k^-)(\mathbf{h}_q - \hat{\mathbf{h}}_k^-)^T \tag{3.15}$$

Again we point out that $\mathbf{h}_q$ represents computational measurements corresponding to simulation runs with input parameter determined by $\boldsymbol{\xi}^q$.

3.3 Bayesian Inference

To describe Bayesian inference framework, let us assume the following sensor model for data observation:

$$\mathbf{y}_k \triangleq \mathbf{y}(t_k) = \mathbf{h}(\mathbf{x}_k, \boldsymbol{\Theta}, \nu_k) \tag{3.16}$$

where, $\mathbf{y}_k \in \mathbb{R}^b$ is observed sensor data and the nonlinear function $\mathbf{h}(\cdot)$ captures the sensor model which is a function of state $\mathbf{x}_k$, parameter $\boldsymbol{\Theta}$, and measurement noise $\boldsymbol{\nu}_k$. Note that probability density function (pdf) of $\boldsymbol{\nu}_k$ is given by $p(\boldsymbol{\nu}_k)$.

One should notice the difference between applied observation model for minimum variance framework and Bayesian framework. In minimum variance estimation framework, the observation model, i.e. Eq. (3.1), is an linear function of noise signal $\boldsymbol{\nu}$. Also, the knowledge of only mean and covariance of the noise signal is enough for parameter estimation. On the other hand, the observation model in Bayesian inference framework, i.e. Eq. (3.16), can be a nonlinear function of noise signal in general. In addition, complete knowledge of probability density function (pdf) of noise signal is required for parameter estimation purposes in Bayesian inference framework.

Given observation data $\mathbf{y}_k$ and using Bayes' theorem, posterior distribution of $\boldsymbol{\Theta}$ can be written as:

$$p(\boldsymbol{\Theta}|\mathbf{y}_k) = \frac{p(\boldsymbol{\Theta})p(\mathbf{y}_k|\boldsymbol{\Theta})}{p(\mathbf{y}_k)} \tag{3.17}$$

where, $p(\boldsymbol{\Theta})$ is prior distribution of parameter $\boldsymbol{\Theta}$, $p(\mathbf{y}_k|\boldsymbol{\Theta})$ is likelihood of measurements given the parameter, and $p(\mathbf{y}_k)$ is probability of measurements, which is equal to:

$$p(\mathbf{y}_k) = \int_{\boldsymbol{\Theta}} p(\mathbf{y}_k|\boldsymbol{\Theta})p(\boldsymbol{\Theta})d\boldsymbol{\Theta} = \mathcal{E}_{\boldsymbol{\Theta}}\{p(\mathbf{y}_k|\boldsymbol{\Theta})\} \tag{3.18}$$

Note that posterior moments of $\boldsymbol{\Theta}$ can be computed by multiplication of appropriate functions of $\boldsymbol{\Theta}$ in Eq. (3.17) and integrating with respect to $\boldsymbol{\Theta}$ [18]. For instance, posterior mean of $\boldsymbol{\Theta}$, denoted by $\hat{\boldsymbol{\Theta}}^+$, can be computed as:

$$\hat{\boldsymbol{\Theta}}^+ = \mathcal{E}_{\boldsymbol{\Theta}}\{\boldsymbol{\Theta}\} = \frac{\int_{\boldsymbol{\Theta}} \boldsymbol{\Theta}p(\boldsymbol{\Theta})p(\mathbf{y}_k|\boldsymbol{\Theta})d\boldsymbol{\Theta}}{\mathcal{E}_{\boldsymbol{\Theta}}\{p(\mathbf{y}_k|\boldsymbol{\Theta})\}} = \frac{\mathcal{E}_{\boldsymbol{\Theta}}\{\boldsymbol{\Theta}p(\mathbf{y}_k|\boldsymbol{\Theta})\}}{\mathcal{E}_{\boldsymbol{\Theta}}\{p(\mathbf{y}_k|\boldsymbol{\Theta})\}} \tag{3.19}$$

Similarly, posterior second order moment of $\boldsymbol{\Theta}$, denoted by $\mathbf{P}^+$, can be obtained as:

$$\mathbf{P}^+ = \int_{\boldsymbol{\Theta}} \boldsymbol{\Theta}\boldsymbol{\Theta}^T p(\boldsymbol{\Theta}|\mathbf{y}_k)d\boldsymbol{\Theta} = \frac{\mathcal{E}_{\boldsymbol{\Theta}}\{\boldsymbol{\Theta}\boldsymbol{\Theta}^T p(\mathbf{y}_k|\boldsymbol{\Theta})\}}{\mathcal{E}_{\boldsymbol{\Theta}}\{p(\mathbf{y}_k|\boldsymbol{\Theta})\}} \tag{3.20}$$

Note that posterior value of covariance of $\boldsymbol{\Theta}$, denoted by $\boldsymbol{\Sigma}^+$, is obtained as:

$$\boldsymbol{\Sigma}^+ = \mathbf{P}^+ - \hat{\boldsymbol{\Theta}}^+\hat{\boldsymbol{\Theta}}^{+^T} \tag{3.21}$$

Likewise, higher order posterior moments of parameter $\boldsymbol{\Theta}$, denoted by $\mathcal{E}^{+}\{\phi(\boldsymbol{\Theta})\}$, can be written as:

$$\mathcal{E}^{+}\{\phi(\boldsymbol{\Theta})\} = \int_{\boldsymbol{\Theta}} \phi(\boldsymbol{\Theta}) p(\boldsymbol{\Theta}|\mathbf{y}_k) d\boldsymbol{\Theta} = \frac{\mathcal{E}_{\boldsymbol{\Theta}}\{\phi(\boldsymbol{\Theta}) p(\mathbf{y}_k|\boldsymbol{\Theta})\}}{\mathcal{E}_{\boldsymbol{\Theta}}\{p(\mathbf{y}_k|\boldsymbol{\Theta})\}} \tag{3.22}$$

where,

$$\phi(\boldsymbol{\Theta}) = \prod_{i=1}^{m} \boldsymbol{\theta}_i^{n_i}, \quad n_i \geq 0 \tag{3.23}$$

Similar to minimum variance estimation method, quadrature points are used to evaluate expected values in Eq. (3.19), Eq. (3.20) and Eq. (3.22). i.e.

$$\mathcal{E}_{\boldsymbol{\Theta}}\{\boldsymbol{\Theta} p(\mathbf{y}_k|\boldsymbol{\Theta})\} \simeq \sum_{q=1}^{N_q} w_q \boldsymbol{\Theta}(\boldsymbol{\xi}^q) p(\mathbf{y}_k|\boldsymbol{\Theta}(\boldsymbol{\xi}^q)) \tag{3.24}$$

$$\mathcal{E}_{\boldsymbol{\Theta}}\{\boldsymbol{\Theta}\boldsymbol{\Theta}^T p(\mathbf{y}_k|\boldsymbol{\Theta})\} \simeq \sum_{q=1}^{N_q} w_q \boldsymbol{\Theta}(\boldsymbol{\xi}^q) \boldsymbol{\Theta}^T(\boldsymbol{\xi}^q) p(\mathbf{y}_k|\boldsymbol{\Theta}(\boldsymbol{\xi}^q)) \tag{3.25}$$

$$\mathcal{E}_{\boldsymbol{\Theta}}\{p(\mathbf{y}_k|\boldsymbol{\Theta})\} \simeq \sum_{q=1}^{N_q} w_q p(\mathbf{y}_k|\boldsymbol{\Theta}(\boldsymbol{\xi}^q)) \tag{3.26}$$

$$\mathcal{E}_{\phi(\boldsymbol{\Theta})}\{\phi(\boldsymbol{\Theta}) p(\mathbf{y}_k|\boldsymbol{\Theta})\} \simeq \sum_{q=1}^{N_q} w_q \boldsymbol{\Theta}(\boldsymbol{\xi}^q) p(\mathbf{y}_k|\boldsymbol{\Theta}(\boldsymbol{\xi}^q)) \tag{3.27}$$

where, $\boldsymbol{\Theta}(\boldsymbol{\xi}^q)$ represents the quadrature value of the parameter vector $\boldsymbol{\Theta}$, which is obtained according to prior pdf $p(\boldsymbol{\Theta})$ and w_q is the weight of quadrature point $\boldsymbol{\Theta}(\boldsymbol{\xi}^q)$, corresponding to the applied quadrature scheme.

3.4 Reconstruction of Posterior Distribution

The posterior statistics obtained from Bayesian Inference or Minimum Variance framework can be used to approximate the posterior distribution of parameter $\boldsymbol{\Theta}$. This can be achieved by making use of Generalized Polynomial Chaos (gPC) Theory [31] [27].

As mentioned in Section 2.3, the uncertain system parameter $\boldsymbol{\Theta}$ can be written as a linear combination of basis functions, $\phi_k(\boldsymbol{\xi})$, which span the stochastic space of random variables

$\boldsymbol{\xi} = [\boldsymbol{\xi}_1, \cdots \boldsymbol{\xi}_m]^T$:

$$\theta_i(\boldsymbol{\xi}) = \sum_{k=0}^{N} \theta_{i_k} \phi_k(\boldsymbol{\xi}) \tag{3.28}$$

The coefficients θ_{i_k} are obtained by making use of following *normal equation*:

$$\theta_{i_k} = \frac{\mathcal{E}[\theta_i(\boldsymbol{\xi})\phi_k(\boldsymbol{\xi})]}{\mathcal{E}[\phi_k(\boldsymbol{\xi})\phi_k(\boldsymbol{\xi})]} \tag{3.29}$$

Note that the integrals involved in Eq. (3.29) can be evaluated using quadrature scheme, i.e.

$$\mathcal{E}[\phi_i(\boldsymbol{\xi}), \phi_j(\boldsymbol{\xi})] \simeq \sum_{q=1}^{M} w_q \phi_i(\boldsymbol{\xi}_q)\phi_j(\boldsymbol{\xi}_q) \tag{3.30}$$

$$\mathcal{E}[\theta_i(\boldsymbol{\xi}), \phi_j(\boldsymbol{\xi})] \simeq \sum_{q=1}^{M} w_q \theta_i(\boldsymbol{\xi}_q)\phi_j(\boldsymbol{\xi}_q) \tag{3.31}$$

One can update the polynomial expansion coefficients of Eq. (3.28) on the arrival of measurement data as described in [18]. To illustrate this in more detail, let us denote denote prior and posterior values of polynomial chaos coefficients of parameter θ_i, with $\theta_{i_k}^-$ and $\theta_{i_k}^+$, respectively. Assuming basis functions $\phi_k(\boldsymbol{\xi})$ to be orthonormal, one can write prior and posterior statistics of parameter θ_i in terms of prior and posterior values of its polynomial chaos coefficients. For instance, prior and posterior mean of θ_i can be written as:

$$\mathcal{E}^-[\theta_i] = \theta_{i_1}^-, \quad i = 1, 2, \cdots, m \tag{3.32}$$

$$\mathcal{E}^+[\theta_i] = \theta_{i_1}^+, \quad i = 1, 2, \cdots, m \tag{3.33}$$

Similarly, each element of prior and posterior covariance of parameter Θ can be written as:

$$\boldsymbol{\Sigma}_k^-(i,j) = \sum_{l=0}^{N} \boldsymbol{\theta}_{i_l}^- \boldsymbol{\theta}_{j_l}^-, \quad i = 1, 2, \cdots, m, , \quad j = 1, 2, \cdots, m \tag{3.34}$$

$$\boldsymbol{\Sigma}_k^+(i,j) = \sum_{l=0}^{N} \boldsymbol{\theta}_{i_l}^+ \boldsymbol{\theta}_{j_l}^+, \quad i = 1, 2, \cdots, m, , \quad j = 1, 2, \cdots, m \tag{3.35}$$

One can equate obtained posterior statistics from gPC expansion of parameter Θ with posterior statistics obtained from minimum variance or Bayesian estimation methods [18]. This results in a set of nonlinear algebraic equations which can be solved to find posterior

coefficients of gPC expansion of $\boldsymbol{\Theta}$:

$$\theta_{i_1}^{+} = \hat{\boldsymbol{\Theta}}_i^{+}, \quad i = 1, 2, \cdots, m \tag{3.36}$$

$$\sum_{l=0}^{N} \boldsymbol{\theta}_{i_l}^{+} \boldsymbol{\theta}_{j_l}^{+} = \boldsymbol{\Sigma}_k^{+}(i,j), \quad i,j = 1, \cdots, m \tag{3.37}$$

Notice that Eq. (3.36) directly results in values of $\boldsymbol{\Theta}_{pc_1}^{+}$ components, while Eq. (3.37) provides m^2 equations for remaining mN unknown coefficients. Depending on order of applied PC expansion and dimension of $\boldsymbol{\Theta}$, obtained set of equations can be over determined, determined, or under determined. Different approaches like nonlinear least square method can be used to solve Eq. (3.37) in case of over determined set of equations, while for under determined set of equations, regularization techniques can be used to find a solution for the posterior coefficients of polynomial expansion.

3.5 Numerical Simulations

3.5.1 Test Case 1

For the first example, once again source parameter estimation of dispersion/advection of propane is simulated over New York area by using Bayesian Inference framework. The domain of interest and the applied wind-field (at one specific time) are shown in Fig. 3.2. Simulation time is considered to be 24 $hrs.$, starting from 00 : 00 of September 1^{st}, 2013. North American Regional Reanalysis wind data at pressure level 100 kpa (height $\simeq$ 100 $m.$) is used as the wind-field for simulation. Three instantaneous mass releases are considered where their location is known and the only uncertain parameters are their amount of mass release. It is assumed that releases happen at the same time, i.e. all source releases happen at 00 : 00 of September 1^{st}. All mass releases are assumed to be uniformly distributed between 100 kg and 300 kg, as shown in Table 3.1.

As explained in Test Case 1 in Section 2.4.2, a set of 59 CUT8 quadrature points are used to quantify the uncertainty involved in concentration of dispersal material. Also, a 6^{th} order gPC expansion is used to reconstruct distribution of parameters after each update. Simulation of dispersion/advection has been performed using SCIPUFF numerical model, where concentration of propane is recorded every 10 $mins.$ and measurement update is implemented every one hour.

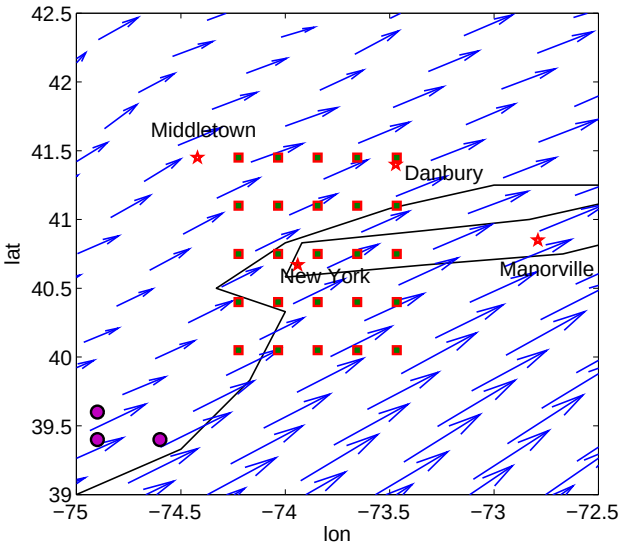


Figure 3.2: Schematic layout of Propane release over New York region, source locations are shown with purple circles, the wind-field (at $t = 0\ hr$ and pressure level $= 100\ kPa$) is shown over the two dimensional domain with blue vector field. Position of stationary sensors are shown with square markers. Geographical location of different cities can be seen on the background map

Table 3.1: Test Case 1: Uncertain source parameters. Actual values of parameters represent the values which are used to generate the synthetic measurement data.

Parameter	Distribution/Range	Actual value (kg)
$m_{1_{src}}$	Uniform $\in [100, 300]$	166.2
$m_{2_{src}}$	Uniform $\in [100, 300]$	246.04
$m_{3_{src}}$	Uniform $\in [100, 300]$	274.3

Sensor Model

The sensor used for the measurements is a bar sensor with a discrete numbers of bars. The number of bars ranges from zero to fifteen. These bar readings indicate the concentration magnitude at sensor location at the instant of data observation. The sensor displays $z = 0, \cdots, 15$, when the internal continuous-valued concentration magnitude $\mathbf{x}_{int}$ is be-

tween thresholds T_z and T_{z+1}, where $0 \leq T_z < T_{z+1}$. The thresholds T_z's are defined on a logarithmic scale, i.e. $T_z \in \{0, 5 \times 10^{-14}, 10^{-13}, 5 \times 10^{-13}, 10^{-12}, \cdots, 5 \times 10^{-7}\}$.

In here, we assume that each sensor reading is polluted with a discrete error, denoted by e, which is uniformly distributed over the set $\{-1, 0, 1\}$. Hence, measurement model can be written as:

$$\mathbf{z}_k \triangleq \mathbf{z}(t_k) = \mathbf{h}(\mathbf{x}_k, \boldsymbol{\Theta}) + e_k \tag{3.38}$$

where,

$$r = E[ee^T] = \frac{2}{3}$$

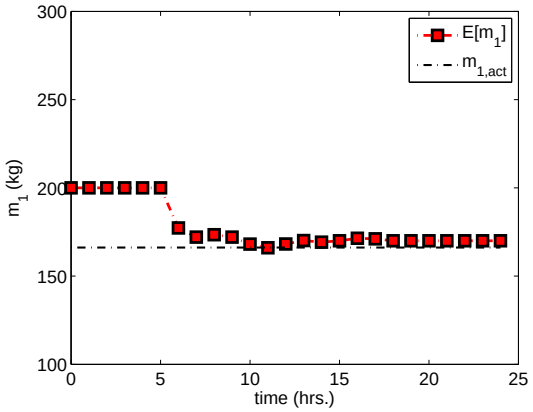
Note that function $\mathbf{h}(.)$ captures all discretizations involved in sensor model, i.e.

$$\mathbf{h}(\mathbf{x}_k, \boldsymbol{\Theta}) \in \{0, 1, 2, \cdots, 15\} \tag{3.39}$$

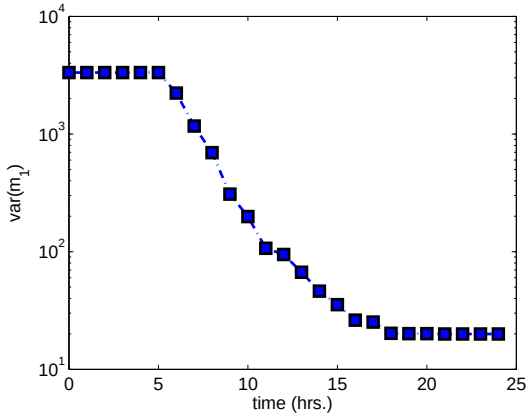
We use the minimum variance framework, described in Section 3.2, to find the first two posterior statistics of source parameters. As set of 25 sensors are used to perform the task of data observation. Fig. 3.5 illustrates source locations and the wind-field (at $t = 0$ $hrs.$) over the two dimensional spatial domain.

Fig. 3.3 illustrates mean and variance of parameter estimates during the time while using minimum variance estimation method. It is clear from from Fig. 3.3 that the mean estimate of the source parameters converges to actual value of the parameters and the variance of parameter estimates reduces during the time, which means more confident estimates source parameters over the time. One should note that parameter estimates don't change after $t = 18$ hrs. This is due to the fact that after $t = 18$ $hrs.$, no concentration of propane is measured by applied sensors and all sensor readings agree with model forecast. Hence, there will not be any change in statistics of parameter estimates after this time.

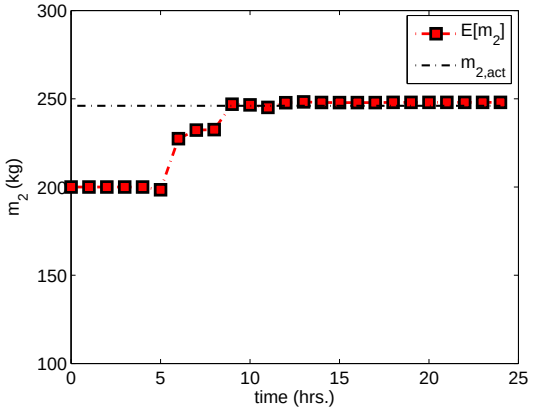
Finally, Fig. 3.4 illustrates distributions of uncertain parameters at final time, which are reconstructed by help of a 6^{th} order polynomial chaos surrogate model, as described in Section 3.4. Comparison of posterior and prior distributions of source parameters clearly demonstrates performance of minimum variance framework in estimation of source parameters.



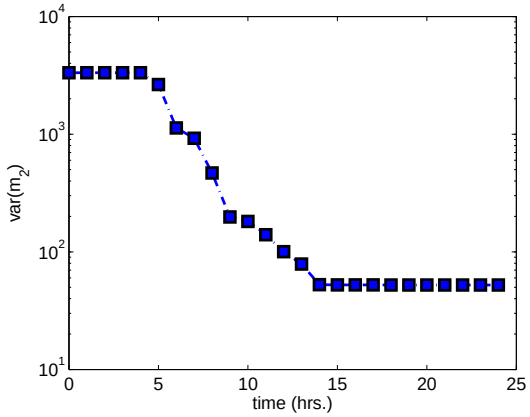
(a) convergence of $\mathcal{E}[m_1]$ over the time



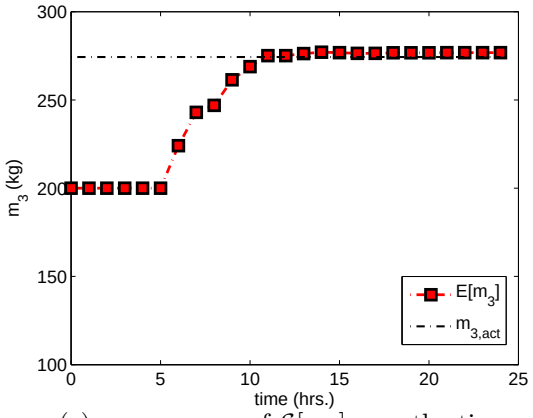
(b) convergence of $var(m_1)$ over the time



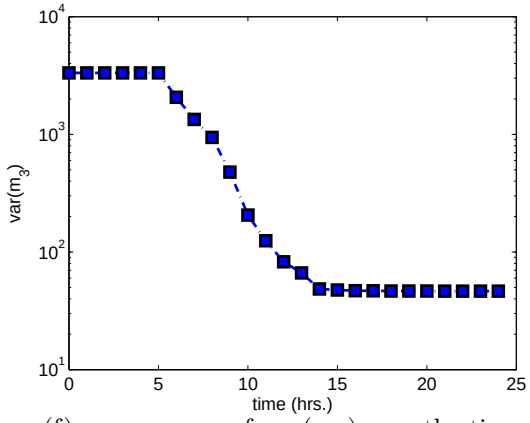
(c) convergence of $\mathcal{E}[m_2]$ over the time



(d) convergence of $var(m_2)$ over the time

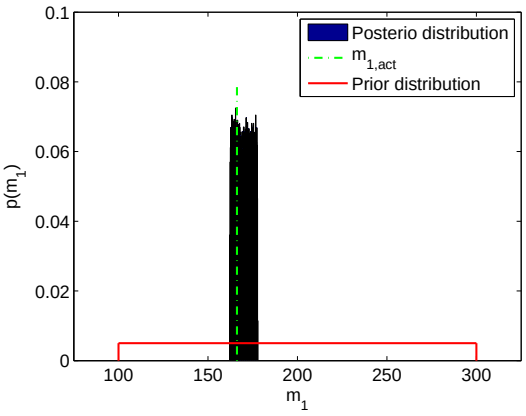


(e) convergence of $\mathcal{E}[m_3]$ over the time

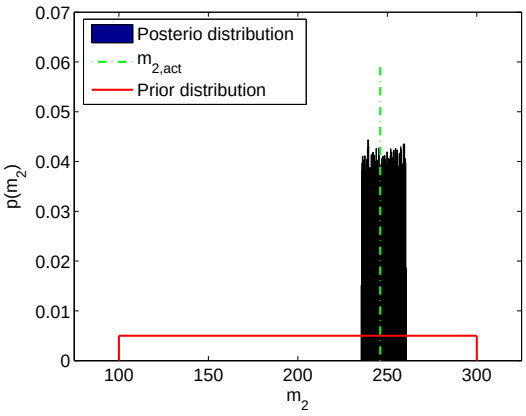


(f) convergence of $var(m_3)$ over the time

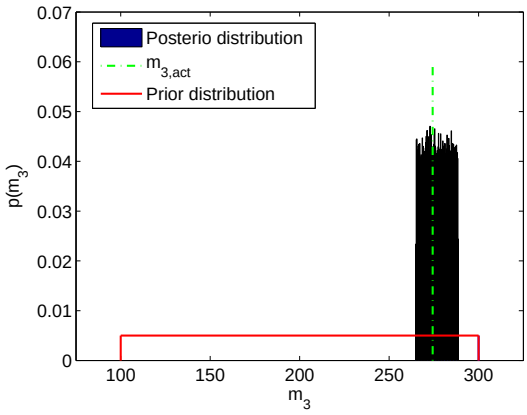
Figure 3.3: Test Case 1: Mean and variance of source parameter estimates during the time, obtained by using minimum variance estimation framework.



(a) m_1



(b) m_2



(c) m_3

Figure 3.4: Test Case 1: Posterior distribution of source parameters at $t = 24\ hr.$ versus their prior distribution, obtained by minimum variance estimation method and a 6^{th} order polynomial chaos surrogate model. Dashed green line shows actual value of parameters

3.5.2 Test Case 2

Similar to first example, again, dispersion/advection of propane over New York area is simulated, but Bayesian Inference framework and different type of data observation sensor are used for source parameter estimation. In the following, we first describe applied sensor model and then simulation results while using Bayesian Inference framework are presented for two different layout of data observation sensors.

Sensor Model

The sensor used for the measurements is a bar sensor with a discrete numbers of bars, which is slightly different from the one used in [80]. The number of bars ranges from zero to fifteen. These bar readings indicate the concentration magnitude at sensor location at the instant of data observation. The sensor displays $z = 0, \cdots, 15$, when the internal continuous-valued concentration magnitude $\mathbf{x}_{int}$ is between thresholds T_z and T_{z+1}, where $0 \leq T_z < T_{z+1}$. The thresholds T_z's are defined on a logarithmic scale, i.e. $T_z \in \{0, 5 \times 10^{-14}, 10^{-13}, 5 \times 10^{-13}, 10^{-12}, \cdots, 5 \times 10^{-7}\}$.

Properties of the sensor are determined by these thresholds and properties of $\mathbf{x}_{int}$, which is *assumed* to be normally distributed about the true concentration $\mathbf{x}$ [80]. Measurement error $v = \mathbf{x}_{int}$ - $\mathbf{x}$ may be considered a combination of multiplicative noise and additive noise with mean zero and standard deviation $\sqrt{R(\mathbf{x})} = \sigma(\mathbf{x}) = a\mathbf{x} + b$ where a is the proportionality constant and b accounts for the thermal motion of the electrons in the components [80]. In practice, because the true value of $\mathbf{x}$ is never known, $\sigma(\mathbf{x})$ is usually approximated by $\sigma(\tilde{\mathbf{x}})$, where $\tilde{\mathbf{x}}$ is an estimate of $\mathbf{x}$. a is equal to 1 and $b = 10^{-15}$ in our simulations. Also, it is assumed that $\tilde{\mathbf{x}} = T_z$, where T_z is the sensor bar corresponding to $\mathbf{x}_{int}$. Probability density function of $\mathbf{x}_{int}$ given the corresponding concentration $\mathbf{x}$ is:

$$p(\mathbf{x}_{int}|\mathbf{x}) = \mathcal{N}(\mathbf{x}_{int}; \mathbf{x}, R) = \frac{1}{\sqrt{2\pi R}} e^{-\frac{(\mathbf{x}_{int}-\mathbf{x})^2}{2R}} \tag{3.40}$$

where, $\mathcal{N}(,;.,.)$ denotes a Gaussian probability density function with the mean and variance specified by its second and third arguments. Strictly speaking, $p(\mathbf{x}_{int}|\mathbf{x})$ is not a Gaussian distribution because it is only defined for non-negative values of $\mathbf{x}_{int}$. Following Eq. (3.40), likelihood function, or simply probability of $\mathbf{z}$ conditioned on $\mathbf{x}$, is determined by the following

integral:

$$P(\mathbf{z}|\mathbf{x}) \propto \int_{T_z}^{T_{z+1}} p(\mathbf{x}_{int}|\mathbf{x}) d\mathbf{x}_{int}, \quad \sum_z P(\mathbf{z}|\mathbf{x}) = 1 \tag{3.41}$$

As mentioned before, sensor readings are polluted with some error, due to internal random variable and quantization effect. It is clear that when the number of thresholds goes to infinity and the length of intervals between two thresholds goes to zero, the sensor output becomes continuous and the likelihood function reduces to the (truncated) Gaussian distribution defined by Eq. (3.40). When the uncertainty due to internal random variable vanishes, likelihood function is flat over the interval $[T_z, T_{z+1}]$ associated with $\mathbf{z}$ and zero elsewhere.

Bayesian Inference is used to find the first two posterior statistics of source parameters. Note that CUT8 quadrature points are used to perform integrations involved in Eq. (3.19) and Eq. (3.20). Two different sensor orientations are used for source parameter estimation. Simulation results for each of these layouts are described in the following.

Orientation A

Fig. 3.5 illustrates source locations and the wind-field (at $t = 0\ hrs.$) over the two dimensional spatial domain. Similar to Test Case 1, a number of 25 sensors are used for data observation purpose, which are shown in Fig. 3.5.

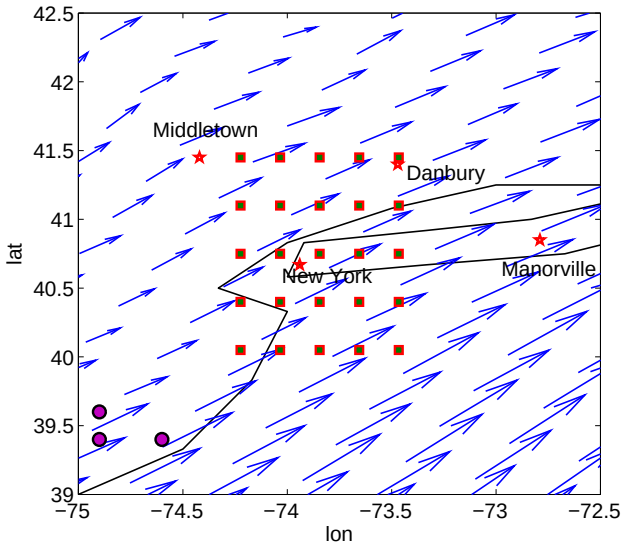
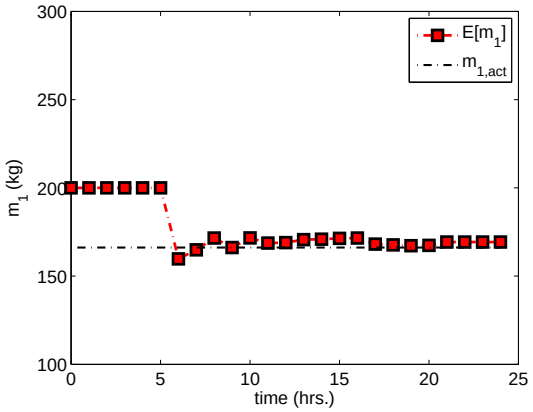


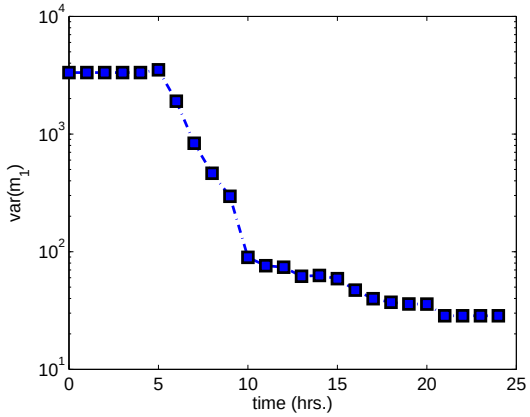
Figure 3.5: Test Case 2: Schematic layout of data observation sensors for orientation A

Convergence behavior for the mean and variance of parameter estimates using 25 stationary sensors are shown in Fig. 3.6. It can be seen from Fig. 3.6 that mean estimate of source parameters converges to their actual values and associated variance with these estimates reduces along the time. Comparison of obtained statistics for parameter estimates, using Bayesian Inference and the minimum variance framework shows that the minimum variance estimation method results in less variance (more confident estimates) for parameter estimates. This is due to the fact that the minimum variance framework is designed to minimize posterior covariance of parameter estimates. Hence, it always results in minimal value of the variance, in comparison with other estimation methods like Bayesian Inference approach.

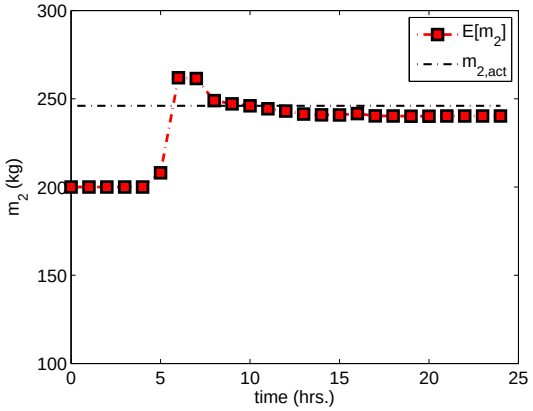
Fig. 3.7 represents distribution of uncertain parameters at the final time, reconstructed by the help of a 6^{th} order polynomial chaos surrogate model, as described in Section 3.4. Similar to the minimum variance estimation method, Fig. 3.7 shows that using Bayesian Inference method improves the confidence of posterior estimates of source parameters.



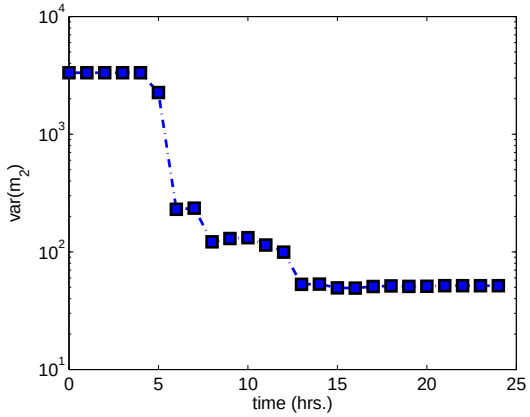
(a) convergence of $\mathcal{E}[m_1]$ over the time



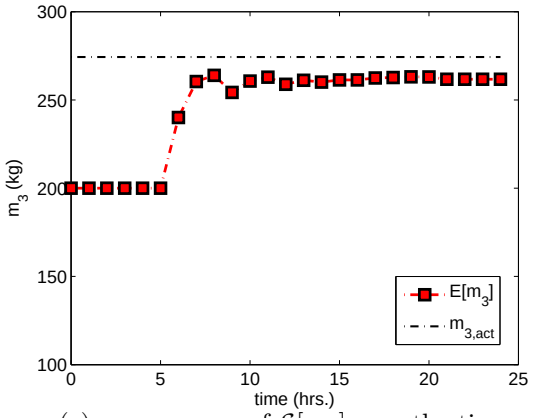
(b) convergence of $var(m_1)$ over the time



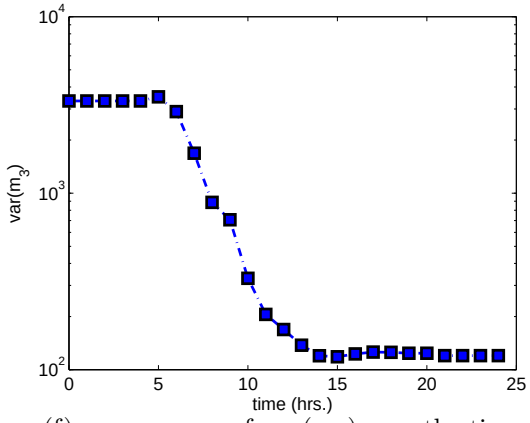
(c) convergence of $\mathcal{E}[m_2]$ over the time



(d) convergence of $var(m_2)$ over the time

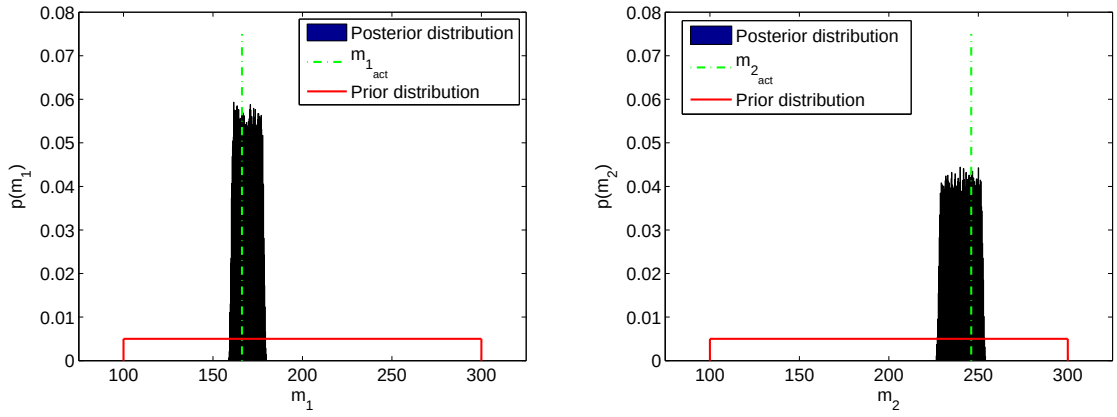


(e) convergence of $\mathcal{E}[m_3]$ over the time

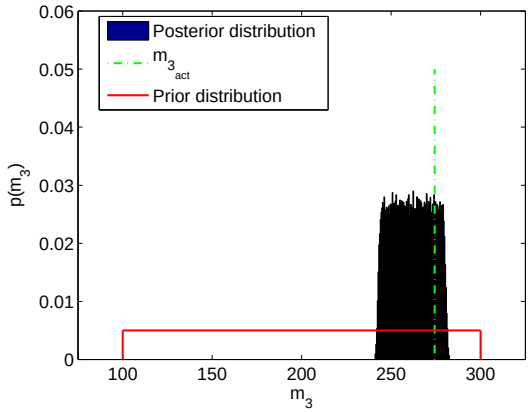


(f) convergence of $var(m_3)$ over the time

Figure 3.6: Test Case 2, Orientation A: Mean and variance of source parameter estimates during the time, obtained by using Bayesian Inference framework and 25 stationary sensors.



(a) m_1 (b) m_2



(c) m_3

Figure 3.7: Test Case 2, Orientation A: Posterior distribution of source parameters at $t = 24\ hr.$ versus their prior distribution, obtained by Bayesian Inference method and a 6^{th} order polynomial chaos surrogate model. Dashed green line shows actual value of parameters

Orientation B

To highlight the role of sensor placement, we have verified performance of estimation approach using different sensor layouts. In particular, we considered just three data observation sensors for Orientation B, whose locations are shown in Fig. 3.8, while all the other parameters are similar to Orientation A.

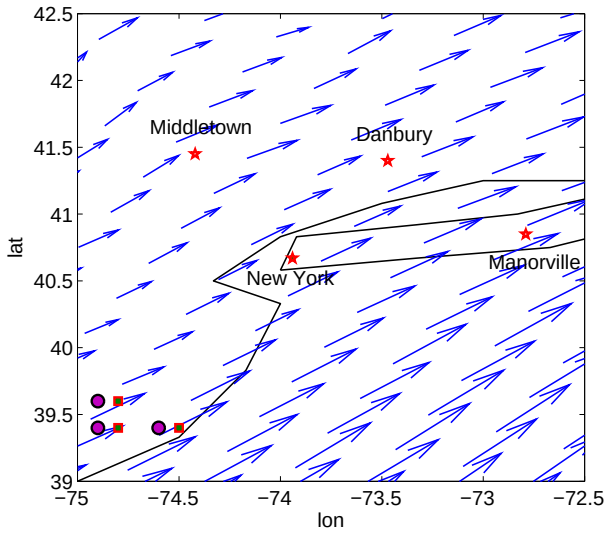


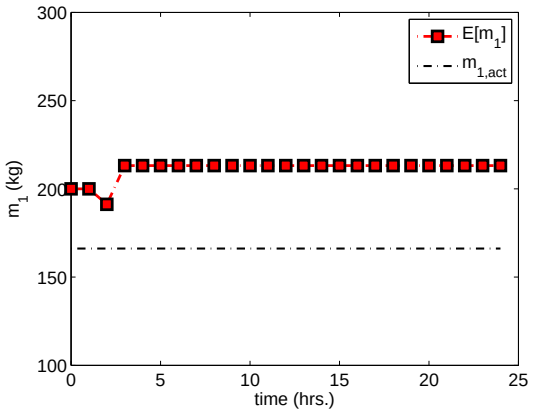
Figure 3.8: Test Case 2: Schematic layout of data observation sensors for orientation B

Convergence behavior for the mean and variance of source parameters estimates using three stationary sensors, is shown in Fig. 3.9. Fig. 3.9 illustrates that using this specific layout of stationary sensors results in poor estimates of the parameters. This is due to the inefficient placement of sensors during the data assimilation process.

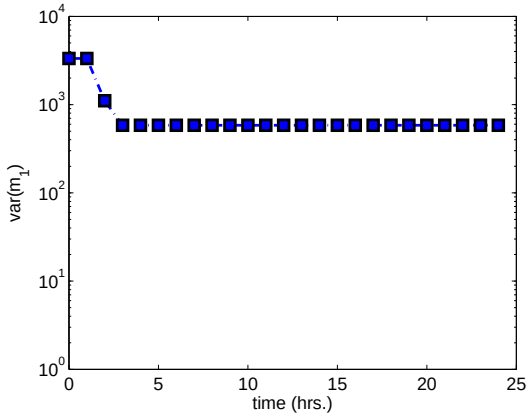
Distribution of uncertain parameters at the final time while using 3 sensors are also shown in Fig. 3.10. As one can see, posterior distributions of parameters are not very confident and there exist large amount of uncertainty involved in these estimates. This is due to poor placement of data observation sensors. Hence, optimal placement of data observation sensors is of high importance for accurate parameter estimation of distributed systems.

3.6 Summary

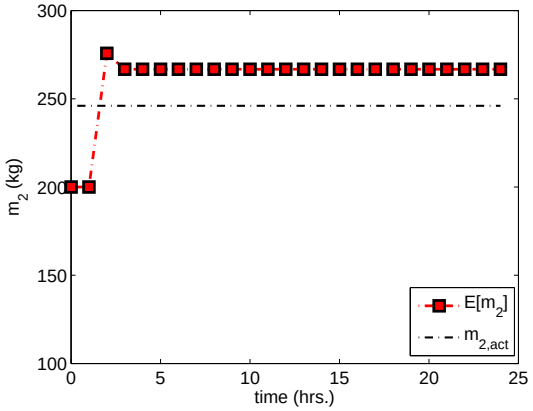
In this chapter, we utilized two recently developed parameter estimation methods by Madankan *et al.* [18] for source characterization of atmospheric release incidents. In the minimum vari-



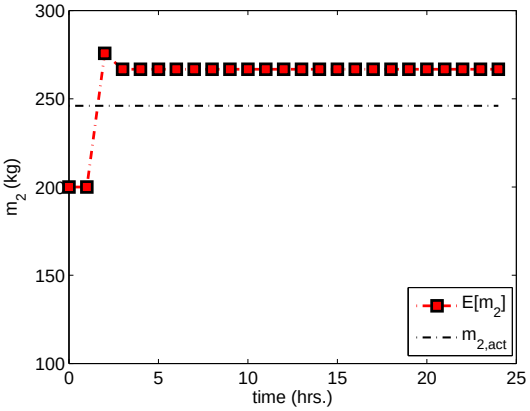
(a) convergence of $\mathcal{E}[m_1]$ over the time



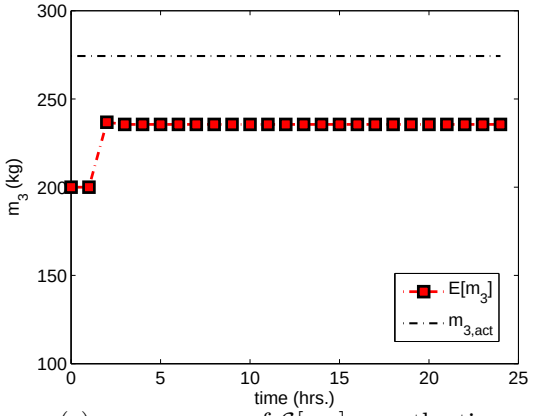
(b) convergence of $var(m_1)$ over the time



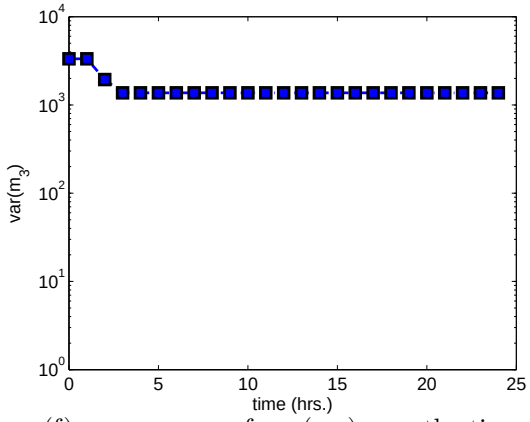
(c) convergence of $\mathcal{E}[m_2]$ over the time



(d) convergence of $var(m_2)$ over the time

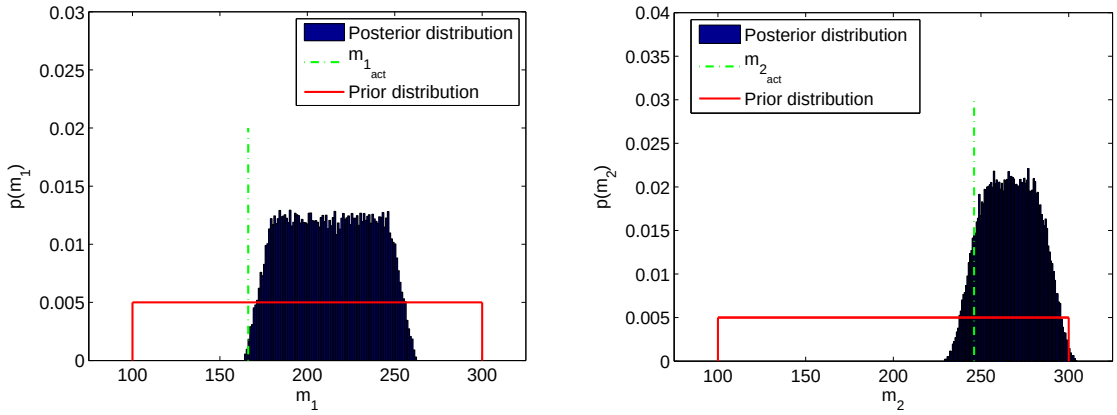


(e) convergence of $\mathcal{E}[m_3]$ over the time

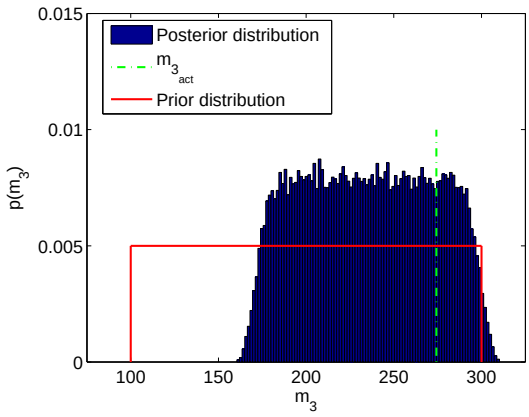


(f) convergence of $var(m_3)$ over the time

Figure 3.9: Test Case 2, Orientation B: Mean and variance of source parameter estimates during the time, obtained by using Bayesian Inference framework and 3 stationary sensors.



(a) m_1 (b) m_2



(c) m_3

Figure 3.10: Test Case 2, Orientation B: Posterior distribution of source parameters at $t = 24\ hr.$ versus their prior distribution, obtained by Bayesian Inference method and a 6^{th} order polynomial chaos surrogate model. Dashed green line shows actual value of parameters

ance estimation framework, posterior estimate of the mean and covariance of uncertain parameters are provided based on model forecast and observational data. Note that there is no restrictive assumption for distribution of uncertain parameters and associated noise in measurement data while making use of the minimum variance estimation framework. The only restriction of the minimum estimation framework is that observation operator should be an affine function of associated noise signal. Hence, using the minimum variance framework is not straightforward in presence of observation operators which are nonlinear function of noise signal. Also, first two statistics of associated noise signal should be known while using

minimum variance estimation method.

On the other hand, Bayesian Inference method can be used to estimate posterior statistics of uncertain parameters up to a desired order. In addition, there is no restriction regarding structure of observation operator while using Bayesian inference framework. Hence, Bayesian inference method can be applied in more general conditions, comparing to minimum variance framework. Note that, distribution of associated noise signal should be known when using Bayesian inference approach, while in minimum variance estimation method, knowledge of just first two moments of noise signal is enough.

One should notice that obtained posterior statistics from either Bayesian Inference or the minimum variance method can be combined with polynomial surrogate model to approximate posterior distribution of estimated parameters, as described in Section 3.4.

In the end, we once again emphasize that performance of applied parameter estimation method crucially depends on locations of applied sensors, as we showed in our second numerical example. Hence, it is necessary to develop a general framework to optimally locate observation sensors over spatial domain to improve performance of the estimation process. In Chapter 5, we will develop a general optimal sensor placement strategy to perform this task. But before that, let us demonstrate performance of proposed uncertainty quantification and estimation methodology for a real world problem. In the next chapter, we show efficiency of developed mathematical tools by applying these methods to the problem of uncertainty quantification and source parameter estimation of Eyjafjallajökull eruption which happened in April 2010 in Iceland.

Chapter 4

Application to Volcanic Ash Plume Dispersion

4.1 Introduction

Ash clouds are produced by the explosive eruptions of volcanoes. These clouds, propagating downwind from a volcano eruption column, are a hazard to aircraft, causing damage to the engines [81]. On December 15, 1989, KLM Flight 867 lost all its engines when the airplane entered a plume of ash originating at the Redoubt Volcano in the Aleutian Islands [82]. That incident caused more than $80 million (US) in damage to the aircraft, but fortunately no lives were lost. The recent eruption of the EyjafjallajökullVolcano in Iceland wreaked havoc on European aviation after the eruption started on April 14, 2010. Decisions about the closure of European air-space, largely based on deterministic ash plume models, resulted in more than $4 billion in economic losses and left more than 10 million stranded passengers [83]. In addition to the large financial consequences of volcanic eruptions, there are significant health and environmental consequences of ash propagation and its subsequent fallout, ranging from inhalation of the ash particles to crop damage from tephra fallout. Clearly, those charged with volcanic risk management need accurate information for decision making. Among other components, this information flow should include a map of the probability of ash being present at a given location at a specified time.

Other hazardous events present similar needs. For example, the accidental release of radioactive gaseous material, such as occurred at the Chernobyl nuclear reactor explosion,

or the oil spill resulting from the Deepwater Horizon accident in Gulf of Mexico, also demand tools and approaches, to accurately forecast the advection and dispersion of a material.

In this chapter, we illustrate performance of the proposed methodology in the previous two chapters for uncertainty quantification and source parameter estimation of Eyjafjallajökull eruption which occurred in April 2010 in Iceland. In the following, we first review concurrent approaches for uncertainty quantification and source parameter estimation of volcanic eruption incidents and briefly describe our methodology. Then, applied numerical model of eruption column and volcanic ash transport and dispersion (VATD) model are explained in section 4.2. Our simulation results for uncertainty quantification and source parameter estimation are then illustrated in Section 4.3 and Section 4.4, respectively. Finally, a discussion of results is presented in Section 4.5.

Current Approaches and Limitations

Often times volcanologists extrapolate information from past eruptions to create maps forecasting future events and areas at risk. Basing forecasts solely on past recorded events does not always provide a reliable estimate of likely eruption scenarios – prior events may have gone unreported, and site-specific conditions may have changed. Computer simulations using physics-based model equations, calibrated using field data, provide additional information on which to base hazard forecasts. To predict ash cloud movement, model systems may incorporate stochastic variability, such as uncertainty in source parameters or randomly varying wind fields, to better capture possible ash particle transport. A major source of uncertainty impacting the location of a volcanic ash cloud are the characteristics of the volcanic eruption column, including the distribution of grain size in the column and the column rise height [84]. Several investigations have tried to quantify the effect of source parameter uncertainty on the position of ash clouds. For example, during the Eyjafjallajökull eruption, the London Volcanic Ash Advisory Centers (VAAC) used the NAME computational model [85] for ash advection/dispersion to make forecasts of the position of the ash cloud, which in turn were used to issue advisories to the airline industry. In related work, Denevish *et al.* [86] applied NAME, with a specified set of input source parameters estimated from measurement data, to study the arrival of the Eyjafjallajökullash cloud over the United Kingdom. Through a sensitivity analysis, this study demonstrated that the position and concentration of ash over a given region of interest were particularly dependent on eruption source parameters such as

the column height and the particle profile within the column. O'Dowd *et al.* [87] simulated the dispersion of ash from Eyjafjallajökull using the REMOTE computational model, for a specified set of source parameters. In another study, Webley *et al.* [88] used the WRF-Chem dispersion and tracking model to forecast the ash cloud position, given the column height, particle grain-size distribution and mass eruption rate. Heinold *et al.* [89] simulated the Eyjafjallajökull emission, transport, and particle deposition over Europe by using the regional chemistry-transport model COSMO-MUSCAT, given the height of ash particles and their size distribution. Dispersion of the ash cloud from the Eyjafjallajökull eruption has also been simulated by using the FALL3D computational model [90], where the input parameters are approximated from the observed height of the eruption column and from the total grain size distribution as reconstructed from field observations. These investigations each apply different computational models to forecast ash cloud position as a function of time, each with its successes and limitations. In each instance, however, a specified set of the eruption source parameters, perhaps obtained retrospectively from radar or satellite data, is used to forecast ash cloud motion. Because there is great uncertainty in the model inputs, deterministic physics based models alone are limited in their ability to make meaningful forecasts.

In order to make accurate long-term forecasts, it is necessary to understand how the uncertainty in source parameters and the variability of wind fields propagate through the numerical advection/dispersion codes. Although a detailed sensitivity analysis can relate the variations in source parameters and wind data to ash cloud motion, uncertainty analysis provides a richer suite of tools, allowing an assessment of one's confidence in making forecasts based on all available information. Of course a successful application of uncertainty analysis must overcome the challenges posed by the large number of uncertain input parameters and the associated cost of computation. Data input and output drive the calculations of uncertainty quantification, and present additional difficulties for any analysis. Importantly, in real-time hazard assessment one is constrained by the need for rapid analysis. Each of these factors affects the trade-off between completeness and speed. In addition, propagating uncertain model inputs leads to forecasts with uncertainties that grow in time and which must be tamed in order to make useful forecasts; assimilating available observational data to refine the model forecast reduces these uncertainties.

Surprisingly, limited research has been done on fusing model forecasts with available measurement data to accurately forecast ash cloud motion. The exceptions are the recent

works of Stohl *et al.* , Denlinger *et al.* and Kristiansen *et al.* [91–93], who use inversion methods to couple *a-priori* source information and the output of dispersion models together with satellite data, to estimate volcanic source parameters. As a consequence, simulations performed by using these *posterior* source estimates result in better correspondence with satellite data. The major drawback of this approach is that the inversion method results in a deterministic *point estimate* for the posterior values of source parameters, and completely neglects prior information and inaccuracies in measurement data. An alternative to simply fitting the measurement data is to exploit sensor noise characteristics. A simple probabilistic approach is to apply a Maximum Likelihood Estimate (MLE) [70] to estimate the parameter values. However, the MLE also provides only point estimates and does not provide any information about one's confidence in those estimates.

A Bayesian method such as the Maximum *a posteriori* (MAP) estimation [94] combines a prior distribution together with information contained in measurements, to provide optimal estimates for source parameters. Like the MLE method, significant computational effort is required to solve the optimization problem resulting from the MAP approach, to determine optimal source parameters. This computational burden restricts the application of the method in large scale dynamical systems.

Our Approach

Figure 4.1 outlines our basic approach. Past knowledge of similar eruption and eruption source observation are used to create an initial probability distribution of the model parameters, for a recently developed model that couples a volcanic eruption column (the `bent` model) with a volcanic ash transport and dispersion (VATD) model (`puff`) [95]. These distributions are then used to generate an ensemble of simulation runs, guided by Conjugate Unscented Transform quadrature scheme [46, 47, 49]. The ensemble outcomes are then integrated to generate a probabilistic map of the ash distribution in space and time by constructing a polynomial chaos surrogate model of the VATD model. As satellite imagery becomes available, this data is used to find a posterior estimate of the volcano source parameters, using a minimum variance estimator as part of the solution of an inverse problem. Furthermore, the satellite data is also used to improve the model parameter distribution by updating the polynomial chaos surrogate model. These refined source parameters estimates can then be used in subsequent propagation and forecast.

In the end, one should note that even though we have employed the `bent` and `puff` models, any other column and VATD model could be used, and the statistical calculations appropriately adapted. Indeed, the framework introduced here provides an approach for developing maps for many hazard scenarios, assuming the cost of simulations is not prohibitive.

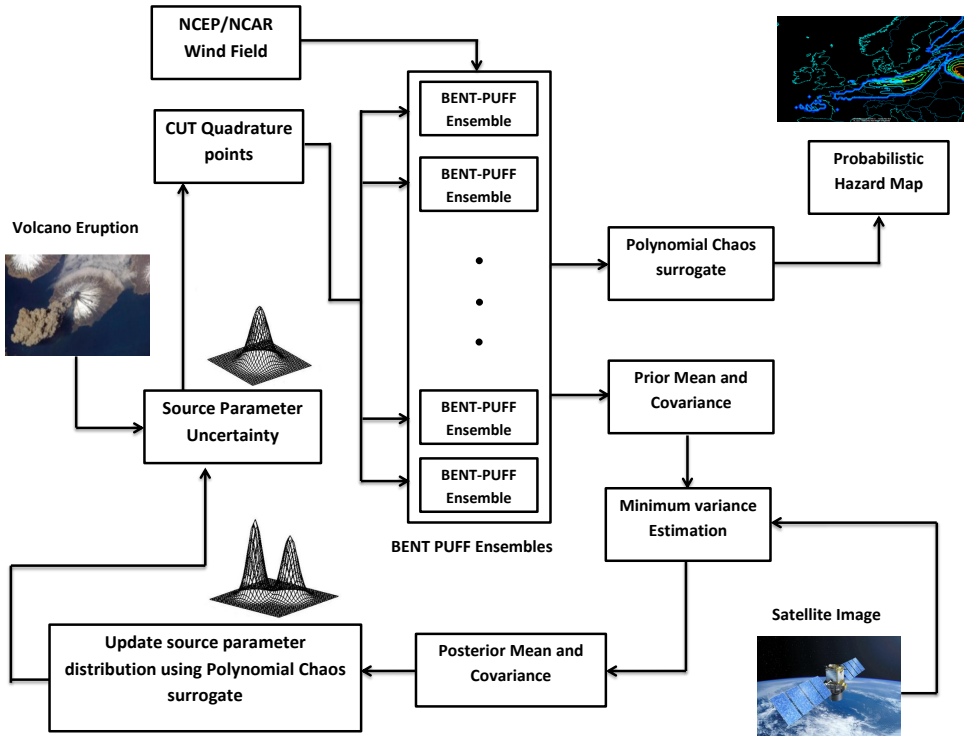


Figure 4.1: Schematic view of probabilistic model forecast and source parameter estimation process while incorporating prior knowledge of source uncertainty and satellite imagery.

4.2 Bent-PUFF Numerical Model

Ash transport models can be divided into two broad categories: those intended to calculate eruption column and tephra fall deposit characteristics based on source vent conditions, as in [96] (eruption column model), and those intended to forecast long-range atmospheric transport, dispersion and fallout, as in [97] (VATD model). All ash transport models rely on the existence of an explicit relationship between the eruption column and atmospheric dynamics, and the resulting transport, dispersion and settling of the ash. The focus of this work is on the long-range movement of ash clouds, and not eruption column dynamics or tephra deposition. Therefore, a simple VATD model, but one that nonetheless contains

several sources of uncertainty, is considered to focus attention on long-range transport and dispersion. Tanaka [98] and Searcy et al. [99] developed `puff`, an ash tracking model for forecasting the paths of incipient volcanic clouds. `puff` simplifies the eruption plume to a vertical source, and uses a Lagrangian pseudo-particle representation of the ash plume in a detailed 3-D regional windfield to determine the trajectory of the cloud. `puff` and other dispersion models have proven extremely useful in modeling the distal transport of ash for aviation safety [99]. During an eruption crisis, `puff` forecasts have been used to forecast ash cloud movement critical to the assessment of potential impacts – for example, on aircraft flight paths. `puff` has been validated against historic volcanic eruptions such as the 1992 Crater Peak vent eruption at Mount Spurr and the 2006 eruption at Mount Augustine with reasonable success [99, 100]. To start a simulation, `puff` requires as inputs the eruption start time and duration, the initial plume height, the vertical distribution of particles of varying size, a representative wind field, and the simulation end time. At first, some of these parameters must be assumed, based on past activity of the volcano, or by using the Eruption Source Parameters (ESP) of Mastin et al. [101].

To initialize a `puff` simulation a collection of particles of different sizes must be specified as a function of altitude, a process that is not well constrained; see [102,103]. It is important to remember that `puff` particles are not simple surrogates for ash concentration, but are representatives of ejecta of a given size at a specified height. As such this number is a user-selected input that affects both simulation time and resolution of the output. In addition to particle grain-size distribution and windfield, other `puff` input parameters include the coefficient of turbulent diffusion, and particle settling speed, both of which are estimated. Instead of guessing the initial particle distribution as a function of height, a volcanic eruption plume model called `bent` is employed to provide initial conditions for the VATD model. The essential features of this coupling between `bent` and `puff` is described in [95]. `bent` solves the equations for mass, momentum and energy balance, averaged over a cross-sectional slice of the eruption column [104]. `bent` assumes a grain-size distribution of pyroclasts and, depending on the volcanic vent size and the speed of the ejecta, the model equations forecast the height distribution of the various sized clasts. `bent` has been tested against plume rise height data and ash dispersal data [103]. In particular, the discussion in that paper (among many others) corroborates that the scaling relationships derived in [105] between energy and plume rise height are valid for energetic volcanic plumes piercing the tropopause.

In a tool we call `bent-puff`, `bent` incorporates important physics of the volcano column and provides initial conditions for `puff`. On the one hand, physics guides the model coupling and determines how outputs from `bent` feed into `puff`. On the other hand, this coupling can be viewed as simply substituting one set of uncertain parameters in `puff` (vent size, velocity, clast size distribution) for an uncertain function of `bent` (particle height distribution). In any event, physically relevant inputs from the volcano source – together with their variability – are modeled and propagated through `bent` and `puff`.

4.3 Uncertainty Quantification

The performance of the proposed methodology for generating accurate hazard maps and for updating model parameter estimates, is assessed using data from the April 2010 Eyjafjallajökull eruption in Iceland. For simulation of ash transport, the `bent` volcanic column model is first used to generate initial ash cloud data, and then `puff` VATD model [106] is utilized to simulate transport of the ash over spatial domain during the time period $14-16$ April 2010. `bent` numerical model produces mass loading, plume height, and grain size distribution, which are used in `puff`, given atmospheric winds and volcanic source conditions. Icelandic Meteorological Office (IMO) Keflavik radiosonde data from 14 April 00Z, where 00Z refers to midnight in Universal Time, was used to generate the atmospheric winds for `bent`. Numerical model `puff`, together with the given windfield, tracks the propagation of ash from Iceland to Europe. `puff` can be run using one of several Numerical Weather Prediction (NWP) windfields [107–110]. These NWP models are available at different levels of spatial and temporal resolution. For this simulation, `puff` uses global NCEP/NCAR Reanalysis windfields to propagate ash, using 6-hr, 2.5° data. These wind fields assimilate observation wind data into model runs. Output from a deterministic `puff` model run consists simply of the position of the representative numerical particles; one can smooth this positions to determine a smoothed concentration field. The outputs are post-processed to extract other quantities of interest, such as maximum height of ash at a given geographical location. The top-height of ash is a useful quantity for the purpose of air traffic routing.

All four volcano source parameters – vent radius b_0, vent velocity w_0, mean grain size Md_ϕ, and standard deviation of grain size σ_ϕ – are assumed to be uncertain, and the prior density functions for these parameters, based upon previous eruption studies, are listed in

Table 4.1: Eruption source parameters based on observations of Eyjafjallajökull volcano and information from other similar eruptions of the past.

Parameter	Value range	PDF	Comment
Vent radius, b_0, (m)	65-150	Uniform, non-negative	Measured from IMO radar image of summit vents on 14 April 2010
Vent velocity, w_0, (m/s)	45-124	Uniform, non-negative	Measured by infrasound [111] 6-21 May, when MER similar to 14-18 April
Mean grain size, Md_φ, φ	3.5-7	Uniform, $\in \mathbb{R}$	[112], Table 1, vulcanian and phreatoplinian. A. Hoskuldsson, Eyjafjallajökull Eruption Workshop, 09/2010, presentation, 'Vulcanian with unusual production of fine ash'.
σ_φ, φ	$0.5 - 3$	Uniform, $\in \mathbb{R}$	[112], Table 1, vulcanian and phreatoplinian
Eruption temperature	1200 C	Fixed	[95]
Erupted water mass fraction	0.017	Fixed	[95]
Eruption duration	3 hr	Fixed	[95]

Table 4.1. The CUT quadrature scheme described in Section 2.2 was used to produce initial ensembles of source parameters. In earlier work [95], it was shown that an 8^{th} order quadrature scheme is sufficient for computing statistics of ash top-height. Hence, 161 CUT8 quadrature points were generated. Following runs of `bent` corresponding to CUT8 quadrature points, each `bent` output was then propagated through `puff`, which was then run for three days. The outputs from `puff` were then used to create a polynomial surrogate model of degree 4 for ash top-height. 5×10^4 secondary sample points of $\boldsymbol{\xi} \in \mathbb{R}^4$ were then used to compute the probabilistic hazard map described in Section 2.3.

Meteosat-9 retrievals of ash-cloud height were used to validate the probabilistic hazard map methodology and to refine prior probability density functions. Volcanic ash was identified in the satellite data using the methodology described in [113] and [114]. The ash loading (mass per unit area) and ash cloud height were retrieved using an optimal estimation approach [115, 116].

Table 4.2: Location 52N, 13.5E: conc is the puff computed absolute air concentration (in mg/m^3) in a grid cell of size $0.5^\circ \times 0.5^\circ \times 2km$ at 1200hours on 16^{th} April, 2010 , and count is the number of `puff` particles in that cell

Number of ash particles	height (m)	conc.	count	height	conc.	count	height	conc.	count
10^5	3000	7.4×10^{-5}	28	5000	4.23×10^{-5}	16	7000	-	-
5×10^5	3000	1.17×10^{-4}	221	5000	3.54×10^{-5}	67	7000	-	-
10^6	3000	1.12×10^{-4}	405	5000	4.12×10^{-5}	156	7000	-	-
2×10^6	3000	1.12×10^{-4}	884	5000	4.03×10^{-5}	305	7000	-	-
4×10^6	3000	1.09×10^{-4}	1655	5000	4.10×10^{-5}	3620	7000	1.32×10^{-7}	2
8×10^6	3000	1.15×10^{-4}	3471	5000	4.15×10^{-5}	1256	7000	1.98×10^{-7}	6
10^7	3000	1.10×10^{-4}	4151	5000	3.99×10^{-5}	1510	7000	2.91×10^{-7}	11

Computing Probability of Ash top-height

Like any Lagrangian model, the accuracy of the `bent-puff` model is greatly influenced by the number of ash particles used in simulation. To understand the convergence of the approach proposed here, it is necessary to understand how the number of ash particles impact the output of `puff`. For this purpose, probabilistic hazard maps were computed corresponding to five different values of the number of ash particles: 4×10^6, 10^7, 2×10^7, 4×10^7, and 8×10^7. For all `puff` runs, the vertical position of ash particles is quantized in 2-km altitude levels. Table 4.2 represents the absolute and relative ash concentration at a particular location, for different altitudes and different number of ash particles used in one deterministic run of `bent-puff`. As expected, both the concentration and the ash top-height are significantly affected by the number of ash particles. The table shows that, by increasing the number of ash particles from 2×10^6 to 4×10^6, the maximum height of ash at location 52N, 13.5E increases from 5000 m to 7000 m.

Figure Fig. 4.2 shows the processor time and estimates of memory required to complete a single run of the `bent-puff` model, as a function of the number of particles. From this plot, it is clear that the computational time increases exponentially with an increase in number of ash particles in a run. Here again the trade-off between desired accuracy and computational cost is evident. These results are consistent with prior studies performed on the convergence of `puff` output [117].

Fig. 4.3 shows the probability of ash top-height being greater than or equal to specified threshold h_{thresh} for a few specific locations and times, as a function of the number of ash particles. It appears that the probability values have converged for lower value of h_{thresh}, but considerable fluctuations for $h_{thresh} = 5000\ m$ and $h_{thresh} = 7000\ m$ remain. This observation

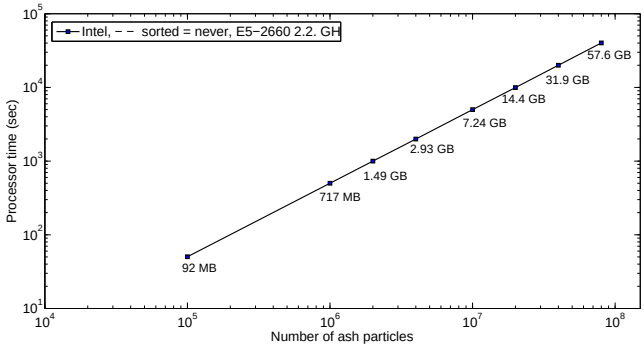
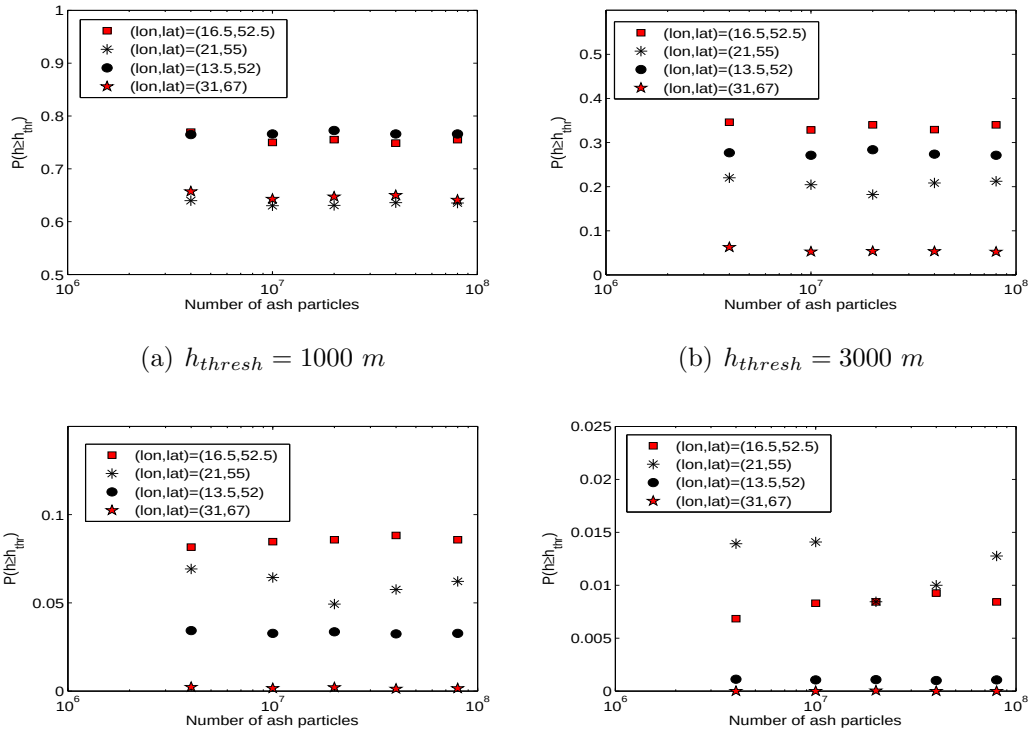


Figure 4.2: Processor Time for Single Deterministic Run of `bent-puff` vs. Number of Ash Particles.

is consistent with the convergence of `bent-puff` model shown in Table 4.2. One surmises that these fluctuations can be attributed to the accuracy of the `puff` model rather than any aliasing error in the convergence of quadrature scheme.



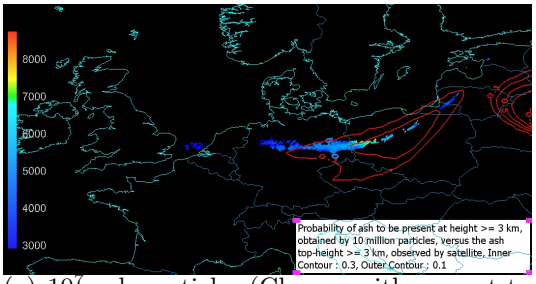
(a) $h_{thresh} = 1000\ m$ (b) $h_{thresh} = 3000\ m$

(c) $h_{thresh} = 5000\ m$ (d) $h_{thresh} = 7000\ m$

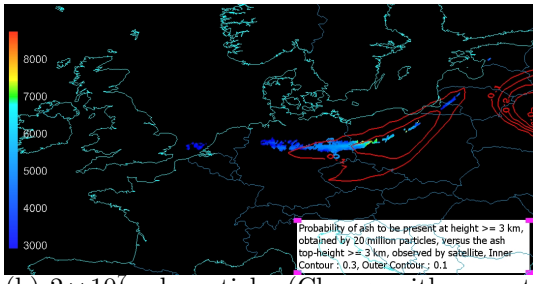
Figure 4.3: Probability of ash top-height $\geq$ h_{thresh} at different spatial locations on April 16^{th}, 0600 hrs

Fig. 4.4 shows the probabilistic hazard map consisting of probability of ash top-height

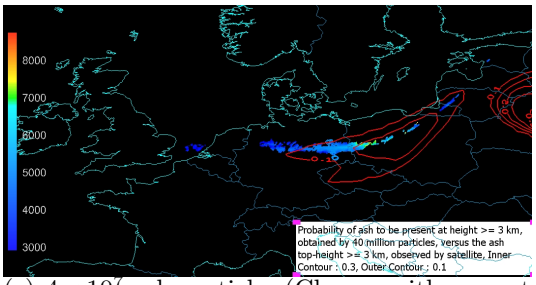
being greater than or equal to $h_{thresh} = 3\ km$ for different numbers of ash particles in `bent-puff` model, overlaid with satellite observed ash top-height greater than or equal to 3 km, on April 16^{th}, 0600 hrs (54 hours after eruption). Fig. 4.5 shows the probability of ash top-height to be greater than or equal to $h_{thresh} = 5\ km$, overlaid with satellite observed ash top-height greater than or equal to 5 km, again on April 16^{th}, 0600 hrs. From these plots, one can conclude that the probabilistic hazard map calculations have converged with respect to number of ash particles used and satellite imagery consistently fall within most probable forecasted region.



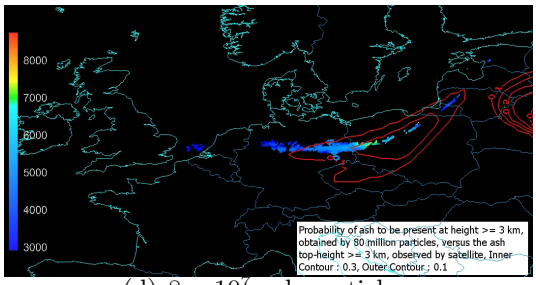
(a) 10^7 ash particles (Change with respect to 80 million ash particles = 1.36%)



(b) 2×10^7 ash particles (Change with respect to 80 million ash particles = 1.13%)



(c) 4×10^7 ash particles (Change with respect to 80 million ash particles = 0.96%)

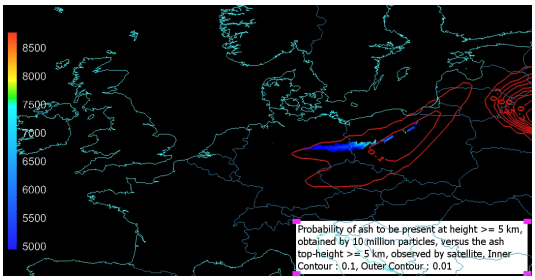


(d) 8×10^7 ash particles

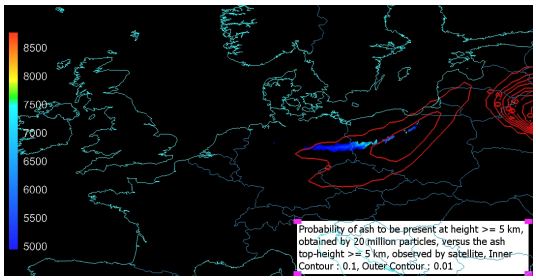
Figure 4.4: Probability of ash top-height $\geq 3\ km$ versus satellite observed ash top-height $\geq 3\ km$ on April 16^{th}, 1200 hrs (60 hours after eruption).

Fig. 4.6 shows the probability map of ash top-height exceeding 1 km overlaid with satellite observed ash top-height at six-hour interval for 16^{th} April. Most of the satellite data lies within the high probability region, although the probable ash cloud footprint is quite large, owing to the large uncertainty in prior source parameters. Note also the predicted ash in the north-east corner of the image is not supported by satellite imagery; further study indicates this area was obscured by meteorological clouds.

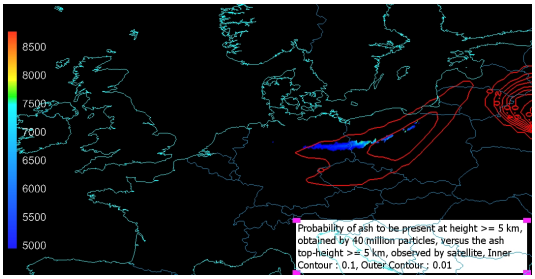
To compare the accuracy of the CUT quadrature scheme, the 8^{th} order Clenshaw-Curtis



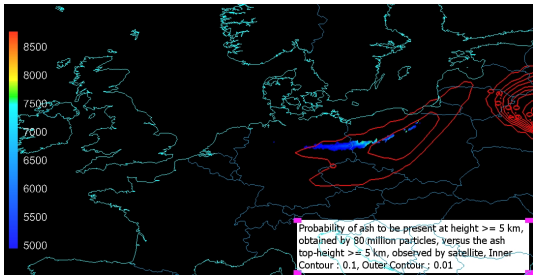
(a) 10^7 ash particles (Change with respect to 80 million ash particles = 1.25%)



(b) 2×10^7 ash particles (Change with respect to 80 million ash particles = 1.10%)



(c) 4×10^7 ash particles (Change with respect to 80 million ash particles = 0.97%)

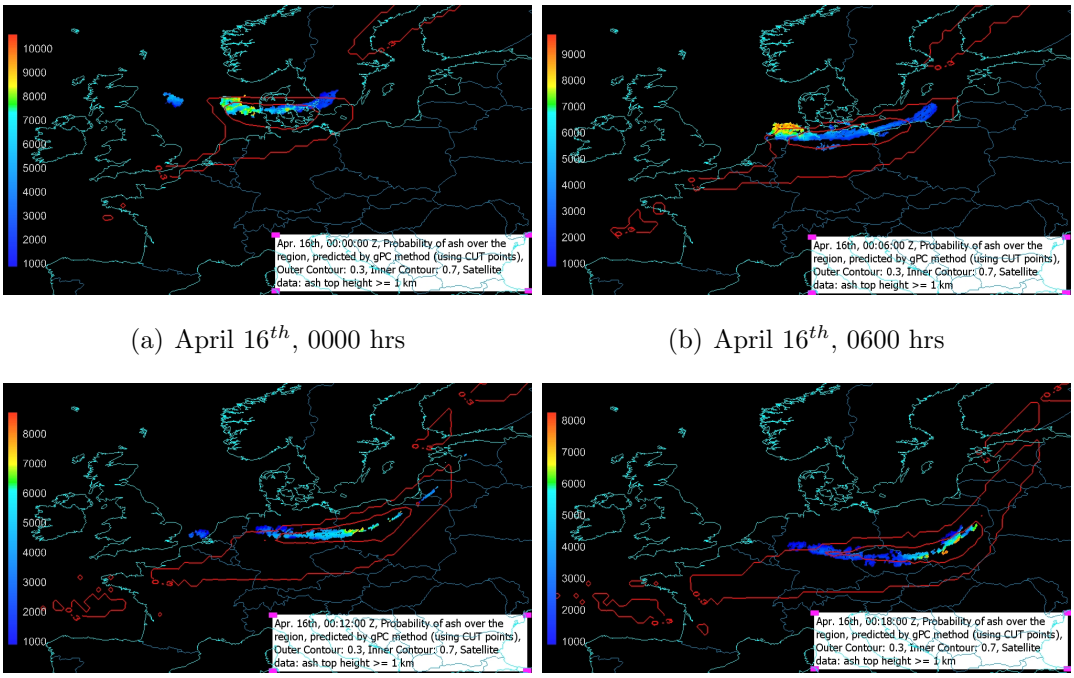


(d) 8×10^7 ash particles

Figure 4.5: Probability of ash top-height ≥ 5 km versus satellite observed ash top-height ≥ 5 km on April 16^{th}, 1200 hrs (60 hours after eruption).

(CC) quadrature scheme with 9^4 quadrature points is employed to compute probabilistic hazard maps. The convergence of the Clenshaw-Curtis quadrature scheme in computing the mean and standard deviation of the ash top-height has been studied in earlier work [95]. Fig. 4.7 shows the probability map of ash top-height exceeding 1 km overlaid with satellite observed ash top-height at six-hour intervals for 16^{th} April. From Fig. 4.6 and Fig. 4.7, it is clear that probability maps computed with the help of CUT and CC quadrature schemes are indistinguishable. We conclude that the CUT methodology provides an order of magnitude computation savings without the loss of any accuracy.

Finally, Fig. 4.8 and Fig. 4.9 illustrate mean and standard deviation of ash top-height, obtained by CUT8 and Clenshaw-Curtis quadrature points, respectively. One can see from comparison of these figures that both quadrature schemes result in similar statistics for ash top-height, while CUT quadrature scheme uses significantly less computational resources due to lower number of quadrature points.



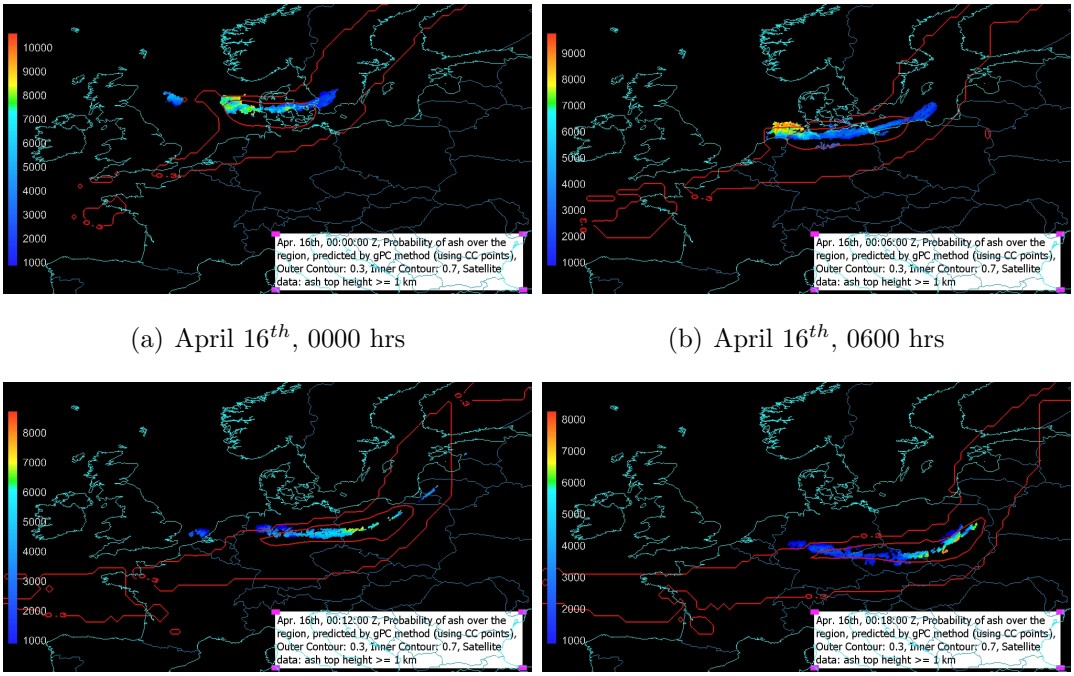
(a) April 16^{th}, 0000 hrs

(b) April 16^{th}, 0600 hrs

(c) April 16^{th}, 1200 hrs

(d) April 16^{th}, 1800 hrs

Figure 4.6: Probability Maps for Ash Top-Height ≥ 1 km and corresponding satellite observed ash top-height.



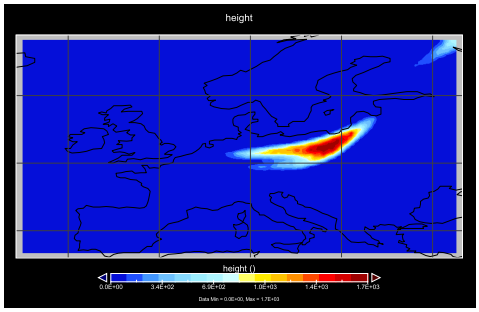
(a) April 16^{th}, 0000 hrs

(b) April 16^{th}, 0600 hrs

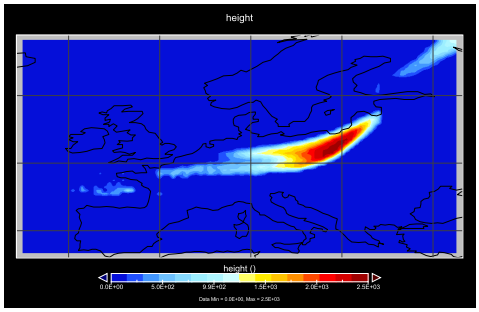
(c) April 16^{th}, 1200 hrs

(d) April 16^{th}, 1800 hrs

Figure 4.7: Probability Maps (obtained through Clenshaw-Curtis quadrature Scheme) for Ash Top-Height ≥ 1 km and corresponding satellite observed ash top-height.

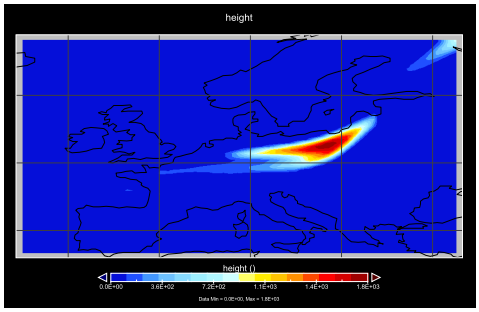


(a) Mean Ash Top-Height

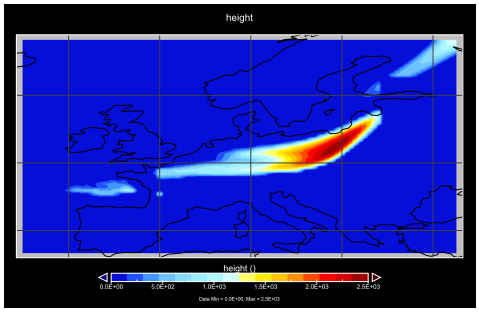


(b) Standard Deviation of Ash Top-Height

Figure 4.8: Mean and Standard deviation of Ash Top-Height on April 16^{th}, 1200 hrs., obtained by CUT quadrature points



(a) Mean Ash Top-Height

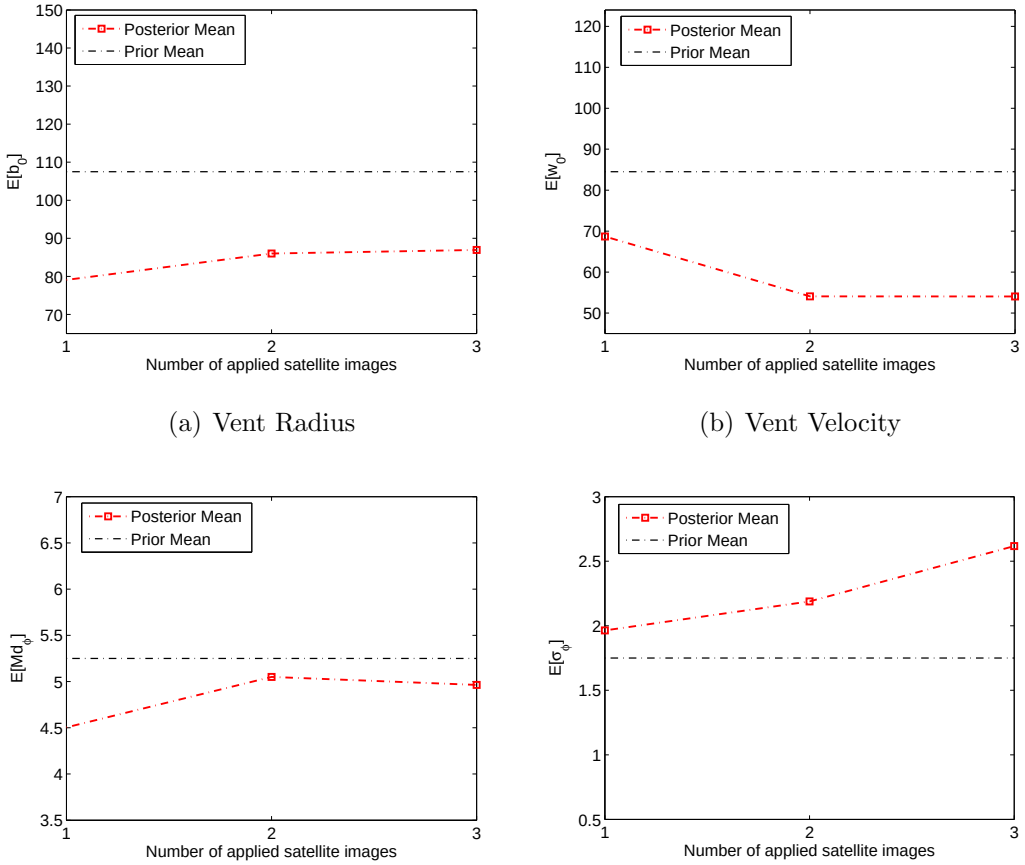


(b) Standard Deviation of Ash Top-Height

Figure 4.9: Mean and Standard deviation of Ash Top-Height on April 16^{th}, 1200 hrs., obtained by Clenshaw-Curtis quadrature points

4.4 Source Parameter Estimation

As just shown, due to the large uncertainty in source parameters, the uncertainty in the probable ash footprint is very high (see Fig. 4.6). This finding suggests we should re-compute the source parameter distributions making use of satellite observations. The procedure listed in Section 3.2 is used to compute posterior estimate for the source parameters and the corresponding probability density function, using satellite data from three different times



(a) Vent Radius

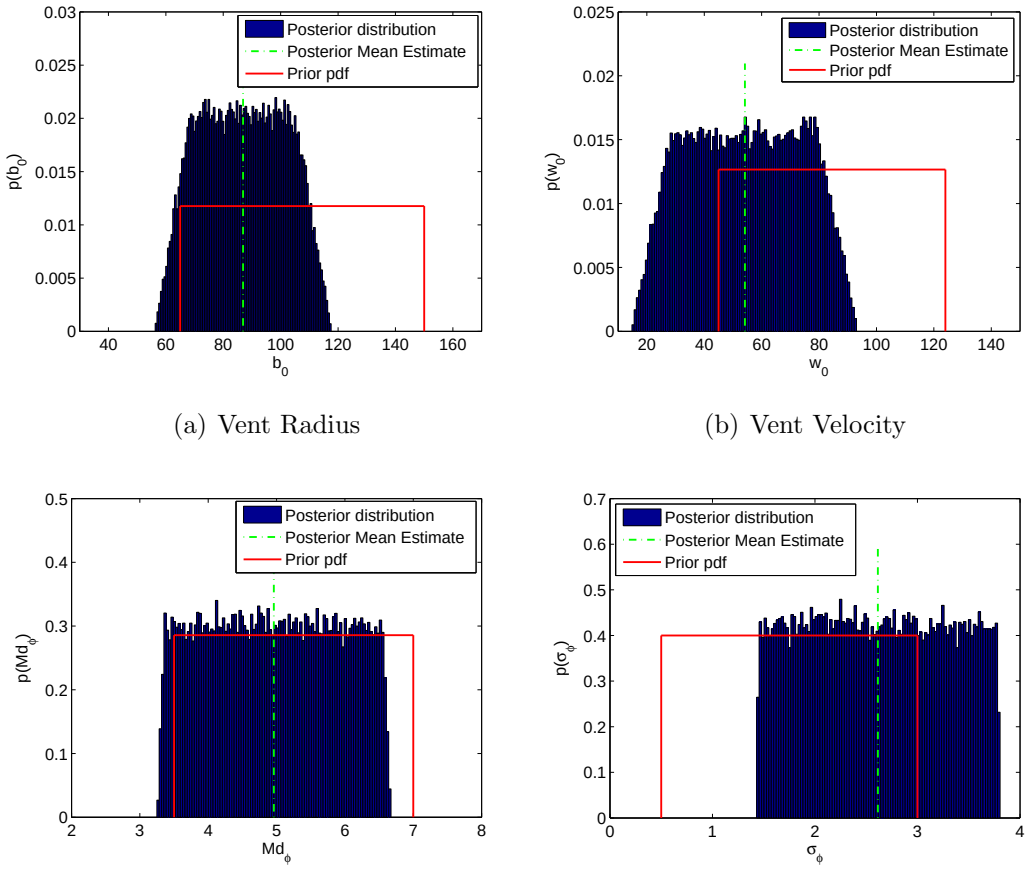
(b) Vent Velocity

(c) Mean Grain Size

(d) σ_ϕ

Figure 4.10: Posterior Mean Estimates for Source Parameters versus number of applied satellite data.

(April 16^{th} at 0600 hrs, 1200 hrs, and 1800 hrs). Satellite observed ash top-heights are estimated to be accurate to within 100 m intervals around the observed height, so the sensor noise $\boldsymbol{\nu}_k$ is taken to be a zero-mean uniform density function over the interval $[-100\ 100]$ m. Due to height quantization in the `bent-puff` model, ash top-height provided by `bent-puff` model is assumed to be polluted with zero-mean uniformly distributed random noise between $-1000\,m$ and $+1000\,m$. Thus $\mathbf{Q}_k$ in Eq. (3.5) is taken to be $3.33\times10^{-1}km^2$ in our simulations. 4×10^7 ash particles were used in the `bent-puff` model to compute different expectation integrals involved in the calculation of posterior source parameter distributions. To reduce the potential source of numerical error, PETSc [118] with two level domain decomposition based algebraic preconditioning (block Jacobi or Additive Schwarz) is used to compute the inverse involved in the computation of $\mathbf{K}_k$ in Eq. (3.5).



(a) Vent Radius

(b) Vent Velocity

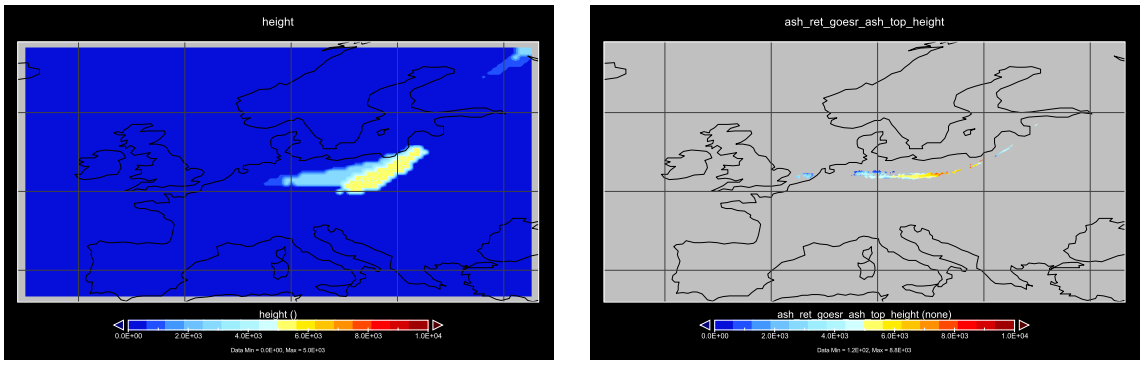
(c) Mean Grain Size

(d) σ_ϕ

Figure 4.11: Prior and posterior estimate of source parameters.

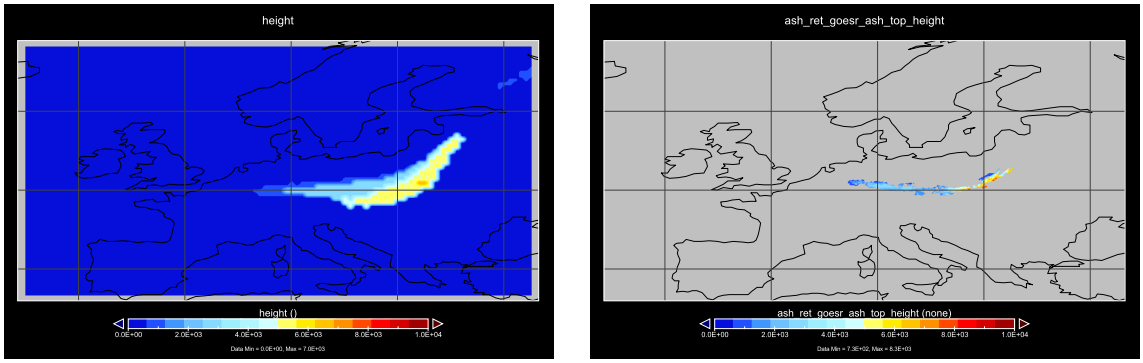
The prior mean and standard deviation of ash top-height are shown in Fig. 4.8. Fig. 4.10 shows the posterior mean of the source parameters computed through Eq. (3.3), versus the number of satellite images considered in calculations of posterior mean and covariance. The expected source parameter values converge as more and more observational data are made available. Fig. 4.11 shows the assumed prior source parameter distributions and the computed posterior distributions based on satellite imagery for all three time-intervals. As expected, the uncertainty in source parameters decreases after the assimilation of satellite imagery. Because vent radius and eruption velocity directly control mass eruption rate, thermal flux and therefore eruption plume height, the fact that the cloud top height estimated from satellite data changes these values is intuitive. The effect of the satellite data on the grain size distribution is less obvious, but nevertheless can be easily understood when one remembers that the particles are settling, and the settling is a function of grain size. The large increase in the standard deviation of the grain size distribution would furthermore seem to be a reflection that the posterior estimate requires a greater number of fine-grained particles that settle only slowly.

Finally, the quality of the source parameter estimates is assessed by performing a single deterministic run of `bent-puff` corresponding to the estimated posterior mean of source parameters and comparing it against satellite observed ash top-height. Fig. 4.12(a) shows the ash top-height forecast at time 1200 hrs on April 16^{th} using posterior estimates for source parameters obtained through incorporating satellite observation available at 600 hrs on April 16^{th}. Similarly, Fig. 4.13(a) shows the ash top-height forecast at time 1800 hrs, obtained through incorporating satellite observations available at 600 hrs and 1200 hrs. Fig. 4.12(b) and Fig. 4.13(b) show the satellite observed ash top-height at 600 hrs and 1200 hrs, respectively. These results indicate that the forecast of ash cloud top-height based on the posterior estimate of source parameters match very well with the observed satellite data. The observed and computed ash top-height differ from each other with an accuracy of ± 2 km, which corresponds to the numerical accuracy of `bent-puff` model.



(a) `bent-puff` forecast (b) Satellite Observation

Figure 4.12: Comparison of Forecast of Ash top-height and Satellite Observation on April 16^{th}, 1200 hrs.



(a) `bent-puff` forecast (b) Satellite Observation

Figure 4.13: Comparison of Forecast of Ash top-height and Satellite Observation on April 16^{th}, 1800 hrs.

Validation of Posterior Estimate of Source Parameters

The prior values for the source parameters used in this study were estimated based on the limited data that was available immediately following the eruption, and provided only a rough guide to true values, but nevertheless reflect the type of data that may be available at the time of eruption. Since the eruption, further studies have been completed and better estimates of the source parameters have become available. We compare these independent estimates of source parameters with the posterior mean estimates obtained here and reported

in Table 4.3.

The vent radius was estimated from an airborne IMO radar image of the Earth's surface in the summit region taken on 14 April 2010 at 1030 UTC during the paroxysmal phase of the eruption, based on our own image analysis. Radar imagery is useful for this because of the ability of radio waves to penetrate eruption clouds. We originally assumed three of the darkest areas on the image to be craters. Later image guidance provided by the Icelandic Institute of Earth Sciences (`http://earthice.hi.is/eruption_eyjafjallajokull_2010`), however, suggests the presence of five roughly elliptical craters at that time, ranging in equivalent circular radius from 21 to 119 m (Table 4.3). Assuming that pressure balance in the plume as it exited the crater(s) developed rapidly based on the lack of atmospheric shocks in videography, and that crater diameter reflects pressure balance, the posterior estimate of 87 m eruption radius suggests that one of the two larger craters was active during the paroxysmal phase of the eruption on the morning of 14 April. This result is consistent with observations from other eruptions that vent activity migrates with time, and that the active vent during the most vigorous phase of an eruption should be that which allows the greatest flux. Measures of the initial velocity are now available for the initial, vigorous 14–18 April phase of the eruption based on a video of the erupting plume that was analyzed using a vortex tracking algorithm. The velocity near the vent was found to correlate with the relative vigor of the discharge from the volcano and plume height, and estimated to be 20–30 m/s on 17 April [119]. Given that this measurement was made slightly above the vent on the outer margin of the plume, in a rapidly decelerating section of the plume, it is probable that this observed velocity is slightly lower than the true exit velocity. The observed velocity of 20–30 m/s slightly above the vent thus is in accord with the mean value for the posterior mean exit velocity of 54m/s, which is on the lower end of the prior range. The grain size distribution was studied mainly for the second intensive phase of the eruption (early May), during which time activity was similar to that seen in the early phase from 14–18 April. From these observations, it was found that the mean grain size, Md_φ, changed from -0.9 to 4.5φ and the standard deviation, σ_φ, from 0.7 to 2.6 φ with distance from the vent for deposits from the cloud found on land [120]. In fact, at a distance of 44km from the vent and for the next $12km$, the observations shown a quasi-constant mean size of 4.5φ and a σ_φ of 2.6. This distance is within one computational cell from the source, and therefore grain sizes measured there should represent initial conditions. Furthermore, once the size of grains falling to the

ground becomes constant, it can be assumed that the depositing grain size reflects the grain size of particles left in the cloud. If these assumptions are valid, the posterior estimate of the initial grain size distribution of $4.96 \pm 2.62\varphi$ correlates well with the measured value of $4.5 \pm 2.6\varphi$.

Table 4.3: Comparison of prior, posterior and new estimates of eruption source parameters based on observations of Eyjafjallajökull eruption and simulations.

Parameter	Prior range	Posterior mean	New	Reference
Vent radius, b_0 (m)	65-150	87	21, 65, 119, 31, 32	Remeasured based on better image guidance
Vent velocity, w_0 (m/s)	45-124	54	$> 20 - 30$	[119]
Mean grain size, Md_φ (φ)	3.5-7	4.96	4.5	[120]
σ_φ, φ	$0.5 - 3$	2.62	2.6	[120]

Note that the grain size unit of geology, φ, is defined as: $\varphi = -\log_2(D/D_0)$, where diameter, D, is measured in mm, and reference diameter $D_0 = 1$ mm.

4.5 Summary

In this chapter, application of developed methodology in previous two chapters to probabilistic forecasting and source parameter estimation of volcanic ash transport is outlined. The method of quadrature points (using recently developed CUT points) is used to propagate parameter uncertainty through the `bent-puff` model. The quadrature ensembles are then used to construct a polynomial chaos surrogate model which is then sampled to provide probabilistic hazard map for ash top-height. Furthermore, the method of quadrature points in conjunction with the minimum variance unbiased linear estimation approach is used to fuse `bent-puff` model forecasts and satellite observational data, to find a posterior estimate of source parameters and to update coefficients of polynomial chaos surrogate model. The updated polynomial chaos surrogate model is used to obtain posterior distribution of source parameters. This methodology is implemented and validated using the 2010 Eyjafjallajökull volcanic eruption as a benchmark problem. Numerical simulations illustrate the computational efficiency of using the CUT method. The source parameter estimation method proposed here provides not only mean estimates, but also a statistical confidence bound for

that mean. Validation of the simulation results shows that the posterior estimate of source parameters corresponds well with values obtained in other references. Hazard maps based on our approach accurately forecast the location of ash, when tested against satellite data.

In this work, we have used the NOAA NCEP Reanalysis 1 wind field to compute the hazard map. The Reanalysis windfield uses observation data to produce a "best" known realization of the wind field consistent with data. Uncertainty introduced into the wind forecast from the NWP model is significant, and incorporating this uncertainty into an enhanced model ensemble is the subject of ongoing work.

Finally, it is important to note that the overall framework for probabilistic model forecast and source estimation described here is not dependent on the choice of VATD or eruption model; other models can easily be used to generate column and plume outputs that are used in the subsequent uncertainty analysis.

Chapter 5

Optimal Information Collection

5.1 Introduction

There is no doubt that proper sensor placement is intimately tied to the source estimation and model uncertainty characterization. Poor data measurements caused by sparse or scattered sensors over the domain of interest results in a poor estimate of source location. Hence, finding optimal locations of the sensors is very important for accurate source parameter estimation. Due to the dynamics of plume dispersion phenomenon, it is much more efficient to apply mobile sensors, instead of static sensors for data monitoring purposes. Different strategies have been suggested to determine the optimal path of mobile sensors for source parameter characterization of the plume dispersion phenomenon. These methods can be categorized as Chemotaxis, Anemotaxis, and Fluxotaxis.

In chemotaxis approach [13,14], mobile sensors follow the concentration gradient. Therefore, the direction of the largest concentration is the goal direction for the chemotaxis. Generally, this direction is estimated by seeking out a neighbor with the largest detected chemical concentration and computing the bearing toward this neighbor. The major problem with chemotaxis approach is that the mobile sensors always try to move in a direction of positive concentration gradient and hence, easily susceptible to be trapped in a local maxima and plateaus of the concentration.

In anemotaxis strategy [15, 16], mobile sensors always move upstream while being inside the plume. Hence, the upstream (upwind) is the goal direction for mobile sensors. The utilization of wind direction as the main control parameter of the mobile sensor's direction

restricts the applicability of this approach to an environment where the terrain map is relatively simple and the wind direction is stable. Also, the lack of communication among different neighbors in this approach restricts each mobile sensor to only local information about the environment, and the team can easily be trapped in local concentration maxima and plateaus.

In the fluxotaxis approach [17], the mobile sensors compute the amount of dispersal material flux passing through virtual surfaces formed by neighboring sensors. Each individual sensor independently calculates the amount of local material flux, relative to the current position of its neighbors. The major advantage of this approach is that it combines information about both fluid velocity and material density to find the true source location. However, the lack of collaboration between mobile sensors is the primary drawback of this approach. Each mobile sensor can only make decisions based on the local information it sensed from its surrounding environment.

The major deficiency of aforementioned approaches lies in their restriction to continuous releases phenomena. In other words, none of chemotaxis, anemotaxis, and fluxotaxis methods are applicable in presence of spontaneous releases. This is due to the presence of discontinuities in plume flow which hinders evaluation of concentration gradient or flux of plume during the time. Furthermore, they do not take into account the information available through the numerical model forecasts.

Besides aforementioned methods, recently multiple researches have been done on the field of optimal sensor placement, based on information theory. Choi *et al.* [121] studied continuous motion planning of mobile sensors for informative forecasting in presence of *linear time varying* systems and Gaussian uncertainty. Bourgault *et al.* [122] suggested a robotic exploration approach based on maximizing mutual information, while linearizing the dynamics and sensor model and assuming *Gaussian* uncertainty involved in the process.

Optimal sensor placement is also widely used in target tracking problem [10, 11, 123]. For instance, in [123], optimal sensor placement and motion coordination of mobile sensor networks are used to tackle target tracking problem. This is achieved by maximizing the Fisher Information matrix, or equivalently minimizing the associated Cramer-Rao lower bound of parameter estimates. Tharmarasa *et al.* [10] studied the problem of selecting a small subset of the available sensors in a large network of sensors in order to track multiple targets. A search scheme based on combination of optimization methods and the Cramer-Rao

lower bound was used to perform this task. The major drawback of these works is that the Fisher Information matrix usually consists of parameters which are to be estimated. Hence, an estimate of unknown parameters is used during maximization of the Fisher Information matrix. Williams *et al.* [9] studied the problem of choosing optimal subset of sensors from a *stationary* sensor network, in presence of *linear* model dynamics and *linear* sensor model, while minimizing the communication cost. Approximate Dynamic Programming techniques were used to achieve this goal. The major drawback of this work is that it doesn't consider any model to represent sensor dynamics. Hence, a large *stationary* sensor network is needed (while a small portion of them is used at each time) to assure the performance of the proposed approach. As well, linearity assumptions in dynamic model and sensor model restrict applicability of this method. In a research by Julian *et al.* [11], an information theoretic framework was presented for distributed control of a set of mobile robots. The basic idea of this work is to move mobile robots along gradient of mutual information to maximize information collection. The major drawback of this approach is that it may not be effective in the case that initial positions of robots are far way from regions of interest (local optimality). Also, importance sampling techniques were used to calculate gradient of mutual information, which could result in computational delay. Hoffmann *et al.* [12] presented a control approach for mobile sensor networks, based on maximizing mutual information. In detail, he used a particle filter framework and Monte Carlo integration method for evaluation of mutual information. Also, an iterative approach was used at each time step to find optimal control signal for each mobile sensor. However, the use of Monte Carlo integration to evaluate mutual information can reduce its computational performance. Another drawback of this approach is that control policy at each time should be solved iteratively, which could result in computational cost.

The key idea of this chapter is to *optimally* locate data monitoring sensors over the spatial domain of interest such that the uncertainty involved in source parameter estimates is minimized. This has been achieved by maximizing the mutual information between the model output and data measurements. A dynamic programming based approach is used to maximize the mutual information between uncertain parameters and observational data. As it will be shown, proposed approach expedites the convergence of estimation process and avoids possible local optimalities while finding mobile sensor locations by maximizing the mutual information, rather moving along its gradient. In other words, proposed approach

moves the mobile sensors toward the plume, even though the sensors are initially located far from the plume area. Along proposed approach for optimal sensor placement, a combination of generalized Polynomial Chaos (gPC) and Bayesian inference is used for data assimilation process which allows us to apply our method in presence of nonlinear dynamics and sensor model and non-Gaussian uncertainties, without using any Monte Carlo sampling. As well, a set of recently developed quadrature points, named as Conjugate Unscented Transform points [44, 46, 47, 49], are used to alleviate the computational complexities associated with evaluation of mutual information, uncertainty propagation, and estimation process. Fig. 5.1 represents schematic view of optimal information collection in estimation process and its connection with other components.

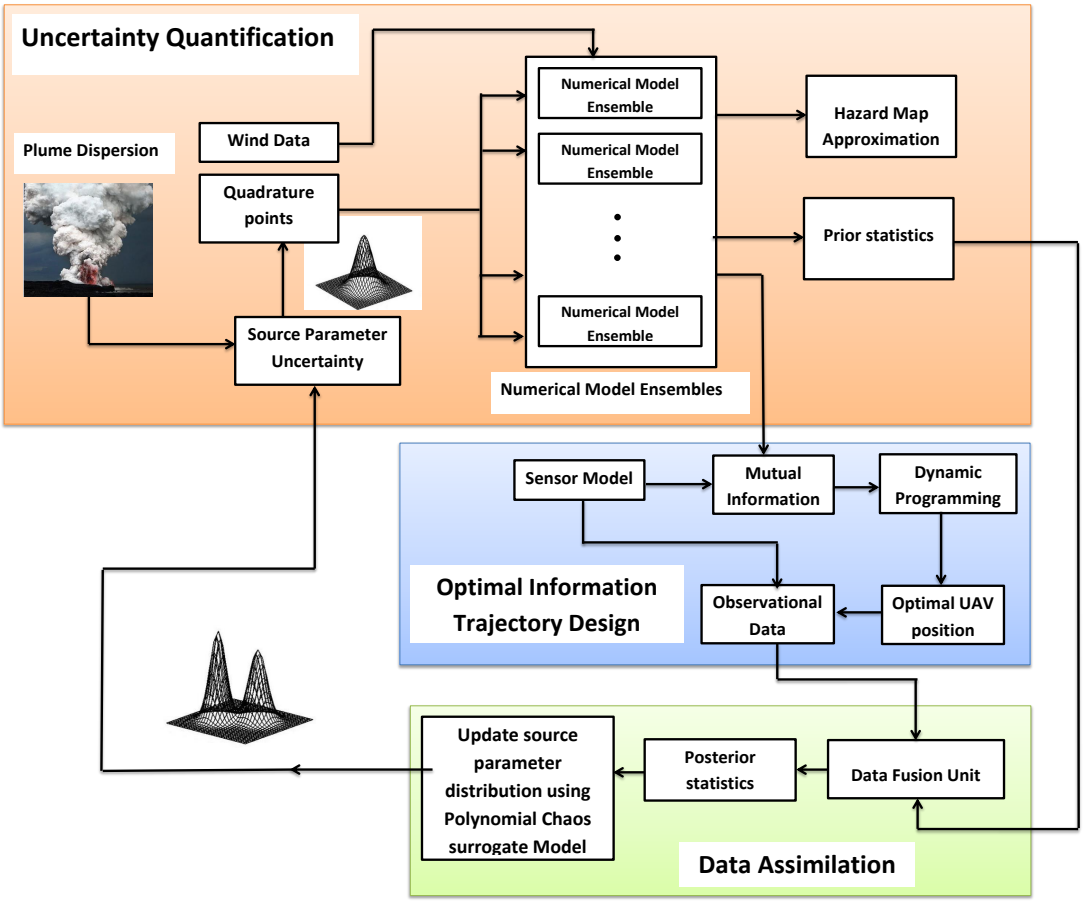


Figure 5.1: Schematic view of Optimal Information Collection estimation process and its connection with other components

5.2 Dynamic Data Monitoring (DDM)

One of the main challenges in data monitoring of large scale dynamical systems, like plume dispersion phenomenon, is to effectively observe the system state at each measurement time step such that it improves accuracy of estimation. This can be done by *optimally* allocating a set of mobile sensors (Unmanned Aerial Vehicles) such that measurements with *more information content* are sought at each time step. Optimal location of each mobile sensor at each time step can be achieved by maximizing the mutual information between model predictions and observed data. To explain this procedure in more detail, assume a set of Unmanned Aerial Vehicles (UAV) which are used for data observation. Each of the UAVs is equipped with a chemical sensor to measure the concentration of pollutant material and the dynamic model of each of the UAVs is given as:

$$\mathbf{s}_{k+1}^{v} = F(\mathbf{s}_k^v, u_k^v), \quad v = 1, 2, \cdots, N_u \tag{5.1}$$

where N_u is total number of UAVs and k denotes k^{th} time step. Also, initial condition of v^{th} UAV is assumed to be given as $\mathbf{s}_0^v$. The state of UAV, $\mathbf{s}_k^v$, consists of (x, y, z) components of position and heading angle. Our objective is to find a sequence of control inputs $U^{1:N_u} = \{u_0^{1:N_u}, u_1^{N_u}, \cdots, u_{N_t-1}^{1:N_u}\}$ such that it maximizes the mutual information between the sequence of expected observational data and the source parameters Θ over the time $t \in [t_1, t_{N_t}]$. Note that due to independence of measurement data from individual sensors or UAVs, the problem of maximizing mutual information can be stated as:

$$\begin{aligned} \min_{U^{1:N_u}=\{u_0^{1:N_u}, u_1^{1:N_u}, \cdots, u_{N_t-1}^{1:N_u}\}} & J(\mathbf{s}_0^{1:N_u}) = \\ \min_{U^{1:N_u}=\{u_0^{1:N_u}, u_1^{1:N_u}, \cdots, u_{N_t-1}^{1:N_u}\}} & \sum_{k=0}^{N_t-1} \left(\sum_{v=1}^{N_u} \left\{ -I(\mathbf{\Theta}^k; \mathbf{z}^k | \mathbf{s}_k^v) + u_k^{vT} B u_k^v \right\} \right) \end{aligned} \tag{5.2}$$

constrained to

$$\begin{cases} \mathbf{s}_{k+1}^v = F(\mathbf{s}_k^v, u_k^v), \quad \mathbf{s}_0^v \qquad v = 1, 2, \cdots, N_u \\ s_k^a \neq \mathbf{s}_k^b, \quad a, b = 1, 2, \cdots, N_u, \quad a \neq b \end{cases} \tag{5.3}$$

In above equations, $\mathbf{z}^k$ and Θ^k are the measurement data and uncertain source parameter at time step t_k, respectively. $I(\mathbf{\Theta}^k; \mathbf{z}^k | \mathbf{s}_k^v)$ represents the *Mutual Information* between the measurement data and source parameter, given v^{th} UAV position at time t_k. B is a constant

positive definite matrix. N_t is the total number of time steps during the estimation process and Δ_t is the time interval between every two consecutive measurement updates. $\mathbf{s}_0^{1:N_u}$ represents initial conditions and $u_k^{1:N_u}$ denotes applied control signal at time t_k for a set of N_u UAVs. Note that the first term in Eq. (5.2) maximizes the mutual information between measurements and source parameters, while the second term corresponds to minimizing the amount of energy used by each UAV. One should notice that Eq. (5.2) is a function of a sequence of control signals $U^{1:N_u} = \{u_0^{1:N_u}, u_1^{1:N_u}, \cdots, u_{N_t-1}^{1:N_u}\}$ which are enforced on a set of N_u UAVs during the time. Also, the cost function J is a function of initial position of UAVs, i.e. $J = J(\mathbf{s}_0^{1:N_u})$. Hence, the above optimization shall be performed for every possible combination for initial positions of N_u UAVs. As shown in Eq. (5.3), the optimization is constrained to the dynamic model for each UAV. Additionally, the second constraint in Eq. (5.3) is considered to avoid redundant observations at the same time instance. In the following, we first explain evaluation of mutual information $I(\boldsymbol{\Theta}^k; \mathbf{z}^k | \mathbf{s}_k^v)$ and then mathematical details for minimization of Eq. (5.2) are presented.

5.2.1 Mutual Information as a measure for sensor performance

According to information theory, the mutual information between source parameter $\boldsymbol{\Theta}$ and measurement $\mathbf{z}$ can be written as [124]:

$$I(\boldsymbol{\Theta}; \mathbf{z}) = \int_{\mathbf{z}} \int_{\boldsymbol{\Theta}} p(\boldsymbol{\Theta}, \mathbf{z}) \ln \left(\frac{p(\boldsymbol{\Theta}, \mathbf{z})}{p(\boldsymbol{\Theta}) p(\mathbf{z})} \right) d\boldsymbol{\Theta} d\mathbf{z} \tag{5.4}$$

Using Bayes' rule, $p(\boldsymbol{\Theta}, \mathbf{z})$ can be written as:

$$p(\boldsymbol{\Theta}, \mathbf{z}) = p(\mathbf{z}|\boldsymbol{\Theta}) p(\boldsymbol{\Theta})$$

Hence $I(\boldsymbol{\Theta}; \mathbf{z})$ will be equal to:

$$\begin{aligned} I(\boldsymbol{\Theta}; \mathbf{z}) = \int_{\mathbf{z}} \int_{\boldsymbol{\Theta}} p(\mathbf{z}|\boldsymbol{\Theta}) p(\boldsymbol{\Theta}) ln \left(\frac{p(\mathbf{z}|\boldsymbol{\Theta}) p(\boldsymbol{\Theta})}{p(\boldsymbol{\Theta}) p(\mathbf{z})} \right) d\boldsymbol{\Theta} d\mathbf{z} = \\ \int_{\mathbf{z}} \int_{\boldsymbol{\Theta}} p(\mathbf{z}|\boldsymbol{\Theta}) p(\boldsymbol{\Theta}) ln \left(p(\mathbf{z}|\boldsymbol{\Theta}) \right) d\boldsymbol{\Theta} d\mathbf{z} - \int_{\mathbf{z}} p(\mathbf{z}) ln \left(p(\mathbf{z}) \right) d\mathbf{z} \end{aligned} \tag{5.5}$$

Mutual Information can be understood as the reduction in uncertainty. Hence, by maximizing mutual information one inherently reduces the uncertainty thus leading to a better measurement and estimate. It is to be noted that the mutual information is evaluated before

the actual measurement is made. The likelihood pdf $p(\mathbf{z}|\mathbf{\Theta})$ and the measurement pdf $p(\mathbf{z})$ that are evaluated from $p(\mathbf{\Theta})$ are used to compute the Mutual Information.

5.2.2 Principle of Optimality

Solution of Eq. (5.2) is based on a very simple idea, which is called *principle of optimality*. Principle of optimality states that if a policy $U^* = \{u_0^*, u_1^*, \cdots, u_{N_t-1}^*\}$ is an optimal policy for problem Eq. (5.2), then truncated policy $\{u_k^*, u_{k+1}^*, \cdots, u_{N_t-1}^*\}$ is optimal for the time interval $[k, N_t - 1]$.

To describe the intuition behind the principle of optimality, lets consider the problem of finding shortest distance from point c_1 to c_9. This has been shown in Fig. 5.2. Assume that we are looking for the shortest path between the starting point c_1 and destination c_9, which is shown by black line in Fig. 5.2. Based on principle of optimality, if we find the optimal path between starting point and destination, then by starting from any point on the path, the rest of the path is also optimal. For instance, given the path passing through the points $c_1 - c_2 - c_5 - c_9$ being the shortest distance between c_1 and c_9. Then the shortest path from c_2 to c_9 is given by $c_2 - c_5 - c_9$.

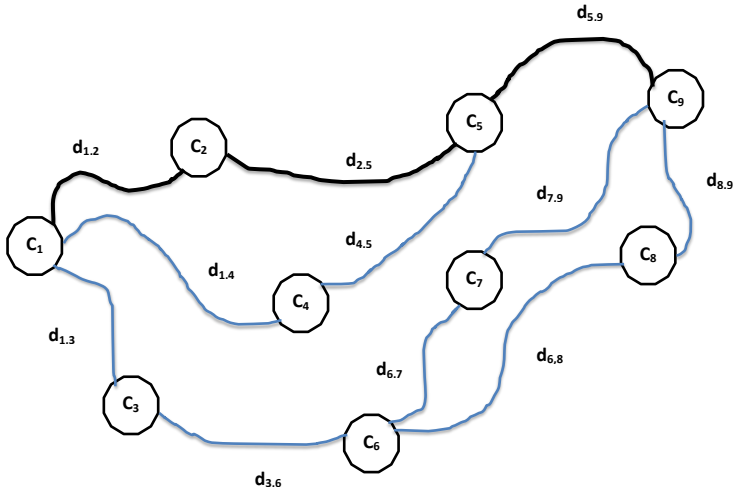


Figure 5.2: Principle of Optimality in the context of finding shortest distance between points c_1 and c_9. The optimal path between starting point and destination is also optimal between any middle point on the path and destination

5.2.3 Dynamic Programming

The simple, but powerful principle of optimality can be used in backward manner to find the optimal policy $U^{1:N_u}$. In detail, to solve Eq. (5.2), one can first find the optimal control signal at time $N_t - 1$ and then use it to find the optimal control signal at time step $N_t - 1$. This process can be performed backward in time, until we find corresponding optimal control signals at all time steps.

This technique is called dynamic programming [26] which is used to find the optimal policy $U^{1:N_u} = \{u_0^{1:N_u}, u_1^{1:N_u}, \cdots, u_{N_t-1}^{1:N_u}\}$ that minimizes Eq. (5.2). According to Dynamic Programming method, optimal policy $U^{1:N_u}$ is computed by backward optimization in time, i.e. first finding the optimal control input $u_{N_t-1}^{1:N_u}$, and then using $u_{N_t-1}^{1:N_u}$ to find optimal control signal $u_{N_t-2}^{1:N_u}$. This procedure will be repeated recursively to find the rest of the control signals u_k^vs. Generally, the following *recursive* algorithm [26] can be used to find optimal policy U^v that minimizes Eq. (5.2) for each UAV:

Given an initial position $\mathbf{s}_0^{1:N_u}$, the optimal cost $J^*(\mathbf{s}_0^{1:N_u})$ is equal to $J_0(\mathbf{s}_0^{1:N_u})$, given by the last step of the following algorithm, which proceeds backward in time from period $N_t - 1$ to 0:

$$J_{N_t}(\mathbf{s}_{N_t}^{1:N_u}) = 0 \tag{5.6}$$

$$J_k(\mathbf{s}_k^{1:N_u}) = \min_{u_k^{1:N_u} \in U_k^{1:N_u}(\mathbf{s}_k)} \left\{ g_k(\mathbf{s}_k^{1:N_u}, u_k^{1:N_u}) + J_{k+1}(\mathbf{s}_{k+1}^{1:N_u}) \right\}, \quad k = N_t - 1, N_t - 2, \cdots, 0 \tag{5.7}$$

constrained to

$$\begin{cases} \mathbf{s}_{k+1}^v = F(\mathbf{s}_k^v, u_k^v), \quad \mathbf{s}_0^v \qquad v = 1, 2, \cdots, N_u \\ s_k^a \neq \mathbf{s}_k^b, \quad a, b = 1, 2, \cdots, N_u, \quad a \neq b \end{cases} \tag{5.8}$$

where,

$$g_k(\mathbf{s}_k^{1:N_u}, u_k^{1:N_u}) = \sum_{v=1}^{N_u} \left\{ -I(\mathbf{\Theta}^k; \mathbf{z}^k | \mathbf{s}_k^v) + u_k^{vT} B u_k^v \right\} \tag{5.9}$$

As mentioned before, this process should be performed for each of the possible values for initial condition $\mathbf{s}_0^v$. After finding the optimal policy U^v, it can be used to optimally locate

the corresponding UAVs such that the mutual information between source parameters and measurement data is maximized.

Implementation of above algorithm using the continuous dynamics of the UAV in Eq. (5.1) requires excessive computational resources and is not achievable in real time. Hence to make the problem tractable, the spatial domain has been discretized into a uniform grid. The mutual information is now only evaluated at these grid nodes and the UAV motion is restricted to these nodes. This lead to a more tractable sub-optimal solution that is implementable in real time. Unfortunately, the problem complexity grows exponentially with the number of UAVs. For instance, even in presence of 2 UAVs and a $100 \times 100 \times 10$ spatial grid points, there will be 10^{10} possible combinations for positions of UAVs for which $J(\mathbf{s}_0^{1:2})$ needs to be minimized. Hence, an enormous computational effort is required to perform such minimization. A simpler alternative approach to overcome these deficiencies in minimizing Eq. (5.2) is to recursively find *sub-optimal* policies for each one of the UAVs individually with slight modifications in original cost function. The idea is to first find optimal position for the first UAV during the time. Then, the *sub-optimal* policies for all other UAVs can be found by minimizing the following modified cost function:

$$\min_{U=\{u_0^v, u_1^v, \cdots, u_{N_t-1}^v\}} J(\mathbf{s}_0^v) \tag{5.10}$$

$$\text{constraint to} = \begin{cases} \mathbf{s}_{k+1}^v = F(\mathbf{s}_k^v, u_k^v) \\ \mathbf{s}_0^v \end{cases} \tag{5.11}$$

where,

$$J(\mathbf{s}_0^v) = \sum_{k=0}^{N_t-1} \left\{ -I(\mathbf{\Theta}_k, \mathbf{z}_k | \mathbf{s}_k^v) + u_k^{vT} B u_k^v + \alpha \sum_{j=1}^{v-1} e^{-[\mathbf{s}_k^v - \mathbf{s}_k^j]^T W [\mathbf{s}_k^v - \mathbf{s}_k^j]} \right\}, \quad v = 2, 3, \cdots, N_u \tag{5.12}$$

with $\alpha > 0$ and the positive definite diagonal matrix W being penalty factors that determine the separation between neighboring UAVs. Hence, the UAVs are made to *spread out* in the spatial domain of interest thus avoiding redundancy in measurements .

One should note that the computational cost involved in recursively finding sub-optimal locations of UAVs is far less than the computational cost involved in solving Eq. (5.7) to Eq. (5.9). For comparison, consider a case where N_u UAVs are used to perform data observation over a $N_g \times N_g \times N_g$ spatial grid points. In this case, the computational cost involved

in solving Eq. (5.7) to Eq. (5.9) will be proportionate to $N_g^{3N_u}$, while the computational cost involved in finding sub-optimal locations of UAVs using Eq. (5.10) to Eq. (5.12) is proportionate to $N_u N_g^3$. Hence, solving Eq. (5.10) results in significantly less computational with respect to solving Eq. (5.7), especially in presence of larger number of UAVs.

5.2.4 Limited Lookahead Policy

Depending on the nature of phenomenon under study, evaluation of mutual information map for all the future time steps can be computationally expensive. This will result in computational complexity while finding optimal control policies for the UAVs during source parameter estimation process. Hence, using Eq. (5.12) restricts real time applications of proposed algorithm. One way to avoid these computational complexities is to approximate the true cost-to-go function J_{k+1} in Eq. (5.12) with some function, denoted by $\tilde{J}_{k+1}$, which is a *limited lookahead* approximation of true cost-to-go function J_{k+1}. For instance, in Eq. (5.12), J_{k+1} can be approximated as:

$$J_{k+1}(\mathbf{s}_k^v, u_k^v) \simeq \tilde{J}_{k+1}(\mathbf{s}_k^v, u_k^v) = \sum_{i=k+1}^{k+1+l} \left\{ -I(\boldsymbol{\Theta}_i, \mathbf{z}_i | F(\mathbf{s}_{i-1}^v, u_{i-1}^v)) + u_i^{vT} B u_i^v + \alpha \sum_{j=1}^{v-1} e^{-[\mathbf{s}_i^v - \mathbf{s}_i^j]^T W [\mathbf{s}_i^v - \mathbf{s}_i^j]} \right\}, \quad v = 2, 3, \cdots, N_u \tag{5.13}$$

where, l is the number of future time steps which are used for approximation of true cost-to-go function J_{k+1}. As one can see, evaluation of Eq. (5.12) requires knowledge of mutual information for all the time steps between $k+1$ and $N_t - 1$. While in *limited lookahead* method, J_{k+1} is approximated by a limited number of future time steps. Similarly, J_{k+1} in Eq. (5.7) can be approximated as:

$$J_{k+1}(\mathbf{s}_k^v, u_k^v) \simeq \tilde{J}_{k+1}(\mathbf{s}_k^v, u_k^v) = \sum_{i=k+1}^{k+1+l} -I(\boldsymbol{\Theta}_i, \mathbf{z}_i | F(\mathbf{s}_{i-1}^v, u_{i-1}^v)) + u_i^{vT} B u_i^v, \quad v = 1 \tag{5.14}$$

For plume dispersion applications, making use of limited lookahead policy has two major benefits with respect to the original dynamic programming algorithm. First, as we mentioned before, depending on the problem in hand, limited lookahead policy could result in considerably less computational cost involved in finding *sub-optimal* control policies. The second benefit of limited lookahead policy is that due to dependence of optimal policies on

future wind data, the optimal policies obtained using the original cost function may be erroneous for distant future time steps. Hence, using limited lookahead policy avoids erroneous optimal policies by approximating the true cost-to-go function with limited number of future time steps.

The only drawback of using limited lookahead policy is that it may be slower in plume detection, especially if the UAVs are located far away from the mutual information map. To illustrate this more clearly, consider the situation shown in Fig. 5.3. As one can see, if initial position of the UAV is far from mutual information map, given maximum speed of UAV and lookahead time step $l = 3$, there will not be any information in range of the UAV. Hence, the UAV does not move and proposed algorithm suggests that the UAV should stay at the same position during the time.

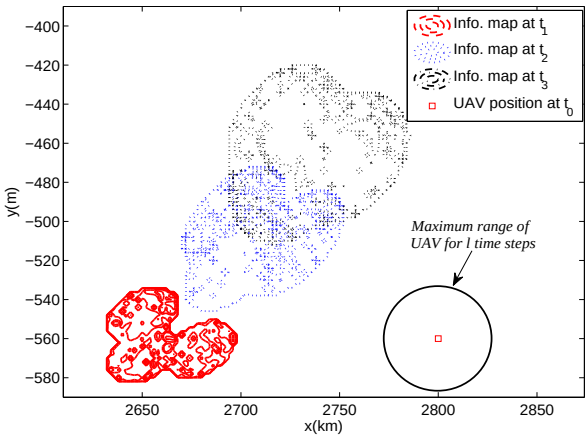


Figure 5.3: Schematic layout of applied UAV sensor and mutual information maps at three consecutive times. Red square shows initial position of UAV and black circle shows its maximum range, given maximum speed of UAV and $l = 3$.

This drawback can be overcome by minimizing the distance between the UAV position and mutual information map, whenever the mutual information inside the range of UAV is zero. In this way, the UAV will move toward the mutual information map, even if mutual information map is out of range of UAV in l time step. This can be mathematically described as the following:

If $\sum_{i=k+1}^{k+1+l} I(\boldsymbol{\Theta}_i, \mathbf{z}_i | F(\mathbf{s}_{i-1}^v, u_{i-1}^v)) = 0$, then

$$\tilde{J}_{k+1}(\mathbf{s}_k^v, u_k^v) = \sum_{i=k+1}^{k+1+l} d(\mathbf{s}_{I_{max}}^i, F(\mathbf{s}_{i-1}^v, u_{i-1}^v)), \quad v = 1 \tag{5.15}$$

and

$$\tilde{J}_{k+1}(\mathbf{s}_k^v, u_k^v) = \sum_{i=k+1}^{k+1+l} d(\mathbf{s}_{I_{max}}^i, F(\mathbf{s}_{i-1}^v, u_{i-1}^v)) + \alpha \sum_{j=1}^{v-1} e^{-[\mathbf{s}_i^v - \mathbf{s}_i^j]^T [\mathbf{s}_i^v - \mathbf{s}_i^j]}, \quad v = 2, 3, \cdots, N_u \tag{5.16}$$

where,

$$d(\mathbf{s}_{I_{max}}^i, F(\mathbf{s}_{i-1}^v, u_{i-1}^v)) = ||\mathbf{s}_{I_{max}}^i - F(\mathbf{s}_{i-1}^v, u_{i-1}^v)||_2 \tag{5.17}$$

is the Euclidean distance between the spatial location where mutual information obtains its maximum (denoted by $\mathbf{s}_{I_{max}}^i$) and location of UAV. Using above algorithm, the UAVs always move toward the mutual information map, independent of initial location of UAVs or lookahead time step l. Note that this property of proposed approach guarantees faster detection of the plume and consequently faster convergence of estimation process.

5.2.5 Simplification of Limited Lookahead Policy

Computational cost of proposed limited lookahead method can be further alleviated by finding optimal control policies just for the spatial points within the maximum range of each UAV. The intuition behind this idea is that since each UAV has a limited range during a specified time, there is no need to find optimal control policies for the points outside the range of the UAV. In this way, there is no need to minimize Eq. (5.13) and Eq. (5.14) over the whole spatial domain. Hence, the computational cost involved in minimizing Eq. (5.13) and Eq. (5.14) will significantly reduce. This can be easily seen in Fig. 5.3. As one can see, the number of spatial points within the maximum range of the UAV is far less than the total number of spatial grid points. Hence, the computational cost involved in finding optimal policy for the points within the maximum range of the UAV will be significantly less than the computational cost required for finding optimal policy for all spatial points.

5.3 Numerical Simulations

To validate the performance of proposed approach two different test cases are implemented. In the first test case, dispersion/ advection of Propane is simulated over New York city. North

American Regional Reanalysis wind data is used for simulation of dispersion phenomenon in this case. Three sources are considered for this scenario where the amount of mass release is assumed to be uncertain. For the second test case, a more complicated dispersion/advection scenario is simulated. For this case, dispersion of Chlorine is simulated over the city of Denver. A three dimensional wind field, accessed from North American Regional Reanalysis data base is used to perform the simulation. Also, a continuous mass release is considered where its location and amount of mass release are uncertain. SCIPUFF numerical model is used to simulate dispersion/advection in all these test cases. Also, a discretized chemical sensor is utilized to perform the task of data observation for all the examples. In the following, we first explain the sensor model in detail and then the description of each example is given.

5.3.1 Sensor Model

The sensor used for the measurements is a bar sensor with discrete numbers of bars used in [80], with slight differences. The number of bars ranges from zero to fifteen. These bar readings indicate the concentration magnitude at the sensor location at the instant; the sensor displays $z = 0, \cdots, 15$, bars when the internal continuous-valued concentration magnitude $\mathbf{x}_{int}$ is between thresholds T_z and T_{z+1}, where $0 \leq T_z < T_{z+1}$. The thresholds T_z's are defined on a logarithmic scale, i.e. $T_z \in \{0, 5\times10^{-14}, 10^{-13}, 5\times10^{-13}, 10^{-12}, \cdots, 5\times10^{-7}\}$.

Properties of the sensor are determined by these thresholds and the properties of $\mathbf{x}_{int}$, which is *assumed* to be normally distributed about the true concentration $\mathbf{x}$ [80]. Measurement error $v = \mathbf{x}_{int}$ - $\mathbf{x}$ may be considered a combination of multiplicative noise and additive noise with mean zero and standard deviation $\sqrt{R(\mathbf{x})} = \sigma(\mathbf{x}) = a\mathbf{x} + b$ where a is the proportionality constant and b accounts for the thermal motion of the electrons in the components [80]. In practice, because the true value of $\mathbf{x}$ is never known, $\sigma(\mathbf{x})$ is usually approximated by $\sigma(\tilde{\mathbf{x}})$, where $\tilde{\mathbf{x}}$ is an estimate of $\mathbf{x}$. a is equal to 1 and $b = 10^{-15}$ in our simulations. Also, it is assumed that $\tilde{\mathbf{x}} = T_z$, where T_z is the sensor bar corresponding to $\mathbf{x}_{int}$. Probability density function of $\mathbf{x}_{int}$ given the corresponding concentration $\mathbf{x}$ is:

$$p(\mathbf{x}_{int}|\mathbf{x}) = \mathcal{N}(\mathbf{x}_{int}; \mathbf{x}, R) = \frac{1}{\sqrt{2\pi R}} e^{-\frac{(\mathbf{x}_{int}-\mathbf{x})^2}{2R}} \tag{5.18}$$

where, $\mathcal{N}(,;.,.)$ denotes a Gaussian probability density function with the mean and variance specified by its second and third arguments. Strictly speaking, $p(\mathbf{x}_{int}|\mathbf{x})$ is not a Gaussian distribution because it is only defined for non-negative values of $\mathbf{x}_{int}$. Following Eq. (5.18),

likelihood function, or simply probability of $\mathbf{z}$ conditioned on $\mathbf{x}$, is determined by the following integral:

$$P(\mathbf{z}|\mathbf{x}) \propto \int_{T_z}^{T_{z+1}} p(\mathbf{x}_{int}|\mathbf{x}) d\mathbf{x}_{int}, \quad \sum_{z} P(\mathbf{z}|\mathbf{x}) = 1 \tag{5.19}$$

As mentioned before, sensor readings are polluted with some error, due to internal random variable and quantization effect. It is clear that when the number of thresholds goes to infinity and the length of intervals between two thresholds goes to zero, the sensor output becomes continuous and the likelihood function reduces to the (truncated) Gaussian distribution defined by Eq. (5.18). When the uncertainty due to internal random variable vanishes, likelihood function is flat over the interval $[T_z, T_{z+1}]$ associated with $\mathbf{z}$ and zero elsewhere.

Note that due to discretization involved in sensor output, the mutual information in Eq. (5.5) will be written as:

$$I(\boldsymbol{\Theta}; \mathbf{z}) = \sum_{z=0}^{N_z} \sum_{q=1}^{M} w_q \Gamma(T_{z+1}, T_z, \boldsymbol{\Theta}(\boldsymbol{\xi}^q), R) \ln \left(\Gamma(T_{z+1}, T_z, \boldsymbol{\Theta}(\boldsymbol{\xi}^q), R) \right) -$$
$$\sum_{z=0}^{N_z} \left(\sum_{q=1}^{M} w_q \Gamma(T_{z+1}, T_z, \boldsymbol{\Theta}(\boldsymbol{\xi}^q), R) \right) \ln \left(\sum_{q=1}^{M} w_q \Gamma(T_{z+1}, T_z, \boldsymbol{\Theta}(\boldsymbol{\xi}^q), R) \right)$$

where,

$$\Gamma(T_{z+1}, T_z, \boldsymbol{\Theta}(\boldsymbol{\xi}^q), R) = \frac{1}{2} \left\{ \operatorname{erf} \left(\frac{T_{z+1} - \mathbf{x}(\boldsymbol{\Theta})}{\sqrt{2R}} \right) - \operatorname{erf} \left(\frac{T_z - \mathbf{x}(\boldsymbol{\Theta})}{\sqrt{2R}} \right) \right\}$$

and $N_z = 15$. Note that quadrature scheme is used to evaluate the integrals in $I(\boldsymbol{\Theta}; \mathbf{z})$.

5.3.2 UAV model

The UAVs are modeled by the following discrete equations

$$\begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ \lambda \end{bmatrix}_{k+1} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ \lambda \end{bmatrix}_{k} + \begin{bmatrix} u_k^{lon} \cos(\lambda_k + \frac{\pi u_k^{\lambda}}{2}) \\ u_k^{lat} \sin(\lambda_k + \frac{\pi u_k^{\lambda}}{2}) \\ 50 u_k^{z} \\ \frac{\pi u_k^{\lambda}}{2} \end{bmatrix} \tag{5.20}$$

where $(s_1, s_2, s_3)_k$ is (lat, lon, z) coordinate of each UAV on spatial domain and λ_k represents heading angle of the UAV at time t_k. Control input of each UAV is composed of four different signals, i.e.

$$U_k = [u_k^{lon}, u_k^{lat}, u_k^{z}, u_k^{\lambda}]^T \tag{5.21}$$

In above equation, u_k^{lon}, u_k^{lat}, and u_k^z determine displacement of UAV in longitude, latitude, and vertical directions, respectively; and u_k^λ determines heading angle for the UAV. In this article, for simplicity we assumed that $u_k^{lat} = u_k^{lon} \equiv u_k^h$. Hence, our simplified model will be written as:

$$\begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ \lambda \end{bmatrix}_{k+1} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ \lambda \end{bmatrix}_k + \begin{bmatrix} u_k^h \cos(\lambda_k + \frac{\pi u_k^\lambda}{2}) \\ u_k^h \sin(\lambda_k + \frac{\pi u_k^\lambda}{2}) \\ 50 u_k^z \\ \frac{\pi u_k^\lambda}{2} \end{bmatrix}, \qquad U_k = [u_k^h, u_k^z, u_k^\lambda]^T \tag{5.22}$$

Note that depending on dynamic of the problem, each of the elements of U_k can take different values. For instance, it is assumed that u_k^λ can take one of the following discrete values:

$$u_k^\lambda = \begin{cases} -1, & \textit{move toward south} \\ 0, & \textit{move toward east} \\ 1, & \textit{move toward north} \\ 2, & \textit{move toward west} \\ 3, & s_{s+1} = s_k \end{cases} \tag{5.23}$$

As mentioned in Eq. (5.23), different values of u_k^λ determine the direction of the UAV at each time step t_k. In this manuscript, it is always assumed that all elements of control signal U_k take discrete values. Other control signals (u_k^h and u_k^z) are described separately for each example.

5.3.3 Test case 1

In this case the dispersion/advection of propane is simulated over New York area. The domain of interest and the applied wind field (at one specific time) are shown in Fig. 5.4. Simulation time is considered to be 24 $hrs.$ starting from 00 : 00 of September 1^{st}, 2013. North American Regional Reanalysis wind data at pressure level 100 kpa (height $\simeq$ 100 $m.$) is used as the windfield for simulation. Three instantaneous mass releases are considered where their location is known and the only uncertain parameters are their amount of mass release. It is assumed that releases happen at the same time, i.e. all source releases happen at 00 : 00 of September 1^{st}. All mass releases are assumed to be uniformly distributed between

100 kg and 300 kg. Fig. 5.4 illustrates source locations and the windfield (at $t = 0$ $hrs.$) over the two dimensional spatial domain.

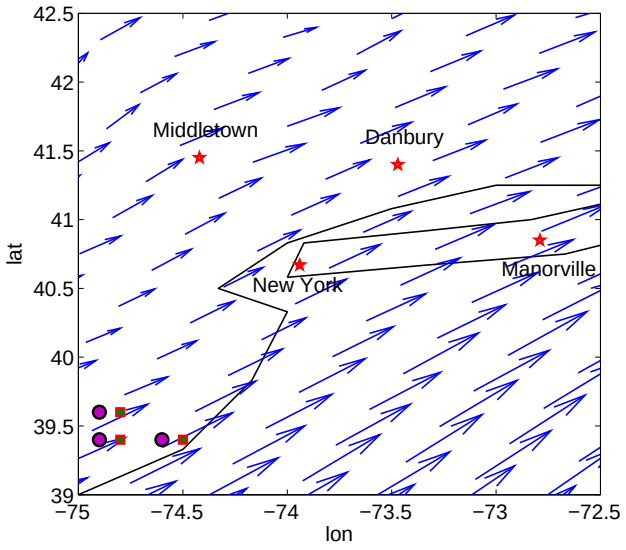


Figure 5.4: Schematic layout of Propane release over New York region, source locations are shown with purple circles, the wind-field (at $t = 0$ hr and pressure level = 100 kPa) is shown over the two dimensional domain with blue vector field. Position of stationary sensors (initial positions of UAVs) are shown with square markers. Geographical location of different cities can be seen on the background map

A set of 59 CUT8 quadrature points are used to quantify the uncertainty involved in concentration of dispersal material. Also, a 6^{th} order gPC expansion is used to reconstruct distribution of parameters after each update. Simulation of dispersion/advection has been performed using SCIPUFF numerical model, where concentration of propane is recorded every 10 $mins$.

Performance of the DDM approach is verified by comparing data assimilation results obtained by mobile and stationary sensors. Three sensors are used for data observation and a limited lookahead policy with $l = 6$ is used for finding location of each mobile sensor during the time. Note that $u_k^h \in \{2, 4, 6\}$ $km.$ and $u_k^z = 2$ in our simulations. Also, u_k^λ is given in Eq. (5.23). Hence, optimal way-points of mobile sensors are found by changing control variables u_k^h and u_k^λ. We considered $\alpha = 5$ and $W = diag([1, 1, 0])$ in our simulation. Fig. 5.3.3 illustrates positions of UAVs during the time over the spatial domain. It can be

observed from Fig. 5.3.3 that the UAVs follow the plume during the time and end up to the locations where the mutual information map obtains its maximum.

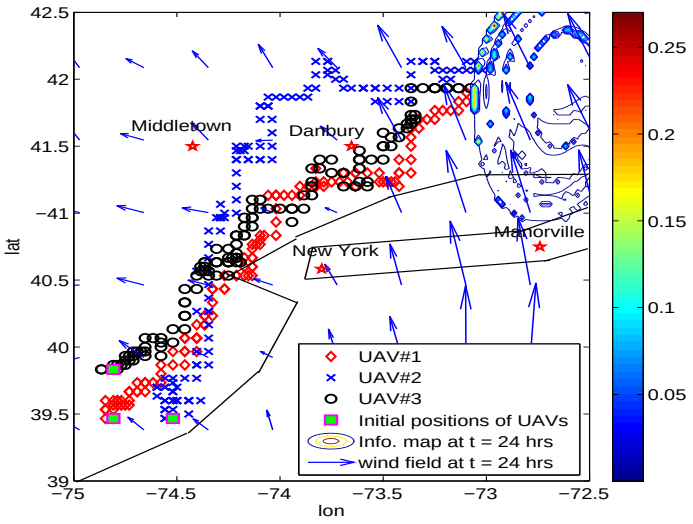
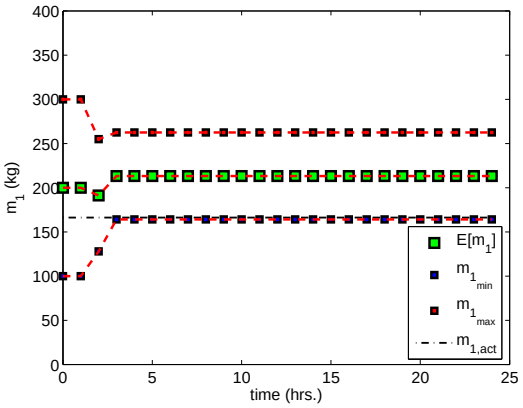


Figure 5.5: Test Case 1: Obtained way-points for mobile sensors at different time steps, contour map and corresponding colorbar represent the information map at the final time ($t_f = 24\ hr$). The windfield at $t = 24\ hr$ (pressure level = 100 kpa) and geographical map of the region are shown in the background.

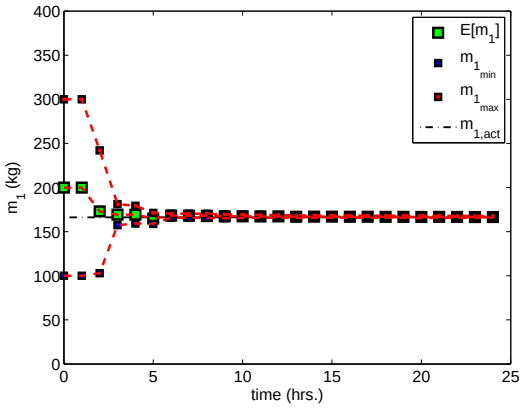
Convergence behavior for mean estimate of source parameters using stationary sensors, along with the minimum and maximum range of estimation are shown in Fig. 5.6. Fig. 5.6 illustrates that using stationary sensors results in poor estimates of the parameters. This is due to the inefficient placement of sensors during data assimilation process.

Convergence behavior for mean estimate of source parameters using mobile sensors are also illustrated in Fig. 5.6 where the DDM approach results in significantly better convergence for mean estimates of the parameters. Also, estimation process is very confident about the source parameters estimates.

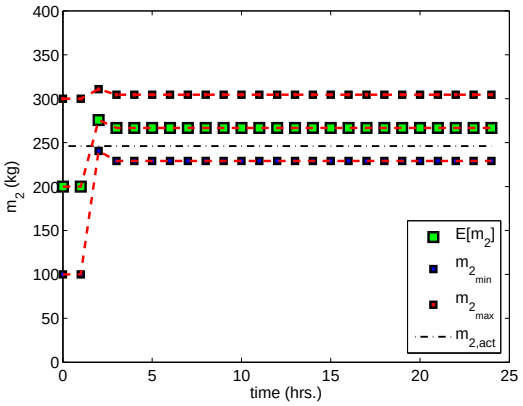
Table 5.1 shows the amount of information $I(\boldsymbol{\Theta}; \mathbf{z})$ collected by each sensor while using stationary and mobile sensors. It is clear from Table 5.1 that using Dynamic Data Monitoring method significantly increases the amount of mutual information collected by each sensor during the data assimilation process, and consequently improves the convergence behavior of the estimation process.



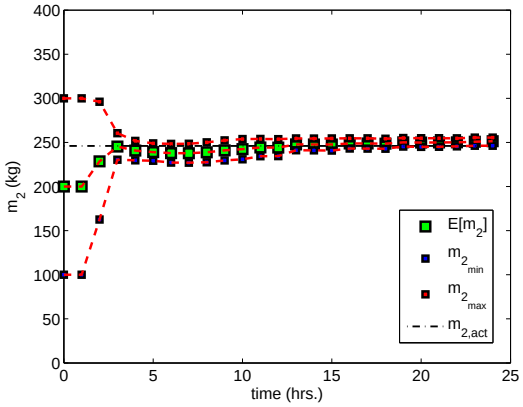
(a) convergence of m_1 using stationary sensors



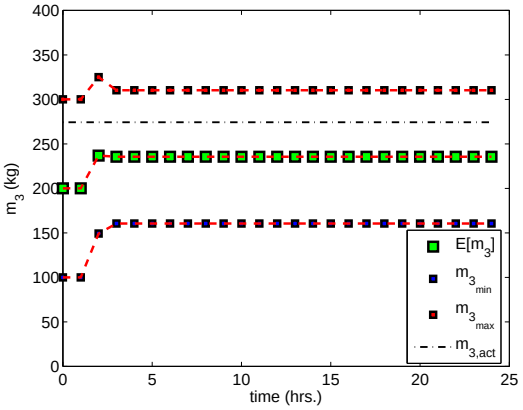
(b) convergence of m_1 using mobile sensors



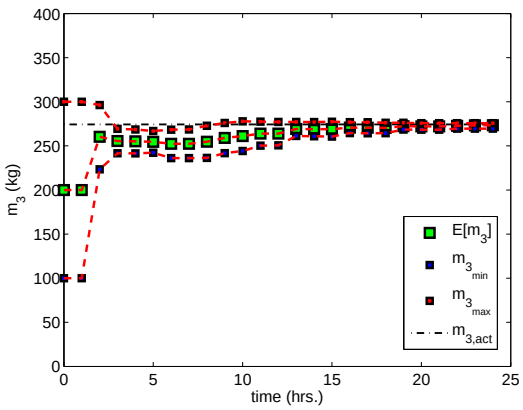
(c) convergence of m_2 using stationary sensors



(d) convergence of m_2 using mobile sensors



(e) convergence of m_3 using stationary sensors



(f) convergence of m_3 using mobile sensors

Figure 5.6: Test Case 1: Source parameter estimates during the time obtained using stationary and mobile sensors

Table 5.1: Test Case 1: Information collected by each sensor while with and without using Dynamic Data Monitoring approach.

sensor number	collected information	
	Stationary	Dynamic
sensor 1	1.8743	8.2208
sensor 2	1.4219	6.8420
sensor 3	1.4009	7.1927

Comparison with chemotaxis strategy

To further illustrate performance of proposed methodology, we have utilized chemotaxis strategy (i.e. moving UAVs along gradient of concentration field) to estimate uncertain source parameters. The same structure for mobile sensors is used while implementing chemotaxis strategy. Obtained way-points for mobile sensors can be seen in Fig. 5.3.3.

Fig. 5.8 illustrate convergence for mean estimate of source parameters using chemotaxis algorithm. As expected, using chemotaxis method results in better convergence comparing with stationary sensors, but its performance is not as good as DDM approach and obtained estimates for source parameters are not as confident as those of DDM method.

To highlight performance of proposed DDM approach, one can compare convergence behavior for parameter estimates obtained by DDM approach with convergence behavior of source parameter estimates using 25 stationary sensors, shown in Fig. 3.6. Comparison of Fig. 5.6 with Fig. 3.6 clearly shows that using DDM approach with just 3 mobile sensors results in much faster convergence for mean estimate of source parameters with respect to the case of using 25 *well located* stationary sensors. Hence, using Dynamic Data Monitoring method expedites convergence of estimation approach while using smaller number of data observation sensors. This can also be seen by comparing overall Root Mean Square Error (RMSE) between mean estimates of source parameters and their actual values while using different sensor networks and different methods. It is clear from Table 5.2 that proposed DDM algorithm outperforms all the other methods / sensor networks and DDM approach results in least amount of RMSE comparing to other sensor networks and chemotaxis al-

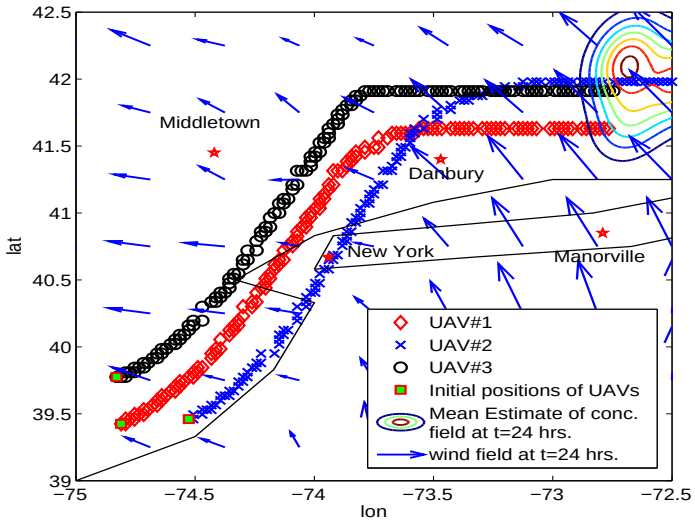
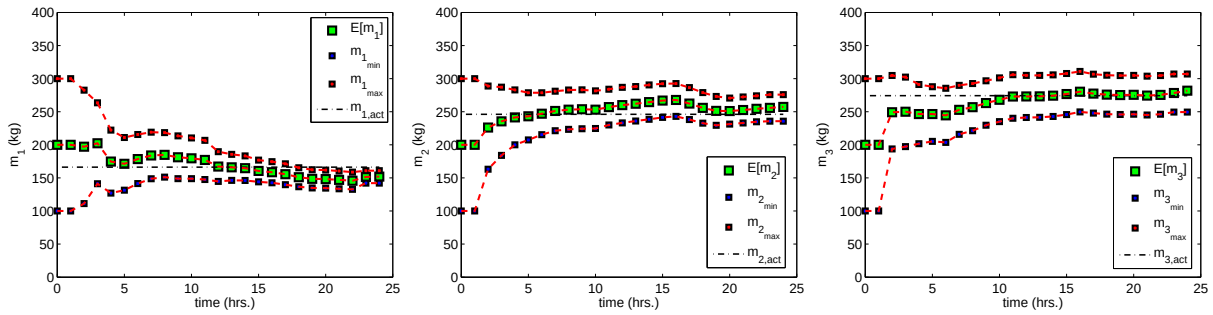


Figure 5.7: Test Case 1: Obtained way-points for mobile sensors at different time steps using Chemotaxis strategy. Contour map represents the mean estimate of concentration field at the final time ($t_f = 24\ hr$). The wind field at $t = 24\ hr$ (pressure level = 100 kpa) and geographical map of the region are shown in the background.



(a) (b) (c)

Figure 5.8: Test Case 1: Source parameter estimates during the time, obtained using Chemotaxis strategy

gorithm and using DDM approach with just 3 sensors results in lower value of RMSE in parameter estimates, comparing with other sensor networks. Hence, proposed DDM method provides more accurate estimates while using less number of data observation sensors.

Table 5.2: Test Case 1: Root Mean Square Error (RMSE) between mean estimate of source parameters and their actual values while using different sensor networks and different methods.

sensor network/method	RMSE		
	m_1	m_2	m_3
3 *stationary* sensors	46.27	21.18	38.67
25 *stationary* sensors	14.49	19.42	33.86
3 *mobile* sensors / chemotaxis	16.11	11.97	14.32
3 *mobile* sensors / DDM	1.93	5.74	12.19

5.3.4 Test Case 2

For the last example, we considered dispersion/advection of Chlorine over a *three dimensional* spatial domain. Simulation time is considered to be 10 $hrs.$ starting from $12:00$ of June 6^{th}, 2012. Three dimensional North American Regional Reanalysis wind data is used as applied windfield for simulation. Also, a $33 \times 46 \times 20$ grid points were used to perform simulations over the spatial domain. In our simulation, height variation is between 0 and 1000 m and a discretization of 50 m in vertical direction is performed to consider the effect of the windfield at different heights. Fig. 5.9 illustrates the source locations and the applied windfield (at $height = 100$ $m.$ and $t = 12:00$ of June 6^{th}) over the two dimensional spatial domain.

Uncertain source parameters are assumed to be source location $(lon_{src}, lat_{src}, z_{src})$ and mass release rate of chlorine (m_{src}). Source location is assumed to be *uniformly* distributed over the region $(lon, lat, z) \in [-103.8, -103.5] \times [39.45, 39.65] \times [300, 400]$ m. The mass release rate of chlorine is considered to be uniformly distributed between 6 kg/min and 10 kg/min and total release time is 1 hr. Actual source location is equal to $(lon_{src}, lat_{src}, z_{scr}) = (-103.73, 39.49, 386.4\ m)$. Projection of actual source location over $lon - lat$ plane is shown with a red circle in Fig. 5.9. The actual amount of mass release rate for chlorine is also assumed to be 8.11 kg/min. Table 5.3 illustrates uncertain parameters and their actual value for Test Case 2.

A set of 161 CUT8 quadrature points are used to quantify the uncertainty involved in

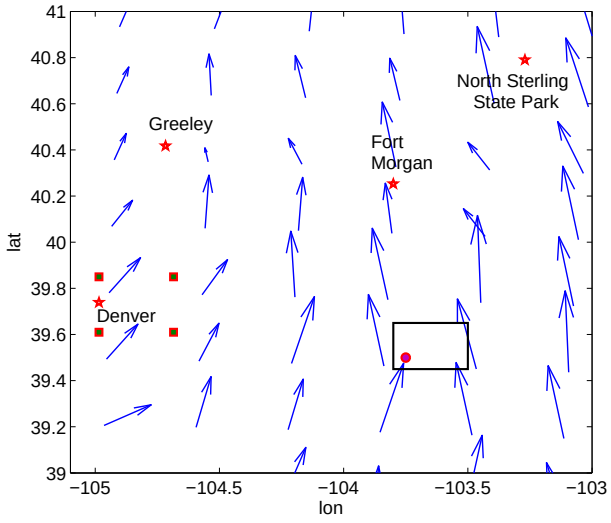


Figure 5.9: Schematic layout of Chlorine release over Denver region, source locations are shown with red circle, the applied wind-field (at initial time and pressure level = 100 kPa) is shown over the two dimensional domain with blue vector field. Black rectangle shows uncertain source location. Position of stationary sensors (initial positions of UAVs) are shown with squares. Geographical location of different cities can be seen on the background map

concentration of pollutant material and a 6^{th} order gPC expansion was used to reconstruct distribution of parameters after each measurement update. Concentration of chlorine is recorded every 10 $mins$. Performance of DDM approach is verified by comparing data assimilation results obtained by mobile and stationary sensors. Four sensors are used for data observation and a limited lookahead policy with $l = 2$ is used for finding location of each mobile sensor during the time. Note that $u_k^h \in \{5, 10\}$ $km.$, $u_k^z \in \{2, \cdots, 15\}$ in our simulations, and u_k^λ is given in Eq. (5.23). Also, $\alpha = 5$ and $W = diag([5, 5, 0.01])$. Fig. 5.10 illustrates positions of UAVs during the time over the $lon - lat$ domain. Fig. 5.10 shows that the UAVs follow the plume during the time and end up to the locations where the mutual information map obtains its maximum.

Convergence behavior for mean estimate of source parameters using stationary sensors, along with the minimum and maximum range of estimation are shown in Fig. 5.11. As one can see, using stationary sensors results in poor estimates of the parameters which is due to

Table 5.3: Test Case 2: Uncertain parameters and their actual value

Parameter	Distribution/Range	Actual value
lon_{src}	Uniform $\in [-103.8, -103.5]$	-103.73
lat_{src}	Uniform $\in [39.55, 39.77]$	39.58
z_{src}	Uniform $\in [300, 400]$ $m.$	386.4 m
m_{src}	Uniform $\in [4, 10]$ kg/min	8.11 kg/min

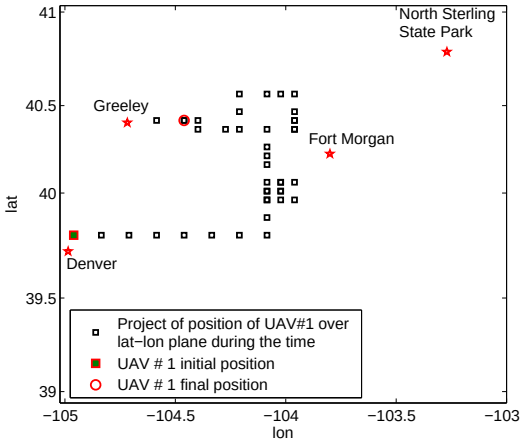
the ineffective placement of sensors.

Convergence behavior for mean estimate of source parameters using mobile sensors are illustrated in Fig. 5.12. It can be observed that the DDM approach results in significantly better estimates of the parameters and estimation process is very confident about the source parameters estimates.

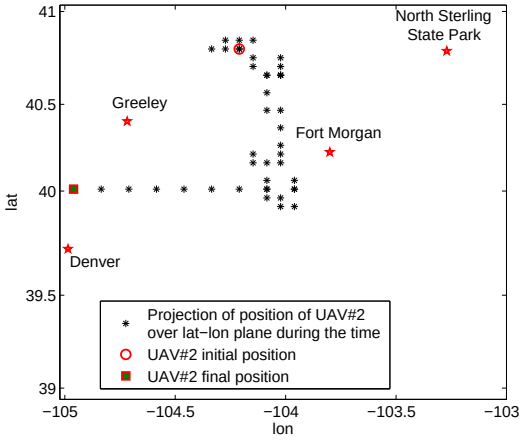
Table 5.4 shows the amount of information $I(\boldsymbol{\Theta}; \mathbf{z})$ collected by each sensor while using stationary and mobile sensors. From Table 5.4 it can be concluded that the Dynamic Data Monitoring method significantly increases the amount of mutual information collected by each sensor during the data assimilation process, and consequently improves the convergence behavior of estimation process.

Table 5.4: Test Case 2: Information collected by each sensor while with and without using Dynamic Data Monitoring approach.

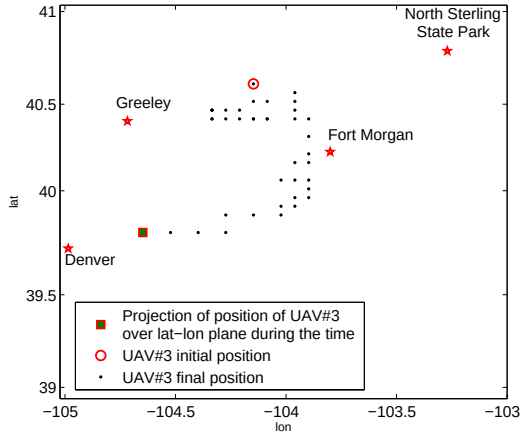
sensor number	collected information	
	Stationary	Dynamic
sensor 1	0	19.2996
sensor 2	0	20.0772
sensor 3	0	21.6315
sensor 4	3.6754	22.1826



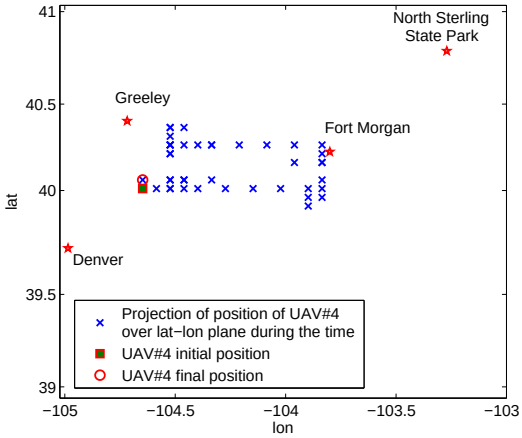
(a) $UAV\#1$



(b) $UAV\#2$



(c) $UAV\#3$

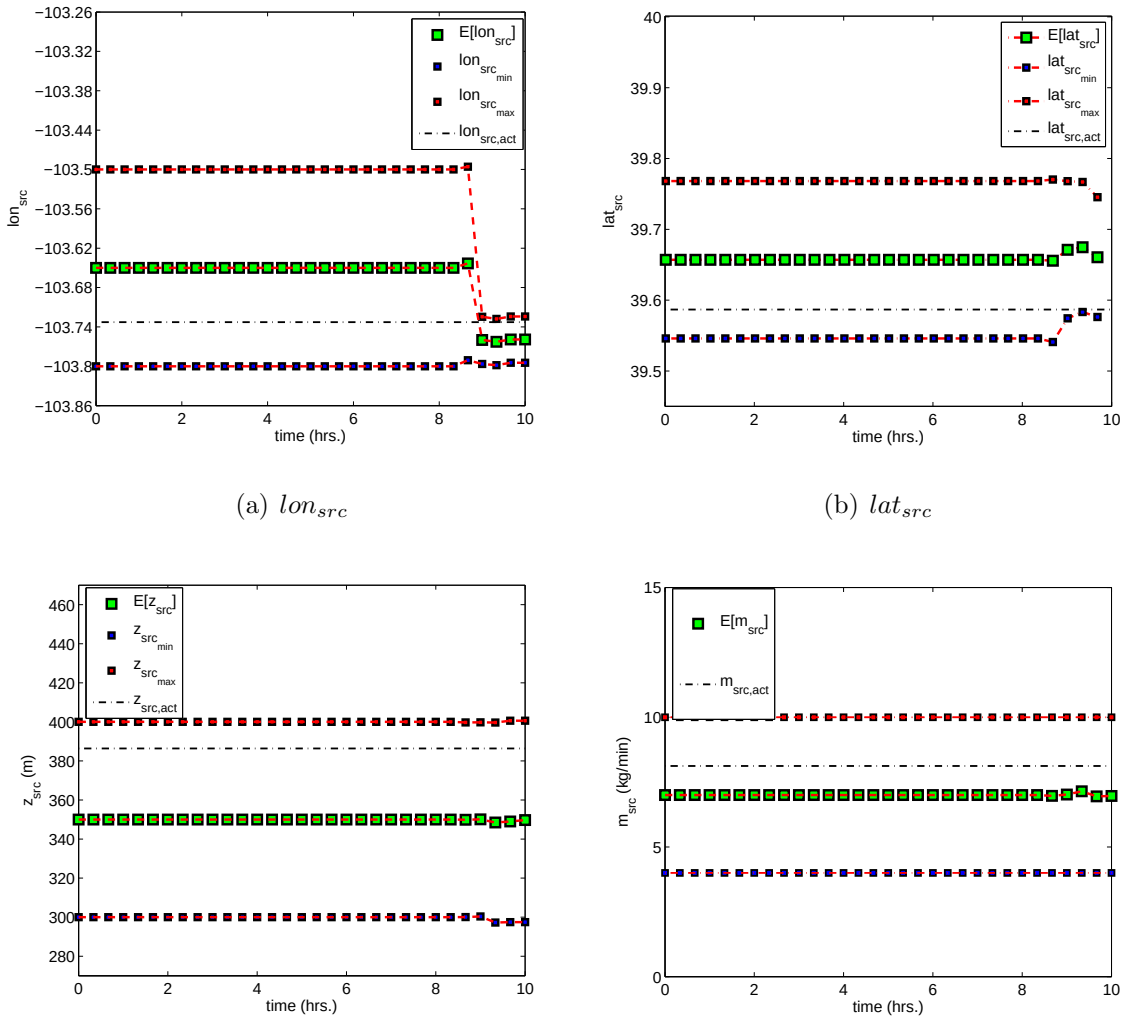


(d) $UAV\#4$

Figure 5.10: Test Case 2: Obtained way-points for mobile sensors at different time steps (a) $UAV\#1$, (b) $UAV\#2$, (c) $UAV\#3$, and (d) $UAV\#4$. Geographical map of the region are shown in the background.

5.4 Summary

In this chapter, we developed a general framework for optimal allocation of data observation sensors to ensure better performance of source characterization approach while source characterization of plume release incidents. The key idea of the presented method is to optimally

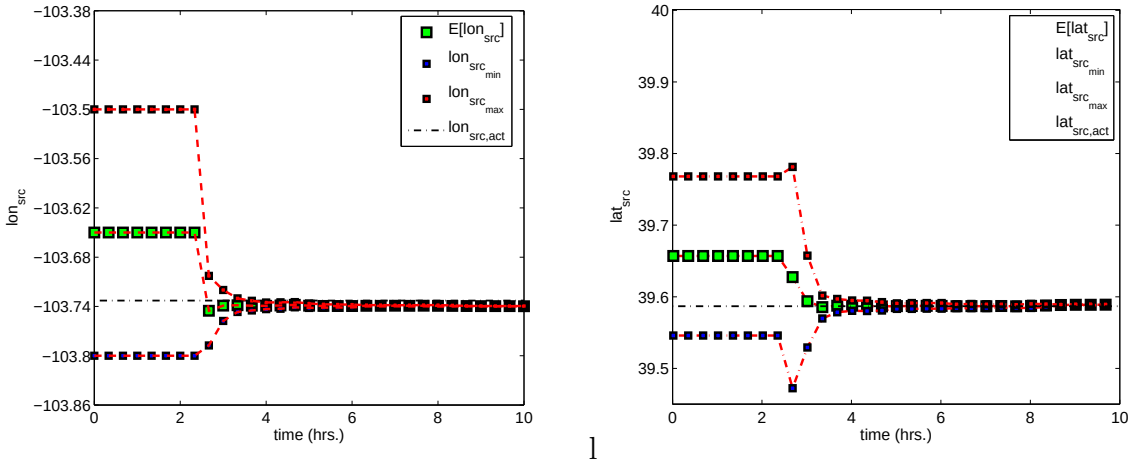


(a) lon_{src} (b) lat_{src}

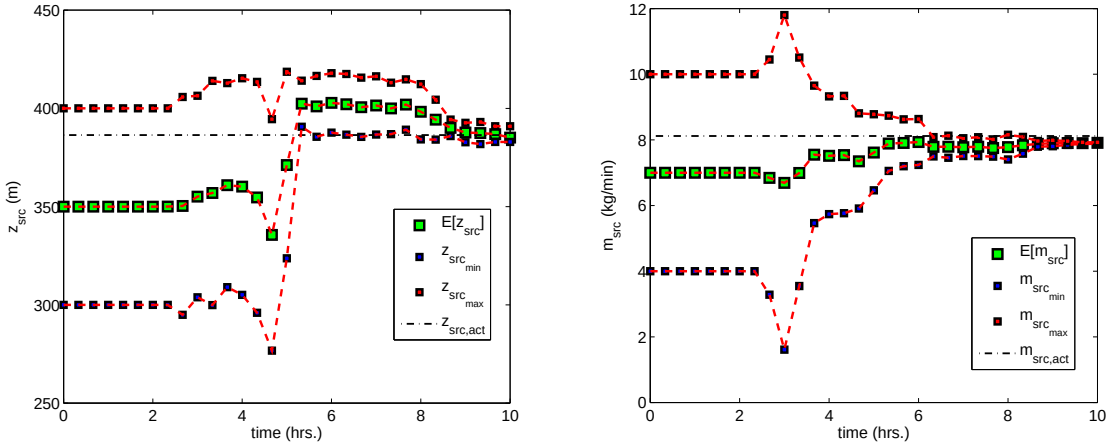
(c) z_{cvr} (d) m_{cvr}

Figure 5.11: Test Case 2: Source parameter estimates during the time, obtained using stationary sensors

locate data monitoring sensors such that the mutual information between model predictions and data measurements is maximized thereby giving a better reduction in uncertainty. The main advantage of this approach is that it *significantly increases* the accuracy of the estimation algorithm, while using *fewer* number of data observation sensors. The limited lookahead dynamic programming approach is well suited for the current problem as it is capable of incorporating or assimilating the updated information of wind data and measurements and thus provides robust UAV trajectories to collect better measurements. Mutual collisions and measurement redundancy are avoided by constantly maintaining a sufficient separation



(a) lon_{src} (b) lat_{src}



(c) z_{cvr} (d) m_{cvr}

Figure 5.12: Test Case 2: Source parameter estimates during the time, obtained using mobile sensors

between UAVs at all times. The numerical simulations validate the proposed methodology where the mobile sensors, when optimally planned to make measurements at specific locations and specific time instances, provide better estimates that significantly outperform the estimates from stationary sensors. One should note that the proposed dynamic data monitoring method can be applied to other types of data assimilation problems, especially in large scale systems with Hidden Markov Models or even black-box models with minor modifications.

The major computational effort in using Optimal Information Trajectory Design ap-

proach lies in propagation of new set of quadrature points after updating source parameter distributions. The computational effort required for propagating these quadrature points varies depending on the set of quadrature points and number of uncertain source parameters. Hence, using an *efficient* and *accurate* set of quadrature points can significantly reduce the computational complexities involved in whole process. As mentioned in Chapter 2, we have used a new quadrature scheme, named as CUT points, to alleviate the issue of computational complexities involved in propagation of new set of quadrature points. Besides using efficient CUT points, one can use parallel computing techniques to propagate the new set of quadrature points in considerably shorter time. This can be easily done due to the fact that each of the quadrature points can be propagated through the model, independent from the other.

The other source of computational complexity lies in the calculation of mutual information and the implementation of Dynamic Programming method for finding control policies to optimally position the UAVs. Note that the computational effort required to evaluate the mutual information is independent of the number of UAVs. However, depending on the number of UAVs the required time for computation of optimal policy increases linearly.

Finally, it should be noted that in absence of parallel computing, the computational time required for evaluation of mutual information and finding optimal policies of UAVs are negligible in comparison with the computational effort needed for propagation of quadrature points. Hence, DDM approach can be implemented in *almost* real time by using parallel computing techniques in propagation of new set of quadrature points.

Chapter 6

Data Assimilation in absence of Sensor Error Characteristics

6.1 Introduction

There has been numerous researches regarding probabilistic parameter estimation of dynamic systems [4, 5, 125, 126], given some measurement data. In all these approaches, there is a crucial need for knowledge of statistics of associated noise in observed data. For instance, information of mean and variance of the noise is required in all minimum variance based methods [125, 126]. Similarly, in Bayesian inference frameworks, distribution of noise signal is needed to construct the likelihood function [5].

A major challenge in estimation of dynamic systems is when information regarding the statistical properties of sensor data are not available or partially available. This frequently happens in data assimilation of atmospheric data by using satellite imagery. This is due to the fact that satellite imagery data can be polluted with noise, depending on weather conditions, clouds, humidity, etc. Unfortunately, there is no accurate procedure to quantify the error due to these factors on the output of satellite data. Hence, the use of the classical data assimilation methods in this situation is not straight forward. One way to proceed in this situation is to assume some statistics for the associated noise in measurement data and perform the estimation. However, obtained estimation results may not be as *accurate* as they should be in reality, depending on assumed values for noise statistics.

There exist numerous works [127–140] regarding data assimilation in absence of sensor

error characteristics. The essence of these works is about the estimation of the noise statistics along with model-data fusion of dynamical systems. These works can be divided into two different categories: i) covariance estimation methods, ii) methods that estimate distribution of the associated noise in measurement data.

The key idea of covariance estimation methods is to estimate the covariance of associated measurement noise along with other unknown parameters. Various covariance estimation techniques [127–137] exist in literature. Mehra [127] developed a parametric method for estimation of the covariance matrices with zero cross-correlation. In detail, this method reduces to estimation of diagonal elements of the covariance matrix of measurement noise. Bagchi [128] proposed a maximum likelihood approach for identification of continuous linear systems with unknown noise covariance, where covariance of the noise signal is estimated along with other unknown parameters. Rong Li *et al.* [129] developed a recursive multiple model approach for noise identification of dynamical systems in presence of non-stationary noise. Oussalah *et al.* [130] developed an adaptive Kalman filter to estimate associated covariance matrices of the noise signals for linear time invariant dynamical systems. Marginalized Particle Filter [141] technique is also used in [131–133] for estimation of associated covariance matrices of present disturbances for nonlinear dynamical systems. Odelson *et al.* [134] and Rajamani *et al.* [135] developed autocovariance least-square based methods to estimate covariance of associated noise signals in linear time invariant dynamical systems. Sarkka *et al.* [136] developed a variational Bayesian adaptive Kalman filter to estimate the covariance matrix of associated noise signal in measurement data for linear time invariant dynamical systems. However, developed methodology in [136] is restricted only to diagonal covariance matrices. Cai *et al.* [137] used a tapering technique for estimation of covariance matrix with non diagonal elements. In this method, a weight is assigned to each element of the matrix and the problem of covariance estimation reduces to estimation of these weights during the time.

One should note that most of the aforementioned works, other than [129, 131–133] , are restricted to linear time invariant dynamical systems. Also, most of the above techniques assume the associated noise signal to be Gaussian, which could be a restrictive assumption for some practical applications. In addition, all of them only concentrate on estimating covariance matrix of associated noise signal, while knowing only covariance of the noise signal may not be enough. One should also consider that all the current approaches for

covariance estimation of noise signal increase computational complexity of the whole model-data fusion process due to the introduction of new uncertain parameters that need to be estimated.

Besides covariance estimation technique, there exist few recent works [138,140] that focus on estimation of the probability density function of the existing noise signal in measurement data. Jaoua *et al.* [138] utilized Dirichlet process mixture to develop an online estimation framework for estimation of state and unknown measurement noise density in nonlinear dynamic state-space models. Ozkan *et al.* [140] developed an online Bayesian estimation for the measurement noise density of a given state space model using particle filter and Dirichlet process mixture. Note that in spite of most of the covariance estimation methods, all the density estimation methods are applicable to nonlinear time varying dynamical systems. In addition, they don't consider any restrictive assumption regarding distribution of the noise signal. However, similar to covariance estimation techniques, density estimation methods significantly increase the computational complexity involved in the whole model-data fusion process by introducing new uncertain parameters that need to be estimated.

The key contribution of this chapter is to develop a new model-data fusion method, with emphasis on large scale systems, that does not require any information regarding statistics of the existing noise signal in measurement data. Hence, avoiding the computational complexities involved in current methods for estimation of the noise signal statistics (e.g. covariance or density function). The key idea of this paper is to not make any assumption regarding the statistical properties of sensor data that are not available. Hence, in the absence of statistical information of the sensor data, we maintain the higher order prior statistics but update the posterior mean to be in compliance with sensor data.

Our Approach

The proposed approach is based on the premises that one should not make any assumption regarding the statistical properties of sensor data which are not available. In the absence of statistical information regarding the sensor data, we maintain the prior statistics but update the posterior mean to be in compliance with sensor data. Another aspect of the proposed approach is to evaluate different error metrics to compare the model output with sensor observation. These metrics are very important specifically for spatial data integration. The method presented in this chapter consists of four different components that are

combined together to perform the task of parameter estimation and state forecasting. These components are i) Uncertainty Quantification, ii) Error Evaluation, iii) Optimization, and iv) Reconstruction of Posterior Quadrature Points . Schematic view of our approach is illustrated in Fig. 6.1. As shown in Fig. 6.1, estimation process starts with a given prior uncertainty in parameters of dispersion/advection incident. Similar to previous chapters, the first step to perform the estimation is to quantify the effect of uncertain source parameters on model output. As described in Chapter 2, the method of quadrature points is used to perform the task of uncertainty quantification and then a weighted average of propagated quadrature points is used to determine *prior statistics* of model output over the time. The method of quadrature points is described in more detail in Chapter 2.

After precise uncertainty quantification of model outputs, the next step of data assimilation process involves optimizing prior weights of each quadrature points. This involves computing the error ensembles by comparing model forecast corresponding to each quadrature point with observation data. The main feature of the proposed approach is to pose the computation of posterior weights as an iterative convex optimization problem with a guaranteed solution.In Section 6.4, we will explain the procedure for error evaluation in more detail.

Obtained error ensembles and prior error statistics are then used in an *convex quadratic optimization* problem to find *posterior* weights of the quadrature points. The intuition behind this optimization is to minimize the expected value of the error between model forecast ensembles and measurement data while preserving higher order statistics of error distribution. This results in posterior values of quadrature weights which are then used to find posterior estimate of source parameters and accurate forecasting of model output. In Section 6.5, detailed information regarding construction of this optimization problem is presented.

To complete the estimation procedure, a set of new quadrature points are then reconstructed based on obtained posterior statistics of source parameters. These new set of quadrature points are then propagated through the dynamical model and whole procedure is repeated recursively for the next time steps. We will explain detailed procedure for reconstruction of posterior quadrature points of source parameters in Section 6.6.

The structure of this chapter is as follows. First, the problem of parameter estimation dynamical systems in absence of sensor error characteristics is formulated in Section 6.2.

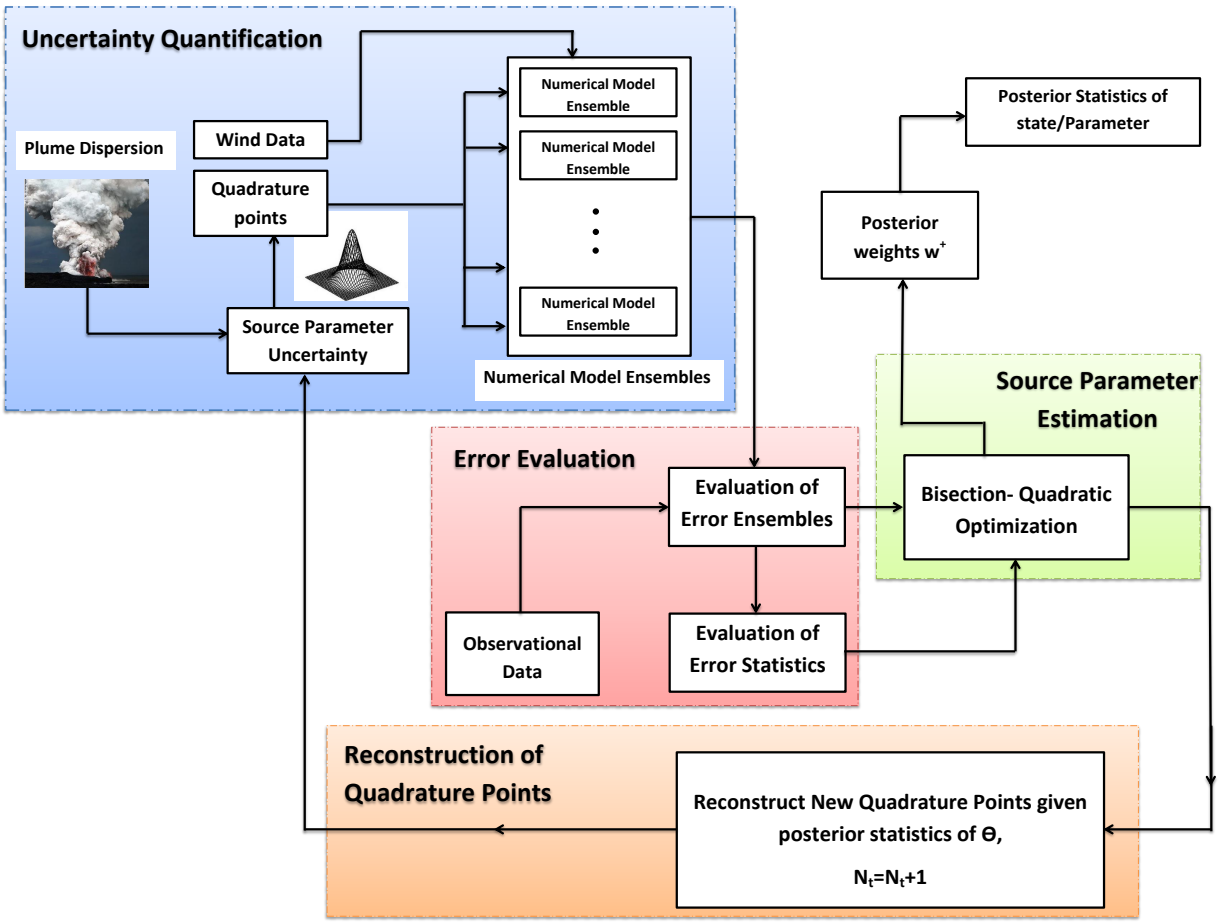


Figure 6.1: Schematic view of estimation process. N_t represents the time step. w^- and w^+ denote prior and posterior weights

We briefly discuss the method of quadrature points in Section 6.3. In Section 6.4, we will discuss about the theoretical basis of evaluation of error ensembles and error statistics. Then in Section 6.5, we construct a convex-quadratic optimization problem whose solution results in posterior value of quadrature weights. We will propose a scheme for reconstruction of quadrature points of source parameters in Section 6.6. We then demonstrate performance of proposed approach by using a simple numerical example in Section 6.7. In the end, summary of the chapter is presented in Section 6.8.

6.2 Problem Statement

Let us consider a dynamic system, given by Eq. (2.1):

$$\dot{\mathbf{x}} = f(t, \mathbf{x}, \Theta)$$

where, t is time, and $\mathbf{x} \in \mathbb{R}^{n\times 1}$ and $\Theta \in \mathbb{R}^{m\times 1}$ represent states and parameters of the system, respectively. Parameter Θ is assumed to be uncertain, which is defined by some prior distribution $p(\Theta)$, i.e. $\Theta \sim p(\Theta)$.

Also, measurement data is given using the following model:

$$\mathbf{y} = \mathbf{h}(\mathbf{x}, \Theta, \nu) \tag{6.1}$$

where, $\mathbf{y} \in \mathbb{R}^{b\times 1}$ denotes observation data, $\mathbf{h}(.) \in \mathbb{R}^{b\times 1}$ is the observation operator, and ν is the associated noise. As compared to Chapter 3, we know relax the assumption regarding the statistical properties of sensor noise, i.e., no information is available regarding the statistical properties of ν.

Our objective is to have a reliable estimate of parameter $\mathbf{\Theta}$, by assimilating measurement data with forward model.

6.3 Uncertainty Quantification

As we previously explained in Chapter 2, the first step to perform estimation process is to quantify the uncertainty associated with forward model output during the time propagation. Similar to previous chapters in this dissertation, the method of quadrature points is used to perform this task, i.e. N quadrature points for parameter Θ are generated based on prior density function, $p(\Theta)$ and Eq. (2.1) is simulated for each of these realizations. Statistics of the model output can be evaluated according to Eq. (2.2), i.e,

$$\mathcal{E}[\mathbf{x}^k] = \int_{\boldsymbol{\xi}} \mathbf{x}^k(\mathbf{\Theta}, t) p(\boldsymbol{\xi}) d\boldsymbol{\xi} \simeq \sum_{q}^{k} w_q \mathbf{x}^N(\mathbf{\Theta}(\boldsymbol{\xi}^q), t), \quad k = 1, 2, \cdots$$

For detail description about the method of quadrature points, please see Chapter 2.

6.4 Error Evaluation

To compare model outputs with measurement data, one needs to provide a measure to compute the difference (error) between model outputs and the measurement data. To describe this in more detail, let us assume that the error between model predictions and given measurement data to be given as:

$$e = d(\mathcal{M}(\mathbf{\Theta}), \mathbf{y}) \tag{6.2}$$

where, error $e \in \mathbb{R}^+$, and $d(\mathcal{M}(\mathbf{\Theta}), \mathbf{y}) \geq \mathbf{0}$ is a metric operator which is used to calculate the difference between model forecast $\mathcal{M}(\mathbf{\Theta}) = \mathbf{h}(\mathbf{x}, \Theta, 0)$ and observation data $\mathbf{y}$. For instance, $d(\mathcal{M}(\mathbf{\Theta}), \mathbf{y})$ can be simply defined as Frobenius norm between $\mathcal{M}(\mathbf{\Theta})$ and $\mathbf{y}$.

We emphasize here that applied operator $d(.,.)$ for calculation of the difference between model forecasts $\mathcal{M}(\mathbf{\Theta})$ and measurement data $\mathbf{y}$ should be a proper metric to measure the difference between model forecasts and measurement data in distinguishable manner. In general, function $d : \mathbb{R}^b \times \mathbb{R}^b \longrightarrow \mathbb{R}^+$ is a metric. If given $A, B, C \in \mathbb{R}^b$, then we have:

- $d(A, B) \geq 0$
- $d(A, B) = 0$, if and only if $A = B$
- $d(A, B) = d(B, A)$
- $d(A, C) \leq d(A, B) + d(B, C)$

A wide variety of error metrics can be used for measuring the difference between model forecast ensembles and observation data. However, for large scale systems like atmospheric dispersion models, which are emphases of this dissertation, we are looking for error metrics which integrate all the information regarding the difference between model output and measurement data into a scalar number. In this way, we can avoid the issue of dimensionality, and at the same time, applied metric will be able to distinguish between each and every model output ensembles. Different metrics can be used for this purpose [142–144]. In the following, we describe few of these metrics.

Hausdorff Metric

Hausdorff distance is defined as the *maximum distance of a set to the nearest point in the other set*. Mathematically, Hausdorff distance from set A to set B is defined as:

$$h(A,B) = \max_{a\in A}\left\{\min_{b\in B}\{dist(a,b)\}\right\} \tag{6.3}$$

where, a and b are points of sets A and B, respectively, and $dist(a,b)$ is any metric distance between these points.

Similarly, Hausdorff distance from set B to set A, denoted by $h(B,A)$, is defined as:

$$h(B,A) = \max_{b\in B}\left\{\min_{a\in A}\{dist(a,b)\}\right\} \tag{6.4}$$

Note that $h(A,B) \neq h(B,A)$ in general. Hence, Hausdorff distance can not be used as a metric. To overcome this drawback, a more general definition of Hausdorff distance is given as

$$H(A,B) = \max\{h(A,B), h(B,A)\} \tag{6.5}$$

Note that Eq. (6.5) represents a proper metric.

To calculate $H(A,B)$ between each model output ensemble and observational data, one needs to perform the following procedure.

- Consider a concentration field which is a discrete function defined over a spatial domain D of size $n_1 \times n_2$ which takes the discrete values $c \in \{0,1,\cdots,G\}$. Note that even though model output is continuous field over spatial domain D, this discretization can be easily performed.

- Apply an appropriate norm to calculate $dist(a,b)$ in Eq. (6.3) and Eq. (6.4). Even though, different norms can be used to calculate $dist(a,b)$ in these equations, we have used the following equation for evaluation of $dist(a_{i,j}, b_{l,m})$ in our simulations [142]:

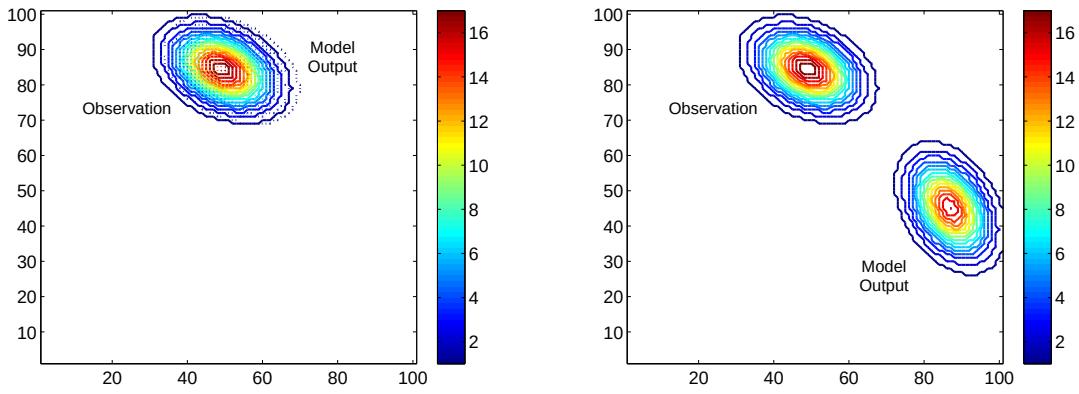
$$dist(a_{i,j}, b_{l,m}) = \max\{\frac{|i-l|}{n_1}, \frac{|j-m|}{n_2}, \frac{|a_{i,j}-b_{l,m}|}{G}\} \tag{6.6}$$

  where, $a_{i,j} \in A$ is value of concentration field A at spatial grid point (i,j). Similarly, $b_{l,m} \in B$ is value of concentration field B at spatial grid point (l,m). Also, n_1 and n_2 are the number of grid points in spatial domain D. Note that $a_{i,j}, b_{l,m} \in \{0,1,\cdots,G\}$.

  The major benefit of using Eq. (6.6) is that given Eq. (6.6) for $d(a,b)$, Hausdorff metric will be always normalized, i.e. $H(A,B) \in [0,1]$.

- Use Eq. (6.3) to Eq. (6.5) to find $H(A, B)$.

Fig. 6.2 shows Hausdorff distance for two samples of model outputs over observational data. As Fig. 6.2(a) represents, model output and observational data are very close to each other, whenever $H(A, B)$ is small. On the other hand, by increasing $H(A, B)$, the difference between model output and observational data increases. This can be seen in Fig. 6.2(b). Note that $n_1 = n_2 = 101$ and $G = 20$ in our simulations.



(a) $H = 0.05$ (b) $H = 0.49$

Figure 6.2: Comparison of different samples of the model outputs and observation data at a given time (a) two images are very close and correspondingly H is small (H=0.05) (b) two images are different from each other and correspondingly H is large (H=0.49)

It should be mentioned that Hausdorff metric can be used to quantify the error between model output ensembles and observation data. Hence, Eq. (6.5) can be used instead of Eq. (6.2), i.e.

$$e(\Theta_q) = H(\mathcal{M}(\Theta_q), \mathbf{y}), \quad q = 1, 2, \cdots, N$$

where, $\mathcal{M}(\mathbf{\Theta}_q) = \mathbf{h}(\mathbf{x}, \Theta_q, \mathbf{0})$ is the model forecast concentration field spread over spatial domain D, which is generated by atmospheric model based on q^{th} realization of uncertain parameter Θ. As well, $\mathbf{y}$ is the observed concentration field, provided by Eq. (6.1). Notice that observed concentration field $\mathbf{y}$ can be noisy in general.

Euclidean Distance

Given two image A and B, Euclidean metric is defined as:

$$d_E(A,B) = \sqrt{\sum_{i,j}(a_{i,j} - b_{i,j})^2} \tag{7.15}$$

where, $a_{i,j}$ and $b_{i,j}$ are the intensity of images A and B, respectively, at grid point (i,j). For more information regarding Euclidean distance, please see [144].

6.4.1 Error Statistics

One can find statistics of the error, after computing the difference between model forecast ensembles and measurement data. Note that there will be N realizations for error, given N quadrature points for model parameter Θ. Statistics of error can be defined using these quadrature values. For instance, expected value of each element of error $e \in \mathbb{R}^+$ can be written as:

$$m_1 = \sum_{q=1}^{N} w_q^- e_k(\Theta_q) \tag{6.7}$$

where, w_q^- is the corresponding weight for Θ_q which is q^{th} quadrature value of parameter Θ, and m_1 denotes expected value of error $e \in \mathbb{R}^+$, i.e. $m_1 = \mathcal{E}[e]$.

Similarly, higher order central moments for error e can be defined as:

$$m_k = \sum_{q=1}^{N} w_q^- \left(e_k(\Theta_q) - m_1\right)^k, \quad k = 2,3,\cdots,N_m \tag{6.8}$$

where, $N_m \in \mathbb{N}$ is the highest order of central moments calculated.

6.5 Source Parameter Estimation

The key idea of source parameter estimation is to find posterior values of quadrature weights w_q,s to minimize expected value of error, denoted by $\mathcal{E}[e]$, while preserving its higher order central moments. This can be mathematically defined as the following:

$$\min_{\mathbf{w}} \mathcal{E}[e] = \min_{w} \sum_{q=1}^{N} w_q e(\Theta_q) \tag{6.9}$$

subject to

$$\sum_{q=1}^{N} w_q \left(e(\Theta_q) - \underbrace{\sum_{q=1}^{N} w_q e(\Theta_q)}_{m_1} \right)^k = m_k, \quad 2 \leq k \leq N_m \tag{6.10}$$

$$\sum_{q=1}^{N} w_q = 1, \quad 0 \leq w_q \leq 1 \tag{6.11}$$

where, $\mathbf{w} = [w_1, w_2, \cdots, w_N]^T$. Note that Eq. (6.9) is an ill-posed optimization problem and minimizing only $\mathcal{E}[e]$, given Eq. (6.10) and Eq. (6.11) has infinite solution. To surpass this issue, one can add a penalty term to Eq. (6.9) to minimize the difference between prior and posterior weights, and at the same time to minimize $\mathcal{E}[e]$. In this way, modified optimization problem can be written as:

$$\min_{\mathbf{w}} \lambda_1 \mathcal{E}[e] + \lambda_2 |\mathbf{w} - \mathbf{w}^-|_2 = \min_{w} \lambda_1 \underbrace{\sum_{q=1}^{N} w_q e(\Theta_q)}_{m_1} + \lambda_2 \sum_{q=1}^{N} (w_q - w_q^-)^2 \tag{6.12}$$

subject to

$$\sum_{q=1}^{N} w_q \left(\prod_{j=1}^{b} \left(e(\Theta_q) - \underbrace{\sum_{q=1}^{N} w_q e(\Theta_q)}_{m_1} \right)^{n_k} \right) = m_k, \quad 2 \leq k \leq N_m \tag{6.13}$$

$$\sum_{q=1}^{N} w_q = 1, \quad 0 \leq w_q \leq 1 \tag{6.14}$$

where, m_1 is the expected value vector of error e obtained by Eq. (6.7), $\lambda_1, \lambda_2 > 0$ are given constants, and $\mathbf{w}^- = [w_1^-, w_2^-, \cdots, w_N^-]^T$ is *prior* values of quadrature weights. Coefficients λ_1 and λ_2 represent the relative weights to the two terms in the cost function. Intuitively, when $\lambda_2 \ll \lambda_1$ Eq. (6.12) focuses on minimizing $\mathcal{E}[e]$, rather than the difference of $\mathbf{w}$ and $\mathbf{w}^-$. On the other hand, when $\lambda_1 \ll \lambda_2$, Eq. (6.12) returns the same values of $\mathbf{w}^-$ for $\mathbf{w}$.

It should be mentioned that proposed approach minimizes the expected value of the difference between model forecasts and measurement data while preserving higher order central moments of error. Schematic view of this process is illustrated in Fig. 6.3. As illustrated in this figure, proposed optimization problem minimizes expected value of the error while

preserving shape of error distribution, i.e. higher order statistics of error distribution.The intuition behind preserving higher order central moments of error is to account for possible inaccuracies of the measurement data.

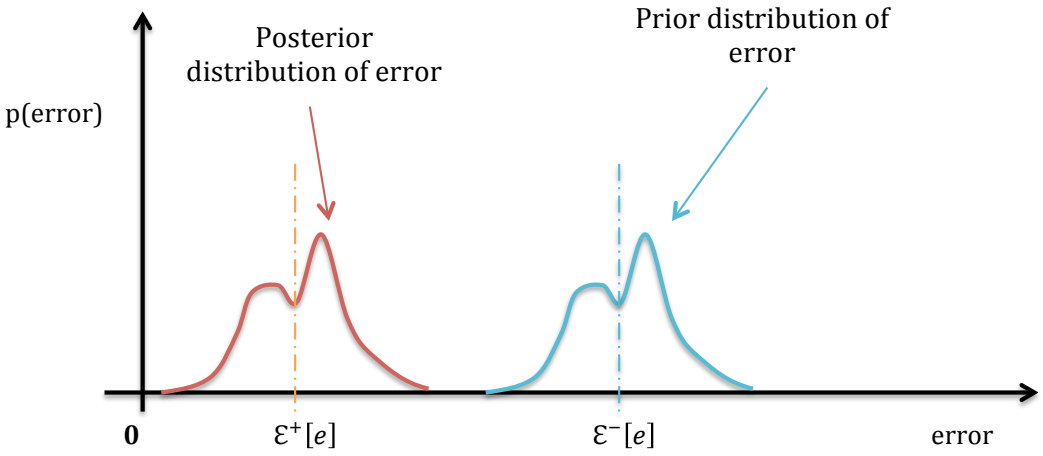


Figure 6.3: Schematic view of convex optimization method used for finding posterior quadrature weights. As illustrated, proposed optimization problem minimizes expected value of the error while preserving shape of error distribution, i.e. higher order statistics of error distribution. Prior and posterior values of expected value of error ensembles are also shown in the figure.

Convex Quadratic Optimization

Eq. (6.12), along with Eq. (6.13) and Eq. (6.14) is a nonlinear optimization problem, due to presence of equality constraints for higher order central moments. In the following, we utilize an iterative approach to transfer this nonlinear optimization problem into a convex quadratic optimization problem, which is much easier to solve. This has been achieved by making use of the bisection method.

Let us assume that the value of m_1 is known, then the nonlinear optimization problem is converted to the following quadratic optimization problem.

$$\min_{\mathbf{w}} |\mathbf{w} - \mathbf{w}^-|_2 = \sum_{q=1}^{N} (w_q - w_q^-)^2 \tag{6.15}$$

subject to

$$\sum_{q=1}^{N} w_q \left(e(\Theta_q) - m_1\right)^k = \quad m_k, \quad 2 \leq k \leq N_m \tag{6.16}$$

$$\sum_{q=1}^{N} w_q e(\Theta_q) = m_1 \tag{6.17}$$

$$\sum_{q=1}^{N} w_q = 1 \tag{6.18}$$

$$0 \leq w_q \leq 1 \tag{6.19}$$

One can combine Eq. (6.17) and Eq. (6.18) to further simplify above optimization problem. This results in the following optimization problem:

$$\min_{\mathbf{w}} |\mathbf{w} - \mathbf{w}^-|_2 = \sum_{q=1}^{N} (w_q - w_q^-)^2 \tag{6.20}$$

subject to

$$\sum_{q=1}^{N} w_q \left(e(\Theta_q) - m_1\right)^k = \quad m_k, \quad 2 \leq k \leq N_m \tag{6.21}$$

$$\sum_{q=2}^{N} w_q \left(e(\Theta_q) - e(\Theta_1)\right) = \qquad m_1 - e(\Theta_1) \tag{6.22}$$

$$0 \leq w_q \leq 1 \tag{6.23}$$

As one can see, Eq. (6.20) along Eq. (6.21) to Eq. (6.23) is a quadratic optimization problem which can be solved much easier than the original optimization problem of Eq. (6.12) to Eq. (6.14). We can iteratively solve resulted quadratic optimization problem for different value of m_1 to minimize the original optimization problem. This can be achieved with the help of the bisection method. In detail, we first assume lower bound and upper bound of m_1 to be zero and prior mean, $\sum_{q=1}^{N} w_q^- e(\Theta_q)$, respectively. Then, we assign m_1 to be average of considered lower and upper bounds and solve the resulted quadratic convex optimization problem in Eq. (6.20)-Eq. (6.23). If the optimization was not successful, then we substitute m_1 with the new value which is given from bisection method. This process is repeated iteratively, until convergence of bisection method is ensured. Applied pseudo-code for this procedure is described in the following:

Iterative algorithm for minimizing Eq. (6.15) subject to Eq. (6.16) to Eq. (6.19)

assign upper bound (ub) of m_1 to be expected value of error, i.e. $ub = \sum_{q=1}^{N} w_q^- e(\Theta_q)$
assign lower bound (lb) of m_1 to be zero.
while (ub − lb) ≥ threshold
    $m_1 = 0.5 * (ub + lb)$;
    Perform quadratic optimization to solve Eq. (6.20) to Eq. (6.23)
    if optimization is feasible
        ub = m_1;
    else
        lb = m_1;
    end
end

Routine quadratic optimization methods like quadprog in Matlab or CVX optimization toolbox [145] can be used to perform quadratic optimization.

One may say that proposed iterative bisection algorithm doesn't provide the smallest value of m_1. In the following, we will prove a theorem which declares that the value of m_1 given by above bisection algorithm is the smallest possible value of m_1 and optimization problem given by Eq. (6.20) to Eq. (6.23) is infeasible for any smaller value of m_1. But before that, we will prove the following two lemmas which help us in proof of mentioned theorem.

Lemma 6.5.1. *Assume that for a given value of m_1, denoted by m_1^0 (e.g. the value of m_1 given by prior mean of error) and for a given value of $2 \leq k = k^* \leq N_m$, there exist $0 \leq w_q \leq 1$ such that Eq. (6.21) holds, i.e.*

$$\exists \quad 0 \leq w_q \leq 1 \quad \textit{such that} \quad \sum_{q=1}^{N} w_q (e(\boldsymbol{\Theta}_q) - m_1^0)^{k^*} = m_{k^*} \tag{6.24}$$

If for any given value of $m_1 = m_1^c < m_1^0$

$$\sum_{q=1}^{N} w_q (e(\boldsymbol{\Theta}_q) - m_1^c)^{k^*} \neq m_{k^*}, \quad \forall \quad 0 \leq w_q \leq 1 \tag{6.25}$$

then, for all $m_1 \leq m_1^c$, we have

$$\sum_{q=1}^{N} w_q (e(\boldsymbol{\Theta}_q) - m_1)^{k^*} \neq m_{k^*}, \quad \forall \quad 0 \leq w_q \leq 1 \tag{6.26}$$

Proof. The proof is composed of two steps. First we show that if for a given k^* and $m_1^c < m_1^0$

$$\sum_{q=1}^{N} w_q (e(\boldsymbol{\Theta}_q) - m_1^c)^{k^*} \neq m_{k^*}, \quad \forall \quad 0 \leq w_q \leq 1$$

then

$$\sum_{q=1}^{N} w_q (e(\boldsymbol{\Theta}_q) - m_1^c)^{k^*} > m_{k^*}, \quad \forall \quad 0 \leq w_q \leq 1$$

This can be easily proved by contradiction. Assume that for all $0 \leq w_q \leq 1$

$$\sum_{q=1}^{N} w_q (e(\boldsymbol{\Theta}_q) - m_1^c)^{k^*} < m_{k^*}$$

On the other hand, we know that $m_1^c < m_1^0$. Hence, we will have:

$$\left.\begin{array}{l} \displaystyle\sum_{q=1}^{N} w_q (e(\boldsymbol{\Theta}_q) - m_1^c)^{k^*} < m_{k^*}, \quad \forall \quad 0 \leq w_q \leq 1 \\ (e(\boldsymbol{\Theta}_q) - m_1^0) < (e(\boldsymbol{\Theta}_q) - m_1^c), \quad \forall \quad 1 \leq q \leq N \end{array}\right\} \Rightarrow \tag{6.27}$$

$$\sum_{q=1}^{N} w_q (e(\boldsymbol{\Theta}_q) - m_1^0)^{k^*} < m_{k^*}, \quad \forall \quad 0 \leq w_q \leq 1$$

which is in contradiction with Eq. (6.24). Thus,

$$\forall \quad m_1^c < m_1^0 \quad \text{and } 0 \leq w_q \leq 1, \quad \sum_{q=1}^{N} w_q (e(\boldsymbol{\Theta}_q) - m_1^c)^{k^*} > m_{k^*} \tag{6.28}$$

In the next step, we just need to prove that

$$\sum_{q=1}^{N} w_q (e(\boldsymbol{\Theta}_q) - m_1)^{k^*} \neq m_{k^*}, \quad \forall \quad 0 \leq w_q \leq 1, \quad \forall \quad m_1 < m_1^c \tag{6.29}$$

This can be easily proved by making use of the fact that $m_1 < m_1^c$ and Eq. (6.28):

$$\forall \quad m_1^c < m_1^0 \quad \text{and } 0 \leq w_q \leq 1, \quad \sum_{q=1}^{N} w_q (e(\boldsymbol{\Theta}_q) - m_1^c)^{k^*} > m_{k^*}, \quad \xRightarrow{m_1 < m_1^c}$$

$$\forall \quad m_1 < m_1^c \quad \text{and } 0 \leq w_q \leq 1, \quad \sum_{q=1}^{N} w_q (e(\boldsymbol{\Theta}_q) - m_1)^{k^*} > m_{k^*}$$

Hence,

$$\forall \quad m_1 < m_1^c \quad \text{and } 0 \leq w_q \leq 1, \quad \sum_{q=1}^{N} w_q (e(\boldsymbol{\Theta}_q) - m_1^c)^{k^*} \neq m_{k^*}$$

This completes the proof for lemma 1.

□

Lemma 6.5.2. *Assume that for a given value of m_1, denoted by m_1^0 (e.g. the value of m_1 given by prior mean of error), there exist $0 \leq w_q \leq 1$ such that Eq. (6.22) holds, i.e.*

$$\exists \quad 0 \leq w_q \leq 1 \quad \textit{such that} \quad \sum_{q=2}^{N} w_q \left(e(\Theta_q) - e(\Theta_1)\right) = m_1^0 - e(\Theta_1) \tag{6.30}$$

If for a given value of $m_1 = m_1^c < m_1^0$

$$\sum_{q=2}^{N} w_q (e(\boldsymbol{\Theta}_q) - e(\boldsymbol{\Theta}_1)) \neq m_1^c - e(\boldsymbol{\Theta}_1), \quad \forall \quad 0 \leq w_q \leq 1 \tag{6.31}$$

then, $\forall \quad m_1 \leq m_1^c$, we have

$$\sum_{q=2}^{N} w_q (e(\boldsymbol{\Theta}_q) - e(\boldsymbol{\Theta}_1)) \neq m_1 - e(\boldsymbol{\Theta}_1), \quad \forall \quad 0 \leq w_q \leq 1 \tag{6.32}$$

Proof. The proof is very similar to the proof for lemma 1. First we show that if for a $m_1^c < m_1^0$

$$\sum_{q=2}^{N} w_q \left(e(\Theta_q) - e(\Theta_1)\right) \neq m_1^c - e(\Theta_1), \quad \forall 0 \leq w_q \leq 1$$

then

$$\sum_{q=2}^{N} w_q \left(e(\Theta_q) - e(\Theta_1)\right) > m_1^c - e(\Theta_1), \quad \forall 0 \leq w_q \leq 1$$

This can be easily proved by contradiction. Assume that for all $0 \leq w_q \leq 1$

$$\sum_{q=2}^{N} w_q \left(e(\Theta_q) - e(\Theta_1)\right) < m_1^c - e(\Theta_1)$$

On the other hand, we know that $m_1^c < m_1^0$. Hence, we will have:

$$\forall \quad 0 \leq w_q \leq 1, \quad \sum_{q=2}^{N} w_q \left(e(\Theta_q) - e(\Theta_1)\right) < m_1^0 - e(\Theta_1)$$

which is in contradiction with Eq. (6.30). Thus,

$$\forall \quad m_1^c < m_1^0 \quad \text{and } 0 \leq w_q \leq 1, \quad \sum_{q=2}^{N} w_q \left(e(\Theta_q) - e(\Theta_1)\right) > m_1^c - e(\Theta_1) \tag{6.33}$$

In the next step, we just need to prove that

$$\sum_{q=2}^{N} w_q \left(e(\Theta_q) - e(\Theta_1)\right) \neq m_1 - e(\Theta_1) \quad \forall \quad 0 \leq w_q \leq 1, \quad \forall \quad m_1 < m_1^c \tag{6.34}$$

This can be easily proved by making use of the fact that $m_1 < m_1^c$ and Eq. (6.33):

$$\forall \quad m_1^c < m_1^0 \quad \text{and } 0 \leq w_q \leq 1, \quad \sum_{q=2}^{N} w_q \left(e(\Theta_q) - e(\Theta_1)\right) > m_1^c - e(\Theta_1), \quad \xRightarrow{m_1 < m_1^c}$$

$$\forall \quad m_1 < m_1^c \quad \text{and } 0 \leq w_q \leq 1, \quad \sum_{q=2}^{N} w_q \left(e(\Theta_q) - e(\Theta_1)\right) > m_1 - e(\Theta_1)$$

Hence,

$$\forall \quad m_1 < m_1^c \quad \text{and } 0 \leq w_q \leq 1, \quad \sum_{q=2}^{N} w_q \left(e(\Theta_q) - e(\Theta_1)\right) \neq m_1 - e(\Theta_1)$$

This completes the proof for lemma 2.

□

Theorem 6.5.3. *Assume that for a given value of $m_1 = m_1^c$, there exists no solution for optimization problem defined by Eq. (6.20) to Eq. (6.23). Then the optimization problem defined by Eq. (6.20) to Eq. (6.23) doesn't have any solution for any values of $m_1 < m_1^c$.*

Proof. We use lemma 1 and lemma 2 to prove this theorem. Assume there exists no solution for the optimization problem, then either one of Eq. (6.21) and Eq. (6.22) is violated.

If Eq. (6.21) is violated, then as we proved in lemma 1, Eq. (6.21) is violated for all values of $m_1 < m_1^c$.

On the other hand, if Eq. (6.22) is violated, then as we proved in lemma 2, Eq. (6.22) is violated for all values of $m_1 < m_1^c$.

Hence, if the optimization problem defined by Eq. (6.20) to Eq. (6.23) returns an infeasible solution for a given value of $m_1 = m_1^c$, then it also returns an infeasible solution for any value of $m_1 < m_1^c$. □

6.5.1 Posterior Statistics of Parameter Θ and State x

After finding posterior values of quadrature weights, one can use these weights to evaluate posterior statistics of Θ.

$$m_k^+(\Theta) = \sum_{i=1}^{N} w_i^+ \left(\prod_{j=1}^{m} \Theta_{i_j}^{n_j} \right), \quad \sum_{j=1}^{m} n_j = k, \quad k = 1, 2, \cdots \tag{6.35}$$

where, $m_k^+(\Theta)$ denotes k^{th} order posterior moment of parameter Θ and Θ_{i_j} is j^{th} element of i^{th} quadrature of vector Θ.

One should note that similar procedure can be used for state estimation, i.e. obtained posterior weights can be used to estimate posterior statistics of state after each update.

$$m_k^+(\mathbf{x}) = \sum_{i=1}^{N} w_i^+ \left(\prod_{j=1}^{n} x_{i_j}^{n_j} \right), \quad \sum_{j=1}^{m} n_j = k, \quad k = 1, 2, \cdots \tag{6.36}$$

6.6 Reconstruction of Quadrature Points

The major drawback of proposed approach for data assimilation lies in possible degeneracy of weights after few time steps. In other words, most of the weights reduces to zero after few measurement updates and proposed data assimilation can result in a erroneous estimate. To surpass this problem, a resampling approach is considered to update the weights and model forecast after each measurement update. Hence, after each measurement update, new quadrature points are reconstructed based on obtained posterior statistics from data assimilation. Numerical model is then simulated to reproduce model forecasts using these new quadrature points.

One can use sophisticated approximation methods like Maximum Entropy Principle [124], Polynomial Chaos surrogate model, etc, for reconstruction new quadrature points. In here, we have used a simple approximation technique which *approximates* posterior pdf with a uniform pdf. Note that mean and covariance of approximated uniform pdf are equal to posterior statistics, obtained from data assimilation. New quadrature points are then generated based on this approximated uniform pdf.

6.7 A Simple Example

To verify the performance of the proposed approach, we consider the dispersion of Propane over Manhattan island, similar to the Test Case 1 in 5.3.3. The domain of interest and applied wind field (at one specific time) are shown in Fig. 6.4. Simulation time is considered to be 10 $hrs.$ starting from 00 : 00 of September 1^{st}, 2013. North American Regional Reanalysis wind data at pressure level 100 kpa (height $\simeq 100\ m.$) is used as the windfield for simulation. Four instantaneous mass releases are considered where their locations are known and the only uncertain parameters are their amount of mass release. It is assumed that releases happen at the same time, i.e. all source releases happen at 00 : 00 of September 1^{st}. All mass releases are assumed to be uniformly distributed between 50 kg and 150 kg. Fig. 6.4 illustrates source locations and the windfield (at $t = 0\ hrs.$) over the two dimensional spatial domain. A set of 161 CUT8 quadrature points are used to quantify the uncertainty involved in concentration of propane. Simulation of dispersion/advection has been performed using SCIPUFF numerical model, where concentration of propane is recorded every 30 $mins.$

We assumed that measurement data is available every 30 $mins.$, starting from 02 : 00 of September 1^{st}. A random realization of mass $[m_1, m_2, m_3, m_4] = [66.2, 137.04, 83.8, 122.1]^T\ kg.$ was used for simulation of dispersion phenomena. Then, obtained concentration field was discretized and polluted with a uniformly distributed integer random field $\omega(lon, lat, t)$ to generate the measurement data. The magnitude of ω at each spatio-temporal location is between -5 and $+5$, i.e. $\omega(lon, lat, t) \in \{-5, -4, \cdots, +5\}$. We use the Hausdorff metric and Euclidean distance to calculate the error between model predictions and measurement data. Note that corresponding discretization has been made in model prediction outputs before calculation of Hausdorff metric between each model output and measurement data.

We have performed resampling after each weight update. In other words, we generate a new set of quadrature points after each weight update, based on the procedure we explained in Section 6.6. In addition, $N_m = 3$ in our simulations.

Fig. 6.5 and Fig. 6.6 represent statistics (mean and variance) of parameter estimates using Hausdorff metric and Euclidean distance. From Fig. 6.5 it is clear that mean estimate of source parameters converge to their actual value and they show similar convergence behavior irrespective of the use of the metrics. However, Fig. 6.6 shows that the significant difference exist between variance of source parameters as computed with the application

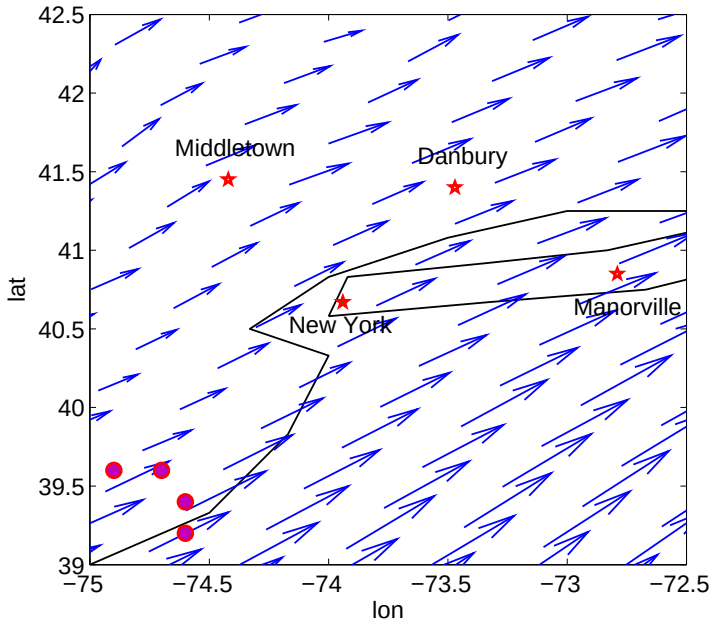


Figure 6.4: Schematic layout of Propane release over New York region, source locations are shown with purple circles, the wind-field (at $t = 0\ hr$ and pressure level $= 100\ kPa$) is shown over the two dimensional domain with blue vector field. Geographical location of different cities can be seen on the background map

of HD and Euclidian metrics. In other words, using Hausdorff metric consistently results in greater variance for source parameter estimates, as compared to the Euclidean distance. This behavior can be easily understood by considering Eq. (6.5). As Eq. (6.5) shows, Hausdorff metric is constructed based on highly nonlinear operators (minimum and maximum). Hence, it is possible that two very close image have large value of Hausdorff distance. On the other hand, Euclidean distance is a smooth function of difference between model forecast and observational data and small changes in either of these images results in small changes in the value of Euclidean distance. Consequently, these inherent properties of applied metrics effects statistics of source parameter estimates and result in completely different statistics for the variance of source parameters.

To summarize, both Hausdorff metric and Euclidean distance result in similar values for mean estimate of source parameters. But obtained estimates using the Euclidean distance are much more confident (almost a point estimate in final time) than obtained estimates

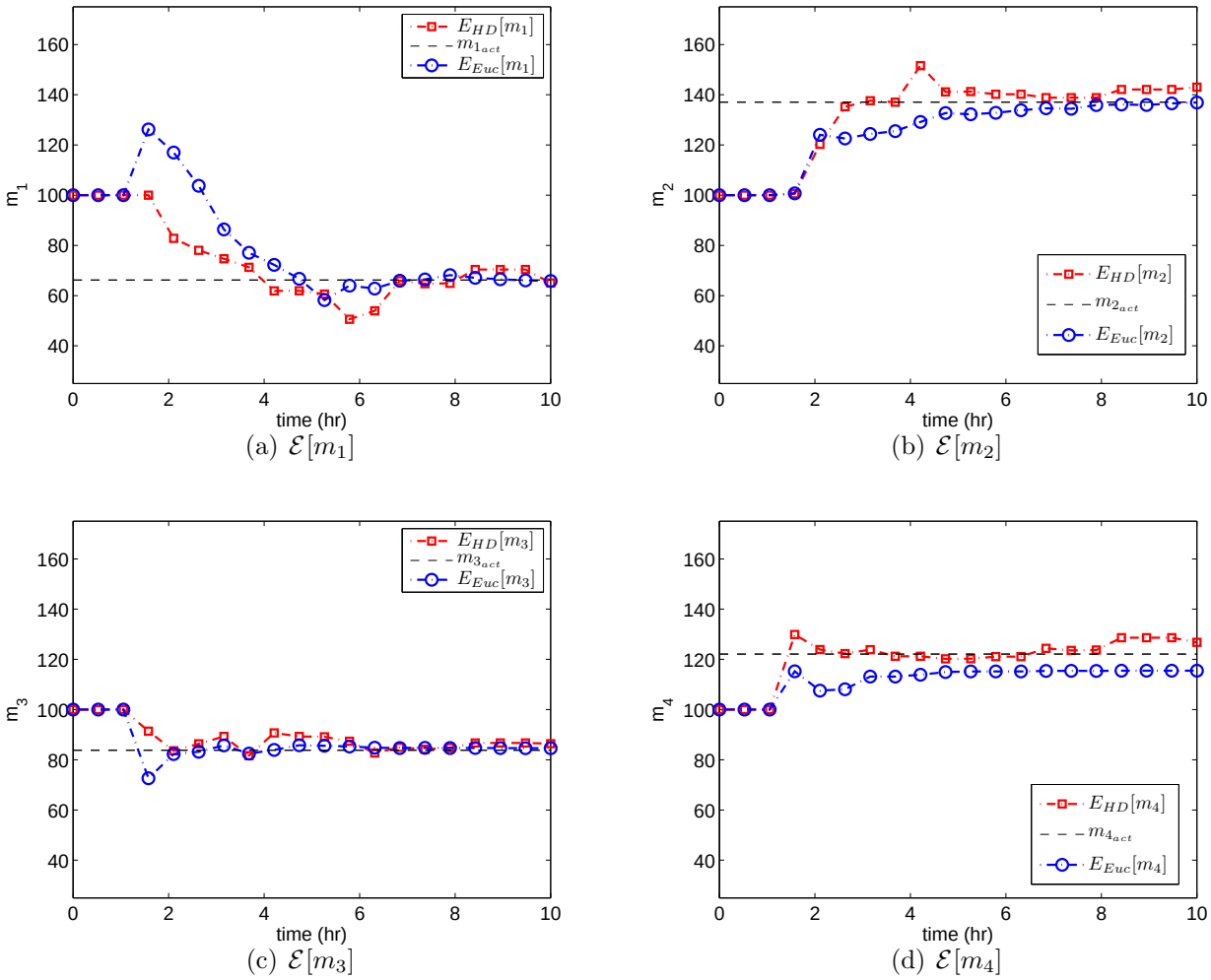
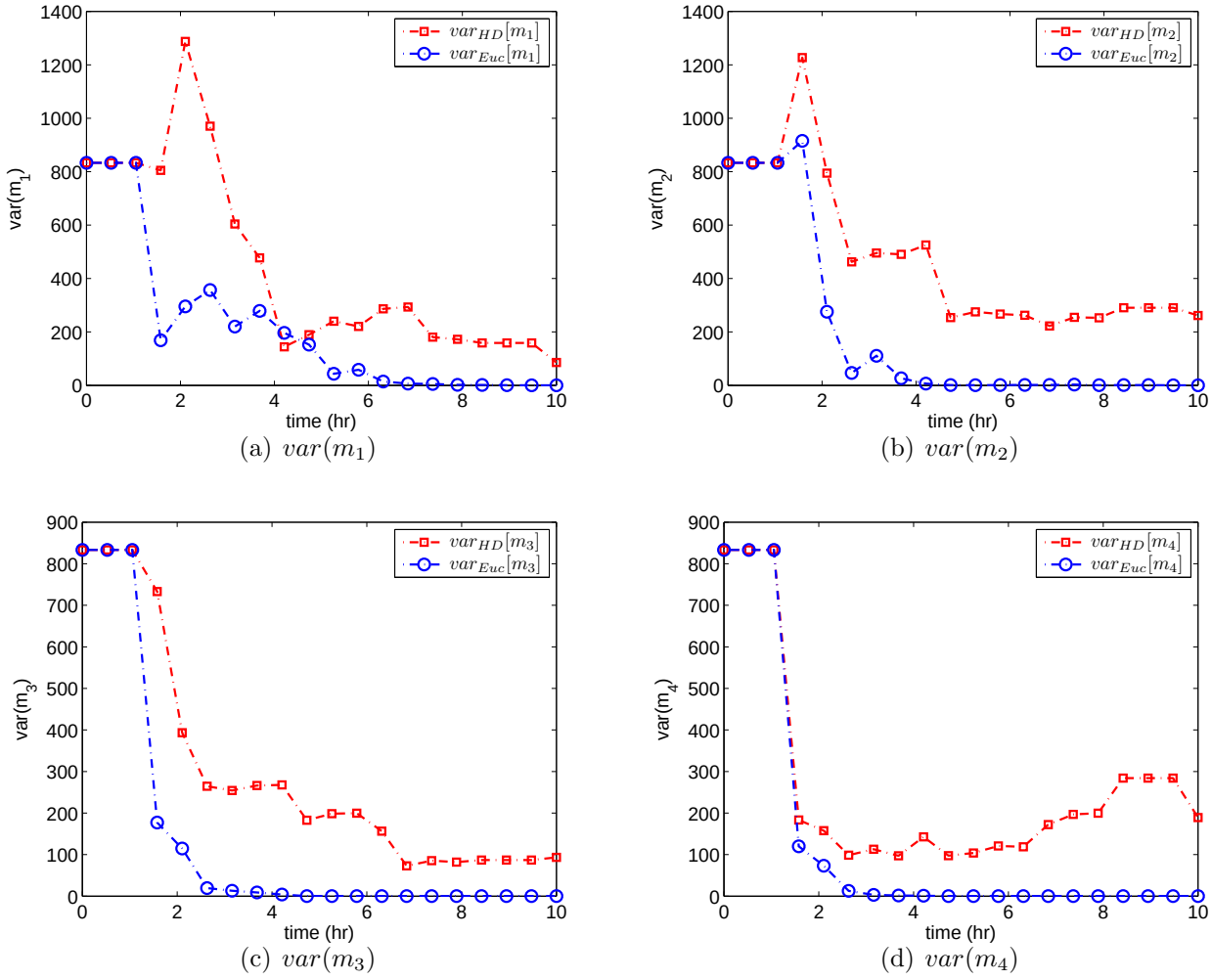


Figure 6.5: Mean estimate of source parameters during the time. Hausdorff (HD) and Euclidean (Euc) metrics are used for evaluation of the error between model forecasts and observation data. Dashed black lines correspond with actual value of parameters.

using the Hausdorff metric.

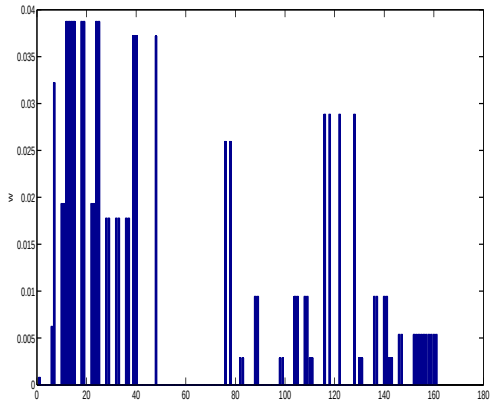
We have also shown posterior weights and corresponding Hausdorff and Euclidean distance in Fig. 6.7 and Fig. 6.8, respectively. As one can see, most of the posterior weights are zero when in the case of the Euclidean distance. On the other hand, the use of the Hausdorff metric results in non-zero values for most of the posterior weights. This is due the fact that Hausdorff metric returns similar values for most of model forecast ensembles, as Fig. 6.7(b), Fig. 6.7(d), and Fig. 6.7(f) shows. In other words, Hausdorff metric is less discriminatory in comparison with the Euclidean distance.



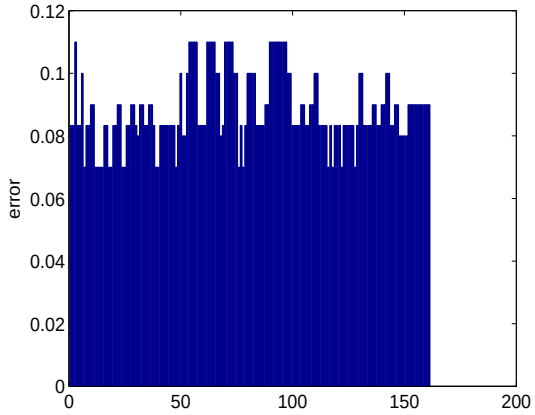
(a) $var(m_1)$ (b) $var(m_2)$

(c) $var(m_3)$ (d) $var(m_4)$

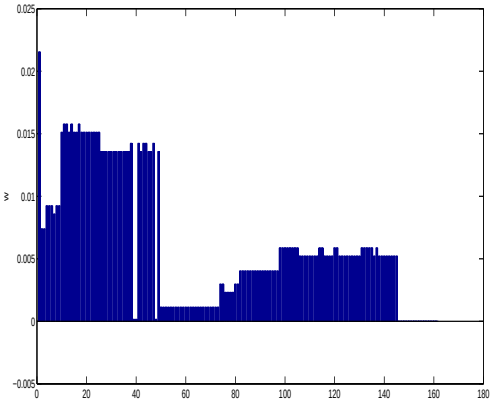
Figure 6.6: Variance of source parameter estimates during the time. Hausdorff (HD) and Euclidean (Euc) metrics are used for evaluation of error between model forecasts and observation data.



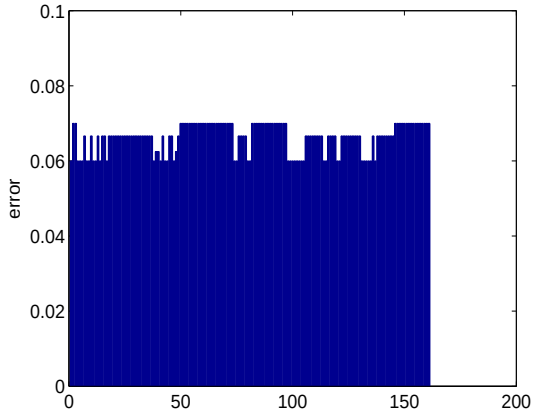
(a) $w, t = 2\ hrs.$



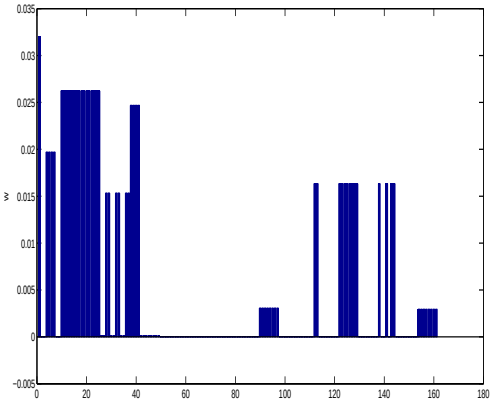
(b) $HD, t = 2\ hrs.$



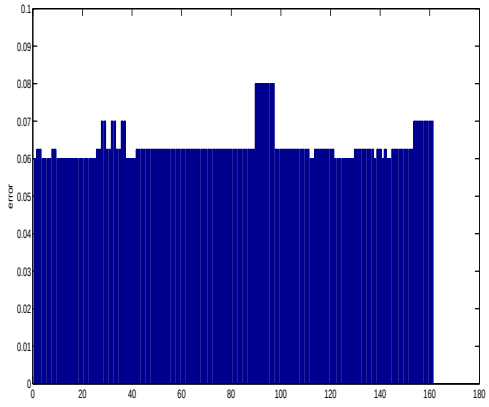
(c) $w, t = 6\ hrs.$



(d) $HD, t = 6\ hrs.$

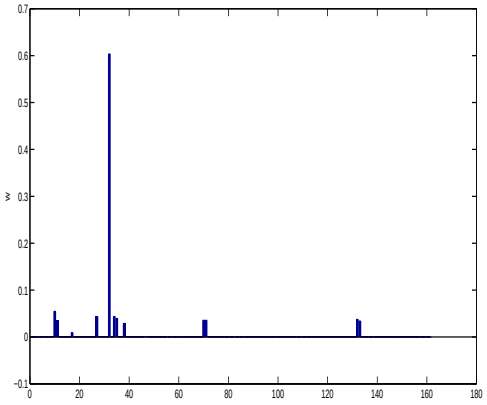


(e) $w, t = 6\ hrs.$

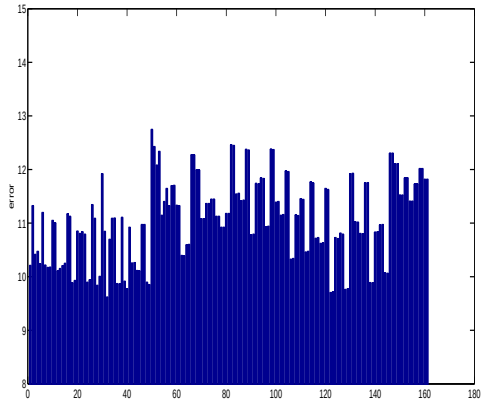


(f) $HD, t = 6\ hrs.$

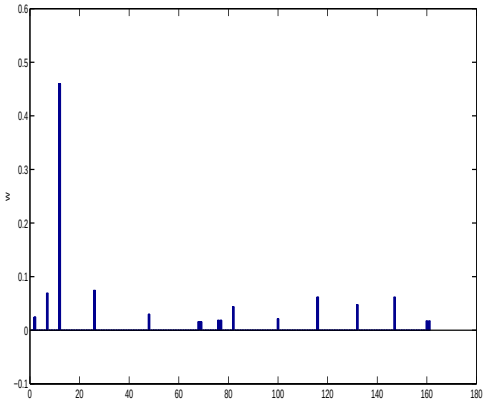
Figure 6.7: Hausdorff distance and corresponding posterior weights at different times a) weights at $t = 2\ hrs.$, b) Hausdorff distance at $t = 2\ hrs.$, c) weights at $t = 6\ hrs.$, d) Hausdorff distance at $t = 6\ hrs.$, e) weights at $t = 10\ hrs.$, e) Hausdorff distance at $t = 10\ hrs.$,



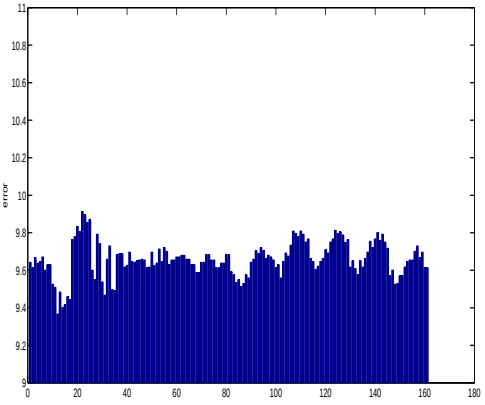
(a) $w, t = 2\ hrs.$



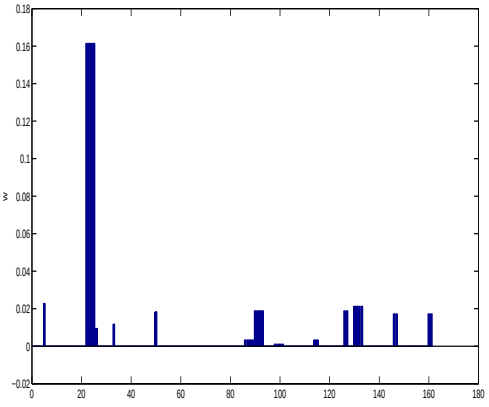
(b) Euclidean distance, $t = 2\ hrs.$



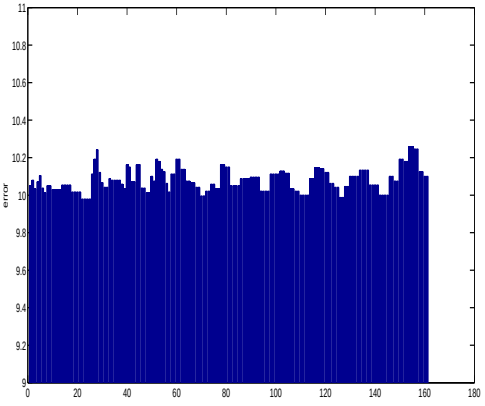
(c) $w, t = 6\ hrs.$



(d) Euclidean distance, $t = 6\ hrs.$



(e) $w, t = 6\ hrs.$

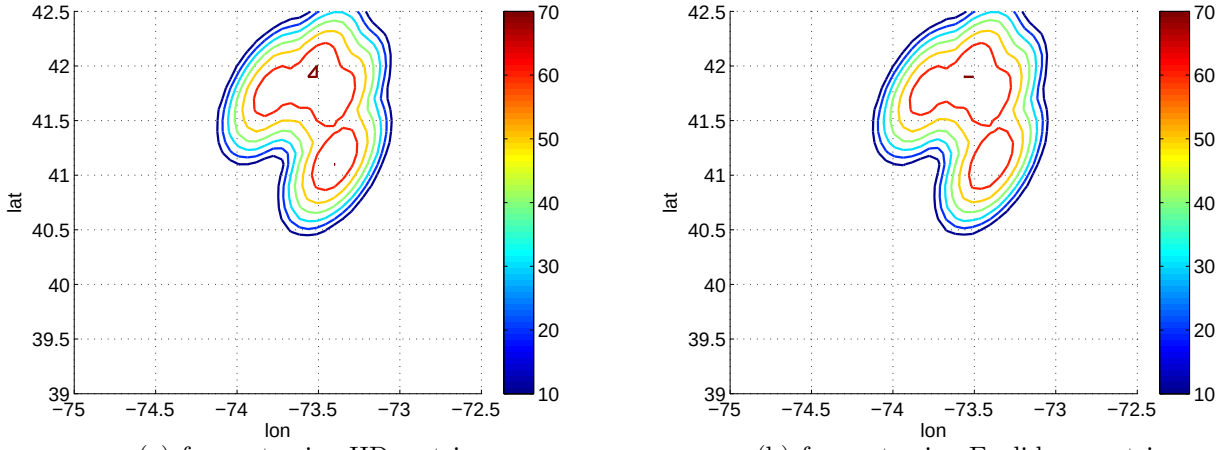


(f) Euclidean distance, $t = 6\ hrs.$

Figure 6.8: Euclidean distance and corresponding posterior weights at different times a) weights at $t = 2\ hrs.$, b) Euclidean distance at $t = 2\ hrs.$, c) weights at $t = 6\ hrs.$, d) Euclidean distance at $t = 6\ hrs.$, e) weights at $t = 10\ hrs.$, e) Euclidean distance at $t = 10\ hrs.$,

Forecast of Concentration Field

The quality of the source parameter estimates can be verified by performing a single deterministic run of SCIPUFF numerical model corresponding to the estimated posterior mean of source parameters and comparing it against observation data. Fig. 6.9 illustrates the forecast of concentration field at $t = 9:00\ hrs.$ after release, obtained using source parameter estimates at $t = 8:30$.



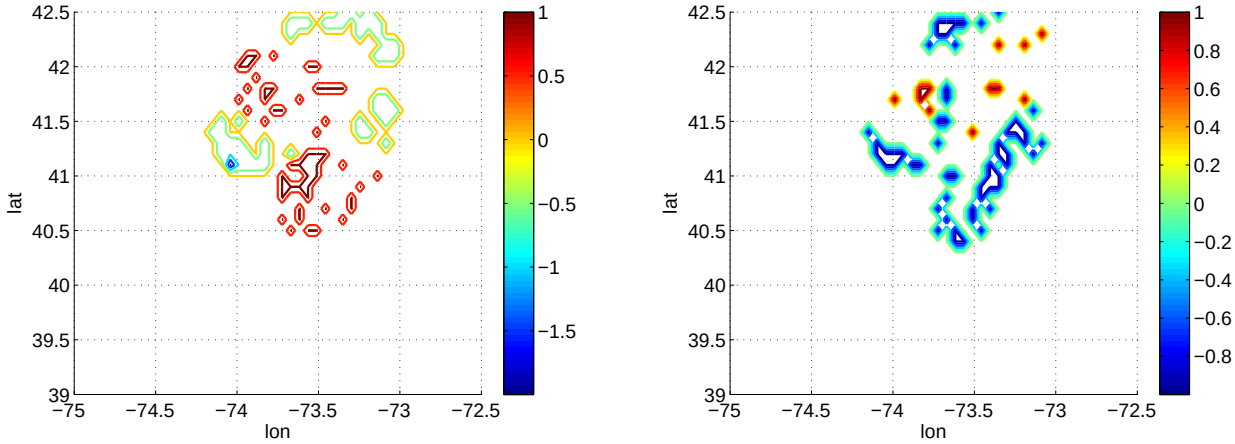
(a) forecast using HD metric (b) forecast using Euclidean metric

Figure 6.9: Forecast of concentration field at $t = 9:00\ hrs.$ using *posterior mean estimate* of source parameters at $t = 8:30$ while using a) Hausdorff metric, b) Euclidean metric

In addition, the discrepancy between forecast and true concentration field is also shown in Fig. 6.10, where correspondence between forecast and true concentration field is indisputable.

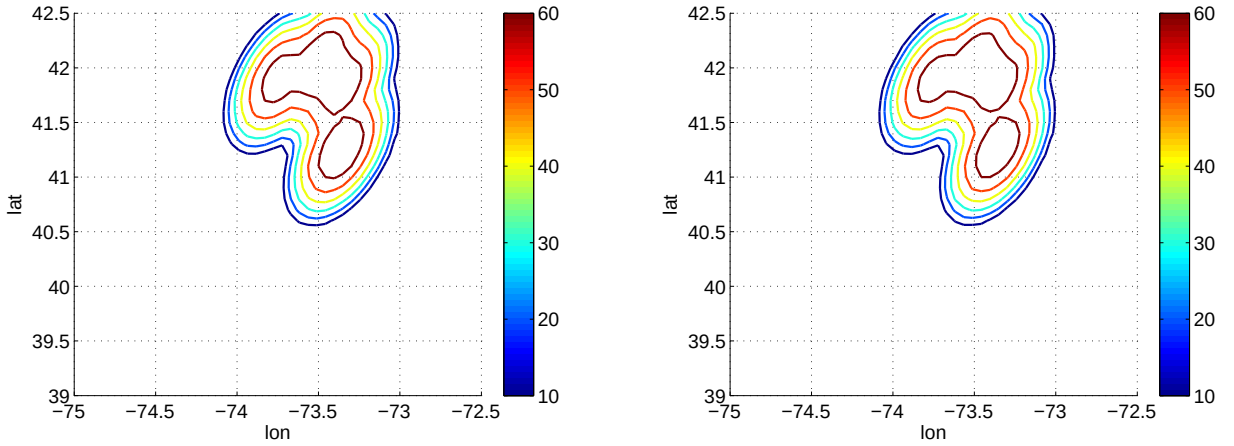
Similarly, Fig. 6.11 shows forecast of concentration field at $t = 9:30$, which is obtained using source parameter estimates at $t = 9:00$ and Fig. 6.12 represents the discrepancy between forecast and true concentration fields. These results clearly indicate that the forecast of concentration field based on the source parameter estimates matches very well with the true concentration field.

To further study performance of our methodology in forecasting concentration field, we have compared the discrepancy between forecast and true concentration fields in Table 6.1 and Table 6.2. We have also shown prior error in these Tables, where prior error denotes the discrepancy between forecast concentration field, obtained using prior mean estimates of source parameters, and true concentration field. Note that corresponding error metric



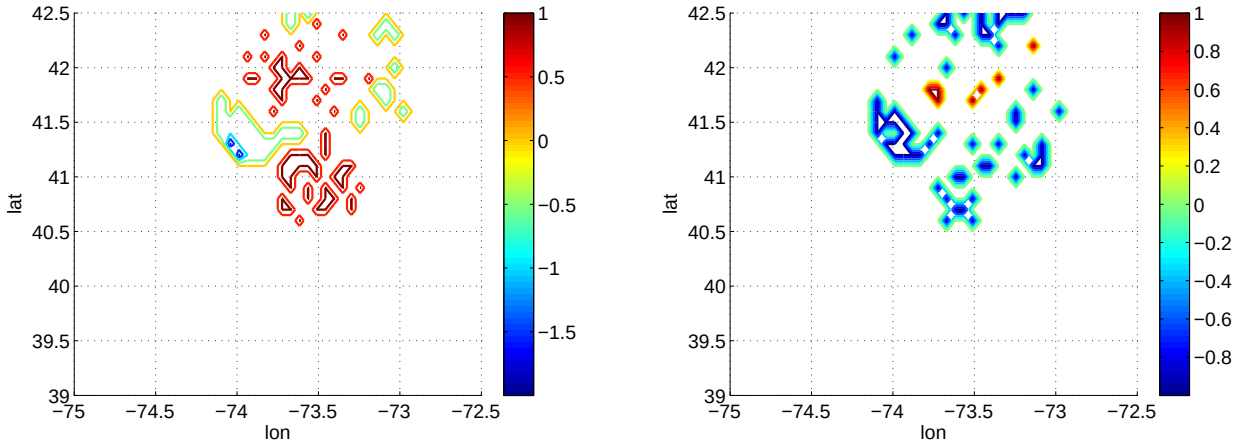
(a) using HD metric (b) using Euclidean metric

Figure 6.10: Discrepancy between forecast concentration field at $t = 9:00\ hrs.$ (obtained using *posterior mean estimate* of source parameters at $t = 8:30$) and true concentration field while using a) Hausdorff metric, b) Euclidean metric



(a) forecast using HD metric (b) forecast using Euclidean metric

Figure 6.11: Forecast of concentration field at $t = 9:30$ using *posterior mean estimate* of source parameters at $t = 9:00\ hrs.$ while using a) Hausdorff metric, b) Euclidean metric



(a) using HD metric (b) using Euclidean metric

Figure 6.12: Discrepancy between forecast concentration field at $t = 9:30$ (obtained using *posterior mean estimate* of source parameters at $t = 9:00\ hrs.$) and true concentration field while using a) Hausdorff metric, b) Euclidean metric

Table 6.1: Comparison of the error between forecast and true concentration fields, obtained using Euclidean distance at different times.

	Prior Error	$e_{Euclidean}$
$t = 9\ hrs.$	1141	69
$t = 9:30\ hrs.$	1079	67

is used for each of the methods to evaluate the discrepancy between model forecast and true concentration field. In other words, Euclidean distance is used to find the discrepancy between model forecasts and true concentration field in Table 6.1. Similarly, Hausdorff metric is used to find the difference between model forecasts and true concentration field in Table 6.2. As Table 6.1 and Table 6.2 represent, the error of model forecast,obtained by posterior mean estimate of parameters, is always less than the error of model forecast obtained by prior mean estimate of source parameters. As well, associated error in model forecast reduces with time.

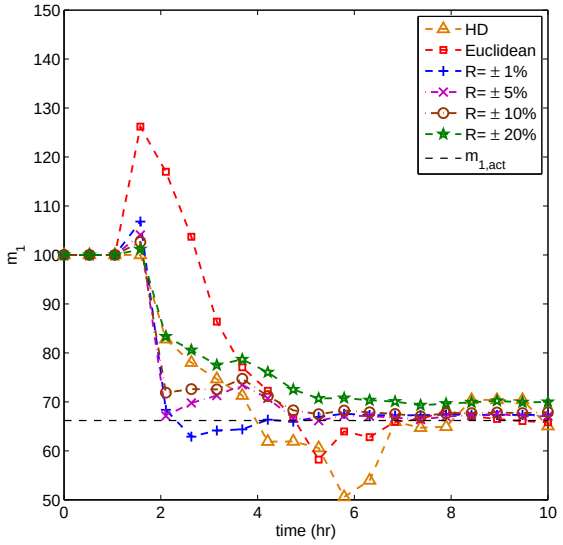
Table 6.2: Comparison of the error between forecast and true concentration fields, obtained using Hausdorff metric at different times.

	Prior Error	e_{HD}
$t = 9\ hrs.$	0.04	0.02
$t = 9:30\ hrs.$	0.04	0.02

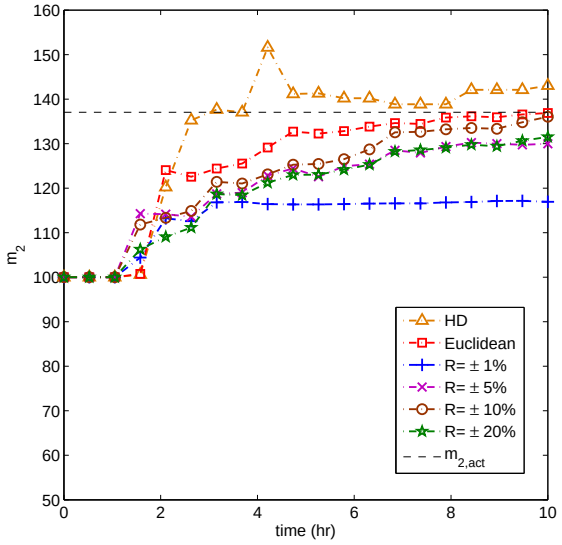
Comparison with Minimum Variance Estimation

To compare the performance of the proposed methodology, we have also shown performance of the minimum variance estimation framework. Note that in the minimum variance framework, we need to *assume* some value for the covariance of associated noise in measurement data. Fig. 6.13 illustrates mean of source parameter estimates while using our proposed method and the minimum variance framework. Same measurement data is used in both methods. We have shown simulation results for the minimum variance framework while considering four different values for covariance of associated noise in measurement data. As one can see, performance of the minimum variance framework crucially depends on the value of covariance matrix $\mathbf{R}$. For instance, when value of $\mathbf{R}$ is underestimated, the minimum variance framework results in inaccurate estimates of parameters. On the other hand, overestimation of the covariance matrix, $\mathbf{R}$ leads to less confident estimates of source parameters, as it can be seen in Fig. 6.14.

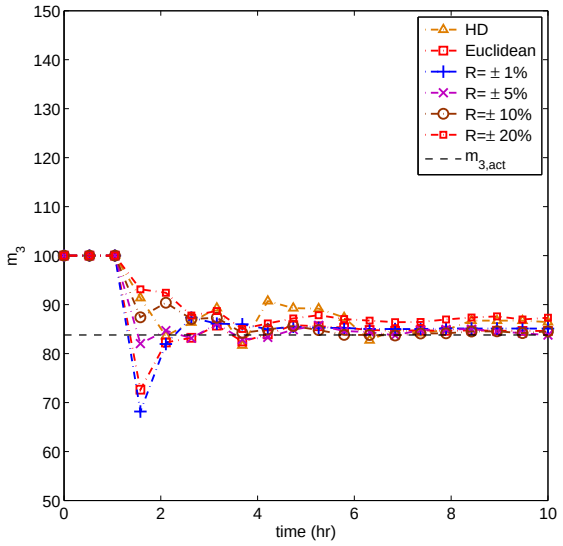
To summarize, Table 6.3 shows the Root Mean Square Error (RMSE) between actual values of source parameters and their estimates, obtained by the proposed approach and the minimum variance framework. Comparison of RMSE for the minimum variance estimation method shows that whenever statistics of the noise is accurately known, the minimum variance framework results in very reliable results with minimal error. But whenever inaccurate estimation of noise statistics is used, the minimum variance framework results in poor performance. Note that this reduction of performance is much sharper when the covariance of the noise is underestimated. On the other hand, comparing the RMSE of proposed approach with the minimum variance framework shows that its performance is comparable with performance of the minimum variance framework at its best performance, while it doesn't make use of any information regarding statistics of associated noise in measurement data.



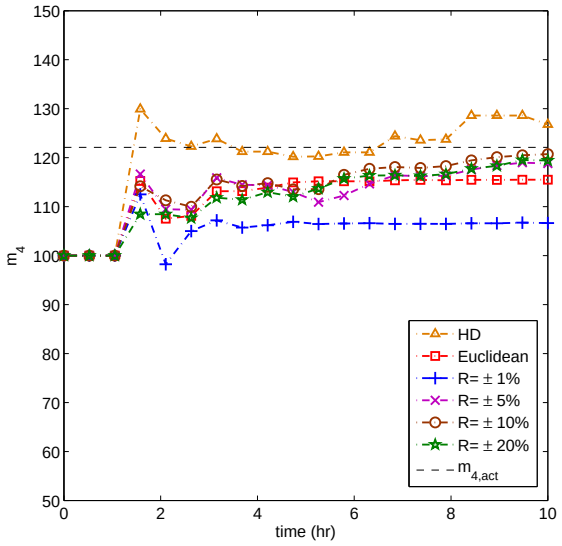
(a) $\mathcal{E}[m_1]$



(b) $\mathcal{E}[m_2]$

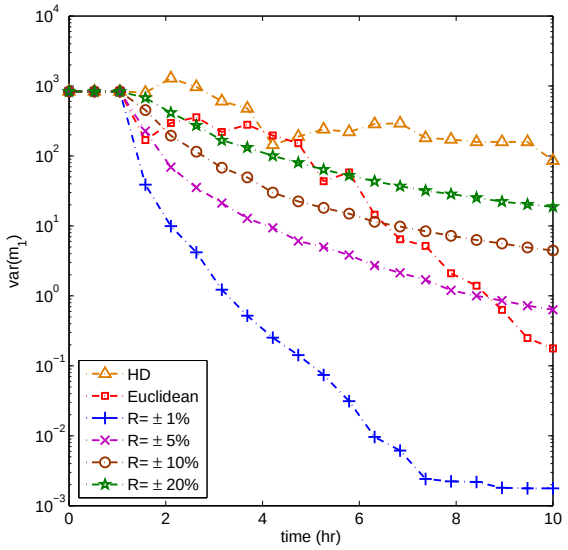


(c) $\mathcal{E}[m_3]$

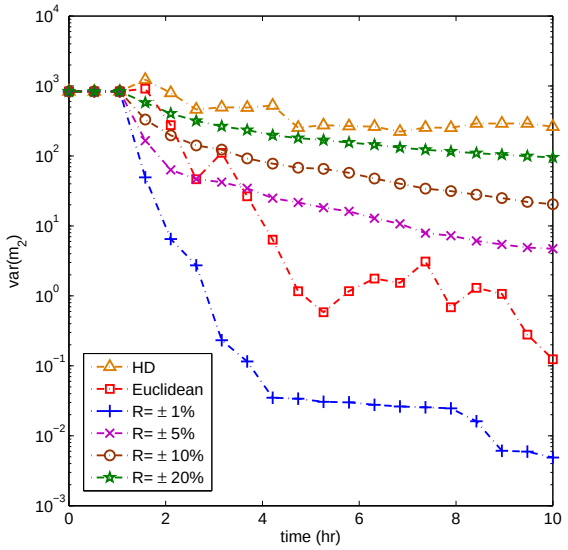


(d) $\mathcal{E}[m_4]$

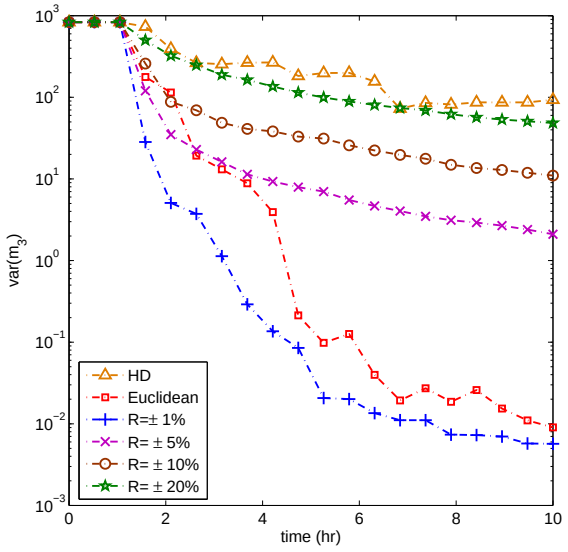
Figure 6.13: Comparison of mean estimate of source parameters using proposed method and minimum variance framework. Assumed values of the covariance matrix $\mathbf{R}$ are shown for each minimum variance simulation. Note that our proposed approach doesn't make use of noise statistics. a) m_1 b) m_2 c) m_3, and d) m_4



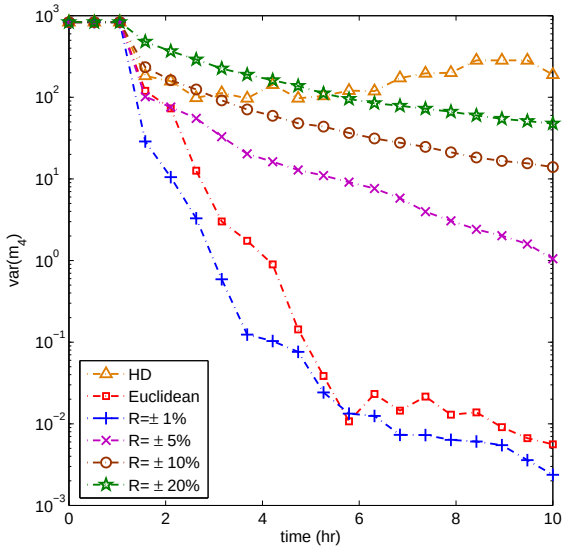
(a) $var(m_1)$



(b) $var(m_2)$



(c) $var(m_3)$



(d) $var(m_4)$

Figure 6.14: Comparison of variance of source parameter estimates using proposed method and minimum variance framework a) m_1 b) m_2 c) m_3, and d) m_4

Table 6.3: Overall RMSE between mean estimate of the source parameters and their actual parameters, obtained by proposed approach and the minimum variance framework.

Method	RMSE
proposed approach (using Hausdorff metric)	3.1715
proposed approach (using Euclidean distance)	2.4510
minimum variance framework ($R = \pm 1\%$)	8.5793
minimum variance framework ($R = \pm 5\%$)	3.5670
minimum variance framework ($R = \pm 10\%$)	2.2483
minimum variance framework ($R = \pm 20\%$)	3.7781

6.8 Summary

In this Chapter, we have developed a novel probabilistic approach for accurate parameter estimation and forecasting of large scale systems in the absence of any knowledge regarding the sensor error characteristics. The key idea of this approach is to minimize the expected value of error between model forecast and measurement data while preserving higher order prior or model forecasted central moments of error distribution. In this way, the proposed method tries to minimize the expected value of the difference between model forecasts and measurement data, while preserving the spread of model forecasts to account for possible inaccuracies of measurement data. We have also presented an iterative approach to convert the original nonlinear optimization problem involved in this process to a convex quadratic optimization problem, which is much easier to solve and computationally affordable.

We have demonstrated performance of the proposed approach for source parameter estimation and forecasting of an atmospheric release incident over New York region, using different error metrics like Hausdorff or Euclidean methods. Clearly, one can perform more comprehensive studies regarding the effect of different error metrics like Template Matching [143], Earth Mover's Distance [146], etc, on convergence behavior of proposed approach.

Once again, we emphasize that despite conventional estimation methods, like minimum variance framework or Bayesian Inference method, suggested approach doesn't require any information regarding statistics of associated noise in measurement data, which makes it a very powerful tool for applications where no accurate information regarding associated noise in measurement data is available.

Chapter 7

Conclusion

The overall goal of this dissertation is to provide an end-to-end computational probabilistic framework for uncertainty quantification, forecasting, and parameter estimation of large scale systems, with emphasis on atmospheric release incidents. Our approach consists of three different components. Uncertainty Quantification, Optimal Information Collection, and Data Fusion are three different components which come together to perform the task of source characterization and forecasting. We explained mathematical details for each of these components in different chapters.

In Chapter 2, we presented mathematical details of the method of quadrature points that is used for uncertainty quantification of dynamical systems in presence of parametric/initial condition uncertainty. Even though, the method of quadrature points can be utilized using any quadrature scheme, we applied a recently developed conjugate unscented transform quadrature points method which expedites the computations involved in uncertainty quantification. We also utilized an algorithm based on polynomial chaos surrogate model to approximate probabilistic hazard map of model output. Finally, we demonstrated performance and accuracy of presented methodology in this chapter by some real world examples.

In Chapter 3, we utilized two probabilistic estimation frameworks for parameter estimation of large scale systems. These two methods include the minimum variance framework and the Bayesian Inference based method. These methods provide posterior statistics of parameters, given model forecast information and available measurement data. As discussed in Section 3.6, each of these methods are applicable to different types of problems with different restrictions. The minimum variance framework is applicable to the problems where

the observation operator is an affine function of associated measurement noise and it only provides posterior estimate of the first two moments of uncertain parameters. Also, the information regarding the first two moments of associated measurement noise is required for the minimum variance method.

On the other hand, observation operator in the applied Bayesian Inference method can be any function of associated noise in measurement data, despite minimum variance framework. Also, against the minimum variance framework, Bayesian Inference method can be used to estimate higher order posterior statistics of uncertain parameters by tuning the order of required moments while using Eq. (3.22). However, the complete information of the likelihood function (equivalently complete information regarding distribution of associated noise in the measurement) is needed to implement the Bayesian Inference method.

We also made use of a polynomial chaos surrogate model in Chapter 3 to approximate the posterior distribution of uncertain parameters based on the obtained posterior moments from estimation algorithms. Finally, we demonstrated performance of the proposed estimation methods for source parameter estimation of atmospheric release incidents by some numerical examples.

In Chapter 5, we developed a general framework for optimal information collection in order to ensure performance of developed estimation methods. The key idea of the proposed approach is to optimally locate mobile sensors over spatial locations such that the mutual information between model forecast and observation data is maximized. This leads to an optimization problem which is solved by making use of the principle of optimality and dynamic programming to optimally locate mobile sensors at each time step. Note that solution of original optimization problem is computationally intractable and we made use of simplification techniques like limited lookahead method to alleviate the computational cost involved optimization problem. We showed performance of the proposed framework by simulation of some realistic atmospheric release incidents.

Finally in Chapter 6, we developed a novel approach to account for source characterization and forecasting of large scale systems in absence of sensor error characteristics. The key idea of the proposed method is to minimize the expected value of error between the model forecasts and measurement data while preserving higher order central moments of error distribution. In this way, proposed approach minimizes the expected value of the difference between model forecasts and measurement data, while preserving the spread of model forecasts to account for

possible inaccuracies of measurement data. We demonstrated performance of the proposed method for source parameter estimation and forecasting of a simulated atmospheric release incident. The key strength of the proposed method is that unlike conventional probabilistic estimation methods, like minimum variance framework or Bayesian Inference method, it doesn't require any information regarding statistics of associated noise in measurement data, which makes it a very powerful tool for applications where no accurate information regarding associated noise in measurement data is available.

Future Work

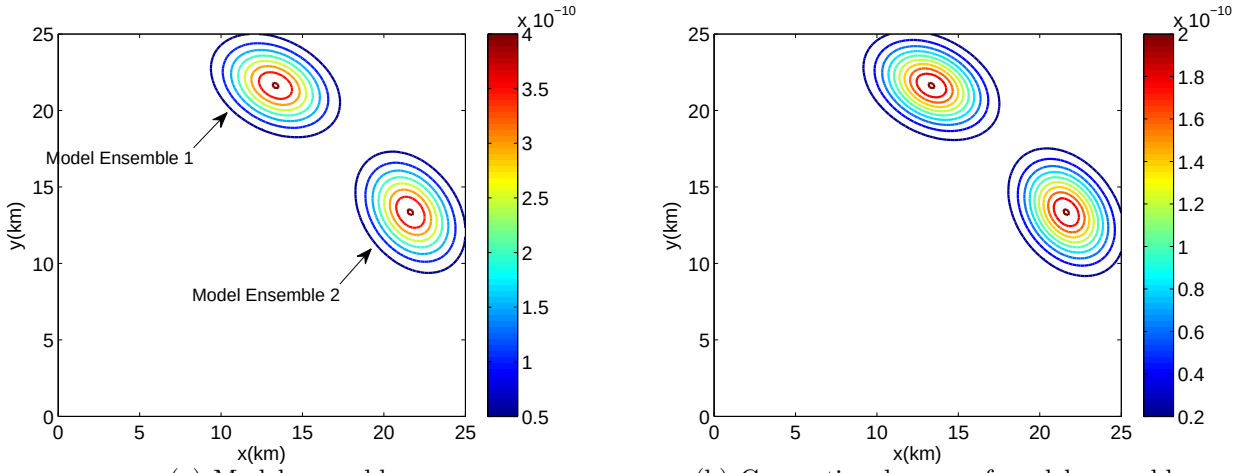
Even though we studied different aspects of the problem of source parameter estimation of large scale systems, still there exists many challenging issues that can be studied in order to expedite, reinforce, and generalize the developed methods. In the following, we briefly describe few of these open problems.

Uncertainty Quantification in Presence of Model Input Uncertainty

Developing new and efficient methods for uncertainty quantification of atmospheric models in presence of model input uncertainties is one of these challenging issues which still needs to be studied. In this dissertation, we didn't consider any model input uncertainties like applied wind-fields, while there can be a considerable source of uncertainty in applied wind-fields, depending on accuracy and resolution of utilized sensors for recording wind data. Hence, studying atmospheric models in presence of parametric and model input uncertainties will improve and generalize the methodology developed in this dissertation.

Shape Preserving Statistics

In this dissertation, we made use of conventional methods for calculating statistics of the model outputs while implementing estimation methods, which may degrade accuracy of estimation methods. In most of atmospheric release incidents all model output ensembles have similar characteristics, while obtained expected value of model output ensembles doesn't preserve these tributes. To explain this in more detail, assume two different model ensembles, as shown in Fig. 7.1(a). As one can see, both these model ensembles are similar in shape, but their spatial location is different. Conventional expected value of these ensembles is shown



(a) Model ensembles (b) Conventional mean of model ensembles

Figure 7.1: a) Two different model ensembles which are similar in shape, but differ from each other in spatial domain, b) Conventional expected value of model ensembles

in Fig. 7.1(b). As one can see, expected value of model ensembles doesn't preserve the shape of each ensemble and results in a bimodal shape which is completely different from original ensembles and consequently may effect performance of estimation and forecasting.

To avoid this drawback, one can perform estimation and forecasting by using *shape preserving statistics*. In this way, obtained expected value preserves the characteristics of model ensembles. Fig. 7 illustrates *shape preserving expected value* of model ensembles.

Optimal Path Planning

We developed a general framework in Chapter 5 to find the optimal way-points for mobile sensors to improve the performance of estimation method. Given optimal way-points, one needs to design the optimal path for mobile sensors which we didn't investigate in this dissertation. Various methods [147–149] have been proposed for optimal path planning of UAVs which can be combined with the proposed optimal information collection method to improve its applicability to realistic experiments.

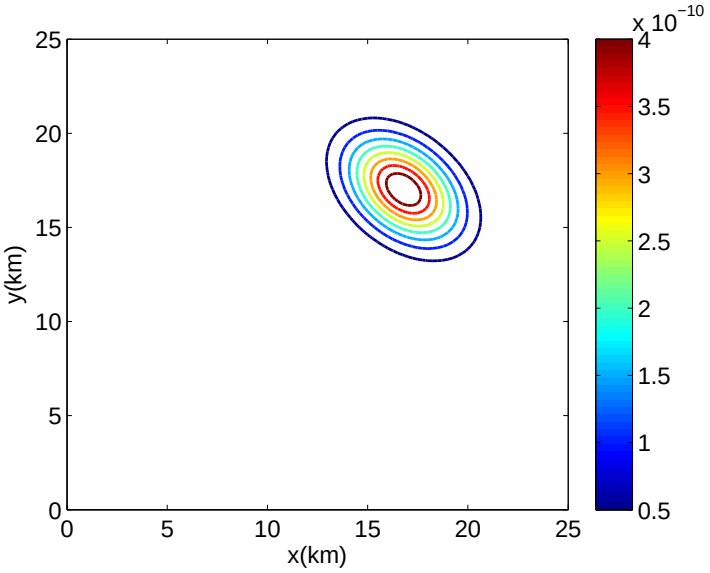


Figure 7.2: Shape preserving expected value of model ensembles (shown in Fig. 7.1(a))

Enhancement of Developed Approach for Data Assimilation in Absence of Sensor Error Characteristics

In Chapter 6, we have developed a novel method which provides us a statistical tool for source parameter estimation and forecasting of large scale systems without the use of any information regarding statistics of associated noise in measurement data. As we showed applied metric in evaluation of the error between model forecast ensembles and measurement data highly influences performance of estimation process. Hence, a comprehensive study regarding performance of the proposed method in presence of different available metrics is highly important. In addition, performance of the proposed approach under different values of N_m is another open area of research which can be investigated in more detail.

Bibliography

[1] R. F. Curtain, H. Zwart, An introduction to infinite-dimensional linear systems theory, Vol. 21, Springer, 1995.

[2] C. B. Vreugdenhil, Numerical methods for shallow-water flow, Vol. 13, Springer, 1994.

[3] K. Shankar Rao, Source estimation methods for atmospheric dispersion, Atmospheric Environment 41 (33) (2007) 6964–6973.

[4] R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems 1, Transactions of the ASME–Journal of Basic Engineering 82 (Series D) (1960) 35–45.

[5] S. E. Cohn, An introduction to estimation theory, JOURNAL-METEOROLOGICAL SOCIETY OF JAPAN SERIES 2 75 (1997) 147–178.

[6] A. Patwardhan, M. J. Small, Bayesian methods for model uncertainty analysis with application to future sea level rise, Risk analysis 12 (4) (1992) 513–523.

[7] D. Gamerman, H. F. Lopes, Markov chain Monte Carlo: stochastic simulation for Bayesian inference, CRC Press, 2006.

[8] A. O. Hero, D. Cochran, Sensor management: Past, present, and future, Sensors Journal, IEEE 11 (12) (2011) 3064–3075.

[9] J. L. Williams, J. W. Fisher, A. S. Willsky, Approximate dynamic programming for communication-constrained sensor network management, Signal Processing, IEEE Transactions on 55 (8) (2007) 4300–4311.

[10] R. Tharmarasa, T. Kirubarajan, M. L. Hernandez, Large-scale optimal sensor array management for multitarget tracking, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 37 (5) (2007) 803–814.

[11] B. J. Julian, M. Angermann, M. Schwager, D. Rus, Distributed robotic sensor networks: An information-theoretic approach, The International Journal of Robotics Research 31 (10) (2012) 1134–1154.

[12] G. M. Hoffmann, C. J. Tomlin, Mobile sensor network control using mutual information methods and particle filters, Automatic Control, IEEE Transactions on 55 (1) (2010) 32–47.

[13] H. Ishida, T. Nakamoto, T. Moriizumi, T. Kikas, J. Janata, Plume-tracking robots: A new application of chemical sensors, The Biological Bulletin 200 (2) (2001) 222–226.

[14] G. Sandini, G. Lucarini, M. Varoli, Gradient driven self-organizing systems, in: Intelligent Robots and Systems' 93, IROS'93. Proceedings of the 1993 IEEE/RSJ International Conference on, Vol. 1, IEEE, 1993, pp. 429–432.

[15] A. T. Hayes, A. Martinoli, R. M. Goodman, Swarm robotic odor localization, in: Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on, Vol. 2, IEEE, 2001, pp. 1073–1078.

[16] S. Kazadi, Extension of plume tracking behavior to robot swarms, in: Proceedings of the SCI2003 Conference, 2003.

[17] D. Zarzhitsky, D. F. Spears, W. M. Spears, Distributed robotics approach to chemical plume tracing, in: Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on, IEEE, 2005, pp. 4034–4039.

[18] R. Madankan, P. Singla, T. Singh, P. Scott, Polynomial chaos based bayesian approach for state and parameter estimation, AIAA Journal of Guidance, Navigation, and Control`doi:10.2514/1.58377`.

[19] A. H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, 1970.

[20] R. Stratonovich, Optimum nonlinear systems which bring about a separation of a signal with constant parameters from noise, Radiofizika 2 (6) (1959) 892–901.

[21] M. D. Sohn, P. Reynolds, N. Singh, A. J. Gadgil, Rapidly locating and characterizing pollutant releases in buildings, Journal of the Air & Waste Management Association 52 (12) (2002) 1422–1432.

[22] P. Sreedharan, M. D. Sohn, A. J. Gadgil, W. W. Nazaroff, Systems approach to evaluating sensor characteristics for real-time monitoring of high-risk indoor contaminant releases, Atmospheric Environment 40 (19) (2006) 3490–3502.

[23] A. Keats, E. Yee, F.-S. Lien, Bayesian inference for source determination with applications to a complex urban environment, Atmospheric environment 41 (3) (2007) 465–479.

[24] G. Johannesson, B. Hanley, J. Nitao, Dynamic bayesian models via monte carlo–an introduction with examples, Lawrence Livermore National Laboratory, UCRL-TR-207173.

[25] R. Bellman, R. Kalaba, Selected Papers on Mathematical Trends in Control Theory, Dover, New York, NY, 1964.

[26] D. P. Bertsekas, Dynamic Programming and Optimal Control, vols I and II, Cambridge, 2000.

[27] D. Xiu, G. E. Karniadakis, The wiener–askey polynomial chaos for stochastic differential equations, SIAM Journal on Scientific Computing 24 (2) (2002) 619–644.

[28] H. G. Matthies, Quantifying uncertainty: modern computational representation of probability and applications, in: Extreme man-made and natural hazards in dynamics of structures, Springer, 2007, pp. 105–135.

[29] R. G. Ghanem, P. D. Spanos, Stochastic finite elements: a spectral approach, Vol. 41, Springer, 1991.

[30] D. Xiu, Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos, Computer Methods in Applied Mechanics and Engineering 191 (43) (2002) 4927–4948. doi:10.1016/S0045-7825(02)00421-8.
URL http://linkinghub.elsevier.com/retrieve/pii/S0045782502004218

[31] N. Wiener, The homogeneous chaos, American Journal of Mathematics 60 (4) (1938) 897–936.

[32] A. Kolmogorov, On analytical methods in the theory of probability, Math. Ann 104 (1931) 415–458.

[33] G. Terejanu, P. Singla, T. Singh, P. D. Scott, Uncertainty propagation for nonlinear dynamic systems using gaussian mixture models, Journal of Guidance, Control, and Dynamics 31 (6) (2008) 1623–1633.

[34] R. Linares, V. Kumar, P. Singla, J. L. Crassidis, Information theoretic space object data association methods using an adaptive gaussian sum filter, Advances in the Astronautical Sci-ences 140 (2011) 665–680.

[35] H. Risken, Fokker-Planck Equation, Springer, 1984.

[36] J. M. Hammersley, D. C. Handscomb, Monte carlo methods, Vol. 1, Springer, 1964.

[37] K. Karhunen, Über lineare Methoden in der Wahrscheinlichkeitsrechnung, Vol. 37, Universitat Helsinki, 1947.

[38] M. Loeve, Probability theory, vol. ii, Graduate texts in mathematics 46 (1978) 0–387.

[39] K. Dalbey, A. K. Patra, E. B. Pitman, M. I. Bursik, M. F. Sheridan, Input uncertainty propagation methods and hazard mapping of geophysical mass flows 113 (2008) 1–16. doi:10.1029/2006JB004471.

[40] R. Madankan, P. Singla, A. Patra, M. Bursik, J. Dehn, M. Jones, M. Pavolonis, B. Pitman, T. Singh, P. Webley, Polynomial chaos quadrature-based minimum variance approach for source parameters estimation, Procedia Computer Science 9 (0) (2012) 1129 – 1138, ¡ce:title¿Proceedings of the International Conference on Computational Science, ICCS 2012¡/ce:title¿. doi:10.1016/j.procs.2012.04.122.

[41] I. A. Abramowitz, Milton; Stegun, Handbook of Mathematical Functions (with Formulas, Graphs, and Mathematical Tables), Dover, New York, 1972, Ch. 25.4, Integration.

[42] J. Waldvogel, Fast construction of the fejér and clenshaw–curtis quadrature rules, BIT Numerical Mathematics 46 (1) (2006) 195–202.

[43] J. Garcke, et al., Sparse grid tutorial, Mathematical Sciences Institute, Australian National University, Canberra Australia.

[44] N. Adurthi, The conjugate unscented transform - a method to evaluate multidimensional expectation integrals, Master's thesis, University at Buffalo.

[45] A. Nagavenkat, P. Singla, T. Singh, The Conjugate Unscented Transform-An Approach to Evaluate Multi-Dimensional Expectation Integrals, in: Proceedings of the American Control Conference, ACC2012, Montreal, Canada, 2012.

[46] A. Nagavenkat, P. Singla, T. Singh, Conjugate unscented transform rules for uniform probability density functions, Proceedings of the American Control Conference, 2013.

[47] A. Nagavenkat, P. Singla, T. Singh, Conjugate unscented transform and its application to filtering and stochastic integral calculation, AIAA Guidance, Navigation, and Control Conference, 2012.

[48] S. Julier, J. Uhlmann, H. Durrant-Whyte, A new method for the nonlinear transformation of means and covariances in filters and estimators, IEEE Transactions on Automatic Control AC-45 (3) (2000) 477–482.

[49] A. Nagavenkat, P. Singla, T. Singh, The conjugate unscented transform-an approach to evaluate multi-dimensional expectation integrals, Proceedings of the American Control Conference, 2012.

[50] O. LeMaitre, O. Knio, H. Najm, R. Ghanem, A stochastic projection method for fluid flow: I. basic formulation, Journal of Computational Physics 173 (2001) 481–511.

[51] D. Xiu, J. Hesthaven, High-order collocation methods for differential equations with random inputs, SIAM Journal of Scientific Computing 27 (2005) 1118–1139.

[52] M. Berveiller, B. Sudret, M. Lemaire, Stochastic finite element: a non intrusive approach by regression, Rev. Eur. Mec. Numer. 15 (2006) 81–92.

[53] R. I. Sykes, S. F. Parker, D. S. Henn, SCIPUFF Version 2.2 Technical Documentation (729).

[54] C. Allen, G. Young, S. Haupt, Improving pollutant source characterization by better estimating wind direction with a genetic algorithm, Atmospheric Environment 41 (11) (2007) 2283–2289. doi:10.1016/j.atmosenv.2006.11.007.

[55] K. J. Long, S. E. Haupt, G. S. Young, Assessing sensitivity of source term estimation, Atmospheric Environment 44 (12) (2010) 1558–1567. doi:10.1016/j.atmosenv.2010.01.003.

[56] L. M. Rodriguez, S. Ellen Haupt, G. S. Young, Impact of sensor characteristics on source characterization for dispersion modeling, Measurement 44 (5) (2011) 802–814. doi:10.1016/j.measurement.2011.01.014.

[57] S. E. Albo, O. O. Oluwole, R. C. Miake-Lye, The Aerodyne Inverse Modeling System (AIMS): Source estimation applied to the FFT 07 experiment and to simulated mobile sensor data, Atmospheric Environment 45 (33) (2011) 6085–6092. doi:10.1016/j.atmosenv.2011.07.033.

[58] S. Haupt, C. Allen, G. Young, A Genetic Algorithm Method for Sensor Data Assimilation and Source Characterization, The 2006 IEEE International Joint Conference on Neural Network Proceedings (2006) 5096–5103doi:10.1109/IJCNN.2006.247238.

[59] C. T. Allen, S. E. Haupt, G. S. Young, Source Characterization with a Genetic AlgorithmCoupled DispersionBackward Model Incorporating SCIPUFF, Journal of Applied Meteorology and Climatology 46 (3) (2007) 273–287. doi:10.1175/JAM2459.1.

[60] K. J. Long, S. E. Haupt, G. S. Young, L. M. Rodriguez, M. M. Iii, Source term estimation using genetic algorithm and scipuff, in: 7 th Conference on Artificial Intelligence and its Applications to the Environmental Sciences, 2009.

[61] J. M. Sohn, M. D., Sextro, R. G., Gadgil, A. J., & Daisey, Responding to sudden pollutant release in office buildings: 1. Framework and analysis tools, Indoor Air 13 (2003) 267–276.

[62] P. Sreedharan, M. D. Sohn, A. J. Gadgil, W. W. Nazaroff, Evaluating sensor characteristics for real-time monitoring of high-risk indoor contaminant releases (2005) 1–32.

[63] S. T. Fotini Katopodes Chow, Branko Kosovic, Chan, Source Inversion for Contaminant Plume Dispersion in Urban Environments Using Building- Resolving Simulations.

[64] D. Hartley, R. Prinn, Feasibility of determining surface emissions of trace gases using an inverse method in a three-dimensional chemical transport model, Journal of Geophysical Research: Atmospheres (1984–2012) 98 (D3) (1993) 5183–5197.

[65] M. E. Chang, D. E. Hartley, C. Cardelino, W.-L. Chang, Inverse modeling of biogenic isoprene emissions, Geophysical Research Letters 23 (21) (1996) 3007–3010.

[66] L. Bruhwiler, A. Michalak, W. Peters, D. Baker, P. Tans, An improved kalman smoother for atmospheric inversions, Atmospheric Chemistry and Physics 5 (10) (2005) 2691–2702.

[67] D. Brunner, S. Henne, C. Keller, S. Reimann, M. Vollmer, S. O'Doherty, M. Maione, An extended kalman-filter for regional scale inverse emission estimation, Atmospheric Chemistry and Physics 12 (7) (2012) 3455–3478.

[68] S. Napelenok, R. Pinder, A. Gilliland, R. Martin, A method for evaluating spatially-resolved no x emissions using kalman filter inversion, direct sensitivities, and space-based no 2 observations, Atmospheric Chemistry and Physics 8 (18) (2008) 5603–5614.

[69] M. Drews, B. Lauritzen, H. Madsen, Analysis of a kalman filter based method for on-line estimation of atmospheric dispersion parameters using radiation monitoring data, Radiation protection dosimetry 113 (1) (2005) 75–89.

[70] A. H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, San Diego, CA, 1970, Ch. 4-7.

[71] L. Deguillaume, M. Beekmann, L. Menut, Bayesian monte carlo analysis applied to regional-scale inverse emission modeling for reactive trace gases, Journal of Geophysical Research: Atmospheres (1984–2012) 112 (D2).

[72] E. Yee, F.-S. Lien, A. Keats, R. DAmours, Bayesian inversion of concentration data: Source reconstruction in the adjoint representation of atmospheric diffusion, Journal of Wind Engineering and Industrial Aerodynamics 96 (10) (2008) 1805–1816.

[73] G. Johannesson, K. Dyer, W. Hanley, B. Kosovic, S. Larsen, G. Loosmore, J. Lundquist, A. Mirin, Sequential monte-carlo based framework for dynamic data-driven event reconstruction for atmospheric release, in: Proc. of the Joint Statistical Meeting, Minneapolis, MN, American Statistical Association and Cosponsors, 2005, pp. 73–80.

[74] I. Senocak, N. W. Hengartner, M. B. Short, W. B. Daniel, Stochastic event reconstruction of atmospheric contaminant dispersion using bayesian inference, Atmospheric Environment 42 (33) (2008) 7718–7727.

[75] P. Robins, P. Thomas, Non-linear bayesian cbrn source term estimation, in: Information Fusion, 2005 8th International Conference on, Vol. 2, IEEE, 2005, pp. 8–pp.

[76] L. Delle Monache, J. K. Lundquist, B. Kosović, G. Johannesson, K. M. Dyer, R. D. Aines, F. K. Chow, R. D. Belles, W. G. Hanley, S. C. Larsen, et al., Bayesian inference and markov chain monte carlo sampling to reconstruct a contaminant source on a continental scale., Journal of Applied Meteorology & Climatology 47 (10).

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

[78] X. Zheng, Z. Chen, Inverse calculation approaches for source determination in hazardous chemical releases, Journal of Loss Prevention in the Process Industries 24 (4) (2011) 293–301. doi:10.1016/j.jlp.2011.01.002.

[79] A. Gelb, Applied Optimal Estimation, MIT Press, 1974, 203-214.

[80] P. Robins, V. Rapley, P. Thomas, A probabilistic chemical sensor model for data fusion, in: Information Fusion, 2005 8th International Conference on, Vol. 2, IEEE, 2005, pp. 7–pp.

[81] D. Schneider, W. Rose, L. Kelley, Tracking of 1992 eruption clouds from Crater Peak vent of Mount Spurr Volcano, Alaska using AVHRR, U. S. Geological Survey Bulletin 2139 (1995) 27–36.

[82] K. F. 867, http://en.wikipedia.org/wiki/klm_flight_867.

[83] CNN, New ash cloud could extend air travel threat, Accessed at http://www.cnn.com/2010/TRAVEL/04/19/volcano.ash/index.html.

[84] S. Gislason, H. Alfredsson, E. Eiriksdottir, T. Hassenkam, S. Stipp, Volcanic ash from the 2010 Eyjafjallajökull eruption, Applied Geochemistry 26, Supplement (0) (2011) S188 – S190. doi:10.1016/j.apgeochem.2011.03.100.

[85] D. B. Ryall, R. H. Maryon, Validation of the uk met. offices name model against the etex dataset, Atmospheric Environment 32 (1998) 4265–4276.

[86] B. Devenish, D. Thomson, F. Marenco, S. Leadbetter, H. Ricketts, H. Dacre, A study of the arrival over the united kingdom in april 2010 of the Eyjafjallajökull ash cloud using ground-based LIDAR and numerical simulations, Atmospheric Environment 48 (2012) 152–164.

[87] C. O'Dowd, S. Varghese, D. Martin, R. Flanagan, A. McKinstry, D. Ceburnis, J. Ovadnevaite, G. Martucci, J. Bialek, C. Monahan, H. Berresheim, A. Vaishya, T. Grigas, Z. McGraw, S. Jennings, B. Langmann, T. Semmler, R. McGrath, The Eyjafjallajökull ash plume part 2: Simulating ash cloud dispersion with REMOTE, Atmospheric Environment 48 (0) (2012) 143 – 151. doi:10.1016/j.atmosenv.2011.10.037.

[88] P. Webley, T. Steensen, M. Stuefer, G. Grell, S. Freitas, M. Pavolonis, Analyzing the Eyjafjallajökull 2010 eruption using satellite remote sensing, lidar and wrf-chem dispersion and tracking model, Journal of Geophysical Research: Atmospheres (1984–2012) 117 (D13).

[89] B. Heinold, I. Tegen, R. Wolke, A. Ansmann, I. Mattis, A. Minikin, U. Schumann, B. Weinzierl, Simulations of the 2010 Eyjafjallajökull volcanic ash dispersal over europe using cosmo–muscat, Atmospheric Environment 48 (2012) 195–204.

[90] A. Folch, A. Costa, S. Basart, Validation of the fall3d ash dispersion model using observations of the 2010 Eyjafjallajökull volcanic ash clouds, Atmospheric Environment 48 (2012) 165–183.

[91] A. Stohl, A. Prata, S. Eckhardt, L. Clarisse, A. Durant, S. Henne, N. Kristiansen, A. Minikin, U. Schumann, P. Seibert, et al., Determination of time-and height-resolved volcanic ash emissions and their use for quantitative ash dispersion modeling: the 2010 Eyjafjallajökull eruption, Atmos. Chem. Phys 11 (9) (2011) 4333–4351.

[92] R. P. Denlinger, M. Pavolonis, J. Sieglaff, A robust method to forecast volcanic ash clouds, Journal of Geophysical Research: Atmospheres 117 (D13) (2012) n/a–n/a. doi:10.1029/2012JD017732.
URL http://dx.doi.org/10.1029/2012JD017732

[93] N. I. Kristiansen, A. Stohl, A. J. Prata, N. Bukowiecki, H. Dacre, S. Eckhardt, S. Henne, M. C. Hort, B. T. Johnson, F. Marenco, B. Neininger, O. Reitebuch, P. Seibert, D. J. Thomson, H. N. Webster, B. Weinzierl, Performance assessment of a volcanic ash transport model mini-ensemble used for inverse modeling of the 2010 eyjafjallajkull eruption, Journal of Geophysical Research: Atmospheres 117 (D20) (2012) n/a–n/a. doi:10.1029/2011JD016844.
URL http://dx.doi.org/10.1029/2011JD016844

[94] H. W. Sorenson, Parameter estimation: principles and problems, Marcel Dekker New York, 1980.

[95] M. Bursik, M. Jones, S. Carn, K. Dean, A. Patra, M. Pavolonis, E. B. Pitman, T. Singh, P. Singla, P. Webley, et al., Estimation and propagation of volcanic source parameter uncertainty in an ash transport and dispersal model: application to the eyjafjallajokull plume of 14–16 april 2010, Bulletin of Volcanology (2012) 1–18.

[96] S. Carey, R. Sparks, Quantitative models of the fallout and dispersal of tephra from volcanic eruption columns, Bull. Volcanology 48 (1986) 109–125.

[97] T. Suzuki, A theoretical model for dispersion of tephra, Terra Scientific Publishing, Tokyo, 2005, pp. 95–116.

[98] H. Tanaka, Development of a prediction scheme for the volcanic ash fall from redoubt volcano, in: First Int'l Symposium on Volcanic Ash and Aviation Safety, Seattle, 1991, p. 58.

[99] C. Searcy, K. Dean, B. Stringer, PUFF: A volcanic ash tracking and prediction model, J. Volcanology and Geophysical Research 80 (1998) 1–16.

[100] P. Webley, K. Dean, J. Dehn, J. Bailey, R. Peterson, Volcanic ash dispersion modeling of the 2006 eruption of Augustine Volcano , USGS Professional Paper: Augustine Volcano 2006 eruption.

[101] L. Mastin, M. Guffanti, R. Servanckx, P. Webley, S. Barostti, K. Dean, R. Denlinger, A. Durant, J. Ewert, C. Gardner, A. Holliday, A. Neri, W. Rose, D. Schneider, L. Siebert, B. Stunder, G. Swanson, A. Tupper, A. Volentik, A. Waythomas, A multidisciplinary effort to assign realistic source parameters to models of volcanic ash-cloud

transport and dispersion during eruptions, J. of Volcanology and Geothermal Research 186 (2009) 10–21, special issue on Volcanic Ash Clouds; L. Mastin and P.W. Webley (eds.).

[102] R. S. J. Sparks, M. I. Bursik, S. N. Carey, J. S. Gilbert, L. S. Glaze, H. Sigurdsson, A. W. Woods, Volcanic Plumes, John Wiley & Sons, London, 1997, 574p.

[103] M. Bursik, S. Kobs, A. Burns, O. Braitseva, L. Bazanova, I. Melekestsev, A. Kurbatov, D. Pieri, Volcanic plumes and the wind: jetstream interaction examples and implications for air traffic, J. of Volcanology and Geothermal Research 186 (2009) 60–67.

[104] M. Bursik, Effect of wind on the rise height of volcanic plumes, Geophys. Res. Lett. 18 (2001) 3621–3624.

[105] B. Morton, J. Turner, G. Taylor, Gravitational turbulent convection from maintained and instantaneous sources, Proceedings Royal Soc. London Ser. A 234 (1956) 1–23.

[106] K. D. C. Searcy, W. Stringer, Puff: A high-resolution volcanic ash tracking model, Journal of Volcanology and Geothermal Research 80 (1998) 1–16.

[107] National Center for Environmental Prediction (2009). Unidata online access to the operational Global Forecasting System (GFS) numerical weather prediction model, `http://motherlode.ucar.edu:8080/thredds/catalog/fmrc/NCEP/GFS/Global_0p5deg/catalog.html`.

[108] National Center for Environmental Prediction (2009). Unidata online access to the operational North American Mesoscale (NAM) numerical weather prediction model, `http://motherlode.ucar.edu:8080/thredds/catalog/fmrc/NCEP/NAM/Alaska$_$11km/catalog.html`.

[109] United States Navy Fleet Numerical Meteorology and Oceanography Center (2009). Online access to the Navy Operational Global Atmospheric Prediction System numerical weather prediction model[link].
URL `https://www.fnmoc.navy.mil/public`

[110] T. W. Research, Forecasting, http://www.wrf-model.org/index.php.

[111] M. Ripepe, S. D. Angelis, G. Lacanna, B. Voight, Observation of infrasonic and gravity waves at soufriere hills volcano, montserrat, Geophysical Research Letters 37 (2010) L00E14, doi:10.1029/2010GL042557.

[112] A. W. Woods, M. I. Bursik, Particle fallout, thermal disequilibrium and volcanic plumes, Bulletin of Volcanology 53 (1991) 559–570.

[113] M. J. Pavolonis, W. F. Feltz, A. K. Heidinger, G. M. Gallina, A daytime complement to the reverse absorption technique for improved automated detection of volcanic ash, J. Atmos. Ocean. Technol. 23 (2006) 1422–1444.

[114] M. J. Pavolonis, Advances in extracting cloud composition information from spaceborne infrared radiances: A robust alternative to brightness temperatures. part i: Theory, J. Applied Meteorology and Climatology 49 (2010) 1992–2012.

[115] A. K. Heidinger, M. J. Pavolonis, Nearly 30 years of gazing at cirrus clouds through a split-window. part i: Methodology, J. Appl. Meteorol. and Climatology 48 (2009) 1110–1116.

[116] A. K. Heidinger, M. J. P. R. E. Holz, B. A. Baum, S. Berthier, Using calipso to explore the sensitivity to cirrus height in the infrared observations from npoess/viirs and goes-r/abi, Journal of Geophysical Research 115, doi:10.1029/2009JD012152.

[117] S. Scollo, M. Prestilippo, M. Coltelli, R. Peterson, G. Spata, A statistical approach to evaluate the tephra deposit and ash concentration from PUFF model forecasts, Journal of Volcanology and Geothermal Research 200 (2011) 129–142.

[118] S. Balay, J. Brown, K. Buschelman, V. Eijkhout, W. Gropp, D. Kaushik, M. Knepley, L. C. McInnes, B. Smith, H. Zhang, Petsc users manual revision 3.3.

[119] G. Petersen, H. Bojornsson, P. Arason, The impact of the atmosphere on the Eyjafjallajökull 2010 eruption plume, Journal of Geophysical Research 117. doi:10.1029/2011JD016762.

[120] C. Bonadonna, R. Genco, M. Gouhier, M. Pistolesi, R. Cioni, F. Alfano, A. Hoskuldsson, M. Ripepe, Tephra sedimentation during the 2010 Eyjafjallajökull eruption (ice-

land) from deposit, radar and satellit obervations, Journal of Geophysical Research 116. doi:10.1029/2011JB008426.

[121] H.-L. Choi, J. P. How, Continuous trajectory planning of mobile sensors for informative forecasting, Automatica 46 (8) (2010) 1266–1275.

[122] F. Bourgault, A. A. Makarenko, S. B. Williams, B. Grocholsky, H. F. Durrant-Whyte, Information based adaptive robotic exploration, in: Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on, Vol. 1, IEEE, 2002, pp. 540–545.

[123] S. MartíNez, F. Bullo, Optimal sensor placement and motion coordination for target tracking, Automatica 42 (4) (2006) 661–668.

[124] T. M. Cover, J. A. Thomas, Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing), Wiley-Interscience, 2006.

[125] A. H. Jazwinski, Stochastic processes and filtering theory, Courier Dover Publications, 2007.

[126] J. L. Crassidis, J. L. Junkins, Optimal estimation of dynamic systems, CRC press, 2011.

[127] R. K. Mehra, On the identification of variances and adaptive kalman filtering, Automatic Control, IEEE Transactions on 15 (2) (1970) 175–184.

[128] A. Bagchi, Continuous time systems identification with unknown noise covariance, Automatica 11 (5) (1975) 533–536.

[129] X. Rong Li, Y. Bar-Shalom, A recursive multiple model approach to noise identification, Aerospace and Electronic Systems, IEEE Transactions on 30 (3) (1994) 671–684.

[130] M. Oussalah, J. De Schutter, Adaptive kalman filter for noise identification, in: PROCEEDINGS OF THE INTERNATIONAL SEMINAR ON MODAL ANALYSIS, Vol. 3, KU Leuven; 1998, 2001, pp. 1225–1232.

[131] V. **S**mídl, On estimation of unknown disturbances of non-linear state-space model using marginalized particle filter, Tech. rep., Citeseer (2008).

[132] S. Saha, E. Ozkan, F. Gustafsson, V. Smidl, Marginalized particle filters for bayesian estimation of gaussian noise parameters, in: Information Fusion (FUSION), 2010 13th Conference on, IEEE, 2010, pp. 1–8.

[133] E. Özkan, V. **S**mídl, S. Saha, C. Lundquist, F. Gustafsson, Marginalized adaptive particle filtering for nonlinear models with unknown time-varying noise parameters, Automatica 49 (6) (2013) 1566–1575.

[134] B. J. Odelson, M. R. Rajamani, J. B. Rawlings, A new autocovariance least-squares method for estimating noise covariances, Automatica 42 (2) (2006) 303–308.

[135] M. R. Rajamani, J. B. Rawlings, Estimation of the disturbance structure from data using semidefinite programming and optimal weighting, Automatica 45 (1) (2009) 142–148.

[136] S. Sarkka, A. Nummenmaa, Recursive noise adaptive kalman filtering by variational bayesian approximations, Automatic Control, IEEE Transactions on 54 (3) (2009) 596–600.

[137] T. T. Cai, C.-H. Zhang, H. H. Zhou, et al., Optimal rates of convergence for covariance matrix estimation, The Annals of Statistics 38 (4) (2010) 2118–2144.

[138] N. Jaoua, F. Septier, E. Duflos, P. Vanheeghe, State and impulsive time-varying measurement noise density estimation in nonlinear dynamic systems using dirichlet process mixtures, in: Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on, IEEE, 2014, pp. 330–334.

[139] M. Darouach, M. Zasadzinski, M. Boutayeb, Extension of minimum variance estimation for systems with unknown inputs, Automatica 39 (5) (2003) 867–876.

[140] E. Ozkan, S. Saha, F. Gustafsson, V. Smidl, Non-parametric bayesian measurement noise density estimation in non-linear filtering, in: Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, IEEE, 2011, pp. 5924–5927.

[141] V. Smidl, A. Quinn, Variational bayesian filtering, Signal Processing, IEEE Transactions on 56 (10) (2008) 5020–5030.

[142] V. Di Gesu, V. Starovoitov, Distance-based functions for image comparison, Pattern Recognition Letters 20 (2) (1999) 207–214.

[143] H. Zhou, M. Chen, M. F. Webster, Comparative evaluation of visualization and experimental results using image comparison metrics, in: Proceedings of the conference on Visualization'02, IEEE Computer Society, 2002, pp. 315–322.

[144] L. Wang, Y. Zhang, J. Feng, On the euclidean distance of images, Pattern Analysis and Machine Intelligence, IEEE Transactions on 27 (8) (2005) 1334–1339.

[145] M. Grant, S. Boyd, Y. Ye, Cvx: Matlab software for disciplined convex programming (2008).

[146] Y. Rubner, C. Tomasi, L. J. Guibas, The earth mover's distance as a metric for image retrieval, International Journal of Computer Vision 40 (2) (2000) 99–121.

[147] J. Bellingham, M. Tillerson, A. Richards, J. P. How, Multi-task allocation and path planning for cooperating uavs, in: Cooperative Control: Models, Applications and Algorithms, Springer, 2003, pp. 23–41.

[148] S. A. Bortoff, Path planning for uavs, in: American Control Conference, 2000. Proceedings of the 2000, Vol. 1, IEEE, 2000, pp. 364–368.

[149] M. Shanmugavel, A. Tsourdos, B. White, R. Żbikowski, Co-operative path planning of multiple uavs using dubins paths with clothoid arcs, Control Engineering Practice 18 (9) (2010) 1084–1092.