The dissertation of Luna Marie Rodriguez was reviewed and approved* by the following:

George S. Young  
Professor of Meteorology and Geo-Environmental Engineering  
Dissertation Advisor  
Co-Chair of Committee

Sue Ellen Haupt  
Adjunct Professor of Meteorology  
Dissertation Advisor  
Co-Chair of Committee

Jenni Evans  
Professor of Meteorology

Jeffrey R. S. Brownson  
Assistant Professor of Energy and Mineral Engineering

Paul E. Bieringer  
Project Scientist at the National Center for Atmospheric Research  
Special Committee Member

Dr. Johannes Verlinde  
Professor of Meteorology  
Associate Head, Graduate Program in Meteorology

*Signatures are on file in the Graduate School
ABSTRACT

We can never quantify the atmospheric state precisely: there will always be an uncertainty associated with measured quantities and model output. This work seeks to understand both the uncertainty introduced by measurements and the uncertainty introduced by approximating the model’s nonlinear terms. We seek both to understand these sources of uncertainty and to incorporate that understanding throughout the scientific process for Atmospheric Transport & Dispersion (AT&D) problems. First we will examine how errors in the input wind fields may translate into AT&D model solution errors. We focus on street-level concentration plume errors that occur in building-aware AT&D models for a set of hazardous release scenarios where the release location varies relative to the building locations and city building configurations. Second, we use Source Term Estimation (STE) techniques to examine how estimates of uncertainty in measurements, e.g. the wind direction, can be used to help bound the problem.

We use two techniques to examine the STE problem. Using the Genetic Algorithm coupled with an AT&D model Variational (GA-Var) method we preform sensitivity analyses to achieve five goals: (1) establish adequate thresholds to filter out noise in our concentration data without decimating the signal; (2) use a robust statistical method to quantify the uncertainty in our predictions; (3) determine the best cost function for each of the variables we seek to retrieve; (4) given that real-time wind direction data are difficult to come by, determine if the GA-estimated wind direction is representative of the advecting wind; and (5) determine the robustness of the GA when a limited number of sensors are available.

To further examine the STE problem we use the Variational Iterative Refinement STE Algorithm (VIRSA). VIRSA is a combined modeling system that includes the Second-order Closure Integrated PUFF model, a hybrid Lagrangian-Eulerian Plume Model (LEPM), and its
formal adjoint. While numerous approaches to the STE problem exist, each with its own strengths and weaknesses, this approach addresses STE in an operational environment where computational resources are limited and a timely solution is critical. VIRSA is an adjoint method, computationally efficient, and fast but like any gradient descent minimization its downfall is that it can fall prey to local minima in the solution space. In the work presented here we incorporate new methods to address issues related to uncertainty and using what we know about that uncertainty to reduce the tendency to find local minima rather than the global minimum. We explore approaches to map the uncertainty in our observations and link it back to the background error covariance matrix utilized by the adjoint minimization.
# TABLE OF CONTENTS

LIST OF FIGURES ........................................................................................................ iv

LIST OF TABLES ........................................................................................................... ix

ACKNOWLEDGEMENTS .............................................................................................. x

Chapter 1  Introduction ............................................................................................... 1

Chapter 2  Urban Transport and Dispersion Model Sensitivity to Wind Direction Errors and Source Location ......................................................................................................................... 5

2.1 Motivation .............................................................................................................. 6

2.2 Methods .................................................................................................................. 8
  2.2.1 Domain Characteristics ................................................................................. 8
  2.2.2 Source Placement .......................................................................................... 12
  2.2.3 Source Characteristics ................................................................................... 13
  2.2.4 Uncertainty Calculation Procedure ................................................................. 15

2.3 Results and Discussion .......................................................................................... 18
  2.3.1 Impact of fidelity/complexity/symmetry of the urban topography ............... 19
  2.3.2 Impact of the source location relative to the downwind building ............... 27
  2.3.3 Results that depart from the generalized wind-error sensitivity behaviors .... 31

2.4 Conclusions ............................................................................................................ 34

Chapter 3 A Genetic Algorithm Approach to Source Term Estimation: Performance Enhancements through Sensitivity Studies .......................................................................................... 36

3.1 Motivation .............................................................................................................. 37

3.2 Algorithm Refinements ......................................................................................... 40
  3.2.1 Thresholds ...................................................................................................... 42
  3.2.2 Overview of Statistical Method ....................................................................... 47
  3.2.3 Cost Function .................................................................................................. 48
  3.2.4 Results of Sensitivities ................................................................................... 49

3.3 Sensitivity to Limited Sensor Data & Wind Data .................................................. 51

3.4 Conclusions ............................................................................................................ 56

3.5 Appendices ............................................................................................................ 57
  3.5.A Appendix A ....................................................................................................... 57
  3.5.B Appendix B ....................................................................................................... 58

Chapter 4 The Use of Ensemble-derived Background Error Covariance Matrices to Improve Gradient Descent Performance in Source Term Estimation ........................................ 61

4.1 Motivation .............................................................................................................. 62

4.2 Cost Function and Background Error Covariance Matrix .................................... 66
4.3 Methodology .................................................................................................................. 70
  4.3.1 Physics-Based Method ............................................................................................... 71
  4.3.2 Ensemble-Based Method ........................................................................................... 74
4.4 Data .................................................................................................................................. 76
4.5 Results ............................................................................................................................. 77
  4.5.1 Physics-Based Approach ......................................................................................... 77
  4.5.2 Ensemble-Based Approach ...................................................................................... 83
  4.5.3 Combination Method ............................................................................................... 89
4.6 Conclusions ..................................................................................................................... 91
4.7 Appendix ......................................................................................................................... 93

Chapter 5  Conclusions ....................................................................................................... 99

References ........................................................................................................................... 101
LIST OF FIGURES

Figure 2-1: Modern city grid domain with dimensions of $4000 \times 4000 \times 100$ m$^3$ ..................9

Figure 2-2: Colonial Era city domain with dimensions of $4000 \times 4000 \times 100$ m$^3$. ...............9

Figure 2-3: Modern city, grid design with mixed building heights used in this study and a zoomed view of the downtown area. .................................................................11

Figure 2-4: Colonial city, hub spoke design with mixed building heights used in this study and a zoomed view of the downtown area. .................................................................11

Figure 2-5: Urban topography from parts of Denver, CO. The area shown was selected so that the domain contained mixed building heights with a grid design. .........................12

Figure 2-6: Urban topography from parts of Washington, DC. The area shown was selected so that the domain contained mixed building heights with both grid and hub-and-spoke urban designs. .................................................................12

Figure 2-7: Example source locations for the generalized Modern city domain of uniform building heights. .............................................................................................................13

Figure 2-8: The schematic of methodology depicted here first determines the intersection of the plume generated by the true wind direction and the plume generated by the error wind. Then the intersection is normalized by the plume generated by the true wind direction as described in Chapter 2 section 2.3. ..............................................16

Figure 2-9: Example of Methodology calculations, the FOO quantifies how well the plume generated from error winds (Error Plume) overlaps with the plume generated from the true wind (Truth Plume). .................................................................17

Figure 2-10: The Fraction Of Overlap (FOO) for the non-urban domain with a function fit that shows a spread of 4.76. ....................................................................................18

Figure 2-11: Plan view of smoke dispersal through an array of staggered cubes (left) and unobstructed fetch (right). Taken from Brown (2004) but originally published in Davidson et al. (1995). .................................................................22

Figure 2-12: Schematic of two plume dispersion patterns expected using small wind direction variations. .............................................................................................................22

Figure 2-13: Fraction Of Overlap (FOO) statistic and spread value when using a wind direction that transports a plume down an urban canyon in a Modern (O) and the Colonial (•) domain for (a.) equal block heights, (b.) varying building heights, and (c.) the case studies. The Non-Urban domain (dotted line) is also shown for comparison.................................................................23
Figure 2-14: Fraction Of Overlap (FOO) statistic and spread value when using a wind direction that advects a plume towards a building obstruction in a Modern (○) and the Colonial (●) domain for (a.) equal block heights, (b.) varying building heights, and (c.) the case studies. The Non-Urban domain (dotted line) is also shown for comparison. .................................................................25

Figure 2-15: Plume function spread for modern, colonial, and case studies for urban canyon and building obstruction scenarios. .................................................................26

Figure 2-16: Fraction Of Overlap (FOO) statistic and spread value when using a wind direction that advects a plume towards an urban canyon (a.), building obstruction (b.), and a building corner (c.) in a Modern (○) and the Colonial (●) domain with equal block heights. The Non-Urban domain (dotted line) is also shown for comparison. .................................................................29

Figure 2-17: Fraction Of Overlap (FOO) statistic and spread value when using a wind direction that advects a plume towards an urban canyon (a.), building obstruction (b.), and a building corner (c.) in a Modern (○) and the Colonial (●) domain with varying building heights. The Non-Urban domain (dotted line) is also shown for comparison. .................................................................30

Figure 2-18: Spread of plume when transported toward modern/colonial blocks and buildings for building obstruction, urban canyon, and building corner scenarios. .........31

Figure 2-19: Fraction Of Overlap (FOO) statistic and spread value when using a wind direction that advects the plume towards a building corner for a Modern (○) and the Colonial (●) domain for (a.) equal block heights and (b.) varying building heights. The Non-Urban domain (dotted line) is also shown for comparison. .....................33

Figure 2-20: Fraction Of Overlap (FOO) statistic and spread value when using a wind direction that advects a plume towards an alternate type of urban canyon in a Modern (○) and the Colonial (●) domain for (a.) varying building heights and (c.) the case studies. The Non-Urban domain (dotted line) is also shown for comparison. .......34

Figure 3-1: The GA begins with a random population of “guesses” for the source variables. The resulting concentration field is then compared to our “true” observations by means of a cost function. .................................................................42

Figure 3-2: Maximum concentration time series of Trial 15, shown is the time threshold ($10^{-1}$ of the maximum concentration detected) and the concentration threshold ($10^{-3}$ of the maximum concentration detected). .................................................................43

Figure 3-3: Maximum concentration time series of Trial 30, shown is the time threshold ($10^{-1}$ of the maximum concentration detected) and the concentration threshold ($10^{-3}$ of the maximum concentration detected). .................................................................44

Figure 3-4: Maximum concentration time series of Trial 54, shown is the time threshold ($10^{-1}$ of the maximum concentration detected) and the concentration threshold ($10^{-3}$ of the maximum concentration detected). .................................................................44
Figure 3-5: On the top is the concentration field for Trial 15 data with loose (a) and strict (b) thresholds, in the middle is the concentration field for Trial 30 data with loose (c) and strict (d) thresholds, and on the bottom is the concentration field for Trial 54 data with loose (e) and strict (f) thresholds. The concentration data is in kg/s, however we have taken the logarithm.

Figure 3-6: We run the GA method 100 times. Then we take the median of a bootstrap sample of 10, 1000 times.

Figure 3-7: With the fixed wind direction and location we now find the emission rate with a linear cost function.

Figure 3-8: Absolute error distance for Trial 15, 30, and 54 when using data limitation. The mean wind column shows the absolute error distance with data limitation when using the observed mean wind as the wind in the AT&D model. The GA-Var estimated wind column shows the absolute error in distance with data limitation when using wind estimated from the GA-Var method.

Figure 3-9: Absolute error emission rate for Trial 15, 30, and 54 when using data limitation. The mean wind column shows the absolute error emission rate with data limitation when using the observed mean wind as the wind in the AT&D model. The GA-Var estimated wind column shows the absolute error in emission rate with data limitation when using wind estimated from the GA-Var method.

Figure 3-10: Absolute errors of distance and emission rate for Trial 30 using stable stability. The mean wind column shows the absolute errors with data limitation when using the observed mean wind as the wind in the AT&D model. The GA-Var estimated wind column shows absolute errors with data limitation when using wind estimated from the GA-Var method.

Figure 4-1: Schematic of the National Center for Atmospheric Research (NCAR) – Sage Management Source Term Estimation (STE) algorithm, known as the Variational Iterative Refinement STE algorithm (VIRSA), which consists of a combination of modeling systems. VIRSA consists of the source type preprocessor, the Second-order Closure Integrated PUFF model (SCIPUFF) and SCIPUFF’s STE model (Step 1), a hybrid Lagrangian-Eulerian Plume Model (LEPM) and the LEPM’s formal adjoint (Step 2), and the Hazard Assessment (Step 3) (Bieringer et. al. 2010).

Figure 4-2: Gradient descent minimization depiction, where the variational data assimilation technique to iteratively refine the first guess solution is shown.

Figure 4-3: Schematic of a modeling a nonlinear process, represented by the gray curve. We can develop a tangent linear model (TLM) by linearizing the nonlinear model. The adjoint of that TLM gives us a mathematical means to go from an end state to an initial state.
Figure 4-4: Schematic of cost via gradient descent minimization to create the cost surface, where the lowest (best) cost value for each point in the domain while minimizing for the other variables of interest is shown. ..................................................68

Figure 4-5: Universe of STE variables of interest in the plume reference frame, and how their uncertainties are related to each other. .................................................................71

Figure 4-6: A 2-dimensional visualization of the cost function surface. The general shape shows how wind variability can translate into uncertainty in location.................................72

Figure 4-7: Schematic of standard forward numerical weather prediction ensemble method. .......................................................................................................................75

Figure 4-8: A “reverse” standard ensemble approach where we take uncertainty in the concentration field and via the adjoint and gradient descent minimization find the uncertainty in our initial conditions .................................................................75

Figure 4-9: Methodology of ensemble-based approach to determine the background error covariance matrix. ........................................................................................................76

Figure 4-10: The wind direction time series of FFT07 Trial 54 for three sonic anemometers and 40 portable weather information displays (PWIDs). .........................79

Figure 4-11: The wind speed time series of FFT07 Trial 54 for three sonic anemometers and 40 portable weather information displays (PWIDs). ........................................79

Figure 4-12: The wind direction time series of FFT07 Trial 54 for the central sonic anemometer. ......................................................................................................................80

Figure 4-13: The wind speed time series of FFT07 Trial 54 for the central sonic anemometer. ......................................................................................................................80

Figure 4-14: Power spectral density for the wind direction, shown is the calculated cutoff frequency and the -5/3 slope that shows we are in the inertial subrange ..................81

Figure 4-15: Power spectral density for the wind speed, shown is the calculated cutoff frequency and the -5/3 slope that shows we are in the inertial subrange ..................81

Figure 4-16: Time series of detrended wind direction and filtered wind direction for FFT07 Trial 54 ........................................................................................................82

Figure 4-17: Time series of detrended wind speed and filtered wind speed for FFT07 Trial 54 ........................................................................................................82

Figure 4-18: Physics-based method results: (a) is an unbounded cost function solution, (b) shows the bounded solution, and (c) shows the scaled solution ..................83
Figure 4-19: The cost surface on a linear scale for each pair of variables that we minimized i.e., Longitude vs. Latitude, Wind Direction vs. Latitude, and Longitude vs. Wind Direction, using the constrained $E_B$ and the unconstrained $E_B$. .................85

Figure 4-20: The cost surface on a logarithmic scale for each pair of variables that we minimized i.e., Longitude vs. Latitude, Wind Direction vs. Latitude, and Longitude vs. Wind Direction, using the constrained $E_B$ and the unconstrained $E_B$. .................86

Figure 4-21: The cost surface on a linear scale starting from a different first guess, for each pair of variables that we minimized i.e., Longitude vs. Latitude, Wind Direction vs. Latitude, and Longitude vs. Wind Direction, using the constrained $E_B$ and the unconstrained $E_B$. ..........................................................87

Figure 4-22: The cost surface on a logarithmic scale starting from a different first guess, for each pair of variables that we minimized i.e., Longitude vs. Latitude, Wind Direction vs. Latitude, and Longitude vs. Wind Direction, using the constrained $E_B$ and the unconstrained $E_B$. ..........................................................88

Figure 4-23: Results with 100 random potential first guesses in the area bound by the physics-based method and evaluated the second component of VIRSA with the unconstrained $E_B$ (a) and the constrained $E_B$ (b). Results with 100 random potential first guesses in a 5km X 5km domain and evaluated the second component of VIRSA with the unconstrained $E_B$ (c) and the constrained $E_B$ (d). .........................90

Figure 4-24: Starting with a different first guess results with 100 random potential first guesses in the area bound by the physics-based method and evaluated the second component of VIRSA with the unconstrained $E_B$ (a) and the constrained $E_B$ (b). Results with 100 random potential first guesses in a 5km X 5km domain and evaluated the second component of VIRSA with the unconstrained $E_B$ (c) and the constrained $E_B$ (d). ..............................................................................91
LIST OF TABLES

Table 2-1: Anthrax characteristics from the Lawrence Berkeley National Laboratory database of physical, chemical, and toxicological properties of chemical and biological (CB) warfare agents for modeling airborne dispersion in and around buildings. .................................................................14

Table 2-2: LCt values for Anthrax using a probit slope, a spore ratio of $3 \times 10^7$ spores/mg, and a light breathing rate of 0.02 m$^3$/mg. These values are calculated using the Lawrence Berkeley National Laboratory database of physical, chemical and toxicological properties of chemical and biological (CB) warfare agents for modeling airborne dispersion in and around buildings. .........................................................15

Table 2-3: The spread values normalized by the non-urban spread (4.76) that were fit to each FOO scenario..............................................................................................................................................24

Table 3-1: Absolute error of the predicted parameters when using the Gaussian AT&D model, strict and loose thresholds for Trials 15, 30, and 54.................................................................49

Table 3-2: Absolute error of the predicted parameters when using the SCIPUFF AT&D model, strict and loose thresholds for Trials 15, 30, and 54.................................................................50

Table 3-3: Mean and standard deviation of error distance when using the mean observed wind. ..................................................................................................................................................59

Table 3-4: Mean and standard deviation of error distance when using the GA-Var estimated wind. .............................................................................................................................................60
ACKNOWLEDGEMENTS

This work was supported by the Defense Threat Reduction Agency under Grant 01-03-D-0010-0012, the Student Research Associate Program (SRAP) funded by DTRA, by the Bunton–Waller Fellowship, the Sloan scholarship, the National Center for Atmospheric Research, and part of this work was performed under the auspices of the Significant Opportunities in Atmospheric Research and Science Program (SOARS). SOARS is managed by the University Corporation for Atmospheric Research and is funded by the National Science Foundation, the National Oceanic and Atmospheric Administration, the Cooperative Institute for Research in Environmental Science, and by the Center for Multiscale Modeling of Atmospheric Processes.

I would like to thank Dr. Andrew Annunzio, Dr. John Hannan, Jonathan Hurst, George Bieberbach, and Dr. Francois Vandenburge for helpful discussions. I would like to thank Dr. George Young, Dr. Sue Ellen Haupt, and Dr. Paul Bieringer for their guidance in this Ph.D. journey, their patience, and mentorship. I would like to thank Dr. Jenni Evans and Dr. Jeffrey Brownson for their time and helpful reviews.

I would also like to thank my support network, my family and friends who in one way or another have helped make this solitary journey feel like we have accomplished it together. I would like to thank Steven Landis, my loving husband. Finally, this would not have been possible without the unconditional love from my 4 lb. Chihuahua, Nova.
Chapter 1

Introduction

To accurately describe atmospheric flow phenomena we use the governing equations; i.e. the conservation of mass, the conservation of momentum, the thermodynamic equation and the equation of state. These governing equations are used to create models by applying simplifying assumptions, boundary, and initial conditions to model the phenomena of choice. Even after these simplifying assumptions are made the atmospheric flow physics typically include nonlinear terms in the model equations. Such nonlinear systems, if they include dissipation, produce “dynamical chaos”. Lorenz (1963, 1996) describes dynamical chaos as a system in which, even if two realizations start from slightly different initial conditions, they will diverge with time; i.e. the realization depends sensitively on the initial conditions of the system.

Wilks (2005) describes the atmosphere as never completely observable; therefore, a mathematical model will never match the “real” atmosphere. Thus, if we use a chaotic dynamical system to describe our phenomena of interest, and recognize the fact that we can never measure the atmosphere precisely, it is clear that there will always be an uncertainty associated with both the measured and predicted atmospheric state. This work seeks to understand both the uncertainty introduced by measurements and the uncertainty introduced by approximating the model’s nonlinear terms.

Marzban (2009) discusses a simple example of the uncertainty introduced by measurement. He states that even if a quantity were to be measured several times in the same way and in the same circumstances, a different measured value is likely to be obtained each time. Assuming that the measuring system has sufficient resolution to distinguish between the values, this uncertainty exists even when measuring something as simple as the length of a table. All
measurements are subject to uncertainty and on top of that atmospheric models also add uncertainty. Examples of measurement uncertainty are digital meteorological and concentration sensors, which need to be adjusted to detect the signal in the presence of noise when used to measure winds and concentrations, respectively.

As mentioned earlier, to model the atmosphere we use the governing equations and approximate the physics of those equations to focus on the phenomena of interest. Part of approximating the physics includes removing unimportant terms and approximating those terms that we choose not to explicitly model. When we choose to approximate terms in the governing equations we introduce further uncertainty. After we approximate the physics we have two options when dealing with the nonlinear terms, we can take an analytical approach or a numerical approach. Many times an analytic approach is not feasible because the equations cannot be solved using known techniques so we are forced to either simplify the physics further still to achieve more tractable forms, or parameterize terms in the equations. Simplifying the physics or parameterizing introduces yet more uncertainty into our solutions. Therefore more often than not, instead of solving for an analytic solution we solve the models numerically by discretizing the model equations in space and time. Discretizing the equations introduces uncertainty as well because we cannot solve all ranges of spatial and temporal scales. This shortcoming introduces both approximation errors in derivatives and the need to parameterize only important processes occurring at unresolvable scales. The work here seeks to understand these sources of uncertainty and to exploit that understanding throughout the scientific process.

We seek to examine uncertainty in the context of Atmospheric Transport and Dispersion (AT&D). Uncertainty plays a key role in AT&D modeling: for example, uncertainty in either input variables such as wind direction, source location (x,y), or contaminant mass will produce uncertainties in predictions of downwind concentrations. To prevent hazardous outcomes in the event of an accidental contaminant release, it becomes important to estimate the source
characteristics (location, wind direction, and mass) from remote measurements of the concentration field. We will use Source Term Estimation (STE) techniques to examine how quantification of uncertainty in measurements, such as the wind direction, can be used to help bound the problem. STE is a difficult problem because, with irregular atmospheric fluctuations due to the stochastic nature of atmospheric turbulence, it is difficult to retrieve useful information from instantaneous concentrations.

The AT&D models used for air-quality and defense applications require input information describing the source parameters and require wind and source information to forecast concentration and dosage fields. Therefore, the models use observational data or mesoscale model-generated forecast winds. First we will examine how errors in these input wind fields may translate into AT&D model solution errors. Chapter 2 focuses on street-level plume errors that occur in building-aware AT&D models for a set of hazardous scenarios where the release location varies relative to the building locations and city building configurations.

Chapter 3 focuses on STE variables and how we can retrieve them utilizing a Genetic Algorithm (GA). One method of estimating source characteristics is coupling a GA with an AT&D model. Allen (2006, 2007), Haupt (2005, 2006, 2007), Long (2010), and Rodriguez (2011) use genetic algorithms to optimize source characteristics and several of these studies examine the effect of adding noise to the data to simulate errors in the sensor data, input parameters, and the effects of atmospheric turbulence. The algorithm used in this section, the GA coupled AT&D retrieval technique, has been previously successful at estimating source characteristics and meteorological variables necessary to predict the AT&D of a contaminant. This section describes via sensitivity analyses how we have tuned the GA, processed the data to optimize the results, and quantified the uncertainty of this method when using real data. The sensitivity analyses were performed to achieve five goals: (1) establish adequate thresholds to filter out noise in our concentration data without decimating the signal; (2) use a robust statistical
method to quantify the uncertainty in our predictions; (3) determine the best cost function for each of the variables we seek to retrieve; (4) given that real-time wind direction data are difficult to come by, determine if the GA predicted wind direction is representative of the advecting wind; (5) determine the robustness of the GA when limited sensors are available. The AT&D models used in this current study are a Gaussian Plume model and the Second-order Closure Integrated PUFF model (SCIPUFF).

Chapter 4 of this work focuses on the ambiguity between STE variables and more importantly, the relationships between their uncertainties. An extremely simple example of this ambiguity including only two variables is how wind direction error can propagate into a source location error, specifically in the crosswind location. The work in this section uses a Variational Iterative Refinement STE Algorithm (VIRSA) that utilizes variational data assimilation techniques to fuse Chem-Bio and meteorological observations to characterize agent release source parameters and provide a refined hazard assessment. The underlying algorithm consists of a combination of modeling systems, including SCIPUFF, SCIPUFF’s STE model, a hybrid Lagrangian-Eulerian Plume Model (LEPM), and the LEPM’s formal adjoint. SCIPUFF and its STE model are used to calculate a source estimate “first guess”. The LEPM and corresponding adjoint are then used to iteratively refine this source estimate using variational data assimilation techniques. In the work presented here we incorporate new methods to address issues related to uncertainty and using what we know about that uncertainty to reduce the tendency to find local minima rather than the global minimum. We explore approaches to map the uncertainty in our observations and link it back to the background error covariance matrix utilized by the adjoint minimization.
Chapter 2

Urban Transport and Dispersion Model Sensitivity to Wind Direction Errors and Source Location

The transport and dispersion (T&D) models used for air-quality and defense applications require information describing the source parameters and meteorological conditions to forecast concentration and dosage fields. In many cases the source parameters are known and the meteorological conditions are based on observational data or mesoscale-model-generated forecast conditions. This research examines how errors in the input wind fields translate into uncertainty in the contaminant concentration predictions. In particular, this study focuses on street-level errors in the dispersion patterns that occur in “building aware” T&D models that are sensitive to urban designs (e.g. road and building patterns) and release locations relative to the buildings. This problem was evaluated by first creating a “truth” plume for a given release location and wind direction. Then the T&D model uncertainty associated with input wind errors were determined by comparing plumes calculated using wind directions varied at 2-degree increments to the truth plume. The uncertainty is quantified as fraction of overlap (FOO). The results are evaluated in a control simulation with no buildings, and in two commonly observed city designs (e.g. a regular grid, and hub and spoke configuration). The analysis examines both idealized building configurations along with the urban topography from cities that represent the regular grid and hub and spoke city designs. Results show that the relative impact of the uncertainty in the meteorological conditions and the corresponding sensitivity of the model to variations in the wind direction vary significantly with the release location and city designs. This suggests that some source locations are less (more) sensitive to uncertainty in meteorological conditions and that this
information can be factored into the confidence that is placed in emergency response decisions based on this information.

2.1 Motivation

Since September 11, 2001, the United States (US) government has made significant investments in sensing and modeling technologies designed to protect the US armed services and homeland against the threats posed by weapons of mass destruction (WMD). These technologies include the development of fast response transport and dispersion (T&D) models that can account for the dispersion of chemical and biological (CB) agents released in urban areas. The need for accurate atmospheric T&D forecasting techniques has become increasingly important because of the threat of an intentional release of hazardous material into the atmosphere, particularly in areas of complex local surface forcing and for longer transport distances (Rife, 2004). Although extensive investments have been made to improve the accuracy of these T&D models in urban settings, the accuracy of the solutions are still highly dependent upon the meteorological conditions used. Chang (2003) determined that in cases where meteorological models were coupled with T&D models, the T&D models were strongly influenced by the diagnostic wind model that was used to generate gridded wind fields from observed winds. Brown (2007) showed that the sensitivities of plume transport in cities to wind direction, including how the street-level flow patterns in cities can be very robust (i.e., unchanging) as the upper-level wind direction changes, and then suddenly shift 180 degrees at critical upper-level wind directions. The research presented here enhances these findings by quantitatively characterizing how errors in the input wind direction translates into street level T&D uncertainty, specifically, into uncertainty in downwind hazard zones.
The urban settings evaluated in this study are designed after two commonly occurring city design characteristics. The first is based on a rectangular grid design and the other is a mix of rectangular grids and a hub-spoke or web-like design. Due to the historic significance, related to its use in the USA and the ready availability of lethal dosage (LD) values, this study uses anthrax and its corresponding lethal exposure concentrations values to define hazard thresholds. Detailed information on the domain characteristics, source placement, and source characteristics can be found in Chapter 2 sections 2.2.1, 2.2.2, and 2.2.3 respectively.

This study uses the Röckle (1990) based Quick Urban Industrial Complex (QUIC) Dispersion Modeling System developed at the Los Alamos National Laboratory (LANL) to evaluate the dispersion pattern variability associated with wind direction errors. QUIC is a fast-response, urban dispersion modeling system capable of computing three-dimensional wind patterns and dispersion of airborne contaminants around clusters of buildings. The system used is comprised of a wind model (QUIC-URB), a Lagrangian dispersion model (QUIC-PLUME), and a graphical user interface (GUI), QUIC-GUI (LANL, 2007). Hanna (2009) determined that the performance of QUIC-URB/PLUME from LANL was comparable to the performance of other Röckle based models like MicroSWIFT/SPRAY (MSS) from the Science Applications International Corporation (SAIC) & ARIA Technologies, Three-Dimensional Wind Field Model (3DWF) from the Army Research Laboratory (ARL), and the Israel Institute for Biological Research (IIBR) Kaplan and Dinar Model. The QUIC URB/PLUME modeling system is used because its performance was representative of the Computational Fluid Dynamics (CFD) models used when evaluated with the New York City (NYC) Midtown tracer measurements (Allwine et. al., 2008) and it is representative of the Röckle class building-aware models and relevant for urban T&D applications like those developed for the Pentagon and surrounding facilities (Warner and Coauthors, 2007).
2.2 Methods

2.2.1 Domain Characteristics

To evaluate the effect of uncertainty in T&D solutions associated with the input winds, this study uses domains that vary from a simple non-urban domain that serves as a control simulation to the complex urban building configurations. For the complex city designs, this study handles both generalized urban environments and actual building locations and heights from central Denver, CO and Washington, DC. The generalized urban designs diagnose the response of the T&D model to wind direction errors relative to the non-urban control simulation, and are used to infer the models response to a basic category of urban building/road network design. The generalized urban designs are chosen because they are representative of two types of city designs found in North America, a modern city grid design (Figure 2-1) and the colonial era, hub-and-spoke design (Figure 2-2). The red outlines in both Figure 2-1 and 2-2 represent individual city blocks where the buildings are 30 m high and cover a spatial footprint of 100 x 100 m. Houston, TX, Portland, OR, and Sacramento, CA are examples of cities with a grid-based building and road structure with city blocks on the order of 100 x 100 m. The generalized urban design illustrated in Figure 2-2 was developed to emulate a common characteristic of a hub and spoke urban design where several major streets converge at a city center or square and are narrower in comparison to a modern city street. This design is common in colonial era cities of the northeastern United States like Boston, MA, Philadelphia, PA, and Washington DC. The city blocks in the colonial era design have a spatial footprint of 100 x 100 m and have a building height of 30 m.
Figure 2-1: Modern city grid domain with dimensions of $4000 \times 4000 \times 100$ m$^3$.

Figure 2-2: Colonial Era city domain with dimensions of $4000 \times 4000 \times 100$ m$^3$. 
The sensitivity of the T&D solution to wind uncertainties in urban environments with mixed building heights is also examined. Many urban centers have a mixture of building heights with tall commercial buildings at the core of the city surrounded by a ring of shorter multi-story buildings and an outer ring of single story buildings. Two generalized urban domains, which incorporate the effects of mixed building heights along with grid vs. hub and spoke urban design, were created for this study. Figure 2-3 illustrates the modern city, grid design and Figure 2-4 illustrates the colonial era urban design with mixed building heights used in this study. The building heights and road networks of the domain in Figure 2-3 and 2-4 respectively were inspired by the buildings from the central business districts of Denver, CO and Boston, MA. The results of the sensitivity studies using the generalized urban domains described above are contrasted with results from a comparable experiment using the actual urban topography from parts of Denver, CO and Washington, DC. The areas of the cities used were selected so that the domains contain mixed building heights with both grid and hub and spoke urban designs. Figures 2-5 and 2-6 illustrate the buildings used in the T&D sensitivity simulations for these two cities. It is important to note that in these case studies we evaluated one area in each domain to demonstrate how a plume in different urban environments, in a general sense, are affected by errors in the input winds.
Figure 2-3: Modern city, grid design with mixed building heights used in this study and a zoomed view of the downtown area.

Figure 2-4: Colonial city, hub spoke design with mixed building heights used in this study and a zoomed view of the downtown area.
Figure 2-5: Urban topography from parts of Denver, CO. The area shown was selected so that the domain contained mixed building heights with a grid design.

Figure 2-6: Urban topography from parts of Washington, DC. The area shown was selected so that the domain contained mixed building heights with both grid and hub-and-spoke urban designs.

2.2.2 Source Placement

The impact of wind uncertainties on an urban T&D solution is also affected by the location of the release location relative to a downwind and upwind building obstruction. To examine this effect, this study uses sources located in three locations: (1) in the middle of the street (or urban canyon), (2) directly upwind of a building obstruction, and (3) upwind of the corner of the building obstruction. Examples illustrating these source locations for the generalized Modern city domain of uniform building heights are shown in Figure 2-7.
2.2.3 Source Characteristics

To define a hazard area that could be used in the sensitivity analysis it was necessary to set a threshold that would define the extent of the hazard area from the T&D simulation. Realism was added to the experiment by choosing anthrax as the material being released; its human response characteristics determine the downwind hazard area. The anthrax characteristics (Table 2-1) and the lethal concentration toxicity (LC50) calculations are derived from the Lawrence Berkeley National Laboratory database of physical, chemical, and toxicological properties of chemical and biological (CB) warfare agents for modeling airborne dispersion in and around buildings (Thatcher et. al., 2000).
Table 2-1: Anthrax characteristics from the Lawrence Berkeley National Laboratory database of physical, chemical, and toxicological properties of chemical and biological (CB) warfare agents for modeling airborne dispersion in and around buildings.

<table>
<thead>
<tr>
<th>BIOLOGICAL CLASS</th>
<th>SPORE FORMING</th>
<th>PERSISTENCE</th>
<th>SIZE (µm)</th>
<th>SHAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacillus anthracis</td>
<td>2 hours</td>
<td>Years</td>
<td>~1 diameter × ~1.5 length</td>
<td>Rod</td>
</tr>
<tr>
<td>DISSEMINATION/ ROUTE OF ENTRY</td>
<td>INCUBATION/ ONSET</td>
<td>CONTAGIOUS</td>
<td>50% INFECTIVE DOSE (ORGANISMS/PERSOM)</td>
<td>UNTREATED LETHALITY (%)</td>
</tr>
<tr>
<td>Spore inhalation, ingestion (rare), broken skin</td>
<td>1-2 hrs, 1-7 days</td>
<td>No</td>
<td>8,000 to 20,000</td>
<td>100</td>
</tr>
</tbody>
</table>

To quantify anthrax exposure mortality we use different dosage thresholds to identify the hazard areas. These areas are defined within the 50 and 10 LCt (Table 2-2). The LCt is defined as

\[ LCt = \frac{LD}{SR \times BR} \quad (2-2a) \]

where: LD is the lethal dosage (spores), SR is the spore ratio (spores per mass of contaminant released), and BR is the breathing rate (m³/s). The thresholds represent the minimum value used to define a hazard zone with an anticipated level of health response within a given population. Any dosage above that value is considered hazardous and any dosage below that value is still hazardous but has a lower probability of lethality relative to this population health response. In this study we deem any dosage below that LCt value to be non-hazardous. Concentration scales linearly, only results that used an LCt10 threshold (1.2 × 10⁻⁵ g s/m³) of anthrax are shown.
Table 2-2: LCt values for Anthrax using a probit slope, a spore ratio of $3 \times 10^7$ spores/mg, and a light breathing rate of 0.02 m$^3$/mg. These values are calculated using the Lawrence Berkeley National Laboratory database of physical, chemical and toxicological properties of chemical and biological (CB) warfare agents for modeling airborne dispersion in and around buildings.

<table>
<thead>
<tr>
<th>PERCENT</th>
<th>LCT VALUE (g s/m$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.00078</td>
</tr>
<tr>
<td>10</td>
<td>0.000012</td>
</tr>
</tbody>
</table>

2.2.4 Uncertainty Calculation Procedure

An identical twin setup using the QUIC system is applied to assess the sensitivity of the downwind hazard area from the T&D model to wind direction uncertainties. To assess this sensitivity we first create a “building aware” wind field using QUIC-URB. The prescribed “truth” wind on the upwind side of the release and buildings is set at 3 m/s at all levels in the model. QUIC-PLUME is then used to compute the dispersion of 10 kg of anthrax through the urban environment. For the purposes of this identical twin experiment, this plume is now considered as truth against which all of the others are compared. In this way, truth plumes are created for each of the urban configurations and release locations being examined. Next, plumes from the same release location are created with the QUIC system using winds that are varied to emulate wind direction uncertainties. These winds depart from the “true” wind direction value in 2° increments to a maximum of 40° in both the counter-clockwise and clockwise directions from the direction used to produce the truth plumes. The process is repeated for “true” wind directions in each urban design to evaluate the sensitivity in the hazard areas for scenarios where the source is located...
upwind of building obstructions, urban canyons, and building corners of the building/road network. The plume sensitivity to wind direction at 2 meters above ground level at LCl10 are then quantified using the fraction of overlap (FOO), figure of merit in space (FMS), measure of effectiveness (MOE), and normalized absolute difference (NAD) metrics (Warner et. al., 2003). The FMS, MOE, and NAD metrics have been tested and evaluated with success against field observations; however, in this study, given that what we wish to quantify is how well the modeled error plumes line up with respect to the model truth plume, the simple FOO suffices as a metric. The FOO quantifies how well the plume generated from error winds (Error Plume) overlaps with the plume generated from the true wind (Truth Plume). This is calculated by determining the intersection of the Truth Plume and Error Plume and normalizes by the Truth Plume, it is defined as

\[
FOO = \frac{\text{area(TruthPlume} \cap \text{ErrorPlume})}{\text{area(TruthPlume)}}
\]

(2-2b)

Thus, we will focus our results in terms of the FOO. Figure 2-8 is a schematic of this methodology using the FOO and Figure 2-9 shows an example that illustrates this approach using results for the modern grid domain from this study. When using the FOO metric, a higher value of FOO corresponds to a higher tolerance to wind direction uncertainty.

Figure 2-8: The schematic of methodology depicted here first determines the intersection of the plume generated by the true wind direction and the plume generated by the error wind. Then the
intersection is normalized by the plume generated by the true wind direction as described in Chapter 2 section 2.3.

![Diagram](image)

Figure 2-9: Example of Methodology calculations, the FOO quantifies how well the plume generated from error winds (Error Plume) overlaps with the plume generated from the true wind (Truth Plume).

Finally, given that we calculate the FOO at 2° increments to a maximum of 40° in both the counter-clockwise and clockwise directions from the direction used to produce the truth plume, we expect an exponential increase and then decrease in the FOO for a non-urban domain. We created a function that is fit to our non-urban FOO so that we can quantify the spread. This function is defined as

\[
Fit = \frac{A}{(2\pi)^{\frac{3}{2}}} \exp\left(\frac{-|\theta|}{2\sigma}\right),
\]  

(2-2c)

where \( \frac{A}{(2\pi)^{\frac{3}{2}}} \) is the amplitude, \( \theta \) are the wind angle increments, and \( \sigma \) is the spread. Then we fit this function to each FOO scenario and optimize the values of the A and \( \sigma \) to compare the
spread in each scenario to the spread of the non-urban FOO. For the purpose of this analysis the spread corresponds to the relative sensitivity of the hazard zone to uncertainty in wind direction. Larger spread values correspond with lower relative sensitivities to uncertainties in wind direction and smaller spread values correspond with higher relative sensitivities to uncertainties in wind direction. An example of the fit to the non-urban FOO is plotted in Figure 2-10.

Figure 2-10: The Fraction Of Overlap (FOO) for the non-urban domain with a function fit that shows a spread of 4.76.

2.3 Results and Discussion

The goal of this study is to characterize the sensitivity of urban dispersion simulations to uncertainties in the winds used to drive these simulations, and to identify scenarios where urban building and road configurations are more (less) susceptible to these uncertainties. The results provide a means to more accurately characterize simulation uncertainty that is critical to guiding emergency response decisions. The presentation of the results is partitioned to illustrate patterns
that correspond with: (2.3.1) the fidelity/complexity/symmetry of the urban building and road network, (2.3.2) the impact of the source location relative to the downwind building, and (2.3.3) how these results depart from the general behaviors discussed in 2.3.1 and 2.3.2, when the release location is moved from a major street intersection to a smaller street further from the urban center. Implications of model sensitivity are discussed for each result and are then summarized to provide generalized guidance that can be used by emergency response personnel that rely on these tools.

### 2.3.1 Impact of fidelity/complexity/symmetry of the urban topography

Simulation speed is critical for emergency response applications. In many cases simplified urban topographies are used to reduce the number of building elements, thereby reducing the amount of time required to compute a solution. While the reduction of the fidelity of the building data clearly results in a less accurate solution since detailed geospatial information is not available, the impact on the solution sensitivity to wind direction uncertainties is not as clear. Key questions addressed here are: (1) how does the tradeoff between the need for a faster simulation and simulation fidelity influence the dispersion solutions’ sensitivity to errors in the wind fields, and (2) does this vary for building and road configurations in cities with modern vs. colonial layouts. Numerous studies of contaminant dispersion in an urban environment have used uniform grids of building obstructions to characterize the properties of urban dispersion. Figure 2-11 is an image of a smoke plume from one such study conducted by Davidson et al. (1995). This figure contrasts smoke plume releases in an urban vs. non-urban domain and illustrates that the plume tends to disperse more in the urban environment of “staggered buildings”. This enhanced dispersion due to the building obstructions also tends to make the staggered scenario more tolerant to wind direction uncertainties because the downwind plume is much wider unlike in the
scenario with no “staggered buildings”. These qualitative findings were also documented by Brown (2007). In that study Brown demonstrated that downwind dispersion in urban environments can be highly variable given small errors or uncertainties in the wind direction. This behavior as depicted in Figure 2-12, results in situations where small variations (uncertainty in the wind direction) at times result in small variations in the downwind dispersion pattern while other situations with small wind direction uncertainties result in large differences to the downwind dispersion pattern. This property of reduced sensitivity of the dispersed solution to wind direction uncertainties for locations when building obstructions are present is also shown quantitatively in Figure 2-13. Here, all of the results for scenarios with buildings indicate a higher FOO score than a corresponding non-urban scenario. This figure also provides an example of how the dispersion solution sensitivity to wind direction uncertainty can vary for different urban topographies. In this example the contaminant release occurs in the center of an urban canyon and the winds are parallel to the street canyon. Figure 2-13a shows the FOO results for simulations where simplified urban terrain representing city blocks with a uniform building height are used. The FOO of the contaminant plumes for cities with both uniform grids and grids with large diagonal boulevards (common in the North America colonial era cities) have similar sensitivities, and are a factor of 2.78 and 2.35 respectively, larger than the spread when no buildings are present implying that the size/location of the plume footprint is less sensitive to wind direction. This difference in sensitivity is due to the presence of more narrow “roads” in the colonial domain. When the complexity of the urban topography is increased by adding buildings of differing heights, the dispersion solution becomes slightly less sensitive to the wind uncertainties than the more simplistic urban topography used in Figure 2-13a. This point is illustrated in Figure 2-13b where FOO results indicate that the simulations for both the modern grid and colonial city designs are a factor of 3.45 and 3.12 respectively, larger than the spread when no buildings are present implying that the size/location of the plume footprint is less sensitive to wind direction.
The addition of non-uniform buildings also results in differences in the error sensitivity between the positive and negative wind direction differences. Figure 2-13b which shows the modern grid indicates more tolerance to wind direction uncertainty because the larger buildings in the downtown area appear to channel the plume more efficiently than in a colonial grid smaller buildings and narrower streets. This Figure 2-13c shows the results for the comparable FOO analysis for the urban topographies of the cities of Denver, CO and Washington, DC where the urban canyon formed in each scenario are between buildings of similar heights. In this example the urban topography is substantially more complex than was used in the simulations shown in Figures 2-13a and 2-13b. Of the two “real-world” cities examined, Washington, DC has smaller but more buildings. The portion of Denver used in the analysis has larger and taller buildings. The inverse relationship between complexity of the urban topography and the sensitivity of the atmospheric dispersion solution to wind direction uncertainties is also evident in Figure 2-13c (Table 2-3 under Urban Canyon, Similar Building Height). The FOO values for both Denver and Washington, DC are larger than those seen in Figure 2-13b with Washington DC environment having the most complex urban topography and was a factor of 6.61 larger than the spread when no buildings are present implying that the size/location of the plume footprint is less sensitive to wind direction. A full list of all of the normalized spread values for each of the building scenarios can be found in Table 2-3.
Figure 2-11: Plan view of smoke dispersal through an array of staggered cubes (left) and unobstructed fetch (right). Taken from Brown (2004) but originally published in Davidson et al. (1995).

Figure 2-12: Schematic of two plume dispersion patterns expected using small wind direction variations.
Figure 2-13: Fraction Of Overlap (FOO) statistic and spread value when using a wind direction that transports a plume down an urban canyon in a Modern (O) and the Colonial (•) domain for (a.) equal block heights, (b.) varying building heights, and (c.) the case studies. The Non-Urban domain (dotted line) is also shown for comparison.
Table 2-3: The spread values normalized by the non-urban spread (4.76) that were fit to each FOO scenario.

| Source Location | Urban Canyon | | | | |
|-----------------|--------------|--------------------------|--------------------------|--------------------------|
|                 | Spread       | Building Corner          | Building Obstruction     | |
|                 | Urban Canyon | Similar Building Height | Varying Building Height | |
| Constant Building Height | Modern | 2.78 | - | 2.09 | 2.21 |
| Colonial | 2.35 | 2.48 | 2.14 | 2.20 |
| Varying Building Height | Modern | 3.45 | 2.77 | 4.74 | 3.86 |
| Colonial | 3.12 | 3.18 | 3.95 | 3.98 |
| Case Studies | Denver | 3.14 | 3.06 | - | 3.54 |
| DC | 6.61 | 3.41 | - | 6.50 |

Figures 2-14a, 2-14b, and 2-14c depict results for a comparable experiment to that shown in Figure 2-13 except that in this case the base wind direction and release location are confined so that a building obstructs the contaminant dispersion directly downwind of the release. In this case an inverse relationship between complexity of the urban topography and the sensitivity of the atmospheric dispersion solution to wind direction uncertainties is again evident. While the results between the modern and colonial era cities tend to be similar to each other for the less complex specifications of the buildings/roads, the FOO results shown in Figure 2-14c show a reversal of the results from 2-13c for errors less than 15° from the true wind. For errors larger than 15° from the true wind we see that Denver is less tolerant to errors because the plume does not interact with and is not channeled by as many buildings as for Washington D.C. This finding suggests that while the results presented here represent the general behavior of the model, there can be specific instances of building configurations where this behavior does not apply.
Figure 2-14: Fraction Of Overlap (FOO) statistic and spread value when using a wind direction that advects a plume towards a building obstruction in a Modern (○) and the Colonial (♦) domain for (a.) equal block heights, (b.) varying building heights, and (c.) the case studies. The Non-Urban domain (dotted line) is also shown for comparison.
This inverse relationship between the complexity of the urban topography and solution sensitivity to wind uncertainties can be more clearly seen in Figure 2-15, which shows the spread of the FOO results plot for the scenarios of increasing complexity in the building/street configurations. The abscissa in this plot ranges from low complexity of configuration, non-urban on the left to the larger complexity configuration, Washington DC results on the right. This figure also includes the results from Figures 2-13a/2-14a, 2-13b/2-14b, and the Denver and Washington DC FOO results from both sets of simulations. The results imply that cities with complex building topographies typically see a broader pattern of dispersion, and consequently, changes to the input winds have less of an impact on the solution. There is only a slight difference between the results when there was and was not a building obstruction downwind of the release location.

![Figure 2-15: Plume function spread for modern, colonial, and case studies for urban canyon and building obstruction scenarios.](image)

Figure 2-15: Plume function spread for modern, colonial, and case studies for urban canyon and building obstruction scenarios.
In summary, these results indicate that the urban dispersion model solutions for North American colonial era cities (common on the US East Coast) are less sensitive to wind direction uncertainties than the cities built more recently that typically follow a relatively uniform grid. This finding is due to the greater complexity in the urban building/road network. However, in order to fully capitalize on this behavior of the urban dispersion model, however, it is necessary to have sufficient fidelity in the urban building/road data sets used in the models.

2.3.2 Impact of the source location relative to the downwind building

The location of a contaminant source coupled with the wind direction and complexity of the urban environment results in a large number of possible release scenarios. This portion of the study attempts to consolidate these scenarios into three basic categories. The first is the situation where the release occurs in the street canyon and the nominal wind direction is such that the contaminant is carried downwind without any short-range building obstructions. The second is a situation where the release again occurs in the street canyon, but this time the nominal wind direction is such that the contaminant is carried downwind directly into the face of a nearby building. The third category of scenario examined is the situation where the contaminant release occurs in an intersection of streets and the nominal wind direction is such that the contaminant is carried downwind into the corner of the building. The question being addressed here is, does the wind direction and release location relative to down-wind building obstacles influence the solutions sensitivity to errors in the prevailing wind direction?

Figures 2-16 and 2-17 provide a comparison of FOO results that address this question and illustrate the impact of source location relative to downwind buildings for both the idealized modern and colonial city designs. The urban topography used in the analysis displayed in Figure 2-16 is the more simplistic representation of the buildings where the building obstacles represent
entire city blocks. The urban topography in Figure 2-17 represents a more complex representation of the buildings where individual buildings of varying height are prescribed. The FOO analysis results in both cases suggest that situations where the corner of the building obstructing is directly downwind of the release is the least sensitive to wind direction uncertainties. In Figures 2-16 and 2-17, the results are mixed for the situations where the release is carried downwind through the urban canyon and the face of a building is directly downwind from the release. In each of these scenarios the model solutions are more sensitive to the wind direction uncertainties than the building corner case but roughly comparable to each other and similar between modern and colonial city designs. In Figure 2-16b, both the Modern and Colonial domains have symmetric FOOs and in Figure 2-16c, they both become less symmetric. The release location being in a "main" avenue for the Colonial domain, better channeled for the corner release (Figure 2-16c) as well as, the release location being in a larger center plaza area. Figure 2-18 summarizes this finding by plotting the spread of the FOO results for the scenarios illustrated in Figures 2-16 and 2-17 where the release locations are varied relative to the downwind building obstructions. The abscissa in this plot ranges on the left from the scenario where there is not a building directly downwind of the release in the truth simulation, to a scenario where a building is directly downwind of the release, to a scenarios where the release occurs directly upwind of the corner of a building on the right.
Figure 2-16: Fraction Of Overlap (FOO) statistic and spread value when using a wind direction that advects a plume towards an urban canyon (a.), building obstruction (b.), and a building corner (c.) in a Modern (O) and the Colonial (*) domain with equal block heights. The Non-Urban domain (dotted line) is also shown for comparison.
Figure 2-17: Fraction Of Overlap (FOO) statistic and spread value when using a wind direction that advects a plume towards an urban canyon (a.), building obstruction (b.), and a building corner (c.) in a Modern (○) and the Colonial (●) domain with varying building heights. The Non-Urban domain (dotted line) is also shown for comparison.
2.3.3 Results that depart from the generalized wind-error sensitivity behaviors

Due to the complexity of many urban landscapes it is difficult to break a release scenario down to a single category. In reality, the real-world urban landscapes combined with the release scenario is likely to be a combination of several categories and their corresponding effect on the sensitivity of the solution to wind direction errors. An example of this response can be seen in Figure 2-19. This figure shows the FOO analysis results for scenarios where the contaminant flows toward a nearby building corner. Here the FOO results for the colonial city layout are skewed higher to the left of 0° and the modern city layout FOO results are skewed much higher to the right of 0°. This illustrates the channeling effects that large buildings near the source location can have on the dispersion patterns and corresponding sensitivities to wind direction.
uncertainties. A similar example can be seen in Figure 2-20, which illustrates the FOO analysis solutions for scenarios where the contaminant flows down urban canyons in Denver and Washington, DC without any direct obstruction from a nearby building, but here the canyon includes significantly varying building heights (Table 2-3 under Urban Canyon, Varying Building Heights). In Figure 2-20a we see that the modern city layout is more sensitive to wind direction uncertainties than the colonial city layout, while 2-20b shows the opposite result when a different baseline wind direction that carries the contaminant down alternative urban canyon is selected. In this situation there is little difference in the Denver results when the baseline wind direction is changed, but a larger change in the Washington DC results. This is largely due to size of the urban canyon. The baseline wind direction carries the contaminant from Dupont Circle down Massachusetts Avenue in Figure 2-20a, which is a much wider urban canyon than the results shown in Figure 2-20b where the baseline winds carry the contaminant from Dupont Circle down the narrower 19th Street.
Figure 2-19: Fraction Of Overlap (FOO) statistic and spread value when using a wind direction that advects the plume towards a building corner for a Modern (○) and the Colonial (●) domain for (a.) equal block heights and (b.) varying building heights. The Non-Urban domain (dotted line) is also shown for comparison.
2.4 Conclusions

Modeling contaminant dispersion in urban environments is a critical capability for the emergency response community. This community relies on the accuracy of the results from these tools to make decisions that impact the health and safety of both the general public and first responders during both intentional and accidental airborne releases of hazardous materials. In
spite of the fact that during these crisis situations, high quality weather information is sometimes not available, critical decisions still need to be made. In these situations, emergency response managers need to have a measure of confidence in the dispersion model results. This work addresses this need by characterizing the impact of wind direction uncertainties on the corresponding dispersion solutions.

The results of this study indicate that there are some general rules of thumb that can be applied to the problem of how sensitive the dispersion solution is to wind direction errors. First, the sensitivity of dispersion solutions to uncertainties in wind direction is inversely related to the complexity and dynamic variability in the urban topography. In cases where the complexity of the urban topography is artificially reduced by simplifying the building database to improve the speed of the solution, the resulting solution becomes more sensitive to wind direction uncertainty. Second, simulation results from releases that occur directly upwind of a corner of a building are less sensitive to wind direction uncertainties than those scenarios where there is no obstruction directly downwind or the face of a building is directly downwind. Third, the presence of large buildings near the release location can act as large barriers to the flow and corresponding material dispersion and strongly influence the sensitivity of the solutions to wind direction. Fourth, the presence or lack of major (wide) streets, which act as large urban canyons, can result in significant differences in the sensitivity of the model solutions to the wind uncertainty. Finally, this analysis indicates that there is a broad range of model sensitivities to wind direction uncertainty. While this may make it challenging to identify generalized solutions to the wind direction uncertainty problem affecting the emergency response community, it does mean that there are release scenarios and locations that are very tolerant to wind direction uncertainties. This finding suggests that for important locations it may be beneficial to pre-compute the wind direction error sensitivities so that they will be readily available should the need unfortunately arise.
Chapter 3

A Genetic Algorithm Approach to Source Term Estimation: Performance Enhancements through Sensitivity Studies

The nature of atmospheric turbulence makes it impossible to ensure that source term estimation algorithms will work in real-world conditions until they are tested against Atmospheric Transport and Dispersion (AT&D) field datasets. The FUsing Sensor Information from Observing Networks (FUSION) Field Trial (FFT07) was conducted to create a reliable dataset for such tests. The Genetic Algorithm (GA) method used in this study couples the AT&D model to the observations to achieve a Variational solution (GA-Var). The GA method has been used successfully in prior studies to estimate source characteristics and meteorological parameters necessary to predict the AT&D of a contaminant. Here the GA is coupled to the AT&D model with observations using a dual cost function approach, rather than the single cost function method because the emission rate is linear in the AT&D model. Using the GA-Var method we performed sensitivity analyses to achieve five goals: (1) establish adequate concentration thresholds to determine the duration of the contaminant passage and to filter out noise in our concentration data without decimating the signal; (2) use a robust statistical method to quantify the uncertainty in our source characteristic retrievals; (3) determine the best cost function for the variables we seek to retrieve; (4) given that representative real-time wind direction data are difficult to come by, determine if the GA retrieved wind direction is representative of the contaminant advecting wind; (5) determine the robustness of the GA when only a limited number of sensors are available. A statistical analysis using bootstrap sampling was used to provide quantitative uncertainty estimates for the cases examined. This current study quantifies the uncertainty of the GA coupled AT&D method using a subset of the FFT07 dataset, that is, those trials that include a single source only, for continuous releases.
3.1 Motivation

An estimate of the release source parameters based on available observations is often required to enact a response to an accidental or intentional release of Chemical, Biological, Nuclear, Radiological (CBNR). Many existing Atmospheric Transport and Dispersion (AT&D) datasets, however report only cumulative dosages, which make it impossible to retrieve information about the instantaneous concentrations given the stochastic nature of atmospheric turbulence. To ensure that Source Term Estimation (STE) algorithms will work in real world conditions they must be tested against AT&D concentration field datasets with sufficient spacial and temporal resolution to capture these turbulence driven fluctuations (Storwold 2007).

There has been extensive work on devising methods for back-calculating source characteristics and Rao (2007) reviews some of these methods, including adjoint and tangent linear models, Kalman filters, and variational data assimilation. Haupt (2008) compares and contrasts different formulations and outlines a general paradigm that encompasses how 26 groups fulfill the elements of the paradigm, including the inversion variables, an AT&D model, inversion technique, and test data that supply the sensor data. Bieringer (2011) discusses a combination of modeling systems, including the Second-Order Closure Integrated PUFF model (SCIPUFF), its corresponding STE model, a hybrid Lagrangian-Eulerian Plume Model (LEPM), its formal adjoint, and the software infrastructure necessary to link them. Shuford (2008) use a Bayesian framework within which hidden Markov models of release schedule serve to calculate the most probable state. Like Shuford (2008), Brown (2010) uses a Monte Carlo Bayesian data fusion algorithm. Cervone (2010) uses a method similar to a GA but with a non-Darwinian evolution approach to STE.

Another method for estimating source characteristics for an atmospheric contaminant release is the GA-Var technique. The GA coupled AT&D model iteratively compares predictions
with observations. Allen (2006, 2007), Haupt (2005, 2006, 2007), Long (2010), Rodriguez (2011), and Schemhl (2012) have used this GA based approach to optimize source characteristics. Several of these studies examine the effect of adding noise to the data to simulate errors in the sensor data, input parameters, and the effects of atmospheric turbulence. The STE algorithm used in this work is the GA-Var, which couples concentration observations with AT&D model forecasts. This method has been used successfully to estimate source characteristics and meteorological parameters necessary to predict the AT&D of a contaminant. The work presented in this paper seeks to describe via sensitivity analyses how we have tuned the GA and processed the data to optimize the results when using real data from FFT07. Using the GA-Var method we performed sensitivity analyses to achieve 5 goals: (1) establish adequate concentration thresholds to determine the duration of the contaminant passage and to filter out noise in our concentration data without decimating the signal; (2) use a robust statistical method to quantify the uncertainty in our source characteristic retrievals; (3) determine the best cost function for the variables we seek to retrieve; (4) given that representative real-time wind direction data are difficult to come by, determine if the GA retrieved wind direction is representative of the contaminant advecting wind; (5) determine the robustness of the GA when only a limited number of sensors are available.

The Defense Threat Reduction Agency (DTRA) sponsored the FUsing Sensor Information from Observing Networks (FUSION) Field Trial 2007 (FFT07) to collect data from an abundance of research-grade tracer, sensor, and meteorological instruments suitable for testing current and future CBNR STE algorithms (Storwold 2007). This dataset was created with both day and night releases in a highly instrumented desert test area at the U.S. Army’s Dugway Proving Ground. Part of the FFT07 data release plan was to make the data available in phases. The FFT07 data is partitioned into two datasets, Trials and Cases. The Trial dataset contains readings from 100 sensors, source information (location and amount), and abundant
meteorological information. These data were made available to test and train the current STE algorithms. The Case dataset contains 104 different release events with limited meteorological data, concentration data for only four or 16 sensors, and no information about the source location or release amount. Platt (2010) discusses the first phase of FFT07 estimations of the source location and release amount of the Case data.

The observations used here are from the FFT07 Trial Data. After examination we determined that Trials 15, 30, and 54 are representative of a single-location continuous release. Trials 15 and 54 were selected because they both were early morning releases. Trial 15 was specifically selected because of the “poor” quality of the data; the data show an accidental gas release in the lower south quadrant of the sensor grid and the upper right quadrant of the sensor grid was offline. Unlike Trial 15, Trial 54 had all but seven sensors online and the sensors offline were at the edges of the grid; thus, this Trial is considered to have “good” quality data. Trial 30 was selected because while the data are considered “good” quality, the release occurred at night in stable meteorological conditions. This set of three cases allows the examination of both data quality effects and those of assuming atmospheric neutrality under varied conditions of actual stability.

Algorithm refinements are discussed in Section 3.2. In Section 3.2.1 an analysis of the sensitivity of the STE results to input thresholds is undertaken. First we determine the duration of the contaminant passage and then strict and loose thresholds are applied to the concentration data to filter noise. Section 3.2.2 describes the bootstrap sampling statistical analysis done to compare the FFT07 data to the predictions from the GA coupled AT&D model; this is done to quantify the uncertainty in the parameters retrieved. Section 3.2.3 discusses the sensitivity to our cost function metric: we use a dual cost function approach, rather than a single cost function method used previously (Rodriguez 2010). Section 3.2.4 discusses the results of the application of 3.2.1, 3.2.2, and 3.2.3 to Trials 15, 30, and 54. Section 3.3 assesses the robustness of the algorithm. First, it
focuses on wind direction where temporally and spatially averaged wind fields were compared to the wind direction retrieved by the GA. Second, it discusses the robustness of the algorithm to the quantity of data. Finally in Section 3.4 we discuss the conclusions of this work.

3.2 Algorithm Refinements

The GA-Var coupled AT&D procedure for source term estimation has been proven to work with identical twin data in Allen et al. (2006, 2007), Haupt (2005), Haupt et. al. (2006) and Long et. al. (2010). As in the aforementioned studies, the GA begins with a set of potential solutions. In this study the potential solutions consist of wind direction, source location (x,y), and emission rate. Also, these solutions are bounded by a priori estimates of the feasible range. We now use the available meteorological data to limit the potential solution of wind direction to within 30 degrees of the mean wind direction. Each set of potential solutions is then fed into an AT&D model, as described below.

The AT&D models used in this current study are a Gaussian Plume model and SCIPUFF. The Gaussian Plume model is a simple model that assumes that the contaminant dispersion has a Gaussian distribution. It is a computationally efficient equation where the concentration of the contaminate is determined by

$$C = \frac{Q}{U} \frac{1}{2\pi \sigma_y \sigma_z} \exp \left( -\frac{(y - y_o)^2}{2\sigma_y^2} \right) \left[ \exp \left( -\frac{(z - z_o)^2}{2\sigma_z^2} \right) + \exp \left( -\frac{(z + z_o)^2}{2\sigma_z^2} \right) \right]$$

C is the concentration, (x, y, z) are the Cartesian coordinates downwind of the plume, Q is the emission rate of the source, U is the wind speed, y_o is the displacement from the plume centerline, z_o is the effective height of the plume centerline, and \( \left( \sigma_y, \sigma_z \right) \) are the dispersion coefficients (Beychok 2005). SCIPUFF is a Lagrangian puff dispersion model that uses Gaussian puffs to
create a three-dimensional time-dependent concentration field. This model is more complicated than the Gaussian model because the diffusion is based on turbulence closure theory, which gives a prediction of the dispersion rate in terms of the measurable turbulent velocity statistics of the wind field (SCIPUFF, 2011).

The resulting concentration fields of these models are then compared to the sensor observations via a cost function that is logarithmic-in-concentration,

\[
Cost_{\text{logarithmic}} = \frac{\sum_{t=1}^{\text{Time}} \left( \sqrt{\sum_{r=1}^{\text{TR}} \left( \log (C_r + \varepsilon) - \log (R_r + \varepsilon) \right)^2} \right)}{\sum_{t=1}^{\text{Time}} \left( \sqrt{\sum_{r=1}^{\text{TR}} \log (R_r + \varepsilon)^2} \right)}
\]  

(3-1b)

where: \( C_r \) is the concentration as predicted by the dispersion model at receptor \( r \), \( R_r \) is the observation data value at receptor \( r \), \( \text{TR} \) is the total number of receptors, and unlike in previous studies \( \varepsilon \) is the minimum concentration detected, added to avoid logarithms of zero and to scale them accordingly. The solutions with the lowest cost mate, mutate, and then this process iterates until it converges to a best solution (Allen, 2006). Figure 3-1 shows a schematic of this process and in the following sections we will describe enhancements we have applied to the GA-Var process.
Figure 3-1: The GA begins with a random population of “guesses” for the source variables. The resulting concentration field is then compared to our “true” observations by means of a cost function.

First we discuss the Thresholding (3.2.1), then the Statistical Method (3.2.2), the application of a second Cost Function (3.2.3), and finally results (3.3) showing the applications of sections 3.2.1-3.2.3.

3.2.1 Thresholds

A time series of the maximum concentration over all sensors is created for use in discerning the duration of contaminant passage and in noise filtering. Subjective analysis of the time and space of the concentration series suggested that the duration threshold be $10^1$ of the maximum concentration detected. The passage duration is then the time between the first concentration detected over the threshold and the last concentration detected over the threshold.
Figures 3-2 through 3-4 show a plot of the duration thresholds for Trials 15, 30, and 54, respectively.

Figure 3-2: Maximum concentration time series of Trial 15, shown is the time threshold ($10^{-1}$ of the maximum concentration detected) and the concentration threshold ($10^{-3}$ of the maximum concentration detected).
Figure 3-3: Maximum concentration time series of Trial 30, shown is the time threshold (10$^{-1}$ of the maximum concentration detected) and the concentration threshold (10$^{-3}$ of the maximum concentration detected).

Figure 3-4: Maximum concentration time series of Trial 54, shown is the time threshold (10$^{-1}$ of the maximum concentration detected) and the concentration threshold (10$^{-3}$ of the maximum concentration detected).
To mitigate the noise present in the concentration sensor measurements, concentration thresholds were applied to all sensors for each Trial. Previous work (Rodriguez 2011) showed that the GA-Var works best when the concentration data spans at least three orders of magnitude; thus the strict noise threshold is chosen to be $10^{-3}$ of the maximum concentration detected and the loose noise threshold is $10^{-4}$ of the maximum concentration detected in each Trial. Similarly to Rodriguez (2011), the noise thresholds are specified with respect to the maximum concentration value detected. Any data under that level is assumed to be observing noise and is set to zero. After thresholding, given that these three Trials are continuous releases, each sensor is averaged over the duration of contaminant passage (determined with the duration threshold), and then the data at each sensor is used as the observation for each Trial. Figure 3-5 shows Trials 15, 30, and 54 with strict and loose thresholds.
Figure 3-5: On the top is the concentration field for Trial 15 data with loose (a) and strict (b) thresholds, in the middle is the concentration field for Trial 30 data with loose (c) and strict (d) thresholds, and on the bottom is the concentration field for Trial 54 data with loose (e) and strict (f) thresholds. The concentration data is in kg/s, however we have taken the logarithm.
3.2.2 Overview of Statistical Method

The GA-Var method is stochastic and depends on a first guess; thus, each path that optimizes the solution will be different. For that reason it should be run multiple times. In an operational scenario where an estimate of source characteristics is required as quickly as possible we do not expect the GA-Var method to be run more than 10 times given the computational cost vs. time to construct an effective warning. To sample and quantify the uncertainty in ten GA-Var runs, the GA-Var method is run 100 times using the logarithmic cost function predicting 100 potential solutions. Each solution contains four STE variables (wind direction, source location (x,y), and source emission rate). Then the median of a bootstrap sample of 10 solutions is computed 1000 times. This is done to achieve stability in the mean and standard deviation of each variable, thus quantifying the uncertainty in an operational scenario (10 GA-Var runs). A simple schematic showing this procedure can be viewed in Figure 3-6.

Figure 3-6: We run the GA method 100 times. Then we take the median of a bootstrap sample of 10, 1000 times.
3.2.3 Cost Function

Because the AT&D models used are linear with respect to mass we opt to apply a second cost function that is linear to achieve a better estimate of the emission rate. To apply this second cost function, the wind direction and source location (x,y) from the mean bootstrap solution are then used in the AT&D model with potential emission rate solutions. The resulting concentration fields are then compared via a cost function that is linear in concentration,

\[
Cost_{\text{linear}} = \frac{\sum_{t=1}^{\text{Time}} \left( \sum_{r=1}^{\text{TR}} (C_r - R_r)^2 \right)_t}{\sum_{t=1}^{\text{Time}} \left( \sum_{r=1}^{\text{TR}} (R_r)^2 \right)_t}
\]  

(3-1c)

where: \(C_r\) is the concentration as predicted by the dispersion model at receptor r, \(R_r\) is the observation data value at receptor r, and \(\text{TR}\) is the total number of receptors. The solution with the lowest cost is then used as the emission rate. Figure 3-7 shows a schematic of this process.

---

**Figure 3-7**: With the fixed wind direction and location we now find the emission rate with a linear cost function.
3.2.4 Results of Sensitivities

The results for Trials 15, 30, and 54 are shown in Table 3-1 using the Gaussian AT&D model and in Table 3-2 using the SCIPUFF AT&D model. The results listed in these tables are the absolute error of the mean predictions for wind direction, source location (x,y), and emission rate. We show both the emission rate errors when using the logarithmic-in-concentration cost function and the linear-in-concentration cost function. For all Trials and thresholds with the exception of one scenario (Trial 15, using SCIPUFF as the dispersion model) using a linear in concentration cost function to determine the emission rate yields smaller error.

Table 3-1: Absolute error of the predicted parameters when using the Gaussian AT&D model, strict and loose thresholds for Trials 15, 30, and 54.

<table>
<thead>
<tr>
<th>Trial</th>
<th>Threshold</th>
<th>Wind Direction [degrees]</th>
<th>Strength (Log Cost) [kg/s]</th>
<th>Location (x) [meters]</th>
<th>Location (y) [meters]</th>
<th>Strength (Linear Cost) [kg/s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>Strict</td>
<td>3</td>
<td>0.216</td>
<td>10</td>
<td>15</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>Loose</td>
<td>12</td>
<td>0.258</td>
<td>12</td>
<td>11</td>
<td>0.033</td>
</tr>
<tr>
<td>30</td>
<td>Strict</td>
<td>14</td>
<td>0.040</td>
<td>122</td>
<td>139</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>Loose</td>
<td>12</td>
<td>0.029</td>
<td>95</td>
<td>113</td>
<td>0.008</td>
</tr>
<tr>
<td>54</td>
<td>Strict</td>
<td>6</td>
<td>0.086</td>
<td>19</td>
<td>30</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>Loose</td>
<td>5</td>
<td>0.065</td>
<td>17</td>
<td>26</td>
<td>0.005</td>
</tr>
</tbody>
</table>
Table 3-2: Absolute error of the predicted parameters when using the SCIPUFF AT&D model, strict and loose thresholds for Trials 15, 30, and 54.

<table>
<thead>
<tr>
<th>Trial</th>
<th>Threshold</th>
<th>Wind Direction [degrees]</th>
<th>Strength (Log Cost) [kg/s]</th>
<th>Location (x) [meters]</th>
<th>Location (y) [meters]</th>
<th>Strength (Linear Cost) [kg/s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>Strict</td>
<td>9</td>
<td>0.429</td>
<td>884</td>
<td>60</td>
<td>0.989</td>
</tr>
<tr>
<td></td>
<td>Loose</td>
<td>4</td>
<td>0.003</td>
<td>41</td>
<td>3</td>
<td>0.035</td>
</tr>
<tr>
<td>30</td>
<td>Strict</td>
<td>7</td>
<td>0.011</td>
<td>22</td>
<td>16</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>Loose</td>
<td>5</td>
<td>0.011</td>
<td>64</td>
<td>181</td>
<td>0.003</td>
</tr>
<tr>
<td>54</td>
<td>Strict</td>
<td>3</td>
<td>0.011</td>
<td>13</td>
<td>120</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Loose</td>
<td>1</td>
<td>0.011</td>
<td>58</td>
<td>175</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Figure 3-5 reveals that Trial 15 is missing the upper right quadrant of the sensors. These sensors were either offline or did not pass the field post-processing quality control test (Storwold, 2007). In this Trial when the strict thresholds are employed, we discard a large quantity of data and achieve worse retrievals for the source term parameters when using the higher quality SCIPUFF model than when using the simple Gaussian model. Given the “poor” quality of data for this Trial, a loose threshold is needed because the strict threshold eliminates concentration signals that are necessary to estimate the source term parameters.

In Figure 3-5, we can see that Trial 30 is missing nine of the sensors, but only three of these sensors effectively measure the contaminant plume. In this Trial when using SCIPUFF as the AT&D model and the strict thresholds are employed, we do not see a significant difference from the case when the loose thresholds are applied. This result can also be seen in the similar retrievals for the source term parameters when using the Gaussian Plume Model. The SCIPUFF model does do better than the Gaussian model at retrieving the source terms when using strict thresholds. Given the better quality of data for Trial 30 compared to Trial 15, the strict threshold can be applied to retrieve the source term parameters.
As previously mentioned, Trial 54 is considered “good” quality data given that only seven sensors were offline and these sensors were towards the edges of the sensor grid. Nonetheless, Trial 54 has smaller errors in the retrieval source parameters for both the strict and loose thresholds when using the Gaussian AT&D model than when using the SCIPUFF model. Moreover, for Trial 54, there is minimal difference when using a strict or loose threshold, or whether applying the Gaussian or the SCIPUFF AT&D model.

Strict thresholds yield the best results when using SCIPUFF as the AT&D model but can only be used when data are of high quality. Loose thresholds yield better results with SCIPUFF when data are sparse. There is no significant difference in model results for strict or loose thresholds when using the Gaussian AT&D model.

### 3.3 Sensitivity to Limited Sensor Data & Wind Data

In the previous section we found that a higher quality model (SCIPUFF) although more computationally expensive, performs at a similar standard as the lower quality model (Gaussian Plume) in the GA-Var setup. In an operational scenario time is of the essence; therefore, we will proceed in this section using only the Gaussian AT&D model. Along these same lines, we also know that we will not have research grade sensors and high quality meteorological data in operational settings and so we choose to apply strict thresholds for our analysis in this section. We also know that up-to-date and spatially representative meteorological information may not be available. For this reason we choose to evaluate the GA-Var methodology with limited wind and stability information using the mean observed wind and a neutral stability. This setup represents a worst-case scenario in terms of the availability of meteorological data.

In this section we would like to address two questions: (1) can the GA-Var methodology retrieve the representative wind direction and (2) can the GA-Var methodology work when using
fewer sensors? To test these two questions we run the GA-Var method as described in Section 2. Our first experiment uses the mean observed wind as our advecting wind and the second experiment uses the GA-Var to determine the advecting wind. Similar to the work of Annunzio (2011), to evaluate if the GA-Var method will work with fewer sensors, we randomly select 15, 30, 45, 60, 75, 90 sensors and then compare against the results achieved using the full 100 sensors. It is important to note that these extractions are random and we may accordingly select sets of sensors that either never reported concentration or were offline. Such cases will, of course, yield invalid results, pulling down the overall results for that sensor count.

The results that address the two questions posed above can be seen in Figures 3-8 and 3-9, for detailed tables see Appendix 3.5.B. Figure 3-8 shows the absolute error in distance in source location for Trials 15, 30, and 54 when using the mean observed wind and the GA-Var estimated wind. Figure 3-9 shows the absolute error in emission rate for Trials 15, 30, and 54 when using the mean observed wind and the GA-Var estimated wind. For all Trials and all sets of sensors, the difference in both distance & emission rate between using the GA-Var estimated wind to when using the mean observed wind is very small. In general the GA-Var estimated wind produces less outliers for both the location error and the emission rate error. For all Trials, the location error is small and consistent until we have fewer than 30 sensors. When using fewer than 30 sensors we have so little information that it becomes too difficult to retrieve a correct location estimate. Trials 15 and 54 show the smallest location errors. This can be attributed to using the correct stability. Trial 30 has larger, location errors because it was a case where a stable atmospheric boundary layer was modeled with an assumed neutral stability. We must note that in an operational scenario we may not know the stability conditions to insert into the AT&D model so a neutral stability was used in all scenarios. Trial 30 shows a release in stable conditions, which explains why Trial 15 and 54 have smaller errors than Trial 30. When the GA-Var method is run with stable condition dispersion coefficients (see Appendix 3.5.A), the errors in location do
decrease but still are not as small as the location errors in Trials 15 and 54. Stable condition dynamics are still not well understood and very difficult to model, which explains the magnitude of the errors in Trial 30 when run with stable conditions. In Figure 3-9 we see that Trials 30 and 54 show the smallest emission rate errors even when using 15 sensors. This result can be attributed to the concentration signal lost when applying the strict threshold. The error in emissions for Trial 15 are larger because a lot of sensor concentration information is denied with the application of the strict threshold.
Figure 3-8: Absolute error distance for Trial 15, 30, and 54 when using data limitation. The mean wind column shows the absolute error distance with data limitation when using the observed mean wind as the wind in the AT&D model. The GA-Var estimated wind column shows the absolute error in distance with data limitation when using wind estimated from the GA-Var method.
Figure 3-9: Absolute error emission rate for Trial 15, 30, and 54 when using data limitation. The mean wind column shows the absolute error emission rate with data limitation when using the observed mean wind as the wind in the AT&D model. The GA-Var estimated wind column shows the absolute error in emission rate with data limitation when using wind estimated from the GA-Var method.

GA-Var methodology can estimate the representative wind direction in an operational scenario even if there is no or little meteorological information available. This is shown by the
small difference between the errors of location and emission rate when using the mean observed wind vs. the GA-Var retrieved wind. This analysis also shows that this method works with limited sensor information given that we at least have 30 concentration sensors.

### 3.4 Conclusions

This study suggests that with an adequate amount of observational data the GA-Var approach provides a viable means to determine both the CBRN release source parameters and the wind direction/speed from concentration observations. The logarithmic-in-concentration cost function does well at capturing the shape and location of the plume and hence the source location and wind direction, while the linear-in-concentration cost function is better able to capture the higher concentrations necessary to retrieve the emission rate. This is true when employing either loose or strict thresholds to the concentration data. When using SCIPUFF as the AT&D model strict thresholds yield the best results but this approach can only be used when data are of high quality. Loose thresholds yield better results when data are “poorer”. There is no significant difference in retrieval results between using the Gaussian AT&D model. The Gaussian Plume-GA-Var method is much less computationally expensive, which is better for the GA-Var method if we choose to use this operationally.

Real-time wind data are difficult to come by so we determined that the GA-Var estimated wind is representative of the advecting wind and that this methodology is still valid when a limited number of sensors are available. In general the GA-Var methodology is robust enough to estimate the wind in an operational scenario if there is no or little meteorological information available. This analysis also shows that this method also works with limited sensor information given that we at least have 30 sensors. Both of these findings are good news for the operational STE community. Less favorable is the finding that STE retrieval quality decreases substantially if
the assumed stability is widely in error. Thus, some reasonable estimate of stability, such as the
deducible from synoptic weather observations is important for achieving high quality STE
retrievals.

3.5 Appendices

3.5.A Appendix A

Here we have the results for when the GA-Var method is run with stable condition
dispersion coefficients. Figure 3-10 shows the absolute error in distance and emission rate for
Trial 30 when using the mean observed wind and the GA-Var estimated wind. The errors in
location do decrease but still are not as small as the location errors in Trials 15 and 54.
Figure 3-10: Absolute errors of distance and emission rate for Trial 30 using stable stability. The mean wind column shows the absolute errors with data limitation when using the observed mean wind as the wind in the AT&D model. The GA-Var estimated wind column shows absolute errors with data limitation when using wind estimated from the GA-Var method.

3.5.B Appendix B

Table 3-3 show the mean absolute error and standard deviation distance for Trials 15, 30, and 54 when using the mean observed wind and the GA-Var estimated wind. Tables 3-4 show the mean absolute error and standard deviation of the emission rate for Trials 15, 30, and 54 when using the mean observed wind and the GA-Var estimated wind.
Table 3-3: Mean and standard deviation of error distance when using the mean observed wind.

<table>
<thead>
<tr>
<th>Mean Observed Winds</th>
<th>Mean Distance (meters)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial</td>
<td>Sensors</td>
</tr>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>257.4</td>
</tr>
<tr>
<td>30</td>
<td>294.0</td>
</tr>
<tr>
<td>54</td>
<td>93.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standard Deviation</th>
<th>Sensors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial</td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>290.9</td>
</tr>
<tr>
<td>30</td>
<td>272.6</td>
</tr>
<tr>
<td>54</td>
<td>67.4</td>
</tr>
</tbody>
</table>
Table 3-4: Mean and standard deviation of error distance when using the GA-Var estimated wind.

<table>
<thead>
<tr>
<th>Trial</th>
<th>Sensors</th>
<th>15</th>
<th>30</th>
<th>45</th>
<th>60</th>
<th>75</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>310.0</td>
<td>93.9</td>
<td>62.0</td>
<td>47.1</td>
<td>37.4</td>
<td>29.1</td>
<td>19.3</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>275.1</td>
<td>206.4</td>
<td>189.4</td>
<td>189.0</td>
<td>189.5</td>
<td>194.4</td>
<td>187.0</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>134.7</td>
<td>56.6</td>
<td>45.2</td>
<td>43.3</td>
<td>41.0</td>
<td>35.9</td>
<td>35.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trial</th>
<th>Sensors</th>
<th>15</th>
<th>30</th>
<th>45</th>
<th>60</th>
<th>75</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>306.9</td>
<td>134.9</td>
<td>40.5</td>
<td>36.6</td>
<td>29.3</td>
<td>26.8</td>
<td>9.8</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>278.4</td>
<td>100.3</td>
<td>72.1</td>
<td>72.4</td>
<td>52.3</td>
<td>44.5</td>
<td>26.3</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>141.0</td>
<td>42.4</td>
<td>29.1</td>
<td>20.7</td>
<td>18.4</td>
<td>14.3</td>
<td>12.8</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 4

The Use of Ensemble-derived Background Error Covariance Matrices to Improve Gradient Descent Performance in Source Term Estimation

Accurate hazard assessments rely heavily on the source term parameters necessary to characterize the release in an Atmospheric Transport and Dispersion (AT&D) simulation. Unfortunately, these source parameters are often not known and must therefore be based on rudimentary assumptions. Source Term Estimation (STE) strives to solve this problem using the available observations. In this work we use an algorithm that utilizes variational data assimilation techniques to fuse Chemical and Biological (CB) and meteorological observations to characterize source parameters and provide a refined hazard assessment. The National Center for Atmospheric Research (NCAR) – Sage Management Source Term Estimation algorithm, known as the Variational Iterative Refinement STE algorithm (VIRSA), consists of a combination of modeling systems. VIRSA consists of the Second-order Closure Integrated PUFF model (SCIPUFF), SCIPUFF’s STE model, a hybrid Lagrangian-Eulerian Plume Model (LEPM), and the LEPM’s formal adjoint. While numerous approaches to the STE problem exist, each with its own strengths and weaknesses, this approach addresses STE in an operational environment where computational resources are limited and a timely solution is critical. VIRSA is an adjoint method, computationally efficient and fast, but like any gradient descent minimization, its downfall is that it can fall prey to local minima of the cost function. In the work presented here we incorporate new methods to address issues related to uncertainty and using what we know about that uncertainty to reduce the tendency to find local minima rather than the global minimum. We explore approaches to map the uncertainty in our observations and link it back to the background error covariance matrix utilized by the adjoint minimization. Results from these approaches use
data from the FUUsing Sensor Information from Observing Networks (FUSION) Field Trial 2007 (FFT07) dispersion trial and an ensemble of “single realization” plume simulations created by the Eulerian/semi-Lagrangian (EULAG) computational fluid dynamics (CFD) model that replicates this FFT07 trial.

4.1 Motivation

Chemical and Biological (CB) agent detection and effective use of these observations in hazard assessment models are key elements of: (1) CB defense programs that seek to ensure that operations are minimally affected by a CB attack and (2) any accidental release like those in the aftermath of a natural disaster. Accurate hazard assessments rely heavily on the source term parameters necessary to characterize the release for AT&D simulation. Unfortunately, source parameters such as the release location and rate are often unknown and therefore based on rudimentary assumptions. Moreover, in an intentional or typical accidental release of a CB agent we also have little to no relevant meteorological information. When available, the meteorological information used for this application are collected at the closest meteorological station or are based on weather forecast model predictions. Meteorological information is necessary because parameters such as wind direction and speed are required to model the transport and dispersion of the contaminant in order to assess a downwind hazard zone. In this work and as in in Chapter 3 we seek to estimate the source term parameters required to assess the downwind hazard zone, e.g. wind direction and source location. In this Chapter, however, we use an adjoint method and exploit uncertainty information so as to reduce the tendency to converge to local minima in the solution space.

Many methods exist for Source Term Estimation (STE). Rao (2007) reviews methods that include adjoint and tangent linear models, Kalman filters, and variational data assimilation
while Haupt and Young (2008) compare and contrast a general paradigm that encompasses the estimated variables, AT&D model, estimation technique, and test data. While numerous approaches to the STE problem exist, each with their own strengths and weaknesses, the method used in this body of work approaches STE assuming that it will be used in a forward deployed operational environment where computational resources are limited and a timely solution is critical. The work presented here uses the National Center for Atmospheric Research (NCAR) – Sage Management STE algorithm, known as the Variational Iterative Refinement STE Algorithm (VIRSA), which consists of a combination of modeling systems (Bieringer et. al. 2010).

VIRSA consists of the source type preprocessor, the Second-order Closure Integrated PUFF model (SCIPUFF) and SCIPUFF’s STE model (Step 1), a hybrid Lagrangian-Eulerian Plume Model (LEPM) and the LEPM’s formal adjoint (Step 2), and the Hazard Assessment (Step 3). The relationship between the VIRSA components is shown in Figure 4-1. The source type preprocessor utilizes time-series observations to estimate the release type (Instantaneous vs. Continuous) and applies thresholds to the concentration data to prepare it for the following components. The first component of VIRSA uses SCIPUFF’s STE model (also known as inverse/reverse SCIPUFF), which is a backward trajectory-based methodology to produce a first guess estimate of the release parameters (Bieringer et. al. 2010). Then the first guess estimate is used as input in a forward SCIPUFF simulation to create a forecast. This forward SCIPUFF simulation provides a time series of the AT&D spread parameters used in the second component of VIRSA. The second component of VIRSA uses a variational data assimilation technique to iteratively refine the first guess solution provided by the inverse SCIPUFF STE. In this component of the algorithm, a Gaussian dispersion model has been altered to use the dispersion time series computed in the Lagrangian reference frame by SCIPUFF. Because this model uses elements of dispersion solutions from both the Lagrangian and Eulerian reference frames it is referred to as the LEPM. A hybrid model such as LEPM provides a forward atmospheric
dispersion model solution of comparable accuracy to the model used for the hazard assessment (in this case SCIPUFF), while remaining simple enough that a numerical adjoint can be derived for it. The LEPM and its adjoint provide the necessary numerical components required to solve the variational minimization problem. The LEPM and its adjoint are used to iteratively refine the SCIPUFF based STE estimate parameters from step 1 using a gradient descent minimization. This is accomplished by using the forward and adjoint LEPM model to identify the source parameter values that minimize the difference between the forward LEPM model predictions of contaminant concentration and the sensor observations. The difference between the LEPM model concentration prediction and concentration observations is quantified via a cost function. A schematic of a cost function and the gradient descent minimization process in the second component of VIRSA is shown in Figure 4-2. In Figure 4-2 we start with a first guess from the first component of VIRSA, then we run the forward LEPM model and get a concentration field. We compute the cost (J, in Figure 4-2) by comparing the predicted concentration field to the observed concentration field. After the cost is calculated the numerical adjoint is evaluated and this gives us information about the cost gradient and a new “first” guess. This component of the algorithm iterates until converges to a solution the final component of VIRSA computes a downwind hazard using SCIPUFF and the refined set of source and meteorological parameters from the second step. The downwind hazard is a map showing the probability of the concentration being over a toxic threshold.
Figure 4-1: Schematic of the National Center for Atmospheric Research (NCAR) – Sage Management Source Term Estimation (STE) algorithm, known as the Variational Iterative Refinement STE algorithm (VIRSA), which consists of a combination of modeling systems. VIRSA consists of the source type preprocessor, the Second-order Closure Integrated PUFF model (SCIPUFF) and SCIPUFF’s STE model (Step 1), a hybrid Lagrangian-Eulerian Plume Model (LEPM) and the LEPM’s formal adjoint (Step 2), and the Hazard Assessment (Step 3) (Bieringer et. al. 2010).

Figure 4-2: Gradient descent minimization depiction, where the variational data assimilation technique to iteratively refine the first guess solution is shown.

The work here focuses on improving the performance of the second component of VIRSA. The advantage of using the variational minimization process is that it is computationally
efficient; however, any gradient descent minimization method can fall prey to local minima. In the work presented here we incorporate new methods to address this issue by using source parameter uncertainty information to reduce the tendency to converge to local minima. Specifically we address the following 2 questions: (1) Can we exploit the relationships between the uncertainties in the STE parameters to better direct the search for the best concentration field correspondence between the model and observations? (2) Knowing one variable’s uncertainty, can we estimate other variables’ uncertainties that are unknown? To answer these questions we have developed methods that exploit the uncertainty relationships to simplify the cost function surface solved by the algorithm. To simplify the cost function we employ two approaches: (1) a physics-based method that uses the variability and uncertainty in wind direction measurements to influence our estimates of the uncertainty in the source location and (2) an ensemble-based method that maps variability in the concentration field to the uncertainty in source characteristics directly via the adjoint model. In Section 2 we discuss the cost function and background error covariance matrix, which provide background information necessary to understand these two approaches. The specific methodology used in each approach is described in Section 3. A discussion of the field and virtual data used to test these approaches is discussed in Section 4. Finally the results will be presented in Section 5 and the conclusions in Section 6.

### 4.2 Cost Function and Background Error Covariance Matrix

We apply an adjoint method to estimate source characteristics, but first we will discuss the STE problem in a broader context. In this work we model a nonlinear process (e.g. atmospheric transport and dispersion), which is represented by the gray curve in Figure 4-3, and we develop a tangent linear model (TLM) by linearizing the nonlinear model. The adjoint of that TLM gives us a mathematical means to go from an end state to an initial state so in the STE
context we can go transform from a concentration field to source characteristics. If we have a problem for which we have information about the output of a model (concentration information) but we have limited information about the inputs (source characteristics), we can use the adjoint to optimize the solution of the problem. Because we do not know the inputs, we use the concentration from the forward model and compare it to the observations with a cost function and then use the adjoint to give us cost gradient information to infer new inputs, i.e. the source characteristics (Figure 4-2). The cost is the difference between the predicted concentration field from the LEPM model and the observed concentration field and the function will be described in more detail below. A way to visualize the solution space is to plot a cost surface such as that shown in Figure 4-4. That figure indicates the lowest (best) cost value for each point (x,y) in the domain while minimizing for the other variables of interest.

Figure 4-3: Schematic of a modeling a nonlinear process, represented by the gray curve. We can develop a tangent linear model (TLM) by linearizing the nonlinear model. The adjoint of that TLM gives us a mathematical means to go from an end state to an initial state.
Figure 4-4: Schematic of cost via gradient descent minimization to create the cost surface, where the lowest (best) cost value for each point in the domain while minimizing for the other variables of interest is shown.

Uncertainty is incorporated into the cost function in the second component of VIRSA.

The cost function \( J \) for this problem is defined as:

\[
J = \left( \frac{1}{2} \left[ C^{\text{obs}}(t) - C(t) \right]^T \left[ R^{-1} \left[ C^{\text{obs}}(t) - C(t) \right] \right] \right) + \left( \frac{1}{2} [A_M]^T [E_B]^{-1} [A_M] \right),
\]

(4-2a)

where \( C^{\text{obs}}(t) \) is the concentration observed at time \( t \), \( C(t) \) is concentration predicted at time \( t \), and \( R \) is the uncertainty associated with each sensor. \( R \) is defined as:

\[
R = \begin{bmatrix}
\sigma_1^2 & 0 & 0 & 0 \\
0 & \sigma_2^2 & 0 & 0 \\
0 & 0 & \ldots & 0 \\
0 & 0 & 0 & \sigma_N^2 \\
\end{bmatrix},
\]

(4-2b)

where \( \sigma_n^2 \) is the uncertainty associated with each observation (1, 2, 3, …N). In this case we are assuming that the uncertainty with respect to one observation is not correlated with the uncertainty of other observations. This lack of correlation between the uncertainties in the observations is represented in \( R \) by defining the individual uncertainties on the diagonal while the off-diagonal terms are 0. The background error covariance matrix \( E_B \) contains the uncertainty in each STE variable we seek to optimize and \( A_m \) is the solution after \( m \) iterations. The STE
variables we examine through this approach are: lateral location (x,y), height of release (z), wind direction (θ), wind speed (U), emission rate (Q), and time of release (t). Am (Equation 4-2c) is a column vector of these STE variables

\[ A_m = \begin{bmatrix} x_m \\ y_m \\ z_m \\ \theta_m \\ S_m \\ Q_m \\ t_m \end{bmatrix}. \tag{4-2c} \]

If one assumes that the uncertainties are uncorrelated between STE variables the background error covariance matrix is defined as,

\[ E_B = \begin{bmatrix} \sigma_{STE1}^2 & 0 & 0 & 0 \\ 0 & \sigma_{STE2}^2 & 0 & 0 \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \sigma_{STE n}^2 \end{bmatrix}. \tag{4-2d} \]

where \( \sigma_{STE n}^2 \) is the uncertainty associated with each STE variable (1, 2, 3, ..n). In the absence of additional information linking the uncertainties between the STE variables the non-diagonal terms are set to 0 implying that these uncertainties are uncorrelated. This is the default error covariance matrix commonly used for this type of problem, however we believe that the uncertainties are in fact correlated. The challenge, therefore, is how to quantify these uncertainty relationships. First we will discuss a method that does not correlate the uncertainties in the off-diagonal terms.

We first demonstrate a method to quantify this relationship with a physics-based method that uses wind measurements to estimate the uncertainty of the source location parameters and thereby redefine the terms on the diagonal. This approach allows us to replace the off-diagonal zeroes with values that define/bound the uncertainty relationships by relating one variable’s uncertainty with another variable’s uncertainty. Using this approach \( E_B \) then becomes,
The terms on the diagonal of this matrix are the uncertainty of one variable with respect to another. In this case it relates the uncertainty in the lateral source location to the uncertainties in the winds.

It is desirable to have a method capable of filling in all of the diagonal terms so that $E_B$ becomes Equation 4-2f,

$$E_B = \begin{bmatrix}
\sigma_{\text{STE} 1}^2 & \ldots & \text{cov}(\text{STE} 1, \text{STE} n) \\
\ldots & \sigma_{\text{STE} 2}^2 & \ldots \\
\text{cov}(\text{STE} n, \text{STE} 1) & \ldots & \sigma_{\text{STE} n}^2
\end{bmatrix}.$$  

(4-2e)

The ensemble-based method described in more detail in section 3.2 and provides us with an approach to define all of the terms in $E_B$.

4.3 Methodology

Because gradient descent minimization methods can fall prey to local minima in the solution space they are prone to provide different STE solutions depending on the first guess location. In this work we examine two approaches that use knowledge of uncertainty in the STE variables to limit or transform the search domain, thereby removing potential solutions associated with local minima. The first method exploits the physics of the problem to quantify uncertainties in the wind information so as to bound the search domain for the source location. The second method uses synthetic observations from an ensemble of single realization concentration fields predicted by an ensemble of Large Eddy Simulations (LES) to create a background error.
covariance matrix. Both approaches allow us to morph the cost surface from having multiple local minima to a solution surface that has a single global minimum.

4.3.1 Physics-Based Method

For the CB problem it is common that a subset, and in some cases, all of the STE variables are unknown. In most cases, however, some information regarding the uncertainties, or in lay terms, the overall range within which the STE variables must lie is known. In the limiting case where little information is known these ranges can be deduced based on our understanding of atmospheric physics and AT&D. Even when high quality, representative observations are available these uncertainties still remain. For example, due to spatial/temporal heterogeneities in the wind field it is difficult, at best, to fully characterize the winds that advect and disperse airborne contaminants. These uncertainties are inter-related by physics of the STE problem as seen in Figure 4-5. The question is, can these uncertainties be quantified and then exploited through the background error covariance matrix to improve the accuracy of the variational STE method.

Figure 4-5: Universe of STE variables of interest in the plume reference frame, and how their uncertainties are related to each other.
As described above and illustrated in Figure 4-6, uncertainties in both the source’s location and the wind speed and direction are related. The input wind information that is used in the forward LEPM model are single values for wind speed and direction, although we know that the wind has variability in both speed and direction. The wind speed variability can translate into along-wind (x) location uncertainty from the first-guess. The wind direction variability can translate into cross-wind (y) variability from the first guess. This is illustrated in Figure 4-6a, which depicts the cost surface that is the lowest (best) cost value for each point in this parameter subspace when minimized for the other variables of interest. The roughly conical cost basin in Figure 4-6a shows the shape of the location uncertainty due to variability of winds depicted in Figure 4-6b. If we look at the uncertainty relationships between location (x,y) and wind speed & direction, we can translate the variability of the wind to bound the uncertainty in location (Figure 4-6b).

Figure 4-6: A 2-dimensional visualization of the cost function surface. The general shape shows how wind variability can translate into uncertainty in location.

Based on Figure 4-6, we can analyze this wind uncertainty and thus bound the uncertainty in location. This task is accomplished in this here by calculating advective eddy time
and using it to distinguish between the wind direction and wind speed changes that cause plume meander and advect the plume and those that simply disperse it. Using similarity theory we determine the advective eddy time scale to be Equation 4-3a,

\[
\text{Advective eddy time scale} = \frac{\text{Mean wind speed}}{\text{Width of plume}}. \tag{4-3a}
\]

We use the advective eddy time scale to filter the spectrum frequencies that contribute to plume meander from the frequencies that contribute to plume diffusion. After we filter the winds we calculate the variance \(\sigma_{\text{STE}}^2\) of wind direction and wind speed time series to define the wind uncertainties in the \(E_B\). The wind uncertainties in \(E_B\) are \(\sigma_U\) and \(\sigma_\theta\). We can use the \(\sigma_U\) and \(\sigma_\theta\) to define uncertainties in location, such as Equations 4-3b and 4-3c:

\[
\sigma_{xx} = \sigma_{uu} \times \sigma_{tt}, \tag{4-3b}
\]

and

\[
\sigma_{yy} = \sigma_{xx} \sin \left(\frac{\sigma_{uu}}{2}\right), \tag{4-3c}
\]

where \(\sigma_x\) is the variance in the along-wind location, \(\sigma_t\) is the uncertainty in the time of the release (first guess location/mean wind speed), and \(\sigma_y\) is the variance in the cross-wind location.

In this scenario we have only resolved the diagonal terms in the \(E_B\). In the next section we test this method by examining the changes in the cost function shape when there are no bounds set on the diagonal and when we use the error covariance matrix set below in Equation 4-3d,

\[
E_B = \begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & \sigma_\theta^2 & 0 \\
0 & 0 & 0 & \sigma_U^2
\end{bmatrix}. \tag{4-3d}
\]
4.3.2 Ensemble-Based Method

While the physics-based approach is a means to relate the uncertainties between some of the STE variables, which can simplify the corresponding cost function, it is difficult to use this approach to relate all of the uncertainties in the problem with each other. To obtain all of the uncertainty values in the $E_B$ we therefore use an ensemble-based approach much like the forward ensemble method illustrated in Figure 4-7. The difference from the standard ensemble approach shown in Figure 4-7 is that here we use the uncertainty in the concentration observations and via the adjoint and gradient descent minimization we assess the uncertainty in the initial conditions (Figure 4-8). To demonstrate the efficacy of the approach we use an ensemble of observations based on a set of ‘20 single realization’ dispersion solutions produced by an LES model where the imbedded advection diffusion equation is used to compute the concentration at the observation locations. The ensemble of observations provides a representative range of concentration variability for a given meteorological condition. We then use the adjoint to determine the corresponding ensemble of potential STE solutions. This way we are directly mapping the uncertainty in the concentration values backwards via the adjoint. The mean of ensemble of potential solutions is subtracted from the vector of potential solutions and then multiplied by the vector transpose, thus calculating the variance and covariance (Figure 4-9). Because the adjoint model inherently contains the physics of the AT&D problem being solved, the relationships between the uncertainties in the STE variables are now mathematically calibrated in $E_B$. We examine the value of this method by contrasting the changes in the cost function shape for cases where bounds on the diagonal uncertainty terms are very large (no uncertainty information) and zeroes on the non-diagonal terms to when we use the new fully populated $E_B$. This method can be tested on however many STE variables we seek to retrieve and can be added as a preprocessing
step in the algorithm. In this work we only examine the source location (x,y) and wind direction information.

**Standard Ensemble**

![Diagram of Standard Ensemble](image)

Figure 4-7: Schematic of standard forward numerical weather prediction ensemble method.

**Ensemble by Adjoint with Gradient Descent**

![Diagram of Ensemble by Adjoint with Gradient Descent](image)

Figure 4-8: A “reverse” standard ensemble approach where we take uncertainty in the concentration field and via the adjoint and gradient descent minimization find the uncertainty in our initial conditions.
Figure 4-9: Methodology of ensemble-based approach to determine the background error covariance matrix.

4.4 Data

The results presented in the next section use two types of concentration and meteorological data, field trial observations and synthetic observations. FFT07, sponsored by the DTRA, collected data from research-grade tracer, sensor, and meteorological instruments suitable for testing current and future CB STE algorithms (Storwold 2007). This field trial occurred in a highly instrumented desert test area at the U.S. Army’s Dugway Proving Ground. The FFT07 Trial dataset contains high frequency readings from 100 concentration sensors, meteorological sensors, and source information (location and amount). The field observations used here are from the FFT07 Trial data. After examination we determined that Trial 54 is representative of a single-location continuous release, as mentioned in Chapter 3.
For the synthetic observations we used the Eulerian/semi-Lagrangian (EULAG) computational fluid dynamics (CFD) model (Prusa et. al., 2008). EULAG uses a Large Eddy Simulation (LES) approach to solve the partial differential equations governing turbulent fluid flow. EULAG is a general multi-scale, multi-physics computational model for simulating thermo-fluid flows across a wide range of scales and physical scenarios (Smolarkiewicz and Prusa 2005, Prusa et al. 2008, Smolarkiewicz et al. 2001).

The specifications of FFT07 Trial 54 were modeled in EULAG and the resulting CFD data simulate the transport and dispersion of the Trial. An ensemble of 20 members was used as the synthetic Trial 54 data.

4.5 Results

First we examine the physics-based method and results, then the ensemble-based method and results, and finally a combination of both the physics-based and ensemble-based method.

4.5.1 Physics-Based Approach

First we assess the variability of wind direction (Figure 4-10) and speed (Figure 4-11) for the FFT07 Trial 54. Figures 4-10 and 4-11 depict a time series of the observations from all of the surface stations deployed during FFT07. For the purpose of demonstrating the approach it is not necessary to use all of the data, so our analysis uses only the data from the central sonic anemometer shown in Figures 4-12 and 4-13. This station was chosen because it was shown to be qualitatively representative of the measurements of the other instruments. Next we computed the power spectra for wind direction and speed of FFT07 – Trial 54. The analysis utilized 50 minutes of measurements collected at a 10 Hz temporal resolution (Figures 4-14 and 4-15). This
time scale can be used to determine a cutoff frequency in the power spectra between the low-frequency eddies responsible for plume meander and the higher frequency eddies responsible for diffusion. This cutoff frequency (advective eddy timescale) is then used to filter the wind direction and speed data to obtain a time series with only meander as shown in Figures 4-16 and 4-17, respectively. The specifications of this filtering can be found in Appendix 4.7. The standard deviation of the raw wind direction data and the filtered wind direction data are computed. We use the standard deviation along with the computed trend in wind direction and speed over the duration of the release to bound the uncertainty in the location (x,y). The uncertainty terms used in the E_B are \( \sigma_x = 151.3 \) meters, \( \sigma_y = 389.7 \) meters, and \( \sigma_\theta = 25.2 \) degrees. By utilizing uncertainty estimates we can constrain the problem such that a search for an optimal solution in a complex multidimensional cost surface becomes significantly simplified as seen in, Figure 4-18. Utilizing this process potentially negates the need to search the entire solution space because it morphs the cost surface into a more concave form so that gradient descent minimization from any starting point can be guaranteed to reach the global optimum.
Figure 4-10: The wind direction time series of FFT07 Trial 54 for three sonic anemometers and 40 portable weather information displays (PWIDs).

Figure 4-11: The wind speed time series of FFT07 Trial 54 for three sonic anemometers and 40 portable weather information displays (PWIDs).
Figure 4-12: The wind direction time series of FFT07 Trial 54 for the central sonic anemometer.

Figure 4-13: The wind speed time series of FFT07 Trial 54 for the central sonic anemometer.
Figure 4-14: Power spectral density for the wind direction, shown is the calculated cutoff frequency and the -5/3 slope that shows we are in the inertial subrange.

Figure 4-15: Power spectral density for the wind speed, shown is the calculated cutoff frequency and the -5/3 slope that shows we are in the inertial subrange.
Figure 4-16: Time series of detrended wind direction and filtered wind direction for FFT07 Trial 54.

Figure 4-17: Time series of detrended wind speed and filtered wind speed for FFT07 Trial 54.
4.5.2 Ensemble-Based Approach

For the ensemble-based method we use 20 concentration fields, one from each LES ensemble member, which generate an ensemble of synthetic observations. Each set of observations is used in the second component of VIRSA with an unconstrained $E_B$ (Figure 4-9); the unconstrained $E_B$ is defined as where the bounds on the diagonal uncertainty terms are very large (no uncertainty information) and zeroes on the off-diagonal terms. The first guess for each member is also kept constant. The ensemble of potential solutions is then detrended (the mean value of all ensemble members is subtracted from each ensemble member) and this vector is then multiplied by its transpose to create the new $E_B$ (constrained). The new $E_B$ is created to constrain the uncertainty of each variable of interest and will be referred to as the constrained $E_B$. To evaluate how well the method formed we create the cost surface using the original data with the unconstrained $E_B$ and another cost surface using the constrained $E_B$ (Figure 4-19). Figures 4-19 and 4-20 show the cost surface but Figure 4-19 is on a linear scale and Figure 4-20 is on a logarithmic scale; in the algorithm the cost is linear but it is easier to visualize with a log scale. These figures show the best cost surface relative to the other variable of interest for each pair of variables i.e., Longitude vs. Latitude, Wind Direction vs. Latitude, and Longitude vs. Wind.
Direction. In both Figures 4-19 and 4-20 we see that the cost surface for the unconstrained variables is more complex and have local minima unlike the constrained cost surface, which is more convex and has no local minima, only a global minimum located over the source release. Figures 4-21 and 4-22 show the same results for a different first guess, but demonstrate similar results.
Figure 4-19: The cost surface on a linear scale for each pair of variables that we minimized i.e., Longitude vs. Latitude, Wind Direction vs. Latitude, and Longitude vs. Wind Direction, using the constrained $E_B$ and the unconstrained $E_B$. 
Figure 4-20: The cost surface on a logarithmic scale for each pair of variables that we minimized i.e., Longitude vs. Latitude, Wind Direction vs. Latitude, and Longitude vs. Wind Direction, using the constrained $E_b$ and the unconstrained $E_b$. 
Figure 4-21: The cost surface on a linear scale starting from a different first guess, for each pair of variables that we minimized i.e., Longitude vs. Latitude, Wind Direction vs. Latitude, and Longitude vs. Wind Direction, using the constrained $E_B$ and the unconstrained $E_B$. 
Figure 4-22: The cost surface on a logarithmic scale starting from a different first guess, for each pair of variables that we minimized i.e., Longitude vs. Latitude, Wind Direction vs. Latitude, and Longitude vs. Wind Direction, using the constrained $E_B$ and the unconstrained $E_B$. 
4.5.3 Combination Method

We can view both the physic-based method and the ensemble-based method as ways of incorporating a soft constraint because the boundaries are not rigid in VIRSA. However, we can use the physics-based method as a hard constraint by using the uncertainty in location as an area of potential first guesses from the first component of VIRSA. We selected 100 random potential first guesses in the area bound by the physics-based method and evaluated the second component of VIRSA with the unconstrained $E_B$ and the constrained $E_B$. The results are shown in Figures 4-23a, 4-23b, 4-24a and 4-24b. These results indicate that given a good original first guess bounding the potential first guesses does not significantly improve our final source location estimate. We do note that the final location estimate using the constrained $E_B$ is closer to the true source location. The robustness of the ensemble-based method is shown by not bounding the potential solutions with the physics-based method. We selected 100 random potential first guesses in a 5km X 5km domain and evaluated the second component of VIRSA with the unconstrained $E_B$ and the constrained $E_B$, Figures 4-23c, 4-23d, 4-24c and 4-24d. When we use the constrained $E_B$ all of the potential solutions fall into the global minimum and when using the unconstrained $E_B$ we have many solutions fall into local minima.
Figure 4-23: Results with 100 random potential first guesses in the area bound by the physics-based method and evaluated the second component of VIRSA with the unconstrained $E_B$ (a) and the constrained $E_B$ (b). Results with 100 random potential first guesses in a 5km X 5km domain and evaluated the second component of VIRSA with the unconstrained $E_B$ (c) and the constrained $E_B$ (d).
Figure 4-24: Starting with a different first guess results with 100 random potential first guesses in the area bound by the physics-based method and evaluated the second component of VIRSA with the unconstrained $E_B$ (a) and the constrained $E_B$ (b). Results with 100 random potential first guesses in a 5km X 5km domain and evaluated the second component of VIRSA with the unconstrained $E_B$ (c) and the constrained $E_B$ (d).

4.6 Conclusions

We need fast and computationally efficient algorithms to retrieve source characteristics for prediction of hazard zones in case of an accidental or intentional release of a CB agent into the atmosphere. VIRSA is fast and computationally efficient but gradient descent minimization can
fall prey to local minima. In this work we use uncertainty estimates to limit and transform the
search domain so as to remove false solutions associated with local minima of the cost function.

In the work presented here we incorporate two new methods to address issues related to
uncertainty and the use of uncertainty estimates to eliminate such local minima. We transform
the cost surface using a physics-based method and then we also transform the cost surface using
the ensemble-based method. Given that on many occasions we will not have such high quality
and high frequency wind data, therefore the ensemble-based method shows more promise for
operational implementation. We expect that with more experiments using the ensemble-based
methods we can redefine $E_B$ for different stabilities and variables of interest. The beauty of using
a method such as the ensemble-based method is that we would no longer need superior
algorithms to solve the STE problem but could use a simple gradient descent minimization.

The ensemble-based method shown in this chapter morphs the cost surface so that the
only minima is a global minimum, though the $E_B$ was created from a reliable first guess. If we
used a less reliable first guess we need to incorporate uncertainties from the other variables. We
expect that if we minimized more variables and created the full $E_B$ from a less reliable first guess
the method could still morph the cost surface to facilitate global optimization. This remains to be
tested and will be in the future.
4.7 Appendix

Here we have the code for filtering our high frequency wind data.

```matlab
% Meander Analysis
% Diagnostic to forecast frequency cutoff
% Determine width of plume to measure diffusion length and calculate mean
% wind to compute frequency cutoff between diffusion and meandering.
% Mean Wind/width of plume = Frequency cutoff
%
% Date       Author            Comment
% 05-18-2011  Luna Rodriguez and George Young

% I. Housekeeping
clear all
close all
clc

% II. Load wind data
load ('output_Wind_Trial_54.mat');

% III. Fill in hard-coded variables
samplerate = 10;
Duration = 10*60;
y = WindDirCen;
y = detrend(y);
n = length(y);
DetrendLine = WindDirCen - y;
SweepRate = ((max(DetrendLine) - min(DetrendLine))/length(DetrendLine))*samplerate; % degrees per
SweepAngle = Duration * SweepRate;

% III.
Y = fft(y,n);
Pyy = Y.*conj(Y)/n;
f = samplerate/n*(0:nfreq-1);

%III.Create power spectra figure
figure
plot(log10(f),log10(Pyy(1:nfreq)))
axis([-3.5 1.0 -5 5])
title('Wind Direction Power spectral density')
xlabel('Frequency (Hz)')

% IV.1
WindSpeed = sqrt((((U_WindsCen(:,1)).^2))+((V_WindsCen(:,1)).^2));
meanWindSpeed = mean(WindSpeed);
x = WindSpeed;
x = detrend(x);
NN = length(x);
DetrendLineWS = WindSpeed - x;
SweepRateWS = ((max(DetrendLineWS) - min(DetrendLineWS))/length(DetrendLineWS))*samplerate; % degrees per
SweepAngleWS = Duration * SweepRateWS;
X = fft(x,NN);
Pxx = X.*conj(X)/NN;

figure
plot(log10(f),log10(Pxx(1:nfreq)))
```
% IV.2 Calculate width of plume *** approximated***
widthPlume = 200;

% IV.3 Calculate cutoff frequency
cutfreq = meanWindSpeed/widthPlume;

% IV.4 Create weight matrix to apply to raw data
Fpass = 0.9 * cutfreq;
Fstop = 1.1 * cutfreq;
Dpass = 0.057501127785;
Dstop = 0.1;
dens = 20;
[N, Fo, Ao, W] = firpmord([Fpass, Fstop]/(samplerate/2), [1,0],[Dpass, Dstop]);
b = firpm(N, Fo, Ao, W, (dens));
a = sum(b);

figure
freqz(b,a,1024); %
saveas(gcf, 'freq.png')
yfilt = filtfilt(b,a,y);
xfilt = filtfilt(b,a,x);

figure
plot(y,'-b')
hold on
plot(yfilt, '-m', 'LineWidth',4)
hold off

figure
plot(x,'-b')
hold on
plot(xfilt, '-m', 'LineWidth',4)
hold off

% IV.5 Compute Statistics
stdevWindDir = std(y)+SweepAngle;
stdevWindDirFil = std(yfilt)+SweepAngle;
stdevWindSpd = std(x)+SweepAngleWS;
stdevWindSpdFil = std(xfilt)+SweepAngleWS;
save(fullfile('analysis_output_Wind_Trial_' num2str(trial) '.mat'))
function Hd = filter

%FILTER Returns a discrete-time filter object.
%
% MATLAB Code
% Generated by MATLAB(R) 7.12 and the Signal Processing Toolbox 6.15.
% Generated on: 29-May-2011 12:41:03
%
% Chebyshev Type II Lowpass filter designed using FDESIGN.LOWPASS.
% All frequency values are in Hz.
Fs = 48000;  % Sampling Frequency
N = 10;      % Order
Fstop = 12000;  % Stopband Frequency
Astop = 20;  % Stopband Attenuation (dB)

% Construct an FDESIGN object and call its CHEBY2 method.
hd = fdesign.lowpass( 'N,Fst,Ast', N, Fstop, Astop, Fs);
Hd = design(hd, 'cheby2');
% [EOF]

ans =

    FilterStructure: 'Direct-Form II, Second-Order Sections'
    Arithmetic: 'double'
    sosMatrix: [5x6 double]
    ScaleValues: [6x1 double]
    OptimizeScaleValues: true
    PersistentMemory: false

Published with MATLAB 7.14
Chapter 5

Conclusions

Uncertainty introduced into source term estimation by measurement uncertainty and that introduced by approximating a transport and dispersion model’s nonlinear terms are examined in this dissertation. It explores both the sources of this uncertainty and describes novel methods to quantify it within the AT&D framework. First, we examined how errors in the input wind fields may translate into AT&D model solution errors. Second, we used a Genetic Algorithm to variationally couple a fitness function with an AT&D model (GA-Var) to examine how estimates of uncertainty in measurements, e.g. the wind direction, can be used to help bound the STE problem. Finally we used the Variational Iterative Refinement STE Algorithm (VIRSA) and uncertainty estimates to bound and transform the solution space so as to remove potential solutions associated with local minima of the fitness function. The highlights for each chapter are described below.

The first component characterizes the impact of wind direction uncertainties on the corresponding dispersion solutions. The results of this study indicate that there are several general rules of thumb that can be applied to quantify the sensitivity of the dispersion solution to wind direction errors. It is challenging to identify generalized solutions to the wind direction uncertainty problem affecting the emergency response community, although there are release scenarios and locations that are very tolerant to wind direction uncertainties. This finding suggests that for important locations it may be beneficial to pre-compute the wind direction error sensitivities so that they will be readily available should the need arise.

The second component of this study suggests that with an adequate amount of observational data, the GA-Var approach provides a viable means to determine both the CBRN
release source parameters and the advecting wind direction/speed from concentration observations. Because real-time wind data are difficult to obtain, it is particularly useful that the GA-Var estimated wind is representative of the advecting wind and that this methodology is still valid when a limited number of sensors are available. In general the GA-Var methodology is robust enough to estimate the wind in an operational scenario if there is little to no meteorological information available. A less favorable finding is that STE retrieval quality decreases substantially if the assumed static stability is widely in error. Thus, some reasonable estimate of stability, such as is deducible from synoptic weather observations, is important for achieving high quality STE retrievals.

In the third component, we use the combination of modeling systems in VIRSA to quantify the ambiguity between STE variables and more importantly, the relationships between their uncertainties. We use uncertainty estimates to bound and transform the search domain so as to remove potential solutions associated with local minima. In the work presented we incorporate two methods to address issues related to uncertainty and using the estimates of uncertainty to eliminate local minima from the solution space. We transform the cost surface using the physics-based method and the ensemble-based method. The beauty of using the ensemble-based method is that we would no longer need advanced optimization algorithms to solve the STE problem, but could rather use a simple gradient descent minimization.
References


Platt, N., and D. Deriggi, 2010: Comparative investigation of source term estimation algorithms using fusion field trial 2007 data. 8th Conference on Artificial Intelligence Applications to Environmental Sciences at AMS Annual Meeting, Atlanta, GA, J1.2.


Second-order Closure Integrated PUFF Model (SCIPUFF), cited 2011: [Available online at http://www.epa.gov/scram001/dispersion_alt.htm#SCIPUFF]


VITA

Luna Marie Rodriguez

National Center for Atmospheric Research
3450 Mitchell Lane
Boulder, CO 80301

Voice: 303-497-2745
Fax: 303-497-8386
lunar@ucar.edu
lunar@starinst.org

Education

The Pennsylvania State University, PA
Ph.D. in Meteorology: 2012
Dissertation: “Uncertainty Propagation Within Source Term Estimation”
Awarded: December 2012

The Pennsylvania State University, PA
M.S. in Meteorology: 2008
Thesis: “Source term estimation using a genetic algorithm and incorporating sensor constraints”
Awarded: August 2008

University of Puerto Rico, PR
B.S. in Physics: 2006

Research Experience

National Center for Atmospheric Research, CO
Associate Scientist II
Research Source Term Estimation (STE) and uncertainty propagation using the Variational Iterative Refinement STE Algorithm. 2011 - Present

The Pennsylvania State University, PA
Research Assistant
Research STE using a Genetic Algorithm (GA) coupled with AT&D models. 2007 - 2010

Peer Reviewed Papers