IMPROVED SOURCE-RECONSTRUCTION THROUGH THE EXPLOITATION OF DOSE-RESPONSE MODELS

A Thesis

by

SAMAR HAMZA EMAM ELKHALIFA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Konstantinos Kakosimos
Committee Members,	Arul Jayaraman
	Reza Tafreshi
Head of Department,	M. Nazmul Karim

August 2016

Major Subject: Chemical Engineering

ABSTRACT

The emergency actions following an accidental or intentional release of a hazardous material (HazMat) are strongly influenced by what is known of the HazMat event and its evolution (e.g. mass flowrate of the release, location of the release, and the duration of the release). Thus, the availability of such information in a timely and accurate manner is of great importance for emergency crews both on-site and in the areas in the proximity of the site. This information aids in predicting the fate of the release and mitigating the possible threats imposed by it on human health.

Characterizing the source of an atmospheric contaminant is typically regarded as an inverse problem, where one has to infer the source characteristics of the released HazMat from concentration or deposition measurements. The inverse solution is obtained by combining atmospheric dispersion models and concentration measurements in an optimal way, where dispersion models are used to produce concentration predictions in space and time using as input initial guesses of the source information. Furthermore, estimating the source parameters needed for atmospheric transport and dispersion modeling requires the application of deterministic and stochastic inversion techniques, such as the Bayesian frames employing Monte Carlo methods.

Locating the source and determining its release rate based on downwind concentration measurements, however, is only viable for some cases like industrial accidents, where measurement data from monitoring stations are typically obtained on-site. For other release cases, such as transport accidents or malevolent attacks, there is a lack of input data (i.e. immediate concentration measurements). Hence, there is a necessity to develop new

approaches for real cases where this information is probably not available to the required extent or at all. This work presents the development, application and assessment of computational algorithms used in reconstructing the source characteristics following an accidental release of an airborne hazard. A promising methodology is the utilization of the resulting health symptoms from exposure as indirect input for emergency response systems. In this work, a total of six scenarios were constructed and analyzed in terms of their ability to reconstruct the source rate in addition to the source location. The results revealed that the source term information can be identified with good agreement with the true source parameters when using the new scheme. However, the complexity of the information used as input was reflected on the minimum requirement of input data needed to reconstruct the source term. The results also revealed the necessity to explore other sophisticated sampling and inversion techniques.

ACKNOWLEDGEMENT

I would like to express my gratitude to my advisor Dr. Konstantinos Kakosimos for his guidance, assistance and support in completing the work described in this document, and for his comments. I am also grateful to my thesis committee members: Dr. Arul Jayaraman and Dr. Reza Tafreshi for serving in my committee and for their comments.

I would like to express my gratitude for the support given to me by members of the Department of the Chemical Engineering at Texas A&M University at Qatar (TAMUQ) and the graduate studies office at TAMUQ, especially for granting me the Itochu fellowship which allowed me to pursue my graduate school studies.

This publication was made possible by a NPRP award [NPRP 7-674-2-252] from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the authors.

NOMENCLATURE

HazMat	Hazardous Material
PBTK	Physiologically-based Toxicocokinetic
ER	Emergency Response
CBRN	Chemical, biological, radiological, and nuclear
SCIPUFF	Second-Order Closure Integrated Puff
MCMC	Markov Chain Monte Carlo
SMC	Sequential Monte Carlo
GA	Genetic Algorithm
C	Concentration of the toxic in the air
T	Exposure duration
$L_{e,p}$/TL/L	Toxic Load
n	Toxic load exponent
ti	Initial time
tf	Final time
Y	Probit
a,b	Constant specific to the chemicals and consequence type
ADME	Adsorption, distribution, metabolism, and elimination
X	Downwind distance (m)
Y	Crosswind distance (m)
Zc	Profile center height (m)
b	half-width parameter (m)
bx	Half-length parameter (m)
sr2	Square root of two
C(t)	Instantaneous concentration at a specific time
AEGL	Acute Exposure Guideline Levels

α	Power law defined for each of AEGL bands
C_{4hr} particular AEGL	USEPA concentration corresponding to the 4 hour duration for a
H_2S	Hydrogen Sulphide
C(X,Y,Z, t)	Downstream mass concentration
Rsq	Residual Squared; normalized men square error
J	Cost function of the correlation coefficient method
Q/Qs	Source Rate
U	Wind Velocity
fminbnd	Matlab's built-in function, used to obtain the source rate
corrcoef	Matlab's built-in function used in obtaining the source coordinates
R^2	Pearson correlation coefficient (Matlab's correcoef), ranges from 0 to 1.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

1. INTRODUCTION

1.1 Background

The release of hazardous materials (HazMat), whether intentional or accidental, poses great threats to the public, especially in densely populated areas. The Sarin gas terrorist attacks in Matsumoto city in 1994 and Tokyo in 1995 (Morita et al., 1995; Okumura et al., 1996), and the famous Bhopal incident in India in 1984 (Bowonder & Miyake, 1988) are examples of such release incidents and their severe consequences (see Figure 1). In the event of a malicious intentional or accidental release of a HazMat, emergency response systems rely on highly uncertain information (i.e. unknown HazMat and release rate), thus resulting in equally uncertain response actions. In order for first responders to take the appropriate action and ensure public safety, areas where human lives are endangered need to be identified. This requires the availability of fast and reliable information on the source characteristics, such as the amount being released, source location and nature of the contaminant. These information enable the accurate prediction of the subsequent atmospheric transport and dispersion, which in turn aid in mitigating the contaminant threat and allow an early identification and treatment of the exposed individuals.

Figure 1. Rows of people who died as a result of exposure to the poisonous Methyl isocyanate (MIC) gas in Bhopal (India) in 1984

Risk assessments are conducted to evaluate the severity of such incidents and the potential hazards to the exposed population. Moreover, data certainty and availability are two major issues that have great impact on the accuracy of the prediction of the risks imposed by hazards on human health (Christopoulos & Roumeliotis, 2005) and consequently on the effectiveness of the emergency response action. Uncertainties in data arise from various sources, such as the meteorological information, source term characteristics, and model errors (Rao, 2005). The source term characteristics (i.e. emission rate over time), however, tend to be the most dominant (Korsakissok et al., 2013). Therefore, the accurate estimation of the unknown source characteristics, which is to be used in atmospheric dispersion models to forecast the fate of the release, is crucial for emergency response systems dealing with hazardous airborne releases.

Characterizing the source of an atmospheric contaminant is approached by solving the inverse problem, where one has to infer the source characteristics of the released HazMat from concentration or deposition measurements (Lushi & Stockie, 2010). The inverse solution is obtained by combining atmospheric dispersion models and concentration measurements in an optimal way (Haupt & Young, 2008). Dispersion models are used to produce concentration predictions in space and time using as input initial guesses of the source information. The difference between the field-monitored data and the model predictions is then compared in a cost function, and accordingly a better guess of the source term is generated and passed to the dispersion model. Further, estimating the source parameters needed for atmospheric transport and dispersion modeling requires the application of deterministic and stochastic inversion techniques, such as the Bayesian frames employing Monte Carlo methods (Rao, 2007).

1.2 Motivation and Scope of Work

Current studies investigating the possible tools and methods of obtaining the solution of the inverse problem of the estimation of the source characteristics focus on the use of downwind concentration measurements. Source reconstruction using concentration measurements is a viable option for some cases like industrial accidents, where measurement data from monitoring stations are typically obtained on-site (Nasstrom et al., 2007). However, for release cases such as transport accidents or malevolent attacks, there is a lack of input data (i.e. immediate concentration measurements) (Vallero & Isukapalli, 2013; National Research Council, 2003). From this stems the need to develop new approaches for real cases where these information are probably not available at the required extent or at all.

Clinical field health observations provided by the emergency response personnel, such as coughing, asthma, and deaths, are among the first information available after such unforeseen

events. The use of clinical data in exposure reconstruction with the aid of biomedical models (e.g. physiologically-based toxicokinetic models) has shown great success in a number of recent applications aiming at resolving the issue of data unavailability following the release of a contaminant. Examples for the application of physiologically-based toxicokinetic (PBTK) models include the works of Georgopoulos et al. (2009) who estimated potential exposure to a contaminant via the use of human biomarker data, and Chen et al. (2012) who utilized Monte Carlo techniques to estimate exposure based on levels of a metabolite in clinical blood and urine samples. In the aforementioned applications, the focus was on the use of biomarkers such as blood measurements, and not on the observed health symptoms. Observed health effects provide an evidence of past exposure, and hence there is a potential for their use in exposure reconstruction. The adversity of these effects on human health is linked to both the exposure and duration, and is dependent on the nature and dosage of the HazMats. The use of a known pattern of disease and weather conditions in identifying the likely point of release of a biological agent was investigated by Meselson et al. (1994) to infer the likely cause of the 1979 Sverdlovsk Anthrax outbreak. However, to date, the use of health effects in source reconstruction has not been fully explored. In this work, we explore the use of health symptoms as input to a new emergency response system that can reconstruct the source characteristics (i.e. emission rate and location).

1.3 Research Problem and Objectives

The overall goal of this thesis is to develop and test an integrated framework which allows source reconstruction using observed health symptoms, and demonstrate the utility of this approach for emergency response applications. In particular, the proposed work aims to solve the problem of lack of information regarding the magnitude and location of a release that are

needed for accurate dispersion predictions of airborne hazards in case of accidental or malevolent releases. Figure 2 illustrates the evolution of the HazMat event versus the flow of a proposed integrated emergency response (ER) and preparedness system.

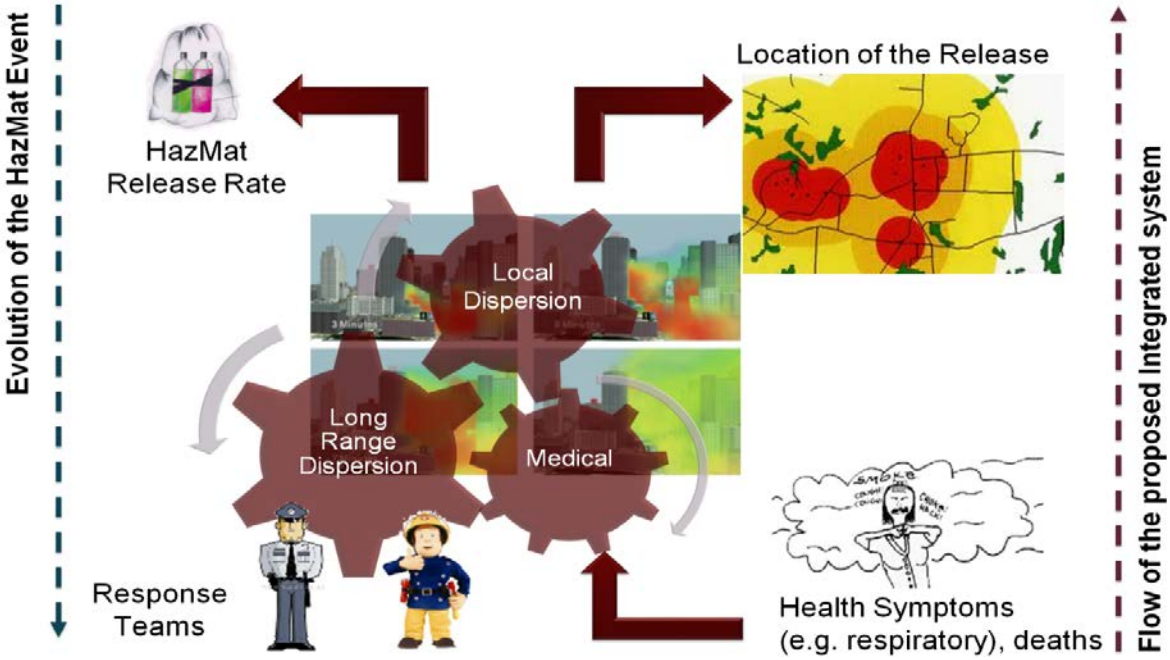

Figure 2. Flow of the HazMat event and the proposed integrated system

In this work, we focus on the medical component (i.e. health outcomes resulting from an accidental toxic gas release) of the integrated emergency response and preparedness system. HazMats released into the atmosphere result in varying health effects among the exposed population as they have specific impacts on human functions and organs. The resulting health effects are implicitly linked to the dosage/exposure, and hence there is a possibility of

exploiting health observations as indirect input data for new Emergency Response System (ERS) that can reconstruct the source (i.e. location and rate).

Dose-response models are employed in order to translate health observations/symptoms to useful exposure and atmospheric concentration information via the application of inversion techniques and algorithms. The inverse problem of locating an unknown source and determining its release rate is approached by coupling atmospheric dispersion modelling with dose response models. The first provides the relationship between the source term and the downwind concentration of an airborne contaminant, while the latter establishes the link between exposure and the corresponding health outcomes.

The specific objectives of this thesis are to:

- Develop and implement computational tools that can be used to locate and quantify the unknown release of a hazardous substance, making it suitable for use in emergency preparedness and response in cases of acute chemical, biological or radiological incidents.
- Assess the capability of the scheme to reconstruct the source over a number of scenarios that prove the validity of the hypothesis while at the same time stress its limits.

1.4 Thesis Organization

This thesis focuses on the development and implementation of computational tools that incorporate dose-response models for use in source reconstruction and emergency response applications. Different cases are presented to demonstrate the capability of this approach in characterizing the source of a HazMat release.

Section 1 presents a brief introduction and overall review of the concepts pertaining to source reconstruction and emergency response in cases of atmospheric hazardous releases, the

motivation behind the current work, as well as a statement of the problem being addressed and the objectives to be achieved upon the completion of this research work.

Section 2 then presents the relevant background information found in the literature on the approaches used to address the source reconstruction problem, from the use of concentration measurements and dispersion models, to the different inversion algorithms being used, and the introduction and use of the biomedical models in deriving potential doses and exposures. It also provides an overview of the different dose-response models that function as a linkage between the observed health outcomes and the exposure suffered by the exposed individuals.

Section 3 describes the methods used in constructing the illustrative scenario that involves the release of a HazMat from an industrial facility, the subsequent dispersion and downwind concentration calculations, and the estimation of the resulting health symptoms. This scenario constitutes the basis on which the inverse problem is then formulated; source reconstruction starting from health observations.

Section 4 presents a preliminary analysis of the proposed framework in which the capability of the scheme to reconstruct the source is assessed over a number of scenarios that prove the validity of the hypothesis while at the same time stress its limits.

Finally, conclusions and recommendations for future work are presented in Section 5.

2. LITERATURE REVIEW

2.1 Previous Studies

Over the last years, there has been a surge of interest in finding means to characterize the source term associated with the release of chemical, biological, radiological and nuclear (CBRN) agents for use in disaster planning and national security. Examples include attempts to calculate the amount of leaked radioactive substances following the Chernobyl (Gudiksen et al., 1989) and Fukushima (Stohl et al., 2012) accidents. The reconstruction of the source term characteristics is commonly approached using inversion algorithms. This type of approach utilizes atmospheric dispersion models, meteorological data, and depending on the complexity of the models other data including description of the topography of the site and other factors that affect dispersion might be required (Haupt & Young, 2008). An atmospheric dispersion model uses the known source release information (i.e. release location, altitude, amount of agent released, and meteorological data) to predict the subsequent dispersion and transport of the contaminant in space and time. Commonly used atmospheric dispersion models in exposure assessment include ALOHA, DEGADIS, and SLAB (Mazzola & Addis, 1995).

There is a considerable prior work on the use of the forward atmospheric dispersion models in source reconstruction. Rudd et al. (2012), for example, applied the Gaussian Plume model in their inversion algorithm to obtain estimates of the source term parameters (i.e. source strength and location) as well as the uncertainty in those estimates from a set of limited concentration measurements obtained from a network of chemical sensors. Furthermore, Allen et al. (2007) incorporated the Second-Order Closure Integrated Puff (SCIPUFF) dispersion model into a Genetic Algorithm-coupled system to estimate the source characteristics. Another

example is the use of a Gaussian Plume solution for the advection-diffusion equation in estimating contaminant emissions from multiple point sources which was explored by Lushi and Stockie (2010). Figure 3 illustrates a typical setup of the inverse problem using concentration measurements.

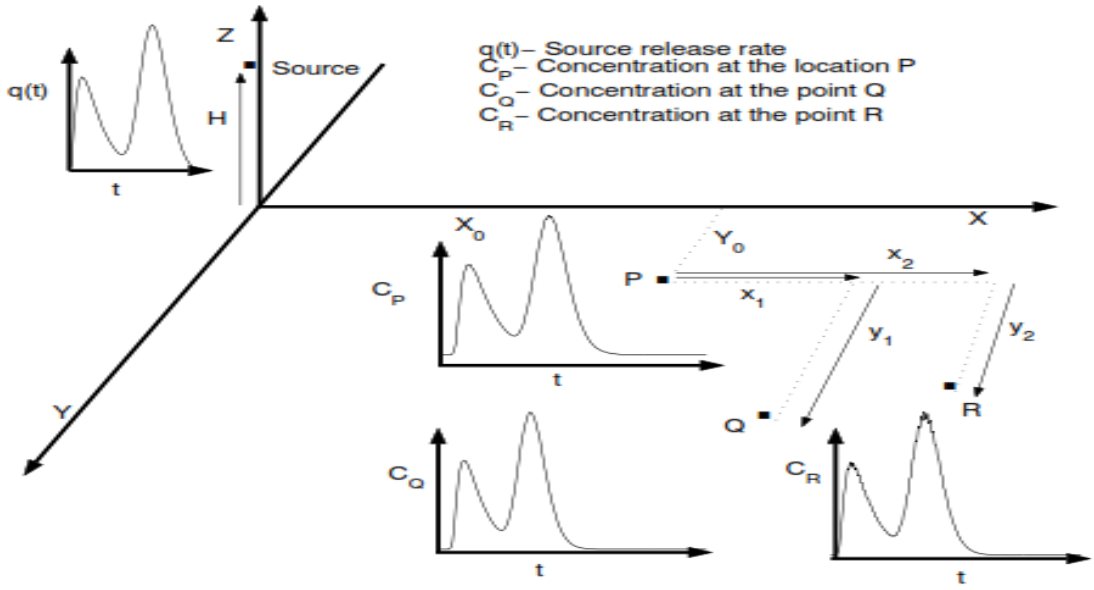

Figure 3. Setup of the inverse problem using concentration measurements (Kathirgamanathan et al., 2003)

Due to the recent advances in computational techniques and tools, the estimation of the source term can be obtained through a number of methods. Source estimation methods applied to atmospheric toxic gas releases can be classified as forward and backward (inverse) modelling methods according to Rao (2007), who provides a review of these methods and examples of their applications. In the forward modelling methods, the atmospheric transport from sources to receptors is analyzed using forward-running transport and dispersion models or computational fluid dynamics models which are run many times until an array of potential

sources is achieved. The resulting model predictions from each run are compared to measurements of observed concentrations from multiple sensors, and accordingly the source conditions are then adjusted until convergence is achieved and an acceptable level of agreement is obtained. Examples of the commonly used forward modelling methods include Bayesian updating and inference schemes using stochastic Monte Carlo or Markov Chain Monte Carlo sampling techniques. The probabilistic approach using a Bayesian inferential scheme for source reconstruction has been developed and implemented in a number of studies, including Wang et al. (2015) and Yee et al. (2014). Wang et al. (2015) evaluated the performance (i.e. accuracy, time consumption, quantification of uncertainty) of three different probabilistic methods (i.e. MCMC, sequential Monte Carlo (SMC), and ensemble Kalman) in terms of their source term estimation of a toxic gas release using synthetic concentration data. Inverse modelling methods, on the other hand, use only one model run in the reverse direction from the receptors to estimate the upwind sources. Inverse modeling methods include adjoint and tangent linear models, Kalman filters, and variational data assimilation, among others. A commonly used technique in inverse modeling is the variational method (e.g. Haupt et al., 2008). This technique is based on minimizing the discrepancy between the model predictions and the measured observations via adjustments of the initial conditions of the forward model. In other words, a cost function is defined as the sum of the squared differences between the modelled and observed quantities. The optimal solution that best explains the measured observations over a given time period is then obtained by searching for the parameters (e.g. source rate and location) that would result in the minimum of the cost function. Furthermore, different approaches to estimate the source term based on the use of forward and backward techniques can be found in the literature. The use of least-squares inversion in obtaining

estimates of the source parameters is implemented in the works of de Foy et al. (2015), Singh and Rani (2014), in addition to Lushi and Stockie (2010). Wawrzynczak et al. (2014) demonstrated the successful application of the Genetic Algorithm (GA) method in estimating the parameters of an abrupt contamination source under a Bayesian frame. The method combines Bayesian inference with GA to find the most probable location of the source and its strength using knowledge of the downwind concentration measurements and the wind field. Other approaches include the use of the artificial neural networks (e.g. Sharma et al., 2011), and the renormalisation inversion (Singh et al., 2015; Issartel et al., 2012; Sharan et al., 2009).

In recent years, there have been notable efforts to incorporate dose-response models in health risk assessment. These models, such as the PBTK models, are known for their scientific soundness and reliability, and hence aid in reducing the inherited uncertainties in risk assessment. The significance of using different toxicity models in estimating causality for military medical planning was investigated by Weinrich et al. (2008). The models explored in the study included: Haber's law, the integrated toxic load (TL) model, the mean concentration TL model, in addition to a two-compartment toxicokinetic model. A number of recent studies aiming at exposure reconstruction (e.g. Christensen et al., 2015; Côté et al., 2014; Georgopoulos et al., 2009) examined the use of PBTK models in estimating potential exposures to chemical agents through the interpretation of human biomarker data (see Figure 4).

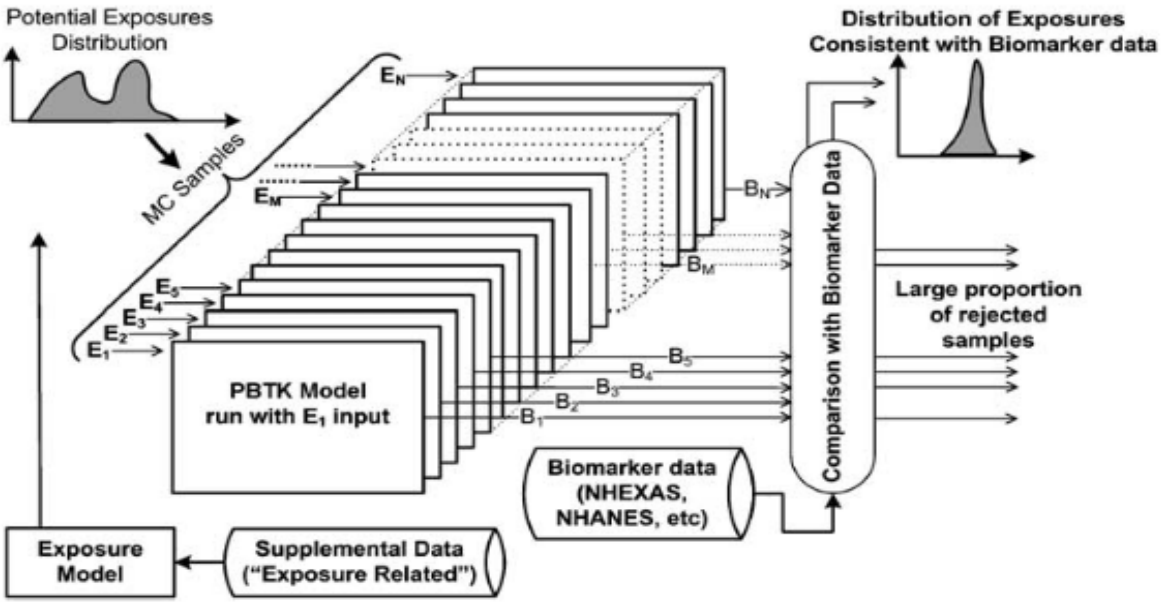

Figure 4. A simplified schematic of a computational framework for source reconstruction using biomarkers and PBTK modeling (Georgopoulos et al., 2009)

In the work of Chen et al. (2010, 2012), the use of clinical data (i.e. blood sampling, and urinary data) in reconstructing past exposures was explored using MCMC techniques. In addition to clinical data, observed health effects (e.g. coughing, dizziness, and death) provide an evidence of past exposure, and hence there is a potential for their use in exposure reconstruction. The adversity of these effects on human health is linked to both the exposure and duration, and is dependent on the nature and dosage of the HazMats. The use of a known pattern of disease and weather conditions in identifying the likely point of release of a biological agent was investigated by Meselson et al. (1994) to infer the likely cause of the 1979 Sverdlovsk Anthrax outbreak.

2.2 Dose-Response Models

2.2.1 Introduction

Accidental or deliberate releases of HazMats into the atmosphere can result in adverse health effects among the exposed individuals. Toxic effects of chemical agents are typically quantified by the dosage from exposure that would produce specific toxic effects (e.g. death). Estimates of acute health effects following exposure to a HazMat can be obtained through the application of a number of dose-response models and relationships that exist in the literature. A brief overview of some of the commonly used relationships is presented next. It should be noted, however, that the current work is aimed at exploring the use of the dose-response models in emergency response applications, hence the details underlying the use of these models are outside the scope of this work.

2.2.2 Toxic Load

Toxicological effects resulting from exposure to hazardous chemicals depend on dosage, which itself is dependent on both the concentration of the toxic chemical and the exposure duration (Reisfeld et al., 2006). This is in line with Haber's law (Witschi, 1999), which states that for a specified percentage of a population, dosage is the key determinant of the resulting toxicological effects. Mathematically, Haber's law can be expressed as follows:

$$CxT = L_{e,p} \tag{2.1}$$

where

C is the concentration of the toxic chemical in the air

T is the exposure duration

Le,p is the toxic load; a constant value that determines the toxicological effect in a specified percentage of the population

As implied by Haber's law, a linear relation exists between the product of toxic concentration and exposure time, and the level of the harm experienced by the exposed individuals. Consequently, for a given dosage, toxicological effects resulting from conditions involving prolonged, low-concentration exposures will be identical to those involving short-duration, high-concentration exposures. In later studies covering wider ranges of exposure times and different toxic chemicals, the validity of Haber's Law was shown to be limited to relatively short exposure times (Witschi, 1999; Stage, 2004). Moreover, Haber's law suffered the inability to address biological effects associated with exposure to a number of acutely toxic gases and aerosols. To account for these limitations, ten Berge et al. (1986) used the method of Probit analysis to re-evaluate raw data obtained from previously published acute inhalation toxicity studies for some volatile industrial chemicals. In their work, ten Berge et al. (1986) varied the concentration and duration of exposure for various toxic vapors, and concluded that the concentration and exposure duration were not equally significant in determining the toxic effect. The study employed the toxic load model in describing the relationship between concentration, exposure duration, and the resulting biological effect. The results of ten Berge et al. (1986) agree with those of Busvine (1938), who also found that the toxic response depends on a parameter L, which is commonly known as the toxic load. The toxic load model can be expressed as:

$$C^n xT = L_{e,p} \tag{2.2}$$

where n here is the toxic load exponent, whose value varies with the type of toxic under consideration as well as the exposure conditions.

Mioduszewski et al. (2002) tested the adequacy of the toxic load models of Haber and ten Berge et al. (1986) in predicting the toxicity of Sarin vapor exposure in rats for different

exposure conditions. The results of the study implied the necessity to expand beyond the dependence on Haber's rule. Additionally, the toxic load can be computed directly for continuous releases in which concentration is constant. For instantaneous (i.e. time-varying) concentrations, a number of extensions of the toxic load model are found in the literature. Table 1 presents examples of four proposed extensions of the TL model, namely the integrated concentration (also known as the instantaneous exposure model), the mean concentration, the concentration intensity, and the peak concentration models. A review of various extensions of the toxic load model for the case of time-varying concentrations and their application can be found in Gunatilaka et al. (2014).

Table 1. Summary of four proposed extensions of the TL model for the case of time-varying concentrations

TL model	Expression	Reference
Integrated Concentration (instantaneous exposure)	$Le,p\ (x) = \int_{ti}^{tf} C(t)^n dt$ (2.3)	ten Berge and van Heemst (1983)
Average Concentration	$Le,p\ (x) = [\frac{\int_{ti}^{tf} C(t)dt}{tf-ti}]^n\ x\ (tf - ti)$ (2.4)	Hilderman (1999)
Concentration Intensity	$Le,p\ (x) = \frac{(\int c(x,t)dt)^{2-n}}{(\int c^2(x,t)dt)^{1-n}}$ (2.5)	Sykes et al. (2004), Czech et al. (2011)
Peak Concentration	$Le,p = \frac{\int c(x,t)dt}{{c_{peak}}^{1-n}\ (x)}$ (2.6)	Stage (2004)

2.2.3 Probits

One of the key aspects of risk analysis is the estimation of exposure to hazardous agents suffered by a given population, and the subsequent formation of a relation between the resulting effects and the estimated exposure (Griffith, 1991). The Probit analysis provides a useful framework for expressing the varying susceptibility of populations exposed to stressing agents. A Probit function relates the concentration and duration of a hazardous exposure to the degree of injury observed in an exposed population. The application of the Probit analysis covers a wide range of stressing agents and populations including blast overpressure and its

effects on structures, in addition to exposure to toxic, radioactive, or flammable materials and the resulting human health effects (Griffith, 1991). For the particular case of toxic health effects, the Probit analysis is used in the assessment of toxic response to irritant gases. The Probit method was first introduced by Bliss (1934a, 1934b) as a tool to linearize cumulative normal distribution dose-response curves. The estimation of the fraction of a population responding to a release is achieved by first calculating the toxic load, L, which is then substituted in the Probit equation to determine the value of the variable Y, which in turn is converted to percentage using the transformation tables/charts as the ones provided by Finney (1971). Moreover, it is understood that one Probit unit corresponds to one standard deviation of the normal distribution, and that the median (50th percentile) is arbitrarily assigned a Probit value of 5. Typical Probit functions are of the following form (HSE, 2006):

$$Y = a + b\,(ln\,L) \tag{2.7}$$

where:

Y = Probit, (value range 2.67 – 8.09 representing 1 – 99.9% response) a measure of the percentage of the vulnerable resource that might sustain damage. Injury probability can then be determined by evaluation of Y on a Probit transformation chart such as that provided by Finney (1971).

a, b are constants specific to the chemicals and consequence type

L= the product of intensity or concentration of received hazardous agent to an exponent “n” and the duration of exposure in seconds or minutes

A number of Probit functions that provide estimates of injury levels resulting from exposure to harmful substances have been generated and published for a variety of chemicals.

Due to the fact that multiple Probit functions may exist for the same hazard in different sources in the literature, dose-effect estimates may not always be consistent for each of the Probit functions being applied. Hence, the selection of the most adequate Probit function is critical as it impacts the rest of the Probit analysis. Another issue with the use of Probits is the paucity of data on human toxic response and the fact that most of these expressions are derived from tests on animals, which dictates cautious application of Probits to human exposures (Griffith, 1991). Furthermore, we are a long way from having a sound basis for the application of Probits to human exposures even for the case of steady exposures. This is further complicated when considering the unresolved issues of dispersion factors such as intermittency and concentration fluctuations.

2.2.4 Physiologically-Based Toxicokinetic (PBTK) Models

The application of PBTK models in dose-response studies has received wide recognition in the field of human health risk assessment, due to their scientific soundness and reliability (Edler & Kitsos, 2005). PBTK models allow the conversion of exposure concentration to the target tissue dose, which provides better scientific basis than the external or exposure dose of the parent chemical when it comes to performing risk assessments and relating observed toxic effects (Andersen et al., 1991; Clewell et al., 2002). The first application of the PBTK models in risk assessment started in the 1980's with Methylene Chloride (Andersen et al., 1987). Since then, the PBTK models have been used in various applications in risk assessment such as the reconstruction of doses that are associated with different routes and media of exposures (Roy et al., 1996; H. V Rao & Ginsberg, 1997; Lyons et al., 2008) . Figure 5 represents sample application of PBTK models in risk assessment.

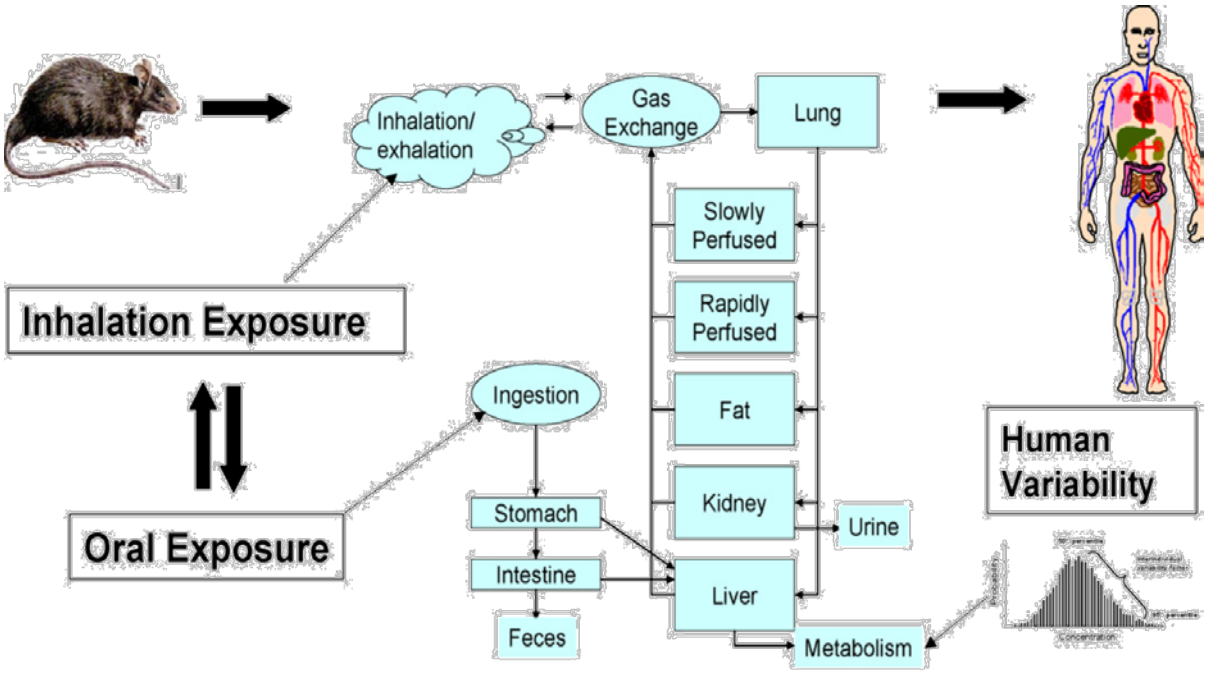

Figure 5. Sample application of PBTK models in risk assessment (Dewoskin, 2007)

In principle, pharmacokineticts stand for the quantitative description of absorption, distribution, metabolism, and elimination (ADME) of xenobiotics in living organisms (Edler & Kitsos, 2005). PBTK models relate external exposure and internal target tissue doses by tracing the flow of an administered chemical or its metabolites through the blood to different organs of the body that are known as the compartments. The level of complexity of PBTK models varies with the amount of details employed. The one-compartment PBTK model is the least detailed model, which consists of a simple relationship between intake, retention, and elimination from the body that is treated as a homogenous whole. On the other hand, Multi-compartment models provide detailed description of distinct body components and tissues, such as blood, liver, fatty tissues, etc.

Furthermore, the embodied biological and mechanistic information in the PBTK models enable them to be used for extrapolation of the kinetic behavior of chemicals across different exposure routes, from high dose to low dose, and from test animal species to humans (Reisfeld et al., 2006). Therefore, the main advantages of the PBTK models are their ability to perform such extrapolations, as well as the detailed description they provide of the concentration-time profiles in individual tissues or organs. Despite their attractive uses, the advantages of the PBTK models come at the expense of requiring a relatively large investment in resources to develop a validated PBTK model with well-found parameter values.

In this work, the instantaneous toxic load model is employed in the dose-response modeling discussed in section 3. This extension of the toxic load model is selected due to its simple application, in addition to being well-suited for cases involving releases with time-varying concentrations.

3. METHODOLOGY

3.1 Introduction

In this section, an integrated modeling system is developed and applied to a synthetic scenario involving the release of a hazardous material in order to test the proposed methodology; reconstruction of the source term using health outcomes as input data. The synthetic scenario is constructed and employed in absence of real data. First, forward modeling based on known source characteristics is carried out to estimate the health effects that are associated with a particular release scenario. The results from the forward model are then used to formulate the inverse problem in which the estimated health outcomes are assumed to be the known information, while the characteristics of the source that are responsible for the observed outcomes need to be back-calculated. The forward modeling system combines forward dispersion simulations with dose-response simulations. In the dispersion modeling stage, an atmospheric dispersion model is used to generate estimates of the downwind concentrations of the chemical agent for a known emission rate and location of a point source. Dose-response models are then used to convert the chemical agent's time histories into health effect predictions across different locations downwind from the release location. A sample of the results generated via the forward model is used as input into the inverse algorithm, in which health outcomes are back-transformed to the potential environmental exposure. The solution of the inverse problem is obtained based on a forward modelling scheme that employs Monte Carlo sampling. The inverse solution derives a posterior estimate of the source parameters based on optimization methods that use guesses of the emission rate and location coordinates as a prior. The inverse algorithm is based on minimizing the deviation between the model

predictions and the actual observations obtained from the synthetic release scenario, which will eventually lead to finding an optimal solution of the release characteristics. Figure 6 summarizes the approach followed in reconstructing the source term characteristics. The details of each of the steps are discussed in the following subsections.

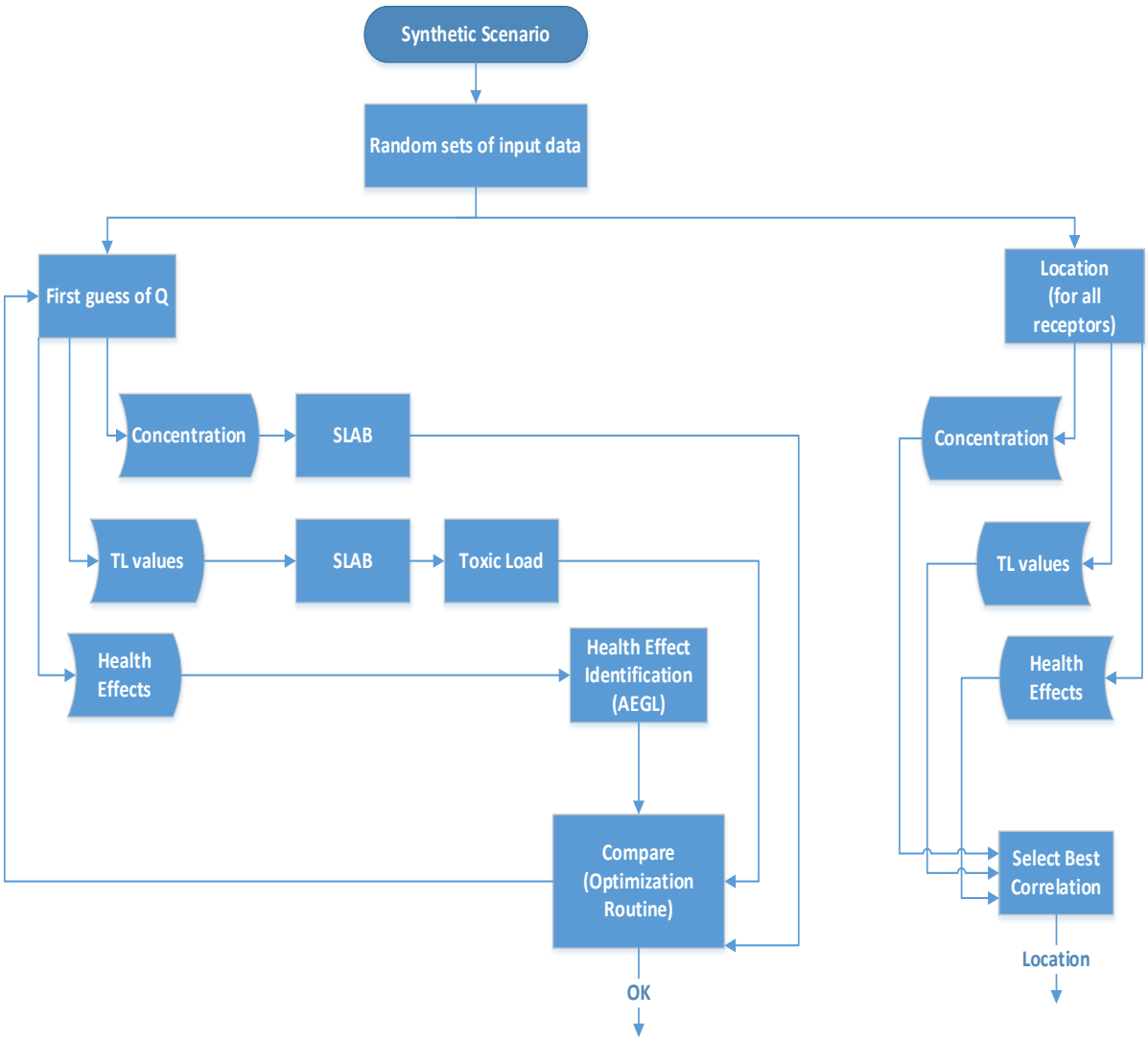

Figure 6. Schematic representation of the proposed methodology for source term reconstruction

3.2 Development of the Forward Model

3.2.1 Atmospheric Dispersion Modeling Using the SLAB Model

Atmospheric releases of hazardous agents can typically be modeled using atmospheric dispersion models, which predict the spread and movement of the cloud of such contaminants. Depending on the complexity of the dispersion model, certain data (e.g. weather conditions, and description of the nature of the site) are needed to facilitate the simulations. In this work, the atmospheric dispersion of the hazardous agent was modelled using the U.S. Environmental Protection Agency's (USEPA) SLAB model (Ermak, 1990) for the case of a vertical jet source release. This model has been used in several studies involving denser-than-air atmospheric releases and simulations, and is widely employed in a number of applications involving accidental releases (e.g. Bhosale et al., 2015; Ni et al., 2014; and Hanna et al., 2012). The results produced by the SLAB model are based on the solution of the one-dimensional (in downwind distance) equations of momentum, conservation of mass and energy, and the equation of state. The SLAB model supports the simulation of the atmospheric transport and dispersion of four different area source releases: ground-level evaporating pool, horizontal and vertical jets at any height, and instantaneous volume sources. In addition, this model can be applied to continuous and instantaneous releases, and accounts for user-defined finite duration release times as well as the effects of averaging times. Moreover, SLAB does not calculate the source emission rate, and hence the source conditions need to be provided as input by the user.

The input data required for the model include the source type, source rate, properties of the source, spill and field properties, meteorological parameters, as well as a numerical substep parameter. SLAB generates an output file containing distributions with downwind distance of the maximum pollutant concentration, the time when the maximum concentration occurs, the

time duration of the cloud, and the parameters that describe cloud geometry (width and depth). The most important output from SLAB is the time-averaged volume concentration, which is of interest to us in this work. The time-averaged volume concentration of a chemical agent is given as a function of the travel time from the source (t), in addition to the three spatial dimensions: downwind distance from the source (x), crosswind distance from the cloud centerline (y), and height above the ground level (z). For a vertical jet source, the time-averaged volume concentration is given by the following equation, which computes the atmospheric concentration C at a particular point (x, y, z) downwind from the release location (Ermak, 1990):

$$C\ (x,y,z,t) = \ cc(x) * (erf(xa) - erf(xb)) * \ (erf(ya) - erf(yb)) * \ (exp(-za^2) + exp(-zb^2)) \quad (3.1)$$

where:

$$ya = \ (y + b(x))/(sr2 * betac(x)) \quad (3.2)$$

$$yb \ = \ (y - b(x))/(sr2 * betac(x)) \quad (3.3)$$

$$za \ = \ (z - zc(x))/(sr2 * sig(x)) \quad (3.4)$$

$$zb = \ (z + zc(x))/(sr2 * sig(x)) \quad (3.5)$$

$$xa = (x - xc(t) + bx(t))/(sr2 * betax(t)) \quad (3.6)$$

$$xb \ = \ (x - xc(t) - bx(t))/(sr2 * betax(t)) \quad (3.7)$$

$$sr2 = \sqrt{2} \tag{3.8}$$

and where *erf* is the error function and *exp* is the exponential function.

For the purpose of this work, the volume fractions (in units of ppm as reported in the USEPA tables) are converted into units of mg/m3 as required for the toxic load calculations, which are discussed next. Further, the mathematical formulations provided by the SLAB model are programmed in MATLAB's m-files in order to facilitate the dispersion simulations.

3.2.2 Dose-Response Modeling: Evaluation of the Onset of the USEPA's AEGL-1,2, and 3 Conditions Using EAGLE Software

Adverse health effects can be the result of an inhalation exposure to toxic airborne contaminants. The USEPA has established three distinct Acute Exposure Guideline Levels (AEGLs) for a number of chemical agents in order to facilitate the prediction of the onset of such adverse health effects in a general population exposed to a specific concentration of a toxic agent for a fixed exposure duration (U.S. Environmental Protection Agency AEGL Program, 2012). However, in an actual release, inhalation exposures may vary with durations, which makes it a challenge to estimate the onset of the health effects using the fixed values provided by USEPA's AEGLs. To resolve this issue, the U.S. Naval Research Laboratory has recently developed and published a science-based software/algorithm that is fast and easy-to-use (Boris & Patnaik, 2014). The software, which is known as EAGLE, can be used in conjunction with a dispersion model to enable the prediction of the onset of AEGL 1 (notable discomfort), AEGL 2 (irreversible or serious adverse impact that could impair the ability to escape) and AEGL 3 (life threatening or death) conditions for time-varying plumes involving chemical agents with tabulated AEGLs.

Initiating the computations of this software does not require the availability of prior knowledge of the exposure history, hence it allows in situ evaluation as the plume is evolving. This, however, requires storing one, two, or three additional variables; the evolving toxic loads at each point where a result is desired. Further, the algorithm is designed to enable the use of realistic concentration time history of a contaminant as they are encountered. Therefore, hazard areas can be updated quickly and incorporated into dispersion models or in emergency incidents.

The software algorithm implements the Induction Parameter Model, which was developed in 1980 for fast combustion (Boris & Oran, 2000). As the software's application of the Induction Parameter Model is aimed at potentially toxic plumes, the nonlinear ten Berge generalization of Haber's law (ten Berge et al., 1986; Stage, 2004) is used to obtain the induction parameter. The onset of the health conditions is determined using the toxic load (TL), which is integrated from zero using the concentration time history of the released contaminant. The time-dependent toxic load is defined by the following integral:

$$TL_k(t) = \int_0^t \left[\frac{dTL_k}{dt'}\right] dt' = \int_0^t TLrate(t')dt' \tag{3.9}$$

When the above integrated toxic load, induction parameter in this application, reaches unity, the corresponding AEGL conditions have occurred, where the subscript k=1, 2, 3, is a particular AEGL condition. Moreover, the algorithm defines a power law that accounts for the rate of accumulation of toxic load spanning the time intervals for each of the three AEGLs. Thus, the toxic load is computed by interpolating the instantaneous concentrations. For instance, a toxic load rate of a given chemical agent for a time duration falling between the 60-min band and the 4-hour band can be computed as follows:

$$TLrate(t) = \frac{dTL}{dt} = \frac{1}{14400s}\left[\frac{C(t)}{C_{4hr}}\right]^{\alpha} \quad (3.10)$$

where:

α is a power law defined for each of the individual AEGL bands

C(t) is the instantaneous concentration at the point in question

C_{4hr} is the USEPA concentration corresponding to the 4 hour duration for a particular AEGL

In this work, the algorithm provided by EAGLE is implemented in MATLAB to evaluate the toxic loads and the corresponding health effects resulting from exposure to Hydrogen Sulphide (H_2S) gas. H_2S is selected due to its toxic nature and the health effects associated with that, in addition to being a well-known industrial hazard (Agency for Toxic Substances and Disease Registry, 2006; Guidotti, 1996).

3.3 Formulation of the Inverse Problem and the Inversion Method

We now consider the inverse problem of determining the source term characteristics from observed health symptoms. Given the observed health symptoms at different downwind locations, the aim now is to find distributions of the source term parameters (i.e. unknown source term in the dispersion model and its location) that will produce health effect predictions consistent with those being observed on site.

The problem is formulated using a Cartesian coordinate system (X, Y, Z) in which the X-axis is aligned parallel to the direction of the wind, the Y-axis is oriented horizontally in the cross-wind direction, and the Z-axis in the vertical direction. The release of the HazMat (i.e. H_2S gas) is assumed to take place at time t=0 from a single point source located at (Xs, Ys, 0),

with a steady emission rate Qs (kg/s). The released material is blown by a wind with mean velocity u=(X, 0, 0). The resulting downstream mass concentration C(X, Y, Z, t) of the cloud of the released material is determined using the forward dispersion model. An illustration of the setup of the inverse problem is shown in Figure 7.

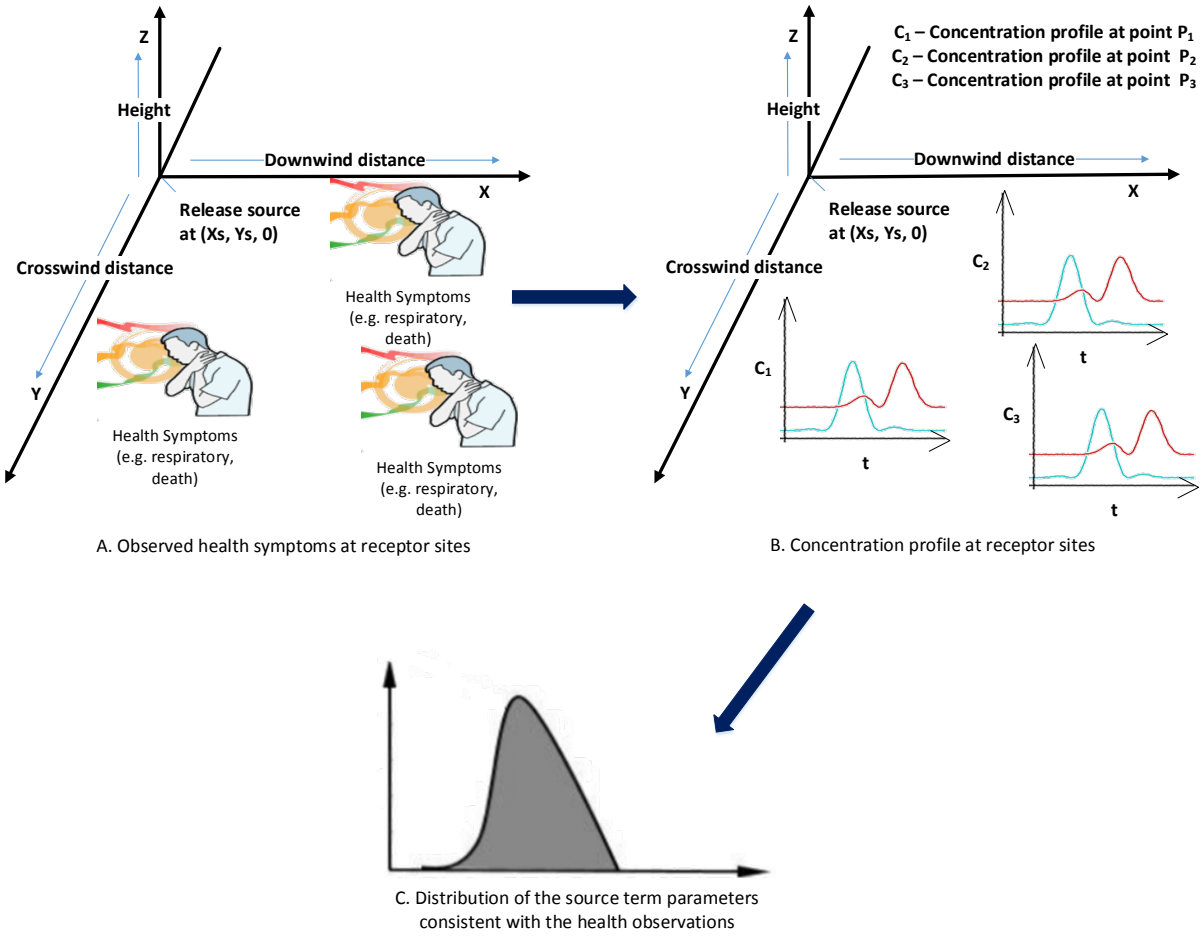

Figure 7. Illustration of the inverse problem

For an unknown source release rate Qs and/or unknown release location (i.e. Xs and Ys), health observations are recorded at multiple receptors (X, Y, 2). The coordinates X and Y are chosen so that the receptors' locations are located downwind of the release point and at a breathing elevation of 2 m as defined by SLAB for the case of a vertical jet release. Also, wind is taken to be constant over the time interval of length Delta t. In practice, the wind velocity varies with time and these variations have a significant impact on the plume dispersion and consequently the toxic load values. As mentioned before, the occurrence of the health effects is linked to the values of the toxic load, where a value of 1 and above implies that the specified health effects are observed at the receptor location.

An optimal estimate of the source rate is obtained through the implementation of an optimization routine that is based on minimizing a cost function. The cost function is minimized according to the difference (normalized residual) between the model predictions and the observations as follows:

$$Rsq = \frac{\sum(Observation - Model\ Prediction)^2}{\sum(Observation - mean(Observation))^2} \tag{3.11}$$

Finding the minimum of the cost function then reduces to making a first guess for Qs and then adjusting the solution iteratively until the estimate converges. The best estimate of the release rate is the value of Qs at convergence.

To obtain a best estimate of the release location, we employ the correlation coefficient method. The correlation coefficient method defines a cost function J:

$$J = \frac{\langle (c^c - \langle c^c \rangle)(c^0 - \langle c^0 \rangle) \rangle}{\sqrt{\langle (c^c - \langle c^c \rangle)^2 \rangle}\sqrt{\langle (c^0 - \langle c^0 \rangle)^2 \rangle}} \rightarrow min \tag{3.12}$$

where ‹› denotes arithmetic averaging over all measurements, while the minimum is sought over all possible source locations. The above cost function is minimized with respect to the source coordinates only, while an arbitrary source rate Qs is used in the procedure of this minimization. It should be noted that since the source rate is excluded from the cost function, the solution obtained is independent of the choice of the source rate. In this algorithm, the source coordinates are assumed to coincide with one of the grid nodes. The location of the corresponding grid node for which the minimum of the correlation coefficient is achieved is considered to be the identified source location.

3.4 Computational Implementation

We begin by defining a 5000 m x 1000 m domain area over which the dispersion calculations, toxic load simulations, and the subsequent inverse calculations are performed. The domain area is then divided into a 200 x 40 grid, with each cell being a square of 25 m x 25 m. The gridpoints begin at 0 m and extend to a downwind distance of 5000 m, and 1000 m in the positive crosswind direction. Further, the release is assumed to take place from a source at ground level (i.e. it is located at 0 m height in the vertical direction). In all of the scenarios considered in this work, the actual location of the source is taken to be at (500, 300, 0) in the Cartesian coordinates, while the release rate of the HazMat is set equal to 100 kg/s for a release duration of 3,440 seconds (approximately 1 hour). The release is assumed to take place in an industrial natural gas facility, from a pipeline transporting natural gas that contains 4% H_2S.

3.4.1 The Forward Model

The forward computations are initiated by the atmospheric dispersion model, SLAB, which predicts mass concentrations of H_2S as a function of time at each gridpoint, and at a breathing zone at 2 m elevation above the ground. The inputs to the dispersion model include

the domain area, source rate and location, as well as the time-steps over which the model calculates the downwind concentrations. Noise is then introduced to the synthetic concentration data in order to verify the robustness of the inversion algorithm under less ideal, but perhaps more realistic conditions. In the literature of atmospheric dispersion modeling (COST ESSEM ES1006, 2015), it is shown that most semi-empirical atmospheric dispersion models (e.g. SLAB) perform better than 40%-50% when compared with measured data. Therefore, we have introduced a random noise spanning -20% to +20% at all SLAB calculations; applied to the modeling stage and the reference case of the synthetic release scenario in this work. This noise accounts also for the atmospheric conditions variability. Additionally, all the source parameters are assumed to be constant over time (e.g. Qs and Us). Also, the geometry of the site at which the release takes place is characterized as a flat terrain in an open environment since the SLAB model assumes a flat terrain that does not involve any obstacles (e.g. buildings and trees).

It should be noted that with the existing version of SLAB, different model runs must be conducted for every single change in the input parameters. This is both computationally extensive and impractical, especially in real time situations, where running a large number of simulations quickly is desired in order to provide rapid information that are needed for planning and emergency response purposes. Hence, the SLAB calculations have been modified to avoid multiple model runs for the different guesses of the source rate. Here, we have defined four different source rates, pre-calculated their parameters from the original SLAB model, and used them as bands for interpolating the parameters of the new guesses of the source rate. As expected, the results of the interpolation are associated with some error when compared to the actual model predictions.

We then proceed with the toxic load calculations for the same 61 output times from SLAB and locations in the area affected by the H_2S cloud. The TL calculations of EAGLE require providing tabulated USEPA values of AEGLs for the chemical of concern, H_2S in this case. Next, a power law of 4.4 is obtained by fitting the AEGL data provided by the USEPA for H_2S. The calculations also require defining a lower and upper time limits over which the USEPA table is extrapolated and extended. Here, we have defined the lower and upper limits to be 30 seconds and 24 hours, respectively. The outputs from the EAGLE's TL calculations are TL values corresponding to the three different AEGLs at various downwind locations and time intervals following the HazMat release.

3.4.2 The Inverse Problem

In the inverse calculations, we are primarily interested in finding the source term characteristics through the use of various combinations of input data and release conditions. To do this, six different cases are tested in this work to reconstruct the source release rate from a known location, and to identify the location of the release when the source rate is assumed to be known.

3.4.2.1 Identification of an Unknown Release Rate

The objective of the first scenario is to determine the release rate of a HazMat emitted from a known source location. Three different classes of data are used as input to the inversion algorithm each time. In the first case, the release rate is back-calculated based on the conventional use of concentration data. The second case utilizes exact receptor measurements of TL reported at specific time-intervals and meeting the criterion that a value of a TL greater than zero is detected (i.e. a release triggers the toxic load computations). The use of the health effects as input is tested in the last case in which health observations are collected at specific

times following the release, with the criterion that at each receptor AEGLk conditions are present. The input data to the modeling system in this case is the inequality TL>=1.

For each of the scenarios considered, random samples of observations (concentration, TL values, or health effects) from the forward model are collected at different candidate locations (xi, yi). MC simulations are performed to generate samples of varying size of observations. The sampled results are then passed into the inverse routine in which they are compared in a cost function against model predictions using different guesses of Qs. The inverse routine computes iteratively Qs along with the residual error associated with each set of observations.

3.4.2.2 Identification of an Unknown Source Location

In the second case, the objective is to identify the unknown location of the source from a series of different observations. Similar to the first case, the location is identified using three different types of input data (i.e. TL values, health observations, and concentration measurements). As discussed earlier, the solution for the source location is obtained through the implementation of the correlation coefficient method, where the minimum of the correlation function corresponds to the identified location.

Atmospheric dispersion and dose-response simulations were run using MATLAB m-files scripts, and samples were collected after convergence was achieved. Moreover, the inversion computations were carried out using MATLAB's fminbnd and corrcoef functions. The code that generates the results consists of the following main routines (see the Appendix for details):

- Read_Predict: Calls all the values of the SLAB parameters for the pre-defined source rates in order to facilitate the interpolation calculations.
- LinIterm: Performs the interpolation calculations for the newly guessed values of the source rate using the parameters in Read_Predict.

- SLAB: Calculates the downwind concentration using the 1D dispersion equation described in 3.2.1.
- Solve: Main program that generates random permutation samples for use as input to the inverse algorithm, and then estimates the release rate based on the use of MATLAB's fminbnd function.
- Solve_Loc: Using this routine, we obtain an estimate of the location coordinates using the correlation coefficient method and MATLAB's corrcoef function.

4. RESULTS AND DISCUSSION

In order to demonstrate the capability of the proposed scheme in predicting the source parameters, calculations were performed following the release of a toxic H_2S gas for conditions where the three different AEGLs take place. The sensitivity of the source term estimation is evaluated against several parameters, namely the type of input information (i.e. concentration, TL values, and health observations), the number of receptors utilized in the inversion and the number of iterations used in generating the results. Multiple tests and model runs were conducted for the different AEGLs and time-steps, however, we present here only selected examples (i.e. the case where AEGL-1 conditions are met and for a specific time-step) to illustrate the results of the source-reconstruction methodology in the following subsections.

4.1 Forward Modeling Results

4.1.1 Concentration Profile at Different Time-steps

Figure 8 shows the profile of the H_2S concentration at some arbitrary receptors downwind from the release location, and at different timesteps following the toxic release.

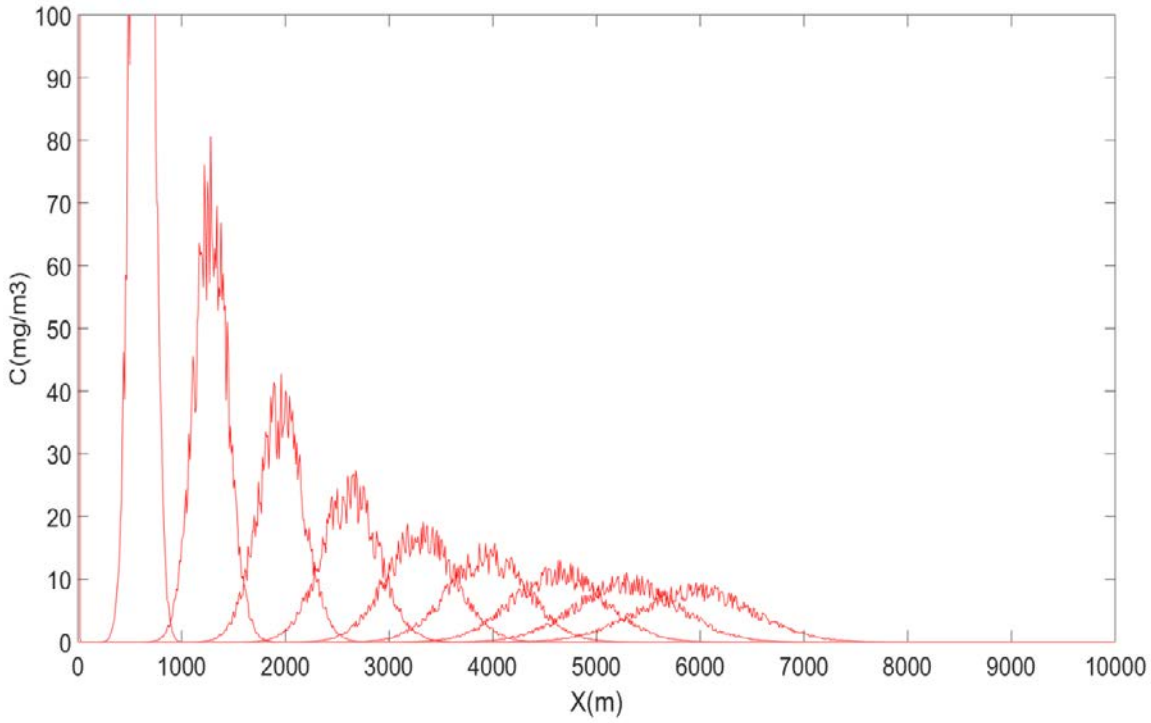

Figure 8. Concentration profile at different time intervals (from left to right: 201s, 401s, 601s, 801s, 1001s, 1201s, 1401s, 1601s, 1801s), downwind from the release location

Concentration profiles have higher peaks (corresponding to high concentrations) at the early stages of the release and cover a smaller downwind distance away from the point of the release. The peak is then seen to decrease with time as the plume disperses and the profile widens to cover larger distances downwind from the release. The time intervals are recorded following the start of the release (i.e. t= 0 seconds), with a time-interval of 201 seconds for instance being the length of time that passed since that. Also, noise is introduced to the synthetic concentration data as discussed in the methodology section.

4.1.2 Toxic Load Profile at Different Downwind Locations

Figure 9 shows the TL profile for the three AEGLs at some arbitrary receptors downwind from the release location. AEGL-1 being the less severe conditions and AEGL-3 conditions being the most severe ones.

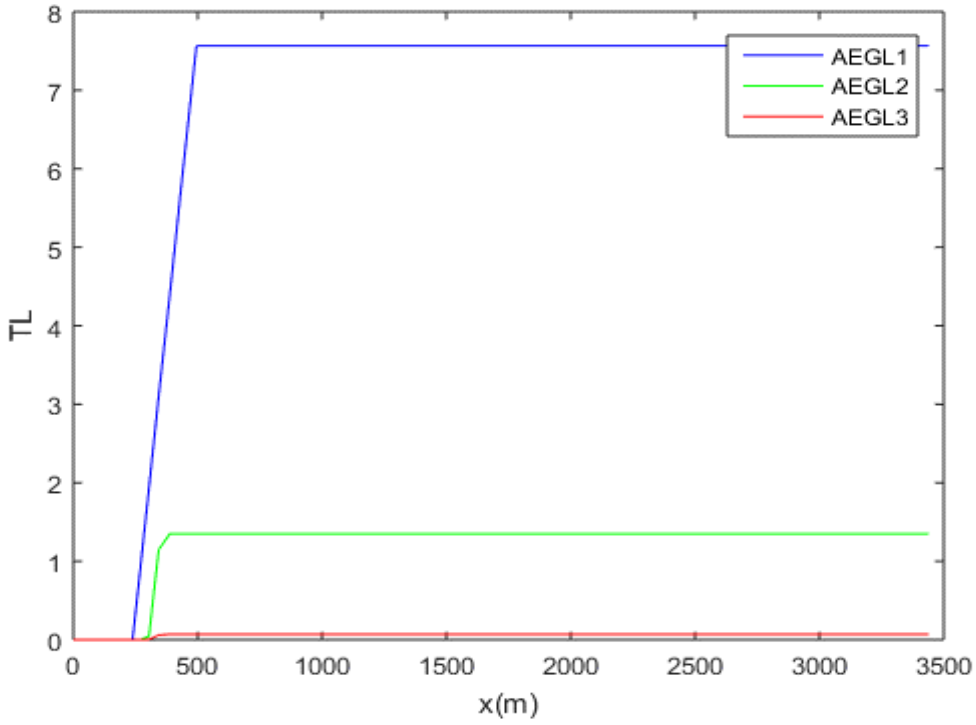

Figure 9. TL at some arbitrary locations downwind from the release location

As shown, both AEGL-1 and AEGL-2 conditions are met (i.e. TL values equal to or greater than 1) and hence indicating the occurrence of the corresponding health observations, whereas AEGL-3 conditions are not achieved for the given exposure conditions and duration.

4.1.3 Regions with Health Effects Observations at a Fixed Timestep

Figure 10 depicts the regions with health effects observations resulting from exposure to a toxic H_2S cloud at a time t=600 seconds. As seen, the affected areas extend up to approximately 2800 m horizontally and 600 m vertically in the domain area, with AEGL-1 conditions being the dominant.

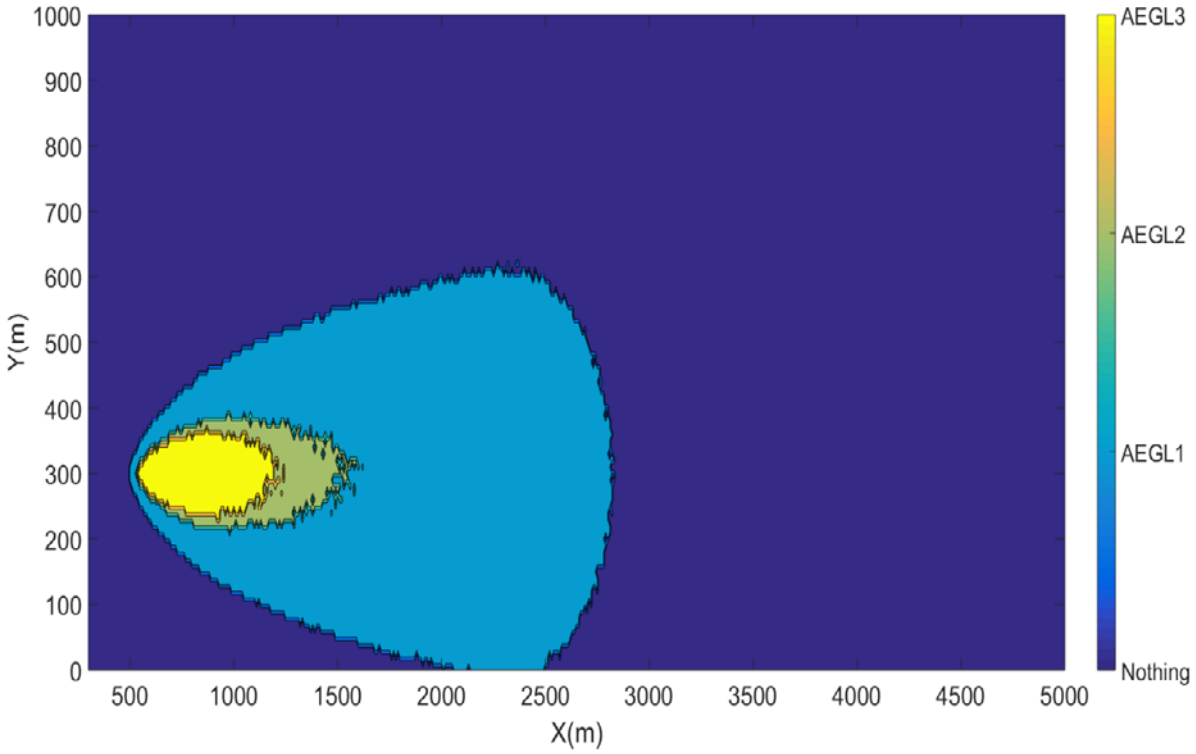

Figure 10. Regions with health symptoms observations at t=600 seconds

4.2 Results of the Inverse Calculations

4.2.1 Unknown Source Rate (Qs)

The source rate was estimated using different sets of receptors ranging from 2 to 100 sample points, and model runs were performed for two different iterations (i.e. 10 and 100) in order to investigate the impact of increasing the magnitude on the convergence of the results. Moreover, the frequencies of the estimates and their corresponding residuals are compared for the case of using 9 receptors and 100 receptors, for 100 iterations in both cases.

4.2.1.1 Qs using Concentration Data

The estimated source rate for a varying size of receptors using two different iterations is shown in Figure 11 for the case of using concentration as input data to the source reconstruction algorithm. As shown, both the size of the receptors used and the number of iterations influence the convergence of the results. For example, using 9 receptors and 100 iterations, the results are shown to range from approximately 80-200 kg/s, with a moving average of 140 kg/s. Whereas, using a sample of 100 receptors for the same number of iterations, the results are shown to exhibit lesser variability and are confined to a range of approximately 88-113 kg/s, with a moving average of 103 kg/s. Further, the number of iterations used seems to have insignificant impact on the results for receptors sizes of 58 and above.

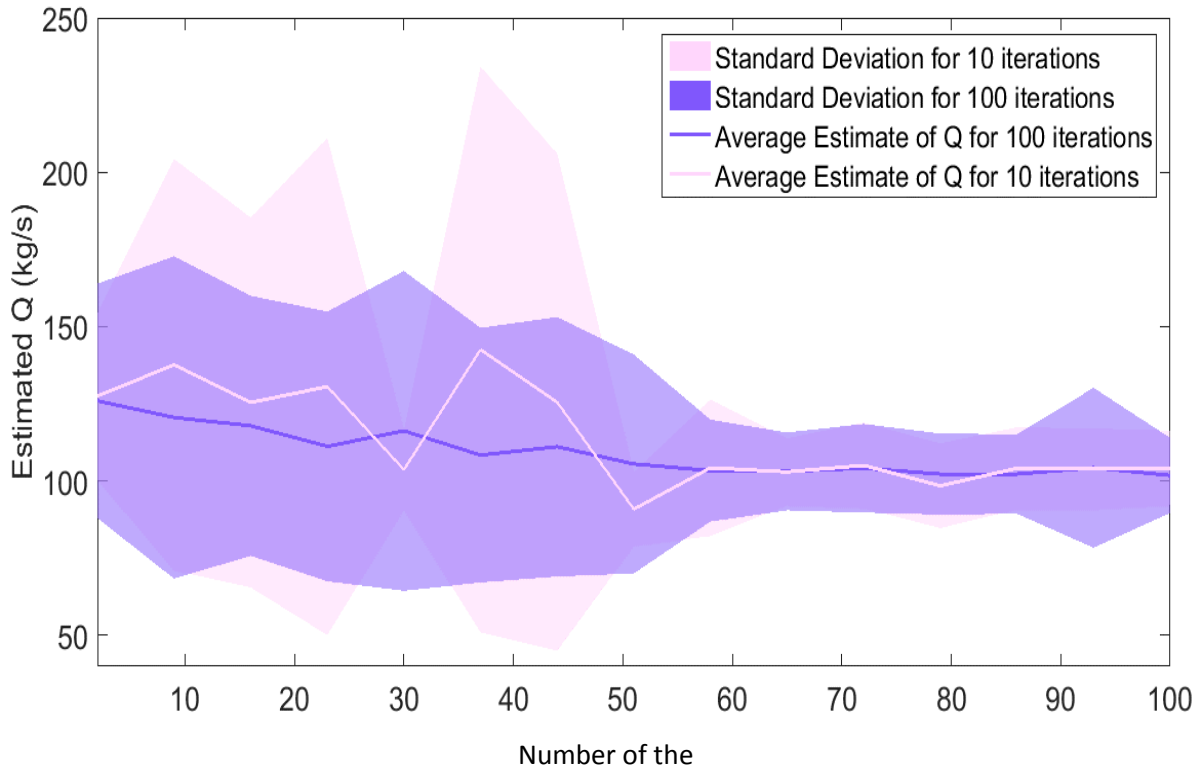

Figure 11. Estimated Qs versus the size of the receptors for 10 and 100 iterations based on the use of concentration as input

Next, the frequencies of the estimated results and their residual errors were estimated for a fixed number of iterations and two different sizes of receptors. Figure 12 and Figure 14 show the frequencies of the source rates estimates when utilizing concentration data and 100 iteration for 9 and 100 receptor points, respectively. The frequency distribution of the corresponding residual errors are shown in Figure 13 and Figure 15.

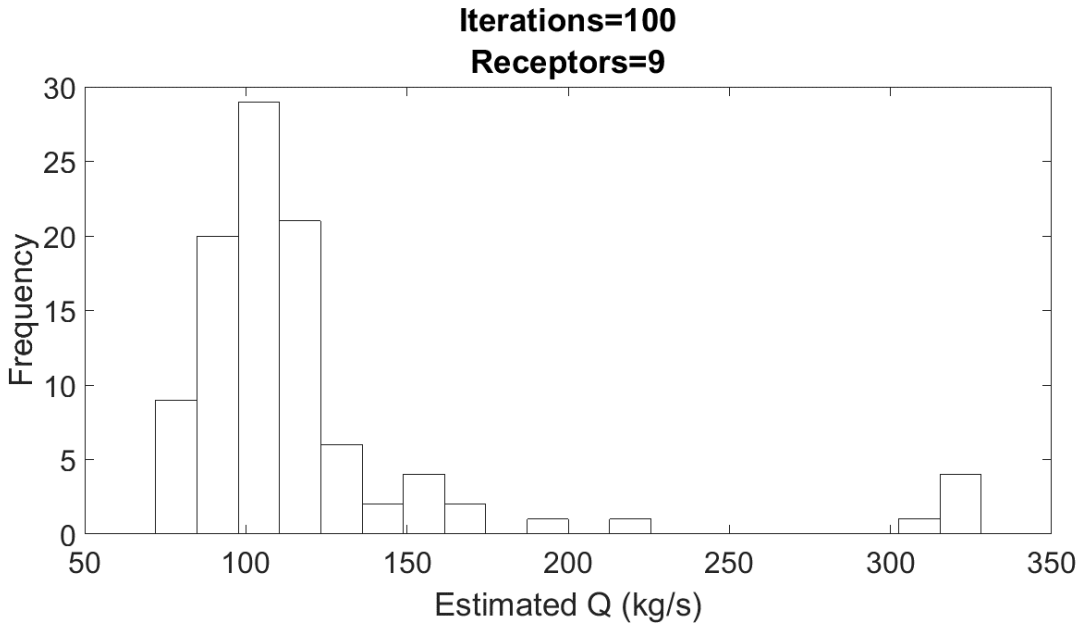

Figure 12. Estimated Qs versus the frequency for 9 receptors when using concentration as the input

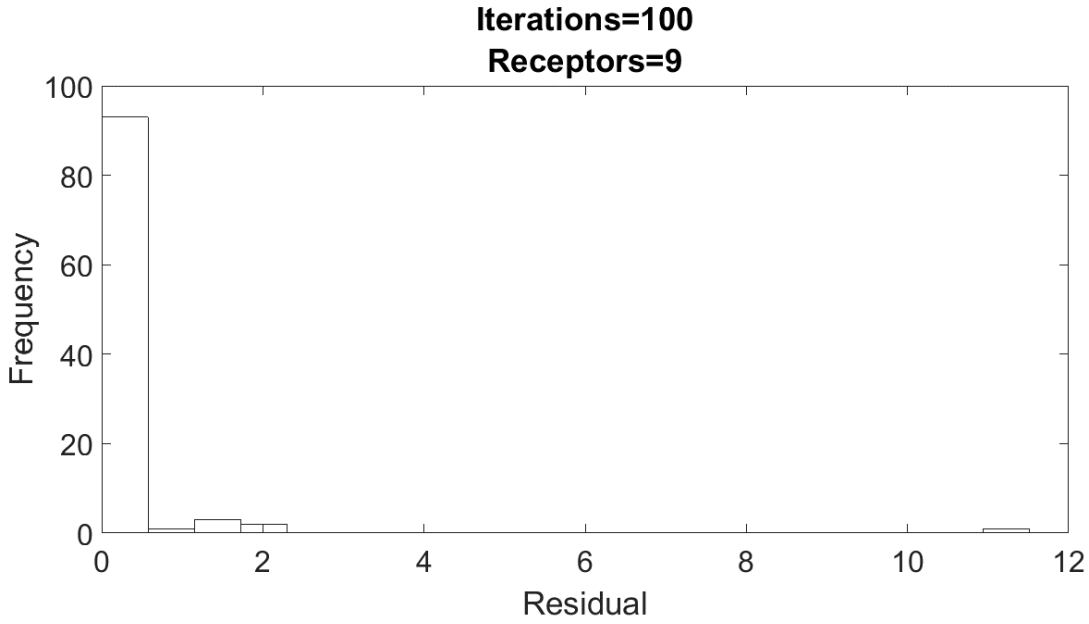

Figure 13. Residual error versus the frequency for 9 receptors when using concentration as the input

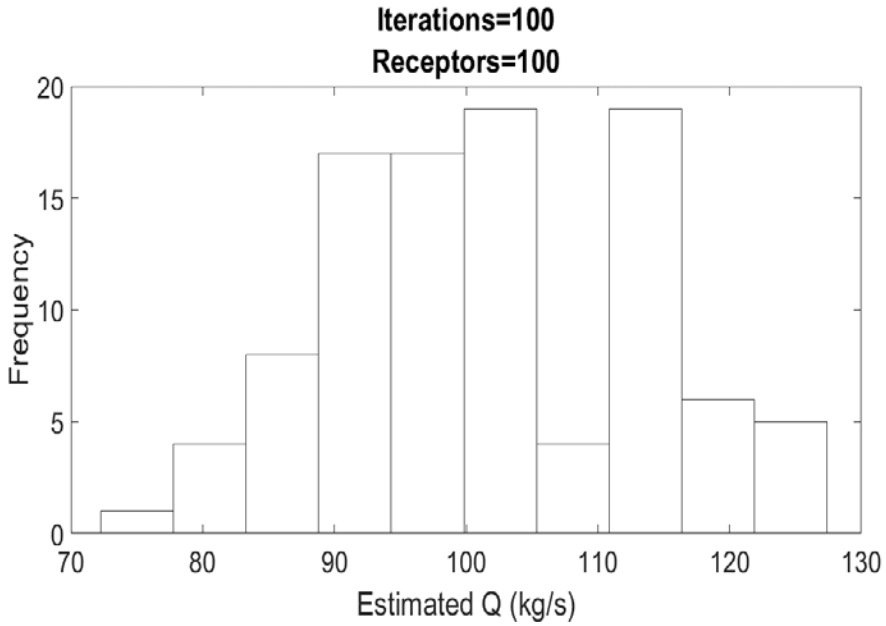

Figure 14. Estimated Qs versus the frequency for 100 receptors when using concentration as the input

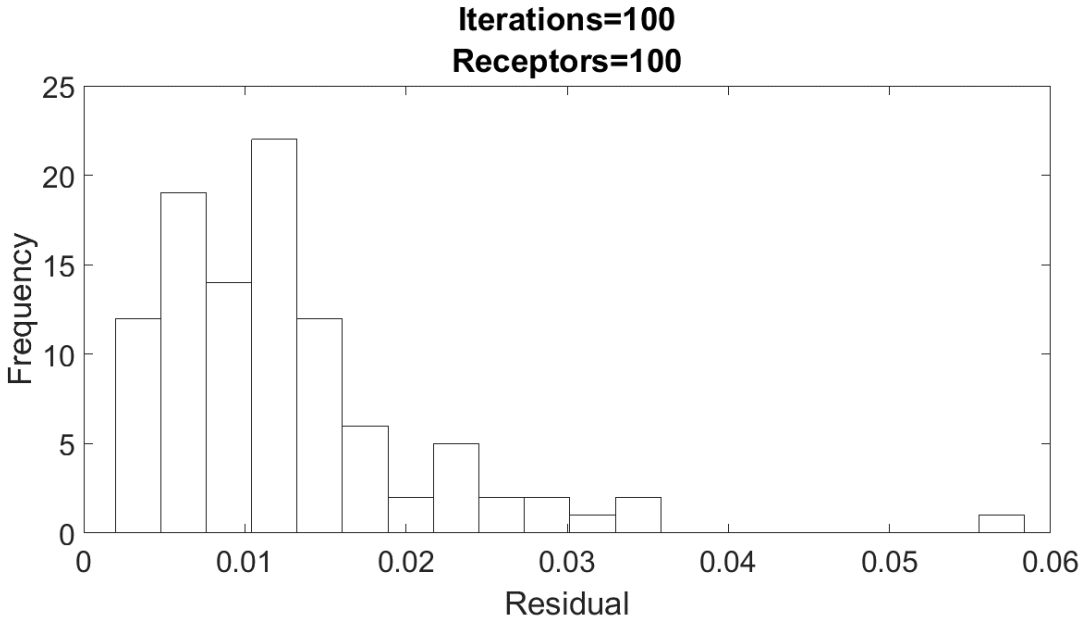

Figure 15. Residual error versus the frequency for 100 receptors when using concentration as the input

As shown in Figure 12, the estimated results of the source rate when using 9 receptors, fall between 72 kg/s and 328 kg/s, with the majority of the estimates being concentrated in the 113.3-116.6 region. Increasing the number of receptor points is shown to result in better convergence of the source rate estimates (see Figure 14), with the results falling in the 72-127 kg/s range, and majority of the estimates being close to the true solution.

4.2.1.2 Qs using TL Values

Figure 16 shows the estimated source rate for the case of using TL values as input data to the source reconstruction algorithm. Interestingly, the results using this type of information have better convergence when compared to the use of concentration data.

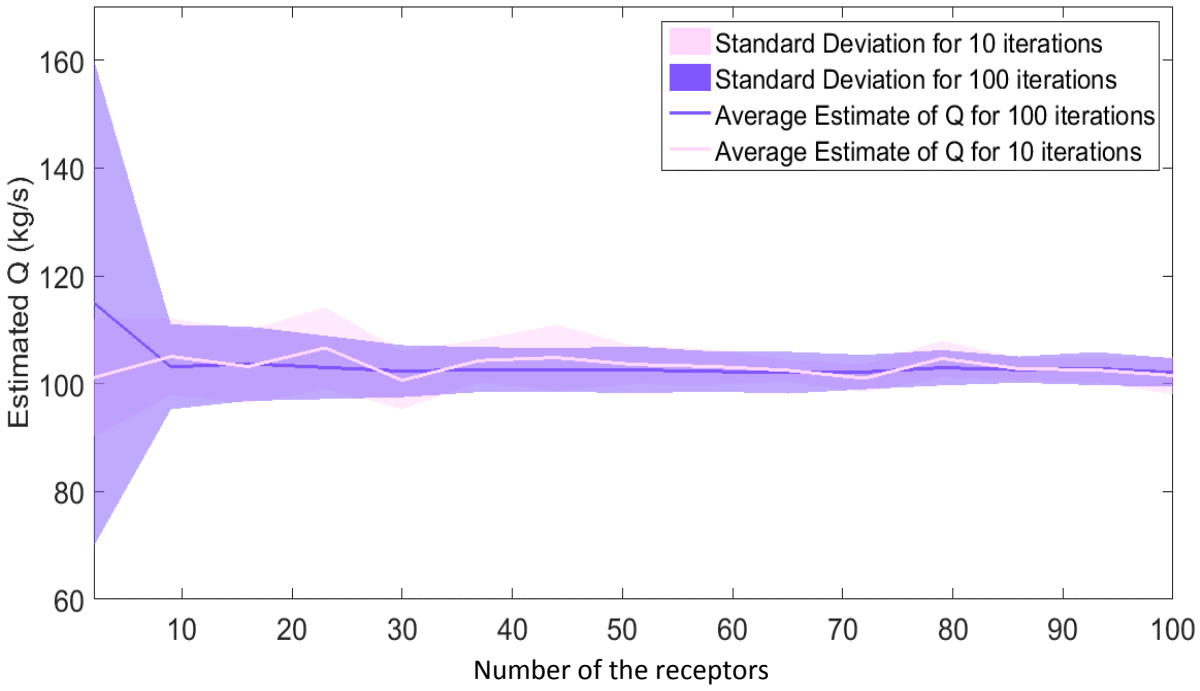

Figure 16. Estimated Qs versus the size of the receptors for 10 and 100 iterations based on the use of TL as input

The frequencies of the estimated source rate and the corresponding residual errors for 100 iterations and two sizes of receptor points are shown in Figures 17-20.

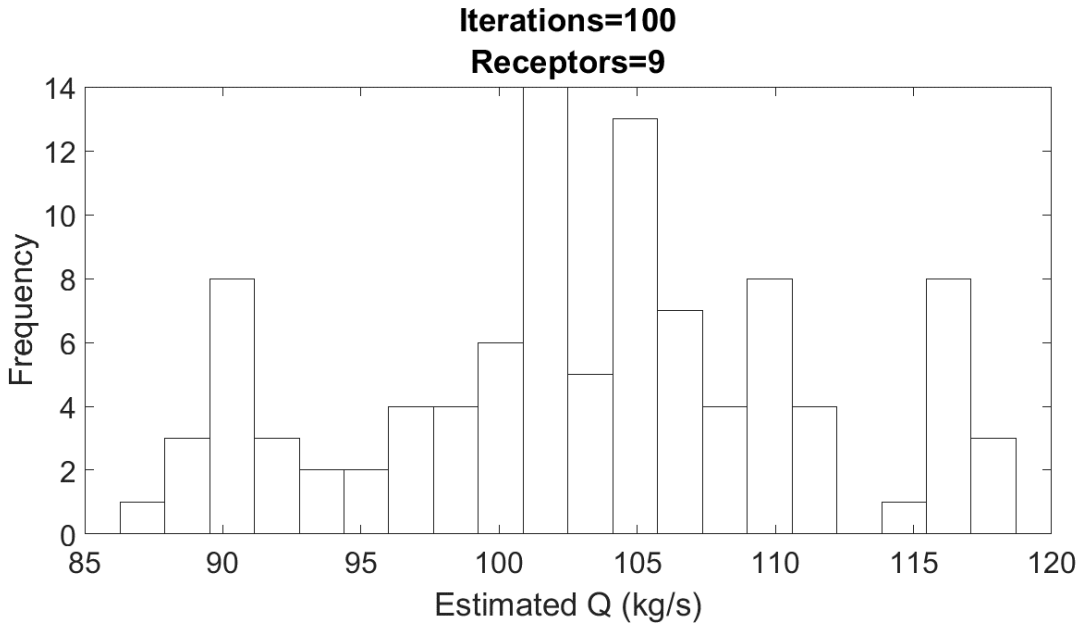

Figure 17. Estimated Qs versus the frequency for 9 receptors when using TL as the input

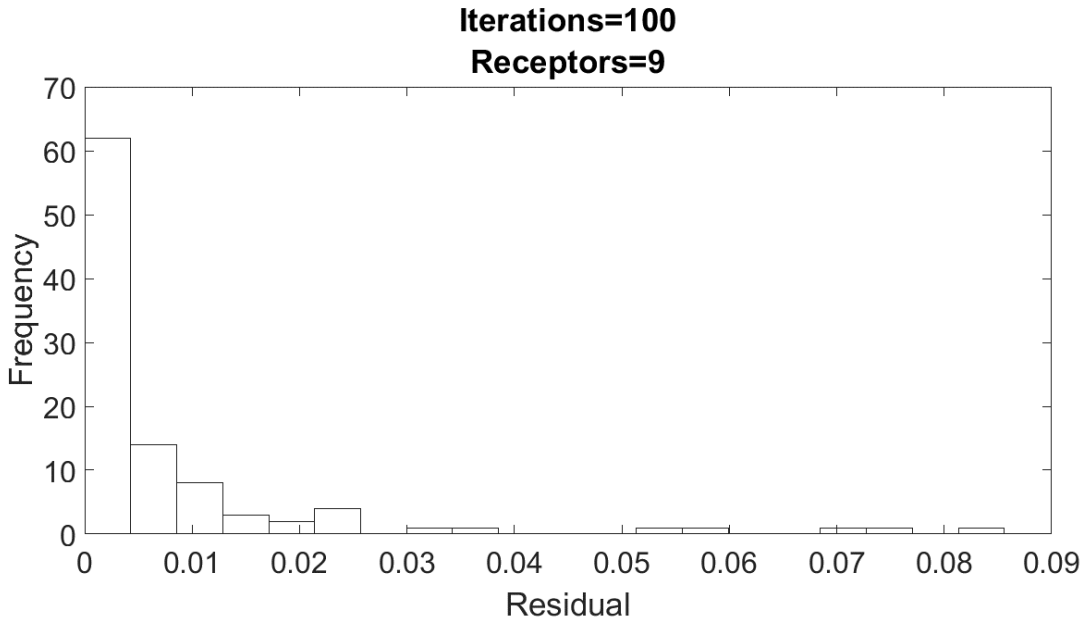

Figure 18. Residual error versus the frequency for 9 receptors when using TL as the input

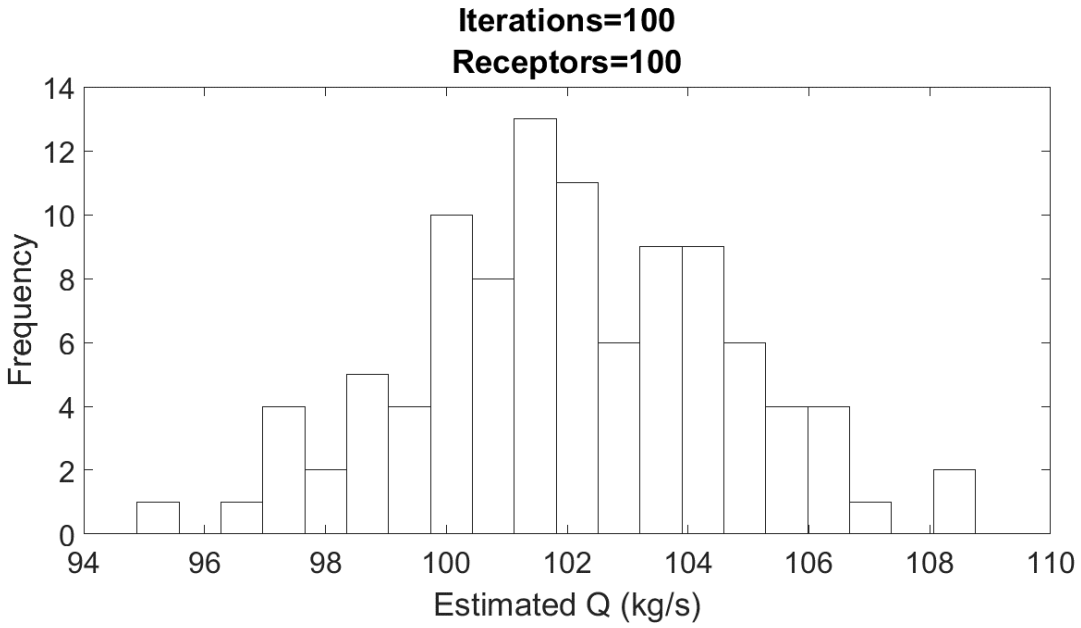

Figure 19. Estimated Qs versus the frequency for 100 receptors when using TL as the input

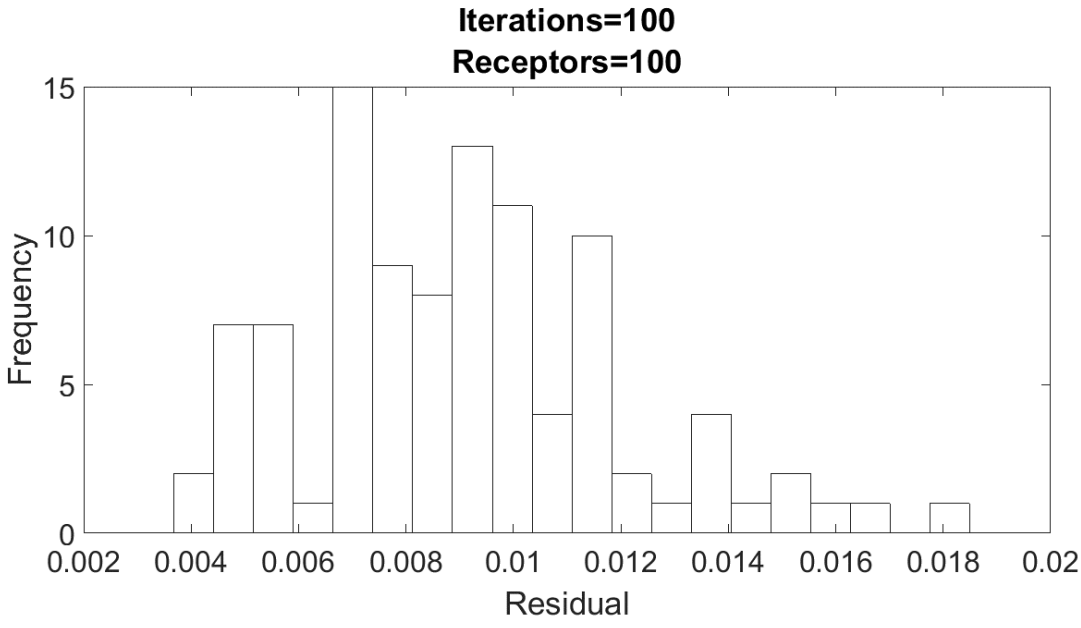

Figure 20. Residual error versus the frequency for 100 receptors when using TL as the input

As shown in Figure 17 and Figure 19, the estimated results of the source rate when using 9 receptors, lie between 86.28 kg/s and 118.71 kg/s, with the majority of the estimates being concentrated in the 101.7-103.4 region. Increasing the number of receptor points does not result in significance improvement in terms of convergence, with the results falling in the 95-108 kg/s range, and majority of the estimates being in the 101.34-102 kg/s region.

4.2.1.3 Qs using Health Effects

Lastly, the source rate was estimated from health effects as the input data. The results of the source estimation are shown in Figure 21. Unlike the use of concentration and TL data, the complexity associated with use of health effects information influences the convergence of the source estimates. As seen, for the two different number of iterations, the results fail to reach reasonable convergence when using small number of sample points and are shown to exhibit greater variations. The results then improve when using a sample size of 44 points and above for both of the iterations used. Also, using smaller number of iterations, the results converge faster using smaller number of points in contrast to the use of the larger number of iterations for which a larger size of receptors is required to reach the same convergence. This is only observed up till 44 points and is a result of the added error. Since model runs represent different scenarios and since health effects are treated as an integrative; it takes values of 0, 1, 2, and 3 and treats 2.1, 2.2, and 2.3 the same. Here, using 100 iterations is like adding another 100 iterations/scenarios and hence adding to the existing error.

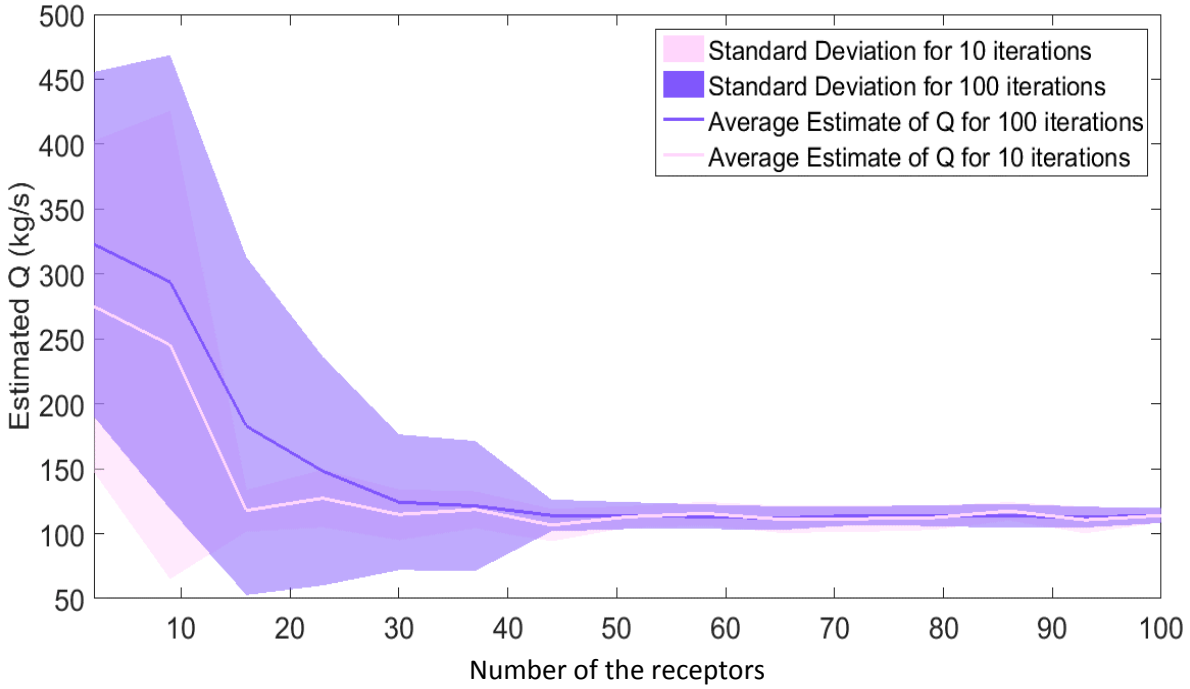

Figure 21. Estimated Qs versus the size of the receptors for 10 and 100 iterations based on the use of health effects as input

The frequencies of the source rate estimates and the corresponding residual errors for 100 iterations and two sizes of receptor points are shown in Figures 22-25.

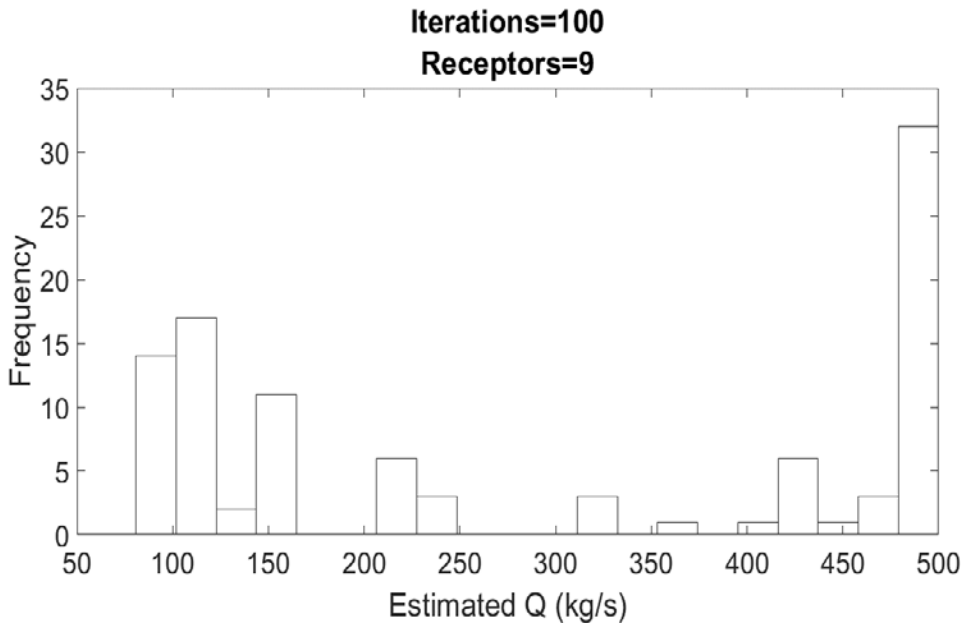

Figure 22. Estimated Qs versus the frequency for 9 receptors when using health effects as the input

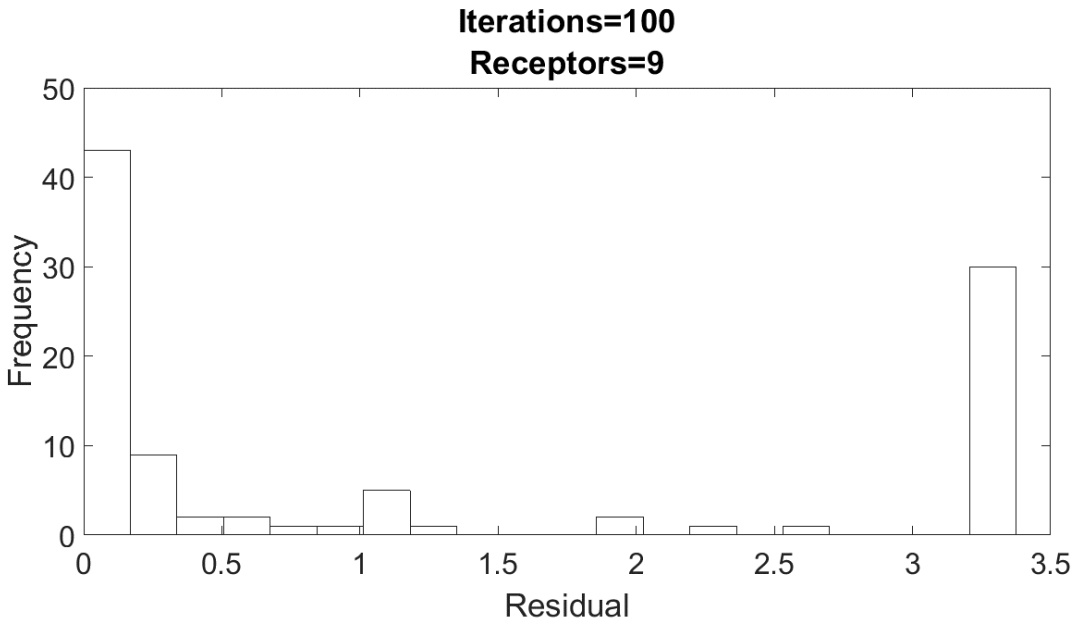

Figure 23. Residual error versus the frequency for 9 receptors when using health effects as the input

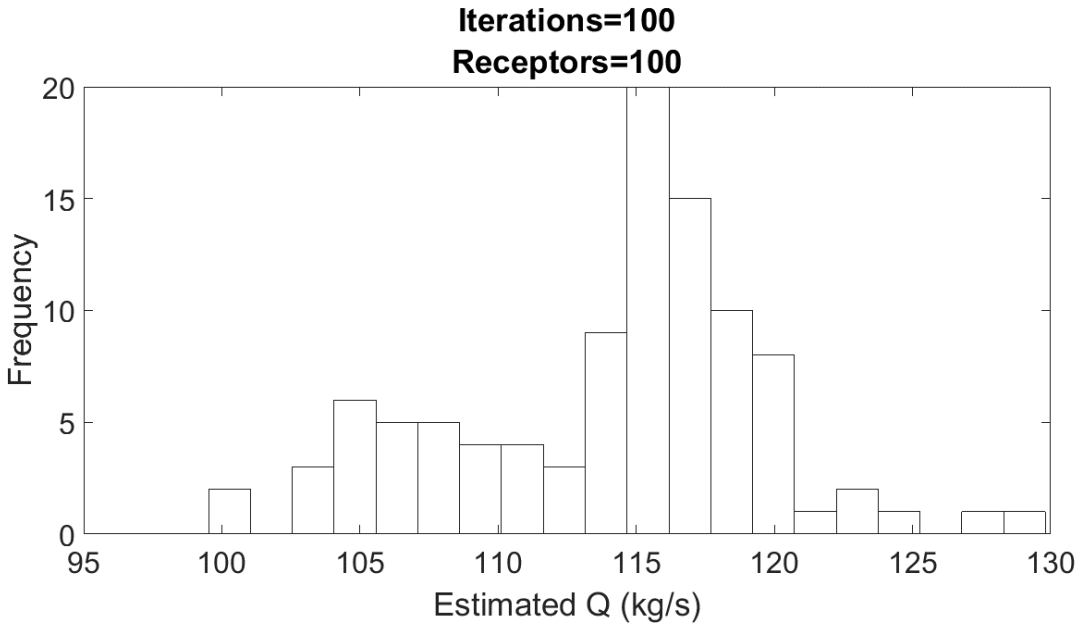

Figure 24. Estimated Qs versus the frequency for 100 receptors when using health effects as the input

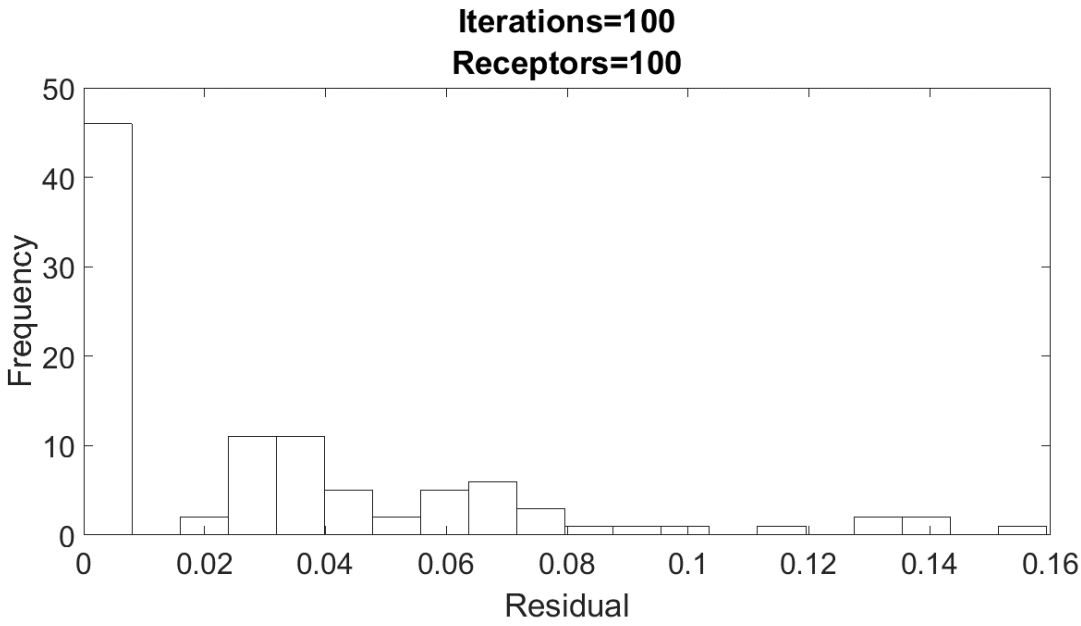

Figure 25. Residual error versus the frequency for 100 receptors when using health effects as the input

As shown in Figure 22, the estimated results of the source rate when using 9 receptors, lie between 83.4 kg/s and 500 kg/s, with the majority of the estimates falling in the 116.67-118.34 region. Increasing the number of receptor points to 100, better convergence is observed when compared to the use of the 9 points, with the results falling in the 100-130 kg/s range, and the majority of the estimates being in the 500 kg/s region (see Figure 24).

4.2.2 Unknown Source Location

Similar to the source rate reconstruction, the location of the release was estimated using the three different classes of input data, and each time the size of the receptors was varied between 2 to 100 points, for 100 iterations. The performance of the modeling system was assessed based on the use of the different types of input data and the size of receptors required to achieve the desired results.

4.2.2.1 Source Location Using Concentration Data

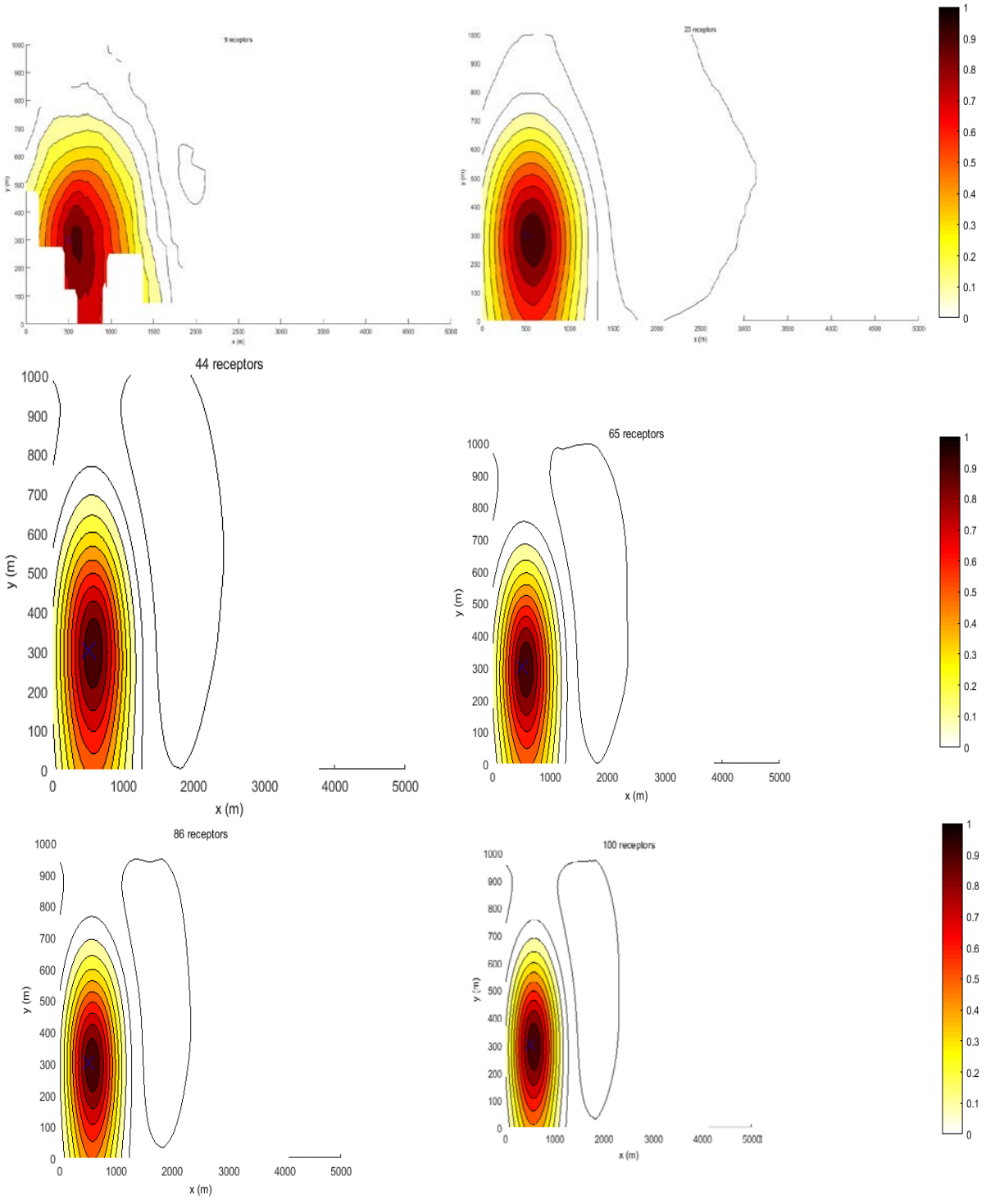

Figure 26. Contours of the estimated location regions for 9, 23, 44, 65, 86, and 100 receptors using concentration observations

For each set of receptors used, Figure 26 shows the regions where estimates of the release location are produced with varying levels of confidence, where a value of R^2 of 1 (see the color bar on the right) indicates that the exact location is reconstructed. The color bar represents the estimated values of R^2 (Pearson correlation coefficient) that are associated with each guessed location coordinates, with values closer to 1 indicating that the solution is closest to the actual release location which is marked by the X symbol.

Although the number of receptors was varied between 2 to 100 points, a minimum of 9 points was required in order to reconstruct the location of the hazardous release when using concentration as input data. The results are shown for six different sizes of receptors in the above mentioned range. As expected, reasonable estimates of the release location are obtained when using concentration measurement/data in pinpointing the location of the toxic H_2S release.

4.2.2.2 Source Location using TL Values

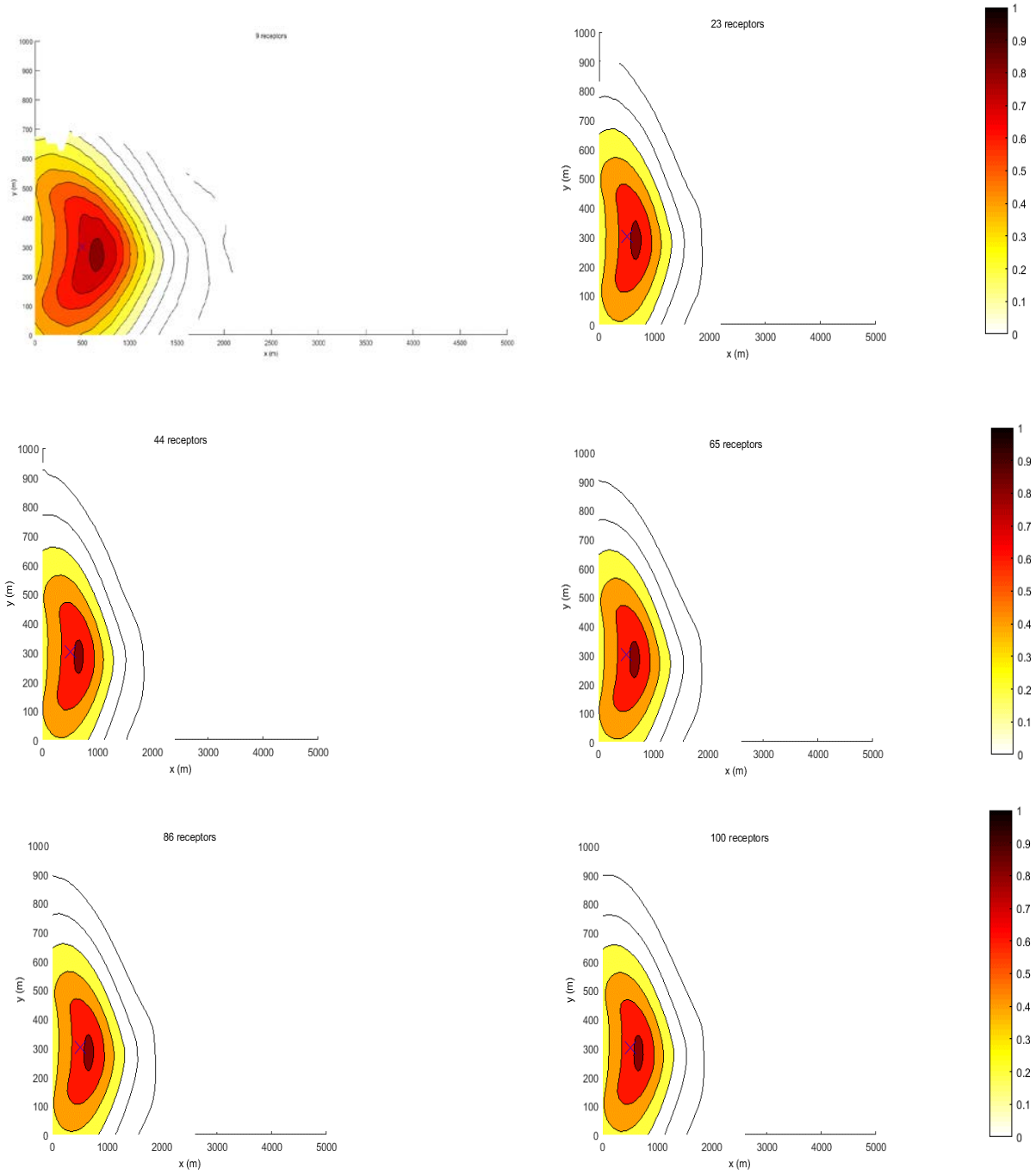

Figure 27. Contours of the estimated location regions for 9, 23, 44, 65, 86, and 100 receptors using TL values

As shown in Figure 27, identifying the location of the release from TL values was achieved in a similar manner to the case of the use of concentration as input data. As seen, the use of the concentration resulted in better estimates for the same number of receptors when compared to the use of TL values.

Similar to the first case, a minimum of 9 receptors was needed to achieve an estimate of the coordinates of the release location. The same six receptors sizes used in the case of concentration being the input data were also used to demonstrate the approach when using the TL values. In this second case, the results are also shown to be reasonably close to the actual/true location of the release location depending on the confidence level being considered. As mentioned before, values of R^2 closer to 1 indicate the true solution is either approached with high accuracy or reconstructed as it is.

4.2.2.3 Source Location using Health Effects

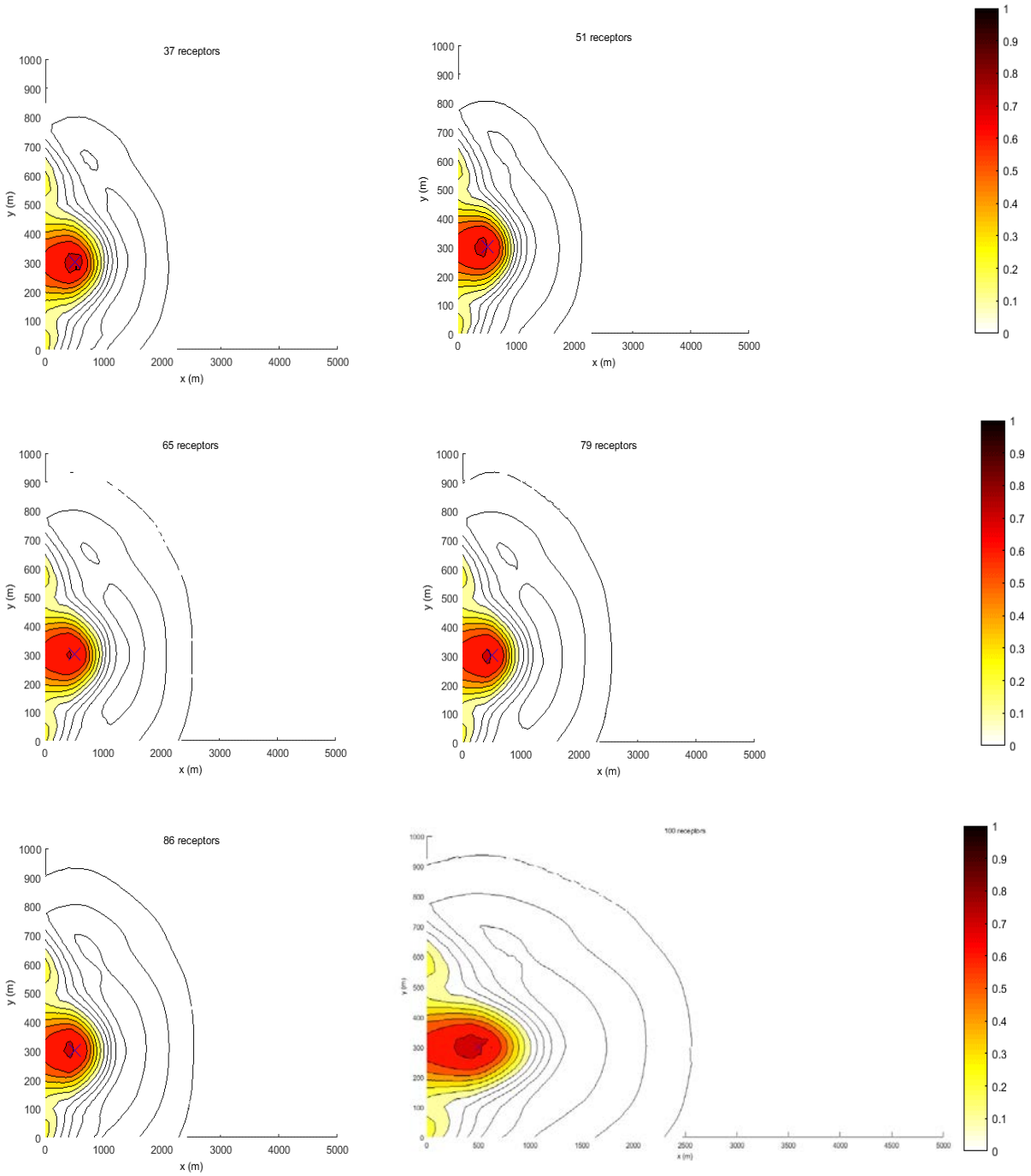

Figure 28. Contours of the estimated location regions for 37, 51, 65, 79, 86, and 100 receptors using health effects observations

Figure 28 shows the results for the source location estimation from observed health effects. Again, the size of the receptors was varied between 2 and 100 sample points. As shown, the level of the complexity associated with the use of health effects as input data was reflected on the minimum number of receptors required to reconstruct the source location when compared to the use of concentration and TL values.

4.3 Discussion

4.3.1 Effect of the Size of the Input Data Used on the Convergence of the Results

The sensitivity of the results to the size of the input data was gauged by varying the number of receptor points used in each of the individual scenarios. With respect to the performance of the proposed scheme in reconstructing the source rate, all of the three types of information used allowed the prediction of "averaged" source term rate to within 40-500 kg/s, and a wide variation in the quality of the algorithm's predictions was seen, particularly with the use of the health effects observations as input data.

It was shown that for the case of the use of the concentration data (see Figure 11), increasing the size of data points used in reconstructing the release rate, regardless of the number of iterations used, leads to better convergence of the results and better agreement with the actual release rate. The necessity to have large size of input data for the source reconstruction algorithm is due to the inherent variability and uncertainty of predicting the actual release rate. Here, this is reflected mostly with the introduced noise to the synthetic data as discussed in the methodology section. Without the noise, this is a simple well-defined problem and for which the solution algorithm converges in few iterations and with a small number of receptors. However, in reality, uncertainty is high and only scarce number of data are typically available for use in the source estimation, thus the considered scenarios are only

for exploratory purposes aiming at assessing the limitations associated with the proposed scheme.

In contrast to the use of the concentration data, the averaged source rate estimates that are obtained from the toxic load values are shown to exhibit less variability and good agreement with the actual release rate when utilizing 9 receptor points and above (see Figure 16). The results then become independent of the size of the receptors used from 51 points and above. The improvement associated with the use of the TL values when compared to the conventional use of the concentration data can be attributed to the fact that the TL is estimated using an integral of time and hence it contains previous history up till the specified time-frame at which the data are collected.

Finally, the moving average when utilizing health effects as input data seems to improve with increasing the number of the points where the observations are made (see Figure 21). Also, greater variations of the results are observed with the use of receptors points within 2 to 44. The variability of the results improves in the range 44 to 100 receptors.

In terms of the retrieval of the release location, the number of the data points used seems to influence the goodness of the estimates. The source location is well-retrieved (R^2 =0.8 within a downwind distance of 480-800 m and crosswind distance of 100-410 m) when utilizing the concentration data as input to the source estimation algorithm with a small number of input data as 9 points (see Figure 26). The estimates then improve when the size of the data points used is increased to 23 points (R^2 =0.8 within a downwind distance of 390-780 m and crosswind distance of 100-410 m) and become independent of the size of the data points used when increasing the size beyond that.

Similar to the use of the concentration data as input, the use of the toxic load values is also shown to result in good estimates of the source location (R^2 =0.8 within a downwind distance of 600-750 m and crosswind distance of 200-330 m) when using 9 points (see Figure 27). Increasing the number of the sampled points in reconstructing the source location, leads to a slight decrease in the quality of the estimates, which then become independent of the size of data points.

The use of health effects observations is accompanied by a great degree of complexity, and hence achieving source location estimates is not possible with smaller number of points as in the previous two cases of concentration and toxic load being the input to the source estimation algorithm. In the constructed release scenario, a minimum of 37 receptors with reported health observations is required to achieve a reasonable estimate (R^2 =0.8 within a downwind distance of 320-640 m and crosswind distance of 260-330 m) of the location coordinates (see Figure 28).

4.3.2 Effect of the Number of Iterations on the Convergence of the Source Rate Results

With respect to the influence of the number of iterations used in generating the source rate estimates on the performance of the scheme and the quality of the results, the results were assessed for all of the three classes of input data for two different number of model iterations.

As seen in Figure 11, increasing the number of iterations used in generating the results of the source rate estimates from concentration data is shown to yield better convergence and lesser variations from the moving average. This is because the use of such a large number of iterations accounts for different possible combinations of a certain receptors size rather than only a single set of a given number of data points; making it more representative. In reality, however, only a limited number of data points are available, hence the 10 iterations scenario

is closer to real-life situations where only the data themselves are used and not the combinations. The use of the random combinations here is for exploratory purposes to assess the performance against the uncertainty associated with the use of points with the same size but collected at different locations. Similar to the use of concentration data, TL information are shown to result in lesser variations from the moving average of the sample and better convergence for receptor sizes falling between 9 to 100 points (see Figure 14). Unlike the first two cases, using health effects as input data, better convergence and lesser deviation of the results from the moving average is observed when using smaller number of iterations in contrast to the use of larger number of iterations, especially for receptors size in the range of 2-44 data points (see Figure 21).

4.3.3 Impact of the Type of Data Used

With respect to the retrieval of the source term parameters, all the three classes of input data enabled the prediction of the source term characteristics with varying degrees of confidence and requirements of the receptor size. As can be seen from the results, the use of the concentration data and toxic load information required the least number of data points to reconstruct the source parameters. As the level of complexity of the provided information increases (i.e. the use of the health observations), more data points are required to achieve reasonable results. The required receptors size for the case of health effects observations was significantly larger than the other two cases, due to the complexity involved with the use of such type of information. The criterion sought for by an observer or an emergency responder is not exact values, instead it is a check of whether or not the health observations are present in a specific location.

5. CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

In this work, we have investigated the feasibility of a new source-reconstruction methodology based on the coupling of atmospheric dispersion and dose-response modeling. The proposed methodology utilizes health symptoms resulting from exposure to a HazMat release as indirect input to an integrated modeling system that back-calculates the responsible source characteristics. As indicated by the results of the case studies presented in this work, the proposed scheme for source reconstruction from health effects observations has managed to achieve good estimates of the source rate and location. The scheme, however, suffers from the inability to reconstruct the source parameters from a small set of observations or input data, which in real-time situations may not be possible as only limited number of observations are available to back-calculate the release parameters.

5.2 Recommendations

- To validate the scheme, it is recommended to explore the use of other dose-response models like the PBTK models, which unlike the toxic load model and the AEGL, account for the physical activities of the victims (such as seeking shelter), as well as the physiological and biological parameters of the exposed individuals and the different routes of exposure.
- Use of other dispersion models which allow the inclusion of variable meteorological information, and site topography (computational fluid dynamics models e.g. QUIC) and more complex information associated with the release of HazMats.

- Use of real data of the release (e.g. H_2S release from a natural gas facility) and health observations to validate the results since the synthetic scenarios are typically free from error and hence are not representative of the actual case scenarios.
- In this work, a simple forward modeling scheme employing MC sampling was utilized in obtaining the preliminary results as the focus is on the hypothesis rather than the mathematical background of the inverse problem. To account for the uncertainty of the results and obtain more realistic information with reduced randomness, it is recommended to utilize MCMC techniques to allow for comparison of the different model runs/iterations, and Bayesian inference to determine the probability of the solutions.
- Other inversion techniques, whether forward or backward modeling, are recommended to be applied and explored such as the artificial neural networks.
- It is also recommended to validate the scheme against other well-known toxic chemicals such as Chlorine.
- The release considered in this work was assumed to take place from a stationary source. It is recommended to study the effect of having a release from a moving source such as the ones occurring from transport accidents and terrorist attacks.
- Other scenarios and case studies need to be explored to validate the hypothesis and its limits, such as the reconstruction of the time at which the release first takes place and the inclusion of more input information to the modeling systems such as two different times at which the observations are recorded or estimated based on the provided information by the victims.

REFERENCES

Agency for Toxic Substances and Disease Registry. (2006). *TOXICOLOGICAL PROFILE FOR HYDROGEN SULFIDE*.

Allen, C. T., Haupt, S. E., & Young, G. S. (2007). Source characterization with a genetic algorithm-coupled dispersion-backward model incorporating SCIPUFF. *Journal of Applied Meteorology and Climatology*, *46*(3), 273–287. http://doi.org/10.1175/JAM2459.1

Andersen, M., Clewell, H. 3rd;, Gargas, M., MacNaughton, M., Reitz, R., Nolan, R., & McKenna, M. (1991). Physiologically based pharmacokinetic modeling with dichloromethane, its metabolite, carbon monoxide, and blood carboxyhemoglobin in rats and humans. *Toxicology and Applied Pharmacology*, *108*, 14–27.

Andersen, M. E., Clewell III, H. J., Gargas, M. L., Smith, F. A., & Reitz, R. H. (1987). Physiologically Based Pharmacokinetics and the Risk Assessment Process for Methylene Chloride, *Toxicology and Applied Pharmacology*, *87(2)*, 185-205.

Bhosale, V. A. ., Patil, K. I. ., & Dehankar, P. B. (2015). Study of accidental releases heavy gas dispersion comparing slab models and screen-3 model. *International Journal of Engineering Sciences & Research Technology*, *4*(1), 765–769.

Bliss, C. I. (1934a). The method of probits. *Science*, *79*, 38–39.

Bliss, C. I. (1934b). The method of probits - a correction. *Science*, *79*, 409–410.

Boris, J. P., & Oran, E. S. (2000). *Numerical Simulation of Reactive Flow* (Second). New York: Cambridge University Press. http://doi.org/http://dx.doi.org/10.1017/CBO9780511574474

Boris, J. P.; Patnaik, G. (2014). Acute Exposure Guideline Levels (AEGLs) for Time Varying Toxic Plumes. Washignton: Naval Research Laboratory.

Bowonder, B., & Miyake, T. (1988). Managing hazardous facilities: Lessons from the Bhopal accident. *Journal of Hazardous Materials*, *19*(3), 237–269. http://doi.org/10.1016/0304-3894(88)80025-6

Busvine, J. R. (1938). The Toxicity of Ethylene Oxide to Calandra Oryzae, C. Granaria, Tribolium Castaneum, and Cimex Lectularius. *Annals of Applied Biology*, *25*, 605–632.

Chen, C. C., Shih, M. C., & Wu, K. Y. (2010). Exposure estimation using repeated blood concentration measurements. *Stochastic Environmental Research and Risk Assessment*, *24*(3), 445–454. http://doi.org/DOI 10.1007/s00477-009-0332-0

Chen, C. C., Shih, M. C., & Wu, K. Y. (2012). Exposure reconstruction using a physiologically based toxicokinetic model with cumulative amount of metabolite in urine: A case study of trichloroethylene inhalation. *Stochastic Environmental Research and Risk Assessment*, *26*(1), 21–31. http://doi.org/10.1007/s00477-011-0502-8

Christensen, K. L. Y., Lorber, M., Ye, X., & Calafat, A. M. (2015). Reconstruction of bisphenol A intake using a simple pharmacokinetic model. *J Expo Sci Environ Epidemiol*, *25*(3), 240–248. http://doi.org/10.1038/jes.2013.81

Christopoulos, V. N., & Roumeliotis, S. (2005). Adaptive sensing for instantaneous gas release parameter estimation. *Proceedings - IEEE International Conference on Robotics and Automation, 2005*, 4450–4456. http://doi.org/10.1109/ROBOT.2005.1570805

Clewell, H. J. 3rd, Andersen, M. E., & Barton, H. A. (2002). A consistent approach for the application of pharmacokinetic modeling in cancer and noncancer risk assessment. *Environmental Health Perspectives*, *110*, 85–93.

COST ESSEM ES1006. (2015). Evaluation, improvement and guidance for the use of local-scale emergency prediction and response tools for airborne hazards in built environments, *COST Offic*.

Czech, C., Platt, N., Urban, J., Bieringer, P., Bieberbach, G., Wyszogrodzki, A., & Weil, J. (2011). A Comparison of Hazard Predictions Based on the Ensemble-Mean Plume Versus Individual Plume Realizations Using Different Toxic Load Models. In *the Special Symposium on Applications of Air Pollution Meteorology, 91st Annual American Meteorological Society Meeting*. Seattle, Washington.

De Foy, B., Cui, Y. Y., Schauer, J. J., Janssen, M., Turner, J. R., & Wiedinmyer, C. (2015). Estimating sources of elemental and organic carbon and their temporal emission patterns using a Least Squares Inverse model and hourly measurements from the St. Louis-Midwest Supersite. *Atmospheric Chemistry and Physics Discussions*, *14*(8), 12019–12070. http://doi.org/10.5194/acpd-14-12019-2014

Dewoskin, R. S. (2007). PBPK models in risk assessment — A focus on chloroprene, *166*(February), 352–359. http://doi.org/10.1016/j.cbi.2007.01.016

Edler, L., & Kitsos, C. P. (2005). *Recent advances in quantitative methods for cancer and human health risk assessment*. John Wiley & Sons.

Ermak, D. L. (1990). *User's Manual for SLAB: An Atmospheric Dispersion Model for Denser-Than-Air Releases*.

Finney, D. J. (1971). Probit Analysis. *Cambridge University Press*, (third edition).

Georgopoulos, P. G., Sasso, A. F., Isukapalli, S. S., Lioy, P. J., Vallero, D. A., Okino, M., & Reiter, L. (2009). Reconstructing population exposures to environmental chemicals

from biomarkers : Challenges and opportunities, 149–171. http://doi.org/10.1038/jes.2008.9

Griffith, R. F. (1991). The use of probit expressions in the assessment of acute population impact of toxic releases, *4*, 49–57.

Gudiksen, P. H., Harvey, T. F., & Lange, R. (1989). Chernobyl source term, atmospheric dispersion, and dose estimation. *Health Physics*, *57*(5), 697–706. http://doi.org/10.1097/00004032-198911000-00001

Guidotti, T. L. (1996). Hydrogen sulphide. *Occupational Medicine*, *46*(5), 367–371. http://doi.org/10.1093/occmed/46.5.367

Gunatilaka, A., Skvortsov, A., & Gailis, R. (2014). A review of toxicity models for realistic atmospheric applications. *Atmospheric Environment*, *84*, 230–243. http://doi.org/10.1016/j.atmosenv.2013.11.051

Hanna, S., Britter, R., Argenta, E., & Chang, J. (2012). The Jack Rabbit chlorine release experiments: Implications of dense gas removal from a depression and downwind concentrations. *Journal of Hazardous Materials*, *213-214*, 406–412. http://doi.org/10.1016/j.jhazmat.2012.02.013

Haupt, S. E., & Young, G. S. (2008). Paradigms for Source Characterization. In *15th Joint Conference on the Applications of Air Pollution Meteorology with the A&WMA* (pp. 20–24). New Orleans, LA.

Haupt, S. E., Young, G. S., Long, K. J., Beyer-Lout, A., & Annunzio, A. J. (2008). Data fusion and prediction for CBRN transport and dispersion for security. *IEEE Aerospace Conference Proceedings*. http://doi.org/10.1109/AERO.2008.4526584

Hilderman, T. L. (1999). A model for effective toxic load from fluctuating gas concentrations, 115–134.

HSE. (2006). Methods of approximation and determination of human vulnerability for offshore major accident hazard assessment, 1–55. Retrieved from http://www.hse.gov.uk/foi/internalops/hid_circs/technical_osd/spc_tech_osd_30/spctecosd30.pdf

Issartel, J. P., Sharan, M., & Singh, S. K. (2012). Identification of a point of release by use of optimally weighted least squares. *Pure and Applied Geophysics*, *169*(3), 467–482. http://doi.org/10.1007/s00024-011-0381-4

Kathirgamanathan, P., Mckibbin, R., & McLachlan, R. I. (2003). Inverse Modelling for Identifying the Origin and Release rate of Atmospheric Pollution- An Optimisation Approach. *MODSIM*, *7*(1), 61–66.

Korsakissok, I., Mathieu, a., & Didier, D. (2013). Atmospheric dispersion and ground deposition induced by the Fukushima Nuclear Power Plant accident: A local-scale simulation and sensitivity study. *Atmospheric Environment*, *70*, 267–279. http://doi.org/10.1016/j.atmosenv.2013.01.002

Lushi, E., & Stockie, J. M. (2010). An inverse Gaussian plume approach for estimating atmospheric pollutant emissions from multiple point sources. *Atmospheric Environment*, *44*(8), 1097–1107. http://doi.org/10.1016/j.atmosenv.2009.11.039

Lyons, M. a., Yang, R. S. H., Mayeno, A. N., & Reisfeld, B. (2008). Computational toxicology of chloroform: Reverse dosimetry using Bayesian inference, Markov chain Monte Carlo simulation, and human biomonitoring data. *Environmental Health Perspectives*, *116*(8), 1040–1046. http://doi.org/10.1289/ehp.11079

Mazzola, C. A., Addis, R. P., & Emergency Management Laboratory (Oak Ridge Institute for Science and Education). (1995). *Atmospheric Dispersion Modeling Resources* (Second).

Meselson, M., Guillemin, J., Hugh-Jones, M., Langmuir, a, Popova, I., Shelokov, a, & Yampolskaya, O. (1994). The Sverdlovsk anthrax outbreak of 1979. *Science (New York, N.Y.)*, *266*(5188), 1202–1208. http://doi.org/10.1126/science.7973702

Mioduszewski, R., Manthei, J., Way, R., Burnett, D., Gaviola, B., Muse, W., … Crosier, R. (2002). Interaction of exposure concentration and duration in determining acute toxic effects of sarin vapor in rats. *Toxicological Sciences*, *66*(2), 176–184. http://doi.org/10.1093/toxsci/66.2.176

Morita, H., Yanagisawa, N., Nakajima, T., Shimizu, M., Hirabayashi, H., Okudera, H., … Mimura, S. Sarin poisoning in Matsumoto, Japan., 346 Lancet 290–293 (1995). http://doi.org/10.3918/jsicm.4.363

Nasstrom, J. S., Sugiyama, G., Baskett, R. L., Larsen, S. C., & Bradley, M. M. (2007). The National Atmospheric Release Advisory Center modelling and decision-support system for radiological and nuclear emergency preparedness and response. *International Journal of Emergency Management*, *4*(3), 524–550. http://doi.org/10.1504/IJEM.2007.014301

National Research Council. (2003). *Tracking and Predicting the Atmospheric Dispersion of Hazardous Material Releases: Implications for Homeland Security*. Washington, D.C.: National Academies Press.

Ni, H., Rui, Y., Wang, J., & Cheng, L. (2014). A Synthetic Method for Atmospheric Diffusion Simulation and Environmental Impact Assessment of Accidental Pollution in the Chemical Industry in a WEBGIS Context. *International Journal of Environmental Research and Public Health*, *11*(9), 9238–9255. http://doi.org/10.3390/ijerph110909238

Okumura, T., Takasu, N., Ishimatsu, S., Miyanoki, S., Mitsuhashi, A., Kumada, K., … Hinohara, S. (1996). Report on 640 Victims of the Tokyo Subway Sarin Attack. *Annals of Emergency Medicine*, *28*(2), 129–135. http://doi.org/10.1016/S0196-0644(96)70052-5

Rao, K. S. (2005). Uncertainty analysis in atmospheric dispersion modeling. *Pure and Applied Geophysics*, *162*(10), 1893–1917. http://doi.org/10.1007/s00024-005-2697-4

Rao, K. S. (2007). Source estimation methods for atmospheric dispersion. *Atmospheric Environment*, *41*(33), 6964–6973. http://doi.org/10.1016/j.atmosenv.2007.04.064

Rao, H. V, & Ginsberg, G. L. (1997). A physiologically-based pharmacokinetic model assessment of methyl t-butyl ether in groundwater for a bathing and showering determination. *Risk Analysis : An Official Publication of the Society for Risk Analysis*, *17*(5), 583–98. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9404048

Reisfeld, B., Mayeno, A. N., Lyons, M. A., & Yang, R. S. H. (2006). Physiologically Based Pharmacokinetic and Pharmacodynamic Modeling. In *Computational Toxicology* (pp. 33–69). John Wiley & Sons. http://doi.org/10.1002/9780470145890.ch3

Roy, A., Weisel, C. P., Lioy, P. J., & Georgopoulos, P. G. (1996). A Distributed Parameter Physiologically-Based Pharmacokinetic Model for Dermal and Inhalation Exposure to Volatile Organic Compounds. *Risk Analysis: An International Journal*, *16*(2), 147–160. Retrieved from http://lib-ezproxy.tamu.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bcr&AN=11945352&site=eds-live

Rudd, A. C., Robins, A. G., Lepley, J. J., & Belcher, S. E. (2012). An Inverse Method for Determining Source Characteristics for Emergency Response Applications. *Boundary-Layer Meteorology*, *144*(1), 1–20. http://doi.org/10.1007/s10546-012-9712-y

Sharan, M., Issartel, J.-P., Singh, S. K., & Kumar, P. (2009). An inversion technique for the retrieval of single-point emissions from atmospheric concentration measurements. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *465*(2107), 2069–2088. http://doi.org/10.1098/rspa.2008.0402

Sharma, P. K., Gera, B., & Ghosh, A. K. (2011). Inverse problems using Artificial Neural Networks in long range atmospheric dispersion. *Kerntechnik*, *76*(2), 115–120. http://doi.org/10.3139/124.110116

Singh, S. K., & Rani, R. (2014). A least-squares inversion technique for identification of a point release: Application to Fusion Field Trials 2007. *Atmospheric Environment*, *92*, 104–117. http://doi.org/10.1016/j.atmosenv.2014.04.012

Singh, S. K., Turbelin, G., Issartel, J.-P., Kumar, P., & Feiz, A. A. (2015). Reconstruction of an atmospheric tracer source in Fusion Field Trials: Analyzing resolution features.

Journal of Geophysical Research: Atmospheres, *120*(12), 6192–6206. http://doi.org/10.1002/2015JD023099

Stage, S. a. (2004). Determination of acute exposure guideline levels in a dispersion model. *Journal of the Air & Waste Management Association (1995)*, *54*(1), 49–59. http://doi.org/10.1080/10473289.2004.10470885

Stohl, A., Seibert, P., Wotawa, G., Arnold, D., Burkhart, J. F., Eckhardt, S., … Yasunari, T. J. (2012). Xenon-133 and caesium-137 releases into the atmosphere from the Fukushima Dai-ichi nuclear power plant: Determination of the source term, atmospheric dispersion, and deposition. *Atmospheric Chemistry and Physics*, *12*(5), 2313–2343. http://doi.org/10.5194/acp-12-2313-2012

Sykes, R. I., Parker, S. F., & Henn, D. S. (2004). *SCIPUFF Version 2.1Technical Documentation*.

ten Berge, W. F., & van Heemst, M. V. (1983). Validity and accuracy of a commonly used toxicity-assessment model in risk analysis. In *Fourth International Symposium on Loss Prevention and Safety Promotion in the Process Industries* (Vol. 1). Institute of Chemical Engineers.

ten Berge, W. F., Zwart, A., & Appelman, L. M. (1986). Concentration-time mortality response relationship of irritant and systemically acting vapours and gases. *Journal of Hazardous Materials*, *13*(3), 301–309. http://doi.org/10.1016/0304-3894(86)85003-8

U.S. Environmental Protection Agency AEGL Program. (2012). No Title. Retrieved from http://www.epa.gov/oppt/aegl/

Vallero, D., & Isukapalli, S. (2013). Simulating real-world exposures during emergency events : studying effects of indoor and outdoor releases in the Urban Dispersion Program in upper Manhattan , NY, (January 2012), 1–11. http://doi.org/10.1038/jes.2013.38

Wang, Y., Huang, H., & Zhu, W. (2015). Stochastic source term estimation of HAZMAT releases : algorithms and uncertainty. In *12th International Conference on Information Systems for Crisis Response and Management*. Kristiansand, Norway.

Wawrzynczak, A., Jaroszyński, M., & Borysiewicz, M. (2014). Data-driven Genetic Algorithm in Bayesian estimation of the abrupt atmospheric contamination, *2*, 519–527. http://doi.org/10.15439/2014F272

Weinrich, S. L., Lawrence, A. E., Burr, J. K., Grotte, J. H., & Laviolet, L. L. (2008). A Comparative Analysis of Toxicity Models. *Institute for Defense Analyses*, (April).

Witschi, H. (1999). Some Notes on the History of Haber's Law, *168*, 164–168.

Yee, E., Hoffman, I., & Ungar, K. (2014). Bayesian Inference for Source Reconstruction : A Real-World Application, *2014*(1).

APPENDIX: MATLAB CODE

1. Solve

```
function
[Qmean,RESmean]=SolveV4(NRecS,NStep,NRecE,NIterS,NItStep,NIterE,CorTL)
%% SET GLOBAL Parameters
x=0:25:5000;
y=0:25:1000;
t=600;
z=2;
Src.x=500;
Src.y=300;

%CorTL='HE';
AEGLk=1;
Agent=AgentSetup('H2S');
%Iter=100;
%NRecS=10;
%NRecE=10;
%NStep=1;
Receptors=NRecS:NStep:NRecE;
Iterations=NIterS:NItStep:NIterE;


%% COLLECT Predict
Q=[50 125 250 500];
for ip=1:4
    P(ip)= Read_PredictV4(strcat('SLAB/predict_0',int2str(ip-1)));
end


%% CREATE REF case
Qx=100;
Pref= Read_PredictV4('SLAB/predict_x100');
%Pref=PrepPV4(Q,100,P);
REF=zeros(length(x),length(y));
for ix=1:length(x)
    for iy=1:length(y)
        REF(ix,iy)=ImpactV4(CorTL,Pref,AEGLk,Agent,Src,x(ix),y(iy),z,t);
    end
end

%parpool(2);
fileNameBase=strcat('output/Q_',CorTL);
fileNameProg=strcat(fileNameBase,'_',
int2str(NRecS),'_',int2str(NRecE),'_',int2str(NIterS),'_',int2str(NIterE))
;
parfor_progress(fileNameProg,NIterE*((NRecE-NRecS)/NStep+1));

%% CREATE random Sample
```

```
poolobj = gcp('nocreate'); % If no pool, do not create new one.
if isempty(poolobj)
    poolsize = 0;
else
    poolsize = poolobj.NumWorkers;
end

for iR=1:numel(Receptors)
    iRec=Receptors(iR);
    iREF=find(REF>0);
    nsample=NIterE;
    if numel(iREF)<iRec
        error( 'Not possible to proceed further, available Receptors:
%4i\n', numel(iREF));
    end
    binomComb=nchoosek(numel(iREF),iRec);
    if binomComb<nsample;nsample=binomComb;end
    randREF=zeros(nsample,iRec);
    for h=1:nsample
        randREF(h,:)=randperm(numel(iREF),iRec);%Random sample of
receptors following TLn filter
    end
    [nzXX,nzY]=ind2sub(size(REF),iREF(randREF));%Linear indixing
convertred back into subscripts
    REFrand=REF(iREF(randREF));
    xrand=x(nzXX);
    yrand=y(nzY);

%% RUN ALL ITERATIONS
Qguess=zeros(1,NIterE);
res=zeros(1,NIterE);
    parfor it=1 : NIterE

[Qguess(it),res(it)]=fminbnd(@(Qs)Stockie_InverseV4(Qs,xrand(it,:),yrand(i
t,:),z,REFrand(it,:),t,P,Q,CorTL,AEGLk,Agent,Src),50,500);
         parfor_progress(fileNameProg);
    end
    % Save the partial files
    for it=1:numel(Iterations)
        NIter=Iterations(it);
        SQguess=Qguess(1: NIter);
        SRes=res(1: NIter);
        Qmean(iR,it)=mean(SQguess);
        Qstdev(iR,it)=std(SQguess);
        RESmean(iR,it)=mean(SRes);

fileNameMat=strcat(fileNameBase,'_i',int2str(NIter),'_r',int2str(iRec));
        save(strcat(fileNameMat, '.mat'),'AEGLk', 't','SQguess', 'SRes');
    end

end

parfor_progress(fileNameProg,0);
drawnow;
```

```
end
```

Consolidate the results after the running the main codes

```
%what to execute
examples='Q';
switch examples
    case 'Solve'
        %% Examples of Running Code
        [Qmean,RESmean]=SolveV4(2,7,100,2,14,100,'C'); % find Q using
concentration
        % [Qmean,RESmean]=SolveV4(2,2,20,10,10,5,30,'TL'); % find Q using
Toxic Load
        % [Qmean,RESmean]=SolveV4(2,2,20,10,10,5,30,'HE'); % find Q using
Toxic Load
        % [Res,FSrc]=SolveV4_loc(2,2,20,10,'C'); % find Location using
concentration
        % [Res,FSrc]=SolveV4_loc(2,2,20,10,'TL'); % find Location using
Toxic Load
        % [Res,FSrc]=SolveV4_loc(2,2,20,10,'HE'); % find Location using
Toxic Load

    case 'Location'
        %% TEST of Location Identification
        %   Runs should be already done

        x=0:25:2000;
        y=0:25:1000;
        Src.x=500;
        Src.y=300;
        [fgh]=figure;
        hold on;

        %GET DATA
        NRecS=2;
        NStep=7;
        NRecE=100;
        NIter=10;
        CorTL='C'; % change to C or TL or HE
        fileNameBase=strcat('output/Loc_',CorTL,'_i',int2str(NIter));

        Receptors=NRecS:NStep:NRecE;
        for iR=1:numel(Receptors)
            iRec=Receptors(iR);
            fileNameMat=strcat(fileNameBase,'_r',int2str(iRec));
            load(strcat(fileNameMat, '.mat'));
            figure(fgh);
            contourf(x,y,Res');
            caxis([0 1]);
            colormap hot;
            colorbar;
            plot(Src.x,Src.y,'-x','MarkerSize',15);
            drawnow;
            Movie(iR)=getframe(gcf);
            minResR(iR)=minRes;
```

```
        end
        hold off;
        plot(Receptors,minResR);
        % myVideo = VideoWriter('myfile.avi');
        % myVideo.FrameRate = 1;  % Default 30
        % myVideo.Quality = 50;    % Default 75
        % open(myVideo);
        % writeVideo(myVideo, Movie);
        % close(myVideo);

    case 'Q'

        %% TEST of Source Rate
        %   Runs should be already done

        x=0:25:5000;
        y=0:25:1000;
        Src.x=500;
        Src.y=300;
        [fghQ]=figure;

        %GET DATA
        NRecS=2;
        NStep=7;
        NRecE=100;
        NIterS=100;
        NItStep=100;
        NIterE=100;
        CorTL='C'; % change to C or TL
        fileNameBase=strcat('output/Q_',CorTL);

        Receptors=NRecS:NStep:NRecE;
        Iterations=NIterS:NItStep:NIterE;

        for iR=1:numel(Receptors)
            iRec=Receptors(iR);
            for it=1:numel(Iterations)
                iTer=Iterations(it);

fileNameMat=strcat(fileNameBase,'_i',int2str(iTer),'_r',int2str(iRec));
                load(strcat(fileNameMat, '.mat'));
                %RES(iR,it,:)=SRes;
                meanRES(iR,it)=mean(SRes);
                %QGUESS(iR,it,:)=SQguess;
                meanQGUESS(iR,it)=mean(SQguess);
                stdevQGUESS(iR,it)=std(SQguess);
%                 figure(fghQ);
%                 hist(Qguess);
%                 drawnow;
%                 MovieQ(iR)=getframe(gcf);
            end
        end
%%
end
```

```
%% Plot Q Mean
figure;
it=100;
Iterations(it)
X=[Receptors,fliplr(Receptors)];                %#create continuous x
value array for plotting
y1=meanQGUESS(:,it)-stdevQGUESS(:,it);
y2=meanQGUESS(:,it)+stdevQGUESS(:,it);
Y=[y2',fliplr(y1)'];              %#create y values for out and then back
h=fill(X,Y,'b');                  %#plot filled area
set(h,'Facecolor',[1 0.84 0.99]);
set(h,'FaceAlpha',0.5);
set(h,'LineStyle','none');
hold on
h=plot(Receptors,meanQGUESS(:,it));
set(h,'Color',[1 0.84 0.99]);
set(h,'LineWidth',2);

it=100;
Iterations(it)
X=[Receptors,fliplr(Receptors)];                %#create continuous x
value array for plotting
y1=meanQGUESS(:,it)-stdevQGUESS(:,it);
y2=meanQGUESS(:,it)+stdevQGUESS(:,it);
Y=[y2',fliplr(y1)'];              %#create y values for out and then back
h2=fill(X,Y,'b');                  %#plot filled area
set(h2,'Facecolor',[0.5 0.34 0.99]);
set(h2,'FaceAlpha',0.5);
set(h2,'LineStyle','none');
h2=plot(Receptors,meanQGUESS(:,it));
set(h2,'Color',[0.5 0.34 0.99]);
set(h2,'LineWidth',2)
```

Solve_Loc

```
function [Res,FSrc]=SolveV4_loc(NRecS,NStep,NRecE,NIter,CorTL)
%Find location of source using correllation

% NRecS=10;
% NStep=10;
% NRecE=10;
% NIter=2;
%CorTL='HE';

%% SET GLOBAL Parameters
x=0:25:5000;
y=0:25:1000;
t=600;
z=2;
Src.x=600;
Src.y=300;


AEGLk=1;
Agent=AgentSetup('H2S');
```

```
%Iter=100;
%NRecS=10;
%NRecE=10;
%NStep=1;
Iterations=NRecS:NStep:NRecE;


%% COLLECT Predict
Q=[50 125 250 500];
for ip=1:4
    P(ip)= Read_PredictV4(strcat('SLAB/predict_0',int2str(ip-1)));
end


%% CREATE REF and BASE case
Qx=100;
Pref= Read_PredictV4('SLAB/predict_x100');
%Pref=PrepPV4(Q,100,P);
REF=zeros(length(x),length(y));
Pbase=PrepPV4(Q,300,P);
BSrc.ixB=round(length(x)/2);
BSrc.iyB=round(length(y)/2);
BSrc.x=x(BSrc.ixB);
BSrc.y=y(BSrc.iyB);
for ix=1:length(x)
    for iy=1:length(y)
        REF(ix,iy)=ImpactV4(CorTL,Pref,AEGLk,Agent,Src,x(ix),y(iy),z,t);

BASE(ix,iy)=ImpactV4(CorTL,Pbase,AEGLk,Agent,BSrc,x(ix),y(iy),z,t);
    end
end

%parpool(2);
fileNameBase=strcat('output/Loc_',CorTL,'_i',int2str(NIter));
fileNameProg=strcat(fileNameBase,'_', int2str(NRecS),'_',int2str(NRecE));
parfor_progress(fileNameProg,numel(Iterations)*length(x));

%% CREATE random Sample

%parpool(2);

for iR=1:numel(Iterations)
iRec=Iterations(iR);

    iREF=find(REF>0);
    nsample=NIter;
    if numel(iREF)<iRec
        error( 'Not possible to proceed further, available Receptors:
%4i\n', numel(iREF));
    end
    binomComb=nchoosek(numel(iREF),iRec);
    if binomComb<nsample;nsample=binomComb;end
    randREF=zeros(nsample,iRec);
    for h=1:nsample
```

```
        randREF(h,:)=randperm(numel(iREF),iRec);%Random sample of
receptors following TLn filter
    end
    [nzXX,nzY]=ind2sub(size(REF),iREF(randREF));%Linear indixing
convertred back into subscripts
    REFrand=REF(iREF(randREF));
    xrand=x(nzXX);
    yrand=y(nzY);

%% RUN ITERATIONS
    N = size(xrand,1);
    Res=zeros(numel(x),numel(y));
    FSrc.x=mean(x);
    FSrc.y=mean(y);
    minRes=100;
    for ix=1:length(x)
         for iy=1:length(y)
            ixB=ix-BSrc.ixB;
            iyB=iy-BSrc.iyB;
            nBASE=shiftmatrix(BASE,ixB, iyB);
            BASErand=nBASE(iREF(randREF));
            CorRes=zeros(1,N);
            for it=1 : N %Different realizations
                tcorcof  = corrcoef(REFrand(it,:),BASErand(it,:));
                CorRes(it)=tcorcof(1,2);
            end
            Res(ix,iy)=mean(CorRes);
            if abs((1-Res(ix,iy))/Res(ix,iy))<minRes
                FSrc.x=x(ix);
                FSrc.y=y(iy);
                minRes=abs((1-Res(ix,iy))/Res(ix,iy));
            end
         end
         parfor_progress(fileNameProg);
    end
    minRes=1/(1+minRes);
    fileNameMat=strcat(fileNameBase,'_r',int2str(iRec));
    save(strcat(fileNameMat, '.mat'),'AEGLk', 't', 'Iterations',
'Res','FSrc','minRes');
end

parfor_progress(fileNameProg,0);
drawnow;
end
```

Consolidate the results after running the main code

```
%what to execute
examples='Location'; %Use 'Q' or 'Location'
switch examples
    case 'Solve'
        %% Examples of Running Code
        [Qmean,RESmean]=SolveV4(2,7,100,10,50,200,'C'); % find Q using
concentration
        %[Qmean,RESmean]=SolveV4(2,2,20,10,50,200,'TL'); % find Q using
Toxic Load
        % [Qmean,RESmean]=SolveV4(2,2,20,10,5,30,'HE'); % find Q using
Toxic Load
        % [Res,FSrc]=SolveV4_loc(2,2,20,10,'C'); % find Location using
concentration
        % [Res,FSrc]=SolveV4_loc(2,2,20,10,'TL'); % find Location using
Toxic Load
        % [Res,FSrc]=SolveV4_loc(2,2,20,10,'HE'); % find Location using
Toxic Load

    case 'Location'
        %% TEST of Location Identification
        %   Runs should be already done

        x=0:25:5000;
        y=0:25:1000;
        Src.x=500;
        Src.y=300;
        [fgh]=figure;
        hold on;

        %GET DATA
        NRecS=2;
        NStep=7;
        NRecE=100;
        NIter=100;
        CorTL='HE'; % change to C or TL or HE
        fileNameBase=strcat('output/Loc_',CorTL,'_i',int2str(NIter));

        Receptors=NRecS:NStep:NRecE;
        for iR=1:numel(Receptors)
            iRec=Receptors(iR);
            fileNameMat=strcat(fileNameBase,'_r',int2str(iRec))
            load(strcat(fileNameMat, '.mat'));
            figure(fgh);
            contourf(x,y,Res');
            caxis([0 1]);
            colormap hot;
            colorbar;
            plot(Src.x,Src.y,'-x','MarkerSize',15); %try to plot a circle
of 25 radius
            drawnow;
            pause;
            Movie(iR)=getframe(gcf);
            minResR(iR)=minRes;
```

```
        end
        hold off;
        plot(Receptors,minResR);
        % myVideo = VideoWriter('myfile.avi');
        % myVideo.FrameRate = 1;  % Default 30
        % myVideo.Quality = 50;    % Default 75
        % open(myVideo);
        % writeVideo(myVideo, Movie);
        % close(myVideo);

    case 'Q'

        %% TEST of Source Rate
        %   Runs should be already done

        x=0:25:2000;
        y=0:25:1000;
        Src.x=500;
        Src.y=300;
        [fghQ]=figure;

        %GET DATA
        NRecS=2;
        NStep=7;
        NRecE=100;
        NIterS=10;
        NItStep=50;
        NIterE=100;
        CorTL='C'; % change to C or TL
        fileNameBase=strcat('output/Q_',CorTL);%Used to be:
('output/A1_t1/Q_',CorTL);

        Receptors=NRecS:NStep:NRecE;
        Iterations=NIterS:NItStep:NIterE;

        for iR=1:numel(Receptors)
            iRec=Receptors(iR);
            for it=1:numel(Iterations)
                iTer=Iterations(it);

fileNameMat=strcat(fileNameBase,'_i',int2str(iTer),'_r',int2str(iRec));
                load(strcat(fileNameMat, '.mat'));
                %RES(iR,it,:)=SRes;
                meanRES(iR,it)=mean(SRes);
                %QGUESS(iR,it,:)=SQguess;
                meanQGUESS(iR,it)=mean(SQguess);
%                 figure(fghQ);
%                 hist(Qguess);
%                 drawnow;
%                 MovieQ(iR)=getframe(gcf);
            end
        end
%%
end
```