Paradigms and Commonalities in Atmospheric Source Term Estimation Methods

¹Paul E. Bieringer*,
²George S. Young, ³Luna M. Rodriguez, ¹Andrew J. Annunzio
³Francois Vandenberghe, and ³Sue Ellen Haupt

¹Aeris LLC
Louisville, CO, USA

²Department of Meteorology and Atmospheric Science
Pennsylvania State University
University Park, PA, USA

³Research Applications Laboratory
National Center for Atmospheric
3450 Mitchell Lane
Boulder, CO, USA

Revised Submission to Atmospheric Environment

December 2016

*Corresponding Author

Address: 1314 Main St. Suite 101, Louisville, CO 80027
Email: paulb@aerisllc.com
Telephone: 720-370-9012

1 **Abstract:**

2 Modeling the downwind hazard area resulting from the unknown release of an
3 atmospheric contaminant requires estimation of the source characteristics of a localized
4 source from concentration or dosage observations and use of this information to model
5 the subsequent transport and dispersion of the contaminant. This source term estimation
6 problem is mathematically challenging because airborne material concentration
7 observations and wind data are typically sparse and the turbulent wind field chaotic.
8 Methods for addressing this problem fall into three general categories: forward modeling,
9 inverse modeling, and nonlinear optimization. Because numerous methods have been
10 developed on various foundations, they often have a disparate nomenclature. This
11 situation poses challenges to those facing a new source term estimation problem,
12 particularly when selecting the best method for the problem at hand. There is, however,
13 much commonality between many of these methods, especially within each category.
14 Here we seek to address the difficulties encountered when selecting an STE method by
15 providing a synthesis of the various methods that highlights commonalities, potential
16 opportunities for component exchange, and lessons learned that can be applied across
17 methods.

18

19 **Key Words:**

20 Atmospheric Transport and Dispersion, Source Term Estimation, Nonlinear
21 Optimization, Adjoint Methods, Inverse Modeling

22

1. Introduction

A variety of applications rely on accurate source term estimates (STE) for materials released into the atmosphere from localized areas. The receptor or observations used in STE problems can come in several forms that include information derived from sensors repeatedly making rapid concentration measurements, time-integrated air samples collected over longer time periods (e.g. dose rate measurements) (Astrup et al., 2004; Drews et al., 2004; Terada et al., 2012; Saunier et al., 2013; Katata et al., 2015) or from deposition measurements (Stohl et al., 2012; Winiarek et al., 2014). This is an active area of research, and numerous methods are being developed, enhanced, and then utilized for a wide range of agricultural, industrial, environmental, public health, transportation, and defense applications. There are numerous types of STE methods/approaches that are used for these applications.

Several of the STE methods are based on the reversal of one or more components of the transport and dispersion physics, so as to work backwards from concentration observations to source parameters. Traditional back-trajectory methods (Stohl, 1998; Jaffe et al., 1999; Hannan et al., 2000) reverse advection and ignore dispersion since these methods investigate back-trajectories on the fluid parcel scale. Entity back-tracking methods involve the back-tracking of larger elements, and thus do not ignore diffusion (Annunzio et al., 2012). For situations where the source and receptor locations are known, some methods use what is called an “inverse-dispersion” technique where a unit release from the known source location is scaled by the ratio of measured and predicted values (Flesch et al., 2005). Inverse models (Draxler and Hess, 1998; Sykes et al., 2008; Wang, 2008) reverse advection, but not dispersion, and so require post-processing to

deduce source parameters from a dispersion solution that expands backward in time. These methods are similar to adjoint models (Kaminski et al., 1999); however, for the former it is difficult to formulate an explicit adjoint. When available, adjoint models represent an elegant numerical tool to relate the sensors measurements to the sources (Pudykiewicz 1998, Krysta and Bocquet, 2007) and to efficiently conduct the corresponding inversion. Adjoint models furthermore form the cornerstone of the variational methods (Robinson and Langner 1998; Elbern and Schmidt 1999; Bieringer et al., 2010; Rodriguez et al., 2012; Bieringer et al., 2015) which go even further, alternating runs of the adjoint model with runs of the forward model to iteratively achieve an improvement of the source term parameters by matching predictions of the forward model to concentration observations. Bocquet (2005) extends the application of the adjoint model in the inversion process even further by generalizing the variational approach to situations when the probability distribution of the source is not Gaussian. This approach seeks to maximize the Entropy, which insures that no other information than the measurements and what is known from the source is used in the inversion (Krysta and Bocquet, 2007).

STE methods can also be based on nonlinear optimization approaches that do not use an adjoint model, but instead combine a forward transport and dispersion model with machine learning algorithms to iteratively refine estimates of the inputs to the forward model so as to improve the match between model output and concentration observations. There exist a wide variety of methods for iteratively improving a population of one or more trial solutions using this approach. These include genetic algorithms (Holland, 1975; Goldberg, 1989; Haupt, 2005; Allen et al., 2006, 2007; Haupt et al., 2006, 2007;

Long et al., 2010; Rodriguez et al., 2011; Schmehl et al., 2012), evolutionary strategy (Beyer and Schwefel, 2002; Cervone et al., 2009), simulated annealing (Press et al., 2002; Thomson et al., 2007), Bayesian inference with Markov Chain Monte Carlo (MCMC) sampling (Braak, 2006; Delle Monache et al., 2008; Robins et al., 2009), gradient descent (Bieringer et al., 2010; Rodriguez, 2012; Bieringer et al., 2015), particle swarm (Kennedy and Eberhart, 2001; Guohua and Longkai, 2014), four-dimension variational (Elbern and Schmidt 1999) and many others. All of these algorithms share the requirement for a component to create the initial trial solution(s), a component for updating them, and a component for evaluating their quality.

Detailed descriptions of these methods are provided in many of the individual papers mentioned above. Furthermore, surveys of many of these approaches exist in the literature and Rao (2007) and Redwood (2011) are examples of papers that summarize a variety of STE methods for localized sources and provide brief descriptions of them. The existing surveys focus on the differences between the methods, but do not examine them in the context of their similarities and how this relates to their utilization for the end application. Further, these surveys typically focus on the details of the implementation of the search for the source parameters, details that can vary widely between methods. What has been given less attention are the striking similarities that exist between this wide array of STE approaches for characterizing the release parameters from localized sources. In this manuscript we first provide a brief description of applications where STE methods are utilized to characterize the source parameters for localized sources. Next we provide a listing of the types of methods used for these applications, a brief description of these methods, and references that the reader can follow to find more detailed

descriptions of each method. In this section we also break these methods into three primary categories (i.e. 1. direct forward simulations; 2. direct inverse modeling; and 3. nonlinear optimization based methods). This categorization of the STE methods is then used to illustrate the similarities between the methods while also highlighting the respective strengths and weaknesses between the categories. We conclude this paper with a discussion that ties the categories of methods back to the applications described in Section 2 and the implications associated with their use relative to the application needs.

2. STE Applications

Over the past several decades the estimation of the source parameters associated with the localized release of a contaminant or other material into the atmosphere has become increasingly important. The applications for which this need arises span a broad set of technical disciplines and can for example range from agricultural, industrial, environmental, public health, transportation, to defense applications. For agricultural/public health applications STE methods are used in retrospective analyses for problems that seek to better understand the source of pollens and pollutants that result in allergic airway responses (Efsthathiou et al., 2011). On November 21, 2016, a severe thunderstorm developed in Melbourne, Australia leaving four dead and thousands requiring respiratory care due to the allergens and pollen. Pollutants and pathogens can also affect the safety of our food supply. Since the introduction of genetically modified (GM) crops, the estimation of pollen sources and dispersion have been the subject of an increasing number of studies due to the need to maintain seed purity and because cross contamination of seeds can have major economic impacts (Jarosz et. al., 2004). In

transportation applications STE is relevant for characterizing volcanic ash concentrations along aircraft flight routes. The volcanic eruption of Eyjafjallajökull in Iceland in 2010 led to extensive cancelations of commercial air travel over Europe and the North Atlantic due to concerns of aircraft engine failures associated with the ingestion of volcanic ash. The quantity of ash emitted from erupting volcanoes like Eyjafjallajökull is poorly understood, can be highly variable, and is critical for developing the ash cloud hazard area estimates. The combination of observations and atmospheric transport and dispersion (AT&D) models has been used to estimate the ash source term and can improve ash concentration forecasts thereby helping to reduce the large uncertainties in airspace closures and improve air-travel safety in impacted regions (Schmehl et al., 2012). Fortunately, no serious incidents were reported from this recent eruption due to aircraft flying through ash; however, the disruption of air travel led to large economic losses for the impacted region and industries.

There are several national defense related problems where STE methods are commonly applied. Because the intentional or accidental release of chemical, biological, radiological and nuclear (CBRN) materials remains an ongoing threat to the safety and security of civilian populations, many nations, including the United States, are investing heavily in rapid detection and enhanced situational awareness technologies to improve the outcomes for these incidents (White House, 2012; DoD, 2012). This problem can be particularly challenging because it is sometimes necessary to assume that some releases will be covert (Norige et al., 2009; Carrano and Jeys, 2010) and nearly all the source parameters are unknown. In situations where decisions associated with contamination avoidance are desired the STE method will be required to rapidly provide source

parameters that can be used in forward dispersion simulations of the downwind hazard prediction (Bieringer et al., 2010; Bieringer et al., 2015).

Another defense application for STE is associated with determining compliance with the Comprehensive Nuclear-Test-Ban Treaty (CTBT). The CTBT bans nuclear explosions in any environment and includes tests that might be conducted in the atmosphere, under water, and underground. Atmospheric modeling and STE methods are used extensively by the CTBT Organization (CTBTO) and affiliated scientists to identify the sources and the characteristics of gasses such as radioxenon that is a by-product of nuclear fission. The results of these analyses are then used to ensure the compliance of the treaty by signature countries and to build trust between participating nations (Wotawa et al., 2003). They have also been applied to evaluation of announced nuclear weapons tests conducted by non-signature nations like the Democratic People's Republic of Korea (DPRK) nuclear test conducted in 2013 (Ringbom et al., 2014).

In addition to the defense applications described above, there are additional high-consequence life and death situations where timely and accurate STE methods are desired. These situations typically involve the AT&D of hazardous pollutants from an industrial facility or a transportation accident where public health and safety are a critical factor. The incident at the Union Carbide India Limited pesticide plant in Bhopal, Madhya Pradesh, India on December 2-3, 1984 where over 500,000 people were exposed to methyl isocyanate (Varma and Varma, 2005), and the Chernobyl nuclear power plant accident in the former Soviet Union on 26 April, 1986 (Medvedev, 1992) are two such examples of high-consequence events that resulted in numerous injuries and fatalities. The breach in nuclear material containment at the Fukushima Dai-ichi nuclear power

plant (FD-NPP) FD-NPP following the Tohoku earthquake and tsunami provides another example of a similar, more recent, high-consequence event where STE played a role in the decisions that impacted the health and safety of large numbers of people (Onishi and Fackler, 2011; Stohl et al., 2012; Terada et al., 2012; Saunier et al., 2013; Pullen et al., 2013; Bieringer et. al., 2013; Winiarek et al., 2014; Katata et al., 2015; Kim et al., 2016). Here too radioxenon observations were used in STE analyses that evaluated the time varying radiation release from the Fukushima Dai-ichi nuclear power plant (FD-NPP) (Eslinger et al., 2014). In many of these high-consequence human health cases, methods for making more accurate source term estimates could have been used to better determine the downwind hazard and corresponding emergency response and radiation remediation decisions.

3. STE Method Categories and Descriptions

As described above, source term estimation has been, and continues to be, utilized in for a variety of problems. Due to the wide range of applications and technical disciplines involved in developing these approaches for STE, many of the methods and software systems for determining source parameters across these disparate efforts were developed independent of one another. The positive outcome from this evolution of STE within these individual communities is that it has led to a relatively wide variety of algorithmic, analytical, and combinations of methods being applied to this problem. The downside, however, is that it has led to a situation where the terminology used is highly variable, it has been difficult to compare and contrast the approaches, and it can be challenging for someone to select the best STE approach for their particular application. In this section

we attempt to bring clarity to this problem by focusing on categories of STE methods and providing examples of specific approaches used to characterize parameters for localized sources.

Here we postulate that STE methods commonly fall into three main categories. The first category involves methods that use direct forward simulations. The use of this method is typically confined to relatively simple scenarios where the locations of both the source and receptor (i.e. the observation) are known and STE is used to determine the release rate. The second category contains methods involving direct inverse modeling. This category of methods starts with the observations and then works backward (by a variety of means) to a source. The third category contains methods that use nonlinear optimization techniques in conjunction with a forward AT&D model solution to iteratively search for an optimal set of source parameters. The nonlinear optimization STE methods iteratively adjust source inputs in the forward model, and in some cases the meteorological inputs (which can include both observations and model based atmospheric reconstructions), to improve the match between model output and concentration observations. There is some commonality even between the three STE categories. Nearly all of them, with the exception of a very simple backward air parcel trajectory method, involve using an AT&D model, and all methods, whether they use an AT&D model or not, are sensitive to the accuracy of the weather conditions used to drive them.

3.1 Direct Forward Simulation Methods

One particularly simple, but common, scenario where STE is used involves situations where the locations of both the source and receptor (i.e. the observation) are known and

either a single instantaneous release or a continuous release at a constant rate can be assumed. For this application, STE is used to determine the release amount or rate. A direct forward simulation method is commonly used in this situation. It involves first running a forward in time AT&D model with a unit continuous release rate or instantaneous release. The dispersion solution from this forward simulation is then used to compute a dilution factor. The dilution factor is the ratio of the concentration computed at the receptor location over the release rate or mass and is meant to represent the decrease in concentration that occurs during the transport and dispersion of the material from the source to the receptor. The STE is then estimated by using an observation or series of observations (n) to effectively scale the unit release from the AT&D simulation and determine the release rate or mass:

$$\text{Average Release} = \text{Unit release} * \frac{\sum_{n=1}^n \frac{\text{Observed Concentration}_n}{\text{Predicted Concentration}_n}}{n} \quad (1)$$

Flesch et al., (2005) provides an example where this method is applied to the problem of characterizing gas emissions from farms. The strengths of this method are its simplicity, ease of implementation, and relatively low computational requirements. This approach can be very effective in situations where the meteorological conditions are well known and the AT&D model has sufficient fidelity to compute an accurate dilution factor. If sufficient computer resources are available, a grid of direct forward simulations can also be used to estimate source locations based on the percent contribution of simulated material transport from each individual source to receptor.

3.2 Inverse and Backtracking Methods

The approach for using backward or inverse computations was first put forth by Gifford (1959) where he provides a quantitative method for estimating urban air pollution concentrations from multiple sources. Since this seminal work by Gifford, inverse modeling has become widely used for STE applications. Within the direct inverse modeling category there are two primary subcategories. The first and most simple are methods that start from the observation location and use backward parcel trajectories to determine the air parcel source as shown in Figure 1.

Figure 1. Back trajectory schematic depicting receptors that did not detect a contaminant (black R), detected a contaminant (light blue R), the trajectories of the contaminant (orange lines), and the source of the contaminant (orange S).

Figure 1 shows a schematic of back trajectories of a contaminant from the triggered receptors (blue Rs) to the original source (orange S). This method has been used in the meteorological community to study long-range pollution transport and determine likely source locations (Stohl, 1998; Jaffe et al., 1999; Hannan et. al., 2000; Young et al., 2009). The strengths of this method are its simplicity, ease of implementation, and low computational requirements. Its weaknesses are that it is *highly* sensitive to the accuracy of the wind information. This category of methods is also not appropriate for estimation

of release rates or precise source locations since this approach typically does not model atmospheric effects beyond parcel transport. Moreover, a time-evolving range of wind directions must be known so that back trajectories initialized from different contaminant observation times can be computed to determine the intersection at the source location.

Figure 2. A direct inverse modeling schematic where in: (2A) we have a chemical release scenario in which a plume crosses a network of sensors; (2B) shows the first step of the STE process where the winds are reversed and contaminant releases are made from the triggered sensor locations; (2C) shows the STE process where again the winds are reversed and contaminant releases are made from the null sensor locations; and (2D) illustrates the final step in the process where the inverse contaminant release information from the triggered and null sensors are combined to triangulate the source parameters as shown in Bieringer et. al. 2015.

The second subcategory within the direct inverse modeling uses models that describe contaminant evolution with both transport and dispersion. Figure 2 shows an example of this methodology as describe in Bieringer et. al. (2015). Here the wind direction is reversed and used in a series of AT&D model simulations shown in Figure 2B and 2C.

Contaminant releases initiated at each observation location expand back towards the source region. Statistical post processing is then used to deduce, from the overlap of concentration fields resulting from these puffs/plumes, the source location as depicted in Figure 2D. If the source location is known, this method can also be used to determine the release mass, timing, and/or release rate.

This approach takes advantage of the mathematical form of the transport and dispersion equation to approximate an adjoint method. A complete mathematical description of the adjoint methods in this context was given by Pudykiewicz (1998). Briefly described, it begins by noting that the sensor measurements μ_i at the i^{th} sensor ($i=1, \dots, N$) can be expressed as a product of the sensor's response function $\pi_i(x, t)$ and the concentration $c(x, t)$:

$$\mu_i = \int \pi_i(x, t) c(x, t) dx dt + \varepsilon_i \quad (2)$$

ε_i is the measurement noise. The integral is over the model grid and the time period. However, the response function usually has the shape of a Dirac distribution in space and a Heaviside function in time (Pudykiewicz, 1998). So, in practice the integral is limited in space and time. The integral on the right hand-side of Equation 1 is akin to a scalar product (Krista and Bocquet, 2007), and can be rewritten into a more generic form: $\mu_i = \langle c, \pi_i \rangle$. Assuming a purely transport and dispersion model, the concentration c follows the equation:

$$\frac{\partial c}{\partial t} + \text{div}(\mathbf{u} c) - \text{div}(\mathbf{K} \nabla c) = \sigma \quad (3)$$

Where σ is the source term, $\text{div}(\mathbf{u} c)$ is the transport term, $\text{div}(\mathbf{K} \nabla c)$ is the dispersion, or diffusion, term. As a linear operator, the model can be written in a more compact form:

291 $Lc = \sigma$. Using the Lagrange duality relation (Pudykiewicz, 1998), there exists an operator
 292 L^* and we can define a state vector c_i^* such that:

$$293 \quad L^* c_i^* = \pi_i \quad \text{and} \quad \langle Lc, c^* \rangle = \langle c, L^* c^* \rangle \quad (4)$$

294 L^* is the adjoint operator and c^* is the adjoint state also called sometime, the
 295 retroplume (Issartel, 2003). Using the scalar product notation and equations (4), the first
 296 term of the right-hand side of (2) can then be re-written as:

$$297 \quad \langle c, \pi_i \rangle = \langle c, Lc_i^* \rangle = \langle Lc, c_i^* \rangle = \langle \sigma, c_i^* \rangle \quad (5)$$

298 or:

$$299 \quad \mu_i = \langle \sigma, c_i^* \rangle + \epsilon_i \quad (6)$$

300 Collecting all the sensor readings into a one vector μ and defining the matrix $H = c_{ik}^*$,
 301 where the indices correspond to the i^{th} sensor measurement and to the k^{th} model grid cell,
 302 equation (6) can be recast in a concise form:

$$303 \quad \mu = H\sigma + \epsilon \quad (7)$$

304 The interest in this formula is that the adjoint state c_i^* depends only on the
 305 measurements and can be pre-computed once for all. Thus, the problem of finding the
 306 sources responsible for the observed sensors measurements (2) and consistent with the
 307 model equations (3) can be reduced to a matrix inversion. An explicit expression of the
 308 adjoint operator L^* can be found using equation (4), see for example Pudykiewicz (1998)
 309 for the derivation of continuous model equations and Krista and Bocquet (2007) for
 310 discretized equations:

$$311 \quad -\frac{\partial c^*}{\partial t} - \text{div}(\mathbf{u} c^*) - \text{div}(\mathbf{K} \nabla c^*) = \pi \quad (8)$$

312 As we can see from equation (8), in the context of pure transport and dispersion, the
 313 adjoint equations can simply be obtained by reversing the wind direction and a backward

integrating from measurement time to release time. The use of the reverse model, therefore, finds its theoretical basis in the application of the adjoint methods.

In real world applications, however, models are not simply limited to transport and dispersion, but contain additional terms accounting for the physics of the contaminant under study (e.g. wet deposition, radioactive decay, etc.). Hourdin and Talagrand (2006) describe how to construct an adjoint operator for the most common terms found in transport and dispersion applications, provided that the corresponding physics is linear. Their approach is physically based: for example they justify the self-adjoint property of the diffusion term by the symmetry in time of atmospheric turbulence. Nonetheless, when it comes to using a computer-based model, the adjoint model must be constructed at the code level. Giering and Kaminski (1996) have established a list of coding rules to derive such an adjoint code from the code of the forward model. These rules are very systemic and have provided the basis for compiler-like automatic adjoint code generators. In addition to non-self-adjoint physical terms, the use of sophisticated numerical methods, e.g. puff splitting, to discretize the AT&D model may also cause the reverse model to differ from the adjoint model. In that case, attempting to invert equation (7) may simply fail or the solution will only approximate the true source.

Note also that, generally the matrix inversion in equation (7) is ill posed, since the number of sensor measurements is always lower than the number of grid cells. Various regularization methods have been proposed. All of these methods rely on additional *prior* information on the source that must be provided or assumed, e.g. smoothness, shape of the *prior* probability distribution, etc. Because the adjoint solution c^* , after normalization, satisfies the definition of a probability function, Pudykiewicz (1998)

337 postulated that the probability of finding the source at a given location and time is
338 proportional to the value of the retroplume. This approach avoids the matrix inversion,
339 but requires some statistical processing as in Bieringer et al., (2015). While
340 Pudykiewicz's hypothesis to relate the source probability density to retroplumes remains
341 unproved, no counter example has been published so far.

342 This is the same adjoint model that is used in variational methods, but with a different
343 forcing. Instead of being forced by the measurement function, the adjoint model is forced
344 by the difference between the observed and model-predicted sensor readings. As a result,
345 the adjoint state is now a difference, in fact a gradient, that can no longer be interpreted
346 as a retroplume, but as a sensitivity (Rabier et al., 1996; Carmichael et al., 1997).
347 Because variational methods are based on an iterative gradient descent, the numerical
348 operator L^* that is used must correspond to the adjoint operator of the discretized forward
349 operator L to machine accuracy. In other words, an approximate reverse model is not
350 valid with variational methods, but instead the computer code of the adjoint operator L^*
351 must be directly constructed from the code of the AT&D model L . Variational methods
352 are ubiquitous in meteorology (Courtier et al. 1994). Applications to AT&D problems
353 can be found in Roberson and Langner (1998), Elbern and Schmidt (1999) or Krista et
354 al., (2006).

355 This method has been implemented in several United States (US) modeling tools used
356 by the Department of Defense and National Oceanic and Atmospheric Administration.
357 Examples include adjoint models of the Second order Closure Integrated PUFF
358 (SCIPUFF) (Sykes et al., 2008) and HYSPLIT (Draxler and Hess, 1998; Stein et al.,
359 2015) AT&D modeling systems. While this approach is more capable than the simple

back trajectory method, it is still very sensitive to inaccuracies in the meteorological data used and can be computationally expensive in situations where there are a large number of concentration observations available for use.

These inversion-based methods have the advantage of being simple, and computationally efficient for the problems that do not require many inverse AT&D calculations. They also work well if the weather observations are accurate and representative. While not as accurate as the nonlinear optimization approaches discussed in the next section, inverse models can yield a trial solution suitable for initializing the iterative optimization methods described next. These later methods also offer the potential for inferring the weather information rather than being sensitive to it.

3.3 Nonlinear Optimization Approaches

In contrast to the inverse methods discussed above, iterative nonlinear optimization methods utilize an atmospheric AT&D model repeatedly run in a forward-in-time configuration. The model is run repeatedly, with the errors in each forecast of the concentration field being used to devise a potentially better estimate of the forward model's input variables (e.g. source and weather parameters). The various optimization-based methods differ primarily in how they obtain the improved estimate of these input variables. While all attempt to move the forward model inputs in a direction that produces less error in the forward model output, some use sampling to estimate this gradient while others must compute it explicitly.

All nonlinear optimization based methods for STE can, therefore, be viewed as searching for a set of forward model input values that maximize the skill of the forward

model (i.e. minimizes error and cost and maximizes fitness). The problem can be thought of as a search for the lowest point in the multi-dimensional surface that plots cost as a function of the STE parameters. Figure 3 from Rodriguez (2012) provides an example of a STE cost surface developed from a tracer gas measurement collected during the Fusing Sensor Information from Observing Networks (FUSION) Field Trial 2007 (FFT 07) (Storwald et al., 2007). Here the cost surface characterizes the degree to which a AT&D model solution matches the observations that were collected from the network of sensors, represented by the lines of dots located at the bottom left of the image, as a function of the source location. This example illustrates how the cost surface can be quite complex and contain multiple local minima (Robins et al., 2009; Rodriguez, 2012). The solution sought, however, is the global cost minimum (i.e. the deepest of the local minima). Each local minimum is surrounded by a range of input values for which a simple gradient descent would lead to that particular local minimum. This region of solution space can be thought of as the attractor basin of the corresponding local minimum. This perspective will provide insight into many aspects of the optimization-based STE methods described below.

There are numerous nonlinear optimization techniques that do not require a direct calculation of the local gradient of skill (i.e. inferring this gradient from a population or history of trial solutions) (Holland, 1975; Press et al., 2002; Sarker et al., 2003; Jones, 2002; Keats et al., 2007). Because of their extensive use in STE we will focus on a family of evolutionary algorithms including simulated annealing, evolutionary strategy, genetic algorithms, and on the related Bayesian inference methods. We conclude this

section with a description of an adjoint model based gradient descent approach and discussion of approaches for tuning these nonlinear optimization approaches.

All of the nonlinear optimization algorithms described below follow the same broad set of steps to iteratively evolve a population of one or more trial solutions:

1. Specify an initial population of trial solutions;
2. Compute the skill metric for each trial solution;
3. Modify the high-skill trial solutions;
4. Repeat steps 2 and 3 until convergence or the computational cost limit reached.

Across these optimization methods, the measure of skill used in step 2 is based on a comparison of the model solution to observations. Thus, the methods differ primarily in population size (step 1), the metric used to quantify skill (step 2), and in how the population is evolved (step 3). It is differences in implementation of these steps that distinguish the methods discussed below.

Figure 3. A three-dimensional cost function surface for a multi-dimensional STE problem that includes: location, time, mass, wind speed, and wind direction.

The solution improvement step (3) is composed of two subcomponents: selecting the trial solution(s) to continue evolving and modifying them in hopes of obtaining an improvement in forward model skill. Improvement in the trial solution results from both exploitation of cost gradient estimates (obtained from existing trial solutions) to design the next generation, and subsequent culling of that population. Methods differ in the emphasis they place on these two aspects. Although the approaches taken to the two tasks vary widely between optimization methods, they are modular enough that potential exists to combine aspects of different methods in order to meet the needs of a particular STE problem. The following sections provide additional detail and examples of different nonlinear methods used for STE.

3.3.1 Simulated Annealing

Simulated annealing follows all the steps associated with nonlinear STE methods described above. It takes the simplest possible approach to step 1, starting with a single expert-selected trial solution (Thomson et al., 2007). For step 2 the skill metric in simulated annealing is often computed as the sum of squared differences between the AT&D model prediction and the concentration observations (called energy). Any metric summarizing this prediction error can, however, be used. For step 3 simulated annealing executes the solution modification first, making a random jump to a nearby point in solution space (i.e. “moving the system”). This operation yields a population of two trial solutions (original and perturbed). The decision as to which member to retain for the next generation is made probabilistically with the solution yielding greatest skill being favored. By allowing occasional jumps to lower skill solutions, simulated annealing can avoid converging into a local, rather than global, optimum. Achieving this while eventually settling into the global optimum requires a gradual decrease in the probability of picking the lower skill solution (Press et al., 2002). The success and efficiency of simulated annealing methods thus depend on having selected an appropriate schedule for this annealing process (Thomson et al., 2007). Because simulated annealing was designed to be analogous to the physics of annealing in metallurgy, the parameter that controls these odds is called *temperature*, its change is called an *annealing schedule* and fitness is called *energy* (Press et al., 2002). Equation 9 illustrates step 3 for simulated annealing:

$$s' = s_i + \text{random jump} \quad (9)$$

$$s_{i+i} = s' \text{ with probability } P(e_i, e', T), \text{ else } s_{i+i} = s_i$$

where s is the trial solution, i is the iteration count, s' is a potential new solution, e is the energy (i.e. skill metric) and T is the temperature.

3.3.2 Evolutionary Strategy and Genetic Algorithms

Evolutionary strategy and genetic algorithm approaches have many characteristics that are similar to simulated annealing. For example, the initial step for the evolutionary strategy approach is much like simulated annealing in that the simplest form of the evolutionary strategy uses a single expert-selected trial solution (Beyer and Schwefel, 2002). More complex evolutionary strategies and genetic algorithms, however use multiple, randomly selected trial solutions within an expert-selected range (Cervone et al., 2010; Rodriguez et al., 2011). The latter approach offers advantages in searching the parameter space of the STE problem (Haupt, 2005; Haupt et al., 2006, 2007; Robins et al., 2009) multiple times, thereby increasing the chances of avoiding convergence to a locally optimal solution rather than to the global optimum.

For step 2 the evolutionary strategy and genetic algorithms can exploit any skill metric, but performance can depend strongly on the metric selected. For these approaches the skill metric is often called a cost function and decreases as the solution improves. Metrics based on the logarithm of concentration yield more skill in finding source location than in determining source strength, as they are sensitive to the overall shape of the concentration field (Allen et al., 2007; Rodriguez et al., 2011). In contrast, Rodriguez et al. (2011) demonstrated that making the cost function linear in concentration yields better estimates of source strength. Both advantages can be

478 achieved by solving the full STE problem with a logarithmic cost function and then
479 solving again for just source strength using a linear cost function (Rodriguez et al., 2011).

480 Sensor error can be accounted for by a term in the cost function (Allen et al., 2007;
481 Long et al., 2010) by setting all concentration predictions and observations less than the
482 sensor detection to that value (Delle Monache et al., 2008). One of these methods must
483 be used if log concentration appears in the cost or likelihood function so as to avoid
484 infinite contributions from sensors outside the contaminant field.

485 The solution modification step (i.e. step 3) in evolutionary strategy is implemented
486 similarly to that in simulated annealing, although the heuristic for evolving the standard
487 deviation of the perturbations is called self-adaptation. Because evolutionary strategy
488 precludes jumps to lower skill solutions, it favors descent over search when compared to
489 simulated annealing. Inclusion of multiple child solutions instead of just the one used in
490 simulated annealing can compensate, thus the number of offspring can be used to tune the
491 balance between search and descent in evolutionary strategy methods.

492 Genetic algorithms and evolutionary strategies retain the random jump feature of
493 simulated annealing, calling it mutation in keeping with the methods' evolutionary basis.
494 Evolutionary strategies differ from simulated annealing, however, in that the best solution
495 is always selected and that multiple perturbed "child" solutions can be included along
496 with the "parent" solution in the population from which this selection is made.
497 Convergence to a local optimum is prevented by this use of multiple offspring rather than
498 probabilistic acceptance of less skilled solutions. Genetic algorithms extend this
499 biological analogy (Haupt, 2005; Haupt et al., 2006, 2007) to include information
500 exchange between parents (usually 2) selected by fitness (i.e. breeding). As with

501 simulated annealing, the genetic algorithm and evolutionary strategy methods select
502 parents based on fitness or fitness ranking. The odds of a trial solution being selected to
503 breed are thus set proportional to either fitness or fitness rank. Fitness ranking helps to
504 avoid premature convergence on one particularly fit, but not optimal, solution. The
505 corresponding drawback fitness/rank proportional breeding, loss of good solutions from
506 prior generations, can be eliminated in the genetic algorithm method by using an elitist
507 strategy that retains for the next generation the best trial solution (or solutions) of the
508 current population (Haupt, 2005; Haupt et al., 2006, 2007). The degree of elitism in the
509 genetic algorithm or evolutionary strategy can be used as a design tradeoff that enables
510 the method to prioritize search vs. solution convergence and be tuned to meet the
511 requirements of the STE method implementation. Population size also impacts this
512 tradeoff with larger populations favoring search (e.g. Goldberg, 1989), although smaller
513 populations also work well if combined with larger mutation rates (Haupt and Haupt,
514 2000; Haupt and Haupt, 2004). As with simulated annealing, the success and efficiency
515 of the genetic algorithm method depends on the weighting given to search and descent.

516 Many methods exist for information exchange between parents, some are better suited
517 than others for particular STE cost surface structures. Linear mixes of parent solutions
518 can be particularly good at moving the population along gently sloping valleys in the cost
519 surface (Rodriguez, 2012) Crossover, where portions of one solution are exchanged with
520 another, sacrifices this valley running prowess for more equitable search of the solution
521 space (Haupt 2005; Haupt et al., 2006; 2007). This operator can be used in any method
522 that exploits a population of trial solutions, e.g. in genetic algorithms (Haupt 2005; Haupt

et al., 2006; 2007) and in evolutionary strategy (Breyer and Schwefel, 2002). Equation 10 is representative of step 3 for evolutionary strategy and genetic algorithm approaches:

$$\text{Step3: } s_{i+1,j} = M(B(s_{i,k}, s_{i,k})) \quad (10)$$

where $s_{i,j}$ is the j th member of the i th generation of a population of trial solutions, k and m indicate the selected parent trial solutions, B the breeding function and M is the mutation function.

3.3.3 Bayesian Inference

Bayesian inference based methods theoretically provide one of the most sophisticated approaches used for STE. The fundament of the Bayesian inference approach is the Bayes theorem:

$$p(\sigma|\mu) = \frac{p(\mu|\sigma)*p(\sigma)}{p(\mu)} \quad (11)$$

Where $p(\sigma|\mu)$ is the probability of the source to be σ given that the measurement that has been observed is μ . This probability can only be known after that the measurement has been taken. For this reason, it named the *posterior* conditional probability. $p(\mu/\sigma)$ is the probability that, for a given source σ , the sensor reading will be μ . This function only depends on the sensor characteristics and the model used to propagate contaminant from the sources at release time to the sensors at later time. This probability can therefore be evaluated even before measurements take place, and is called the likelihood function. $p(\sigma)$ is the probability for the source to be σ . It is called the *prior*, because it encapsulates all *prior* information available on the source. $p(\mu)$ is the probability for the sensor reading to be μ . Because $p(\mu)$ is positive and independent of any sources, the *posterior* probability

function, considered as a function of the source σ , is proportional to the product of the likelihood and prior functions:

$$p(\sigma|\mu) \propto p(\mu|\sigma) * p(\sigma) \quad (12)$$

This relation provides the basis for the estimation of the source σ . Although, theoretically, the best linear unbiased estimator will be the mean of $p(\sigma/\mu)$, most source inversion algorithms take advantage of techniques from optimization and rather search for the mode of $p(\sigma/\mu)$. This leads to the so-called Maximum *a posteriori* estimator. When the likelihood and the prior probability functions are Gaussian, then the posterior is also Gaussian and the mode is also the mean of the function. Under those hypotheses, the estimation is optimal.

Bayesian inference, while using a single trial solution, is sometimes run as an ensemble of non-interacting runs (Delle Monache et al., 2008) or interacting runs (Braak, 2006; Robins et al., 2009) to yield this same parallel search advantage for step 1. The disadvantage of using multi-member populations in situations where this parallel search is unneeded is unnecessary computational cost. For step 2 the Bayesian methods follow the same sequence as evolutionary strategy and genetic algorithms and differ only in the details of population size and how high-skill trial solutions are selected and modified (Delle Monache et al., 2008; Robins et al., 2009). As with genetic algorithms, Bayesian inference using a skill metric based on the logarithm of concentration yields more skill in finding source location than in determining source strength. When a skill metric linear in concentration is used it yields better estimates of source strength (Delle Monache et al., 2008). Sensor error can be accounted for by inclusion of a sensor error model in the likelihood function of Bayesian inference Monte Carlo Markov chain (Delle Monache et

al., 2008; Robins et al., 2009) or by setting all concentration predictions and observations less than the sensor detection to that value (Delle Monache et al., 2008).

For step 3 in the Bayesian inference method, the odds of a new trial solution being accepted increases with the likelihood of the observations given that trial solution (Delle Monache et al., 2008). This approach is more theoretically sophisticated than the otherwise similar probabilistic acceptance used in simulated annealing and fitness/rank proportional breeding used in genetic algorithms. The probability of accepting a less skilled solution at a particular iteration can be a predefined function of iteration count or be made to depend on skill history of the evolving solution up to that point (Delle Monache et al., 2008).

Differential Evolution Monte Carlo takes the breeding approach one step further than genetic algorithms as it uses a three-parent breeding scheme exchanging information between multiple Markov chains (Braak, 2006; Robins et al., 2009). This advantage would also accrue to any other breeding method that distributes the offspring preferentially along the valley floor (i.e. in directions for which the cost did not increase). Such cost valleys are frequent in localized STE problems where they result from two or more STE parameters having a similar impact on the concentration field (Rodriguez, 2012; Bieringer et al., 2015).

3.3.4 Explicit Gradient Descent Methods

Many of the same steps described above apply to iterative methods that explicitly compute the cost gradient. Explicitly-calculated gradient descent methods differ from the sample based methods that will be discussed above in that step 2 becomes the explicit

calculation of the local gradient of skill with respect to all of the STE parameters, rather than calculation of the skill itself. The other three steps listed above are effectively same as the other nonlinear methods. Because step 3 in gradient descent methods yields new solutions within the same solution basin, step 1 must provide these methods with an initial solution that lies within the attractor basin of the global minimum. This requirement places a considerable burden on the expert tasked with selecting the initial solution, particularly since there is no way to tell if the wrong basin was selected without a computationally expensive sampling of the solution space (Allen et al., 2007). A back-trajectory approach (Bieringer et al., 2010; Rodriguez, 2012; Bieringer et al., 2015) or another nonlinear optimization approach can, however, be used with a gradient descent method (Long et al., 2010) to increase the odds of starting within the correct basin.

The explicit gradient descent approach for STE relies on the utilization of variational methods that have been successfully applied to nonlinear optimization problems (Roberston and Langner, 1998 and Krysta and Bocquet, 2006). In general, though, the extension of the variational approach to nonlinear problems prerequisites the linearization of the model equations (Courtier et al., 1994). The linearized, or tangent, model and its adjoint counterpart are applied onto perturbations around the linearized state. The optimal perturbation resulting from the variational solution of the linear problem is then added to the nonlinear state. After linearization again, a new optimal perturbation is sought. The approach is, therefore, comprised of two levels of iterations: an outer loop where the full nonlinear model computes the exact solution and an inner loop in which the adjoint-based variational method is applied. In general, a gradient descent-like algorithm is used to solve the linear variational problem. Bieringer et al, (2015) used this incremental

variational approach; with a nuance in the sense that they use a simplified model, with its adjoint, in the inner loops. This is similar to the ECMWF operational numerical weather prediction system, in which a model with simplified physics and lower resolution has been developed for the inner loops. Usually, those surrogate models are intended to have a better linear behavior and be computationally more efficient.

The explicit gradient methods execute step 3 by simply shifting the trial solution down the local cost gradient with each iteration. As described above, this is conceptually similar to the methods that use an AT&D model where the winds are reversed (Draxler and Hess, 1998; Sykes et al., 2008) except that here an explicit mathematical adjoint model is used in place of reversing the winds and the heuristic post processing that is required to estimate the source. The main strength of the gradient descent approach is that it is a computationally inexpensive search method in which large quantities of data can be utilized to efficiently search through a parameter space containing multiple STE parameters (e.g. location, time, release rate/mass, wind speed, direction, etc.). For this reason, it is a good method for scenarios where little prior STE information is known, there is significant uncertainty in the meteorological conditions, and large quantities of data are available. The weakness of the gradient descent method is that it can be difficult to develop and maintain the tangent linear and numerical adjoint models and that it must be provided with an initial solution that lies within the attractor basin of the global optimum. As mentioned above, this requirement places a considerable burden on the expert tasked with selecting the initial solution, particularly since there is no way to unequivocally tell if the wrong basin was selected without a computationally expensive sampling of solution space. The method, however, has been shown to be effective at

providing rapid response STE solutions when coupled with a back trajectory based approach to provide a sufficiently accurate first guess (Bieringer et al., 2010; Bieringer et al., 2015).

Variational data assimilation methods (Krysta and Bocquet, 2006; Rodriguez, 2012; Bieringer et al., 2015) use both forward-in-time and backward-in-time models of AT&D and so are in some ways hybrids of inverse models and nonlinear optimization methods. The Variational Iterative Refinement STE Algorithm (VIRSA) (Bieringer et. al., 2015), for example, uses a hybrid Lagrangian-Eulerian Plume Model (LEPM), and the LEPM's formal adjoint. The variational method is initialized with a single trial solution (Step 1). It then iteratively refines the trial solution by using the forward AT&D model (Step 2) and its adjoint (Step 3) to identify the source parameter values that minimize the difference between the forward model predictions of contaminant concentration and the sensor observations. As in all iterative STE methods, this difference is quantified via a cost function (Step 2). The choice of forward model is limited only by the requirement to derive the adjoint of its tangent-linear approximation. Likewise, the cost function can be adapted to fit a particular STE problem. This method has also been used to refine meteorological parameters such as wind speed and direction along with the more traditional STE variables (i.e. release rate/mass, time, location) (Bieringer et al., 2015).

3.3.5 Tuning Nonlinear-Optimization

All of these optimization methods start from an expert-defined first guess at the region of solution space to search, but, with exception of the gradient descent methods, can break out of a sub-optimum cost basin if properly tuned. The challenge in this tuning

is to find the proper balance between searching for the attractor basin of the global optimum and descent of the current basin towards its minimum-cost solution (Haupt, 2005; Haupt et al., 2006, 2007). All nonlinear optimization methods used for localized STE have one or more parameters, which can be tuned to adjust this balance. For those methods with multi-member populations of trial solutions, achieving the correct balance may not be required to find the global optimum, since members can start in various attractor basins, but not finding this balance can significantly increase computational requirements (Delle Monache et al., 2008).

All of the nonlinear optimization methods described above require that the skill of the forward model linking STE parameters to the concentration field be assessed via a scalar metric (i.e. a skill score, error statistic, cost function or likelihood function). The skill metric should be chosen for computational efficiency, particularly if a large number of forward model runs are required by the solution algorithm chosen. Equally important is to select a skill metric that is both sensitive to the STE parameters sought and relatively insensitive to noise in the concentration observations (Rodriguez et al., 2011). Meeting the latter constraint often requires considerable thought and experimentation, but can pay great dividends in terms of accuracy and efficiency of the STE method.

Most of the evolutionary algorithms are robust in the face of local optima, but only if the correct balance of search and descent is achieved. Modes of failure thus include premature convergence if information is lost too fast in the selection process (only a problem with the genetic algorithm methods) or the mutation rate decreases too rapidly (primarily an issue in the simulated annealing and evolutionary strategy methods). These algorithms can also converge to local optima if descent is emphasized over search, or

wander forever if search is emphasized over descent. Tuning of the search versus descent is thus crucial and problem dependent. Expert knowledge helps for this tuning, but is not as crucial for picking the initial solution(s) as it is for gradient descent computing methods.

Evolutionary algorithms tend to be computationally expensive because the skill metric must be computed for each trial solution. For example, it is not uncommon for a genetic algorithm to require the evaluation of 10^3 to 10^4 trial solutions on an STE problem (Kuroki et al., 2010; Robins et al., 2009). This is also true for the Bayesian inference methods and often limits their use to less complex STE problems where fewer unknown variables reduces the number of trial solutions and associated computational requirements. For simulated annealing, evolutionary algorithms, and Bayesian inference methods, each of these solution evaluations involves running the forward model, so computational cost can also be large if the forward model is complex.

4. Conclusions

In our presentation of source term estimation methods utilized for localized STE problems, we postulate that the approaches fall into three broad categories, forward modeling, inverse modeling, and nonlinear optimization. All of these methods work from environmental and contaminant concentration, dosage, or deposition observations to achieve estimates of some combination of source variables such location, time, and quantity. The methods primarily differ in how they achieve this mapping. The purely forward method, appropriate for use in situations where the source and receptor locations are known, uses the AT&D model to compute a dilution factor from which an estimate of

707 the release mass can be determined. When an ensemble of forward simulations using
708 varied source locations is used, this purely forward approach can be extended to also
709 estimate location, albeit at considerable computational expense. The basic inverse
710 methods rely on time reversing one or more of the physical processes in the AT&D
711 model. Trajectory methods and inverse models merely reverse the transport and are
712 conceptually analogous to methods that use a formal mathematical adjoint. The
713 information content of multiple concentration observations can be exploited by forming a
714 statistical consensus of backward-in-time runs from each concentration observation, as in
715 the trajectory and inverse model methods. Formal mathematical adjoint models can also
716 be utilized in STE and have the advantage of being a computationally efficient method.
717 Significant effort, however, is typically required to develop and maintain the linked
718 combination of forward and tangent linear models required to implement a mathematical
719 adjoint for STE. These trajectory and inverse model methods are typically
720 computationally efficient, but are very sensitive to the accuracy of the available weather
721 data.

722 Nonlinear optimization provides a means of source term estimation using only a
723 forward model of transport and dispersion. These iterative methods implement a
724 common set of steps: specify an initial population of trial solutions, compute the skill
725 metric for each trial solution, modify the high-skill trial solutions, repeat until
726 convergence. Although the various methods differ greatly in how each of these steps is
727 implemented, the steps themselves are largely modular allowing a mix-and-match
728 approach to achieve good performance on a particular problem. Indeed, the latest
729 evolutionary strategy and Bayesian inference Monte Carlo methods implement many of

the features of genetic algorithms including the multi-member trial solution populations and exchange of information between members of this population to create the next generation of trial solutions. These tactics help ensure that the iteration converges to the globally optimal solution to the STE value rather than become trapped in a locally optimum solution. The price for this robustness is increased computational cost, so these more advanced methods should be used only when the problem's cost surface is so convoluted that the initial population of trial solutions cannot be guaranteed to lie within the cost basin as the global optimum, and time / computational resources are available.

The variational method can, like the other nonlinear methods, be used to iteratively improve an initial guess at the source parameters by alternately running the forward AT&D model and its backward-in-time adjoint until the forward model forecast resembles the concentration observations. The variational method is also computationally efficient, but like any method involving gradient descent, can become trapped in local optima of the cost function and so requires a good initial guess from some other source. Used in combination, these approaches can succeed with the inverse method providing the initial guess for the variational method.

Inverse modeling time reverses the physics of transport and dispersion while iterative optimization runs the physics forward in time for an iteratively improved set of estimated source variables. Some methods, such as the variational approach, use both forward and backward models iteratively. The meteorological parameters of AT&D can be either provided with the concentration data or deduced from it along with the source variables. The choice depends on the relative availability of meteorological and concentration data. Given the component interchangeability discussed above, the basic framework for STE is

less relevant except for the Bayesian inference methods' inherent advantage in providing a straightforward means of estimating solution uncertainty.

Acknowledgements:

The authors would like to acknowledge the United States (US) National Science Foundation (NSF) for providing funding to support the 2012 Workshop on Source Term Estimation Methods for Estimating the Atmospheric Radiation Release from the Fukushima Daiichi Nuclear Power Plant. We would also like to thank Dr. Branko Kosovic, Dr. Steve Hanna, and Dr. Ryohii Ohba and all of the participants of the workshop for contributing their time and expertise to that meeting. The contributions of the workshop participants with their diverse expertise clarified many of the commonalities in the approaches and served as the inspiration for this manuscript.

5. References

- Allen C.T., Young G.S., Haupt S.E., 2006. Improving Pollutant Source Characterization by Optimizing Meteorological Data with a Genetic Algorithm. *Atmos. Environ.* **41**: 2283-2289.
- Allen C.T., Haupt S.E., Young G.S., 2007. Source Characterization with a Receptor/Dispersion Model Coupled with a Genetic Algorithm. *J. Appl. Meteor. Climatol.* **46**: 273-287.
- Annunzio A.J., Young G.S., Haupt S.E., 2012. Utilizing state estimation to determine the source location for a contaminant. *Atmos. Environ.* **46**: 580-589.

779 Astrup, P., Turcanu, C., Puch, R., Palma, C. R., and Mikkelsen, T., 2004. Data
 780 assimilation in the early phase: Kalman filtering RIMPUFF, Riso National
 781 Laboratory report Riso-R-1466(EN).
 782 Beyer H-G, Schwefel H-P., 2002. Evolution Strategies: A Comprehensive Introduction.
 783 *Nat. Comput.* **1**: 3-52.
 784 Bieringer P.E., Sykes I., Vandenberghe F., Hurst J., Weil J., Bieberbach G., Parker S.,
 785 Cabell R., 2010. Automated source parameter estimation for atmospheric transport
 786 and dispersion applications. *13th International Conference on Harmonization within*
 787 *Atmospheric Dispersion Modeling for Regulatory Purposes*. 1-4, June, Paris, France.
 788 Bieringer P.E., Hanna S., Young G.S., Kosovic B., Hannan J., Ohba R., 2013. Methods
 789 for estimating the radiation release from the Fukushima Dai-ichi nuclear power plant.
 790 *Bull. Amer. Meteor. Soc.*, **94**: ES1-ES4.
 791 Bieringer, P.E., Vandenberghe F., Sykes I., Hannan J., Hurst J., Bieberbach G.,
 792 Rodriguez L.M., and Fry R., 2015. Automated Source Parameter and Low Level
 793 Wind Estimation for Atmospheric Transport and Dispersion Applications., *Atmos.*
 794 *Environ.* **122**, pp. 206-219.
 795 Beyer H-G, Schwefel H-P., 2002. Evolution Strategies: A Comprehensive Introduction.
 796 *Nat. Comput.* **1**: 3-52.
 797 Bocquet M., 2005. Reconstruction of an atmospheric tracer source using the principle of
 798 maximum entropy. II: Applications. *Q. J. R. Meteorol. Soc.* **131**: 2209–2223.
 799 Bocquet M., 2007. High-resolution reconstruction of a tracer dispersion event:
 800 application to ETEX. *Q. J. R. Meteorol. Soc.* **133**: 1013–1026.

801 Braak CJ., 2006. A Markov chain Monte Carlo version of the genetic algorithm
802 Differential Evolution: easy Bayesian computing for real parameter spaces. *Stat.*
803 *Comput.* **16**: 230–249.

804 Carmichael GR, A. Sandu, F. A. Potras, 1997: Sensitivity analysis for atmospheric
805 chemistry models via automatic differentiation. *Atmos. Environ.* 31: 475–489.

806 Carrano J, Jeys T., 2010. Chemical and Biological Sensor Standards Study II. Defense
807 Threat Reduction Agency Rep., 76 pp. [Available online at
808 [http://www.dtra.mil/docs/system-](http://www.dtra.mil/docs/system-documents/Chem_Bio_Sensor_Standards_Study_Vol_2_Oct_2010.pdf?sfvrsn50)
809 [documents/Chem_Bio_Sensor_Standards_Study_Vol_2_Oct_2010.pdf?sfvrsn50.](http://www.dtra.mil/docs/system-documents/Chem_Bio_Sensor_Standards_Study_Vol_2_Oct_2010.pdf?sfvrsn50)]

810 Cervone G., Franzese P., Grajdeanu A., 2010. Characterization of atmospheric
811 contaminant sources using adaptive evolutionary algorithms. *Atmos. Environ.* **44**:
812 3787-3796.

813 Courtier P., J.-N. Thepaut, and A. Hollingsworth, 1994: A strategy for operational
814 implementation of 4D-Var using an incremental approach. *Q. J. R. Meteorol. Soc.*
815 **120**, 1367-1387.

816 Delle Monache L., Lundquist J.K., Kosović B., Johannesson G., Dyer K.M., Aines R.D.,
817 Chow F.K., Belles R.D., Hanley W.G., Larsen S.C., Loosmore G.A., Nitao J.J.,
818 Sugiyama G.A., Vogt P.J., 2008. Bayesian Inference and Markov Chain Monte Carlo
819 Sampling to Reconstruct a Contaminant Source on a Continental Scale. *J. Appl.*
820 *Meteor. Climatol.* **47**: 2600–2613.

821 DoD., 2012. The Department of Defense Fiscal Year (FY) 2013 President’s Budget
822 Submission. [Available online at:

http://comptroller.defense.gov/defbudget/fy2013/budget_justification/pdfs/02_Procurement/Chemical_and_Biological_Defense_Program_PB_2013.pdf].

Draxler R.R., Hess G.D., 1998. An Overview of the HYSPLIT_4 Modelling System for Trajectories, Dispersion, and Deposition. *Aust. Meteorol. Mag.* **47**: 295-308.

Drews, M., Lauritzen, B., Madsen, H., and Smith, J. Q., 2004. Kalman filtration of radiation monitoring data from atmospheric dispersion of radioactive materials, *Radiat. Prot. Dosim.*, **111**, 257–269.

Efsthathiou C., Isukapalli S., Georgopoulos P., 2011. A mechanistic modeling system for estimating large-scale emissions and transport of pollen and co-allergens. *Atmos. Environ.* **45**: 2260-2276.

Elbern H, H. Schmidt, 1999: A four-dimensional variational chemistry data assimilation scheme for eulerian chemistry transport modeling. *J. Geophys. Res.* **104**: 18583–18598.

Eslinger, P.W., Biegalski S.R., Bowyer T.W., Cooper M.W., Haas D.A., Hayes J.C., Hoffman I., Korpach E., Yi J., Miley H.S., Rishel J.P., Ungar K., White B., Woods V.T., 2014. Source term estimation of radioxenon released from the Fukushima Dai-ichi nuclear reactors using measured air concentrations and atmospheric transport modeling. *Journal of Environmental Radioactivity*, **127**: 127-132.

Errico R.M., 1997. What is an adjoint model? *Bull. Amer. Meteor. Soc.* **78**: 2577–2591.

Flesch T.K., Wilson J.D., Harper L.A., Crenna B.P., 2005. Estimating gas emissions from a farm with an inverse-dispersion technique. *Atmos. Environ.* **39**: 4863-4874.

844 Giering, R. and T. Kaminski, 1996: *Recipes for adjoint code construction*. Internal
845 Report 212 from Max-Planck Institute fur Meteorologie, Hamburg, Germany.
846 35pp.

847 Gifford F., 1959. Computation of pollution from several sources. *Int. J. Air Poll.*, **2**, 109-
848 110.

849 Gilks W.R., Richardson S., Spiegelhalter D.J., 1996. Markov Chain Monte Carlo in
850 Practice. Chapman & Hall, London, UK.

851 Goldberg D.E., 1989. Genetic Algorithms in Search, Optimization, and Machine
852 Learning. Addison-Wesley Longman Publishing Co., Inc. Boston, MA.

853 Guohua C., Longkai C., 2014. Enhancing situation awareness of chemical release through
854 source inversion. *Procedia Eng.* **84**: 742-751.

855 Hannan J.R., Fuelberg H.E., Thompson A.M., Bieberbach Jr. G., Knabb R.D., Kondo Y.,
856 Anderson B.E., Browell E.V., Gregory G.L., Sachse G.W., Singh H.B., 2000.
857 Atmospheric chemical transport based on high-resolution model-derived winds: A
858 case study. *J. Geophys. Res.* **105**: 3807-3820.

859 Hastings W.K., 1970. Monte carlo sampling methods using markov chains and their
860 applications. *Biometrika*, **57**: 97-109.

861 Haupt, R.L. and S.E. Haupt, 2000: Optimum Population Size and Mutation Rate for a
862 Simple Real Genetic Algorithm that Optimizes Array Factors, *Appl. Comput.*
863 *Electrom.*, **15**, No. 2, pp. 94-102.

864 Haupt, R.L., and S.E. Haupt, 2004: *Practical Genetic Algorithms, Second Edition with*
865 *CD*, John Wiley and Sons, New York, NY, 255 pp.

866 Haupt S.E., 2005. A demonstration of coupled receptor/dispersion modeling with a
 867 genetic algorithm. *Atmos. Environ.* **39**: 7181-7189.

868 Haupt S.E., Young G.S., Allen C.T., 2006. Validation of a receptor/dispersion model
 869 with a genetic algorithm using synthetic data. *J. Appl. Meteor.* **45**: 476-490.

870 Haupt S.E., Young G.S., Allen C.T., 2007. A genetic algorithm method to assimilate
 871 sensor data for a toxic contaminant release. *J. Cp.* **2**: 85-93.

872 Holland J.H., 1975. Adaptation in Natural and Artificial Systems. The University of
 873 Michigan Press, Ann Arbor, MI, USA.

874 Hourdin, F. and O. Talagrand, 2006: Eulerian backtracking of atmospheric tracers: I
 875 Adjoint derivation and parametrization of subgrid-scale transport, *Q. J. R. Meteorol.*
 876 *Soc* **132** : 567-583.

877 Issartel J-P. 2003: Rebuilding sources of linear tracers after atmospheric concentration
 878 measurements. *Atmos. Chem. Phys.* **3**: 2111 – 2125.

879 Jaffe D., Anderson T., Covert D., Kotchenruther R., Barbara B., Danielson J., Simpson W.,
 880 Bernsten T., Karlsdottir S., Blake D., Harris J., Carmichael G., Uno I., 1999.
 881 Transport of Asian air pollution to North America. *Geophys. Res. Lett.* **26**: 711-714.

882 Jarosz N, Loubet B, Hubert, 2004. Modelling airborne concentration and deposition rate of
 883 maize pollen. *Atmos. Environ.* **38**: 5555-5566.

884 Jones G., 2002. Genetic and evolutionary algorithms, in *Encyclopedia of Computational*
 885 *Chemistry*, John Wiley and Sons, Hoboken, NJ, USA.

886 Kaminski T., Heimann M., Giering R., 1998. A coarse grid three-dimensional global inverse
 887 model of the atmospheric transport: 1. Adjoint model and Jacobian matrix. *J.*
 888 *Geophys. Res.-Atmos.* **104**: 18535-18553.

889 Katata G., Chino M., Kobayashi T., Terada H., Ota M., Nagai H., Kajino M., Draxler R.,
 890 Hort M.C., Malo A., Torii T., Sanada Y., 2015. Detailed source term estimation
 891 of the atmospheric release for the Fukushima Daiichi Nuclear Power Station
 892 accident by coupling simulations of an atmospheric dispersion model with an
 893 improved deposition scheme and oceanic dispersion model. *Atmos. Chem. Phys.*,
 894 **15**, 1029–1070.

895 Kim, M., Ohba R., Oura M., Kato S., Takigawa M., Bieringer P.E., Lauritzen B., Drews
 896 M., 2016. A source term estimation method for a nuclear accident, using
 897 atmospheric dispersion models., *Int. J. Environ. Pollut.* **58**: 39-51.

898 Keats A., Yee E., Liena F-S., 2007. Bayesian inference for source determination with
 899 applications to a complex urban environment. *Atmos. Environ.* **41**: 465–479.

900 Kennedy J., Eberhart RC., 2001. Swarm Intelligence. Morgan Kaufmann, San Francisco, CA,
 901 USA.

902 Krysta, M., M. Bocquet, B. Sportisse, O. Isnard, 2006: Data assimilation for short- range
 903 dispersion of radionuclides: an application to wind tunnel data. *Atmos. Environ.* **40**,
 904 7267–7279.

905 Krysta M., Bocquet M., 2007. Source reconstruction of an accidental radionuclide release at
 906 European scale. *Q. J. Roy. Meteorol. Soc.* **133**: 529–544.

907 Kuroki Y., Young G.S., Haupt S.E., 2010. UAV navigation by an expert system for
 908 contaminant mapping with a genetic algorithm. *Expert Syst. Appl.* **37**: 4687-4697.

909 Long K., Haupt S.E., Young G.S., 2010. Assessing sensitivity of source term estimation.
 910 *Atmos. Environ.* **44**: 1558–1567.

911 Medvedev Z.A., 1992. The Legacy of Chernobyl, W. W. Norton & Company Inc., New
 912 York, New York, USA.

913 Metropolis N., Rosenbluth A.W., Teller A.H., Teller E., 1953. Equations of state
 914 calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087-1092.

915 Norige A., Thorton J., Schiefelbein C., Rudzinski C., 2009. High density distributed
 916 sensing for chemical and biological defense. *Linc. Lab. J.* **18**: 25–40. [Available
 917 online at [https://www.ll.mit.](https://www.ll.mit.edu/publications/journal/pdf/vol18_no1/18_1_1_Norige.pdf)
 918 [edu/publications/journal/pdf/vol18_no1/18_1_1_Norige.pdf](https://www.ll.mit.edu/publications/journal/pdf/vol18_no1/18_1_1_Norige.pdf).]

919 Onishi N., Fackler M., 2011. Japan held nuclear data, leaving evacuees in peril. *New*
 920 *York Times*, 8 August. [Available online at
 921 www.nytimes.com/2011/08/09/world/asia/09japan.html].

922 Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P., 2002. Numerical Recipes in
 923 C++: The Art of Scientific Computing (Second Edition). Cambridge University Press,
 924 Cambridge, UK.

925 Pullen J., Chang J., Hanna S., 2013. Air–sea transport, dispersion, and fate modeling in
 926 the vicinity of the Fukushima nuclear power plant: A special conference session
 927 summary. *Bull. Amer. Meteor. Soc.* **94**: 31–39.

928 Rabier, F., E. Klinker, P. Courtier, and A. Hollingsworth, 1996: Sensitivity of forecast
 929 errors to initial conditions. *Quart. J. Roy. Meteor. Soc.*, **122**, 121-150.

930 Rao K.S., 2007. Source estimation methods for atmospheric dispersion, *Atmos. Environ.*
 931 **41**: 6964-6973.

932 Redwood M., 2011. Source term estimation and event reconstruction: a survey, 1-21.
 933 Report for the Atmospheric Dispersion Modelling Liaison Committee. [Available
 934 online at: <http://www.admlc.org.uk/documents/ADMLC-R6-2011-1.pdf>]

935 Ringbom, A., Axelsson A., Aldener M., Auer M., Bowyer T.W., Fritioff T., Hoffman I.,
 936 Khrustalev K., Nikkinen M., Popov V., Popov Y., Ungar K., Wotawa G., 2014.
 937 Radioxenon detections in the CTBT international monitoring system likely related
 938 to the announced nuclear test in North Korea on February 12, 2013. *Journal of*
 939 *Environmental Radioactivity*, **128**: 47-63.

940 Robertson L, J. Langner, 1998: Source function estimate by means of adjoint variational
 941 data assimilation applied to the ETEX-I tracer experiment. *Atmos. Environ.* **32**:
 942 24, 4219–4225.

943 Robins P., Rapley V.E., Green N., 2009. Real-time sequential inference of static
 944 parameters with expensive likelihood calculations. *J. R. Stat. Soc.: Series C*
 945 *(Appl. Stat.)*. 58: 641–662,

946 Rodriguez L.M., Haupt S.E., Young G.S., 2011. Impact of sensor characteristics on
 947 source characterization for dispersion modeling, *Measurement*. **44**: 802-814.

948 Rodriguez L.M., 2012. Uncertainty Propagation within Source Term Estimation. Ph. D.
 949 dissertation, The Pennsylvania State University, State College, PA. USA.

950 Sarker R., Kamruzzaman J., Newton C., 2003. Evolutionary optimization (EvOpt): a brief
 951 review and analysis. *Int. J. Comp. Intel. Appl.* **3**: 311-330.

952 Saunier, O., Mathieu A., Didier D., Tombette M., Quélo D., Winiarek V., Bocquet M.,
 953 2013. An inverse modeling method to assess the source term of the Fukushima
 954 Nuclear Power Plant accident using gamma dose rate observations. *Atmos. Chem.*
 955 *Phys.*, **13**, 11403–11421.

956 Stohl A., 1998. Computation, accuracy and applications of trajectories: A review and
 957 bibliography. *Atmos. Environ.* **32**: 947–966.

958

- 959 Stohl, A., Seibert, P., Wotawa, G., Arnold, D., Burkhardt, J. F., Eckhardt, S., Tapia, C.,
960 Vargas, A., and Yasunari, T. J., 2012. Xenon- 133 and caesium-137 releases into
961 the atmosphere from the Fukushima Dai-ichi nuclear power plant: determination
962 of the source term, atmospheric dispersion, and deposition, *Atmos. Chem. Phys.*,
963 **12**, 2313–2343, doi:10.5194/acp-12-2313-2012.
- 964 Schmehl K.J., Haupt S.E., Pavolonis M., 2012. A genetic algorithm variational approach
965 to data assimilation and application to volcanic emissions. *Pure and Appl.*
966 *Geophys.* **169**: 519-537.
- 967 Stein, A. F., Draxler R.R., Rolph G.D., Stunder B.J.B., Cohen M.D., Ngan F., 2015.
968 NOAA's Hysplit atmospheric transport and dispersion modeling system. *Bull.*
969 *Amer. Meteor. Soc.*, **96**, 2059–2077.
- 970 Storwold, D.P., 2007. Detailed test plan for the fusing sensor information from observing
971 networks (FUSION) field trial 2007 (FFT 07). West Desert Test Center, U.S.
972 Army Dugway Proving Ground. WDTC Document No. WDTC-TP-07- 078.
- 973 Sykes R.I., Parker S.F., Henn D.S., Chowdhury B., 2008. SCIPUFF Version 2.4
974 Technical Documentation.
- 975 Terada, H., Katata, G., Chino, M., and Nagai, H., 2012. Atmospheric dis- charge and
976 dispersion of radionuclides during the Fukushima Dai-ichi Nuclear Power plant
977 accident. Part II: verification of the source term and analysis of regional-scale
978 atmospheric disper- sion, *J. Environ. Radioact.*, **112**, 141–154.

979 Thomson L.C., Hirst B., Gibson G., Gillespie S., Jonathan P., Skeldo K.I.D., Padgett
 980 M.L., 2007. An improved algorithm for locating a gas source using inverse methods.
 981 *Atmos. Environ.* **41**: 1128-1134.

982 Varma R., Varma D.R., 2005. The Bhopal disaster of 1984. *B. Sci. Technol. Soc.* **25**: 37-
 983 45.

984 Wang Y., 2008. An inversion method to backtrack source parameters and associated
 985 concentration field for an inert gas release in urban environments, *Army Research*
 986 *Laboratory*, **ARL-TN-304**, 11pp.

987 Wotawa, G., De Geer L., Denier P., Kalinowski M., Toivonen H., D'Amours R., Desiato
 988 F., Issartel J., Langer M., Seibert P., Frank A., Sloan C., Yamazawa H., 2003,
 989 Atmospheric transport modeling in support of CTBT verification – overview and
 990 basic concepts. *Atmos. Environ.* **37**: 2529-2537.

991 White House. 2012. National Strategy for Biosurveillance, July 2012, [Available online
 992 at [http://www.whitehouse.gov/the-press-office/2012/07/31/national-strategy-](http://www.whitehouse.gov/the-press-office/2012/07/31/national-strategy-biosurveillance)
 993 [biosurveillance](http://www.whitehouse.gov/the-press-office/2012/07/31/national-strategy-biosurveillance)].

994 Winiarek, V., Bocquet, M., Duhanyan, N., Roustan, Y., Saunier, O., and Mathieu, A.,
 995 2014. Estimation of the caesium-137 source term from the Fukushima Daiichi nuclear
 996 power plant using a consistent joint assimilation of air concentration and deposition
 997 observations, *Atmos. Environ.*, **82**, 268–279, doi:10.1016/j.atmosenv.2013.10.017.

998 Young G.S., Limbacher J., Haupt S.E., 2009. Back trajectories for hazard origin
 999 estimation: BackHOE. *Seventh Conference on Artificial Intelligence and its*
 1000 *Applications to the Environmental Sciences*, 10-16, January, Phoenix, AZ, USA.