J6.1 PARADIGMS FOR SOURCE CHARACTERIZATION

Sue Ellen Haupt*
George S. Young

Applied Research Laboratory and Meteorology Department,
The Pennsylvania State University, State College, PA

1. INTRODUCTION

A common problem in atmospheric dispersion is identifying and characterizing the source of an atmospheric contaminant. Such characterizations have been important in quantifying the emissions from sources, in spite of the fact that the contaminant transport and dispersion may occur over a regional domain. Such problems range from identifying the source of particulates through attributing one region's pollution problem to sources upstream.

A recent application of source characterization is in the context of homeland and defense security. In the event of the release of a hazardous contaminant (either accidentally or intentionally), decision makers are responsible for appropriate action, which might include evacuation, protection, and mitigation procedures, as well as summoning medical help.

Various groups have been devising techniques to assimilate concentration data in order to characterize the source of the contaminant. Such techniques include Bayesian approaches, Monte Carlo Markov Chain methods, four dimensional variational and adjoint assimilation methods, Kalman Filter, and statistical learning approaches, among others. Rao (2007) has recently reviewed some of the techniques. His primary categorization is in terms of forward models that predict primarily from the source to the receptor versus backward techniques that invert from the receptor to the source. This present work takes a different point of view: we break the stages of the source estimation process into the elements that are necessary for all algorithms to accomplish the same goal. We then build a matrix of what type of algorithm or data fills the role of each element for source characterization techniques found in the literature.

* *Corresponding author address:* Sue Ellen Haupt, Applied Research Laboratory, P.O. Box 30, Pennsylvania State University, State College, PA 16804; e-mail: haupts2@asme.org

The goal of this paper is to compare and contrast the different formulations and to outline a general paradigm that encompasses all the methods. This paradigm is developed in section 2, which also presents how the various methods fulfill each of the primary elements of the paradigm. Section 3 discusses some of the issues that must be considered when tuning each technique and discusses the problem of confounding variables more deeply. Section 4 gives a detailed example of the method formulated by the authors. Finally, conclusions are made in section 5. That section also summarizes and discusses how the segments of the different methods might be combined to optimize the process. We conclude with an analysis of the prospects for accurately characterizing sources of contaminant.

2. PROBLEM DEFINITION

2.1 The Paradigm

Certain elements of the source characterization process is common to all techniques. These elements appear in Figure 1. Some methods explicitly treat the problem as one in optimization, while for others, the parallel with an optimization problem is inferred. The key aspect is to obtain both field monitored concentrations and estimates of the predicted concentrations at common times and locations. Sometimes interpolation is necessary to get the data on a common grid. The field monitored data is obtained from sensor(s) located in the field. The estimated predicted concentration is produced by a dispersion model. That model requires input parameters including source information and meteorological data. Since source information is typically what the entire method is designed to compute, it is necessary to make "guesses" at such data to obtain a concentration estimate. The atmospheric transport and dispersion (AT&D) models also require information about the meteorological conditions as well as information about the site (terrain, land usage, etc.) Although such data are often assumed to be known, the AT&D model can be quite sensitive to errors or changes in those conditions as discussed later.

The difference between the predicted concentrations and those observed are compared quantitatively in a cost function. This explicit comparison guides the algorithm to generate new, better guesses at the input data. The ways to produce these new estimates are the primary differences in the various source estimation methods. Each method includes some algorithm to generate these improved source parameters. One can further break down this "inner" paradigm, but it is beyond the scope of this present analysis.

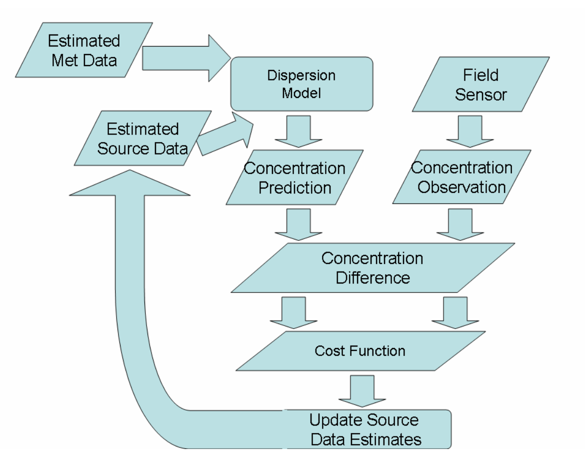

Figure 1. Flowchart of source inversion process.

2.2 Comparison of Techniques

Quite a few research groups have formulated and tested methods for characterizing source variables for dispersion based on downwind concentration measurements. Table 1 (see end of paper) delineates how each of these methods fills the elements of the paradigm presented above. That table lists which of the source and other parameters are sought, the AT&D model used to produce predicted concentrations, the inversion techniques, and the concentration data used for comparison. There is a wide range of complexity in the methods to match the range of applications. Some methods invert for only source strength but on a regional scale with sparse field data while others seek several meteorological variables in addition to characterizing source strength, location, and time of release but on a small scale construed grid in the context of an identical twin numerical experiment (that is, synthetic test data is created using the same model filling the role of the AT&D model in the algorithm). The AT&D models range from simple back-trajectories and basic Gaussian plume models through high resolution Computational Fluid Dynamics (CFD) models. The inversion techniques themselves tend to be quite innovative. The main inversion categories include: 1) Bayesian methods coupled with statistical sampling; 2) adjoint and variational methods that use a tangent linear version of the AT&D model; 3) Kalman Filter and similar analytic update approaches; and 4) statistical learning techniques such as genetic algorithms and simulated annealing. Each method has its own system for evolving better solutions to the nonlinear problem without falling prone to finding the incorrect local minima to the problem. In fact, such systems constitute an "inner paradigm" that will be discussed in future papers.

3. CROSSCUTTING ASPECTS

3.1 Identifying the Issues

As seen above, we can define a general paradigm for the source characterization process. There are, however, various "cross-cutting" issues that must be addressed by each of the methods. These crosscutting aspects represent problems that can make the inversion process difficult. The outlook on these issues constitutes the rationale for many of the differences in the algorithms. Some of those issues are discussed in general here to pave the way for understanding the rationale behind the decisions made by the algorithm developers. Many of the differences are in these details.

The cross-cutting issues include:

- Some source variables are coupled, making the characterization problem ill-posed.
- There are various ways to formulate the cost function that can produce different convergence properties.
- Sufficient data are required to accurately back-calculate source variables.
- Noise in the sensor data and in the modeled prediction can contaminate the results.
- Constructing good verification methods that clearly quantify the error yet are sufficiently realistic is difficult.
- There are confounding influences from nonlinear sensitivities to variables that may not be part of the optimization problem.

3.2 Sensitivity to Meteorological Variables

We choose to elaborate on one very important cross-cutting issue, the final bullet. One confounding issue is the fact that concentration predictions are quite sensitive to meteorological conditions. It is typical for modelers to assume specific meteorological conditions for making the concentration prediction with the AT&D model. Meteorology, however, is highly influenced by local conditions such as variations in terrain and localized heat fluxes. Errors in such variables can produce errors in wind speed and direction that, when used to drive an AT&D model, can make a large change in the predicted concentration field. In addition, there is an inherent uncertainty in the turbulence that drives dispersion. The stochastic nature of turbulence makes it difficult to define a sufficiently accurate spatially and temporally varying wind field to characterize the actual AT&D. In addition, the actual AT&D is a single realization of an ensemble of possible realizations. Sensitivity to initial conditions of the nonlinear fluid transport problem makes this a formidable problem.

Krysta, et al. (2006) noted that, even in a wind tunnel environment, the wind variability can cause a plume prediction with the concentration peak phase shifted from that monitored. Their solution was to make the wind components variables of the problem. That approach produced a better match to the data. That philosophy is also adopted by Allen et al. (2007), Haupt et al. (2007, 2008), and Long et al. (2008) as described below. This approach is explicitly added to our paradigm flow chart in Figure 2. Note that the difference between this figure and Figure 1 is that now the meteorological data are also found by the inversion routine and used to update the next iteration of the concentration prediction.

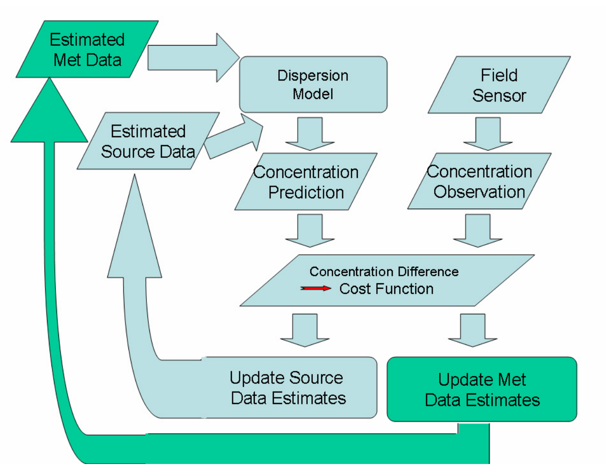

Figure 2. Flowchart of Paradigm augmented by characterizing meteorological data.

4. AN EXAMPLE TECHNIQUE

4.1 Model Formulation

We describe here how one particular method fulfills the elements of the paradigm as well as how it deals with the confounding issue of sensitivity to meteorological variables. We choose the genetic algorithm-based method developed by the authors and their students for further elaboration (Allen et al. 2007, Haupt et al. 2007, 2008, Long 2007, Long et al. 2008). Each of the primary elements of Table 1 is described in more detail after an overview of some elements intrinsic to the model formulation.

The core of the model is the comparison of the sensor data with the predicted concentrations in terms of a log-normal cost function formulation:

$$\text{cost} = \frac{\sqrt{\sum_{r=1}^{TR}\left[\ln(aC_r+\varepsilon)-\ln(aR_r+\varepsilon)\right]^2}}{\sqrt{\sum_{r=1}^{TR}\left[\ln(aR_r+\varepsilon)\right]^2}} \quad (1)$$

where: C_r= forecast concentration as predicted by the transport and dispersion model at receptor *r*,

R_r=observed concentration retrieved from receptor *r*,

TR is the total number of receptors, and

a and *ε* are constants used to avoid taking the logarithm of zero (*a* = 1, $\varepsilon = 1\times10^{-13}$ here).

4.2 Sensor Data

The applications tested by this technique have been in terms of identical twin experiments; that is, the transport and dispersion model used to optimize agreement with the sensor data is the same model that produces the synthetic sensor data. The identical twin approach is convenient for formulating and testing problems because it removes several of the sources of error from consideration: we no longer expect inherent fluctuations due to turbulence and we are assured that the sensor data is not contaminated with noise. Since it allows us to compare to data that we know are exact, it allows evaluation of the inversion algorithm alone rather than the combination of model and data. The disadvantage, of course, is that a level of realism is lost in this approach. Thus, the identical twin approach to the data element is best suited for algorithm development and analysis rather than for estimating absolute performance of an algorithm in the real-world setting. As described below, these issues can be addressed by adding noise to the twin data – we still know the correct solution and can evaluate the algorithm. Current plans include testing the model with fraternal twin experiments (constructed test data with a more refined AT&D model) as well as with real sensor data.

4.3 Concentration Prediction

A Gaussian puff AT&D model is used to forecast the contaminant concentration. This model is used because it is an exact solution of the ensemble averaged diffusion equation. The problem is formulated in a Cartesian domain. Wind is assumed to blow in the positive *x*-direction (i.e. the domain is rotated to so that the x axis aligns with the direction of the wind). This model is formulated as:

$$C_r = \frac{Q\Delta t}{(2\pi)^{1.5}\sigma_x\sigma_y\sigma_z}\exp\left(\frac{-(x_r-Ut)^2}{2\sigma_x^2}\right)\exp\left(\frac{-y_r^2}{2\sigma_y^2}\right)\times \left[\exp\left(\frac{-(z_r-H_e)^2}{2\sigma_z^2}\right)+\exp\left(\frac{-(z_r+H_e)^2}{2\sigma_z^2}\right)\right] \quad (2)$$

where: C_r is the concentration at receptor *r*,
(x_r, y_r, z_r) are the Cartesian coordinates downwind of the puff,
Q is the emission rate,
Δt is the length of time of the release itself,
t is the elapsed time since the release,
U is the wind speed,
H_e is the effective height of the puff centerline, and
$(\sigma_x, \sigma_y, \sigma_z)$ are the standard deviations of the concentration distribution in the *x*-, *y*-, and *z*-directions, respectively.

The transport is in the *x*-direction at wind speed, *U* and the contaminant dispersed in the *y*- and *z*-directions with standard deviation of the spread given by the dispersion coefficients. The dispersion coefficients are computed according to Beychok (1994).

$$\sigma = \exp\left\{I + J\left[\ln(x) + K(\ln(x)\right]^2\right\} \quad (3)$$

where x is the downwind distance (in km) and I, J, and K are empirical coefficients dependent on the Pasquill Stability Class, which characterizes the atmospheric turbulence scales. The coefficients are then looked up in tables to produce σ_y and σ_z Beychok (1994). The dispersed pollutant form the C_r of (2) and the monitored data, R_r are the concentration values that are compared in the cost function (1).

4.4 Sensitivity to Meteorological Variables

As noted above, there is a recognized sensitivity of the concentration forecasts to errors in meteorological variables. The approach to ameliorating this problem taken here is to address it directly by including those variables in the solution process. Like Krysta et al. (2006), we formulate the problem to additionally solve for wind speed and wind direction. Note that in other work, we study the value of adding more variables to the back-calculation, including atmospheric stability class (Haupt et al. 2008) and depth of the planetary boundary layer (Annunzio et al. 2008). For the work reported here, we concentrate on solving for six variables: source location (*x,y*), source strength, time of release, wind speed, and wind direction.

4.5 Solution Technique

For this problem we chose a continuous parameter GA, that is, one in which the parameters are real numbers. Figure 1 flowcharts the GA solution process. The genetic algorithm starts with a population of random vectors (i.e. chromosomes) that are evaluated using the forward model and cost function (1). Here we use a uniform crossover scheme that blends all parameters, not just a single parameter at the crossover point. This blending scheme has the advantage of simultaneously changing all parameters, which can improve performance when the response of the cost function to some of the parameters is correlated. In this case, the response to source location and wind vector are highly correlated. The number of new chromosomes produced by mating is determined by the selection rate, which is the fraction of the population retained in each generation.

A single guess of the variables to be optimized is placed in a row vector called a chromosome. The GA works with many such guesses at once, so a matrix of trial solutions is formed with chromosomes as the rows. Initially, all of the chromosomes in the population matrix contain random values taken from each variables physically possible range. This matrix is passed to the cost function and a column vector of costs is returned.

Two primary operations drive the solution to evolve toward the optimal solution: mating and mutation. The operation of mating combines the variable values from the best trial solutions to produce a new population of improved variable estimates. The GA is quite robust at solving difficult nonlinear coupled optimization problems with a multitude of local minima that are difficult for traditional techniques. More details of the GA technique are described by Haupt and Haupt (2004).

The chromosome population is further modified through mutations. Mutations replace individual values with new random values. The mutation operator generates new solutions to maintain an adequate sampling of the parameter space, preventing premature convergence to a suboptimal set of parameter values. The number of mutations in each generation is controlled by the mutation rate.

Each round of mating and mutating constitutes one GA generation. We run the GA for a pre-determined number of iterations, or until convergence has occurred. We employ elitism, which prevents the best solution computed in each generation from being changed until it is supplanted.

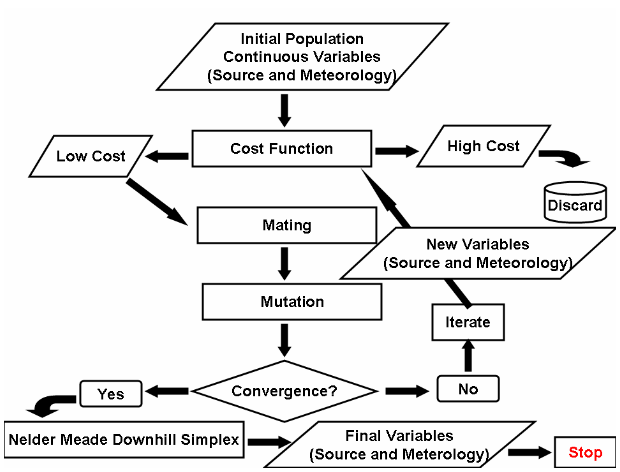

Figure 3. Flowchart of the continuous GA.

We use a hybrid GA, which uses the GA to find the correct solution basin, then applies the Nelder-Meade Downhill Simplex (NMDS) method to complete finding the minimum point of that basin. The rationale for this combination is that the GA is sufficiently robust to usually find the basin of the global minima. Once that basin is identified, however, the NMDS finds the bottom of that basin more rapidly. As demonstrated in in Allen et al. (2007), the NMDS method alone is not reliable for finding the global minimum.

4.6 Evaluation

Since this model was formulated to develop academic insight, it has been tested in a multitude of formulations (Haupt 2005; Haupt et al. 2006; Allen et al 2006, 2007; Haupt et al 2007; Long 2007; Long et al 2008) including in papers presented at this conference (Annunzio et al. 2008, Rodriguez et al. 2008; Haupt et al. 2008; Long et al. 2008b). For the six parameter version described here, see the results of Long et al (2008) and Haupt et al (2008). Its performance has been quantified by including a large number of Monte Carlo runs (Haupt et al. 2006, 2007, Long et al. 2008a). We have included noise in the data to determine the ability of such a model to reconstruct the source and meteorological data in a realistic noisy scenario (Haupt et al. 2006, 2007, Long et al. 2008a). We have even attempted to quantify how many sensors are necessary for a given level of noise in the data and a given correlation between the succeeding time steps (Long et al. 2008a, Haupt et al. 2008). The reader is referred to those other works to see details of model performance.

5. CONCLUSIONS

This paper defines a paradigm for addressing the source characterization problem. Given field sensor data, various techniques have been developed to back-calculate source variables. Each of these techniques requires some method to predict concentrations to compare them to the measured sensor data, then to update the source variables to produce a better prediction. Table 1 delineates how each of those techniques fulfills the elements of the paradigm, including the inversion variables, an AT&D model, inversion technique, and test data that supply the sensor data. Some cross-cutting issues are described and one specific issue that deals with the confounding effects of meteorological conditions is elaborated on. More details of a GA-based technique are used as an example of the paradigm.

In future work, we plan to delve deeper into this paradigm by dissecting the inversion techniques. The inversion techniques themselves have several commonalities that are discerned by studying them in more detail. There are additional cross-cutting issues that are relevant at that level.

Again, each group of technique developers have devised ways to deal with those issues.

In summary, there are many ways to approach the source characterization problem. Various teams have developed viable solution methods. Each of those methods displays strengths and limitations. It is instructive to analyze their relative merits. A "best" technique might combine the strengths of several of these techniques.

ACKNOWLEGEMENTS – This research is supported by DTRA (grant number W911NF-06-C-0162). We particularly thank John Hannan and Christopher Kiley for valuable feedback throughout this project.

REFERENCES

Allen, C.T., S.E. Haupt, and G.S. Young, 2007: Source Characterization with a Receptor/ Dispersion Model Coupled With A Genetic Algorithm, *Journal of Applied Meteorology and Climatology,* **46**, 273-287.

Allen, C.T., G.S. Young, and S.E. Haupt, 2006: Improving Pollutant Source Characterization by Optimizing Meteorological Data with a Genetic Algorithm, *Atmospheric Environment,* **41**, 2283-2289.

Annunzio, A.J., S.E. Haupt, and G.S. Young, 2008: Source Characterization and Meteorology Retrieval Including Atmospheric Boundary Layer Depth using a Genetic Algorithm, 15th Joint Conference on the Applications of Air Pollution Meteorology with the A&WMA, New Orleans, LA, Jan. 20-24.

Bergin, M.S. and J.B. Milford, 2000: Application of Bayesian Monte Carlo analysis to a Lagrangian photochemical air quality model. *Atmospheric Environment*, **34**, 781-792.

Beychok, M. R., 1994: Fundamentals of Gas Stack Dispersion, 3rd ed. Milton Beychok, pub., Irvine, CA, 193 pp.

Bocquet, M., 2005a: Reconstruction of an atmospheric tracer source using the principle of maximum entropy. I: Applications. *Q.J.R. Meteorolo. Soc*., **131**, 2209-2223.

Bocquet, M., 2005b: Reconstruction of an atmospheric tracer source using the principle of maximum entropy. I: Theory. *Q.J.R. Meteorolo. Soc*., **131**, 2191-2208.

Brown, R.C., D.M. Dussault, and R.C. Miake-Lye, P. Heimback, 2005: SCIPUFF Adjoint model for atmospheric transport and dispersion optimization, Report RR-1392, Aerodyne Research, Billerica, MA, 24pp.

Camelli, F. and R. Lohner, 2004: Assessing maximum possible damage for contaminant release, *Engineering Computations*, **21**, 748-760.

Chang, M.E., D.E. Hartley, C. Cardelino, D. Haas-Laursen, and W.-L. Chang, 1997: On using inverse methods for resolving emissions with large spatial inhomogeneities. *J. Geophysical Research*, **102**, 16,023-16,036.

Cheng Y., Umanaheswara Reddy K.V., Singh T., Scott P.D., 2007. CBRN Data Fusion Using Puff-Based Model and Bar-Reading Sensor Data. 10th International Conference on Information Fusion, Quebec City, Canada, July 9-12th, 2007.

Chow, F.K., B. Kosovic, and S. Chan, 2007: Source inversion for contaminant plume dispersion in urban environments using building-resolving simulations, submitted to *J. Appl. Meteorol. and Climatol.*

Elbern, H. H. Schmidt, and A. Ebel, 1997: Variational data assimilation for tropospheric chemistry modeling, *J. Geophysical Research*, **102**, 15,967-15,985.

Kaminski, T. and M. Heimann, 1999: A coarse grid three-dimensional global inverse model of the atmospheric transport. 2. Inversion of the transport of CO2 in the 1980s. *J. Geophysical Research*, **104**, 18,555-18,581.

Haas-Larsen, D.E. and D.E. Hartley, 1996: Optimizing an inverse method to deduce time-varying emissions of trace gases. *J. Geophysical Research*, **101**, 22,823-22,831.

Haupt, S.E., 2005: A Demonstration of Coupled Receptor/Dispersion Modeling with a Genetic Algorithm, *Atmospheric Environment*, **39**, 7181-7189.

Haupt, S.E., G.S. Young, K.J. Long, and A. Beyer-Lout, 2008: Data Requirements for Assimilating

Concentration Data with a Genetic Algorithm, Sixth Conference on Artificial Intelligence Application to Environmental Science, New Orleans, LA, Jan. 20-24.

Haupt R. L., Haupt S. E., 2004: Practical Genetic Algorithms, 2nd edition with CD. John Wiley & Sons, New York, NY.

Haupt, S.E., R.L. Haupt, and G.S. Young, 2008: A Mixed Integer Genetic Algorithm used in Chem-Bio Defense Applications, submitted to *Journal of Soft Computing*.

Haupt, S.E., G.S. Young, and C.T. Allen, 2007: A Genetic Algorithm Method to Assimilate Sensor Data for a Toxic Contaminant Release, *Journal of Computers*, **2**, 85-93.

Haupt, S.E. G.S. Young, and C.T. Allen, 2006: Validation Of A Receptor/Dispersion Model Coupled With A Genetic Algorithm, *Journal of Applied Meteorology,* **45**, 476–490.

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

Hourdin, F., O. Talagrand, and A. Idelkadi, 2006: Eulerian backtracking of atmospheric tracers. II: Numerical aspects. *Q.J.R. Meteorolo. Soc*., **132**, 585-603.

Houweling, S. T. Kaminski, F. Dentener, J. Lelieveld, and M. Heimann, 1999: Inverse modeling of methane sources and sinks using the adjoint of a global transport model. *J. Geophysical Research*, **104**, 26,137-26,160.

Issartel, J.-P., 2005: Emergence of a tracer source from air concentration measurements, a new strategy for linear assimilation. *Atmos. Chem. Phys.*, **5**, 249-273.

Issartel, J.-P. and J. Baverel, 2003: Inverse transport for the verification of the Comprehensive Nuclear Test Ban Treaty. *Atmos. Chem. Phys.*, **3**, 475-486.

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

Keats, A., E. Yee, and F.S. Lien, 2007b: Reply to 'Comments on: "Bayesian inference for source determination with applications to a complex urban environment." ' *Atmospheric Environment*, **41**, 5547-5551.

Krysta, M. and M. Bocquet, 2007: Source reconstruction of an accidental radionuclide release at European scale. *Q. J. R. Meteorol. Soc.* **133**, 529-544.

Krysta, M., M. Bocquet, B. Sportisse, and O. Isnard, 2006: Data assimilation for short-range dispersion of radionuclides: An application to wind tunnel data. *Atmospheric Environment,* **40**, 7267-7279.

Long, K.J., S.E. Haupt, and G.S. Young, 2008a: Improving Meteorological Forcing and Contaminant Source Characterization Using a Genetic Algorithm. To be submitted to *Optimization and Engineering.*

Long, K.J., S.E. Haupt, G.S. Young, 2008b: Source Characterization and Meteorology Retrieval using a Genetic Algorithm with SCIPUFF, 15th Joint Conference on the Applications of Air Pollution Meteorology with the A&WMA, New Orleans, LA, Jan. 20-24.

Monache, L.D., J.K. Lundquist, B. Kosovic, G. Johannesson, K.M. Dyer, R.D. Aines, F. K. Chow, R.D. Belles, W.G. Hanley, S.C. Larsen, G.A. Loosmore, J.J. Nitao, G.A. Sugiyama, P.J. Vogt, 2007: Bayesian inference and Markov Chain Monte Caro sampling to reconstruct a contaminant source at continental scale, submitted to *J. Appl. Meteorol. and Climatol.*

Pudykiewicz, J.A., 1998: Application of adjoint tracer transport equations for evaluating source parameters. *Atmospheric Environment*, **32,** 3039-3050.

Pudykiewicz, J.A. and A. S. Koziol, 1998: An application of the theory of kinematics of mixing to the study of tropospheric dispersion. *Atmospheric Environment*, **32,** 4227-4244.

Rao, K.S., 2007: Source estimation methods for atmospheric dispersion, *Atmos. Env*. **41**, 6963-6973.

Robertson, L. and J. Langner, 1998: Source function estimate by means of variational data assimilation applied to the ETEX-I tracer experiment. *Atmospheric Environment*, **32,** 4219-4225.

Robins, P. and V.E. Rapley, 2006: Real-time sequential inference of static parameters with expensive likelihood calculations, Dstl.

Robins, P., Rapley, V., Thomas, P., 2005: A Probabilistic Chemical Sensor Model for Data Fusion, Dstl, Salisbury, UK.

Robins, P. and P. Thomas, 2007: Non-linear Bayesian CBRN source term estimation, Dstl.

Rodriguez, L.M., S.E. Haupt, and G.S. Young, 2008: Adding Realism to Source Characterization with a Genetic Algorithm, 15th Joint Conference on the Applications of Air Pollution Meteorology with the A&WMA, New Orleans, LA, Jan. 20-24.

Stohl, A. S. Eckhardt, C. Forster, P. James, N. Spichtinger, P. Seibert, 2002: A replacement for simple back trajectory calculations in the interpretation of atmospheric trace substance measurements. *Atmospheric Environment*, **36,** 4635-4648.

Storch, R.B., L.C.G. Pimententel, and H.R.B. Orlande, 2007: Identification of atmospheric boundary layer parameters by inverse problem, *Atmos. Env*., **41**, 1417-1425.

Sykes, R.I., 2007: Source estimation using sensor data and reverse transport. Chemical and Biological Information Systems Conference, Austin, TX, Jan. 8-12.

Terejanu G., Singh T., Scott P.D., 2007. Unscented Kalman Filter/Smoother for a CBRN Puff-Based Dispersion Model. 10th International Conference on Information Fusion, Quebec City, Canada, July 9-12th, 2007.

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

Umamaheswara Reddy K.V., Cheng Y., Singh T., Scott P.D., 2007. Data Assimilation in Variable Dimension Dispersion Models using Particle Filters. 10th International Conference on Information Fusion, Quebec City, Canada, July 9-12th, 2007.

Vukicevic, T. and P. Hess, 2000: Analysis of tropospheric transport in the Pacific Basin using the adjoint technique. *J. Geophysical Research*, **105**, 7213-7230.

Wotawa, G., L.-E. DeGeer, P. Denier, M. Kainowski, H. Toivonen, R.D'Amours, F. Desiato, J.-P. Issartel, M. Langer, P. Seibert, A. Frank, C. Sloan, H. Yamazawa, 2003: Atmospheric transport modeling in support of CTBT verification – overview and basic concepts. *Atmospheric Environment*, **37,** 2529-2537.

Table 1. Elements of Paradigm Model for development groups

Who	Inversion Variables	AT&D Model	Inversion Technique	Application (data)
Allen et al 2006	Source Strength Time of Release	SCIPUFF	Genetic Algorithm	Dipole Pride 26
Allen et al (2007) Haupt et al. (2007) Long et al. (2008) Haupt et al. (2008)	Strength Location (x,y,z) Time of Release Wind speed Wind direction Stability category	Gaussian Puff	Genetic Algorithm	Identical Twin Data with Noise
Bergin&Milford (2000)	"uncertainty" of concentration	Lagrangian photochemical air quality model	Bayesian Monte Carlo	2 cases of 2-day ozone trajectories
Bouquet & Krysta (2007) Bouquet (2005a,b)	Activities (source strength)	Eulerian dispersion model (Advection-diffusion	Maximum entropy	Twin synthetic
Brown, et al. (2005)	Location (x,y,z) Mass Time	SCIPUFF	Adjoint/tangent linear	DP26
Camelli & Lohner (2004)	Location (x,y)	CFD FEFLO	GA	1. 6 building simulated urban scenario -2. Subway station
Chang, et al. (1997)	Emission scaling factors	Urban Airshed Model	Kalman Filter	Atlanta – 2 days Aug. '92
Chow, et al (2007)	Location (x,y) Release Rate	3D CFD FEM3MP	Stochastic Sampling (MCMC) with Bayesian inference	1. Flow about cube 2. JU2000
Elbern, et al. (1997)	Source strength (multiple sources)	Regional chemical transport model (RADM2)	4D variational/ Adjoint	1. Simulated 2. Real
Haas-Laursen, et al. (1996)	Sources and sinks of chemicals (flux)	Effectively Sampled Region (ESR) 2D AT&D &chem model	Kalman filter	Synthetic
Haupt (2005) Haupt et al. (2006)	Source Strength (multiple sources)	Gaussian Plume	Genetic Algorithm	Identical twin synthetic data with noise
Hourdin & Talagrand (2006)	Strength	LMDZ – GCM w/ chemistry 1D advection	Adjoint	ETEX
Howeling, et al. (1999)	Sources and sinks of methane (flux regions)	Global transport 3D model (TM2)	Adjoint	NOAA cruise data
Issartel (200% Issatel & Baverel (2003)	Location Strength Time	Advection equation	Adjoint functions with illumination	ETEX

Who	Inversion Variables	AT&D Model	Inversion Technique	Application (data)
Kaminski & Hermann (1999)	CO2 fluxes on grid	3D course grid transport model	Bayesian synthesis	Global CO2 observations
Keats, et al. (2007a,b)	Location (x,y,z) Strength	Advection equation	Bayesian inference & MCMC	Urban field data (MUST & JU2003)
Krysta, et al. (2006)	Emission rate Wind (u,v)	Gaussian Puff	Variational - minimization	Wind tunnel data – near nuclear plant
Monache, et al. (2007)	Location (x,y) Release Rate	Lagrangian particle model	Stochastic Sampling (MCMC) with Bayesian inference	European release
Pudykiewicz (1998)	Strength Location (x,y,z) Time of release	Advection eqn w/ Gaussian diffusion	Adjoint	Chernobyl nuclear accident
Robertson& Langner ((1998)	Location (x,y,z)	MATCH (off-line Eulerian)	"poor mans" Variational Method	ETEX
Robins, Thomas, Rapley (2005a,b; 2006)	Mass Location (x,,y) Time of release Material type	Gaussian	Bayesian stats & Diff Evolution MC	Simulated data
Stohl, et al. (2002)	Probability of initial location	Lagrangian Particle Model(FLEXPART)	Back Trajectories and cluster analysis	1. Synthetic lidar profiles 2. synthetic surface concentration
Sykes (2007)	Mass(scaling) Location (x,y)	HPAC	Adjoint Method	1. ETEX 2. Dugway ensemble expt: case 10
Thomson, et al. (2007)	Strength Position	Gaussian plume	Random search w/ simulated annealing	Desert field data
Wotawa, et al. (2003)	Emission factors	Conceptual (Lagrangian)	Source Receptor Matrix	Synthetic scenarios
Vukicevic & Hess (2000)	Transport (Sensitivity of mixing ration)	Transport & Chem. Transformation model (HANK)	Adjoint	Chemical concentration at Hawaii -Field data