Source Identification of Puff-Based Dispersion Models Using Convex Optimization

Umamaheswara Konda
Dept. of MAE
University at Buffalo
Buffalo, NY, U.S.A.
venkatar@buffalo.edu

Yang Cheng
Dept. of AE
Mississippi State University
Mississippi State, MS, USA
cheng@ae.msstate.edu

Tarunraj Singh
Dept. of MAE
University at Buffalo
Buffalo, NY, USA
tsingh@eng.buffalo.edu

Peter D. Scott
Dept. of CSE
University at Buffalo
Buffalo, NY, USA
peter@buffalo.edu

Abstract – *A convex optimization based source estimation method is presented for dynamic models. The effectiveness of the method is illustrated in the context of a simple atmospheric puff-based dispersion model. Source estimation is the process of inferring the source parameters from the sensor measurements and the physical model. In dispersion, the most important source parameters include the locations and strengths of the sources as well as their number. A source identification method usually involves global search of the multidimensional parameter space, including a large area of possible source locations based on a batch of sensor data gathered over a reasonably long time interval. In this work, a grid-based algorithm is presented for efficient source identification where the number of sources is unknown and may be large. The source identification problem is formulated as a convex optimization problem in the ℓ_1 metric, which exploits the sparse nature of the solution to efficiently estimate the source characteristics.*

Keywords: Source estimation, multiple sources, dispersion models, convex optimization, L_1 minimization.

1 Introduction

High-fidelity dispersion forecasting requires good initialization of the dispersion model, accurate propagation of the dominant modes of the dispersion model, and effective data assimilation in the presence of high nonlinearity and uncertainty. Source estimation is the process of inferring the source parameters from the sensor measurements. The source parameters include the locations and strengths of the sources as well as the number of sources. A source estimation method usually involves global search of a large area of possible source locations based on a batch of sensor data gathered over a reasonably long time interval. That is in contrast with most data assimilation and fusion algorithms, especially the recursive algorithms, which are usually based on local search or local update techniques. A survey of source estimation algorithms for atmospheric dispersion was recently presented by Rao [1], including forward and backward modeling methods. Forward modeling methods include stochastic Monte Carlo or Markov Chain Monte Carlo sampling techniques; backward or inverse modeling methods include adjoint and tangent linear models, Kalman filters, and variational data assimilation.

Most source estimation algorithms assume that the number of sources is known. When the number of sources is unknown, source estimation algorithms usually employ exhaustive comparison of all hypotheses about the number of sources in terms of accuracy and complexity, which is computationally very expensive. A novel convex optimization-based algorithm for source estimation where the number of sources is unknown, is presented in [2], for radioactive source identification. The objective of the current work is to extend the algorithm to atmospheric dispersion scenarios where the release concentration observations are available at several time-steps and the release durations might be different for different sources. The algorithm is based on the observation that the contributions of multiple sources to a (noiseless) sensor observation satisfy the linear superposition principle. This partially linear structure of the system is exploited by discretizing the nonlinear part of the system. The source estimation problem can then be approximately formulated as a convex programming problem, which can be solved efficiently using open-source or proprietary software even when the size of the problem is large [3]. This method has the potential to provide near real-time source estimation. It is noteworthy that the method is not optimal in the Bayes sense because of the grid approximation of the continuous space as well as the cost function we choose to minimize, but many strategies exist to further improve the estimation result.

The remainder of the paper proceeds as follows. First, the dispersion and sensor models are reviewed. Then, the convex problem formulation is presented, followed by an overview of the ℓ_1 minimization method. Finally, the numerical test results and discussions are given.

2 Puff-Based Dispersion Model

A two-dimensional atmospheric dispersion model, based on the RIMPUFF [4] (Riso Mesoscale PUFF) model which was designed to calculate the concentration and doses result-

ing from the dispersion of airborne particles, is used. It is a Lagrangian mesoscale atmospheric dispersion puff model, which applies both to homogeneous and inhomogeneous terrain with moderate topography on a horizontal scale of up to 50 km, and responds to changing (non-stationary) meteorological conditions [4]. The model simulates the continuous release (emission) of airborne materials by sequentially releasing a series of Gaussian shaped puffs at a fixed rate. The amount of airborne materials allocated to individual puffs equals the release rate times the time elapsed between puff releases. At each time step, the model advects, diffuses and deposits the individual puffs according to local meteorological parameter values.

The concentration distribution in each puff is Gaussian in the two dimensional space. The mean of the Gaussian represents the location of the puff center, and the standard deviations represent the size of the puff in various directions. For simplicity, the Gaussian is assumed circular in our case. The standard deviation σ_x in the downwind direction is used as a mathematical tool and is made equal to the standard deviation σ_y in the crosswind direction [4]. The standard deviation of this circular Gaussian is denoted σ_{xy}.

Each Gaussian puff has four parameters: $[X, Y, \sigma_{xy}, Q]$, where

$$\begin{aligned} [X, Y] &= \text{Centroid of the Gaussian puff} \\ \sigma_{xy} &= \text{Puff size (std. deviation)} \\ Q &= \text{Activity (mass) of the puff} \end{aligned} \tag{1}$$

Note that the subscript for puff indexing has been dropped for simplicity. The state vector representing all the puffs at time t_k is of size $4N_k$, where N_k is the number of puffs at time t_k. The number N_k of puffs increases with time for continuous release. It is assumed that all the puffs originating from the same source have the same initial centroid (identical to the source location), size, and mass. When there are L sources, $4L$ independent parameters are needed to initialize the puffs. Note that $L \leq N_k$ and L is constant. Fewer than $4L$ independent parameters are needed when certain parameters, for example, the initial puff sizes, are assumed to be identical for all sources. The number of independent source parameters is greater than $4L$ when additional release-related parameters such as release rates and release duration times are included in source estimation.

The evolution of the parameters of the Gaussian puffs is described by the puff dynamics. In the absence of dispersion model error, the mass of each puff remains constant. The other puff parameters are updated based on the local meteorological parameters. A puff splitting scheme is usually employed in the puff based dispersion models like RIMPUFF, to account for dispersion over complex terrains which might result in plume splitting. This scheme is included in our model to represent a similar structure for testing purpose. When an initially small puff is grown to the size comparable with the grid spacing of the flow model, the original puff is replaced with five new smaller puffs, under the following constraints [4]:

1. $\sum_{j=1}^{5} Q_j = Q$
2. $\sum_{j=1}^{5} Q_j (d_j^2 + \sigma_{xy(j)}^2) = Q\sigma_{xy}^2$
3. $\sum_{j=1}^{5} C_{j(X,Y)} = C_{(X,Y)}$
4. $\sigma_{xy(j)} = \frac{1}{2}\sigma_{xy}$

where $j = 1, \ldots, 5$. Up to second moments of the puff are preserved by the splitting process. We assume that the perturbation to the material concentration distribution due to puff splitting is small.

3 Sensor Measurement model

The concentration at a sensor grid point $[x_g, y_g]$ is available at discrete times. At each sampling instant t_k, it is calculated by summing the contributions of all the puffs at that instant.

$$C_{gk} = \sum_{i_k=1}^{N_k} \frac{Q_{i_k}}{2\pi\sigma_{xy(i_k)}^2} \exp\left[-\frac{1}{2}\left(\frac{(X_{i_k} - x_g)^2 + (Y_{i_k} - y_g)^2}{\sigma_{xy(i_k)}^2}\right)\right]$$

where N_k is the number of puffs. Note that C_{gk} is linear in Q_{i_k}. The sensor output of the concentration is contaminated by zero-mean white Gaussian noise v_k with known variance, i.e.,

$$\tilde{C}_{gk} = C_{gk} + v_{gk} \tag{2}$$

4 Grid-Based Problem Formulation

A simple source identification problem is used to illustrate how the problem is formulated. The following assumptions are made in the problem:

1. All release rate and time parameters are known;
2. The error of the dispersion forecasting model is negligible.

Let the puff state (of size $4N_k$) at t_k be partitioned as

$$\mathcal{X}_k = \begin{bmatrix} (\mathcal{X}_k^C)^T & (\mathcal{X}_k^Q)^T \end{bmatrix}^T \tag{3}$$

where $\mathcal{X}_k^Q$ is a collection of the puff masses at t_k and $\mathcal{X}_k^C$ denotes the other states at t_k. The initial condition of the puff-based dispersion model is denoted by

$$\mathbf{x}_0 = \begin{bmatrix} \mathbf{x}_0^C & \mathbf{x}_0^Q \end{bmatrix} \tag{4}$$

Note that while $\mathcal{X}_k$ is a large vector of time-varying length $4N_k$ due to continuous release of puffs and puff splitting, the size of $\mathbf{x}_0$ is fixed (at most $4L$) and much smaller. The concatenation of the concentrations at all sensors at t_k will be denoted by $\mathcal{Y}_k$. Formally, the puff-based forecast model is given by

$$\mathcal{X}_{k+1}^C = \mathcal{F}(\mathcal{X}_k^C, \mathcal{X}_k^Q) \tag{5}$$

and

$$\mathcal{X}_{k+1}^Q = \mathcal{X}_k^Q \tag{6}$$

where $\mathcal{F}$ is a vector of nonlinear functions of $\mathcal{X}_k^C$ and $\mathcal{X}_k^Q$. The measurement model is given by

$$\mathcal{Y}_k = \mathcal{G}(\mathcal{X}_k^C) \cdot \mathcal{X}_k^Q \tag{7}$$

where $\mathcal{G}(\mathcal{X}_k^C)$ is a matrix of nonlinear functions of $\mathcal{X}_k^C$. Because both the dynamic model and the observation model are linear in $\mathcal{X}_k^Q$ and $\mathbf{x}_0^Q$, we must have

$$\mathcal{Y}_k = A'(\mathbf{x}_0^C) \cdot \mathbf{x}_0^Q \tag{8}$$

where $A'(\mathbf{x}_0^C)$ is a matrix of nonlinear functions of $\mathbf{x}_0^C$. Collecting all the measurements up to t_n, we have formally

$$\mathcal{Y}^n = A(\mathbf{x}_0^C) \cdot \mathbf{x}_0^Q \tag{9}$$

Now we introduce a grid over the parameter subspace $\mathbf{x}_0^C$. The concatenation of the grid points is denoted by $\mathbf{X}_0^C$. If the initial sizes of the puffs are known, we only need to grid the space of initial puff centroids (identical to source locations). The above equation is then rewritten as

$$\mathcal{Y}^n = A(\mathbf{X}_0^C) \cdot \mathbf{X}_0^Q \tag{10}$$

A zero component of $\mathbf{X}_0^Q$ means that there is no actual source at the corresponding location. This system of equations is underdetermined, but we can solve it using the ℓ_1 minimization method. In ideal cases, $\mathbf{X}_0^Q$ can be perfectly reconstructed as the sparsest vector satisfying $\mathcal{Y}^n = \mathcal{A} \cdot \mathbf{X}_0^Q$.

5 ℓ_1 Minimization

The p-norm of a vector is defined by

$$\|\mathbf{X}\|_p = (\sum_{i=1}^{N} |X_i|^p)^{(1/p)} \tag{11}$$

The optimal estimate $\hat{\mathbf{X}}$ is defined as the solution to the ℓ_1 minimization problem for the noiseless case in [5]:

$$\hat{\mathbf{X}} = \arg\min_X \|D\mathbf{X}\|_1 \quad \text{subject to } \mathcal{A}\hat{\mathbf{X}} = \mathcal{Y}^{\mathbf{n}},\ \mathbf{0} \le \hat{\mathbf{X}} \le \mathbf{X}^{UB} \tag{12}$$

where D is a diagonal matrix with the diagonal elements given by the 2-norms of the columns of A, $A\mathbf{X} = \hat{A}(D\hat{\mathbf{X}}) = (AD^{-1})(D\hat{\mathbf{X}})$, with $\hat{A} = AD^{-1}$ a matrix of unit (2-norm) column vectors. This is called an ℓ_1 minimization problem because the cost is defined by the 1-norm of the vector. In the original ℓ_1 minimization problem in [5], the matrix A in the constraint $A\mathbf{X} = \mathcal{Y}^n$ is assumed to be a matrix of unit column vectors. That is why we introduced the normalized version of A as well as D. The constraint $\mathbf{0} \le \hat{\mathbf{X}} \le \mathbf{X}^{UB}$ requires that all the elements of $\hat{\mathbf{X}}$ be nonnegative and upper bounded by $\mathbf{X}^{UB}$. This constraint is not in the ℓ_1 minimization problem in [5] but is important in our source term estimation problem, as will be shown in the next section.

The solution of minimum 1-norm is preferred to the well-known least-square solution, defined by

$$\hat{\mathbf{X}}^{LS} = \arg\min_X \|\mathbf{X}\|_2 \quad \text{subject to } A\hat{\mathbf{X}}^{LS} = \mathcal{Y}^{\mathbf{n}} \tag{13}$$

not only because the latter cannot guarantee nonnegative $\hat{\mathbf{X}}^{LS}$ but because we know the truth is represented by a sparse vector and the solution of minimum 1-norm is sparser than the solution of the minimum 2-norm. Were the truth $\mathbf{X}$ not a sparse vector, there would be no reason to prefer a sparse solution $\hat{\mathbf{X}}$ of our problem. To see that the 1-norm is a better measure of sparsity than the 2-norm, consider two N-dimensional unit vectors $[1, 0, \ldots, 0]^T$ and $[1/\sqrt{N}, \ldots, 1/\sqrt{N}]^T$. The former is sparse while the latter is not. The 2-norms of the two vectors are identical and the 1-norms of them are 1 and $\sqrt{N}$, respectively.

It should also be noted that although the solution of minimum 1-norm is sparser than that of minimum 2-norm, the solution of minimum 1-norm is not necessarily the sparsest. For example, if there are only two vectors $[1, 1, 1, 0, \ldots, 0]^T$ and $[10, 0, \ldots, 0]^T$ satisfying $A\mathbf{X} = \mathcal{Y}^n$, the former, because its 1-norm is smaller, will be chosen as the solution of the ℓ_1 minimization problem, but the latter is actually sparser. However, finding out the sparsest solution is NP-hard [6]. An iterative reweighting scheme in [6] helps improve the sparsity of the solution, but there is no absolute guarantee that the sparsest solution can be found by using the iterative reweighting scheme. The basic idea of iterative reweighting is to minimize a cost in favor of a sparser solution than that of minimum 1-norm. Formally, the new cost is given by

$$\hat{\mathbf{X}} = \arg\min_X \|[\mathrm{diag}(\mathbf{X})]^{-1}\mathbf{X}\|_1 = \arg\min_X \sum_{i=1}^{N} \frac{X_i}{|X_i|}$$

with $0/0 \triangleq 0$ in the summation. The new cost corresponds to the 0-norm of $\mathbf{X}$ or the total number of nonzero elements of $\mathbf{X}$. To avoid the "divide by zero" problem in practice, $X_i/|X_i|$ is replaced by $X_i/(|X_i| + \epsilon)$, where ϵ is a small positive number. The minimization problem is solved iteratively using the ℓ_1 minimization method, with the cost given by

$$\hat{\mathbf{X}}^{(l+1)} = \arg\min_X \sum_{i=1}^{N} \frac{X_i}{|\hat{X}_i^{(l)}| + \epsilon}$$

where $\hat{\mathbf{X}}^{(l+1)}$ is the optimal estimate in the $(l+1)^{\text{th}}$ iteration and $\hat{X}_i^{(l)}$ is the i^{th} element of the optimal estimate from the l^{th} iteration.

In the present work, the normalized version of $\tilde{A}$ is used in the problem definition, given by $\hat{\tilde{A}} = \tilde{A}\tilde{D}^{-1}$, where $\tilde{D}$ is the diagonal matrix of the 2-norms of the columns of $\tilde{A}$. The source estimation problem is formulated as:

$$\hat{\mathbf{X}} = \arg\min_X \|\tilde{D}\mathbf{X}\|_1$$
$$\text{subject to } \|W(\tilde{A}\hat{\mathbf{X}} - \mathcal{Y}^n)\|_2 \le \lambda,\ \mathbf{0} \le \hat{\mathbf{X}} \le \mathbf{X}^{UB} \tag{14}$$

where λ is a control parameter. A lower bound of λ is obtained by solving

$$\hat{\mathbf{X}} = \arg\min_{X} \|W(\tilde{A}\mathbf{X} - \mathcal{Y}^n)\|_2 \quad \text{subject to } \mathbf{0} \leq \hat{\mathbf{X}} \leq \mathbf{X}^{UB} \tag{15}$$

The tuning parameter λ is determined by simulations. When the control parameter is too small, no feasible solution exists. When it is too large, the optimal solution may be nothing but the null vector. We have used 1.1 times the lower bound in the second formulation. For the weighting matrix W, we have used $W = I_{M\times M}$ because of its simplicity. The optimal weighting matrix W requires knowledge about the noise as well as the error.

6 Numerical Tests

The performance of the source estimation technique using the above method, is studied for the puff-based dispersion model, using simulated concentration measurements. Various source configurations are considered for the purpose of illustration.

6.1 Setup

The sampling interval for the dispersion model is $2s$ and the release from the sources continues for about $50s$ with a puff being released every three time-steps ($6s$). The units of activity can be chosen according to the species under study. Note that the units of concentration are then $ActivityUnits/m^2$. The wind speed is $5m/s$ and blows at $\frac{\pi}{4}$ across the region.

In the simulation experiments, the sources are assumed to be anywhere in the $400m \times 400m$ square region around the origin, shown as a shaded region in figure 1. The activity of the sources is indicated by the activity of each puff released which, at any given source location, varies uniformly between 1×10^5 and 50×10^5. The true source locations and activities are modeled as uniformly distributed random variables. The concentration measurement sensors are located $100m$ apart along each of the axes as shown in figure 1. The noise in each of the sensors is modeled as a random Gaussian distribution, with the standard deviation as 1% of the average concentration across the region.

6.2 Compute A and $\mathcal{Y}^n$

The matrix A is determined by the sensor configuration and the grid resolution of the uncertainty region. Its size is $n \times m$, where n is the number of concentration measurements used for the purpose of estimation, and m is the total number of grid points into which the $400m \times 400m$ source location uncertainty region is resolved. Each column of A is a collection of all the n predicted measurements due to a source at a particular grid point in the uncertainty region. The vector $\mathcal{Y}^n$ is a collection of all the noisy concentration measurements during the truth simulation, at various time instants.

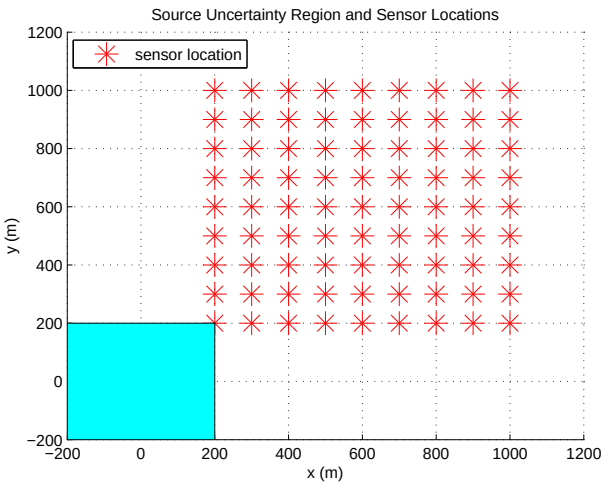

Figure 1: Source uncertainty region and Sensor locations on the grid

6.3 Results for source estimation

For the purpose of computing A, the source uncertainty region is uniformly partitioned with a resolution of $10m$ resulting in $m = 41^2$ grid points. The concentration measurements at all sensor locations, at $60s$, $80s$ and $100s$ are used to construct $\mathcal{Y}^n$. The predicted concentration measurements for a source of puff activity 1×10^5, at each of the grid points make up the columns of A.

Fig. 2 shows the estimation result for a case of a single random source, with an uncertain strength and location. It can be seen that estimated locations and strengths of the source are close to the truth, even though two separate but nearby sources are predicted. Similarly the estimation result for a case of three random sources is illustrated in fig. 3. Even though six different sources are estimated, the estimated locations are close to the actual sources and the estimated strengths are comparable.

6.4 Results with uncertain source release durations

The method is further extended to estimate the source characteristics, when the duration of source release is uncertain. The duration of source release is characterized by a parameter α, the fraction of the total simulation time ($T = 240s$), after which the source release stops. In the above simulations, α is fixed at 0.2, meaning the source release stops after $48s$. In this example, α is modeled as a uniformly distributed random variable between 0.15 and 0.25, which means the source release stops anywhere between $36s$ and $60s$. To accommodate for varying α, A is recomputed for various α. The uncertainty interval of α is uniformly partitioned into 5 grid points, and each original column of A is now replaced by 5 columns corresponding to uncertain α. This results in $m = 5 \times 41^2$ columns for A in this example. Fig. 4 shows the estimation result for a case of two random

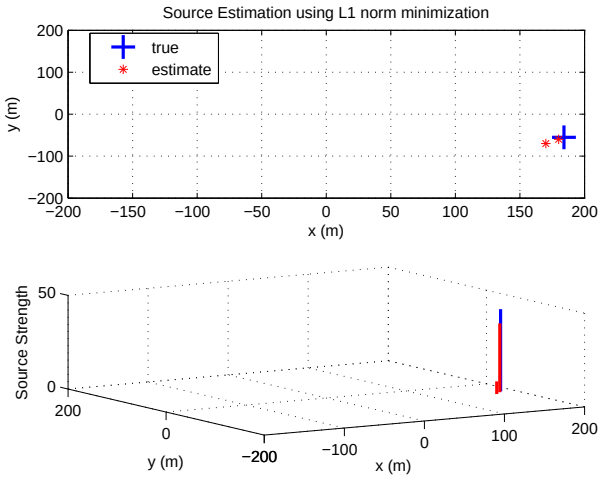

Figure 2: Source estimation (1 source)

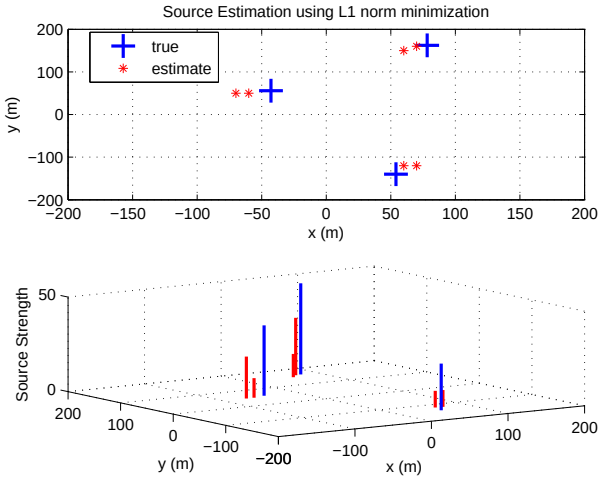

Figure 3: Source estimation (3 sources)

sources, with an uncertain location, strength and release duration.

7 Discussion of the results

The performance of the method can be described using the root mean squared error (RMSE) between estimated and true source locations and strengths. Several scenarios are considered as shown in Table 1. The RMSE for each scenario is evaluated using 50 Monte Carlo runs for each of them. For the purpose of evaluation of RMSE, the estimated sources are clustered into the same number of clusters as is the number of true sources. This is done by assigning each estimated source to its nearest true source. Note that there might be a few runs (typically $5-6$) in which a true source might not have any assigned estimated source. In such a case, the estimated source is assumed to be at the true source

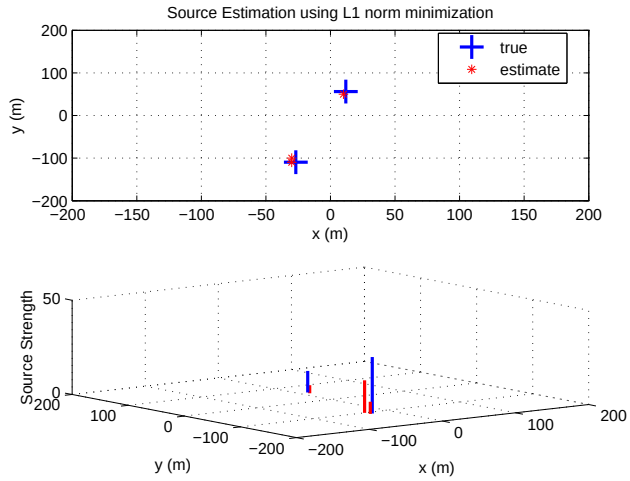

Figure 4: Source estimation with variable α (2 sources)

location with zero strength. Then the estimated source location, corresponding to a particular true source, is obtained by taking the strength-weighted mean of the estimated source locations in that cluster. The estimated source strength is the sum of the strengths in that cluster. The RMSE is evaluated by comparing these cluster means and strengths with the corresponding true source locations and strengths. The RMSE is obtained by taking square-root of the mean of all the squared errors.

Table 1 shows the results for different scenarios. All scenarios have the sensor observation noise as described before in section 6.1. It can be seen that the estimation performance drops as expected, with the increase in process model uncertainty. The wind noise is modeled as a random Gaussian variable with a standard deviation of $1m/s$ (SNR = $20dB$).

8 Conclusion

A convex optimization based method for assimilation of sensor data using atmospheric dispersion models to estimate the source parameters in chemical, biological, radiological and nuclear (CBRN) incidents is presented. The effectiveness of the method is illustrated in the context of atmospheric chemical dispersion, using a nonlinear Lagrangian puff-based dispersion model and a nonlinear sensor model measuring concentration. The source identification problem is formulated as a convex optimization problem in the ℓ_1 metric, which exploits the sparse nature of the solution to efficiently estimate the source characteristics, even when the number of sources is large. The current work illustrates the effectiveness of the method using simulated measurements which are similar to the model predictions. While the approach works quite well in predicting uncertainty in multiple sources when the model predictions and observations are close, further work needs to be done to extend this method to combine actual field observations with dispersion models.

No. of Sources	RMSE position (m)	RMSE strength ($\times 10^5$)
Fixed α	no wind noise	
1	1.2598	0.4315
2	5.0801	5.0502
3	10.2308	5.7835
Uncertain α	no wind noise	
1	1.6736	0.5821
2	10.4144	2.4652
3	13.1449	6.6882
Fixed α	with wind noise	
1	8.5436	1.3590
2	15.1855	3.3051
3	25.7999	8.5507
Uncertain α	with wind noise	
1	11.5242	2.1974
2	23.0967	7.4102
3	32.1244	10.5595

Table 1: RMSE for 50 Monte Carlo runs

References

[1] K. S. Rao, "Source estimation methods for atmospheric dispersion," *Atmospheric Environment*, vol. 41, pp. 6964–6973, 2007.

[2] Y. Cheng and T. Singh, "Source term estimation using convex optimization," in *The 11th International Conference on Information Fusion*, Cologne, Germany, 2008.

[3] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004, available at http://www.stanford.edu/ boyd/cvxbook. html.

[4] S. Thykier-Nielsen, S.Deme, and T. Mikkelsen, "Description of the atmospheric dispersion module rimpuff," Riso National Laboratory, P.O.Box 49, DK-4000 Roskilde, Denmark, Tech. Rep. RODOS(WG2)-TN(98)-02, 1999.

[5] E. J. Candès, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, Madrid, Spain, 2006.

[6] E. J. Candès, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," preprint.