J1.5: COMBINED METHODS FROM ENTITY AND FIELD FRAMEWORKS TO DETERMINE THE SOURCE CHARACTERISTICS OF A CONTAMINANT

Andrew J. Annunzio*, Sue Ellen Haupt, George S. Young, and Luna Marie Rodriguez
The Pennsylvania State University, State College PA

1. Introduction

Determining the source characteristics of a contaminant is an important issue in homeland and defense security in order to accurately predict subsequent atmospheric transport and dispersion (AT&D). In previous work, we have utilized the field approach to obtaining the contaminant source information. There, the source information is obtained by matching observed surface concentration values with predicted surface concentrations from a dispersion model that has the source information as an input via a Genetic Algorithm (GA) (Haupt 2005; Allen et al. 2007, 2008; Long et al. 2010, Rodriguez et al. 2010). In other work, we investigated the implications of Lagrangian and Eulerian frameworks on the AT&D problem (Annunzio et al., 2010). There, it is determined that the Lagrangian framework has advantages over the Eulerian framework when it is possible to describe the concentration data by an entity. Here, we apply this Lagrangian/entity approach to the source term estimation (STE) problem. This approach is also utilized in other work where parcel trajectories are traced back to the source location; however, this method requires ample meteorological wind and concentration data (Young et al. 2009). The Lagrangian method developed here does not require wind data because the relevant wind information is diagnosed from the concentration observations. Two separate methods are developed: a strictly Lagrangian method for an instantaneous release and a mixed Eulerian/Lagrangian approach for a continuous release. The methods described here approximate the concentration by a single entity in a single entity field approximation (SEFA). For this formulation, we analyze the state of a contaminant plume and a time series of contaminant puff states to find the plume/puff source location. The components of the state are the puff/plume axis and spread; we focus on these components because the contaminant source is located at the point on the puff/plume axis where the spread equals zero. Many Eulerian techniques exist for STE, most of which match surface concentration observations with surface concentration predictions (Rao 2007). Our Lagrangian approach for instantaneous, single sources is advantageous because for this scenario it is not necessary to match the concentration field with predictions of the concentration field to obtain the source information.

For a continuous release, contaminants appear to remain stationary with respect to the sensor grid, and thus, a Lagrangian formulation is difficult for a contaminant plume. Therefore, a mixed Eulerian-Lagrangian approach is adopted for a continuous release: this approach is mixed Eulerian-Lagrangian because surface concentration observations are matched with surface concentration predictions to determine a Lagrangian quantity, the spread. A further advantage of our techniques for single sources is that meteorological data are not required to accurately estimate the source location.

2. Methods

In this section we develop the Lagrangian based source term estimation algorithms for both an instantaneous and continuous release.

a. Instantaneous release

To determine the source location for an instantaneous release, the evolution of the puff state must follow a dynamical system. Here, we assume a simple dynamical system given by

$$x_p = x_o + \bar{u}(t - t_o) \tag{1a}$$

$$y_p = y_o + \bar{v}(t - t_o) \tag{1b}$$

$$S_i = a(t - t_o)^b \tag{1c}$$

where x and y describe the position of the puff centroid at a time t, t_o is the initial release time, $\bar{u}$ and $\bar{v}$ represents the mean zonal and meridional velocity of the contaminant puff, S_i is the puff spread, and a and b are constants that are inputs to the power law equation (1c) that describes the puff spread. Without a time series of surface concentration observations, none of the variables are known; however, it is possible to determine x_p, y_p, and S from these concentration observations. The puff

* Corresponding author address: Andrew Joseph Annunzio, The Pennsylvania State University, The Applied Research Lab and the Department of Meteorology, State College PA, 16804, email: aja199@psu.edu

centroid, represented by x_p and y_p is determined from a concentration weighted location

$$x_p = \frac{\sum_{i=1}^{N_x} C_i x_{g,i}}{\sum_{i=1}^{N_x} C_i} \qquad (2a)$$

$$y_p = \frac{\sum_{i=1}^{N_y} C_i y_{g,i}}{\sum_{i=1}^{N_y} C_i} \qquad (2b)$$

where $x_{g,i}$, $y_{g,i}$ is the horizontal location of a grid point in the x, y plane, N_x and N_y represents the number of grid points in each direction, and C_i is the concentration values at the i^{th} grid point. This calculation is computed at discrete times when concentration observations are available, and from these calculations there is a time series of the puff centroid location. Further, with knowledge of the puff centroid, determination of the puff spread is possible. The puff spread is calculated through a concentration weighted sum.

$$S_i = \frac{\sum_{j=1}^{N_g} r_j C_j}{\sum_{j=1}^{N_g} C_j} \qquad (3)$$

where $r = \left((x_p - x_j)^2 + (y_p - y_j)^2\right)^{\frac{1}{2}}$ is the Euclidian distance from the puff centroid to a grid point, and N_g describes the number of grid points. From these calculations, a time series of the puff spread is available. The puff spread is assumed to be a power law function the depends on the unknown variables a, b, and t_o. Because the power law function is nonlinear, optimization is required to determine these variables. The optimization used for the fit is a Genetic Algorithm (GA), a robust optimization technique that mimics the natural selection process (Haupt and Haupt 2004, Haupt 2005). The GA minimizes a cost function given by

$$\frac{\left(\sum_{i=1}^{N} \left(S - a(t - t_o)^b\right)^2\right)^{\frac{1}{2}}}{\left(\sum_{i=1}^{N} S^2\right)^{\frac{1}{2}}} \qquad (4)$$

and this fit provides a best estimate of the contaminant release time, t_o.

From the time series of the puff centroid, we can calculate the mean puff translation velocity and direction (or puff axis) in the dynamical system (1a)-(1c).

$$\bar{u} = \frac{1}{(N_t - 1)} \left(\sum_{i=2}^{N_t} x_i - x_{i-1}\right) \qquad (5a)$$

$$\bar{v} = \frac{1}{(N_t - 1)} \left(\sum_{i=2}^{N_t} y_i - y_{i-1}\right) \qquad (5b)$$

where N_t represents the number of puff centroid observations. After this calculation, only two unknowns remain in the dynamical system (1), namely the two dimensional source location. To calculate the source location, we can easily invert equations (1a) and (1b). This inversion requires input from the first puff centroid observation.

$$x_o = x_1 - \bar{u}(t_1 - t_o) \qquad (6a)$$

$$y_o = y_1 - \bar{v}(t_1 - t_o) \qquad (6b)$$

b. Plume

For a continuous release, we take a more parametric, mixed Eulerian/Lagrangian approach to obtaining the source information. The plume axis is determined from the concentration data while the plume spread is obtained by matching the concentration field. The first crucial step in this method is time averaging the contaminant concentration field. The averaging period is determined as the time from when the contaminant first enters the sensor domain until the contaminant exits the sensor domain. The next step in determining the entity state is fitting the plume axis. This is accomplished with a concentration weighted least squares fit

$$\left(\frac{\partial}{\partial a} \left(\sum_{i=1}^{N_g} C_g x_g - (ax+b) \right)^2 \right) = 0 \tag{7a}$$

$$\left(\frac{\partial}{\partial b} \left(\sum_{i=1}^{N_g} C_g x_g - (ax+b) \right)^2 \right) = 0 \tag{7b}$$

The concentration weighted least squares fit is used so that the location of the plume axis is heavily weighted by the highest concentration values. The concentration weighted approach has one drawback in that it does not account for horizontal and vertical diffusion that decreases concentration values on the plume axis downwind from the source. However, because the plume spread and atmospheric turbulence statistics are unknown, it is not possible to accurately include the effects of atmospheric diffusion for this fit. The effective plume axis provides information on the influencing wind because the plume axis lies parallel to the mean direction of contaminant travel. With the information on the mean plume axis, it is now possible to determine the contaminant spread. Taking a similar approach to that used for the instantaneous case, we assume that the plume spread follows a power law

$$S_c = c(x' - x_o')^d \tag{8}$$

Where x' represents points that lie on the plume axis, x_o' is the unknown source location that lies on the plume axis, and c and d are constants to be determined. Unlike the instantaneous release, we cannot compute a time series of the spread or calculate values of *S* as a function of *x*. Therefore, these three unknowns must be determined from the concentration field. This is possible if S_c is an input to a dispersion model and the unknown variables are determined by matching the concentration observations with predictions of surface concentration observations. For this work, the dispersion model used is the Gaussian dispersion model written as

$$C = \frac{M}{U 2\pi S_c^2} \exp\left(\frac{(y-y')^2}{2 S_c^2} \right) \tag{9}$$

The process of matching the data is accomplished by minimizing the difference between the concentration observations and the concentration forecasts with a Genetic Algorithm (GA) similar to Haupt (2005), Allen et al (2007 and 2008), and Long et al (2010). The GA optimization process is discussed in those works, and thus, a discussion is omitted here. The implementation of a GA is not crucial for this work, however, this optimization technique is chosen because of its robustness compared to other optimization techniques (Haupt and Haupt 2004).

3. Results

To test these algorithms, we used trial data from the FUSION Field Trial 2007 (FFT 07) dataset. FFT 07 was a field experiment where surface contaminant concentration observations were recorded from continuous and instantaneous contaminant releases. These data, as well as meteorological observations taken during the contaminant release, are available for researchers to test their source term estimation algorithms. For each trial, 100 concentration sensors recorded data on the contaminant, 40 meteorological sensors recorded wind data, and 3 towers were available to calculate turbulence statistics. We tested these algorithms on trials 15 and 71, which are single release plume and puff situations respectively. For both trials, we used none of the meteorological data, because the relevant meteorological information is inferred by the algorithms. Further, we randomly removed sensors from the 100 sensors available to test our algorithms when less data are available. We perform this process in 100 separate Monte Carlo simulations

Instantaneous release results for 80, 60, 40, and 20 sensors are shown in figures 1, 2, 3 and 4 respectively. Figure1 shows that our algorithm can consistently determine the puff axis and spread, and hence accurately estimate the source location of the contaminant when 80 sensors report concentration data. As more data are removed, it becomes more difficult to determine the puff axis and spread yielding less accurate source terms estimates exemplified by figures 2 and 3. When only 20 sensors report surface concentration data, the source term estimates are even less accurate as expected. This is indicated in figure 4: the inability to locate the contaminant source is due to the difficulty in computing the puff axis and spread. This inability to determine these variables is because of insufficient data. For the 20 sensor cases, a median of only 2 sensors sense the contaminant puff throughout the entire simulation making it difficult for any algorithm to determine the unknown variables.

Figures 5 through 8 display results for the continuous release when 80, 60, 40 and 20 sensors are available to sense the plume. Results for the 80 sensor cases are very promising; the mean source estimate error of the 100 simulations is 55.4 m. This is seen in figure 5 with the cluster of source estimates near the actual source location. Similar to the instantaneous release, as the sensor density decreases for the continuous release, the errors in source location estimates increase. In figures 6 and 7, the cluster of estimates near the origin expands, and we also have several estimates that diverge to the far edge of our search domain. With less sensor data, it becomes more difficult to accurately determine the plume axis and spread, and hence, less accurate source location estimates. Despite less data, the mean source location error is 118.5 m and 165.2 m for the 60 and 40 sensor simulations respectively and the median error is 57.9 m and 88.5 m. Figure 8 shows results when only 20 concentration sensors are available. For this scenario it is increasingly more difficult to determine the plume axis and spread, because only a median of 2 sensors report contaminant concentrations.

4. Conclusions

Our results show that when sufficient concentration data is available to compute the plume/puff axis and spread, we are able to determine the source of continuous and instantaneous releases. As the sensor density decreases, it becomes more difficult to determine these variables. For these scenarios, it is possible that only one or two sensors actually report contaminant concentrations, making it very difficult to find the source location of the contaminant. It is worthy of note, that accurate source term estimates are still possible when only few sensors report contaminant data. Accurate estimates are possible if these few sensors are located such that information is available on the plume/puff axis and spread, however, not every sensor configuration will provide this necessary information.

This algorithm was compared with other source term estimation algorithms by submitting results for FFT 07 case data. The FFT 07 case data has less wind and concentration data than the trials, and the cases we tested had 16 concentration sensors. The source term of the contaminant was unknown for the cases, and results were submitted for validation; several researchers submitted results from their source term estimation algorithms. The results of the case data show that for simulations with a single puff or plume release, our algorithms, on average, can estimate the source location better than most competing algorithms (Platt and Deriggi 2010).

In future work we will extend this model to a multi-entity field approximation (MEFA). A MEFA is appropriate when more than one contaminant release occurs or when the turbulence or flow obstacles cause an entity to split.

5. References

Annunzio, A.J., S.E. Haupt, and G.S. Young, 2010; Comparison of Multisensor Data Fusion and Data Assimilation applied to contaminant dispersion in a turbulent flow field, *Submitted to J. Atmos. Sci.*

Allen C.T., G.S. Young, and S.E. Haupt, 2007: Improving Pollutant source characterization by better estimating wind direction with a genetic algorithm. *Atmos. Environ.* **41,** 2283-2289.

Haupt R.L., S.E. Haupt, 2004: *Practical Genetic Algorithms, second ed. with CD*. Wiley, New York, NY, 255 pp.

Haupt S.E., 2005: A Demonstration of Coupled Receptor/Dispersion Modeling with a Genetic Algorithm *Atmos. Environ.* **39**, 1781-1788.

Long, K.J., S.E. Haupt, and G.S. Young, 2010: Assessing Sensitivity of Source Term Estimation. *Submitted to Atmos. Environ, accepted.*

Platt, N., and D. Deriggi, 2010: Comparative investigation of source term estimation algorithms using fusion field trial 2007 data. 8th Conference on Artificial Intelligence Applications to Environmental Sciences at AMS Annual Meeting, Atlanta, GA, Jan 17-21.

Rao, K.S., 2007: Source estimation methods for atmospheric dispersion, *Atmos. Environ*., **41**, 6983-6973.

Rodriguez, L.M., S.E. Haupt, and G.S. Young, 2010: Impact of Sensor Characteristics on Source Characterization for dispersion modeling, submitted to *Measurement.*

Young, G.S., J.A. Limbacher, S.E. Haupt, and A.J. Annunzio, 2009: Back trajectories for Hazard Origin Estimation, Seventh Conference on Artificial Intelligence and its Applications to the Environmental Sciences at AMS Annual Meeting, Phoenix, AZ, Jan. 11-15.

6. Figures

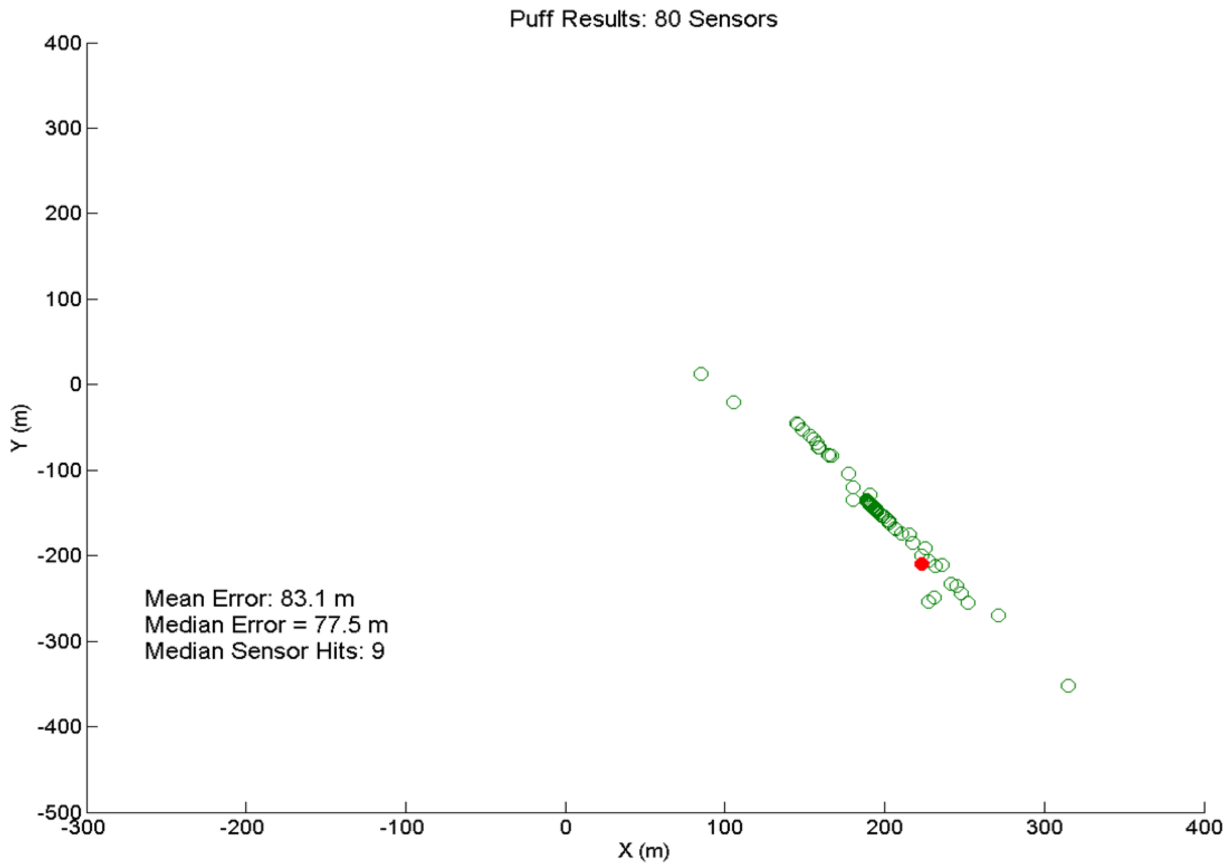

Figure 1: Source location estimates (green) and the actual source location (red) when 80 sensors are available to sense a contaminant puff

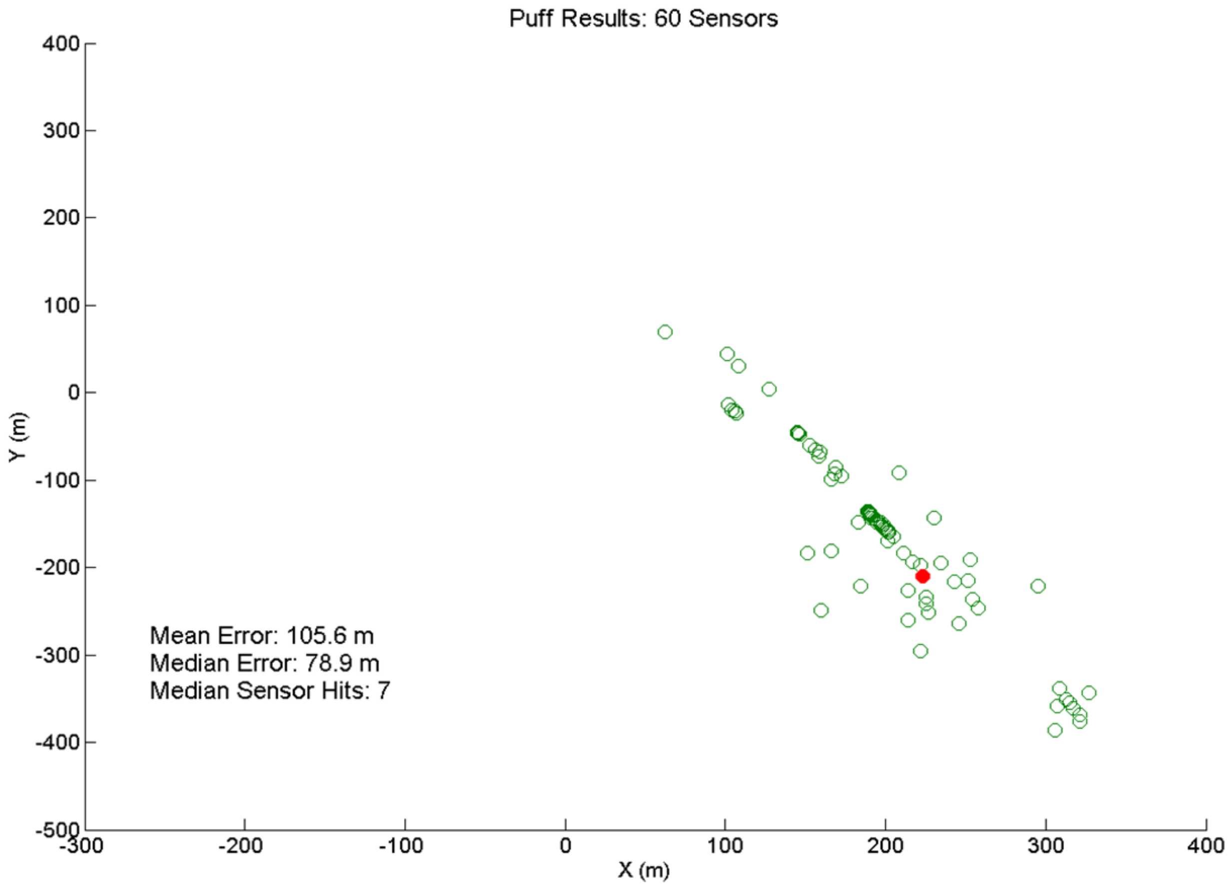

Figure 2: Source location estimates (green) and the actual source location (red) when 60 sensors are available to sense a contaminant puff

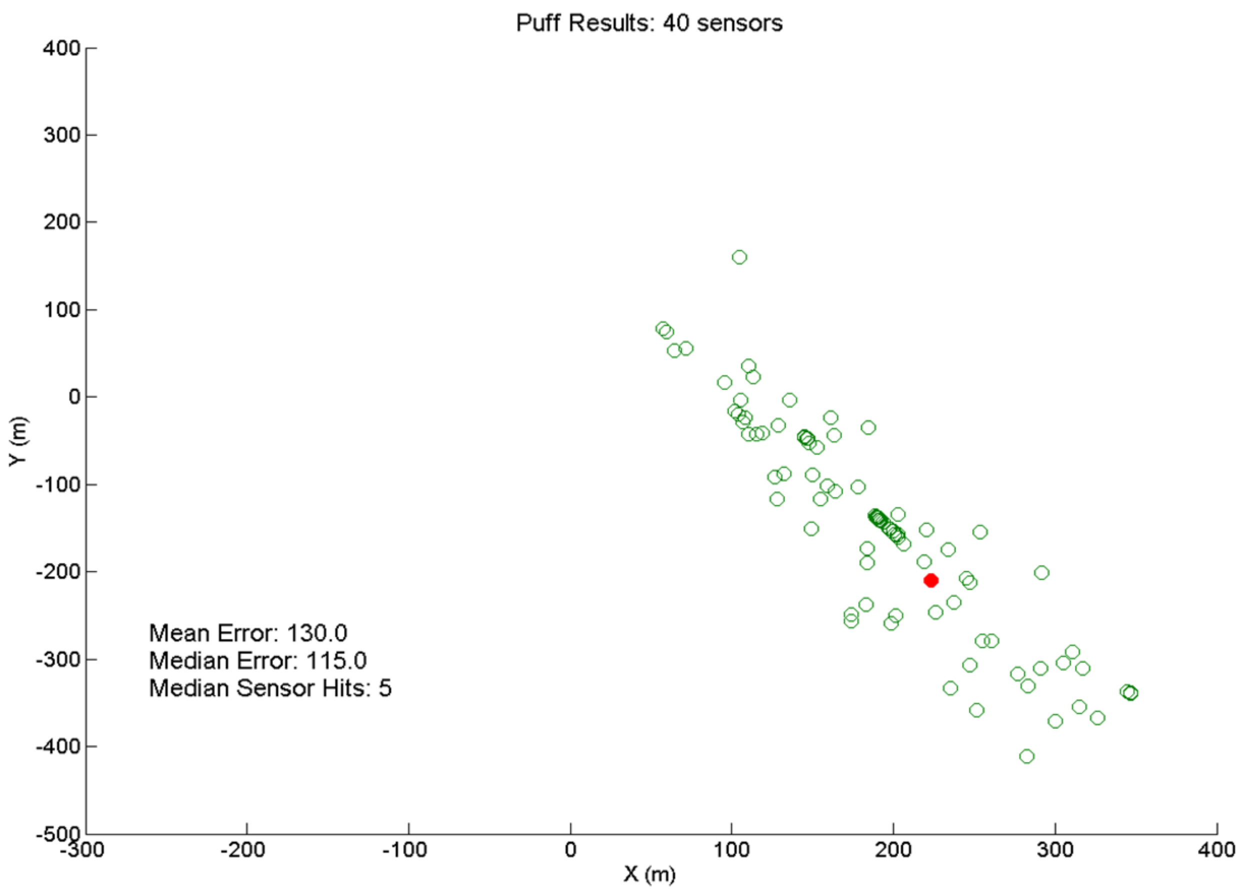

Figure 3: Source location estimates (green) and the actual source location (red) when 40 sensors are available to sense a contaminant puff

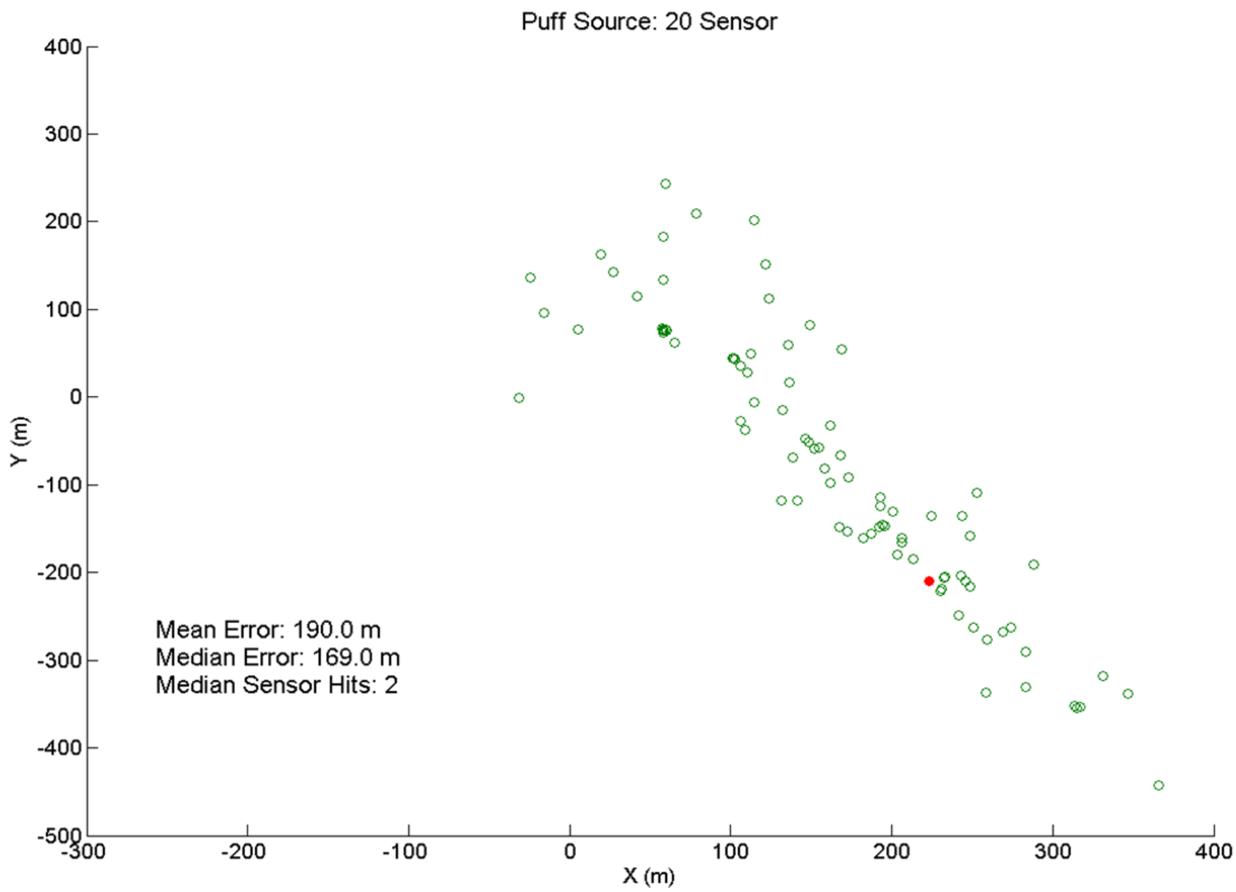

Figure 4: Source location estimates (green) and the actual source location (red) when 80 sensors are available to sense a contaminant puff

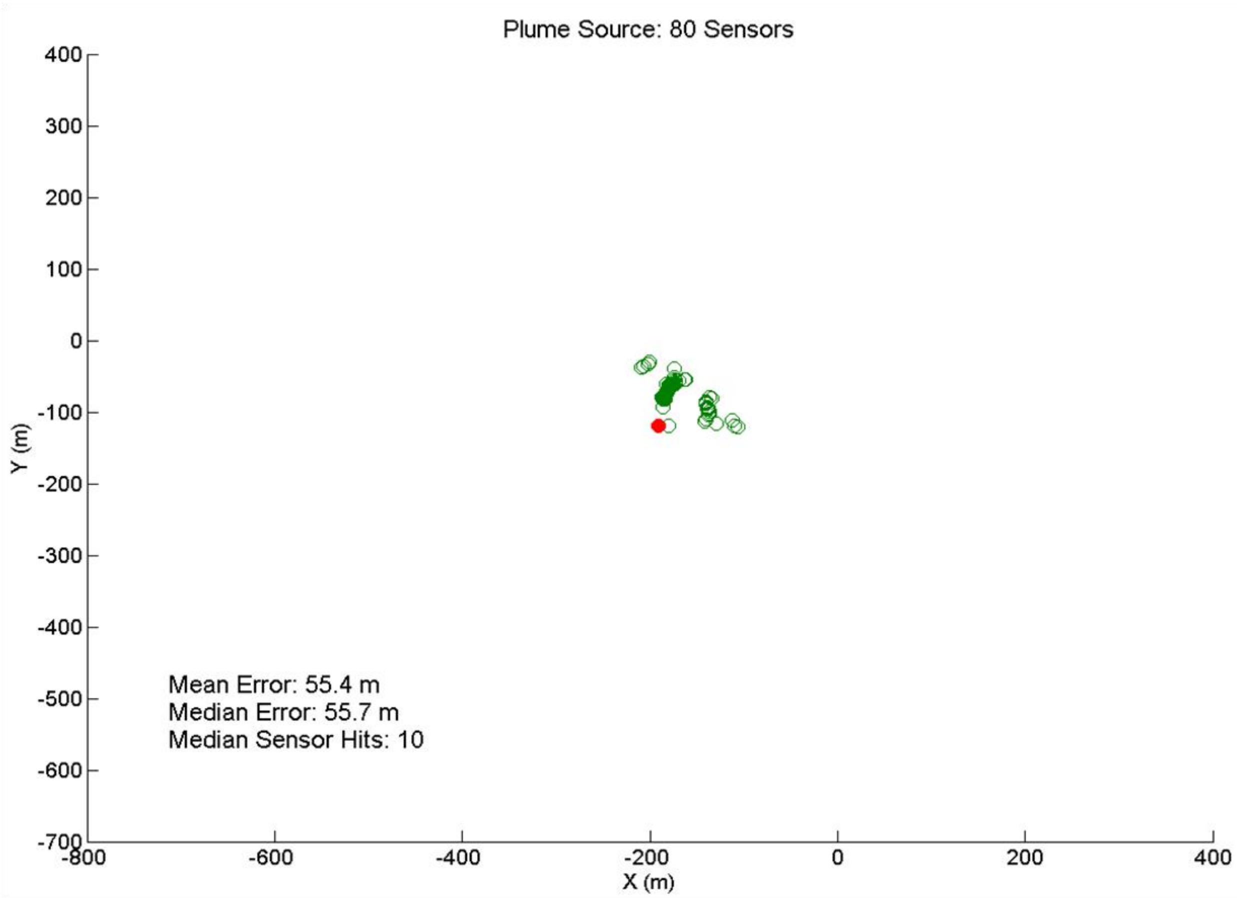

Figure 5: Source location estimates (green) and the actual source location (red) when 80 sensors are available to sense a contaminant plume

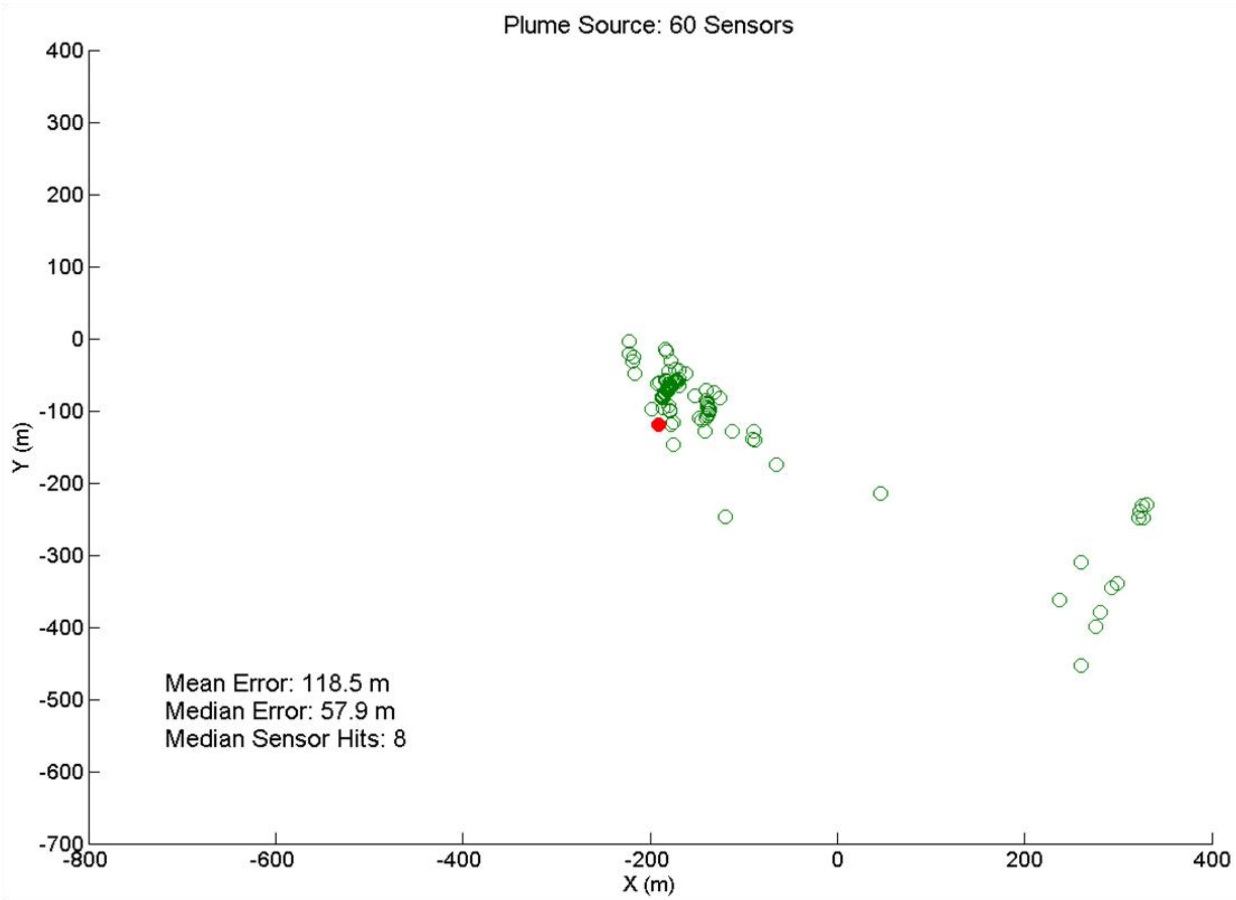

Figure 6: Source location estimates (green) and the actual source location (red) when 60 sensors are available to sense a contaminant plume

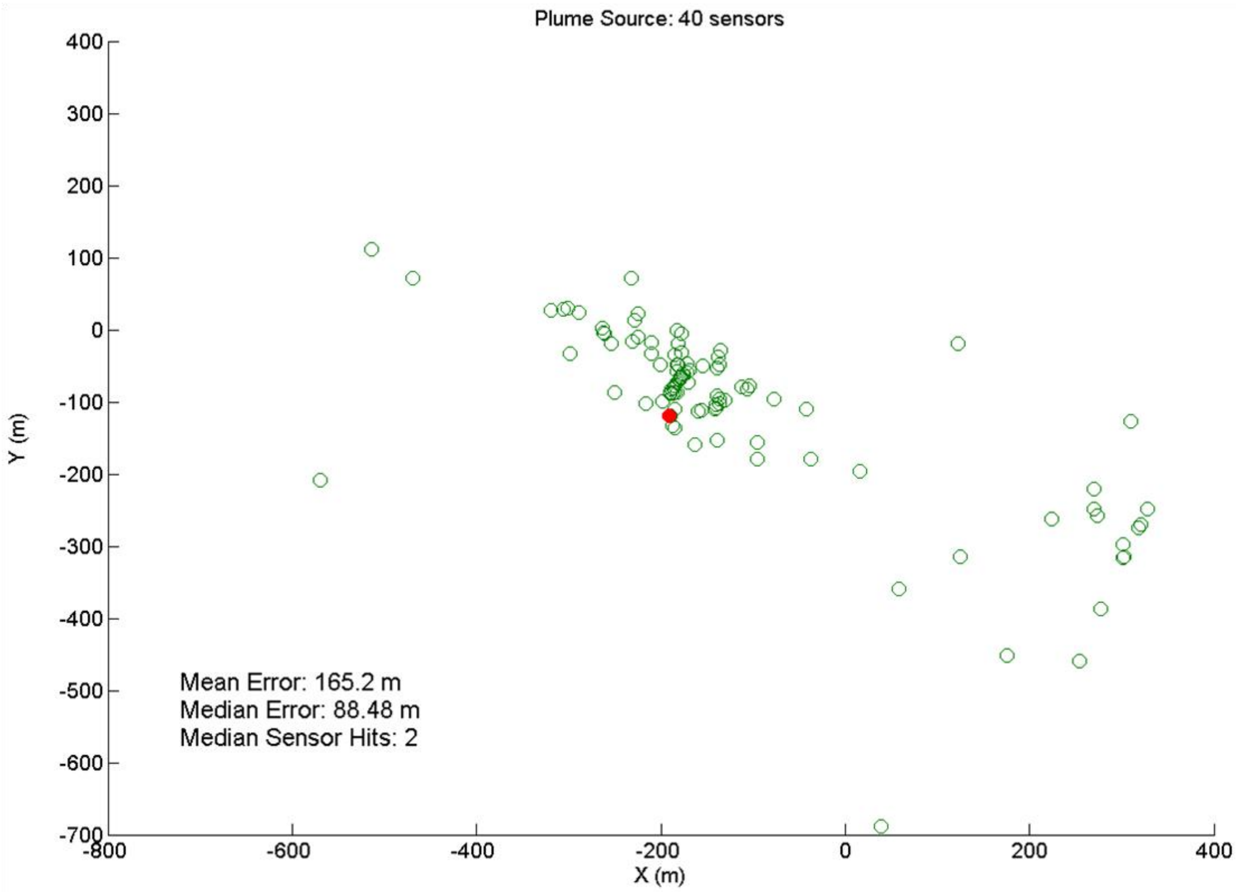

Figure 7: Source location estimates (green) and the actual source location (red) when 40 sensors are available to sense a contaminant plume

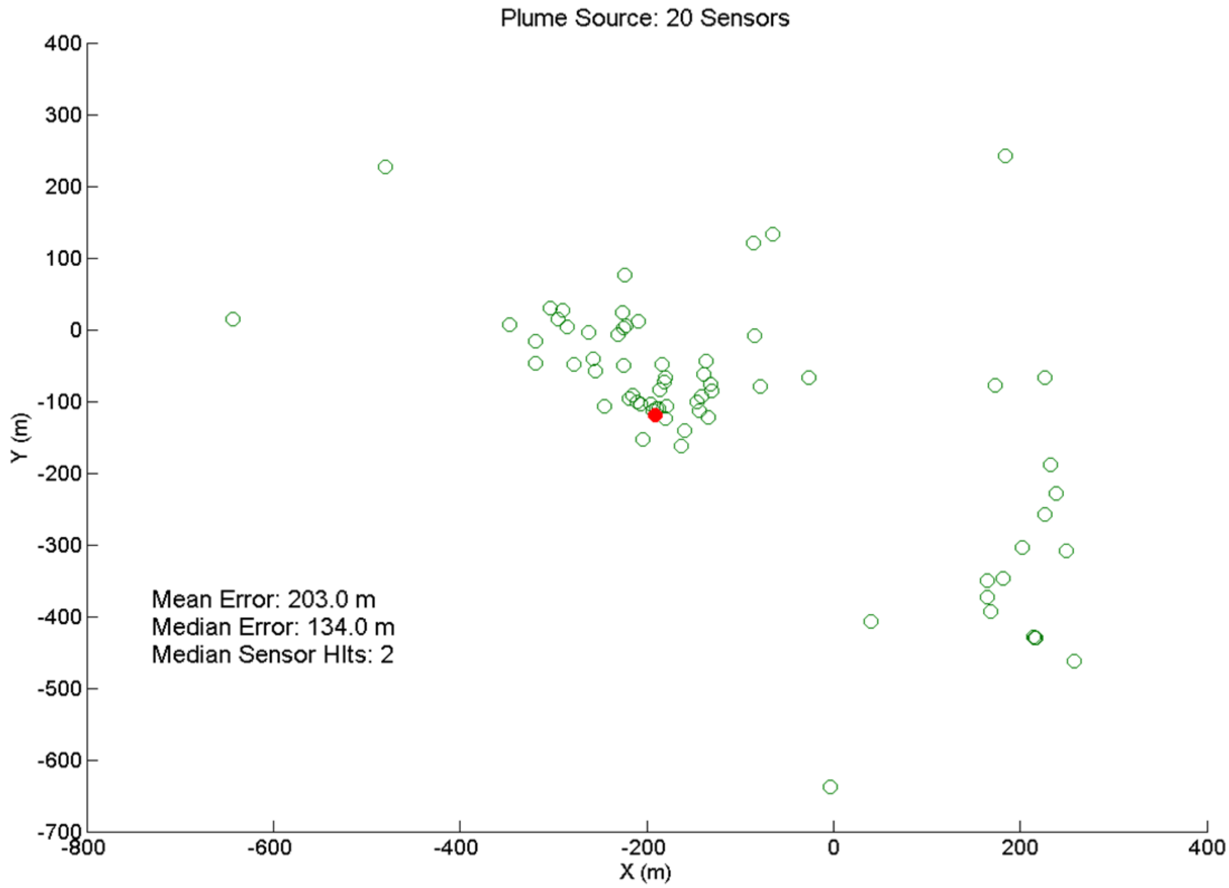

Figure 8: Source location estimates (green) and the actual source location (red) when 20 sensors are available to sense a contaminant plume