3.3 SOURCE INVERSION IN CITIES USING THE COLLECTOR FOOTPRINT METHODOLOGY

Dragan Zajic and Michael J. Brown
Los Alamos National Laboratory, Los Alamos, NM

1. Introduction

Source inversion modeling in urban areas has received a lot of attention recently due to its application in homeland security chemical, biological, and radiological air monitoring networks, as well as in identifying the potential industrial pollution sources in rapidly expanding urban areas. The goal of source inversion modeling is to find efficient methods to estimate the location, strength and number of pollution sources using sensor network measurements and known meteorological conditions. In this paper, we describe a relatively fast method for doing source inversion in cities where the effects of buildings on transport and dispersion are explicitly modeled.

2. Background

In recent years, a popular approach for performing source inversion and estimating the probability distributions of source parameter values has been Bayesian inference in combination with Markov Chain Monte Carlo (MCMC, e.g., Keats et al., 2006; Chow et al., 2006; Yee, 2007; Senocak et al., 2007). The probability distribution is obtained by running a dispersion model with a series of guesses for the source location and source strength and based on agreement or disagreement with the sensor measurements the new source parameters are either accepted or rejected based on comparison with the results from the previous set of source parameters. The new source parameters are accepted if it gives better agreement (i.e., has a lower cost function) than the previous set of source parameters, but with a certain probability they may also be accepted even if it gives worse agreement. The MCMC approach provides the methodology for choosing the new source parameters based on the old parameters.

Other optimization methods have been used for source characterization. For example, Thompson et. al. (2006) used simulated annealing with a Gaussian plume model to estimate source location and strength in a desert area as one way of finding new oil and gas reserves. They also tested the model's performance for different cost functions as well as its sensitivity to randomly introduced noise and offsets to the concentration data. Another recently used stochastic search approach for these types of problems is genetic algorithms (Allen and Haupt, 2007; Allen et al., 2007). They used this method to estimate locations and strengths of multiple sources. This approach is very useful for estimation of pollution emission from multiple industrial facilities in urban areas. The disadvantage of these optimization methods as compared to the Bayesian/MCMC approach is that it only provides a single solution instead of a probability distribution.

All of these approaches require a large number of runs of the forward dispersion model during the search for the best source parameters. This can be impractical for fast response applications that require quick results. The Gaussian plume model is very useful for these approaches since it is fast, but it is not very accurate in urban areas since it does not account for the presence of buildings. Computational fluid dynamics (CFD) models can be used that give more details about the flow and concentration fields, but they are computationally expensive and therefore not practical for fast response applications. One way to reduce computational time is to run CFD models in advance and create a "look-up" database for different wind directions and source locations and retrieve these values during the inverse model's calculations.

The different optimization methods for finding the source parameters are computationally expensive since they require many transport and dispersion calculations to be performed. In urban environments, the use of CFD modeling to account for effects of the buildings on the flow field will result in the need for massively parallel computer platforms to accomplish source inversion. In the next section, we describe an approach which is much faster, but still accounts for the effects of buildings. The approach uses the collector footprint concept in which the number of plume calculations is equal to the number of concentration sensors and the dispersion model is a fast-running empirical-diagnostic code. We also discuss using simulated annealing with the footprint approach in order to find the optimal solution for problems where there is large uncertainty in the measurements.

3. Source Inversion Method for Releases in Cities

3.1. *Transport and dispersion model*

The Quick Urban & Industrial Complex (QUIC) dispersion modeling system is used for the fast computation of the three dimensional flow and concentration fields around buildings. It consists of three basic components: the QUIC-URB model (Pardyjak and Brown, 2002 & 2003) that calculates 3D wind fields based on Röckle (1990) by determining flow regimes from building-spacing logic, using empirically-obtained building-flow parameterizations and imposing mass conservation, the QUIC-PLUME Lagrangian random-walk model that computes turbulence parameters and concentration fields (Williams et al 2004), and the QUIC-GUI graphical user interface that was developed to

*Corresponding author address: Dragan Zajic, Los Alamos National Laboratory, Group D-3, Los Alamos, NM 87545. E-mail:dzajic@lanl.gov.

facilitate user input and visualization of results. Later additions to the modeling system include the QUIC Pressure Solver that enables calculations of the pressure field (Gowardhan et al., 2006a), a sensor siting tool for optimum placement of sensors within a given area, and dense gas dispersion capabilities (Williams et al., 2004). The code has been shown to perform fairly well when tested against urban tracer data and wind-tunnel building experiments (e.g., Gowardhan et al., 2006b; Pol et al., 2006).

3.2. *Detector footprint approach*

The source inversion code is based on the concept of using detector footprints – a region upwind of each detector in which a release of a certain size would be detected by the detector – in order to quickly determine regions where the release might have occurred. The number of plume dispersion calculations that need to be solved in this approach is equal to the number of detectors. This technique has been implemented into an operational tool called the Biological Agent Event Reconstruction Tool (BERT) for homeland security applications (see Linger et al., 2005).

The concept of the detector footprint is illustrated in Fig. 1. Figure 1a shows the detector and its upwind footprint which in this example is the area where a release between 5 and 10 g would cause the measured values of dosage (or concentration) at the detector within a factor of two. If the measured concentration at a second detector is below the sensitivity of the sensor, i.e., the threshold detection limit, then its footprint can be used to eliminate certain areas as potential source locations. This is illustrated in Fig. 1b where the so-called null footprint is formed using the detector threshold value. When more than one sensor detects the agent, then the potential source area is defined by the intersection of footprints (see Fig. 1c).

Footprints are computed for different release amounts (e.g., 1.25 to 2.5 g, 2.5 to 5 g, 5 to 10 g, and so on) and intersections of footprints from hit detectors and elimination of areas from null footprints are calculated to determine potential source areas for a range of release mass quantities. If the footprints from the hit detectors do not overlap, then a lower bound on the release amount can be determined. If the null footprints completely overlap the hit footprints, then an upper bound on the release amount can be estimated.

The detector footprints are computed by simply specifying the source location at the detector location, by reversing the wind field, and solving for the source strength field Q by inputting the concentration (or dosage) measured at the collector into the plume dispersion model. Note that one can equivalently release an amount Q at the detector location and then create the footprints from the concentration contours that agree with the detector measurement – this is often easier to implement in an existing dispersion model. One can show that this is possible through use of the Gaussian plume model equations, that is, a release of an amount Q from location (x1, y1) results in a concentration at location (x2, y2) that is equivalent to the concentration at (x1, y1) from a release Q at location (x2, y2) by simply reversing the wind direction (Fig. 2). More information on the collector footprint approach can be found in Brown et al. (2007).

The ratio between the lower and upper bound of each footprint (two in the example shown above) can be interpreted as an estimate of the uncertainty in the measurements or in the plume dispersion model fidelity. For complex building configurations or for cases where the sensors are known to contain large uncertainty, a larger ratio should be used. As compared to many approaches that calculate a single best source location (x,y), the footprint method computes an area where the release might have been possible. The approach enables fast estimation of the area where the airborne contaminant might have come from which could be useful for emergency responders and clean-up efforts.

The approach does have a downside in that it doesn't always find a solution, either owing to uncertainties in the measured concentrations or errors in the dispersion model. The optimization methods mentioned in Section 2 do not have this problem. We have combined an optimization method using simulated annealing with the detector footprint approach to get around this problem. Simulated annealing gives the values of the location and the strength of the source using the footprints by minimizing a cost function based on the difference between modeled and measured concentrations.

3.3. *Simulated annealing*

Simulated annealing (SA) is a stochastic optimization method based on analogy with annealing process in materials (Kirkpatrick et al., 1983). The advantage of SA when compared to some "classical" optimization methods is its success in finding global minimum, thus avoiding ending up in a local minimum.

For source inversion applications, the first step in the SA method is to compute concentration fields for sources at each of the collector sites using reversed winds. The SA routine then randomly picks a source location and strength and new concentration fields are obtained by linear scaling of the already calculated field at each collector site. The cost function is defined as:

$$CF = \sum_{i=1}^{N} \frac{\left(C_{iM} - C_{iC}\right)^2}{C_{iM}^2} \tag{1}$$

Where summation is performed over all N sensors, C_{iM} is measured concentration at sensor i and C_{iC} is concentration at proposed location when the model is run in inverse mode with source at location of sensor i for proposed value of source strength.

The SA algorithm iteratively randomly picks the possible solutions and if the new guess is better i.e. has the lower cost function it is accepted. In a case the new guess is worse then previous value it is not necessarily rejected. It is accepted when the following expression is satisfied:

$$r < \exp\left[-\left(CF_n - CF_{n-1}\right)/T\right] \tag{2}$$

where r is the random number from interval [0,1], CF_n and CF_{n-1} are values of cost functions at *n*th and (*n-1*)th steps and T is the synthetic temperature. This possibility of acceptance of a worse solution gives the SA method the ability to search a wider solution space and avoid getting stuck in a local minimum of a cost function. The cost function is a measure of the goodness of the proposed solution and its definition can have influence on efficiency of the search algorithm. The name "temperature" for parameter T comes from analogy with statistical mechanics and which decrease reduces the probability of worse proposals to be accepted in later stages of calculation.

3.4. *Implementation*

The original dispersion model employed within the BERT source inversion model used a segmented Gaussian plume model, while in this work we employ the QUIC dispersion modeling system which is more appropriate within dense urban areas due to its ability to resolve buildings and estimate complex flow and dispersion patterns occurring within cities. The footprint for each sensor was obtained by running the QUIC-PLUME model after reversing the flow field obtained by QUIC-URB and placing the source at the sensor locations. Only one calculation at each detector location is required. When a footprint for a different source strength is needed, the previously calculated concentration field is scaled accordingly. The footprint for each detector is obtained by going through each grid element in the computed concentration field and checking if the concentration is within the given range. If the footprint is calculated using the factor of 2 range in the release amount, then a given point belongs to the footprint if the concentration at that point is within the factor of 2 of the sensor measurement. The grid element that belongs to all the hit detector's footprints and is not within a null footprint is a potential source location. If none of the grid elements satisfy these conditions then the solution does not exist for the tested source strength and a calculation with a new value is attempted.

4. Results

4.1. *Footprint Approach*

In order to evaluate the ability of the described methods to estimate source parameters, calculations were performed using the QUIC model to create synthetic detector measurements for a mock city consisting of a small group of buildings (see Fig. 3). The first set of runs had no noise introduced, i.e., the resulting concentrations from the forward runs were used as measurement data for the mock sensors. In the second set of runs, Gaussian noise was added to the detector measurements since real sensor measurements contain errors due to inaccuracy or poor performance of the measurement device and because the dispersion model does not perfectly represent real-world dispersion.

Assuming that the source height is the same as the height of the sensors (1m), the concentration field from one horizontal plane is used in our analyses. This simplification is reasonable since it is very probable that in a real event the release could be at street level either due to a traffic accident or an intentional release. Also, it is important to mention that we assume existence of only one source.

In the evaluation case presented here, the release is continuous with a source strength of 100 g/s and location (xs, ys, zs)=(29m, 15m, 1m). Figure 4 shows the simulated plume and the "measurements" computed at each detector. Since the two left-most detectors measure very low concentrations, we first obtained results using the other four sensors. Figure 5 shows the upwind footprints for the four detectors, which depict the regions where a release of 85-170 g would result in a detector measurement to within a factor of two. The intersection of the four footprints is where a single release could have occurred that would have resulted in the measurements at all four detectors. Figure 6 shows the number of footprints that intersect at each given location (grid element) in the domain for a source strength in the range of 85-170 g/s. The two dark red regions indicate the most probable locations of the source and one of them is exactly at the source location. Also, the correct value of source strength 100 g/s is within the range of source strength that would cause given concentration measurements.

When we used the footprints from all six sensors we were not able to obtain results using the factor of 2 as the level of uncertainty, i.e., there was no grid element that belonged to all six footprints. Using a factor of 3, we obtained the results presented in Fig. 7. The location is two gird elements (4m) from the correct solution and the exact source strength was not within the estimated source strength interval, but very close. The corresponding footprints used to obtain the solution in this case are given in Fig. 8. The likely reason for the difficulty with finding a solution for this six detector case is due to the stochastic uncertainty in the very low concentrations at two of the detectors. The random-walk model produces mean concentrations in the "tails" that are not statistically converged. This behavior is similar to real sensors which often have less accuracy at very low concentrations. This suggests that source inversion models should consider an option to throw-out measurements near detector threshold.

Performance of the footprint approach was tested with introduction of noise to the synthetic data. The introduced noise had a Gaussian distribution with zero mean and standard deviation of 10%, 20% and 50% of the "measured" concentration values at each given detector. For the 10% noise case, when we use 4 sensors and a factor of 2 as the level of uncertainty the results are almost identical to the case without noise. In the case of 20% noise, the estimated location of the source is missed by one grid element (2m) in many cases. Also, in a small number of cases the model was not able to give solutions. Introduction of 50% noise caused inability of model to find solution in most of the cases indicating that at least one of the detectors had a greater than two difference from the original model-produced measurement value. When we used a factor of 3 to obtain footprints for all six detectors for 10% and

20% noise, the solution is obtained in most of the runs and results are identical to the case without noise, i.e., the source location and range of possible source strengths is the same. For largest noise (50%) the model was not able to find the solution in most of the cases even when the factor of 3 uncertainty is used.

4.2. *Footprint Approach with Simulated Annealing*

Figure 9 represents the source location at every thousandth step as the simulated annealing method iteratively searches for the optimal solution for the case of 6 detectors. The solution obtained using the simulated annealing optimization algorithm for the no-noise case was: location (28.97, 11.74), source strength 64.21 g/s. The solution obtained by SA is shown in Fig. 7 and is denoted by the white triangle. Thus, the location was estimated pretty well, while the source strength was within a factor of 2. It is important to point out here that since this is the stochastic method each run of SA algorithm will give slightly different though very close results (for example location (29.35, 10.82) and source strength 64.10 g/s).

In the case of simulated annealing calculations, noise with standard deviation of 10% and 20% does not cause significant changes in the results, while 50% standard deviation gives much larger span of solutions for different runs especially with respect to source strength which in a number of runs changed between 30 and 70 g/s.

5. Conclusions

In this work we presented two approaches for fast estimation of the source location and strength in cities using detector measurements. Both approaches, detector footprint and detector footprint with simulated annealing, gave similar and satisfactory results for a synthetically-created data set. The location was estimated very precisely with and without introduced Gaussian noise on top of the detector measurements (for standard deviations of noise distribution 10% and 20% of measured values). Larger noise (50%) introduced more uncertainty and in many cases the collector footprint approach was not able to find a solution unless the uncertainty factor was increased. Further work will have the goal of testing the model using real data measured during different field campaigns within urban areas and implementing the null footprint scheme.

Acknowledgments

The authors would like to thank Dr. Michael W. Williams, Dr. Matthew A. Nelson and Akshay A. Gowardhan for useful comments and discussion.

References

Allen, C. T., and S. E. Haupt, 2007: Source characterization with a genetic algorithm-coupled dispersion-backward model incorporating SCIPUFF. *J. Appl. Meteor*. **41**, 465–479.

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.

Brown M. J., M. D. Williams, G. E. Streit, M. Nelson and S. Linger, 2007: An operational Event Reconstruction Tool: Source Inversion for Biological Agent Detectors. *7th Symposium on the Urban Environment, American Meteorological Society*, *San Diego, CA*.

Chow, F. K., B. Kosovic, and S. T. Chan, 2006: Source inversion for contaminant plume dispersion in urban environments using building-resolving simulations. *6th Symposium on the Urban Environment, American Meteorological Society*.

Gowardhan, A., M. Brown, M. Nelson, E. Pardyjak, D. DeCroix, D., 2006a: Evaluation of the QUIC Pressure Solver using wind-tunnel data from single and multi-building experiments. 6th AMS Symp. Urban Env., Atlanta, GA, LA-UR-05-9016, 12 pp.

Gowardhan, A., M. Brown, M. Williams, E. Pardyjak, 2006b: Evaluation of the QUIC Urban Dispersion Model using the Salt Lake City URBAN 2000 Tracer Experiment Data- IOP 10. 6th AMS Symp. Urban Env., Atlanta, GA, LA-UR-05-9017, 13 pp.

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

Kirkpatrick S., C. D. Gelatt, Jr., M. P. Vecchi, 1983: Optimization by Simulated Annealing, Science, Volume 220, Number 4598.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953: J. Chem. Phys., **21**, 1087-1092.

Linger, S., M. Brown, W.B. Daniel, A. Ivey, T. McPherson, M. Williams, and K. Omberg, 2005: Sensing threats with BioWatch, GeoIntelligence, Mar/Apr 2005, pp. 16-21, LA-UR-05-0799.

Neumann, S., L. Glascoe, B. Kosovic, K. Dyer, W. Hanley, and J. Nitao, 2006: Event reconstruction for atmospheric releases employing urban puff model UDM with stochastic inversion methodology. *6th Symposium on the Urban Environment, American Meteorological Society*.

Pardyjak, E.R. and M.J. Brown, 2002: Fast response modeling of a two building urban street canyon. *Proceedings of the 4th Symposium on the Urban Environment*, May 20-24, 2002, *American Meteorological Society*. Norfolk, VA.

Pardyjak, E.R. and M. Brown, 2003: QUIC-URB v. 1.1: Theory and User's Guide, LA-UR-07-3181, 22 pp.

Pol S., N. Bagal, B. Singh, M. Brown, and E. Pardyjak, 2006: Implementation of a rooftop recirculation parameterization into the QUIC fast response wind model, 6th AMS Urb. Env. Symp., Atlanta, GA, LA-UR-05-8631, 19 pp.

Röckle, R. (1990), "Bestimmung der Stomungsverhaltnisse im Bereich komplexer Bebauungsstrukturen" *Ph.D. thesis, Vom Fachbereich Mechanik, der Technischen Hochschule Darmstadt, Germany.*

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

Senocak I., N. W. Hengartner, M. B. Short, and B. W. Daniel 2007: Stochastic event reconstruction of atmospheric contaminant dispersion. *7th Symposium on the Urban Environment, American Meteorological Society*, *San Diego, CA*.

Yee E. 2007: Bayesian probabilistic approach for inverse source determination from limited and noisy chemical or biological sensor concentration measurements. Chemical and Biological Sensing VIII, edited by Augustus W. Fountain III. Proc. Of SPIE Vol. 6554, 65540W.

Williams, M., M. Brown, B. Singh, & D. Boswell, 2004: QUIC-PLUME Theory Guide, LA-UR-04-0561, 22 pp.

Williams, M., M. Brown, and A. Gowardhan, 2005: Adaptation of the QUIC-PLUME model for heavy gas dispersion around buildings, AMS ASAAQ San Francisco, LA-UR-05-2255, 9 pp.

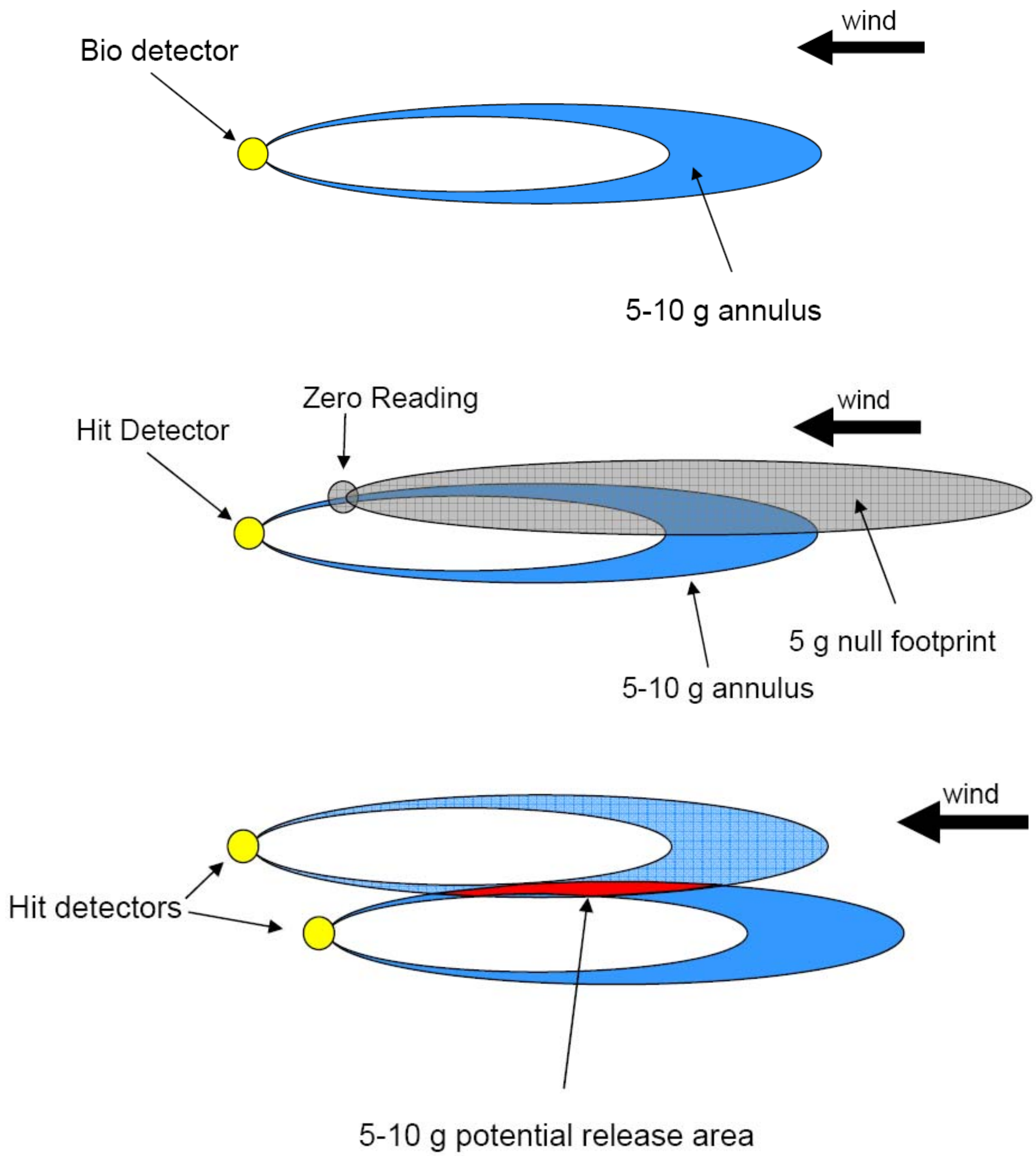

Figure 1. Illustration of the collector footprint methodology: (a) detector and its footprint showing where a release of 5 to 10 g would result in the measured dosage to within a factor of 2, (b) detectors that read zero dosage can be used to create null footprints and thus eliminate certain areas as possible source locations, (c) overlap region is the area from which a release of 5 to 10 g would cause both detectors to measure respective dosages or concentrations within a factor of 2 (from Brown et. al. 2007).

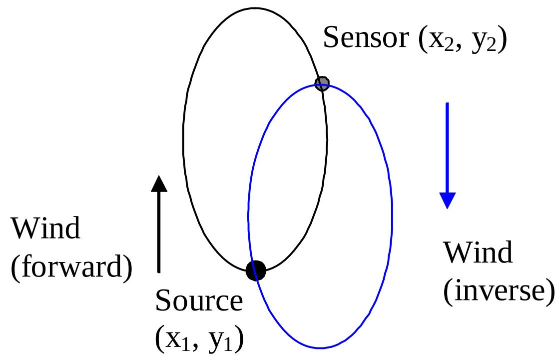

Figure 2. During the 'forward run' source is at (x_1, y_1) and sensor measures C_{forw} (x_2, y_2). During the 'inverse run' the wind field is inverted and the source is at (x_2, y_2). For the same source strength during the inverse run, the concentration $C_{inv}(x_1, y_1)$ is equal to $C_{forw}(x_2, y_2)$, i.e. $C_{inv}(x_1, y_1) = C_{forw}(x_2, y_2)$.

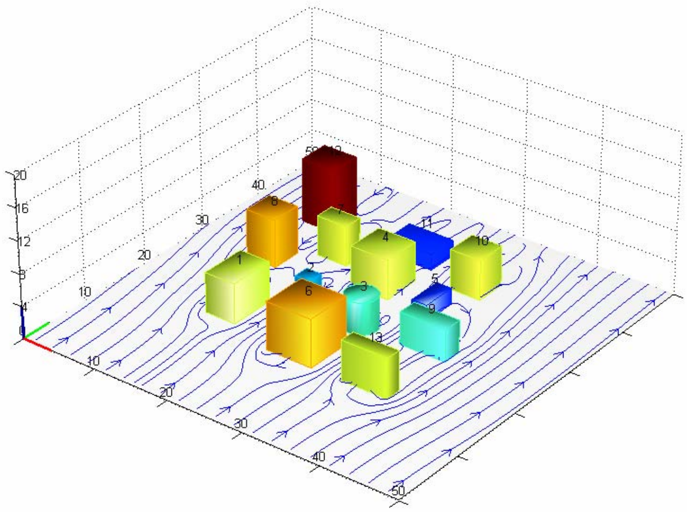

Figure 3. The mock city consisting of 13 buildings used to illustrate the inversion methodology. Streamlines from the QUIC-URB model are overlaid.

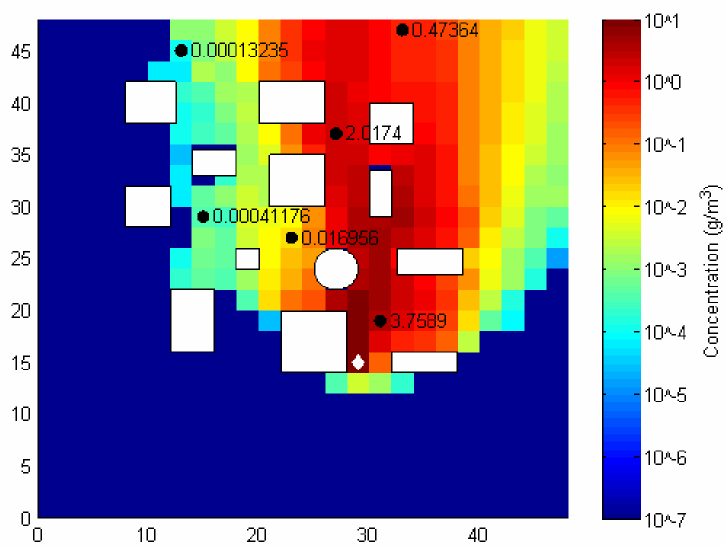

Figure 4. Plume obtained using the QUIC model for continuous source from location (xs, ys, zs)=(29m, 15m, 1m) and source strength Qs=100 g/s. Diamond represents the location of the source while black dots indicate sensor locations with values of “measured” concentrations. These “measurements” will be used to test the source inversion model.

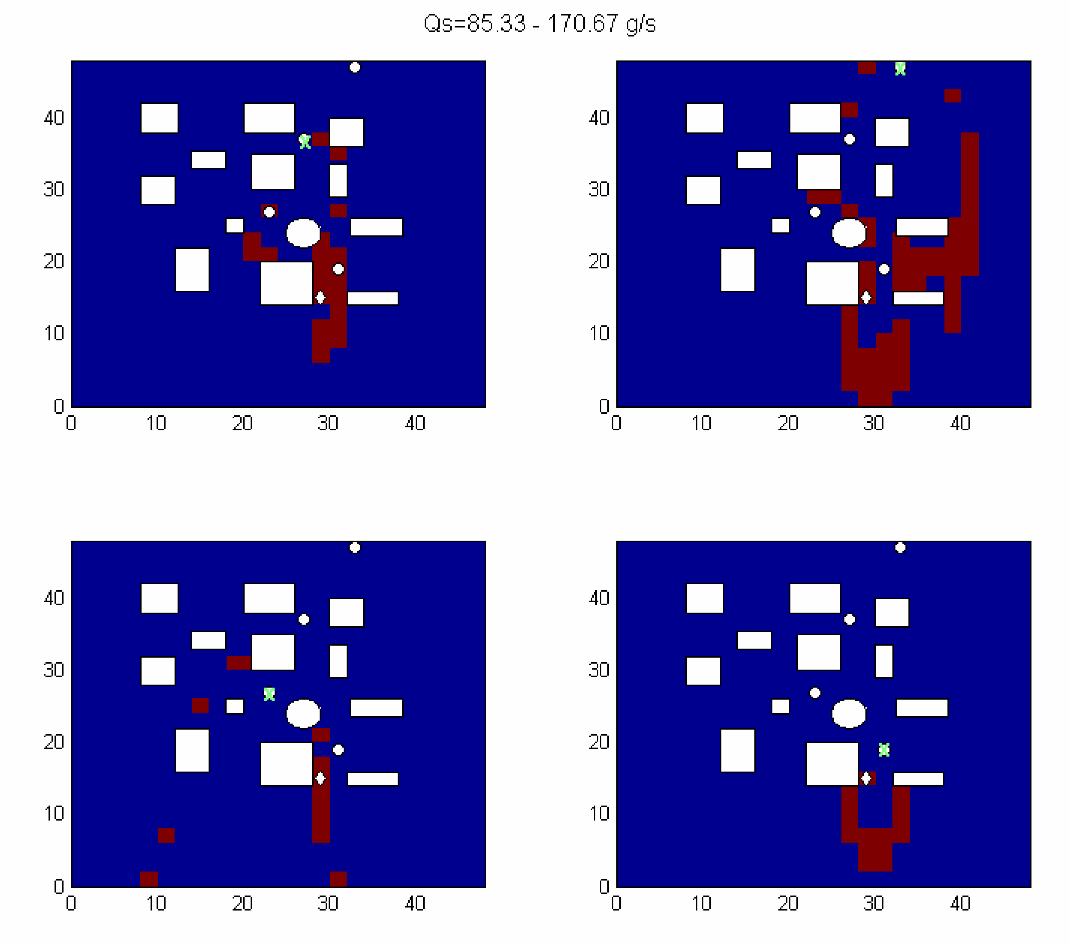

Figure 5. The upwind footprints (red) of the four sensors used to obtain the potential source locations. The green symbol X indicates the sensor used to create the corresponding footprint. The intersection of the four footprints represents where a release could potentially occur.

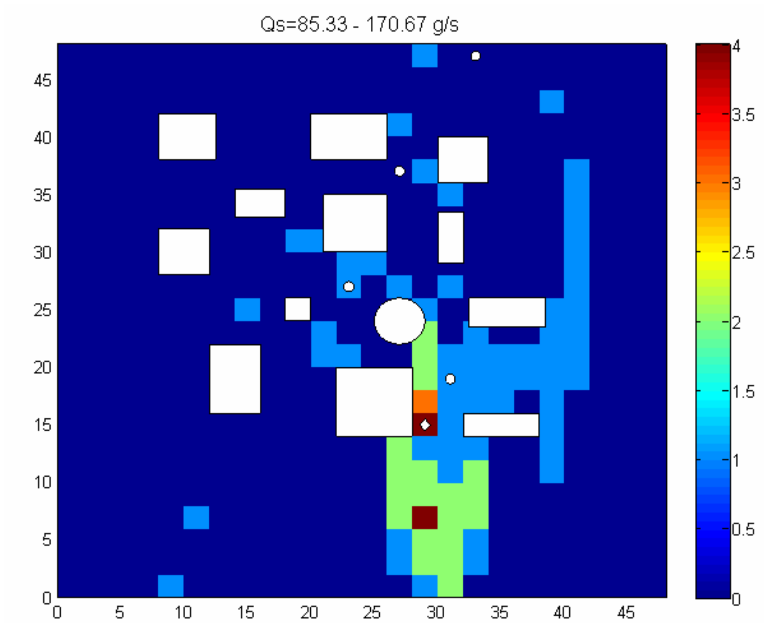

Figure 6. The computational domain where color of the grid element indicates how many footprints intersect at each given point (grid element) for release rates within the range 85-170 g/s. Footprints were obtained for 4 sensors using a factor of 2 uncertainty. Circles denote sensor locations, the diamond is the source location and the dark red grid cells represent where all collector footprints intersect, i.e., potential release locations.

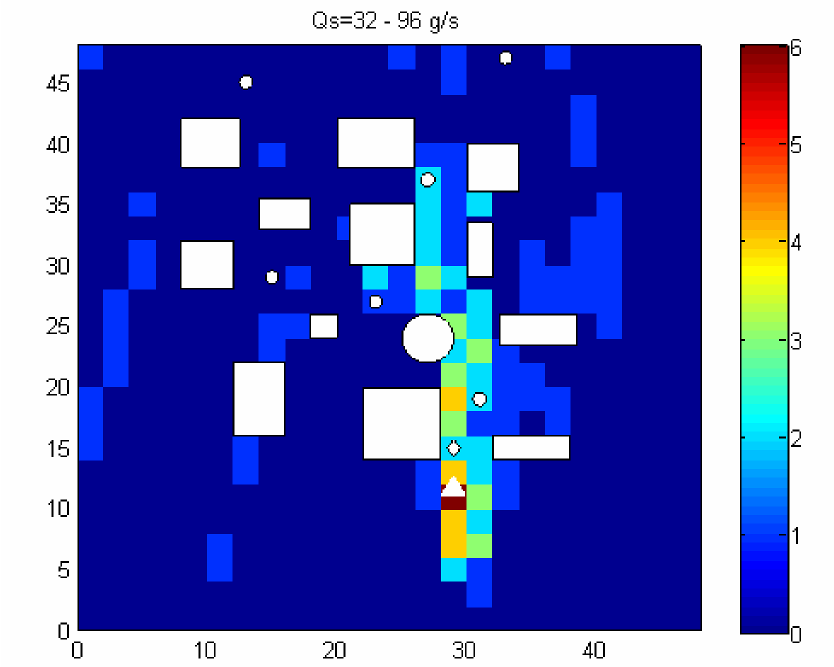

Figure 7. The intersection of all six footprints for release rates within range 32-96 g/s. Circles denote sensor locations, the diamond is the source location and dark red grid cells represent where all six collector footprints intersect, i.e., the potential source locations. These results were obtained using a factor of 3 uncertainty, as a factor of 2 did not give solutions. The triangle symbol indicates the estimated source location using simulated annealing.

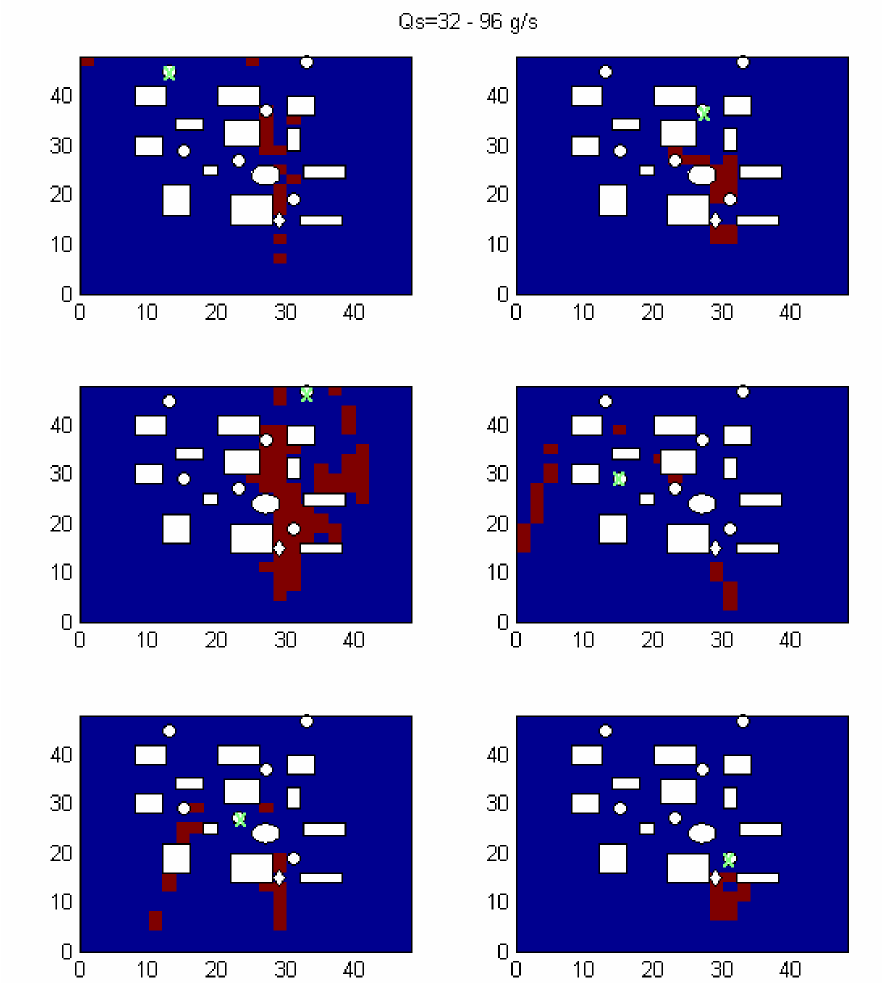

Figure 8. The footprints (using a factor of 3 uncertainty) for all six sensors for the solution given in Fig. 7. The green symbol X indicates the sensor used to create the corresponding footprint.

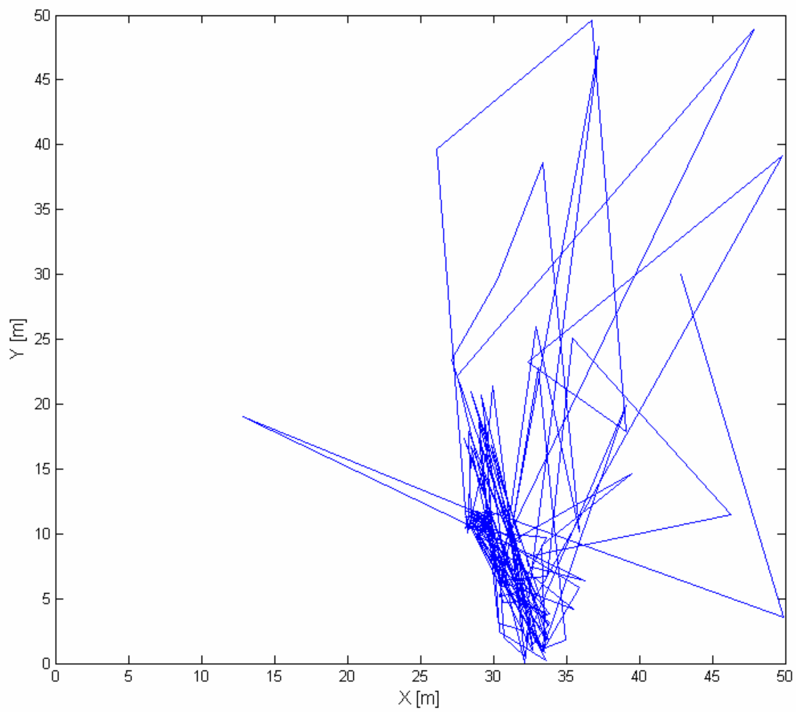

Figure 9. The simulated annealing search for the source location (every thousandth step).