

SOURCE TERM ESTIMATION USING A GENETIC ALGORITHM AND SCIPUFF

Kerrie J. Long*, Sue Ellen Haupt, George S. Young, Luna M. Rodriguez and McKenzie McNeal III
Applied Research Lab, The Pennsylvania State University, State College, Pennsylvania

1. INTRODUCTION

The unintentional or intentional release of a harmful contaminant is a serious threat to civilians and to our nation's military forces. Responding quickly and with the best information possible is critical to mitigating the threat and minimizing the impact on population and property. Characterizing the source accurately is vital to driving atmospheric transport and dispersion (AT&D) models and thus predicting the future state of the puff. Unfortunately, the meteorological and chemical data is likely to be spatially and temporally sparse so efficient use of the data is critical to determining an accurate solution.

Forecasts are extremely sensitive to initial conditions and even slight differences in initial conditions can lead to drastically different realizations such as those depicted by the two contoured concentration fields superimposed on the array of sensors in Fig. 1. Inaccurate initial conditions can lead to poor forecast quality and thus greatly hamper mitigation and evacuation procedures. But suppose we have sensors in the field, as illustrated by the black circles in Fig. 1. Can we use the time dependent concentration information made available by these sensors to determine the characteristics of the source? We demonstrate that a genetic algorithm back-calculation model coupled with SCIPUFF is successful

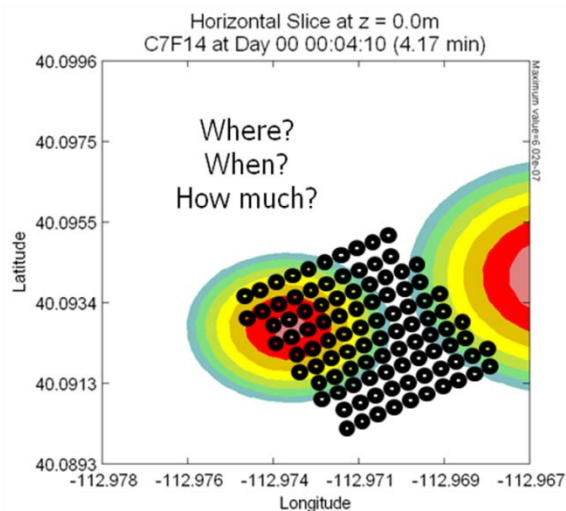


Figure 1: The source term estimation problem. Colored contoured concentration fields suggest different realizations resulting from varied initial conditions. A sensor network, indicated by black circles, reports concentration amounts.

*Corresponding author address: Kerrie J. Long, Applied Research Lab, The Pennsylvania State University, P.O. Box 30, State College, PA 16804-0030; kjl203@psu.edu

at identifying the basic source term information. Section 2 describes the GA back-calculation model and SCIPUFF. The observation data is also described in that section. The results are described in section 3. Section 4 summarizes the work and discusses future prospects.

2. EXPERIMENTAL DESIGN

Our approach to the source term estimation problem uses a genetic algorithm variational technique, GA-Var. GA-Var avoids the backward integration step of traditional four-dimensional variational techniques by directly optimizing the unknown variables using forward integration and solution evolution. In this study we couple the GA with the Second-order Closure Integrated Puff (SCIPUFF) model to back-calculate several parameters describing a contaminant release. Specifically, a set of trial solutions, each representing a possible source term, is randomly initialized. The GA evolves the population of potential solutions through mating and mutation operators and for each new trial solution a new forecast is created via SCIPUFF. This resulting forecast concentration field is compared to the observed concentration field via a cost function. The observed concentration data is obtained from NCAR's Eulerian/semi-Lagrangian (EULAG) numerical model.

2.1 Observation Data

The observation data for this study is generated from NCAR's Eulerian/semi-Lagrangian (EULAG) numerical model and is meant to simulate the FFT07 field experiment. The EULAG model uses a Large Eddy Simulation (LES) approach to solve the partial differential equation governing the flow. It is coupled with a global climatology analysis tool (GCAT).

The source (indicated by the yellow burst in Fig. 2) is located at 40.0971N, -112.9755E. The release is an instantaneous release (a single puff) in a convective boundary layer with wind generally out of the North Northwest. A network of 100 sensors (indicated by the black circles in Fig. 2) is located just south and east of the source. The sensors are approximately 100 m apart. We use the concentration value from a height of 10 m above the ground. The observation data are collected every 10 seconds for 300 seconds.

2.2 SCIPUFF

SCIPUFF is a sophisticated Gaussian puff based model that splits and merges the puffs according to prescribed conditions (Sykes et al. 2004). The inputs for SCIPUFF were derived from the CFD wind field data. The most representative wind direction and speed was obtained from the observation data and is implemented as a single fixed wind field with a uniform wind direction.

The release was set up in SCIPUFF as a daytime release beginning at noon local time. The boundary layer is modeled as simple diurnal and large scale variability is set to none. The release is modeled as continuous with 0.74 kg s^{-1} released for 10 s. A generic gas, C7F14, with the same density as air is chosen to mimic the material used in the LES release. The material is released from the surface and the 100 sensors retrieve concentration in the field from a height of 10 m (indicated in Fig. 2). The release is run for 300 s with data output every 10 seconds. The maximum time step for internal SCIPUFF calculations is set to 10 seconds. Note that domain for the GA is much smaller than the domain for SCIPUFF so that even if the source is placed on the extreme most point in the GA domain, it will still have ample room to grow and disperse in the 5 minute simulation before the puff moves off the SCIPUFF domain. The GA and SCIPUFF domains are described in Table 1.

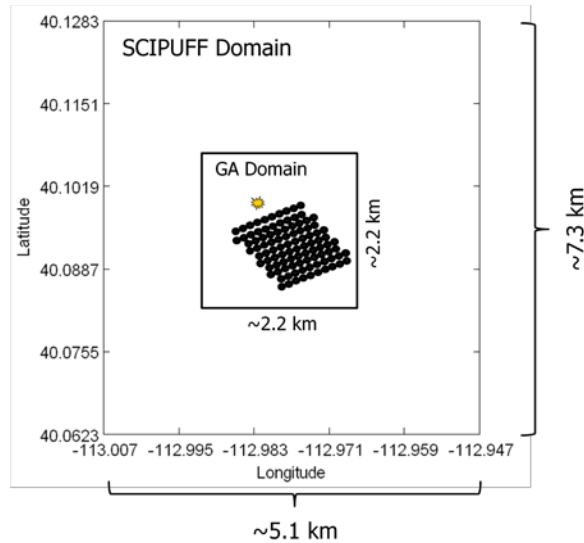


Figure 2: The experimental setup with the release indicated by the yellow burst and the sensor network indicated by the black circles. The trial solution can be located anywhere within the GA domain and the puff can advect and disperse anywhere within the larger SCIPUFF domain.

Table 1: GA & SCIPUFF Domains

		SW Corner	NE Corner
SCIPUFF	Latitude	40.0623°N	40.1283°N
	Longitude	-113.0070°E	-112.9470°E
GA	Latitude	40.0871°N	40.1072°N
	Longitude	-112.9901°E	-112.9639°E

2.3 Genetic Algorithm

The genetic algorithm is an artificial intelligence optimization technique inspired by the biological processes of genetic recombination and evolution. It begins with a population of trial solutions or chromosomes and evolves them closer to the true solution through mating and mutation operators as

illustrated in Figure 3. The GA used in this study is described in more detail in Haupt and Haupt (2004).

The GA is initialized with a random population of trial solutions. In this example, the trial solutions are a vector of source locations. These potential source locations are then given to SCIPUFF and a set of resulting concentrations fields are calculated.

The fitness of each member of the population is determined by comparing the forecast field to the observed concentration field by evaluating the cost function. The cost function is simply the difference between the observations and the forecast.

$$cost\ function = \frac{\sum_{t=1}^{30} \sqrt{\sum_{r=1}^{TR} (\log_{10}(C_r + \epsilon) - \log_{10}(R_r + \epsilon))^2}}{\sum_{t=1}^{30} \sqrt{\sum_{r=1}^{TR} (\log_{10}(R_r + \epsilon))^2}} \quad (1)$$

where C_r is the concentration as predicted by SCIPUFF, R_r is the receptor data value at receptor r , TR is the total number of receptors (in this case 100), a and ϵ are constants, and the cost function is summed over all thirty time steps. To avoid taking the logarithm of zero, ϵ is added to C_r and R_r quantities. The value of ϵ is 10 orders of magnitude smaller than the largest concentrations of C_r and R_r . If ϵ approaches one, then it will dwarf the concentration values, C_r and R_r , thus rendering the cost function meaningless. Cost function values closer to zero are desirable.

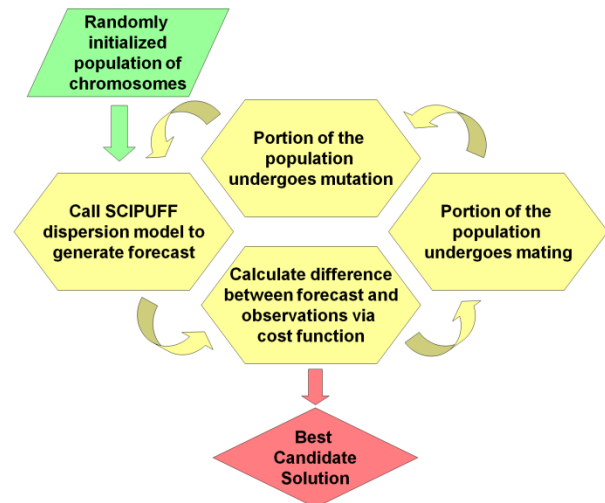


Figure 3: Schematic of the GA process.

After the cost function is evaluated, the population is then resorted such that solutions with lower cost function values are more likely to participate in natural selection. Next, a portion of the population is modified through the mating operation where two parent chromosomes are recombined to produce two new children chromosomes. Then a portion of the population is modified through the mutation operation where a random mutation is introduced into a percentage of the solutions. This cycle continues and the population is evolved until convergence is reached.

3. RESULTS

Initially, a sensitivity study was conducted to determine the optimal parameters for the GA, which were found to be a population size of 64, a mutation rate of 0.05 and 10 iterations. Results were not sensitive to an increase in the number of iterations, an increase in population size, or varying mutation rates, suggesting that the solution converges under this configuration.

The back-calculation algorithm is run for several different wind speeds and directions to illustrate the high sensitivity to those meteorological parameters. Fig. 4 illustrates the differences in retrieval with only a 1 m s^{-1} difference in wind speed. Note that for plotting purposes a wind direction of 10° is labeled 370° and so on. For a wind direction of 350° , the error in source location is reduced from 174 m to 45 m by a change in wind speed of only 1 m s^{-1} . This study illustrated the importance of knowing the wind data correctly.

If the results are further decomposed into the x and y components as illustrated in Fig. 5, we see that the y source location, in this case the latitude, is more accurately retrieved by the GA. In particular notice the extreme sensitivity to the error in the x source location. In some instances, the reduction the MAE is reduced by nearly an order of magnitude. Notice that the best retrieval for the error in the y location is at 325° but for the x error the best retrieval occurs at 360° . Note that the best location magnitude error in Fig. 4 more closely agrees with 360° .

4. SUMMARY & DISCUSSION

With SCIPUFF as our forward dispersion model we are able to use the GA to back-calculate several source parameters including the location of the release in the x- and y-directions. Now we have the source information necessary to create a forecast. The CFD data provides a critical bridge to understanding the details of the flow when moving from an identical twin synthetic data experiment to a field experiment.

As pointed out in Krysta et al. (2006) and Allen et al. (2007) back-calculating the most representative wind speed and direction is critical to the success of retrieval in the model. Future directions include extending the model to characterize more complicated release scenarios involving multiple releases and more complex meteorology. We also plan on expanding the model to retrieve meteorological variables such as wind direction and speed.

ACKNOWLEDGEMENTS

This work was supported by DTRA under grant number W911NF-06-C-0162 and DTRA01-03-D-0010-0012. The authors wish to thank NCAR/RAL for allowing use of their data generated from NCARs GCAT/EULAG model and for their assistance in interpreting the data.

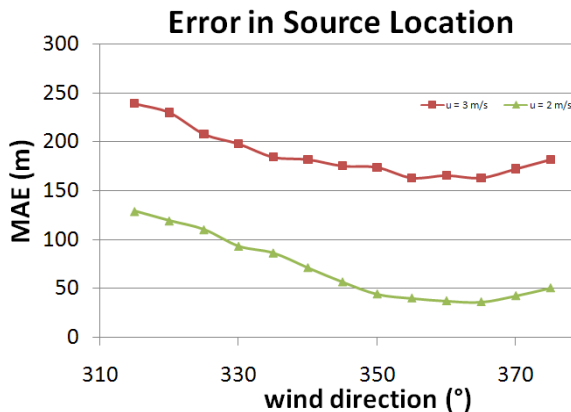


Figure 4: Mean absolute error (MAE) in source location as a function of wind direction and wind speed.

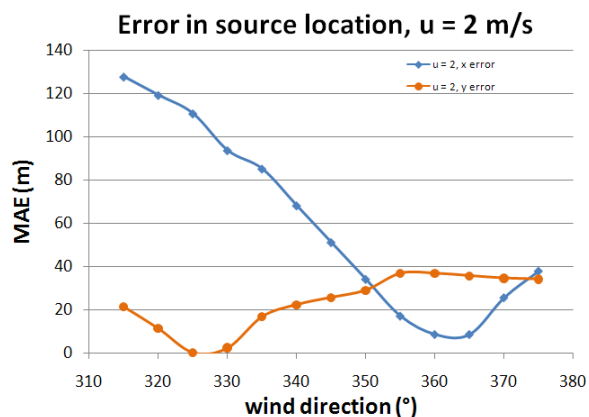


Figure 5: Mean absolute error (MAE) in x and y source location as a function of wind direction and wind speed.

REFERENCES

- Allen C.T., Young G.S., Haupt S.E., 2007: Improving pollutant source characterization by optimizing meteorological data with a genetic algorithm. *Atmos. Environ.* 41: 2283-2289
- Haupt, R.L. and S.E. Haupt, 2004: *Practical Genetic Algorithms*. John Wiley & Sons, Inc. Hoboken, NJ, 2 ed. 253 pp.
- Krysta M., Bocquet M., Sportisse B., Isnard O., 2006: Data assimilation for short-range dispersion of radionuclides: an application to wind tunnel data. *Atmos. Environ.* 40: 7267-7279.
- Sykes R.I., et al., 2004: SCIPUFF Version 2.0, Technical Documentation; A.R.A.P. Report no. 727, Titan Corp. Princeton, NJ.