Source Characterization and Meteorology Retrieval Including Atmospheric Boundary Layer Depth Using a Genetic Algorithm

Andrew J. Annunzio*, Sue E Haupt, and George S. Young
The Pennsylvania State University, Department of Meteorology, University Park, Pennsylvania

1. Introduction

An important issue in homeland and military security is characterizing the source of a harmful contaminant in the atmosphere. Characterizing the source allows appropriate response to and mitigation of the contaminant. Further, finding the appropriate initial conditions permits proper determination of where it will disperse. Characterizing the source provides initial conditions for the location and strength of the release, but it does not give enough insight to model the spread of the contaminant. Ultimately, meteorological parameters such as wind speed and wind direction govern the dispersion of contaminants allowing one to compute surface concentrations. The depth of the boundary layer is another meteorological parameter that impacts surface concentration by reflecting contaminants back to the surface. The capping inversion at the top of the boundary layer acts as a lid to rising thermals that result from surface heating and also limits the domain of turbulence (Stull, 1998). The impact of reflected contaminants on surface concentration depends on the depth of the boundary layer and atmospheric stability. A parameter that helps assess the influence that a capping inversion has on the dispersion of a contaminant is the ventilation factor given by the product of wind speed and boundary layer depth (Hsu, 2003). A shallow boundary layer coupled with calm winds implies stagnant air and hence low ventilation. Deep boundary layers with fast winds, on the other hand entail rigorous motions and high values of ventilation (Eagleman, 1996). While ventilation may describe how efficiently the atmosphere disperses contaminants, it does not provide insight into how the boundary layer depth impacts surface concentrations. Further, before use of the ventilation factor, the height of the capping inversion and wind speed must be determined, which are unknowns for this problem.

This work uses the concept of assimilating surface concentration measurements to back calculate the source characteristics of the release as well as meteorological parameters. This is an extension of prior work where wind speed and direction were back calculated along with the (x,y) source location, source strength, and time of release (Allen 2006, Long 2007).

*Corresponding author address: Andrew J. Annunzio, Department of Meteorology, The Pennsylvania State University, University Park, PA 16802; aja199@psu.edu

This current work is unique, because back calculation includes boundary layer depth that impacts surface concentration further from the source.

2. Model Formulation

This goal of this study is to determine the source of a contaminant and meteorological parameters, including wind speed, wind direction, and boundary layer depth, in order to accurately model the transport and dispersion of a contaminant. The data inversion is accomplished with an optimization technique. Such a calculation requires a robust optimization technique; therefore we implement a hybrid Genetic Algorithm (GA) for the task. Prior to any back calculation, a forward model computes concentration data at grid points in the domain. The GA optimizes the number of parameters in question in the forward model to best match the concentration data at grid points.

2.1 Forward model

For this problem we use an identical twin experiment where the model itself creates the observations (Daley, 1991). This approach is useful for technique development because it eliminates model errors that are a source of uncertainty. The model chosen to create synthetic data is the three dimensional trapped Gaussian puff model given by

$$
\begin{aligned}
C = {} & \frac{q\Delta t}{u\sigma_x\sigma_y\sigma_z(2\pi)^{\frac{3}{2}}}\left(\exp\left(-\frac{(x-ut)^2}{2\sigma_x^2}\right)\times\exp\left(-\frac{y^2}{2\sigma_y^2}\right)\right. \\
& \times\left(\exp-\frac{(z_{sensor}-z_{release})^2}{2\sigma_z^2}\right)+\exp\left(\frac{(z_{sensor}+z_{release})^2}{2\sigma_z^2}\right) \qquad (1) \\
& +\sum_{n=1}^{N}\left(\exp\left(\frac{(z_{sensor}+z_{release}+2nz_i)^2}{2\sigma_z^2}\right)+\exp\left(\frac{(z_{sensor}+z_{release}-2nz_i)^2}{2\sigma_z^2}\right)\right. \\
& \left.\left.+\exp\left(\frac{(z_{sensor}-z_{release}+2nz_i)^2}{2\sigma_z^2}\right)+\exp\left(\frac{(z_{sensor}-z_{release}-2nz_i)^2}{2\sigma_x^2}\right)\right)\right)
\end{aligned}
$$

Where σ_x, σ_y, and σ_z represent the standard deviations of the puff in each of the three dimensions, z_i is the boundary layer depth, z_{sensor} is the height of the sensor, $z_{release}$ is the height of the release, q is the release rate over a time span Δt, u is the wind speed, and N is the number of reflection terms included. Experiments show that two or three reflection terms describe almost all of the dispersion, however, we will use five reflection terms in creating the synthetic data (Beychock, 2005). The forward model computes the concentration at the sensors given "guesses" of the seven parameters to be optimized. The hybrid GA optimizes the parameters by minimizing the squared difference between the concentrations produced by the synthetic data and the concentrations calculated with the current guess to the variables summed over each grid point and each time step.

2.2 Domain Considerations

In the Gaussian Puff model, at a distance near

$$\frac{\sigma_z}{z_i} = 1.2 \quad (2)$$

vertical dispersion becomes uniform (Beychok, 2005). In (1) and (2) the vertical dispersion coefficient σ_z describes the vertical spread of the puff which is given by

$$\sigma_z = \exp(I + J\ln(x) + K(\ln(x))^2) \quad (3)$$

The constants I, J, and K in (3) are functions of Pasquill stability classification. In general a more stable atmosphere produces dispersion coefficients that limit the spread of the puff (Beychok, 2005). Rigorous vertical motions accompany unstable conditions causing the puff to spread very quickly: the dispersion coefficients are larger in this case. Uniform concentration close to the source is the result for the most unstable case, which poses a problem for the back calculation because many sensors possess the same value. Sensors reporting the same reading make the back calculation more difficult since the GA has less distinguishing information to work with. To combat this problem, we can adjust the size of the domain in the forward model depending on the boundary layer depth and the atmospheric stability. Initially, the domain size is small so that contaminant concentration does not become uniform. Then we will expand our domain so that there is a large portion of the domain where dispersion is uniform. This situation only occurs in the unstable cases (Stability classes A and B). In the nearly unstable and neutral cases (Stability classes C and D), we need to enlarge instead of shorten the domain so that back calculation is possible. In order to determine domain size we use the Turner approximation

$$z_i = z_{release} + 2.15\sigma_z \quad (4)$$

to compute the horizontal distance at which the reflection terms start to play a role or the distance at which the upper portion of the plume is no longer Gaussian (Beychok, 2005). In order to determine where contaminants reflecting off the base of the inversion impact surface concentrations, we extend Turner's equation to

$$2z_i - z_{release} = 2.15\sigma_z \quad (5)$$

If we plug (3) into (5) and solve for *x*, we obtain the equation

$$x = \exp\left(-\frac{J}{2K} \pm \left(\left(\frac{I}{K}\ln\left(\frac{2z_i - z_{release}}{2.15}\right) - \frac{I}{K}\right) + \left(\frac{J}{2K}\right)^2\right)^{\frac{1}{2}}\right) \quad (6)$$

Then, the distance x becomes the radius of our domain if we only want a minimal amount of reflected contaminants in surface concentration. For stability classes C and D this distance can be large, and therefore the distance *x* is taken as the domain radius. For stability classes A and B, *x* is small, and making the domain larger necessitates adjusting (5) to

$$2nz_i - z_{release} = 2.15\sigma_z \quad (7)$$

Although our domain is now larger, the risk of having uniform concentration at the surface is now greater. To ascertain this distance, we plug (2) into (7) and divide by z_i to obtain the *n* that determines the distance where concentration is uniform. From this relation, one can also predict how many of the sensors will report equal concentration values.

2.3 The Hybrid Genetic Algorithm

A genetic algorithm is a global optimization technique that is quite robust (Haupt, 2004). The crossover scheme used here blends the variables instead of using the more standard single point crossover. The cost function employed is logarithmic and is given by

$$\frac{\sum_{t=1}^{5}\left(\sum_{r=1}^{R}\log_{10}(aC_r + \varepsilon) - \log_{10}(aR_r + \varepsilon))^2\right)^{\frac{1}{2}}}{\sum_{t=1}^{5}\left(\sum_{r=1}^{R}(\log_{10}(aR_r + \varepsilon))^2\right)^{\frac{1}{2}}} \quad (8)$$

where C_r is the concentration given by placing current guesses to the variables into (1) and R_r is the receptor concentration from the identical twin experiment. The constant ε is placed in the logarithm to avoid taking the logarithm of zero. After the GA finds the region of the global minimum, the Nelder Meade Downhill Simplex Algorithm cascades down the global minimum to the solution (Allen, 2006, Long, 2007).

2.4 Testing the model

In the identical twin experiment the "truth" data is computed with the (x,y) source set at (0.0 ,0.0) m, the source strength set at 1.0 kg/m^3, while the source height, $z_{release}$, is allowed two different values in different runs: 1 m and 50 m. Although a source height is possible above 1000 m, the GA only searches for a source height between 0 and 1000 m since sources above this range are unlikely to require an immediate response. For the meteorological parameters, wind speed is 5 m/s (which is changed with stability) and the wind direction is 180 degrees. The boundary layer depth is given several different test values to allow for a wide range of possibilities and scenarios. The range of the boundary layer depth is between zero and five km, because any boundary layer lid above five km is unrealistic for the homogeneous and stationary turbulence assumption that the Gaussian puff model invokes. We also vary stability by changing the Pasquill stability class between A, the very unstable case, and D the neutral case. It is unnecessary to test stable, classes E and F, because the vertical plume dispersion is small enough that it does not reach the top of the boundary layer.

There are several objectives for our numerical experiments. First, we ascertain how many reflection terms are needed in order for the GA to converge to a solution. In future work, for stability classes A and B, we will vary the size of the domain to consider whether back calculation is possible when contaminant concentration is uniform for a large portion of the domain. Lastly, we will determine a threshold for noise that the model can withstand before back calculation for all parameters becomes unachievable.

3. Results

3.1 Four Variable Sensitivity test

The sensitivity test yields the number of reflection terms necessary to back calculate wind direction, source strength, boundary layer depth, and source height for the four stability classes studied here. We compute skill scores to determine whether the inversion is successful or not; a skill score greater than 0.1 implies a less successful retrieval and a skill score less than 0.1 implies a quite successful retrieval. Figure 1 shows results for a 1 m source height and for all stability classes. The sensitivity test for stability class A presents some interesting results. For shallower boundary layer depths, the back calculation is successful for any number of reflection terms. As the depth of the boundary layer increases, more reflection terms are necessary to accurately ascertain all four parameters. For stability class D, the back calculation is successful for all boundary layer depths considered when two or more reflections terms are considered. The back calculation for Stability class B yielded skill scores greater than .1 when two and three reflection terms are considered. The unsuccessful result is attributed to the model finding a source height of 0 m instead of 1 m. When a 50 m release is considered results improve. Figure 2 illustrates the results for stability classes A, B, C, and D for a 50 m source height. Back calculation is successful for all lid heights no matter the number of reflection terms, however, results are better when more reflection terms are included. For the 50 m release, stability classes A and D yield nearly the same results as the 1 meter release.

4. Summary and Conclusions

The results indicate that it is possible to determine an appropriate set of initial conditions required to accurately predict the dispersion of contaminants in the atmosphere, including the depth of the boundary layer, source height, source strength, and wind direction. Back calculation is attained even during conditions where vertical transport of concentration is great and boundary layer depth is low. Retrieval becomes difficult, however, when the source height is near the top of the boundary layer. We plan to resolve this problem, and possibly to set up boundary layer scenarios such as release in a desert. This work determines the limit as to what one can accomplish in back calculating the source and meteorological variables with the Gaussian Puff model in an identical twin experiment using a GA and a Nelder Meade Downhill Simplex. The next step is to employ real data and to consider more advanced dispersion models such as SCIPUFF.

Acknowledgements

This work was supported by DTRA under grant number W911NF-06-C-0162 and Penn State's Applied Research Laboratory as internal research and development. We would like to extend a special thanks to Anke Beyer-Lout, Luna Rodriguez, Kerrie Long, and Yuki Kuroki for frequent discussions on this project.

References

Allen, C.T., 2006: Source Characterization and Meteorological Data Optimization With a Genetic Algorithm-Coupled Dispersion/Backward Model. M.S. Thesis, Department of Meteorology, The Pennsylvania State University, 79pp.

Allen, C.T., S.E. Haupt., G.S. Young, 2006. Source Characterization with a Genetic Algorithm-Coupled Dispersion/Backward Model Incorporating SCIPUFF. *Journal of Applied Meteorology and Climatology*, **46**, 273-287.

Beychok, M.R., 2005: *Fundamentals of Stack Gas Dispersion, 4th Edition.* Irvine, CA, 201 pp.

Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, New York, NY, 457 pp.

.

Eagleman,, J.R., 1996: *Air Pollution Meteorology 2nd Edition*. Trimedia Publishing Co., 258 pp.

Haupt, R.L. and S.E. Haupt, 2004: *Practical Genetic Algorithms, 2nd edition.* Wiley, New York, NY, 255 pp.

Hsu, S.A., 2003: Nowcasting Mixing Height and Ventillation Factor for Rapid Atmospheric Dispersion Estimates on Land, *National Weather Digest,* December 2003,37, 75-78.

Long, K.J., 2007: Improving Contaminant Source Characterization and Meteorological Data Forcing with a Genetic Agorithm. M.S. Thesis, Department of Meteorology, The Pennsylvania State University.

Stull, R.B., 1988: An *Introduction to Boundary Layer Meteorology*. Kluwar Academic, The Netherlands, 666 pp.

Figures

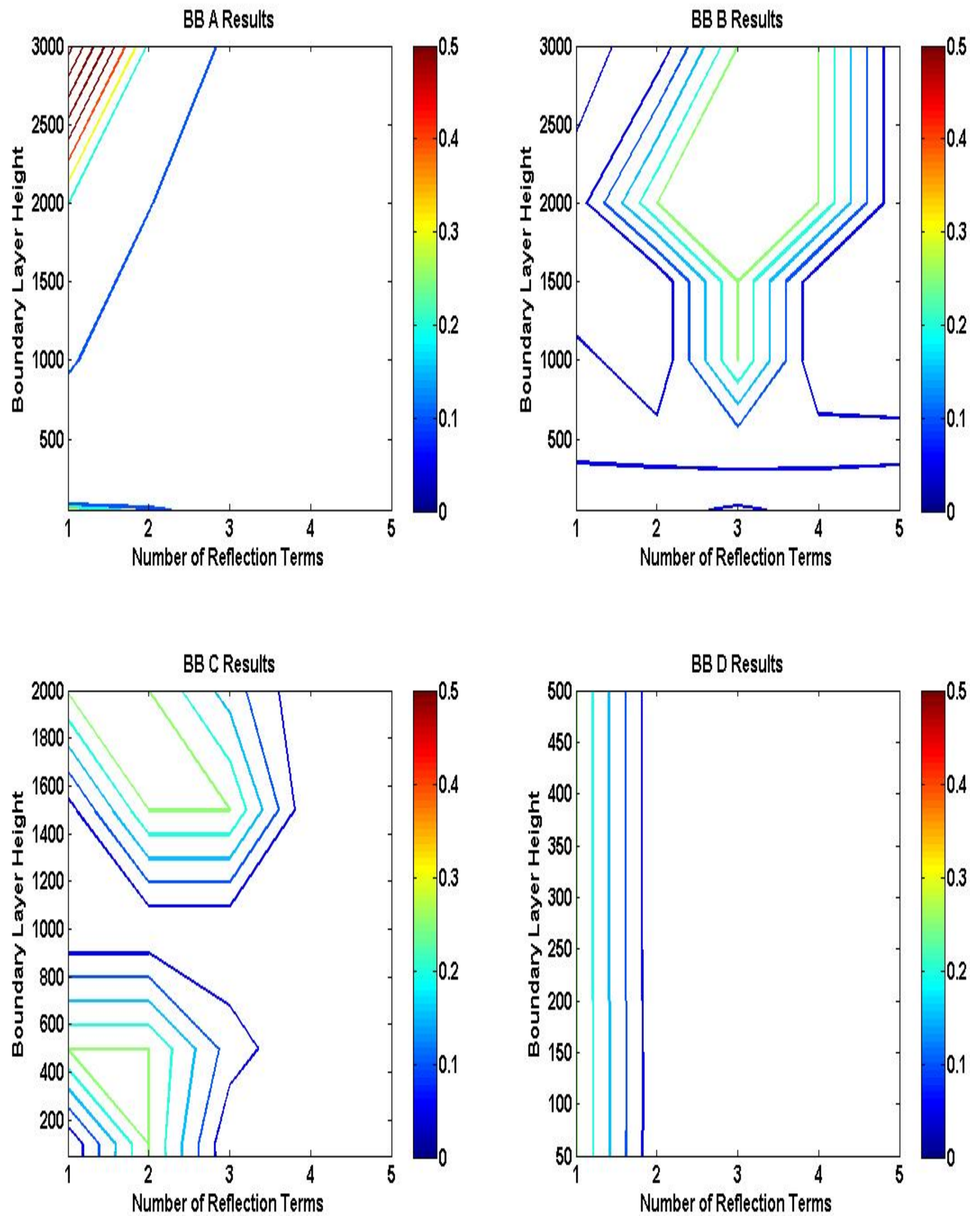

Figure.1 Results for a 1 m release for stability classes A through D for an 8x8 grid.

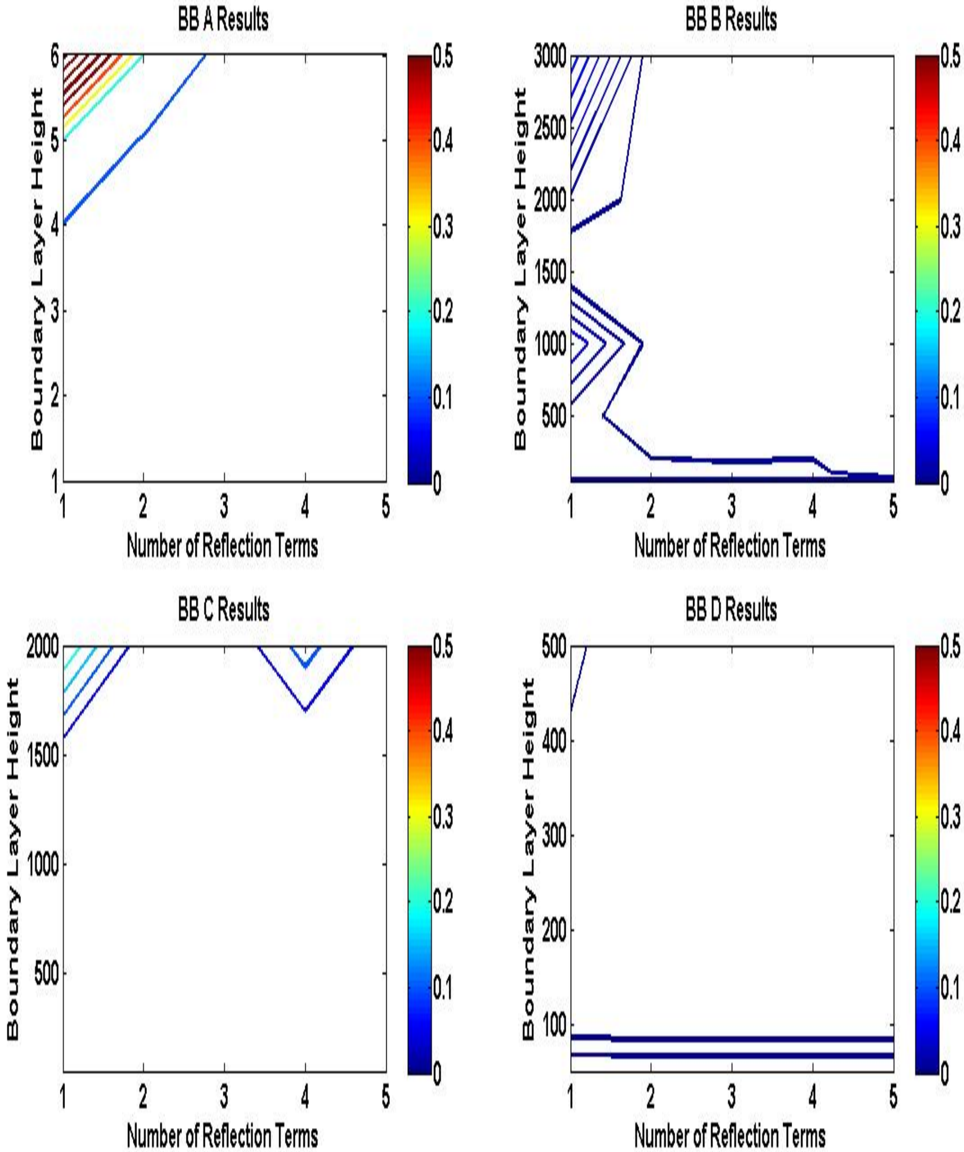

Figure.2 Results for a 50 m release for stability classes A through D for an 8x8 grid.