J4.1 DATA REQUIREMENTS FOR ASSIMILATING CONCENTRATION DATA WITH A GENETIC ALGORITHM

Sue Ellen Haupt*
Kerrie J. Long
George S. Young
Anke Beyer-Lout

Applied Research Laboratory and Meteorology Department,
The Pennsylvania State University, State College, PA

1. INTRODUCTION

In the event of a contaminant release, atmospheric transport and dispersion (AT&D) models would be used to predict the path of the contaminant plume. If monitored contaminant concentration data are available, various assimilation techniques can be applied to incorporate the data into the transport and dispersion model, and thus, more accurately predict the plume path. We refer to this as the forward assimilation problem.

The AT&D models can also be combined with other techniques to estimate unknown source characteristics or to retrieve meteorological data, the backward assimilation problem. Recovering such data is equally important for AT&D prediction.

Both the back-calculation techniques and the forward assimilation methods rely on obtaining sufficient concentration data monitored by either a stationary or a mobile sensor network. In addition, how much data is required is an open question. To be useful, the sensor network must be sited strategically or should be evolvable to follow the plume of contaminant.

A second critical issue for AT&D is determining accurate local meteorological data. A relatively small error in wind direction can produce a large error the concentration field since the transport could be in the wrong direction (Peltier et al. 2008). Even when the wind direction is known, local effects can lead to large errors (Krysta et al. 2006). Therefore, our assimilation methods emphasize using field monitored concentration data to infer the correct wind data. Note that there is only one-way coupling between the AT&D concentration tendency equation and the wind field evolution equations. Although the wind field forces the dispersion equations, the concentrations have no impact on the wind field. Thus, a resourceful method is required to infer the wind field from the concentration field. This work uses a genetic algorithm (GA) for that purpose.

This paper discusses the requirements for developing a sensor network for dispersion assimilation and assesses data needs for defining the plume and for back-calculating source characteristics and meteorological data. Concepts from information theory are used to delineate the minimum requirements for the backward assimilation process.

Section 2 discusses the data requirements for the forward assimilation problem while section 3 treats the back-calculation for source characterization. A discussion of the data requirements in the presence of noise appears in section 4. Section 5 summarizes and discusses the results.

2. DATA REQUIREMENTS FOR FORWARD ASSIMILATION

For the forward problem, we wish to assimilate chemical, biological, radiological, or nuclear (CBRN) concentrations into a simple wind model that then forces a transport and dispersion model to forecast contaminant concentration. There is a long history of assimilating monitored data into meteorological models (Daley 1991, Kalnay 2003). In most cases, the goal is to assimilate data observations into the model fields so that the analyzed field is consistent with the model physics. Most methods work with observed quantities that are either the same fields as those being predicted or ones that can be readily transformed into the predicted quantities. For the CBRN problem, however, the observed quantity is concentration, but the wind field must be modeled to predict a concentration closer to that observed.

* *Corresponding author address:* Sue Ellen Haupt, Applied Research Laboratory, P.O. Box 30, Pennsylvania State University, State College, PA 16804; e-mail: haupts2@asme.org

Therefore, it requires inverting a full transport and dispersion model to relate wind field to concentration. Thus, using concentration data to assimilate the wind field is a complex problem. The goal here is to use concentration data to infer a time varying wind field and to use that field to predict the subsequent transport and dispersion.

A basic example of assimilating concentration data into a dispersion model is a continuous release modeled with a Gaussian puff dispersion model in a meandering flow field. Not only does it vary smoothly in time and space, but it also represents an important realizable state of the atmosphere. It is well documented that meandering wind conditions are common during nocturnal stable boundary layer conditions (Hanna 1983, Mahrt 1999, among others). It is also analogous to vertical plume meandering observed in unstable direct numerical simulations in the convective boundary layer (Liu and Leung 2005).

We choose to evaluate concentration assimilation methods in a varying wind field in the context of an identical twin experiment; that is, the monitored concentration data is "created" using the same transport and dispersion model that will be used for each assimilation step (Daley 1991). Such an approach has the advantage that we generate a "truth" that can be used for comparing our results without the need to consider sensor errors, model errors, or background noise at this initial stage of technique development.

We concentrate on an instantaneous release of contaminant in a neutrally buoyant atmosphere, which can be modeled with a Gaussian puff equation:

$$C_r = \frac{Q\Delta t}{(2\pi)^{1.5}\sigma_x\sigma_y\sigma_z}\exp\left(\frac{-(x_r-Ut)^2}{2\sigma_x^{\,2}}\right)\exp\left(\frac{-y_r^{\,2}}{2\sigma_y^{\,2}}\right)\times \left[\exp\left(\frac{-(z_r-H_e)^2}{2\sigma_z^{\,2}}\right)+\exp\left(\frac{-(z_r+H_e)^2}{2\sigma_z^{\,2}}\right)\right] \quad (1)$$

where: C_r is the concentration at receptor r,
(x_r, y_r, z_r) are the Cartesian coordinates downwind of the puff,
Q is the emission rate,
Δt is the length of time of the release itself,
t is the elapsed time since the release,
U is the wind speed,
H_e is the effective height of the puff centerline, and
$(\sigma_x, \sigma_y, \sigma_z)$ are the standard deviations of the concentration distribution in the *x*-, *y*-, and *z*-directions, respectively.

The standard deviations of the model are computed according to Beychok (1994):

$$\sigma = \exp\left[I + J\ln(x) + K\left(\ln(x)\right)^2\right] \quad (2)$$

where: x is the downwind distance in km, and
I, J, and *K* are coefficients determined by Pasquill stability class for both σ_x, and σ_z and can be found in Beychok (1994).

The puff transport and dispersion occurs in a sinusoidally varying wind field with a constant wind speed of 5 ms $^{-1}$ and direction, θ, defined as:

$$\theta = \theta_0 \sin(2\pi\omega t) \quad (3)$$

where: θ_0 is the maximum amplitude (set at 20°),
ω is the oscillation frequency (set at 1/600 s $^{-1}$), and
t is the time variable, measured from the time of release.

For an instantaneous release, it is equivalent to view this sinusoidal wind variation as either varying in time (the entire field with a single wind that varies in time) or in space (meandering concentration field, as would be the case where local terrain affected flow). The goal is to use our inversion routine to recover this time series of wind direction from concentration measurements. Then we will determine whether this process can be accomplished in a data-sparse situation.

The static approach to this problem uses the spatially varying wind field to produce a meandering concentration field, as would be the case where local terrain affects the flow. In this case a fit to the Lagrangian evolution of the plume concentration field is accomplished in a single GA solution. We presume that the entire field is available *a priori* and seek to find the wind direction at each time. In this case, the wind speed is held constant.

To solve the problem, it is posed as one in optimization. We construct a cost function that compares the log normal monitored concentration with those predicted by the current guess at the correct wind direction. The cost function is

$$\text{cost function} = \frac{\sum_{t=1}^{5}\sqrt{\sum_{r=1}^{\mathrm{TR}}\left(\ln(aC_r+\varepsilon)-\ln(aR_r+\varepsilon)\right)^2}}{\sum_{t=1}^{5}\sqrt{\sum_{r=1}^{TR}\left(\ln(aR_r+\varepsilon)\right)^2}} \quad (4)$$

where: C_r is the concentration as predicted by the dispersion model given by (1),

R_r is the observation data value at receptor r,

TR is the total number of receptors,

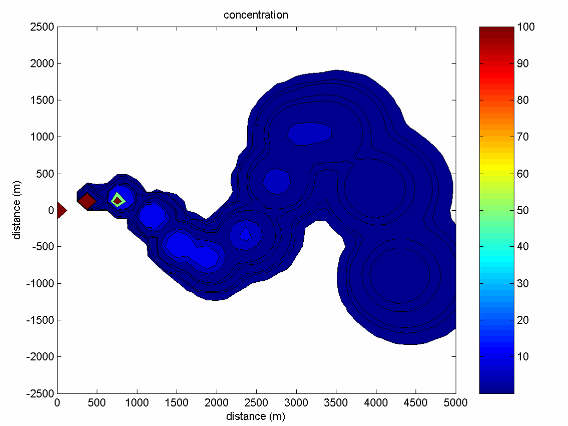
a and ε are constants.

Because the concentration data is log-normally distributed, taking the natural logarithm allows for a more meaningful comparison between the forecast and observation values. The value of the normalization scale factor, *a,* is determined by taking the sum of every concentration value, R_r, over the entire domain and over all time steps and dividing one by that total. To avoid taking the natural logarithm of 0, a small ε (1×10^{-13}) is added to the aC_r and aR_r quantities.

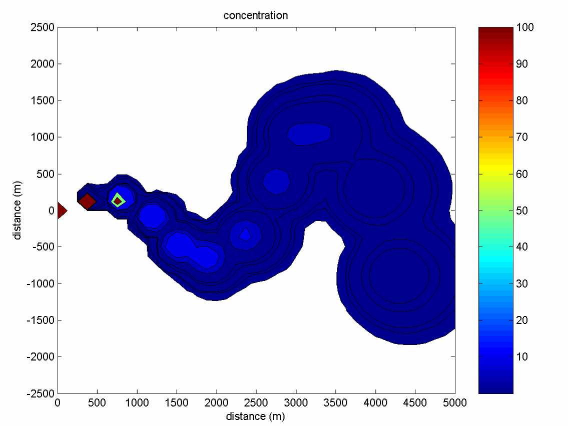
The problem is solved using a continuous parameter genetic algorithm (Haupt and Haupt 2004, Haupt et al. 2006). The GA is elitist and retains the best candidate solution.

The first experiment uses all of the sensor data in the 41x41 grid and sampled the plume every four time steps (80s). Figure 1b illustrates the resulting plume. The plots are nearly indistinguishable from the sampled data in Figure 1a. The GA match to the actual wind direction is indicated in Figure 2. Except for the two initial time steps, the wind direction found by the GA and the truth are nearly identical. The lack of agreement at the initial time reflects the fact that the plume at that time had not yet produced any concentrations – i.e., the problem at that time is ill-posed. The second calculation overshoots to push the computed wind direction toward the exact solution. Figure 3 shows the GA convergence. The cost function continues to decrease, indicating that we could continue to iterate the GA to get increasingly more accurate solutions.

What if we do not have such a dense sensor network? Figure 4a shows the results of sampling one in four of the sensors (an 11x11 grid), every 500m. Although the plume is coarse, it accurately reproduces the location of the plume when compared to Figure 1a. The extremely coarse network in Figure 4b shows the limitations of the model when sampling every eight points on a 6x6 grid (1000m). Even in this data starved situation, however, the basic shape of the plume is captured.



(a)



(b)

Figure 1. Static Assimilation of a meandering plume. a) truth and b) GA computed plume.

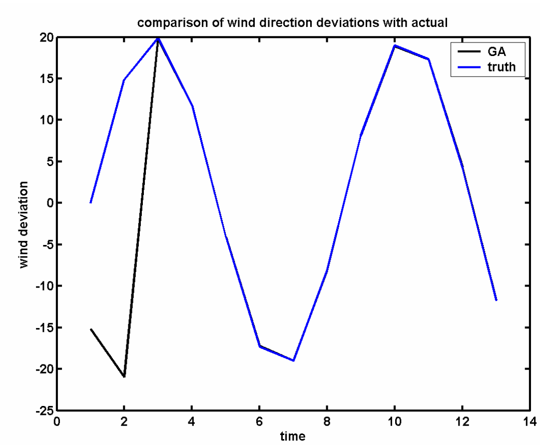


Figure 2. GA match to the actual wind direction.

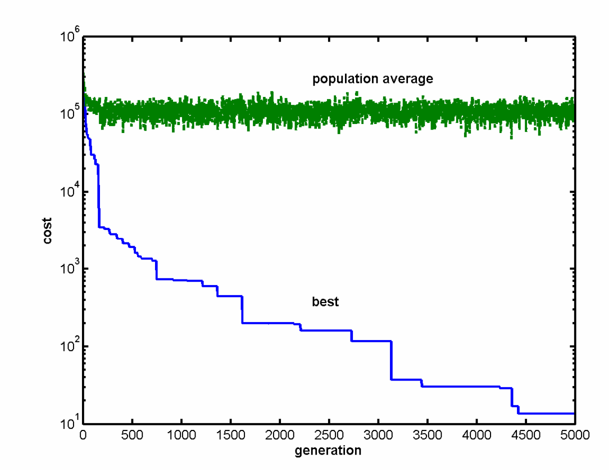
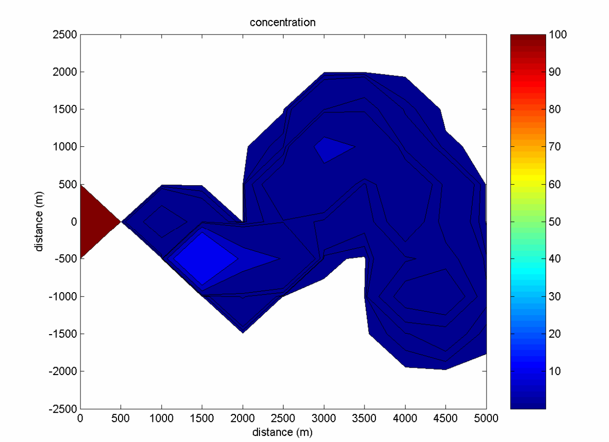
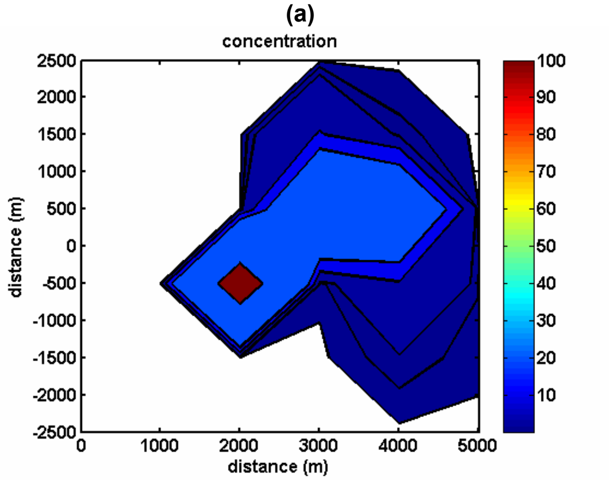


Figure 3. Convergence of GA solution.



(a)



(b)

Figure 4. GA computed plume when the spatial sampling is a) 1:4 and b) 1:8.

3. DATA REQUIREMENTS FOR BACK-CALCULATING SOURCE INFORMATION

3.1 Problem Formulation

The second related problem in the accurate prediction of a dispersed contaminant is the characterization of the source, or the backward problem. What if our field sensors detect significant levels of a contaminant but we don't know where it was emitted or the strength of the source? Emergency managers want to be able to predict the transport and dispersion of the contaminant, but they need good estimates of the input parameters to make that prediction. At times they may not even have access to site specific meteorological data, which is essential for accurate predictions.

The solution to the source and wind characterization problem requires a robust optimization method such as a genetic algorithm (GA). The method presented here uses a GA to find the combination of two dimensional source location, source strength, time of release, surface wind direction, and surface wind speed that best matches the monitored receptor data with the forecast concentration data. This approach is also validated with an identical twin experiment.

This study builds on our previous work. Haupt (2005) first demonstrated that a GA-coupled model was capable of back-calculating source strength for several sources given data from a single field monitor (source apportionment). The model relied on a forward computation of dispersed contaminant from a Gaussian plume model. This model was demonstrated to work well for a circular source configuration and for the actual source/receptor configuration from Logan, UT (Haupt 2005). This model was further validated using Monte Carlo statistical techniques and in the presence of noisy observations by Haupt, et al. (2006). Allen et al. (2006) extended their analysis with a more sophisticated dispersion model, SCIPUFF. With this method they correctly identified time of release, source location, and strength contributions from multiple sources for configurations with moderate amounts of white noise. Allen et al. (2007) reconfigured the genetic algorithm model to directly identify four parameters: source location, source strength, and wind direction. When noise was incorporated into this scheme, the results were excellent for grid sizes 8 x 8 and larger. The model was further enhanced (Haupt et al. 2007) and more meteorological variables were added to the back-calculation (Long 2007, Long et al. 2008a, Haupt et al. 2008).

The Gaussian puff model is the same one used in the meandering plume forward assimilation problem above (1). The simulated source release is located at the center of a 16,000 m by 16,000 m equally spaced grid. The receptors are located at the intersections and corners of the grid. The release location is the center of the grid to allow assessing model skill with any wind direction. Note, however, that means that if we knew the wind direction *a priori*, we would only rely on about one-quarter of the sensors. Since we wish to assess how much data is necessary, we consider seven different grid spacings (Table 1). Based on a 16,000 m by 16,000 m domain size and a wind speed transporting the puff at 5 m s^{-1}, the best time intervals are 6, 12, 18, 24, and 30 minutes following the release. The height of release is 10 m above the surface. We assume Pasquill stability class D (neutral). Figure 5 shows the progression of the puff at the five times for two different wind directions considered here: 180°and 225°.

Table 1. Seven different grid sizes and corresponding grid-spacing studied.

Grid Size	Grid-spacing
2x2	8000 m
4x4	4000 m
6x6	2667 m
8x8	2000 m
16x16	1000 m
32x32	500 m
64x64	250 m

A population of 1200 chromosomes is initialized with random values. The cost function measures how close each forecast concentration as predicted by (1) is to the observed concentration and uses the same cost function (summed over all five time steps) as used for the forward assimilation (4).

The GA is generally good at finding the correct solution basin: so by coupling it with a gradient descent method, we are able to further optimize the solution. Such a combination of algorithms is known as a hybrid GA. The best candidate solution found by the GA after 100 iterations is used as the first guess for a Nelder-Mead downhill simplex algorithm (Nelder and Mead 1965), which performs a local search to find the minimum of that basin. The GA searches the

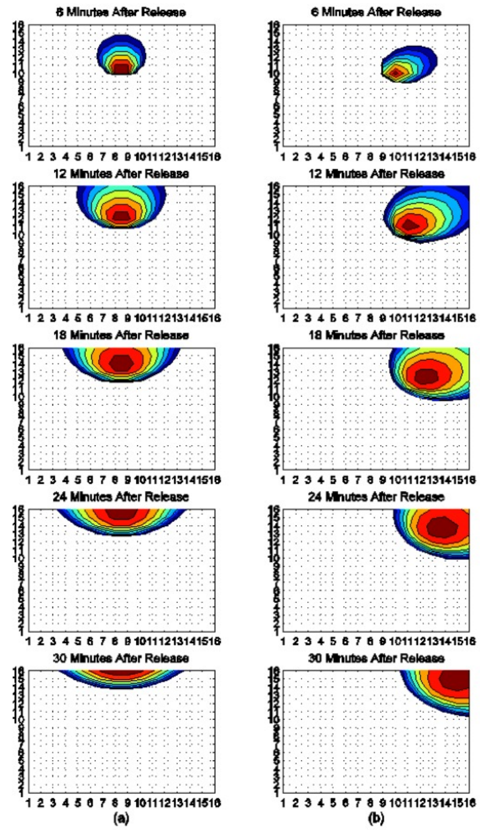


Figure 5. Evolution of the puff in time for (a) 180° wind direction and (b) 225° wind direction on a 16×16 grid. The source is located is the center of the domain, between receptors 8 and 9.

solution space of each parameter according to the bounds dictated in Table 2: however, the Nelder-Mead simplex algorithm used here does not allow setting bounds. As a result, parameters such as source strength can become negative despite the physical impossibility.

Table 2. Searchable solution space for the GA

Parameter	Minimum Value	Maximum Value
Location (x,y)	- 8000 m	+ 8000 m
Source Strength	0 Kg s^{-1}	5 Kg s^{-1}
Time of Release	-300 s	+ 300 s
Wind Direction	0 °	360 °
Wind Speed	0 m s^{-1}	20 m s^{-1}

Hereafter, a solution generated by the GA or the Nelder-Mead downhill simplex algorithm is said to be *within tolerance* when it meets the criteria given by Table 3. As shown there, this study seeks to retrieve six parameters: source strength, source location (x,y), time of release, wind speed, and wind direction. The first set of runs is conducted without noise. The model is run ten times for both wind directions for each of the grid size/grid-spacing configurations described in Table 1.

Table 3. Tolerances for an acceptable solution

Parameter	Tolerance
Location (x,y)	± 10 m
Source Strength	± 0.05 Kg s^{-1}
Time of Release	± 5 s
Wind Direction	± 0.1 °
Wind Speed	± 0.05 m s^{-1}

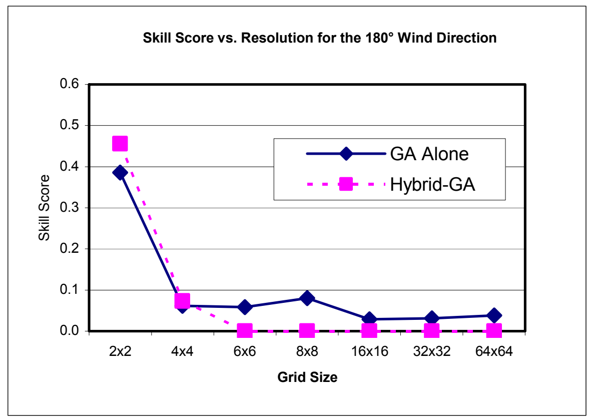
3.2 Results in a Noiseless Environment

The first set of runs assume that the data are error free. Later additive and multiplicative noise will be incorporated into the model in order to simulate errors. Table 4 (see end of paper) displays the results for the 180° wind direction case. We see that even with a 4x4 grid, the GA coupled with the Nelder-Mead optimization characterizes each parameter well with the only exception being the location, which is found to be 20 m south of the true source. The model characterizes the source within tolerance for the 6x6 grid sizes and larger for the 180° wind direction. Recall that if we knew the wind direction, these grids would require less that one-quarter the number of sensors used.

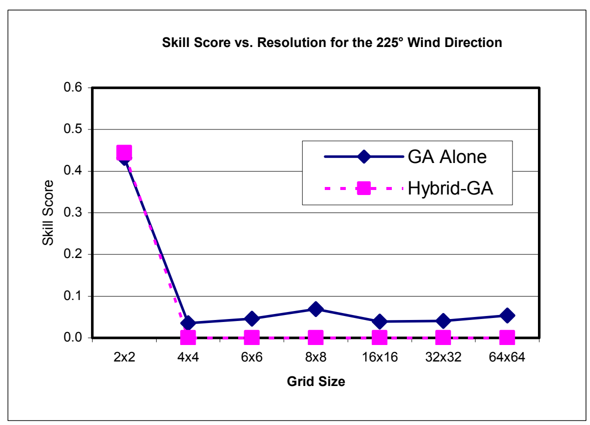
In the 225° wind direction case (see Table 5 at the end of the paper), the model is able to identify nearly every parameter in the 4x4 grid also. The only exception is that the wind direction is found to be slightly larger than 225°. As with the 180° wind direction, the model characterizes the source within tolerance for grid sizes 6x6 and larger for the 225° wind direction.

Because we are using an identical twin experiment, skill scores can be used to quantify the closeness of each solution to the known answers. The skill score is evaluated using five component equations, one to quantify the accuracy of each parameter (i.e. source strength, source location, time of release, wind speed, and wind direction). The most desirable skill score is 0; the least desirable skill score is 1. Skill scores that fall below 0.1 correspond to meeting the tight tolerances defined in Table 3.

Figure 6 illustrates that the skill scores fall below 0.1 at relatively low resolutions. For the 180° wind direction, the GA finds a very good solution and the Nelder-Mead search optimizes the solution with the 6x6 grid size and larger (Figure 6a). For the 225° wind direction, the GA finds a good solution for grid sizes as small as 4x4 and the Nelder-Mead method is able to refine the solution further (Figure 6b). Thus we conclude that if we do not know the wind direction, we require at least 16 sensors. If wind direction was known, four sensors would be sufficient.



(a)



(b)

Figure 6. Skill score versus resolution for (a) 180° and (b) 225° wind direction.

3.3 Results with Noisy Data

Observed concentration data are likely to be fraught with uncertainty. Many chemical sensors currently in use display a rectangular bar-shaped readout for the concentration that provides only an order of magnitude precision (Robins, et al. 2005). Errors in observations can also result from uncertainty in the meteorological data as well as from the chaotic nature inherent in turbulent flow as discussed in Haupt et al. (2006). In order to simulate a more realistic environment, we corrupt the observation data with both additive and multiplicative noise at six signal-to-noise ratios (SNRs): 100, 10, 5, 2, 1, and 0.1. The two validations presented previously were run without noise, that is at an SNR of infinity. The signal and the noise are of equal magnitudes for an SNR of 1. For an SNR of 0.1, the noise is ten times greater than the signal and at this point we expect the model to fail. The complete results for both additive and multiplicative noise can be found in Long (2007).

Gaussian additive noise has a mean of zero. For this study, we've used a clipped Gaussian additive noise: that is, concentrations below 0 are set to 0. Figure 7 plots the skill score for additive noise at every grid size for the 180° wind direction. Results were quite similar for the 225° wind direction (Long 2007). Again, skill score values of 0.1 or lower correspond to good results whereas skill score values of 0.2 or greater indicate less accurate solutions. The results for SNR of 100, 10, and 5 are very similar to the model results without noise. When the noise reaches 50% of the signal (SNR = 2), many of the solutions found by the hybrid GA are outside the domain and the model begins to fail. At an SNR of 1 and 0.1, solutions are often unphysical and the model fails. Figure 7 also illustrates that 2x2 grid sizes are too small to produce good solutions and the model fails for that resolution. Again, these conclusions assume that wind direction is unknown.

We also considered the impact of multiplicative noise. The multiplicative noise is a clipped Gaussian with a mean of 1. Figure 8 plots the skill score as computed by the GA for multiplicative noise for each grid size for the 180° wind direction. Again, the results for SNR of 100, 10, and 5 are very similar to the model results without noise. When the noise approaches the same magnitude as the signal (SNR = 2), many of the solutions are outside the domain and the model begins to fail. Again the model is unsuccessful at SNR of 1 and 0.1 and for 2x2 grid sizes at all noise levels. Model behavior is similar for both additive and multiplicative noise, which demonstrates consistent performance. As before, the behavior in the presence of multiplicative noise was quite similar for a wind direction of 225° (Long 2007).

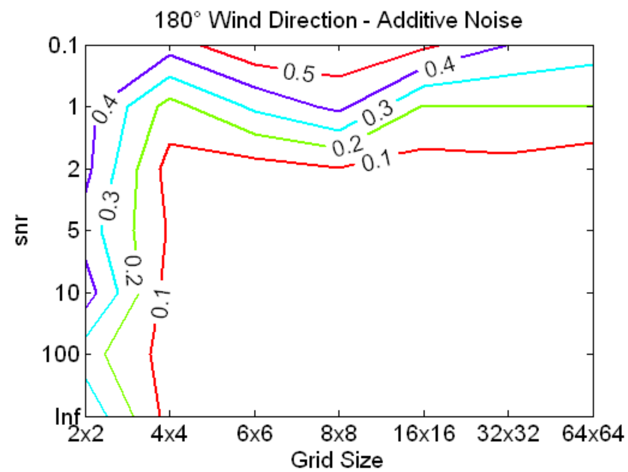


Figure 7. Skill score plot with additive noise, wind direction = 180°.

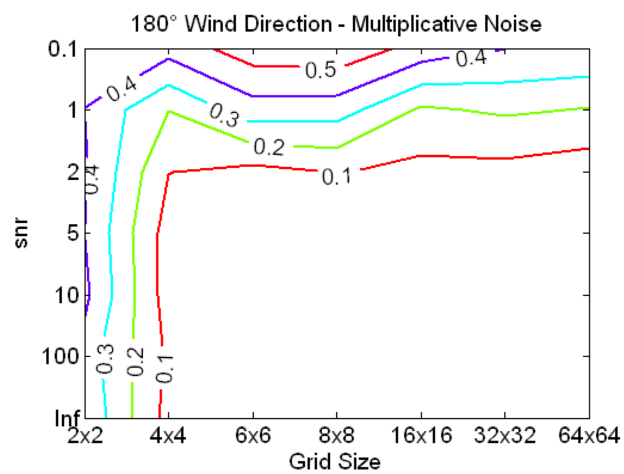


Figure 8. Skill score plot with multiplicative noise, wind direction = 180°.

4. DATA REQUIREMENTS IN A NOISY ENVIRONMENT

In order to quantify how many receptors are necessary to obtain a good solution for this full problem of solving for both meteorological and source variables, we examine heuristic methods for developing information measures. This work is inspired by the field of information theory (IT) initiated by Shannon (1948). The goal is to define an information measure that quantifies the amount of independent information in a data analysis. In this case, it means that we want to define how fine the grid resolution must be for a given amount of noise in the data in order for our method to obtain a sufficiently close solution. We do that by defining two correlations: $\overline{r^2_{t,t+1}}$ is the squared correlation between the noiseless concentration field with the one at the next time step (where *t* is the time step) and $\overline{r^2_{c,c+n}}$ is the squared correlation between the concentration field with and without noise, both averaged over all time steps. Both correlations are functions of grid resolution and signal to noise ratio and are determined from the model runs that produced Figures 7 and 8 plus similar plots for the 225° wind direction case. We define an information measure, FIT, as:

$$FIT = \left[\left(1 - \overline{r^2_{t,t+1}}\right)^{\alpha} \left(\overline{r^2_{c,c+n}}\right)^{\beta} \right] \geq \gamma \qquad (5)$$

where α and β are powers to be determined and γ is the threshold value to be fit such that the quantity yielded produces a successful solution. The first factor is designed to measure the amount of information that can be extracted from the puff transport and dispersion and the second factor is designed to measure the degree to which the pattern remains uncontaminated by noise.

FIT generates a binary matrix that is a function of grid resolution and signal to noise ratio. It is then compared with the successful model configurations matrix, SUS. Successful configurations are defined as those grid size/noise combinations where the ten run median value of every parameter is found within the strict tolerances defined in Table 3. The successful configurations for the 180° wind direction are indicated in Table 4 (see end of paper). A similar table is constructed for the 225° wind direction (not presented here).

By minimizing the difference between the FIT and SUS, we are able to determine the values of α, β, and γ that will guarantee a model setup that produces solutions within the strict tolerances defined previously. Thus, we are able to determine the minimum number of receptors needed to obtain a solution within tolerance in a specified noise environment.

Note that the fit to the parameters is not unique here. We wish to optimize the two powers and the threshold value so that the number of matches is maximized. There are ranges of values of α, β, and γ that meet this goal. When jointly considering both wind directions, a possible solution was computed with a GA to be α=0.1, β=4, and γ=0.85. This means that the squared correlation between the concentration field with and without noise is a critical factor in determining how many receptors are necessary. In a noiseless or low noise environment, a 6×6 grid is sufficient to back-calculate all wind and source variables. When the noise is within a factor of two of the signal, the GA can no longer distinguish the source and meteorological parameters accurately. This analysis provides guidance for configuring a receptor grid to provide enough data for this GA-based model to back-calculate source and meteorological parameters. It is problem dependent, however, and should be further developed for other configurations.

5. CONCLUSIONS

We have assessed using a GA to assimilate data into an AT&D model that predicts concentration. The goal is to infer meteorological data in the context of both a forward assimilation problem and a problem of back-calculating the model input parameters. In both cases, the GA has successfully computed the required parameters. In both cases, we considered whether the algorithm could still perform well when fewer sensor observations were available. In general, the algorithm works well even when only two or three sensors are impacted by the concentration. That observation implies, however, that the sensors must be strategically sited to be able to intercept the concentration plume. Several methods could be used to facilitate such a process. First, one could optimize sensor siting if the likely meteorological conditions are known *a priori*. Since that is unlikely, one could use historical climate data to estimate the most probable wind conditions and use that information to guide sensor siting. The most useful approach, though, would be to have an evolvable sensor network available, perhaps mounted on Unmanned Aerial Vehicles (UAVs). In other work (not shown here) we have developed methods to guide a single UAV with a mounted CBRN sensor through a concentration plume so that it makes optimal use of available concentration data for the back-calculation problem.

We have also considered the impact of including either additive or multiplicative noise on the ability of the back-calculation GA model to invert for the AT&D input parameters. Such noise simulates errors due to model error, monitoring error, and the inherent error due to the stochastic nature of turbulence. The model is robust and can withstand additive and multiplicative signal-to-noise ratios of 100, 10, and 5. Success drops significantly as the noise reaches about 50% of the signal (SNR=2). Further cases (not shown) confirm that this result is not sensitive to wind direction or cost function metric.

These results have been validated using an identical twin experimental approach to construct the synthetic data. Such data are ideal for developing and testing purposes. We plan to further test our model with field test data and are currently testing the model with CFD-produced data. We are additionally in the process of upgrading the model by replacing the Gaussian puff equation with a more sophisticated dispersion model (Long et al. 2008b). Finally, we are also evaluating the impact of more realistic sensor models, including thresholds and saturation values (Rodriguez et al 2008).

In closing, the implications of this work are that when planning a sensor network, it is imperative to consider the physics of the situation if one expects the resulting data to be useful. It is never adequate to assume a constant known wind and simple dispersion characteristics. Therefore, groups planning how to design or evolve such networks must work closely with transport and dispersion modelers to assess the problem. As future networks become mounted on Unmanned Aerial Vehicles (UAVs), further research will be needed to coordinate the control of evolvable sensor networks in a simulated setting. We suspect that such considerations extend beyond this single problem and that the control of evolvable networks often depends on the physics of the situation.

ACKNOWLEDGEMENTS – This research is supported by DTRA (grant number W911NF-06-C-0162). We especially thank our project monitors, John Hannan and Christopher Kiley.

REFERENCES

Allen, C.T., S.E. Haupt, and G.S. Young, 2007: Source Characterization with a Receptor/ Dispersion Model Coupled With A Genetic Algorithm, *Journal of Applied Meteorology and Climatology,* **46**, 273-287.

Allen, C.T., G.S. Young, and S.E. Haupt, 2006: Improving Pollutant Source Characterization by Optimizing Meteorological Data with a Genetic Algorithm, *Atmos. Env.,* **41**, 2283-2289.

Beychok, M. R., 1994: Fundamentals of Gas Stack Dispersion, 3rd ed. Milton Beychok, pub., Irvine, CA, 193 pp.

Daley, R., 1991: *Atmospheric Data Assimilation*. Cambridge University Press, Cambridge, 457 pp.

Hanna S.R., 1983: Lateral turbulence intensity and plume meandering during stable conditions, *Journal of Climate and Applied Meteorology*, **22**, 1424-1430.

Haupt R. L., Haupt S. E., 2004: Practical Genetic Algorithms, 2nd edition with CD. John Wiley & Sons, New York, NY.

Haupt, S.E., R.L. Haupt, and G.S. Young, 2008: A Mixed Integer Genetic Algorithm used in Chem-Bio Defense Applications, submitted to *Journal of Soft Computing*.

Haupt, S.E., G.S. Young, and C.T. Allen, 2007: A Genetic Algorithm Method to Assimilate Sensor Data for a Toxic Contaminant Release, *Journal of Computers*, **2**, 85-93.

Haupt, S.E. G.S. Young, and C.T. Allen, 2006: Validation Of A Receptor/Dispersion Model Coupled With A Genetic Algorithm, *Journal of Applied Meteorology,* **45**, 476–490.

Haupt, S.E., 2005: A Demonstration of Coupled Receptor/Dispersion Modeling with a Genetic Algorithm, *Atmos. Env.*, **39**, 7181-7189.

Kalnay, Eugenia, 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, Cambridge, 136-204.

Krysta, M., Bocquet, M., Sportisse, B., Isnard, O., 2006: Data Assimilation for Short-Range Dispersion of Radionuclides: An Application to Wind Tunnel Data. *Atmos. Env*. **40**, 7267-7279.

Liu D.-H., Leung D.Y.C., 2005: On plume meandering in unstable stratification, *Atmos. Env.*, **39**, 2995-2999.

Long, K.J., 2007: Improving contaminant source characterization and meteorological data forcing with a genetic algorithm. Master's Thesis, The Pennsylvania State University.

Long, K.J., S.E. Haupt, and G.S. Young, 2008a: Improving Meteorological Forcing and Contaminant Source Characterization Using a Genetic Algorithm. To be submitted to *Optimization and Engineering.*

Long, K.J., S.E. Haupt, G.S. Young, 2008b: Source Characterization and Meteorology Retrieval using a Genetic Algorithm with SCIPUFF, 15th Joint Conference on the Applications of Air Pollution Meteorology with the A&WMA, New Orleans, LA, Jan. 20-24.

Mahrt L., 1999: Stratified atmospheric boundary layers, *Boundary Layer Meteorology*, **90**, 375-396.

Nelder, J.A., Mead, R., 1965: A Simplex Method for Function Minimization, *Computer Journal*, **7**, 308-313.

Peltier, L.J., S.E. Haupt, J.C. Wyngaard, D.R. Stauffer, A. Deng, and J. Lee, 2008: Parameterization of NWP uncertainty for dispersion modeling, 15th Joint Conference on the Applications of Air Pollution Meteorology with the A&WMA, New Orleans, LA, Jan. 20-24.

Robins, P., Rapley, V., Thomas, P., 2005: A Probabilistic Chemical Sensor Model for Data Fusion, Dstl, Salisbury, UK.

Rodriguez, L.M., S.E. Haupt, and G.S. Young, 2008: Adding Realism to Source Characterization with a Genetic Algorithm, 15th Joint Conference on the Applications of Air Pollution Meteorology with the A&WMA, New Orleans, LA, Jan. 20-24.

Shannon, C.E. 1948: A mathematical theory of communication, *Bell System Technical Journal*, **27**, 379-423 and 623-656.

Table 4. Six parameter results for the 180° wind direction – ten run average.

	Grid Size	Found θ (°)	Strength (Kg s^{-1})	(x,y) (m,m)	Release Time (s)	Speed (m s^{-1})	Cost Function
Actual Solution		180.00	1.00	(0,0)	0	5.0	1.0e-3
GA Alone	2x2	179.12	2.94	(120,-730)	172	7.6	2.0e-1
Hybrid GA	2x2	178.10	5.05	(290,-760)	184	7.9	1.0e-1
GA Alone	4x4	179.69	1.30	(40,-40)	26	5.2	4.1e-3
Hybrid GA	4x4	180.00	1.00	(0,-20)	0	5.0	2.0e-9
GA Alone	6x6	179.91	1.69	(10,80)	28	5.0	2.2e-3
Hybrid GA	6x6	180.00	1.00	(0,0)	0	5.0	3.2e-9
GA Alone	8x8	179.18	1.90	(80,170)	39	5.0	6.0e-3
Hybrid GA	8x8	180.00	1.00	(0,0)	0	5.0	3.1e-9
GA Alone	16x16	179.96	1.35	(0,40)	13	5.0	1.6e-3
Hybrid GA	16x16	180.00	1.00	(0,0)	0	5.0	3.4e-9
GA Alone	32x32	180.07	1.39	(-10,40)	13	5.0	1.8e-3
Hybrid GA	32x32	180.00	1.00	(0,0)	0	5.0	3.6e-8
GA Alone	64x64	179.96	1.45	(0,80)	18	5.0	2.1e-3
Hybrid GA	64x64	180.00	1.00	(0,0)	0	5.0	3.0e-9

Table 5. Six parameter results for the 225° wind direction – ten run average.

	Grid Size	Found θ (°)	Strength (Kg s^{-1})	(x,y) (m,m)	Release Time (s)	Speed (m s^{-1})	Cost Function
Actual Solution		225.00	1.00	(0,0)	0	5.0	1.0e-3
GA Alone	2x2	225.92	4.42	(-3340,-3070)	-83	8.6	7.3e-4
Hybrid GA	2x2	225.81	4.76	(-3290,-3040)	-81	8.6	5.5e-10
GA Alone	4x4	224.68	1.30	(80,-40)	-2	5.0	9.6e-4
Hybrid GA	4x4	225.02	1.00	(0,0)	0	5.0	1.1e-7
GA Alone	6x6	225.50	1.43	(90,180)	17	4.9	2.8e-3
Hybrid GA	6x6	225.00	1.00	(0,0)	0	5.0	1.2e-9
GA Alone	8x8	225.25	1.71	(70,120)	36	5.0	4.1e-3
Hybrid GA	8x8	225.00	1.00	(0,0)	0	5.0	2.0e-9
GA Alone	16x16	225.16	1.46	(60,80)	17	5.0	2.9e-3
Hybrid GA	16x16	225.00	1.00	(0,0)	0	5.0	3.5e-9
GA Alone	32x32	225.01	1.46	(60,70)	20	5.0	2.8e-3
Hybrid GA	32x32	225.00	1.00	(0,0)	0	5.0	2.6e-9
GA Alone	64x64	224.91	1.61	(90,80)	27	5.0	3.7e-3
Hybrid GA	64x64	225.00	1.00	(0,0)	0	5.0	3.3e-9

Table 4. Success Table: Check marks '✓' indicate model configurations where the solutions were found within tolerance.

	No Noise	Additive Noise						Multiplicative Noise					
SNR	Inf	100	10	5	2	1	0.1	100	10	5	2	1	0.1
2x2													
4x4													
6x6	✓	✓	✓					✓	✓				
8x8	✓	✓	✓	✓				✓	✓	✓			
16x16	✓	✓	✓	✓				✓	✓	✓			
32x32	✓	✓	✓	✓				✓	✓	✓			
64x64	✓	✓	✓	✓				✓	✓	✓			