J1.4

COUPLING A GENETIC ALGORITHM WITH AN ATMOSPHERIC TRANSPORT AND DISPERSION MODEL

Luna M. Rodriguez[1,2,3], Andrew J. Annunzio[1,2], Kerrie J. Schmehl[2],
Sue Ellen Haupt[1,2,3], and George S. Young[1]
[1]The Pennsylvania State University Department of Meteorology,
the [2]Applied Research Laboratory at The Pennsylvania State University,
and the [3]Research Applications Laboratory at the National Center for Atmospheric Research

1. INTRODUCTION

An accidental or intentional release of Chemical, Biological, Nuclear, or Radioactive (CBNR) material into the atmosphere will require the source of that release to be estimated from remote measurements of the concentration field. However, many existing Atmospheric Transport and Dispersion (AT&D) datasets measure dosages, which make it impossible to retrieve information about the instantaneous concentrations given the stochastic nature of atmospheric turbulence. To ensure that source term estimation (STE) algorithms will work in real world conditions they must be tested against AT&D concentration field datasets. The FUsing Sensor Information from Observing Networks (FUSION) Field Trial 2007 (FFT07) was executed to collect data from an abundance of research-grade tracer, sensor, and meteorological instruments suitable for testing current and future CBNR algorithms (Storwold 2007).

The FFT07 data is partitioned into two datasets. The Trial dataset contains readings from 100 sensors, source information (location and amount), and abundant meteorological information. These data were made available to test and train the current CBNR algorithms. The Case dataset contains 104 different release events with limited meteorological data, concentration data for only four or 16 sensors, and no information about the source location or release amount. For the first phase of FFT07 predictions of the source location and release amount of the Case data were submitted using a Genetic Algorithm (GA) coupled with an AT&D model.

The CBNR algorithm used in this work is the GA coupled AT&D model, and this method has been previously successful at estimating source characteristics and meteorological parameters necessary to predict the AT&D of a contaminant. The AT&D models used in this current study are a Gaussian Plume model and the Second-Order Closure Integrated Puff Model (SCIPuff). The GA coupled AT&D model iteratively compares predictions with observations. For this work the observations are from the FFT07 Trial Data. The comparison of these data to the predictions uses a new dual cost function approach, rather than a single cost function method used previously (Rodriguez 2010). Additionally, strict and loose thresholds are applied to the concentration data to filter noise and a statistical analysis uses bootstrap sampling to quantify the uncertainty in the estimates of the individual releases.

2. DATA

The FFT07 Trial data were examined and Trials 15, 30, and 54 were determined to be representative of the group of single continuous releases in the dataset. The Trial data is available at 20 ms., therefore, to reduce noise two thresholds were applied to each Trial. Previous work (Rodriguez 2011) showed that the GA coupled AT&D model works best when the concentration data spans at least three orders of magnitude; thus the strict threshold is chosen to be 10^{-3} of the maximum concentration detected and the loose threshold is 10^{-4} of the maximum concentration detected in each Trial. After thresholding, each sensor is averaged over the duration of the release and finally, these data are used as the observation at each sensor for each Trial. Only results for Trial 15 and 30 will be shown here and their threshold concentration fields can be viewed in Figure 1.

3. EXPERIMENTAL METHODS

The GA coupled AT&D procedure for source characterization has been proven to work with identical twin data in Allen et al. (2006, 2007), Haupt (2005), Haupt et. al. (2006), and Long et. al. (2010). Like the aforementioned

Corresponding author address: Luna M. Rodriguez, Research Applications Laboratory National Center for Atmospheric Research 3450 Mitchell Lane Boulder, CO 80301; e-mail: lmr257@psu.edu

studies, the GA begins with a set of potential solutions; in this study the potential solutions consist of wind direction, source location (x,y), and emission rate. Also these solutions are bounded, however we now use meteorological data available to limit the potential solution of wind direction to within 30 degrees clockwise or counterclockwise of the mean wind direction. Each set of potential solutions is then fed into an AT&D model, either the Gaussian Plume Model or SCIPUFF. The resulting concentration fields of these models are then compared via a cost function that is logarithmic in concentration,

$$\text{Cost} = \frac{\sum_{t=1}^{Time}\left(\sqrt{\sum_{r=1}^{TR}\left(\log\left(C_r+\varepsilon\right)-\log\left(R_r+\varepsilon\right)\right)^2}\right)_t}{\sum_{t=1}^{Time}\left(\sqrt{\sum_{r=1}^{TR}\left(\log\left(R_r+\varepsilon\right)\right)^2}\right)_t}$$

where: C_r is the concentration as predicted by the dispersion model at receptor r, R_r is the observation data value at receptor r, TR is the total number of receptors, and ε is the minimum concentration detected, added to avoid logarithms of zero and to scale them accordingly. The solutions with the lowest cost mate, mutate, and then this process iterates until it converges to a best solution.

To quantify the uncertainty, the GA coupled AT&D method is run 100 times. Then the median of a bootstrap sample of 10 solutions is computed 1000 times. Finally the mean and standard deviation are then computed. The wind direction and location (x,y) from the mean bootstrap solution are then inserted in the AT&D model with potential emission rate solutions. The resulting concentration fields are then compared via a cost function that is linear in concentration,

$$\text{Cost} = \frac{\sum_{t=1}^{Time}\left(\sum_{r=1}^{TR}\left(C_r-R_r\right)^2\right)_t}{\sum_{t=1}^{Time}\left(\sum_{r=1}^{TR}\left(R_r\right)^2\right)_t}$$

where: C_r is the concentration as predicted by the dispersion model at receptor r, R_r is the observation data value at receptor r, and TR is the total number of receptors. The solution with the lowest cost is then used as the emission rate.

4. RESULTS

The results for Trials 15 and 30 can be seen in Tables 1 and 2, respectively. The results listed in these tables are the absolute error of the mean predictions for wind direction, location (x,y), and emission rate. We show the emission rate errors when using the logarithmic in concentration cost function and the linear in concentration cost function. For all Trials and thresholds with the exception of one scenario (Trial 15 with loose threshold, using SCIPuff as the dispersion model) using a linear in concentration cost function to determine the emission rate shows a smaller error, therefore a better prediction.

In Figure 1 we notice that Trial 15 is missing the upper right quadrant of the sensors. These sensors were either offline or did not pass the post processing quality control test. In this Trial when the strict thresholds are employed we lose a large quantity of data and have worse predictions for the source term parameters. Given the poor quality of data for this Trial a loose threshold is needed because the strict threshold eliminates concentration signals that are necessary to estimate the source term parameters.

In Figure 1, Trial 30 is missing nine of the sensors but only three of these sensors are downwind of the source. In this Trial when the strict thresholds are employed we do not see a significant difference from the case when the loose thresholds are applied. This can also be seen in the similar predictions for the source term parameters when using the Gaussian Plume Model. The SCIPuff model does do better at predicting the source terms when using strict thresholds. Given the better quality of data for this Trial the strict threshold can be applied to estimate the source term parameters.

5. CONCLUSIONS

When using this approach with the dual cost functions our model results improve from previous experiments (Rodriguez et. al. 2010). The logarithmic in concentration cost function does well at capturing the shape and location of the plume, while the linear in concentration cost function is able to capture the higher concentrations necessary to predict the emission rate. This is true when employing loose or strict thresholds. Loose thresholds yield better results when data are sparse. Strict thresholds yield the best results but can only be used when data are of high quality. There is no significant difference in model results between AT&D

model used when the GA method is employed. However, the Gaussian Plume-GA method is much less computationally expensive.

ACKNOWLEDGEMENTS

This work was supported by the Defense Threat Reduction Agency under grants W911NF-06-C-0162, 01-03-D-0010-0012, by the Bunton-Waller Fellowship, and professional development by the Minorities Striving and Pursuing Higher Degrees of Success Professional Development Program.

REFERENCES

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

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

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

Haupt, S. E., G. S. Young, and C. T. Allen, 2006: Validation of a Receptor/Dispersion Model with a Genetic Algorithm Using Synthetic Data. J. Appl. Meteor. 45, 476-490.

Long, K.J., S.E. Haupt, and G.S. Young, 2010: Assessing Sensitivity of Source Term Estimation. Atmospheric Environment, in press.

Rao, .S., 2007: Source estimation methods for atmospheric dispersion, Atmos. Env. 41, 6963-6973.

Rodriguez, L.M., Annunzio, A.J., Long, K.J., Haupt, S.E., Young, G.S., Source Term Characterization of FFT07 data using a Genetic Algorithm. Paper in 16th Conference on Air Pollution Meteorology of the 90th American Meteorological Society Annual Meeting.

Rodriguez, L.M., Haupt, S.E., Young, G.S., Impact of Sensor Characteristics on Source Characterization for Dispersion Modeling. Paper accepted January 2011.

Storwold, D. P., 2007: Detailed Test Plan for the FUsing Sensor Information from Observing Networks (FUSION) Field Trial 2007 (FFT07), 46.

Figure:

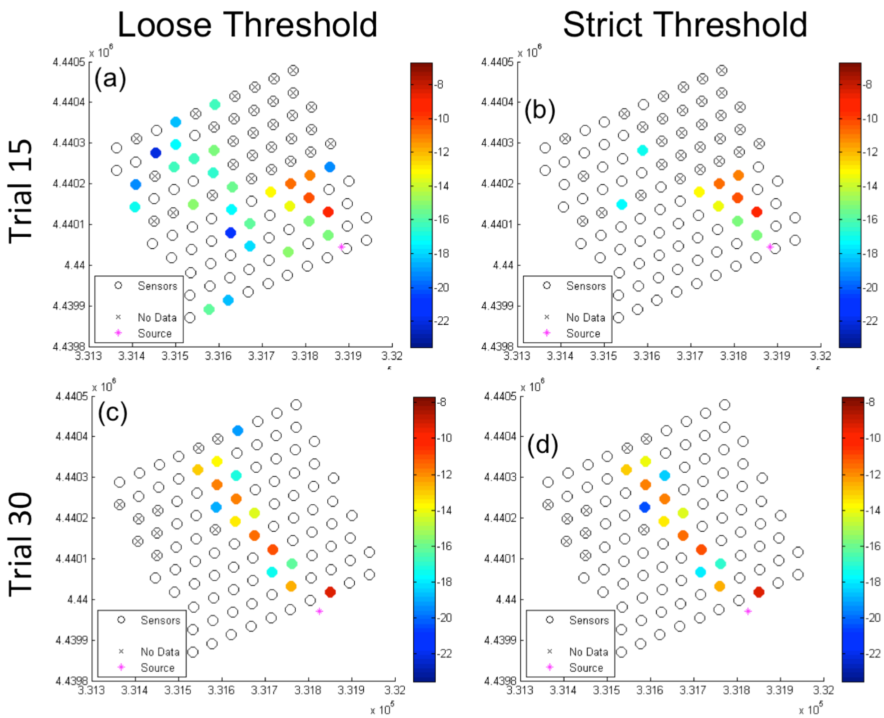

Figure 1. On the top is the concentration field for Trial 15 data with loose (a) and strict (b) thresholds and on the bottom is the concentration field for Trial 30 data with loose (c) and strict (d) thresholds. The concentration data is in kg/s, however it has been logged.

Tables:

Table 1. Absolute error of the prediction parameters for Trial 15.

		Wind Direction [degrees]	Strength Log Cost [kg/s]	Location (X) [meters]	Location (Y) [meters]	Strength Linear Cost [kg/s]
Gaussian	Strict Threshold	3	0.216	10	15	0.037
	Loose Threshold	12	0.258	12	11	0.033
SCIPuff	Strict Threshold	9	0.429	884	60	0.989
	Loose Threshold	4	0.003	41	3	0.035

Table 2. Absolute error of the prediction parameters for Trial 30.

		Wind Direction [degrees]	Strength Log Cost [kg/s]	Location (X) [meters]	Location (Y) [meters]	Strength Linear Cost [kg/s]
Gaussian	Strict Threshold	14	0.040	122	139	0.007
	Loose Threshold	12	0.029	95	113	0.008
SCIPuff	Strict Threshold	7	0.011	22	16	0.000
	Loose Threshold	5	0.011	64	181	0.003