7B.3 METHODS OF MITIGATING UNCERTAINTY IN CONTAMINANT DISPERSION IN A TURBULENT FLOW: DATA ASSIMILATION VS. MUTISENOSR DATA FUSION

Andrew J. Annunzio*, Sue Ellen Haupt, George S. Young
The Pennsylvania State University, University Park, Pennsylvania

1. INTRODUCTION

In atmospheric transport and dispersion (AT&D) there are two different methods of incorporating data into dispersion models: Data Assimilation (DA) and Multi-sensor Data Fusion (MSDF). These methods rely on the premise that sensors report concentration data, wind data, or both types within the same domain as the model. DA and MSDF both have the same goal: to incorporate data from sensors into a model to improve the prediction of an unknown (Hall 2004, Kalnay 2003). Although these methods share a common goal, their approaches to defining the unknown variables differ. In DA, the unknown is a field variable, while in MSDF the unknown is an entity. In some situations, one can pose the unknowns as either a field or an entity. In AT&D this is possible and one can implement either method because contaminant concentration can be considered a field variable or a contaminant filled puff can be considered an entity. The specification of a field or an entity changes the framework for incorporating data: with a contaminant puff, one takes a Langrangian approach to incorporating data, while for contaminant concentration one uses an Eulerian approach to incorporating data. While both methods have been used extensively in AT&D, it is instructive to directly compare and contrast their use for modeling a contaminant release in a turbulent flow, which is the purpose of this study.

Both MSDF and DA include several techniques by which data is incorporated into dispersion models. While the frameworks for the methods are disparate, the techniques available to incorporate the data are strikingly similar. This overlap enables comparison, and here we compare the most basic DA technique, Newtonian Relaxation (Nudging), with the most basic MSDF technique, Alpha Filtering. These techniques are the most basic because model computation time increases marginally when augmenting the model concentration field prediction (DA) or puff characteristics (MSDF) with an observation.

2. PROCEDURES

Our objective is an unbiased comparison between the Nudging and Alpha Filtering techniques through an analytic formulation. We assume stationary, homogeneous turbulence to maintain consistency between frameworks, equations that govern puff dispersion under these meteorological conditions remain the same. Further, the unknowns for each remain the same: the two-dimensional source location and total integrated mass of the puff.

2.1 Data Assimilation: Nudging

Data Assimilation techniques combine information from all available resources to better predict a field variable (Kalney 2003). DA techniques come in two forms, static and dynamic techniques. (Daley 1991, Lewis et al 2006, Kalnay 2003). Static techniques describe the field at a certain time and do not aid in numerical weather prediction (Lewis et al, 2006). Dynamic techniques, on the other hand, are implemented in numerical weather prediction and estimate the appropriate solution of a field variable at a later time (Lewis et al, 2006). Within dynamic techniques exist sequential estimation, in which a model is updated in parallel with the observations. This is done by augmenting the governing equations with an innovation vector consisting of a function multiplied by the difference between the model and observation at each grid point in the domain.

For the comparison we consider the most basic sequential estimation technique, DA Nudging. This technique is most basic because the function in the innovation vector becomes a constant 1/τ. For this formulation, the specification of concentration as a field variable leads to the scalar conservation equation in turbulent flow. In order to simplify this analysis we require that the sensors lie on the grid points so that interpolation techniques are not necessary. We allow the equation to match the Partial Differential Equation (PDE) that produces the Gaussian puff model augmented with the innovation vector. For this situation, the governing equation describing the evolution of a passive contaminant is given by

$$\frac{\partial C}{\partial t}+U\frac{\partial C}{\partial x}-K_x\frac{\partial^2 C}{\partial x^2}-K_y\frac{\partial^2 C}{\partial y^2}-K_z\frac{\partial^2 C}{\partial x^2} = \frac{1}{\tau}(\vec{C}_{obs}(t)g(x,y,z,t)-C) \quad (1)$$

where U is the mean wind speed, x, y, and z are the spatial variables, t is time, K_x, K_y, and K_z are eddy diffusivities in the *x, y,* and *z* directions respectively, τ is the nudging coefficient, $\vec{C}_{obs}(t)$ accounts for the observations at the grid points at a certain time, and $g(x,y,z,t)$ is a function that allows the observations to influence other grid points throughout the domain, forming the innovation. These observations act as a source for concentration values at sensors throughout

* Corresponding author address: Andrew J. Annunzio, The Pennsylvania State University, Department of Meteorology and Applied Research Lab, 503 Walker Building, University Park PA, 16803. E-mail: aja199@psu.edu

the domain. In order to solve (1) we first obtain a Green's function from the homogeneous equation and then add the effects that the initial conditions and source terms have on the Green's function to obtain the nudged concentration equation.

$$
\begin{aligned}
C_\tau = & \iiint_\Omega \frac{f(x,y,z)}{U(K_x K_y K_z)^{1/2}(4\pi)^{3/2}(t)^{3/2}} \exp\left(-\frac{(x-x'-Ut)^2}{4K_x t}\right) \\
& \exp\left(-\frac{(y-y')^2}{4K_y t}\right)\exp\left(-\frac{(z-z')^2}{4K_z t}\right)\exp\left(-\frac{t}{\tau}\right)dx'\,dy'\,dz' \\
& \iiint_\Omega \int_{t_{ob}}^{t_f} \frac{\vec{C}_{obs}(t)g(x,y,z,t)}{(K_x K_y K_z)^{1/2}(4\pi)^{3/2}(t-t')^{3/2}} \exp\left(-\frac{(x-x'-U(t-t'))^2}{4K_x(t-t')}\right) \\
& \exp\left(-\frac{(y-y')^2}{4K_y(t-t')}\right)\exp\left(-\frac{(z-z')^2}{4K_z(t-t')}\right)\exp\left(\frac{t-t'}{\tau}\right)dx'\,dy'\,dz'\,dt'
\end{aligned}
\tag{2}
$$

where Ω represents the spatial domain and all other variables are the same as previously defined. The unknowns in this problem are introduced through the initial condition which is given by,

$$
f(x,y,z) = \frac{M_u}{U}\,\delta(x - x_{o,u}, y - y_{o,u}, z - z_o) \tag{3}
$$

where δ is the Kronecker delta function, $x_{o,u}$ and $y_{o,u}$ represent the guessed initial source location and M_u represents the guessed mass of the puff. To complete the analytic formulation for Nudging, we then compute the first integral in (2) to obtain,

$$
\begin{aligned}
C_\tau = & \frac{M_u}{U(K_x K_y K_z)^{1/2}(4\pi)^{3/2}(t)^{3/2}} \exp\left(-\frac{(x-x_{o,u}-Ut)^2}{4K_x t}\right) \\
& \exp\left(-\frac{(y-y_{o,u})^2}{4K_y t}\right)\exp\left(-\frac{(z-z_o)^2}{4K_z t}\right)\exp\left(-\frac{t}{\tau}\right) + \\
& \iiint_\Omega \int_{t_{ob}}^{t_f} \frac{\vec{C}_{obs}(t)g(x,y,z,t)}{(K_x K_y K_z)^{1/2}(4\pi)^{3/2}(t-t')^{3/2}} \exp\left(-\frac{(x-x'-U(t-t'))^2}{4K_x(t-t')}\right) \\
& \exp\left(-\frac{(y-y')^2}{4K_y(t-t')}\right)\exp\left(-\frac{(z-z')^2}{4K_z(t-t')}\right)\exp\left(\frac{t-t'}{\tau}\right)dx'\,dy'\,dz'\,dt'
\end{aligned}
\tag{4}
$$

Then, if one considers $g(x,y,z,t)$ to be Gaussian, then (4) can be integrated to obtain

$$
\begin{aligned}
C_\tau = & \frac{M_u}{(K_x K_y K_z)^{1/2}(4\pi)^{3/2}(t)^{3/2}} \exp\left(-\frac{(x-x_{o,u}-Ut)^2}{4K_x t}\right) \\
& \exp\left(-\frac{(y-y_{o,u})^2}{4K_y t}\right)\exp\left(-\frac{(z-z')^2}{4K_z t}\right)\exp\left(-\frac{t}{\tau}\right) + \\
& \int_{t_{ob}}^{t_f} \frac{\vec{C}_{obs}(t)}{(K_x K_y K_z)^{1/2}(4\pi)^{3/2}(t-t_{ob})^{3/2}} \exp\left(-\frac{(x-x_{ob}-U(t-t_{ob}))^2}{4K_x(t-t_{ob})}\right) \\
& \exp\left(-\frac{(y-y_{o,u})^2}{4K_y(t-t_{ob})}\right)\exp\left(-\frac{(z-z')^2}{4K_z(t-t_{ob})}\right)\exp\left(-\frac{t-t'}{\tau}\right)dt'
\end{aligned}
\tag{5}
$$

.

Equation (5) predicts surface concentration values via a decaying Gaussian added with a superposition of decaying Gaussians multiplied by input from observations.

2.2 Multi-Sensor Data Fusion: Alpha Filtering

MSDF shares the same goal as DA and combines data from multiple sensors to better characterize an entity (Hall 2004). Similar to DA, MSDF techniques fall into two categories, batch and sequential estimation (Hall 2004). Batch estimation is an offline technique, while sequential estimation is an online technique. Unlike Data Assimilation, for sequential techniques in MSDF rather than augmenting the governing equation with an innovation vector, one instead updates an unknown after a prediction of the variable is attained (Hall 2004).

Traditionally, the Alpha Filter is designed to update the position of an entity. Here we extend the filter to update characteristics of an entity as well (Painter et al 1990). For our formulation of Alpha Filtering, a PDE no longer needs to be solved. The problem is posed as one in optimization: puff characteristics are optimized given contaminant concentrations. The characteristics of the puff that we wish to ascertain are total integrated mass of the puff and the two-dimensional puff location at each time step. In order to derive these characteristics, we fit a Gaussian puff to the data. Note that with the assumption of stationary, homogeneous turbulence, and aligning our x coordinate with the mean wind direction, we only need to find the downwind location of the puff when optimizing the two-dimensional location. We pose the optimization problem in terms of least squares, taking derivatives with respect to the unknowns of the squared difference between the observations and the values computed from the Gaussian puff equation. Because the Gaussian puff equation involves exponential terms, it is easier to take the natural logarithm of both.

$$
\begin{aligned}
& \frac{\partial}{\partial(x_o + Ut)}\left(\left(\ln(C_{obs}) - \ln(C(M,(x_o+Ut))\right)^2\right) \\
& \frac{\partial}{\partial(M)}\left(\left(\ln(C_{obs}) - \ln(C(M,(x_o+Ut))\right)^2\right)
\end{aligned}
\tag{6}
$$

where $C(M,(x_o + Ut)$ is given by

$$
\begin{aligned}
C(M,(x_o+ut)) = & \frac{M}{(K_x K_y K_z)^{1/2}(2\pi t)^{3/2}U} \\
& \exp\left(\frac{(x-(x'+Ut))^2}{4K_x t}\right)\exp\left(\frac{(y-y')^2}{4K_y t}\right)\exp\left(\frac{(z-z_o)^2}{4K_z t}\right)
\end{aligned}
\tag{7}
$$

The alpha filter updates the guessed characteristics with the ascertained characteristics at each time step with an equation given as a linear combination of the two (Hall 2004):

$$M_f = M_g + \alpha(M_a - M_g)$$
$$(x_o + Ut)_f = (x_o + Ut)_g + \alpha\left((x_o + Ut)_a - (x_o + Ut)_g\right) \quad (8)$$

where the subscript *g* denotes the guessed characteristic, the subscript *a* indicates the ascertained characteristic, and the subscript *f* designates the final characteristic, and *α* is a constant whose value is between zero and one and plays a similar role to the *τ* in nudging. To complete the analytic formulation for the Alpha Filter, we replace the unknowns in the Gaussian puff equation with (8) to obtain,

$$C_\alpha = \frac{M_f}{(K_x K_y K_z)^{1/2}(2\pi t)^{3/2}U} \exp\left(-\frac{\left(x-(x'+Ut)_f\right)^2}{4K_x t}\right)\exp\left(-\frac{(y-y')^2}{4K_y t}\right)\exp\left(-\frac{(z-z_o)^2}{4K_z t}\right) \quad (9)$$

Surface concentration values are predicted by (9), which is a Gaussian puff equation with updated unknowns.

3. RESULTS

The final piece of the puzzle needed to enable the comparison is data. Here, we implement an identical twin experiment to compute the data, where the forward model creates the observations (Daley 1991). This is advantageous for model development as we can check whether our analytic methods are working properly. Because contaminant concentration will not exactly follow a Gaussian distribution, we add white clipped Gaussian noise to the truth data to account for atmospheric and sensor noise. This is done for several signal to noise ratio (SNR) levels. The noisy data is then taken as the truth in the experiment.

The result of the comparison is given in Table 1. As one can see, the Alpha Filter outperforms Nudging for all SNRs tested. For Nudging, the RMSE remains nearly constant for every case. This occurs as the function $g(x, y, z, t)$ spreads the observations out over the domain, and therefore acts to spread and smooth the noise. Thus, one should not expect the RMSE for Nudging to change significantly as the SNR increases. The same notion is not true for the Alpha Filter, as the Alpha Filter predicts surface concentration values with a Gaussian model; therefore it is expected that accuracy should deteriorate with increasing noise. When the SNR is infinite, the optimized values of the unknowns are very accurate and thus the Alpha Filter almost exactly reproduces the data. As the SNR decreases, the RMSE remains smaller than the RMSE for Nudging implying that the optimization is still relatively successful implying more success for the Alpha Filter than Nudging.

Table 1: RMSE at the last time step summed over all the grid points

Signal to Noise Ratio	INF	100	50	10	5	2
RMSE for Nudging (10^-9)	.430	1.28	.430	1.11	.530	.866
RMSE for Alpha Filtering (10^-9)	0.0	.00083	.0079	.016	.047	.079

4. CONCLUSIONS

This work has analytically and numerically compared and contrasted the methods of DA and MSDF for incorporating data into models using Gaussian puff transport and dispersion as an example. The Eulerian framework of DA treats the concentration values as a field variable and leads to solving a PDE. In contrast, the Lagrangian framework of MSDF casts the puff as an entity with features that can be optimized. When the optimization is approached analytically with a least squares technique, one obtains an expression for a puff characteristic. The comparison results in the Alpha Filtering being champion over Nudging. Because Alpha Filtering and Nudging act similarly but within formulations suited to their specific framework, the conclusion is not that Alpha Filtering is better than Nudging, but that the entity formulation is better in this situation than the field approximation. In future research we plan to extend the analytic comparison to other MSDF and DA techniques as well as to progress beyond the identical twin experiment to incorporate real data. The latter extension will require numeric formulations to the problem instead of analytics. Further, we wish to explore the implications of the synergistic application of entity and field frameworks in AT&D modeling.

ACKNOWLEDGEMENTS

This work was funded by DTRA under contract number W911NF-06-C-0162 and DTRA01-03-D-0010-0012.

REFERENCES

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

Hall, D. and McMullen, S., 2004: *Mathematical Techniques in Multisensor Data Fusion.* Artech House, INC., Norwood, MA 02062

Kalnay, E,. 2003: *Atmospheric modeling, data assimilation and predictability.* Cambridge Univ. Press, 341 pp.

Lewis, J. Lakshmivarahan, S. and Dhall, S., 2006: *Dynamic Data Assimilation: A Least Squares*

Approach. Cambridge University Press, New York, NY, 651 pp.

Painter, J., Kerstetter, D., and Jowers, S. "Reconcilling Steady-State Kalman and Alpha-Beta Filter Design." IEEE Transactions on Aerospace and Electronic Systems Vol. 26, No. 6, Nov. 1991, (986-991)