3.13 ON USING DATA ASSIMILATION IN DISPERSION MODELING

Anke Beyer*, George S. Young, Sue Ellen Haupt
The Pennsylvania State University, University Park, Pennsylvania

1. INTRODUCTION

Transport and dispersion models are important tools for addressing the issue of a chemical, biological, radiological, or nuclear (CBRN) release. Since the 1950's transport and dispersion modeling with application to CBRN emergencies is a topic of intense research (Fast et al. 1995). The Chernobyl nuclear accident, for example, promoted the development of new long-range transport models and applications of existing air quality models to radioactivity in many research facilities (Ishikawa 1994). A number of field experiments have been conducted in the last 40 years, for example the 1980 Great Plains Mesoscale Tracer Field experiment (Moran and Pielke 1995) and the Dipole Pride experiments (Watson et al. 1998).

The dispersion of hazardous puffs or plumes is the result of three factors: the transport by the wind field, dispersion by turbulence, and chemical reactions. The basic data requirements for a dispersion modeling system are hazard source characterization, surface topography, and meteorological data. Using these three elements, and the equations of transport, dispersion, and chemical reaction, it becomes possible to accurately forecast the spread and evolution of hazardous materials. Currently available modeling systems range from relatively simple to highly complex (Arya 1999). A simple example is the Gaussian plume model, which has been used for almost a century to predict dispersion from continuous point sources in air quality applications (Weil et al. 1992).

2. DATA

Observations play a key role in numerical weather prediction (NWP) as well as in transport and dispersion modeling. In operational NWP over 11,000 land surface observations, over 7,000 marine surface observations and approximately 900 upper air soundings are available each day (WMO 2005). This conventional observational network is augmented by satellites, radar and aircraft. With the average scale of synoptic eddies being approximately 2000 km and their evolution measured in days, the data coverage and the time resolution for NWP forecasts is generally sufficient.

In contrast; the characteristic length scale of a fatal hazardous release is much smaller than that of a synoptic disturbance: instead it will interact most strongly with the boundary layer eddies. These eddies scale with the boundary layer depth (Kaimal et al. 1976) so their typical length scales range from 1 to 5 km and their evolution takes less than an hour. Thus, much higher spatial and temporal sampling rates are required to map a hazardous release and its interaction with the dominant eddies. As with NWP, it is expected that several sensors will be needed per eddy to successfully initialize a model of the transport and dispersion of a plume. Unfortunately, the current State and Local Monitoring Network for air quality in the USA contains only approximately 4,000 fixed monitoring stations that measure criteria pollutants (particulate matter, sulfur dioxide, carbon monoxide, nitrogen dioxide, ozone, and lead) (EPA). An additional 500 air-sampling devices have been deployed in 31 US cities in the framework of BioWatch, a program of the Department of Homeland Security (Bridis 2003). The BioWatch sensor network detects biological agents by combining state-of-the-art laboratory testing and traditional filter sampling methods (Bridis 2003). But, even with multiple mobile instruments being rapidly deployed to the site of events there is no guaranteeing that all relevant turbulent eddies are observed. Also, the traditional types of data collection used in the EPA and BioWatch network have the disadvantage of a long sampling time. In an immediate emergency situation real-time data is needed.

An alternative to gridded ground based measurements are Unmanned Aerial Vehicles (UAVs). UAVs share the advantage of an aircraft in rapid deployment to arbitrary locations and heights, but reduce the cost and safety restrictions a piloted plane would present (Watai et al. 2006).

* *Corresponding author address:* Anke Beyer, Dept. of Meteorology, 503 Walker Building, Pennsylvania State Univ., University Park, PA 16802-5013; e-mail: aub166@psu.edu.

Moreover, the UAV can 'spaghetti-sample' a four dimensional volume efficiently.

Remote sensing provides another option to gather observational data for dispersion applications. The radar can measure wind velocities and differentiate the precipitation type (Serafin and Wilson 2000), enabling the user to infer the transport and deposition of pollutants. Lidar systems can estimate the turbulent mixing depth, by using particulate matter and hydrometeors as tracers (Beran and Hall 1974).

An advantage of many remote sensing devices is their ability to rapidly scan large regions in three dimensions (Beran and Hal 1974). But the difference in the output data of remote sensors and UAVs compared to the classical gridded surface observations must be considered in the design of a data assimilation system. The assimilation method applied to the transport and dispersion problem should be able to continuously incorporate observational data into the model.

3. PROCEDURES

Data Assimilation provides a methodology to deal with under-resolved phenomena for which a model is to be run. Data assimilation techniques allow combining all available information, with observations possibly sampled at different times and intervals and different locations into a unified description of the system consistent with the model physics (Kalnay 2003). This analysis can then be used to initialize forecast models. In the last three decades the field has evolved to where complex statistical and constrained optimization techniques are used for the initialization of operational NWP forecast models (Ide et al. 1997; Houtekamer and Mitchell 1998). These and other assimilation methods now have to be tested and evaluated in the context of transport and dispersion models.

The need for these tests becomes clear looking at Figure 1. Barnes successive corrections algorithm* has been used to assimilate a 500 hPa geopotential height field and a Gaussian plume based on the same number of true data points and the same number of random observations. A comparison suggests that a basic function fitting assimilation method such as the successive corrections algorithm attuned for NWP does not perform as well for the dispersion problem.

Some of the more advanced, frequently used data assimilation techniques in NWP applications include optimal interpolation (Gandin 1963), three dimensional (Sasaki 1970) and four dimensional variational analysis (3D- and 4D-Var) (Lewis and Derber 1985), extended Kalman filtering (Kalman and Bucy 1961), ensemble Kalman filtering (Houtekamer and Mitchell 1998), and Newtonian relaxation (data nudging) (Hoke and Anthes 1976). For our purposes, i.e. the prediction of transport and dispersion of a CBRN release, a flexible assimilation method is needed, not just to improve the accuracy of a forecast with observational data, but to use the CBRN concentration data to modify and correct the predicted wind field and vice versa.

As mentioned before, time dependent observational data requires an assimilation method that is able to assimilate continuously. Both of the Kalman filter techniques, 4D-Var, and Nudging seem to fit our needs. For further evaluation, Table 1 outlines the methodology, possibilities for dispersion modeling applications, advantages, and disadvantages of these three data assimilation techniques.

Four dimensional variational analysis is an extension of the 3D-Var that allows use of observations distributed within a time interval (Kalnay 2003). In NWP applications the prediction model is used as a strong constraint, but numerous other constraints are possible, for example, temporal smoothing (Daley 1991). The cost function is evaluated by integrating the full nonlinear model forward in time. Then the adjoint model is integrated backwards in time to determine the variation in the cost function. This information is used iteratively in a descent algorithm. 4D-Var can assimilate modeled and non-modeled variables, but is computationally expensive. It requires the derivation of the adjoint model and *a priori* knowledge of the forecast error covariance, which is the most difficult error covariance to estimate and has a crucial impact on the accuracy of the forecast.

Unlike 4D-Var, the Kalman Filter technique assimilates the observations sequentially each time step (Caya et al. 2005). The Extended Kalman Filter is a variant of the Kalman Filter that can be used for nonlinear problems (Miller et al. 1994) and is hence suitable for transport and dispersion applications. The forecast error covariance is advanced in time using the model itself rather than estimating it prior to the assimilation step. Lacking a reasonable assumption for linking the error covariance matrix to eddy size and mixing length, this fact might

* Barnes successive corrections algorithm is a basic fitting technique with *a priori* weights decreasing with each successive iteration cycle and with distance between observation and analysis (Daley, 1991).

come in handy for dealing with transport and dispersion models. But, like 4D-Var, the Extended Kalman Filter is computationally expensive. A promising simplification is the Ensemble Kalman Filter (Kalnay 2003). In this approach an ensemble of data assimilation cycles is carried out simultaneously. The Ensemble Kalman Filter does not specifically integrate the forecast error covariance, but instead computes it diagnostically from the spread of the model states across the ensemble.

Nudging is a computationally efficient technique, which relaxes the model state toward the observations by directly adding artificial tendency terms to the prognostic equations (Stauffer and Seaman 1993). The relaxation time scale, which acts as a proportionality constant in the tendency terms is always positive and has to be chosen based on scaling arguments so that the nudging tendencies are relatively small compared to the other terms in the prognostic equations (Stauffer and Seaman 1993). Nudging requires *a priori* knowledge of the forecast error covariance to correctly estimate the relaxation timescale and assimilates only those variables that are explicitly modeled.

A hybrid data assimilation technique may help reduce the computational cost without a drastic impact on the quality of the forecast, by using the advantages of certain methods, but avoiding the disadvantages. Hamill and Snyder (2000) demonstrated how to construct an Ensemble Kalman Filter and 3D-Var hybrid analysis scheme. Lili Lei (personal communication) suggested an Ensemble Kalman filter – Nudging hybrid technique, which uses the diagonal of the error covariance matrix computed with the Ensemble Kalman Filter to estimate the relaxation timescale, then used to nudge the forecast towards the observations.

Given the unresolved nature of puffs or plumes in transport with expected fixed sensor densities a data assimilation strategy has to be developed for transport and dispersion models. While remote sensors and UAVs can provide greater observation densities, their asynchronous sampling strategy imposes further requirements on the assimilation strategy. The methods outlined above can handle both low sensor density and asynchronous observations. What remains is to develop a methodology for testing and tuning them in the dispersion and transport model initialization problem. The choice of method and tuning is expected to vary with the sophistication of the transport and diffusion model. A Gaussian Plume model (Arya 1999), for example, is steady state and so cannot make use of the time dependence of 4D-Var or Kalman Filtering. In contrast, a Gaussian Puff (Arya 1999) or eddy-resolving particle tracking model (Böning and Cox 1988) can assimilate time-varying observations via either of these techniques.

The Gaussian Puff model for an instantaneous release suggests itself for initial testing as it is the simplest time-dependent transport and dispersion model. Being the oldest and simplest example of time dependent ensemble-averaged models, it requires a minimum of input information and is easily implemented. Keeping in mind that computational resources are limited in real-time CBRN situations, Nudging or the Ensemble Kalman Filter-Nudging-Hybrid seem likely to be the best assimilation techniques. The testing approach will be to measure the accuracy of the analysis at the end of the observation interval as a function of observation density in space and time. Comparison will be facilitated by normalizing the time and space scales by the puff half-width and half-width transit time (past an observation site) at the end of the observation interval. Subsequent work will involve more sophisticated time-dependent models such as SCIPUFF (Sykes et al. 1998).

4. CONCLUSIONS

To evaluate emergency response measures in the case of a CBRN release, the prediction of the transport and dispersion of these hazardous materials is necessary. *In situ* observations are most likely to be sparse, so optimum use must be made of available data. Three data assimilation techniques have been proposed to be suitable for dealing with the under resolved observational data. The advantages and disadvantages of the assimilation methods have been discussed and Nudging or a Hybrid technique are most likely to be most adequate for a CBRN scenario. Although the Kalman Filter and 4D-Var are more complex and advanced methods, for immediate emergency response computational time is crucial.

For further assessment of the proposed techniques, a series of tests has to be conducted. Identical twin experiments provide an ideal first level of testing, allowing the methodologies to be tested and compared in a tightly controlled situation. In this type of experiment the test data is created using the same dynamical model used for the simulation, eliminating one source of uncertainty. In particular, the amount and character of noise in the observational data can be controlled in an identical twin experiment, whereas it cannot in real-world testing. Thus, identical twin experiments will allow the robustness of each method to be computed in the face of varying amounts of observational data and varying degrees of noise in that data. Once the best system for the expected data environment has been selected and tuned, its performance can be further evaluated in second-level tests involving real-world observations from experimental releases of non-hazardous tracers.

Acknowledgements. This research is supported by DTRA (grant number W911NF-06-C-0162). The Authors would like to express their thanks to Kerrie J. Long and Lili Lei for helpful discussions relating to the implementation of several data assimilation techniques and the possibilities of hybrid methods.

5. REFERENCES

Arya, S.P., 1999: Air Pollution Meteorology and Dispersion. Oxford University Press, Oxford.

Beran, D.W., F.F. Hall, Jr., 1974: Remote Sensing for Air Pollution Meteorology. *Bull. Amer. Meteor. Soc*., **55**, 1097-1105.

Böning, C.W., and M.D. Cox, 1988: Particle Dispersion and Mixing of Conservative Properties in an Eddy-Resolving Model. *J. Phy. Ocean.*, **18**, 320-338.

Bridis, T., 2003: 'Biowatch' sensors could save thousands of lives. *ASSOCIATED PRESS/Miami Herald*, 15th Nov. 2003

Caya, A., J. Sun, and C. Snyder, 2005: A Comparison between 4D Var and the Ensemble Kalman Filter Technique for Radar Data Assimilation. *Mon. Wea. Rev*., **133**, 3081-3093.

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

Fast, J.D., B.L. O'Steen, and R.P. Addis, 1995: Advanced Atmospheric Modeling for Emergency Response. *J. Appl. Meteor*., **34**, 626-648.

Gandin, L. S., 1963: Objective analysis of meteorological fields. Gidrometeor. Isdaty., Leningrad. [Israel Program for Scientific Translations, Jerusalem, 1965, 242 pp.]

Hamill, T.H., and C. Snyder, 2000: A Hybrid Ensemble Kalman Filter–3D Variational Analysis Scheme. *Mon. Wea. Rev*., **128**, 2905-2919.

Hoke, J.E., and R.A. Anthes, 1976: The Initialization of Numerical Models by a Dynamic Initialization Technique. *Mon. Wea. Rev*., **104**, 1551-1556.

Houtekamer, P.L., and H.L. Mitchell, 1998: Data Assimilation Using an Ensemble Kalman Filter Technique. *Mon. Wea. Rev*., **126**, 796-811.

Ide, K., P. Courtier, M. Ghil, and A.C. Lorenc, 1997: Unified Notation for Data Assimilation: Operational, Sequential and Variational. *J. Met. Soc. Japan*, **75**, 181-189.

Ishikawa, H., 1995: Evaluation of the Effect of Horizontal Diffusion on the Long-Range Atmospheric Transport Simulation with Chernobyl Data. *J. Appl. Meteor*., **34**, 1653-1665.

Kaimal, J.C., J.C. Wyngaard, D.A. Haugen, O.R. Coté, Y. Izumi, S.J. Caughey, and C.J. Readings: Turbulence Structure in the Convective Boundary Layer. *J. Appl. Meteor*., **33**, 2152-2169.

Kalman, R., and Bucy, R., 1961: New results in linear prediction and filtering theory. Trans. AMSE, *J. Basic Eng*., **83D**, 95–108.

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

Lewis, J.M., and J.C. Derber, 1985: The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, **37**, 309-327.

Miller, R.N., M. Ghil, and F. Gauthiez, 1994: Advanced Data Assimilation in Strongly Nonlinear Dynamical Systems. *J. Atmos. Sci*, **51**, 1037-1056.

Moran, M.D., and R.G. Pielke, 1996: Evaluation of a Mesoscale Atmospheric Dispersion Modeling System with Observations from the 1980 Great Plains Mesoscale Tracer Field Experiment. Part I: Datasets and Meteorological Simulations. *J. Appl. Meteor*., **35**, 281-307.

Sasaki, Y., 1970: Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev*., **98**, 875-883.

Serafin R. J., and J. W. Wilson, 2000: Operational weather radar in the United States: Progress and opportunity. *Bull. Amer. Meteor. Soc*., **81**, 501–518.

Stauffer, D.R., and N.L. Seaman, 1993: Multiscale Four-Dimensional Data Assimilation. J*. Appl. Meteor*., **33**, 416-434.

Sykes, R. I., S. F. Parker, D. S. Henn, C. P. Cerasoli, and L. P. Santos, 1998: PC-SCIPUFF version 1.2PD technical documentation. ARAP Rep. 718, Titan

Research and Technology Division, Titan Corp., Princeton, NJ, 172 pp. available at: http://www.titan.com/products-services/336/download_scipuff.html

Watai, T., T. Machida, N. Ishizaki, and G. Inoue, 2006: A Lightweight Observation System for Atmospheric Carbon Dioxide Concentration Using a Small Unmanned Aerial Vehicle, *J. Atmosph. Ocean. Tech*., **23**, 700-710.

Watson, T. B., R. E. Keislar, B. Reese, D. H. George, and C. A. Biltoft, 1998: The Defense Special Weapons Agency Dipole Pride 26 field experiment. NOAA Air Resources Laboratory Tech. Memo. ERL ARL-225, 90 pp.

Weil, J.C., R.I. Sykes, and A. Venkatram, 1992: Evaluating Air-Quality Models: Review and Outlook. *J. Appl. Meteor*., **31**, 1121-1145.

World Meteorological Organization (WMO), 2005: Twenty-Second Status Report on Implementation of the World Weather Watch. WMO-No. 986 (2005).

6. FIGURES AND TABLES

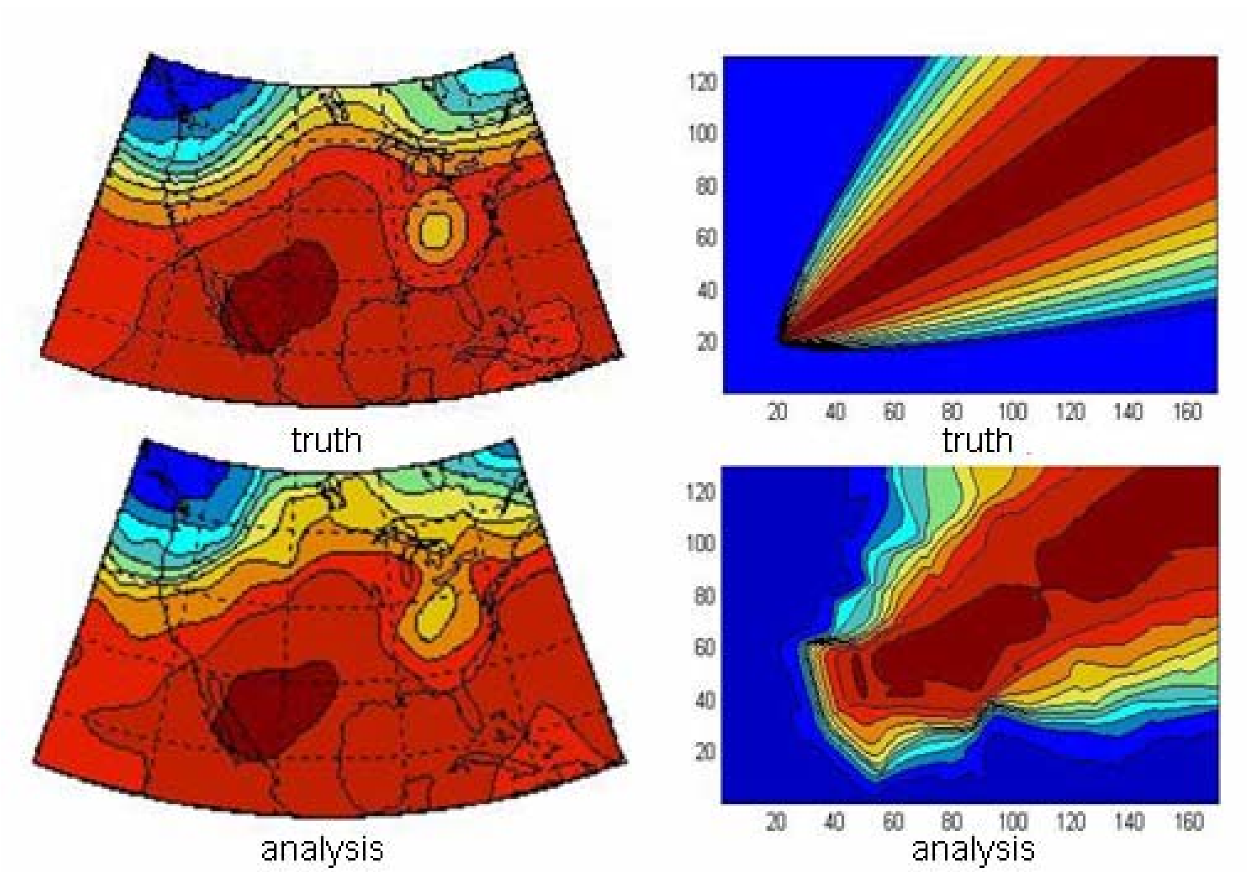

Figure 1: Barnes successive corrections algorithm was used to compute the analysis with 110 random observations taken from the 22,000 true data points. Right hand side: 500 hPa geopotential height field in dm; left hand side: Gaussian plume of pollutant. First characteristic length scale is 50 data points, γ= 0.99 and 4 iterations where used.

Table 1: Comparison between data assimilation methods.

	4D VAR	Kalman Filter	Nudging
How It Works	• Approaches the data filtering problem by updating a time dependent background field • Model produced background field • Model is run forward and backward, gradually adjusting the initial condition • until the model run fits the observations	• Approaches the data filtering problem by updating a time dependent background field • Model produced background field • Subtract an optimized fraction of the model error at each time step • Many variations proposed for improving generality and efficiency	• Approaches the data filtering problem using the model itself • Add term to the model's time dependent budget equations • Subtract a fixed fraction of model error at each time step • Modeled fields gradually adjust to observations
Dispersion Modeling Application	• Can jointly assimilate observations of meteorological and concentration variables • Can assimilate observations which are sparse in space and time	• Can assimilate meteorological and concentration data • Suitable for long range forecasting • Maintains the quality of a forecast over time	• Can assimilate meteorological and concentration data • Suitable for long range forecasting • Maintains the quality of a forecast over time
Advantages	• Assimilates modeled and non-modeled variables • The complete model is used as a constraint • Other constraints can be applied • Older observations retain value to the analysis	• Actual model dynamics used to constrain analysis • Able to assimilate asynchronous observations of multiple variables • Computes time dependant background error covariance using model itself	• Actual model dynamics used to constrain analysis • Able to assimilate asynchronous observations of multiple variables • Does not require knowledge of error covariances • Simple to implement • Computationally efficient
Disadvantages	• Requires derivation of the model's adjoint • Background error covariance must be known • Computationally expensive	• Requires cleverness to compute error covariance accurately • Computationally expensive even when elegant mathematics are applied to improve efficiency	• Assimilates only those variables which are explicitly modeled • Requires reasonable background field or long integration to produce a good analysis