DEVELOPMENT AND FIRST TESTS OF A DATA ASSIMILATION ALGORITHM IN A LAGRANGIAN PUFF ATMOSPERIC DISPERSION MODEL

V. Tsiouri [1,3], I. Kovalets [2], S. Andronopoulos [1], J.G. Bartzis [3]
[1]Environmental Reasearch Laboratory, Institute of Nuclear Technology and Radiation Protection, NCSR 'Demokritos', 15310 Aghia Paraskevi Attikis, Greece
[2]Department of Environmental Modelling, Institute of Mathematics Mashine & System Problems, National Academy of Sciences of Ukraine, pr. Glushkova-42, Kiev 03187, Ukraine
[3]Department of Engineering and Management of Energy Resources, University of Western Macedonia, Bakola & Sialvera str., 50100, Kozani, Greece

INTRODUCTION

Data assimilation (DA) in atmospheric dispersion models has recently become one of the most challenging problems. Data assimilation aims at improving a model's predictions by merging measurements with model results in an optimal way. The basic types of the data assimilation approaches used in the ADMs are the same as of those used in the NWP models, i.e., the algorithms based on: Kalman Filtering (KF), or Ensemble Kalman Filtering (EnKF) approach (*Segers, 2002*) and the variational algorithms, which often lead to the adjoint equations (e.g., *Le Dimet and Talagrand, 1986*). In variational data assimilation one attempts to find optimal parameters (the so-called control variables, for instance initial conditions of the model), that minimize a discrepancy between model results and measurements for a chosen analysis period.

The DA procedures for the ADMs were mainly implemented in the Eulerian models (e.g., *Segers, 2002*). From the literature it can be seen an obvious lack of data assimilation algorithms applied in puff and Lagrangian particle models, which are widely used in the emergency response systems. In Lagrangian (and "puff") models (e.g. *Bartzis et al., 2000*) the pollutant is emitted in parcels which move with the local wind speed. The concentration is calculated by summing the contribution of all parcels. In *Astrup et al.* (2004) the Extended Kalman Filtering (EKF) approach was used. The possibilities to overcome some difficulties using the Ensemble Kalman Filtering approach were studied by *Zheng et al.,* (2006). The paper of *Fisher and Lary* (1995) concerned the implementation of the variational DA approach in a Lagrangian particle model. In the works of *Jeong et al.*, (2005), and *Quelo et al,* (2005), the variational methodologies used are capable to deal only with constant wind speeds and constant source rate. In the work of *Drews et al.*, (2005), that implements the Extended Kalman Filtering approach, the case of variable source function and constant wind speed is considered.

The present work is concerned with extending the variational methodology to the puff model, which in principal is capable of dealing with variable source function and wind speed. The developed methodologies are closely related to the work of *Kathiragamatan et al.*, (2004). He developed a variational approach to a Eulerian model. However, his method is limited to constant wind speed. Other important differences with the present approach are in the cost functional formulation. In the present work we aim at estimating the unknown source emission rate. Tests with constant source function and constant wind speed as well as with several forms of source term functions and variable wind speed are performed. The results showed that the formulation of the methodology is more general and can be extended to take into account variable meteorological conditions. In addition, other assumptions i.e., uncorrelated errors of source function estimation, root mean square error of the observations either constant or proportional to the values of the observed concentration, are also discussed.

DESCRIPTION OF METHOD

The problem of atmospheric dispersion forecast with the assimilation of the data of available concentration measurements is considered. In the model used in this study (a simplified version of the DIPCOT model, *Bartzis et al.*, 2000) the concentration C^M at a point with coordinates (X, Y, Z) is calculated as the contribution of all the L puffs present in the computational domain:

$$C^M(X,Y,Z) = \frac{1}{(2p)^{3/2}} \sum_{p=1}^{L} \frac{q_p \Delta t}{s_{xp} s_{yp} s_{zp}} \exp\left[-\frac{1}{2}\frac{\left(X_p - X\right)^2}{s_{xp}^2}\right] \exp\left[-\frac{1}{2}\frac{\left(Y_p - Y\right)^2}{s_{yp}^2}\right]$$
$$\left\{\exp\left[-\frac{1}{2}\frac{\left(Z_p - Z\right)^2}{s_{zp}^2}\right] + \exp\left[-\frac{1}{2}\frac{\left(Z_p + Z - 2Z_g\right)^2}{s_{zp}^2}\right]\right\} \qquad (1)$$

Where (X_p, Y_p, Z_p) are the coordinates of the centre of each puff p, (s_{xp}, s_{yp}, s_{zp}) are the standard deviations of the distribution of matter in each puff, representing the puff size in the alongwind, crosswind and vertical directions respectively and Z_g is the ground height at the location (X, Y). M_p is the "load" (e.g., mass, radioactivity) of each puff. It is assumed that the puffs are released at regular time intervals Δt, q_p is the (average) source emission rate during the release of puff p.

In atmospheric dispersion the pollutant emission rate is of particular importance since the pollutant concentration is linearly dependent on it. However, in the event of an accidental hazardous pollutant release usually there is a large uncertainty regarding the time evolution and the magnitude of the source term. Based on the above facts the vector of control variables in the present study has been selected to include the source emission rates corresponding to the release of each model puff. By the above selection it is implied that other parameters, such as the duration of the release or the puffs coordinates, are assumed known and they are not adjusted by the data assimilation procedure.

More specific, the model-calculated concentrations C_{nk}^M (the measurement locations $1 = k = K$, and the measurement times $1 = n = N$) are defined as:

$$\mathbf{C}^M = \mathbf{G} \cdot \mathbf{q} \qquad (2)$$

$\mathbf{G}$ is a matrix with dimensions $[(N \times K) \times L]$, such as its i^{th} row is defined as follows:

$$\mathbf{G}(i,*) = \mathbf{R}_{nk} \qquad (3)$$

Where: $i = (k-1) \times N + n$, $(1 \le i \le K \times N)$, $\mathbf{R}_{nk} = [\Delta t \times f(p,n,k), 1 \le p \le L]$, and f is obviously defined by equation 1. Finally the cost functional concerning our problem is defined as:

$$J = \left(\mathbf{C}^O - \mathbf{G}\cdot\mathbf{q}\right)^T \mathbf{O}^{-1}\left(\mathbf{C}^O - \mathbf{G}\cdot\mathbf{q}\right) + \left(\mathbf{q} - \mathbf{q}_B\right)^T \mathbf{B}^{-1}\left(\mathbf{q} - \mathbf{q}_B\right) \qquad (4)$$

Where, $\mathbf{q}_B$ is the first guess of the source rates, $\mathbf{O}$ is the covariance matrix of the observation errors, $\mathbf{B}$ is the covariance matrix of the background errors and $\mathbf{C}^O$ is the observations matrix. In formula (4) the first term characterizes the difference of the analyzed vector of control variables with the measurements. The second term characterizes its difference with the first guess.

In order to calculate the values of **q** that minimize J, the functional J is differentiated with respect to vector **q** and the derivative is set equal to zero , $\partial J/\partial q = 0$.
And, consequently the next equation is obtained:

$$\left(\mathbf{I} + \left(\mathbf{O}^{-1}\mathbf{GB}\right)^T \mathbf{G}\right)\cdot \mathbf{q} = \left(\mathbf{O}^{-1}\mathbf{GB}\right)^T \mathbf{C}^O + \mathbf{q}_B \quad (5)$$

If observational errors are uncorrelated (*Daley,* (1991)) and constant equation (5) becomes:

$$\left(s^{2} + \mathbf{G}^T\mathbf{G}\right)\cdot \mathbf{q} = \mathbf{G}^T\mathbf{C}_O + s^{2}\mathbf{q}_B \quad (6)$$

Where, $s = s_O^2 / s_B^2$, If the observational errors are assumed proportional to the values of observed concentration e.g. $s_{O,nk}^{2} = \left(a\, C_{nk}^O\right)^2$ (as in *Lary and Fisher,* (1995)), then from equation (5) the following equation is obtained:

$$\left(s_{\text{mod}}^{2} + \mathbf{G}_{\text{mod}}^T\mathbf{G}\right)\cdot \mathbf{q} = \mathbf{G}_{\text{mod}}^T\mathbf{C}^O + s_{\text{mod}}^{2}\mathbf{q}_B \quad (7)$$

Where, $\mathbf{G}_{\text{mod}} = \mathbf{C}_{\text{new}}^{-1}\cdot \mathbf{G}$ and $s_{\text{mod}}^{2} = a^2/s_B^2$.
From the equations (6, 7) **q** is calculated.

APPLICATIONS-RESULTS

For the evaluation of the data assimilation algorithm performance, “identical twin” experiments were used, due to lack of real experimental data. The dispersion model generates “concentration observations” using a “true” source term function. Then the model is run again using an “assumed” source term and assimilating the observations with the aim to evaluate the true source function. For simplicity reasons, the applications concern 1-dimensional dispersion. First some test results are presented in the figures 1 and 2 having constant emission rate constant wind speed and one observation point with coordinates $\left(X_m, Y_m, Z_m\right) = \left(5000\text{m},0,0\right)$. Figure 1 indicates the important dependence of the quality of adjustment on values of s . The adjustment of source function improves as the values of s decrease. This is expected, because small s means small error of measurements, and in this case the measurements are given higher weights. At this point it should be noted that the maximum value of s for which results can be considered satisfactory, i.e., $s = 10^{-17}$, would not produce as good results, if concentrations were measured at a point with greater distance from the source. This is because the concentrations there are significantly (possibly by orders of magnitude) smaller. Therefore, for them to be given a considerable weight s should also be correspondingly smaller. The way to overcome this undesirable behaviour is to abandon the assumption of constant s and assume that s is proportional to observed concentrations. In the following test cases the root mean square error of the observations, is proportional to the values of concentration observed and the improvement of results it can be clearly seen (figure 2). In figure 3 the test results for variable in time source term and variable in space wind speed are presented having three measurement points with coordinates:
$\left(X_{m1}, Y_{m1}, Z_{m1}\right) = \left(5000\text{m},0,0\right), \left(X_{m2}, Y_{m2}, Z_{m2}\right) = \left(10000\text{m},0,0\right)\ \left(X_{m3}, Y_{m3}, Z_{m3}\right) = \left(20000\text{m},0,0\right)$

CONCLUSIONS

In the current work the algorithm of variational data assimilation is developed, that allows adjustment of source function in puff/particle model for non-stationary (wind and source) conditions. In the present work the test case with constant wind speed and constant source function was considered in the first place. Additionally constant relative error (error of observation measurements to the error of source function estimation) was assumed with assimilated data from a single measurement point. The results showed that the constant relative error cannot be used in real world problems, and the assumption of error of observations proportional to the concentrations was adopted. Additionally tests with variable in time source term function and with three measurements points were carried out. Finally tests with variable wind speed showed that the formulation of the methodology is more general and can be extended to take into account variable meteorological conditions.

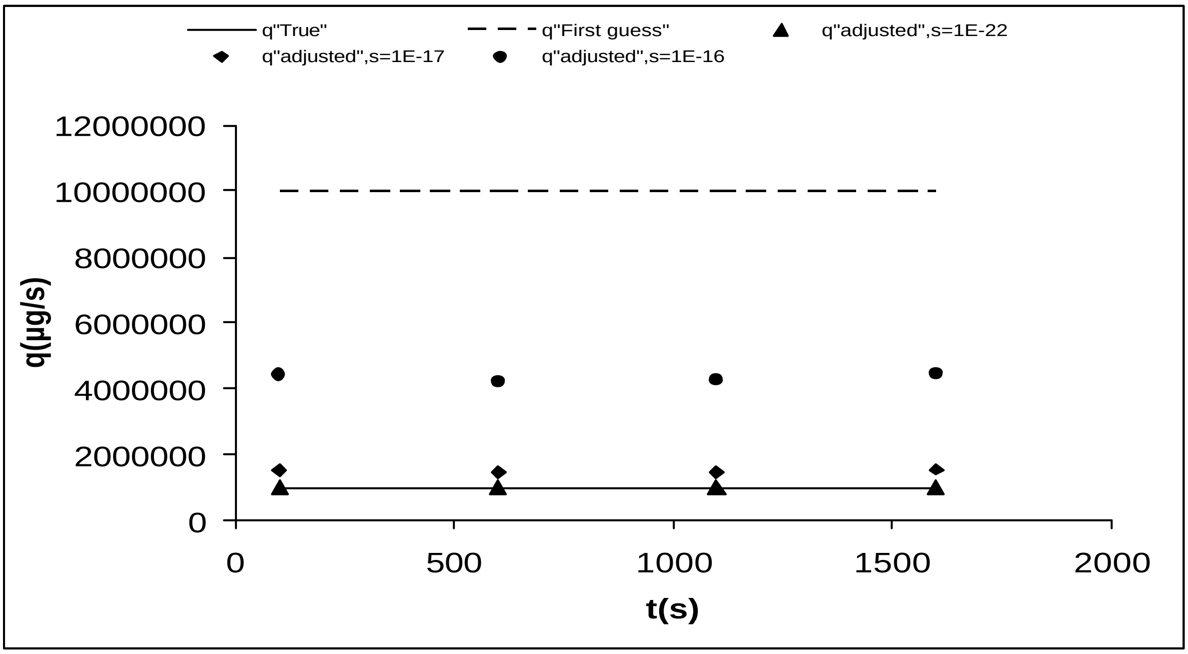

Figure 1: Estimation of "true" source function by the simulation runs for different values of s. The "true" source function (q_T) (at the times of puffs releases), the background (q_B) and the adjusted source functions (q_A) are presented.

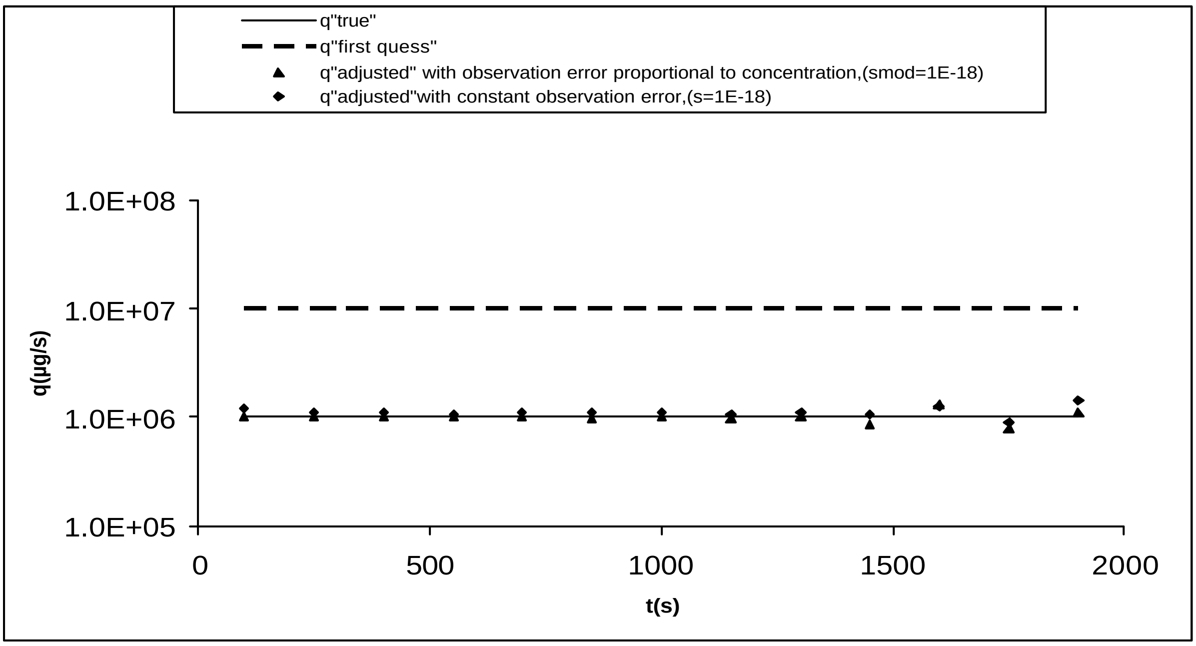

Figure 2: The adjusted source function (q_A) for the case that the RMS error of the observations is constant and for the case that the RMS error of the observations is proportional to the values of concentration are compared with the "true" (q_B) and the background source functions (q_T) (at the times of puffs releases)

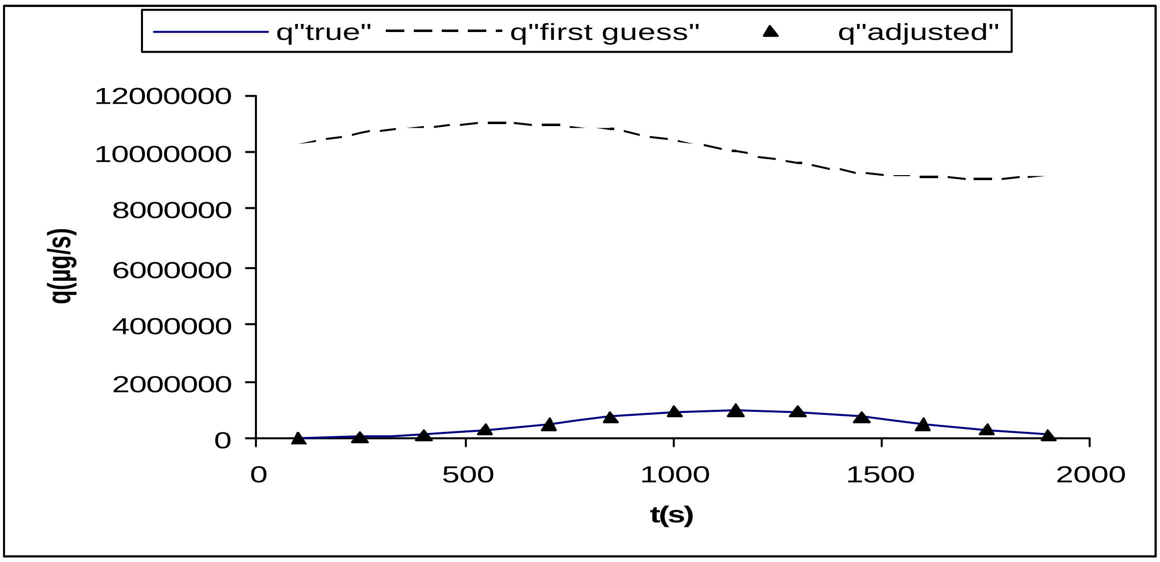

Figure 3: Estimation of source function in the case of variable in time emission rate and variable wind speed. The "true" source function (q_T) (at the times of puffs release), the background (q_B) and the adjusted source functions (q_A) are presented

REFERENCES

Astrup P., C. Turcanu, R.O. Puch,C. Rojas Palma, T.Mikkelsen, 2004: Data Assimilation in the early phase: Kalman Filtering RIMPUFF, RODOS(RA5)-TN(04)-01

Bartzis J., Davakis E., Andronopoulos S., 2000: DIPCOT: A Lagrangian model for atmospheric dispersion over complex terrain, Model Description, "DEMOKRITOS" Report.

Daley, 1991: Atmospheric Data Analysis, Cambridge University Press

Drews, M., Lauritzen, B., Madsen, H., 2005: Analysis of a Kalman filter based method for on-line estimation of atmospheric dispersion parameters using radiation monitoring data, Radiation Protection Dosimetry 113 (1), pp. 75-89

Fisher M., Lary D.J., 1995: Lagrangian four-dimensional variational data assimilation of chemical species, Quarterly Journal - Royal Meteorological Society 121 (527), pp. 1681-1704

Jeong, H.-J., Kim, E.-H., Suh, K.-S., Hwang, W.-T., Han, M.-H., Lee, H.-K, 2005: Determination of the source rate released into the environment from a nuclear power plant, Radiation Protection Dosimetry 113 (3), pp. 308-313

Kathirgamanathan, P., McKibbin, R., McLachlan, R.I., 2004: Source release-rate estimation of atmospheric pollution from a non-steady point source at a known location, Environmental Modeling and Assessment 9 (1), pp. 33-42

Le Dimet F.X., and Talagrand O., 1986: Variational algorithms for analyses and assimilation of meteorological observations: theoretical aspects. Tellus B, 38A, p.97-110

Quélo, D., Sportisse, B., Isnard, O.,2005: Data assimilation for short range atmospheric dispersion of radionuclides: A case study of second-order sensitivity, Journal of Environmental Radioactivity 84 (3), pp. 393-408

Segers A., 2002: Data assimilation in atmospheric chemistry models using Kalman Filtering, PhD Thesis, Delft University, published by DUP Science, 220 p.

Zheng D. Q., Leung J. K. C., Lee J. K. C., Lam H. Y., 2006: Data assimilation in the atmospheric dispersionmodel applied to nuclear accident assessments, Atmospheric Environment (submitted)