The Impacts of Data Error and Model Resolution on the Result of Variational Data Assimilation

S. Yang and Q. Xu
Cooperative Institute of Mesoscale Meteorological Studies
University of Oklahoma
Norman, Oklahoma

Introduction

The representativeness and accuracy of the measurements or estimates of the lateral boundary fluxes and surface fluxes are crucial for the single-column model and budget studies of climatic variables over Atmospheric Radiation Measurement (ARM) sites. Since the direct measurements of these fluxes have limited resolution, it is desirable to use high-resolution Doppler-radar data to retrieve the three-dimensional wind and temperature fields and improve the representativeness and accuracy of the estimated fluxes.

To this end, a number of techniques of variational data assimilation (VDA) are currently being developed (Sun et al. 1991, Qiu and Xu 1992, Xu et al. 1993). This paper is intended to address the following specific issues:

- How is the retrieval error affected by the observational error?
- Could the observational error cause a bias in the retrieval?
- Does the retrieval error depend on the model grid resolution and how?

Theoretical Analysis

Let us assume that the physical process is governed by the one-dimensional advection equation

$$\frac{\partial Q}{\partial t} + u(x)\frac{\partial Q}{\partial x} = 0 \quad (1)$$

The variational data assimilation method seeks the best estimate of u(x) in Equation (1) that minimizes the following cost function:

$$J = \int_0^T \int_0^L \left[q_o(x,t) - q(x,t)\right]^2 dx\,dt + \varepsilon T^3 \int_0^L \left[u(x) - u_b(x)\right]^2 dx \quad (2)$$

where

T = the assimilation time interval

L = the assimilation domain

ε = a small but positive parameter

$u_b(x)$ = a known background of u(x)

$q_o(x,t)$ = the observed value of Q

q(x,t) = the predicted value of Q from Equation (1) with the observed initial condition.

To facilitate the theoretical analysis, we assume that Q is observed at two time levels, $t = t_1$ and $t = t_2$, and the values of $q_o(x,t)$ at t_1 and t_2 are denoted by q_1 and q_2, respectively, so linear interpolation can be used for the discrete forms of $\frac{\partial q}{\partial x}$ and q_o. With these approximations, the minimum condition of the cost function, $\frac{\partial J}{\partial u_i} = 0$, leads to the following explicit expression for the retrieved velocity, at a given discrete point $x=x_i$:

$$u_i = -\frac{q_2 - q_1}{T}\frac{B_i}{2\varepsilon + A_i} + \frac{2\varepsilon}{2\varepsilon + A_i} u_b(x_i) \quad (3)$$

where $u_i=u(x_i)$, A_i and B_i are functions of the first order finite differences of q_1 and q_2 associated with the grid spacing Δx.

Real observational data are not free of noise. For simplicity, we assume that the noise is uncorrelated in time and in space and obeys the Gaussian probability distribution. From Equation (3), the statistic mean and variance of the retrieved u_i can be derived as follows:

$$\bar{u}_i = -\frac{\overline{q_2} - \overline{q_1}}{T} F_1 \tag{4}$$

$$\sigma_u^2 = \overline{(u_i - \bar{u}_i)^2} = \overline{u_i^2} - \bar{u}_i^2$$

$$= \frac{2\sigma^2}{T^2} F_2 + \frac{\left(\overline{q_2} - \overline{q_1}\right)^2}{T^2} F_3 \tag{5}$$

where the overbar represents statistical mean; F_1, F_2 and F_3 are functions of $\partial \bar{q}_j / \partial x$ (j = 1, and the standard deviation of noise (denoted by σ) in the observations of q_1 and q_2.

To measure the systematic bias of the retrieved velocity, we define

$$r = \bar{u}_i / u_a \quad \text{and} \quad d = r - 1 \tag{6}$$

where u_a is the true advection velocity. To measure the RMS difference between the retrieved u_i and the true advection velocity u_a, we define

$$\sigma_a^2 = \sigma_u^2 + \left(\bar{u}_i - u_a\right)^2 = \sigma_u^2 + (1-r)^2 u_a^2 \tag{7}$$

The theoretical results are shown in Figure 1. As shown, u_i can be seriously underestimated when Δx is small or slightly overestimated when Δx is large. For a given σ level of observational noise, the standard deviation σ_u of the retrieval increases with Δx first and then decreases and approaches a constant value when Δx is large. When σ is large, the position of maximum σ_u shifts to a larger value on the Δx axis.

Numerical Experiments

The above theoretical analysis is obtained for a simplified VDA in which only linear steps of time integration are considered at a given grid point. In this section, the robustness of the theoretical results is tested against the statistics of retrieved velocities obtained in a large number of numerical experiments. A leap-frog central differential scheme is used with a time filter for the time integration of Equation (1). This numerical scheme is also used to create the "true" (noise-free) field of Q(x,t), denoted by Q_a. With a given true advection velocity $u_a(x)$, Q_a is integrated from an initial condition, then the "observational data" for q_1 and q_2 are "generated" by adding random noise on Q_a. For each set of q_1 and q_2, a sample of retrieved velocity field, denoted by $u_r(x)$, is obtained through the numerical VDA. In each numerical experiment, 200 samples of the retrieved fields are collected.

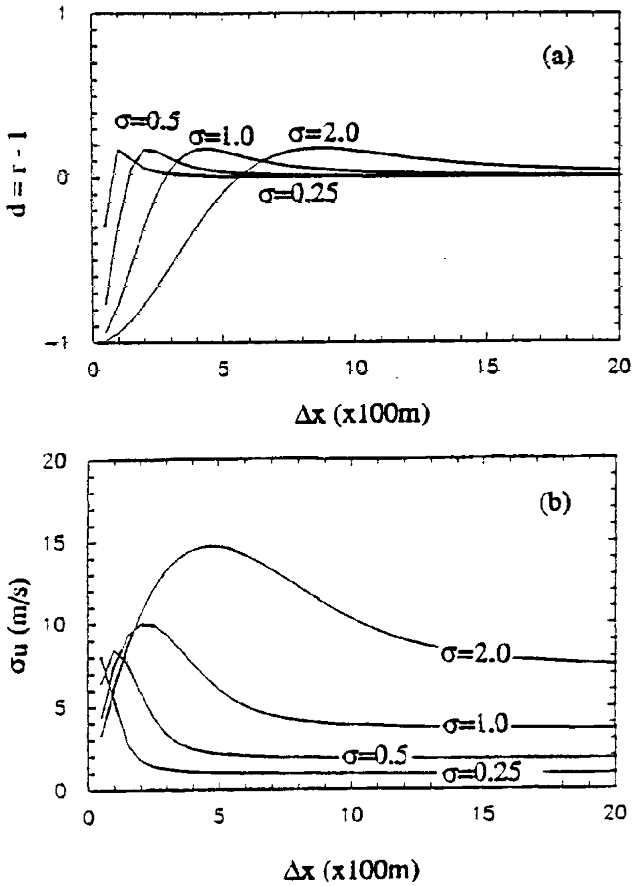

Figure 1. Theoretical estimates of (a) the retrieval bias d = r -1, and (b) the retrieval error standard deviation σ_u versus grid size Δx for different values of the standard deviation of measurement noise σ. Here, u_a = 5.0ms^{-1}, q_x = 4.0x10^{-3} m^{-1}, T = 100s, ε = 2.0x10^{-8} m^{-2}.

The domain-averaged relative bias of the retrieved velocity $\langle u_r \rangle$ with respect to true velocity u_a is defined by

$$r_d = \frac{\overline{\left| \langle u_r(x) \rangle - u_a(x) \right|}^x}{\overline{\left| u_a(x) \right|}^x} \quad (8)$$

where the angle braces denote averaging over 200 samples, and the overbar with superscript x denotes domain-averaging along the x axis. The theoretical estimate is closely compared with the numerical result of r_d in Figure 2, where σ=0.5, $u_a(x)$ = 5.0 m/s, and the initial value of Q_a is given by $A\sin(2\pi x/\lambda)$ with $0 \leq x \leq L$, A=5.0, and $\lambda=10^4$m. The theoretical estimate of the variance of the retrieved velocity is also verified by the numerical experiments (not shown). The robustness of the theoretical results is tested by 60 numerical experiments with different settings of A,σ and u_a.

Error Suppression

The power spectrum of the variance of the retrieved velocity and its spatial correlation coefficient suggest (not shown) that the error field of the retrieved velocity behaves very much like a white noise field. Thus, the error can be effectively suppressed if a three-point filter is used as a weak constraint for the retrieved velocity field (see C2 in Figure 3).

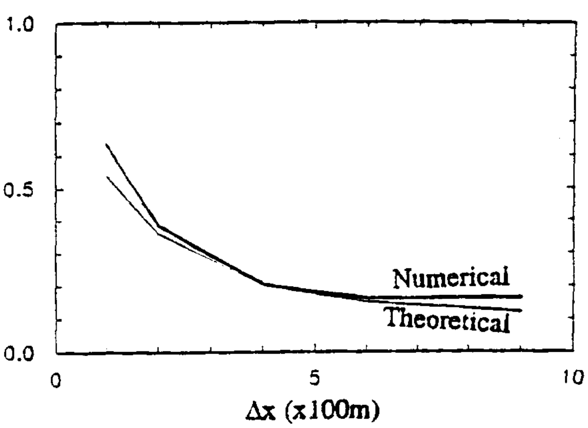

Figure 2. Theoretical estimate and numerical result for the domain-averaged relative deviation of the retrievals versus grid size Δx. The noise standard deviation is σ = 0.5.

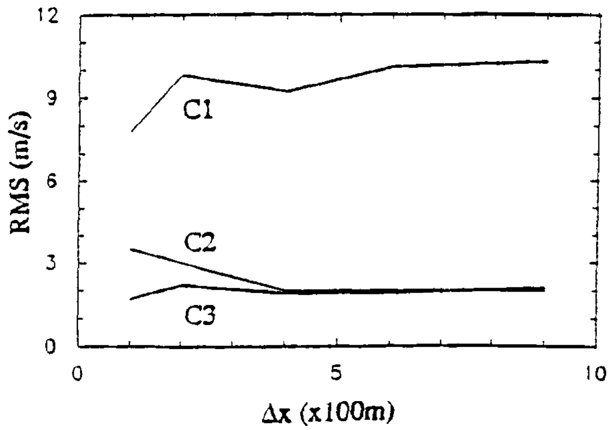

Figure 3. Domain-averaged RMS difference between the retrieved and true velocities. C1 is fixed frame without spatial filter; C2 is fixed frame with spatial filter; C3 is moving frame with spatial filter.

Furthermore, Equation (7) indicates that σ_a^2 is proportional to the true velocity u_a, so σ_a^2 should decrease with u_a. This result implies that the error can be further reduced if the moving reference frame technique (Gal-Chen 1982) is used in the VDA retrieval, because u_a can be reduced in a moving reference frame (see C3 in Figure 3).

Summary and Conclusions

The model has a critical grid size below which the retrieved velocity can be systematically smaller than the true value (see Figure 1a). The variance of the retrieved velocity is proportional to the variance of noise and the time tendency of the observed variables; it also depends on the model resolution (see Equation [5] and Figure 1b). The retrieval can be improved by applying a spatial filter to the retrieved field and/or by performing the VDA retrieval in a moving reference frame.

Acknowledgment

This work is supported by National Oceanic and Atmospheric Administration (NOAA) contract NA90-RAH00078 and U.S. Department of Energy/Pacific

Northwest Laboratory contract 144880-A-Q1 at the Cooperative Institute of Mesoscale Meteorological Studies, University of Oklahoma.

References

Gal-Chen, T. 1982. Errors in fixed and moving frame of references: Applications for conventional and Doppler radar analysis. *J. Atmos. Sci.* **39**:2279-2300.

Sun, J., D. W. Flicker, and D. K. Lilly. 1991. Recovery of three-dimensional wind and temperature fields from simulated single-Doppler radar data. *J. Atmos. Sci.* **48**:876-890.

Qiu, C., and Q. Xu. 1992. A simple adjoint method of wind analysis for single-doppler data. *J. Atmos. Ocean. Tech.* **9**:588-598.

Xu, Q., C. J. Qiu, J. X. Yu., H. D. Gu, and M. Wolfson. 1993. Adjoint-method retrievals of microburst winds from TDWR data. Preprints, *26th International Conference on Radar Meteorology*, pp. 433-434. American Meteorological Society, Boston, Massachusetts.