JP1J.3 THE IMPACT OF DIFFERENT DATA FIELDS ON STORM-SSCALE DATA ASSIMILATION

Guoqing Ge[1,2,*], Jidong Gao[1], Kelvin Droegemeier[1,2]
Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, OK
School of Meteorology, University of Oklahoma, Norman, OK

1. Introduction

Many data assimilation techniques have been developed to assimilate radar velocity and reflectivity data into storm scale numerical prediction model (i.e. Sun et al. 1991; Sun and Crook 1997,1998; Gao et al. 2002; 2004; Snyder and Zhang 2003; Zhang et al. 2004; Tong and Xue 2005). Although these research efforts have been generally successful, a clear understanding of which data fields have the strongest impact on and how much information is needed for storm scale data assimilation is lacking. Such knowledge is very important for designing storm-scale observing systems and developing new data assimilation procedures. Weygandt et al.(1998) have successfully done some experiments to study the relative importance of different data fields in a numerically-simulated convective storm by withdrawing information about each model variable and then rerunning the simulation. It is found that the perturbation horizontal velocity exerts the greatest influence on the convective evolution. Nascimento and Droegemeier (2002) conducted experiments using the same method, but with an idealized numerically-simulated bow echo, instead of supercell storm. It confirms that horizontal wind field is important on the simulation. Furthermore, it is also shown that the water vapor field should be correctly initialized in the model initialization in order to get reasonable results. Sun (2005) examined the relative importance of the initial analysis fields in a supercell storm observed during STEPS using same method as before and found that the prediction is most sensitive to the initialization of wind, water vapor and temperature perturbations. These results are inspiring, but some important questions are still left unresolved: What's the impact of different data fields with different intervals? How much information, or how many variables are needed to get better results?

In this study, these issues are addressed by assimilating the pseudo-observations from different model variables separately or simultaneously to determine the impact of different data field, or a group of data fields on storm-scale data assimilation. This is accomplished by first creating a control simulation of a thunderstorm using a nonhydrostatic stormscale numerical model; then obtaining different type of simulated observations from the control run. The observational data are assumed to have some types of errors, exist at every model grid point. The assimilation experiments are performed using a simple variational method for each data assimilation cycle. By using different intervals (1 minutes, 5 minutes, 10 minutes), the different types of pseudo observations, and the different amount of pseudo-observations, the impact of each data fields with different intervals can be quantitatively evaluated.

2. Methodology

a. Description of variational methodology

The standard formulation of variational methods was derived from first principles by Lorenc using Baysian probabilities and assuming

Gaussian error distributions (Lorenc 1986). The concept of a variational method is to determine the analysis by direct minimization of a cost function. The cost function can be written as:

$$J(x) = J_B + J_O = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b)$$
$$+\frac{1}{2}[H(x) - y_o]^T R^{-1}[H(x) - y_o]$$

where J_B measures the departure of the analysis *x* from the background, and is weighted by the inverse of the background error covariance matrix B; J_O measures the departure $H(x)$, which is the projection of the analysis x in observational space, from the observation, y_o and is weighted by the inverse of the observational error covariance matrix R.

In this study, the observations are directly drawn from the model variables, so no projection or interpolation is needed. Both B and R are assumed uncorrelated and are constants for different variables based on their corresponding standard error deviation.

b. The prediction model and truth simulation

In this study, we use simulated data from a classic May 20, 1977 Del City, Oklahoma supercell storm case (Ray et al. 1981). The Advanced Regional Prediction System (ARPS) (Xue et al 2000; 2001; 2003) is used to simulate such a deep convective storm within a 64 x 64 x 16 km physical domain. The model grid comprises of 67 x 67 x 35 grid points. Horizontal resolution of 1km and vertical resolution of 0.5km are used. The truth simulation is initialized from a modified real sounding plus a +4K ellipsoidal thermal bubble centered at x=48, y=16 and z=1.5km, with radii of 10km in x and y and 1.5km in z direction. Open conditions are used at the lateral boundaries. A wave radiation condition is also applied at the top boundary. Free-slip conditions are applied to the bottom boundary. The length of simulation is up to three hours. A constant wind of u=3ms^{-1}and v= ms^{-1} is subtracted from the observed sounding to keep the primary storm cell near the center of model grid. The evolution of the simulated storms is similar to those documented in Xue et al. (2001).

During the truth run, the supercell strengthens over the first 20 minutes. The strength of the cell then decreases thereafter. At around 55 minutes, the cell splits into two. The right moving cell tends to dominate the system. Another cell moves northwestward and splits again at 95 minutes (Tong and Xue 2005).

c. Data assimilation procedure

One or several of the following simulated observations are assumed to be available on the grid points in different data assimilation experiments: three components of wind field, *u*, *v*, and *w*, perturbation potential temperature θ', perturbation water vapor mixing ration (q_v) and rain water mixing ratio (q_r).

The observations are assimilated every 1 minute, 5 minutes, and 10 minutes separately. Observations are obtained by adding Gaussian noises with zero mean and standard deviation of σ_1 to the control run (Table 1).

Table 1. The standard deviation of observation error (σ_1) and background error (σ_2).

	σ_1	σ_2
u	1m/s	3m/s
v	1m/s	3m/s
w	0.667m/s	2m/s
θ'	0.667K	2K
q_r	0.167g/kg	0.5g/kg
q_v	0.033g/kg	0.1g/kg

The background error is also assumed same type with zero and standard deviation of σ_2 (Table 1). The pseudo observations are obtained from 30 minutes of control run to 120 minutes and all assimilation experiments are started from the environmental sounding.

RMS error of composite reflectivity is used to analyze the results of different data assimilation schemes in order to compare them with truth run. When calculating the RMS error, only grids which are located in the cloudy region (where observed Z ⩾ 10dBz) are taken into account.

3. The assimilation experiments

In order to illustrate effects of different assimilation intervals, three types of data assimilation experiments are performed with 1 minute (1m), 5 minutes (5m), and 10 minutes (10m) interval respectively. In each type, 14 experiments—*u, v, w, pt*, q_r, q_v, *uv, upt*, uq_r, uq_v, *uvw, uvwpt*, $uvwq_r$, $uvwq_v$—are conducted separately to show the effect of different amount of variables used in observations.

The changes of RMS error for reflectivity with time are shown in Fig 1. At the beginning, the background is horizontally homogeneous and the reflectivity of assimilation run is zero everywhere. The RMS error at this time is 35dBz. When the observations derived from control run is assimilated into model, the RMS error decreases and the structure of the storm is gradually recovered.

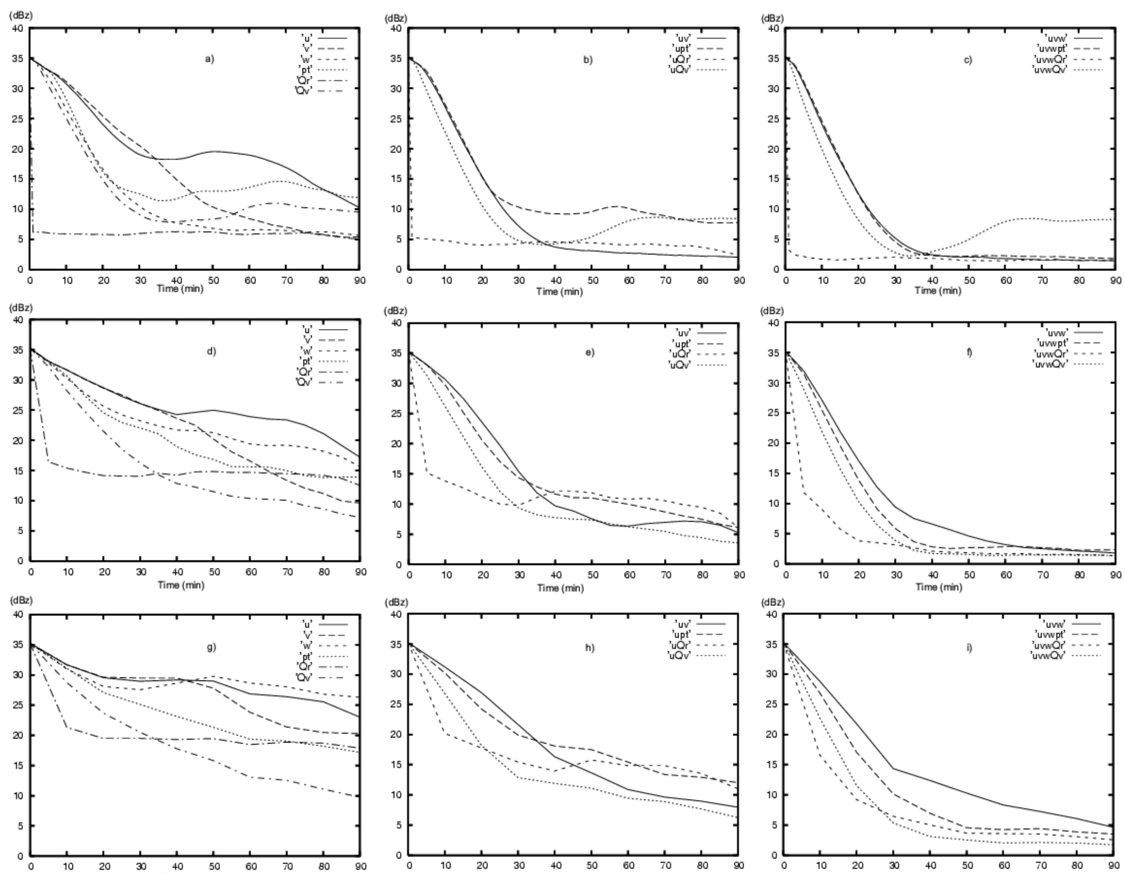

Fig 1. The reflectivity RMS errors of all experiments, averaged over points at which the reflectivity is greater than 10 dBz. a) b) c) with 1 minute interval; c) d) e) with 5 minutes interval; f) g) h) with 10 minutes interval.

Table 2, Table 3, Table 4 show all experiments and the spin-up time RMS error at the end of assimilation cycle. Herein spin-up time is defined as the time when RMS error first decrease to 10dBz, i. e., the assimilated fields are closed enough to the truth run. Blank cells in "Spin-up Time" columns means RMS error of the corresponding experiment is always above 10dBz during the experiments.

In table 2, the RMS errors of experiments *u, v, w, pt,* q_r, q_v at the end of each experiment are generally large, mostly larger than 10dBz. This indicates that just assimilating one variable is often not enough to recover the full storm, or get good results. But when shorten the intervals, experiments v, w, q_v give relative low RMS errors. It is found that in experiment q_r with 1 minute interval the RMS errors quickly decrease to 10dBz. This is because q_r is directly related to the reflectivity, but it does not mean the structure of the storm will be constructed very well at that time (figures not shown). Another thing is that assimilating q_v can get better results than other variables no matter what the assimilation interval is. This shows that q_v has great impact on stormscale data assimilation.

Table 2. Experiments and spin-up time, RMS error for experiment u, v, w, pt, q_r, q_v *(spin-up time is defined as the time when RMS error first decrease to 10dBz and RMS error is the value at the end of each experiment; blank cells in "Spin-up Time" columns means RMS error of the corresponding experiment doesn't reach below 10dBz during the experiments).*

	Category 1m (interval=1 minute)		Category 5m (interval=5 minutes)		Category 10m (interval=10 minutes)	
	Spin-up Time (minutes)	RMS error (dBz)	Spin-up Time (minutes)	RMS error (dBz)	Spin-up Time (minutes)	RMS error (dBz)
U		10.3		17.2		27.5
V	51	5.4	85	9.5		20.3
W	31	5.6		15.6		26.3
Pt		11.9		13.9		17.2
q_r	1	5.0		12.5		17.9
q_v	28	9.5	70	7.2	88	9.7

Table 3. Same as Table 2, but when two variables are assimilated.

	Category 1m (interval=1 minute)		Category 5m (interval=5 minutes)		Category 10m (interval=10 minutes)	
	Spin-up Time (minutes)	RMS error (dBz)	Spin-up Time (minutes)	RMS error (dBz)	Spin-up Time (minutes)	RMS error (dBz)
uv	26	2.0	40	5.3	68	8.0
upt	31	7.8	60	6.1		12.0
uq_r	1	2.5	25	6.1		11.0
uq_v	21	8.5	29	3.6	58	6.2

Table 4. Same as Table 2, but when three and more variables are assimilated.

	Category 1m (interval=1 minute)		Category 5m (interval=5 minutes)		Category 10m (interval=10 minutes)	
	Spin-up Time (minutes)	RMS error (dBz)	Spin-up Time (minutes)	RMS error (dBz)	Spin-up Time (minutes)	RMS error (dBz)
uvw	22	1.5	29	1.8	52	4.6
uvwpt	22	1.8	24	2.3	31	3.5
$uvwq_r$	1	1.6	9	1.3	19	2.6
$uvwq_v$	18	8.3	21	1.4	23	1.7

In Table 3, the RMS errors of experiments uv, upt, uQr, uQv at the end of each experiment are generally below 10dBz. It takes longer for RMS error of 10 minutes interval assimilation to decrease to 10dBz, comparing to 5 minutes interval assimilation, and so does 5 minutes interval comparing to 1 minutes interval. So shorter assimilation interval can speed up the spin-up time.

In Table 4, the RMS error of each experiment is rather low, mainly smaller than 2 dBz. It is not surprise that more variables are assimilated, better results can be gotten. From Table 4 we can also conclude that given only the full wind field is sufficient to recover all the detail of the storm. This indicates that the wind field is very important for storm scale data assimilation (Weygandt et al. 1998; Nascimento and Droegemeier 2002; Sun et al. 2005).

4. Conclusion

In this study we tested the impact of

different data fields in storm-scale data assimilation and the impact of different data assimilation window. It is founded that observation for only single variable may not be sufficient to get the reasonable assimilation results. The accurate full wind field contains sufficient information for stormscale data assimilation. Adding potential temperature (θ'), rainwater mixing ratio (q_r) or water vapor mixing ratio (q_v) into observations can speed up spin-up time, but may increase the RMS error afterward.

5. Acknowledgements

This work was supported by NSF grants ATM-0331756, ATM-0129892, ATM-0331594 and EEC-0313747.

6. References

Gao, J., M. Xue, K. Brewster, F. Carr, and K. K. Droegemeier, 2002: New Development of a 3DVAR system for a nonhydrostatic NWP model. P*reprint, 15th Conf. Num. Wea. Pred.and 19th Conf. Wea. Anal. Forecasting,* SanAntonio, TX, Amer. Meteor. Soc., 339-341.

Gao, J.-D., M. Xue, K. Brewster, and K. K. Droegemeier, 2004: A three-dimensional variational data analysis method with recursive filter for Doppler radars. J. *Atmos. Ocean. Tech.,* 21, 457-469.

Nascimento, E. L and K. K. Droegemeier, 2002: Dynamic adjustment within an idealized numerically-simulated bow echo: implications for data assimilation. *Preprint, Symposium on Observations, Data Assimilation, and Probabilistic Prediction.*

Ray, P. S., B. Johnson, K. W. Johnson, J. S. Bradberry, J. J. Stephens, K. K. Wagner, R. B. Wilhelmson, and J. B. Klemp, 1981: The morphology of severe tornadic storms on 20 May 1977. *J. Atmos. Sci.*, 38, 1643-1663.

Snyder, C. and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. M*on. Wea. Rev.,* 131, 1663-1677.

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.

Sun, J. and N. A. Crook, 1997: Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint. Part I: Model development and simulated data experiments. J. *Atmos. Sci.,* 54,1642-1661.

_____, 1998: Dynamical and Microphysical Retrieval from Doppler Radar Observations Using a Cloud Model and Its Adjoint. Part II: Retrieval Experiments of an Observed Florida Convective Storm. J. *Atmos. Sci.,* 55, 835-852.

Sun, J. 2005: Initialization and numerical forecasting of a supercell storm observed during STEPS. M*on. Wea. Rev.,* 133, 793-813.

Tong, M. and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS Experiments. M*on. Wea. Rev.,* 133, 1789-1807.

Weygandt, S and P. Nutter, E. Kalnay, S. K. Park, and K. K. Droegemeier.1999. The relative importance of different data fields in a numerically-simulated convective storm. *Preprint, 8th Conf. on Mesoscale Process. 310-315*

Xue, M., K. K. Droegemeier, and V. Wong, 2000: The Advanced Regional Prediction System (ARPS) - A multiscale nonhydrostatic atmospheric simulation and prediction tool. Part I: Model dynamics and verification. *Meteor. Atmos. Physics*, 75, 161-193.

Xue, M., D.-H. Wang, J.-D. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and data assimilation. *Meteor. Atmos. Physics*, 82, 139-170.

Xue, M., K. K. Droegemeier, V. Wong, A. Shapiro, K. Brewster, F. Carr, D. Weber, Y. Liu, and D. Wang, 2001: The Advanced Regional Prediction System (ARPS) - A multi-scale nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications. *Meteor. Atmos. Phys.*, 76, 143-166.

Xue, M., M. Tong, and K. K. Droegemeier, 2005: An OSSE framework based on the ensemble square-root Kalman filter for evaluating impact of data from radar networks on thunderstorm analysis and forecast. J. *Atmos. Ocean Tech.,* Accepted.

Zhang, F., C., Snyder, and J. Sun, 2004: Impacts of initial estimate and observations on the convective-scale data assimilation with an ensemble Kalman filter. Mon*. Wea. Rev., 1*32, 1238-1253.