Proceedings of ERAD 2006

Phased-array radar data assimilation at the National Weather Radar Testbed – Theoretical issues and practical solutions

Qin Xu[1], Li Wei[2], Huijuan Lu[2], Kang Nai[2], and Qingyun Zhao[3]

[1] NOAA/National Severe Storms Laboratory (NSSL), Norman, OK, USA.

[2] CIMMS, University of Oklahoma, Norman, OK, USA.

[3] Naval Research Laboratory, Monterey, CA 93943, USA.

1 Introduction

A phased-array weather radar has been constructed at NSSL Norman Oklahoma. This establishes the first National Weather Radar Testbed (NWRT) equipped with the state-of-the-art (solid-state) phased-array antenna (Forsyth et al. 2005). An important and yet very challenging research goal is to optimally design and utilize the electronically-controlled agile beam scans for meteorological applications, such as assimilating phased-array radar observations to improve numerical analysis and prediction of severe storms and other hazardous weather conditions. This paper reports our recent research progress in this direction with particular attentions to the following important issues: (i) how to design phased-array scan strategies to enhance radar observation information content for data assimilation; (ii) how to take the advantages of phased-array rapid and flexible scan capabilities to improve error covariance estimation for radar data assimilation. These issues will be addressed theoretically, and some practical solutions will be proposed and demonstrated by numerical experiments.

2 Information content extracted from observations

When observations are assimilated into a numerical weather prediction (NWP) model by an optimal analysis system, the background state, denoted by vector **b**, is provided by the prediction of the NWP model. The background probability density function (pdf), denoted by $q(\mathbf{x})$, is assumed to be Gaussian with a pre-estimated or predicted covariance matrix **B**. The information content from the observations can be measured by the relative entropy defined by $R(p, q) = \int d\mathbf{x} p(\mathbf{x}) \ln[p(\mathbf{x})/q(\mathbf{x})]$, where $p(\mathbf{x})$ is the analysis pdf and **x** is the state vector. For Gaussian pdfs, this gives

$$R(p, q) = (\mathbf{a} - \mathbf{b})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{a} - \mathbf{b})/2 + [\ln|\mathbf{G}| + \mathrm{Tr}(\mathbf{G}^{-1}) - n]/2, \quad (1)$$

where **a** is the analysis mean, $(\ )^{\mathrm{T}}$ denotes the transpose of ( ), $|\mathbf{G}|$ is the determinant of $\mathbf{G} = \mathbf{B}^{1/2}\mathbf{A}^{-1}\mathbf{B}^{1/2}$, **A** is the analysis covariance matrix, Tr( ) denotes the trace of ( ), and n is the dimension of **b**. The first and second terms on the right-hand side of (1) are the signal and dispersion parts of the information content, respectively (Majda et al. 2002). The above results lead to the following two points:

(a) The information content extracted from observations by an optimal analysis can be measured only indirectly in terms of the innovation (defined by **a** – **b**) and pdf changes (from q to p) produced by the analysis, so the information content depends on both the observation and background pdfs.

(b) Since the relative entropy is invariant with respect to smooth invertible variable transformations, the information content measured by (1) is invariant when the Gaussian pdfs are transformed to non-Gaussian or vice versa.

Note that **A** is related to **B** by $\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$ or, equivalently, $\mathbf{A} = \mathbf{B} - \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}$, where **R** is the observation error covariance matrix and **H** is the tangent-linearization of the observation operator $H(\ )$ at $\mathbf{x} = \mathbf{b}$. Substituting these relationships into (1) gives

$$R(p, q) = \sum[d_i^2\lambda_i^2/(1+\lambda_i^2)^2 + \ln(1+\lambda_i^2) - \lambda_i^2/(1+\lambda_i^2)]/2. \quad (2)$$

Here, d_i is the i-th element of $\mathbf{d} = \mathbf{U}^{\mathrm{T}}\mathbf{R}^{-1/2}[\mathbf{y} - H(\mathbf{b})]$; **y** is the observation vector; **U** is the left orthogonal matrix given by the singular value decomposition (SVD) of the scaled observation operator: $\mathbf{M} = \mathbf{R}^{-1//2}\mathbf{H}\mathbf{B}^{1/2} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\mathbf{T}}$; λ_i is the i-th diagonal element of the diagonal matrix **Λ**; and the summation is over i from 1 to r = rank(**M**). In (2), the observation space is transformed by $\mathbf{U}^{\mathrm{T}}\mathbf{R}^{-1/2}$. In this transformed observation space, the information content

Correspondence to: Qin Xu.

qin.xu@noaa.gov

becomes separable between components associated with different singular values of **M**. Based on this understanding, two additional points can be made as follows:

(c) Observations can be compressed into surper-observations by applying the truncated transformation $\mathbf{I}_s\mathbf{U}^T\mathbf{R}^{-1/2}$ to **y**. The super-observation vector is given by $\mathbf{y}_s = \mathbf{I}_s\mathbf{U}^T\mathbf{R}^{-1/2}\mathbf{y}$, where $\mathbf{I}_s$ is the $s\times s$ identity matrix that projects R^m onto R^s ($s < m$) and m is the dimension of **y**. This compression causes no information loss as long as $s \geq r = \text{rank}(\mathbf{M})$.

(d) If $\mathbf{y}_s = \mathbf{I}_s\mathbf{U}^T\mathbf{R}^{-1/2}\mathbf{y}$ is further truncated to $s < r$, then the compression will cause an information loss. The dispersion part of the information loss caused by the SVD-based compression is the minimum loss for a given truncation number s ($< r$). The signal part of the information loss depends on the truncated non-zero singular values and associated components of **d**.

Denote by n' the dimension of the subspace in which the model state vector is affected by the observations through the analysis. If the background resolution is much coarser than the radar observations and the background covariance is local or localized (nonzero only within a certain range of spatial separation), then $r \leq n' << m$ and there can be considerable information redundancies (quantified by the dimension of the null space of **M**, that is, $m - r$). Redundant observations impose unnecessary computational burdens on the analysis system and can also cause the analysis ill conditioned. Redundant radar observations can be compressed based on the SVD of **M** as in (c)-(d). This compression can be efficient only if observations are assimilated serially in small batches with the background covariance updated also serially by using a Kalman filter. If the observations are analyzed in a single batch without updating the background covariance as in most operational data assimilations, then the observation space will be too large to implement the SVD-based compression unless the compression is localized with an additional information loss.

With the phased-array radar, the flexible agile-beam allows adaptive scan strategies to reduce or eliminate information redundancy and thus enhance the information content. The relative entropy in (2) can be used to measure the optimality of radar scanning strategy in terms of maximizing the information content from observations for a given data assimilation system. This problem is examined with idealized observations and background fields in simple one-dimensional settings. The results (not shown) suggest that the operational WSR-88D radar scans may have excessive spatial resolutions (0.25 km in radial range and about 1° in azimuthal) in radial-velocity observations even for a storm-scale data assimilation system with a horizontal resolution of $\Delta x = 2$ km. The rapid and flexible agile-beam scans from the phased-array radar can be configured to properly reduce the spatial resolutions and enhance the temporal resolution and/or measurement accuracy (by repeated and uncorrelated sampling). Trade-offs between the spatial resolution, temporal resolution and measurement accuracy will be examined in the next section by performing observing system simulation experiments (OSSEs) with an ensemble Kalman filter (EnKF).

3 Assimilation experiments with simulated radar scans

The assimilation system used in this study is the same as in Tong and Xue (2005, referred to as TX05 hereafter) except that it uses the ensemble square-root Kalman filter (Whitaker and Hamill 2002). As in TX05, the Advanced Regional Prediction System (ARPS, Xue et al. 2003) is used to generate the "true" fields for a classic supercell storm case. The physical domain is 64×64×16 km^3. For the assimilation and forecast runs, the model grid comprises 35×35×35 points with grid resolutions of 2 km in the horizontal directions and 0.5 km in the vertical. However, unlike in TX05, the "truth" fields are generated by simulations on a 131×131×35 grid with the horizontal resolution enhanced to 0.5 km, so the model is not perfect here. During the truth simulation, the initial convective cell strengthens over the first 20 min and then splits into two at around 55 min. The updraft of the right-moving cell reaches a peak value of 45 ms^{-1} at 90 min, while the left-moving cell starts to split again at t = 95 min. The evolution is roughly the same as shown in Fig. 1 of TX05.

Radial-velocity observations are sampled from the "true" fields in precipitation regions (with reflectivity > 10 dBZ) to simulate phased-array radar scans in four different modes:

Mode-1: 5 min per volume scan with spatial resolutions of 0.25 km in radial range and 1° in azimuthal on 14 tilts with elevation angles of 0.48, 1.45, 2.4, 3.3, 4.3, 5.2, 6.2, 7.5, 8.7, 10.0, 12, 14.0, 16.7 and 19.5 degree. Observation errors are assumed to be uncorrelated (or de-correlated) with a standard deviation of $\sigma_o = 4$ ms^{-1} (as estimated in section 4).

Mode-2: As mode-1, but the observations are thinned in radial range with the range resolution increased to 1 km.

Mode-3: As mode-1, except that the spatial resolutions are 2 km in radial range and 2° in azimuthal. Here, the range resolution is reduced 16 times and each observation is produced by 16 times as many individual measurements as those in mode-1, so the error variance is reduced 16 times. This gives $\sigma_o = 1$ ms^{-1}.

Mode-4: 2 min per volume scan with spatial resolutions of 2 km in radial range and 2° in azimuthal on the same tilts as in mode-1. Here, the spatial resolution is reduced 16 times but the temporal resolution is increased by 2.5 times, so the error variance is reduced 6.4 times and $\sigma_o = 1.58$ ms^{-1}.

The assimilation starts at t = 25 min and ends at 100 min, and 40 ensemble members are used. The rms errors of ensemble mean forecasts and analyses are plotted in Fig. 1 for each scan mode. The drops in the error curves at specific times correspond to the error reductions produced by the analyses. The error curves for model-1 and model-2 are very close to each other, so model-1 has excessive spatial resolutions. When the scan mode-1 changes to mode-3, and then to mode-4, the rms errors are reduced progressively. Note that the scan mode-1, mode-3 and mode-4 require essentially the same amount of measurement capabilities but they impact the analyses and predictions differently. Thus, it

is possible to design an optimal or nearly optimal scan strategy with proper trade-offs between the spatial resolution, temporal resolution and measurement accuracy.

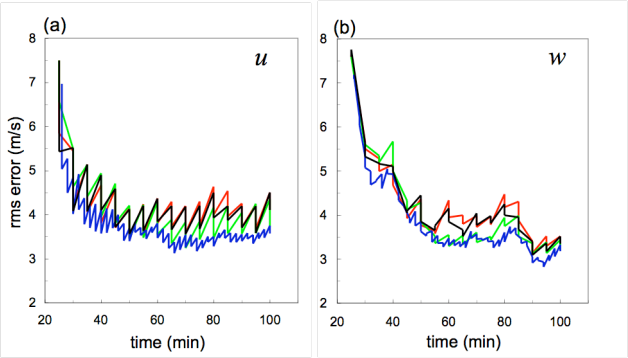

Fig. 1. The rms errors of the ensemble-mean forecast and analysis, averaged over points at which the reflectivity is greater than 10 dBZ for (a) *u* and (b) *w*. The black, red, green, and blue curves are for scan mode-1, mode-2, mode-3 and mode-4, respectively. The drops in the error curves at specific times correspond to the reduction of error by analysis.

4 Radar wind observation error estimation

As shown in section 2, the information content from observations depends on the observation and background pdfs. An optimal analysis system can produce truly optimal analyses only if the observation and background error statistics are accurately estimated to represent the true uncertainties described by their pdfs (or joint pdf if they are correlated). It is thus necessary and important to estimate these error statistics as accurately as possible. To this end, the method of Xu and Wei (2001) was modified to estimate the observation and background error statistics from radar radial-velocity innovations (Xu et al. 2003, henceforth referred to as X03). The method was used to estimate the phased-array radar wind observation error variance and background wind error covariance (Xu et al. 2005, henceforth referred to as X05). The phased-array radar data were collected on 2 June 2004 when a squall line moved southeastward through the central Oklahoma area in the 140 km radial range of the radar scans (Fig. 1 of X05). The background data were provided by the Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS®, Hodur 1997). These data will be used here with a refined method.

The refined method uses the radial-velocity correlation function based on (2.6) of Xu and Gong (2003):

$$R_{vr}(\mathbf{x}_i, \mathbf{x}_j) = [(R_{ll} + R_{tt})\cos(\beta_i - \beta_j) + (R_{ll} - R_{tt})\cos(\beta_i + \beta_j)]/2. \quad (3)$$

Here, $\mathbf{x}_i$ and $\mathbf{x}_j$ are the coordinates of two correlated points with respect to the radar; $R_{ll} = R_{ll}(r)$ and $R_{tt} = R_{tt}(r)$ are the correlation functions for the longitudinal- and transverse-component velocities with respect to $\mathbf{r} = \mathbf{x}_j - \mathbf{x}_i$, respectively; $r = |\mathbf{r}|$; and β_i (or β_j) is the angle of vector $\mathbf{x}_i$ (or $\mathbf{x}_j$) with respect to $\mathbf{x}_j - \mathbf{x}_i$. In X05, (3) was used to model the background error correlation function R_b, while $R_{ll}(r)$ and $R_{tt}(r)$ were expressed by truncated spectral expansions with the correlation range set to $D = D_b = 160$ km [see (4.1) of Xu and Wei 2001]. The refined method makes a further use of (3) to model the observation error correlation function R_o by setting the correlation range to $D = D_o = 3$ km (based on the innovation correlation computed below).

The innovation correlation is computed by $<d_i d_j>$, where d_i is the bias-removed and normalized radial-velocity innovation (observation minus background) at the i-th observation point $\mathbf{x}_i$ and $< >$ denotes the statistical mean (computed from time series). The innovation correlation can be partitioned into

$$<d_i d_j> = \begin{cases} s_b^2 R_b & \text{for } D_b > r \geq D_o = 3 \text{ km} \\ s_b^2 R_b + s_o^2 R_o & \text{for } D_o > r \geq 0, \end{cases} \quad (4)$$

where $s_b = \sigma_b/\sigma_d$, $s_o = \sigma_o/\sigma_d$, σ_b^2 and σ_o^2 are the background and observation error variances, respectively, and σ_d^2 is the innovation variance. Note that $R_b = R_o = 1$ at $r = 0$ and $D_o << D_b$, so $s_b^2 R_b$ can be estimated from $<d_i d_j>$ by the least-squares fitting over $D_b > r \geq D_o$ with s_b^2 given by the interception of the fitted curve on the vertical coordinate at $r = 0$ according to the first equation in (4).

In X03 and X05, s_b^2 and R_b were estimated by the least-squares fitting in a single step. This single-step approach tends to fit $s_b^2 R_b$ over the entire range of $D_b > r \geq D$, so s_b^2 is not always accurately estimated (by extrapolating the fitted curve to $r = 0$). The refined method estimates $s_b^2 R_b$ in two steps: (i) Estimate s_b^2 by fitting a simplified quadratic form of $s_b^2 R_b$ [derived from (3) in the limit of small r] over the range of 10 km $> r \geq D_o$; (ii) Estimate R_b with s_b^2 fixed by fitting the truncated spectral form of R_b over the full range of $D_b > r \geq D_o$. In addition to this two-step approach, the innovation data binning strategy is also refined to facilitate the observation correlation estimation. In X03, innovation data pairs were binned coarsely in the three-dimensional parameter space of $\{r, \cos(\beta_i - \beta_j), \cos(\beta_i + \beta_j)\}$. To enhance the computational efficiency, innovation data pairs were selected and binned along each beam, each pair of two opposite beams, and five selected range circles in X05. Here, with the refined binning strategy, innovation data pairs are binned in two ways: (i) every range gate spacing (0.24 km) along each beam, and (ii) every 1° along each range circle.

The refined method is applied to 25 volume scans from the phased-array radar (2 min per volume from 2100 to 2200 UTC on 2 June 2004), and the data were processed through quality controls (see X05). The background error correlation functions estimated here are similar to those in X05, but the estimated error variances are somewhat different form those in X05. In particular, $s_b^2 = 0.81$ and $\sigma_d^2 = 81.0$ m^{2}s^{-2} are estimated here from innovation data pairs binned along each radar beam between 50 km $\leq r \leq$ 70 km (over 80 range gates) on the lowest tilt (0.75° elevation angle). This gives $\sigma_b = 8.1$ ms^{-1} and $\sigma_o = 3.9$ ms^{-1}. These estimates are considered to be more accurate than those ($\sigma_b = 8.4$ ms^{-1} and $\sigma_o = 2.5$ ms^{-1}) estimated by the single-step fitting in X05.

COAMPS® is a registered trademark of the Naval Research Laboratory.

The estimated observation error correlation functions are plotted in Fig. 2 in comparison with those estimated from the NSSL KOUN radar in Xu et al. (2006). As shown, the phased-array radar radial-velocity observation errors are correlated up to r = 3 km, while the KOUN radial-velocity observation errors are correlated only up to 2 km and the error standard deviation is only 2.1 ms^{-1}. The differences between the two types of observations can be explained in terms of instrumentation error and sampling error. First, the current phased-array radar at NWRT has less detection power and resolution than the KOUN. Limited by the size of the antenna, the phased-array radar beam is wider (about 1.7 degree depending on the viewing angle with respect to the antenna facing direction) than the KOUN beam (about 1 degree). This explains the above differences in terms of instrumentation error. Secondly, the phased-array radar observations were collected for a squall line on 2 June 2004, while the KOUN observations were collected for calm weather on 9 May 2004, so the above differences can be also partially due to different sampling errors.

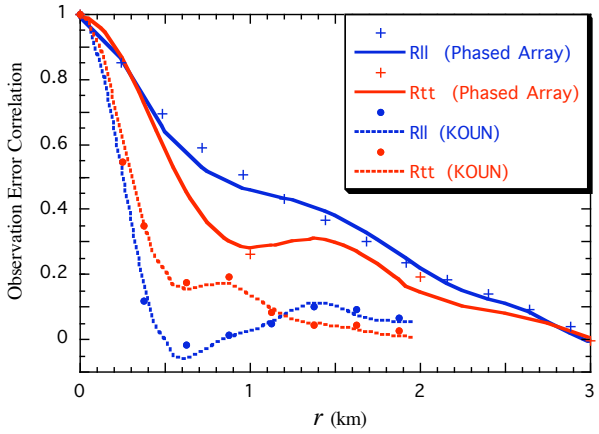

Fig. 2. Observation error correlation data points: + for phased-array radar, and • for KOUN. Estimated observation error correlation functions $R_{ll}(r)$ (blue) and $R_{tt}(r)$ (red): solid for the phased-array radar, and dashed for KOUN.

5 Summary

The theoretical analysis in section 2 and the assimilation experiments in section 3 suggest that the operational WSR-88D radar scans may have excessive spatial resolutions (0.25 km in radial range and about 1° in azimuthal) in radial-velocity observations even for a storm-scale data assimilation system with a resolution of Δx = 2 km. With the phased-array radar, the rapid and flexible agile-beam scans can be configured adaptively to reduce the spatial resolutions and enhance the temporal resolution and/or measurement accuracy. This will reduce or eliminate information redundancy and enhance the information content.

In principle, the singular-value form of the relative entropy derived in (2) can be used to measure the optimality of radar scanning strategy in terms of maximizing the information content from observations for a given data assimilation system. This problem deserves further investigation, especially for practical implementations with an EnKF assimilation system.

The innovation method refined in section 4 can estimate the radar radial-velocity observation covariance as well as the background wind error covariance. The refined method is now being tested with a variety of radar radial-velocity innovation data. The method will be further improved, so it can take as much as possible the advantages of phased-array rapid and flexible scan capabilities to estimate the observation and background error covariances in real time for phased-array radar data assimilation.

Acknowledgements: Drs. Ming Xue and Mingjing Tong kindly provided the ARPS EnKF. The research work was supported by the FAA contract IA# DTFA03-01-X-9007 to NSSL and by ONR Grant N000140410312 to the University of Oklahoma.

References

Forsyth, D. E., J. F. Kimpel, D. S. Zrnic, R. Ferek, J. F. Heimmer, T. J. McNellis, J. E. Crain, A. M. Shapiro, R. J. Vogt, and W. Benner, 2005: The national weather radar testbed (Phased-Array). *32nd Conference on Radar Meteorology*, 24-29 October 2005, Albuquerque, New Mexico, Amer. Meteor. Soc., 12R.3.

Hodur, R. M., 1997: The Naval Research Laboratory's coupled ocean/atmosphere mesoscale prediction system (COAMPS). *Mon. Wea. Rev.*, **125**, 1414-1430.

Majda, A. J., R. Kleeman, and D. Cai, 2002. A mathematical framework for quantifying predictability through relative entropy. *Methods Appl. Anal.* **9**, 425-444.

Whitaker, J. S. and T. M. Hamill, 2002. Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.* **130**, 1913-1924.

Tong, M. and M. Xue, 2005: Ensemble Kalman Filter Assimilation of Doppler Radar Data with a Compressible Nonhydrostatic Model: OSS Experiments. *Mon. Wea. Rev.* **133**, 1789-1807.

Xu, Q. and J. Gong, 2003: Background error covariance functions for Doppler radial-wind analysis. *Q. J. R. Meteorol. Soc.* **129**, 1703-1720.

Xu, Q., K. Nai, and L. Wei, 2006: An innovation method for estimating radar radial-velocity observation error and background wind error covariances. *Quart. J. Roy. Meteor. Soc.*, (accepted subject to minor revision).

Xu, Q., K. Nai, L. Wei, P. Zhang, L. Wang, H. Lu, and Q. Zhao, 2005: Progress in Doppler radar data assimilation. *32nd Conference on Radar Meteorology*, 24-29 October 2005, Albuquerque, New Mexico, Amer. Meteor. Soc., CD-ROM, JP1J7.

Xu, Q., L. Wang, and K. Nai, 2003: Error covariance estimation for Doppler wind data assimilation. Preprints, *31th Conference on Radar Meteorology*, 6–12 August 2003, Seattle, Washington, Amer. Meteor. Soc., 108-109.

Xu, Q., and L. Wei, 2001: Estimation of three-dimensional error covariances. Part II: Analysis of wind innovation vectors. *Mon. Wea. Rev.* **129**, 2939-2954.

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