Doppler Radar Data Assimilation with a Local SVD-En3DVar Method

XU Daosheng[1,2] (徐道生), SHAO Aimei[1,2*] (邵爱梅), and QIU Chongjian[1] (邱崇践)

1 *Key Laboratory for Semi-Arid Climate Change of the Ministry of Education, Key Laboratory of Arid Climate Change and Disaster Reduction of Gansu Province, College of Atmospheric Sciences, Lanzhou University, Lanzhou* 730000

2 *State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing* 100081

(Received December 12, 2011; in final form March 25, 2012)

ABSTRACT

An observation localization scheme is introduced into an ensemble-based three-dimensional variational (3DVar) assimilation method based on the singular value decomposition technique (SVD-En3DVar) to improve assimilation skill. A point-by-point analysis technique is adopted in which the weight of each observation decreases with increasing distance between the analysis point and the observation point. A set of numerical experiments, in which simulated Doppler radar data are assimilated into the Weather Research and Forecasting (WRF) model, is designed to test the scheme. The results are compared with those obtained using the original global and local patch schemes in SVD-En3DVar, neither of which includes this type of observation localization. The observation localization scheme not only eliminates spurious analysis increments in areas of missing data, but also avoids the discontinuous analysis fields that arise from the local patch scheme. The new scheme provides better analysis fields and a more reasonable short-range rainfall forecast than the original schemes. Additional forecast experiments that assimilate real data from 10 radars indicate that the short-term precipitation forecast skill can be improved by assimilating radar data and the observation localization scheme provides a better forecast than the other two schemes.

Key words: Doppler radar, ensemble, data assimilation, 3DVar (three-dimensional variational) method, SVD (singular value decomposition), localization

Citation: Xu Daosheng, Shao Aimei, and Qiu Chongjian, 2012: Doppler radar data assimilation with a local SVD-En3DVar method. *Acta Meteor. Sinica*, **26**(6), 717–734, doi: 10.1007/s13351-012-0604-3.

1. Introduction

Variational data assimilation, both four-dimensional variational (4DVar) and three-dimensional variational (3DVar), and ensemble Kalman filtering (EnKF) are the two most important approaches in current atmospheric data assimilation. One major drawback of the current variational method is the assumption that the background error covariances are static, nearly homogeneous and isotropic (Parrish and Derber, 1998; Courtier, 1997; Cohn et al., 1998; Lorenc, 2003). The EnKF approach (Evensen, 1994; Houtekamer and Mitchell, 2001; Anderson, 2001) provides an alternative to variational data assimilation. The EnKF estimates the flow-dependent background error covariances from an ensemble of short-term forecasts, and it is easy to implement without tangent-linear and adjoint models. The development of hybrid methods that combine the advantages of the variational and EnKF methods has aroused wide attention in recent years. For example, hybrid approaches that combine EnKF and 3DVar (e.g., Hamill and Snyder, 2000; Zupanski, 2005; Buehner, 2005) use flow-dependent background error covariances constructed by ensemble forecasts within a variational framework. Several algorithms have also been designed to combine forecast ensembles and 4DVar (Qiu et al., 2007; Liu et al., 2008; Zhang et al.,

Supported by the Open Project Fund of the State Key Laboratory of Severe Weather of Chinese Academy of Meteorological Sciences, National Natural Science Foundation of China (40875063 and 41275102), and Fundamental Research Fund for Central Universities of China (lzujbky-2010-9).

*Corresponding author: sam@lzu.edu.cn.

2009; Tian et al., 2008; Wang et al., 2009). Experiments using these algorithms have shown that hybrid schemes are able to produce better results than either 3DVar or 4DVar.

Qiu et al. (2007) proposed an ensemble-based 4DVar approach that uses the singular value decomposition (SVD) technique (hereafter called SVD-En4DVar for short). This was later extended by Shao et al. (2009). SVD-En4DVar is a hybrid method that avoids the use of tangent-linear and adjoint models. The SVD technique is used to extract the leading singular vectors from an ensemble of four-dimensional (4D) perturbation fields produced by the model, then a linear combination of the extracted singular vectors is used to fit 4D innovation data (observation minus background) to produce an incremental analysis. Qiu et al. (2007) showed that this method is robust even when the model is imperfect and the observations are incomplete, although a smaller ensemble will produce larger truncation errors in the analysis variables during the expansion process. In addition, this expansion algorithm allows an observation at any point to influence the analysis at all grid points. Spurious correlations caused by sample errors will lead to larger analysis errors in the absence of observations. This phenomenon is especially obvious for radar data assimilation with incomplete observations. The analysis variables are expanded on a series of base vectors (the number of which cannot exceed the number of forecast members) using the SVD technique.

For the aforementioned reasons, a localization technique is imperative for application of the SVD-En4DVar method to meso- and micro-scale data assimilation. The forecast error covariances are not used directly in the SVD-En4DVar method; it is therefore difficult to apply covariance localization techniques such as the Schur product method (Gaspari and Cohn, 1999). Xu et al. (2011) recently adopted a local patch localization scheme in an emsemble-based 3DVar approach (SVD-En3DVar) to reduce the influence of spurious long-distance correlations among error covariances. In this localization scheme, the analysis domain is separated into many subdomains (called local patches). The analysis is then performed independently in each local patch. This procedure limits the influence of each individual observation to a smaller region. This scheme is equivalent to achieving localization with a heaviside weighting function (i.e., the weight of an observation is equal to 1 within the local patch and 0 outside of the local patch). As a result, the analysis increment is discontinuous at the edges of each local patch. Furthermore, observations that are far from the center of the local patch may still have a disproportionately large contribution to the analysis increment at the center point.

Hunt et al. (2007) solved similar problems in the local EnKF by applying an observation localization scheme in which the observational error covariance was weighted according to distance from the center of the local patch. In this paper, we introduce a similar observation localization scheme into SVD-En3DVar to evaluate the resulting improvement in assimilation skill. The SVD-En3DVar results with the global scheme and the local patch scheme are provided for comparison. The impacts of the different localization schemes are tested by conducting a series of assimilation and forecast experiments in which the SVD-En3DVar method is applied to simulated and real radar observations within the Weather Research and Forecasting (WRF) model framework.

This paper is organized as follows. The SVD-En3DVar method and two localization schemes are described in Section 2. The model and experimental design are described in Section 3. Test results using a single radar observation are shown in Section 4. Model experiments using simulated radar data are reported and the sensitivity of the results to various localization parameters is discussed in Section 5. Experiments using real radar data are presented in Section 6. The results are summarized and discussed in Section 7.

2. SVD-En3DVar method and localization schemes

Consider an assimilation step performed at time $t = t_0$. A suitable model integration time (denoted by τ) is chosen to generate the perturbed forecast ensemble. Perturbation fields are produced for each member of the forecast ensemble according to the method described by Xu et al. (2011). In this method, pseudorandom perturbations in temperature and specific humidity fields are taken as observation innovations. A 3DVar system is then used to assimilate these observation innovations and generate the perturbation fields for all variables. These perturbation fields are then added to the initial fields at time $t = t_0 - \tau$, yielding a set of perturbed initial fields. The perturbed forecast ensemble (which contains M members denoted by $\boldsymbol{u}_m$ with $m = 1, \cdots, M$) is obtained by integrating the model with this set of perturbed initial fields from $t = t_0 - \tau$ to $t = t_0$. The forecast with unperturbed initial fields is taken as the background state (denoted by $\boldsymbol{u}_{\mathrm{b}}$). Two sets of perturbations are then calculated, $\Delta\boldsymbol{u}_m = \boldsymbol{u}_m - \boldsymbol{u}_{\mathrm{b}}$ in grid space and $\Delta\boldsymbol{d}_m = H(\boldsymbol{u}_m) - H(\boldsymbol{u}_{\mathrm{b}})$ in observation space, where H is the observation operator. Combining the perturbations in grid space and observation space yields the vector

$$\boldsymbol{a}_m = (\Delta\boldsymbol{u}_m^{\mathrm{T}}, \Delta\boldsymbol{d}_m^{\mathrm{T}})^{\mathrm{T}}. \tag{1}$$

The dimension of vector $\boldsymbol{a}_m$ is $N_v \times N_x \times N_{\mathrm{o}}$, where N_v is the number of the model variables, N_x is the number of the spatial grid points, and N_{o} is the number of observations. A matrix $\boldsymbol{A}$ is constructed from these deviation vectors as

$$\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_M). \tag{2}$$

The singular value decomposition (SVD) of $\boldsymbol{A}$ yields:

$$\boldsymbol{A} = \boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}, \tag{3}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix composed of the singular values of $\boldsymbol{A}$ with $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_M > 0$. $\boldsymbol{B}$ is an $N \times M$ rectangular matrix composed of the first M left singular vectors of $\boldsymbol{A}$, and $\boldsymbol{V}$ is an orthogonal matrix composed of the right singular vectors of $\boldsymbol{A}$. Similar to the mth column vector of $\boldsymbol{A}$ in Eq. (1), the mth column vector of $\boldsymbol{B}$ can be written in the partitioned form

$$\boldsymbol{b}_m = (\boldsymbol{b}_m^{u\mathrm{T}}, \boldsymbol{b}_m^{d\mathrm{T}})^{\mathrm{T}}, \tag{4}$$

where $\boldsymbol{b}_m^u$ and $\boldsymbol{b}_m^d$ correspond to $\Delta\boldsymbol{u}_m$ and $\Delta\boldsymbol{d}_m$ in Eq. (1), respectively.

The vector $\boldsymbol{x} = (\Delta\boldsymbol{u}^{\mathrm{T}}, \boldsymbol{d}^{\mathrm{T}})^{\mathrm{T}}$ can be expressed as the linear combination of the leading p singular vectors in $\boldsymbol{B}$:

$$\boldsymbol{x} = \sum_{r=1}^{p} \boldsymbol{\alpha}_r \boldsymbol{b}_r = \boldsymbol{b}\boldsymbol{\alpha}. \tag{5}$$

From Eq. (4), we obtain

$$\Delta\boldsymbol{u} = \sum_{r=1}^{p} \boldsymbol{\alpha}_r \boldsymbol{b}_r^u = \boldsymbol{b}^u\boldsymbol{\alpha} \tag{6}$$

and

$$\Delta\boldsymbol{d} = \sum_{r=1}^{p} \boldsymbol{\alpha}_r \boldsymbol{b}_r^d = \boldsymbol{b}^d\boldsymbol{\alpha}. \tag{7}$$

The 3DVar cost function is

$$\begin{aligned}\boldsymbol{J}(\Delta\boldsymbol{u}) = {} & \Delta\boldsymbol{u}^{\mathrm{T}}\boldsymbol{P}^{-1}\Delta\boldsymbol{u} + (\boldsymbol{H}\Delta\boldsymbol{u} - \Delta\boldsymbol{y})^{\mathrm{T}} \\ & \cdot\boldsymbol{O}^{-1}(\boldsymbol{H}\Delta\boldsymbol{u} - \Delta\boldsymbol{y}),\end{aligned} \tag{8}$$

where $\boldsymbol{P}$ is the background error covariance matrix. This matrix is similar to that used in conventional EnKF, where $\boldsymbol{P} \approx \boldsymbol{b}^u\boldsymbol{\Lambda}_P^2(\boldsymbol{b}^u)^{\mathrm{T}}/(M-1)$. $\boldsymbol{O}$ is the observation error covariance matrix, $\boldsymbol{y}$ is the observation and $\Delta\boldsymbol{y} = \boldsymbol{y} - \boldsymbol{H}\boldsymbol{u}_b$. Using Eqs. (6) and (7), the cost function in control variable $\boldsymbol{\alpha}$ space can be written as:

$$\begin{aligned}\boldsymbol{J}(\boldsymbol{\alpha}) = {} & (M-1)\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\Lambda}_P^{-2}\boldsymbol{\alpha} \\ & + \sum_{r=1}^{p}(\boldsymbol{\alpha}_r\boldsymbol{b}_r^d - \Delta\boldsymbol{y})^{\mathrm{T}}\boldsymbol{O}^{-1}(\boldsymbol{\alpha}_r\boldsymbol{b}_r^d - \Delta\boldsymbol{y}).\end{aligned} \tag{9}$$

The control variable $\boldsymbol{\alpha}$ is obtained by minimizing this cost function, and the analysis increment in the patch can be computed according to Eq. (6).

The observation operators for radial velocity V_{r} and reflectivity Z are

$$V_{\mathrm{r}} = u\frac{x - x_i}{r_i} + v\frac{y - y_i}{r_i} + (w - v_T)\frac{z - z_i}{r_i} \tag{10}$$

and

$$Z = 43.1 + 17.5\lg(\rho q_{\mathrm{r}}), \tag{11}$$

respectively, where u, v, and w represent the zonal, meridional, and vertical components of wind, (x, y, z) is the location of the radar, (x_i, y_i, z_i) is the location of observation i, r_i is the distance between the radar and the observation, v_T (m s^{-1}) is the mass-weighted fall velocity of rainwater, ρ is the density of air, and $q_{\rm r}$ is the density of rainwater. The control variables are the three wind components (u, v, and w), the perturbed potential temperature ($\theta_{\rm p}$), the perturbed geopotential ($H_{\rm p}$), and the mixing ratios of water vapor ($q_{\rm v}$), rain water ($q_{\rm r}$), and cloud water ($q_{\rm c}$).

In the original SVD-En4DVar (Shao et al., 2009), the analysis field is treated as a whole and all of the observations are assimilated simultaneously. This approach (hereafter called the global (GB) scheme) means that p-dimensional orthogonal vectors are used to fit the entire set of observations and analysis fields. Many studies of EnKF have demonstrated that localization techniques are necessary if the size of the ensemble used to estimate background error covariance is small. The Schur product method (Gaspari and Cohn, 1999) is widely used for localization in EnKF. The Schur product method reduces covariances through an elementwise multiplication of the background error covariance matrix and a correlation function with local support. This procedure acts to smooth the increment fields because it reduces the background error covariance in a distance-dependent manner, with greater reductions farther from the observation. The Schur product method could be applied in SVD-En3DVar by multiplying the correlation function matrix and the matrix $\boldsymbol{A}\boldsymbol{A}^{\rm T}$; however, the dimension of matrix $\boldsymbol{A}\boldsymbol{A}^{\rm T}$ is very high ($N = N_v \times N_x \times N_{\rm o}$). The eigenvector decomposition of $\boldsymbol{A}\boldsymbol{A}^{\rm T}$ would be very difficult even for a small local patch. The dimension of the matrix $\boldsymbol{A}^{\rm T}\boldsymbol{A}$ ($M \times M$) is much lower than that of $\boldsymbol{A}\boldsymbol{A}^{\rm T}$; therefore, the eigenvector decomposition of $\boldsymbol{A}^{\rm T}\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}^2\boldsymbol{V}^2$ could be used to obtain $\boldsymbol{V}$ and $\boldsymbol{\Lambda}$. $\boldsymbol{B}$ would then be computed as $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{V}\boldsymbol{\Lambda}^{-1}$ according to Eq. (3). Even so, the Schur product method is still not suitable for application within SVD-En3DVar.

Xu et al. (2010) introduced a local patch scheme (hereafter called the LP scheme) into SVD-En3DVar. The LP scheme is a point-to-point assimilation algorithm that separates the global grid into independent local patches with horizontal and vertical grid lengths $l_{\rm h}$ and $l_{\rm v}$, respectively. Each grid point has its own local patch and is located at the center of a local patch. The assimilation process is performed independently for each patch, then the analysis values at the center point of each local patch are combined to obtain the global analysis field. In this way, the impact of any given observation is confined to its own local patch, and the dimension of matrix $\boldsymbol{A}$ is much lower in each local patch than in the full domain. The defect of this scheme is the fixed weight assigned to each observation, which leads to discontinuous analysis increments at the edge of each local patch and does not completely eliminate the influence of spurious long-distance correlations within a local patch. Fortunately, it is convenient to adopt observation localization rather than covariance localization under this point-to-point assimilation algorithm (Hunt et al., 2007; Miyoshi and Yamane, 2007). Observation localization is realized by multiplying observational error variance by the inverse of a localization weighting function (such as the Gaussian function). In this paper we introduce an observation localization scheme (hereafter called the OL scheme) into SVD-En3DVar. The Gaussian function is used as the weighting function for observation localization:

$$w(\sigma_{\rm h}, \sigma_{\rm v}) = \begin{cases} \exp\Big[\Big(-\dfrac{r_{\rm h}^2}{\sigma_{\rm h}^2}\Big)\cdot\Big(-\dfrac{r_{\rm v}^2}{\sigma_{\rm v}^2}\Big)\Big], \\ \qquad\qquad r_{\rm h} \leqslant l_{\rm h} \text{ and } r_{\rm v} \leqslant l_{\rm v}, \\ 0, \qquad\quad r_{\rm h} > l_{\rm h} \text{ or } r_{\rm v} > l_{\rm v}, \end{cases} \tag{12}$$

where $r_{\rm h}$ and $r_{\rm v}$ denote the horizontal and vertical distance between the observation point and the center of the local patch, respectively; $l_{\rm h}$ and $l_{\rm v}$ are the same horizontal and vertical grid lengths of the local patch as in the LP scheme; and $\sigma_{\rm h}$ and $\sigma_{\rm v}$ are the localization scale parameters in the horizontal and vertical directions, respectively. As in the LP scheme, the assimilation is performed independently for each local patch in the OL scheme; however, the singular vectors adopted during the assimilation are the same as those used in the GB scheme (i.e., those obtained from the decomposition of matrix $\boldsymbol{A}$ over the whole domain). This approach results in considerable savings in computational cost relative to the LP scheme (in which the singular vectors are calculated independently for each local patch).

3. Model and experimental design

The advanced WRF model is used to perform the assimilation and forecast experiments, with initial and boundary conditions provided by NCEP-FNL 1-degree analysis data. All experiments were conducted on a 180×180 grid mesh with a grid spacing of 7 km. The center of the analysis domain is located at 22.8°N, 112.8°E. The model is integrated on 27 vertical layers, with the model top at 50 hPa. The Ferrier scheme is used to model microphysical processes in all experiments. The assimilation time is set to 1200 UTC 6 June 2008. The background state is generated according to a 6-h forecast integrated from 0600 UTC 6 June 2008 to the analysis time. The background horizontal wind field on the 8th model level ($\sigma = 0.86$; approximately 850 hPa) is shown in Fig. 1. The flow is strong and southwesterly near the Leizhou Peninsula, with obvious convergence near the center of Guangdong Province.

The ensemble samples are produced according to the method described in Section 2, with the size of the ensemble fixed at 30. Horizontal wind perturbations on the 8th model level of the first ensemble member at the initial time and analysis time are shown in Fig. 2. The magnitude of the initial perturbation is large in most areas except for Jiangxi, with a number of eddies (Fig. 2a). Following the 6-h forecast, the perturbation is mainly located in the areas of strongest rainfall over Guangdong and the sea south of Guangdong; the perturbation is relatively weak in all other places (Fig. 2b).

The radial velocity and reflectivity are assimilated into SVD-En3DVar using the three schemes outlined above (the GB, LP, and OL schemes). Three sets of assimilation experiments are performed for each scheme,

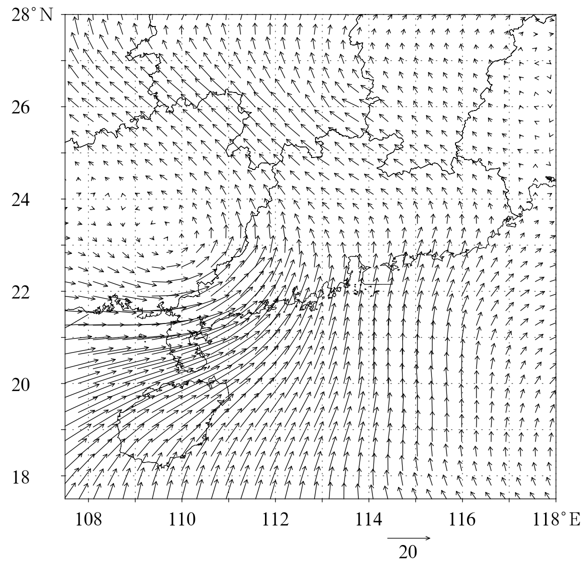

Fig. 1. Background horizontal wind field on the 8th model level at 1200 UTC 6 June 2008.

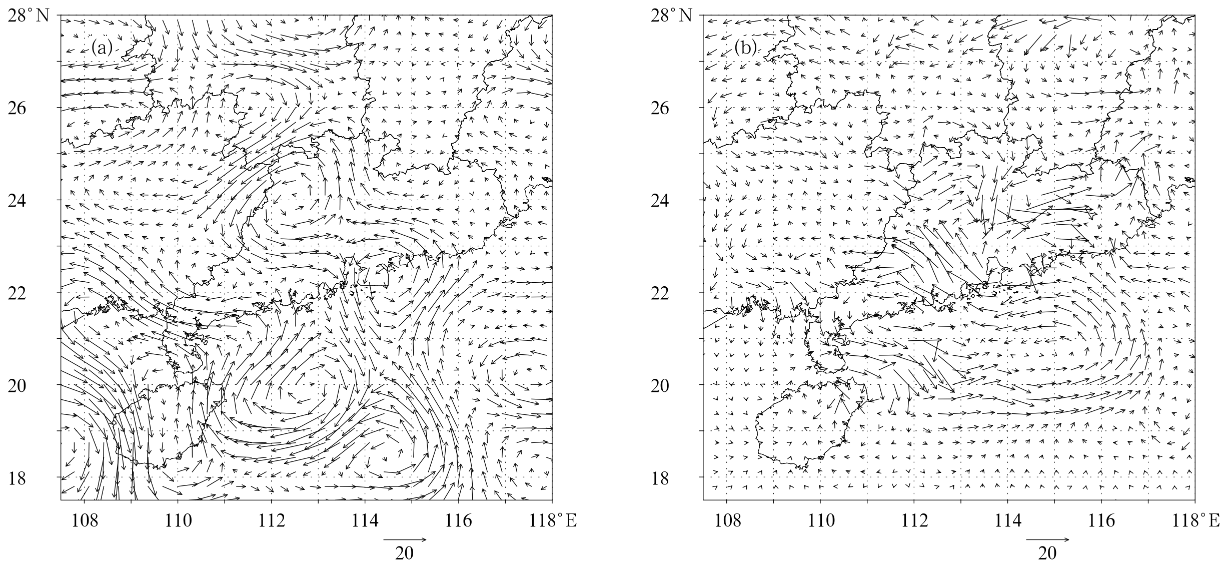

Fig. 2. Perturbed horizontal wind field on the 8th model level according to the first ensemble member at (a) the initial time (0600 UTC 6 June 2008) and (b) the analysis time (1200 UTC 6 June 2008).

i.e., a single observation experiment, a simulated observation experiment, and a real observation experiment. The results of these experiments are analyzed and compared in the following sections.

4. Experiments with a single radar observation

A single observation experiment is performed to test the efficiency of the OL scheme in SVD-En3DVar, with the other two schemes (the GB and LP schemes) used as a comparison. The grid point (98, 102) on the 8th model level is selected as the position of the single observation. This grid point is located to the north of the Guangzhou radar (grid point (98, 94)). The observation innovation is set to 5 m s^{-1}.

The observation is located directly northward of the radar site; the increment of the radar radial velocity observation is therefore very close to the meridional component (v) of horizontal wind. Figure 3 shows the analysis increment of v at the 8th model level after the single observation is assimilated using each of the above three schemes. The assimilation of a single observation impacts the whole analysis field when the (GB) scheme is used (Fig. 3a), with relatively large analysis increments even at grid points located far from the observation. Non-zero analysis increments are confined to the local patch when the LP scheme is used (Fig. 3b), with all analysis increments outside of the local patch equal to zero. This distribution results in a discontinuity at the edge of the local patch. Non-zero analysis increments are mainly distributed near the observation when the OL scheme is used (Fig. 3c), with reduced magnitudes at increasing distance from the observation. The sharp discontinuity intrinsic to the LP scheme is greatly alleviated when the OL scheme is applied instead.

5. Experiments with simulated radar data

5.1 Simulated radar data

The simulated observations used in the observing system simulation experiments (OSSEs) are distributed according to the real radar network. The radar network consists of 10 S-band Doppler radars located at Haikou, Guangzhou, Shaoguan, Meizhou, Shantou, Yangjiang, Liuzhou, Nanning, Guilin, and Longyan, respectively. This network covers most of the southern part of South China. Each radar volume scan covers 9 elevation angles: 0.5°, 1.5°, 2.4°, 3.4°, 4.3°, 6.0°, 9.9°, 14.6°, and 19.5°. The gate spacing is 1.0 km for reflectivity and 0.25 km for radial velocity. The locations of the 10 radars are indicated in Fig. 4a, and the distribution of available observational data on the 8th model level is shown in Fig. 4b. The simulated radar data are interpolated to the model grid for analysis.

The simulated radar data are calculated by adding random errors to the "true" field. The standard deviation of the random errors is 1.0 m s^{-1} for radial velocity and 5 dBZ for reflectivity. The "true" field is generated by integrating the WRF model from an initial state to the analysis time, where the initial state is obtained by adding a perturbation field to the background field at 0600 UTC 6 June 2008. The perturbation field is obtained by subtracting the 24-h forecast valid at 0600 UTC 6 June 2008 from the 12-h forecast valid at the same time. Figure 5 shows the true horizontal wind field on the 8th model level. The center of the cyclonic vortex is strengthened and shifted eastward (to the middle of Guangdong Province) relative to the background field.

5.2 Sensitivity to the localization parameters

Several sets of OSSEs were conducted to examine the sensitivity of the OL scheme to the values of the observation localization parameters, with particular focus on the localization scales in the horizontal and vertical directions. In one set of experiments, the horizontal localization scales were set variously to 3, 5, 7, or 10 grid lengths without vertical localization. This set of experiments was then used to investigate the sensitivity of the assimilation to the horizontal localization parameters, and was compared with results obtained using the LP scheme with $l_{\mathrm{h}} = 20$ and $l_{\mathrm{v}} = 20$. The results are listed in Table 1. No matter which of the four localization scales was used, the root-mean-square errors (RMSEs) from the OL scheme are smaller than those from the LP scheme. The optimal

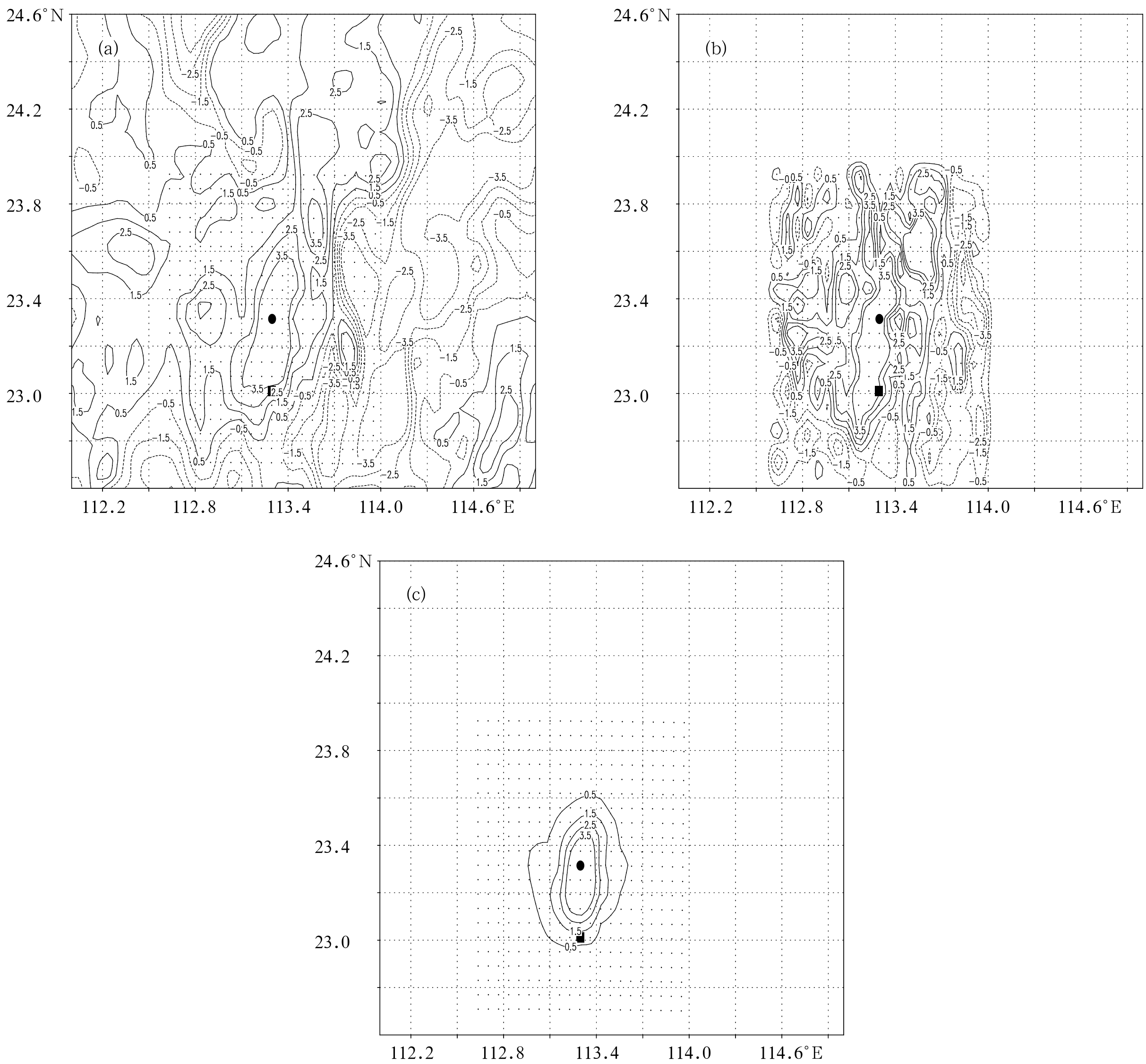

Fig. 3. Distribution of analysis increments with the assimilation of a single observation on the 8th model level. Assimilation is accomplished using the three schemes outline in Section 3: (a) the GB scheme; (b) the LP scheme; and (c) the OL scheme. The dotted area indicates the area of the local patch. The solid circle represents the observation point and the solid square represents the location of the Guangzhou radar site.

assimilation is obtained with localization scale $\sigma_{\rm h} = 5$. The accuracy of the analysis is slightly sensitive to the choice of horizontal localization scales.

The horizontal localization scale $\sigma_{\rm h}$ was then fixed to 5 grid spaces and the vertical localization scale varied to 3, 5, and 9 grid spaces. This set of experiments was used to investigate the sensitivity of the analysis to the vertical localization scale. The analysis errors decrease slightly when vertical observation localization is used (Table 2). The result is best when $\sigma_{\rm v} = 5$. Therefore, $\sigma_{\rm h}$ and $\sigma_{\rm v}$ are both fixed to 5 grid spaces in the following experiments.

5.3 Analysis increments in horizontal winds

The analysis increment in the horizontal wind fields is compared with the true increments (i.e., the difference between the true field and the background; Fig. 6a) to illustrate the performance of the three

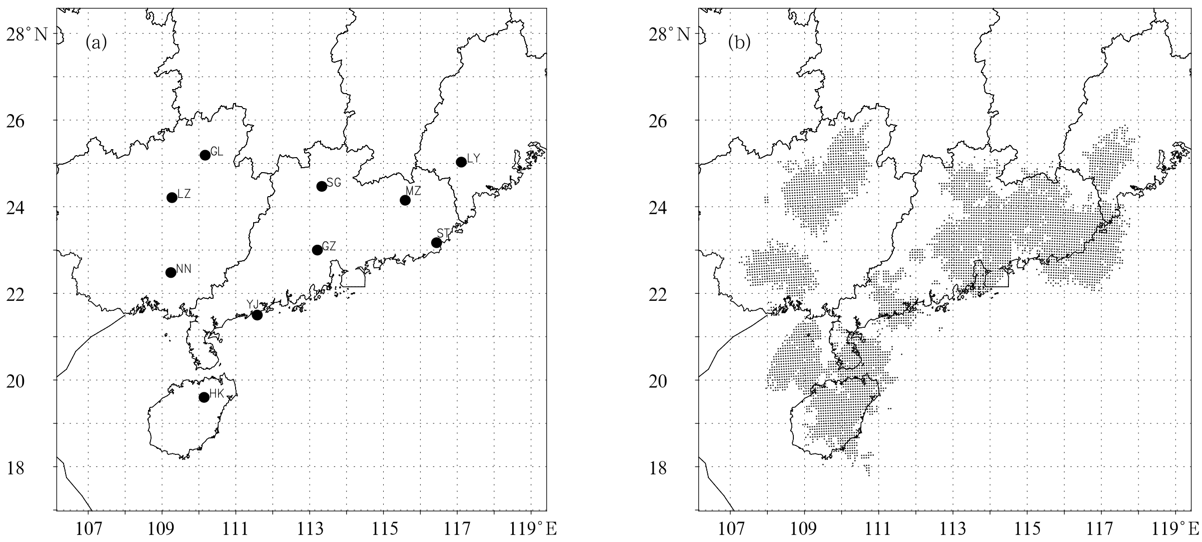

Fig. 4. (a) Location of the 10 radars in the real radar network in South China and (b) distribution of available simulated observation data on the 8th model level.

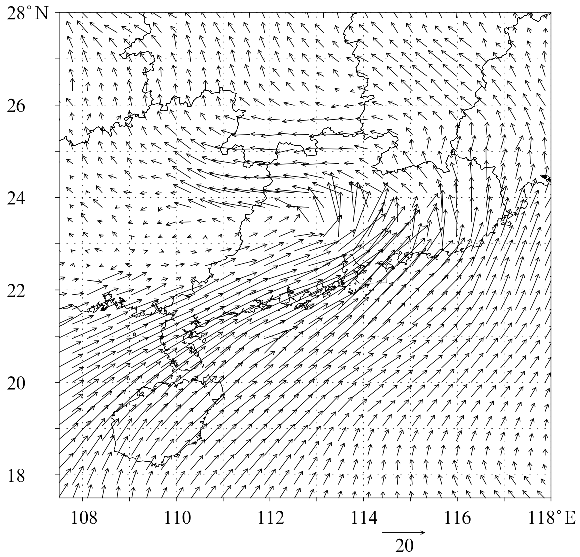

Fig. 5. True horizontal wind field on the 8th model level at the analysis time (1200 UTC 6 June 2008).

Table 1. Root-mean-square error (RMSE) of analysis assimilated using different horizontal localization scales ($\sigma_{\rm h}$)

	u (m s^{-1})	v (m s^{-1})	$\theta_{\rm p}$ (K)	$q_{\rm v}$ (kg kg^{-1})
Background	3.47	3.44	0.886	7.69E-04
LP	3.17	3.21	0.879	7.63E-04
OL ($\sigma_{\rm h}$=3)	3.02	3.13	0.877	7.35E-04
OL ($\sigma_{\rm h}$=5)	2.93	3.07	0.870	7.32E-04
OL ($\sigma_{\rm h}$=7)	2.95	3.08	0.873	7.34E-04
OL ($\sigma_{\rm h}$=10)	2.96	3.10	0.874	7.36E-04

Table 2. Root-mean-square error (RMSE) of analysis field with different vertical localization scales ($\sigma_{\rm v}$) and fixed horizontal localization scale ($\sigma_{\rm h}$ = 5)

	Without vertical localization	With vertical localization		
		$\sigma_{\rm v}$=3	$\sigma_{\rm v}$=5	$\sigma_{\rm v}$=9
u (m s^{-1})	2.93	2.91	2.87	2.91
v (m s^{-1})	3.07	3.04	3.00	3.03
$\theta_{\rm p}$ (K)	0.870	0.858	0.857	0.863
$q_{\rm v}$ (kg kg^{-1})	7.32E-04	7.32E-04	7.28E-04	7.28E-04

localization schemes in SVD-En3DVar. The analysis increment for the GB scheme (Fig. 6b) is consistent with the true increment in some data-void regions, such as in the area of the cyclonic vortex over the sea southeast of Guangdong. This consistency indicates that the GB scheme has some ability to reasonably extend observed information to data-void areas. However, larger spurious analysis increments arise in some regions far from the observation points, such as west of Guangxi, Hunan, Jiangxi, Fujian, and the South China Sea, suggesting that error covariance between these long-distance grid points is suspect. These spurious analysis increments are reduced substantially when the OS and LP schemes are used. Some spurious analysis increments still persist over the sea south of Guangdong when the LP scheme is used (Fig. 6c), and the analysis increment is discontinuous in southern Jiangxi Province. This discontinuity is caused by the sudden change in the weight assigned to specific observations at the edge of a local patch in the LP scheme (in contrast to the fixed weight of observations inside the local patch). These limitations are solved in the OL scheme by the introduction of a distance-dependent observation weighting scheme.

In areas where observations are dense, such as Guangdong Province, the analysis increment is consistent with the true increment in all the three schemes. Figure 6 indicates that the analysis increment is smaller than the true increment when the GB scheme is used. The primary reason for this discrepancy is that the GB scheme uses limited singular vectors to fit all observations over the experimental domain, leading to the loss of observed information. These problems are further explored by calculating the relative observation innovation

$$\varepsilon = \frac{\sqrt{\sum\limits_{i=1}^{n}[y_i - H(u_{\mathrm{a}i})]^2}}{\sqrt{\sum\limits_{i=1}^{n}[y_i - H(u_{\mathrm{b}i})]^2}}, \tag{13}$$

where y is the observation, u_{a} and u_{b} are analysis field

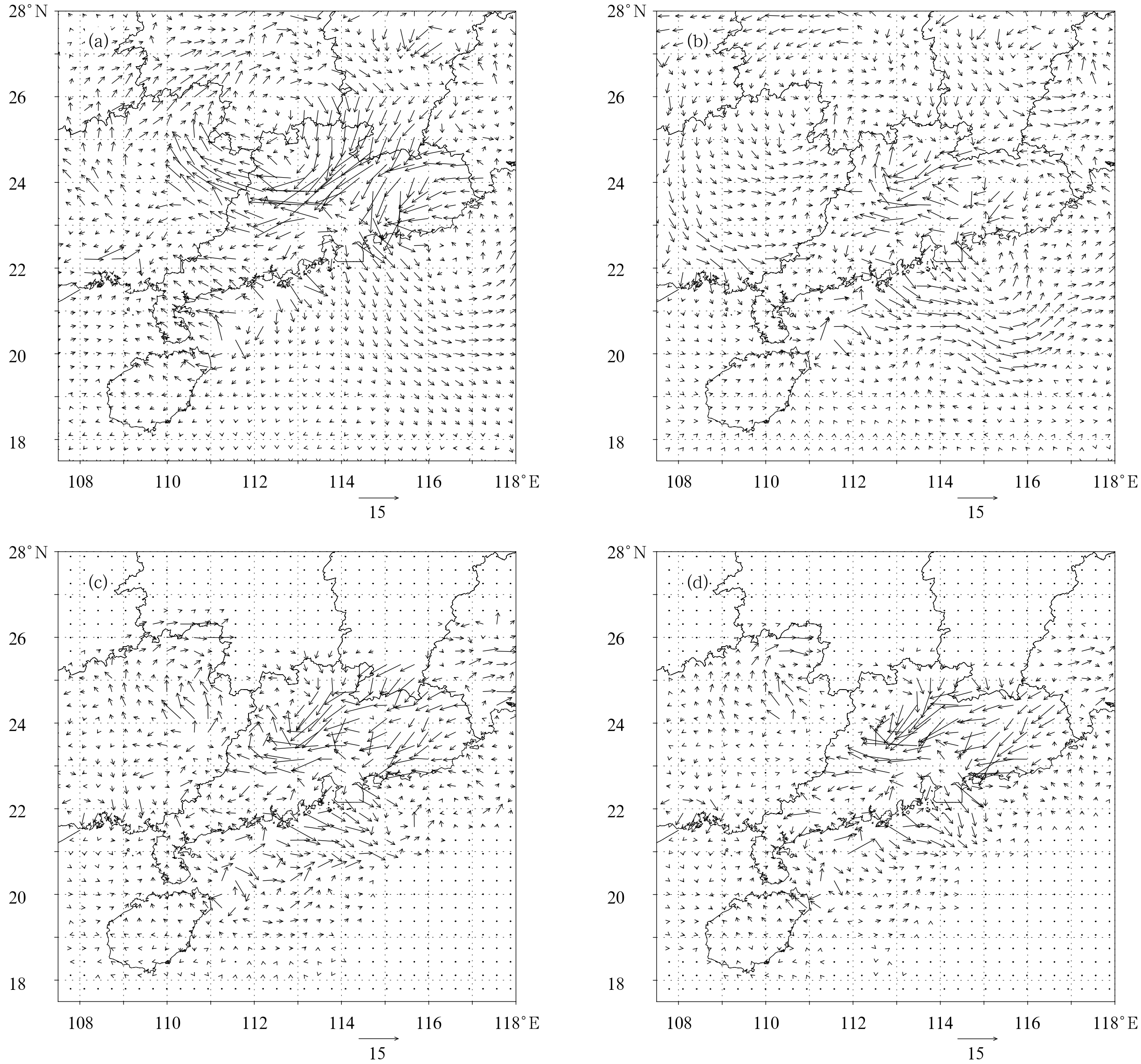

Fig. 6. Increment of horizontal winds on the 8th model level according to (a) the true state of the model, (b) analysis using the GB scheme, (c) analysis using the LP scheme, and (d) analysis using the OL scheme.

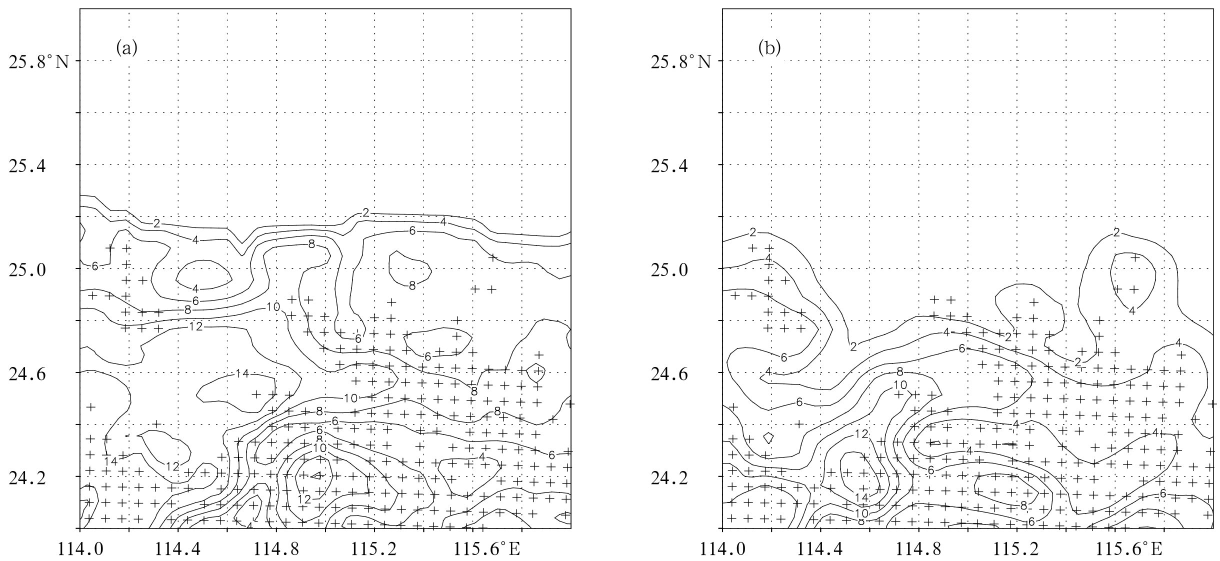

Fig. 7. Analysis increment of horizontal winds on the 8th model level using (a) the LP scheme and (b) the OL scheme. The symbol "+" denotes an observation point.

and background field, respectively, and H denotes the observational operator. The magnitude of the relative observation innovation reflects the contribution of the observations to the analysis. That is to say, the larger relative observation innovations mean that less observational information is merged into the analysis. The relative observation innovation is 0.891 for the GB scheme, 0.616 for the LP scheme, and 0.628 for the OL scheme. The relatively high value for the GB scheme implies that relatively little observational information is absorbed when the GB scheme is used; this situation is obviously improved when either of the two localization schemes (LP or OL) is introduced.

Figure 7 shows the analysis increment for horizontal winds in a smaller domain (24°–26°N, 114°–116°E) with only radial velocity observations on the 8th model level assimilated to better understand the performance of the LP and OL localization schemes. The LP and OL schemes are consistent in areas where observations are dense; however, a large gradient in the analysis increment occurs at the edge of the observing network when the LP scheme is used. This gradient is substantially reduced when the OL scheme is used, as mentioned earlier.

5.4 Precipitation forecast

The previous section shows that introducing a localization technique (especially the OL scheme) can improve the assimilation skill of SVD-En3DVar. In this section, a series of forecasting experiments are conducted both with and without the above assimilation schemes to investigate the impacts of the localization technique on precipitation forecasts. The 12-h forecast results are shown at 6-h intervals in Figs. 8 and 9. The actual 6-h cumulative precipitation from 1200 to 1800 UTC 6 June 2008 was mainly distributed in the middle of Guangdong Province (Fig. 8a). A small amount of precipitation occurred in southeastern Guangxi. The intensity and distribution of forecasted precipitation without observational data assimilation is obviously different from the actual intensity and distribution. For example, the forecast underestimates the intensity of rainfall in eastern Guangdong and predicts spurious rainfall in southern Jiangxi, while the distribution of predicted rainfall in southeastern Guangxi is inconsistent with the actual distribution. The precipitation forecast can be improved with assimilation of observational data (Figs. 8c, d). When observational data are assimilated, the intensity of the predicted rainfall in eastern Guangdong is stronger and the spurious rainfall forecast for southern Jiangxi is weaker. The precipitation forecast in eastern Guangdong is much closer to the truth when the OL localization scheme is used than when the LP

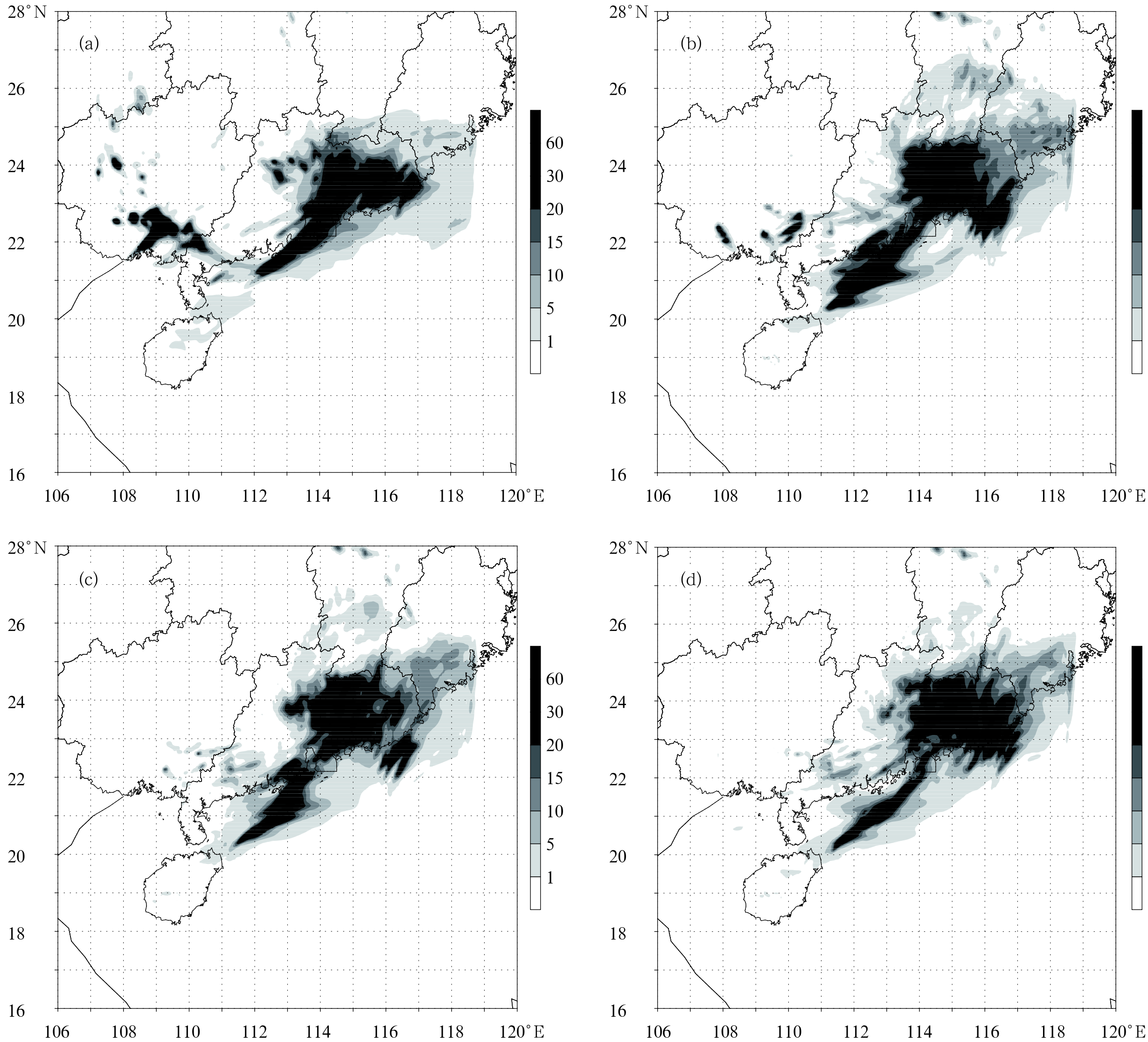

Fig. 8. 6-h cumulative precipitation from 1200 to 1800 UTC 6 June 2008 in experiments with simulated radar data. (a) Observation, (b) forecast without data assimilation, (c) forecast with data assimilation using the LP scheme, and (d) forecast with data assimilation using the OL scheme.

localization scheme is used. The spurious precipitation in the forecast for southern Jiangxi persists when the LP scheme is used; it is eliminated when the OL scheme is used. The application of the OL scheme in SVD-En3DVar substantially improves the precipitation forecast in this case. The results for the subsequent 6 hours (Fig. 9) are similar to those shown in Fig. 8, although the improvement of the forecast with data assimilation is less clear.

The quality of the rainfall forecast is quantitatively evaluated using the threat score (TS). Figure 10 shows TSs calculated using different precipitation thresholds. The TS for cumulative precipitation over the first 6-h period is higher when data assimilation is performed and is the highest when the OL scheme is used. When data assimilation is performed for the subsequent 6-h period, the TS is slightly higher than or equal to that without assimilation.

5.5 Extended assimilation experiments

The experiments described above indicate that the OL scheme represents an improvement over the

LP scheme. The effects of the assimilation are not always obvious, however, especially for non-observed variables such as temperature and humidity. This may result from the limited nature of the background field at the start time of assimilation process, as this field seldom contains mesoscale information, or an insufficiently long integration time for the perturbation samples. In this case, the estimated background error covariance cannot accurately reflect the structure of the forecast error. A pair of assimilation cycle experiments using the OL and LP localization schemes is performed to improve the effects of the data assimilation. These experiments use the assimilation cycle technique to institute a continuous 2-h assimilation process at 30-min intervals from the original assimilation time (1200 UTC 06 June 2008) to 1400 UTC 6 June 2008. The procedure is similar to that used in the traditional EnKF. The perturbed forecast ensemble contains 30 members, obtained using the method described in Section 2. These ensemble members are used as initial fields for 30-min forecasts conducted using the WRF model. These forecasts represent a new ensemble of initial conditions for the next assimilation cycle. Five of these assimilation cycles are performed using each of the OL and LP localization schemes. A 6-h precipitation forecast is subsequently performed to

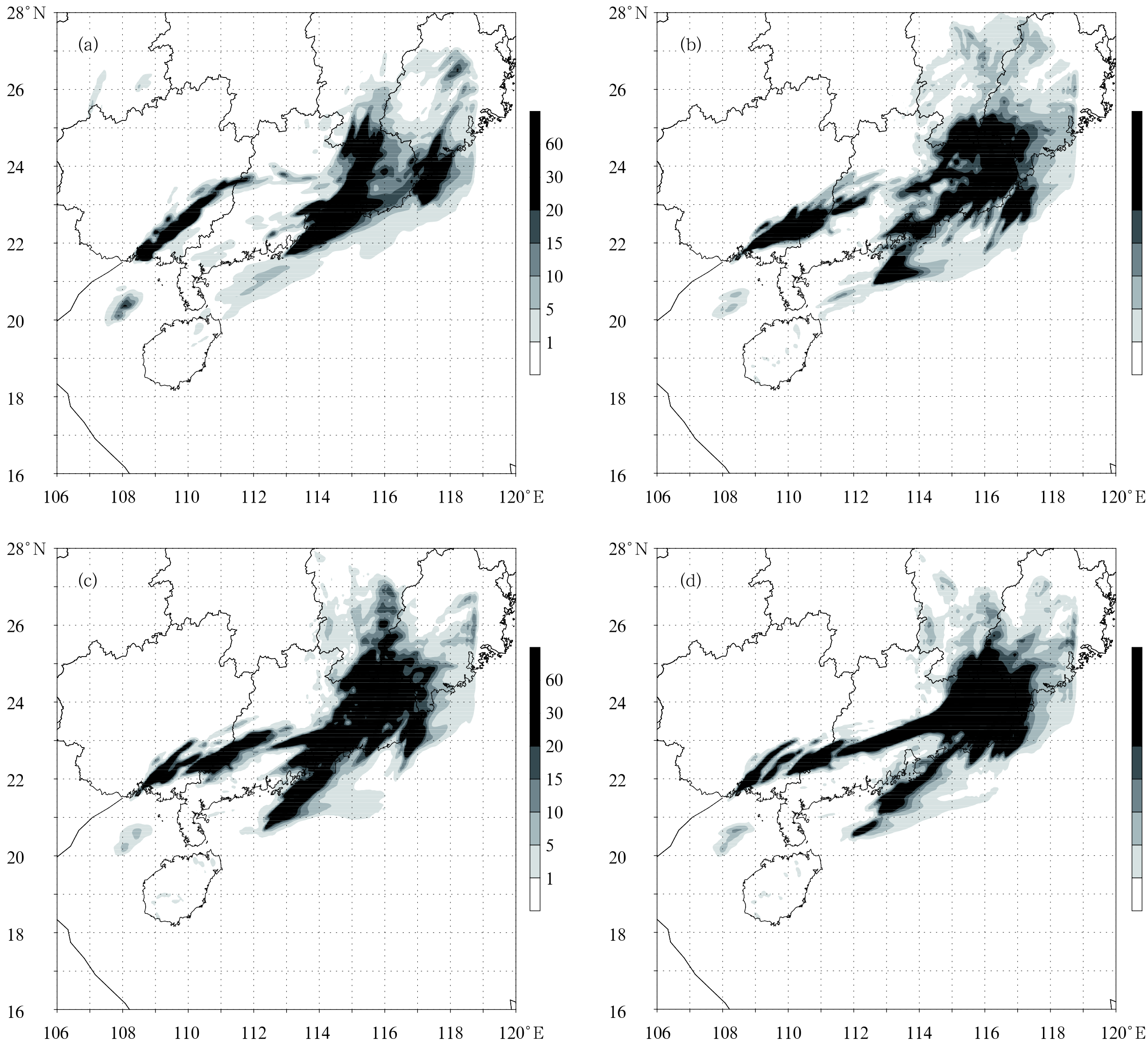

Fig. 9. As in Fig. 8, but for the 6-h period from 1800 UTC 6 to 0000 UTC 7 June 2008.

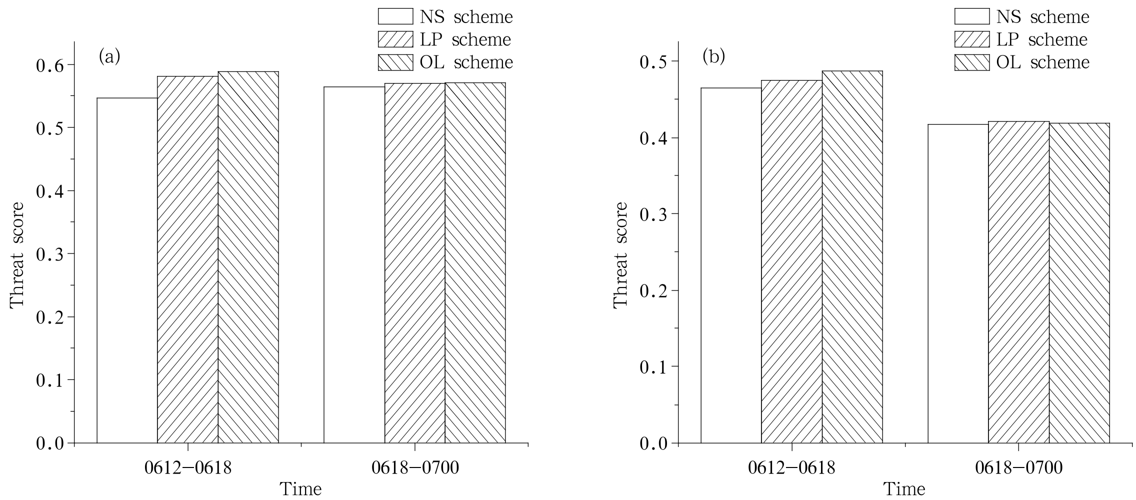

Fig. 10. Threat scores of 6-h cumulative rainfall forecasts without assimilation of simulated observations (NS) and with data assimilation using either the LP or OL localization scheme (LP and OL, respectively). (a) Precipitation threshold = 1 mm and (b) precipitation threshold = 15 mm.

test and compare the effect of the assimilation when the analysis field at 1400 UTC 6 June 2008 is used as the initial field.

The RMSEs of the analysis fields using the LP and OL schemes at 1400 UTC 6 June are listed in Table 3. The results without data assimilation also listed here for comparison (first row). The RMSEs of the analysis fields are substantially reduced after five assimilation cycles, especially the errors of temperature and humidity. The assimilation provides a greater improvement when the OL scheme is used than when the LP scheme is used. These results indicate that the quality of the estimated error covariance is improved by instituting assimilation cycles in which observed information is absorbed continuously. The assimilation is in turn improved by this improvement in the estimated error covariance. The improvement of the 6-h precipitation forecast associated with data assimilation is more pronounced when assimilation cycles are used (Fig. 11) than when they are not used (Fig. 8). Relative to forecasts without data assimilation (Fig. 11b), the area of precipitation is reduced by including data assimilation. The area and pattern of rainfall are consistent with the observations.

Table 3. Root-mean-square error (RMSE) of background or analysis fields after 5 assimilation cycles

	u (m s^{-1})	v (m s^{-1})	$\theta_{\rm p}$ (K)	$q_{\rm v}$ (kg kg^{-1})
Background	3.00	2.93	0.808	7.72E-04
LP	2.35	2.46	0.635	6.34E-04
OL	2.08	2.12	0.578	5.71E-04

6. Experiments with real radar data

In this section, a pair of assimilation and forecasting experiments is performed using real observations to further test the performance of the OL scheme. The observations come from the same radar network as in the OSSE, but are actual radar observations (rather than simulation). The forecast length is 12 h. The observed precipitation comes from MICAPS data. Based on the OSSE results, the observation localization scale parameters are fixed to 5 grid spaces in both the horizontal and vertical directions.

The observations (Fig. 12a) indicate that precipitation occurred mainly in southern Guangdong Province. The forecasts without assimilation show a broader distribution of rainfall, with the center of maximum precipitation situated to the east of the observed precipitation in Guangdong. Moreover, these forecasts predict spurious rainfall in southern Jiangxi and southern Fujian. The precipitation forecasts using the two data assimilation schemes are more consistent with the observations. The horizontal area and intensity of the strongest rainfall are reduced relative to the forecast without data assimilation, and the center of

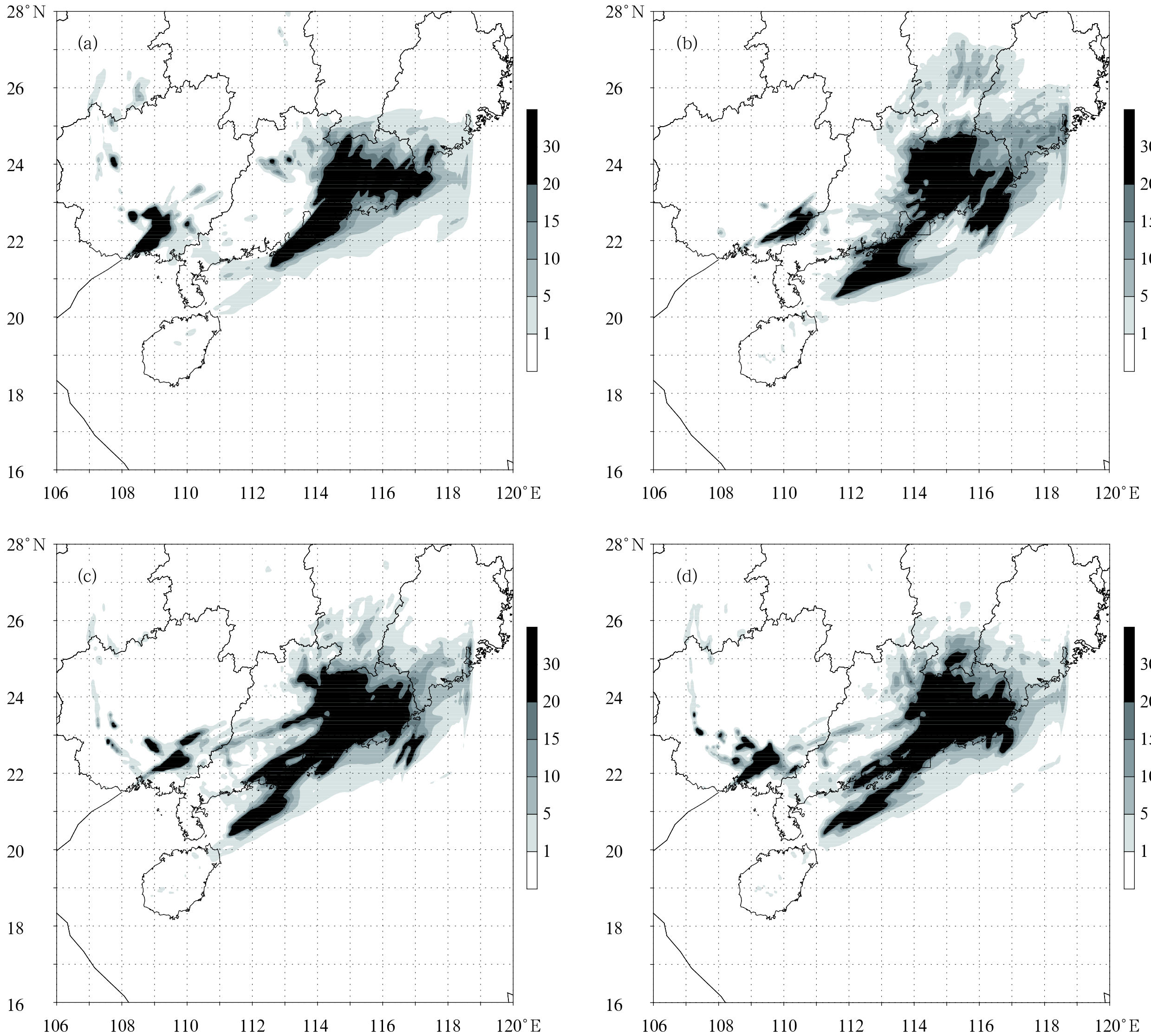

Fig. 11. As in Fig. 8, but for the 6-h period from 1400 to 2000 UTC 6 June 2008.

rainfall is shifted toward the west. The predicted horizontal extent and intensity of spurious precipitation in southern Jiangxi and southern Fujian are smaller with the OL localization scheme relative to the LP scheme. The area of spurious rainfall in southeastern Guangxi is also alleviated (see Fig. 12d). The observed rainfall over the following six hours was located mainly in southeastern Guangdong (Fig. 13a). The forecast without data assimilation indicates strong rainfall far to the east of the observed rainfall, and the intensity of the precipitation center is weaker than observed. The forecast predicts substantial amounts of precipitation in southern Jiangxi and Fujian. After data assimilation, the intensity of the forecasted rainfall in Guangdong is stronger and closer to that observed. A small area of spurious rainfall in southwestern Guangdong appears in the forecast using the LP scheme, but not in the forecast using the OL scheme.

The TSs of the rainfall forecasts before and after assimilation are shown in Fig. 14. These scores are substantially improved by data assimilation, with the best scores resulting from the use of the OL localization scheme.

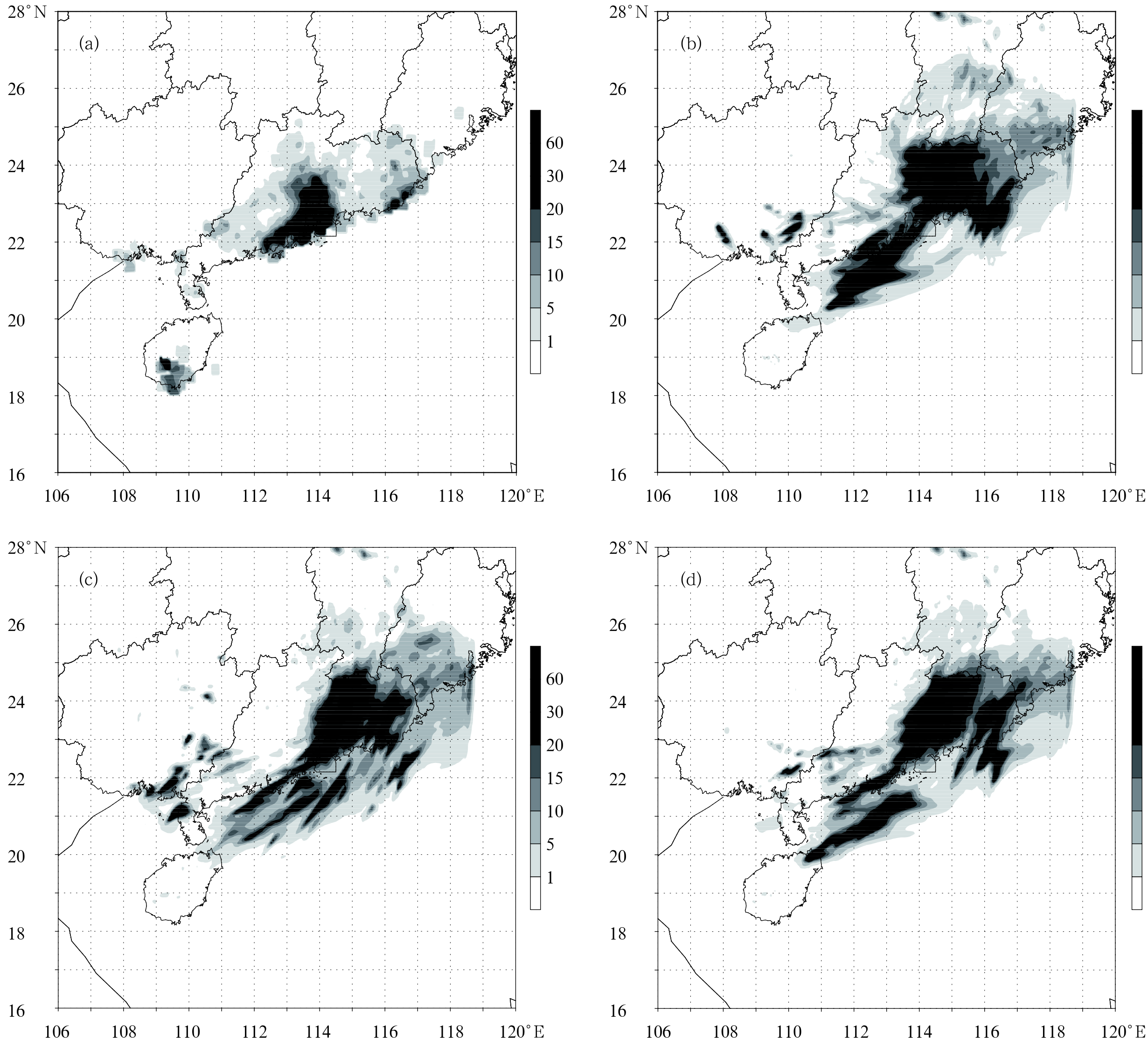

Fig. 12. As in Fig. 8, but for the 6-h period from 1200 to 1800 UTC 06 June 2008 in experiments with real radar data.

7. Conclusions

An observation localization scheme (OL) is introduced into the SVD-En3DVar system for radar data assimilation. This scheme is based on the local patch (LP) localization scheme. Both the global (GB) and LP schemes have been used as comparisons to evaluate the OL scheme. The primary conclusions are drawn as follows.

(1) In areas where observational data are dense, the analysis fields using the three schemes are consistent; however, the differences between the three schemes are often substantial in regions where observational data are sparse. Considerable analysis increments occur far from the observational position when the GB scheme is used. These analysis increments lack credibility particularly when the ensemble size is small. The analysis increments are confined to limited regions when the LP method is used; however, this method yields discontinuities in analysis increments in the transitional zones between areas with dense observations and areas with sparse observations. Both

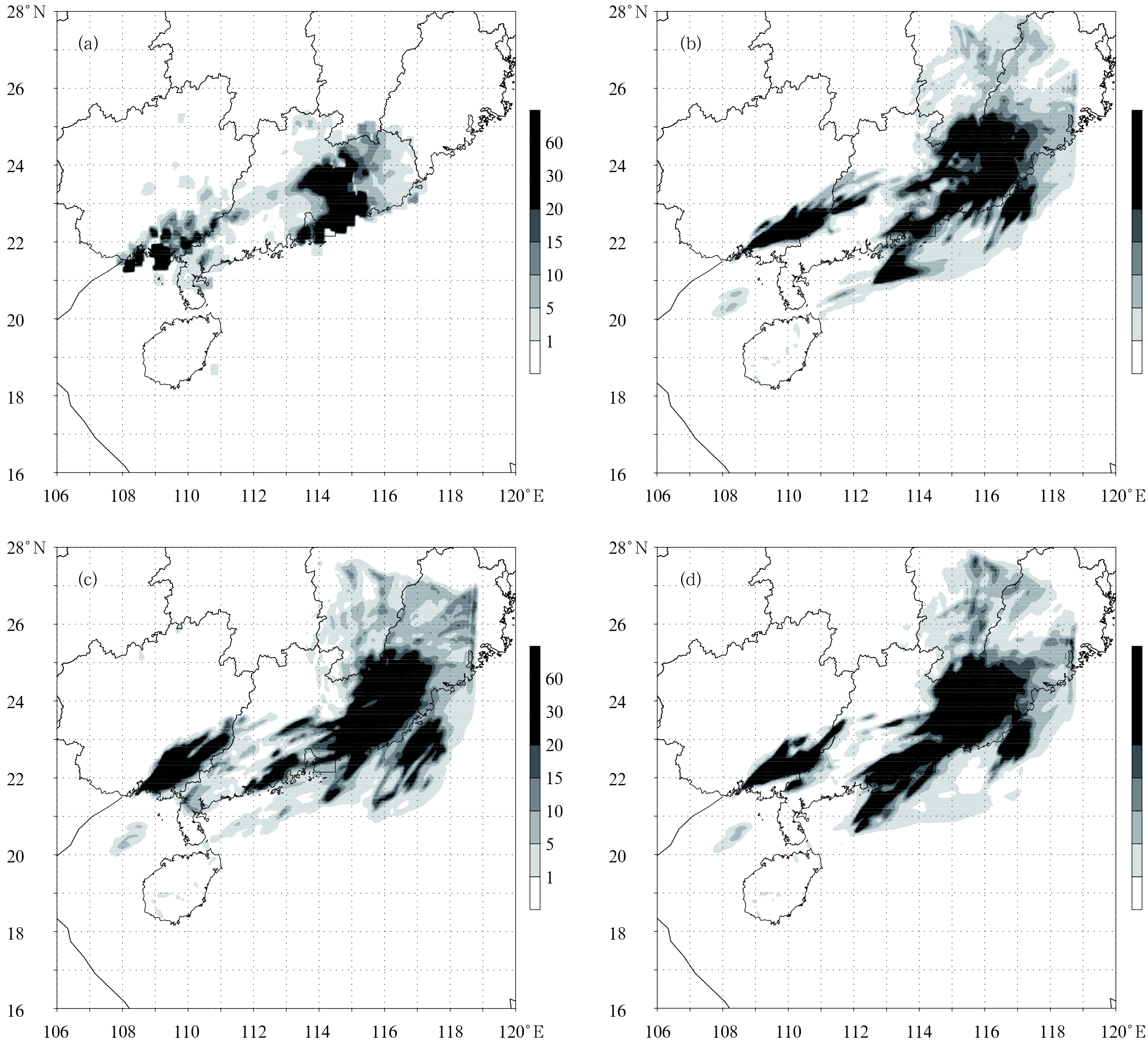

Fig. 13. As in Fig. 8, but for the 6-h period from 1800 UTC 6 June to 0000 UTC 7 June 2008 in experiments with real radar data.

of these problems are alleviated when the OL scheme is used. Furthermore, statistical analysis indicates that the RMSEs in analysis fields are less when the OL scheme is used than when either of the other two schemes is used. Radar data assimilation can improve the skill of precipitation forecasts. Of these three schemes, the OL scheme provides the greatest improvement in precipitation forecast skill.

(2) The efficiency of the OL scheme is related to the localization scales; however, sensitivity experiments show that the assimilation skill is not sensitive to localization scales within a certain range. Assimilation and forecast experiments with real radar data from a heavy rainfall event also indicate that radar data assimilation can improve the skill of precipitation forecasts. The OL scheme offers more improvement than the LP scheme.

(3) Observational information can be continuously absorbed using assimilation cycles. This procedure can substantially improve the quality of analysis fields, resulting in additional improvements in the skill of precipitation forecasts. The full potential of assimilation cycles should be further explored in future work.

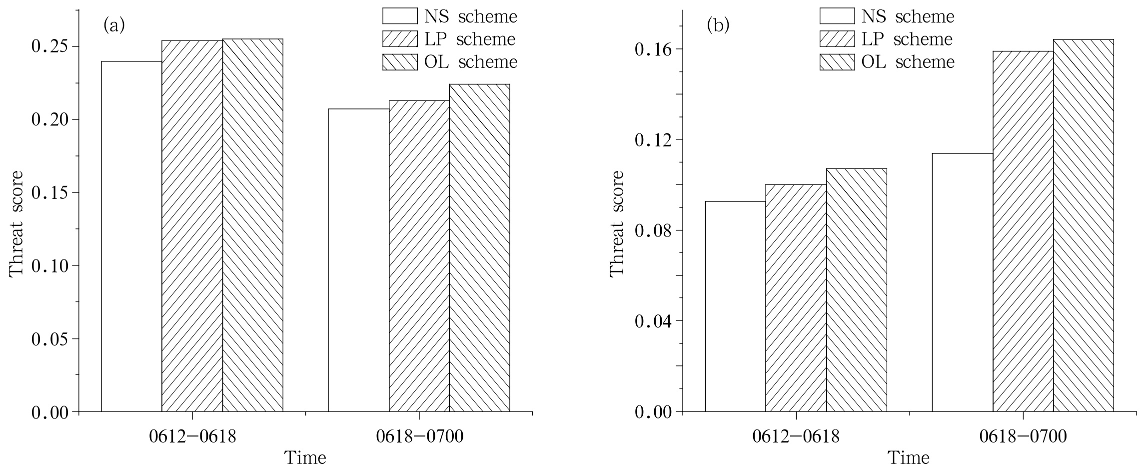

Fig. 14. As in Fig. 10, but for experiments with or without real radar data assimilation.

REFERENCES

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.

Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background error covariances: Evaluation in a quasi-operation NWP setting. *Quart. J. Roy. Meteor. Soc.*, **131**, 1013–1043.

Cohn, S. E., A. da Silva, J. Guo, et al., 1998: Assessing the effects of data selection with the DAO physical-space statistical analysis system. *Mon. Wea. Rev.*, **126**, 2913–2926.

Courtier, P., 1997: Dual formulation of four dimensional variational assimilation. *Quart. J. Roy. Meteor. Soc.*, **123**, 2449–2461.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143–10162.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.

Hamill, T. M., and C. F. Snyder, 2000: A hybrid ensemble Kalman filter–3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919.

Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter. *Physica D*, **230**, 112–126.

Liu, C., Q. Xiao, and B. Wang, 2008: An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Mon. Wea. Rev.*, **136**, 0063–3373.

Lorenc, A., 2003: The potential of the ensemble Kalman filter for NWP: A comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183–3203.

Parrish, D., and J. Derber, 1992: The National Meterological Center's spectral statistical analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.

Qiu, C., A. Shao, Q. Xu, et al., 2007: Fitting model fields to observations by using singular value decomposition: An ensemble-based 4DVar approach. *J. Geophys. Res.*, **112**, nsemble-based 4-D variational data assimilation. *J. Geophys. Res.*, **112**, D11105.

Shao, A., S. Xi, C. Qiu, et al., 2009: A hybrid space approach for ensemble-based 4D variational data assimilation. *J. Geophys. Res.*, **114**, D17114

Tian, X., Z. H. Xie, and A. G. Dai, 2008: An ensemble-based explicit four-dimensional variational assimilation method. *J. Geophys. Res.*, **113**, D21124.

Miyoshi, T., and S. Yamane, 2007: Local ensemble transform kalman filtering with an AGCM at T159/L48 resolution. *Mon. Wea. Rev.*, **135**, 3841–3861.

Wang Bin, Liu Juanjuan, Wang Shudong, et al., 2009: An economical approach to four-dimensional variational data assimilation. *Adv. Atmos. Sci.*, **27**, 715–727.

Xu Daosheng, Shao Aimei, and Qiu Chongjian, 2010: Using SVD-En3DVar for radar data assimilation: comparing local and global scheme. The International Conference on Multimedia Technology 2010.

Ningbo, 29–31 October, China, 3, 1504–1507.

—–, —–, and —–, 2011: Assimilation of Doppler radar velocity observations with SVD-En3DVar method. Part I: Simulated data experimrnts. *Chinese J. Atmos. Sci.*, **35**(4), 753–766. (in Chinese)

Zhang, F. Q., M. Zhang, and J. A. Hansen, et al., 2009: Coupling ensemble kalman filter with four-dimensional variational data assimilation. *Adv. Atmos. Sci.*, **26**, 1–8.

Zupanski, M., 2005: Maximum likelihood ensemble filter: Theoretical aspects. *Mon. Wea. Rev.*, **133**, 17010–1726.