# A Local Data Assimilation Method (Local DA v1.0) and its Application in a Simulated Typhoon Case

Shizhang Wang [1], Xiaoshi Qiao [1]

[1] Nanjing Joint Institute for Atmospheric Sciences, Nanjing, 210000, China

5   *Correspondence to*: Shizhang Wang (164293231@qq.com)

**Abstract.** A local data assimilation method, Local DA, is introduced. The proposed algorithm aims to perform hybrid and multiscale analyses simultaneously yet independently for each grid, vertical column or column group and aims to flexibly perform analyses with or without ensemble perturbations. To achieve these goals, an error sample matrix is constructed by explicitly computing the localized background error correlation matrix of model variables that are projected onto

10   observation-associated grids (e.g., radar velocity) or columns (e.g., precipitable water vapor). This error sample matrix allows Local DA to apply the conjugate gradient (CG) method to solve the cost function and to perform localization in the model-variable space, the observation-variable space, or both spaces (double-space localization). To assess the Local DA performance, a typhoon case is simulated, and a multiscale observation network comprising sounding, wind profiler, precipitable water vapor, and radar data is built; additionally, a time-lagged ensemble is employed. The results show that

15   experiments using the hybrid covariance and double-space localization yield smaller analysis errors than experiments without the static covariance and experiments without double-space localization. Moreover, the hybrid covariance plays a more important role than does localization when a poor time-lagged ensemble is used. The results further indicate that applying the CG method for each local analysis does not result in a discontinuity issue, and the wall clock time of Local DA implemented in parallel is halved as the number of cores doubles.

20   **1 Introduction**

Data assimilation (DA), which estimates the atmospheric state by ingesting information from model predictions, observations, and background error covariances, has been shown to be crucial for the success of numerical weather prediction (Bonavita et al., 2017). Therefore, many previous studies on DA have focused primarily on how to utilize observations and how to estimate background error covariances (e.g., Huang et al., 2021; Wang et al., 2021; Lei et al., 2021;

25   Zhang et al., 2009; Brousseau et al., 2011, 2012; Wang et al., 2013; Wang et al., 2012; Kalnay and Yang, 2008; Buehner and Shlyaeva, 2015). At present, there are two prevailing approaches of DA: hybrid analysis, which concerns the background error covariance, and multiscale analysis, which often addresses the difference in observation scales.

Hybrid analysis was originally proposed to combine the advantages of two kinds of DA methods, namely, stand-alone ensemble-based DA and stand-alone variational DA. The former uses short-term ensemble forecasts as error samples

30  (dynamic samples) to estimate the error covariance (Evensen, 1994), while the latter uses the statistical information extracted from historical forecasts, such as the scale of correlation  extracted from the differences between monthly 12- and 24-h forecasts (e.g., the NMC method, Parrish and Derber, 1992). Using both dynamic and static error covariances leverages the advantages of flow-dependent error information and prevents the analysis from degenerating as a result of large sampling errors arising from a limited ensemble size (Wang et al., 2009; Etherton and Bishop, 2004).

35  Many approaches are applied to conduct hybrid analyses. A widely used approach is to add an ensemble-associated control variable to a variational DA framework (Lorenc, 2003; Wang et al., 2008). An alternative is to explicitly combine the dynamic and static covariances (Hamill and Snyder, 2000). However, these two approaches are equivalent (Wang et al., 2007). Another method is to average the analyses yielded by the ensemble Kalman filter (EnKF) and the variational method (Bonavita et al., 2017; Penny, 2014). Recently, a hybrid scheme based on the EnKF framework was developed (Lei et al.,

40  2021) that uses a large ensemble size (=800) to simulate the static error covariance. Nevertheless, given the variety of hybrid approaches available, how to conduct hybrid DA is still a matter of debate. In this study, a hybrid scheme is implemented following Hamill and Snyder (2000), though the proposed scheme differs in regard to the details.

The multiscale DA aims to properly utilize observations on a variety of scales. A fundamental component of multiscale DA is performing scale-dependent localization. Localization is inevitable due to sampling errors (e.g., distant spurious

45  correlations) in ensemble-based DA, including in hybrid DA (e.g., Huang et al., 2021; Wang et al., 2021). There are two ways to localize the covariance, either explicitly or implicitly. The explicit multiscale DA often performs localization in observation space or model-observation space. The local ensemble transform Kalman filter (LETKF, Hunt et al., 2007) is the algorithm that performs localization in observation space. In model-observation space, the multiscale localization is often done sequentially; examples include the assimilation of synoptic-scale observations with a large localization radius and then

50  performing radar DA with a small radius of influence (e.g., Zhang et al., 2009; Johnson et al., 2015). In comparison, the implicit method is done in model-variable space, requiring error samples that apply several scales of localization to the ensemble members (Buehner and Shlyaeva, 2015). With the implicit method, all observations on different scales can be ingested simultaneously. Recent studies have shown that multiscale DA outperforms DA with fixed localization (Caron and Buehner, 2018; Caron et al., 2019; Huang et al., 2021).

55  In addition to the quality of the hybrid or multiscale DA analysis, the computational efficiency should also be considered (Bonavita et al., 2017). A highly parallelized DA scheme is preferable due to the continuously increasing model resolution and number of available observations. One DA scheme that can be highly parallelized is the LETKF algorithm, whose analysis is grid-independent.

In brief, both hybrid DA and multiscale DA are necessary, and the parallel computation efficiency of the LETKF is

60  attractive. Thus, it is desirable to utilize all their advantages. A straightforward idea for achieving the hybrid DA with the LETKF is to use a large-size static ensemble, similar to the EnKF-based hybrid scheme proposed by Lei et al. (2021). However, due to the limited availability of computational and storage resources, a large static ensemble (>=800) is not always feasible. The LETKF, however, performs the analysis in the ensemble space, which implies that a static ensemble is

necessary. In this situation, a grid-independent analysis scheme is needed that can be flexibly run with or without a large

65  static ensemble. Motivated by this necessity, we propose a DA scheme that follows the grid-independent analysis of the LETKF and can perform both hybrid and multiscale analysis with or without static ensemble members, similar to other variational-based hybrid schemes. The scheme is named Local DA hereinafter.

Compared with other hybrid schemes, Local DA has two features. First, the proposed scheme constructs an error sample matrix that replaces the ensemble perturbation in the LETKF. To construct this matrix, Local DA explicitly computes the

70  local background error correlation matrix of model variables that are projected onto the observation-associated grids (for scalar observations) and/or columns (for observations that measure an integrated quantity of the atmosphere, such as precipitable water vapor (PWV)). The explicit approach requires much more memory than the LETKF, but we will show that the cost is acceptable. Since the correlation matrix is explicitly constructed, we can determine how to compute the matrix components (in terms of using an ensemble or not) according to the need. Moreover, we can straightforwardly realize the

75  hybrid idea of Hamill and Snyder (2000) that is often utilized with a simple model (Kleist and Ide, 2015; Penny, 2014; Etherton and Bishop, 2004; Lei et al., 2021) because it explicitly computes and directly combines the background error covariance matrices. In this study, we attempt to evaluate the hybrid idea of Hamill and Snyder (2000) in a realistic complicated scenario; this is a secondary goal of this study..

Second, Local DA can perform the multiscale analysis in the model-variable space, the observation-variable space, or in both

80  spaces (double-space localization). In the model-variable space, Local DA adopts a scale-aware localization approach for the multiscale analysis that applies a bandpass filter to create filtered samples and individually performs localization for each waveband. In this approach, the localization radius increases as the wavelength increases, and we attempt to retain the long-distance covariances of large-scale errors while constraining the small-scale covariances within a smaller radius. A similar idea (i.e., lacking cross-waveband covariance) is the scale-dependent localization technique proposed by Buehner (2012).

85  Although the use of cross-waveband covariance is likely to further improve the multiscale analysis, but the relative performance depends on ensemble size (Caron et al., 2019). To simplify this study, we leave this issue to be addressed in future work. In contrast, in the observation-variable space, the localization method is similar to the LETKF. The localization radius varies according to the scale of an observation: large-scale observations are assigned a large radius, while small-scale observations are assigned a small radius. Because model-variable space localization and observation-variable space

90  localization are employed for the background error covariance and observation error covariance, respectively, it is possible to perform localization in both spaces. Although double-space localization may result in a double penalty, it would be interesting to note the localization performances of both hybrid and multiscale DA. Note that the LETKF of Wang et al. (2021) can also realize double-space localization, but this application has not yet been investigated.

As the first paper to report on Local DA, this study focuses on the following main issues: i) how to realize both hybrid and

95  multiscale local analysis, ii) the spatial continuity of local analysis, iii) the impact of Local DA on the subsequent forecast, iv) the sensitivity of Local DA to localization (either in the model space or in the observation space) and to the ensemble size, and v) comparing Local DA with the LETKF in terms of single deterministic analysis and forecasting. Updating the

ensemble perturbation is important for the ensemble forecast and cycling DA, but we plan to investigate this issue in future work. To avoid issues associated with the quality control of observations when evaluating the performance of Local DA, we

100 adopt observing system simulation experiments (OSSEs). The simulated multiscale observing system consists of sounding, wind profiler, PWV, and radar (radial velocity and reflectivity) observations on scales varying from the synoptic scale to the convective scale. A simulated typhoon case is selected for the evaluation.

The remainder of this paper is organized as follows. In Sect. 2, Local DA and its associated multiscale localization technique are described, including the formula, workflow, and other details. Sect. 3 describes the numerical experiments, and Sect. 4

105 discusses the results. A summary and conclusions are given in Sect. 5.

## 2 Method

### 2.1 The Local DA scheme

Local DA is implemented in the LETKF framework and comprises three steps. In the first step, Local DA maps the model state variables to the observation-associated grids or columns; these variables include the background model state ($\mathbf{x}^f$) and

110 the ensemble perturbations ($\mathbf{X}$). The observation prior can be written as $h(\mathbf{H}_i\mathbf{x}^f)$, where $\mathbf{H}_i$ is an operator that interpolates model variables to the observation-associated grids or columns and $h$ is the observation operator that converts model variables into observation variables. Similarly, the tangent linear of $h(\mathbf{H}_i\mathbf{x}^f)$ can be written as $\mathbf{H}_o\mathbf{H}_i\mathbf{X}$, where $\mathbf{H}_o$ is the linear operator of $h$. The cost function of Local DA is written in the space of $\mathbf{H}_i\mathbf{x}^f$ and $\mathbf{H}_i\mathbf{X}$. In this space, the size of the local background error correlation matrix is much smaller than that in the model-grid space, which allows us to explicitly compute

115 the correlation matrix.

In the second step, Local DA computes the minimization of the cost function

$$J = \frac{1}{2}\mathbf{v}_o^T\mathbf{v}_o + \frac{1}{2}(\mathbf{H}_o\widehat{\mathbf{X}}_o\mathbf{v}_o - \mathbf{d})^T\mathbf{R}^{-1}(\mathbf{H}_o\widehat{\mathbf{X}}_o\mathbf{v}_o - \mathbf{d}),\tag{1}$$

where $\mathbf{v}_o$ is the control variable (or a combination of error samples in the LETKF), the observation error covariance is denoted by $\mathbf{R}$, which is a diagonal matrix in this study, and $\mathbf{d}$ is the observation innovation vector. $\widehat{\mathbf{X}}_o$ ($=\alpha\mathbf{S}_o\mathbf{C}_{oo}$) represents

120 a constructed error-sample matrix, where $\mathbf{C}_{oo}$ is the local background error correlation matrix (its dynamic part is computed using $\mathbf{H}_i\mathbf{X}$, while its static part is computed with a distance-correlation function in the current version of Local DA), $\mathbf{S}_o$ stores the standard deviations (STDs) of the model variables, and $\alpha$ is a parameter that adjusts the trace of $\mathbf{C}_{oo}$. We will discuss $\alpha$ later.

In the third step, Local DA computes the model state increment $\mathbf{x}_m$ on the model grids according to the linear combination of

125 $\widehat{\mathbf{X}}_{mo}$ such that

$$\mathbf{x}_m = \widehat{\mathbf{X}}_{mo}\mathbf{v}_o,\tag{2}$$

where $\hat{\mathbf{X}}_{mo} = \alpha \mathbf{S}_m \mathbf{C}_{mo}$, $\mathbf{S}_m$ contains the STDs of the model variables on the model grids, and $\mathbf{C}_{mo}$ is a correlation matrix that contains the correlation coefficients between the model variables in the model-grid space and those in the $\mathbf{H}_i\mathbf{x}$ space. Details regarding $\mathbf{C}_{oo}$ and $\mathbf{C}_{mo}$ will be given later. As long as $\mathbf{x}_m$ is obtained, we can perform the deterministic analysis in accordance

130 with $\mathbf{x}^a = \mathbf{x}^f + \mathbf{x}_m$.

By using the hybrid and localization approaches, the rank of $\hat{\mathbf{X}}_o$ is much higher than that of $\mathbf{X}$ of which the rank is not

higher than the ensemble size. Our early test (not shown) indicates that the rank of $\hat{\mathbf{X}}_o$ is a full rank matrix in most cases.

For those rank-deficient cases, the rank of $\hat{\mathbf{X}}_o$ is often greater than 97% of the full rank value. The high-rank correlation

matrix is expected to result in a small analysis error (Huang et al., 2019).

135 To use Eqs. (1) and (2), several questions need to be answered: i) How do we construct $\mathbf{C}_{oo}$ and $\mathbf{C}_{mo}$? ii) How do we calculate $\alpha$? iii) How do we solve Eq. (1) to obtain $\mathbf{v}_o$? iv) How do we deal with nonlinear observation operators? The subsequent subsections present the answers to these questions.

### 2.1.1 The local background error correlation matrix

In Local DA, the actual correlation matrix $\tilde{\mathbf{C}}$ is the square of $\mathbf{C}_{oo}$ multiplied by a rescaling parameter $\alpha^2$:

140 $$\tilde{\mathbf{C}} = \alpha^2 \mathbf{C}_{oo} \mathbf{C}_{oo}^T. \tag{3}$$

By using the rescaling parameter, the trace of $\tilde{\mathbf{C}}$ is equivalent to that of $\mathbf{C}_{oo}$. $\alpha$ is computed according to

$$\alpha = \sqrt{tr(\mathbf{C}_{oo}) \Big/ tr(\mathbf{C}_{oo}\mathbf{C}_{oo}^T)}, \tag{4}$$

where $tr(\ )$ denotes the calculation of the trace of a matrix. Notably, $tr(\mathbf{C}_{oo}\mathbf{C}_{oo}^T)$ is equal to the sum of squares of all elements in $\mathbf{C}_{oo}$. There is no need to compute $\mathbf{C}_{oo}\mathbf{C}_{oo}^T$. $\tilde{\mathbf{C}}$ and $\mathbf{C}_{oo}$ are identical in terms of Eigenvectors and the trace of the

145 matrix (total variance). The eigenvalues of $\tilde{\mathbf{C}}$ are the squares of the corresponding Eigenvalues of $\mathbf{C}_{oo}$ multiplied by $\alpha^2$.

Therefore, $\tilde{\mathbf{C}}$ is an approximation of $\mathbf{C}_{oo}$. Storto and Andriopoulos (2021) proposed a hybrid DA scheme that also used the rescaling parameter to tune the trace of a matrix (see their Eq. (15)), but they constructed the background error covariance in a way differing from ours.

$\mathbf{C}_{oo}$ is a $K \times K$ matrix, where $K$ is the number of model variables associated with the operators of all ambient observations. $K$

150 is computed according to

$$K = \sum_{i=1}^{N_t} (N_o N_{op})_i, \tag{5}$$

where $N_\mathrm{t}$ is the number of observation types, such as the zonal wind from soundings and/or the radial velocity from radars, $N_\mathrm{o}$ is the number of observations of the $i$th type, and $N_\mathrm{op}$ is the number of analysis variables used in the observation operator of the $i$th type. Assuming there are three available observations (two zonal wind observations and a surface pressure observation), the background error correlation matrix is

$$\mathbf{C}_{oo} = \begin{pmatrix} c_{u1u1} & c_{u1u2} & c_{u1ps1} \\ c_{u2u1} & c_{u2u2} & c_{u2ps1} \\ c_{ps1u1} & c_{ps1u2} & c_{ps1ps1} \end{pmatrix}, \qquad (6)$$

where $c$ is the correlation coefficient between the model variables projected onto observation-associated grids and the subscripts "$u1$", "$u2$", and "$ps$" represent the two zonal wind observations and the surface pressure observation, respectively. If the model variables to be updated are the zonal wind ($u$), potential temperature ($\theta$), and water vapor mixing ratio ($q$) located where observation "$u1$" is, the corresponding $\mathbf{C}_\mathrm{mo}$ is

$$\mathbf{C}_{mo} = \begin{pmatrix} c_{\theta1u1} & c_{\theta1u2} & c_{\theta1ps1} \\ c_{q1u1} & c_{q1u2} & c_{q1ps1} \\ c_{u1u1} & c_{u1u2} & c_{u1ps1} \end{pmatrix}. \qquad (7)$$

where the subscripts "$\theta1$" and "$q1$" indicate that variables are on the same grid as observation "$u1$". For $\mathbf{C}_{oo}$, the corresponding $\mathbf{S}_\mathrm{o}$ is

$$\mathbf{S}_o = \begin{pmatrix} s_{u1} & & \\ & s_{u2} & \\ & & s_{ps1} \end{pmatrix}, \qquad (8)$$

where $s$ denotes the STDs of the model variables projected onto observation-associated grids. $\mathbf{S}_\mathrm{m}$ contains the STDs on the model grids, which are $s_{u1}$, $s_{\theta1}$, and $s_{q1}$ for the aforementioned example.

Note that the size of $\mathbf{C}_{oo}$ grows rapidly as $K$ increases. However, given that $\mathbf{C}_{oo}$ is used for the local analysis, the memory requirement is still affordable. For high-resolution observations, thinning can help reduce the cost, which is also necessary to ensure that the observation errors are uncorrelated (e.g., Hoeflinger et al., 2001). For observations at the same place that are associated with the same model variables, such as the radar reflectivity and differential reflectivity, the same $\mathbf{C}_{oo}$ components

are used. This strategy is also valid for passive microwave observations at different frequencies obtained by one satellite. Therefore, the size of $\mathbf{C}_{oo}$ is controllable. In this study, we use a simple thinning approach to control the matrix size, as described in the Appendix.

### 2.1.2 The solution of Local DA

175   There are two approaches to solve the gradient of Eq. (1): i) matrix decomposition and ii) an iterative algorithm. The first approach is straightforward, but it is time consuming if $\mathbf{C}_{oo}$ is large. Therefore, Local DA adopts an iterative algorithm, namely, the conjugate gradient (CG) method (Shewchuk, 1994). Theoretically, using the CG method requires the background error covariance matrix to be positive definite. However, with the control variable transform, a positive semidefinite covariance matrix is sufficient to obtain the best linear unbiased estimate (Ménétrier and Auligné, 2015). A

180   strictly diagonally dominant matrix with nonnegative diagonal elements s is positive semidefinite. Performing covariance localization is a way to generate the matrix; the details of this localization will be given later.

Note that Local DA performs the CG step locally, unlike other variational-based DA methods that apply the CG method globally. Therefore, it is necessary to investigate whether the local application of the CG method causes a nonnegligible spatial discontinuity, which will be discussed in Sect. 4. For computational efficiency, the maximum number of iterations is

185   100. If the error tolerance $\varepsilon^2$ defined in (Shewchuk, 1994) cannot reach $1\times10^{-6}$ by the $100^{th}$ step, the CG method is stopped.

### 2.1.3 The observation operator

The EnKF algorithm often approximates the linear projection ($\mathbf{H}$ in Eq. (1)) of the model variables onto the observation variables according to the departure of the observation priors from their ensemble mean. It is straightforward for Local DA to use the ensemble approximation approach. However, for nonlinear observation operators, there is an alternative to the

190   ensemble mean, namely, the observation prior calculated by using the ensemble mean of the model variables. Tang et al. (2014) demonstrated that this alternative could lead to better results. Furthermore, Yang et al. (2015) examined the application of this alternative in radar DA and showed that the alternative approach produced smaller analysis errors for the model variables associated with radial velocity (three wind components) and reflectivity (mixing ratios of rain, snow, and graupel). Given that remote sensing observations such as those obtained by radars and satellites are important parts of any

195   multiscale observation network, Local DA adopts the alternative approach proposed by Tang et al. (2014).

Local DA approximates the linear projection $\tilde{\mathbf{Y}} = \mathbf{H}_o\widehat{\mathbf{X}}_o$ according to

$$\tilde{\mathbf{Y}} \approx h(\mathbf{x}^{f} + \widehat{\mathbf{X}}_o) - h(\mathbf{x}^{f} + \overline{\widehat{\mathbf{X}}_o}) , \tag{9}$$

where $h$ is the nonlinear observation operator and $\mathbf{x}^{f}$ is the background model state vector. Note that Eq. (9) is written for a deterministic forecast in this study. Compared with the results using the ensemble mean of observation priors, Eq. (9)

200   reduces the analysis error of reflectivity by approximately 2 dBZ in our early test (not shown). This result is consistent with that of Yang et al. (2015).

## 2.2 The multiscale localization analysis

To realize multiscale localization, Local DA first performs scale decomposition with a bandpass filter. The decomposed perturbation, $\mathbf{X}'_b$, is

205  $$\mathbf{X}'_b = \sum_{l=1}^{N_b} \mathbf{X}^l_b, \tag{10}$$

where the superscript "$l$" represents the $l$th waveband and $N_b$ is the number of bands. As a localization approach lacking cross-waveband covariance, Local DA computes the STD of the perturbation, $s$, according to

$$s(i) = \sqrt{\sum_{l=1}^{N_b} \frac{1}{N} \sum_{m=1}^{N} \left[ \mathbf{X}^l_b(i,m) \right]^2}, \tag{11}$$

where $i$ and $m$ denote the $i$th model variable and the $m$th sample, respectively, and $N$ is the sample size. Compared with the

210  raw STD, $\sqrt{\dfrac{1}{N} \sum\limits_{m=1}^{N} \left[ \sum\limits_{l=1}^{N_b} \mathbf{X}^l_b(i,m) \right]^2}$, the cross influence among different scales of $\mathbf{X}'_b$ is ignored in Eq. (11). Nevertheless,

we acknowledge the importance of the cross influence of these perturbations and plan to investigate this issue with regard to Local DA in our future work.

The multiscale correlation coefficient $c(i,j)$ is calculated according to

$$c(i,j) = \sum_{l=1}^{N_b} \frac{\mathrm{cov}\left[ \mathbf{X}^l_b(i), \mathbf{X}^l_b(j) \right]}{s(i)s(j)}, \tag{12}$$

215  where $i$ and $j$ denote the $i$th and $j$th variables, respectively. For the case of $i=j$, Eq. (12) ensures $c(i,j)=1.0$.

We perform localization for each waveband independently to construct the multiscale correlation matrix In principle, our multiscale localization method trusts the correlation coefficient of each waveband when the distance between two variables is smaller than the lower bound of the waveband. For instance, for the waveband of 50 km – 100 km, Local DA starts the localization when the distance $d$ is greater than 50 km. The decorrelation coefficient $r(l,i,j)$ for the $l$th waveband and $c(i,j)$ is

220  calculated according to

$$\begin{cases} r(l,i,j) = 1.0, \ d <= d_{\min}(l) \\ \\ r(l,i,j) = e^{-8\left[\frac{d - d_{\min}(l)}{d_r(l)}\right]^2}, \ d > d_{\min}(l) \ , \\ \\ r(l,i,j) = 0.0, \ d > d_{\max}(l) \end{cases} \tag{13}$$

where $d_{\min}(l)$ and $d_{\max}(l)$ are the lower and upper bounds of the $l$th waveband, respectively, and $d_r(l)$ is the localization radius for the $l$th waveband. Note that how to optimally localize the background error covariance is still an open question; rather, Eq. (13) is simply a preliminary implementation of multiscale localization for Local DA.

225   In addition to localization in the model-variable space, Local DA can perform localization in the observation-variable space, similar to the LETKF. Observation-space localization is conducted by enlarging the observation error as the distance between variables increases. The localization coefficient in the observation space is calculated according to the second formula of Eq. (13), but $d$-$d_{\min}(l)$ and $d_{\min}(l)$ are replaced by $d$ and $d_o$, respectively, where $d_o$ is the localization radius that varies among different observation types. In practice, it is possible for Local DA to perform localization in both the model

230   space and the observation space. Although performing localization in both spaces may result in a double penalty, it would be interesting to note the performance of the double-space localization approach, which has not yet been investigated. The related experiments and results are given in the following sections.

### 2.3 The hybrid analysis

The current version of Local DA calculates a simple "static" correlation by using the second formula of Eq. (13), except that

235   $d$-$d_{\min}(l)$ and $d_{\min}(l)$ are replaced by $d$ and $d_s$, respectively, where $d_s$ is a fixed localization radius. For the $i$th and $j$th variables, the hybrid $c(i,j)$ in $\mathbf{C}_{oo}$ is computed according to

$$c(i, j) = \gamma \sum_{l=1}^{N_b} r(l,i,j) \frac{\mathrm{cov}\left[\mathbf{X}_b^l(i), \mathbf{X}_b^l(j)\right]}{s(i)s(j)} + (1-\gamma)e^{-8\left[\frac{d}{d_s}\right]^2} , \qquad (14)$$

where $\gamma$ is the weight of the dynamic correlation. The hybrid $c(i,j)$ in $\mathbf{C}_{mo}$ is also computed according to Eq. (14), but $\mathbf{X}_b^l(i)$

and $s(i)$ store the variables at a model grid point. In the case where the $i$th and $j$th variables are different types of variables,

240   such as a wind component variable and a temperature variable, the second term on the right-hand side of Eq. (14) is forced to zero. In other words, the current version of Local DA suppresses the cross-variable covariance when the hybrid analysis is performed. To prevent $s(i)$ and $s(j)$ in Eq. (14) from being forced to zero (as often occurs for convective-related variables such as the mixing ratios of rainwater, snow and graupel), we add small, random perturbations with an STD of $1\times10^{-7}$ to the variables for which the STDs are smaller than $1\times10^{-7}$.

### 2.4 The workflow of Local DA

245

Here, we present a step-by-step description of how the hybrid and multiscale analyses described in the previous sections are performed for all the model variables. Note that Local DA can perform the analysis in a much faster way; we will discuss this method later.

1) Apply a bandpass filter to generate filtered samples on $N_b$ scales from $\mathbf{X}_b'$ .

250 2) Store the initial condition, filtered perturbation samples, and observations in separate arrays denoted $\mathbf{x}^f$, $\mathbf{X}'_b$, and $\mathbf{y}^o$,

respectively.

3) For each model variable to be updated, search its ambient observations according to their scales and store these

observations in array $\widehat{\mathbf{y}}^o$; for example, search for sounding data within 300 km while searching for radar data within 15 km.

In addition, according to the observation operators of $\widehat{\mathbf{y}}^o$, store the observation-associated model variables that have been

255 projected onto observation-associated grids or columns into arrays denoted $\widehat{\mathbf{x}}^f$ and $\widehat{\mathbf{X}}'_b$, respectively.

4) Calculate the vector $\mathbf{d}$ in Eq. (1) with $\widehat{\mathbf{y}}^o$ and $\widehat{\mathbf{x}}^f$ ..

5) Use $\widehat{\mathbf{X}}'_b$ to generate $\mathbf{S}_o$, $\mathbf{C}_{oo}$, $\mathbf{S}_m$, and $\mathbf{C}_{mo}$ according to Eqs. (11), (13), and (14).

6) Compute $\alpha$ for $\mathbf{C}_{oo}$ by using Eq. (4).

7) Compute $\widehat{\mathbf{X}}_o = \alpha \mathbf{S}_o \mathbf{C}_{oo}$.

260 8) Calculate $\mathbf{Y} = \mathbf{R}^{-0.5} \mathbf{H}_o \widehat{\mathbf{X}}_o$ by using Eq. (9).

9) Use the CG method to solve $(\mathbf{I} + \mathbf{Y}^T \mathbf{Y}) \mathbf{v}_o = \mathbf{Y}^T \mathbf{R}^{-0.5} \mathbf{d}$ and obtain $\mathbf{v}_o$.

10) Compute the model state increment $\mathbf{x}_m$ according to Eq. (2).

Considering that $\mathbf{S}_m$, $\mathbf{C}_{mo}$ and $\mathbf{x}_m$ can be applied to all variables influenced by $\widehat{\mathbf{y}}^o$, it is not necessary to compute $\mathbf{C}_{oo}$ for each

model variable. In the current version of Local DA, $\mathbf{S}_m$, $\mathbf{C}_{mo}$ and $\mathbf{x}_m$ may contain all the model variables at a model grid

265 point, in one column, or in 5 neighboring columns (N-column analysis). For N-column analysis, $\widehat{\mathbf{y}}^o$ contains ambient

observations at all vertical levels. Due to using the same $\mathbf{C}_{oo}$ for neighboring columns, the 5-column analysis is slightly

rasterized (not shown), leading to slightly larger errors than the 1-column analysis. However, the extent of this degeneration

is acceptable. On the other hand, the wall clock time of the 5-column analysis is close to 1/5 of that of the 1-column analysis.

In this study, all Local DA results are generated using 5-column analysis. A similar N-column analysis approach is the

270 weighted interpolation technique in the LETKF (Yang et al., 2009), which performs the analysis, for example, every 3 grid

points in both the zonal direction and the meridional direction.

## 3 Experimental design

### 3.1 The simulated typhoon

The third typhoon of the 2021 western Pacific season, In-Fa, is selected for the OSSEs performed herein. The true

275 simulation, starting at 00 UTC on 25 July 2021 and ending at 18 UTC on 26 July 2021, simulates the stage in which In-Fa

approaches China. The Weather Research and Forecast (WRF, Skamarock et al., 2018) model V3.9.1 is used for the

simulation. The central latitude and longitude of the forecast domain are 30.5 ° and 122.0 °, respectively. The size of the domain is 201 grids ×201 grids×34 levels with a horizontal resolution of 5 km and a model top pressure of 50 hPa. The physical parameterization schemes are as follows. The WRF single-moment 6-class ice scheme (Hong and Lim, 2006) is

280    adopted for microphysical processes. For longwave and shortwave radiation, the rapid radiative transfer model (RRTM) scheme (Mlawer et al., 1997) and Dudhia scheme (Dudhia, 1989), respectively, are used. The Yonsei University (YSU) scheme (Hong et al., 2006) is employed for the planetary boundary layer simulation. For the cumulus parameterization, the Kain–Fritsch (new Eta) scheme (Kain, 2004) is enabled. The unified Noah land surface model is used to simulate the land surface. We adopt the global forecast system (GFS) analysis at 00 UTC on 25 July 2021 as the initial condition of the true

285    simulation.

### 3.2 The multiscale observation network

The multiscale observation network (Figure 1) comprises simulated sounding, wind profiler, PWV, and radar data. Soundings are available at 00 UTC and 12 UTC on 26 July 2021, whereas the other types of observations are available hourly on 26 July 2021.

290    For each sounding, we simply extract the perturbed model variables, $u$, $v$, $\theta$, and $q_v$, every 2 model levels as the observations. The simulated soundings also include records of the perturbed surface pressure, $ps$. The sounding perturbations follow a Gaussian distribution with zero mean. The perturbation STDs are 0.5 m s$^{-1}$, 5 m s$^{-1}$, 0.5 K, 5×10$^{-5}$ kg kg$^{-1}$, and 10 Pa for $u$, $v$, $\theta$, $q_v$, and $ps$, respectively. To better reflect the reality, no simulated soundings are available over the ocean, and the horizontal resolution of each sounding is 100 km.

295    The simulated wind profiler provides data on horizontal wind components, $u$ and $v$, at all model levels. The perturbations added to the wind profiler data follow a Gaussian distribution with zero mean and an STD of 0.5 m s$^{-1}$. The wind profilers, the data from which have a horizontal resolution of 50 km, provide data only on land.

The PWV observations are computed according to

$$PWV = \frac{1}{g}\int_{p1}^{p2} q_v dp, \qquad\qquad (15)$$

300    where $g$ is the gravitational constant of acceleration and $p1$ and $p2$ represent the bottom and top of a model column, respectively. Perturbations with zero mean and an STD of 0.5 kg m$^{-2}$ are added to the PWV observations. Because the PWV is observed by satellites, this type of observation is available for the whole forecast domain, and the horizontal observation interval is 50 km in both the $x$ and the $y$ directions.

The radar data to be assimilated are radial velocity and reflectivity measurements. We adopt Eq. (3) of Xiao and Sun (2007)

305    to compute the radial velocity, but we ignore the terminal velocity in the OSSE. For reflectivity, the operator proposed by Gao and Stensrud (2012) is employed. Three radars located approximately at Shanghai (121.48 ° E, 31.23 °N), Hangzhou (120.16 ° E, 30.28 °N), and Ningbo (121.55 ° E, 29.88 °N) are simulated with a maximum observation range of 230 km. The

simulated radars work on the volume coverage pattern (VCP) 11 mode, which has 14 elevation levels from 0.5° to 19.5°. Radar data are created on volume-scan elevations, but they are on model grids in the horizontal direction, as shown in Xue et al. (2006). The radial velocity and reflectivity observation errors are 1.0 m s$^{-1}$ and 2.0 dBZ, respectively. The horizontal resolution of the radar data is identical to the model grid spacing.

In total, 2795 simulated soundings, 400 PWV data points, 5332 wind profiler observations, and 391618 radar observations (including radial velocity and reflectivity) are utilized in this study.

### 3.3 The DA experiments

In this study, we design several experiments for Local DA to investigate its performance in single deterministic analyses and to examine the impacts of the hybrid and multiscale analysis approaches. We also run several cycling DA experiments for Local DA, but the results of these cycling experiments are used only for reference because the current version of Local DA does not update the ensemble perturbations and does not seriously consider the cross-variable balance during/after the analysis, which are important for cycling DA experiments (e.g., Zeng et al., 2021). To perform the hybrid analysis, it is necessary to generate the ensemble first. Therefore, in this subsection, we first describe the ensemble and then introduce the experimental design.

### 3.3.1 The time-lagged ensemble

The time-lagged approach (e.g., Branković et al., 1990) is employed to generate the ensemble perturbations, which are created by using deterministic forecasts with different initial times and varying GFS data. For example, the first sample at 00 UTC on 26 July 2021 stores the difference between two deterministic forecasts initialized at 06 UTC on 25 July 2021 and 12 UTC on 25 July 2021. To distinguish these forecasts from the forecasts of the DA experiments, the forecasts used to produce ensemble members are referred to as sample forecasts. The sample forecasts used in this study are shown in Figure 2. Note that some sample forecasts are initialized by the 3-h or 6-h GFS forecast data (highlighted by the thick tick marks in Figure 2). Two ensembles differing in size are created: the large ensemble has 36 members and combines 9 sample forecasts according to $C(9,2) = \dfrac{9!}{2!7!}$, while the small ensemble excludes sample forecasts initialized at 03 UTC and 06 UTC on 25 July 2021 and thus has 15 members ($C(6,2) = \dfrac{6!}{2!4!}$).

Compared with the root mean square error (RMSE) of the initial condition, the spreads of both ensembles are deficient, particularly at lower levels (Figure 3) for $u$, $v$, $w$, $\theta$, and $q_v$. The 36-member ensemble has a larger spread than the 15-member ensemble, but the difference between the two is not distinct even though the ensemble size of the large ensemble is twice that of the small ensemble. This result implies that many members in the ensembles are similar, leading to a small ensemble spread. Therefore, we emphasize that the aforementioned time-lagged ensemble is only valid for the purpose of the

deterministic analysis in this study. For a general ensemble-based algorithm, let alone for the cycling DA experiments, the above ensemble may not be good enough.

### 3.3.2 The DA configurations

340 Two sets of experiments are designed: one for the single deterministic analysis at 00 UTC on 26 July 2021 and one for the cycling analysis from 00 UTC to 18 UTC on 26 July 2021 with a cycling interval of 6 h. For the single deterministic analysis, we first study the influences of the hybrid covariance and multiscale localization on Local DA. Experiment LDA_ctrl is conducted with only the dynamic covariance and fixed localization, whereas LDA_HBC_MSL is the experiment using the hybrid covariance (HBC) and multiscale localization (MSL) simultaneously, and LDA_HBC disables

345 multiscale localization. Second, we investigate the impacts of the localization space. Compared with LDA_HBC_MSL, experiment LDA_DS adopts double-space localization (DS). LDA_OS uses observation-space localization (OS) and disables the static covariance to exclude the impact of horizontal model-space localization. Moreover, due to the use of 5-column analysis in LDA_OS, the observation-space localization is valid only in the horizontal direction, while in the vertical direction, the localization is in the model-variable space.

350 Given that LDA_OS disables the static covariance and performs observation-space localization, we seek to evaluate whether LDA_OS produces a deterministic analysis similar to or close to that yielded by the LETKF. Therefore, a LETKF test, called LETKF_OS, is run for this purpose. Because Local DA approximately follows the LETKF procedure, it is not difficult to implement the LETKF for a deterministic analysis within the Local DA framework.

LDA_HBC_MSL_cyc, LDA_ctrl_cyc, and LDA_DS_cyc are the cycling DA experiments that use the same configurations

355 of their counterparts named without the suffix "_cyc". These experiments perform cycling DA at 00 UTC, 06 UTC, 12 UTC, and 18 UTC on 26 July 2021. A shorter cycling interval has been tested (not shown), and the resulting forecast errors of most variables during the cycling are large, as expected. The shock to the model is simply a cause, as will be shown in Sect. 4. Again, we emphasize that the cycling DA experiments of Local DA are just for reference.

The aforementioned experiments are conducted with the 15-member ensemble. Thus, to examine the sensitivity of Local DA

360 to the ensemble size, we rerun LDA_HBC_MSL, LDA_OS, and LETKF_OS with the 36-member ensemble, and the rerun experiments are referred to as LDA_HBC_MSL_36m, LDA_OS_36m, and LETKF_OS_36m, respectively. In addition, LDA_DS is rerun with only the static correlation matrix to discern the contribution of static information to Local DA; the experiment is named LDA_DS_noENS. For convenience, all experiments are listed in Table 1, where "M" and "O" denote the model-variable and observation-variable spaces, respectively.

365 The model variables to be updated are the three wind components ($u$, $v$, $w$), potential temperature ($\theta$), water vapor mixing ratio ($q_v$), dry-air mass in column ($mu$), and hydrometeor mixing ratios ($q_c$, $q_r$, $q_i$, $q_s$, and $q_g$). The DA-related parameters are listed as follows. The fixed localization radius is 200 km. Multiscale localization uses five bands of samples: <20 km, 20 km – 50 km, 50 km – 100 km, 100 km – 200 km, and >200 km; the localization of these bands is enabled as the distance reaches 0 km, 20 km, 50 km, 100 km, and 200 km, respectively, with corresponding radii of 20 km, 50 km, 100 km, 200 km, and 500

370  km. For observation-space localization, the horizontal radii are 360 km, 150 km, 120 km, and 15 km for sounding, wind profiler, PWV, and radar data, respectively. The vertical radius for all observations is 5 km. The vertical location of a PWV observation is assumed to be 4 km above sea level, which is used in LETKF_OS. For Local DA, we specifically assign fixed localization radii of 1000 km and 20 km for the model variable $ps$ and the reflectivity-related variables, respectively. These values are tuned for the case in which Typhoon In-Fa made landfall in this study. The other DA parameters are 0.5 for the

375  hybrid parameter $\gamma$ and 1.5 for covariance inflation; both are empirically assigned.

# 4 Results and discussion

## 4.1 The convergence of minimization

We examine the minimization convergence by using the data extracted from LDA_HBC_MSL. Figure 4 shows the number of iterations and the ratio of the final value of the cost function ($J_{final}$) to the initial value ($J_{initial}$). Fewer than 100 iterations

380  indicates that the tolerance $\varepsilon^2$ reaches $1 \times 10^{-6}$ within 100 steps. If the minimization does not converge within 100 steps, the CG iteration is stopped by the program. The number of iterations is large near the center of the forecast domain but decreases rapidly outward. According to the distribution of observations (Figure 1), the results (Figure 4a) indicate that the minimization converges more slowly as the number of observations to be assimilated increases.

Although the minimization fails to converge within 100 steps in the area where the observation density is high, the cost

385  function is still reduced by 70% or 80% (Figure 4b). In contrast, near the northeastern and southeastern corners of the domain, where the minimization converges within 10 steps, the final value of the cost function is larger than 70% of its initial value. However, in those areas of the domain, the initial cost function is small, implying no need for a large correction. The results also indicate that no serious discontinuity occurs in LDA_HBC_MSL, which is desired. Similar to the LETKF, the use of slightly different $\mathbf{C}_{oo}$ between neighboring columns does not yield remarkably different analyses.

390  Further investigation (for data within the yellow rectangle plotted in Figure 4a) indicates that approximately 25% minimizations fail to converge within 100 steps (Figure 5a), all associated with the application of radar data. Therefore, we rerun LDA_HBC_MSL using only radar data and observe that only 4% of all minimizations require more than 100 steps to converge. In the case of setting the maximum number of iterations to 500 for LDA_HBC_MSL, all minimizations converge within 300 iteration steps. The results also show that assimilating only radar data produces a smaller ratio of $J_{final}$ to $J_{initial}$

395  than the case using all observations (Figure 5b). Therefore, we speculate that it may be difficult for the current version of Local DA to converge within a few iteration steps when multiple types of observations are ingested. Nevertheless, despite this difficulty, Local DA reduces the cost function by more than 70% within 100 iteration steps in most cases (Figure 5b). Further suppressing the error may require a better background error covariance, which we plan to seek in future work.

## 4.2 The DA performance

400    The RMSE is used to evaluate the analysis quality. For the horizontal wind components, all experiments using the 15-
member ensemble produce an analysis error smaller than the background error from the lower level to the upper level
(Figure 6a, b). Moreover, the RMSEs of the experiments using the hybrid method (LDA_HBC_MSL, LDA_HBC, and
LDA_DS) are consistently smaller than those of the experiments without the static covariance (LDA_ctrl, LDA_OS, and
LETKF_OS). The RMSE profiles of LDA_HBC_MSL and LDA_HBC are similar, implying that the hybrid method plays a

405    more important role than multiscale localization. Huang et al. (2019) has demonstrated that the high-rank ensemble
transform Kalman filter is insensitive to the localization length scale, thus the high-rank background correlation matrix, $\mathbf{C}_{oo}$,
is likely another cause of the relatively small impact of the multiscale localization. Using double-space localization
(LDA_DS) further decreases the RMSE of LDA_HBC_MSL at approximately all levels, which is an indicator of using a
poor ensemble in the analysis because performing localization in the model-variable space alone is insufficient to suppress

410    spurious correlations. These results highlight the necessity of using the hybrid covariance when the ensemble is poor. Similar
results can be observed for hydrometeor variables (Figure 6f-i). Although multiscale localization does not produce distinctly
different wind components between LDA_HBC_MSL and LDA_HBC, it does result in a smaller RMSE for the hydrometeor
variables in LDA_HBC_MSL. Thus, multiscale localization is beneficial for the analysis performance. For other variables
($w$, $\theta$, and $q_v$), LDA_DS produces consistently smaller errors than the other experiments, while the differences between the

415    other experiments are not as distinct. In most cases, LDA_ctrl exhibits the worst performance. Using a fixed radius of 200
km and a poor ensemble are the main causes of the poor performance of LDA_ctrl.

To qualitatively assess the analysis error, we compute the difference in total energy (DTE, Meng and Zhang, 2007).
LDA_ctrl (Figure 7 d-f) decreases the large background errors (Figure 7 a-c) at 850 hPa and 500 hPa but generates many
spurious increments over the ocean, increasing the error there; this problem is more pronounced at 200 hPa. Accordingly,

420    LDA_ctrl fails to substantially reduce the error for the entire domain. Because the spurious increment is a result of using a
poor ensemble, utilizing the hybrid method that suppresses the impact of the ensemble yields smaller errors
(LDA_HBC_MSL, Figure 7 g-i) from the lower to the upper levels. The spurious increment is further reduced in LDA_DS,
especially at 850 hPa and 500 hPa, corresponding to the smallest analysis error in Figure 6. For LDA_OS and LETKF_OS,
we consider mainly whether their analyses are similar. According to the results shown in Figure 8, the two experiments

425    produce similar analyses, but the Local DA analysis is noisier, which implies that the LETKF is better than Local DA as a
stand-alone ensemble-based algorithm.

For the perturbation of the dry-air mass in column ($mu$), although LDA_ctrl corrects the overestimate of the central pressure,
it severely underestimates $mu$ in the typhoon center (Figure 9b), leading to a larger RMSE after DA. In addition, the analysis
of LDA_ctrl is noisy, which is attributable to the unreliable error correlation estimated by the ensemble because

430    LDA_HBC_MSL and LDA_DS (Figure 9c, d), which weaken the impact of the ensemble, generate smaller analysis errors
and less noise. Many spurious increments are still observed in LDA_HBC_MSL and LDA_DS, which may require a better

covariance or a static constraint to resolve. The analysis errors of *mu* in LDA_OS and LETKF_OS (Figure 9e, f) do not differ substantially, as expected. In general, Local DA with the hybrid method and multiscale localization is able to reduce the surface pressure error, but the results are not good enough in terms of the spurious increment.

435   Figure 10 shows the evolutions of the forecast error for the aforementioned experiments. All the DA experiments produce smaller forecast errors than the experiment without DA (BAK) throughout the 18-h forecast in terms of the observation variables. For nearly all investigated variables, the experiments using the hybrid covariance produce smaller errors than the experiments using only the ensemble-based covariance. This finding is consistent with the results in Figure 6. LDA_DS often yields the smallest forecast errors for the wind-related variables (radial velocity and wind profiler observations)

440   because its analysis errors for *u* and *v* are the smallest. Despite the small forecast errors yielded by all the DA experiments, the shock to the model cannot be ignored, particularly for *Z* and *ps* (Figure 10b, f). For this issue, the noisy and imbalanced *mu* analysis is one contributor; the other contributor is likely the lack of a cross-variable-balance adjustment after DA (e.g., Zeng et al., 2021). The spin-up time for the shock is approximately 3 h; thus, the current version of Local DA cannot perform hourly cycling DA. Although the forecast results are encouraging, we emphasize that these results are obtained

445   assuming a perfect model; hence, we plan to conduct experiments with an imperfect model or using real observations in future work to further assess the capabilities of Local DA.

### 4.2.1 The sensitivity to the ensemble size

Since the 15-member ensemble has large sampling errors that degrade the Local DA analysis quality, in this subsection, we investigate whether the 36-member ensemble is helpful for reducing the analysis error. Specifically, LDA_HBC_MSL,

450   LDA_OS, and LETKF_OS are rerun by using the 36-member ensemble. Figure 11 shows comparisons between the experiments using the 15-member and 36-member ensembles. For *u*, *v*, $q_c$, $q_r$, and $q_i$ (Figure 11a, b, f, g, and h), the positive impact of using a large-size ensemble is notable. The reduction in the analysis error is larger in LDA_OS_36m and LETKF_OS_36m than in LDA_HBC_MSL_36m; a straightforward reason for this distinction is that the impact of the ensemble is reduced by 50% in LDA_HBC_MSL_36m due to the use of the hybrid covariance. Additionally, even though

455   the analysis errors of the aforementioned variables are substantially reduced in LDA_OS_36m and LETKF_OS_36m, the errors are still larger than or comparable to those in LDA_HBC_MSL, further indicating the importance of utilizing the hybrid covariance. For the other variables, the impacts of using the 36-member ensemble are relatively small and inconsistent.

The forecast results shown in Figure 12 are consistent with the analysis results in Figure 11; namely, the errors in

460   LDA_HBC_MSL_36m, LDA_OS_36m, and LETKF_OS_36m are smaller than their counterparts using the 15-member ensemble for all the investigated variables. For $V_r$, *Z*, *u*, *v*, and *ps*, the forecast errors of LDA_HBC_MSL_36m are often the smallest (Figure 12). Although the analysis error difference between LDA_HBC_MSL_36m and LDA_HBC_MSL is not large, the difference becomes distinct by the end of the 18-h forecast. Moreover, the shock to the model is alleviated in

LDA_HBC_MSL_36m in terms of *Z* and *ps*, indicating that the large sampling error is a contributor to the imbalanced

465 analysis. In general, Local DA can utilize a large ensemble size to improve the analysis.

### 4.2.2 The impact of static covariance

The difference between LDA_ctrl and LDA_HBC highlights the importance of blending the static covariance, even though the static covariance is composed only of distance-correlation information. In this subsection, we assess the performance of Local DA in the case of using the static covariance only. To achieve this goal, we rerun LDA_DS with the weight of static

470 covariance being 1.0 (LDA_DS_noENS). For brevity, only the forecast results are shown in Figure 13, revealing that LDA_DS yields smaller analysis errors for all the evaluated variables than does LDA_DS_noENS; again, this finding highlights the importance of hybrid DA. However, the main contributor to the small analysis error is the static covariance. Without the ensemble-based covariance, the analysis error of LDA_DS_noENS is not much larger than that of LDA_DS. For *PWV* and *ps*, the forecast errors of both experiments are much closer; this result, on the other hand, demonstrates the

475 ability of Local DA to run without the ensemble. Nevertheless, considering the poor quality of the time-lagged ensemble used in this study, the results shown in Figure 13 do not represent the superiority of the static covariance over the dynamic covariance.

### 4.2.3 The cycling DA performance

In this subsection, the cycling DA performances are evaluated. For the sake of a fair comparison, the LETKF is not involved

480 because we conduct only the single deterministic analysis and forecast. Figure 14 shows that the forecast errors at 12 UTC and 18 UTC in the cycling DA experiments are smaller than those in the forecasts with a single deterministic analysis, indicating that Local DA can be used for cycling DA. LDA_HBC_MSL_cyc and LDA_DS_cyc produce consistently smaller forecast errors than LDA_ctrl_cyc for most of the observation variables, except for the radar observations at 18 UTC; this outcome is expected because the ensemble is not updated, so the pure ensemble-based Local DA analysis cannot produce

485 errors comparable to those of the hybrid approach through cycling DA. Given the poor ensemble used in this study, we expect that using a well-sampled and updated ensemble would yield a lower forecast error during cycling DA.

### 4.3 The computational cost and efficiency

The computational cost and efficiency of Local DA are discussed in this subsection. All tests are conducted on a 36-core workstation with an Intel Xeon Gold 6139 CPU (the maximum frequency is set to 2.30 GHz) and 48 gigabytes of available

490 memory. Heretofore, we have implemented the parallel Local DA with OpenMP, which is not suitable for large-scale parallel computing; however, for the purpose of this study, OpenMP is sufficient. The parallel efficiency is examined first. LDA_HBC_MSL is selected as an example. Figure 15 shows the wall clock time as a function of the number of cores. The wall clock time covers Local DA steps 3) through 9) (as described in Sect. 2d). As expected, the wall clock time is reduced by approximately 50% upon doubling the number of cores, which is valid if the number of cores is not greater than 16. In

495    contrast, increasing the number of cores from 16 to 32 does not shorten the wall clock time; this is attributable to the fact that OpenMP is suitable only when the number of processors is small (<16) (Hoeflinger et al., 2001). Given that no messages need to be passed between the cores for steps 3) through 9), the parallel efficiency of Local DA is likely insensitive to the number of cores. In general, the results demonstrate that Local DA can be highly parallelized.

In addition to its parallelization, the computational speed of Local DA is also investigated. LDA_HBC_MSL takes 225 s to
500    complete all local analyses when 16 cores are used. Note that the number of horizontal grid points within the forecast domain is 40000, and more than 200000 observations are assimilated. Given that the processors work at a frequency of 2.30 GHz, the computational speed of Local DA is acceptable. On average, nearly 70% of the computational time is used to compute $C_{oo}$ and $C_{mo}$; for the minimization using the CG method, the corresponding percentage is approximately 18%.

We also assess the memory consumption of Local DA. To complete Local DA steps 3) through 9), LDA_HBC_MSL uses
505    approximately 4 gigabytes when 16 cores are engaged to store $C_{oo}$ and the associated matrices. By contrast, the LETKF uses only hundreds of megabytes. For each 5-column analysis, the $C_{oo}$ size approximately varies from 2000×2000 to 4500×4500, which is affordable. However, for a much larger size, such as 9000×9000, OpenMP is insufficient; under these circumstances, the MPI-OpenMP hybrid scheme is likely a viable solution for both the computational speed and the memory consumption, which is what we are working on. In general, the total computational cost of Local DA is high, but the cost of
510    each local analysis is affordable. With proper data thinning and parallelization, it is possible for Local DA to be used in the assimilation of real observations.

## 5 Summary and conclusions

This study proposed a local data assimilation scheme (Local DA) that can utilize the hybrid covariance and multiscale localization simultaneously. Local DA explicitly computes a local background error correlation matrix and uses the
515    correlation matrix to construct a local error sample matrix. After localization, the error sample matrix allows Local DA to adopt the conjugate gradient (CG) method to solve the cost function. The constructed matrix also allows Local DA to perform analyses by using only static correlation information. Local DA is assessed assuming a perfect model and by using simulated multiscale observations of a typhoon case. We examine the performance of Local DA in both a single deterministic forecast and cycling DA. Only deterministic analyses and forecasts are considered. Several conclusions can be
520    drawn from the results of the DA experiments:

i) Applying the CG method independently for each column group does not result in a serious discontinuity in the Local DA analysis;

ii) Explicitly computing the background correlation matrix projected onto observation-associated grids/columns is computationally affordable if the observations have been properly thinned;

525    iii) Local DA can effectively utilize the hybrid covariance to produce better analysis with a small ensemble size compared with an analysis using ensemble error information only;

iv) In the case of using a poor ensemble, the hybrid covariance plays a more important role than does multiscale localization; in such a scenario, performing localization in both the model-variable space and the observation-variable space can further decrease the analysis error compared with Local DA featuring only localization of the model-variable space;

530    v) Local DA requires a large amount of memory, but its computational efficiency is acceptable.

As the first study to present Local DA, this paper focuses on describing its basic formulation and algorithm. Future efforts to enhance the algorithm will include developing an MPI-OpenMP hybrid parallel scheme, a static covariance scheme that objectively determines the error variance and scales, and a better multiscale localization scheme. Furthermore, the current version of Local DA introduces a strong shock to the model, which limits the applicability of Local DA in cycling DA.

535    Therefore, we plan to add a cross-variable balance procedure after Local DA to improve the cycling DA performance. Moreover, many parameters of Local DA have yet to be tested; hence, the sensitivity of Local DA to each of these parameters will also be discussed in a future investigation.

**Code and data availability**

540    The code of Local DA v1.0 and the scripts for running the experiments in this study is available at the following link:https://doi.org/10.5281/zenodo.6609906 or by contacting the corresponding author via e-mail. The GFS data are available at https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs.

**Author contributions**

Shizhang Wang performed the coding and designed the data assimilation experiments. Xiaoshi Qiao analyzed the
545    experimental results. Both authors contributed to the writing of the paper.

**Competing interests**

The authors declare that they have no conflicts of interest.

**Acknowledgments**

This work is jointly sponsored by the National Science and Technology Major Project of the Ministry of Science and
550    Technology of China (2021YFC3000902) and the National Natural Science Foundation of China (41875129, 41505090, and 42105006).

**Appendix**

This section provides an example of the procedure used to thin the observations (as mentioned in Sect. 2b). The observations are thinned horizontally, whereas thinning does not occur in the vertical direction. First, we set several rings with different

555 radii at the center point or column of the model variables to be updated. For the 5-column analysis, the center coordinates of the variable-radius rings are the mean latitude and mean longitude of the 5 columns. The radius of the outer ring is the observation search radius mentioned in Sect. 2d (e.g., 300 km for sounding data and 15 km for radar data). From small to large, the radii of the rings are denoted $rr_1$, $rr_2$,..., $rr_{Nr}$, where Nr is the number of rings. We successively search the observations from the inner ring to the outer ring. Within the smallest ring, all ambient observations are selected; this is

560 equivalent to no thinning. For the observations located between two rings (between $rr_i$ and $rr_{i-1}$), we select one observation for each quadrant of the space between the two rings. There are four quadrants: the upper-right, lower-right, lower-left and upper-left quadrants (numbered I, II, III, and IV, respectively). A schematic plot is shown in Figure A1. If no observation is available in the smallest ring, the second ring is treated as the first ring.

Because no thinning occurs in the smallest ring, in a 1-column analysis, we still utilize all observations throughout the

565 forecast domain when Local DA is conducted at a single point. In 5-column analysis, the thinning approach discards some observations and slightly increases the analysis error relative to the 1-column analysis. Our early test (not shown) indicates that Local DA becomes very time-consuming when the thinning process is disabled, as expected. Moreover, the resulting analysis error increases because the assumption of observation errors being uncorrelated is not valid, which is not desired.

**References**

570 Bonavita, M., Trémolet, Y., Holm, E., Lang, S. T., Chrust, M., Janisková, M., Lopez, P., Laloyaux, P., de Rosnay, P., and Fisher, M.: A strategy for data assimilation, European Centre for Medium Range Weather Forecasts Reading, UK2017.

Branković, Č., Palmer, T., Molteni, F., Tibaldi, S., and Cubasch, U.: Extended-range predictions with ECMWF models: Time-lagged ensemble forecasting, Quarterly Journal of the Royal Meteorological Society, 116, 867-912, 1990.

Brousseau, P., Berre, L., Bouttier, F., and Desroziers, G.: Background-error covariances for a convective-scale data-

575 assimilation system: AROME–France 3D-Var, Quarterly Journal of the Royal Meteorological Society, 137, 409-422, 2011.

Brousseau, P., Berre, L., Bouttier, F., and Desroziers, G.: Flow-dependent background-error covariances for a convective-scale data assimilation system, Quarterly Journal of the Royal Meteorological Society, 138, 310-322, 2012.

Buehner, M.: Evaluation of a spatial/spectral covariance localization approach for atmospheric data assimilation, Mon Weather Rev, 140, 617-636, 2012.

580 Buehner, M. and Shlyaeva, A.: Scale-dependent background-error covariance localisation, Tellus A: Dynamic Meteorology and Oceanography, 67, 28027, 2015.

Caron, J.-F. and Buehner, M.: Scale-dependent background error covariance localization: Evaluation in a global deterministic weather forecasting system, Mon Weather Rev, 146, 1367-1381, 2018.

Caron, J.-F., Michel, Y., Montmerle, T., and Arbogast, É.: Improving background error covariances in a 3D ensemble–

585 variational data assimilation system for regional NWP, Mon Weather Rev, 147, 135-151, 2019.

Dudhia, J.: Numerical study of convection observed during the winter monsoon experiment using a mesoscale, two-dimensional model, J. Atmos. Sci., 46, 3077-3107, 1989.

Etherton, B. J. and Bishop, C. H.: Resilience of hybrid ensemble/3DVAR analysis schemes to model error and ensemble covariance error, Mon Weather Rev, 132, 1065-1080, 2004.

590 Evensen, G.: Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using Monte-Carlo Methods to Forecast Error Statistics, J Geophys Res-Oceans, 99, 10143-10162, 1994.

Gao, J. and Stensrud, D. J.: Assimilation of reflectivity data in a convective-scale, cycled 3DVAR framework with hydrometeor classification, Journal of the Atmospheric Sciences, 69, 1054-1065, 2012.

Hamill, T. M. and Snyder, C.: A hybrid ensemble Kalman filter–3D variational analysis scheme, Mon Weather Rev, 128, 595 2905-2919, 2000.

Hoeflinger, J., Alavilli, P., Jackson, T., and Kuhn, B.: Producing scalable performance with OpenMP: Experiments with two CFD applications, Parallel Computing, 27, 391-413, 2001.

Hong, S.-Y. and Lim, J.-O. J.: The WRF single-moment 6-class microphysics scheme (WSM6), Asia-Pacific Journal of Atmospheric Sciences, 42, 129-151, 2006.

600 Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Mon Weather Rev, 134, 2318-2341, 2006.

Huang, B., Wang, X., and Bishop, C. H.: The High-Rank Ensemble Transform Kalman Filter, Mon Weather Rev, 147, 3025-3043, 10.1175/mwr-d-18-0210.1, 2019.

Huang, B., Wang, X., Kleist, D. T., and Lei, T.: A simultaneous multiscale data assimilation using scale-dependent 605 localization in GSI-based hybrid 4DEnVar for NCEP FV3-based GFS, Mon Weather Rev, 149, 479-501, 2021.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, Physica D, 230, 112-126, DOI 10.1016/j.physd.2006.11.008, 2007.

Johnson, A., Wang, X., Carley, J. R., Wicker, L. J., and Karstens, C.: A comparison of multiscale GSI-based EnKF and 3DVar data assimilation using radar and conventional observations for midlatitude convective-scale precipitation forecasts, 610 Mon Weather Rev, 143, 3087-3108, 2015.

Kain, J. S.: The Kain–Fritsch convective parameterization: an update, Journal of applied meteorology, 43, 170-181, 2004.

Kalnay, E. and Yang, S. C.: Accelerating the spin-up of Ensemble Kalman Filtering, Quart. J. Roy. Meteor. Soc., submitted, 2008.

Kleist, D. T. and Ide, K.: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. 615 Part I: System description and 3D-hybrid results, Mon Weather Rev, 143, 433-451, 2015.

Lei, L., Wang, Z., and Tan, Z.-M.: Integrated Hybrid Data Assimilation for an Ensemble Kalman Filter, Mon Weather Rev, 149, 4091-4105, 2021.

Lorenc, A.: The potential of the ensemble Kalman filter for NWP - a comparison with 4D-Var, Quart. J. Roy. Meteor. Soc., 129, 3183-3204, 2003.

620 Ménétrier, B. and Auligné, T.: An Overlooked Issue of Variational Data Assimilation, Mon Weather Rev, 143, 3925-3930, 10.1175/mwr-d-14-00404.1, 2015.

Meng, Z. Y. and Zhang, F. Q.: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect model experiments, Mon. Wea. Rev., 135, 1403-1423, Doi 10.1175/Mwr3352.1, 2007.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer for inhomogeneous 625 atmospheres: RRTM, a validated correlated-k model for the longwave, Journal of Geophysical Research: Atmospheres, 102, 16663-16682, 1997.

Parrish, D. F. and Derber, J. C.: The National Meteorological Center's spectral statistical-interpolation analysis system, Mon Weather Rev, 120, 1747-1763, 1992.

Penny, S. G.: The hybrid local ensemble transform Kalman filter, Mon Weather Rev, 142, 2139-2149, 2014.

630 Shewchuk, J. R.: An introduction to the conjugate gradient method without the agonizing, Edition, School of Computer Science, Carnegie Mellon University,

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., and Powers, J. G.: A description of the advanced research WRF version 3, National Center For Atmospheric Research Boulder CoNCAR/TN-475+STR, 91, 2018.

635 Storto, A. and Andriopoulos, P.: A new stochastic ocean physics package and its application to hybrid-covariance data assimilation, Quarterly Journal of the Royal Meteorological Society, 147, 1691-1725, 2021.

Tang, Y., Ambandan, J., and Chen, D.: Nonlinear measurement function in the ensemble Kalman filter, Advances in Atmospheric Sciences, 31, 551-558, 2014.

Wang, S., Xue, M., and Min, J.: A four-dimensional asynchronous ensemble square-root filter (4DEnSRF) algorithm and
640    tests with simulated radar data, Quarterly Journal of the Royal Meteorological Society, DOI:10.1002/qj.1987,
10.1002/qj.1987, 2012.

Wang, S., Xue, M., Schenkman, A. D., and Min, J.: An iterative ensemble square root filter and tests with simulated radar
data for storm-scale data assimilation, Quarterly Journal of the Royal Meteorological Society, 139, 1888-1903, 2013.

Wang, X., Barker, D. M., Snyder, C., and Hamill, T. M.: A hybrid ETKF–3DVAR data assimilation scheme for the WRF
645    model. Part I: Observing system simulation experiment, Mon Weather Rev, 136, 5116-5131, 2008.

Wang, X., Hamill, T. M., Whitaker, J. S., and Bishop, C. H.: On the theoretical equivalence of differently proposed
ensemble - 3DVAR hybrid analysis scheme, Mon. Wea. Rev., 135, 1055-1076, 2007.

Wang, X., Hamill, T. M., Whitaker, J. S., and Bishop, C. H.: A comparison of the hybrid and EnSRF analysis schemes in the
presence of model errors due to unresolved scales, Mon Weather Rev, 137, 3219-3232, 2009.

650    Wang, X., Chipilski, H. G., Bishop, C. H., Satterfield, E., Baker, N., and Whitaker, J. S.: A multiscale local gain form
ensemble transform Kalman filter (MLGETKF), Mon Weather Rev, 149, 605-622, 2021.

Xiao, Q. and Sun, J.: Multiple-radar data assimilation and short-range quantitative precipitation forecasting of a squall line
observed during IHOP_2002, Mon Weather Rev, 135, 3381-3404, 2007.

Xue, M., Tong, M. J., and Droegemeier, K. K.: An OSSE framework based on the ensemble square root Kalman filter for
655    evaluating the impact of data from radar networks on thunderstorm analysis and forecasting, J. Atmos. Oceanic Technol., 23,
46-66, 2006.

Yang, C., Min, J., and Tang, Y.: Evaluation of two modified Kalman gain algorithms for radar data assimilation in the WRF
model, Tellus A: Dynamic Meteorology and Oceanography, 67, 25950, 2015.

Yang, S. C., Kalnay, E., Hunt, B., and E. Bowler, N.: Weight interpolation for efficient data assimilation with the local
660    ensemble transform Kalman filter, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric
sciences, applied meteorology and physical oceanography, 135, 251-262, 2009.

Zeng, Y., de Lozar, A., Janjic, T., and Seifert, A.: Applying a new integrated mass-flux adjustment filter in rapid update
cycling of convective-scale data assimilation for the COSMO model (v5. 07), Geoscientific Model Development, 14, 1295-
1307, 2021.

665    Zhang, F., Weng, Y., Sippel, J. A., Meng, Z., and Bishop, C. H.: Cloud-resolving hurricane initialization and prediction
through assimilation of Doppler radar observations with an ensemble Kalman filter, Mon Weather Rev, 137, 2105-2125,
2009.

670

**Tables**

**Table 1 DA experiment configurations.**

| Experiment names | DA scheme | Static covariance | Dynamic covariance | Localization space | Multiscale localization | Cycling DA | Ensemble size |
|---|---|---|---|---|---|---|---|
| LDA_ctrl | Local DA | No | Yes | M | No | No | 15 |
| LDA_HBC_MSL | Local DA | Yes | Yes | M | Yes | No | 15 |
| LDA_HBC | Local DA | Yes | Yes | M | No | No | 15 |
| LDA_DS | Local DA | Yes | Yes | M+O | Yes | No | 15 |
| LDA_OS | Local DA | No | Yes | O | Yes | No | 15 |
| LETKF_OS | LETKF | No | Yes | O | Yes | No | 15 |
| LDA_HBC_MSL_cyc | Local DA | Yes | Yes | M | Yes | Yes | 15 |
| LDA_ctrl_cyc | Local DA | No | Yes | M | No | Yes | 15 |
| LDA_DS_cyc | Local DA | Yes | Yes | M+O | Yes | No | 15 |
| LDA_HBC_MSL_36m | Local DA | Yes | Yes | M | Yes | Yes | 36 |
| LDA_OS_36m | Local DA | No | Yes | O | Yes | No | 36 |
| LETKF_OS_36m | Local DA | No | Yes | O | Yes | No | 36 |
| LDA_DS_noENS | Local DA | Yes | No | M+O | Yes | No | 15 |

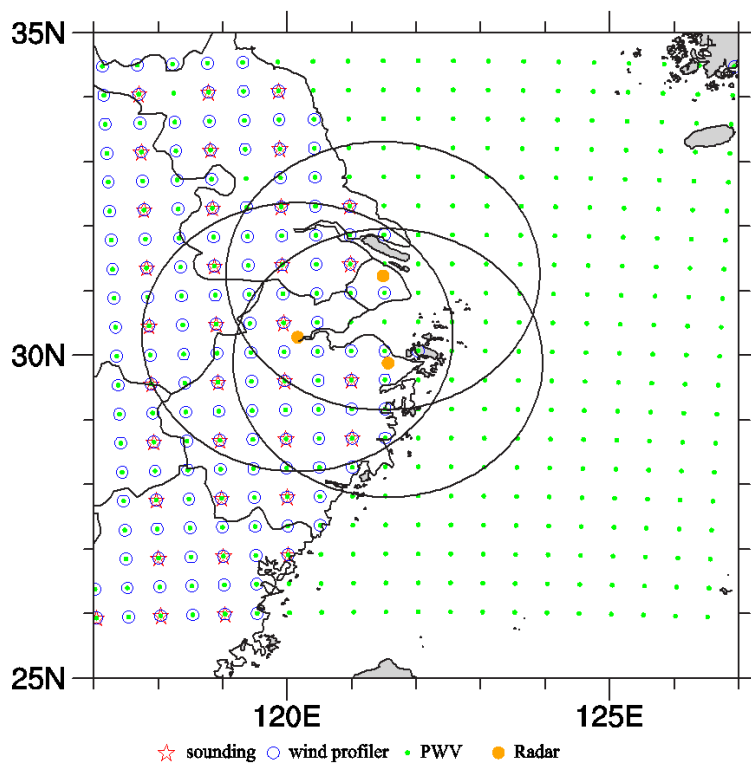
675

**Figures**

**Figure 1 The distribution of simulated observations, where the black rings denote the maximum observation ranges of radars.**

680

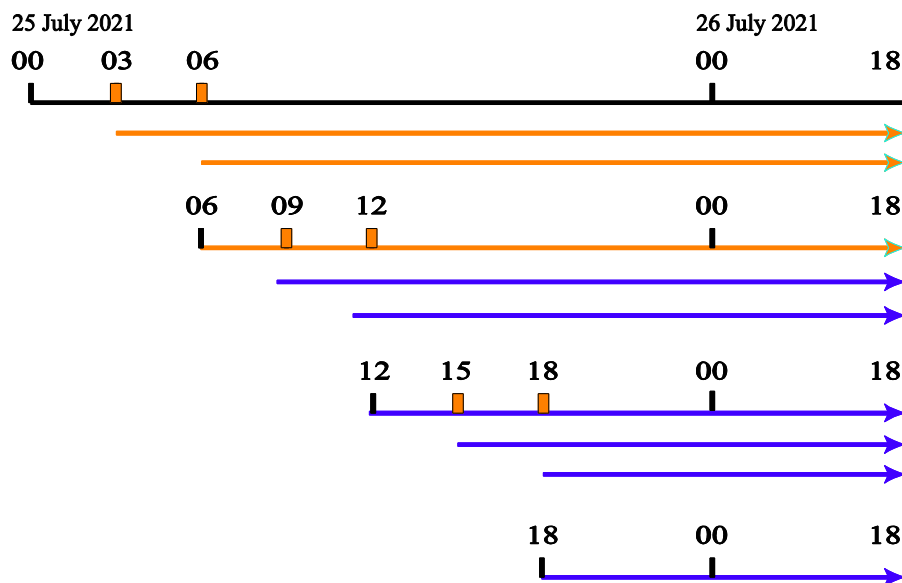



**Figure 2 A flow chart of the time-lagged ensemble generation, where the colored arrows represent the sample forecasts used by the 36-member ensemble. The 15-member ensemble excludes the forecasts shown with orange arrows. The sample forecasts initialized by using the GFS forecast data are highlighted with orange tick marks. The ensembles are available at 00 UTC, 06 UTC, 12 UTC, and 18 UTC on 26 July 2021.**

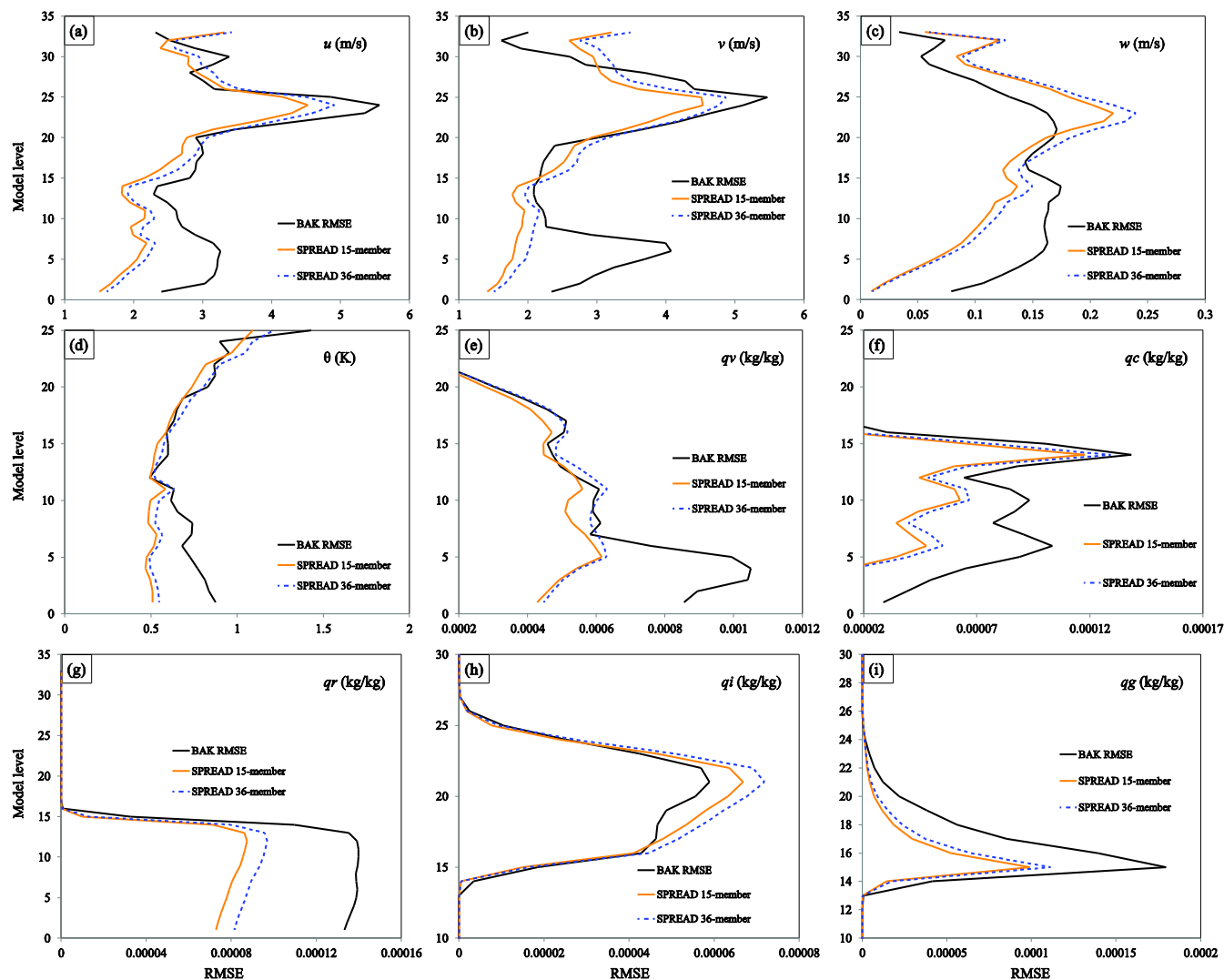
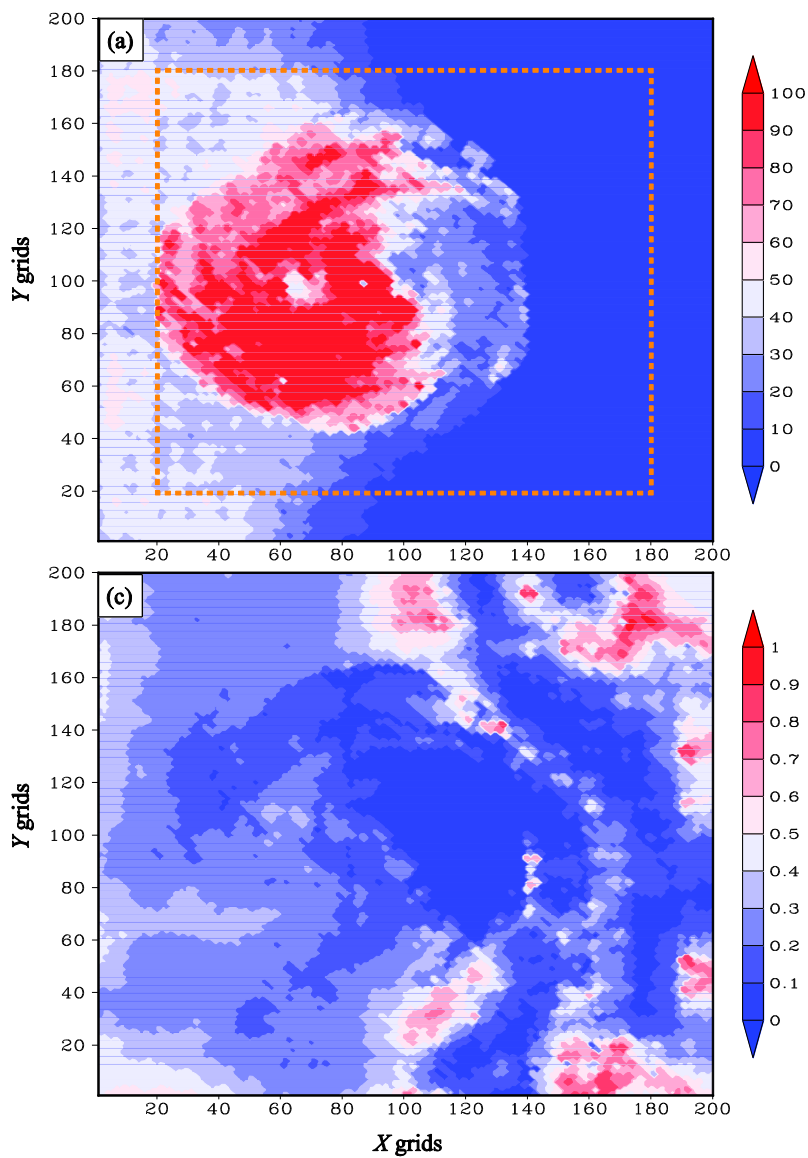
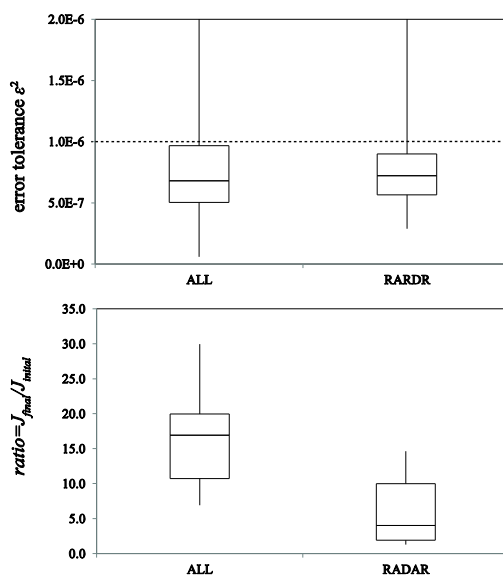
685

690 **Figure 3 The ensemble spread profiles for the 15-member ensemble (orange lines) and the 36-member ensemble (dashed blue lines), where the model levels of some plots are truncated to highlight the differences. "BAK" (black lines) denotes the RMSE of the initial condition without DA.**

695

**Figure 4** The spatial distributions of (a) the number of iterations and (b) the ratio of the final value of the cost function to the initial value.

700 **Figure 5 The boxplots of (a) $\varepsilon^2$ and (b) the ratio of the final $J$ to the initial $J$ in the dahsed rectangle area shown in Figure 1, where "ALL" denotes the DA using all observations and "RADAR" corresponds to the DA using radar data only. The upper and lower bounds of the boxes are the 75th and 25th percentiles, respectively. The middle line indicates the median.**

705

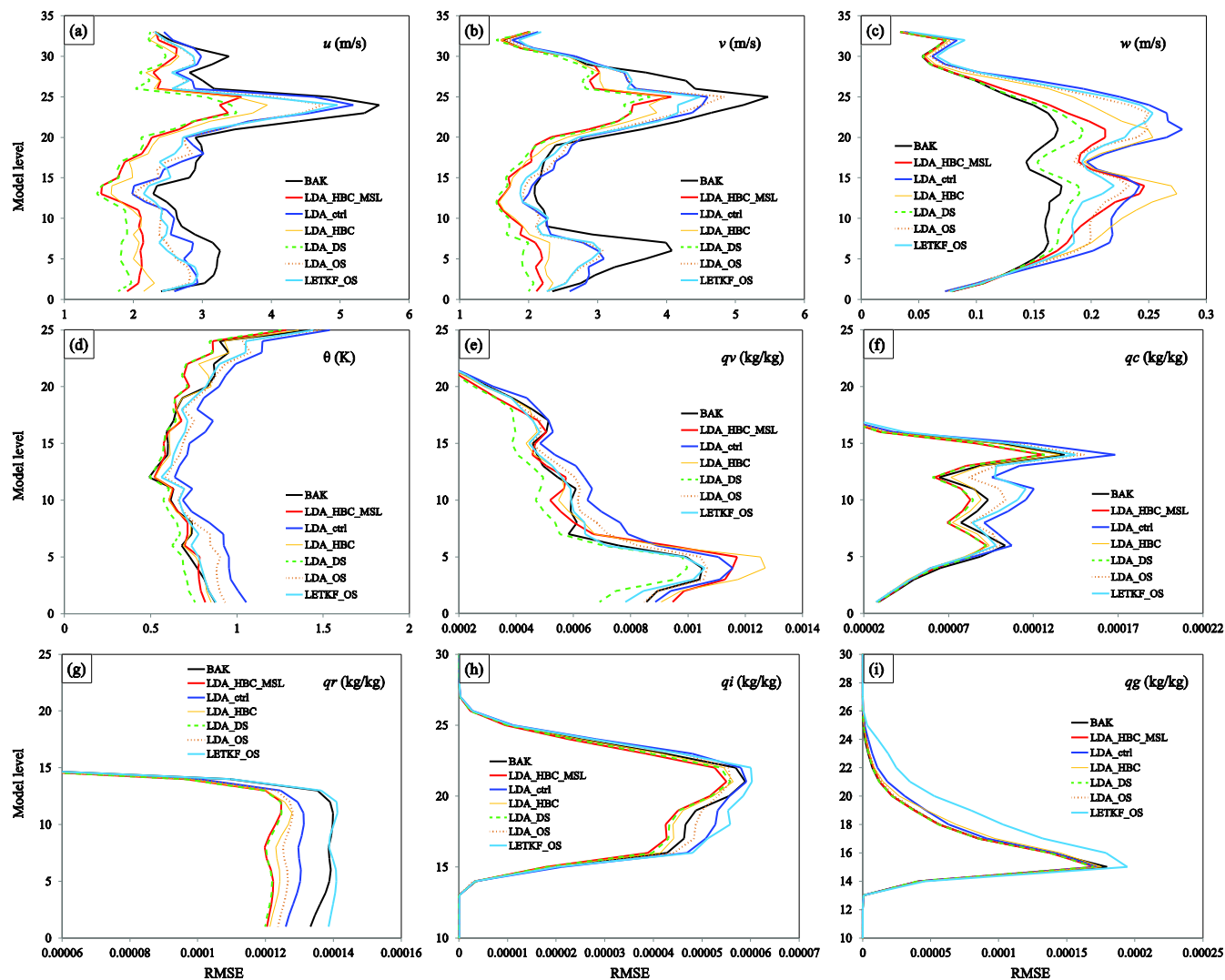



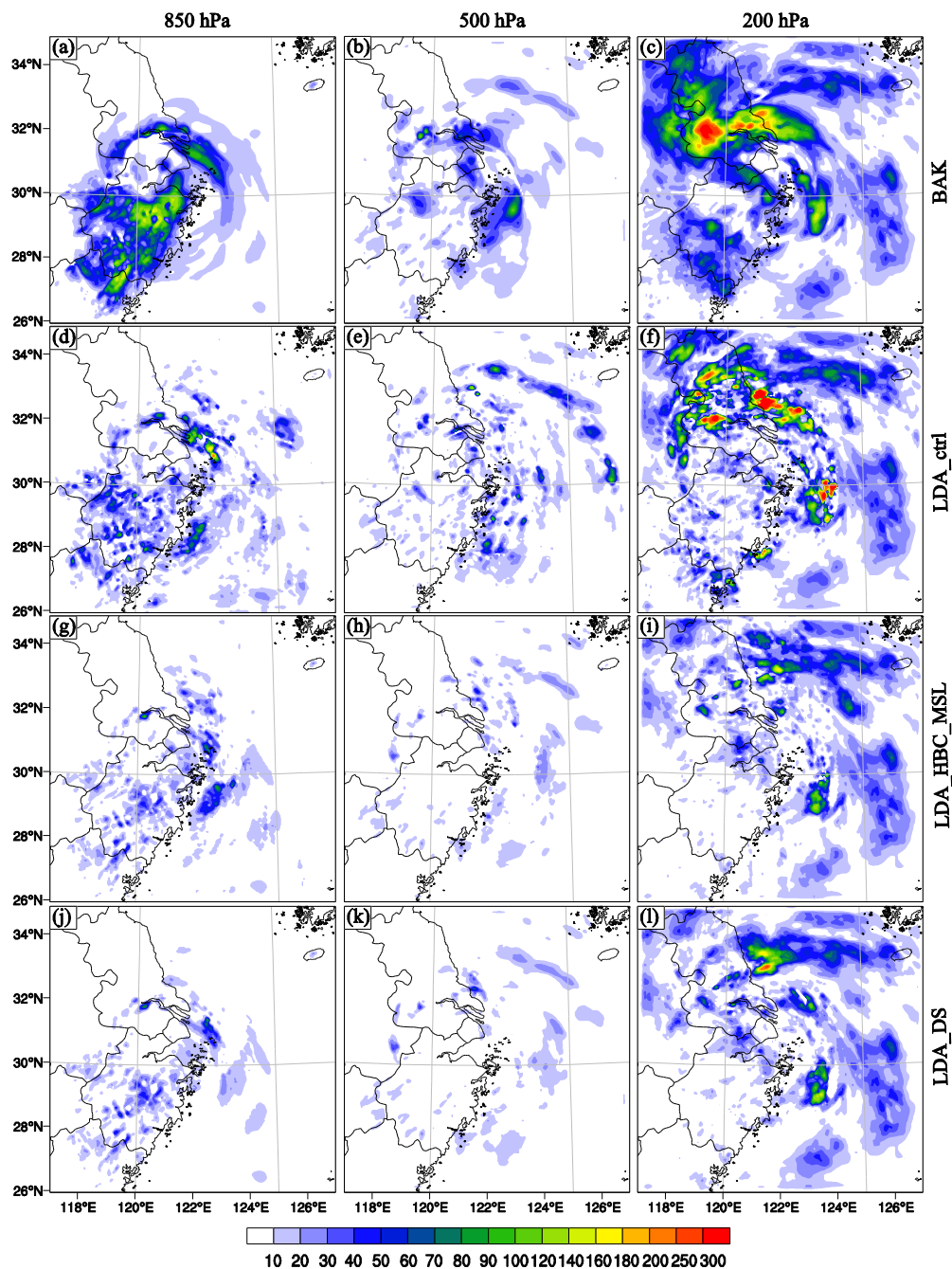
**Figure 6 The RMSE analysis profiles for experiments LDA_HBC_MSL (red lines), LDA_ctrl (blue lines), LDA_HBC (thin yellow lines), LDA_DS (dashed green lines), LDA_OS (dot brown lines), and LETKF_OS (cyan lines), where the model levels of some plots are truncated to highlight the main differences among the experiments. "BAK" (black lines) denotes the initial condition without DA.**

715

**Figure 7 The DTE at 850 hPa (left column), 500 hPa (middle column), and 200 hPa (right column) for (a-c) BAK, (d-f) LDA_ctrl, (g-i) LDA_HBC_MSL, and (j-l) LDA_DS.**

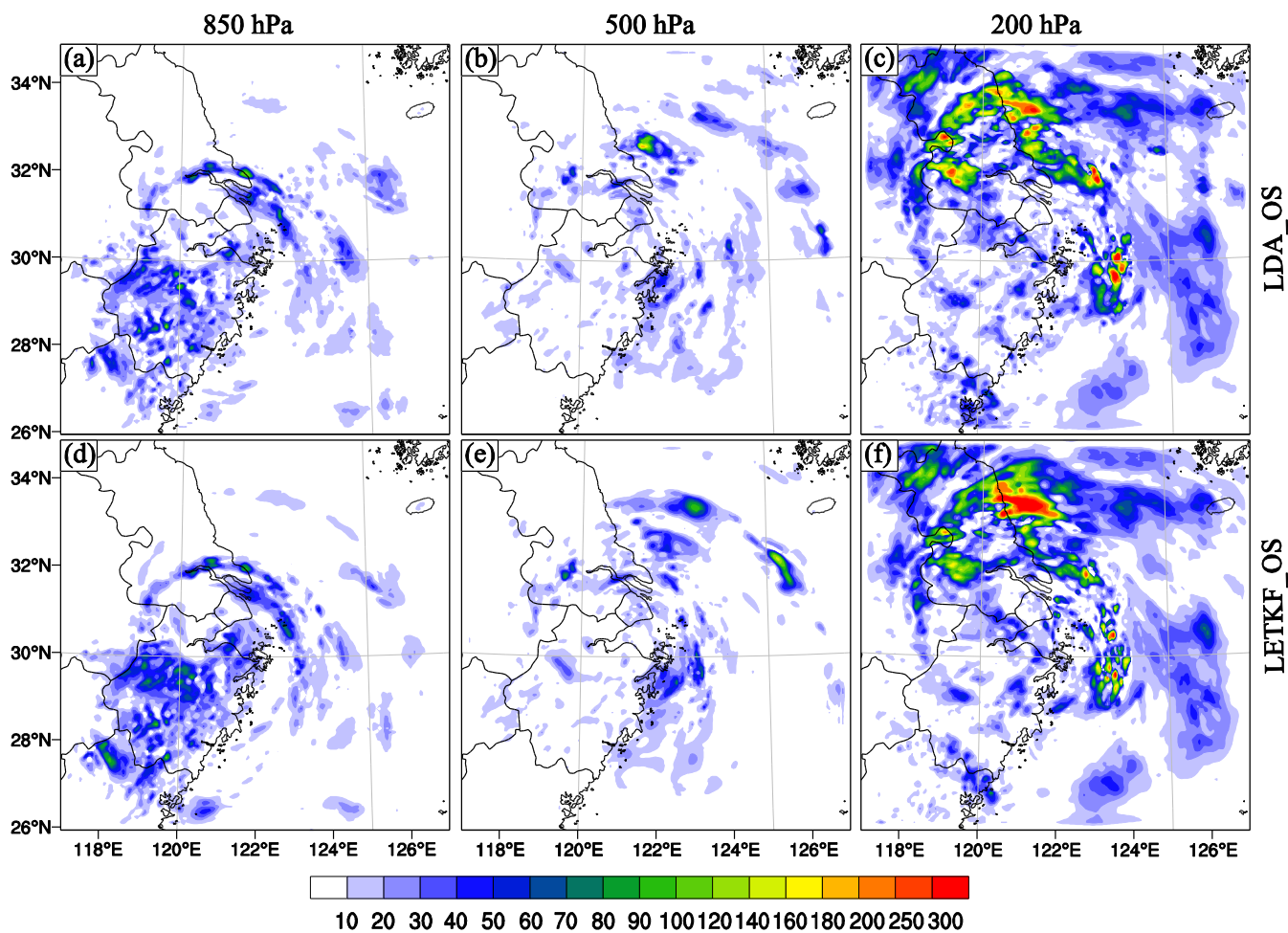



**Figure 8** The same as in Figure 7 but for (a-c) LDA_OS and (d-f) LETKF_OS.

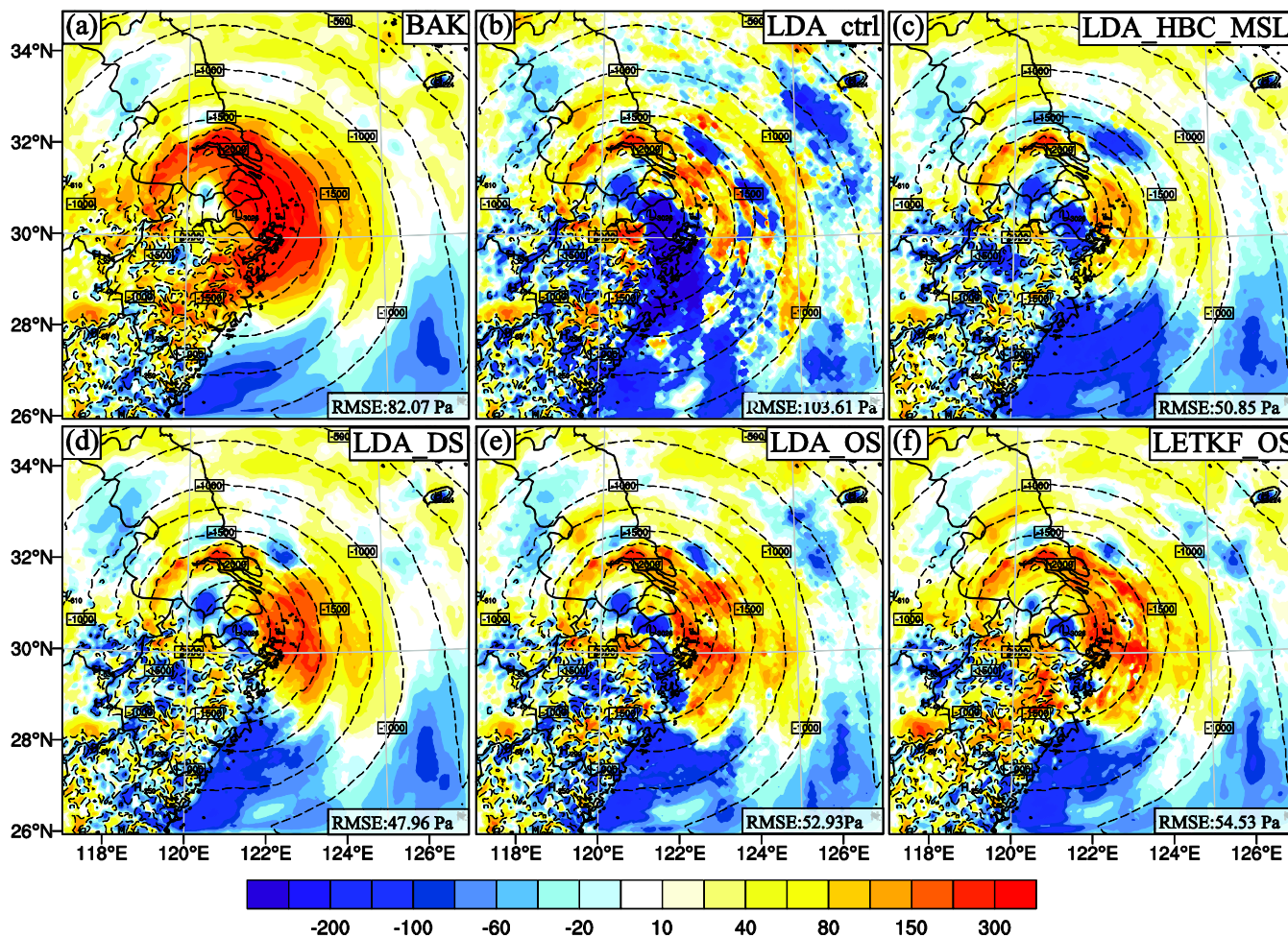
720

725

**Figure 9 The difference in the dry-air mass in column between the truth (contours) and analysis (shading) for (a) BAK, (b) LDA_ctrl, (c) LDA_HBC_MSL, (d) LDA_DS, (e) LDA_OS, and (f) LETKF_OS.**

730

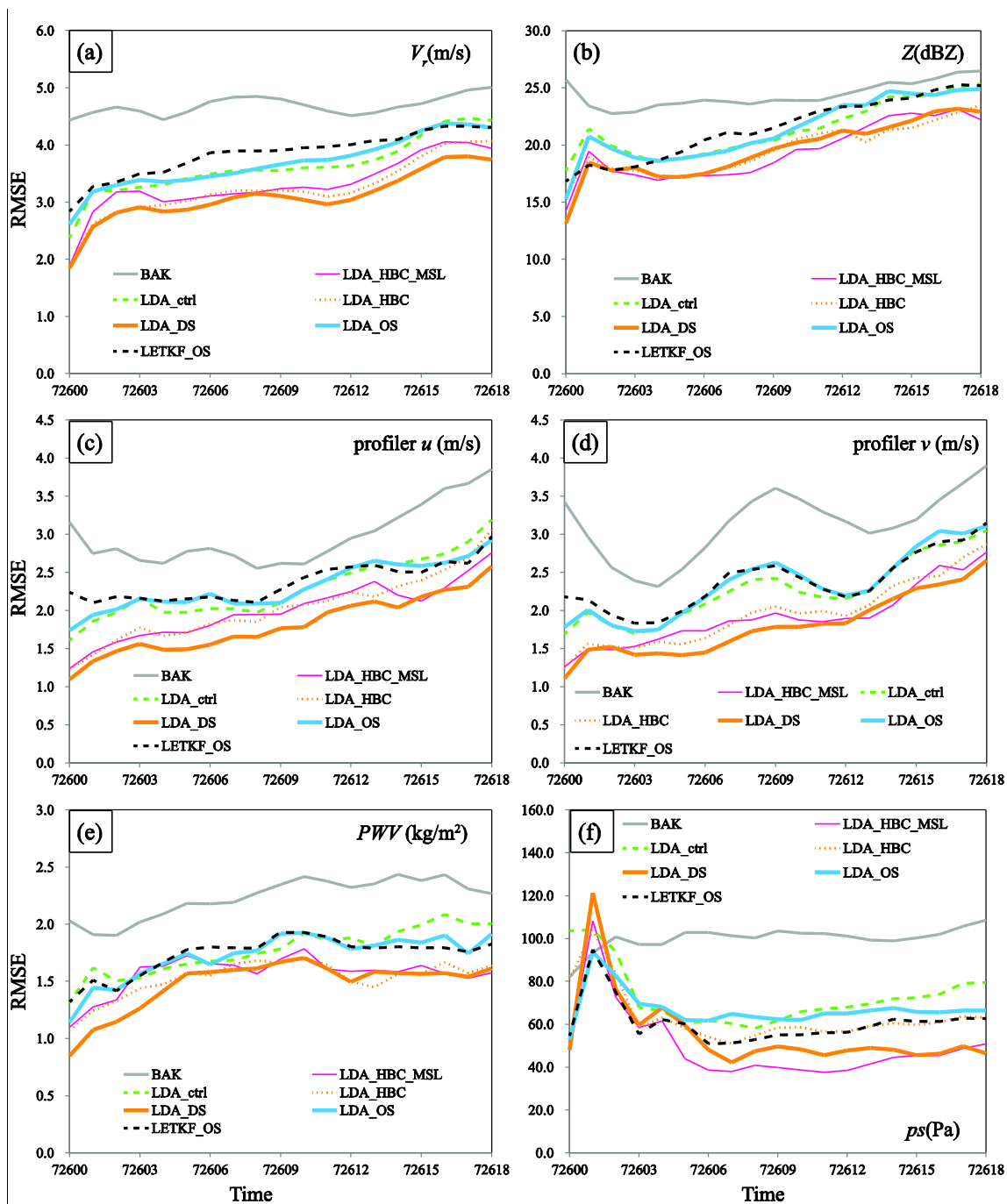



**Figure 10 The evolutions of the forecast RMSEs in terms of the (a) radial velocity, (b) reflectivity, (c) _u_ component (wind profiler),**
**(d) _v_ component (wind profiler), (e) _PWV_, and (f) _ps_ for BAK (gray lines), LDA_HBC_MSL (thin magenta lines), LDA_ctrl**
**(dashed green lines), LDA_HBC (dot red lines), LDA_DS (thick orange lines), LDA_OS (cyan lines), and LETKF_OS (dashed**
**black lines).**

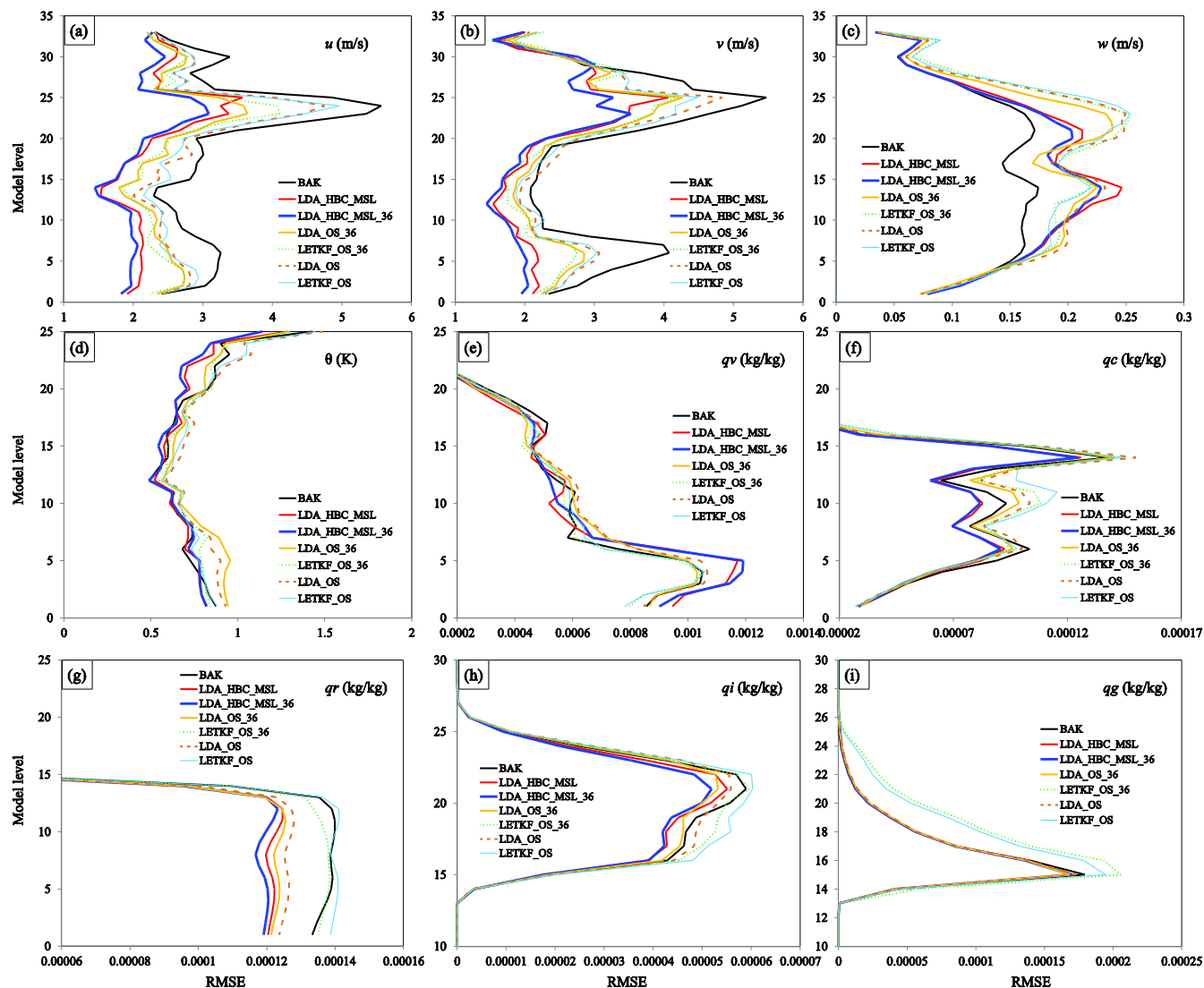



**Figure 11 The same as in Figure 6 but for LDA_HBC_MSL_36m (blue lines), LDA_OS_36m (orange lines), and LETKF_OS_36m (dot green lines). The results of BAK, LDA_HBC_MSL, LDA_OS, and LETKF_OS are plotted for comparison.**

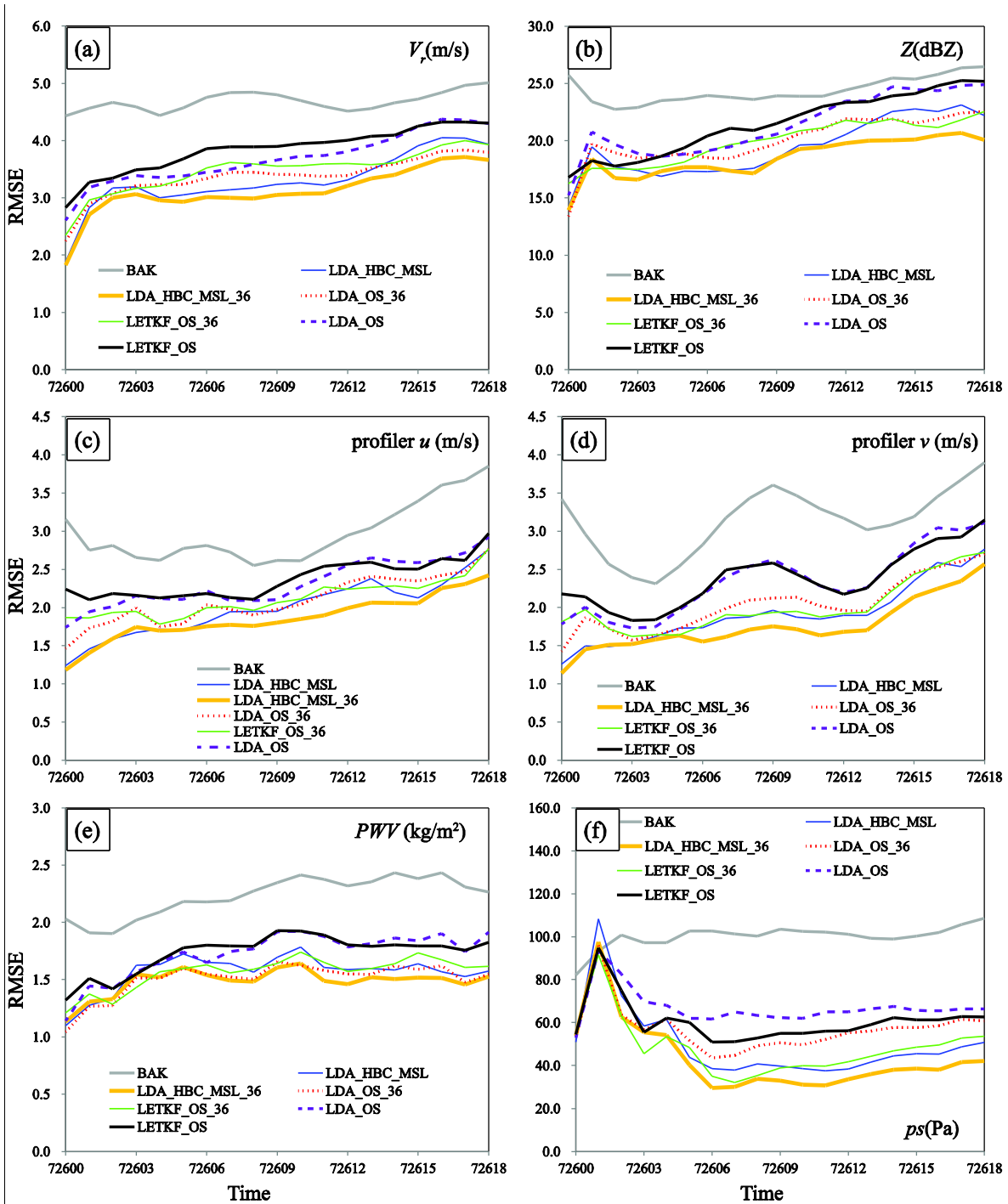



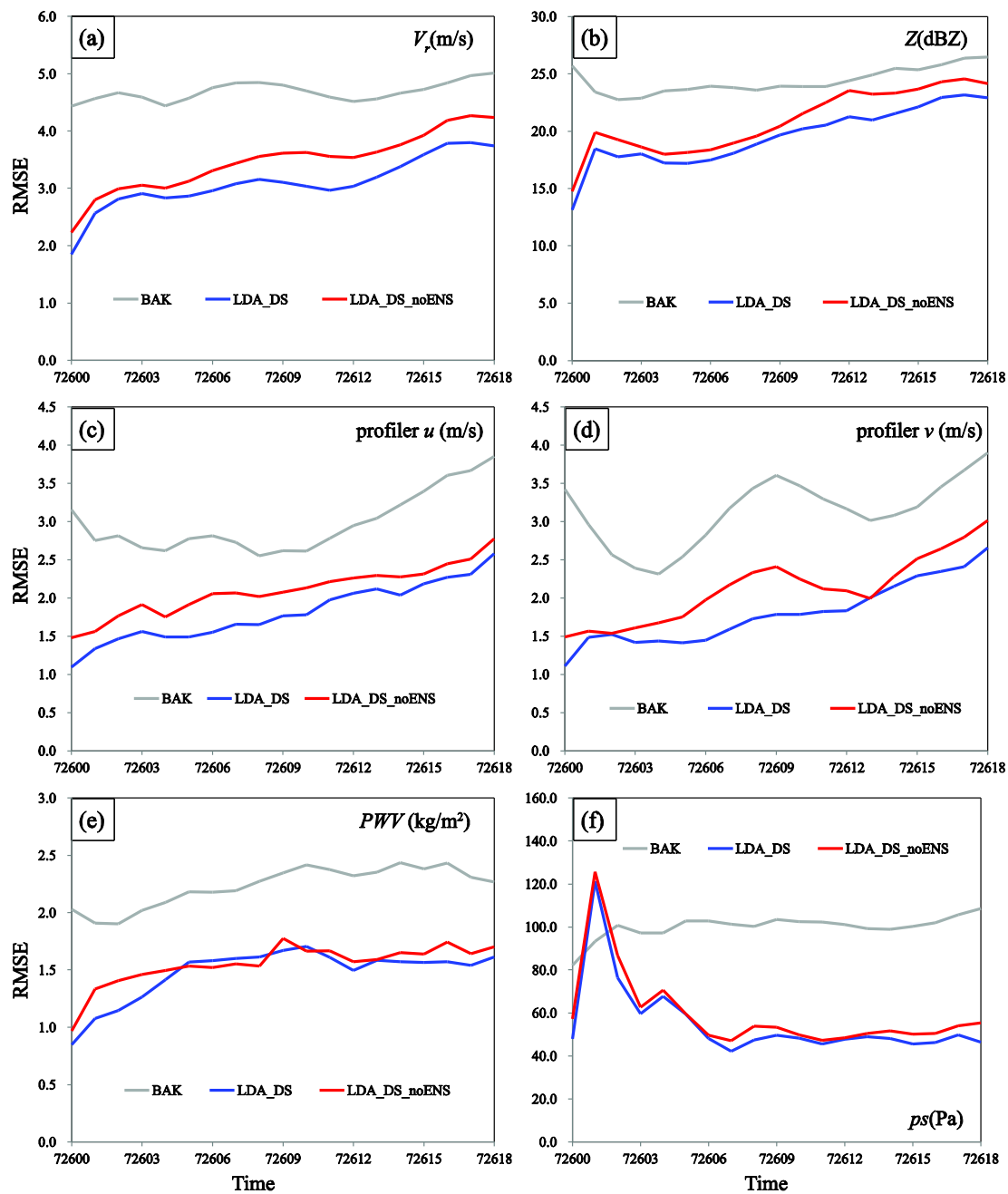
**Figure 12 The same as in Figure 10 but for LDA_HBC_MSL_36m (thick orange lines), LDA_OS_36m (dashed red lines), and LETKF_OS_36m (thin green lines). The results of BAK, LDA_HBC_MSL, LDA_OS, and LETKF_OS are plotted for reference.**

750

**Figure 13 The same as in Figure 10 but to compare LDA_DS_noENS (red lines) and LDA_DS (blue lines). The result of BAK is plotted for reference.**

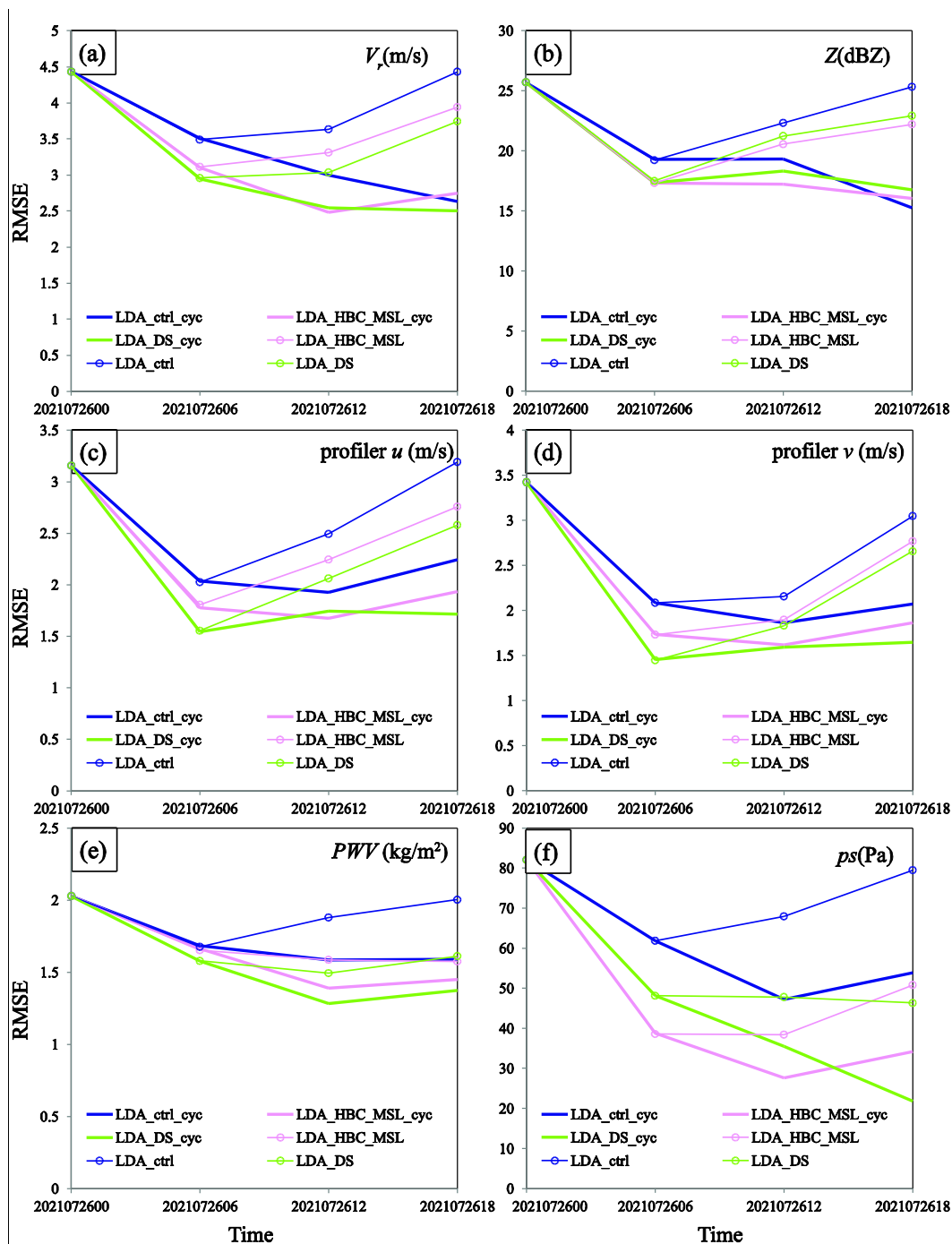
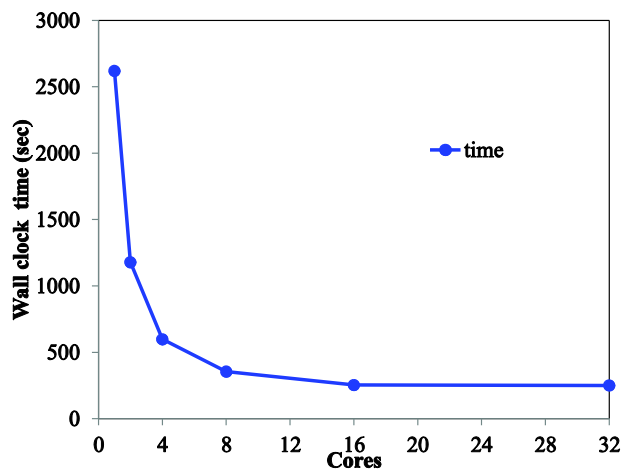



**Figure 14** The evolutions of the forecast RMSE in terms of the (a) radial velocity, (b) reflectivity, (c) *u* component (wind profiler), (d) *v* component (wind profiler), (e) PWV, and (f) *ps* for LDA_ctrl_cyc (blue), LDA_HBC_MSL_cyc (magenta), and LDA_DS_cyc (green). The forecasts of LDA_ctrl, LDA_HBC_MSL, and LDA_DS are plotted for reference.

760 **Figure 15 The wall clock time as a function of the number of cores used in the parallel test.**

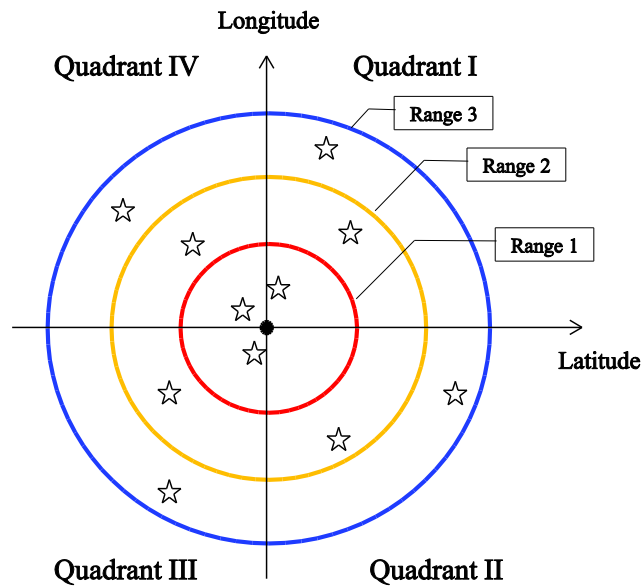



**Figure A1 A schematic of the observation searching approach used in Local DA, where stars represent the selected observations near the grid point (dark solid dot) to be analyzed.**

765