Improving precipitation forecast with hybrid 3DVar and time-lagged ensembles in a heavy rainfall event

Yuanbing Wang a, Jinzhong Min a,⁎, Yao-deng Chen a, Xiang-Yu Huang b, Mingjian Zeng c, Xin Li c

a Key Laboratory of Meteorological Disaster of Ministry of Education (KLME)/Joint International Research Laboratory of Climate and Environment Change (ILCEC)/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science & Technology, Nanjing 210044, China

b Centre for Climate Research Singapore, Meteorological Service, Singapore 537054, Singapore

c Jiangsu Research Institute of Meteorological Sciences, Nanjing 210009, China

A R T I C L E   I N F O

Article history:
Received 14 December 2015
Received in revised form 21 June 2016
Accepted 23 July 2016
Available online 26 July 2016

Keywords:
Data assimilation
Hybrid
Time-lagged ensemble
Precipitation forecast

A B S T R A C T

This study evaluates the performance of three-dimensional variational (3DVar) and a hybrid data assimilation system using time-lagged ensembles in a heavy rainfall event. The time-lagged ensembles are constructed by sampling from a moving time window of 3 h along a model trajectory, which is economical and easy to implement. The proposed hybrid data assimilation system introduces flow-dependent error covariance derived from time-lagged ensemble into variational cost function without significantly increasing computational cost. Single observation tests are performed to document characteristic of the hybrid system. The sensitivity of precipitation forecasts to ensemble covariance weight and localization scale is investigated. Additionally, the TLEn-Var is evaluated and compared to the ETKF (ensemble transformed Kalman filter)–based hybrid assimilation within a continuously cycling framework, through which new hybrid analyses are produced every 3 h over 10 days. The 24 h accumulated precipitation, moisture, wind are analyzed between 3DVar and the hybrid assimilation using time-lagged ensembles.

Results show that model states and precipitation forecast skill are improved by the hybrid assimilation using time-lagged ensembles compared with 3DVar. Simulation of the precipitable water and structure of the wind are also improved. Cyclonic wind increments are generated near the rainfall center, leading to an improved precipitation forecast. This study indicates that the hybrid data assimilation using time-lagged ensembles seems like a viable alternative or supplement in the complex models for some weather service agencies that have limited computing resources to conduct large size of ensembles.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The development of society and economy has great requirement for numerical weather prediction (NWP), especially precipitation forecast. Accuracy of initial condition is one of the main conditions influencing forecast skill. The operational NWP centers usually enhance initial condition with data assimilation technique. In order to introduce effective synoptic information into initial condition, it is necessary to frequently assimilate recent observations for short-range NWP (Benjamin et al., 2004).

Several data assimilation methods have been proposed and applied in NWP, such as three-dimensional variational (3DVar), four-dimensional variational (4DVar), ensemble Kalman filter (EnKF) and the hybrid 3DVar or 4DVar (Hybrid) methods. Among these methods, 3DVar has been most widely used because of its advantage in computational cost and convenience to implement (Wu et al., 2002; Barker et al., 2004). However, 3DVar assumes that background error covariance is time-invariant, homogeneous and isotropic, which conflicts with the reality of “error of the day”, i.e. flow-dependence. 4DVar allows implicit evolution of the background error covariance along with the adjoint model, but the use of static background error covariance at the start of each 4DVAR assimilation window represents a major limitation (Huang et al., 2009; Zhang et al., 2014). Furthermore, the computational cost of 4DVar is huge because of the adjoint and tangent linear models and development and maintenance of them is also a hard work. The EnKF method, in which the background error covariance is estimated from an ensemble of short-term forecasts, provides an alternative to variational data assimilation systems for its flow-dependence as well as convenience to implement (Evensen, 1994; Anderson, 2001; Bishop et al., 2001; Whitaker and Hamill, 2002; Hunt et al., 2007). For smaller ensembles, however, the EnKF is rank deficient and its background error covariance estimation suffers from a variety of sampling errors, including spurious correlations for widely separated locations (Hamill and Snyder, 2000).

The hybrid data assimilation method that couples ensemble-based and variational data assimilation systems has emerged as an alternative method (Hamill and Snyder, 2000) and become one of the research...
focuses in data assimilation field (Wang et al., 2008a,b; Zhang et al., 2013; Schwartz and Liu, 2014). The hybrid assimilation method incorporates flow-dependent background error covariance derived from ensemble into the variational cost function, so that the new background error covariance comes from a combination of traditional static error covariance with ensemble error covariance which can be used to update fields of observed model variables as well as fields of unobserved model variables (Keppenne et al., 2014). The hybrid covariance takes advantage of the relative strengths of both the EnKF and variational methods and ameliorates the problem of rank deficiency caused by limitation of ensemble size as well as isotropos, homogeneous and static covariance caused by assumptions in 3DVar. Hybrid data assimilation has been shown to be more robust than conventional ensemble data assimilation schemes, especially when ensemble size is small or the model error is large (Wang et al., 2007; Zhang et al., 2013) and has been implemented in several variational assimilation systems such as the Weather Research and Forecasting (WRF) model data assimilation system (WRFDA) (Barker et al., 2012) and the Gridpoint Statistical Interpolation (GSI) system which has become operational at National Centers for Environmental Prediction (NCEP) since 2012 (Wang et al., 2013). However, the cost is significantly higher than the one by none ensemble methods. For some weather service agencies, it is probably difficult to afford the cost of ensemble model integrations, thus implementation of hybrid assimilation may compromise between size of the ensemble and the resolution of the model.

Therefore, to avoid huge cost of ensemble integrations in hybrid assimilation system which assimilates observations and outputs forecast products with high frequency, the time-lagged forecasts of short range interval which are inherited from different past analysis times but verify at a same forecast time are directly pulled together from history storage to construct an ensemble in this study, namely time-lagged ensemble (Zhou et al., 2010). The time-lagged ensemble initially proposed as an alternative to Monte Carlo ensemble method (Hoffman and Kalnay, 1983) can be interpreted as forecasts obtained from a set of perturbed initial conditions. The initial conditions which initialize the time-lagged forecasts, the observations at different analysis times, the integration time and the lateral boundary conditions are all different for the time-lagged ensemble members, thus the flow-dependent forecast error can be a result of those conditions which cause the uncertainties (Lu et al., 2007; Vogel et al., 2014). This kind of ensemble is built at very low computational cost which does not require multiple integrations of the numerical model and holds promise for high-resolution applications. It has been widely used in many research and operational ensemble forecast systems (Yuan et al., 2008, 2009; Mittermaier, 2007; Trilaksono et al., 2012; Y. Chen et al., 2013; M. Chen et al., 2013; Jie et al., 2014, 2015). It is noted that the traditional and standard EnKF-based hybrid method should be a better choice in the presence of sufficient computing resource. However, if the computational cost associated with data assimilation like EnKF and ensemble integrations is not affordable, a compromise has to be made between the efficiency and accuracy. In such scenario, a hybrid approach by merging the time-lagged ensemble and 3DVar can be a choice because of its efficiency and flow-dependent feature.

In this study, we construct a hybrid data assimilation system based upon WRFDA using the time-lagged ensembles. To evaluate the effectiveness and flow-dependence of the background error covariance derived from time-lagged ensembles in precipitation forecast, this system is applied and tested in a heavy rainfall event occurred in east China and compared with ETKF-based hybrid assimilation with a continuously cycling framework over 10 days. Details are present in the rest of the paper which is organized as follows. In Section 2, the basic methodology of the WRFDA-based hybrid assimilation using time-lagged ensembles (“TLEn-Var” for short) is introduced. Section 3 describes the rainfall event and Section 4 details the model and data assimilation configurations as well as the experiment design. Results are presented in Section 5 before we conclude in Section 6.

2. Methodology

The cost function of hybrid data assimilation in WRFDA is defined as

$$
J(\delta x_1, \alpha) = \beta_1 \frac{1}{2} \delta x_1^T B^{-1} \delta x_1 + \beta_2 \frac{1}{2} \alpha^T A^{-1} \alpha + \frac{1}{2} (y^0 - H\alpha)^T R^{-1} (y^0 - H\alpha)
$$

(1)

In Eq. (1), the first term of right hand is the background term associated with the static covariance B. The second term is associated with the ensemble covariance. α is the ensemble extended control variable. A defines the spatial covariance of α. The third term is the observation term, $$y^0 = y - H\alpha$$ is the innovation, y denotes the observation, xo is the background forecast, and H is the nonlinear observation operator. H is the linearized observation operator, and R is the observation error covariance. Factors $$\beta_1$$ and $$\beta_2$$ respectively define the weights placed on the static background error covariance and the ensemble covariance. $$\beta_1$$ and $$\beta_2$$ are constrained by $$1/\beta_1 + 1/\beta_2 = 1$$ to conserve the total background error variance.

The analysis increment of the hybrid is a sum of two terms, defined as

$$
\delta x = \delta x_1 + \sum_{i=1}^{N} (\alpha_i - x_i) = \sum_{i=1}^{N} (x_{i,n} - x_i)
$$

(2)

where $$\delta x_1$$ is the increment associated with the static background covariance, and the second term is the increment associated with the flow-dependent ensemble covariance. N is the ensemble size. For traditional (EnKF-based) hybrid assimilation, $$x_{i,n}$$ is the nth ensemble perturbation normalized by $$\sqrt{N-1}$$:

$$
x_{i,n} = (x_{i,n} - x_i)/\sqrt{N-1}
$$

(3)

where $$x_{i,n}$$ is the nth ensemble member and $$x_i$$ is the ensemble mean, provided by the ensemble forecast that is usually initialized by an EnKF data assimilation system.

However, for the hybrid assimilation using time-lagged ensembles, the flow-dependent ensemble error covariance are computed by differences of N previous instances of the model state vectors sampled from the recent history of the current model run. The differences between the time-lagged forecasts launched at different analysis times but verify at the same leading time are calculated and normalized by $$\sqrt{N-1}$$:

$$
x_{i,n} = (x_{j} - x_i)/\sqrt{N-1}, 0 < i < j \leq N
$$

(4)

In Eq. (4), $$x_i$$ and $$x_j$$ are time-lagged ensemble members, the total number of $$x_{i,n}$$ is calculated by $$\sum_{i=1}^{N-1} i$$. For example, in this study the deterministic forecast range is 48 h with an output interval of 3 h, then we can obtain 16 time-lagged ensemble members at each analysis time. The differences between time-lagged ensemble members are $$x_2 - x_1, x_3 - x_1, ..., x_{16} - x_1$$, $$x_3 - x_2, x_4 - x_2, ..., x_{16} - x_2$$. Thus $$\sum_{i=1}^{15} i = 120$$ differences are generated. Next these differences can be used as the ensemble perturbations in a hybrid assimilation run (Fig. 1). The perturbations introduced here come from the time-lagged differences in a single model integration and the main goal of this method is to create ensembles with low computational cost, which are used for the calculation of flow-dependent error covariance.

The underlying assumption in the TLEn-Var is that forecast errors in data assimilation are primarily phase errors in time (Keppenne et al., 2014). The time-lagged forecasts in an ensemble are launched at different analysis times after assimilating different observations but verifies at the same leading time, and the integration time and lateral boundary between members are also different, taking the time evolution of forecast error of different time-lagged ensemble members into account.
The latter member differs from ahead ones significantly while it still contains information that is useful at analysis time. The ensemble size is increased without significantly increasing computational cost using the time-lagged ensemble method. This may provide the hybrid assimilation with reasonable ensemble error covariance information. This method does not require the integrations of multiple model trajectories. Instead, all the necessary covariance information is obtained from a single model integration. Therefore, compared with the traditional hybrid assimilation method, the TLEn-Var is significantly economical. Although its quality may be poorer than the EnKF, from the viewpoint of economy and simplicity, this method is always worth pursuing so long as it surpasses 3DVar.

3. The heavy rainfall event

In this study, the heavy rainfall event over China in the year 2014 is selected for model simulation and further analysis. During 3–5 July, there was a large range of persistent heavy rainfall in southwest China and Jianghuai area. Particularly in 4–5 July the 24 h accumulated rainfall in Jianghuai area exceeded 100 mm. Fig. 2a shows the distribution of 24 h accumulated rainfall during 4–5 July from China Hourly Merged Precipitation Analysis (CMORPH: Shen et al., 2014). The rain band lies from southwest to northeast. The main heavy rainfall area is located at the black box (28°N–34°N, 114°E–122°E) in Fig. 2a. The averaged wind field on 400 hPa from the NCEP FNL (Final) Operational Global Analysis data during 4–5 July in Fig. 2b shows that there was a shear line lied along the rain band. Furthermore, the wind and geopotential height at multiple isobars at 0000 UTC 4 July 2014 are shown in Fig. 3. A low vortex system was located over southwest area at 850 hPa and 700 hPa. The southwesterly flow from Bengal Bay strengthened under the influence of subtropical high peripheral airflow, so that water vapor can be transported to the middle and lower reaches of the Yangtze river region. The subtropical ridge stayed steady around 25°N at 500 hPa, leading to the persistent precipitation. This indicates a synoptic situation, favorable for widespread rainfall over the region being situated along the shear line.

4. Model and experiments design

The hybrid assimilation and forecast system constructed in this study is based on version 3.7.1 of the Advanced Research Weather Research and Forecasting (ARW). All experiments are conducted over a single domain, which covers main area of east China (Fig. 2) with a
181 × 133 horizontal mesh grid using 20-km spacing and 45 vertical levels up to 50 hPa. The WRF single-moment 6-class microphysics scheme (WSM6; Hong and Lim, 2006), the Kain–Fritsch cumulus scheme (Kain, 2004), the Dudhia Shortwave Scheme (Dudhia, 1989), the RRTM longwave scheme (Mlawer et al., 1997), the Noah–MP land surface scheme (Yang et al., 2011), and the Mellor–Yamada–Janjic (MYJ) planetary boundary layer (PBL) scheme (Janjic, 1994) are used for all deterministic forecasts.

This paper presents three groups of experiments. The first group which is denoted as Covariance contains five experiments (Table 1). The experiments in this group are designed based on the weights of ensemble covariance for TLEn-Var, while the localization scale is fixed at 200 km. Experiments in the second group which is denoted as Localization are designed based on localization scales of ensemble error covariance for TLEn-Var, while weight of ensemble covariance is fixed at 0.5 (Table 2). Both of the first and second group include 3DVar as the control experiment. The two groups of experiments are designed respectively for TLEn-Var to examine the sensitivity of analyses and forecasts to different weight of flow-dependent covariance versus static covariance and localization scale, so that the best configuration can be selected as a setup for the subsequent quasi-operational system. In the third group, two configurations of ETKF-based hybrid assimilation experiments are also conducted, which contain 16 and 32 ensemble members respectively. The localization scale, covariance weight, and inflation factors for the ETKF-based hybrid experiments are based on the author’s previous study (Wang et al., 2016). Then the ETKF-based hybrid assimilation (ETKF-Var) are compared to the best configuration of the TLEn-Var to show to what degree the computational cost of the time-lagged method would be reduced relative to the traditional complicated methods and how would the precipitation forecast accuracy be downgraded in the alternative approach relative to the complex hybrid methods. This group of experiments cycled for 10 days from 0000 UTC 1 to 0000 UTC 11 July 2014. The assimilation frequency is 3 h, so totally 81 assimilation cases are tested.

The first forecast cycle of each experiment uses the deterministic 0.5° × 0.5° Global Forecast System (GFS) analysis from NCEP interpolated onto the model domain (Fig. 1) at 0000 UTC 1 July 2014 to create the initial conditions. The first analyses are at 0600 UTC 1 July 2014. The initial prior ensemble for the first ETKF ensemble is constructed by taking Gaussian random draws with zero mean and static

---

**Table 1**

<table>
<thead>
<tr>
<th>Experiment</th>
<th>HE1.0</th>
<th>HE0.75</th>
<th>HE0.5</th>
<th>HE0.25</th>
<th>3DVar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariance</td>
<td>1.0</td>
<td>0.75</td>
<td>0.5</td>
<td>0.25</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2**

<table>
<thead>
<tr>
<th>Experiment</th>
<th>HL300</th>
<th>HL200</th>
<th>HL100</th>
<th>HL50</th>
<th>3DVar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Localization</td>
<td>300 km</td>
<td>200 km</td>
<td>100 km</td>
<td>50 km</td>
<td></td>
</tr>
</tbody>
</table>
background error covariance (Barker, 2005) and adding them to the 20 km 0000 UTC 01 July 2014 background. In the following cycles, lateral boundary conditions are interpolated from the GFS forecasts every 3 h, while initial conditions are provided by analyses produced by the data assimilation experiments. The assimilation cycle with a 3 h period of each experiment continued until 0000 UTC 11 July 2014.

The static background error covariance for the variational system is estimated using the National Meteorological Center (known as NMC) method (Parrish and Derber, 1992), which uses differences between 24 h and 12 h forecasts valid at the same time (i.e., every 0000 and 1200 UTC) over the preceding month namely July 2014. The covariance matrix is formulated using option 5 (Barker, 2005) in WRFDA for a set of control variables that include stream function, unbalanced velocity potential, unbalanced temperature, pseudo relative humidity, and unbalanced surface pressure.

The data assimilation experiments make use of various types of meteorological observations, including conventional observations from radiosondes, pilot balloons, land surface stations, ships and buoys, aircrafts, and satellite products which are mostly satellite winds and derived temperature profiles. Data sorting, quality control, and observational error assignment for all experiments are performed through the observation preprocessing module of WRFDA.

5. Results and discussion

In this section, evaluations of single observation tests, initialization and subsequent forecasts are conducted. Diagnostics including computational cost, ensemble performance, increments analysis, verification against observations, precipitation forecast score, precipitable water and wind fields analysis are shown to reveal how the TLEn-Var contributes to the precipitation forecasts compared with 3DVar.

5.1. Computational cost

The main reason for using the time-lagged ensembles in the hybrid data assimilation is the reduced computational cost. Compared with 3DVar, the additional cost of traditional hybrid assimilation methods mainly comes from three steps: 1) the Kalman filter analysis; 2) the ensemble forecasts; 3) the computation of extended control variables in variational cost function. The first two steps take up most of the additional cost. However, the TLEn-Var avoids these two steps, and the

<table>
<thead>
<tr>
<th>Experiment</th>
<th>3DVar</th>
<th>TLEn-Var</th>
<th>ETKF-Var-16</th>
<th>ETKF-Var-32</th>
</tr>
</thead>
<tbody>
<tr>
<td>cost</td>
<td>71 min 33 s</td>
<td>91 min 53 s</td>
<td>652 min 49 s</td>
<td>1256 min 15 s</td>
</tr>
</tbody>
</table>

Fig. 4. The horizontal distribution of time-lagged ensemble spread (shaded contours) at 800 hPa at 0000 UTC 4 July 2014. a) Zonal wind (m/s); b) meridional wind (m/s); c) temperature (K); d) qvapor (g/g).
computational cost savings are very important for real-time implementations of the deterministic forecast and the pre-process for gridded GFS data are not included because all experiments are same in these two steps. We can see that the 3DVar only uses 71 min 33 s of the wall clock time. Compared to 3DVar, the TLEn-Var adds only about 20 min to the wall clock time because of the computation of extended control variables. The ETKF-based hybrid assimilation method adds the ensemble forecast run and the corresponding EnKF analysis, as a result ETKF-based method with 16 members uses 652 min 49 s of wall clock time, which is about 9 times that of 3DVar and 7 times that of TLEn-Var. The ETKF-based method with 32 members uses 1256 min 15 s of wall clock time, which is about 18 times that of 3DVar. Such computational cost savings are very important for real-time implementations of operational systems. Furthermore, the use of time-lagged ensemble also eases the burden of the vast storage for EnKF-related ensemble forecasts.

5.2. Ensemble performance

In this sub-section, the performance of the time-lagged ensemble is evaluated. In Fig. 4, the horizontal distribution of time-lagged ensemble spread for horizontal winds (U, V), temperature (T), and specific humidity (Q) at 800 hPa at 0000 UTC 4 July 2014 are shown. The spread distributes roughly along the rain band presented in Fig. 2a, indicating the flow-dependence of the time-lagged ensemble. In addition, the root-mean-square error (RMSE), ensemble spread of the time-lagged ensemble and its comparison with the ETKF ensemble with 32 members at each analysis time from 0000 UTC 3 to 0000 UTC 10 July 2014 are presented in Fig. 5. Here the initial perturbations for ETKF ensemble are made by taking Gaussian random draws with the WRFDA 3DVar system using static background error covariance. The ETKF system which has been built in WRFDA utilizes the same observation operator as 3DVar assimilation, and the same set of observations for each ETKF member are assimilated. The inflation scheme for ETKF which is described in Wang et al. (2007) is used. In Fig. 5, the X-axis presents analysis time, Y-axis presents values of ensemble spread and RMSE. a) Zonal wind (m/s); b) meridional wind (m/s); c) temperature (K); d) q vapor (g/kg).

Table 3 lists the wall clock time used by each configuration over 10-days period on a Linux workstation with 36 processors. The wall clock time of the deterministic forecast and the pre-process for gridded GFS data are not included because all experiments are same in these two steps. We can see that the 3DVar only uses 71 min 33 s of the wall clock time. Compared to 3DVar, the TLEn-Var adds only about 20 min to the wall clock time because of the computation of extended control variables. The ETKF-based hybrid assimilation method adds the ensemble forecast run and the corresponding EnKF analysis, as a result ETKF-based method with 16 members uses 652 min 49 s of wall clock time, which is about 9 times that of 3DVar and 7 times that of TLEn-Var. The ETKF-based method with 32 members uses 1256 min 15 s of wall clock time, which is about 18 times that of 3DVar. Such computational cost savings are very important for real-time implementations of operational systems. Furthermore, the use of time-lagged ensemble also eases the burden of the vast storage for EnKF-related ensemble forecasts.

**Fig. 5.** Spread and RMSE of the time-lagged ensemble against observations at each analysis time from 0000 UTC 3 to 0000 UTC 10 July 2014. The postfix "TL" means time-lagged ensemble and "ET" means ETKF ensemble. The X-axis presents analysis time, Y-axis presents value of spread and RMSE. a) Zonal wind (m/s); b) meridional wind (m/s); c) temperature (K); d) q vapor (g/kg).
Fig. 6. Temperature increments sensitivity to ensemble covariance weight and horizontal localization length scale, a) 3DVar; b) HE0.25; c) HE0.5; d) HE0.75; e) HE1.0; f) HL100; g) HL300. The solid black lines are contours of the background temperature at analysis time, and the shaded contours are the temperature increments, unit: K. "HE0.5" means the weight of time-lagged ensemble covariance is 0.5; "HL100" means the localization scale of time-lagged ensemble covariance is 100 km.
5.3. Single observation test

In this sub-section, several single observation tests which assimilate single ideal temperature observation are performed to examine the flow-dependent structure of hybrid background error covariance composed of static background error covariance and ensemble error covariance derived from time-lagged ensembles, as well as inner relationship between variables. For those tests, we placed a single temperature observation of 1 K with observation error of 1 K at ~800 hPa at the horizontal location (32°N, 108°E). The background forecasts and ensembles are taken from 0000 UTC 4 July 2014.

Fig. 6 shows the sensitivity of temperature increments in the single observation tests to ensemble covariance weight factors varying between 0 and 1, as well as horizontal localization length scales of ensemble covariance. For those tests, we placed a single temperature observation of 1 K with observation error of 1 K at ~800 hPa at the horizontal location (32°N, 108°E). The background forecasts and ensembles are taken from 0000 UTC 4 July 2014.

Fig. 6 shows the sensitivity of temperature increments in the single observation tests to ensemble covariance weight factors varying between 0 and 1, as well as horizontal localization length scales of ensemble covariance. For those tests, we placed a single temperature observation of 1 K with observation error of 1 K at ~800 hPa at the horizontal location (32°N, 108°E). The background forecasts and ensembles are taken from 0000 UTC 4 July 2014.

5.4. Verification against observations

In this sub-section, the RMSE of horizontal winds (U, V), temperature (T), and specific humidity (Q) are calculated between model fields and observations. We evaluate the impact of different covariance weights and localization scales in TLEn-Var on analyses and forecasts. In close or equal to 1.0, the horizontal structure of increments is stretched along the temperature contours, because of the impact of flow-dependent ensemble increments. As factor approaches 0, the horizontal structure of the increments lessens and the increments more closely resemble the stretched variational increments, showing characteristics of isotropy and homogeneity. Fig. 6 also displays a reduction in the lateral fine structure as scale decreases from 300 km to 100 km. It is obvious that the smaller horizontal localization length scale restricts the increments to a local area around the position of single observation.

5.4. Verification against observations

In this sub-section, the RMSE of horizontal winds (U, V), temperature (T), and specific humidity (Q) are calculated between model fields and observations. We evaluate the impact of different covariance weights and localization scales in TLEn-Var on analyses and forecasts. In
addition, the optimal configuration of the TLEn-Var is compared to the ETKF-based hybrid assimilation over a 10-days period. Fig. 7a–d displays the vertical profiles of mean RMSE between the analyses of different covariance weights and the observations for U, V, T, Q in TLEn-Var experiments. These differences measure the extent to which the analyses fit observations that have been assimilated at each verifying time. It shows that the experiment HE0.5 fits the wind observations the closest, followed by HE0.75 and HE0.25, and the experiments HE1.0 and 3DVar yield the largest differences between the wind observations throughout the vertical domain. However, the experiment 3DVar fits the temperature and humidity observations as well as HE0.5 and HE0.25, followed by HE0.75, while the HE1.0 fits the farthest. We then examine the vertical distribution of the mean RMSE for 24 h forecasts (Fig. 7e–h). For wind fields, the experiments HE0.75, HE0.5 and HE0.25 perform similar, but has noticeably smaller RMSE than experiments HE1.0 and 3DVar. Fig. 7i–l shows the vertical distribution of the mean RMSE for 48 h forecasts. It shows that the hybrid experiments have substantially smaller RMSE than the control experiment 3DVar throughout the troposphere. The 48 h forecasts of the hybrid experiments have larger superiority on 3DVar than 24 h forecast, indicating that compared with purely static covariance, the hybrid covariance has the advantage of holding an positive impact on longer leading time. Fig. 8a–d displays the vertical profiles of mean RMSE between the analyses of different localization scales and the observations. It shows all the hybrid experiments fit the wind observations closer than control experiment 3DVar, among which the experiments HL50 and HL100 fit observations the closest. The experiment 3DVar fits the temperature and humidity observations the best, especially for near-surface, followed by experiment HL100. The experiment HL300 fits the farthest, although results of all the experiments are very close. Fig. 8e–l shows the vertical distribution of mean RMSE of 24 h and 48 h forecasts respectively of different localization scale experiments. For wind field, the experiment HL200 perform the best, followed by experiments HL100 and HL300. The experiments HL50 and 3DVar get the biggest RMSE. Similar results for temperature and the humidity are obtained.
In this study, 3DVar seems to fit the observations the closest for temperature and humidity. This may be because the 3DVar has an overestimated climatologic background error variance for temperature and humidity, and the ensemble method usually under-estimated the background error variance because of the limited ensemble size. As a result, the 3DVar applies a larger weight to the observations and fits the observations closer. However, the ensemble covariance derived from these time-lagged differences has a more significant advantage of holding a positive impact on longer leading times than 3DVar. This phenomenon results from two competing goals in data assimilation, namely the achievement of a physically consistent state and having a state close to the observations (Bick et al., 2016). Furthermore, it is generally undesirable to closely fit observations during data assimilation, since it may cause valuable information provided by the forecast model to be disregarded and imbalance of the analyses. Overweighting observations may also over-fit observation noise, rather than the signals (Zhang et al., 2013). Our study suggests that the TLEn-Var, although produces less or comparable accurate analyses than 3DVar for temperature and humidity, leads to more balanced model states which eventually leads to better forecasts for longer leading times.

In the first two groups of experiments, a weight factor of ensemble covariance at 0.5 and a distance of localization at 100 km is considered to be the optimal configuration. Then this optimal configuration of the TLEn-Var is compared to the ETKF-based hybrid assimilation over 10-days period to show how would the analysis and forecast accuracy perform in the TLEn-Var relative to the complex hybrid method. Fig. 9 shows the vertical profiles of mean RMSE of the analyses and forecasts against the observations. It is presented that the analyses of time-lagged hybrid experiment fit the observations closer for all variables (U, V, T, Q) than ETKF-based hybrid assimilation. This is because the time-lagged ensemble has bigger spread which has been shown in Sub-section 5.2. However, for the 24 h forecasts, the ETKF-based method with 32 members is slightly better than the time-lagged method, while the time-lagged method performs comparable to the ETKF-based method with 16 members. For the 48 h forecasts, the three experiments tend to have the similar results.

Fig. 9. Vertical profiles for TLEn-Var and ETKF-based hybrid with different ensemble members of the a–d) averaged analysis RMSE against observations; e–h) averaged 24 h forecast RMSE against observations; i–l) averaged 48 h forecast RMSE against observations. The numbers of observations are shown on the right of each panel.
5.5. Precipitation score

Fractions Skill Score (FSS) is one of the neighborhood verification approaches which is calculated to evaluate the precipitation forecast skill (Roberts and Lean, 2008). The FSS is defined as

$$ FSS = 1 - \frac{1}{N} \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{P_{i,j} - P_{o,i,j}}{P_{i,j} + P_{o,i,j}} \right)^2 \right] $$

where $N$ is the number of neighborhoods; $P_{i,j}$ is the proportion of grid boxes within a forecast neighborhood where the prescribed threshold was exceeded (i.e., the proportion of grid boxes that have forecast events); and $P_{o,i,j}$ is the proportion of grid boxes within an observed neighborhood where the prescribed threshold is exceeded (i.e., the proportion of grid boxes that have observed events). In the formula the denominator represents the worst possible forecast (i.e., with no overlap between forecast and observed events). FSS ranges between 0 and 1, with 0 representing no overlap and 1 representing complete overlap between forecast and observed events, respectively. The observed precipitation products is the China Hourly Merged Precipitation Analysis (CHMPA) from China Meteorological Administration.

The FSS with a threshold of 12.5 mm/6 h in the TLEn-Var experiments with different weighted coefficients and 3DVar is shown in Fig. 10a. In general, the scores of the experiments HE0.5 and HE1.0 are greater than others for all 24 h forecast times. When the forecast time is between 36 and 48 h, the FSS values of experiment HE0.75 is clearly greater than the other experiments. Fig. 10b presents the FSS of 3DVar and the TLEn-Var experiments with different horizontal localized scales. The FSS of experiment HL100 is better than others in 24 h forecast time, and the FSS of experiment HL200 is the best between 36 and 48 h. Fig. 10c,d shows the variety of FSS along thresholds for 24 h accumulation precipitation of different experiments. It is clear in Fig. 10c that the FSS scores in the TLEn-Var experiments with different weighted coefficients are greater than that of 3DVar, especially between threshold 20 mm and 45 mm. Fig. 10d shows 3DVar and TLEn-Var experiments with different horizontal localized scales, indicating that all the experiments have little difference when threshold is less than 20 mm; all

Fig. 10. The averaged FSS with a threshold of 12.5 mm/6 h along the forecast time. (a) The 3DVar and TLEn-Var experiments with different weighted coefficients, (b) the 3DVar and the TLEn-Var experiments with different localized length scales. And the averaged FSS of 24 h accumulated precipitation along the thresholds. (c) The 3DVar and TLEn-Var experiments with different weighted coefficients, (d) the 3DVar and TLEn-Var experiments with different localized length scales. The horizontal axis is the forecast time, the vertical coordinates are the averaged value of FSS. “HE0.5” means the weight of time-lagged ensemble covariance is 0.5; “HL100” means the localization scale of time-lagged ensemble covariance is 100 km.

Fig. 11. a. The averaged FSS with a threshold of 12.5 mm/6 h along the forecast time; b. The averaged FSS of 24 h accumulated precipitation along the thresholds.
configurations of TLEn-Var experiments are better than 3DVar between 20 mm and 40 mm; when threshold is bigger than 35 mm, HL100 experiment perform the best.

To show how would precipitation accuracy be downgraded in the TLEn-Var relative to the complex hybrid methods, the best configuration of the TLEn-Var for precipitation forecast is also compared to the ETKF-based hybrid assimilation. Fig. 11a shows the FSS with a threshold of 12.5 mm/6 h along the forecast time for time-lagged method, ETKF-based method with 32 members and 16 members respectively. It is presented that the ETKF-based method with 32 members is better than that with 16 members and the time-lagged method, and the latter two method perform comparable. The FSS of 24 h accumulated precipitation along the thresholds is also presented in Fig. 11b, which shows the similar results as Fig. 11a.

5.6. Precipitation field

The best configuration of the TLEn-Var system in FSS verification, in which the weight of ensemble covariance is 0.5 and the localization scale is 100Km, is selected to be analyzed and compared with 3DVar. Fig. 12a shows the averaged 24 h accumulated precipitation from CHMPA (observed precipitation). The observed heavy rainfall is distributed roughly from southwest to the northeast across the pattern. It is showed that the maximum center is located near (31°N, 117°E). Fig. 12b,c shows the simulated averaged 24 h accumulated precipitation from 3DVar and time-lagged experiments initialized at all 17 analysis times when the heavy rainfall event occurred. It can be seen in Fig. 12b that the accumulated precipitation amount in 3DVar is greatly lower than CHMPA, missing the intensity and coverage of observed precipitation areas. In comparison, the TLEn-Var shows much better intensity and coverage in precipitation forecasts. Fig. 12c shows that heavy precipitation in the maximum center is well captured although the hybrid experiment tends to under-predict the magnitudes of precipitation at the observed maximum regions too.

5.7. Precipitable water and wind

To better understand how the TLEn-Var improve the precipitation simulation, additional variables such as precipitable water and wind are analyzed in this sub-section. The averaged analyses and forecasts during the heavy rainfall event are respectively compared to the final analysis of NCEP Global Forecast System namely FNL, which assimilates almost all the available observations including satellite radiances.

Precipitable water plays a critical role in the maintenance of a lasting heavy rainfall. The precipitable water at the analysis time and the 12 h and 24 h forecast times are presented in Fig. 13 The 3DVar shows an obvious dry bias in the region that is close to the rainfall center (Fig. 13a–e), which explains the weaker model-simulated precipitation in 3DVar. Compared to 3DVar, TLEn-Var (Fig. 13b–f) significantly increases the amount of precipitable water in the dry bias area which is beneficial to formation of the precipitation.

Besides precipitable water condition, the wind condition is also one of the main factors that has significant impact on the precipitation forecast. We further focus on the characteristics of the horizontal and vertical wind fields. The difference of the averaged horizontal wind fields during the heavy rainfall event at 800 hPa are presented in Fig. 14. Compared with the FNL, 3DVar shows a decrease of wind speed in the rainfall area (Fig. 14a), leading to the weaker model simulated precipitation. Fig. 14b presents the difference of averaged wind analysis between 3DVar and TLEn-Var. It shows that the latter enhances the wind speed in the rainfall area. It is clear in Fig. 14b that TLEn-Var generates cyclonic
wind increments near the rainfall area, in favor of the convergence and rise of moisture field as well as the formation of precipitation. The 12 h and 24 h wind forecasts differences (Fig. 14c–f) show that the TLEn-Var promotes wind convergence to maintain a favorable environment for precipitation.

The cross sections of the 12 h and 24 h forecast for vertical velocity along 117°E, are shown in Fig. 15. It is presented that the vertical velocity of 3DVar is weaker in the rainfall center, which leads to the weaker model-simulated precipitation. However, the TLEn-Var (Fig. 15b, d) strengthens the uplift vertical velocity which partly contributes to the improvement of precipitation simulation.

6. Summary

This study evaluates the performance of 3DVar and a hybrid system using time-lagged ensembles for a case study which involves heavy rainfall. The time-lagged ensemble is constructed by sampling from a 3 h moving window along a model trajectory. It is economical and simple to implement, and has the ability to adaptively estimate the spatial distribution of background error.

Single observation tests are performed to document some characteristics of the TLEn-Var system. It is showed that as the weight of ensemble error covariance approaches 1.0, the horizontal structure lessened.
and the increments tended toward a stretched version of the variational increments, and smaller horizontal localization length scale restricts the increments to a local area around the position of single observation. The TLEn-Var also provides estimates of the flow-dependent cross-field covariance between observed and unobserved model fields that can be used to update unobserved variables.

In real observations experiments, the impact of different covariance weights and localization scales on analyses and forecasts in TLEn-Var are investigated. A weight of ensemble covariance at 0.5 and a distance of localization at 100 km is considered to be the optimal configuration in this study. The TLEn-Var is slightly worse than ETKF-based method with 32 members and performs comparable to that with 16 members. Compared with 3DVar, the TLEn-Var improves the simulation of vapor flux, precipitable water, structure of the wind in model. Obvious cyclonic wind increments are generated near the rainfall center, which is advantageous for convergence uplift and formation of precipitation. Precipitation forecast skill is improved by the TLEn-Var.

In general, the hybrid data assimilation using time-lagged ensembles significantly reduces computational cost compared with traditional methods. The forecast skill is improved without significantly increasing
computational cost compared with 3DVar. It also eases the burden of the vast storage of EnKF-related ensemble forecasts. In complex models where running an EnKF requires that the ensemble size or model resolution is severely limited, the time-lagged ensembles seem like a viable and important alternative or supplement. Furthermore, the saving cost can be used to increase the forecast resolution, leading time or coverage domain.

This work may does not represent a major advancement in data assimilation, and the method applied in this paper may be not up to quality with traditional ensemble method like EnKF when the computational resource is sufficient, but the general idea of using forecast differences of time-lagged forecasts verified at the same leading time needs to be investigated because this method is economical and simple to implement. It is not proposed to take the place of the advanced assimilation methods, but to be viewed as a alternative or supplement to those who can not afford the expensive cost of the complicated methods. In the long term, we believe the EnKF-based methods, even much more complicated methods such as the particle filter, four-dimensional ensemble filter, and four-dimensional ensemble-variation will be the optimal choices. However, these complicated methods are still under development or have only be applied in national-level weather service agencies for the moment. At a lower-level such as prefecture level or even some provincial level, the application of these complicated methods will be a big challenge, which is a general problem in most of China or other developing countries. Therefore, in this situation, the method we proposed and tested in this study may supply as a compromising choice during the period of transition. In addition, the TLEn-Var can be also viewed as an evidence to show how hybrid data assimilation systems, in whatever form, compare to traditional variational systems.

Although the time-lagged ensemble used in hybrid data assimilation is significantly economical and easy to implement, the inflation scheme and physical perturbation can not be easily applied in this method because there are no practical ensemble filter and integration steps in the time-lagged ensembles. Furthermore, the resulting error covariance may be dominated by the instances furthest away from the center of the time window. To prevent this from occurring and improve the assimilation performance, a preprocessing procedure of the time-lagged ensemble may be needed. For example, the lagged state instances can be first high-pass filtered and then resampled to remove the remaining sequential ordering information (Keppenne et al., 2014). Furthermore, one can also use time-lagged ensemble to boost the ensemble size in an EnKF framework, that is several time-lagged forecasts can be run concurrently and the resulting time-lagged ensembles combined into a traditional ensemble. These will be further investigated in our future work.

**Acknowledgements**

This work is jointly sponsored by the National Basic Research Program of China (2013CB430102), the Special Fund for Meteorological Scientific Research in Public Interest (GYHY201506002), National Natural Science Foundation of China (41430427, 41205082, 41675102), Key University Science Research Project of Jiangsu Province (11KJA170001) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).