atmosphere

MDPI

Article

Efficient Regional Hybrid Ensemble-Variational Data Assimilation using the Global-Ensemble-Model-Augmented Error Covariance for Numerical Weather Prediction over Eastern China

Yuanbing Wang, Yaodeng Chen * and Jinzhong Min

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; wyb@nuist.edu.cn (Y.W.); minjz@nuist.edu.cn (J.M.)

* Correspondence: keyu@nuist.edu.cn

Received: 29 February 2020; Accepted: 7 April 2020; Published: 9 April 2020

Abstract: An efficient regional hybrid ensemble-variational (EnVar) data assimilation method using the global-ensemble-model-augmented error covariance is proposed and preliminarily tested in this study. This method uses the global ensemble error covariance as the complementary low-resolution regional ensemble error covariance. The high-resolution dynamic ensemble mean is used as the first guess in hybrid EnVar and then re-centered to the updated high-resolution dynamic ensemble perturbations after minimization analysis. In this study, the proposed method is implemented into the Weather Research and Forecasting Model's (WRF) data assimilation system coupled with the ensemble transform Kalman filter (ETKF) and preliminarily tested for numerical weather prediction during the Mei-Yu season over eastern China. It is found that the experiment containing fewer regional dynamic ensemble members but augmented with global ensemble error covariance obtains similar results to the experiment containing many more regional dynamic ensemble members. However, the former experiment only takes up one third of the latter experiment's computational cost. The method proposed in this study also outperforms the 3DVar, hybrid EnVar using the pure global ensemble error covariance, as well as the hybrid EnVar using regional ETKF ensemble with a smaller size. The method proposed in this paper effectively combines the contributions of the ensemble error covariance from both the global and the regional models to produce better initial conditions for the regional WRF data assimilation system.

Keywords: hybrid EnVar; background error covariance; global-ensemble-model-augmented error covariance; WRFDA

1. Introduction

Data assimilation aims to optimally combine the model background with observations, then produce initial conditions for numeric weather forecast (NWP). Modern data assimilation includes variational and ensemble methods [1]. The variational method updates the first guess using the static background error covariance, which is full rank but usually derived with assumptions of isotropy and homogeneity in space and time [2–4]. The ensemble method with flow-dependent background error covariance is fully nonlinear because of the cross-variable covariance but suffers the sample error problem. Thus, a hybrid ensemble-variational (EnVar) method that combines two kinds of background error covariance in a variational framework has been proven to be significantly beneficial to NWP [5,6]. On one hand, there have been many studies of different hybrid schemes in global NWP [7–9]. On the

other hand, it has been also demonstrated that the hybrid EnVar method is suitable for regional models [10–13].

The ensemble error covariance is usually severely rank deficient due to the computational limitations that cause a much smaller ensemble size than model state vector [14–17]. To compensate for the rank deficiency problem, several studies focused on ensemble sampling methods that increase ensemble size without significantly increasing the computational cost have been conducted. It was found that the time-expanded or time-lagged sampling methods can be used to introduce flow-dependent error covariance by reducing the number of integration runs needed to produce ensembles with the desired sample size [18–21]. Kretschmer et al. (2015) improved the performance of an ensemble Kalman filter by adding a collection of 'climatological' perturbations to the forecast ensemble mean to increase the size of the ensemble at analysis time [22]. Some researchers also replaced the high-resolution ensemble with a low-resolution one to reduce the computational cost. For example, Gao and Xue (2008) proposed EnKF with a single high-resolution forecast and a low-resolution ensemble [23], and then the dual-resolution ensemble technique was extended to the hybrid EnVar framework [13,24]. The background error covariance estimated from the low-resolution ensemble is introduced to update the deterministic high-resolution forecast, which is used as the first guess. The ensemble analysis mean is then replaced by the high-resolution analysis after the ensemble perturbations are updated separately. Wu et al. (2017) applied the global ensemble forecasts from the EnKF component of the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) in a regional hybrid EnVar data assimilation system [25]. They found that the global ensemble error covariance has the dynamical consistency to be used in the regional model. Thus, the regional system completely avoids running ensemble forecasts itself and improved the forecast compared to the pure 3DVar. A similar method was later used in the Regional Deterministic Prediction System in Environment Canada [8]. However, although evolving the ensemble at a lower resolution significantly reduces the computational cost, it may decrease the accuracy of the analysis. In addition, most ensemble data assimilation develops and advances algorithms for ensembles with all members having the same resolution and using the same model. Rainwater and Hunt (2013) presented a mixed-resolution local ensemble transform Kalman filter which takes its background error covariance from a linear combination of a low-resolution ensemble and a high-resolution ensemble without the computational cost of running the entire ensemble at a high resolution [26], which was similar to the hybrid EnVar method except that they used the sample covariance of the mixed-resolution ensembles instead of using a static climatological background error covariance.

Inspired by Rainwater and Hunt (2013) [26], Kretschmer et al. (2015) [22], and Wu et al. (2017) [25], an efficient regional hybrid EnVar data assimilation method using the global-ensemble-model-augmented error covariance was proposed and preliminarily tested in this study. This work used the global ensemble error covariance as the low-resolution one, extending Rainwater and Hunt (2013) [26] from the EnKF framework to the hybrid EnVar framework. Similarly to Kretschmer et al. (2015) [22], our work used the high-resolution dynamic ensemble forecast mean as the first guess in the hybrid EnVar data assimilation and then re-centered the analysis to the updated high-resolution dynamic ensemble perturbations. In addition, differently from the work of Kretschmer et al. (2015) [22] that used the climatological perturbations derived from the static background error covariance to augment the regional dynamic ensemble error covariance in an EnKF framework, this study applied the global EnKF ensemble error covariance as the augment, which has been proven to have a similar dynamic consistence to the regional ensemble error covariance [25]. Furthermore, this study is mainly focused on the hybrid EnVar framework. While the global ensemble used in their study has a resolution of approximately 36 km, our regional analysis resolution was 12 km. The global low-resolution part of the mixed-resolution method was functionally similar to the dual-resolution hybrid EnVar data assimilation.

The mixed-resolution data assimilation can be used to combine the information in a small, high-resolution ensemble with a large, low-resolution global ensemble, which can produce a better

analysis than either resolution produces by itself [26]. In this study, besides the regional ensembles, the flow dependent error covariance contributed from the global ensembles will be applied to help produce better initial conditions for the regional data assimilation system, with the assumption that they can provide useful information to improve the large-scale components of the regional background error covariance. Meanwhile, both the ensemble error covariance and the ensemble mean from the regional ensemble will be used in the regional model to help define the smaller-scale part of the background error covariance.

In this study, we implemented the proposed method into the Weather Research and Forecasting Model's Data Assimilation (WRFDA) system coupled with the ensemble transform Kalman filter (ETKF) scheme and tested it for numerical weather prediction over eastern China using the regional model Weather Research and Forecasting Model (WRF). The rest of the paper is organized as follows. In Section 2, the methodology of the proposed method and the global ensemble used in this study are introduced. Section 3 details the experimental set-up and design. Section 4 presents the results of the experiments. Finally, the conclusion and discussion are provided in Section 5.

2. Methodology

2.1. The Hybrid EnVar and ETKF Schemes

The formula of the hybrid EnVar built in WRFDA is written as the following [10]:

$$J(\delta \mathbf{x}_1, \boldsymbol{\alpha}) = \frac{1}{\beta} \cdot \frac{1}{2} \delta \mathbf{x}_1^T \mathbf{B}^{-1} \delta \mathbf{x}_1 + \frac{1}{1-\beta} \cdot \frac{1}{2} \boldsymbol{\alpha}^T \mathbf{A}^{-1} \boldsymbol{\alpha} + \frac{1}{2} (\mathbf{d} - \mathbf{H}\delta \mathbf{x})^T \mathbf{R}^{-1} (\mathbf{d} - \mathbf{H}\delta \mathbf{x}) \quad (1)$$

where $\delta \boldsymbol{x}_1$ is the analysis increment associated with the static background error covariance. The second term is associated with the ensemble background error covariance. $\boldsymbol{\alpha}$ is the ensemble extended control variable. $\boldsymbol{A}$ defines the spatial covariance of $\boldsymbol{\alpha}$. $\boldsymbol{d} = y^o - H\boldsymbol{x}_b$ is the innovation. It is noted that the analysis increment of the hybrid EnVar is the sum of two terms, defined as

$$\delta \boldsymbol{x} = \delta \boldsymbol{x}_1 + \sum_{n=1}^{N} \left(\boldsymbol{\alpha}_n \circ \boldsymbol{x}_{n,b}^e \right) \quad (2)$$

where the second term is the analysis increment associated with the flow-dependent background error covariance and the symbol $\circ$ is an element-wise multiplication or Schur product. N is the ensemble size. $\mathbf{x}_{n,b}^e$ is the n_{th} ensemble perturbation normalized by $\sqrt{N-1}$:

$$\mathbf{x}_{n,b}^e = (\mathbf{x}_{n,b} - \overline{\mathbf{x}_b}) / \sqrt{N-1} \quad (3)$$

in which $\mathbf{x}_{n,b}$ is the n_{th} ensemble member and $\overline{\mathbf{x}_b}$ is the dynamic ensemble mean.

The WRFDA system also includes an ETKF scheme, which is used to generate the ensemble perturbations [10]. The equation is as follows:

$$\boldsymbol{x} = \boldsymbol{x}^e \Pi \boldsymbol{C} (\boldsymbol{\Gamma} + \boldsymbol{I})^{-1/2} \boldsymbol{C}^T \quad (4)$$

where $\mathbf{C}$ and $\boldsymbol{\Gamma}$ are the eigenvector matrix and the eigenvalue matrix of the $(HX^e)^T R^{-1} HX^e$, respectively. $\mathbf{I}$ is the identity matrix. $\Pi = \sqrt{c_1 c_2 \ldots c_i}$ is the inflation factors, and c_i satisfies the equation

$$\widetilde{d}_i^T \widetilde{d}_i = Tr\left(\boldsymbol{R}^{-1} \boldsymbol{H} c_i \boldsymbol{P}_i^e \boldsymbol{H}^T + I \right) \quad (5)$$

where $\widetilde{d}_i = R^{-1/2} y_i - H\overline{X}_i^b$ is the "innovation" normalized by observation error covariance matrix $\boldsymbol{R}$. $\boldsymbol{P}_i^e$ is the ensemble background error covariance matrix. Tr represents the trace of a matrix.

2.2. The Globally Augmented Regional Hybrid EnVar Method

Inspired by Equations (13) and (14) in Rainwater and Hunt (2013), we formulated the globally augmented hybrid EnVar method proposed in this study as the following:

$$\begin{aligned} J(\delta\mathbf{x}_1, \boldsymbol{\alpha}) = \tfrac{1}{\beta}\tfrac{1}{2}\delta\mathbf{x}_1^T\mathbf{B}^{-1}\delta\mathbf{x}_1 + \tfrac{1}{1-\beta}\tfrac{1}{2} < \boldsymbol{\alpha}_r, \boldsymbol{\alpha}_g >^T \mathbf{A}^{-1} < \boldsymbol{\alpha}_r, \boldsymbol{\alpha}_g > \\ + \tfrac{1}{2}(\mathbf{d} - \mathbf{H}\delta\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\delta\mathbf{x}) \end{aligned} \tag{6}$$

where $\boldsymbol{\alpha}_r$ and $\boldsymbol{\alpha}_g$ are the regional and global augmented ensemble error covariance, respectively. Specifically, the total analysis increments in the proposed method are modified to the following equation:

$$\delta\mathbf{x} = \delta\mathbf{x}_1 + \sum_{n=1}^{N_r}(\boldsymbol{\alpha}_{n,r} \circ \mathbf{x}_{n,r}^e) + \mathbf{L}\sum_{n=1}^{N_g}(\boldsymbol{\alpha}_{n,g} \circ \mathbf{x}_{n,g}^e) \tag{7}$$

Here, N_r represents the number of regional ensembles and N_g represents the number of global ensembles. $\boldsymbol{\alpha}_r$ and $\boldsymbol{\alpha}_g$ are the ensemble extended control variable vectors for regional and global applications respectively. $\mathbf{x}_{n,r}^e$ and $\mathbf{x}_{n,g}^e$ are ensemble perturbations calculated using the regional ensemble and global ensemble, respectively. $\mathbf{L}$ is a transformation operator that maps the global ensemble error covariance from the spectral space of global model (NCEP GFS in this study) to the grid space of regional model (WRF in this study). Thus, the proposed method will obtain benefits from increased ensemble size, without correspondingly increasing the number of forecasts carried out.

A complete regional hybrid EnVar data assimilation system with global-ensemble-model-augmented error covariance (Figure 1) can be described at the practical level as follows: (1) the high-resolution regional ensemble is first created and evolved to the analysis time; (2) the global ensemble forecasts valid at the analysis time are also obtained and processed for the regional model (as will be described in next subsection); (3) the high-resolution regional ensemble mean is then calculated and updated using the global-ensemble-model-augmented error covariance according to Equations (4) and (5); (4) meanwhile, the regional high-resolution ensemble forecasts are also updated using the ETKF scheme; (5) the dynamic analysis ensemble members updated by the regional ETKF are then re-centered with the hybrid EnVar analysis and evolved to the next analysis time; and (6) the cycle is repeated.

Although this method is proposed to compensate for the rank deficient problems of the ensemble error covariance by increasing the ensemble members at a low computational cost, it has some potential added benefits for regional data assimilation. Since the global ensemble forecasts are initialized from the global ensemble data assimilation system which assimilates all of the available observations—including the satellite radiance covering most of the earth, especially the ocean area—and has been best tuned, the global ensemble error covariance includes much more accurate large-scale information but lacks mesoscale information. On the contrary, the regional ensemble error covariance has the mesoscale information that is important for the prediction of high impact weather, but the local error may increase rapidly during the cycling run. The combination of the global and regional ensemble error covariance not only reduces the computational cost required by the ensemble forecasts, but also increases the degrees of freedom of the ensemble error covariance and introduces more accurate large-scale error information that can better constrain the regional data assimilation. In addition, an ensemble composed of forecasts evolved at different resolutions and even with different models may better characterize forecast uncertainty than a single resolution ensemble and single model ensemble within given computation constraints [26].

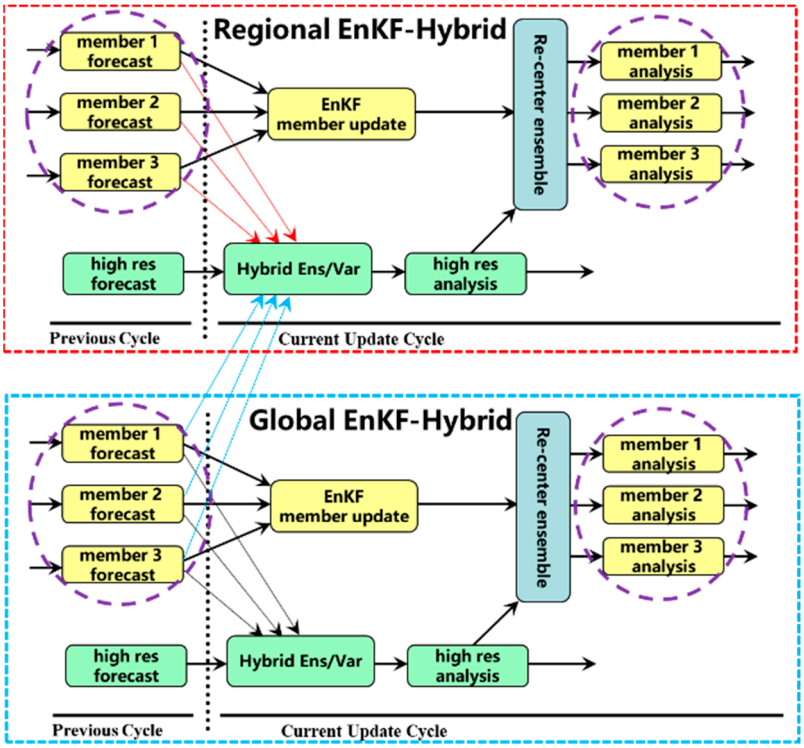

Figure 1. A schematic diagram of the regional hybrid data assimilation using global-ensemble-model -augmented error covariance.

2.3. The NCEP Global Ensemble and Its Process for WRFDA

In this study, we used the operational NCEP Global Ensemble Forecast System (GEFS) data generated from the Global Data Assimilation System (GDAS) to produce the augmented ensemble error covariance for the regional hybrid EnVar data assimilation. The NCEP eighty-member T574 global EnKF ensemble is of about 36 km in grid spacing, with a vertical resolution of 64 levels [27,28]. The ensemble is updated using a "state of art" EnKF system four times a day (every 6 h), which assimilates all available observations including satellite radiance and has been best tuned. Then, the analysis of the 80 EnKF ensemble members from the previous cycle is re-centered by a global hybrid EnVar analysis and used to initialize the global ensemble perturbations.

The extended control variables in this study for the hybrid EnVar of WRFDA system include the horizontal wind components, potential temperature, specific humidity, and surface pressure. Thus, only the atmospheric component of the GEFS output is needed. Since the NCEP GFS model and the WRF model use different data formats, a transformation operator is used to map the ensemble perturbations in spectral space to grid space.

3. Experiment Designs and Configurations

In this study, version 3.9 of the Advanced Research WRF [29] (Skamarock et al., 2008) was used to produce a numerical weather forecast over a computational domain spanning the eastern China area (Figure 2). The horizontal grid spacing is 12 km (225 × 225 grid points). The domain is configured with 45 vertical levels and a 50 hPa model top. The physical parameterizations used in this study include the Morrison double-moment microphysics scheme, the Rapid Radiative Transfer Model for Global Climate Models longwave and shortwave radiation schemes with aerosol and ozone climatology, the Mellor–Yamada–Janjic planetary boundary layer scheme, the Noah land surface model, and the Tiedtke cumulus parameterization.

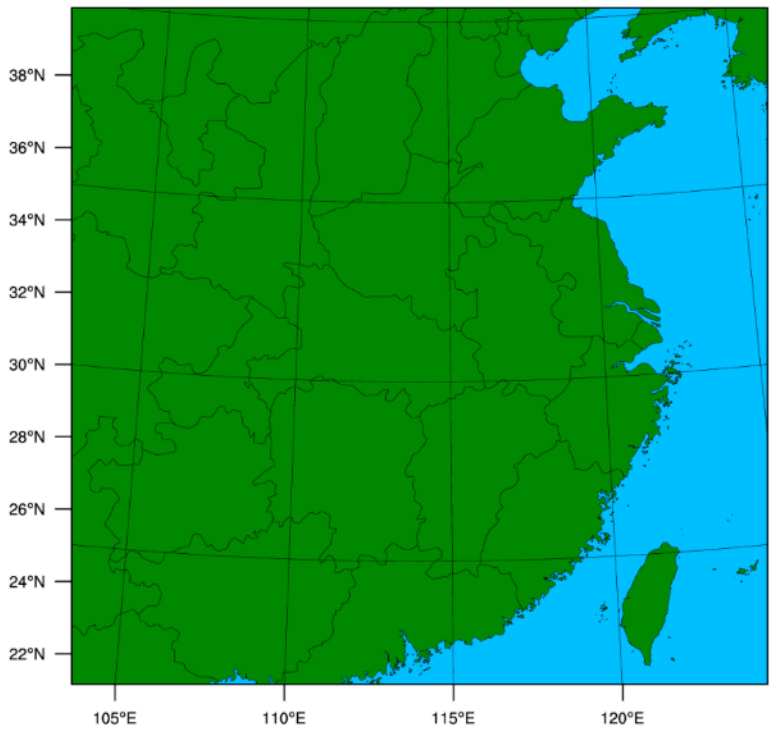

Figure 2. The experimental domain.

Five parallel experiments employing different methodologies were conducted. The first experiment used the basic 3DVar method with the static background error covariance to better show the advantage of hybrid EnVar in the different forms presented in this study (hereafter "3DVar"). The second experiment used the hybrid EnVar method with the flow-dependent background error covariance coming from the 80-memeber global ensemble forecasts only (hereafter "GE-HDA"). The third experiment used the regional hybrid EnVar with the flow-dependent covariance contributed from the regional ETKF ensemble forecasts using 20 ensemble members (hereafter "RE20-HDA"). The fourth experiment used the method proposed in this study, which combined the 20-member regional ensemble and the 80-member global ensemble (hereafter "GE/RE20-HDA"). Finally, as a reference, the fifth experiment used the regional ETKF to provide the flow-dependent ensemble error covariance but with 60 dynamic ensemble forecasts for the hybrid EnVar data assimilation system (hereafter "RE60-HDA").

To create the initial conditions in the first forecast cycle, the NCEP GFS analysis data were interpolated onto the 12 km experimental domain (Figure 2) at 0000 UTC 20 June 2017. In the following cycles, the initial conditions were provided by analyses generated by the data assimilation experiments; the lateral boundary conditions were provided by the 6-hourly GFS analysis. The initial prior ensemble members for the first ETKF analysis were 6 h WRF forecasts initialized at 0000 UTC 20 June 2017 using the so-called "random_CV" method, by adding random noises to the 3DVar analysis in control variable space [30] to develop a flow-dependent structure of background error covariance. The first analyses occurred at 0600 UTC 20 June 2017 using the previous 6 h forecasts as backgrounds. The data assimilation cycle with a 6 h interval for each experiment continued until 1200 UTC 30 June 2017 (Figure 3).

The static background error covariance used in this study was constructed by the so-called NMC (National Meteorology Center) method [2], which takes the differences between 12 h forecasts and 24 h forecasts valid at the same times averaged over at least a month to compute the static background error covariance. In this study, covariance localization was applied to suppress the impact of ensemble error covariance on the analysis increments and reduce the spurious correlations due to sampling error in ensemble-based data assimilation. For the regional ETKF scheme, an adaptive inflation method was also applied to the posterior analysis perturbations [10].

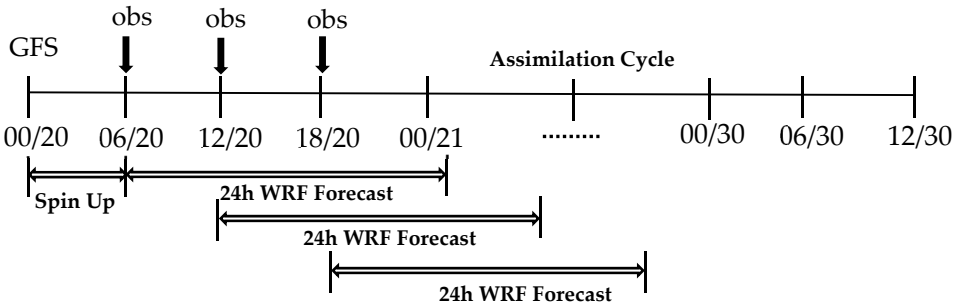

Figure 3. A flow chart of the cycling data assimilation experiment.

A variety of upper-air and surface observations were assimilated in this study (Figure 4). Radiosonde observations of temperature, pressure, specific humidity, and wind were assimilated as well as aircraft reports (AIREP) of temperature and wind. Besides, satellite-derived wind (SATOP), surface observations from surface synoptic observation (SYNOP), and aviation routine weather report (METAR) platforms were also assimilated. Observations were taken within a 1.5 h data assimilation window for each analysis, and all observations were assumed to be valid at the analysis time. Data sorting, quality control, and observational error assignment for each experiment were performed through the observation preprocessing module (i. e. OBSPROC) of WRFDA. The observations of innovations that exceeded five times the observation error were rejected before the minimization iterations.

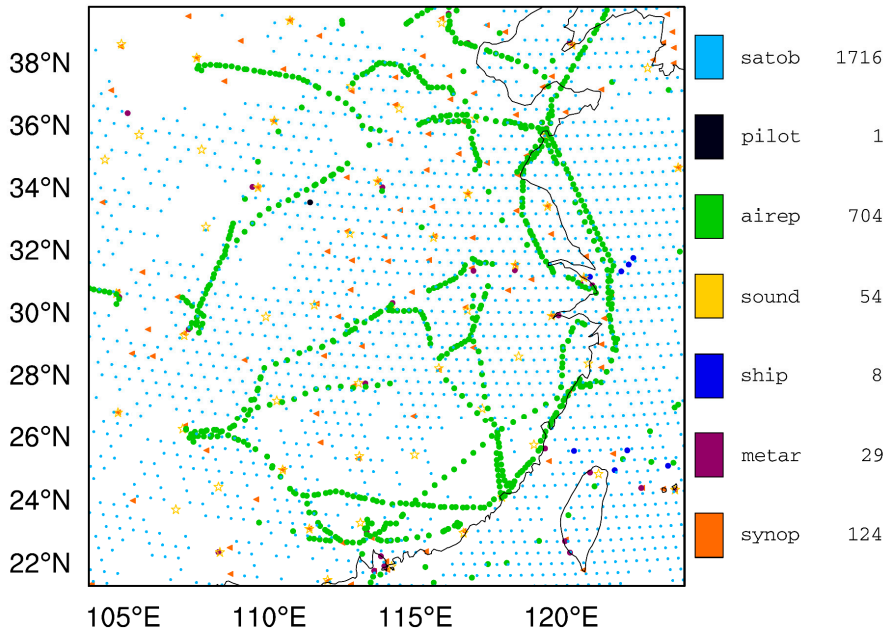

Figure 4. Conventional observations valid at 0000 UTC 23 June 2017.

4. Results

4.1. Single Observation Test

The single observation test can help in understanding the working principles of different data assimilation schemes; this is because their analysis increments can reflect the dynamic structure of the background error covariance used in the data assimilation. In this subsection, the analysis increments of the single observation tests for different experiments were investigated. A single observation of specific humidity was assumed to be located at the center of the model domain at the 850 hPa level at 0000 UTC 23 June 2017. The innovation (observation minus background) of the water vapor mixing ratio was 1 kg/kg. The observation error was set to 0.001 kg/kg. The 6 h forecast after 3 days of 6-hourly cycle assimilating the full set of observations was used as the background (i.e., the first guess). It can be seen the 3DVar humidity increment (Figure 5) is isotropic and uniform, showing little correlation with the background weather situation. The four hybrid data assimilation methods are characterized by anisotropy and a non-uniformity of the analysis increments, which indicates the flow-dependent characteristics corresponding to the humidity contours of the background field with varying degrees. It can also be seen that the distribution range and magnitude of the GE-HDA analysis increments are the smallest, which may be because the flow-dependent background error covariance from the global ensemble contains less mesoscale information. The incremental distribution range of RE20-HDA experiment is larger but with much more noise. The incremental magnitude of RE20-HDA is small, which can be caused by the underestimation of the ensemble error covariance. The analysis increments of the RE60-HDA experiment have the largest distribution range and magnitude, reflecting the relatively larger ensemble error covariance. The analysis increments of GE/RE20-HDA are close to RE60-HDA in terms of distribution range and magnitude, showing the advantage of a larger ensemble and the combination of global and regional ensembles.

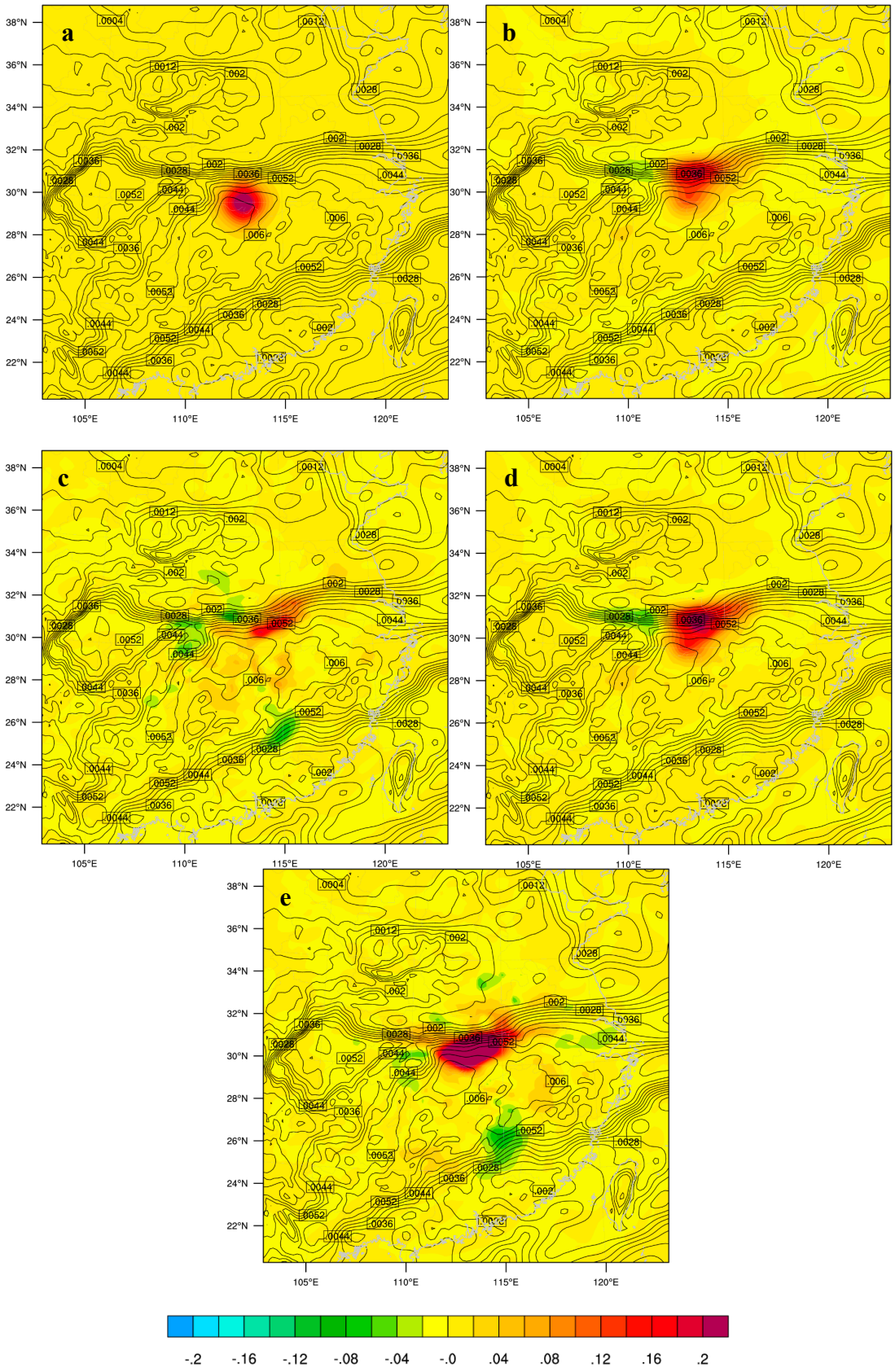

Figure 5. The analysis increments of the water vapor mixing ratio (shaded; kg/kg) of the single observation test. The weight coefficient of the ensemble error covariance is 100%. The localization scale is 400 km. The solid black lines are contours of the background water vapor mixing ratio at analysis time. ((**a**) 3DVar; (**b**) RE20-HDA; (**c**) GE-HDA; (**d**) GE/RE-HDA; (**e**) RE60-HDA).

4.2. Verification against the European Centre for Medium-Range Weather Forecasts (ECMWF) Analysis

Forecast experiments are commonly used to investigate the performance of a data assimilation system in a statistical framework. Thus, the root mean squared error (RMSE) performance of the deterministic forecasts of all data assimilation experiments over the 10-day-long period was investigated in this subsection. The deterministic forecast of each experiment was initialized from the ensemble analysis mean or 3DVar analysis sharing the same physical parameterization schemes. The RMSE of the forecast against the ECMWF analysis was calculated. For the deterministic forecast, the bottom boundary condition and lateral boundary were created from the 6-hourly GFS analysis data. Furthermore, the lateral boundary conditions were also updated according to the atmospheric analysis update.

Figure 6 shows the average analysis RMSEs profiles for the five configurations, which measure the differences of wind components U and V, temperature T, and water vapor Q between the experimental analyses and the ECMWF analyses, respectively. It can be seen that for the wind field, temperature field, and water vapor fields, the analysis RMSEs of the four hybrid EnVar experiments are clearly smaller than the 3DVar. The results of the GE-HDA and RE20-HDA are generally close, but the former are better at the upper levels while the latter are better at lower levels. This indicates that the global ensemble error covariance is more accurate at upper levels but lacks mesoscale information at lower levels. Conversely, the regional ensemble error covariance can provide more accurate mesoscale information at lower levels. This may be because of the finer surface conditions, higher resolution, greater number of local observations, and better suited physical parameterization tuned for the local area in the regional model. These conditions have much more positive impact on the model at lower levels. The RE60-HDA and GE/RE20-HDA schemes are apparently better than the other three data assimilation schemes, obtaining similar results at lower levels. Besides, the GE/RE20-HDA has a clearly lower RMSE at upper levels (100–500 hPa), while the temperature and water vapor fields of RE60-HDA are slightly better.

The error of the first guesses from the short-term forecasts during the cycling period is usually used to evaluate the performance of a data assimilation system. It is necessary to verify the 6 h deterministic forecasts that are similar to the first guess for the next cycle since the cycling interval is 6 h in this study. It can be seen from Figure 7 that for the wind field and the water vapor field, the basic characteristics of the analysis field are generally continued. However, the difference between experiments becomes closer than the analysis fields because of the influence of the model error. For the temperature field, the GE/RE20-HDA and RE60-HDA work best at upper levels. However, the temperature field of RE20-HDA is, overall, the worst for the 6 h forecast field, which is different from the analysis field. For the humidity field, the RMSE of GE-HDA and RE20-HDA is basically the same. For the wind field, however, RE20-HDA has no advantages at lower levels compared with GE-HDA, indicating that the impact of sampling error and the underestimation of covariance caused by the limited ensemble size on the forecast gradually becomes larger.

Shown in Figure 8 are the time series of average analysis RMSE against ECMWF analysis at 500 hPa for U, V, T, and Q of five experiments from 20 to 30, June 2017. It can be seen that the RMSE of the four hybrid EnVar analysis fields is significantly smaller than the RMSE of the 3DVar data assimilation experiments. It is also found that RE60-HDA and GE/RE20-HDA are obviously superior to the other three experiments. For wind and humidity fields, GE/RE20-HDA is clearly better than RE60-HDA; for temperature fields, these two methods are equally effective. In addition, it can be seen that the RMSE of the humidity field is gradually increasing with the analysis cycle, but the humidity error of the GE/RE20-HDA method is relatively much slower, showing its advantage in producing better initial conditions.

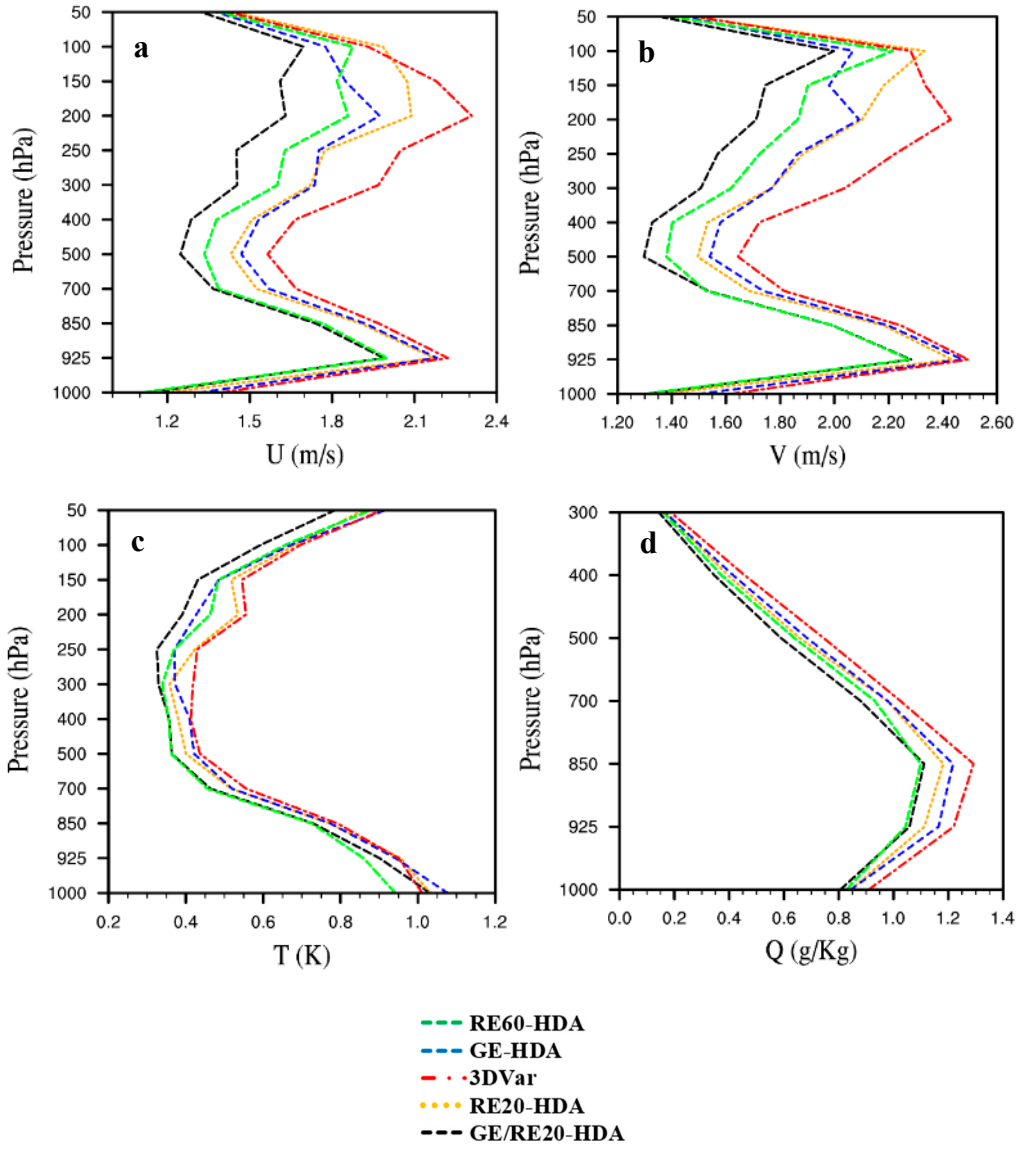

Figure 6. The vertical profiles of the average analysis root mean squared error (RMSE) against the European Centre for Medium-Range Weather Forecasts (ECMWF) analysis. ((**a**) Zonal wind (m/s); (**b**) Meridional wind (m/s); (**c**) Temperature (K); (**d**) Water vapor mixing ratio (g/kg).)

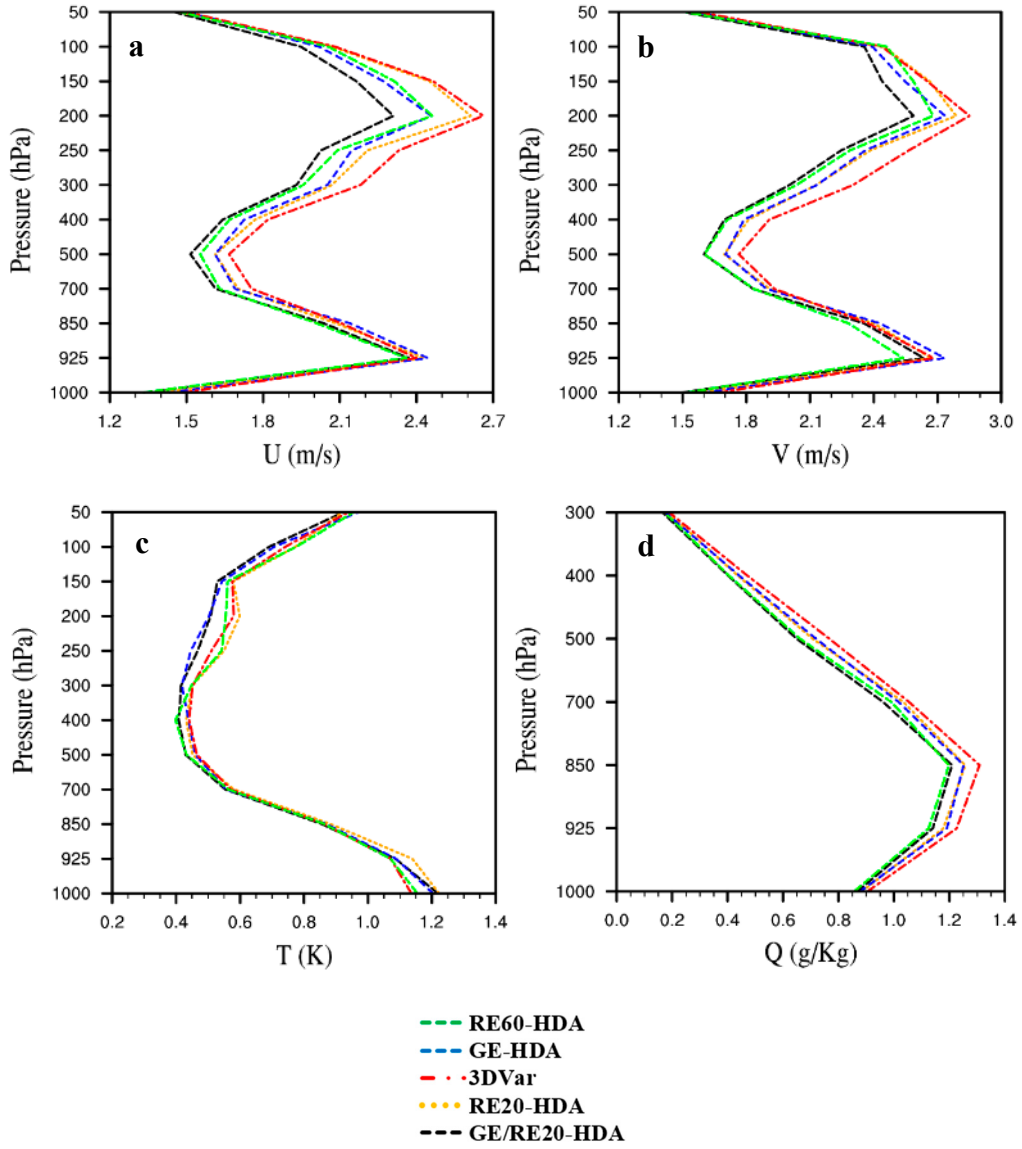

Figure 7. The vertical profiles of the average 6 h forecast root mean squared error (RMSE) against the European Centre for Medium-Range Weather Forecasts (ECMWF) analysis. ((**a**) Zonal wind (m/s); (**b**) Meridional wind (m/s); (**c**) Temperature (K); (**d**) Water vapor mixing ratio (g/kg).)

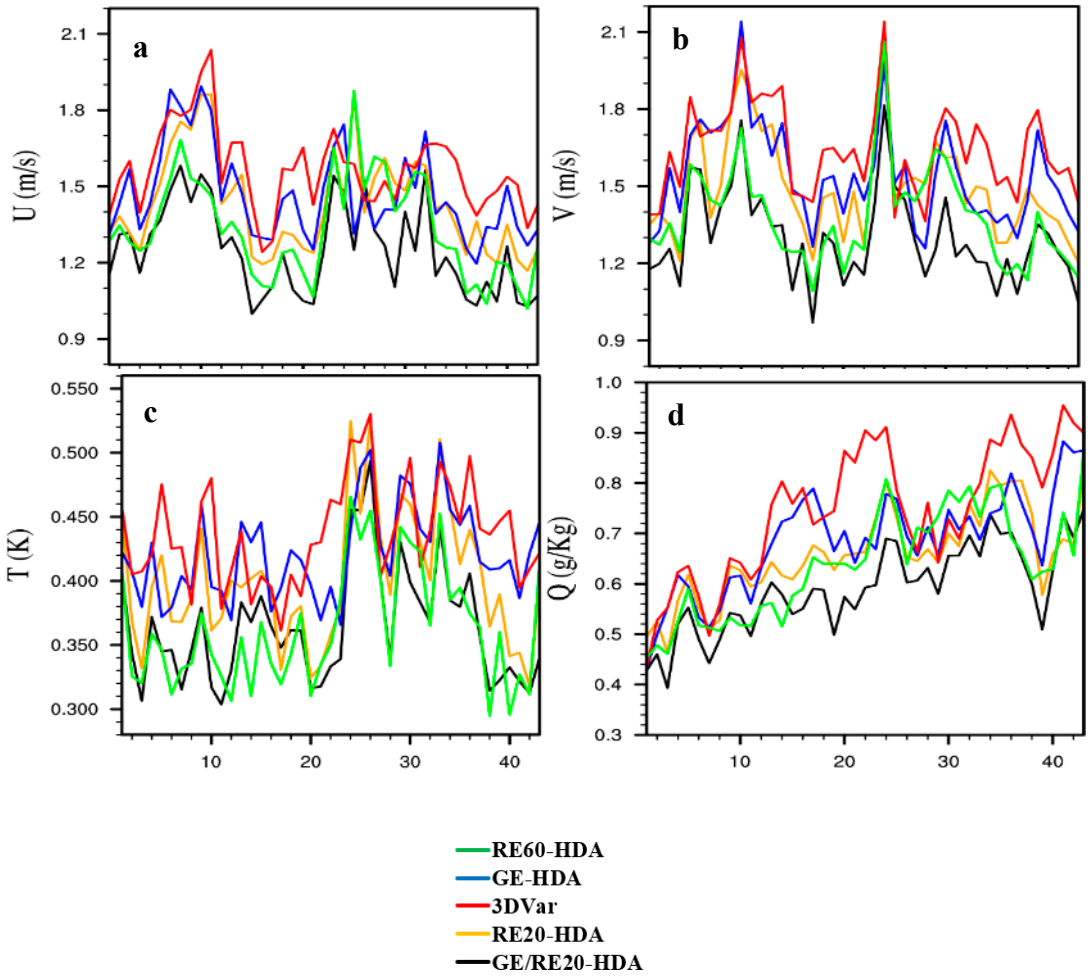

Figure 8. Time series of the average analysis RMSE at 500 hPa. Values on the x axis represent the analysis cycle numbers. ((**a**) Zonal wind (m/s); (**b**) Meridional wind (m/s); (**c**) Temperature (K); (**d**) Water vapor mixing ratio (g/kg).)

4.3. Rainfall Forecast Skill Scores

Three rainfall forecast verification metrics were used in this study to evaluate the rainfall forecast skill of five experiments. The rainfall forecasts were verified using the rainfall observations from the China Hourly Merged Precipitation Analysis Data at 0.1° × 0.1° grid (Shen et al., 2014). The rainfall scores were aggregated over the forty-two 24 h forecasts during the experimental period. This is a persistent heavy rainfall case lasting throughout the experimental period during the Meiyu season that occurred over Jianghuai (the middle and lower reaches of the Yangze River) area. The first verification metric is the Fractions Skill Score (FSS), which ranges between 0 and 1, with 0 representing no overlap and 1 representing complete overlap between forecast and observed events, respectively. The FSS is one of the neighborhood verification methods (Roberts & Lean, 2008), and the influence distance of the neighborhood used in this study is set to 20 km. The second metric is the Equitable Threat Score (ETS), which is commonly known as the Gilbert Skill Score (GSS). The ETS ranges from −1/3 to 1, with 0 or negative values indicating no skill and 1 a perfect score. Different from the FSS, the ETS measures the fraction of observed events that are correctly predicted, adjusted for the frequency of hits that would be expected to occur simply by random chance. The third metric is the Bias Score (BS, also known as Frequency Bias). It ranges from 0 to infinity with 1 representing the perfect score of BS. The BS measures the ratio of the frequency of forecast events to the frequency of observed events, indicating whether the forecast system has the tendency to overpredict (BS > 1) or underpredict (BS < 1) rainfall events [31].

Figure 9 shows the FSS, ETS, and BS as a function of threshold for 24 h accumulated rainfall. For the FSS score, it can be seen that the GE/RE20-HDA generates the best results; for the ETS score, the RE60-HDA obtains the highest score overall. For torrential rain with a threshold larger than 100 mm, the ETS and FSS of GE/RE20-HDA are the highest, but its bias is also the largest; the rainfall forecast bias of the GE/RE20-HDA is the smallest within the precipitation threshold of 100 mm. It can be seen that, for a precipitation threshold larger than 120 mm, the RE20-HDA becomes worse than 3DVar, which may be caused by the poor representation of the weather error structure (i.e., sampling error) for heavy rainfall due to the limited ensemble size. The rainfall forecast skill scores, as a function of forecast range with a threshold of 25 mm every 6 h, are presented in Figure 10. It can be seen that the GE/RE20-HDA experiment still obtains the highest score, and produces the smallest forecast bias. The RE60-HDA is slightly worse than the GE/RE20-HDA scheme, but their results become relatively much closer. The 3DVar is the worst among all the data assimilation schemes, while the RE20-HDA and GE-HDA schemes are equally effective and better than 3DVar.

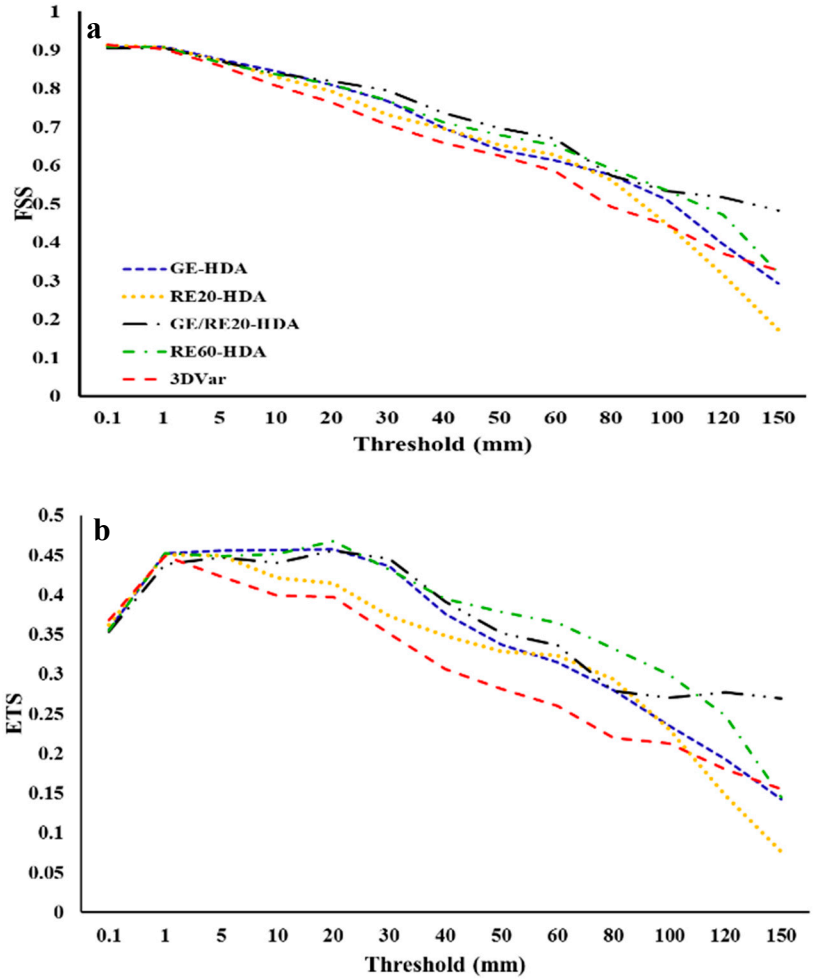

Figure 9. *Cont.*

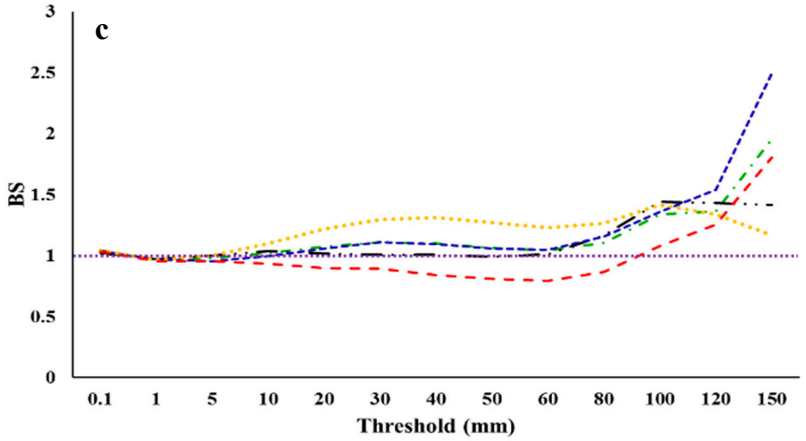

Figure 9. The 24 h accumulated rainfall forecasting scores as a function of forecast leading time for the (**a**) Fraction Skill Score (FSS), (**b**) Equitable Threat Score (ETS), and (**c**) Bias Score (BS).

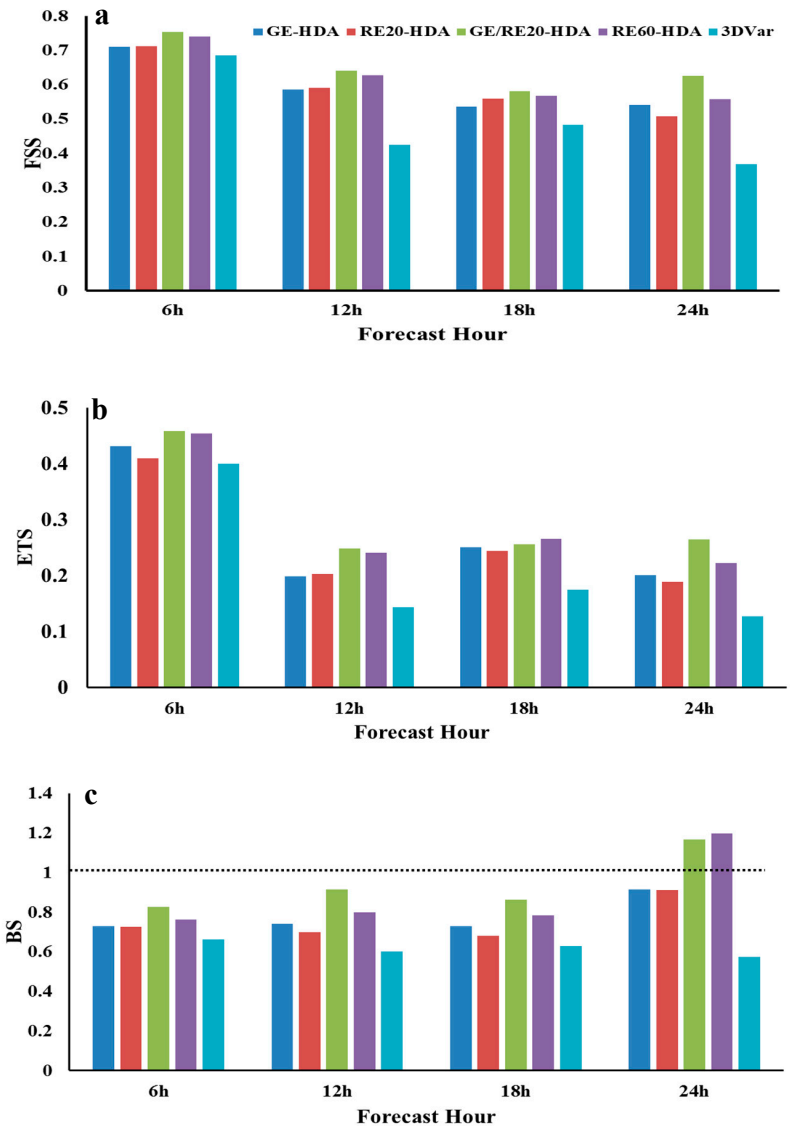

Figure 10. The rainfall forecasting scores with a threshold of 12.5 mm/6 h as a function of forecast leading time for the (**a**) Fraction Skill Score (FSS), (**b**) Equitable Threat Score (ETS), and (**c**) Bias Score (BS).

4.4. Computational Cost Analysis

The main reason for using the global-ensemble-model-augmented error covariance in the hybrid EnVar data assimilation is to increase the ensemble members without significantly increasing the computational cost. Compared with 3DVar, the additional computational cost of traditional hybrid EnVar data assimilation methods usually comes from the following three steps: (1) the Kalman filter analysis; (2) the ensemble integrations; and (3) the computation of extended control variables in the variational cost function. The first two steps take up most of the additional computational cost. However, the augmented global ensemble makes the first two steps much easier and helps produce better analysis with similar or even reduced computational time. Furthermore, the computational cost of extended control variables in the variational cost function is negligible compared with the EnKF analyses and ensemble integrations.

Table 1 lists the wall clock time used by each configuration in a single data assimilation cycle run including the computation of ensemble error covariance, the variational run, and the ensemble run using 120 CPU processors on a Linux workstation. The wall clock time of the deterministic forecast and the pre-process for gribbed GFS data are not included because all experiments share the same time cost in these two steps. We can see that the 3DVar only uses 1 min 16 s of the wall clock time. Compared to 3DVar, the GE-HDA adds only about 4 min to the wall clock time because of the computation of extended control variables and the computing of the global error covariance, as well as the format conversion of the global ensemble data, but the improvement to the quality of analysis is significant, which has been shown in the above sub-sections. However, the ETKF-based experiments (RE20-HDA, GE/RE20-HDA and RE60-HDA) add the ensemble forecast run and the corresponding Kalman filter analysis; as a result, the RE20-HDA produces a result comparable to that with GE-HDA method but uses 32 min 49 s of wall clock time, while GE/RE20-HDA produces a much better result but uses a similar time to RE20-HDA. In this study, the experiment containing 20 regional ensemble members (which is usually viewed as the minimal size for an ensemble data assimilation system) augmented with 80 global ensemble members obtains results similar to those from the experiment containing 60 regional members. However, the former only takes up about one third of the cost of the latter experiment. Such computational cost savings are very important for real-time implementations of operational NWP systems. The cost savings can be used to increase the model resolution, leading time or domain coverage. Obviously, besides the augmented global ensemble, increasing the regional ensemble members can help to further improve the results because of the much more accurate mesoscale information and lesser rank deficiency. However, this may bring an extra burden for computational cost. The regional ensemble members for operational applications may depend on the practical computing resources of operational centers case by case.

Table 1. The computational cost of each data assimilation configuration.

Experiment	3DVar	GE-HDA	RE20-HDA	GE/RE20-HDA	RE60-HDA
cost	1 min 16 s	5 min 38 s	32 min 49 s	36 min 15 s	103 min 28 s

5. Conclusions and Discussion

An efficient regional hybrid EnVar data assimilation method using the global-ensemble-model-augmented error covariance was proposed and preliminarily tested in this study. This work used the global ensemble error covariance as the low-resolution covariance, and used the high-resolution dynamic ensemble forecast mean as the first guess in hybrid EnVar data assimilation, and then re-centered the analysis to the updated high-resolution dynamic ensemble perturbations. We implemented the proposed method into the WRFDA coupled with ETKF scheme and tested it for numerical weather prediction over eastern China. In this study, the experiment containing small regional ensemble members augmented with global ensemble members obtains results similar to those from the experiment containing relatively larger regional members. However,

the former only takes up one third of the computational cost of latter experiment. The method proposed in this study also outperforms the 3DVar, hybrid EnVar with pure global ensemble error covariance, as well as the hybrid EnVar with a small size ETKF ensemble. The method proposed in this paper effectively combines contributions from both the global and the regional ensembles to produce better initial conditions for the regional WRF data assimilation system.

The aim of this paper is to propose a method for improving the performance of hybrid EnVar data assimilation through effectively increasing the rank of the flow-dependent part of the background error covariance by including the information from a global ensemble. In this study, the large-scale part of the ensemble error covariance was contributed by the global ensemble. On a practical level, this part can also be provided by a larger domain ensemble of lower resolution that covers the high-resolution area. One of the main limitations of the global ensemble error covariance used in regional hybrid data assimilation is that the global ensembles are not updated frequently. For example, we can obtain the 3-hourly global ensemble from the NCEP GEFS. But for the hourly rapid update system, the 3-hourly global ensemble is not enough. In this situation, we can utilize the interpolation of the 3-hourly global ensemble to the hourly global ensemble as Yang et al. (2017) [32] did, the time-expanded ensembles as Zhao et al. 2015 [19] did, or a time-lagged ensemble as Wang et al. 2017 [21] did. Furthermore, as the resolution of global ensemble in time and space will be increased in the future, the augmented global error covariance can be used for the rapid updated regional convective-permitting NWP. In addition, we used a combined background error covariance from different ensembles to update the regional analysis mean but not the regional ensemble perturbations. How the high-resolution regional ensemble perturbations will be influenced by the global or low-resolution regional ensemble error covariance may need further study.

Author Contributions: Conceptualization, Y.C. and Y.W.; Formal analysis, Y.W.; Resources, Y.C and J.M.; Writing—original draft, Y.W.; Writing—review & editing, Y.C. and J.M.; Supervision, Y.C. and J.M.; Project administration, Y.C. and J.M.; Funding acquisition, Y.C. and J.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work is jointly sponsored by the National Key Research and Development Program of China (2018YFC1506803), the National Natural Science Foundation of China (41675102, 41805071). The Startup Foundation for Introducing Talent of NUIST (2018r065, 2017r058).

Acknowledgments: The authors thank the editors and three anonymous reviewers for their constructive comments and useful suggestions which helps significantly improve this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bannister, R.N. A review of operational methods of variational and ensemble-variational data assimilation. *Q. J. Roy. Meteor. Soc.* **2017**, *143*, 607–633. [CrossRef]
2. Parrish, D.F.; Derber, J.C. The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Weather Rev.* **1992**, *120*, 1747–1763. [CrossRef]
3. Barker, D.M. Southern high-latitude ensemble data assimilation in the Antarctic mesoscale prediction system. *Mon. Weather Rev.* **2005**, *133*, 3431–3449. [CrossRef]
4. Descombes, G.; Aulign'e, T.; Vandenberghe, F.; Barker, D.M.; Barr´e, J. Generalized background error covariance matrix model (GEN BE v2.0). *Geosci. Model Dev.* **2015**, *8*, 669–696. [CrossRef]
5. Hamill, T.M.; Snyder, C. A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Weather Rev.* **2000**, *128*, 2905–2919. [CrossRef]
6. Lorenc, A.C. The potential of the ensemble Kalman filter for NWP-a comparison with 4D-Var. *Q. J. R. Meteorol. Soc.* **2003**, *129*, 3183–3203. [CrossRef]
7. Wang, X.G.; Parrish, D.; Kleist, D.; Whitaker, J. GSI 3DVar-based ensemble–variational hybrid data assimilation for NCEP global forecast system: Single-resolution experiments. *Mon. Weather Rev.* **2013**, *141*, 4098–4117. [CrossRef]

8. Caron, J.F.; Milewski, T.; Buehner, M.; Fillion, L.; Reszka, M.; Macpherson, S.; St-James, J. Implementation of deterministic weather forecasting systems based on ensemble–variational data assimilation at Environment Canada. Part II: The regional system. *Mon. Weather Rev.* **2015**, *143*, 2560–2580. [CrossRef]
9. Kleist, D.T.; Ide, K. An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part II: 4DEnVar and hybrid variants. *Mon. Weather Rev.* **2015**, *143*, 452–470. [CrossRef]
10. Wang, X.G.; Barker, D.M.; Snyder, C.; Hamill, T.M. A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. *Mon. Weather Rev.* **2008**, *136*, 5132–5147. [CrossRef]
11. Zhang, F.; Zhang, M.; Poterjoy, J. E3DVar: Coupling an Ensemble Kalman Filter with Three-Dimensional Variational Data Assimilation in a Limited-Area Weather Prediction Model and Comparison to E4DVar. *Mon. Weather Rev.* **2013**, *141*, 900–917. [CrossRef]
12. Schwartz, C.S.; Liu, Z. Convection-permitting forecasts initialized with continuously cycling limited-area 3DVar, ensemble Kalman filter, and "Hybrid" variational–ensemble data assimilation systems. *Mon. Weather Rev.* **2014**, *142*, 716–738. [CrossRef]
13. Schwartz, C.S.; Liu, Z.; Huang, X.-Y. Sensitivity of limited-area hybrid variational-ensemble analyses and forecasts to ensemble perturbation resolution. *Mon. Weather Rev.* **2015**, *143*, 3454–3477. [CrossRef]
14. Hamill, T.M.; Whitaker, J.S.; Snyder, C. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Weather Rev.* **2001**, *129*, 2776–2790. [CrossRef]
15. Miyoshi, T.; Kondo, K.; Imamura, T. The 10240-member ensemble Kalman filtering with an intermediate AGCM. *Geophys. Res. Lett.* **2014**, *41*, 5264–5271. [CrossRef]
16. Kondo, K.; Miyoshi, T. Impact of removing covariance localization in an ensemble Kalman filter: Experiments with 10240 members using an intermediate AGCM. *Mon. Weather Rev.* **2016**, *144*, 4849–4865. [CrossRef]
17. Anderson, J.L. Reducing correlation sampling error in Ensemble Kalman Filter data assimilation. *Mon. Weather Rev.* **2016**, *144*, 913–925. [CrossRef]
18. Xu, Q.; Lu, H.; Gao, S.; Xue, M.; Tong, M. Time-expanded sampling for ensemble Kalman filter: Assimilation experiments with simulated radar observations. *Mon. Weather Rev.* **2008**, *136*, 2651–2667. [CrossRef]
19. Zhao, Q.; Xu, Q.; Jin, Y.; McLay, J.; Reynolds, C. Time-expanded sampling for ensemble-based data assimilation applied to conventional and satellite observations. *Weather Forecast.* **2015**, *30*, 855–872. [CrossRef]
20. Gustafsson, N.; Bojarova, J.; Vignes, O. A hybrid variational ensemble data assimilation for the HIgh Resolution Limited Area Model (HIRLAM). *Nonlinear Proc. Geoph.* **2014**, *21*, 303–323. [CrossRef]
21. Wang, Y.; Min, J.; Chen, Y.; Huang, X.-Y.; Zeng, M.; Li, X. Improving precipitation forecast with hybrid 3DVar and time-lagged ensembles in a heavy rainfall event. *Atmos. Res.* **2017**, *183*, 1–16. [CrossRef]
22. Kretschmer, M.; Hunt, B.R.; Ott, E. Data assimilation using a climatologically augmented local ensemble transform Kalman filter. *Tellus A* **2015**, *67*, 26617. [CrossRef]
23. Gao, J.; Xue, M. An efficient dual-resolution approach for ensemble data assimilation and tests with simulated Doppler radar data. *Mon. Weather Rev.* **2008**, *136*, 945–963. [CrossRef]
24. Pan, Y.; Zhu, K.; Xue, M.; Wang, X.; Hu, M.; Benjamin, S.G.; Whitaker, J.S. A GSI-based coupled EnSRF–En3DVar hybrid data assimilation system for the operational rapid refresh model: Tests at a reduced resolution. *Mon. Weather Rev.* **2014**, *142*, 3756–3780. [CrossRef]
25. Wu, W.S.; Parrish, D.F.; Rogers, E.; Lin, Y. Regional Ensemble–Variational Data Assimilation Using Global Ensemble Forecasts. *Weather Forecast.* **2017**, *32*, 83–96. [CrossRef]
26. Rainwater, S.; Hunt, B. Mixed-resolution ensemble data assimilation. *Mon. Weather Rev.* **2013**, *141*, 3007–3021. [CrossRef]
27. Whitaker, J.S.; Hamill, T.M.; Wei, X.; Song, Y.; Toth, Z. Ensemble data assimilation with the NCEP global forecast system. *Mon. Weather Rev.* **2008**, *136*, 463–482. [CrossRef]
28. Kleist, D.T.; Parrish, D.F.; Derber, J.C.; Treadon, R.; Wu, W.S.; Lord, S. Introduction of the GSI into the NCEP global data assimilation system. *Weather Forecast.* **2009**, *24*, 1691–1705. [CrossRef]
29. Skamarock, W.C.; Klemp, J.B.; Dudhia, J.; Gill, D.O.; Barker, D.M.; Wang, W.; Powers, J.G. *A Description of the Advanced Research WRF Version 3*; NCAR Technical Note; Citeseer: Boulder, CO, USA, 2008.
30. Barker, D.M.; Huang, W.; Guo, Y.R.; Bourgeois, A.J.; Xiao, Q.N. A three-dimensional variational data assimilation system for MM5: Implementation and initial results. *Mon. Weather Rev.* **2004**, *132*, 897–914. [CrossRef]

31. Wang, Y.; Liu, Z.; Yang, S.; Min, J.; Chen, L.; Chen, Y.; Zhang, T. Added value of assimilating Himawari-8 AHI water vapor radiances on analyses and forecasts for "7.19" severe storm over north China. *J. Geophys Res.-Atmos.* **2018**, *123*, 3374–3394. [CrossRef]
32. Yang, C.; Liu, Z.; Gao, F.; Childs, P.P.; Min, J. Impact of assimilating GOES imager clear-sky radiance with a rapid refresh assimilation system for convection-permitting forecast over Mexico. *J. Geophys. Res.-Atmos.* **2017**, *122*, 5472–5490. [CrossRef]