Cloud-Dependent Piecewise Assimilation Based on a Hydrometeor-Included Background Error Covariance and Its Impact on Regional Numerical Weather Prediction

DEMING MENG,[a,b] YAODENG CHEN,[a] JUN LI,[b] HONGLI WANG,[c,d] YUANBING WANG,[a] AND TAO SUN[a]

[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 and Technology, Nanjing, China*
[b] *Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin–Madison, Madison, Wisconsin*
[c] *Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado*
[d] *NOAA/OAR/Earth System Research Laboratory/Global Systems Division, Boulder, Colorado*

ABSTRACT: The background error covariance (**B**) behaves differently and needs to be carefully defined in cloudy areas due to larger uncertainties caused by the models' inability to correctly represent complex physical processes. This study proposes a new cloud-dependent **B** strategy by adaptively adjusting the hydrometeor-included **B** in the cloudy areas according to the cloud index (CI) derived from the satellite-based cloud products. The adjustment coefficient is determined by comparing the error statistics of **B** for the clear and cloudy areas based on the two-dimensional geographical masks. The comparison highlights the larger forecast errors and manifests the necessity of using appropriate **B** in cloudy areas. The cloud-dependent **B** is then evaluated by a series of single observation tests and 3-week cycling assimilation and forecasting experiments. The single observation experiments confirm that the cloud-dependent **B** allows cloud dependency for the multivariate analysis increments and alleviates the discontinuities at the cloud mask borders by treating the CI as an exponent. The impact study on regional numerical weather prediction (NWP) demonstrates that the application of the cloud-dependent **B** reduces analysis and forecast bias and increases precipitation forecast skills. Diagnostics of a heavy rainfall case indicate that the application of the cloud-dependent **B** enhances the moisture, wind, and hydrometeors analyses and forecasts, resulting in more accurate forecasts of accumulated precipitation. The cloud-dependent piecewise analysis scheme proposed herein is extensible, and a more precise definition of CI can improve the analysis, which deserves future investigation.

SIGNIFICANCE STATEMENT: Background error covariance (**B**) provides important statistics to quantify errors in model background fields for data assimilation. The majority of variational-based assimilation systems use the same **B** for the cloudy and clear skies. However, the numerical weather prediction models (NWP) typically have less accuracy in cloudy areas; hence, discrepancies in **B** are expected between the cloudy and clear areas. To use appropriate **B** in cloudy areas, this study proposes a cloud-dependent piecewise assimilation scheme in which **B** is adaptively adjusted according to a real-time two-dimensional cloud index (CI) generated by satellite-based cloud products. The newly proposed scheme is evaluated by cycling assimilation and forecast experiments over a 3-week period. Analysis and forecasts reveal that the cloud-dependent **B** is able to represent the larger errors in cloudy regions, resulting in improved fitting to observations and improved precipitation forecasts.

KEYWORDS: Satellite observations; Clouds; Data assimilation; Regional models

1. Introduction

It is a continuing challenge for the research community and the operational centers to improve the atmosphere's initial state (Li et al. 2016; Bannister 2017). Various data assimilation (DA) methods have been demonstrated to be able to improve the initial condition for numerical weather prediction (NWP), among which the variational framework DA, such as three- and four-dimensional variational assimilation (3D-Var and 4D-Var, respectively) and ensemble-based variational assimilation (EnVar), are widely applied operationally (e.g., Brousseau et al. 2011; Ingleby et al. 2013; Aranami et al. 2015; Wu et al. 2017) due to their strong extensibility to assimilate multiple kinds of observations including nonconventional remote sensing observations. One of the indispensable components of the variational framework is background error covariance (**B**) because it provides spatial and cross-variable correlations (Daley 1991; Bannister 2008a,b; Chen et al. 2013; Descombes et al. 2015). Although various EnVar methods (e.g., Wang et al. 2008a,b; Wang et al. 2017, 2018) have been proposed to update **B** with flow dependence, the static and homogeneous **B** is still included in operational DA.

In reality, **B** can behave significantly differently between cloudy and clear areas due to large differences in physical processes. Cloud and precipitation systems are generally associated with significant nongeostrophic motion, resulting in larger variances, stronger cross correlation between control variables (CVs), and smaller horizontal decorrelation length scales, both for the hydrometeor and nonhydrometeor variables (Caron and Fillion 2010; Montmerle and Berre 2010; Ménétrier and Montmerle 2011; Michel et al. 2011). These different **B** behaviors may be attributed

Corresponding author: Yaodeng Chen, keyu@nuist.edu.cn

to the complexity of the physical processes when cloud and precipitation occur due to their direct and indirect influence on the atmosphere (Errico et al. 2007) and highly approximate parameterization schemes of nonlinear microphysical processes (Lopez 2007). These findings lead us to believe that different **B** should be applied for the assimilation in the cloudy and precipitation areas.

To use different **B** in the assimilation under the cloudy and precipitation area, Montmerle and Berre (2010) developed a heterogeneous **B** matrix by applying geographical masks and incorporated it in the Météo-France operational model (AROME) to enable piecewise analysis. Their approaches conduct analysis with different balances and a spatial distribution according to the weather types. Ménétrier and Montmerle (2011) further addressed the issue of mask-borders discontinuities by proposing a convolution of normalized Gaussian kernel and applied this heterogeneous **B** to the analysis and prediction of fog cases. Their studies demonstrate the feasibility of piecewise assimilation and the positive impacts of heterogeneous **B** on the assimilation system. However, the absence of the hydrometeor CVs makes this heterogeneous **B** less appropriate for analyzing the cloud components.

The hydrometeor CVs were introduced into heterogeneous **B** by Michel et al. (2011). Detailed **B** structures for hydrometeor variables and nonhydrometeor variables were diagnosed for clear and precipitation areas in their study. Chen et al. (2016) further applied the hydrometeor-included **B** to assimilate satellite cloud water/ice path for the real precipitation cases. Assimilating the cloud observations with the hydrometeor-included **B** in their study showed great potential for improving precipitation forecast. However, the **B** in their study is isotropic and homogeneous for all CVs, ignoring the different error characteristics under the cloudy and clear skies. Hydrometeor forecast errors vary significantly in space due to the intrinsically heterogeneous nature of hydrometeor fields (Meng et al. 2019; Destouches et al. 2020), suggesting the necessity of using different **B** for clear and cloudy areas.

To achieve the piecewise analysis for different weather types, a new cloud-dependent **B** strategy is proposed in this paper and implemented for the Weather Research and Forecasting (WRF) DA system (WRFDA) variational framework. With the inclusion of hydrometeors, **B** is modeled separately for the cloudy and clear areas in spectral space (also known as the empirical orthogonal function (EOF) space). The WRFDA is then updated to implement the adaptive adjustment of **B** for each CV according to the real-time cloud index (CI), which is quantified as a normalized two-dimensional field using the real-time cloud amount from satellite-based cloud products. This strategy provides an alternative option for piecewise analysis aside from redesigning the heterogeneous **B** in Montmerle and Berre (2010). Moreover, cloud dependencies are also introduced into the assimilation system by using the cloud-dependent **B**. Although various forms of Ensemble-based data assimilation (EnKF) could provide the real flow-dependent **B** in an ensemble subspace based on an ensemble of forecasts (Houtekamer and Mitchell 1998; Ott et al. 2004; Anderson 2012; Barker et al. 2012; Wang et al. 2017, 2018), the heavy

computational pressure associated with ensemble forecasts still hinders its popularization in some operational meteorological centers (Hamill and Snyder 2000). From this perspective, a flow-dependent scheme in a variational framework with a very low computational cost is still valuable to be developed. The difference is that the flow-dependent information comes from satellite cloud observations rather than ensemble forecasts.

Section 2 provides details on the implementation of the cloud-dependent piecewise analysis. Section 3 gives the statistics for **B** with the inclusion of hydrometeors and their comparisons over the clear and cloudy areas. The single-observation experiments using the cloud-dependent **B** are presented in section 4. Section 5 presents the performance of the cloud-dependent **B** application on 3-week rainfall events. A heavy rainfall event diagnosis is described in section 6. Section 7 summarizes conclusions and discussion.

2. Cloud-dependent piecewise assimilation implementation

Following the studies of Montmerle and Berre (2010) and Ménétrier and Montmerle (2011), we develop a heterogeneous **B** to enable piecewise analysis when clouds exist. Different from the previous studies, an alternative technique is developed to deduce the real-time normalized 2D CI, namely using the cloud liquid (LWP) and ice (IWP) water path from the Langley Research Center (LaRC) at National Aeronautics and Space Administration (NASA) (Minnis 2007; Minnis et al. 2008). This cloud product could provide hourly global cloud water path (CWP; the sum of LWP and IWP) data. The reason why observations are introduced to adjust the **B** is that the hydrometeors from model simulations are currently insufficient to reasonably characterize the distribution of the real cloud, especially the inaccuracies in position and magnitude. Similar idea is also used by Geer and Bauer (2011) to develop the symmetric observation error model for all-sky radiance assimilation by combining information from observations and model predictions.

The **B** can be adaptively adjusted according to the normalized CI, referred to as the cloud-dependent **B**. The key technologies of the cloud-dependent piecewise assimilation contain the following three main aspects, 1) the normalized CI setup, 2) the cloud-dependent **B** adjustment in cloudy areas, and 3) the determination of the adjustment coefficient.

a. The normalized CI setup

To match the cloudiness from observation space to model space, CWP is first interpolated to the model grid points. To ensure smooth adjustment of **B**, CI needs to be normalized. Hence, CI is then expressed as a ratio of the CWP at each grid point to the maximum CWP of the entire experiment domain (Fig. 1) to ensure that the CI value is from 0 to 1:

$$\left\{ \begin{array}{l} \mathrm{CWP}| = \mathrm{LWP} + \mathrm{IWP} \\ \mathrm{CI}| = \dfrac{\mathrm{CWP}}{\max(\mathrm{CWP})} \end{array} \right. . \qquad (1)$$

Therefore, the CI is normalized depending on CWP, which varies in real time, ranging from 0 to 1. It is worth noting that if
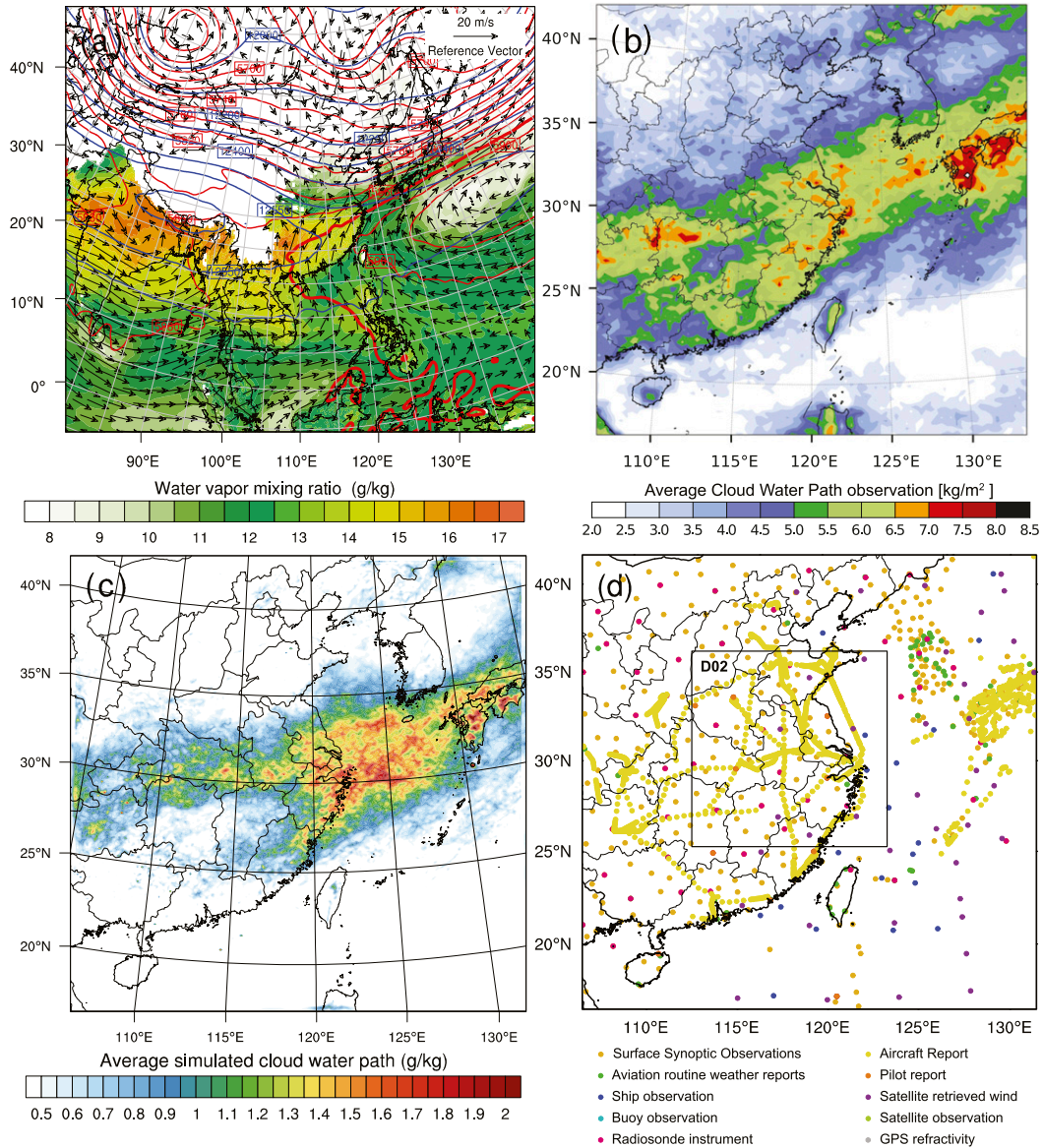
FIG. 1. (a) Meteorological situations for the **B** statistics. The 200-hPa geopotential height field (blue lines; unit: gpm), 500-hPa geopotential height field with the thickened line of 5880 gpm (red lines; unit: gpm), and 850-hPa wind vector field (vectors; unit: m s^{-1}) overlaid with 850-hPa water vapor mixing ratios (shaded; units: g kg^{-1}) of the NCEP FNL (Final) operational global analysis averaged from 30 May to 30 Jun 2016. (b) Distribution of the average NASA cloud water path observation from 30 May to 30 Jun 2016, (c) distribution of the average simulated cloud water path (Qi + Qc) in the samples used for **B** statistics, and (d) experimental domain with the distributions of the conventional observations at 1800 UTC 30 Jun 2016 used in this study. The **B** statistics are performed on D01, while the cycling assimilation and forecast are operated on D01 and D02.

the experiment domain is completely cloud-free, even if this happens rarely, the CI will be set to 0.

The CI can then delineate the clear and the cloudy 2D geographical masks, denoted by F_{clear} and F_{cloud} operator, respectively:

$$\begin{cases} F_{clear}, & CI < 10^{-3} \\ F_{cloud}, & CI \geq 10^{-3} \end{cases}. \tag{2}$$

The current study focuses on verifying the feasibility of the cloud-dependency scheme, so the current definition of CI only considers spatial dependency. Although the time dependency is also crucial for cloud variables, it will be left for further studies.

b. The cloud-dependent **B** adjustment in cloudy areas

In the variational framework assimilation system, the analysis increment δx can be expressed as

$$\delta\mathbf{x} = (\mathbf{x} - \mathbf{x}_b) = \mathbf{B}^{1/2}\mathbf{u}, \tag{3}$$

where background state vector is defined as $\mathbf{x}_b$, and $\mathbf{x}$ is the state vector to be analyzed. The different CVs are defined as $\mathbf{u}$. The square root of $\mathbf{B}$ is represented as $\mathbf{B}^{1/2}$, which can be decomposed to a series of submatrices:

$$\mathbf{B}^{1/2} = \mathbf{U}_p \mathbf{U}_v \mathbf{U}_h, \tag{4}$$

where the physical transform ($\mathbf{U}_p$), vertical transform ($\mathbf{U}_v$), and horizontal transform ($\mathbf{U}_h$) compose the process of control variable transform.

According to Courtier et al. (1998), the two isotropic covariance tensors can be simultaneously applied to two complementary subdomains. Therefore, $\mathbf{B}^{1/2}$ can be decomposed into two terms, with each term corresponding to the clear or cloudy skies, thus,

$$\delta\mathbf{x} = F_{\text{clear}} \mathbf{B}^{1/2}_{\text{clear}} \mathbf{u} + F_{\text{cloud}} \mathbf{B}^{1/2}_{\text{cloud}} \mathbf{u}. \tag{5}$$

In Eq. (5), $\mathbf{B}_{\text{clear}}$ is the homogeneous $\mathbf{B}$ computed using the NMC method over all grid points in the domain, as commonly applied in operation models, which is also applied in Ménétrier and Montmerle (2011). For the grid points defined as cloudy (F_{cloud}), $\mathbf{B}_{\text{cloud}}$ will be adjusted based on the homogeneous $\mathbf{B}$ and a set of adjustment coefficients I_u when transforming from physical space to spectral space (or EOF space) in $\mathbf{U}_v$:

$$\begin{cases} \mathbf{B}^{1/2}_{\text{clear}}| = \mathbf{U}_p \mathbf{U}_v \mathbf{U}_h \\ \mathbf{B}^{1/2}_{\text{cloud}}| = \mathbf{U}_p I_u \mathbf{U}_v \mathbf{U}_h \end{cases} \tag{6}$$

where I_u are a set of scalar variables larger than or equal to 0 and vary according to different CVs. It should be noted that the adjustment of $\mathbf{B}_{\text{cloud}}$ is implemented during the $\mathbf{U}_v$ process, which is a built-in part of WRFDA. Thus, even though Eqs. (5) and (6) demonstrate two $\mathbf{B}$, the iteration only requires the participation of the homogeneous $\mathbf{B}$. When the presence of a cloud is detected, the adjustment would be applied adaptively on the vertical transform of the homogeneous $\mathbf{B}$ by the system. The adjustment can also influence balance relationships and length scales; however, their discussions are beyond this paper's scope.

One may notice that the covariance function will be discontinuous at the cloud mask border if $\mathbf{B}^{1/2}$ is directly adjusted by I_u in cloudy areas. The discontinuity of the covariance function makes the convergence of the cost function difficult. To solve this issue, the equation is updated by taking I_u as the base of the logarithm and CI as the exponent:

$$\mathbf{B}^{1/2}_{\text{cloud}} = \mathbf{U}_p I_u^{\text{CI}} \mathbf{U}_v \mathbf{U}_h. \tag{7}$$

There are two benefits for $\mathbf{B}^{1/2}_{\text{cloud}}$ being formulated as the exponential function of CI instead of a linear function: 1) Since CI is normalized and approaches 1 with the increasing amount of clouds, the usage of I_u^{CI} blurs the boundary of the mask, allowing a smooth adjustment of $\mathbf{B}^{1/2}_{\text{cloud}}$ at the border of the cloud mask, 2) Treating CI as an exponent also reduces the difficulty of coding implementation. This is because CI varies

from 0 to 1, and $\mathbf{B}^{1/2}_{\text{cloud}}$ turns to be $\mathbf{B}^{1/2}_{\text{clear}}$ when CI is very small ($I_u^{\text{CI}} = 1$).

Different from the redesigned heterogeneous $\mathbf{B}$ matrix, in which the size of the CVs is doubled due to the multiplication by two different $\mathbf{B}$ matrices (Montmerle and Berre 2010), this method keeps the same sizes of CVs and cost function gradient as the traditional schemes by directly adjusting the homogeneous $\mathbf{B}$ in cloudy areas. Thus, little extra computational load is added.

The proposed scheme enables the cloud-dependent piecewise analysis according to cloud magnitudes. This scheme may not be as effective as a full heterogeneous scheme obtained from a flow-dependent forecast ensemble in an EnVar system. Nevertheless, the scheme costs substantially less than the EnVar scheme while accounting for the main statistical characteristics in cloud regions.

c. Determination of the adjustment coefficient

To determine the adjustment coefficients I_u for each CV, it is helpful to know the different characteristics of $\mathbf{B}$ for the clear and cloudy areas. To characterize the $\mathbf{B}$ matrix in cloudy areas separately, a two-dimensional cloud-dependent bin is devised following the approach of Michel et al. (2011), so that the cloudy samples can be separated as

$$\varepsilon_b = P_{\text{clear}}\varepsilon_b + P_{\text{cloud}}\varepsilon_b + P_{\text{mixed}}\varepsilon_b, \tag{8}$$

where ε_b is the background errors, and P_{clear}, P_{cloud}, and P_{mixed} are projection operators, which denote clear, cloudy, and mixed bins, respectively. Different from vertically averaged rain content ($\int_{\text{bot}}^{\text{top}} Q_r$) used in Michel et al. (2011), here, vertically averaged cloud content $[\int_{\text{bot}}^{\text{top}}(Q_c + Q_i)]$ is used to define the projection operators because broader coverage in cloud content can provide more cloudy samples. Operator P_{cloud} is determined only when the $\int_{\text{bot}}^{\text{top}}(Q_c + Q_i)$ is greater than 0.01 g kg^{-1} at the same points for both the 12- and 24-h forecasts which are used in the NMC (National Meteorology Center; Parrish and Derber 1992) perturbation calculation, i.e., $\int_{\text{bot}}^{\text{top}}(Q_c + Q_i) \geq 0.01 \text{ g kg}^{-1}$. Similarly, when $\int_{\text{bot}}^{\text{top}}(Q_c + Q_i) < 0.01 \text{ g kg}^{-1}$ for 12- and 24-h forecasts, the bin is classified as P_{clear}. The samples are classified as a "mixed" bin when only one of 12- and 24-h forecasts has $\int_{\text{bot}}^{\text{top}}(Q_c + Q_i) > 0.01 \text{ g kg}^{-1}$. Therefore, $\mathbf{B}$ can be expressed as

$$\mathbf{B} = \overline{\varepsilon_b \varepsilon_b^{\text{T}}} = \overline{P_{\text{clear}}\varepsilon_b \varepsilon_b^T P_{\text{clear}}^{\text{T}}} + \overline{P_{\text{cloud}}\varepsilon_b \varepsilon_b^T P_{\text{cloud}}^{\text{T}}} + \overline{P_{\text{mixed}}\varepsilon_b \varepsilon_b^{\text{T}} P_{\text{mixed}}^{\text{T}}} = \mathbf{B}_{\text{clear}} + \mathbf{B}_{\text{cloud}} + \mathbf{B}_{\text{mixed}}. \tag{9}$$

It should be noted that the mixed situation may be resulted from cloud displacement error and usually occur in the boundary of clouds with small sample sizes, which tends to produce sampling errors in the statistics. Therefore, the focus will be on the characteristics of the $\mathbf{B}$ statistics from clear and cloudy bins.

In addition to the nonhydrometeor CVs in the $\mathbf{B}$, which include the streamfunction (Ψ), unbalanced velocity potential function (X_u), unbalanced temperature (T_u), unbalanced relative humidity (RH_u), and unbalanced surface pressure (P_{su}), including hydrometeor CVs is also indispensable, especially

when performing the assimilation in cloudy areas. In this study, the hydrometeor-included **B** is constructed as Chen et al. (2016).

Then, the errors in both physical and spectral spaces, i.e., standard deviations (SDs) and eigenvalues, in the **B**$_{clear}$ and **B**$_{cloud}$ are used to determine I_u for each CVs:

$$I_u = \frac{\text{cloud}\_\sigma_u^b}{\text{clear}\_\sigma_u^b}, \qquad (10)$$

where cloud_σ_u^b and clear_σ_u^b represents SDs or eigenvalues of **B**$_{cloud}$ and **B**$_{clear}$, respectively.

Theoretically, the definition of I_u should vary with the cloud amount, so that the definition of I_u can be consistent with CI. However, finer classification can lead to fewer samples for each class, thus causing sample errors. It should be noted that the calculation only provides a general magnitude for the ratio of the error between the clear and the cloudy areas, and they can be tuned in practice. Defining I_u more objectively should also be considered in future studies.

3. Comparison of **B** in clear and cloudy areas

In this section, **B** over the clear (BE_CLR) and cloudy (BE_CLD) areas will be generated, and then the errors of physical and spectral space for each CV in **B**$_{clear}$ and **B**$_{cloud}$, respectively, are diagnosed. Finally, the adjustment coefficient I_u will be calculated for each CV.

a. Meteorological situations and the statistical datasets

Mei-yu is an essential component of spring–summer precipitation in China, during which heavy rainfall frequently happens. The Yangtze River basin experienced a more widespread and more intense rainfall process from May to July 2016 compared to an average year. The mei-yu in 2016 was the third largest since 1951, with a total rainfall amount of 584.3 mm over the Jianghuai area during the mei-yu period, which is 108% above average and led to severe flooding (Yuan et al. 2017). The NCEP FNL (Final) operational global analysis (Fig. 1a) shows that a strong anticyclone at 200 hPa (South Asian high pressure; blue lines) is maintained over the Yangtze River basin. The ridgeline of the western Pacific subtropical high (5880 gpm; thick red lines) extends from southern China to southern Japan (500 hPa; red lines). The wind (vectors) at 850 hPa showed clear shear lines in the Yangtze River basin. With this stable atmospheric circulation, the upper-level air diverged, while the lower-level air converged, which favored water vapor transportation and convergence, lead to persistent precipitation. As is shown in Fig. 1b, the average hourly NASA CWP observation from 30 May to 30 June shows that the average cloud observation was concentrated in South China and the Yangtze River basin, which is consistent with the distribution of main rainfall areas (not shown).

The sample datasets for **B** statistics are generated from the WRF Model version 3.9 (Skamarock and Klemp 2008) configured with 321 × 321 grid points. The horizontal resolution is set to 9 km (Fig. 1d, D01), and the terrain-following coordinates are used for 51 vertical layers with the model top at 50 hPa. The analysis at 0.5° horizontal grid resolution from the

National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) is used to drive WRF forecasts. The WRF double-moment 6-class microphysics scheme (WDM6; Lim and Hong 2010) is adopted for the microphysics parameterizations pairing with the Kaine–Fritsch cumulus parameterization scheme (Kain and Michael Fritsch 1990). The boundary layer scheme comes from the Yonsei University scheme (Hong et al. 2006). The shortwave and longwave radiation schemes are the Dudhia shortwave schemes (Dudhia 1989) and the Rapid Radiative Transfer Model (Mlawer et al. 1997).

A total of 63 perturbations files (12-h minus 24-h WRF forecasts) generated for 1 month from 30 May to 30 June are calculated as inputs to generate the **B** statistics using the NMC method. The average simulated cloud water paths (Fig. 1c), which represent the amount of Qc and Qi in an integrated column, show that the distribution of clouds is generally consistent with the CWP observation (Fig. 1b). For comparison, **B**$_{clear}$ based on the samples under the clear skies is also calculated.

b. Diagnostics of **B** over clear and cloudy areas

The variance part of **B** weights the importance of background in the assimilation. In this subsection, SDs in physical space and eigenvalues in spectral space of BE_CLR and BE_CLD are compared, respectively.

The SD vertical profiles of background errors (BE_CLR and BE_CLD) for the nonhydrometeor and hydrometeor CVs are shown in Fig. 2. For wind variables (Ψ and X_u, Figs. 2a,b), there are two extremes located around 5th level (925 hPa) and 35th level (200 hPa) meaning that the WRF model has a larger error on these levels. The lower-level extreme values may be influenced by the boundary layer where turbulent motion is strong because of the complex topography and surface type; the upper-level extreme values may be related to the westerly jet stream, which has high wind speed, strong shear, and high variability with the season, resulting in large simulation errors. The temperature (Fig. 2c) also shows two error extremes, while the relative humidity (Fig. 2d) SDs peak in the midtroposphere. The vertical profiles of SDs for hydrometeors are positively correlated to the distribution of those hydrometeors themselves (not shown). It should be noted that the SDs of the precipitable hydrometeors (Q_r and Q_s) are larger than those of the nonprecipitable hydrometeors (Q_c and Q_i). This may be related to the magnitude of the precipitable hydrometeors themselves, as precipitation has a larger value than clouds. Besides, the SDs of the BE_CLD are larger than those of the BE_CLR for all the CVs. It is indicated that the errors of all the CVs in cloudy areas will increase. Compared to BE_CLR, the SDs in BE_CLD for nonhydrometeor CVs increase by ~1.5 times while those for hydrometeor CVs increase by ~2 times. The analysis field should be closer to the observation when the background error is large. Therefore, the larger SDs in the cloudy area will reduce weight in the background and increase weight in observation during the minimization of the cost function.

The eigenvalue is another critical parameter that reflects the errors but in spectral space, which provides other insights
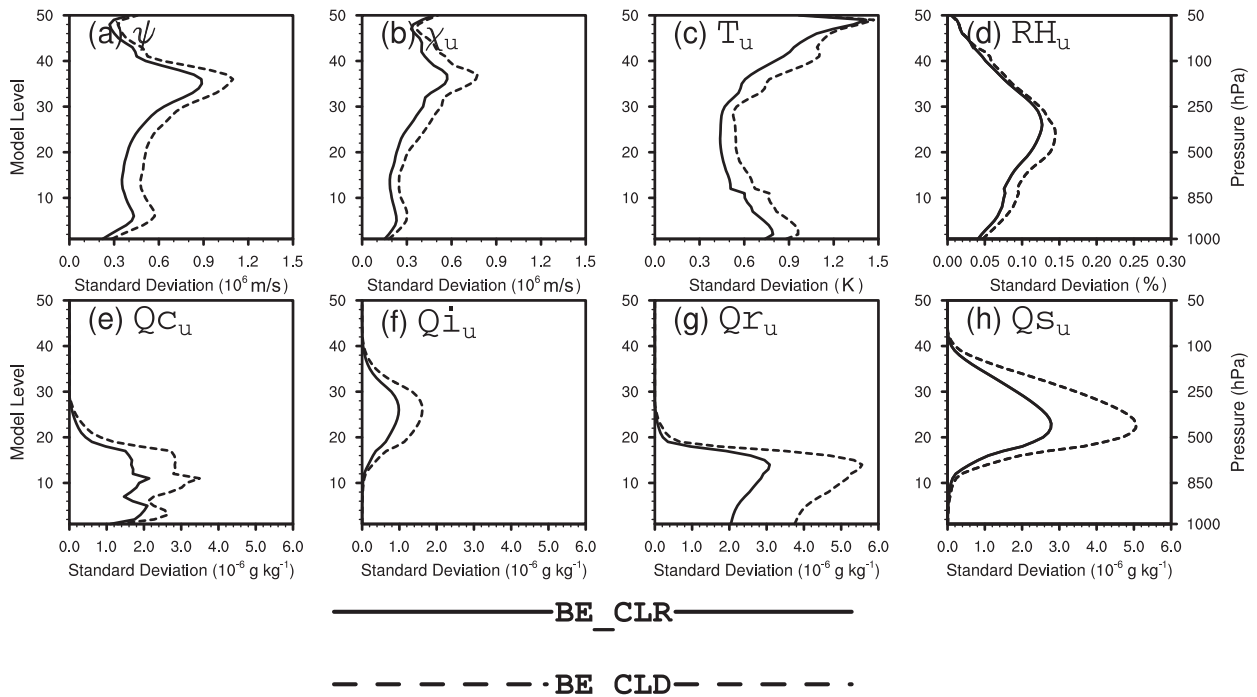
FIG. 2. Vertical profile of error standard deviations for (a) Ψ, (b) X_u, (c) T_u, (d) RH_u, (e) Qc_u, (f) Qi_u, (g) Qr_u, and (h) Qs_u as a function of model level for BE_CLR (solid lines) and BE_CLD (dashed lines). Standard deviations represent the model background errors in physical space, and a larger standard deviation indicates a larger model background error.

besides the above comparison of the BE_CLR and the BE_CLD in physical space. The eigenvalues for the nonhydrometeor and hydrometeor CVs are shown in Fig. 3. They are plotted with respect to the EOF mode index, which represents the variance each EOF model explains. It is clear that the eigenvalues of all the CVs decrease as the EOF mode index increases, which means the smaller the index is, the more variance is explained. The eigenvalues of the BE_CLD are larger than those of the BE_CLR both for nonhydrometeor (Figs. 3a–d) and hydrometeor (Figs. 3e–h) variables which is consistent with the variance characteristics found in physical space. Compared to BE_CLR, the eigenvalues of BE_CLD are increased by ~1.5 times for the nonhydrometeor variables, while they are increased by ~3 times for the hydrometeor CVs, except for Q_i.

Another notable feature is that the eigenvectors for BE_CLR and BE_CLD are almost identical (not shown), indicating that the vertical error distributions of the CVs have little dependence on the clouds after EOF decomposition. However, their vertical autopropagation (indicated by the autocovariance, not shown) is also enhanced over the cloudy areas.

From the above comparisons, it can be seen that the errors vary depending on the weather types. It further confirms the necessity of using **B** separately over clear and cloudy areas. Overall, the errors for nonhydrometeor CVs are increased about 1.5 times in cloudy areas, while those for hydrometeor CVs are increased even more, up to 3 times. Therefore, according to Eq. (10), I_u will be set as 1.5 for the nonhydrometeor CVs and 3 for the hydrometeor CVs, respectively. It is worth noting that vertical cloud dependency has not been considered

in the current adjustment scheme due to the lack of vertical information in the CWP observations, which will be updated in future work.

4. Single-observation experiments using cloud-dependent B

Pseudo-observations are assimilated to understand the response of the assimilation system to the cloud-dependent **B**. As shown in Fig. 4a, a set of pseudo–temperature observations are placed on the latitude line (32°N) with a 1.5°-longitude interval, crossing the region of the thickest cloud (represented by CI) at 0000 UTC 1 July 2016. At the 10th model layer, the innovation is assigned to 1.0 K with an observation error of 1.0 K for each temperature observation. Three experiments are conducted with assimilating each single pseudo–temperature observation independently. No adjustment is made to **B** in cloudy areas in the first experiment (hereafter, No_Adjust). Adaptive adjustment is made to **B** according to CI but without the mask boundary blurring in the second experiment (hereafter, Adjust_no_Blur). The value of **B** is adjusted adaptively according to CI with the mask boundary blurring (hereafter, Adjust_Blur) in the final experiment.

The temperature (T) increments along the west–east, which go through each single observation point on 32°N, are shown in Figs. 4b–d. Analysis increment from each observation in the No_Adjust experiment (Fig. 4b) has the unimodal shapes with the same value because of the isotropic and homogeneous **B**. The increments in cloudy areas are significantly larger than
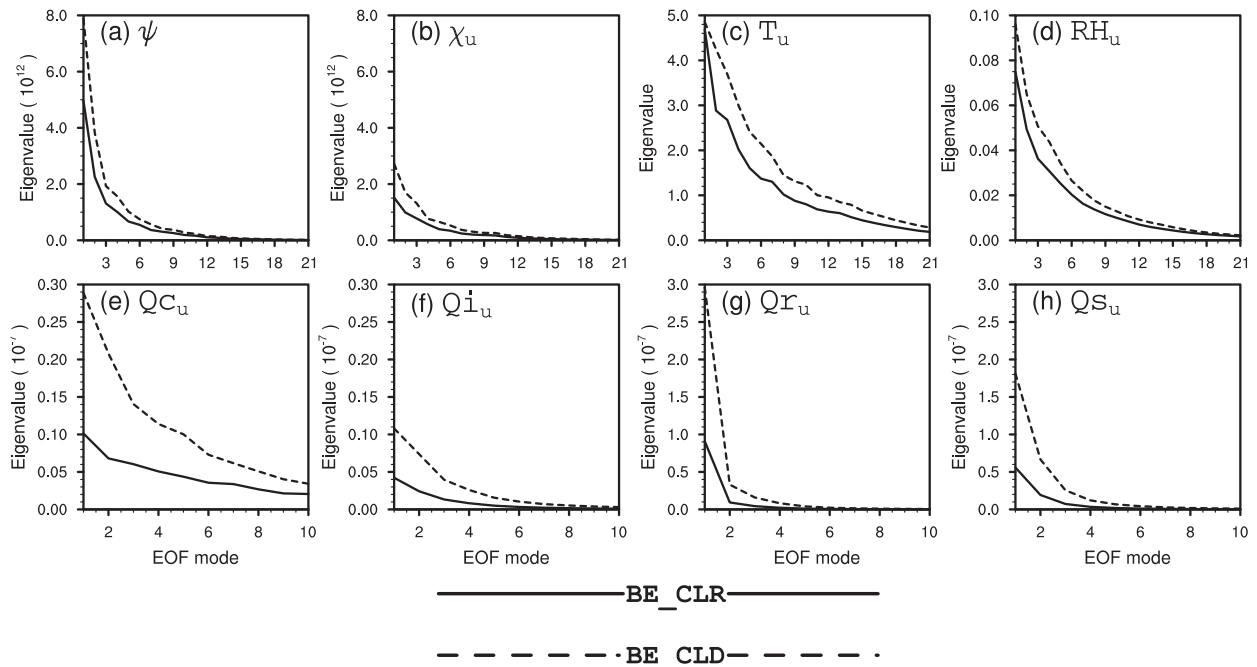
FIG. 3. The eigenvalues calculated as a function of EOF mode for (a) Ψ, (b) X_u, (c) T_u, (d) RH_u, (e) Qc_u, (f) Qi_u, (g) Qr_u, and (h) Qs_u for BE_CLR (solid lines) and BE_CLD (dashed lines). The eigenvalues represent the model background errors in spectral space. The larger the eigenvalue, the greater the model error.

those under the clear skies when **B** is adaptively adjusted with CI in the Adjust_no_Blur experiment (Fig. 4c). However, there are very sharp changes at the cloud/clear border, resulting in discontinuities in this area. These discontinuities are alleviated in the Adjust_Blur experiment (Fig. 4d) by applying the cloud-dependent **B** with the blurred cloud mask border. The **B** can be adjusted smoothly from the clear skies to the cloudy skies, moderating the horizontal gradient at the cloud–clear boundary.

The temperature and Qc increments in horizontal and vertical for three representative points (A: clear sky; B: cloud mask boundary; C: cloudy sky) are shown in Fig. 5. Three approximately circular temperature increments with the same magnitude and scale range are obtained in the No_Adjust experiment (Fig. 5a). They also have almost the same vertical increments, except for point A, which is influenced by the terrain (Fig. 5b). The same horizontal and vertical increments are attributed to the use of isotropic and homogeneous **B**. Similar Qc increments are also created via the covariance between temperature and Qc. In the Adjust_no_Blur experiment, the increments in cloudy areas (C points) are larger than those in the clear skies (A points) both in temperature and Qc. However, the apparent discontinuities exist in the cloud–clear boundary (B points). The Adjust_Blur experiment achieves larger increments in cloudy areas while mitigates the cloud–clear boundary discontinuities and generates some cloud dependency (C point).

5. Impacts of cloud-dependent B on the precipitation forecast

The mei-yu from 20 June to 10 July 2016 is selected for the 3-week cycling data assimilation and forecast experiments to

evaluate the impact of the cloud-dependent **B** on real cases. Numerous continuous heavy rainfall processes occurred during the mei-yu period in 2016.

a. Experiment configurations

The model grid and physics configuration consistent with the **B** statistic sample dataset are used for the evaluation experiments (section 2c). Besides, a 3-km nested grid (D02) is added for more accurate forecasting. To facilitate understanding of the workflow of the cloud-dependent **B**, Fig. 6 illustrates the update of cloud-dependent **B** in the cycled assimilation. As discussed above, CWP is used to generate the normalized CI, which is updated in real time. Then, **B** will be adaptively adjusted in cloudy areas according to CI during the assimilation. The cloud dependency and adjusted value are constant in iterations for a cycle.

To evaluate the impacts of the cloud-dependent **B,** two experiments are carried out. The isotropic and homogeneous **B** without adjustment in cloudy areas for the first experiment (hereafter ''EXP_AVE''), while the second experiment uses the cloud-dependent **B** with the blurred cloud/clear border (hereafter ''EXP_CLD''). Both experiments assimilate the conventional observations datasets, including upper air and surface weather observations (Fig. 1d) and LWP and IWP produced by the NASA LaRC. The observation operators for LWP and IWP developed by Chen et al. (2015) for the WRFDA system are used to assimilate the cloud products.

For both experiments, a total of 157 partial cycling analyses and 12-h forecasts are conducted from 1800 UTC 20 June to 0600 UTC 10 July 2016 at an interval of 3 h. Due to the cloud's rapid evolution, the 3-hourly rapid update cycle is expected to
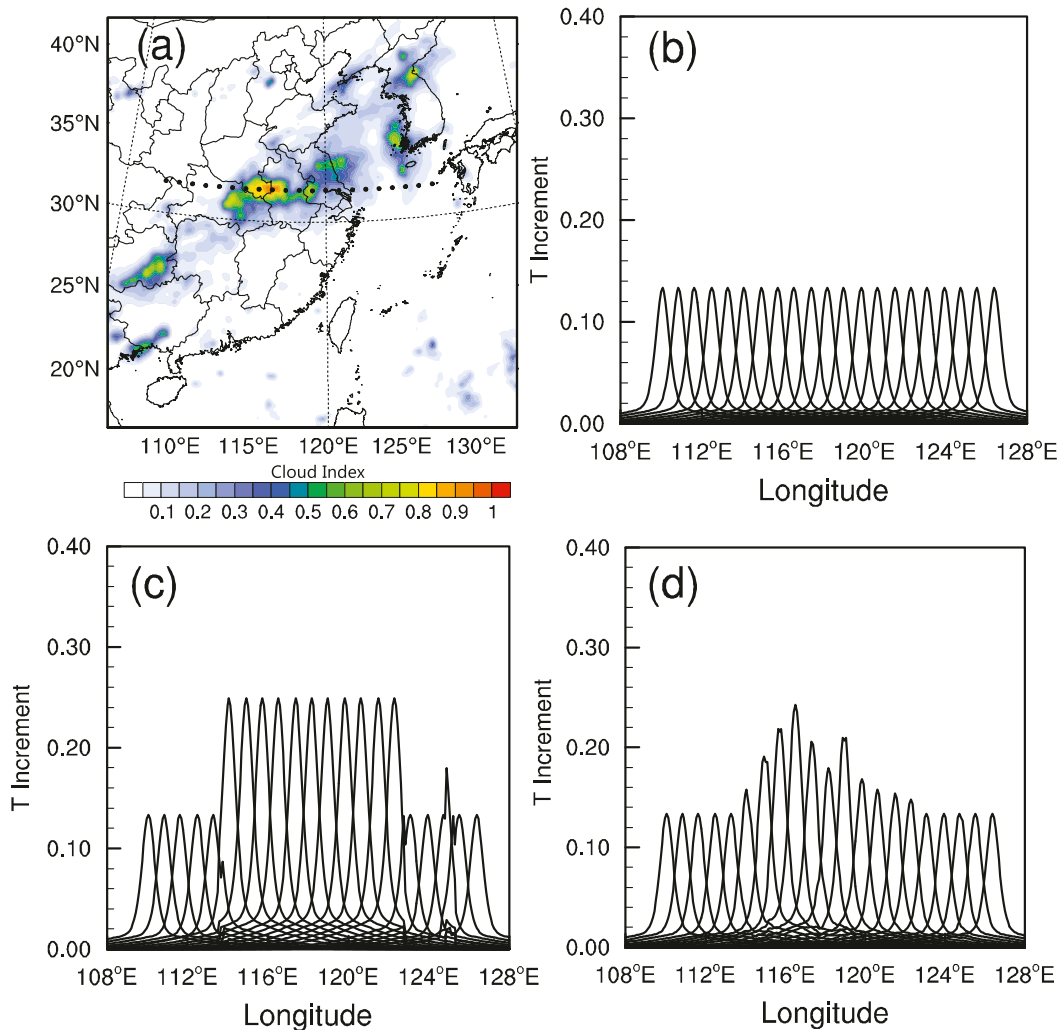
FIG. 4. (a) The cloud index distribution (shaded) calculated from cloud water path at 0000 UTC 1 Jul 2016, and the distribution of the pseudo-observation-point locations (black dots) for single observation tests. Each pseudo-observation point is located at 32°N, with a 1.5°-longitude interval. The temperature increments (unit: K) along the line of latitude (32°N) that go through each single observation point at the 10th level of (b) the No_Adjust experiment, (c) the Adjust_no_Blur experiment, and (d) the Adjust_Blur experiment.

contribute positively to the analysis results. As is shown in Fig. 7, each partial cycling starts at 1800 UTC with the background field taken from a 6-h spinup initialized at 1200 UTC every day. It should be noted that all the assimilations are only carried out at D01.

b. Verification against radiosonde observations

Based on all available radiosonde observations, the average BIAS of analysis and 12-h forecast during the cycling period is calculated to evaluate the two experiments. Figure 8 shows the bias profiles for the prognostic variables, U-wind (U), V-wind (V), temperature (T), and specific humidity (Q), with respect to vertical level. The wind analysis in the EXP_CLD experiment (Figs. 8a,c) fits the wind observations more than that in the EXP_AVE experiment above 850 hPa, especially for V (Fig. 8c), which could result in

more accurate moisture advection and vertical transport. In contrast to the EXP_AVE experiment, the improvement for temperature (Fig. 8e) in the EXP_CLD experiment is only achieved in the lower troposphere. The improvement in specific humidity is pronounced with considerably smaller moisture bias at the lower troposphere in the EXP_CLD experiment (Fig. 8g). After a 12-h forecast, the improvement on wind fields (Figs. 8b,d) and temperature (Fig. 8f) become neutral, showing that the impact of cloud-dependent **B** is limited on 12-h forecasts for wind and temperature. For humidity, most levels above 925 hPa have a smaller bias (Fig. 8h) in the EXP_CLD experiment, indicating that the humidity field is more sensitive to the cloud-dependent **B**.

In addition to the bias, the average root means square error (RMSE) is also calculated to assess the errors in the analysis and 12-h forecast fields. Figure 9 shows the RMSE profiles for
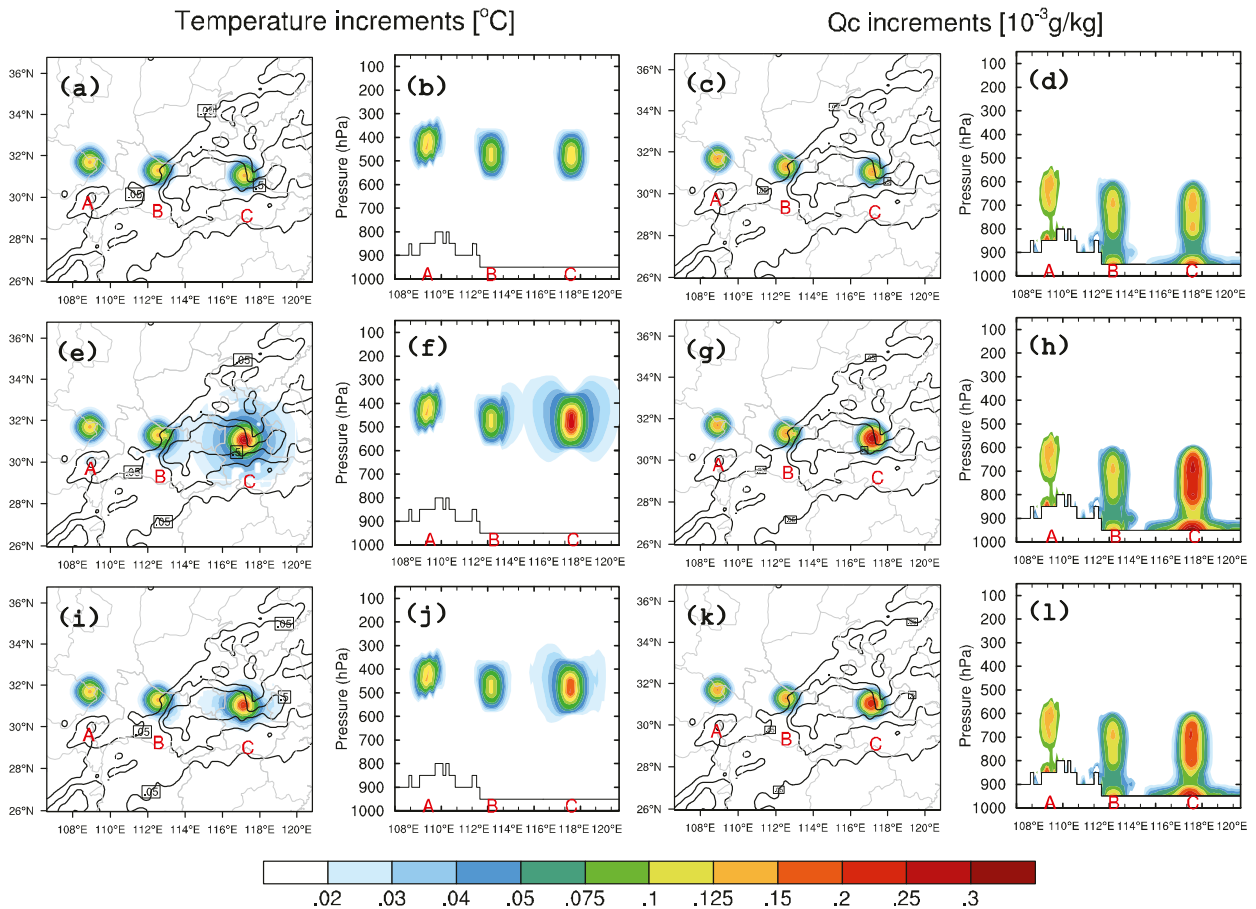
Temperature increments [°C]

Qc increments [10⁻³g/kg]

FIG. 5. (a),(e),(i) The temperature increments (shaded; unit: K) at the 10th level as a result of assimilating three pseudo–temperature observations located at A (clear-sky area), B (edge of cloudy area), and C (cloudy area), respectively. Contours indicate the CI. (b),(f),(j) The vertical cross sections of the temperature increment is along 32°N. (c),(g),(k) and (d),(h),(l) As in (a),(e),(i) and (b),(f),(j), respectively, but for Qc increments. The results are from (a)–(d) the No_Adjust experiment, (e)–(h) the Adjust_no_Blur experiment, and (i)–(l) the Adjust_Blur experiment.

the same prognostic variables (U, V, T, and Q). There is a larger RMSE for the 12-h forecast fields than the analysis fields for the four variables, but no obvious difference is found between the EXP_CLD and the EXP_AVE experiments. Little differences may be explained by the fact that the radiosonde observations used for the verification are too sparse compared to the small-scale cloud areas, resulting in slight differences in RMSE between the two experiments.

The bias and RMSE results demonstrate that the EXP_CLD experiment with the cloud-dependent **B** is overall superior to EXP_AVE with the static **B**. It is expected that the cloud-dependent **B** gives higher weights to observations in cloudy areas and reflects the dynamic changes more appropriately, making the analysis and forecast more accurate.

c. Rainfall forecast skill scores

Three rainfall verification metrics are applied to evaluate the impact of the cloud-dependent **B** on precipitation forecasts. The first verification metric is the equitable threat score (ETS), a measurement of the portion at which the observed

event is correctly predicted, effectively removing the effect of random precipitation on the score. The ETS varies from −1/3 to 1. The higher the ETS score, the better the forecasting skill, and a value of 0 or negative means no forecasting skill. The second method is the False Alarm Ratio (FAR), which measures the ratio of the false alarm number to the total predicted events number. FAR = 1 indicates 100% false prediction, while FAR = 0 indicates the forecast captures all the observed rainfall for the given magnitude. The third method is the probability of detection (POD), which is used to measure the percentage of correctly forecast events. The range of POD is 0–1 with a perfect value of 1. The observation from the China Hourly Merged Precipitation Analysis (CHMPA) with 0.05° × 0.05° spatial resolution (Pan et al. 2018 in Chinese) is used as a reference for the D02 (Fig. 1d) Rainfall skill scores calculation.

Figure 10 shows the ETS, FAR, and POD scores of 6- and 12-h accumulated precipitation for thresholds commonly used by the operational centers in China (Figs. 10a–f) and the improvement ratios of the EXP_CLD experiment compared to the EXP_AVE experiment (Figs. 10g,f). Compared to the
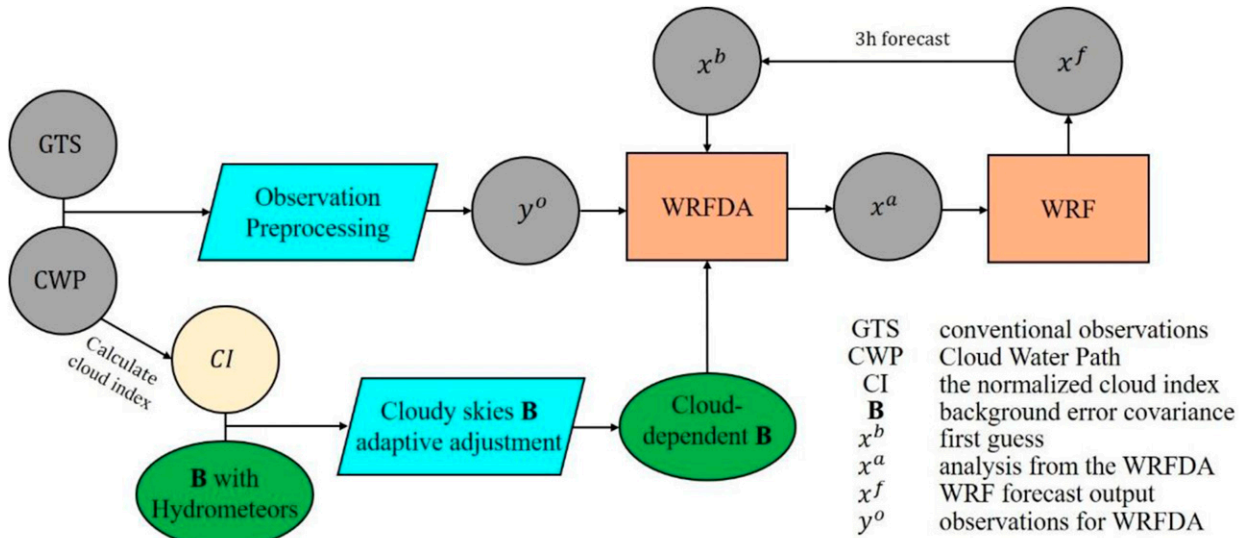
FIG. 6. Schematic diagrams for the update of the cloud-dependent **B**.

EXP_AVE experiment, the ETS in the EXP_CLD experiment is improved clearly at almost all verification thresholds except for a slight deterioration (less than 1%) in the moderate rainfall category in 6 and 12-h forecast. The average improvement in ETS is about 4% for the 6-h forecast and 3% for the 12-h forecast. With increasing thresholds of scores, POD decreases gradually in both 6-h (Fig. 10e) and 12-h (Fig. 10f) results, but the POD of the EXP_CLD is better than the EXP_AVE overall. The improvement ratios in POD share similar ETS patterns, but their values are more pronounced, especially for extreme heavy rain. The improvement ratios in both ETS and POD gradually increase as the verification magnitude increases. It indicates that the cloud-dependent **B** application can effectively improve precipitation forecasting skills for heavy and extremely heavy rainfalls. However, there is a minor difference in the FAR between the two experiments, with only a slight improvement in the 6-h forecast.

Hourly ETS, FAR, and POD are also calculated to hourly accumulated precipitation for light rainfall (with 1-mm threshold values) and heavy rainfall (with 2-mm threshold values). For the 1 mm threshold, the ETS metric (Fig. 11a dashed lines) shows that the EXP_CLD experiment is better than the

EXP_AVE experiment for the first 9 forecast hours. For the 2-mm threshold, even though ETS is lower than the light rainfall scores, the positive impacts are still obtained in the EXP_CLD experiment for the first 9 forecast hours. For the FAR, the result for the 1-mm threshold is significantly lower than for the 2 mm, but the two experiments are comparable. The improvement in POD mainly contributed from the first 6 h, after which the differences between the two experiments are minimal. The results for other larger thresholds are similar to the patterns of 1- and 2-mm thresholds, but with different values (not shown).

Figures 10 and 11 show that cloud-dependent **B** can effectively improve the precipitation scores in the first 6–9 h. The cloud-dependent **B** used in the EXP_CLD experiment allows more critical information in cloudy areas, especially cloud observations, to be assimilated to improve the initial fields, leading to more accurate precipitation forecasts.

6. Diagnostics for an extremely heavy rainfall event

To further investigate how the cloud-dependent **B** contributes to analysis and forecast, an extremely heavy rainfall event
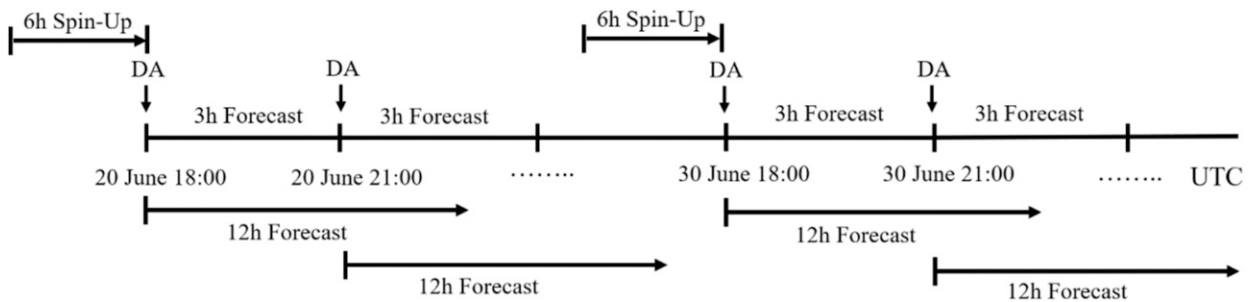

FIG. 7. Flowchart of the partial data assimilation experiments. "DA" means data assimilation.

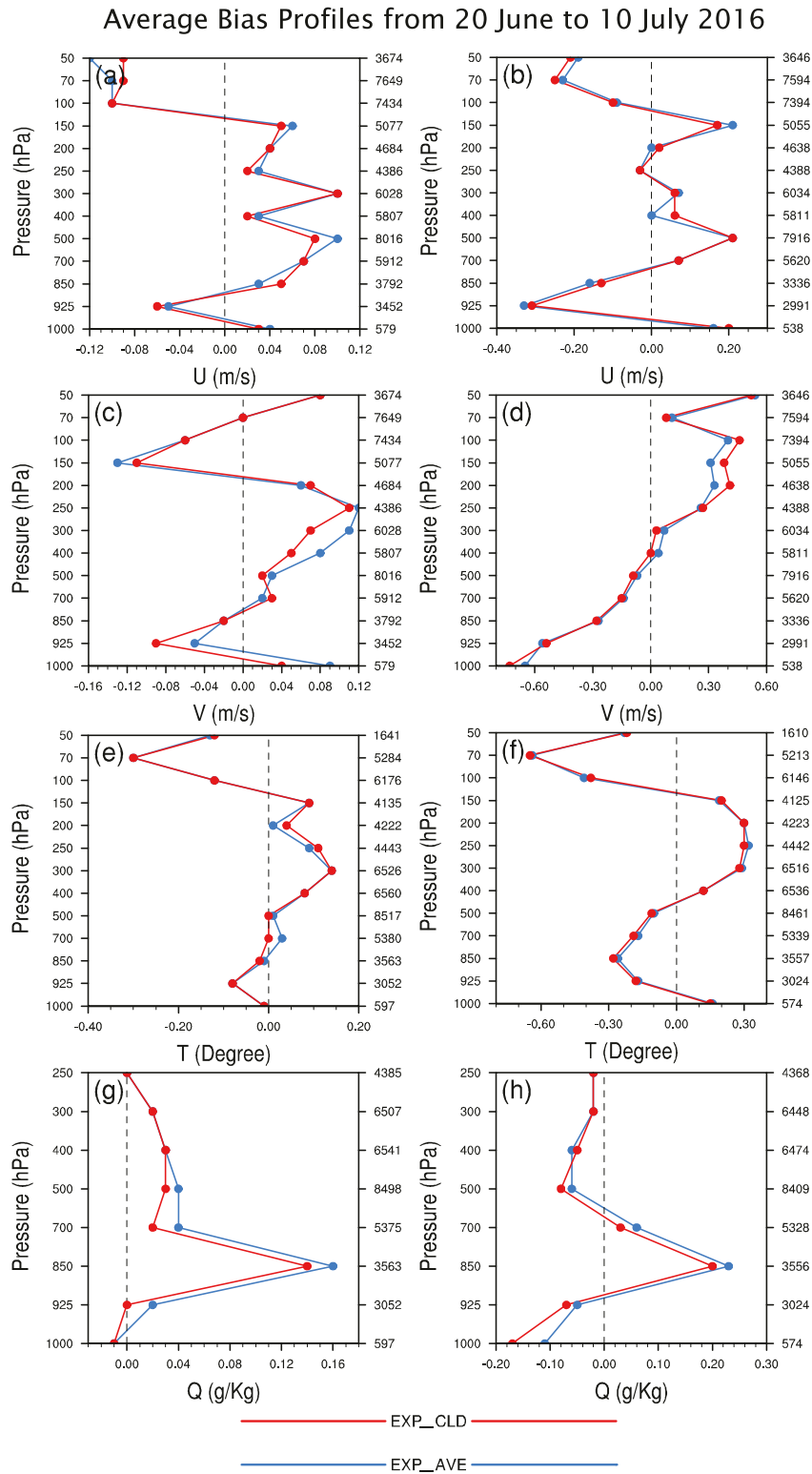# Average Bias Profiles from 20 June to 10 July 2016

FIG. 8. The average BIAS vertical profiles of (a),(c),(e),(g) analysis and (b),(d),(f),(h) 12-h forecasts verified against radiosonde observations for the EXP_AVE (blue lines) and EXP_CLD (red lines) experiments. (a),(b) U wind (unit: $m\,s^{-1}$), (c),(d) V wind (unit: $m\,s^{-1}$), (e),(f) temperature (unit: °C), and (g),(h) specific humidity (unit: $g\,kg^{-1}$). The right ordinates represent the total number of observations used for verification from 1800 UTC 20 Jun to 0600 UTC 10 Jul 2016.

RMSE Profiles 20 June - 10 July 2016
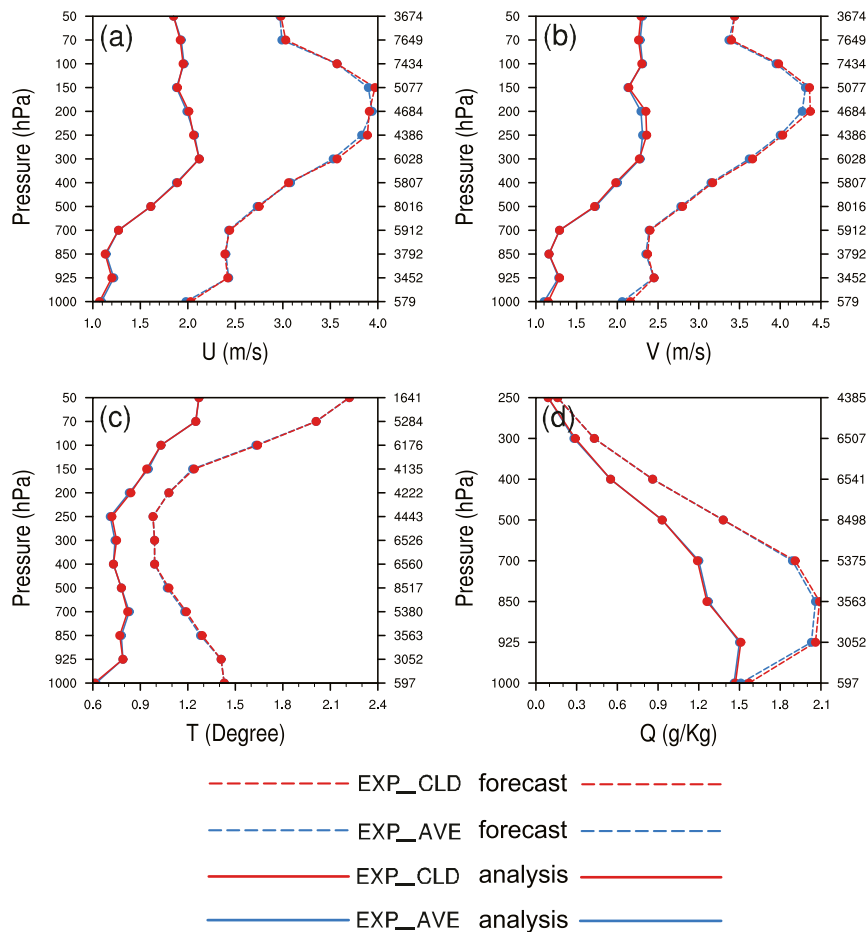
FIG. 9. The average RMSE vertical profiles verified against radiosonde observations for the EXP_AVE (blue lines) and EXP_CLD (red lines) experiments for (a) U wind (unit: m s^{-1}), (b) V wind (unit: m s^{-1}), (c) temperature (unit: °C), and (d) specific humidity (unit: g kg^{-1}). The right ordinates represent the total number of observations used for verification from 1800 UTC 20 Jun to 0600 UTC 10 Jul 2016. The solid line represents 6-h forecast results, and the dashed line represents 12-h forecast results.

from 30 June to 1 July 2016 is chosen for diagnosis. The water vapor is transported along the edge of the subtropical high to the Yangtze–Huaihe River valley, where it rises in a convergence zone. As the southwest vortex (SWV) moves eastward along the low shear line (700 hPa), the water vapor convergence is enhanced (not shown), resulting in this extremely heavy rainfall. In other words, sufficient water vapor and strong upward motion dominated this heavy precipitation event.

a. 12-h accumulated precipitation

The 12-h accumulated precipitation displayed in Fig. 12 is based on the results of the subdomain D02. It can be seen that the observations from the CHMPA show a predominantly east–west band, and the heavy rainfall centers are divided into two areas (A and B). Both EXP_AVE and EXP_CLD experiments perform well for the rainfall forecast in area A. However, the EXP_AVE experiment has a

negative simulation bias at B point compared to the observations. Besides, the simulated rainfall in northeast of the major rainband (marked by C) from the EXP_AVE experiment is stronger than the observations, reaching the torrential rainfall category. As a comparison, the maximum rainfall center in area A is well captured with the rainfall centers in area B improved in the EXP_CLD experiment, and the rainfall magnitude is more consistent with the observations. Besides, the false torrential rainfall in the C area is also reduced.

b. Humidity, wind, and hydrometeor analysis and forecasts

Cross sections of the relative humidity and cloud content (Qc + Qi) analysis along 117.5°E (shown in Fig. 12a dashed line) at 1800 UTC 30 June 2016, and the associated 3h forecasts (additionally add 2D wind vectors) are shown in Fig. 13. The relative humidity in lower layers (below 700 hPa) is enhanced
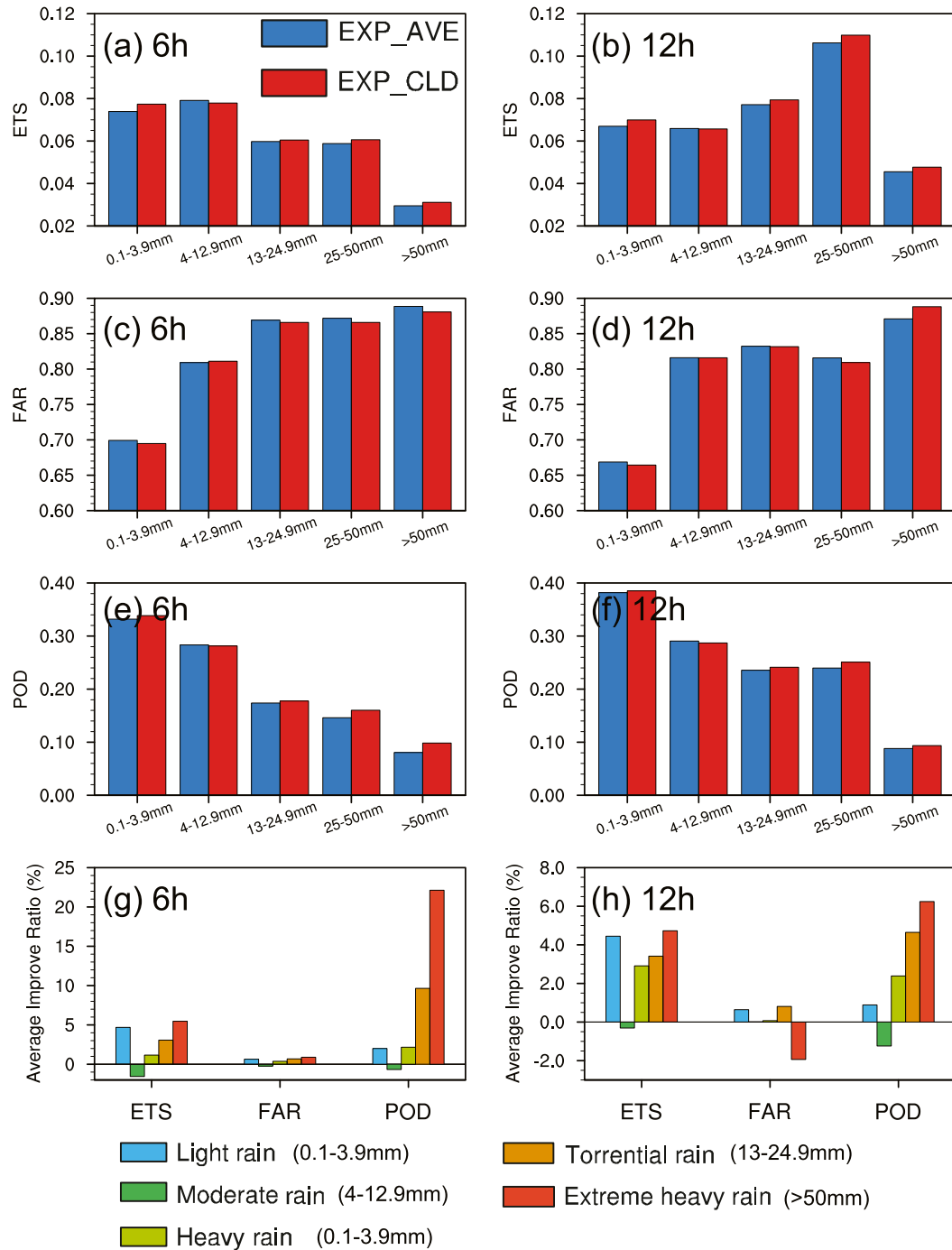
FIG. 10. Rainfall forecast scores for (a),(c),(e) 6- and (b),(d),(f) 12-h accumulated rainfall according to operational precipitation standard thresholds for (a),(b) equitable threat score (ETS); (c),(d) false alarm ratio (FAR); and (e),(f) probability of detection (POD). Average improvements in the percentage of the EXP_CLD experiment compared to the EXP_AVE experiment in different standard thresholds for (g) 6- and (h) 12-h accumulated rainfall.

in the EXP_CLD experiment (Fig. 13c) compared to the EXP_AVE experiment (Fig. 13a) in the precipitation centers. Compared to EXP_AVE, the hydrometeor analyses in the EXP_CLD experiment shows higher cloud

amounts and thicker clouds in the main precipitation area (black dashed box). After the 3-h forecast, moisture conditions over the central rainfall region in the EXP_CLD experiment (Fig. 13d) remain stronger than those in the
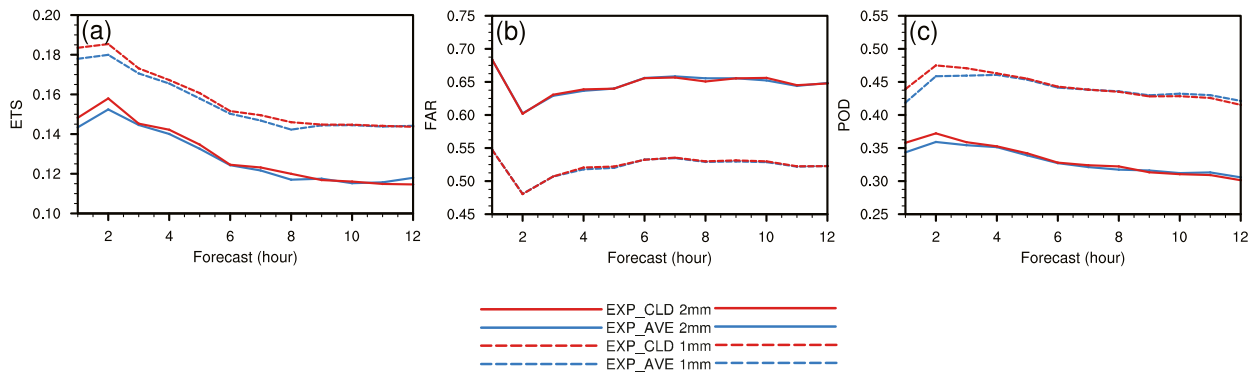
FIG. 11. Rainfall forecast scores with a threshold of 1mm (dashed line) and 2 mm (solid lines) as a function of forecast time for (a) ETS, (b) FAR, and (c) POD.

EXP_AVE experiment (Fig. 13b), with a wider area being saturated. The forecast also produces more hydrometeors over the central rainfall region. The vertical wind circulation in the EXP_CLD experiment is broader and stronger over the central rainfall region, which transports more water vapor from the lower to the upper levels, providing favorable precipitation conditions.

Thus, the cloud-dependent **B** application improves the analysis of moisture and hydrometeors, which produces more favorable conditions for precipitation.

c. Precipitable water and divergence analyses and forecasts

Precipitation magnitudes are closely associated with atmospheric precipitable water and the convergence and divergence of wind. Figure 14 shows the distributions of atmospheric precipitable water differences and divergence at 850 hPa for the analysis and associated 3-h forecast. For the difference in analysis fields, the EXP_CLD experiment is slightly wetter than the EXP_AVE experiment in the main precipitation area. Some convergence of lower-level wind fields is achieved simultaneously. After the 3-h forecast, with the influence of southerly winds (not shown), the region of high precipitable water moves northward. The precipitable water in

the EXP_CLD experiment is enhanced compared to the EXP_AVE experiment, while the convergence of low-level air gradually strengthens with the forecast. The cloud-dependent **B** application improves the analysis fitting to water vapor and wind observations, resulting in stronger precipitable water and greater airflow convergence in the lower layers, favoring the precipitation.

7. Conclusions and discussion

Various studies have pointed out that the complicated physical processes in cloud regions can lead to very different **B** characteristics between the clear and cloudy regions. However, most variational framework assimilation studies remain to use an isotropic and homogeneous **B** without considering the surrounding meteorological conditions.

The horizontally scale-dependent correlations across variables, as noted by Caron and Fillion (2010), motivated the development of heterogeneous **B**. Built on previous works, this study proposes a cloud-dependent **B** with the inclusion of hydrometeor CVs to achieve a better piecewise analysis of nonhydrometeor variables and hydrometeors. The WRFDA system is modified to enable adaptive adjustment
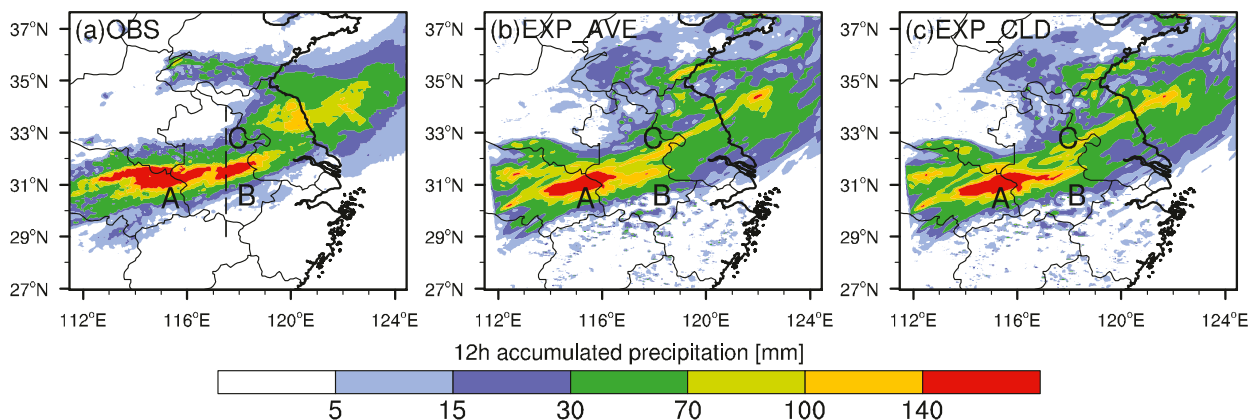

FIG. 12. The 12-h accumulated precipitation beginning at 1800 UTC 30 Jun 2016 for (a) the precipitation observation, (b) the EXP_AVE experiment, and (c) the EXP_CLD experiment. The simulated precipitation results come from the subdomain D02.
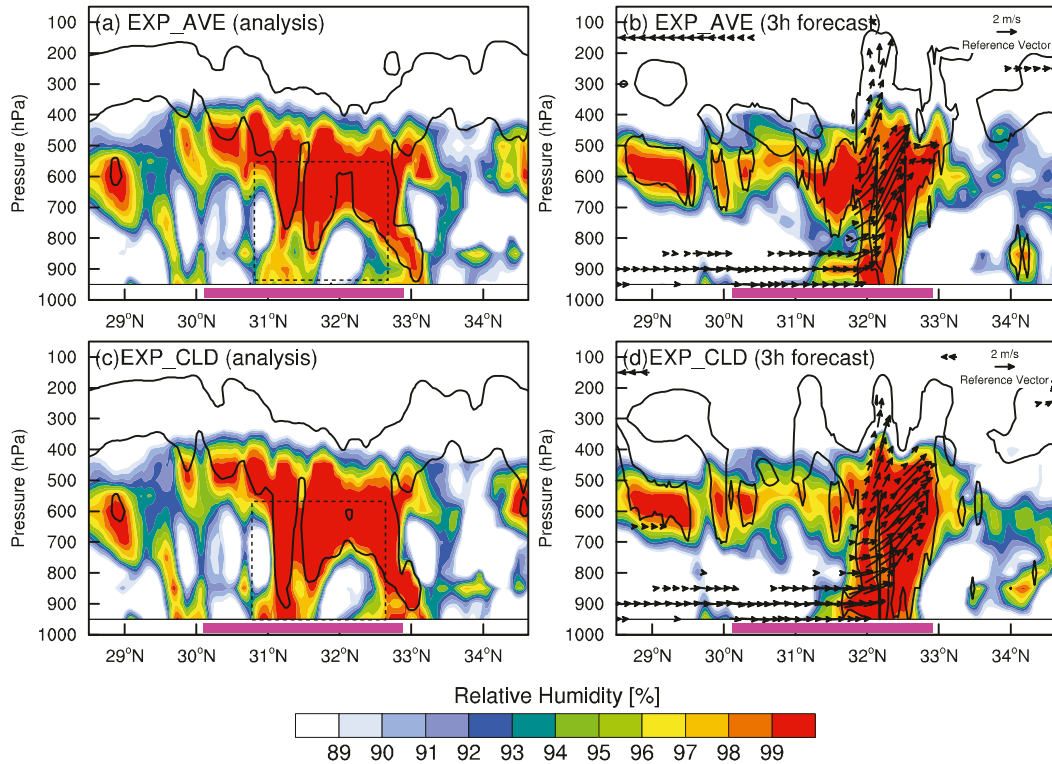
FIG. 13. Cross sections along the dashed line in Fig. 12a of (a),(c) analysis at 1800 UTC 30 Jun 2016 and associated (b),(d) 3-h forecasts for relative humidity (shaded; %), cloud content (contours; hydrometeor boundary defined by a threshold mixing ratio of 0.05 g kg^{-1}), and 2D wind vectors (vectors; units: m s^{-1}; only available in forecasts). (a),(b) The EXP_AVE experiment and (c),(d) the EXP_CLD experiment. The pink rectangle shows the main precipitation area.

of **B** over the cloudy areas according to a 2D CI deduced from satellite-based cloud products. The cloud-dependent **B** for each horizontal grid defined as cloudy can be adaptively adjusted based on the amount of cloudiness and the adjustment coefficients, which are calculated by comparing the characteristics of **B** over clear and cloudy areas. In this way, the analysis is closer to observation where cloud or rainfall occurs, enabling multivariate analysis that includes hydrometeors. A series of single observation tests confirm the advantages of this cloud-dependent piecewise analysis scheme. The multivariate analysis increments show that the cloud-dependent **B** gives higher weightings to temperatures observation in cloudy areas due to the cloud-dependent error statistics.

A 3-week cycling DA and its impact study reveal that the cloud-dependent **B** can be applied to make more efficient use of observations and reduce average bias in analyses and forecasts. The positive impacts can also be found in ETS, FAR, and POD rainfall forecast skill scores, the improvement is primarily contributed by the first 6 h, with main improvements in the ETS and POD scores, leaving little improvement in the FAR. The detailed diagnostics further show that the DA system places more weight on the observations under the cloud skies by using the cloud-dependent **B**, resulting in the enhanced water vapor, wind, and hydrometeors over the central precipitation region. These favorable conditions facilitate the development of precipitation.

Even if EnVar can currently provide real flow-dependent analysis, relying solely on cloud information to achieve cloud-dependent analysis in a variational framework is still a worthy exploration for institutions with insufficient computational resources. The cloud-dependent piecewise analysis proposed in this study is based on the specific cloud and precipitation environment. This method can be extrapolated to different weather environments with local variabilities, such as aerosol assimilation. The flexibility of the CI allows for defining CI on a variable-by-variable basis, which is very important in the combined assimilation of multisource observations. Since the core component in our cloud-dependent scheme is the CI specification, improved CI specification can further improve the analysis and convective forecasting. One way to improve the CI specification is to consider the inclusion of the background hydrometeors, which may mitigate displacement problems, such as for cloudy background with clear observations and vice versa. As discussed in section 2a, adding time dependence to the CI definition is also one of the goals to be explored in future. Future research also can include machine learning, using artificial intelligence (AI) algorithms to provide CI. Besides, the assimilation of high-density satellite radiance with the cloud-dependent **B** and the symmetric observation error model (Geer and Bauer 2011) deserves further investigations.
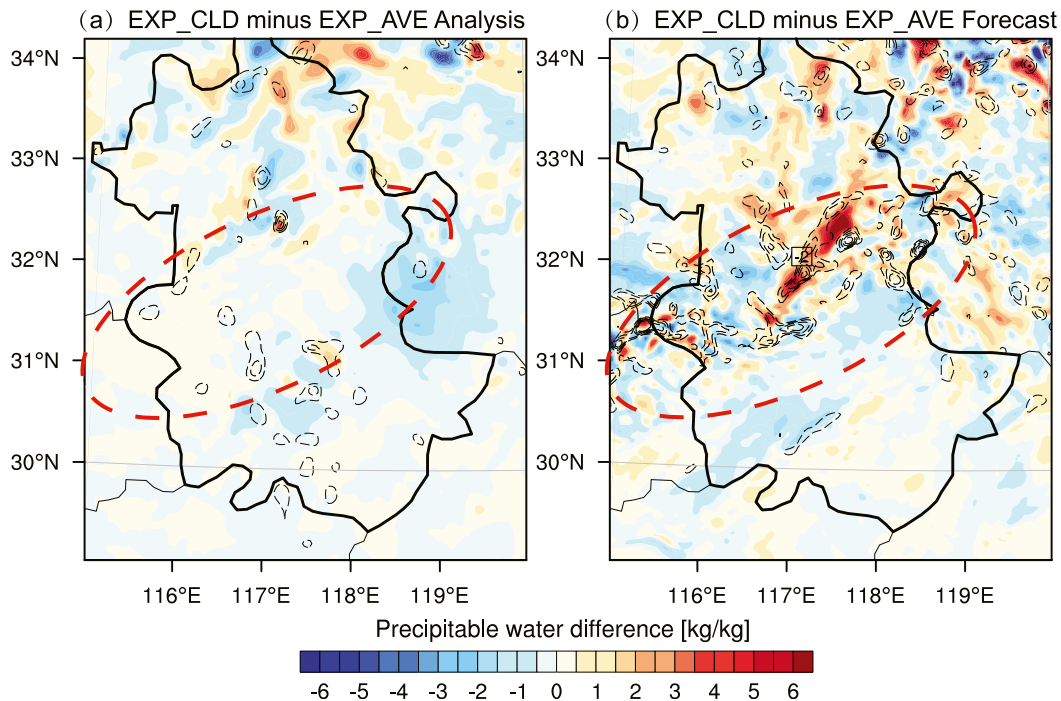
FIG. 14. The differences (EXP_CLD minus EXP_AVE) of precipitable water (shaded; kg kg^{-1}) overlaid with divergence (contours; ranging from -12 to -2 s^{-1}) at 850 hPa over the main precipitation area (Fig. 12, B point) for (a) analysis at 1800 UTC 30 Jun 2016 and associated (b) 3-h forecasts. The red dashed box indicates the main rainfall area that occurred in Anhui Province.

REFERENCES

Anderson, J. L., 2012: Localization and sampling error correction in ensemble Kalman filter data assimilation. *Mon. Wea. Rev.*, **140**, 2359–2371, https://doi.org/10.1175/MWR-D-11-00013.1.

Aranami, K., and Coauthors, 2015: A new operational regional model for convection-permitting numerical weather prediction at JMA. *CAS/JSC WGNE Res. Act. Atmos. Oceanic Model.*, **45**, 505–506.

Bannister, R. N., 2008a: A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error co-variances. *Quart. J. Roy. Meteor. Soc.*, **134**, 1951–1970, https://doi.org/10.1002/qj.339.

——, 2008b: A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. *Quart. J. Roy. Meteor. Soc.*, **134**, 1971–1996, https://doi.org/10.1002/qj.340.

——, 2017: A review of operational methods of variational and ensemble-variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **143**, 607–633, https://doi.org/10.1002/qj.2982.

Barker, D., and Coauthors, 2012: The Weather Research and Forecasting Model's Community Variational/Ensemble Data Assimilation System: WRFDA. *Bull. Amer. Meteor. Soc.*, **93**, 831–843, https://doi.org/10.1175/BAMS-D-11-00167.1.

Brousseau, P., L. Berre, F. Bouttier, and G. Desroziers, 2011: Background-error covariances for a convective-scale data-assimilation system: AROME-France 3D-Var. *Quart. J. Roy. Meteor. Soc.*, **137**, 409–422, https://doi.org/10.1002/qj.750.

Caron, J.-F., and L. Fillion, 2010: An examination of background error correlations between mass and rotational wind over precipitation regions. *Mon. Wea. Rev.*, **138**, 563–578, https://doi.org/10.1175/2009MWR2998.1.

Chen, Y., S. R. H. Rizvi, X.-Y. Huang, J. Min, and X. Zhang, 2013: Balance characteristics of multivariate background error co-variances and their impact on analyses and forecasts in tropical and Arctic regions. *Meteor. Atmos. Phys.*, **121**, 79–98, https://doi.org/10.1007/s00703-013-0251-y.

——, H. Wang, J. Min, X.-Y. Huang, P. Minnis, R. Zhang, J. Haggerty, and R. Palikonda, 2015: Variational assimilation of

cloud liquid/ice water path and its impact on NWP. *J. Appl. Meteor. Climatol.*, **54**, 1809–1825, https://doi.org/10.1175/JAMC-D-14-0243.1.

——, R. Zhang, D. Meng, J. Min, and L. Zhang, 2016: Variational assimilation of satellite cloud water/ice path and microphysics scheme sensitivity to the assimilation of a rainfall case. *Adv. Atmos. Sci.*, **33**, 1158–1170, https://doi.org/10.1007/s00376-016-6004-3.

Courtier, P., and Coauthors, 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quart. J. Roy. Meteor. Soc.*, **124**, 1783–1807, https://doi.org/10.1002/qj.49712455002.

Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.

Descombes, G., T. Auligné, F. Vandenberghe, D. M. Barker, and J. Barré, 2015: Generalized background error covariance matrix model (GEN_BE v2.0). *Geosci. Model Dev.*, **8**, 669–696, https://doi.org/10.5194/gmd-8-669-2015.

Destouches, M., T. Montmerle, Y. Michel, and B. Ménétrier, 2020: Estimating optimal localization for sampled background error covariances of hydrometeor variables. *Quart. J. Roy. Meteor. Soc.*, **147**, 74–93, https://doi.org/10.1002/qj.3906.

Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107, https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2.

Errico, R. M., P. Bauer, and J.-F. Mahfouf, 2007: Issues regarding the assimilation of cloud and precipitation data. *J. Atmos. Sci.*, **64**, 3785–3798, https://doi.org/10.1175/2006JAS2044.1.

Geer, A. J., and P. Bauer, 2011: Observation errors in all-sky data assimilation. *Quart. J. Roy. Meteor. Soc.*, **137**, 2024–2037, https://doi.org/10.1002/qj.830.

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter–3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919, https://doi.org/10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2.

Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, https://doi.org/10.1175/MWR3199.1.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811, https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

Ingleby, N. B., A. C. Lorenc, K. Ngan, F. Rawlins, and D. R. Jackson, 2013: Improved variational analyses using a nonlinear humidity control variable. *Quart. J. Roy. Meteor. Soc.*, **139**, 1875–1887, https://doi.org/10.1002/qj.2073.

Kain, J. S., and J. M. Fritsch, 1990: A one-dimensional entraining/detraining plume model and its application in convective parameterization. *J. Atmos. Sci.*, **47**, 2784–2802, https://doi.org/10.1175/1520-0469(1990)047<2784:AODEPM>2.0.CO;2.

Li, J., P. Wang, H. Han, J. Li, and J. Zheng, 2016: On the assimilation of satellite sounder data in cloudy skies in numerical weather prediction models. *J. Meteor. Res.*, **30**, 169–182, https://doi.org/10.1007/s13351-016-5114-2.

Lim, K.-S. S., and S.-Y. Hong, 2010: Development of an effective double-moment cloud microphysics scheme with prognostic Cloud Condensation Nuclei (CCN) for weather and climate models. *Mon. Wea. Rev.*, **138**, 1587–1612, https://doi.org/10.1175/2009MWR2968.1.

Lopez, P., 2007: Cloud and precipitation parameterizations in modeling and variational data assimilation: A review. *J. Atmos. Sci.*, **64**, 3766–3784, https://doi.org/10.1175/2006JAS2030.1.

Ménétrier, B., and T. Montmerle, 2011: Heterogeneous background-error covariances for the analysis and forecast of fog events. *Quart. J. Roy. Meteor. Soc.*, **137**, 2004–2013, https://doi.org/10.1002/qj.802.

Meng, D., Y. Chen, H. Wang, Y. Gao, R. Potthast, and Y. Wang, 2019: The evaluation of EnVar method including hydrometeors analysis variables for assimilating cloud liquid/ice water path on prediction of rainfall events. *Atmos. Res.*, **219**, 1–12, https://doi.org/10.1016/j.atmosres.2018.12.017.

Michel, Y., T. Auligné, and T. Montmerle, 2011: Heterogeneous convective-scale background error covariances with the inclusion of hydrometeor variables. *Mon. Wea. Rev.*, **139**, 2994–3015, https://doi.org/10.1175/2011MWR3632.1.

Minnis, P., 2007: Cloud retrievals from GOES-R. *Hyperspectral Imaging and Sounding of the Environment*, Santa Fe, NM, OSA, https://doi.org/10.1364/HISE.2007.HWC3.

——, and Coauthors, 2008: Near-real time cloud retrievals from operational and research meteorological satellites. *Proc. SPIE*, **7107**, 710703, https://doi.org/10.1117/12.800344.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16663–16682, https://doi.org/10.1029/97JD00237.

Montmerle, T., and L. Berre, 2010: Diagnosis and formulation of heterogeneous background-error covariances at the mesoscale. *Quart. J. Roy. Meteor. Soc.*, **136**, 1408–1420, https://doi.org/10.1002/qj.655.

Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415–428, https://doi.org/10.3402/tellusa.v56i5.14462.

Pan, Y., J. Gu, J. Yu, Y. Shen, C. Shi, and Z. Zhou, 2018: Test of merging methods for multi-source observed precipitation products at high resolution over China (in Chinese). *Acta Meteor. Sin.*, **76** (5), 755–766.

Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763, https://doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2.

Skamarock, W. C., and J. B. Klemp, 2008: A time-split non-hydrostatic atmospheric model for weather research and forecasting applications. *J. Comput. Phys.*, **227**, 3465–3485, https://doi.org/10.1016/j.jcp.2007.01.037.

Wang, X., D. M. Barker, C. Snyder, and T. M. Hamill, 2008a: A hybrid ETKF–3DVAR data assimilation scheme for the WRF Model. Part I: Observing system simulation experiment. *Mon. Wea. Rev.*, **136**, 5116–5131, https://doi.org/10.1175/2008MWR2444.1.

——, ——, ——, and ——, 2008b: A hybrid ETKF–3DVAR data assimilation scheme for the WRF Model. Part II: Real observation experiments. *Mon. Wea. Rev.*, **136**, 5132–5147, https://doi.org/10.1175/2008MWR2445.1.

Wang, Y., J. Min, Y. Chen, X.-Y. Huang, M. Zeng, and X. Li, 2017: Improving precipitation forecast with hybrid 3DVar and time-lagged ensembles in a heavy rainfall event. *Atmos. Res.*, **183**, 1–16, https://doi.org/10.1016/j.atmosres.2016.07.026.

——, ——, and X.-Y. Huang, 2018: Impact of the hybrid gain ensemble data assimilation on meso-scale numerical weather prediction over east China. *Atmos. Res.*, **206**, 30–45, https://doi.org/10.1016/j.atmosres.2018.02.014.

Wu, W.-S, D. F. Parrish, E. Rogers, and Y. Lin, 2017: Regional ensemble–variational data assimilation using global ensemble forecasts. *Wea. Forecasting*, **32**, 83–96, https://doi.org/10.1175/WAF-D-16-0045.1.

Yuan, Y., H. Gao, W. Li, Y. Liu, L. Chen, B. Zhou, and Y. Ding, 2017: The 2016 summer floods in China and associated physical mechanisms: A comparison with 1998. *J. Meteor. Res.*, **31**, 261–277, https://doi.org/10.1007/s13351-017-6192-5.