Received: 14 September 2022 | Revised: 17 May 2023 | Accepted: 4 June 2023

DOI: 10.1002/qj.4507

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Spatiotemporal estimation of analysis errors in the operational global data assimilation system at the China Meteorological Administration using a modified SAFE method

Jie Feng[1,2] | **Jincheng Wang**[3,4] | **Guokun Dai**[1,2,5] | **Feifan Zhou**[6,7] | **Wansuo Duan**[8]

[1]Department of Atmospheric and Oceanic Sciences and Institute of Atmospheric Sciences, Fudan University, Shanghai, China

[2]Shanghai Frontiers Science Center of Atmosphere-Ocean Interaction, Shanghai, China

[3]CMA Earth System Modeling and Prediction Centre, China Meteorological Administration, Beijing, China

[4]State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing, China

[5]CMA-FDU Joint Laboratory of Marine Meteorology, Shanghai, China

[6]Key Laboratory of Cloud–Precipitation Physics and Severe Storms, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

[7]University of Chinese Academy of Sciences, Beijing, China

[8]LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Bejjing, China

Correspondence
Jincheng Wang, CMA Earth System Modeling and Prediction Centre, China Meteorological Administration, Beijing, China.
Email: wangjc@cma.gov.cn

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 12241103, 42105054; China Postdoctoral Science Foundation, Grant/Award Number: 2022M723468

Abstract

Quantification of the uncertainties in initial analyses against the real atmosphere ("reality") provides a fundamental reference for the evaluation and development of operational data assimilation (DA) systems. Due to the unknown reality, most existing methods for analysis error estimation use reanalysis datasets or observations as a proxy for reality, which are empirical, non-objective, and biased. Unlike these methods, our study adopted a modified Statistical Analysis and Forecast Error (SAFE) estimation method to objectively and directly quantify spatiotemporal errors in analyses compared to reality based on unbiased assumptions. In the present study, the SAFE method was first applied to estimate the annual variation and spatial distribution of analysis errors in the Global Forecast System of Global/Regional Assimilation and PrEdiction System (GRAPES_GFS) at the China Meteorological Administration (CMA) since the beginning of its operational implementation (i.e., 2016–2021). Qualitative comparison to analysis error estimations in previous studies showed that SAFE can provide more reasonable spatial-mean analysis error profiles than can the estimation with the ERA-5 reanalysis as a reference (the approach hereafter called "ERAv"). Moreover, ERAv overestimates (underestimates) the spatial-mean analysis error below (above) approximately 500 hPa compared to SAFE because it neglects the uncertainties inherent in reanalysis. Overall, the SAFE estimation reveals that relative reductions of about 12.5%, 29%, and 24.5% were achieved for the spatial-mean analysis errors of wind, temperature, and geopotential height, respectively, in the GRAPES_GFS throughout the six-year study period. These results can largely be attributed to the DA scheme being upgraded from 3D-Var to 4D-Var. SAFE can also provide more reasonable and accurate pointwise analysis errors than ERAv can.

KEYWORDS

analysis error estimation, data assimilation, GRAPES

 © 2023 Royal Meteorological Society

1 | INTRODUCTION

The goal of data assimilation (DA) is to produce an optimally estimated state of the real atmosphere that can then be used as the initial condition (or analysis) of Numerical Weather Prediction (NWP) models (Daley, 1991; Evensen, 1994; Kalnay, 2003; Tang *et al.*, 2016). Owing to the strong sensitivity of the chaotic atmospheric model to initial conditions, the key to improving NWP skill is reducing the errors in the initial analyses (i.e., the differences from the real atmospheric state or "reality"; Lorenz, 1963; Mu *et al.*, 2002; Ding and Li, 2007). Accurate quantitative estimation of the spatiotemporal variations in analysis errors is key for evaluating and improving DA and forecasting performance at operational meteorological centers.

After nearly 10 years of development, the Global Forecast System of the Global/Regional Assimilation and PrEdiction System for global medium-range NWP (GRAPES_GFS) at the China Meteorological Administration (CMA) began operation using the 3D Variational (3D-Var) DA scheme in June 2016 (Wang *et al.*, 2017; Zhang *et al.*, 2019). In the years since its deployment, major upgrades have been made to the DA and prediction system, for example, the operational implementation of the 4D Variational (4D-Var) scheme in July 2018 (Zhang *et al.*, 2019) that considers the observational time more accurately, modifications to the model physics, and the assimilation of additional satellite observations like those from the FengYun series (Yang *et al.*, 2017; Yin *et al.*, 2020; Yin *et al.*, 2021; see more details in Table 1 below). Consequently, it is crucial to quantitatively evaluate the annual variation and spatial distribution of analysis uncertainties in the GRAPES_GFS DA system with such updates. These assessments can provide objective reference for major updates in the DA and prediction systems.

Up to this point, the accuracy of the GRAPES_GFS DA system's analysis has usually been estimated by verifying its output against the ERA-5 reanalysis (Hersbach and Dee, 2016) produced at the European Centre for Medium-Range Weather Forecasts (ECMWF; e.g., Zhang *et al.*, 2019; Wang *et al.*, 2021). This approach is also often used to verify analyses in many other operational DA systems (e.g., Langland *et al.*, 2008; Swanson and Roebber, 2008). Although the reference reanalysis (or analysis) dataset is generally selected from a separate operational center, the codependence between the verified analyses and their reference data cannot be avoided because most of the assimilated observations in the two systems are identical (Peña and Toth, 2014; Feng *et al.*, 2017). In addition, the uncertainties in the reference analysis, despite being neglected for verification purposes, are comparable in magnitude to those in the verified analysis. If the errors in the verified analyses and references are independent, such verification will significantly overestimate the analysis errors.

There are similar limitations in verifying analyses against observations (e.g., Decker *et al.*, 2012; Bao and Zhang, 2013; Wang *et al.*, 2013) as there are in verifying them against analyses in separate systems; in other words, the errors in the reference observations are not considered. Moreover, observational data are sparsely distributed, resulting in incomplete verification of variables within the model space. In addition to the two approaches mentioned above, ensemble-based DA (e.g., using an ensemble Kalman filter) provides a natural framework for the quantification of initial uncertainties (e.g., Houtekamer and Mitchell, 1998; Whitaker and Hamill, 2002). The uncertainties in the analyses can be estimated by calculating the ensemble spread of the updated initial members (or posteriors; e.g., Buizza *et al.*, 2005; Hopson, 2014; Liu *et al.*, 2022) or analyses at multiple centers (e.g., Park *et al.*, 2008; Wei *et al.*, 2010). However, multiple tunable parameters are involved in ensemble-based DA systems, for example, the localization (e.g., Lorenc, 2003) and inflation (e.g., Anderson and Anderson, 1999) factors for the background error covariance, which can often lead to biased estimations of analysis errors. The ensemble-based DA is also computationally expensive since a huge number of observations need to be assimilated for individual members of the high-dimensional atmospheric system.

Peña and Toth (2014; PT14) proposed a Statistical Analysis and Forecast Error (SAFE) estimation method that can provide unbiased, objective, and time-efficient estimation of analysis and forecast error variances relative to the real atmosphere in the model space. The SAFE method estimates the analysis error variance by solving an inverse equation regarding the fitting between the modeled perceived error variance and the actual perceived error variance (i.e., forecast minus analysis). Feng *et al.* (2017) introduced an additional constraint into SAFE to minimize its sensitivity to sampling errors, facilitating the extended application of SAFE to spatial (or grid-point) estimation of analysis error variances. Feng *et al.* (2017) further demonstrated the effectiveness of SAFE in quantifying grid-point analysis error variance in an observing simulation system experiment (OSSE) based on a quasi-baroclinic model. However, SAFE has not yet been applied to the spatiotemporal estimation of analysis errors in operational NWP systems.

This study will apply the modified SAFE method to quantify the annual variation and spatial distribution of analysis errors for the GRAPES_GFS DA system at CMA. The accuracy of the SAFE estimation will be evaluated and examined by comparing it to traditional verification methods against radiosonde observations and ERA-5 reanalysis.

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4507 by Fudan University, Wiley Online Library on [07/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Relative changes (%) in the spatial-mean analysis errors for UV, *T*, and GH at 250, 500, and 850 hPa in the NH and SH estimated by SAFE for successive years

Parameter	pressure	Domain	2017	2018	2019	2020	2021
UV	250 hPa	NH	−2%	−10%*	−14%*	7%*	3%
			0%	−7%*	−9%*	3%	−2%
	500 hPa						
			5%	−9%*	−8%*	0%	−2%
	850 hPa	SH	9%*	−18%*	−4%	−6%	1%
			3%	−10%*	−3%	1%	3%
			6%	−6%	−1%	−2%	−2%
T	250 hPa	NH	1%	−23%*	−5%	−6%	9%*
			0%	−16%*	−10%*	−3%	2%
	500 hPa						
			5%	−21%*	−3%	5%	1%
	850 hPa	SH	−5%	−32%*	27%*	−37%*	6%
			−4%	−23%*	0%	−9%*	1%
			18%*	−24%*	−20%*	3%	−9%*
GH	250 hPa	NH	1%	−20%*	−15%*	2%	10%*
			4%	−15%*	−9%*	−3%	5%
	500 hPa		2%	−15%*	−7%*	−9%*	3%
		SH	−4%	−28%*	12%*	−23%*	23%*
	850 hPa		5%	−22%*	3%	−21%*	9%*
			3%	−13%*	−3%	−16%*	−1%

Note: "2017," "2018," "2019," "2020," and "2021" denote the relative changes for 2016, 2017, 2018, 2019, and 2020, respectively. For example, the relative change for "2017" is calculated by $(e_{2017} - e_{2016})/e_{2016}$, where e_{2017} and e_{2016} denote the analysis RMSEs for 2017 and 2016, respectively. The asterisk indicates statistically significant differences at the 0.05 level.

Section 2 briefly introduces the SAFE method and the data used in this study. The results are presented in Section 3 in detail, mainly focused on the annual variation of the spatial-mean analysis errors and the spatial distribution of the analysis errors. Section 4 provides a discussion and conclusions.

2 | METHODOLOGY AND DATA

2.1 | SAFE method

This section briefly introduces the SAFE method; more details can be found in PT14 and Feng *et al.* (2017). Let us denote $\boldsymbol{f_i^2}$ as the perceived error variance between the forecast $\mathbf{F_i}$ at lead time i and the analysis $\mathbf{A}$ at the same valid time:

$$f_i^2 = |\mathbf{F}_i - \mathbf{A}|^2, \tag{1}$$

where $|\cdot|$ denotes the L2 norm. Considering the errors in $\mathbf{F}_i$ and $\mathbf{A}$ from the real atmosphere, that is, x_i and x_0, respectively, f_i^2 can be converted to:

$$f_i^2 = |x_i - x_0|^2 = x_i^2 + x_0^2 - 2 \cdot \rho_i \cdot x_i \cdot x_0, \tag{2}$$

where x_i^2 and x_0^2 are the error variances of x_i and x_0 at the same valid time, respectively, and ρ_i is the correlation between x_i and x_0 at the same valid time. SAFE is generally used to estimate the temporal-mean (e.g., one season for global NWP models) error variance for a region or a single grid point, so x_i^2 and x_0^2 designate the temporal-mean error variances, and ρ_i is their correlation in time.

PT14 introduces two basic assumptions to reduce the number of unknown variables on the right-hand side of Equation (2). One is that the sample-mean true error variance grows exponentially in the short range (e.g., 2.5 days

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4507 by Fudan University, Wiley Online Library on [07/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

for global NWP models), expressed as:

$$x_i^2 = x_0^2 \cdot e^{\alpha \cdot i \cdot \Delta t}, \quad (3)$$

where Δt is the DA cycle (six hours for typical global NWP models). The other assumption is that ρ_i decays exponentially, that is:

$$\rho_i = \rho_1^i, \quad (4)$$

where ρ_1 is the error correlation at lead time 1 (generally six hours). Equation (4) indicates that the true forecast error x_i and true analysis error x_0 at the same valid time gradually decorrelate with the forecast lead time (see more discussion in PT14). Feng *et al.* (2020) examined the appropriateness of these two assumptions in an OSSE environment of the operational Global Forecast System at the National Centers for Environmental Prediction (NCEP). Given the two basic assumptions (Equations (3) and (4)), the modeled perceived error variance $f_i^{2'}$ can be written as:

$$f_i^{2'} = x_0^2 \cdot e^{\alpha \cdot i \cdot \Delta t} + x_0^2 - 2 \cdot \rho_1^i \cdot x_0^2 \cdot \sqrt{e^{\alpha \cdot i \cdot \Delta t}}. \quad (5)$$

PT14 constructs a cost function J regarding the fitting of the modeled ($f_i^{2'}$) and actual (f_i^2) perceived error variances and solves the three unknown variables (i.e., x_0^2, α, and ρ_1) by minimizing J:

$$J = \max\left(\left|f_i^{2'} - f_i^2\right| \cdot w_i\right), \quad (6)$$

where max(·) denotes the L-infinite norm, and w_i is the weight on $\left|f_i^{2'} - f_i^2\right|$ which decreases with larger variabilities of $\left|f_i^{2'} - f_i^2\right|$ at longer lead times (see PT14 for more details on the calculation of w_i).

To lessen the iterative sensitivity to sampling errors when solving the minimization, Feng *et al.* (2017) introduced an additional constraint into the cost function:

$$\begin{aligned} J = &\max\left(\left|f_i^{2'} - f_i^2\right| \cdot w_i\right) \ (i = 0, 1, \ldots, n) \\ &+ \max\left(\left|f_{i,j}^{2'} - f_{i,j}^2\right| \cdot w_{i,j}\right) \ (i = 0, 1, \ldots, m), \end{aligned} \quad (7)$$

where $f_{i,j}^2$ is the error variance between the lagged forecasts at i and j lead times at the same valid time, and $w_{i,j}$ is the weight on $\left|f_{i,j}^{2'} - f_{i,j}^2\right|$, calculated in a manner similar to that for w_i. $f_{i,j}^{2'}$ is simulated by a simple exponential growth formula:

$$f_{i,j}^{2'} = y_0^2 \cdot e^{\alpha \cdot i \cdot \Delta t}. \quad (8)$$

Note that the fitting of $f_{i,j}^2$ (Equation 8) involves only two parameters (y_0^2 and α) following a simple monotonous relationship and thus is much less sensitive to sampling errors than the fitting of f_i^2 is (Equation 5). More importantly, Equation (8) has a common error growth rate α as the true forecast error variance in the modeling of f_i^2 (see Equations 3 and 5), which significantly decreases the degrees of freedom of the estimated variables and constrains the convergence space of the cost function.

2.2 | Data

This study uses the $0.25^\circ \times 0.25^\circ$ real-time forecast data of the operational GRAPES_GFS global model at CMA in the winter season (December 1 to the following February 28) from 2016 to 2021. The boreal winter was selected because it is the season in which the baroclinic instability at the synoptic scale dominates and frequently influences the weather and climate in the mid-latitudes of the northern hemisphere (Molteni *et al.*, 1990; Molteni and Palmer, 1993), causing extreme events like cold snaps, freezing temperatures, and blizzards. Many previous studies have also used the boreal winter to evaluate the performance of operational assimilation and NWP (e.g., Simmons and Hollingsworth, 2002; Whitaker *et al.*, 2008). The initial times of the forecast samples and lead times are both at 12-hr intervals in this study, which means $j = i + 2$, and i and j vary by 2 in Equation (7). The perceived error variance used in the SAFE estimation is the difference between the GRAPES_GFS forecasts from the corresponding initial analysis at the same valid time. The sample-mean (over the winter season of each year) true forecast error variance is assumed to grow linearly within the first 2.5 days, which means $n = 10$ and $m = 8$ for Equation (7).

To demonstrate the validity of SAFE, its results are quantitatively compared to those of traditional approaches. One traditional approach is assessing the quality of initial analyses by verifying them against the operational global radiosonde observations within the corresponding assimilation window (this approach is hereafter called "OBSv"). The radiosonde observations (see the distribution in Figure 1) have relatively high accuracy and dense horizontal and vertical distribution over the land and thus are often used as the verification reference for analysis and forecast (e.g., Whitaker *et al.*, 2008; Wang *et al.*, 2013). The other approach to estimate the analysis error of the GRAPES_GFS is verifying it against the ERA-5 reanalysis (Hersbach and Dee, 2016) with a $0.25^\circ \times 0.25^\circ$ resolution (the approach is hereafter called "ERAv"). The use of ERAv allows for a comparison of grid-point error estimations with SAFE for different regions and multiple variables.

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4507 by Fudan University, Wiley Online Library on [07/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

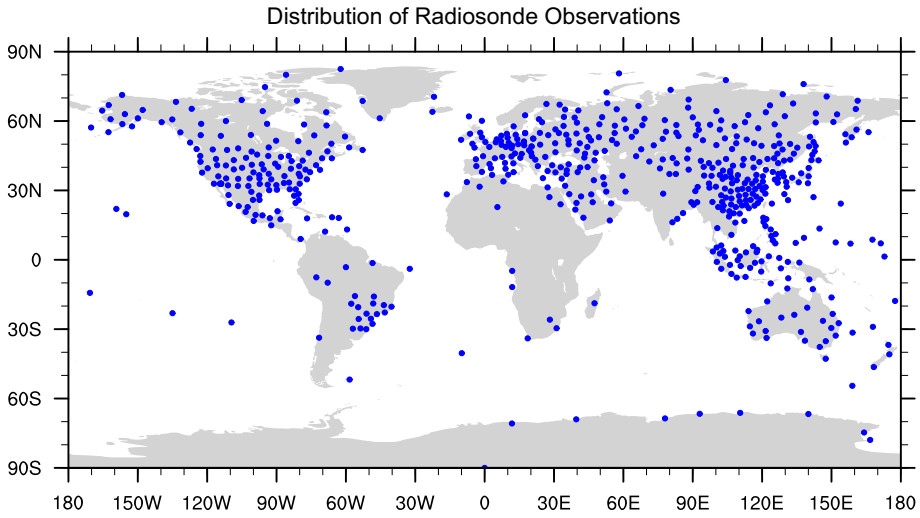

FIGURE 1 Distribution of radiosonde observations assimilated in the GRAPES_GFS data assimilation system

The verification variables include the typical model variables of temperature (T), wind (UV), and geopotential height (GH) at multiple levels from 1,000 hPa to 200 hPa. The UV error variance is calculated by averaging the error variances of the zonal and meridional wind components. Because the lagged forecast difference constraint term (Equation 7) relies on the assumption that the short-range forecast error growth of the model-to-model and model-to-reality comparisons are consistent (i.e., both are α), SAFE is most suitable for regions with negligible model deficiencies at short lead times. Our study therefore only estimates the spatiotemporal error variance in the mid-latitudes (30°–70°) of the northern and southern hemispheres (NH and SH) where systematic model errors are generally negligible (PT14; Li *et al.*, 2018). These regions with major synoptic-scale variabilities are also the most crucial for verifying the model performance.

3 | RESULTS

3.1 | Spatial-mean analysis error for the winter of 2021

The results of the SAFE analysis error estimation are first shown for the spatial mean in the most recent winter considered (2021). Figure 2 displays the temporal evolution of the perceived (black) and estimated true (red) error variances at 500 hPa for UV, T, and GH within the first 2.5 days. It is noticeable that the modeled (or simulated) perceived error variance (black solid lines) has a very good fit with the actual (or measured) values (hollow circles). The modeled perceived error variances are within the 95% confidence interval (black bars) throughout all lead times for all variables in the NH and SH, suggesting the feasibility of the SAFE approach in simulating the evolution of perceived error variances (Equation 5). The estimated true error variances (red solid lines) are derived by fitting the perceived error variance. It is worth noting that the estimated true error variance presents deviations from the perceived error variance and is especially large from 0 to 0.5 days, indicating the importance of considering analysis errors for the verification of short-range forecasts. The relative deviations between the true and perceived error variances become smaller as the forecast errors expand for longer lead times (cf. red and black solid lines).

Figure 3 shows the vertical profiles (black solid lines) of the sample-mean root-mean-square error (RMSE) of the UV, T, and GH analyses over the NH and SH for 2021 estimated by the SAFE method. The analysis error profiles of UV are similar for the NH and SH (Figure 3a,d), which show a maximum at about 300 hPa. This may be related to the fast error growth (see Figure 4) induced by the strong baroclinic instability associated with the upper-level jet stream (Holton and Hakim, 2012). The RMSE of the UV analysis gradually decreases from 300 hPa to both the lower and upper levels and reaches its minimum near the surface. The analysis error profile of the GH displays a pattern that is generally similar to that of UV, except with a slight increase in the RMSE from 300 hPa to 200 hPa. In contrast, the T analysis manifests a rather distinct error profile. The analysis error of T in the NH is minimal in the mid-troposphere (~500 hPa) and gradually increases above and below that level. Unlike in the UV analysis, the RMSE of T in the NH reaches its maximum near the surface (~1,000 hPa), possibly due to the complexity of the lower atmosphere as a result of land use, topography, and surface heat-flux transport in the boundary layer (Wolyn and McKee, 1989; Shafran *et al.*, 2000; Cheng and Steenburgh, 2005). Compared to the NH, the T analysis error in the SH shows much less vertical

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4507 by Fudan University, Wiley Online Library on [07/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

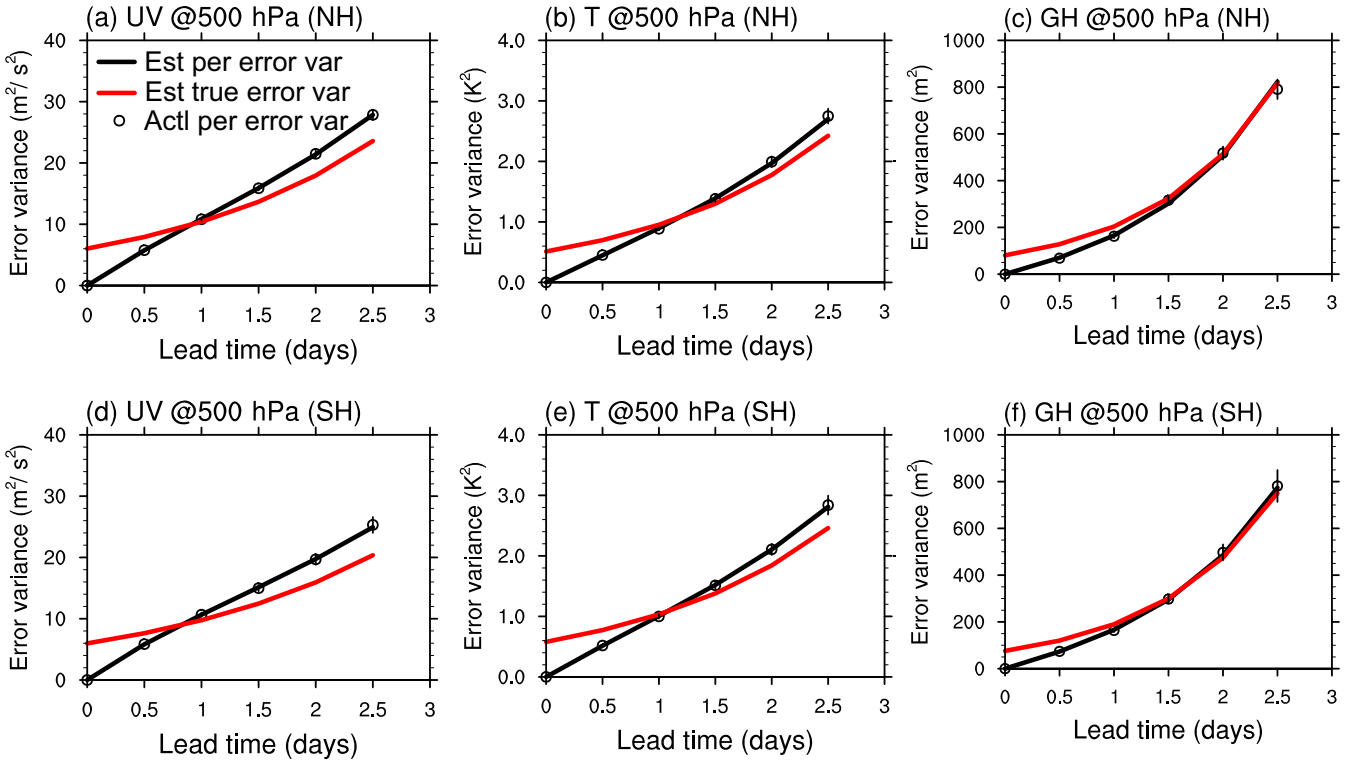

FIGURE 2 Temporal evolution of the actual (hollow black circle) and modeled (black line) perceived error variances and the SAFE-estimated true forecast error variance (red line) for the 500-hPa (a) UV, (b) T, and (c) GH in the NH for the winter of 2021. (d–f) The same as (a–c) but for the SH. The black bars represent the 95% confidence interval

variability and a much lower magnitude near the surface, possibly due to the much weaker orographic effect.

To demonstrate the validity of the SAFE method, its estimated analysis error profiles (dark green solid lines) are compared to those of the OBSv (dark green dashed lines) in the same observation space. Note that the SH 30°–70° region only contains approximately 4% of the global radiosonde observations, while the same region in the NH contains approximately 65% of the global radiosonde observations; the analysis error estimation of the OBSv is thus only shown for two analysis variables – UV and T – in the NH (i.e., Figure 3a,b). The radiosonde observations, as typical conventional observations, have been assimilated into the daily analysis in almost all the operational DA systems worldwide like they have been in the GRAPES_GFS; thus, the comparison against the radiosonde observations does not independently verify the analysis of the GRAPES_GFS. Figure 3a,b show that the analysis error profiles estimated by SAFE are very close in the model and in the observation space (cf. black and dark green solid lines) due to the dense coverage of the radiosonde observations over the NH land. The OBSv estimations present relatively small differences from those of SAFE for UV (~0.4 m·s^{-1}) and T (~0.3 K) above 500 hPa, but they have much larger analysis errors below 500 hPa, especially near the surface. As observational errors are neglected for the OBSv and contain larger representativeness errors at lower levels due to the more complex topographic effects, most of the OBSv verifications overestimate the analysis errors compared to SAFE, especially at lower levels. Nevertheless, the SAFE and OBSv approaches present qualitatively similar vertical patterns of analysis errors for UV and T in the NH; for example, the maximum (minimum) analysis error exists at nearly 300 (500) hPa for UV (T), suggesting that the analysis error estimation made by SAFE was reasonable. In addition, the estimated analysis RMSEs of UV, T, and GH also display vertical profiles that are qualitatively similar to those measured in the OSSE framework (e.g., Figure 5d in Privé *et al.*, 2013; Figure 1a,d in Privé and Errico, 2013; Figure 6 in Wang *et al.*, 2008).

One critical parameter estimated by SAFE is the exponential growth rate (α) of short-range true forecast errors. Figure 4 shows the growth rate of the true forecast RMSEs per six hours for UV, T, and GH in the NH and SH for 2021, as estimated by SAFE. It is noteworthy that the 300-hPa UV and GH both present local maxima of error growth (Figure 4a,c,d,f) associated with the upper-level jet stream in the mid-latitudes of the NH and SH (Holton and Hakim, 2012), leading to large analysis errors (see Figure 3). The UV in both hemispheres and the GH in the SH exhibit additional local maxima of error growth near

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4507 by Fudan University, Wiley Online Library on [07/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

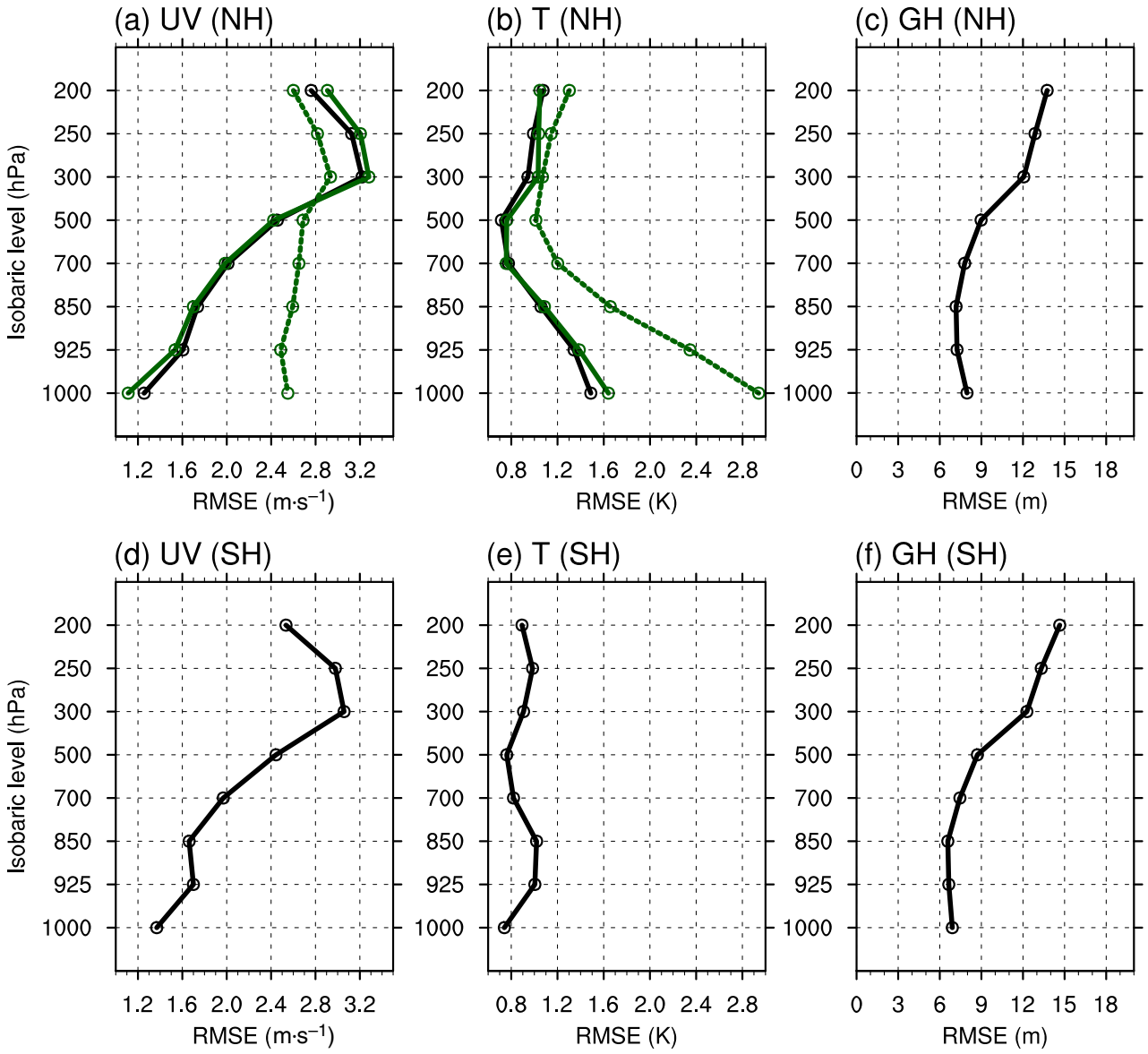

FIGURE 3 Vertical profiles of the estimated analysis RMSE for (a) UV, (b) T, and (c) GH in the NH for the winter of 2021 by the SAFE method. (d–f) The same as (a–c), but for the SH. The analysis error estimation by SAFE (dark green solid lines) and the verification analysis against the radiosonde observations (i.e., the OBSv; dark green dashed lines) are also shown for UV (a) and T (b) in the observation space of the NH

925 hPa; however, their analysis errors are much lower than at 300 hPa (see Figure 3), probably due to the more accurate and denser observations available for the DA at lower levels of the atmosphere. In comparison, T shows a maximum error growth rate at approximately 500 hPa and a much slower error growth rate near the surface. Interestingly, the error growth rates of UV and T are overall below 1.08 per six hours – much lower than that of GH, which ranges from 1.1 to 1.13. This is qualitatively consistent with the results evaluated in an OSSE context by Feng *et al.* (2020); see their Table 2).

Another parameter estimated by SAFE is ρ_1, which indicates the extent to which the six-hour first guess (FG) error is adjusted by the ingestion of observations. Generally, more observations assimilated or more effective assimilation of observations would result in a lower value of ρ_1, that is, a lower correlation between the FG and analysis errors. The results shown in Figure 5 are qualitatively similar for all the variables, indicating that the value of ρ_1 gradually decreases from the upper to lower levels. This is because the observations available for assimilation become more abundant and more accurate from the upper to lower levels and thus exert more remarkable effects on modifying the FG errors during the assimilation, leading to a lower correlation between the FG and analysis errors.

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4507 by Fudan University, Wiley Online Library on [07/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

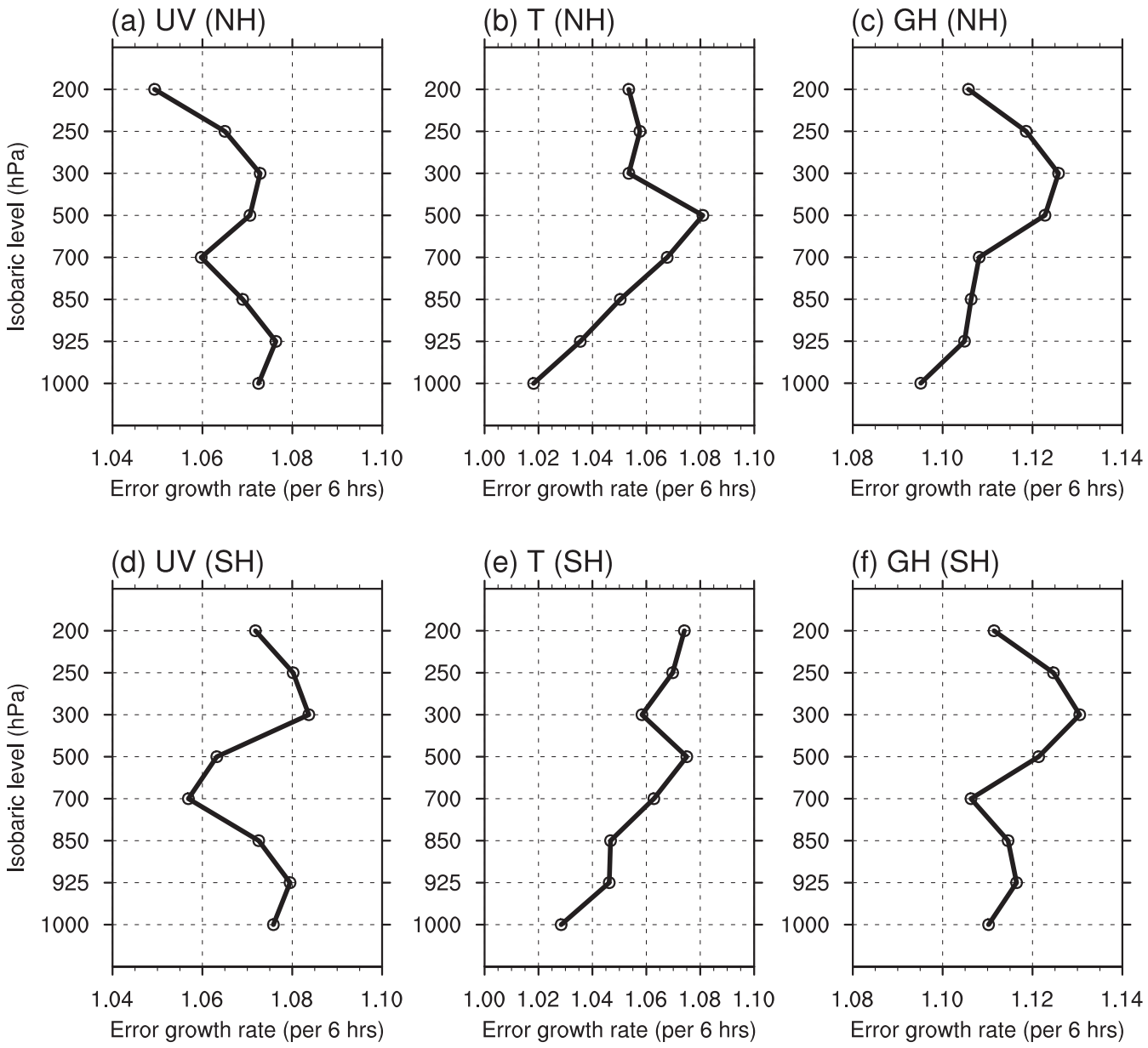

FIGURE 4 Same as Figure 3 but for the growth rate of the true forecast RMSE per six hours

TABLE 2 Comparison of the spatial-mean analysis error estimation ("Regional est") and spatial mean of the grid-point analysis error estimation ("Mean of grid est") by SAFE for UV and *T* at 500 and 850 hPa, and their relative differences

		Regional Est	Mean of Grid Est	Relative difference
		NH; SH	NH; SH	NH; SH
UV ($m\,s^{-1}$)	500 hPa	2.46; 2.44	2.21; 2.21	10%; 9.4%
	850 hPa	1.74; 1.66	1.61; 1.60	7.5%; 3.6%
T (K)	500 hPa	0.71; 0.76	0.67; 0.70	5.6%; 7.9%
	850 hPa	1.06; 1.02	0.97; 0.92	8.5%; 9.8%

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4507 by Fudan University, Wiley Online Library on [07/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

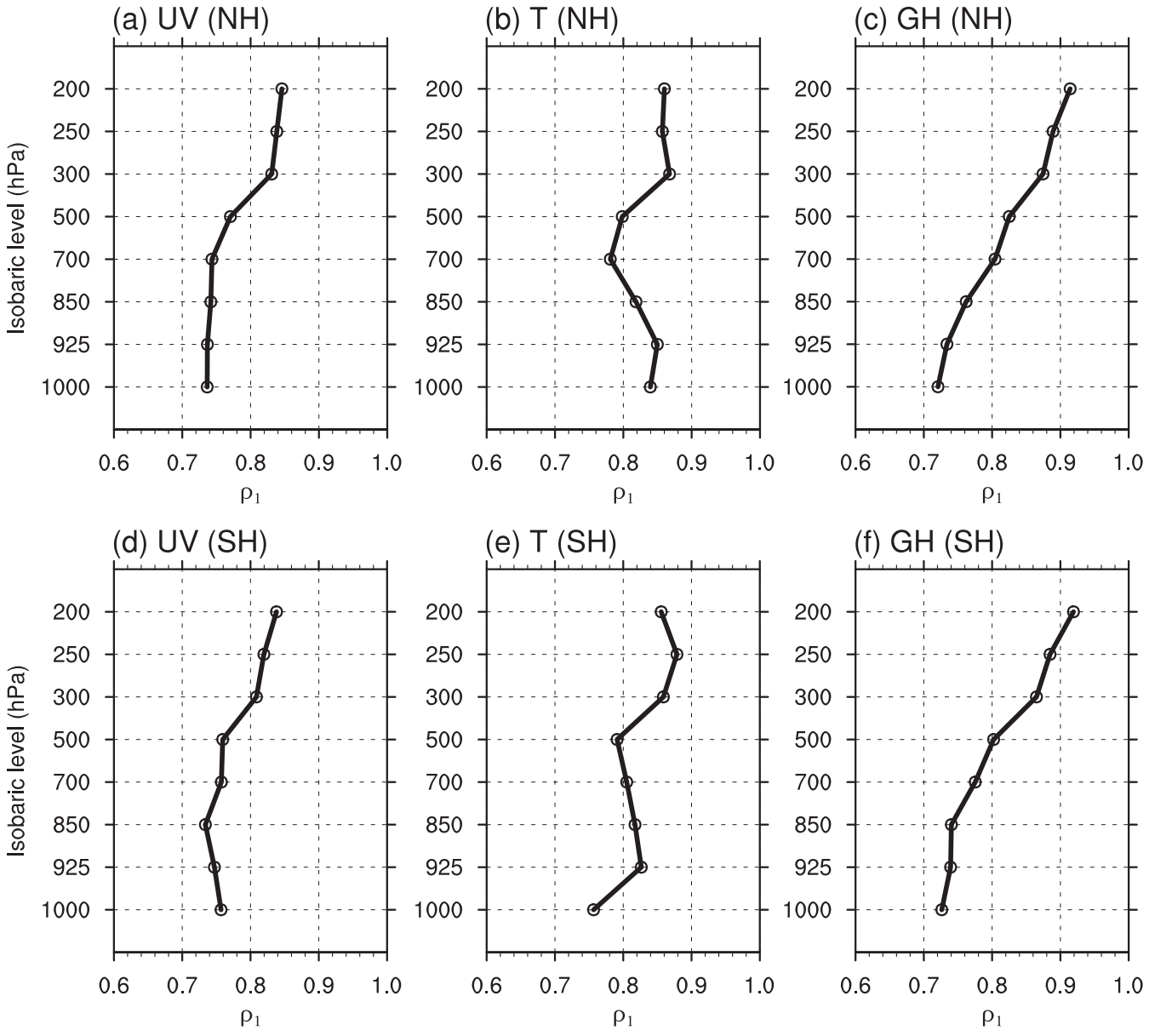

FIGURE 5 Same as Figure 4, but for the correlation ρ_1 between the first guess (six hours) and the analysis errors

3.2 | Annual variation of the spatial-mean analysis error

As one of the two major purposes of this study, the annual variation of the vertical profiles of the spatial-mean analysis error is estimated by SAFE for 2016–2021 (solid lines) in Figure 6. The patterns of the analysis error profiles are overall consistent through the years for individual variables. The analysis RMSEs clearly tend to decrease with time for all the variables at different levels, especially before and after 2017 (see more discussion on Figure 8).

To examine the validity of the SAFE estimation, the yearly variations of the analysis errors (solid lines) estimated using the ERAv are also displayed for comparison (Figure 7). The SAFE analysis error estimation for 2020 is also given in Figure 7 (dashed orange lines). The analysis error profiles of the ERAv display patterns that are qualitatively similar to those from SAFE (cf. solid and dashed orange lines); for example, the maximum (minimum) analysis error of UV (T) exists at nearly 300 (500) hPa in the NH (Figure 7a,b). An unreasonable result for the ERAv, however, is the UV analysis error at 500 hPa being very close to or even larger than (e.g., 2017 and 2019) that at 300 hPa in the NH (Figure 7a). For the GH in the NH, the analysis RMSEs of ERAv at 300 and 500 hPa are also very close for most years. These results are in distinct contrast to those from the OBSv (see Figure 3 and also Lei *et al.*, 2018), other studies using the OSSE (e.g., Wang *et al.*, 2008; Privé *et al.*, 2013; Privé and

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4507 by Fudan University, Wiley Online Library on [07/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

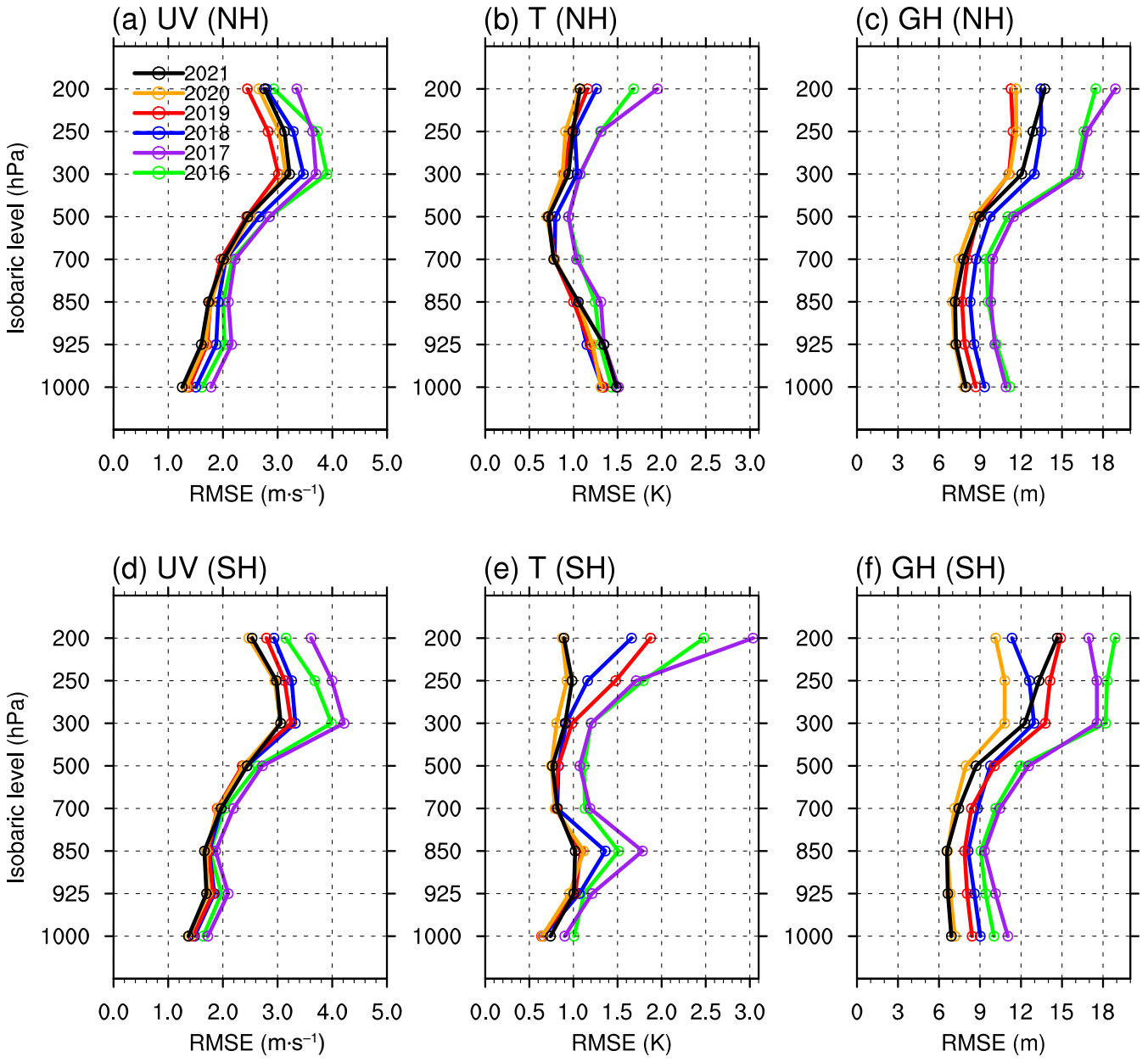

FIGURE 6 Same as Figure 3, but for the annual SAFE estimation from 2016 to 2021

Errico, 2013), and the SAFE estimation (Figure 6a,c), all of which consistently show dramatically lower analysis errors at 500 hPa than at 300 hPa for the UV and GH values in the NH. The unreasonable estimation of UV and GH for the ERAv may be attributed to the codependence between the analysis errors of the GRAPES_GFS and ERA-5.

A remarkable difference in the estimated analysis error profile by SAFE and ERAv is that the former has lower analysis errors overall below 500 hPa but larger analysis errors above 500 hPa for all variables except for the GH in the NH (cf. dashed and solid orange lines in Figure 7). The larger analysis error estimation for ERAv at mid and low levels may be because of more random components in the low-level analysis errors that originated from a variety of aspects, for example, the noise ingested by the assimilation of more low-level observations (e.g., Hunt *et al.*, 2007; Stewart *et al.*, 2013) and the approximations used in representing topographic and boundary-layer effects (e.g., Wolyn and McKee, 1989; Cheng and Steenburgh, 2005). In this case, the errors in the GRAPES_GFS analysis and the ERA-5 reanalysis are nearly independent (or orthogonal), resulting in the overestimation of analysis errors by the ERAv relative to SAFE. In contrast, upper-level variable fields are dominated by more slowly varying synoptic scales and have fewer observations

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4507 by Fudan University, Wiley Online Library on [07/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

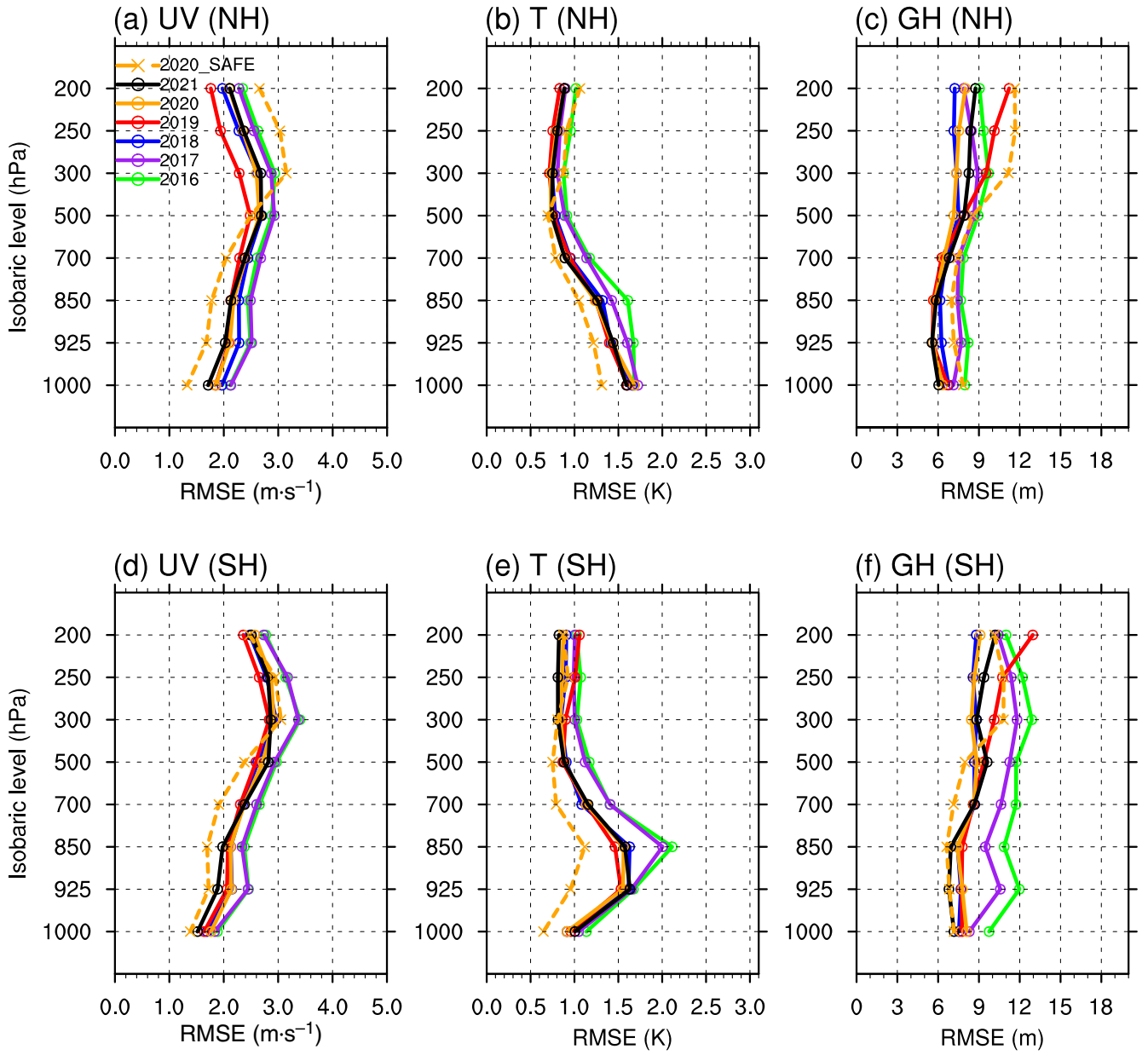

FIGURE 7 Same as Figure 6 but for the analysis errors estimated by ERAv (solid lines). The SAFE-estimated analysis errors for 2020 (dashed lines) are also given for reference

to assimilate than lower levels do, rendering more codependent analysis errors between the GRAPES_GFS analysis and the ERA-5 reanalyses. The ERAv may thus underestimate analysis errors relative to the SAFE estimation.

Figure 8 further shows the annual variation in the analysis RMSEs at three levels (250, 500, and 850 hPa) for the SAFE (solid lines) and ERAv (dashed lines) estimations. The annual variations in the estimated analysis errors by SAFE present qualitatively similar declining trends with time as those by ERAv but manifest systematic differences (as discussed in Figure 7). Specifically, for the SAFE estimations, the overall analysis errors show slight changes from 2016 to 2017 (4% on average in Table 1), most of which are statistically insignificant at the 0.05 level, and then a remarkable reduction (17% on average) in 2018 that is statistically significant at the 0.05 level for almost all the variables, especially at 250 hPa, by nearly 22% on average. Relative reductions of approximately 10%, 23%, and 19% of the analysis RMSEs (Table 1) can be achieved for UV, *T*, and GH, respectively, from 2017 to 2018. This can be attributed to the major upgrade of the GRAPES_GFS DA system from 3D-Var to 4D-Var in Aug 2018 (Zhang *et al.*, 2019).

There is another decrease in the analysis error from 2018 to 2019, although it is less dramatic than that

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4507 by Fudan University, Wiley Online Library on [07/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

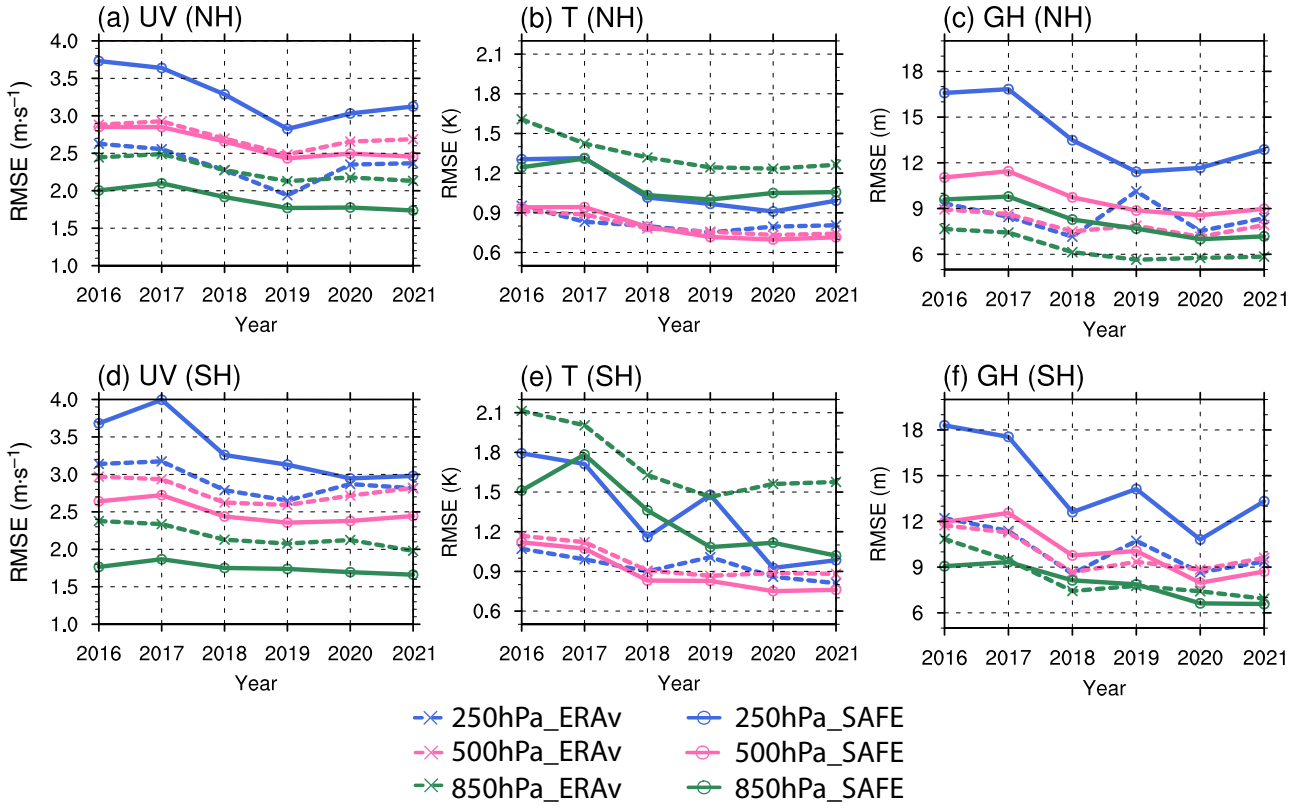

FIGURE 8 Annual variation in the spatial-mean analysis errors for (a) UV, (b) *T*, and (c) GH at 250, 500, and 850 hPa in the NH estimated by SAFE (solid lines) and ERAv (dashed lines). (d–f) The same as (a–c) but for the SH

from 2017 to 2018, mainly for the variables in the NH (about 9% on average). This could primarily be related to the application of a modified planetary boundary-layer (PBL) scheme on the Charney–Phillips (CP) grid in the GRAPES_GFS model and a new cloud scheme implemented on the CP grid (CLOUD_CP; see more details in Chen *et al.*, 2020). The improvement could also be partially attributed to the assimilation of additional observations into the operational system, including the Geostationary Interferometric Infrared Sounder (GIIRS) observations from the FengYun-4A satellite and the Microwave Humidity Sounder-2 (MWHS-2) radiances onboard the FengYun-3D satellite. From 2019 to 2020, the variables only present a slight change in the analysis errors in the NH but show a statistically significant reduction for *T* and GH in the SH by about 18%, possibly due to the additional assimilation of satellite observations such as the Hyperspectral Infrared Atmospheric Sounder (HIRAS), the Microwave Radiation Imager-I (MWRI-I) onboard the FengYun-3D satellite, and some other satellite observations. From 2020 to 2021, most variables remained unchanged or even showed a slight increase. Overall, the analysis RMSE reached a relative reduction of about 12.5%, 29%, and 24.5% for UV, *T*, and GH, respectively, during the six years.

3.3 | Grid-point estimation of analysis errors for the winter of 2021

In addition to estimating the spatial-mean analysis error, SAFE can also provide a grid-point estimation of analysis errors. Figure 9 shows the spatial distribution of analysis errors estimated by SAFE and ERAv for 500-hPa analysis variables UV and *T* in the NH and SH for the winter of 2021 (shading). The 500-hPa GH (green contour) averaged over the winter of 2021 is also shown for reference. Theoretically, the spatial-mean analysis error estimation by SAFE (i.e., by fitting the spatial-mean perceived error variances; see Sections 3.1 and 3.2) should be consistent with the spatial-mean grid-point analysis error estimation (i.e., by fitting the grid-point perceived error variances) within the same region. Table 2 shows a comparison of the spatial-mean analysis error estimation and the spatial-mean grid-point analysis error estimation based on SAFE for the winter of 2021. These two estimations show overall small differences ranging from 3.6% to 10% for variables UV and *T* at 500 and 850 hPa. Since the spatial-mean analysis error estimation by SAFE (i.e., the results in Sections 3.1 and 3.2) is much less influenced by sampling errors than the grid-point estimation is, the magnitude of the grid-point analysis errors by SAFE

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4507 by Fudan University, Wiley Online Library on [07/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

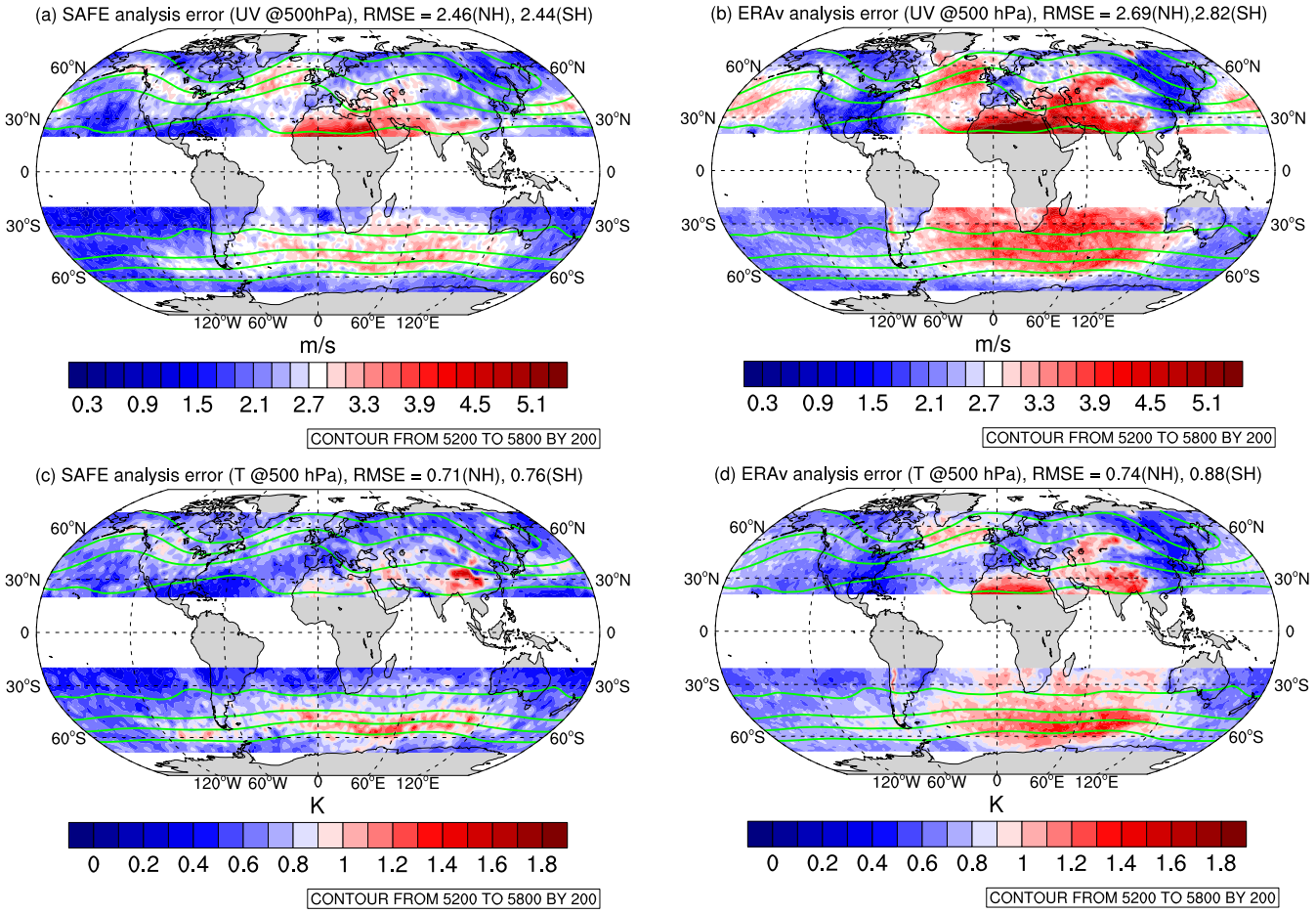

FIGURE 9 Comparison of the grid-point analysis errors (shading, units: $m \cdot s^{-1}$) estimated by (a) SAFE and (b) ERAv for UV at 500 hPa. (c,d) The same as (a,b), but for *T* (units: K) at 500 hPa. Green contours indicate the winter mean 500-hPa GH in 2021 (units: gpm)

for each hemisphere in Figures 9 and 10 is rescaled by the spatial-mean analysis error estimations for each hemisphere without changing the spatial pattern.

In Figure 9, the analysis error estimated by ERAv has a similar spatial pattern to that of the SAFE estimation for both the 500-hPa UV and *T*. However, ERAv displays a general overestimation bias relative to the SAFE estimation for both UV (2.69 vs 2.46 $m \cdot s^{-1}$ for NH, 2.82 vs 2.44 $m \cdot s^{-1}$ for SH) and *T* (0.74 vs 0.71 K for NH, 0.88 vs 0.76 K for SH), as suggested by Figures 7 and 8. For the SAFE estimation of the UV wind (Figure 9a), the maximum analysis errors (4–5 $m \cdot s^{-1}$) mainly occurred over North Africa at about approximately 30° N, which may be related to the sparse distribution of mid-level observing networks in the area (e.g., fewer radiosonde observations, as shown in Figure 1). The downstream regions extending from north of the Arabian Peninsula to Pakistan also exhibit a band of high analysis errors (3–3.5 $m \cdot s^{-1}$), possibly influenced by the low upstream analysis quality.

The south to southeast of the Tibetan Plateau (TP) also presents large analysis errors, which may be related to the complex effects of topography, such as intense wind shear, large friction, and representativeness errors. The large analysis errors around the Caspian Sea (2.7–3.1 $m \cdot s^{-1}$) may be attributed to the baroclinic instability ahead of the trough near Eastern Europe (see the green contour in Figure 9a). Other regions with large analysis errors (~3 $m \cdot s^{-1}$) include the North Atlantic Ocean and the North Pacific Ocean in the "storm track" regions of the maximum variance of GH (e.g., Wallace *et al.*, 1988). Western North America also presents large analysis errors (2.7–3.0 $m \cdot s^{-1}$), possibly due to the strong dynamic instability of the atmospheric river moisture (Wilson *et al.*, 2022). In contrast, the regions with large analysis errors (~3 $m \cdot s^{-1}$) in the SH are concentrated over the Indian and Atlantic Oceans within 30°–60° S, which may be associated with the sparse distribution of available observations there.

The 500-hPa *T* (Figure 9c) has an overall similar distribution of estimated analysis errors as the 500-hPa UV does, except the southern TP presents the largest analysis error of nearly 1.5 K near 30° N. This may be related to the complex thermodynamics influenced by the steep temperature

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4507 by Fudan University, Wiley Online Library on [07/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

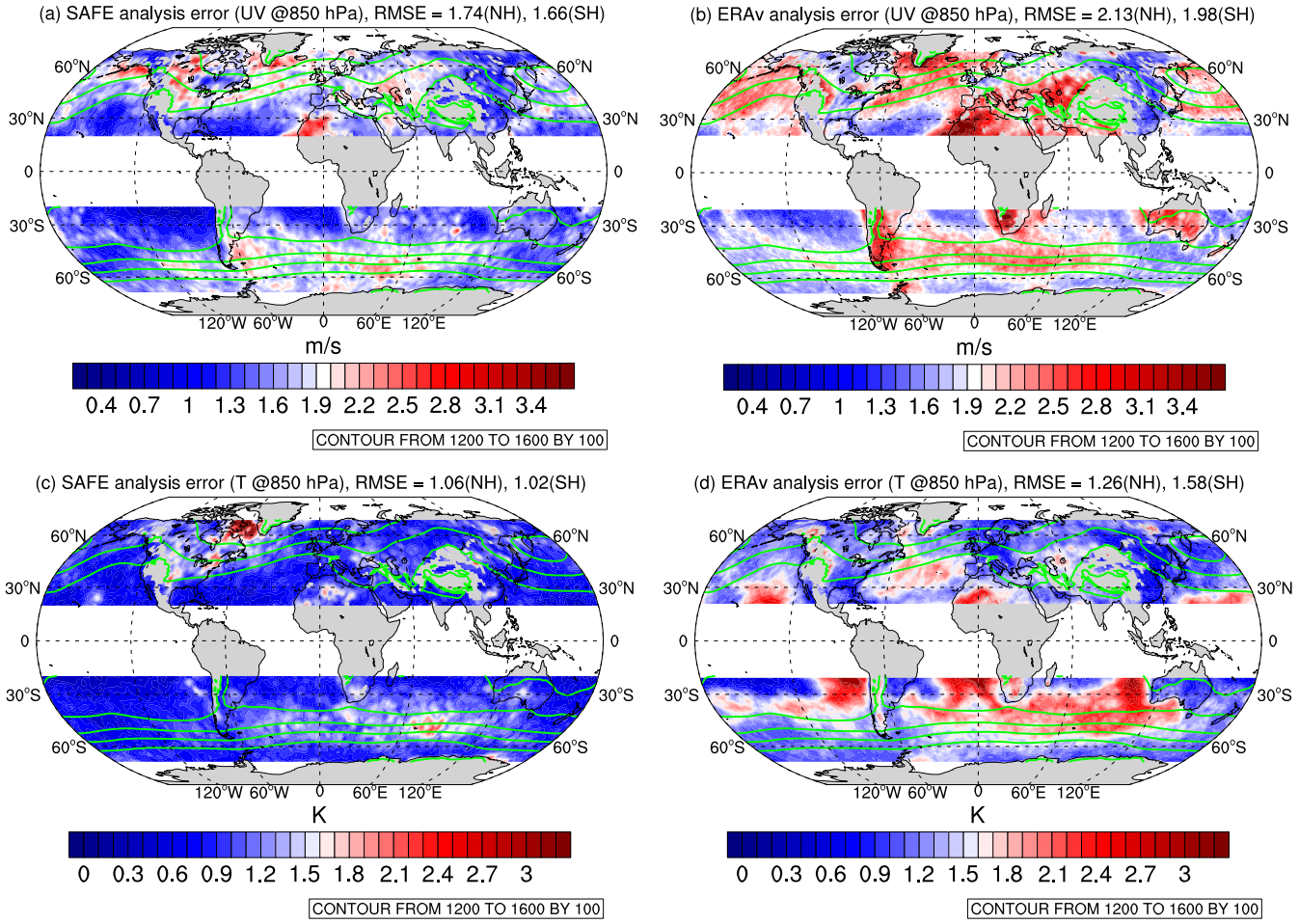

FIGURE 10 Same as Figure 9, but for UV and T at 850 hPa

and moisture gradient at the lateral boundary of the TP. Noticeably, for both UV and T, the overestimated analysis error of the ERAv relative to that from SAFE is particularly dramatic over the oceans extending from the mid-latitudes to the tropics (not shown). This could be attributed to the large systematic bias over the oceans between the global models of the GRAPES_GFS and ERA-5 for the ERAv algorithm.

Figure 10 is similar to Figure 9 but for the 850-hPa level. Despite being much more influenced by the topography compared to 500 hPa, the analysis errors of the 850-hPa UV estimated by SAFE share similarities with those of the 500-hPa UV (cf. Figures 9 and 10). These similarities include the larger analysis errors in North Africa and to the north of the Arabian Peninsula (possibly associated with relatively sparse observations), the region around the Caspian Sea associated with baroclinic instability in front of the trough, the North Atlantic and Pacific Oceans in the storm track regions, northwestern North America being influenced by the atmospheric river, and the South Atlantic and Indian Oceans. The ERAv estimation can capture the major centers of the large analysis errors of UV indicated by SAFE but presents a general overestimation relative to the results of SAFE for both the NH and SH; this is similar to the results for 500 hPa. The overestimation bias of the ERAv is particularly remarkable for UV over the SH land at 850 hPa (cf. Figure 10a,b), for example, southern Africa, southern South America, and Australia, leading to unreasonably larger analysis errors over the land than over the ocean. This is possibly due to the differences between the models used for producing the GRAPES-GFS analysis and the ERA-5 reanalysis in dealing with the topography-related effects and processes in the planetary boundary layer.

The analysis error estimation of the 850-hPa T by SAFE has a roughly similar distribution to that of the 850-hPa UV except for the maximum analysis errors near Baffin Island to the southwest of Greenland Island (~3 K). In contrast to the ERAv result for the 850-hPa UV (Figure 10b), the ERAv estimation for the 850-hPa T (Figure 10d) has prominent analysis errors over the oceans to the west of the continents near 20° N and 20° S, which is probably related to the large systematic differences between the two forecast models. In contrast, SAFE is not affected by systematic errors in the

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4507 by Fudan University, Wiley Online Library on [07/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

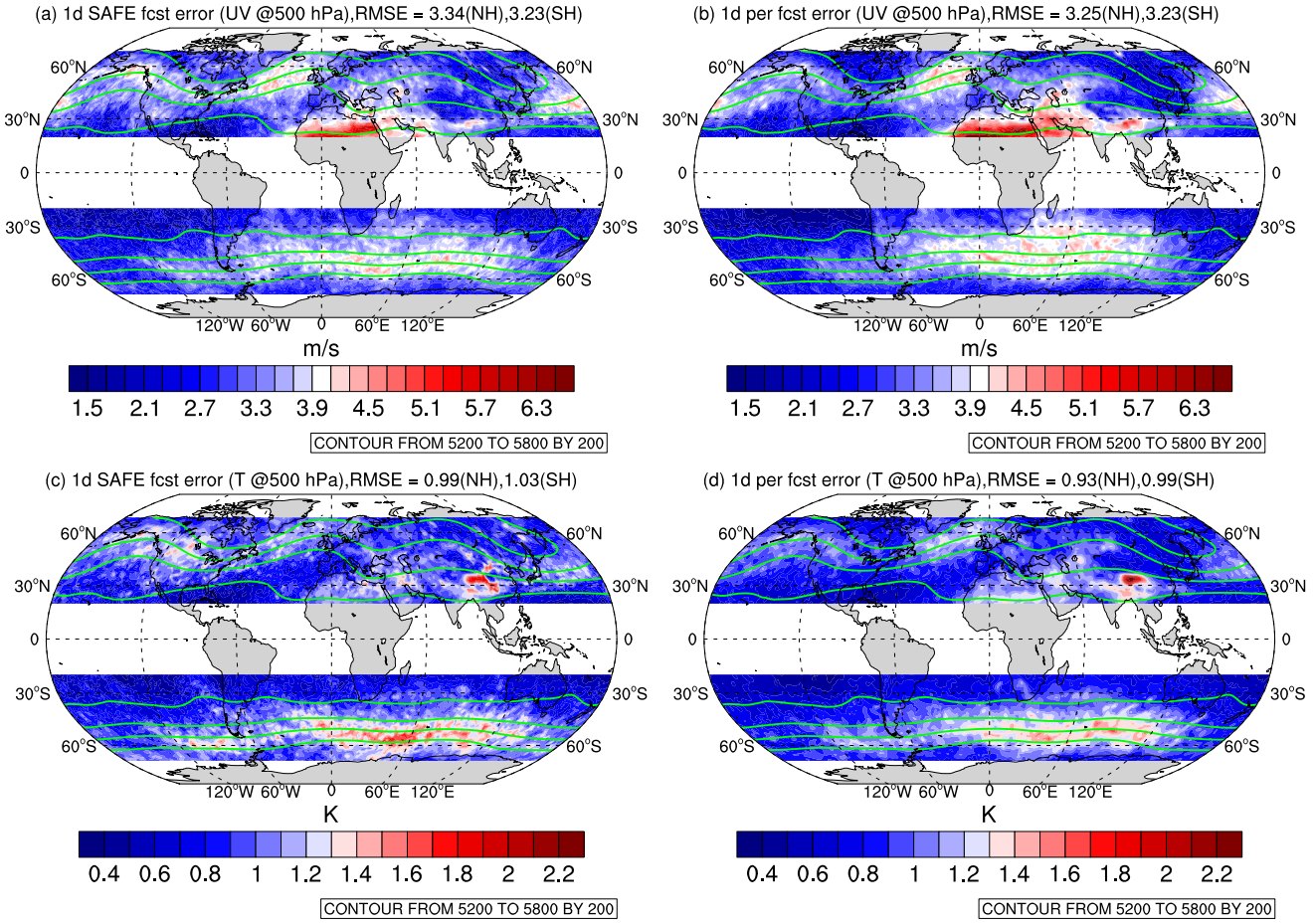

FIGURE 11 Comparison of grid-point forecast error estimates at Day 1 by (a) SAFE and (b) perceived errors for UV at 500 hPa (shading, units: $m \cdot s^{-1}$). (c,d) The same as (a,b), but for T at 500 hPa (shading, units: K). Green contours indicate the winter mean 500-hPa GH in 2021 (units: gpm). RMSEs over the NH and SH are given in the title of each panel

forecast models and manifests much lower analysis errors than the ERAv estimation does over these oceanic regions.

SAFE can also estimate short-range true forecast errors (i.e., the forecast compared to reality), which are assumed to be the result of exponential growth of analysis errors (i.e., $x_0^2 \cdot e^{\alpha \cdot i \cdot \Delta t}$). True forecast errors become closer to the perceived errors beyond an initial short period during which the analysis error cannot be neglected (see Figure 2). For example, the spatial-mean forecast errors from SAFE and the perceived errors are nearly consistent at Day 1 for the 500-hPa UV and T, in contrast to their big differences at initialization (see Figure 2). Therefore, we further compared the spatial distribution of the estimated true forecast errors by SAFE and the perceived errors at Day 1 for the UV and T at 500 hPa (Figure 11) and 850 hPa (Figure 12). This can offer an indirect assessment of the accuracy of the SAFE analysis error estimation.

Figure 11 shows that the estimated Day 1 true forecast errors by SAFE present spatial patterns that are overall similar to the estimated analysis errors for both UV (cf. Figures 8a and 11a) and T (cf. Figures 9c and 11c). This indicates that the uncertainties in the short-range background forecasts are a critical factor affecting the analysis uncertainties in the DA. Noticeably, the SAFE estimation and the perceived error display very similar spatial *distributions* of the Day 1 forecast errors with close spatial-mean RMSEs (less than 6% difference) and high spatial correlations for UV (0.78 for the NH and 0.88 for the SH) and T (0.64 for NH and 0.86 for SH). This indirectly confirms SAFE's reasonable estimation of the grid-point analysis errors.

Figure 12 presents the same information as in Figure 11 but for 850 hPa. The spatial-mean RMSEs of the true forecast errors for the SAFE estimation and the perceived errors at Day 1 are still close for 850-hPa UV, like the variables at 500 hPa (see Figure 11), but show slightly larger (~10%) relative differences for T. The spatial correlations between the SAFE estimation and the perceived errors at 850 hPa are similar to those at 500 hPa: 0.84 (NH) and 0.82 (SH) for UV, and 0.63 (NH) and 0.82 (SH) for T. The lower correlation for T over the NH can

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4507 by Fudan University, Wiley Online Library on [07/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

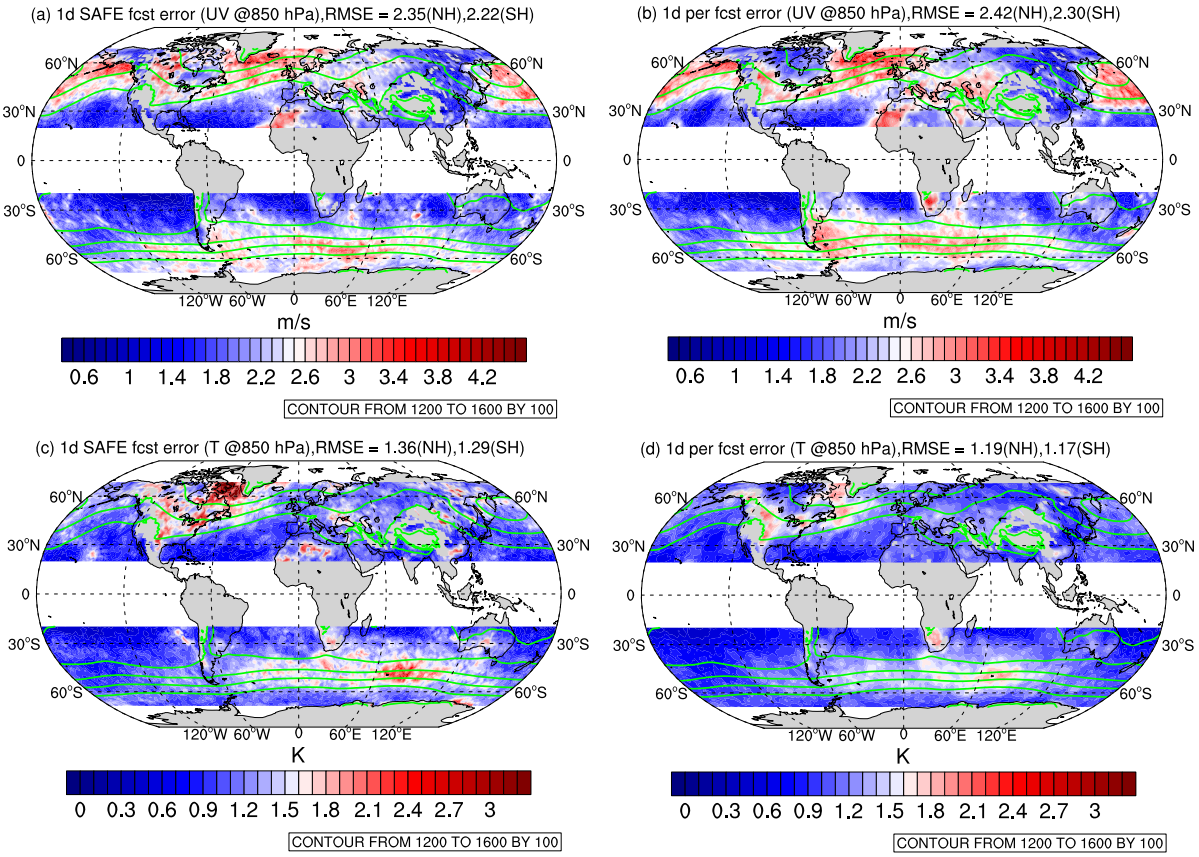

FIGURE 12 Same as Figure 11, but for UV and T at 850 hPa

possibly be attributed to the large forecast error differences between the two approaches in some regions, for example, to the southwest of Greenland Island where the SAFE estimation is almost twice that of the perceived errors (cf. Figure 12c,d). Such differences may be because the perceived errors are calculated using forecasts and analyses produced with the same model. SAFE, however, measures the differences between the model trajectory and the real atmosphere, which are generally more influenced by the model deficiencies associated with complex topography, boundary-layer parameterizations, unresolved scales, and so forth. Overall, the similarities of the Day 1 forecast errors derived from the perceived errors and the SAFE estimation offer indirect evidence of the validity of SAFE in estimating analysis errors.

4 | DISCUSSION AND CONCLUSIONS

Accurate spatiotemporal estimation of analysis errors (i.e., the analysis compared to reality) is particularly crucial to the objective evaluation and upgrading of the performance of operational DA and prediction systems. This study applied a modified SAFE estimation algorithm to quantify the annual variation and spatial distribution of the analysis errors in the operational GRAPES_GFS system at CMA for the winter seasons from 2016 to 2021. Unlike traditional approaches that compare analyses against reference data with errors (e.g., reanalysis at separate operational centers and the observations), SAFE quantifies error variances in analyses relative to the real atmosphere by solving an inverse problem based on unbiased assumptions. This is the first time the modified SAFE has been applied to estimate the spatiotemporal variation of analysis errors in practical operational DA systems.

The spatial-mean error estimation shows that SAFE offers a statistically reliable simulation of the temporal evolution of perceived error variances. The estimated true forecast error variances display remarkable deviations from the perceived error variances at short ranges, indicating the importance of considering the analysis uncertainties for the verification of short-range forecasts. The estimated analysis error profiles by SAFE present similar patterns through the years, that is, the maximum analysis error being at nearly 300 hPa and the minimum being near the surface for UV and GH, while the maximum is near the surface and the minimum is at approximately 500 hPa for T. The SAFE estimations of the analysis error profiles are qualitatively similar to those deduced by other methods

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4507 by Fudan University, Wiley Online Library on [07/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

such as verification against the observations, OSSE, and ensemble spread.

Despite similar profiles of the estimated analysis errors through the years, the magnitude of the analysis errors estimated by SAFE present a clear decreasing trend from 2016 to 2021, with relative reductions of approximately 12.5%, 28%, and 24.5% on average for UV, *T*, and GH, respectively. The most significant improvement in the analysis accuracy can be seen from 2017 to 2018 as a result of an upgrade to the operational DA scheme from 3D-Var to 4D-Var. The other less significant improvement from 2018 to 2020 can probably be attributed to improvements in the PBL and cloud microphysics schemes and the assimilation of additional satellite observations, such as those onboard the FengYun series satellites. The ERAv estimates created by verifying the GRAPES_GFS analysis against the ERA-5 reanalysis display an analysis error trend that is qualitatively similar to that of SAFE. However, the ERAv gives closer or even larger analysis RMSEs for UV and GH at 500 hPa than at 300 hPa for the NH, which is unreasonable when compared to the estimations made with other approaches. In addition, because the errors in the ERA-5 reanalysis are not considered, ERAv displays a general overestimation of analysis errors below 500 hPa and an underestimation above 500 hPa relative to those from SAFE.

As for the grid-point estimation of analysis errors, SAFE can capture the major centers with relatively large analysis errors, which are roughly consistent for UV and *T* at 500 and 850 hPa. These include North Africa, extending eastward to the south of the TP at about 30° N; regions adjacent to the Caspian Sea; northwestern North America; the North Atlantic and Pacific Oceans in storm track regions for the NH; and the South Atlantic and Indian Oceans for the SH. ERAv can also capture these major regions with high analysis uncertainties but presents an overall overestimation, especially for the lower levels that experience stronger orographic effects, possibly due to the systematic bias between the operational global models used for producing the GRAPES_GFS analysis and the ERA-5 reanalysis. The accuracy of the SAFE analysis error estimation is further assessed indirectly by comparing the Day 1 forecast error distributions derived from SAFE and the perceived error. These two methods provide overall similar spatial-mean forecast errors with less than 6% differences (except for ~10% differences for *T* over the 850-hPa NH). Moreover, the spatial correlations of the forecast error distribution between the two approaches exceed 0.8 for most variables and regions. These indirect assessments further validate the accuracy of the spatial analysis errors estimated by SAFE.

This study has demonstrated that SAFE performs better than ERAv does in estimating spatiotemporal analysis errors. Nevertheless, it is worthwhile to further evaluate the accuracy of the SAFE algorithm by comparing the SAFE estimations to verification methods using high-precision independent observations for reference. In the future, the SAFE method can also be applied to analysis error estimations for other DA and prediction systems of the weather, climate, and ocean and to the comparison of analysis accuracy among different seasons. The SAFE estimation of short-range forecast errors (e.g., six hours) also has the potential to improve the accuracy of the specification of static background error covariance and the performance of DA systems. Due to the highly variable concentration and complex phase change of water vapor, its short-range true forecast errors may not satisfy the exponential growth assumption in SAFE. Therefore, the application of SAFE to variables like water vapor needs further consideration and exploration.

AUTHOR CONTRIBUTIONS

Jie Feng: Methodology; visualization; writing – original draft; writing – review and editing. **Jincheng Wang:** Conceptualization; data curation; investigation; supervision. **Guokun Dai:** Validation. **Feifan Zhou:** Writing – review and editing. **Wansuo Duan:** Writing – review and editing.

ACKNOWLEDGEMENTS

The authors are grateful to Drs Yongzhu Liu, Lin Zhang, Suhong Ma at China Meteorological Administration, Profs Mu Mu (Fudan University), Lili Lei (Nanjing University), Qiang Wang (Hohai University), and Dr. Kun Zhang (the Institute of Oceanology) in China for their input and helpful discussions. This study was supported by the National Natural Science Foundation of China (Grant No. 12241103 and 42105054) and China Postdoctoral Science Foundation (2022M723468).

DATA AVAILABILITY STATEMENT

The real-time global forecast products of the GRAPES_GFS at CMA and the ERA-5 reanalysis at ECMWF for verification were provided by the co-author Dr. Jincheng Wang. These data are available from the authors upon request. The ERA-5 reanalysis can also be downloaded from the ECMWF dataset archive at. https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=form

ORCID

Jie Feng https://orcid.org/0000-0002-2480-2003
Jincheng Wang https://orcid.org/0000-0003-2442-9760
Guokun Dai https://orcid.org/0000-0001-5303-2952
Wansuo Duan https://orcid.org/0000-0002-0122-2794

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4507 by Fudan University, Wiley Online Library on [07/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

REFERENCES

Anderson, J.L. and Anderson, S.L. (1999) A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127, 2741–2758. https://doi.org/10.1175/1520-0493(1999)127%3C2741:AMCIOT%3E2.0.CO;2.

Bao, X. and Zhang, F. (2013) Evaluation of NCEP-CFSR, NCEP-NCAR, ERA-interim, and ERA-40 reanalysis datasets against independent sounding observations over the Tibetan plateau. *Journal of Climate*, 26(1), 206–214. https://doi.org/10.1175/JCLI-D-12-00056.1.

Buizza, R., Houtekamer, P.L., Toth, Z., Pellerin, P., Wei, M. and Zhu, Y. (2005) A comparison of the ECMWF, MSC and NCEP global ensemble pre-diction systems. *Monthly Weather Review*, 133, 1076–1097.

Chen, J., Ma, Z., Li, Z., Shen, X., Su, Y., Chen, Q. and Liu, Y.Z. (2020) Vertical diffusion and cloud scheme coupling to the Charney–Phillips vertical grid in GRAPES global forecast system. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 2191–2204. https://doi.org/10.1002/qj.3787.

Cheng, W.Y.Y. and Steenburgh, W.J. (2005) Evaluation of surface sensible weather forecasts by the WRF and the eta models over the western United States. *Weather Forecasting*, 20, 812–821.

Daley, R. (1991) *Atmospheric Data Analysis*. Cambridge: Cambridge University Press, p. 457.

Decker, M., Brunke, M.A., Wang, Z., Sakaguchi, K., Zeng, X. and Bosilovich, M.G. (2012) Evaluation of the reanalysis products from GSFC, NCEP, and ECMWF using flux tower observations. *Journal of Climate*, 25(6), 1916–1944. https://doi.org/10.1175/JCLI-D-11-00004.1.

Ding, R.Q. and Li, J.P. (2007) Nonlinear finite–time Lyapunov exponent and predictability. *Physics Letters A*, 364, 396–400.

Evensen, G. (1994) Sequential data assimilation with a nonlinear quasi–geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5), 10143–10162.

Feng, J., Toth, Z. and Peña, M. (2017) Spatial extended estimates of analysis and short-range forecast error variances. *Tellus A*, 69(1), 1325301.

Feng, J., Toth, Z. and Peña, M. (2020) Partition of analysis and forecast error variance into growing and decaying components. *Quarterly Journal of the Royal Meteorological Society*, 146(728), 1302–1321.

Hersbach, H. and Dee, D. (2016) *ERA5 Reanalysis Is in Production, ECMWF Newsletter 147*. Reading: ECMWF.

Holton, J.R. and Hakim, G.J. (2012) An introduction to dynamic meteorology. In: *An Introduction to Dynamic Meteorology: Fifth Edition*, Fifth edition. Singapore: Academic Press, pp. 1–532. https://doi.org/10.1016/C2009-0-63394-8.

Hopson, T.M. (2014) Assessing the ensemble spread-error relationship. *Monthly Weather Review*, 142(3), 1125–1142. https://doi.org/10.1175/MWR-D-12-00111.1.

Houtekamer, P.L. and Mitchell, H.L. (1998) Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126, 796–811.

Hunt, B., Kostelich, E. and Szunyogh, I. (2007) Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica D*, 230, 112–126.

Kalnay, E. (2003) *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge: Cambridge University Press, pp. 175–184.

Langland, R.H., Maue, R.N. and Bishop, C.H. (2008) Uncertainty in atmospheric temperature analysis. *Tellus*, 60A, 598–603.

Lei, L., Whitaker, J.S. and Bishop, C. (2018) Improving assimilation of radiance observations by implementing model space localization in an ensemble Kalman filter. *Journal of Advances in Modeling Earth Systems*, 10(12), 3221–3232. https://doi.org/10.1029/2018MS001468.

Li, J.P., Feng, J. and Ding, R.Q. (2018) Attractor radius and global attractor radius and their application to the quantification of predictability limits. *Climate Dynamics*, 51, 2359–2374. https://doi.org/10.1007/s00382-017-4017-y.

Liu, D., Huang, C. and Feng, J. (2022) 2022: influence of assimilating wind profiling radar observations in distinct dynamic instability regions on the analysis and forecast of an extreme rainstorm event in southern China. *Remote Sensing*, 14, 3478. https://doi.org/10.3390/rs14143478.

Lorenc, A.C. (2003) The potential of the ensemble Kalman filter for NWP: a comparison with 4D-var. *Quarterly Journal of the Royal Meteorological Society*, 129, 3183–3203. https://doi.org/10.1256/qj.02.132.

Lorenz, E.N. (1963) Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20, 130–141.

Molteni, F. and Palmer, T.N. (1993) Predictability and finite-time instability of the northern winter circulation. *Quarterly Journal of the Royal Meteorological Society*, 119(510), 269–298. https://doi.org/10.1002/qj.49711951004.

Molteni, F., Tibaldi, S. and Palmer, T.N. (1990) Regimes in the wintertime circulation over northern extratropics. I: observational evidence. *Quarterly Journal of the Royal Meteorological Society*, 116(491), 31–67. https://doi.org/10.1002/qj.49711649103.

Mu, M., Duan, W. and Wang, J. (2002) The predictability problems in numerical weather and climate prediction. *Advances in Atmospheric Sciences*, 19(2), 191–204. https://doi.org/10.1007/s00376-002-0016-x.

Park, Y.-Y., Buizza, R. and Leutbecher, M. (2008) TIGGE: preliminary results on comparing and combining ensembles. *Quarterly Journal of the Royal Meteorological Society*, 134, 2029–2050.

Peña, M. and Toth, Z. (2014) Estimation of analysis and forecast error variances. *Tellus*, 66A, 21767. https://doi.org/10.3402/tellusa.v66.21767.

Privé, N. and Errico, R.M. (2013) The role of model and initial condition error in numerical weather forecasting investigated with an observing system simulation experiment. *Tellus-A*, 65, 21740. https://doi.org/10.3402/tellusa.v65i0.21740.

Privé, N.C., Errico, R.M. and Tai, K.-S. (2013) The influence of observation errors on analysis error and forecast skill investigated with an observing system simulation experiment. *Journal of Geophysical Research – Atmospheres*, 118, 5332–5346.

Shafran, P.C., Seaman, N.L. and Gayno, G.A. (2000) Evaluation of numerical predictions of boundary layer structure during the Lake Michigan ozone study (LMOS). *Journal of Applied Meteorology*, 39, 412–426.

Simmons, A.J. and Hollingsworth, A. (2002) Some aspects of the improvement in skill of numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 128, 647–677.

Stewart, L.M., Dance, S.L. and Nichols, N.K. (2013) Data assimilation with correlated observation errors: experiments with a 1-D shallowwater model. *Tellus*, 65A, 19546. https://doi.org/10.3402/tellusa.v65i0.19546.

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4507 by Fudan University, Wiley Online Library on [07/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Swanson, K.L. and Roebber, P. (2008) The impact of analysis error on medium-range weather forecasts. *Monthly Weather Review*, 136, 3425–3431.

Tang, Y.M., Shen, Z.Q. and Gao, Y.Q. (2016) An introduction to ensemble-based data assimilation method in the earth sciences. In: Lee, D., Burg, T.C. and Volos, C.K. (Eds.) *Nonlinear Systems-Design, Analysis, Estimation and Control*. Oregon: InTech.

Wallace, J.M., Lim, G.-H. and Blackmon, M.L. (1988) Relationship between cyclone tracks, anticyclone tracks and baroclinic waveguides. *Journal of the Atmospheric Sciences*, 45(3), 439–462. https://doi.org/10.1175/1520-0469(1988)045<0439:RBCTAT>2.0.CO;2.

Wang, J., Chen, J., Zhang, H., Tian, H. and Shi, Y. (2021) Initial perturbations based on ensemble transform kalman filter with rescaling method for ensemble forecasting. *Weather and Forecasting*, 36(3), 823–842. https://doi.org/10.1175/WAF-D-20-0176.1.

Wang, J.C., Zhuang, Z.R., Han, W., Lu, H.J. and Wang, R.C. (2017) Improvements and performances of the operational GRAPES_GFS 3DVar system (in Chinese). *Journal of Applied Meteorological Science*, 28(1), 11–24. https://doi.org/10.11898/1001-7313.20170102.

Wang, X., Barker, D., Snyder, C. and Hamill, T.M. (2008) A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part I: observing system simulation experiment. *Monthly Weather Review*, 136, 5116–5131. https://doi.org/10.1175/2008MWR2444.1.

Wang, X., Parrish, D., Kleist, D. and Whitaker, J. (2013) GSI 3Dvar-based ensemble-variational hybrid data assimilation for NCEP global forecast system: single-resolution experiments. *Monthly Weather Review*, 141(11), 4098–4117. https://doi.org/10.1175/MWR-D-12-00141.1.

Wei, M.Z., Toth, Z. and Zhu, Y. (2010) Analysis differences and error variance estimates from multi-centre analysis data. *Australian Meteorological and Oceanographic Journal*, 59, 25–34.

Whitaker, J.S. and Hamill, T.M. (2002) Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130, 1913–1924.

Whitaker, J.S., Hamill, T.M., Wei, X., Song, Y. and Toth, Z. (2008) Ensemble data assimilation with the NCEP global forecast system. *Monthly Weather Review*, 136(2), 463–482. https://doi.org/10.1175/2007MWR2018.1.

Wilson, A.M., Cobb, A., Martin Ralph, F., Tallapragada, V., Davis, C. and Doyle, J. (2022) Atmospheric river reconnaissance workshop promotes research and operations partnership. *Bulletin of the American Meteorological Society*, 103, E810–E816. https://doi.org/10.1175/BAMS-D-21-0259.1.

Wolyn, P.G. and McKee, T.B. (1989) Deep stable layers in the inter mountain western United States. *Monthly Weather Review*, 117, 461–472.

Yang, J., Zhang, Z., Wei, C., Lu, F. and Guo, Q. (2017) Introducing the new generation of Chinese geostationary weather satellites, Fengyun-4. *Bulletin of the American Meteorological Society*, 98, 1637–1658. https://doi.org/10.1175/BAMS-D-16-0065.1.

Yin, R.-Y., Han, W., Gao, Z. and Li, J. (2021) Impact of high temporal resolution FY-4A geostationary interferometric infrared sounder (GIIRS) radiance measurements on typhoon forecasts: Maria (2018) case with GRAPES global 4D-Var assimilation system. *Geophysical Research Letters*, 48(15), e2021GL093672. https://doi.org/10.1029/2021GL093672.

Yin, R.Y., Han, W., Gao, Z.Q. and Di, D. (2020) The evaluation of FY4A's geostationary interferometric infrared sounder (GIIRS) long-wave temperature sounding channels using the GRAPES global 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 146, 1459–1476. https://doi.org/10.1002/qj.3746.

Zhang, L., Liu, Y. and Liu, Y. (2019) The operational global four-dimensional variational data assimilation system at the China Meteorological Administration. *Quarterly Journal of the Royal Meteorological Society*, 145(722), 1882–1896. https://doi.org/10.1002/qj.3533.

How to cite this article: Feng, J., Wang, J., Dai, G., Zhou, F. & Duan, W. (2023) Spatiotemporal estimation of analysis errors in the operational global data assimilation system at the China Meteorological Administration using a modified SAFE method. *Quarterly Journal of the Royal Meteorological Society*, 1–19. Available from: https://doi.org/10.1002/qj.4507

1477870x, 0, Downloaded from https://rmets.onlinelibrary.wiley.com/doi/10.1002/qj.4507 by Fudan University, Wiley Online Library on [07/08/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License