Diagnostic Analysis of Various Observation Impacts in the 3DVAR Assimilation System of Global GRAPES

LIHONG ZHANG

Nanjing University of Information Science and Technology, Nanjing, Jiangsu, and Chengdu Institute of Plateau Meteorology, China Meteorological Administration, Chengdu, and National Meteorological Center, China Meteorological Administration, Beijing, China

JIANDONG GONG AND RUICHUN WANG

National Meteorological Center, China Meteorological Administration, Beijing, China

(Manuscript received 18 June 2017, in final form 12 July 2018)

ABSTRACT

Observation impact studies have received increasing amounts of research attention. The impacts of observations on numerical weather prediction (NWP) are highly dependent on assimilation algorithm, prediction system, and observation source. Therefore, the major NWP centers worldwide have each developed their own diagnostic techniques to assess observation impacts. However, similar diagnostic techniques have not yet been developed in China. In this study, a diagnostic technique was exploited with the randomized perturbation method in the Global/Regional Assimilation and Prediction System (GRAPES) 3DVAR system, and then applied to evaluate observation impacts for various regions of the world. It was found that a reasonable and stable estimation could be obtained when the number of perturbations was greater than 15. Because of differences in observations in the Northern and Southern Hemispheres, refractivity data from GNSS radio occultation (GNSS-RO), satellite radiance, and atmospheric motion vector data had more impact in the Southern Hemisphere than in the Northern Hemisphere. However, radiosonde data, aircraft, and surface data were more important in the Northern Hemisphere. Low-impact observation points were located in data-rich areas, whereas high-impact observation points were located in data-poor areas. In the equatorial region, the contributions of observations to the analysis were smaller than those in the nonequatorial regions because of the lack of proper mass–wind balance relationship. Radiosondes contributed the largest impact in China and its surrounding regions, with contributions of radiosondes and GNSS-RO data exceeding 60% of the total contributions, except for wind speed below 700 hPa.

1. Introduction

The objective of atmospheric data assimilation is to produce an accurate representation of the state of the atmosphere based on all existing information (Talagrand 1997). This method can provide an improved initial field for numerical models by combining various observations with an NWP product (the first-guess or background forecast) and their respective error statistics, and thus reduces errors in the initial field. With assimilation methods becoming increasingly sophisticated (Parrish

and Derber 1992; Rabier and Courtier 1992; Rabier et al. 2000; Xue et al. 2008; Huang et al. 2009; Zhang et al. 2013; Gao and Stensrud 2014; Xu et al. 2015) and the number of assimilated observations increasing rapidly, the assessment of the contribution of each observation to the analysis has become one of the most challenging topics in data assimilation and NWP (Cardinali and Healy 2014).

At present, there are two major methods for assessing the impacts of observations on the analysis. One method uses a Lanczos-conjugate gradient method to calculate the degree of freedom for signals (DFS; Fisher 2003;

Denotes content that is immediately available upon publication as open access.

Corresponding author: Jiandong Gong, gongjd@cma.gov.cn

This article is licensed under a [Creative Commons Attribution 4.0 license](http://creativecommons.org/licenses/by/4.0/) (<http://creativecommons.org/licenses/by/4.0/>).

DOI: 10.1175/MWR-D-17-0182.1

Cardinali et al. 2004; Cardinali 2009). The other method uses a randomization procedure whereby observations are perturbed, to estimate the error variance reduction caused by observations (Desroziers et al. 2005). Desroziers and Ivanov (2001) and Chapnik et al. (2004) used a randomization procedure to compute a particular matrix, which was closely associated with the DFS. Desroziers et al. (2005) applied the randomization procedure to diagnose the contributions of observations to error reduction in the analysis, and successfully conducted a diagnostic analysis in a 4DVAR system developed in France. Chapnik et al. (2006) found that in a simplified data assimilation system, the randomization procedure can be extended to a slightly nonlinear case. Brousseau et al. (2014) found that the randomization method and DFS had similar results and suggested that the variance reduction could be used as a diagnostic to improve the understanding of how observations are used.

The observation impacts on the model forecast are very important and are measured mainly by observing system experiments (OSEs; Atlas 1997; English et al. 2004; Radnoti et al. 2012) and forecast sensitivity to observations (FSO; Baker and Daley 2000; Liu et al. 2009). The OSEs (Kelly et al. 2007) can measure the impacts of these assimilated observations on all forecast metrics and can estimate the observation impacts on longer range forecasts without linear approximation constraints. The FSO approach is to calculate the forecast error sensitivity to observations by adjoint-based (Langland and Baker 2004; Cardinali 2009) or ensemble-based (Li et al. 2010; Sommer and Weissmann 2014) techniques. The FSO approach only computes the variation in the short-range forecast error caused by the assimilated observations, where the forecast error is represented and measured by a scalar function of the model parameters. The adjoint-based FSO requires development of the adjoint of a forecast model, which is complicated and is also constrained by linear assumptions. Although the ensemble-based FSO does not use the adjoint model, it requires a large number of ensemble members to assess the observation impacts on the forecast. Thus, the FSO method is complementary to the OSEs for estimating observation impacts in a forecasting system (English et al. 2004; Radnoti et al. 2012).

Since an observation impact study can provide valuable information for the improvement of NWP and the observing system, almost all the major NWP centers worldwide have developed their own diagnostic techniques to assess the impacts of observations on NWP (Cardinali and Prates 2011; Cardinali 2013; Auligné et al. 2011; Brousseau et al. 2014; Joo et al. 2013; Cucurull and Anthes 2014). At present, the OSE is the

most important method for estimating the observation impacts in the operational Global/Regional Assimilation and Prediction System (GRAPES) global system. This method, however, requires a large amount of computational resources. The observation impacts on NWP are highly dependent on the assimilation algorithm, prediction system, and observation sources. Thus, it is necessary to perform diagnostic analysis for a specific assimilation and prediction system.

Since the observation changes first led to the variations of model initial fields and then caused the variations of model forecast, estimation of the impacts of the observations on the analysis is the basis for measuring the observation impacts on the forecast. Therefore, this paper focuses on the observation impacts on the analysis. The operational GRAPES three-dimensional variational (3DVAR) assimilation system and the randomization method with perturbed observations are described in section 2, which are important components of the diagnostic technique in this paper. The technique is validated in section 3 and applied to evaluate the impacts of various observations on different regions in section 4. Final conclusions and discussions are presented in section 5.

2. Diagnostic method, assimilation system, and observations

a. Randomization method

In this paper, a diagnostic technique is developed, which is based on a randomization method (Desroziers et al. 2005) and the GRAPES 3DVAR system. A detailed introduction to this method with perturbed observations is provided here, including the transformation operator and terms related to error variance reduction (EVR) of state variables.

1) THEORETICAL EQUATIONS

Various observations contribute differently to the reduction of the background error variance. Impacts of various observations on the analysis can be obtained through the randomization method with perturbed observations.

Cohn (1997) pointed out that when the background error covariance and observation error covariance are reasonably specified, the analysis error covariance matrix could be expressed as

$$\mathbf{A} = \mathbf{B} - \mathbf{KHB} = \mathbf{B} - \mathbf{AH}^T \mathbf{R}^{-1} \mathbf{HB}, \quad (1)$$

where $\mathbf{A}$ and $\mathbf{B}$ are the analysis and background error covariance matrices, respectively; $\mathbf{H}$ is the tangent linear observation operator; $\mathbf{K}$ is the gain matrix; and

$\mathbf{R}$ is the observation error covariance matrix. $\mathbf{KHB}$ or $\mathbf{AH}^T \mathbf{R}^{-1} \mathbf{HB}$ represents the error reduction term after all observations are assimilated. In the GRAPES 3DVAR system, $\mathbf{R}$ is assumed to be a diagonal matrix. $\mathbf{AH}^T \mathbf{R}^{-1} \mathbf{HB}$ can be expressed as the sum of error covariance reduction of each individual observation dataset:

$$\mathbf{AH}^T \mathbf{R}^{-1} \mathbf{HB} = \sum_i \mathbf{AH}_i^T \mathbf{R}_i^{-1} \mathbf{H}_i \mathbf{B}, \quad (2)$$

where $\mathbf{H}_i$ and $\mathbf{R}_i$, respectively, are the linearized observation operator and observation error matrix of the observation subset i . If defining a projection operator π_i , by which the observation subset i can be obtained from the complete set of observations, $\mathbf{H}_i$ and $\mathbf{R}_i$ can be written as $\mathbf{H}_i = \pi_i \mathbf{H}$ and $\mathbf{R}_i = \pi_i \mathbf{R} \pi_i^T$.

Define $\mathbf{S}_i$ as $\mathbf{S}_i = \pi_i^T \pi_i$ in which the diagonal element s_{ii} is equal to 1 and the remaining elements are 0, and then we have:

$$\mathbf{AH}_i^T \mathbf{R}_i^{-1} \mathbf{H}_i \mathbf{B} = \mathbf{AH}^T \mathbf{R}^{-1} \mathbf{S}_i \mathbf{HB} = \mathbf{KS}_i \mathbf{HB}. \quad (3)$$

Based on Eqs. (1) and (3), it is clear that $\mathbf{KHB}$ and $\mathbf{KS}_i \mathbf{HB}$ are the error reductions caused by assimilating all observations and the observation subset i , respectively. According to the matrix trace definition, the EVR can be obtained by the matrix traces of $\mathbf{KHB}$ and $\mathbf{KS}_i \mathbf{HB}$. The EVR of all observations can be written as

$$r = \text{Tr}(\mathbf{KHB}).$$

Similarly, the EVR of the observation subset i can be written as

$$r_i = \text{Tr}(\mathbf{KS}_i \mathbf{HB}).$$

Here r and r_i can only be expressed in theory. Because $\mathbf{KHB}$ and $\mathbf{KS}_i \mathbf{HB}$ contain various state variables along the diagonal, which have different units, it is impossible to solve $\mathbf{KHB}$ and $\mathbf{KS}_i \mathbf{HB}$ directly in the operational 3DVAR system. Therefore, a transformation operator and an approximate method are applied to obtain the estimates of variance reduction. By using the transform operator $\mathbf{L}$, the EVR for the variable of interest is obtained from that of multiple state variables, and can be written as

$$t = \text{Tr}(\mathbf{LKHBL}^T). \quad (4)$$

Similarly, the EVR caused by the observation subset i is written as

$$t_i = \text{Tr}(\mathbf{LKS}_i \mathbf{HBL}^T). \quad (5)$$

The definition of $\mathbf{L}$ is very flexible, that is, it can be a space projection operator for projections to an isobaric

surface or projections over a certain horizontal range, or it can be a variable transform operator from state variables to other variables.

In the next section, a randomization method is used to solve Eqs. (4) and (5).

2) SOLUTIONS OBTAINED BY THE RANDOMIZATION METHOD

Equation (4) is transformed first in terms of the nature of matrix trace, that is, $\text{Tr}(\mathbf{PQ}) = \text{Tr}(\mathbf{QP})$, and $\mathbf{R}^{1/2}$ is also introduced. We have

$$t = \text{Tr}(\mathbf{R}^{-1/2} \mathbf{HBL}^T \mathbf{LKR}^{1/2}). \quad (6)$$

Girard (1989) suggested that a large matrix trace could be estimated using the Monte Carlo method. That is, for the matrix $\mathbf{Q}$, $\text{Tr}(\mathbf{Q}) \cong \boldsymbol{\eta}^T \mathbf{Q} \boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is a random number vector that satisfies the standard normal distribution. Then, a randomized estimation of t is given by

$$t \cong \boldsymbol{\eta}^T \mathbf{R}^{-1/2} \mathbf{HBL}^T \mathbf{LKR}^{1/2} \boldsymbol{\eta}. \quad (7)$$

In an operational 3DVAR data assimilation system, the gain matrix $\mathbf{K}$ is not explicitly known and computed. It is necessary to transform all those terms that include $\mathbf{K}$. For a linear system or a weakly nonlinear system, after the assimilation of observations $\mathbf{y}^o$, the analysis field $\mathbf{x}^a$ can be written as

$$\mathbf{x}^a(\mathbf{y}^o) = \mathbf{x}^b + \mathbf{Kd},$$

where $\mathbf{x}^b$ is the background field, $\mathbf{d}$ is the innovation vector, and $\mathbf{d} = \mathbf{y}^o - \mathbf{Kx}^b$. With the assimilation of perturbed observations, the analysis field is expressed as

$$\begin{aligned} \mathbf{x}^a(\mathbf{y}^o + \delta \mathbf{y}^o) &\cong \mathbf{x}^a(\mathbf{y}^o) + \mathbf{K} \delta \mathbf{y}^o \\ &\cong \mathbf{x}^a(\mathbf{y}^o) + \delta \mathbf{x}^a(\delta \mathbf{y}^o), \end{aligned} \quad (8)$$

where $\mathbf{K} \delta \mathbf{y}^o$ is the analysis increments caused by observation perturbations $\delta \mathbf{y}^o$. According to Eq. (8), although $\mathbf{K}$ is unknown, $\mathbf{K} \delta \mathbf{y}^o$ can be approximated by the difference between analyses with perturbed observations $[\mathbf{x}^a(\mathbf{y}^o + \delta \mathbf{y}^o)]$ and unperturbed observations $[\mathbf{x}^a(\mathbf{y}^o)]$. Defining $\delta \mathbf{y}^o$ as $\delta \mathbf{y}^o = \mathbf{R}^{1/2} \boldsymbol{\eta}$, allows Eq. (7) to be written as

$$\begin{aligned} t &\cong \delta \mathbf{y}^{oT} \mathbf{R}^{-1} \mathbf{HBL}^T \mathbf{LKR}^{1/2} \boldsymbol{\eta} \\ &= \delta \mathbf{y}^{oT} \mathbf{R}^{-1} \mathbf{HB}^{1/2} \mathbf{B}^{T/2} \mathbf{L}^T \mathbf{L} \delta \mathbf{x}^a(\delta \mathbf{y}^o). \end{aligned} \quad (9)$$

Similarly, the perturbation of the observation subset i is defined as $\delta \mathbf{y}_i^o = \pi_i \delta \mathbf{y}^o$, and the EVR caused by the observation subset i can be expressed as

$$t_i \cong \delta \mathbf{y}^o \mathbf{R}_i^{-1} \mathbf{S}_i \mathbf{HBL}^T \mathbf{LK} \delta \mathbf{y}^o \\ = \delta \mathbf{y}_i^{oT} \mathbf{R}_i^{-1} \boldsymbol{\pi}_i \mathbf{HB}^{1/2} \mathbf{B}^{T/2} \mathbf{L}^T \mathbf{L} \delta \mathbf{x}^a (\delta \mathbf{y}^o). \quad (10)$$

The EVR of the variable of interest can be obtained from all these known variables. The randomization method with perturbed observations can be applied to both linear and weakly nonlinear systems (Desroziers et al. 2009).

Finally, if the assimilated observations have positive contributions to the analysis, the EVR will be greater than zero and equal to or smaller than the background error variance, that is,

$$0 < \text{Tr}(\mathbf{LKHBL}^T) \leq \text{Tr}(\mathbf{LBL}^T). \quad (11)$$

Equation (11) is an important diagnostic formula. Whether or not the EVR caused by assimilation of observations is reasonable can be determined in terms of the ratio of $\text{Tr}(\mathbf{LKHBL}^T)$ to $\text{Tr}(\mathbf{LBL}^T)$. The percentage of EVR $\text{Tr}(\mathbf{LKHBL}^T)$ in total background error variance $\text{Tr}(\mathbf{LBL}^T)$ can be obtained, which can improve understanding of the diagnostic results.

b. GRAPES global 3DVAR system

GRAPES is a global and regional NWP system developed by the China Meteorological Administration (CMA; Chen et al. 2008, 2012; Xue et al. 2003, 2008). The data assimilation system and the forecast model are the most important components of GRAPES. Since 1 June 2016, the global GRAPES has been used operationally in the Numerical Weather Prediction Center (NWPC) of the CMA (Liu et al. 2016; J.-C. Wang et al. 2017).

The incremental form of the 3DVAR assimilation scheme is used in the data assimilation system (Xue et al. 2008), and the GRAPES 4DVAR is currently under development (Liu et al. 2017; Zhang and Liu 2017). The objective function is as follows:

$$J(\delta \mathbf{x}) = J^b(\delta \mathbf{x}) + J^o(\delta \mathbf{x}) = \frac{1}{2} \delta \mathbf{x}^T \mathbf{B}^{-1} \delta \mathbf{x} \\ + \frac{1}{2} (\mathbf{H} \delta \mathbf{x} + \mathbf{d})^T \mathbf{R}^{-1} (\mathbf{H} \delta \mathbf{x} + \mathbf{d}). \quad (12)$$

This assimilation system employs the same height-based terrain-following vertical coordinate and atmospheric state variables as those of the GRAPES model. Variables are staggered in the vertical direction following the Charney–Phillips approach (Charney and Phillips 1953; Cullen et al. 1997). An Arakawa C grid staggering (Arakawa and Lamb 1977) is applied in the horizontal direction. In the operational system, there are 60 levels in the vertical direction from the surface to the model top (about 37.7 km) and the horizontal resolution is 0.25°. There is a single iteration in the outer loop. The

inner loop ends when the norm of the gradient obtained in the minimization process is 10^{-3} of the original norm or the maximum number of iterations is 150. Because of the limitation of computational resources, 0.5° horizontal resolution was used in this study.

To effectively reduce the tremendous cost of minimization and make convergence easier, a series of transforms (Wang et al. 2012; R.-C. Wang et al. 2015; Gong et al. 2016) were performed, such as physical variable transform, balance constraints, and preconditioning transform. J.-C. Wang et al. (2014) used the National Meteorological Center (NMC) method (Parrish and Derber 1992) to re-estimate the background error covariance of the GRAPES 3DVAR system.

There are the corresponding observation operators for various observations in the GRAPES 3DVAR system. The observation error covariance matrix is assumed to be a diagonal matrix and is mainly obtained by inheriting the previous operational settings (Gong et al. 2006; Tao et al. 2006, 2008; Zhuang et al. 2006) or using the innovation vector method (Xue et al. 2013; J.-C. Wang et al. 2015). There are some quality control schemes for observations in the GRAPES assimilation system, such as the background check, the variational quality control for conventional observations (Hao et al. 2013), and the radiosonde observation blacklisting check (Zhuang et al. 2014). Quality control for satellite radiance data is performed from five aspects: channel selection (G. Wang et al. 2014a), outlier elimination (G. Wang et al. 2014b; Wang et al. 2016), bias correction (Han 2014), cloud detection (Chen and Li 2011; Zhu et al. 2013), and data thinning (G. Wang et al. 2017).

c. Observations

At present, the operational assimilation system can assimilate upper-air observations of wind, temperature, and humidity by radiosondes (TEMP); surface pressure from land stations and ships (SYNOP); wind data from aircraft measurements (AIREP); atmospheric motion vector data inferred from cloud motions detected by geostationary satellites (AMV); refractivity radio occultation data from the Global Navigation Satellite System (GNSS-RO); and radiance data from Advanced Microwave Sounding Unit-A measurements (AMSU-A) carried by polar-orbiting satellites. In this paper, clear-sky AMSU-A radiances are assimilated, including the observations over the ocean of channels 5 and 6, and the global observations of channels 7–10. Many studies have shown that the forecast accuracy of the GRAPES can be improved by assimilating these observations (Zhang et al. 2004; Zhang et al. 2007; Ding et al. 2010; Li et al. 2012; J.-C. Wang et al. 2015; Zhuang et al. 2014; Liu and Xue 2014).

FIG. 1. Horizontal distributions of all assimilated observations at 0000 UTC 20 Jun 2014. The black rectangle denotes the target verification domain (0° – 55° N, 70° – 136° E). TEMP, observations from radiosondes; SYNOP, surface pressure from land stations and ships; AIREP, wind data from aircraft measurements; GNSS-RO, refractivity data of radio occultation from the Global Navigation Satellite System (GNSS); AMV, atmospheric motion vector; and AMSU-A, satellite microwave sounder radiances.

As a result of differences in the various measurement methods, the spatial distribution of each observation type has its own characteristics. The observations at 0000 UTC 20 June 2014 were taken as an example to show the horizontal distribution of various observations assimilated by the GRAPES 3DVAR system (Fig. 1) during a 6-h assimilation window. TEMP, SYNOP, and AIREP data are mainly distributed over land, and more observations are available in the Northern Hemisphere than in the Southern Hemisphere and the equatorial region. TEMP and SYNOP data are particularly dense in China and its surrounding area, especially in East China. However, more AMV data are collected over the ocean than over land, and most of the AMV data are concentrated between 60° S and 60° N. GNSS-RO and AMSU-A data are widely distributed over land and the ocean. Table 1 lists the number of assimilated

observations for each observation type shown in Fig. 1. Although the vertical resolutions of TEMP and GNSS-RO data are higher than other observations, there are major differences between the two. TEMP data only cover land and contain three observation types (wind, temperature, and humidity) that can be assimilated, whereas GNSS-RO data are distributed over both land and ocean and are assimilated as the refractivity information. Thus, there are more TEMP data than GNSS-RO data in the Northern Hemisphere, and the opposite is found in the Southern Hemisphere.

3. Stability and rationality of the diagnostics

The randomization method with perturbed observations can only provide approximate estimations. Comparative experiments need to be conducted for any

TABLE 1. Number of assimilated observations for each observation type at 0000 UTC 20 Jun 2014. Note: w is wind, t is temperature, rh is relative humidity, N is refractivity, ps is surface pressure, and ra is satellite radiance. The temperature observations of AIREP are not assimilated because of the decoding errors during 1–30 Jun 2014.

Region	TEMP			GNSS-RO	SYNOP	AIREP	AMV	AMSU-A
	w	t	rh	N	ps	w	w	ra
Southern Hemisphere	2118	1475	718	12 539	895	2405	4619	36 174
Equatorial region	4097	3468	2390	4274	871	2963	5800	23 114
Northern Hemisphere	16 496	17 054	10 256	10 082	4500	32 301	5414	26 358
China and its surrounding area	7396	5623	3700	1704	1840	5720	509	4337

terms that may affect the results to ensure the stability and rationality of the estimation. These experiments were performed for 0000 UTC 20 June 2014 over the region (0° – 55° N, 70° – 136° E) with a focus on wind, temperature, and specific humidity at standard isobaric surfaces.

a. Appropriate number of perturbations

When using the Monte Carlo method to estimate the influence-matrix trace (Desroziers et al. 2005), multiple perturbations are required to obtain stable estimation. It is critical to determine the number of perturbations that can ensure the stability of estimation results but also preserve computational resources. In this section, the impact of the perturbation number on the estimation is discussed and the appropriate number of perturbations is determined based on all observations and different observation types.

Figure 2 presents the percentage of the estimate of the error variance reduction obtained with various numbers of perturbations compared with the mean estimate

obtained with 50 perturbations for all the assimilated observations (Fig. 2a) and different observation-type subsets (Fig. 2b). Here, the average of 50 different estimates is used as a reference, which can be denoted by t_{50} , and the mean estimate of k perturbations ($k \leq 50$) can be denoted by t_k . The closer the percentage of t_k in t_{50} is to 100%, the more stable the mean estimate t_k , and vice versa. Figure 2a displays the percentage results of all the assimilated observations. When $k > 10$, the mean estimates of all the observations for wind speed and specific humidity are within $\pm 10\%$ of t_{50} ; when $k > 15$, the mean estimates for wind, temperature, and specific humidity stay within $\pm 5\%$ of t_{50} as k is increased. Figure 2b presents the results for various observation types. When the observations are divided into subsets of different types, the mean estimates for each observation subset cannot converge as fast as those for all observations. For $k > 15$, the mean estimates for these observation subsets, except AMSU-A, are within $\pm 10\%$ of t_{50} for the same observation subset, but mainly outside $\pm 5\%$ of t_{50} as k is continuously increased.

FIG. 2. Percentage of the estimate of the error variance reduction obtained with various numbers of perturbations compared with the mean estimate obtained with 50 perturbations for (a) all assimilated observations and (b) different observation-type subsets at 0000 UTC 20 Jun 2014. The abscissa is the number of perturbed analyses used to obtain the estimate. The numbers shown in each panel indicate pressure levels with the unit of hPa: W , wind speed; T , temperature; Q , specific humidity; and other abbreviations are as in Fig. 1.

TABLE 2. Experimental design for the verification domain.

Expt	Observations	No. of stations	Target domain
Exp_out	Radiosondes outside the verification domain	485	0°–55°N, 70°–136°E
Exp_in	Radiosondes inside the verification domain	237	
Exp_all	All radiosondes	722	

b. Rationality and flexibility of the diagnostics

When the randomization method with perturbed observations are first introduced (Desroziers et al. 2005), it is suggested that the operator $\mathbf{L}$ can be a variable transform operator or a space projection operator. In the study of observation contributions to EVR in the verification domain, the operator $\mathbf{L}$ consists of variable transform operator ($\mathbf{P}$) and space projection operator ($\mathbf{G}$) simultaneously. Equation (7) can be written as

$$t \cong \delta \mathbf{y}^o \mathbf{T}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{B}^{1/2} \mathbf{B}^{T/2} \mathbf{P}^T \mathbf{G}^T \mathbf{G} \mathbf{P} \delta \mathbf{x}^a (\delta \mathbf{y}^o). \quad (13)$$

The observation impact study can be conducted in a more flexible way with more details using Eq. (13). For example, in the observation impact study, the observations and the analysis variables affected can be in the same region or in different regions. Using an ideal model, Desroziers et al. (2005) investigated the impacts of observations inside and outside the verification domain. In the present study, three comparative experiments with assimilation of TEMP data were designed (Exp_in, Exp_out, and Exp_all; Table 2) in the operational GRAPES global 3DVAR assimilation system.

The vertical distributions of EVR for different analysis variables in the three experiments are presented in Fig. 3, and clearly shows that the EVR in Exp_out is much smaller than those in Exp_in and Exp_all, although there are 485 sounding stations outside the verification domain. Results in Exp_in are only slightly lower than those in Exp_all, suggesting that observations in the verification domain make the largest contribution to background EVR in the GRAPES 3DVAR system. This supports the findings of Desroziers et al. (2005). The EVR distributions (not shown) also indicate that the observations in the verification domain have the largest impacts on the analysis.

These results are consistent with the 3DVAR theory and indicate that the diagnostic tools are consistent. The findings also reflect the following characteristics of 3DVAR method: the spatial spreading of information from observation points to analysis grids can only be performed in a finite domain surrounding observation points since the $\mathbf{B}$ matrix is usually derived from assumptions of homogeneity and isotropy, and without flow-dependent attributes, and the decorrelation length is limited within a few hundred kilometers.

FIG. 3. Vertical distributions (hPa) of error variance reductions in the three experiments for (a) wind speed, (b) temperature, and (c) specific humidity analysis at 0000 UTC 20 Jun 2014. See Table 2 for details of the three experiments.

4. Impact estimation of observations for various regions

At present, GRAPES has been running operationally in the NWPC (CMA). The impacts of various observations in different regions will be helpful for further improvement of assimilation technique and better application of observations. In the following, daily analyses at 0000 UTC throughout June 2014 are used as examples to analyze the impacts of different observation types for various regions. Four groups of experiments were conducted: Exp_nh (the Northern Hemisphere, 20°–90°N), Exp_sh (the Southern Hemisphere, 20° to 90°S), Exp_eq (the equatorial region, 20°S–20°N), and Exp_cn (China and its surrounding area, 0°–55°N, 70°–136°E). China and its surrounding area were examined because TEMP and SYNOP data are densely distributed and the terrain is very complicated in this region.

a. Global impacts

1) IMPACTS OF ALL OBSERVATIONS ASSIMILATED

To investigate the impacts of all observations assimilated in the GRAPES 3DVAR system, Figs. 4a–c display the total EVR and the percentage of $[\text{Tr}(\mathbf{LKHBL}^T)]$ compared with the background error variance $\text{Tr}(\mathbf{LBL}^T)$ for different variables in the three zones at 0000 UTC during 1–30 June 2014. For the wind speed analysis, the maximum EVR of the three zones was mainly between 200 and 300 hPa, which is basically the same as the locations of their maximum values of $\text{Tr}(\mathbf{LBL}^T)$. Below 300 hPa, the EVR was similar in the Southern and Northern Hemisphere and considerably larger than that in the equatorial region. Above 300 hPa, however, the EVR in the Northern Hemisphere was larger than that in the Southern Hemisphere. From 300 to 200 hPa, the contribution to wind speed grew faster in the equatorial region than in the Southern Hemisphere and thus total EVR in the equatorial region was larger than that in the Southern Hemisphere above 250 hPa. The percentage results (lines in Fig. 4a) show that total EVR in the Northern Hemisphere reached 25%–50% of its $\text{Tr}(\mathbf{LBL}^T)$, which was the largest of the three regions. Below 250 hPa, the percentage of the Southern Hemisphere was between the percentages of the other two regions, but above 250 hPa, less than the percentages of the two regions. The results presented in Fig. 4a may be related to the high number of observations of SYNOP, TEMP, AMV, and AIREP in the Northern Hemisphere. In addition, the number of AMV (5999) and AIREP (2598) observations in the equatorial region is larger than that of AMV (4753) and AIREP (2293)

(Fig. 4d) in the Southern Hemisphere, and AMV data are mainly distributed at 300–200 hPa.

The percentage changes in temperature (Fig. 4b) varied among zones with the largest values in the Northern Hemisphere, followed by the Southern Hemisphere, and the smallest values in the equatorial region. This result is consistent with the number of mass field observations in the three regions.

The total EVR and $\text{Tr}(\mathbf{LBL}^T)$ for the specific humidity analysis (Fig. 4c) were mainly below 300 hPa. Therefore, percentage results for specific humidity below 300 hPa were further examined. Figure 4c shows that below 300 hPa, where water vapor is concentrated, the minimum percentages of the EVR for specific humidity were in the equatorial region and the EVR percentages in the Southern Hemisphere were slightly larger than those in the Northern Hemisphere. This is probably because there were more GNSS-RO data assimilated in the Southern Hemisphere (11 108) than in the Northern Hemisphere (9126) or because it was summer in the Northern Hemisphere during the study period and there were more uncertainty in the moist field. However, in these three regions, the maximum percentages for specific humidity were less than 20% and the mean values were also less than those for wind speed and temperature. This was mainly because humidity information assimilated in the operational assimilation system were provided by TEMP and GNSS-RO data and less than wind and temperature information provided by various observations.

The results of the wind and temperatures analyses are in good qualitative agreement with previous studies indicating that reanalysis errors are smaller in the Northern Hemisphere than those in the Southern Hemisphere (Kalnay et al. 1996; Kistler et al. 2001). The GRAPES 3DVAR provided a similar finding as ECMWF 4DVAR (Cardinali et al. 2004) that the average contribution of all observations to the analysis of the Northern Hemisphere was larger than that of the Southern Hemisphere. There is a difference regarding observation impacts in the equatorial between the GRAPES 3DVAR and ECMWF 4DVAR that the average contribution in the equatorial regions in the GRAPES 3DVAR was smaller than that in the Northern and Southern Hemispheres, whereas in ECMWF 4DVAR, the average contribution of all observations in the equatorial regions was larger than that in the Northern Hemisphere. This is closely related with the difference in the assimilation method and **B** matrix used by the two NWP centers. It has been previously indicated that there are more complicated mass–wind balance relationships in the equatorial region than at higher latitudes (Zagar et al. 2004; R.-C. Wang et al. 2015). Thus, the NMC method does not yield

FIG. 4. Total error variance reduction $[Tr(LKHBL^T)]$ due to the complete observation set, the percentage of $[Tr(LKHBL^T)]$ compared with the background error variance $Tr(LBL^T)$ for (a) wind speed, (b) temperature, and (c) specific humidity analysis, and the average number of each observation type in three different regions at 0000 UTC throughout June 2014. The abbreviations used in this figure are as follows: eq, the equatorial region; nh, Northern Hemisphere; sh, the Southern Hemisphere; %, the percentage of $[Tr(LKHBL^T)]$ compared with $Tr(LBL^T)$; TEMP.W, TEMP.T, and TEMP.RH are the wind, temperature, and relative humidity observations by radiosondes (TEMP), respectively. Other abbreviations are as in Fig. 1.

good estimates for the desired background error statistics (Errico et al. 2015), leading to inefficient data assimilation in the equatorial region. Therefore, it is necessary to improve the **B** matrix from the balance-related and multivariate aspects, use a more complicated assimilation method, or assimilate more observations related to humidity.

2) IMPACTS OF VARIOUS OBSERVATION TYPES

Different observation types have different contributions to the total EVR. Figure 5 shows the percentage of wind speed EVR compared with the total EVR at various isobaric surfaces caused by different observation types at 0000 UTC during 1–30 June 2014. In the Southern Hemisphere at 150 hPa and below (Fig. 5a), in terms of the average contribution to the wind speed from large to small, observation subsets were as follows: AMV, AMSU-A, GNSS-RO, TEMP, SYNOP, and AIREP. The contribution of SYNOP was mainly below 500 hPa, and the largest contributions of GNSS-RO and AIREP

occurred at around 200 hPa. Above 150 hPa, only three observation subsets showed important contributions, in descending order: TEMP, AMSU-A, and GNSS-RO. In the Northern Hemisphere (Fig. 5c), the types of observation subsets that contribute to the wind speed analysis were the same as those in the Southern Hemisphere, but the percentages of contributions were different. The percentages of AMV, GNSS-RO, and AMSU-A decreased, and those of SYNOP, TEMP, and AIREP clearly increased in the Northern Hemisphere as the number of conventional observations increased. TEMP data became the largest contributor to the improvement of wind analysis. In the equatorial region (Fig. 5b), AMV made the largest contribution at 150 hPa and below, accounting for more than 42% of the total EVR, with the remainder from TEMP and AIREP. In contrast, above 150 hPa, TEMP accounted for more than 93% of the total EVR. Figure 5b shows that in the equatorial region, the contributions of SYNOP, GNSS-RO, and AMSU-A data were very small, which

FIG. 5. Percentage of wind speed error variance reduction compared with the total variance reduction at various isobaric surfaces due to different observation subsets at 0000 UTC throughout June 2014 in (a) the Southern Hemisphere, (b) the equatorial region, and (c) the Northern Hemisphere. All the abbreviations are as in Fig. 1.

was different to the findings for the Northern and Southern Hemispheres because of the different dynamic balance constraints (R.-C. Wang et al. 2015) and because of fewer SYNOP, GNSS-RO, and AMSU-A data in the equatorial region (Fig. 4d). This indicates that data density is one of the major reasons for the small impact.

For the temperature EVR analysis, in the Southern Hemisphere (Fig. 6a) where conventional observations are sparse, GNSS-RO data were the largest contributor in most cases. The contribution of GNSS-RO increased with height to more than 60% above 500 hPa. The AMSU-A data had the second largest contribution, with a minimum value still larger than 20%. The SYNOP data were the main contributor at the lowest atmospheric levels. The TEMP data had some contribution at all isobaric surfaces but the maximum value was less than 15%. The Northern Hemisphere was different from the Southern Hemisphere because of the higher density of conventional observations. TEMP data were the largest contributor to temperature EVR in the Northern Hemisphere, whereas the second largest contributor was SYNOP data below 500 hPa or GNSS-RO data above 500 hPa. The contribution of AMSU-A data was about 15% at all isobaric surfaces (Fig. 6c). The impacts of AMV and AIREP on temperature were indirect and small in both hemispheres because only wind

data were assimilated from these subsets and their impacts on the temperature analysis can only be realized through the cross correlations of the GRAPES 3DVAR **B** matrix. In the equatorial region (Fig. 6b), the averages of percentage contributions of the observation subsets decreased in the order TEMP > AMSU-A > SYNOP > GNSS-RO below 300 hPa. In contrast, the GNSS-RO becomes the largest contributor to temperature EVR, followed by AMSU-A and TEMP above 300 hPa.

The specific humidity EVR contributions (figure not shown) in the three regions below 300 hPa showed that the same types of observations that contribute to the specific humidity analysis and the orders of the percentage contributions of these observations is similar, with the largest contribution of GNSS-RO, followed by that from TEMP, AMSU-A, and SYNOP. The percentage values of the same type observations in the three regions, however, were different. For example, below 300 hPa, the contributions of TEMP were less than 10% in the Southern Hemisphere, 10%–20% in the equatorial region, and more than 25% in the Northern Hemisphere.

Here, the contributions of TEMP and AMSU-A to temperature supported some of the following findings of the French operational ARPEGE 4DVAR assimilation system (Desroziers et al. 2005). The contribution of TEMP data, in descending order, was largest in the

FIG. 6. As in Fig. 5, but for the temperature analysis.

Northern Hemisphere and smallest in the Southern Hemisphere, whereas the opposite was found for the AMSU-A data. No covariances between humidity and other fields were considered in the **B** matrix of the GRAPES 3DVAR and ARPEGE 4DVAR systems. In the ARPEGE 4DVAR system, if the assimilated observations did not contain humidity information, these observations had no effect on the humidity analysis, such as SYNOP. However, the observation operator of the GRAPES 3DVAR takes it into account that temperature changes can cause humidity adjustments, so AMSU-A and SYNOP data, which do not contain the humidity observation information, have impacts on humidity analysis. In addition, those observations related to mass fields (such as surface pressure, AMSU-A, and GNSS-RO data), can influence the wind analysis in the extra-equatorial region, but have little effect on the wind analysis in the equatorial region because of lacking proper mass–wind dynamic balance relations in this region. These conclusions indicate that the **B** matrix plays an important role in variational data assimilation systems.

3) IMPACT OF SINGLE OBSERVATION

The number of observations of the same type is different in the three regions, so it is necessary to estimate the impact of single observation. Figures 7a–c shows the mean EVR for the wind speed analysis of single observation of the same type in the different regions. The

mean impact of single observation for wind speed in the Southern Hemisphere was always larger than that in the Northern Hemisphere at the same isobaric surface. In conjunction with Fig. 4d, it can be seen that the average number of the same types of observations in the Northern Hemisphere was larger than that in the Southern Hemisphere, except for GNSS-RO and AMSU-A data. These results indicate that low-impact data points occur in data-rich areas and high-impact data points are in data-sparse regions, which is consistent with the findings of ECMWF (Cardinali 2013). In addition, the EVR change of single observation in the Northern and Southern Hemispheres was significantly different from that in the equatorial region, and the impact of single observation in the equatorial region is smaller than that in the Northern and Southern Hemispheres in most cases below 300 hPa, but considerably larger than that in the Northern Hemisphere above 300 hPa. The conclusion above may be related to the dynamic balance constraint scheme of the **B** matrix and observation density.

Similar features were found in the contribution of a single observation for the temperature analysis in the three regions (figure not shown). It is noteworthy that although the impact of all AMSU-A data for the temperature (Fig. 6) was evident in the three regions, the impact of single AMSU-A was very small, which is consistent with the conclusions from ECMWF (Cardinali 2013).

FIG. 7. Average error variance reduction for the wind speed analysis of a single observation of each observation type subset in (a) the Southern Hemisphere, (b) the equatorial region, and (c) the Northern Hemisphere at 0000 UTC throughout June 2014. All abbreviations are as in Fig. 1.

Figure 8 displays the average EVR of single observation for the specific humidity analysis. Although the contributions to the specific humidity analysis again mainly came from TEMP and GNSS-RO data in the three regions, the differences between the contributions of single TEMP and single GNSS-RO were obvious. In the isobaric surface where water vapor is more concentrated (925–500 hPa), the contribution of single GNSS-RO data was considerably greater than that of TEMP in the equatorial region and the Northern Hemisphere where the density of TEMP data is large. In contrast, the contribution of single GNSS-RO was close to the contribution of single TEMP in the Southern Hemisphere where TEMP data are sparse.

b. Situation in China and its surrounding area

Vertical profiles of average EVR by different observation types in the verification domain at 0000 UTC during 1–30 June 2014 are shown in Fig. 9. It is shown that TEMP makes the largest contribution to wind speed

EVR and the maximum reduction occurred at 10 hPa, whereas AIREP and AMV made the second largest contribution and their maxima were at around 200 hPa. The above three observation types can reduce the background wind speed error by half. The contribution of observation subsets to temperature EVR (Fig. 9b) decreased in the order TEMP > GNSS-RO > AMSU-A > SYNOP data. After assimilating all observations, the background temperature error variance could be reduced by more than 60% at 700 hPa and above, of which the TEMP data contributed more than two-thirds. For specific humidity EVR (Fig. 9c), TEMP and GNSS-RO data made the largest contributions, mainly below 300 hPa with the maximum reduction occurring at 700 hPa. However, the total EVR reached less than 20% of the background error variance of specific humidity.

Figure 10 presents the average EVR of single observation in each observation subset at various isobaric surfaces. Figure 10a indicates that although the TEMP subset made the largest contribution to wind speed

FIG. 8. As in Fig. 7, but for the specific humidity analysis.

EVR, the contribution of single TEMP observation was less than that of single AMV observation in most cases. This is likely because the average density of TEMP in China is larger than that of AMV, and TEMP observations only cover land, while AMV data are more distributed over the ocean. The largest contributions of single TEMP, AMV, and AIREP data all occurred at around 200–250 hPa, which is because the maximum value of wind speed error variance in the **B** matrix below 50 hPa was concentrated at 200–250 hPa and the horizontal correlation length scale of the related state variables in **B** was larger at the higher level. Figure 10b suggests that single GNSS-RO observation had the largest contribution to the temperature analysis above 300 hPa and single TEMP observation had the largest contribution below 300 hPa. In the middle and lower troposphere, although the contribution of the SYNOP subset was not distinct, the contribution of single SYNOP was more important mainly because most of the SYNOP observations were concentrated below 700 hPa. The contribution of single observation to specific humidity EVR was smaller than that to wind speed and

temperature EVR. Single GNSS-RO observation had the largest contribution to the specific humidity analysis, followed by single TEMP. Single observation of other subsets made little contribution to the specific humidity analysis.

5. Conclusions and discussion

Observation impacts on NWP is highly dependent on assimilation algorithm, prediction system, and observation sources. In the present study, the EVR estimation method proposed in Desroziers et al. (2005) was applied in the global GRAPES operational 3DVAR system to evaluate impacts of observations on the analysis results for various regions of the world. The major conclusions are as follows.

In the GRAPES global 3DVAR operational system, 16 perturbed analyses are necessary to obtain a stable estimation of the EVR and also sufficient for estimating the EVR of every observation type subset. To investigate the impacts of all assimilated observations on the analysis, the total EVR was compared with the background error variance in different regions. For the wind

FIG. 9. Vertical profiles of average error variance reduction due to different observation subsets in the target verification domain for (a) wind speed, (b) temperature, and (c) specific humidity analysis at 0000 UTC throughout June 2014. The target verification domain and all the abbreviations are as in Fig. 1. Tr(LBL^T) is the background error variance of the variable of interest.

speed and temperature analysis, the percentage of total EVR was larger in the Northern Hemisphere than that in the Southern Hemisphere, because there was a greater amount of TEMP, AIREP, and SYNOP data in the Northern Hemisphere. This result is consistent with Kalnay et al. (1996) and Kistler et al. (2001) where it is found that the reanalysis fields of the Northern Hemisphere are more accurate than those of the Southern Hemisphere and similar conclusions have been also obtained by ECMWF (Cardinali et al. 2004). However, for the specific humidity analysis, the total EVR percentage of the Southern Hemisphere is larger than that of the Northern Hemisphere at the isobaric surfaces where water vapor is concentrated, and the percentage of the equatorial region is the smallest in most cases.

The contributions of different observation types to the total EVR were investigated in the global GRAPES operational 3DVAR system. In the Southern Hemisphere, 60%–90% of the EVR for wind speed, temperature, and specific humidity was attributed to space-based GNSS-RO, AMSU-A, and AMV data in most cases. In the equatorial region, more than 50% of the EVR was attributed to space-based observations except for a few isobaric surfaces. In the Northern Hemisphere, however, more than 50% of the EVR originated from TEMP, AIREP, and SYNOP data contributions; the contribution of AMSU-A data was not dominant. The results show that GNSS-RO and TEMP data were the most important observation types in GRAPES, which agrees with the results of OSEs (Liu and Xue 2014).

FIG. 10. Average error variance reduction of a single observation of each observation-type subset in the target verification domain at 0000 UTC throughout June 2014 for (a) wind speed, (b) temperature, and (c) specific humidity analysis. The target verification domain and all the abbreviations are as in Fig. 1.

and other verification methods (Liu et al. 2016). For temperature analysis, the contribution of TEMP data in the Northern Hemisphere was greater than that in the Southern Hemisphere, whereas the opposite was shown for AMSU-A data, which was consistent with the findings under the French operational ARPEGE 4DVAR assimilation system (Desroziers et al. 2005).

The estimation of impact of single observation indicated that low-impact data points occurred in data-rich regions, whereas high-impact data points were located in data-poor regions. Although the AMSU-A data have a big impact on the temperature analysis, the impact of a single AMSU-A observation is very small. The conclusions are consistent with the findings from ECWMF (Cardinali 2013).

In China and its surrounding area, TEMP made the largest contribution to the analysis, followed by GNSS-RO, whose contribution to specific humidity EVR was equivalent to that of TEMP.

All experiments in this paper were conducted under the assumption that the observation error and background error covariance are reasonably close to their true values. Some sensitivity experiments were performed. When the observation error variances of all observations were magnified 3 times and were not optimal, the contribution of every observation subset was smaller than that obtained with the original observation error variances, but the order of contributions of different observation subsets was unchanged (figures not shown). Brousseau et al. (2014) found that when climatological **B** was used in their 4DVAR assimilation system, the EVR nevertheless provided reasonable information for the analysis. These studies indicated that when **B** and **R** are not optimal, the EVR obtained with the randomization method can still provide useful diagnostic information about the assimilation process, but the closer **B** and **R** get to the truth, the more accurate the estimated EVR.

In this paper, many diagnostic results illustrate the weaknesses of the operational GRAPES system. First, the contributions of all observations in the equatorial region were less than those in the extra-equatorial region. This may be because there are more complicated balance relationships in this region, but these important balances may not be captured by the **B** matrix in 3DVAR, and it is thus difficult for the observations to have large impacts on the analysis field in the equatorial region. Second, the contribution of satellite data in the GRAPES 3DVAR operational system is very small, compared with the diagnostic results in the other operational centers (Kistler et al. 2001; Kelly and Thépaut 2007; Cardinali 2013), especially in humidity analysis. These results are consistent with the diagnostic results obtained based on other verification methods in GRAPES 3DVAR (Liu et al. 2016). Therefore, future work is needed to improve the representation of background error and observation error covariance matrices, especially in the equatorial region, in order to extract maximum information from the observations. In addition, it is essential to increase the number and type of satellite data under cloud conditions assimilated in the global GRAPES operational system. It must be emphasized that the order in which the observations affect the analysis is not completely consistent with the order in which the observations affect the forecast, which has been pointed out in previous studies (Langland and Baker 2004; Cardinali 2009, 2013). Diagnosing will be performed in the future to evaluate the impacts of observations on forecasts under the global GRAPES system.

In summary, although there are fewer observations assimilated in the global GRAPES operational system compared with other NWP operational centers, the above diagnostic results are very positive for the development of the GRAPES assimilation system.

Acknowledgments. We are very grateful to the editor and three reviewers for their valuable suggestions. We also appreciate the suggestions and help provided by experts in NWPC of CMA, especially by Wang Jincheng, Zhang Lin, and Zhuang Zhaorong when we investigated the details of satellite data assimilation and background error processing in the GRAPES 3DVAR system. The excellent research environment and high-efficiency computing platform in the NWPC ensured the accomplishment of this study. This research is jointly sponsored by the special Research Program for Public Welfare (Meteorology) of China (GYHY201506003), the National Natural Science Foundation of China (91437113), the Research Innovation Program for College Graduates of Jiangsu Province (KYLX15_0861),

the Development Program of the Ministry of Science and Technology of China (2015BAC03B01), and the National Natural Science Foundation of China (41705085).

REFERENCES

- Arakawa, A., and V. R. Lamb, 1977: Computational design of the basic dynamical processes of the UCLA general circulation model. *Methods in Computational Physics*, J. Chang, Ed., Academic Press, 174–267.
- Atlas, R., 1997: Atmospheric observations and experiments to assess their usefulness in data assimilation. *J. Meteor. Soc. Japan*, **75**, 111–130, https://doi.org/10.2151/jmsj1965.75.1B_111.
- Auligné, T., H. Huang, H. Lin, H. Wang, Q. Xiao, X. Zhang, and X. Zhang, 2011: Forecast sensitivity to observations (FSO): WRF/WRFPLUS/WRFDA v3.3 user's guide. NCAR, 8 pp., http://www2.mmm.ucar.edu/wrf/users/wrfda/Tutorials/2012_July/docs/README_FSO_v3.3.pdf.
- Baker, N. L., and R. Daley, 2000: Observation and background adjoint sensitivity in the adaptive observation-targeting problem. *Quart. J. Roy. Meteor. Soc.*, **126**, 1431–1454, <https://doi.org/10.1002/qj.4971265611>.
- Brousseau, P., G. Desroziers, F. Bouttier, and B. Chapnik, 2014: A posteriori diagnostics of the impact of observations on the AROME-France convective-scale data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **140**, 982–994, <https://doi.org/10.1002/qj.2179>.
- Cardinali, C., 2009: Monitoring the observation impact on the short-range forecast. *Quart. J. Roy. Meteor. Soc.*, **135**, 239–250, <https://doi.org/10.1002/qj.366>.
- , 2013: Data assimilation: Observation influence diagnostic of a data assimilation system. ECMWF, 22 pp., <https://www.ecmwf.int/sites/default/files/elibrary/2013/16938-observation-influence-diagnostic-data-assimilation-system.pdf>.
- , and F. Prates, 2011: Performance measurement with advanced diagnostic tools of all-sky microwave imager radiances in 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **137**, 2038–2046, <https://doi.org/10.1002/qj.865>.
- , and S. Healy, 2014: Impact of GPS radio occultation measurements in the ECMWF system using adjoint-based diagnostics. *Quart. J. Roy. Meteor. Soc.*, **140**, 2315–2320, <https://doi.org/10.1002/qj.2300>.
- , S. Pezzuli, and E. Andersson, 2004: Influence matrix diagnostic of a data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **130**, 2767–2786, <https://doi.org/10.1256/qj.03.205>.
- Chapnik, B., G. Desroziers, F. Rabier, and O. Talagrand, 2004: Properties and first application of an error statistics tuning method in a variational assimilation. *Quart. J. Roy. Meteor. Soc.*, **130**, 2253–2275, <https://doi.org/10.1256/qj.03.26>.
- , —, —, and —, 2006: Diagnosis and tuning of observational error in a quasi-operational data assimilation setting. *Quart. J. Roy. Meteor. Soc.*, **132**, 543–565, <https://doi.org/10.1256/qj.04.102>.
- Charney, J. G., and N. A. Phillips, 1953: Numerical integration of the quasi-geostrophic equations for barotropic and simple baroclinic flows. *J. Meteor.*, **10**, 71–99, [https://doi.org/10.1175/1520-0469\(1953\)010<0071:NIOTQG>2.0.CO;2](https://doi.org/10.1175/1520-0469(1953)010<0071:NIOTQG>2.0.CO;2).
- Chen, D.-H., and Coauthors, 2008: New generation of multi-scale NWP system (GRAPES): General scientific design (in Chinese). *Chin. Sci. Bull.*, **53**, 3433–3445.
- , J.-S. Xue, X.-S. Shen, J. Sun, Q.-L. Wan, Z.-Y. Jin, and X.-L. Li, 2012: Application and prospect of new generation of

- multi-scale NWP system (GRAPES) (in Chinese). *Chin. Eng. Sci.*, **2012**, 46–54.
- Chen, J., and G. Li, 2011: Application of cloud detection to assimilation of AIRS radiance data (in Chinese). *Meteor. Monogr.*, **37** (5), 555–563.
- Cohn, S. E., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan*, **75**, 257–288, <https://doi.org/10.2151/jmsj1965.75.1B.257>.
- Cucurull, L., and R. A. Anthes, 2014: Impact of infrared, microwave, and radio occultation satellite observations on operational numerical weather prediction. *Mon. Wea. Rev.*, **142**, 4164–4186, <https://doi.org/10.1175/MWR-D-14-00101.1>.
- Cullen, M. J. P., T. Davies, M. H. Mawson, J. A. James, S. C. Coulter, and A. Malcolm, 1997: An overview of numerical methods for the next generation U.K. NWP and climate model. *Atmos.–Ocean*, **35**, 425–444, <https://doi.org/10.1080/07055900.1997.9687359>.
- Desroziers, G., and S. Ivanov, 2001: Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Quart. J. Roy. Meteor. Soc.*, **127**, 1433–1452, <https://doi.org/10.1002/qj.49712757417>.
- , P. Brousseau, and B. Chapnik, 2005: Use of randomization to diagnose the impact of observations on analyses and forecasts. *Quart. J. Roy. Meteor. Soc.*, **131**, 2821–2837, <https://doi.org/10.1256/qj.04.151>.
- , L. Berre, V. Chabot, and B. Chapnik, 2009: A posteriori diagnostics in an ensemble of perturbed analyses. *Mon. Wea. Rev.*, **137**, 3420–3436, <https://doi.org/10.1175/2009MWR2778.1>.
- Ding, Y., S.-Y. Zhuang, and J.-F. Gu, 2010: Assimilation of observed surface wind with GRAPES. *J. Trop. Meteor.*, **16**, 96–100.
- English, S., R. Saunders, B. Candy, M. Forsythe, and A. Collard, 2004: Met Office satellite data OSEs. *Third WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction*, WMO/TD-1228, Alpbach, Austria, WMO, 146–156.
- Errico, R. M., N. C. Prive, and W. Gu, 2015: Use of an OSSE to evaluate background-error covariances estimated by the NMC method. *Quart. J. Roy. Meteor. Soc.*, **141**, 611–618, <https://doi.org/10.1002/qj.2384>.
- Fisher, M., 2003: Estimation of entropy reduction and degrees of freedom for signal for large variational analysis systems. ECMWF Tech. Rep. 397, 20 pp., <https://www.ecmwf.int/sites/default/files/elibrary/2003/9402-estimation-entropy-reduction-and-degrees-freedom-signal-large-variational-analysis-systems.pdf>.
- Gao, J., and D. Stensrud, 2014: Some observing system simulation experiments with a hybrid 3D-EnVAR system for storm-scale radar data assimilation. *Mon. Wea. Rev.*, **142**, 3326–3346, <https://doi.org/10.1175/MWR-D-14-00025.1>.
- Girard, D., 1989: A fast Monte Carlo cross-validation procedure for large least squares problems with noisy data. *Numer. Math.*, **56**, 1–23, <https://doi.org/10.1007/BF01395775>.
- Gong, J.-D., L. Wei, S. W. Tao, G. Zhao, and F. Wan, 2006: Accurate estimation and application of 3-d error covariance structures in global data assimilation. Part I: Accurate estimation of error covariance in observation space (in Chinese). *Acta Meteor. Sin.*, **64** (3), 669–683.
- , R.-C. Wang, and M. Hao, 2016: The impact of a balance constraint between temperature and humidity on the global humidity analysis in GRAPES (in Chinese). *Acta Meteor. Sin.*, **74** (3), 380–396.
- Han, W., 2014: Constrained variational bias correction for satellite radiance assimilation. *19th Int. TOVS Study Conf. (ITSC-19)*, Jeju Island, South Korea, WMO, 10.01.
- Hao, M., H. Zhang, S.-W. Tao, and J.-D. Gong, 2013: Application of variational quality control to regional GRAPES-3DVAR (in Chinese). *Plat. Meteor.*, **32** (1), 122–132.
- Huang, X., and Coauthors, 2009: Four-dimensional variational data assimilation for WRF: Formulation and preliminary results. *Mon. Wea. Rev.*, **137**, 299–314, <https://doi.org/10.1175/2008MWR2577.1>.
- Joo, S., J. Eyre, and R. Marriott, 2013: The impact of MetOp and other satellite data within the Met Office global NWP system using an adjoint-based sensitivity method. *Mon. Wea. Rev.*, **141**, 3331–3342, <https://doi.org/10.1175/MWR-D-12-00232.1>.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Kelly, G., and J. N. Thépaut, 2007: Evaluation of the impact of the space component of the Global Observing System through Observing System Experiments. ECMWF Newsletter, No. 113, ECMWF, Reading, United Kingdom, 16–28, <https://www.ecmwf.int/sites/default/files/elibrary/2007/17527-evaluation-impact-space-component-global-observing-system-through-observing-system-experiments.pdf>.
- , —, R. Buizza, and C. Cardinali, 2007: The value of observations. I: Data denial experiments for the Atlantic and the Pacific. *Quart. J. Roy. Meteor. Soc.*, **133**, 1803–1815, <https://doi.org/10.1002/qj.150>.
- Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247–267, [https://doi.org/10.1175/1520-0477\(2001\)082<0247:TNNYRM>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2).
- Langland, R. H., and N. L. Baker, 2004: Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus*, **56A**, 189–201, <https://doi.org/10.1111/j.1600-0870.2004.00056.x>.
- Li, H., J.-J. Liu, and E. Kalnay, 2010: Correction of ‘Estimating observation impact without adjoint model in an ensemble Kalman filter.’ *Quart. J. Roy. Meteor. Soc.*, **136**, 1652–1654, <https://doi.org/10.1002/qj.658>.
- , W.-Y. Ding, J.-S. Xue, Z.-T. Chen, and Y.-D. Gao, 2012: A study on the application of FY-2E cloud drift wind height reassignment in numerical forecast of Typhoon Chanthu (2003) track. *J. Trop. Meteor.*, **28**, 348–356.
- Liu, J., E. Kalnay, T. Miyoshi, and C. Cardinali, 2009: Analysis sensitivity calculation in an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **135**, 1842–1851, <https://doi.org/10.1002/qj.511>.
- Liu, Y., and J. Xue, 2014: Assimilation of global navigation satellite radio occultation observations in GRAPES: Operational implementation. *J. Meteor. Res.*, **28**, 1061–1074, <https://doi.org/10.1007/s13351-014-4028-0>.
- , J.-S. Xue, L. Zhang, and H.-J. Lu, 2016: Verification and diagnostics for data assimilation system of global GRAPES (in Chinese). *J. Appl. Meteor. Sci.*, **27**, 1–15.
- , L. Zhang, and Z.-Y. Jin, 2017: The optimization of GRAPES global tangent linear model and adjoint model (in Chinese). *J. Appl. Meteor. Sci.*, **28** (1), 62–71.
- Parrish, D., 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](https://doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2).
- Rabier, F., and P. Courtier, 1992: Four-dimensional assimilation in the presence of baroclinic instability. *Quart. J. Roy. Meteor. Soc.*, **118**, 649–672, <https://doi.org/10.1002/qj.49711850604>.

- , H. Jarvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143–1170, <https://doi.org/10.1002/qj.49712656415>.
- Radnoti, G., P. Bauer, A. P. McNally and A. Horányi, 2012: ECMWF study to quantify the interaction between terrestrial and space-based observing systems on Numerical Weather Prediction skill. Tech. Memo. 679, ECMWF, 98 pp., <https://www.ecmwf.int/sites/default/files/elibrary/2012/11814-ecmwf-study-quantify-interaction-between-terrestrial-and-space-based-observing-systems.pdf>.
- Sommer, M., and M. Weissmann, 2014: Observation impact in a convective-scale localized ensemble transform Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **140**, 2672–2679, <https://doi.org/10.1002/qj.2343>.
- Talagrand, O., 1997: Assimilation of observations, an introduction. *J. Meteor. Soc. Japan*, **75**, 191–209, https://doi.org/10.2151/jmsj1965.75.1B_191.
- Tao, S. W., X. H. Chen, and J. D. Gong, 2006: Error analyses for temperature of L band radiosonde (in Chinese). *Meteor. Monogr.*, **32** (10), 47–51.
- , M. Hao, J. S. Xue, and Y. W. Dou, 2008: Analyses of observation data error in numerical forecasting assimilation system (in Chinese). *Meteor. Monogr.*, **34** (7), 34–40.
- Wang, G., Q.-F. Lu, J.-W. Zhang, and H.-Y. Wen, 2014a: Study on method and experiment of hyper-spectral atmospheric sounder channel selection (in Chinese). *Remote Sens. Technol. Appl.*, **29** (5), 795–802.
- , X.-B. Liu, Y. Yang, and Q.-Q. Kong, 2014b: Application of bi-weighting quality control method to FY-3B/IRAS data (in Chinese). *J. Chengdu Univ. Info. Technol.*, **29** (6), 609–615.
- , H. Zhang, and Y. Yang, 2017: Research progress of quality control for AIRS data (in Chinese). *Diqiu Kexue Jinzhan*, **32** (2), 139–150.
- Wang, J.-C., Z.-R. Zhuang, W. Han, and H.-J. Lu, 2014: An improvement of background error covariance in the global GRAPES variational data assimilation and its impact on the analysis and prediction: Statistics of the three-dimensional structure of background error covariance (in Chinese). *Acta Meteor. Sin.*, **72** (1), 62–78.
- , J.-D. Gong, and B. Zhao, 2015: A new method for estimating observation error of the COSMIC refractivity data and its impacts on GRAPES-GFS model weather forecasts (in Chinese). *Acta Meteor. Sin.*, **73** (1), 142–158.
- , —, and R.-C. Wang, 2016: Estimation of background error for brightness temperature in GRAPES 3DVar and its application in radiance data background quality control (in Chinese). *Acta Meteor. Sin.*, **74** (3), 397–409.
- , and Coauthors, 2017: Improvements and performances of the operational GRAPES_GFS 3DVar System (in Chinese). *J. Appl. Meteor. Sci.*, **28** (1), 11–24.
- Wang, R.-C., J.-D. Gong, and L. Zhang, 2012: Statistical estimation of dynamic balance constraints in the GRAPES variational data assimilation system (in Chinese). *J. Appl. Meteor. Sci.*, **23** (2), 129–138.
- , —, —, and H.-J. Lu, 2015: Tropical balance characteristics between mass and wind fields and their impact on analyses and forecasts of GRAPES system. Part II: Application of linear balance equation–regression hybrid constraint scheme (in Chinese). *Chin. J. Atmos. Sci.*, **39** (6), 1225–1236.
- Xu, D., X.-Y. Huang, H. Wang, A. P. Mizzi, and J. Min, 2015: Impact of assimilating radiances with the WRFDA ETKF/3DVAR hybrid system on prediction of two typhoons in 2012. *J. Meteor. Res.*, **29**, 28–40, <https://doi.org/10.1007/s13351-014-4053-z>.
- Xue, C.-B., J.-D. Gong, C.-F. He, and R.-C. Wang, 2013: Quality control of cloud derived wind vectors from geostationary meteorological satellites with its application to data assimilation system (in Chinese). *J. Appl. Meteor. Sci.*, **24** (3), 356–364.
- Xue, J.-S., and Coauthors, 2003: Development of 3D Variational Assimilation System for ATOVS data in China (in Chinese). *Proc. 13th Int. TOVS Study Conf.*, Ste. Adele, Canada, WMO, 30–36.
- , S.-Y. Zhuang, G.-F. Zhu, H. Zhang, Z.-Q. Liu, Y. Liu, and Z.-R. Zhuang, 2008: Scientific design and preliminary results of three-dimensional variational data assimilation system of GRAPES (in Chinese). *Chin. Sci. Bull.*, **53**, 3446–3457.
- Zagar, N., N. Gustafsson, and E. Kallen, 2004: Variational data assimilation in the tropics: The impact of a background-error constraint. *Quart. J. Roy. Meteor. Soc.*, **130**, 103–125, <https://doi.org/10.1256/qj.03.13>.
- Zhang, F., M. Zhang, and J. Poterjoy, 2013: E3DVar: Coupling an ensemble Kalman filter with three-dimensional variational data assimilation in a limited-area weather prediction model and comparison to E4DVar. *Mon. Wea. Rev.*, **141**, 900–917, <https://doi.org/10.1175/MWR-D-12-00075.1>.
- Zhang, H., J.-S. Xue, S.-Y. Zhuang, G.-F. Zhu, and Z.-S. Zhu, 2004: Idea experiments of GRAPES three-dimensional variational data assimilation system. *Acta Meteor. Sin.*, **62**, 31–41.
- Zhang, L., and Y.-Z. Liu, 2017: The preconditioning of minimization algorithm in GRAPES global four-dimensional variational data assimilation system (in Chinese). *J. Appl. Meteor. Sci.*, **28** (2), 168–176.
- Zhang, L.-H., T.-L. Shen, and H.-L. Wang, 2007: Variational assimilation of AMSU data and its application to numerical simulation of heavy rainfall. *Plat. Meteor.*, **26**, 1004–1012.
- Zhu, W.-G., G. Li, H. Zhang, D.-Z. Jin, G. Wang, and Y.-M. Zhong, 2013: Study on application technique of cloud detection and clear channels hyperspectral atmospheric infrared detector AIRS data (in Chinese). *Meteor. Monogr.*, **39** (5), 633–644.
- Zhuang, Z.-R., J.-S. Xue, S. Y. Zhuang, and G. F. Zhu, 2006: A study of the statistical analysis of the geopotential height background errors in the data assimilation (in Chinese). *Chin. J. Atmos. Sci.*, **30**, 533–544.
- , —, W. Han, and Y. Liu, 2014: The application of radiosonde observation blacklisting check to variable data assimilation system (in Chinese). *J. Appl. Meteor. Sci.*, **25**, 274–283.