

Assimilation of TRMM Multisatellite Precipitation Analysis with a Low-Resolution NCEP Global Forecast System

GUO-YUAN LIEN

*Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, Maryland,
and RIKEN Advanced Institute for Computational Science, Kobe, Japan*

TAKEMASA MIYOSHI

*Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, Maryland,
and RIKEN Advanced Institute for Computational Science, Kobe, and Japan Agency for Marine-Earth Science
and Technology, Yokohama, Japan*

EUGENIA KALNAY

Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, Maryland

(Manuscript received 13 April 2015, in final form 27 August 2015)

ABSTRACT


Current methods of assimilation of precipitation into numerical weather prediction models are able to make the model precipitation become similar to the observed precipitation during the assimilation, but the model forecasts tend to return to their original solution after a few hours. To facilitate the precipitation assimilation, a logarithm transformation has been used in several past studies. Lien et al. proposed instead to assimilate precipitation using the local ensemble transform Kalman filter (LETKF) with a Gaussian transformation technique and succeeded in improving the model forecasts in perfect-model observing system simulation experiments (OSSEs).

In this study, the method of Lien et al. is tested within a more realistic configuration: the TRMM Multisatellite Precipitation Analysis (TMPA) data are assimilated into a low-resolution version of the NCEP Global Forecast System (GFS). With guidance from a statistical study comparing the GFS model background precipitation and the TMPA data, some modifications of the assimilation methods proposed in Lien et al. are made, including 1) applying separate Gaussian transformations to model and to observational precipitation based on their own cumulative distribution functions; 2) adopting a quality control criterion based on the correlation between the long-term model and observed precipitation data at the observation location; and 3) proposing a new method to define the transformation of zero precipitation that takes into account the zero precipitation probability in the background ensemble rather than the climatology. With these modifications, the assimilation of the TMPA precipitation data improves both the analysis and 5-day model forecasts when compared with a control experiment assimilating only rawinsonde data.

1. Introduction

Because of the large impact that effective assimilation of precipitation could have in forecasting severe weather, many efforts to assimilate precipitation

observations have been made (e.g., Tsuyuki 1996, 1997; Falkovich et al. 2000; Davolio and Buzzi 2004; Koizumi et al. 2005; Mesinger et al. 2006; Miyoshi and Aranami 2006; Lopez 2011, 2013; Zupanski et al. 2011; Zhang et al. 2013). However, there are many difficulties in assimilating precipitation data, including the nonlinearity of the precipitation process, the non-Gaussian error distribution associated with precipitation, and the large and unknown model and observation errors. These issues have been discussed in several studies (e.g., Errico et al. 2007; Bauer et al. 2011; Lien et al. 2013, hereafter LKM13), leading to the widely shared conclusion that the models cannot “remember” the assimilation changes

 Denotes Open Access content.

Corresponding author address: Guo-Yuan Lien, Data Assimilation Research Team, RIKEN Advanced Institute for Computational Science, 7-1-26, Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan.
E-mail: guo-yuan.lien@riken.jp

DOI: 10.1175/MWR-D-15-0149.1

after a few forecast hours, so medium-range model forecasts are not improved (e.g., Falkovich et al. 2000; Davolio and Buzzi 2004; Tsuyuki and Miyoshi 2007).

LKM13 proposed use of the local ensemble transform Kalman filter (LETKF; Hunt et al. 2007), an efficient type of ensemble Kalman filter (EnKF), and a Gaussian transformation (anamorphosis) method to assimilate precipitation. Unlike other methods, such as nudging to change the moisture and temperature to force the model to precipitate as observed (e.g., Mesinger et al. 2006), the LETKF is able to update the model dynamical variables by the assimilation of precipitation through the flow-dependent error covariance estimated by the ensemble. In addition, the Gaussian anamorphosis transforms original precipitation into a more Gaussian variable and mitigates the non-Gaussianity issue. The use of the Gaussian anamorphosis in ensemble data assimilation was also studied by Amezcua and van Leeuwen (2014), who concluded that the variable transformation cannot exactly reconstruct the Bayesian posterior but can lead to a better approximation of the solution. LKM13 tested the use of the LETKF and of the Gaussian transformation with a simplified general circulation model, known as the Simplified Parameterizations, Primitive Equation Dynamics (SPEEDY; Molteni 2003) model, in a perfect-model simulation. They obtained promising results: the assimilation of precipitation significantly improved not only the analyses but also the 0–5-day model forecasts.

Given the success of these idealized experiments, we perform similar precipitation assimilation experiments with a more realistic configuration. Considering the data availability and the limited computational resources, we choose to assimilate the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) version 7 (Huffman et al. 2012, 2007, 2010) into a low-resolution version of the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS). As indicated above, the LKM13 experiments were conducted within an identical-twin observing system simulation experiment (OSSE) framework so that their most serious approximation is the assumption of a perfect model. In reality, we know that the precipitation parameterizations in the models are far from perfect, with large model errors deteriorating the precipitation forecasts. As to the observation errors, the errors in satellite retrievals may not only be large, but their error characteristics are also mostly unknown. As a result, it is much more difficult to obtain positive impacts by precipitation assimilation with real models and observations. Therefore, before performing the assimilation, we carried out a statistical study investigating characteristics of precipitation in our companion paper,

Lien et al. (2016, hereafter LKMH16). LKMH16 compared a large sample of the short-term GFS model forecasted precipitation and the TMPA dataset for the data assimilation and estimated their statistical differences. We note that, in the present data assimilation study, we neither attempt to improve the model nor the observations. Our goal is to best use this imperfect observation dataset in this imperfect model so that we improve the model forecasts of other variables.

With respect to the observation errors of precipitation, we do not investigate this issue but use tuned constant observation errors in our experiments. There have been several studies to validate the satellite precipitation estimates and to quantify the biases and errors (Bauer et al. 2002; Bowman 2005; Ebert et al. 2007). The error of observed precipitation is typically large compared to other observations used in data assimilation, and it can vary with different grid sizes and different validation time intervals. As an example, for individual TRMM satellite overpasses averaged over a $1^\circ \times 1^\circ$ box, the relative root-mean-square (RMS) difference with respect to a rain gauge centered in the box is 200%–300% (Bowman 2005), much larger than the error of 20% or 50% that LKM13 used. However, by combining information from multiple satellite sensors and averaging raw data in space and/or in time, the errors are reduced. Tian and Peters-Lidard (2010) estimated the lower bound of the uncertainties of satellite-based precipitation measurements in each 0.25° grid over the globe by computing the variance from six different satellite precipitation datasets. They concluded that the uncertainties are relatively small (40%–60%) over the oceans, especially in the tropics and over the lower latitudes of South America. Larger uncertainties (100%–140%) exist over high latitudes, especially during the cold season. Uncertainties are also high over complex terrain, such as the Tibetan Plateau, the Rockies, and the Andes, and near the coastline region. However, the observation error in data assimilation should also include more components, such as the representativeness errors (Errico et al. 2007). Therefore, to simplify the problem with precipitation observation errors in this study, we follow a strategy similar to Lopez (2011, 2013) and LKM13: a constant value is used for the observation error of all precipitation observations after the variable transformation (either the logarithm transformation or the Gaussian transformation). The underlying hypothesis is that, after the precipitation transformation, the observation errors are more uniform (Mahfouf et al. 2007).

Accordingly, following LKM13 and LKMH16, here we conduct the assimilation of global large-scale TMPA data in the GFS model run at T62 resolution. The main objectives of this study are to explore the usefulness of

the LKM13 method in an intermediate-complexity system (i.e., state-of-the-art numerical weather prediction models but at low resolution) and to find the limitations with real models and data. We focus on the comparison between different transformation methods, including a new method proposed here that modifies the Gaussian transformation for zero precipitation values. In addition, in section 5, we verify the validity of the transformation methods by examining the Gaussianity of the actual background error distribution represented by the ensemble. Note that this examination, absent in LKM13, is very important in order to show that the variable transformation can improve the assimilation results by making the errors more Gaussian.

The paper is organized as follows. The precipitation transformation method used in data assimilation is reviewed in section 2, including a detailed discussion about the issue of the zero precipitation transformation. Section 3 provides the experimental settings. Section 4 shows the results of assimilating real precipitation with different transformations as well as no transformation. Section 5 examines the Gaussianity of the background errors of precipitation in our data assimilation experiments with and without the use of the transformation methods. Section 6 provides further discussion and conclusions.

2. Transformation of precipitation

There are several ways to deal with a non-Gaussian variable in data assimilation. Bocquet et al. (2010) provided a comprehensive review of the methods to deal with the non-Gaussianity in various data assimilation schemes. The approaches that do not rely on the Gaussian error assumption, such as the particle filter (van Leeuwen 2009), the maximum entropy method (Eyink and Kim 2006), and the rank histogram filter (Anderson 2010), are generally too expensive or complicated. As a result, these methods have only been applied and tested with simpler systems. On the other hand, a cheaper and more feasible solution would be to do a variable transformation. When non-Gaussian observations are being assimilated, an appropriate transformation of observables can make the error more Gaussian with only a small additional computational cost. Either analytical or empirical formulas can be used for the transformation. For the precipitation assimilation, a logarithm transformation has been widely used (e.g., Lopez 2011, 2013), and the transformation based on Gaussian anamorphosis has been used for precipitation assimilation (LKM13) and for other geophysical assimilation studies (Simon and Bertino 2009; Schöniger et al. 2012).

The main formulation of the logarithm and Gaussian transformation methods has been described in LKM16.

In the following, we will only briefly review these methods. Please refer to LKM13 and LKM16 for more details. However, here we will describe in detail the treatment of the zero precipitation in the Gaussian transformation, which is a critical issue for the precipitation transformation. In addition to the “median of climatological zero precipitation probability” method proposed in LKM13, new methods for the zero precipitation transformation will be introduced and tested in this study.

a. Logarithm transformation

The widely used logarithm transformation is formulated as

$$\tilde{y} = \ln(y + \alpha), \quad (1)$$

where y is the original variable, \tilde{y} is the transformed variable, and the constant α is added to prevent the singularity at zero precipitation ($y = 0$) and can be tuned for different applications. The logarithm transformation enhances the discrimination of small precipitation amounts and damps the contribution of large precipitation amounts, leading to a modification of the dynamical range of the precipitation distribution (Mahfouf et al. 2007). It is also a good choice in the case that the magnitude of precipitation errors varies as a power law of the precipitation amount (Bauer et al. 2002) and that the error distribution is close to a lognormal distribution (Errico et al. 2001). Therefore, we can reasonably assume a constant observation error for precipitation observations in the transformed space.

b. Gaussian transformation

The Gaussian transformation method (LKM13) employs a Gaussian anamorphosis (Wackernagel 2003) from the original variable y to the Gaussian variable \tilde{y} with zero mean and unit variance. It is defined by

$$F^G(\tilde{y}) = F(y), \quad \text{or} \quad (2)$$

$$\tilde{y} = F^{G-1}[F(y)], \quad (3)$$

where F is the empirical cumulative distribution function (CDF) of y that can be determined based on the climatology of the variable, F^G is the CDF of the target Gaussian variable \tilde{y} , and F^{G-1} is the inverse function of F^G . Since we require \tilde{y} to be a Gaussian variable with zero mean and unit variance, its inverse CDF F^{G-1} is

$$F^{G-1}(P) = \sqrt{2} \operatorname{erf}^{-1}(2P - 1), \quad (4)$$

where P is the cumulative probability and erf^{-1} is the inverse error function. An illustration of the transformation procedure with the real data can be found in Fig. 2 of LKM16.

As proposed in LKM16, since the model and observations are imperfect, the Gaussian transformations to the

GFS model and to the TMPA data are separately defined based on their own climatologies. In this way, because the model and observational precipitation are first converted to the same 0–1 scale (cumulative distribution) before the same inverse CDF is applied, it effectively reduces the amplitude-dependent bias between these two variables. Thus, following the steps described in section 4 of LKMH16, 10-yr (2001–10) samples of model precipitation and observational precipitation are constructed, and the CDFs of model and TMPA precipitations are calculated for each T62 GFS grid point and each 10-day period of the year (3 periods per month; 36 periods in total). These CDFs are used for defining the transformations.

Since the transformation is determined based on the climatological samples, it transforms the climatological distribution of the variable into a Gaussian distribution, but this does not necessarily imply that the background error distributions become Gaussian as required in the EnKF data assimilation (e.g., Ott et al. 2004). The background error distributions change with time of the year and location, and in the EnKF they are estimated by the ensemble of model forecasts from the analyses of the previous cycle, typically 6 h. However, because of the small ensemble size used in the EnKF (typically 20–100 members), if no additional assumptions are made, it would be difficult to define the transformation to the entire variable space for every background error distribution at each observational time and location purely based on the background ensemble (however, it is possible to define just the zero precipitation transformation based on the background ensemble as we discuss in section 2c). Therefore, using the climatological sample and defining the climatological Gaussian transformation is a practicable choice. Nevertheless, it is reasonable to assume that the error distributions from a variable with more Gaussian climatological distribution are also more Gaussian, which should benefit the EnKF (LKMH13; LKMH16). Although it may be difficult to theoretically validate this assumption with real problems, in section 5 we test this assumption experimentally by checking the Gaussianity of the transformed errors using the actual samples of background ensembles generated by a realistic model.

c. Zero precipitation in the Gaussian transformation

The Gaussian transformation described above can establish a one-to-one relation between the original variable and the transformed variable if their CDFs are continuous. However, the precipitation CDF is discontinuous at zero (because of the delta-function characteristics of the zero precipitation probability distribution), so the zero precipitation transformation needs to be specially treated. In practice, considering Eqs. (2) or (3), the cumulative probability level of zero precipitation $[F(0)]$ is not well

defined. The essence of the problem is thus to determine a specific value for $F(0)$ in the transformation procedure. LKM13 already discussed this zero precipitation problem and proposed to transform all zero values to the median of the zero precipitation probability in the climatology. It was a very simple method and worked well in the idealized precipitation assimilation experiments. Here, using real data, we propose two other possible methods to define the zero precipitation transformation: the median of the zero precipitation probability determined by the model background ensemble and the uniformly random transformation.

As in LKMH16, we regard all precipitation values smaller than $0.06 \text{ mm (6 h)}^{-1}$ as “zero precipitation”: that is, we define a precipitation value, $y_{\text{trace}} = 0.06 \text{ mm (6 h)}^{-1}$, which is the minimum measurable trace of precipitation. Any precipitation value smaller than y_{trace} is thus converted to zero:

$$y \leftarrow \begin{cases} 0, & \text{if } y < y_{\text{trace}} \\ y, & \text{if } y \geq y_{\text{trace}} \end{cases}. \quad (5)$$

A proper choice of y_{trace} can improve the data assimilation results, since it removes the impact of irrelevantly small values. However, a value of y_{trace} that is too large would diminish the useful information contained in the small precipitation amounts. This choice of y_{trace} was found to lead to better results in our system.

1) METHOD 1: CLIMATOLOGICAL MEDIAN OF ZERO

This is the simple method proposed and used in LKM13. The zero precipitation values are transformed to the median of the climatological zero precipitation probability (CZ method). This means we assign $F(0)$ to be the middle value of the zero precipitation cumulative probability:

$$F(0) = \frac{1}{2} P_c \quad (6)$$

$$\text{so that } \tilde{y}_{\text{zero(CZ)}} = F^{G-1}[F(0)] = F^{G-1}\left(\frac{P_c}{2}\right), \quad (7)$$

where $P_c = F(y_{\text{trace}})$ is the probability of zero precipitation in the climatology.

2) METHOD 2: BACKGROUND MEDIAN OF ZERO

The CZ method is an easy way to transform the zero precipitation, but it may not be appropriate if we consider that it is the background error distribution, not the climatological distribution, that should be Gaussian. Recall the discussion in the end of section 2b that the small size of the background ensemble (typically

20–100) may preclude defining the Gaussian transformation for every observational time and location, so we have to define the Gaussian transformation climatologically. However, in this subsection, we show that it is possible to define the zero precipitation transformation (instead of the entire transformation) based on the background error distribution. By introducing some assumptions, we transform the zero precipitation values to the median of the zero precipitation probability represented by the background ensemble (BZ method). Note that in this way the transformation for all nonzero precipitation values is still defined from the climatology [Eq. (3)], but only the transformation for zero is defined based on the background error distribution that varies in every observational time and location. This is a compromise we can do with the limited ensemble size in the EnKF.¹

Figure 1 illustrates the relation between the CZ and the BZ methods. First, following our assumption, if there is no zero precipitation, the climatological distribution (black curves) and the background error distribution (red curves) in transformed variables are

Gaussian. When the boundary of zero and nonzero precipitation is present (y_{trace} ; thin vertical lines in Fig. 1), the background error distribution can cross this boundary, resulting in a distribution wherein only the nonzero precipitation part (red shaded areas; right side to the boundary) is fitted into the Gaussian envelope, while the zero precipitation part remains undetermined. We note the similarity of the gray shaded area and the red shaded area: they are both parts of the Gaussian distribution, but the ratios of the zero precipitation areas to the whole Gaussian areas are different. In the CZ method, we determine the value of the transformed zero precipitation based on the gray shaded area; in this BZ method, we determine the value of transformed zero precipitation based on the red shaded area instead. The zero precipitation is therefore transformed to the median of the zero precipitation probability under the red Gaussian curve (red vertical thick line in Fig. 1b). In the LETKF analysis calculation, this probability of zero precipitation in the background P_b can be defined by the background ensemble:

$$P_b = \frac{(\text{number of background members with zero precipitation})}{(\text{ensemble size})}, \quad (8)$$

in contrast to P_c which is defined by the climatology.

Based on the concept expressed above, the new formulation of the zero precipitation transformation is derived. We summarize the solution of the zero precipitation transformation of the BZ method as follows but leave the detailed derivation in the appendix:

$$\tilde{y}_{\text{zero(BZ)}} = \bar{y} + \sigma F^{G-1} \left(\frac{P_b}{2} \right), \quad (9)$$

where

$$\bar{y} = \frac{\alpha \tilde{y}_{\text{trace}} + \beta F^{G-1}(P_b)}{\alpha + (1 - P_b) F^{G-1}(P_b)}, \quad (10)$$

$$\sigma = \frac{(1 - P_b) \tilde{y}_{\text{trace}} - \beta}{\alpha + (1 - P_b) F^{G-1}(P_b)}, \quad \text{and} \quad (11)$$

$$\alpha = -\frac{e^{-[F^{G-1}(P_b)]^2/2}}{\sqrt{2\pi}}, \quad \beta = \frac{1}{N} \sum_{i=n_1+1}^N \tilde{y}_i. \quad (12)$$

In these equations, \bar{y} and σ are the mean and standard deviation of the background error distribution that is assumed to be normal; \tilde{y}_i is the transformed background precipitation value in the i th member, assuming these ensemble values have been sorted in ascending order; N is the ensemble size; and n_1 is the number of ensemble members with zero precipitation [i.e., the summation in Eq. (12) only includes the precipitating members].

This new formulation of the zero precipitation transformation is ill posed when there are too few precipitating members, since in this case we are trying to define the transformation of many zero precipitation members by the rest of the few precipitating members. However, as in LKM13, we adopted a criterion that the precipitation observations are assimilated only when there are enough nonzero precipitation members in the background; thus, this problem is automatically prevented. An important characteristic of the BZ method is that the transformed value of the zero precipitation \tilde{y}_{zero} is always closer to the trace precipitation \tilde{y}_{trace} than that found by the CZ method and is even closer when there are many background members having nonzero precipitation.

3) METHOD 3: RANDOM TRANSFORMATION

We may also propose a probabilistic method to transform the zero precipitation in order to expand the

¹ Amezcua and van Leeuwen (2014) defined the anamorphosis to the entire variable space simply based on the background error distribution, but it was an idealized study in which the error distribution functions are known analytically. For realistic EnKF data assimilation, their method is not applicable.

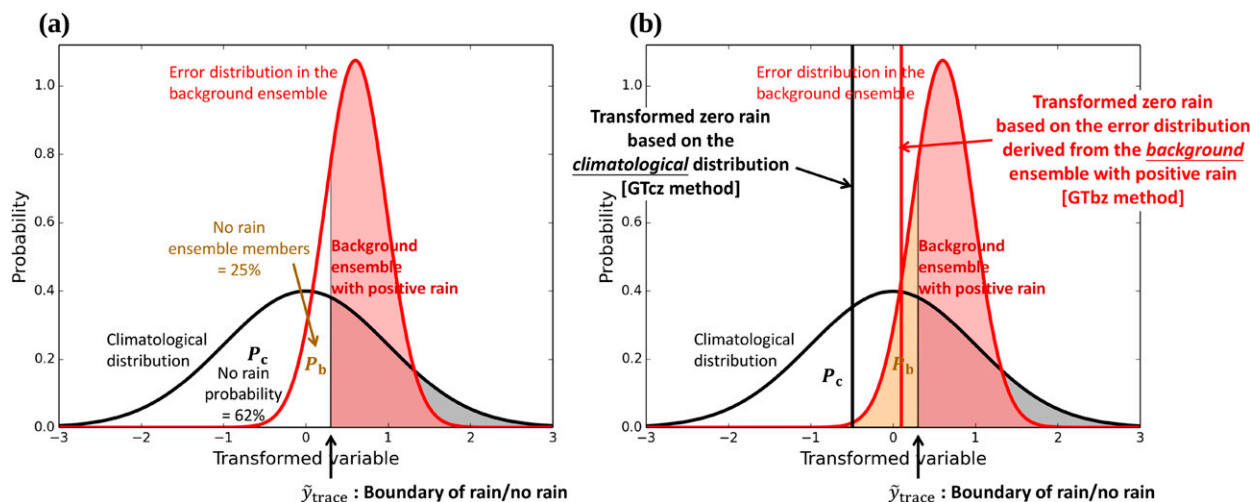


FIG. 1. (a) The relation between the climatological distribution and the error distribution in the background ensemble. The boundary of zero precipitation and nonzero precipitation \tilde{y}_{trace} , the zero precipitation probability in the climatology P_c , and that in the background ensemble P_b are shown. (b) As in (a), but the median of the zero precipitation probability in the CZ (black thick vertical line) and BZ (red thick vertical line) methods are additionally shown.

delta function into a continuous distribution; that is, $F(0)$ is assigned to be a random value from a uniform distribution:

$$\tilde{y}_{\text{zero}(\text{random})} = F^{G-1}(X), \quad X \sim \mathcal{U}(0, P_c), \quad (13)$$

where $\mathcal{U}(0, P_c)$ stands for a uniform distribution from 0 to P_c , the zero precipitation probability in the climatology. In this way, the zero precipitation part of the transformed probability density function (PDF) can be filled, and a perfect (climatological) Gaussian variable can be generated. However, the idea of random observations does not have a sound theoretical basis, and it may lead to additional sampling errors. We tested this idea in our data assimilation experiments, but its experimental impact was worse than either the CZ or BZ approaches. Therefore, the results using this random transformation will not be shown in this article.

3. Experimental design

a. The GFS-LETKF system

We developed the GFS-LETKF system that performs 4D-LETKF analysis (Hunt et al. 2007; Miyoshi and Yamane 2007) with a version of the NCEP GFS model that is low resolution because of computational constraints. It assimilates the conventional observation dataset processed in the NCEP operational systems (i.e., the NCEP PREPBUFR dataset) with a basic thinning function, which keeps at most one observation datum per three-dimensional model grid. We have made the GFS-LETKF code publicly available

online (<http://code.google.com/p/miyoshi/>). Documentation is also provided on the website.

b. Experimental settings

The GFS model is run at a T62 resolution (equivalent to about 200-km horizontal resolution) with 64 vertical hybrid sigma/pressure levels. It uses 32 ensemble members. The initial ensemble at 0000 UTC 1 November 2007 is created by taking a random series of operational GFS analyses in different years but with a similar season and the same time of the day (0000 UTC) in order to prevent the discrepancy caused by the seasonal and diurnal cycles. All conventional (nonradiance) observations taken from the NCEP PREPBUFR dataset are assimilated in the first month in order to spin up the system, evolving the ensemble members so that they represent the errors of the day. After this one-month spinup, the analyses at 0000 UTC 1 December 2007 are used as the initial condition for all experiments. The experimental settings are summarized in Table 1. In the rawinsonde observations (RAOBS) experiment, only the rawinsonde observations are assimilated; in all the other experiments, the global TMPA data are assimilated as well as the rawinsonde observations. In particular, the precipitation is assimilated without using a variable transformation in the no transformation (NT) experiment, and it is assimilated using the logarithm transformation (LOG), the Gaussian transformation with the CZ method (transforming zero precipitation using the climatological zero precipitation probability) (GTcz), and the Gaussian transformation with the BZ method (transform zero precipitation using the background zero precipitation probability) (GTbz) in the other experiments.

TABLE 1. Design of all experiments.

Expt	Observations		Transformation	Observed error of precipitation	Localization scale of precipitation (km)
	Rawinsonde	Precipitation			
RAOBS	×				
NT	×	×		50%, min 0.3	350
LOG	×	×	Log, $\alpha = 0.6$	0.5	350
GTcz	×	×	GTcz	0.5	350
GTbz	×	×	GTbz	0.5	350
GTbz_err0.3	×	×	GTbz	0.3	350
GTbz_err0.7	×	×	GTbz	0.7	350
GTbz_loc500	×	×	GTbz	0.5	500
GTbz_loc200	×	×	GTbz	0.5	200

The effects of the Gaussian transformation applied to the real GFS model precipitation and the TMPA data are examined in LKMH16. We upscale the 0.25° longitude/latitude TMPA data to the grids used by the T62 GFS model so that each precipitation observation corresponds to one model grid point. Besides, different Gaussian transformations are separately defined for the model and observational precipitation based on their own CDFs so that the amplitude-dependent bias between the model and observations can be effectively reduced. The 6-h accumulated precipitation computed from the 3–9-h GFS model forecasts, which is the assimilation window of the 4D-LETKF system, is used as the assimilation variable. The use of the accumulated precipitation amount instead of instantaneous precipitation rate mitigates the effects of precipitation timing errors (LKMH16).

The constant α in the logarithm transformation [Eq. (1)] is set to 0.6 mm (equivalent to 0.1 mm h^{-1} average precipitation rate), which was found to be optimal based on several tests. When no transformation is used, the observation error of precipitation is set to 50% of the observed values but with a minimum error of 0.3 mm (equivalent to 0.05 mm h^{-1} average precipitation rate). In the logarithm transformation experiment, the error is set to a nondimensional constant of 0.5 in the logarithmically transformed variable. In the Gaussian transformation experiments, the error is also set to a constant of 0.5. All of these choices are made to optimize the experimental results in each experiment. The horizontal localization length scale [R localization in Greybush et al. (2011)] of the precipitation observations is 350 km in most experiments, but it is 500 km for all other observations. These localization settings are based on the finding in the LKM13 idealized experiments that the optimal localization scale of precipitation observations is smaller than for regular observations. The vertical localization length scale is 0.4 in natural logarithm of pressure for regular observations and precipitation

observations, while the nominal height of all precipitation observations used for localization is set to 850 hPa. The adaptive inflation (Miyoshi 2011) and the “relaxation to prior” scheme proposed by Zhang et al. (2004) are used together for covariance inflation. Their relaxation constant α for the covariance² is chosen to be 0.5. Note that by combining these two methods, the multiplicative inflation factor is adaptively estimated and the relaxation scheme smooths the cycle-by-cycle evolution of the ensemble spreads at the same time.

The five main experiments (RAOBS, NT, LOG, GTcz, and GTbz) are conducted for a 13-month cycling run until 0000 UTC 1 January 2009, and 5-day free forecasts initialized from each 6-hourly ensemble mean analysis are conducted to quantify the forecast impacts of the assimilation of precipitation. In addition, four other sensitivity experiments are conducted in the same way (but only 3 months long) to examine the sensitivities to the precipitation observation errors (GTbz_err0.3 and GTbz_err0.7) and the localization lengths (GTbz_loc500 and GTbz_loc200). The details of these sensitivity experiments will be described in section 4d. The European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim) dataset is used to verify our results. The one-month period from 1 December 2007 to 1 January 2008 is regarded as an additional spinup period because a certain period is required for the adaptive inflation scheme to adjust to the change of observing systems from the previous conventional observation dataset to the new configurations in each experiment.

² Zhang et al. (2004) defined their covariance relaxation with the ensemble square root filter formulation. In the LETKF, it is equivalent to replacing the weight matrix \mathbf{W}^a with a weighted average of it and the identity matrix: $\mathbf{W}^a \leftarrow (1 - \alpha)\mathbf{W}^a + \alpha\mathbf{I}$ (Hunt et al. 2007). Here, the meaning of the constant α is the same as in Zhang et al. (2004).

c. Quality control criteria for the TMPA assimilation

As suggested in LKM13, with ensemble data assimilation systems, it is better to assimilate precipitation only when the number of the background members with nonzero precipitation is greater than a threshold. With the idealized experiments, they experimentally found that such model background-based criterion is needed to ensure the quality of the precipitation assimilation. Here, with the GFS model and 32 ensemble members, in all precipitation assimilation experiment we require at least 24 (out of 32) precipitating background members to assimilate the precipitation (24mR). Our results testing the Gaussianity of the ensemble forecast errors suggest that this empirically obtained constraint can be justified by the fact that more precipitating members in the background result in more Gaussian background error distributions (see section 5 and Fig. 8).

In addition to the 24mR criterion, a new quality control criterion, which is a geographic mask based on the correlations between long-term model precipitation and observational precipitation, is introduced. The maps of the correlations are shown in Fig. 10 in the companion paper LKMH16, which are calculated using the same 10-yr samples of the model and observational precipitation as we used to compute the CDFs. In the precipitation assimilation experiments, we require that the precipitation be assimilated only where the correlation ≥ 0.35 (shown in greenish colors in Fig. 10, LKMH16). The intention in using this criterion is to avoid using observations over areas where the model and observational precipitation are climatologically inconsistent (LKMH16). Oceanic precipitation data mostly pass the criterion, except for the marine stratocumulus regions in the west of North and South America and west of Australia and Africa. For continental precipitation, the precipitation data are mostly rejected over the entirety of Africa and the Tibetan Plateau. The eastern United States precipitation data are used, while the western United States precipitation data are rejected in winter and in summer. The data over tropical South America are mostly rejected, and the data over southern South America are generally used.

We also test a precipitation assimilation experiment without using this correlation-based geographic mask. It also leads to positive impacts (not shown) similar to the experiments using the correlation mask that we will mostly describe in the next section, but the improvement is slightly smaller than the experiments shown in this study. We think that the impact is small because this correlation-based quality control criterion may overlap with other quality control criteria, such as the 24mR criterion; that is, the precipitation data at those very bad

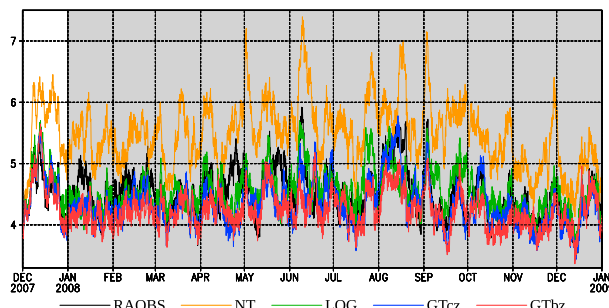


FIG. 2. The 13-month evolution of the global root-mean-square analysis errors of the 500-hPa u wind (m s^{-1}) verified against the ERA-Interim in RAOBS, NT, LOG, GTcz, and GTbz. The shaded period from 1 Jan 2008 to 1 Jan 2009 is the verification period.

areas could be already rejected by other criteria, so the impact is not large.

4. Results

a. Global analysis and forecast errors

Figure 2 shows the evolution of the global analysis RMS errors (RMSEs) of the 500-hPa u wind verified against the ERA-Interim over the 13-month period. Even though the temporal variation is large, it is apparent that the precipitation assimilation experiment without transformation (NT; orange) is clearly worse than RAOBS (black) and that the experiments using Gaussian transformation of precipitation are better, with the BZ method (red) leading to the best analysis. The gray shade indicates the verification period of the entire year 2008 that will be used to compute the average errors and biases in later figures.

Figures 3 and 4 show the summary of the 1-yr results of the main experiments (RAOBS, NT, LOG, GTcz, and GTbz). Figure 3 shows the average 5-day RMS forecast errors for different verification regions (Figs. 3a–l) and biases (Figs. 3m–o) in the 1-yr verification period versus forecast time. Figure 4 focuses on the comparison of the 24-h RMS forecast errors among the 5 experiments verified in the same way. First, for the global results (Figs. 3a–c and 4a–c), the positive impacts by precipitation assimilation using the Gaussian transformation are clear: With either Gaussian transformation (GTcz: blue; GTbz: red), the GFS model analyses ($t = 0$) and forecasts are improved throughout the 5-day range in terms of the 500-hPa u wind (Fig. 3a), the 500-hPa temperature (Fig. 3b), and the 700-hPa specific humidity (Fig. 3c). The more sophisticated zero precipitation transformation method using the information from the background ensemble (GTbz) leads to better results in this global measure, but even the simple GTcz method is usable for effective assimilation of precipitation. The

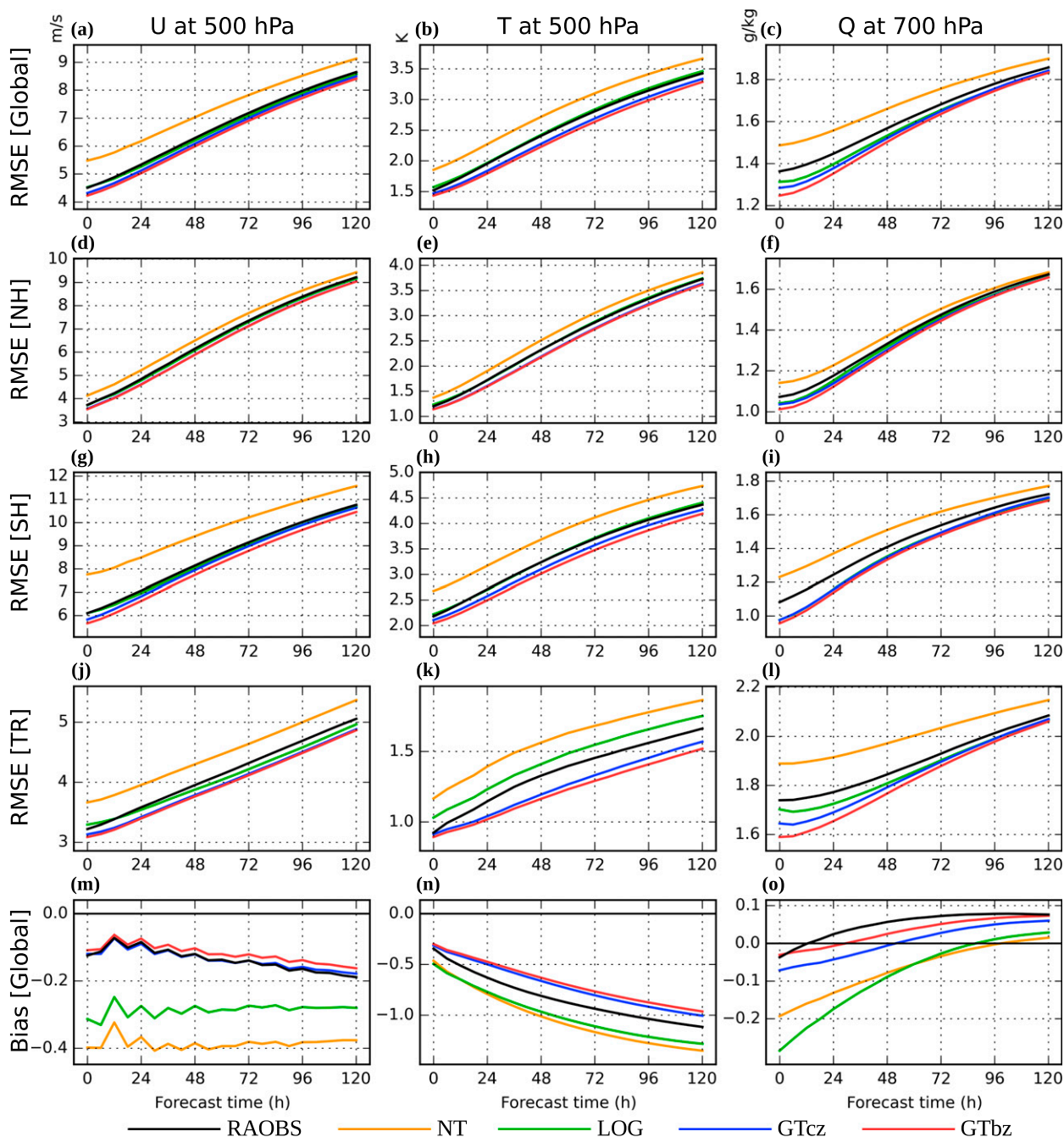


FIG. 3. (a)–(l) The average RMSEs and (m)–(o) biases of (left) 500-hPa u wind (m s^{-1}), (middle) 500-hPa temperature (K), and (right) 700-hPa specific humidity (g kg^{-1}) in 2008 (verified against ERA-Interim) vs forecast time in RAOBS (black lines), NT (orange lines), LOG (green lines), GTcz (blue lines), and GTbz (red lines): (a)–(c) global RMSE, (d)–(f) RMSE in the Northern Hemisphere extratropics (20° – 90° N), (g)–(i) RMSE in the Southern Hemisphere extratropics (20° – 90° S), (j)–(l) RMSE in the tropical region (20° N– 20° S), and (m)–(o) global biases.

improvements in the analysis and forecast errors between both the GTcz and the GTbz experiments to the RAOBS experiments are statistically significant over the entire 0–120-h forecast period at a 99.9% confidence level using hypothesis testing for paired samples. For the

GTbz, the improvements of the average 24-h forecast RMSEs by precipitation assimilation are 5.8%, 8.3%, and 6.5% for 500-hPa u wind, 500-hPa temperature, and 700-hPa moisture, respectively (Figs. 4a–c). The improvement lasts over the whole 5-day forecast, indicating

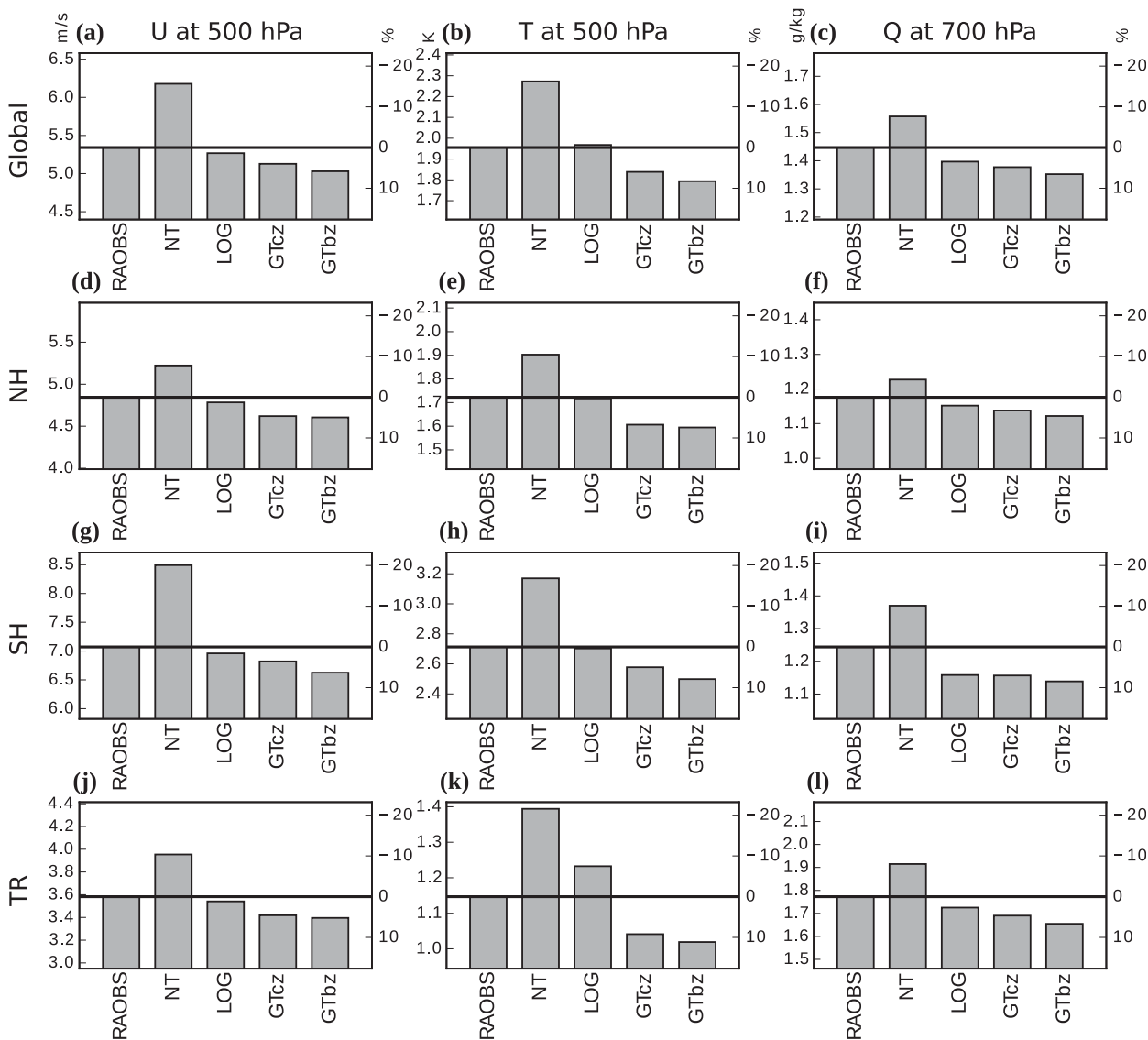


FIG. 4. The 1-yr (2008) average 24-h forecast RMSEs of (left) 500-hPa u wind, (middle) 500-hPa temperature, and (right) 700-hPa specific humidity for (a)–(c) the global, (d)–(f) the Northern Hemisphere extratropical (20° – 90° N), (g)–(i) the Southern Hemisphere extratropical (20° – 90° S), and (j)–(l) the tropical (20° N– 20° S) regions in RAOBS, NT, LOG, GTcz, and GTbz experiments. The left axes show the absolute values for winds (m s^{-1}), temperature (K), and humidity (g kg^{-1}); the right axes show the relative improvement to the RAOBS (%). The thick horizontal lines indicate the RMSE values in RAOBS as a reference.

that the master dynamical variables (i.e., potential vorticity) are improved by the precipitation assimilation using the LETKF. This is consistent with the results obtained by LKM13 with perfect-model OSSEs but very different from other past experiences that the model tends to quickly forget the changes from the assimilation of precipitation (e.g., Mesinger et al. 2006).

In contrast to the Gaussian transformation, with the logarithm transformation (LOG; green in Fig. 3), the impacts are smaller. It shows a similar analysis error as RAOBS in the 500-hPa u wind and then a gradual improvement in

forecast errors over the forecast period. For the 500-hPa temperature, no improvement is seen. For the 700-hPa moisture, the improvement in both analysis and forecasts is clear but is only about half of the improvement obtained using GTbz. The improvements of the average 24-h forecast RMSEs in LOG are 1.4%, -0.6% (negative impact), and 3.5% for 500-hPa u wind, 500-hPa temperature, and 700-hPa moisture, respectively (Figs. 4a–c). If no transformation of precipitation is used (NT; orange in Fig. 3), very large negative impacts by precipitation assimilation are seen with all variables. The negative impacts in LOG and

NT experiments are also seen in the biases (Figs. 3m–o): the precipitation assimilation by these two methods tends to increase the model biases in all three variables.

b. Regional dependence

The regional dependence of the precipitation assimilation is investigated by computing the RMS errors for three separate regions: the Northern Hemisphere extratropics (20°–90°N; NH), the tropics (20°S–20°N; TR), and the Southern Hemisphere extratropics (20°–90°S; SH). The results are shown in Figs. 3d–l and Figs. 4d–l. As shown in LKM13, the analyses and forecasts over the NH region are more accurate than the SH region because of its better rawinsonde observing network, and the NH and SH regions have larger error growth rates than the TR region because of the stronger growth rates of midlatitude baroclinic instabilities. Still, with the Gaussian transformation, the improvement by precipitation assimilation is observed over all three regions. In GTbz, the improvements of the average 24-h forecast RMSEs of 500-hPa u wind by precipitation assimilation in the NH, SH, and TR regions are 4.9%, 6.3%, and 5.2%, respectively (Figs. 4d,g,j). The SH region is improved the most, resulting in about an additional 12-h forecast skill in u wind (Fig. 3g). In GTcz the improvement is smaller but qualitatively similar to GTbz. These results are consistent to what were found with the idealized system (LKM13). In addition, the benefit of the precipitation assimilation in the 700-hPa moisture field is also clear. The precipitation assimilation results in large differences in terms of the 700-hPa specific humidity between RAOBS and GTcz/GTbz at the analysis time ($t = 0$), especially in the SH region (Fig. 3i), although the difference becomes smaller with forecast time. The benefits of the BZ method over the CZ method are larger in the NH and TR regions for the 700-hPa moisture, and in the SH regions for both the 500-hPa u wind and temperature.

In all regions, LOG leads to no or negative impact in the 500-hPa u -wind analysis but, interestingly, to a slight improvement over the 5-day forecast period, especially in the SH region, where the impact turns from negative to positive after 12-h forecasts. In terms of the 500-hPa temperature, the LOG experiment leads to marginal impacts as does RAOBS in the NH and SH region, but it clearly degrades the temperature in the TR region (Figs. 3k and 4k). In terms of the 700-hPa moisture, the LOG experiment, however, brings large positive impacts, showing the particular benefit of precipitation assimilation on the moisture. The impact by LOG is as large as that of GTcz, but only for the 700-hPa moisture in the SH region. By contrast, the NT experiment results

are much worse than those of RAOBS in all regions for all variables.

c. Vertical profiles of the errors

The vertical profiles of the 24-h forecast errors are plotted in Fig. 5. The u -wind error (Figs. 5a,c,e,g) is largest at 200–300 hPa, at the jet levels. The improvement or degradation of the 24-h forecasts by assimilating the TMPA data in GTcz, GTbz, LOG, and NT experiments are consistent at all levels from the surface to higher than 200 hPa. The TR region has different profiles of the precipitation assimilation impacts compared to the other regions. The positive impact for the u wind is smaller at low levels (700–1000 hPa) but much larger at the mid- and upper troposphere (500–200 hPa). As to the moisture profiles of the 24-h forecasts (Figs. 5b,d,f,h), with the logarithm and Gaussian (both GTcz and GTbz) transformations, the moisture forecasts are improved the most at the midlevels (500–700 hPa), which is already shown in Fig. 3. GTbz leads to the best results, showing improvements at all levels in all regions except for the neutral results near the surface in the tropical region. However, when the simpler GTcz method or the LOG method are used, they show some degradation at the lower levels (850–1000 hPa), especially in the tropical region. Again, the NT result is very poor.

d. Sensitivity experiments

Four additional experiments are conducted in order to examine the sensitivity of the assimilation of precipitation to the precipitation observation errors and the localization lengths. To save the computational cost, these sensitivity experiments are conducted only for 3 months ending at 0000 UTC 1 March 2008, and the average period for the forecast verification is 2 months. The experimental settings of these experiments are also listed in Table 1. They are all designed based on the GTbz experiment. In the experiments GTbz_err0.3 and GTbz_err0.7, the observation errors for precipitation are changed to 0.3 and 0.7, respectively, instead of 0.5 in GTbz. In the experiments GTbz_loc500 and GTbz_loc200, the localization length scales for precipitation observations are changed to 500 and 200 km, respectively, from the 350 km used in GTbz. The global results (average 24-h forecast errors over the globe) of these sensitivity experiments are shown in Fig. 6.

1) SENSITIVITY TO OBSERVATION ERRORS

The first set of the sensitivity experiments shows the sensitivity of the 5-day forecast errors to the precipitation observation errors (GTbz_err0.3 and GTbz_err0.7 in Fig. 6). Recall that in this study we use

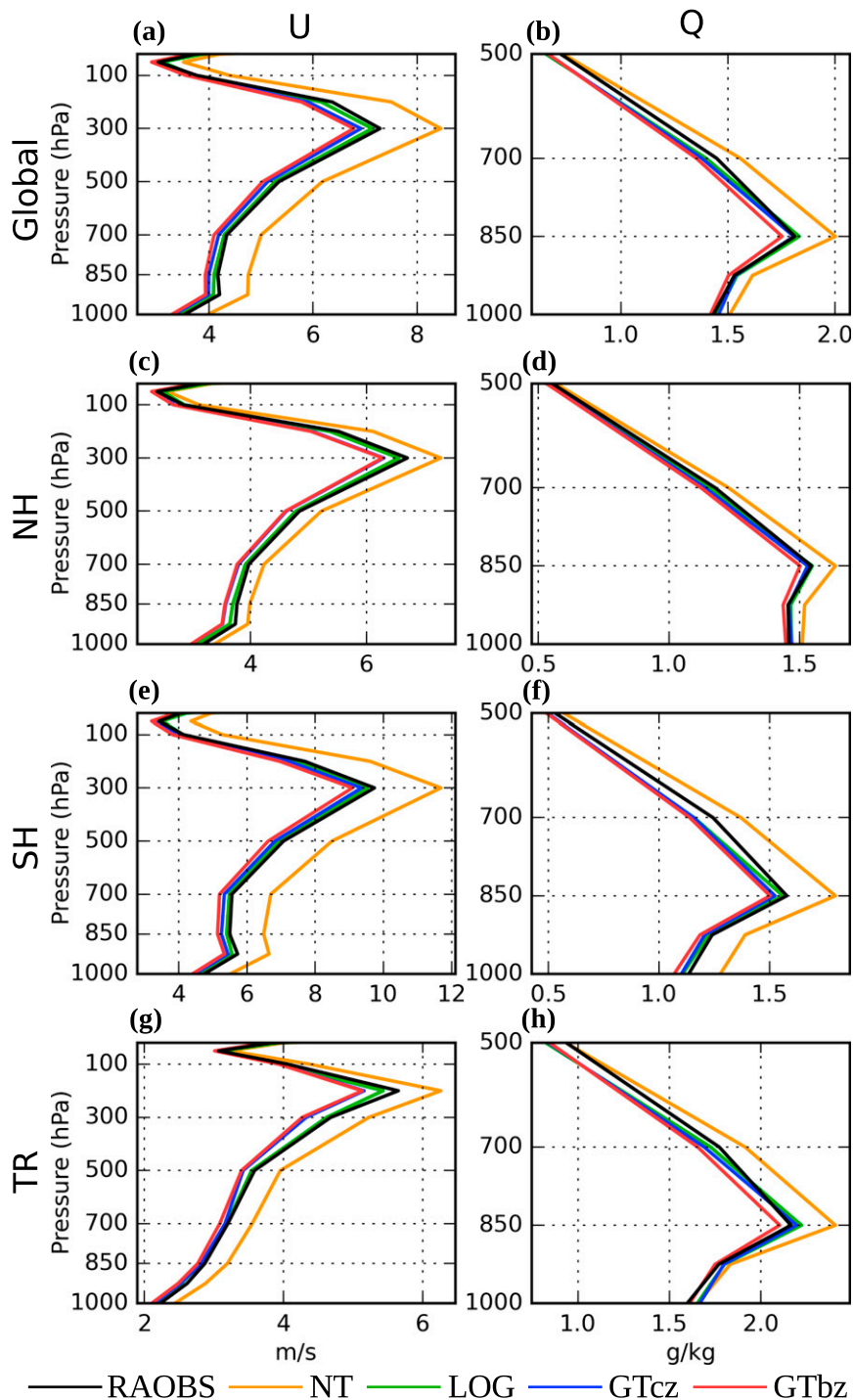


FIG. 5. The vertical profiles of the average root-mean-square 24-h forecast errors of (left) u wind (m s^{-1}) and (right) specific humidity (g kg^{-1}) in 2008 (verified against ERA-Interim) in RAOBS, NT, LOG, GTcz, and GTbz. The verification regions are (a),(b) the globe; (c),(d) the Northern Hemisphere extratropics (20° – 90°N); (e),(f) the Southern Hemisphere extratropics (20° – 90°S); and (g),(h) the tropics (20°N – 20°S). Note that the vertical scales are different for winds and humidity.

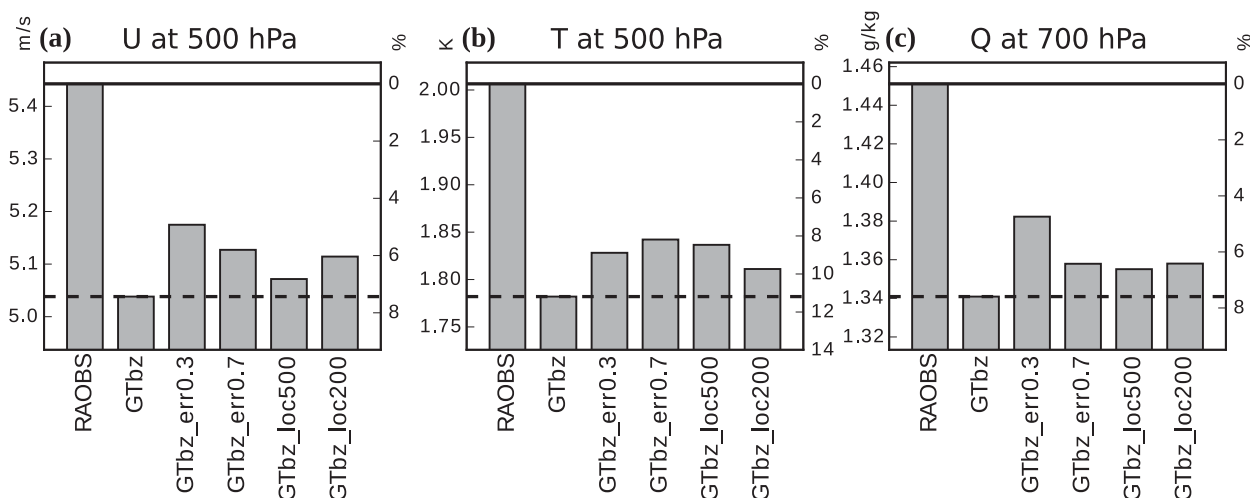


FIG. 6. As in Fig. 4, but for the first 2-month (1 Jan–1 Mar 2008) average 24-h forecast RMSEs for the global region in RAOBS, GTbz, GTbz_err0.3, GTbz_err0.7, GTbz_loc500, and GTbz_loc200 experiments. The thick solid and dashed horizontal lines indicate the RMSE values in RAOBS and GTbz, respectively, as references.

nondimensional constant values for the observation errors of precipitation after the Gaussian transformation is applied. Among the three values, 0.3, 0.5, and 0.7, the observation error of 0.5 as in the control experiment (GTbz) results in the best 24-h forecast errors. When the values of 0.3 (GTbz_err0.3) or 0.7 (GTbz_err0.7) are used, the precipitation assimilation still leads to improvements in the forecasts (compared to RAOBS), but the improvements are smaller than that in GTbz when the value of 0.5 is used.

2) SENSITIVITY TO LOCALIZATION SCALES

The second set of the sensitivity experiments shows the sensitivity of the 5-day forecast errors to the horizontal localization length scales for precipitation observations (GTbz_loc500 and GTbz_loc200 in Fig. 6). It is shown that the control setting (GTbz), the 350-km horizontal localization length scale, also leads to the best result. Note that we use a 500-km horizontal localization scale to assimilation rawinsonde observations, so the optimal localization scale for precipitation assimilation is smaller than that for the rawinsonde observations, which is consistent with the results in LKM13.

5. Examination of the Gaussianity of the background errors

When we define the empirical CDF based on the climatological samples from models or observations, this method transforms the climatological distribution of the original variable into a Gaussian distribution as a whole but not the background error distribution in each observational time and location. As explained in the end of section 2b, a key assumption of our experiments is that the error distributions from a variable with more Gaussian climatological distribution are also more Gaussian. Thus, it is essential to show the validity of this assumption experimentally. Since we can explicitly compute the sample Gaussianity given an ensemble of precipitations, we can verify whether the Gaussian transformation of precipitation makes the ensemble background errors more Gaussian in our real precipitation assimilation experiments. The way to collect the ensemble background precipitation samples is shown in Fig. 7. As in the 4D-LETKF assimilation experiments, we conduct a series of 9-h ensemble GFS forecasts at the T62 resolution initialized from the ensemble analyses of the RAOBS experiment, and then

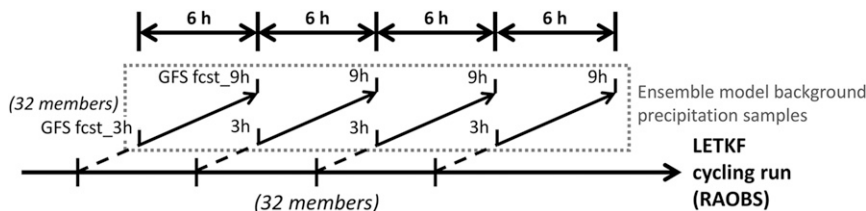


FIG. 7. A schematic of the preparation of precipitation samples for the Gaussianity examination, based on the GFS model ensemble forecasts. Note that in this study we only make the sample every 5 cycles in order to save computational time.

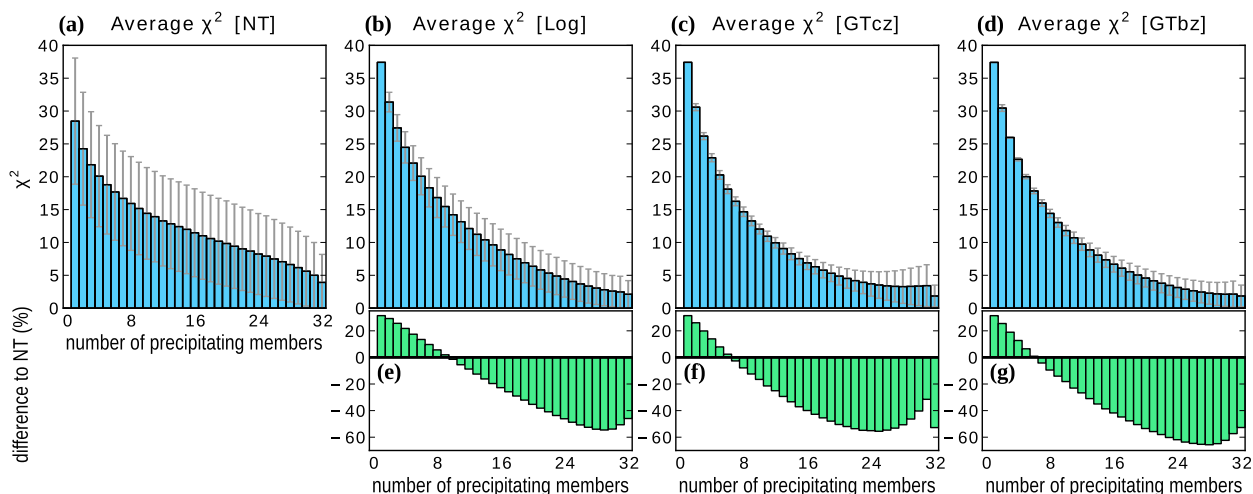


FIG. 8. The average non-Gaussianity χ^2 of background precipitation errors with respect to the number of precipitating background members (a) without transformation; (b) with the logarithm transformation; and with the Gaussian transformation using the (c) CZ and (d) BZ methods. The gray error bars indicate one standard deviation. (e)–(g) The percentage differences of the average χ^2 in each method compared to that without transformation.

the 6-h accumulated precipitation computed from the 3–9-h forecasts (the same as used in the data assimilation experiment) are examined for their Gaussianity. However, instead of computing the statistics every 6-h cycle, here we only compute the Gaussianity statistics for the year 2008 and every 30 h (5 data assimilation cycles) because of the computational burden of running the ensemble forecasts. The use of 30 h instead of a multiple of a day is to avoid always computing the statistics in the same time of the diurnal cycle.

The Gaussianity can be defined by several different measures, such as the sophisticated relative entropy method, simpler methods based on hypothesis testing, or the even simpler skewness and kurtosis (Bocquet et al. 2010). Here, we use the χ^2 value similar to the hypothesis testing method to measure the deviation from the Gaussian distribution for a given distribution:

$$\chi^2 = \sum_{k=1}^K \frac{(\tilde{y}^{(k)} - \tilde{y}_{\text{expected}})^2}{\sigma^2}, \quad (14)$$

where K is the ensemble size, $\tilde{y}^{(k)}$ is the (transformed) observation value in the k th member that has been sorted in ascending order, and

$$\tilde{y}_{\text{expected}}^{(k)} = \bar{y} + \sigma F^{G-1} \left(\frac{k-0.5}{K} \right) \quad (15)$$

represents a realization of the expected Gaussian distribution. The expected mean \bar{y} and the expected variance σ^2 are determined from the background members $\tilde{y}^{(k)}$. The χ^2 value is a measure of the deviation from the expected Gaussian distribution, so it is a measure of the

non-Gaussianity. A larger χ^2 value means a more non-Gaussian background error distribution, and vice versa. After computing the χ^2 value for each precipitation observation at each cycle, they can be averaged in time or in any group to obtain meaningful statistics. Note that in this section all data points with at least one non-precipitation member are included in the statistics. The quality control criteria that were used in the data assimilation experiments are not applied here.

Figure 8 shows the average χ^2 values with respect to the number of precipitating background members, not using and using different transformation methods. The transformation methods here include the logarithm transformation with $\alpha = 0.6$ mm, and the Gaussian transformation with the CZ and BZ methods for zero precipitation transformations. The Gaussianity increases (χ^2 decreases) with the number of precipitating members, regardless of the transformation methods. Therefore, the current Gaussianity statistic provides a compelling reason for implementing the model background-based quality control criterion as proposed in LKM13. Comparing the χ^2 values computed with transformed precipitation to those computed with the original precipitation, it is found that the Gaussianity is considerably increased by using any of the three transformation methods when there are more than 8–10 precipitating members (Figs. 8e–g), while the transformation can also make the error distribution more non-Gaussian when there are too few precipitating members. Although the transformation methods seem to be deteriorating in the latter case, it is not a problem because, as indicated in section 3c, the quality control

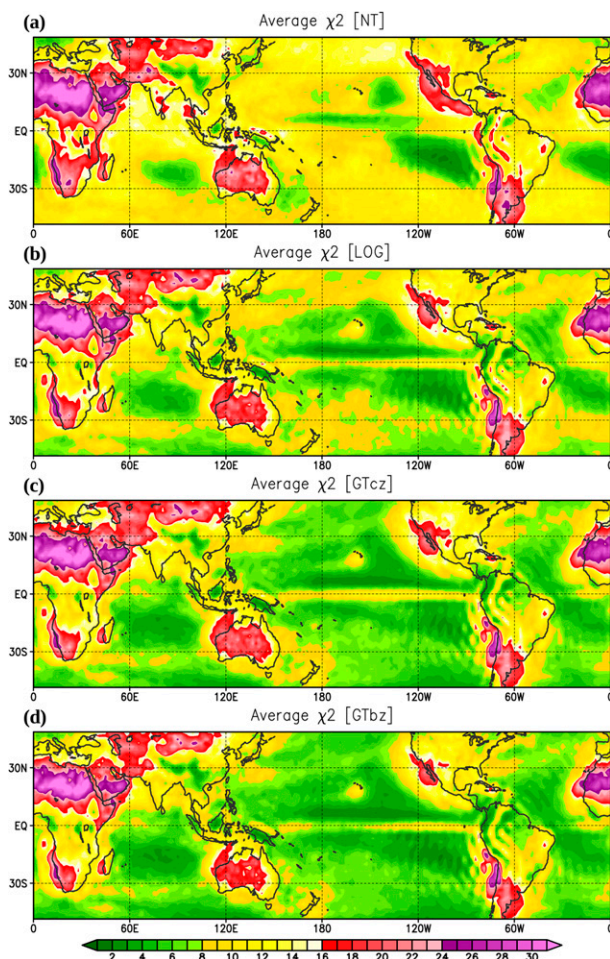


FIG. 9. The maps of the average non-Gaussianity χ^2 of background precipitation errors (a) without transformation; (b) with the logarithm transformation; and with the Gaussian transformation using the (c) CZ and (d) BZ methods. All data points with at least one nonprecipitation member are included in the statistics.

criterion monitoring the number of precipitating members (i.e., the 24mR criterion) prevents the observations from being assimilated in this situation. Comparing the different transformation methods, the Gaussian transformations are generally more effective than the logarithm transformation, and the GTbz method results in the most Gaussian background errors of precipitation, leading to as much as 60% improvement over using the original precipitation values when there are more than 24 precipitating members (Fig. 8g). The simpler GTcz method shows a similar effect as GTbz but suffers a little when most but not all members are precipitating (Fig. 8f).

Figure 9 shows the global distribution of the χ^2 values averaged in time. It is shown that the Gaussianity of the background precipitation error distribution can be very different in different regions. Generally, the Gaussianity of the oceanic precipitation is better than the continental

precipitation. In addition, wet areas show better Gaussianity than dry areas, and the infrequent precipitation in desert areas results in very non-Gaussian errors. However, one exception is that the marine stratocumulus regions show surprisingly good Gaussianity, probably because the model produces frequent drizzle over these regions (LKM16).

Comparing the results with no transformation and with the LOG, GTcz, and GTbz transformations (Fig. 9) we find that transforming the precipitation by any of these methods can improve the Gaussianity over most wet areas, but all of the methods fail to improve the Gaussianity over some dry areas, such as the desert in the Sahara and central Asia. Therefore, it would not be a good idea to carry out assimilation of precipitation over areas with very infrequent precipitation. The two Gaussian transformation methods (GTcz and GTbz) can improve the Gaussianity by 40%–60% in terms of the χ^2 values over the broad oceanic precipitation areas, which is better than the logarithm transformation method. In addition, GTbz is superior to GTcz, not only for its greater average amount of improvement, but also for its broader area of improvement.

6. Conclusions and discussion

This article is the second part of the GFS/TMPA precipitation data assimilation study. In the first part (LKM16), we studied the statistical properties of the model and observational precipitation in preparation for the actual assimilation of real TMPA precipitation observations. In this second part, we assimilated the TMPA data into the GFS model using the LETKF method, using the guidance we obtained in LKM16. This paper also follows our LKM13 study, where several new concepts of precipitation assimilation were tested using an idealized configuration (i.e., identical-twin OSSEs with a simplified atmospheric model) and promising results were first obtained. Therefore, the main goal of this study is to examine whether the same ideas are also applicable to a more realistic system with real satellite precipitation data, which is significantly more challenging than that with the idealized experiments. Another focus of this study is the comparison between different precipitation transformation methods: the Gaussian transformations with two different methods of handling zero values, the commonly used logarithm transformation, and no transformation.

The experiments with the Gaussian transformation of precipitation show clear positive impacts in all variables and all regions by assimilating the TMPA data. For the u winds, the forecast skill can be extended as much as 12 h by assimilating the TMPA data, meaning that the

model remembers the assimilation change over the entire 5-day forecasts. This is consistent to the results obtained with an idealized system (LKM13). In contrast, the precipitation assimilation without transformations leads to much degraded analyses and forecasts, as it did with the OSSEs. The LOG experiment shows smaller improvement than the Gaussian transformation experiments. It improves the wind and moisture fields but fails to improve the temperature field.

In addition to the Gaussian transformation and the quality control criterion based on the number of precipitating members in the ensemble background (i.e., the XmR criterion) proposed in LKM13, additional modifications were made in this study, without which the results would not have been as good as in the current experiments. The new modifications include the following:

- 1) Applying separate Gaussian transformations to the model precipitation and the observational precipitation independently based on their own different CDFs. An important advantage of this is that the amplitude-dependent bias between the model precipitation and observational precipitation can be corrected (LKM16).
- 2) Adopting a new quality control criterion based on the correlation between the long-term model background precipitation and the observation data in each grid point and each period of the year. The motivation for this criterion is to filter out the precipitation observations made at the locations and seasons where the model background and the observation are climatologically inconsistent. The inconsistency can arise from the deficient precipitation parameterization in the model and/or the problematic precipitation retrievals, but in either case, such inconsistent observational data would not be useful for assimilation.
- 3) The important issues of the zero precipitation transformation are also investigated in this real data assimilation study. A new method transforming zero precipitation to the median of the background zero probability, instead of the climatological zero, is proposed and tested. This new method considers the precipitation error distribution in the background ensemble, rather than the climatology, to define the transformation of the zero precipitation.

The Gaussianity of the background error distributions of precipitation before and after several transformation methods is also investigated. This is an important examination to verify whether our proposed (climatological) Gaussian transformation can actually improve the Gaussianity in the background errors at each observational

time and location. The conclusion is that, although it does not become perfectly Gaussian, the background error distribution is much more Gaussian after the variable transformations. This is consistent to the finding by Amezcua and van Leeuwen (2014) that the Gaussian anamorphosis is not a perfect solution of the EnKF assimilation of non-Gaussian variables, but it can be useful if properly implemented. Among the transformation methods used in this study, the GTbz method, which considers the background ensemble at the assimilation point, leads to the largest improvement of the Gaussianity of background errors. The GTcz method follows. The logarithm transformation is also helpful but much less effective.

Regarding the limitations of this study, it is important to note that the complexity of the current configuration is still intermediate between the OSSEs with simplified models and the real operational numerical weather prediction.

First, the model resolution is doubled from T30 in the SPEEDY model to the T62 GFS model, but this resolution is still low compared to state-of-the-art operational numerical weather prediction models. When the model resolution is further increased, the characteristics of the model precipitation errors could be very different from this study; thus, the way to specify the precipitation observation errors may need to be modified in the higher-resolution experiments. Considering the complexity of the nature of the precipitation errors, the application to precipitation assimilation of some kind of adaptive methods to objectively determine observation errors, such as those in Li et al. (2009), may be useful. On the other hand, the larger random error in high-resolution satellite precipitation datasets may also add more difficulties. In this study, the upscaling procedure (i.e., spatial average) plays a role in reducing the random errors in the observation dataset. For studies with the goal of improving medium-range forecasts, an interesting test would be to use a high-resolution model but still include an average operator in the observation operator to see if sacrificing the resolution of the precipitation observation can still help improve the high-resolution model forecasts.

Second, our baseline experiment, RAOBS, assimilates only rawinsonde observations, while operational weather forecasts assimilate much more. This provided our experiments with a large room for improvement to identify positive impacts from the additional assimilation of precipitation, so it does not prove that the precipitation assimilation will still be beneficial when more conventional and satellite observation data are assimilated. We expect that the overall improvement by precipitation assimilation may become smaller when more

observations are assimilated. Nevertheless, this work is certainly an important step forward toward the assimilation of global large-scale satellite precipitation estimates. Obtaining positive impacts by assimilating precipitation on top of a more accurate baseline experiment would be an important future goal.

Finally, the Global Precipitation Measurement (GPM) mission has been launched. It can provide more accurate real-time precipitation estimates at much better spatial and temporal coverage. With this higher-quality precipitation data, we expect that larger impacts of the assimilation of precipitation can be achieved.

Acknowledgments. This study was done as part of Guo-Yuan Lien's Ph.D. thesis work at the University of Maryland, partially supported by NASA Grants NNX11AH39G, NNX11AL25G, and NNX13AG68G; NOAA Grants NA100OAR4310248 and CICS-PAEK-LETKF11; and the Office of Naval Research (ONR) Grant N000141010149 under the National Oceanographic Partnership Program (NOPP). George Huffman provided valuable guidance on the characteristics of the TMPA data. We also gratefully acknowledge the support from the Japan Aerospace Exploration Agency (JAXA) Precipitation Measuring Mission (PMM).

APPENDIX

The Background Median of Zero (BZ) Method of Transformation for Zero Precipitation

As shown in Fig. 1, the boundary of zero and nonzero precipitation is associated with two probability distributions, the climatological distribution and the background error distribution, through the zero precipitation probability in the climatology and in the background ensemble. Both these two distributions are assumed to be Gaussian, where the climatological one is known but the one representing the background error distribution is hypothesized. Therefore, we can write the first equality in the form of inverse CDFs of normal distributions:

$$\tilde{y}_{\text{trace}} = F^{G^{-1}}(P_c) = F^{G'^{-1}}(P_b), \quad (\text{A1})$$

where \tilde{y}_{trace} is the trace value of precipitation (i.e., the boundary of zero and nonzero precipitation) in the transformed variable; $F^{G^{-1}}$ is the inverse CDF of a normal distribution with zero mean and unit standard deviation as we defined in Eq. (4); $F^{G'^{-1}}$ is the inverse CDF of an unknown normal distribution; $P_c = F(y_{\text{trace}})$ is the zero precipitation probability determined from the climatology; and P_b is the zero precipitation probability determined from the background ensemble as in Eq. (8).

Assuming the unknown normal distribution representing the background error distribution has a mean \bar{y} and a standard deviation σ so that $F^{G'^{-1}}(P_b) = \bar{y} + \sigma F^{G^{-1}}(P_b)$, Eq. (A1) becomes

$$\tilde{y}_{\text{trace}} = F^{G^{-1}}(P_c) = \bar{y} + \sigma F^{G^{-1}}(P_b). \quad (\text{A2})$$

To determine the hypothetic background error distribution, what we need to do is to solve these two unknown variables (\bar{y} and σ) so that the zero precipitation transformation defined by the median of the background zero precipitation probability can be represented by Eq. (9).

Now we are going to solve \bar{y} and σ . To solve them, we need another equality, which can be constructed through the calculation of the expected value for the transformed background error distribution. Since the distribution is hypothesized to be Gaussian, its expected value, consisting of the zero precipitation part and the nonzero precipitation part, should be equal to its mean \bar{y} :

$$P_b E[\tilde{y} | \tilde{y} < \tilde{y}_{\text{trace}}] + (1 - P_b) E[\tilde{y} | \tilde{y} \geq \tilde{y}_{\text{trace}}] = \bar{y}, \quad (\text{A3})$$

where $E[\cdot]$ is the expected value of truncated distributions. Assuming an exact Gaussian distribution for the zero precipitation part, the analytical representation of the first term in Eq. (A3) is

$$P_b E[\tilde{y} | \tilde{y} < \tilde{y}_{\text{trace}}] = \int_{-\infty}^{\tilde{y}_{\text{trace}}} \tilde{y} f(\tilde{y}) d\tilde{y}, \quad (\text{A4})$$

where $f(\tilde{y})$ is the PDF of the hypothetic Gaussian background error distribution with the transformed variable, and by definition it can be represented by

$$f(\tilde{y}) = \frac{1}{\sigma} \phi\left(\frac{\tilde{y} - \bar{y}}{\sigma}\right), \quad \phi(y') = \frac{e^{-y'^2/2}}{\sqrt{2\pi}}. \quad (\text{A5})$$

Let $y' = (\tilde{y} - \bar{y})/\sigma$ and substitute it into the integral; Eq. (A4) becomes

$$\begin{aligned} \int_{-\infty}^{\tilde{y}_{\text{trace}}} \tilde{y} f(\tilde{y}) d\tilde{y} &= \int_{-\infty}^{F^{G^{-1}}(P_b)} (\bar{y} + \sigma y') \phi(y') dy' \\ &= \bar{y} \int_{-\infty}^{F^{G^{-1}}(P_b)} \phi(y') dy' + \sigma \int_{-\infty}^{F^{G^{-1}}(P_b)} y' \phi(y') dy'. \end{aligned} \quad (\text{A6})$$

Note the Eq. (A2) is used to substitute the upper bound of the integral. Meanwhile, performing a similar derivation, we can obtain

$$P_b = \int_{-\infty}^{\tilde{y}_{\text{trace}}} f(\tilde{y}) d\tilde{y} = \int_{-\infty}^{F^{G^{-1}}(P_b)} \phi(y') dy' \quad (\text{A7})$$

so that

$$\begin{aligned} \int_{-\infty}^{\tilde{y}_{\text{trace}}} \tilde{y} f(\tilde{y}) d\tilde{y} &= P_b \bar{y} + \sigma \int_{-\infty}^{F^{G^{-1}}(P_b)} y' \phi(y') dy' \\ &= P_b \bar{y} - \sigma \phi[F^{G^{-1}}(P_b)]. \end{aligned} \quad (\text{A8})$$

The second term in Eq. (A3) can be computed during the LETKF assimilation computation using the discrete precipitating background ensemble members:

$$(1 - P_b)E[\tilde{y} | \tilde{y} \geq \tilde{y}_{\text{trace}}] = \frac{1}{N} \sum_{i=n_1+1}^N \tilde{y}_i, \quad (\text{A9})$$

where \tilde{y}_i is the transformed background precipitation value in the i th member, assuming these ensemble values have been sorted in ascending order; N is the ensemble size; and n_1 is the number of ensemble members with zero precipitation (i.e., this summation only includes the precipitating members). Finally, we can represent the Eq. (A3) as

$$P_b \bar{y} + \sigma \alpha + \beta = \bar{y}, \quad (\text{A10})$$

where the definitions of α and β are the same as in Eq. (12).

From Eqs. (A2) and (A10), we can solve \bar{y} and σ as in Eqs. (10) and (11). Plugging in these two values into Eq. (9), we thus obtain the value to which zero precipitation should be transformed in this BZ method. The clean solution is summarized in Eqs. (8)–(12).

REFERENCES

- Amezcu, J., and P. J. van Leeuwen, 2014: Gaussian anamorphosis in the analysis step of the EnKF: A joint state-variable/observation approach. *Tellus*, **66A**, 23493, doi:10.3402/tellusa.v66.23493.
- Anderson, J. L., 2010: A non-Gaussian ensemble filter update for data assimilation. *Mon. Wea. Rev.*, **138**, 4186–4198, doi:10.1175/2010MWR3253.1.
- Bauer, P., J.-F. Mahfouf, W. S. Olson, F. S. Marzano, S. D. Michele, A. Tassa, and A. Mugnai, 2002: Error analysis of TMI rainfall estimates over ocean for variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **128**, 2129–2144, doi:10.1256/003590002320603575.
- , G. Ohning, C. Kummerow, and T. Auligne, 2011: Assimilating satellite observations of clouds and precipitation into NWP models. *Bull. Amer. Meteor. Soc.*, **92**, ES25–ES28, doi:10.1175/2011BAMS3182.1.
- Bocquet, M., C. A. Pires, and L. Wu, 2010: Beyond Gaussian statistical modeling in geophysical data assimilation. *Mon. Wea. Rev.*, **138**, 2997–3023, doi:10.1175/2010MWR3164.1.
- Bowman, K. P., 2005: Comparison of TRMM precipitation retrievals with rain gauge data from ocean buoys. *J. Climate*, **18**, 178–190, doi:10.1175/JCLI3259.1.
- Davolio, S., and A. Buzzi, 2004: A nudging scheme for the assimilation of precipitation data into a mesoscale model. *Wea. Forecasting*, **19**, 855–871, doi:10.1175/1520-0434(2004)019<0855:ANSFTA>2.0.CO;2.
- Ebert, E. E., J. E. Janowiak, and C. Kidd, 2007: Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bull. Amer. Meteor. Soc.*, **88**, 47–64, doi:10.1175/BAMS-88-1-47.
- Errico, R. M., D. J. Stensrud, and K. D. Raeder, 2001: Estimation of the error distributions of precipitation produced by convective parametrization schemes. *Quart. J. Roy. Meteor. Soc.*, **127**, 2495–2512, doi:10.1002/qj.49712757802.
- , P. Bauer, and J.-F. Mahfouf, 2007: Issues regarding the assimilation of cloud and precipitation data. *J. Atmos. Sci.*, **64**, 3785–3798, doi:10.1175/2006JAS2044.1.
- Eyink, G. L., and S. Kim, 2006: A maximum entropy method for particle filtering. *J. Stat. Phys.*, **123**, 1071–1128, doi:10.1007/s10955-006-9124-9.
- Falkovich, A., E. Kalnay, S. Lord, and M. B. Mathur, 2000: A new method of observed rainfall assimilation in forecast models. *J. Appl. Meteor.*, **39**, 1282–1298, doi:10.1175/1520-0450(2000)039<1282:ANMOOR>2.0.CO;2.
- Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt, 2011: Balance and ensemble Kalman filter localization techniques. *Mon. Wea. Rev.*, **139**, 511–522, doi:10.1175/2010MWR3328.1.
- Huffman, G. J., and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.*, **8**, 38–55, doi:10.1175/JHM560.1.
- , R. Adler, D. Bolvin, and E. Nelkin, 2010: The TRMM Multi-Satellite Precipitation Analysis (TMPA). *Satellite Rainfall Applications for Surface Hydrology*, M. Gebremichael and F. Hossain, Eds., Springer, 3–22.
- , E. F. Stocker, D. T. Bolvin, and E. J. Nelkin, 2012: TRMM Multisatellite Precipitation Analysis. TRMM_3B42, version 7. NASA Goddard Space Flight Center, accessed 25 July 2012. [Available online at http://mirador.gsfc.nasa.gov/collections/TRMM_3B42_007.shtml.]
- Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112–126, doi:10.1016/j.physd.2006.11.008.
- Koizumi, K., Y. Ishikawa, and T. Tsuyuki, 2005: Assimilation of precipitation data to the JMA mesoscale model with a four-dimensional variational method and its impact on precipitation forecasts. *SOLEA*, **1**, 45–48, doi:10.2151/solea.2005-013.
- Li, H., E. Kalnay, and T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **135**, 523–533, doi:10.1002/qj.371.
- Lien, G.-Y., E. Kalnay, and T. Miyoshi, 2013: Effective assimilation of global precipitation: Simulation experiments. *Tellus*, **65A**, 19915, doi:10.3402/tellusa.v65i0.19915.
- , —, —, and G. J. Huffman, 2016: Statistical properties of global precipitation in the NCEP GFS model and TMPA observations for data assimilation. *Mon. Wea. Rev.*, **144**, 663–679, doi:10.1175/MWR-D-15-0150.1.
- Lopez, P., 2011: Direct 4D-Var assimilation of NCEP stage IV radar and gauge precipitation data at ECMWF. *Mon. Wea. Rev.*, **139**, 2098–2116, doi:10.1175/2010MWR3565.1.
- , 2013: Experimental 4D-Var assimilation of SYNOP rain gauge data at ECMWF. *Mon. Wea. Rev.*, **141**, 1527–1544, doi:10.1175/MWR-D-12-00024.1.

- Mahfouf, J., B. Brasnett, and S. Gagnon, 2007: A Canadian precipitation analysis (CaPA) project: Description and preliminary results. *Atmos.–Ocean*, **45**, 1–17, doi:[10.3137/ao.v450101](https://doi.org/10.3137/ao.v450101).
- Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, doi:[10.1175/BAMS-87-3-343](https://doi.org/10.1175/BAMS-87-3-343).
- Miyoshi, T., 2011: The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Mon. Wea. Rev.*, **139**, 1519–1535, doi:[10.1175/2010MWR3570.1](https://doi.org/10.1175/2010MWR3570.1).
- , and K. Aranami, 2006: Applying a four-dimensional local ensemble transform Kalman filter (4D-LETKF) to the JMA nonhydrostatic model (NHM). *SOLA*, **2**, 128–131, doi:[10.2151/sola.2006-033](https://doi.org/10.2151/sola.2006-033).
- , and S. Yamane, 2007: Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. *Mon. Wea. Rev.*, **135**, 3841–3861, doi:[10.1175/2007MWR1873.1](https://doi.org/10.1175/2007MWR1873.1).
- Molteni, F., 2003: Atmospheric simulations using a GCM with simplified physical parametrizations. I: Model climatology and variability in multi-decadal experiments. *Climate Dyn.*, **20**, 175–191, doi:[10.1007/s00382-002-0268-2](https://doi.org/10.1007/s00382-002-0268-2).
- Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415–428, doi:[10.1111/j.1600-0870.2004.00076.x](https://doi.org/10.1111/j.1600-0870.2004.00076.x).
- Schöniger, A., W. Nowak, and H.-J. Hendricks Franssen, 2012: Parameter estimation by ensemble Kalman filters with transformed data: Approach and application to hydraulic tomography. *Water Resour. Res.*, **48**, W04502, doi:[10.1029/2011WR010462](https://doi.org/10.1029/2011WR010462).
- Simon, E., and L. Bertino, 2009: Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: A twin experiment. *Ocean Sci.*, **5**, 495–510, doi:[10.5194/os-5-495-2009](https://doi.org/10.5194/os-5-495-2009).
- Tian, Y., and C. D. Peters-Lidard, 2010: A global map of uncertainties in satellite-based precipitation measurements. *Geophys. Res. Lett.*, **37**, L24407, doi:[10.1029/2010GL046008](https://doi.org/10.1029/2010GL046008).
- Tsuyuki, T., 1996: Variational data assimilation in the tropics using precipitation data. Part II: 3D model. *Mon. Wea. Rev.*, **124**, 2545–2561, doi:[10.1175/1520-0493\(1996\)124<2545:VDAITT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<2545:VDAITT>2.0.CO;2).
- , 1997: Variational data assimilation in the tropics using precipitation data. Part III: Assimilation of SSM/I precipitation rates. *Mon. Wea. Rev.*, **125**, 1447–1464, doi:[10.1175/1520-0493\(1997\)125<1447:VDAITT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<1447:VDAITT>2.0.CO;2).
- , and T. Miyoshi, 2007: Recent progress of data assimilation methods in meteorology. *J. Meteor. Soc. Japan*, **85B**, 331–361, doi:[10.2151/jmsj.85B.331](https://doi.org/10.2151/jmsj.85B.331).
- van Leeuwen, P. J., 2009: Particle filtering in geophysical systems. *Mon. Wea. Rev.*, **137**, 4089–4114, doi:[10.1175/2009MWR2835.1](https://doi.org/10.1175/2009MWR2835.1).
- Wackernagel, H., 2003: *Multivariate Geostatistics*. Springer, 408 pp.
- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, doi:[10.1175/1520-0493\(2004\)132<1238:IOIEAO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2).
- Zhang, S. Q., M. Zupanski, A. Y. Hou, X. Lin, and S. H. Cheung, 2013: Assimilation of precipitation-affected radiances in a cloud-resolving WRF ensemble data assimilation system. *Mon. Wea. Rev.*, **141**, 754–772, doi:[10.1175/MWR-D-12-00055.1](https://doi.org/10.1175/MWR-D-12-00055.1).
- Zupanski, D., S. Q. Zhang, M. Zupanski, A. Y. Hou, and S. H. Cheung, 2011: A prototype WRF-based ensemble data assimilation system for dynamically downscaling satellite precipitation observations. *J. Hydrometeor.*, **12**, 118–134, doi:[10.1175/2010JHM1271.1](https://doi.org/10.1175/2010JHM1271.1).