

## Statistical Properties of Global Precipitation in the NCEP GFS Model and TMPA Observations for Data Assimilation

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*

EUGENIA KALNAY

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

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*

GEORGE J. HUFFMAN

*Mesoscale Atmospheric Processes Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland*


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

### ABSTRACT

Assimilation of satellite precipitation data into numerical models presents several difficulties, with two of the most important being the non-Gaussian error distributions associated with precipitation, and large model and observation errors. As a result, improving the model forecast beyond a few hours by assimilating precipitation has been found to be difficult. To identify the challenges and propose practical solutions to assimilation of precipitation, statistics are calculated for global precipitation in a low-resolution NCEP Global Forecast System (GFS) model and the TRMM Multisatellite Precipitation Analysis (TMPA). The samples are constructed using the same model with the same forecast period, observation variables, and resolution as in the follow-on GFS/TMPA precipitation assimilation experiments presented in the companion paper.

The statistical results indicate that the T62 and T126 GFS models generally have positive bias in precipitation compared to the TMPA observations, and that the simulation of the marine stratocumulus precipitation is not realistic in the T62 GFS model. It is necessary to apply to precipitation either the commonly used logarithm transformation or the newly proposed Gaussian transformation to obtain a better relationship between the model and observational precipitation. When the Gaussian transformations are separately applied to the model and observational precipitation, they serve as a bias correction that corrects the amplitude-dependent biases. In addition, using a spatially and/or temporally averaged precipitation variable, such as the 6-h accumulated precipitation, should be advantageous for precipitation assimilation.

---

 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-15-0150.1

© 2016 American Meteorological Society

### 1. Introduction

In recent years, several global precipitation estimations from a variety of remote sensing platforms have become available, such as the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA; Huffman et al. 2007, 2010) and the Global Satellite Mapping of Precipitation (GSMaP; Ushio et al. 2009).

Meanwhile, many efforts to assimilate precipitation observations have also been made (e.g., Tsuyuki 1996, 1997; Falkovich et al. 2000; Davolio and Buzzi 2004; Koizumi et al. 2005; Mesinger et al. 2006). However, serious difficulties still remain in assimilating the precipitation data. For example, most data assimilation schemes, including the variational methods and the ensemble Kalman filter (EnKF) methods, assume Gaussian error distributions for both observation and model background. If the error distribution is not Gaussian, the analysis may not be optimal. Since the precipitation-related variables are far from Gaussian, the non-Gaussianity issue becomes a severe problem for precipitation assimilation. Besides, both the model errors and observation errors are important issues for precipitation assimilation. As a consequence, a widely shared experience is that the precipitation assimilation can be useful in improving the model analyses, but the forecast improvement is usually limited to the first few forecast hours (e.g., Falkovich et al. 2000; Davolio and Buzzi 2004; Tsuyuki and Miyoshi 2007). These issues have been discussed and summarized in several articles, such as Errico et al. (2007), Bauer et al. (2011), and Lien et al. (2013, hereafter LKM13). Notwithstanding these difficulties, several recent studies have shown some usefulness of precipitation assimilation (Lopez 2011, 2013; Zupanski et al. 2011; Zhang et al. 2013).

Data assimilation methods that do not assume Gaussian errors, such as a particle filter, are usually too computationally expensive. Alternatively, a variable transformation technique is a computationally cheaper and practical solution to mitigate the non-Gaussianity problem in realistic geophysical data assimilation systems (Bocquet et al. 2010; LKM13; Amezcua and van Leeuwen 2014). For precipitation data assimilation, the precipitation values are usually transformed by a logarithmic function before assimilating them into the model (e.g., Lopez 2011). Instead of the logarithmic transformation, LKM13 proposed to apply the Gaussian anamorphosis method to precipitation based on its model climatology, under the assumption that a forecast variable with more Gaussian climatological distribution would result in a more Gaussian error distribution. With this transformation, they succeeded in showing effective assimilation of global precipitation in their proof-of-concept observing system simulation experiments (OSSEs), using a simplified general circulation model and the local ensemble transform Kalman filter (LETKF). In their experiments, precipitation assimilation not only improves the analyses but also improves the model forecasts over the entire 5-day forecast period in their experiments.

Although a significant forecast improvement by precipitation assimilation was demonstrated in LKM13 with an idealized system, in real systems improvements

are generally very limited or even absent. The distinct challenges associated with the use of realistic model and real observations include the large and unknown errors related not only to the moist physical parameterization in the model but also to the observations. Since both the model precipitation and the observations could have large errors, the long-term statistics of these two quantities may be very different, which is harmful to the data assimilation use. Therefore, before performing real precipitation data assimilation, it is worthwhile to first investigate the statistical characteristics of precipitation in both model and observation datasets that we would like to use, which are presented in this paper.

We investigate the differences in probability distributions between the precipitation in a series of short-term model forecasts and a precipitation observation dataset, to isolate the different characteristics of the real model and observations. It is noted that the challenges introduced by these differences could not be addressed in LKM13 since they used an identical-twin OSSE method. Here we use more realistic settings: the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS), running at a low resolution, and the TMPA data as the precipitation observations. Given the low resolution feasible in our study, the main focus of our work is assimilation of the global large-scale precipitation, which could be particularly important for improving medium-range model forecasts. Since the probability distributions are dependent on the use (or lack of use) of variable transformations, the results with different transformation methods are investigated. We also show the correlation between model forecasts and observations at each grid point on a map. Several suggestions for real-data precipitation assimilation are made in the concluding section of this article. Although we choose to use the NCEP GFS model and the TMPA data to study the precipitation data assimilation, the same analysis can also be performed with other models and observation datasets.

The paper is organized as follows. The GFS model and TMPA observations are briefly introduced in section 2. Section 3 describes the transformation methods we will use in the precipitation statistics. A series of statistical results are then presented in the following sections. Section 4 shows the cumulative distribution functions (CDFs) of the precipitation data, which will be used to define the Gaussian transformation of precipitation. Section 5 shows the joint probability distribution diagrams between the model and observational precipitation and compares the results in terms of the transformation methods, the temporal integration of precipitation, and the resolution of precipitation data. Section 6 presents the geographic distribution of correlation scores between these two

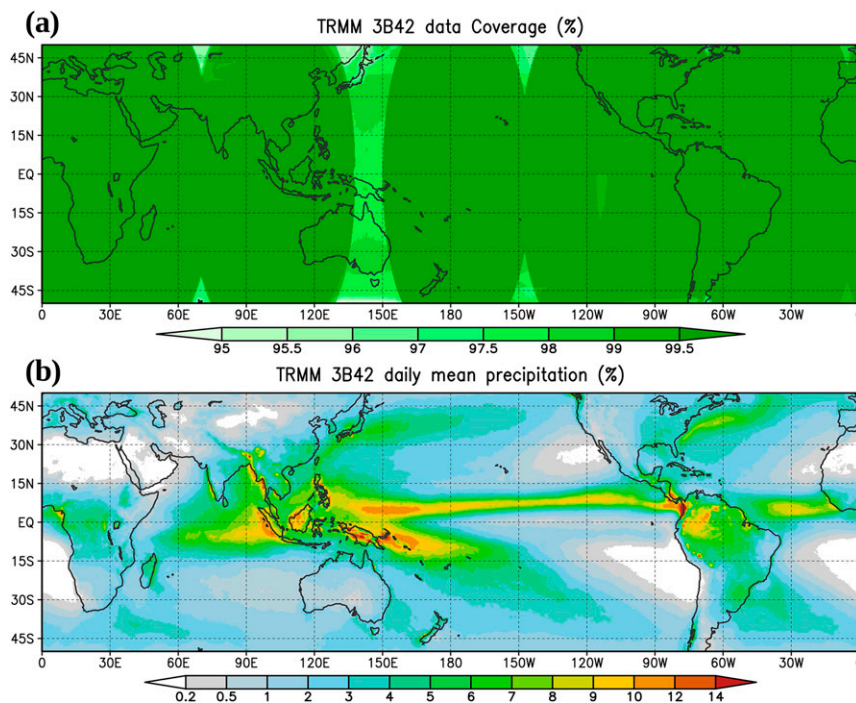


FIG. 1. (a) The data coverage rate (%) and (b) the mean daily precipitation (mm) of the 14-yr (1998–2011) TMPA.

variables. Conclusions and suggestions for the precipitation assimilation are given in section 7. In addition, the successful assimilation of the TMPA data following the guidance derived from this study is presented in a separate paper (Lien et al. 2016, hereafter LMK16).

## 2. The model and observations

The GFS model is the operational global numerical weather prediction (NWP) model used at the NCEP. It is one of the major world state-of-the-art operational NWP model. The GFS model can be run at various spectral resolutions on a hybrid sigma/pressure coordinate. In this study we focus on the large-scale global precipitation and also consider the computational constraints, so the experiments and analyses are done with two lower-resolution configurations: T62 and T126 (roughly equivalent to 200- and 100-km horizontal resolutions) with 64 vertical levels (L64). Convective precipitation is parameterized using a modified Simplified Arakawa–Schubert (SAS) scheme (Pan and Wu 1995; Han and Pan 2011), considering both deep and shallow convection.

The TMPA (Huffman et al. 2007, 2010) is a gridded precipitation dataset compiled from multiple satellite sensors. It has a global coverage from 50°S to 50°N with 0.25° spatial resolution and 3-h temporal resolution. The estimated surface precipitation rate is provided. The

primary data sources are the low-earth-orbit satellites such as the TRMM Microwave Imager (TMI), the Special Sensor Microwave Imager (SSM/I) and Special Sensor Microwave Imager/Sounder (SSMIS) on the Defense Meteorological Satellite Program (DMSP) satellites, the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) on *Aqua*, the Advanced Microwave Sounding Unit-B (AMSU-B) on the National Oceanic and Atmospheric Administration (NOAA) satellite series, and the Microwave Humidity Sounder (MHS) on both the NOAA and the EUMETSAT MetOp series. The microwave satellite observations have a strong physical relationship to the hydrometeors and thus the surface precipitation, but they are spatially and temporally inhomogeneous. To fill the gaps left from the low-earth-orbit sensors, the infrared (IR) data collected by the geosynchronous-earth-orbit satellites are used as the secondary data sources with calibration by the microwave precipitation estimates, though the accuracy of precipitation derived from the IR data is lower. For the research version (i.e., not in real time) of the TMPA, these satellite-derived precipitation amounts are further rescaled based on monthly rain gauge analyses to achieve accurate statistics in the climatological scale, while in the real-time version the satellite-derived precipitation is rescaled with a climatological correction to the research version. With the above data processing procedure, the TMPA has a very high (>95%) data coverage rate

(Fig. 1a), thus becoming a potential good observational source for the assimilation of global precipitation. In this study, we use version 7 of the TMPA research products, labeled as 3B42, released in 2012 (Huffman et al. 2012). The climatological mean daily precipitation computed from the 14-yr TMPA data (1998–2011) is shown in Fig. 1b.

To make the 0.25°-resolution TMPA data correspond to the lower resolutions of the T62/T126 GFS model, we preprocess the precipitation rate data, upscaling the original TMPA grids to the T62 or T126 Gaussian grids used by the GFS model using an area-conserving remapping.

### 3. Transformation of precipitation

In this section, several transformations for precipitation assimilation are described, including the widely used logarithm transformation, and the transformation based on Gaussian anamorphosis used in previous studies such as Simon and Bertino (2009), Schöniger et al. (2012), and LKM13. The transformations have a profound impact on the statistical results shown in later sections.

#### a. Logarithm transformation

The logarithm transformation

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

is a simple and frequently used method to transform precipitation. Here,  $y$  is the original variable,  $\tilde{y}$  is the transformed variable, and  $\alpha$  is a tunable constant added to prevent the singularity at zero precipitation ( $y = 0$ ). Using the logarithm transformation, Lopez (2011) successfully assimilated the NCEP stage-IV precipitation analysis over the eastern United States, and Lopez (2013) presented experimental results of assimilation of the 6-hourly accumulated precipitation observations measured by the rain gauges at synoptic stations.

#### b. Gaussian transformation

The logarithm transformation may be helpful for precipitation assimilation in some regions, seasons, or precipitation types, but a globally invariant analytical transformation may not be applicable to every case. Therefore, following LKM13, we will also examine the effect of the Gaussian transformation on the precipitation statistics. Here we briefly summarize the formulation of the Gaussian transformation in LKM13 and explain the changes made in this study after LKM13.

##### 1) GENERAL FORMULA

The transformation is made by equating the two CDFs of the original variable  $y$  and the transformed variable  $\tilde{y}$ :

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

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

where  $F$  is the CDF of  $y$ ,  $\tilde{F}$  is the CDF of  $\tilde{y}$ , and  $\tilde{F}^{-1}$  is the inverse function of  $\tilde{F}$ . By definition, the CDFs are bounded within  $[0, 1]$ . The CDF of the original variable  $F$  is empirically determined from samples, and the CDF of the transformed variable  $\tilde{F}$  can be arbitrarily chosen so that the transformed variable can have any desired distribution. If we choose

$$\tilde{F}(\tilde{y}) = F^G(\tilde{y}) = \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{\tilde{y}}{\sqrt{2}} \right) \right], \quad (4)$$

which is the CDF of a standard normal distribution with zero mean and unit variance, and  $\operatorname{erf}$  is the error function, then

$$F^{G^{-1}}(P) = \sqrt{2} \operatorname{erf}^{-1}(2P - 1) \quad (5)$$

where  $P$  is the cumulative probability, so that it becomes a ‘‘Gaussian anamorphosis’’ (e.g., Wackernagel 2003):

$$\tilde{y} = F^{G^{-1}}[F(y)]. \quad (6)$$

In this way, the transformed variable  $\tilde{y}$  becomes a Gaussian variable. The use of the Gaussian anamorphosis has appeared in several geophysical data assimilation studies (e.g., Simon and Bertino 2009, 2012; Schöniger et al. 2012). We call this method ‘‘Gaussian transformation’’ hereafter.

Figure 2 provides an illustration of the Gaussian transformation procedure. It displays the 10-yr climatological probability density function (PDF) and CDF of the original and transformed precipitation in both the GFS model forecasts and the TMPA dataset, at three selected locations for the 11–20 January period. The collection of the model and observational precipitation samples will be discussed in later sections, but here we first use the plots to visualize the method. The transformation starts from Figs. 2a, 2e, and 2i, which are the very non-Gaussian PDFs of the original variables. The red color stands for the model precipitation and the green color stands for the observational precipitation. Their CDFs are then calculated (Figs. 2c,g,k). Using the inverse CDF of the standard normal distribution  $F^{G^{-1}}$ , the cumulative probability values are converted into the transformed variables  $\tilde{y}$ , whose CDFs shown in Figs. 2d, 2h, and 2l and PDFs are shown in Figs. 2b, 2f, and 2j. It is important to note that the precipitation distribution contains a great portion of zero values, shown as a delta function in the PDFs and a discontinuity in the CDFs, which need to be treated in a special

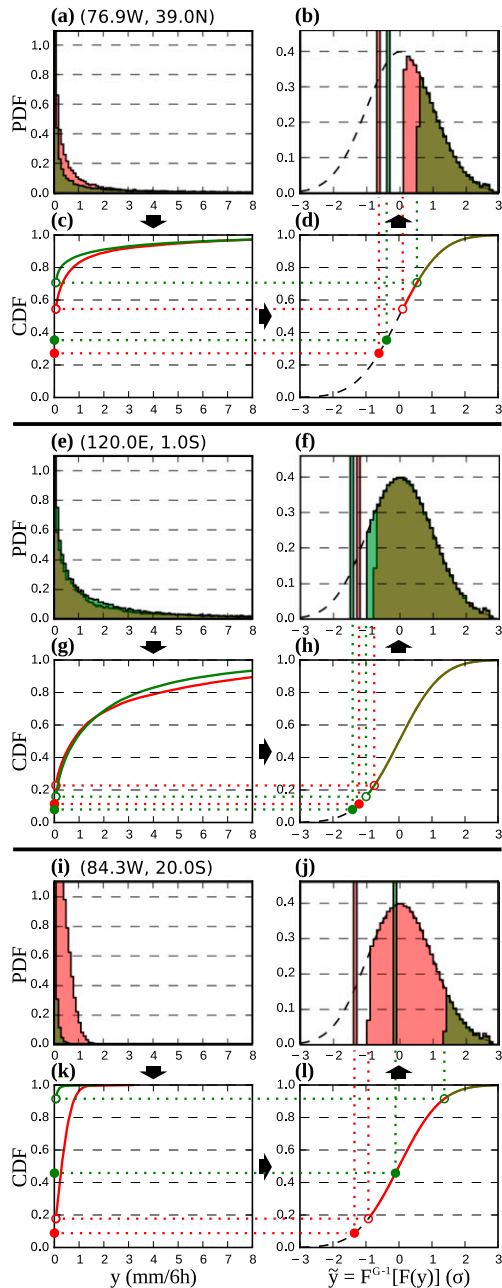


FIG. 2. The PDF and CDF of the original precipitation and the transformed precipitation based on the 10-yr (2001–10) model (red) and observation (green) climatologies for (a)–(d) a grid point in the extratropics (39.0°N, 76.9°W; near Maryland); (e)–(h) a grid point in the tropics (1.0°S, 120.0°E); and (i)–(l) a grid point in a marine stratocumulus region west of South America (20.0°S, 84.3°W). All plots correspond to the 11–20 Jan period. The procedure of the Gaussian transformation is indicated by the arrows [i.e., (a) to (c) to (d) to (b)]. The open circles correspond to the zero precipitation probability and the solid circles correspond to the half value (median) of the zero precipitation probability.

manner. Following LKM13, all the zero values are represented by half of the zero precipitation cumulative probability (i.e., the median; solid circles in Fig. 2) during the transformation:

$$F(0) = \frac{1}{2} P_c, \tag{7}$$

where  $P_c$  is the zero precipitation probability in the climatology. In this way, the zero precipitation is still a delta function in the transformed variable, but it is located at a certain distance away from the trace precipitation values.

This method transforms the climatological distribution of the model forecast variable into a Gaussian distribution, but this does not necessarily make the *background error distributions* Gaussian, as required in the EnKF data assimilation (e.g., Ott et al. 2004). However, it is reasonable to assume that a forecast with a more Gaussian climatological distribution would also result in more Gaussian forecast error distribution (LKM13). It is difficult to validate this assumption using the climatological data in this study but we do provide a validation of this assumption in the companion paper (LMK16) using the actual experimental data from the cycling LETKF data assimilation.

It is worth mentioning that this CDF-based transformation of precipitation has also been used in some climate studies, though they are not related to data assimilation. For example, the standardized precipitation index (SPI) (McKee et al. 1993; Guttman 1999) commonly used to study drought is defined based on a similar method, but the time scales of precipitation accumulations they have focused on are much longer than the 6 h used in weather data assimilation.

## 2) COMPUTATION OF THE CDFS AND TRANSFORMATIONS

Some technical details are described in this subsection. First, we regard all precipitation values smaller than  $0.06 \text{ mm (6 h)}^{-1}$  as “zero precipitation” because such small values in the model or observational precipitation data are irrelevant. This choice of the threshold value leads to good experimental results with the current GFS/TMPA assimilation system. Note that the optimal value of the threshold could change with different models, observations, or assimilation techniques; however, the current choice is close to the threshold used in LKM13,  $0.1 \text{ mm (6 h)}^{-1}$ .

Second, extreme values with cumulative distribution less than 0.001 and greater than 0.999 are set to 0.001 and 0.999, respectively. Consequently, when the original values fall outside the range in the climatological

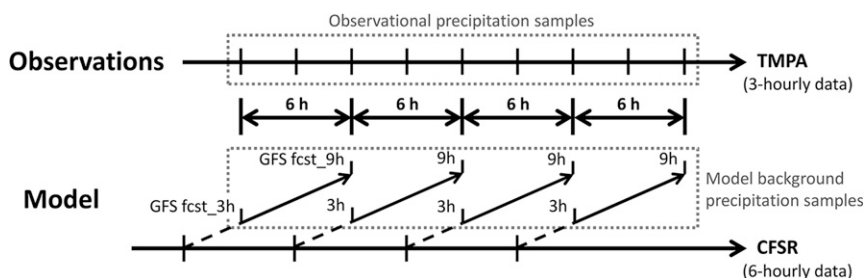


FIG. 3. Schematic of the preparation of precipitation samples from the (top) TMPA observation dataset and the (bottom) GFS model forecasts. For precipitation observations, a 10-yr series of the 3-hourly TMPA data is collected; for model background precipitation, equivalent 10-yr data are formed from a series of 9-h GFS model forecasts every 6 h initialized from the 10-yr CFSR reanalysis. In each forecast cycle, the forecast is conducted with the desired model configuration and resolutions (T62 and T126 in this study), and only the 3–9-h forecasts are used.

samples, they will be transformed to  $-3.09$  and  $3.09$ . It is noted for reference that [Simon and Bertino \(2012\)](#) also discussed this problem and they used parametric linear tails to form their transformation.

Third, we derive the CDFs from precipitation samples using constant-width bins with respect to the cumulative probability in  $[0, 1]$ , not with respect to the precipitation amount as it might be intuitively done. A total of 200 bins are used. The CDFs are thus represented by the 201 (including 0 and 1) discretized precipitation amounts at each cumulative distribution levels at a 0.005 increment. When we need to compute  $F(y)$  for a given precipitation value  $y$ , we perform a linear interpolation from the two nearby data points. Compared to binning with respect to the precipitation amount, this method can more precisely represent the CDF curves using the same number of the bins, particularly for large precipitation values.

### 3) SEPARATE GAUSSIAN TRANSFORMATION APPLIED TO MODEL BACKGROUND AND OBSERVATIONS

Following the methods described above, we can apply the Gaussian transformation to the GFS model and the TMPA data. However, there is an important difference between the Gaussian transformation used in [LKM13](#) and in this study. In [LKM13](#), the transformation was defined purely based on the 10-yr model precipitation climatology, and the same transformation was used for both the model and observational precipitation. There was no need to consider the transformations of the model and observational precipitation separately because the work used an identical-twin configuration so that the two CDFs are identical. In contrast, in this study with a realistic model and real observations, the transformations need to be defined separately for model precipitation and observations (see red and green colors in [Fig. 2](#)). Specifically, the

transformation of the model precipitation is performed based on the CDF computed from the model climatology; and the transformation of the precipitation observations is performed based on the CDF computed from the observation climatology. In this way, the model climatology and the observation climatology are first converted to the same 0–1 scale of their cumulative distribution using the corresponding transformation ([Fig. 2d](#)), then the same  $F^{G-1}$  is applied to obtain the Gaussian variables ([Fig. 2b](#)). Therefore, this method can essentially remove the climatological bias between these two variables that is dependent on the precipitation values, referred to as the “amplitude-dependent bias.” The effect of the separate transformations can be large because the precipitation distribution of the model and observational precipitation can be very different at some regions (e.g., [Figs. 2i–l](#)), which will be discussed in later sections.

### 4. Cumulative distribution functions of the climatological precipitation data

We first construct the empirical CDFs for both the GFS model background precipitation and the TMPA observations, based on their climatological samples. These model and observational CDFs will be compared, and they will also be used in defining the Gaussian precipitation transformation. For a relevant comparison useful for guiding the assimilation of precipitation, we examine the quantities that are used in the data assimilation, which depend on the design of the specific data assimilation system. We now describe how we collect the 10-yr samples of the model background precipitation and observations in correspondence with our proposed 4D-LETKF experiments.

[Figure 3](#) shows a schematic of the sample preparation. First, for the model precipitation, we would like to have the “background values,” which are usually the

short-term (e.g., 6 h) forecasts from the analyses. In our system of 4D-LETKF, forecast variables within the period from 3 to 9 h will be used as the model background (Hunt et al. 2004; Miyoshi and Yamane 2007). Therefore, we conduct a series of 9-h GFS model forecasts at desired resolutions (T62 and T126 in this study) every 6 h initialized from 10-yr (2001–10) CFSR reanalysis data, then the 3–9-h forecasts are collected to form a series of model backgrounds. The GFS model outputs forecast fields every hour in the form of the instantaneous precipitation rate; thus, we can either pick up the precipitation rates every 3 h corresponding to the TMPA observations or compute the 6-h accumulated precipitation centered at time  $t$  by

$$P(6h)_t = \frac{1}{2}Pr_{t-3} + \sum_{t'=t-2}^{t+2} Pr_{t'} + \frac{1}{2}Pr_{t+3}, \quad (8)$$

where  $Pr_t$  is the precipitation rate ( $\text{mm h}^{-1}$ ) at time  $t$ . Note that although we could directly use reanalysis precipitation as the model precipitation samples without performing the short-term forecasts, running the forecasts using the proposed data assimilation system is preferable because the existing reanalysis dataset may be produced in a way that is different from the current data assimilation system (e.g., different configurations of the forecast model), and the specific variable used in the data assimilation, such as the accumulated precipitation within the 3–9-h forecast, may not be provided in the reanalysis dataset.

For the observations, the same 10-yr (2001–10) data should be collected to form a series of equivalent observational data. The original TMPA data are provided with the 3-hourly precipitation rate at a  $0.25^\circ$  longitude–latitude resolution. After upscaling the TMPA data to the Gaussian grids used by the T62/T126 GFS model, either the instantaneous precipitation rate as in its original form, or the 6-h accumulated precipitation amount can be used to compute the statistics. The 6-h accumulated precipitation centered at time  $t$  is computed by a weighted average:

$$P(6h)_t = \frac{3}{2}Pr_{t-3} + 3Pr_t + \frac{3}{2}Pr_{t+3}. \quad (9)$$

After collecting large samples of model background and observational precipitation values, their CDFs are computed using the method described in section 3b, for each T62 grid point and each 10-day period of year (3 periods per month; 36 periods in total); that is,

$$F = F(y; \text{location, period of year}), \quad (10)$$

where  $y$  can be either model or observed 6-h accumulated precipitation in their original value, and  $F$  is the

CDF, as previously defined in Eqs. (2) and (3). The real data contain large spatial and temporal variabilities. Therefore, to create a more “continuous” CDF field smoothly varying in space and time, we include all data within 500-km radius and  $\pm 2$  periods ( $\pm 20$  days) when computing the CDF at each grid point and each period. This choice also increases the sample sizes and thus reduces the sampling errors. The grid numbers within the 500-km radius are about 20 for the T62 resolution and 80 for the T126 resolution (changing with the geographical location), so the total grid numbers used to construct the CDF for each point are roughly  $10 \text{ (yr)} \times 365 \text{ (days yr}^{-1}) \times 4 \text{ (cycles day}^{-1}) \times (5 \text{ periods } 36 \text{ periods}^{-1}) \times \{20, 80\} \cong \{4 \times 10^4, 1.6 \times 10^5\}$  for the {T62, T126} resolution, respectively.

We already presented in Fig. 2 the examples of CDFs at three arbitrary locations in different climatological regions—the extratropics, the tropics, and the marine stratocumulus region—for demonstrating how to construct the Gaussian transformation. The marine stratocumulus region shows a large discrepancy between the CDFs of the model and observational precipitation (Figs. 2i–l). To visualize the entire CDF field as a function of the geographic location, we plot the maps of precipitation amounts at various cumulative distribution levels also for the period of 11–20 January for both the TMPA data and the T62 GFS model backgrounds (Fig. 4). By comparing the fields at the same cumulative distribution levels, it is clearly found that the model has a positive bias compared to the observations since the amounts in Figs. 4b, 4d, and 4f are generally greater than those in Figs. 4a, 4c, and 4e. Positive biases are also generally seen in other seasons (not shown). In terms of geographical patterns, the CDF fields of the model and observations agree reasonably well in most regions. However, in some particular regions, they actually have a large disagreement. For example, the GFS forecast shows a local maximum in the precipitation amount at both the 30% and 60% cumulative distribution levels (Figs. 4b,d) in the Pacific Ocean west of South America (at about  $20^\circ\text{S}$ ), but this local maximum does not appear in the TMPA data (Figs. 4a,c,e). This is the region corresponding to the marine stratocumulus precipitation.

This discrepancy in these regions is most apparent in maps showing the probability of zero precipitation. As shown in Fig. 5, the most significant differences in the zero precipitation probability between the model and observations are found over the regions where the marine stratocumulus are formed over cold waters, including the subtropical eastern Pacific in both the Northern and Southern Hemispheres (west of North and South America), and west of Australia and Africa. In the TMPA data, it rarely rains in these regions, typically with 90% probability of zero precipitation or 10%

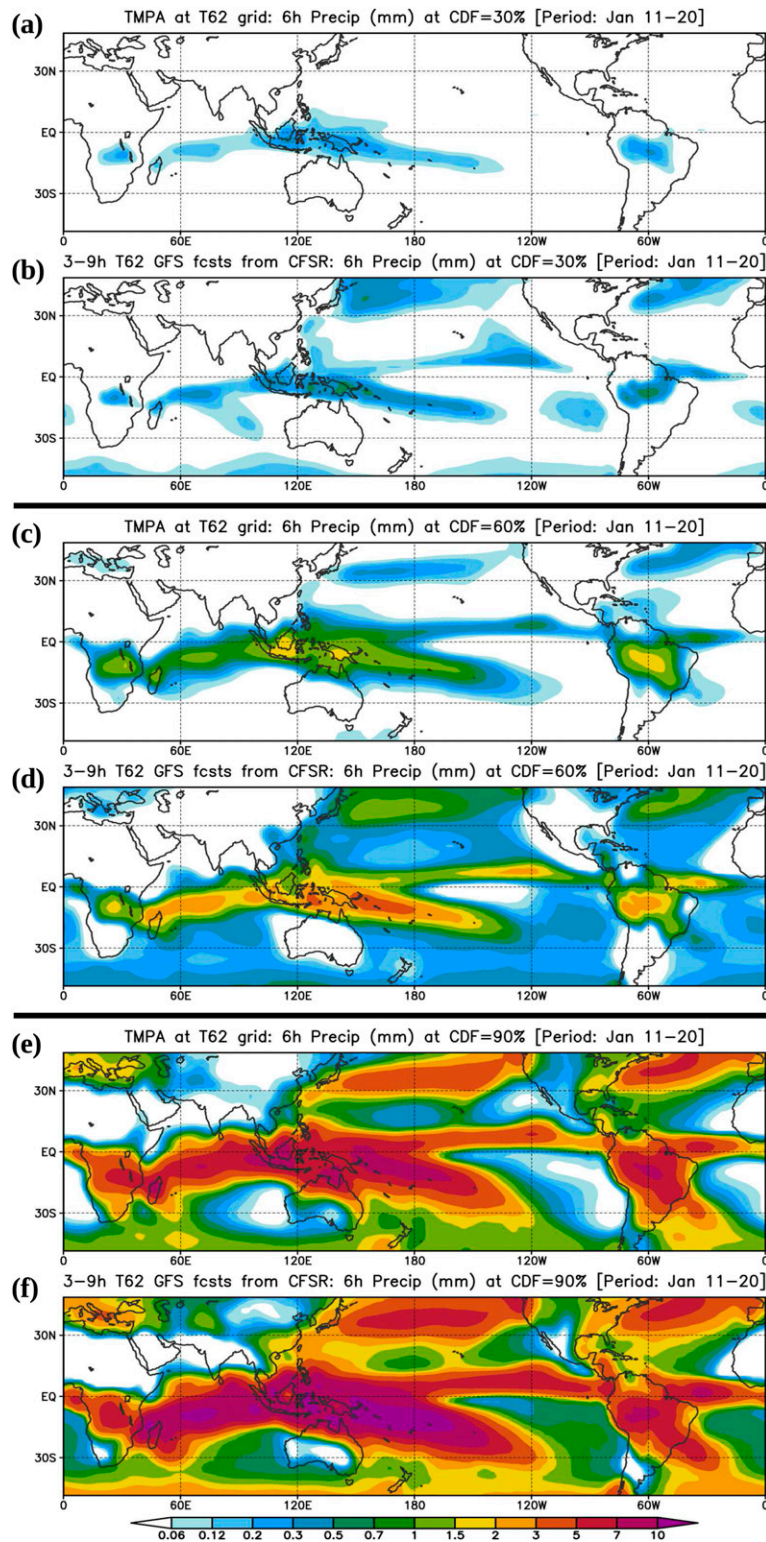


FIG. 4. Comparison of TMPA and GFS precipitation amounts (mm) for different levels of precipitation CDFs: (a),(b) 30%; (c),(d) 60%; and (e),(f) 90% cumulative distribution levels during the 11–20 Jan period for the 10-yr (2001–10) data. (a),(c),(e) TMPA data and (b),(d),(f) T62 GFS model forecasts.



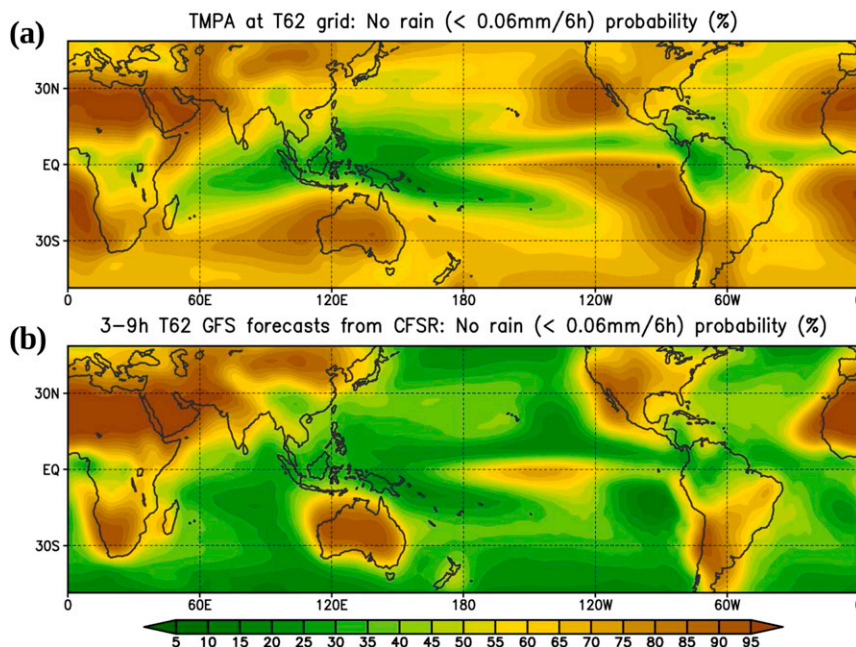


FIG. 5. The maps of (all season) zero precipitation probability (%) in (a) the TMPA data and (b) the T62 GFS model forecasts for the 10-yr (2001–10) data.

probability of nonzero precipitation (green open circle in Figs. 2k and 2l). In contrast, the model drizzle is too frequent, with typically 80% probability of nonzero precipitation (red open circle in Figs. 2k and 2l). Several studies of the marine stratocumulus (vanZanten et al. 2005; Leon et al. 2008) indicate that the real nonzero precipitation probability is not as high as the model climatology here, favoring the TMPA data. The precipitation parameterization in the low-resolution T62 GFS model may be unable to correctly simulate the low level of marine stratocumulus precipitation. However, it has also been documented that the lack of sensitivity of IR and microwave imagers to light precipitation can lead to a low precipitation occurrence bias over the ocean in the satellite precipitation estimates (Huffman et al. 2007; Behrangi et al. 2012). Therefore, these large differences could come from both high bias in the model and low bias in the TMPA data. Since in this paper we do not attempt to improve either the model or the observations, a reasonable strategy is not to assimilate the precipitation data in regions where the disagreement between the model background and the observations is large.

## 5. Joint probability distributions

In this section we use the joint probability distribution diagrams to more clearly show the relationship between the model background precipitation and the precipitation observations. All data points in the 10-yr samples are

included in the statistics. Results with different transformation methods, different variables (i.e., precipitation rate versus accumulated precipitation), and different resolutions will be shown and discussed.

### a. Original data versus logarithm transformed precipitation

Figure 6 shows the joint probability distribution diagrams between the 6-h accumulated precipitation in the T62 GFS model background and in the TMPA data upscaled to the same T62 grids. Different transformation methods are used in each subplot. Only nonzero precipitation is shown in the figures because when the zero precipitation is also plotted, it just adds two saturated lines along the  $x$  axis ( $\bar{y}$ ,  $\bar{y}_{\text{zero}}$ ) and  $y$  axis ( $\bar{y}_{\text{zero}}$ ,  $\bar{y}$ ) representing the abundance of zero precipitation in either the model background or the observation data (not shown). One would expect that the maximum probability regions should be located along the one-to-one diagonal line for a variable that is useful for data assimilation. However, when the joint probability distribution diagram is plotted without a transformation method (Fig. 6a), we barely see any correlation in precipitation between the model background and the observations.<sup>1</sup> The

<sup>1</sup> In this case, the  $R^2$  value computed from linear regression shown in the figure may not be particularly meaningful, since the correlation largely comes from the off-diagonal regions.

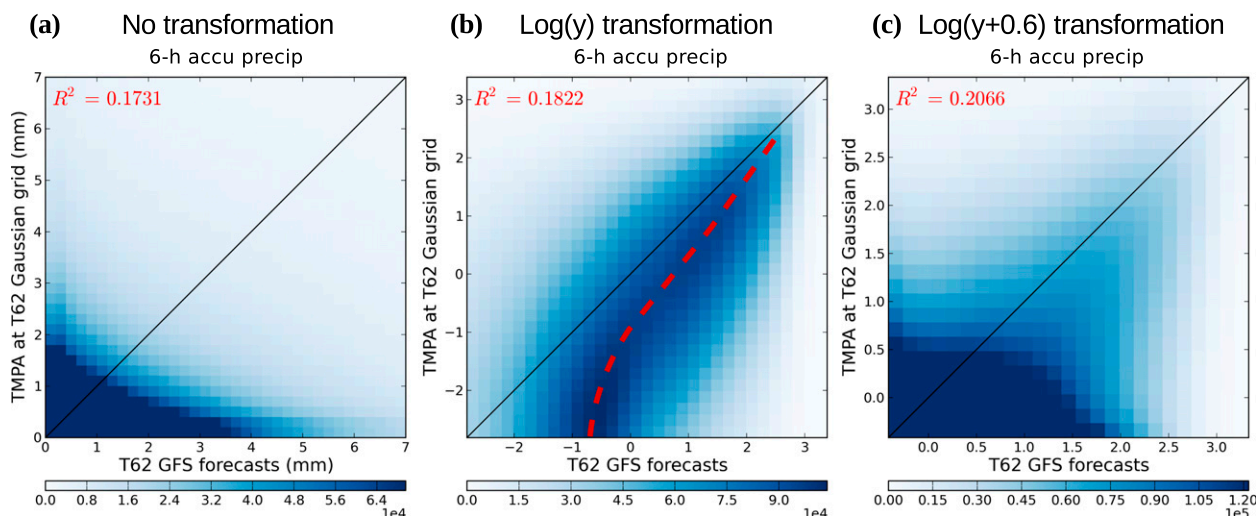


FIG. 6. Joint probability distributions of the 6-h accumulated precipitation with different transformation methods between the T62 GFS model background and the TMPA data upscaled to the same T62 grids. (a) No transformation (mm), (b) exact logarithm transformation [ $\alpha = 0$  in Eq. (1)], and (c) “modified” logarithm transformation ( $\alpha = 0.6$  mm) applied to the precipitation variables. Samples are collected for the 10-yr (2001–10) period, and only positive precipitation is shown.

probability of small precipitation amounts is saturated and not oriented along the one-to-one line. In addition, the Gaussianity of the original precipitation variable is obviously very poor (see skewness and kurtosis shown in Table 1). These statistical properties partly explain why the original precipitation is not a good variable for data assimilation and an appropriate transformation of precipitation is needed.

When we calculate the joint probability using logarithm transformed precipitation [without adding a constant in the logarithmic function;  $\alpha = 0$  in Eq. (1)] (Fig. 6b), the curved line of the maximum probability (indicated with a red dashed curve) is clearly seen. This maximum probability curve is to the right of the one-to-one line, indicating an amplitude-dependent positive bias of the model precipitation when compared to the TMPA data. In this data assimilation study, we do not consider whether the model precipitation or the TMPA data is more correct, but it is clearly better to remove this bias before data assimilation. For example, bias correction schemes have been widely used in the modern satellite radiance data assimilation (e.g., Derber and Wu 1998; Dee 2005). In addition, the skewness and kurtosis calculation indicates that the climatological precipitation distribution after the logarithm transformation becomes much more Gaussian than the original distribution (Table 1).

An interesting fact is found when the “modified” logarithm is used [i.e., a constant  $\alpha = 0.6$  mm (6 h) $^{-1}$  is added in the transformation; Eq. (1)]. In Fig. 6c, saturation in the small precipitation amounts, as in Fig. 6a, is seen again. The maximum probability curve near the

one-to-one line is still retained but it is less obvious than that in Fig. 6b. Therefore, from this joint probability distribution diagram, it is inferred that the use of a too large constant  $\alpha$  in the logarithm transformation may not be a good solution, since it makes the behavior of the transformed variable in the small precipitation amounts similar to the original variable, and thus reduces the discrimination for small amounts. A careful choice of the  $\alpha$  value is thus essential.

#### b. Precipitation rate versus accumulated precipitation

Figure 7a shows the same diagrams but for the instantaneous precipitation rate [in a unit of mm (6 h) $^{-1}$  before the logarithm transformation, and  $\alpha = 0$  in the logarithm transformation]. Comparing it with Fig. 6b, it is clear that the correlation with the precipitation rate is worse than that with the accumulated precipitation amount. In particular, a multimodal feature is seen in the model precipitation. The precipitation rate

TABLE 1. The skewness and kurtosis for the precipitation distributions in the GFS model and TMPA observations using different transformation methods. A Gaussian distribution has zero skewness and kurtosis. Samples are collected for the 10-yr (2001–10) period and within 50°N–50°S, and only positive precipitation is used.

Transformation methods	Skewness		Kurtosis	
	GFS	TMPA	GFS	TMPA
No transformation	2.32	4.67	7.50	38.5
Logarithm transformation	−0.16	0.24	−0.81	−0.87
Gaussian transformation	0.22	0.24	−0.41	−0.14

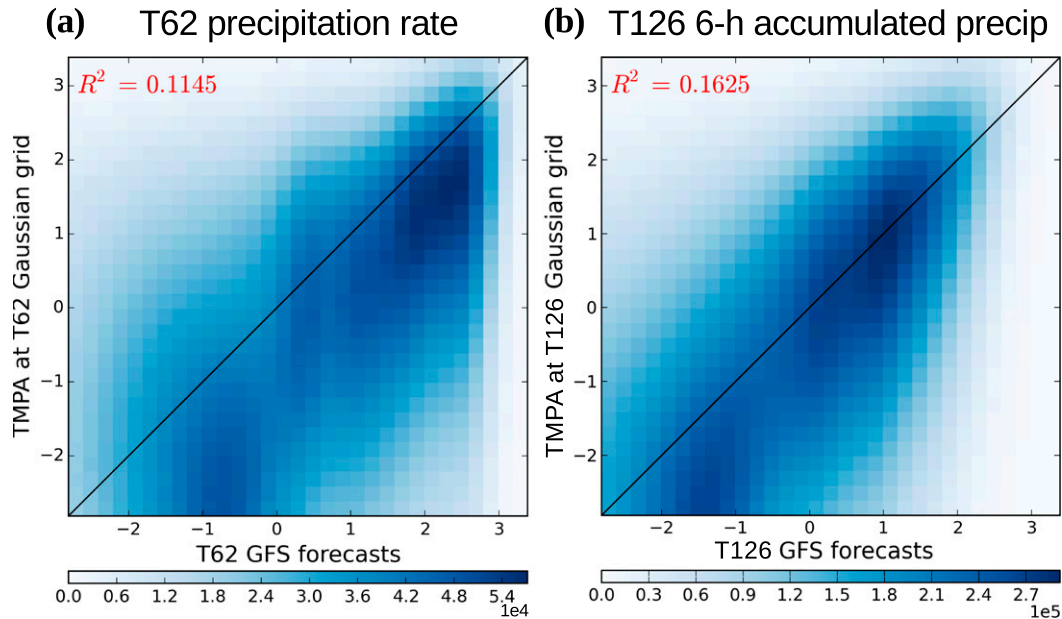


FIG. 7. As in Fig. 6b, but for the logarithm transformed (a) instantaneous precipitation rate [ $\text{mm} (6\text{ h})^{-1}$  before the transformation] at the T62 resolution and (b) 6-h accumulated precipitation (mm before the transformation) at the T126 resolution in both the GFS model background and the TMPA data.

produced from the T62 GFS model tends to be concentrated at several ranges (roughly  $[-1.2, -0.2]$ ,  $[0.2, 0.6]$ , and  $[1.6, 2.6]$  in the logarithm transformed value), which could be related to some deficiencies of the precipitation parameterization at this low resolution. The lower correlation may also be a result of the timing error of the precipitation parameterization scheme. The instantaneous precipitation rate is too sensitive to the timing error, which is common for the precipitation produced from cumulus parameterizations. For example, Chao (2013) showed that cumulus precipitation schemes can have large systematic errors in the precipitation diurnal cycle over the land. Therefore, although the accumulation of precipitation discards the information of the time variations of the precipitation within the 6-h assimilation window, the 6-h accumulated value of precipitation would be still better for the assimilation than the precipitation rate. The successful assimilation of precipitation demonstrated by Lopez (2011, 2013) also used the 6-h accumulated precipitation. Nevertheless, we note that the model resolution we use is fairly coarse, and the model precipitation could perform better in a higher-resolution model.

c. Resolution (T62 versus T126)

The same diagram of Fig. 6b but based on higher-resolution results (6-h accumulated precipitation) is shown in Fig. 7b. We carry out all the same processes

illustrated in Fig. 3 at T126 resolution. At this resolution, the bias between the model and observational precipitation is clearly smaller than that at the T62 resolution as seen in the joint probability distribution diagrams (i.e., the deviation of the maximum probability line from the one-to-one line in Fig. 7b is smaller than that in Fig. 6b); however, the correlation between the model and observations also becomes slightly lower than that at T62 (i.e., 0.1625 versus 0.1822 in  $R^2$ ). This is probably due to the larger random error in the higher-resolution model and observation data. By spatially averaging the field, this random error can be reduced (Huffman et al. 2010), which may be easier for the precipitation assimilation.

However, there is certainly some loss of information caused by upscaling the observation data to a lower resolution, and also a reduction in the accuracy of numerical models by using the low-resolution configuration. Therefore, the choice of the resolution may depend on the specific purpose of the work. In this study, we propose that, for the purpose of improving large-scale medium-range forecasts, using the spatially averaged (i.e., upscaled) TMPA data would be a reasonable choice. Indeed, we show in the companion paper (LMK16) that the assimilation of the global large-scale (lower resolution) precipitation field at the T62 resolution is able to improve the 5-day model forecasts. We do not argue that the higher-resolution model or observations are useless in precipitation assimilation, but that there is a “trade-off” between the resolution and errors.

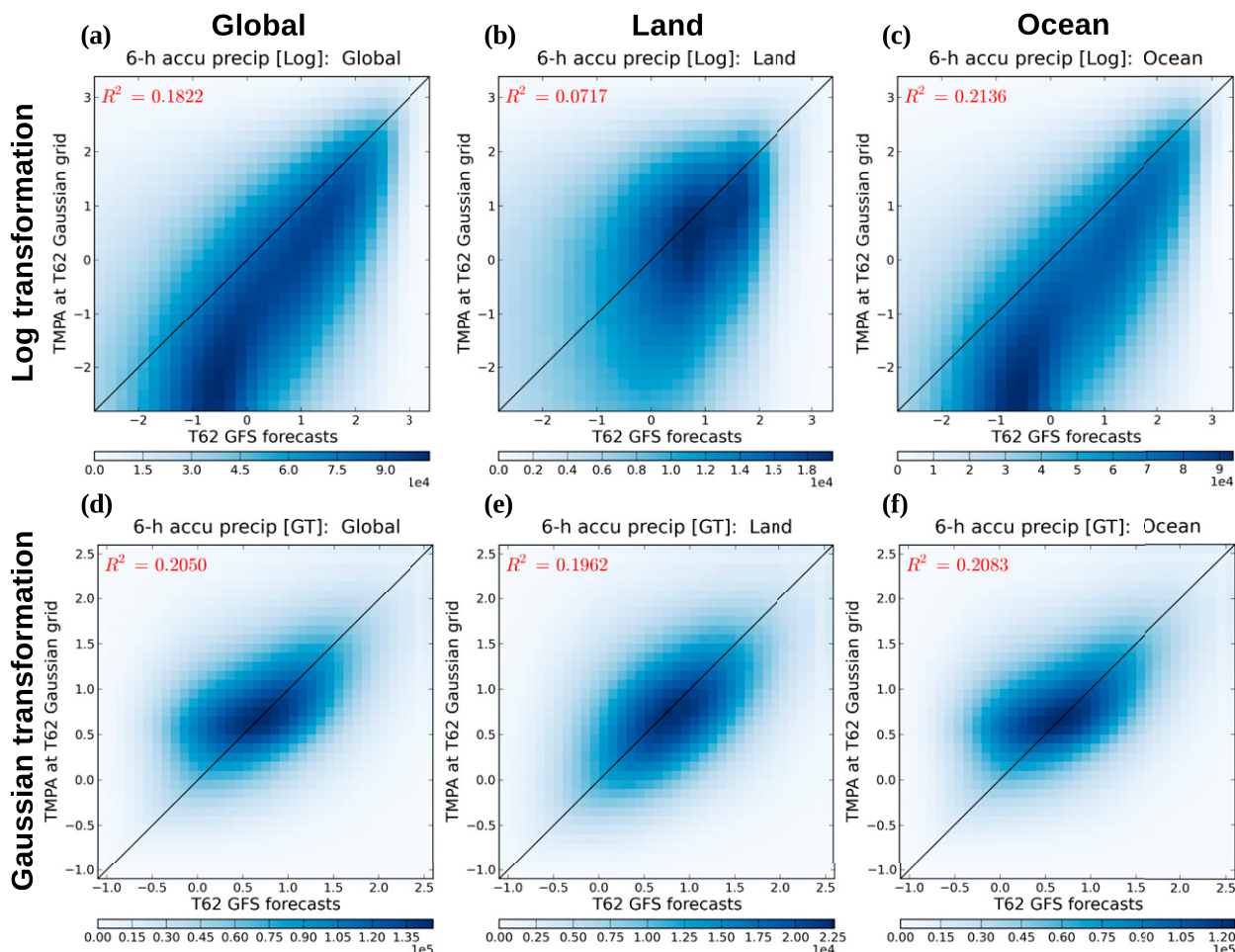


FIG. 8. The joint probability distribution of (a)–(c) the logarithm transformed ( $\alpha = 0$ ) and (d)–(f) the Gaussian transformed 6-h accumulated precipitation between the T62 GFS model background and the TMPA data upscaled to the same T62 grids. (a),(d) Global results; (b),(e) only the precipitation over the land; and (c),(f) only the precipitation over the ocean. Samples are collected for the 10-yr (2001–10) period, and only positive precipitation is shown.

Since it has been shown that model resolution leads to a large impact on the precipitation forecasts (e.g., Wen et al. 2012), assimilating higher-resolution precipitation data and solving the issues regarding the random errors would be important research. Using a higher-resolution model that has better representation of precipitation processes but still employing the spatial average in the observation operator could also be considered.

#### d. Gaussian transformed precipitation

Using the CDFs constructed in section 4, we can define the Gaussian transformations of the GFS model precipitation and the TMPA data following section 3b. Note again that the CDFs are computed for each T62 grid point and each 10-day period of year, and smoothed by including the nearby grids and times. Although this smoothing helps to construct a smooth CDF field and thus a more continuous definition of the Gaussian

transformation, the disadvantage of this method is that the transformation would not be good in regions with an intrinsically large gradient of precipitation climatology, such as regions with complex terrain and orographic precipitation.

With the Gaussian transformation, the joint probability distribution diagrams are shown in Fig. 8. Figures 8a and 8d are the global results. Figure 8a uses the logarithm transformation already shown before (Fig. 6b), and Fig. 8d is the same figure plotted with the Gaussian transformed variables. First, the figure shows that with the Gaussian transformation, the joint distribution of the precipitation variables become more normal. It is found from the skewness and kurtosis statistics (Table 1) that the skewness for the climatological precipitation distributions in the GFS model and TMPA observations using the logarithm and Gaussian transformation methods are similar, while

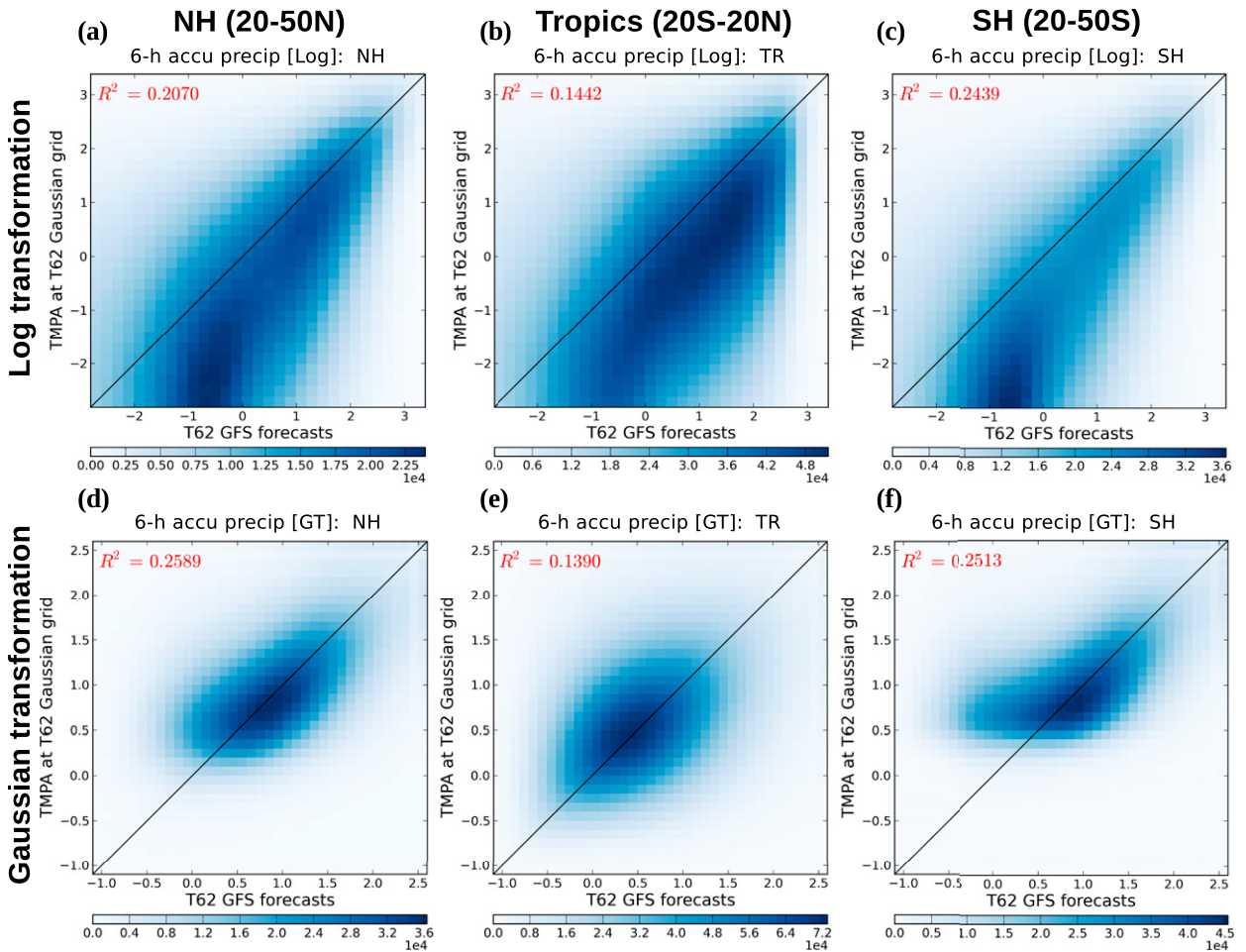


FIG. 9. As in Fig. 8, but for (a),(d) the Northern Hemisphere extratropics (20°–50°N); (b),(e) the tropical regions (20°N–20°S); and (c),(f) the Southern Hemisphere extratropics (20°–50°S).

the Gaussian transformation method particularly improves the kurtosis. However, as mentioned in section 3b, what is most important is the Gaussianity of the error distribution, not of the climatological distribution examined here. The examination of the Gaussianity of the real error distribution is shown in LMK16, and it further confirms the improving Gaussianity of the background precipitation errors in the LETKF data assimilation.

Second, it is also important to indicate that the maximum probability curve becomes more collocated with the one-to-one line (i.e., the biases are reduced), and the correlation square ( $R^2$ ) value increases slightly. In our transformation method defined for model and observations separately, the model climatology and the observation climatology are first converted to the same 0–1 scale (cumulative distribution), and then the same  $F^{G-1}$  is applied to obtain the Gaussian variables. Therefore, this method can effectively reduce the amplitude-dependent

bias as seen in Fig. 8a. We call this method a “CDF-based bias correction.”

The same diagrams are then plotted with land data only (Figs. 8b,e), ocean data only (Figs. 8c,f), the Northern Hemisphere extratropics (20°–50°N; Figs. 9a,d), the tropical regions (20°N–20°S; Figs. 9b,e), and the Southern Hemisphere extratropics (20°–50°S; Figs. 9c,f). Note that the TMPA only covers from 50°S to 50°N so the statistics are done within this extent. Overall, the improvements in the normality, centeredness, and correlations that we found in the global results are also generally found over the separate validation regions. The amplitude-dependent biases are reduced in all regions. The skewness and kurtosis calculations in every region indicate that the skewness for logarithm and Gaussian transformations are of similar magnitude, but the kurtosis of the Gaussian transformed precipitation distribution is much reduced compared to the logarithm transformation (not shown). Regarding the correlation

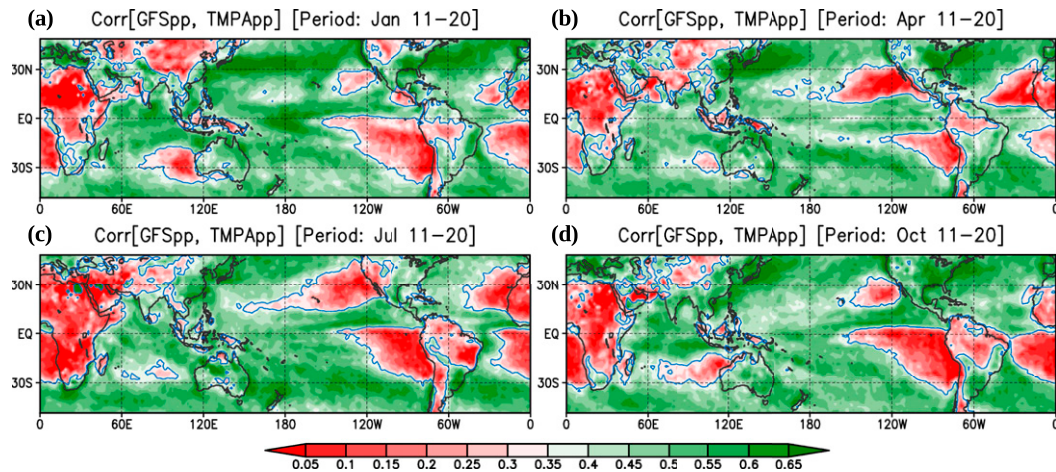


FIG. 10. The maps of correlation between precipitation in the GFS model background and in the TMPA observations during the periods of (a) 11–20 Jan, (b) 11–20 Apr, (c) 11–20 Jul, and (d) 11–20 Oct for the 10-yr (2001–10) data. The blue contours indicate correlations equal to 0.35, which is the threshold used for the precipitation assimilation in LMK16.

between model and observed precipitation, the increase of the correlation using the Gaussian transformation is particularly notable in the land region (Figs. 8b,e) and in the Northern Hemisphere extratropics (Figs. 9a,d). It slightly decreases over the ocean (Figs. 8c,f) and the tropics (Figs. 9b,e), but the change is small. In summary, we find that using separate Gaussian transformations applied to model background precipitation and observations, defined in terms of each grid point and each period of year, the statistical properties of the precipitation variable become significantly more suitable for data assimilation.

## 6. Time correlation maps

Using the same 10-yr samples of data, and the same Gaussian transformation, we also calculate the time correlations between the 6-h accumulated model and observational precipitation at each grid point and each 10-day period of year so that their geographical distributions can be displayed. Similar to the CDF calculation, when computing the correlation at each grid point, the data within  $\pm 2$  periods ( $\pm 20$  days) are considered together to obtain the temporally smoothed field. Thus this correlation score is a simple measure of the statistical “consistency” between the model and the observation climatologies. Figure 10 shows the global correlation maps in four different periods in January, April, July, and October. Overall, the dry area shows smaller correlations, which is expected because it may not be easy to capture the small or infrequent precipitation amounts by the moist physical parameterization in the model. Besides, the correlation over ocean is generally

much higher than that over land, except for the marine stratocumulus region, where the correlations are very low as shown from the discrepancy of the CDF statistics discussed in section 4. Over land, the desert areas (such as the Sahara) show persistent low correlations over the year probably because of the infrequent precipitation events and small precipitation values. The mountainous areas such as the Tibetan Plateau also show low correlations, which could be partly due to the problem of orographic precipitation in the satellite-based estimates (Shige et al. 2013). Over the United States, the eastern area has higher correlation than the western area.

These time correlation maps suggest that the precipitation data distributed over the regions with reasonable correlations can be useful in the data assimilation to improve the model analyses and forecasts, but we hypothesize that the data over regions with very low correlation would be difficult to use, because of shortcomings in the precipitation parameterization in the model. Therefore, the results suggest to set up a threshold of the correlation value below which the observations are rejected in the data assimilation process. We employed this approach, rejecting observations in regions where the climatological correlation was less than 0.35, in the precipitation assimilation experiments and obtained a small improvement over not using this criterion (LMK16).

## 7. Conclusions and suggestions for precipitation assimilation

This article is the first part of our GFS/TMPA precipitation data assimilation study. We calculate statistics with the precipitation variable in the model background

and observations from the point of view of data assimilation. To achieve meaningful statistics, the samples are carefully constructed using the same model with the same forecast period, observation variables, and resolution, as we use in the real precipitation assimilation experiments (LMK16). These statistical results can indicate how to extract more useful information from the precipitation observations.

First, the errors of precipitation in numerical models can contribute to a substantial amount of the difficulties observed in the precipitation assimilation. For example, our statistical results indicate that the GFS model at both T62 and T126 resolution generally has positive bias in precipitation as compared to the TMPA observations, and that it has a severe problem in parameterizing the marine stratocumulus precipitation. In particular, the “precipitation scale” is one of the key points of the problem. The precipitation in the model is simulated by a cumulus parameterization and/or a microphysics parameterization scheme, but the behavior of such methods depends intrinsically on different grid resolutions. In addition, precipitation usually appears in random patches, especially for convective precipitation, leading to large random errors at higher resolution. The timing of the convective precipitation is also difficult to be simulated by models. The high spatial and temporal variability further leads to large representativeness errors, which are also dependent upon resolution and important to data assimilation. In this study, we find the GFS model precipitation at T126 resolution to be less biased than that at T62 resolution, but the correlation to the observations is slightly lower, presumably due to the increasing difficulty in collocating forecasted and observed precipitation that comes with model resolution.

Performing spatial and/or temporal averages can effectively reduce these errors. Huffman et al. (2010) recommended TMPA users to create time/space averages that are appropriate to their application from the original finescale data. Bauer et al. (2011) also pointed out that using spatially/temporally smoothed precipitation data in assimilation can be beneficial. Based on similar arguments, accumulated precipitation (equivalent to a time average) is expected to be a preferable variable to be used in the data assimilation, rather than the instantaneous precipitation rate. However, this strategy may seem to contradict the continued pursuit of higher resolution, especially if we are able to afford high-resolution models and take high-resolution observations. We consider that this is a trade-off between resolution and errors. If the main goal is to improve the medium-range model forecasts, using a smoothed lower-resolution precipitation to improve the large-scale analysis can be a reasonable choice. We note that the

strategy needed for effective assimilation of convective-scale precipitation such as meteorological radar observations could be quite different from the current discussion (e.g., Fabry and Sun 2010; Yussouf et al. 2013).

The ultimate solution to overcome the above problems would be attained by the improvement of the model precipitation parameterization and the satellite precipitation estimates. Strenuous efforts have been made by the modeling (e.g., Han and Pan 2011) and remote sensing retrieval communities (e.g., Tapiador et al. 2012). However, within the scope of our data assimilation study, we do not attempt to improve the model or the observations. Our goal is to optimally use this imperfect observation dataset in this imperfect model, to improve the model forecasts of both precipitation and nonprecipitation variables, such as wind, temperature, and pressure, by using appropriate error covariances in the data assimilation. To achieve this goal, we suggest applying separate Gaussian transformations to model background and observational precipitation, which can improve the Gaussianity of the variables while also effectively removing the amplitude-dependent biases between them. This idea is an extension of the Gaussian precipitation transformation proposed for a perfect model by LKM13 in which the same transformation was applied to both model precipitation and observations.

However, since the transformation method is just an approximate way to mitigate the non-Gaussianity issue in the data assimilation, and both the transformation and the bias correction are constructed based only on the climatologies, there should be some limits of these transformation and correction approaches. Therefore, precipitation observations that are deemed to be “too bad to be assimilated” may need to be rejected. Note that the statement “an observation is bad for assimilation” is not necessarily because the observation itself is bad, but because the model is not capable of making use of this observation in that location and time. The samples of long-term model and observational precipitation data prepared in this study could be a useful reference to define appropriate quality control criteria to assimilate only the “useful” precipitation observations.

Based on the discussion above, we suggest that the problems associated with the assimilation of large-scale satellite precipitation data with the goal to improve the medium-range model forecasts can be addressed as follows:

- Non-Gaussianity of the precipitation variable: apply the Gaussian transformation to both model and observational precipitation. In LKM13, this was shown to be essential for effective assimilation of precipitation using the LETKF in perfect-model simulation experiments.

LKM13 also suggested performing the assimilation only when there are enough background members with nonzero precipitation.

- Inconsistent probability distributions of precipitation in model climatology and observation climatology: define the Gaussian transformations for the model precipitation and the observational precipitation separately based on their own CDFs so that the amplitude-dependent bias is reduced. We call this method a “CDF-based bias correction.”
- Timing errors of the precipitation: use 6-h accumulated amounts.
- Deficient precipitation parameterization: do not assimilate observations where the model is deficient. Appropriate quality control criteria (e.g., a minimum climatological correlation score between the model and the observational precipitation) can be considered to keep only the precipitation observations that the model can effectively use.
- High-resolution observations contain large random errors: perform spatial and/or temporal averages to reduce the random errors; upscale the observations to large-scale grids.

These suggestions to precipitation assimilation based on the statistical approaches were implemented and found to significantly improve the T62 5-day model forecasts, as shown in LMK16.

*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). We obtained a version of the GFS model from NOAA’s Environmental Modeling Center (EMC) with the kind help of Henry Huang and Daryl Kleist, and the model was ported to our Linux cluster with the contribution by Tetsuro Miyachi. We also gratefully acknowledge the support from the Japan Aerospace Exploration Agency (JAXA) Precipitation Measuring Mission (PMM).

#### REFERENCES

- Amezcuca, 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.
- Bauer, P., G. Ohring, 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.
- Behrangi, A., M. Lebsock, S. Wong, and B. Lambriksen, 2012: On the quantification of oceanic rainfall using spaceborne sensors. *J. Geophys. Res.*, **117**, D20105, doi:10.1029/2012JD017979.
- 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.
- Chao, W. C., 2013: Catastrophe-concept-based cumulus parameterization: Correction of systematic errors in the precipitation diurnal cycle over land in a GCM. *J. Atmos. Sci.*, **70**, 3599–3614, doi:10.1175/JAS-D-13-022.1.
- 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.
- Dee, D. P., 2005: Bias and data assimilation. *Quart. J. Roy. Meteor. Soc.*, **131**, 3323–3343, doi:10.1256/qj.05.137.
- Derber, J. C., and W.-S. Wu, 1998: The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Mon. Wea. Rev.*, **126**, 2287–2299, doi:10.1175/1520-0493(1998)126<2287:TUOTCC>2.0.CO;2.
- Errico, R. M., P. Bauer, and J.-F. Mahfouf, 2007: Issues regarding the assimilation of cloud and precipitation data. *J. Atmos. Sci.*, **64**, 3785–3798, doi:10.1175/2006JAS2044.1.
- Fabry, F., and J. Sun, 2010: For how long should what data be assimilated for the mesoscale forecasting of convection and why? Part I: On the propagation of initial condition errors and their implications for data assimilation. *Mon. Wea. Rev.*, **138**, 242–255, doi:10.1175/2009MWR2883.1.
- 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.
- Guttman, N. B., 1999: Accepting the standardized precipitation index: A calculation algorithm. *J. Amer. Water Resour. Assoc.*, **35**, 311–322, doi:10.1111/j.1752-1688.1999.tb03592.x.
- Han, J., and H.-L. Pan, 2011: Revision of convection and vertical diffusion schemes in the NCEP Global Forecast System. *Wea. Forecasting*, **26**, 520–533, doi:10.1175/WAF-D-10-05038.1.
- 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](http://mirador.gsfc.nasa.gov/collections/TRMM_3B42_007.shtml).]
- Hunt, B. R., and Coauthors, 2004: Four-dimensional ensemble Kalman filtering. *Tellus*, **56A**, 273–277, doi:10.1111/j.1600-0870.2004.00066.x.
- 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/sola.2005-013.
- Leon, D. C., Z. Wang, and D. Liu, 2008: Climatology of drizzle in marine boundary layer clouds based on 1 year of data from CloudSat and Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO). *J. Geophys. Res.*, **113**, D00A14, doi:10.1029/2008JD009835.



- 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.
- , T. Miyoshi, and E. Kalnay, 2016: Assimilation of TRMM Multisatellite Precipitation Analysis with a low-resolution NCEP Global Forecast System. *Mon. Wea. Rev.* **144**, 643–661, doi:10.1175/MWR-D-15-0149.1.
- 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.
- McKee, T. B., N. J. Doesken, and J. Kleist, 1993: The relationship of drought frequency and duration to time scales. Preprints, *Eighth Conf. on Applied Climatology*, Anaheim, CA, Amer. Meteor. Soc., 179–183.
- Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, doi:10.1175/BAMS-87-3-343.
- Miyoshi, T., 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.
- 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.
- Pan, H.-L., and W.-S. Wu, 1995: Implementing a mass flux convection parameterization package for the NMC medium-range forecast model. NMC Office Note 409, 43 pp. [Available online at <http://www.lib.ncep.noaa.gov/nceppofficenotes/files/01408A42.pdf>.]
- 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.
- Shige, S., S. Kida, H. Ashiwake, T. Kubota, and K. Aonashi, 2013: Improvement of TMI rain retrievals in mountainous areas. *J. Appl. Meteor. Climatol.*, **52**, 242–254, doi:10.1175/JAMC-D-12-074.1.
- 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.
- , and —, 2012: Gaussian anamorphosis extension of the DEnKF for combined state parameter estimation: Application to a 1D ocean ecosystem model. *J. Mar. Syst.*, **89**, 1–18, doi:10.1016/j.jmarsys.2011.07.007.
- Tapiador, F. J., and Coauthors, 2012: Global precipitation measurement: Methods, datasets and applications. *Atmos. Res.*, **104–105**, 70–97, doi:10.1016/j.atmosres.2011.10.021.
- 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.
- , 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.
- , 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.
- Ushio, T., and Coauthors, 2009: A Kalman filter approach to the Global Satellite Mapping of Precipitation (GSMaP) from combined passive microwave and infrared radiometric data. *J. Meteor. Soc. Japan*, **87A**, 137–151, doi:10.2151/jmsj.87A.137.
- vanZanten, M. C., B. Stevens, G. Vali, and D. H. Lenschow, 2005: Observations of drizzle and nocturnal marine stratocumulus. *J. Atmos. Sci.*, **62**, 88–106, doi:10.1175/JAS-3355.1.
- Wackernagel, H., 2003: *Multivariate Geostatistics*. Springer, 408 pp.
- Wen, M., S. Yang, A. Vintzileos, W. Higgins, and R. Zhang, 2012: Impacts of model resolutions and initial conditions on predictions of the Asian summer monsoon by the NCEP Climate Forecast System. *Wea. Forecasting*, **27**, 629–646, doi:10.1175/WAF-D-11-00128.1.
- Yussouf, N., E. R. Mansell, L. J. Wicker, D. M. Wheatley, and D. J. Stensrud, 2013: The ensemble Kalman filter analyses and forecasts of the 8 May 2003 Oklahoma City tornadic supercell storm using single- and double-moment microphysics schemes. *Mon. Wea. Rev.*, **141**, 3388–3412, doi:10.1175/MWR-D-12-00237.1.
- 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.
- 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. Hydrometeorol.*, **12**, 118–134, doi:10.1175/2010JHM1271.1.