Effective Assimilation of Global Precipitation: Simulation Experiments

Guo-Yuan Lien, Eugenia Kalnay, and Takemasa Miyoshi

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

Submitted to Tellus A

(October 23, 2012)

Corresponding author's address: Guo-Yuan Lien, Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD 20742-2425, USA

E-mail: Guo-Yuan Lien, gylien@atmos.umd.edu
Eugenia Kalnay, ekalnay@atmos.umd.edu
Takemasa Miyoshi, miyoshi@atmos.umd.edu

Abstract

Past attempts to assimilate precipitation by nudging or variational methods have succeeded in forcing the model precipitation to be close to the observed values. However, the model forecasts tend to lose their additional skill after few forecast hours. In this study, a local ensemble transform Kalman filter (LETKF) is used to effectively assimilate precipitation by allowing ensemble members with better precipitation to receive higher weights. In addition, two other changes in the precipitation assimilation process are proposed to solve the problems related to the non-Gaussianity of the precipitation variable: a) transform the precipitation variable into a Gaussian distribution based on its climatological distribution, and b) only assimilate precipitation at the location where at least some ensemble members have positive precipitation. Unlike most current approaches, both positive and zero rain observations are assimilated effectively.

Observing system simulation experiments (OSSEs) are conducted using the SPEEDY model, a simplified but realistic general circulation model. When the global precipitation is assimilated in addition to rawinsonde observations, both the analyses and the medium range forecasts are significantly improved as compared to only having rawinsonde observations. The improvement is much reduced when only modifying the moisture field by precipitation observations with the same approach. The effect of precipitation assimilation on the analyses is retained on the medium-range forecasts, and is larger in the Southern Hemisphere than that in the Northern Hemisphere because the NH analyses are already accurate by the denser rawinsonde stations. Both the Gaussian transformation and the new observation selection criterion are shown to be beneficial to the precipitation assimilation especially in the case of large observation errors. Assigning smaller horizontal localization length scales for precipitation observations further improves the LETKF analysis. The new approach could be used in the assimilation of other non-Gaussian observations.

Key words: ensemble Kalman filter, data assimilation, precipitation, non-Gaussianity

1. Introduction

Precipitation has long been one of the most important and useful meteorological observations. The traditional rain gauge measurement of precipitation can be traced back to 19th century before the rawinsonde network was established yet (e.g., Jones and Bradley 1992). In recent years, more advanced precipitation estimations from a variety of remote sensing platforms, such as satellite and ground-based precipitation radar, have also become available. For example, The Tropical Rainfall Measuring Mission (TRMM) has been producing a set of high-quality, high-resolution global (50S–50N) precipitation estimates (Huffman et al. 2007) which have been widely used in many research areas. The Global Precipitation Measurement (GPM; Hou et al. 2008) mission is scheduled for launch in 2014 as the successor to TRMM. Because of the large impact that effective assimilation of precipitation could have in forecasting severe weather (Bauer et al. 2011), many efforts to assimilate precipitation observations have been made.

Nudging or variational methods have been used previously to assimilate precipitation by modifying the model's moisture and sometimes temperature profiles as well, in order to either enhance or reduce short-term precipitation according to the model parameterization of rain (e.g., Tsuyuki 1996, 1997; Falkovich et al. 2000; Davolio and Buzzi 2004; Koizumi et al. 2005; Mesinger et al. 2006). They are generally successful in forcing the forecasts of precipitation to be close to the observed precipitation during the assimilation, but they revert to the regular forecasts soon after the assimilation of rain ceases. For example, a nudging method was applied to the North American Regional Reanalysis (NARR), and achieved the objective of making the Eta NARR 3 hour forecasts essentially identical to the observed precipitation used to nudge the model. As a result, in the hydrological cycle of the NARR, the model precipitation was extremely close to the observed precipitation (Mesinger et al. 2006). However, the Eta forecasts from the NARR were not superior to the operational forecasts beyond a few hours. Nudging was not effective presumably because it is not an efficient way to update the potential vorticity field, which is the "master" dynamical variable that primarily determines the evolution of the forecast in NWP models.

There also have been a number of essential issues for the precipitation assimilation in the variational framework. Precipitation processes parameterized by the model physics are usually very nonlinear and even discontinuous at some "thresholds" (Zupanski and Mesinger 1995). Therefore, it is problematic to create and use the linearized version of the forward model which is required in the 4D-Var assimilation of precipitation variables (Errico et al. 2007). An inaccurate tangent linear model and adjoint model would yield a poor estimate of the evolution of finite perturbations and degrade the 4D-Var analyses. Sometimes an alternative moist physical parameterization scheme that is more linear and continuous has been used to reduce the nonlinearity problem (e.g., Zupanski and Mesinger 1995; Lopez and Moreau 2005). In addition, the highly non-Gaussian distribution of the precipitation observations seriously violates the basic assumption of normal error statistics made in most data assimilation schemes. The flowindependent background error covariance that is usually used in variational methods cannot describe the relation between precipitation and other state variables. All of the above problems have contributed to the difficulties of the precipitation assimilation, leading to a widely shared experience that forecasts starting from analyses with precipitation assimilation lose their extra skill after just a few forecast hours (e.g., Tsuyuki and Miyoshi 2007; Davolio and Buzzi 2004; Errico et al. 2007). One notable exception is Hou et al. (2004) who used forecast tendency corrections of temperature and moisture as control variables in variational data assimilation in the assimilation of hurricane observed precipitation. They were able to show that large changes in precipitation had long-lasting positive impacts on a hurricane forecast, presumably because the release of latent heat corrected the potential vorticity.

Bauer et al. (2011) recently reviewed the current status of precipitation assimilation and concluded that there are still major difficulties related to (1) the moist physical processes in NWP models and their linear representation and (2) the non-Gaussianity of both precipitation observations and model perturbations. Here we use the EnKF framework to address these critical issues. First, the EnKF method does not require linearization of the model, and it should be able to more efficiently change the potential vorticity field by allowing ensemble members with better precipitation (due to presumably better dynamics) to receive higher weights. Second, a general variable transformation is introduced to solve the problem that precipitation is highly non-Gaussian. Recognizing this non-Gaussianity, transformations such as a logarithmic transformation have been previously applied to the precipitation assimilation (e.g., Hou et al. 2004; Lopez 2011). The logarithmic transformation can alleviate the non-Gaussianity of positive precipitation, whereas our proposed algorithm can transform any continuously distributed variable into a Gaussian distribution. In addition, we also address the issue of zero precipitation. Zero precipitation observations can be successfully assimilated by using a criterion that requires that at least several background ensemble members have positive precipitation in order to assimilate the precipitation observation. We also note that within the variational methods, considerable efforts have focused on the more accurate microphysical parameterizations (e.g, Treadon et al. 2003; Li and Mecikalski 2010, 2012). In the ensemble framework, the precipitation determined by the ensemble model variables can be used in the assimilation without having to account for details in the physical processes. Several pioneering experiments of precipitation assimilation using ensemble data assimilation methods have been conducted, in which promising results have been obtained (Miyoshi and Aranami 2006; Zupanski et al. 2011; Zhang et al. 2012).

In this paper we carry out observing system simulation experiments (OSSEs) using the same system that we had previously tested unsuccessfully before introducing the Gaussian Transform of precipitation and the new criterion for precipitation assimilation. The paper is organized as follows. The methodology including the Gaussian transformation and special treatment of zero precipitation observations are introduced in section 2. Section 3 describes the model, the local ensemble transform Kalman filter (LETKF) used in this study, and the detailed settings of the OSSEs. Section 4 shows the results of the precipitation assimilation. Concluding remarks and further issues are in section 5.

2. Proposed methodology for an effective assimilation of precipitation

(a) Gaussian transformation

In order to satisfy the basic assumption of Gaussian distribution and error statistics in data assimilation, we seek a general transformation algorithm to transform any variable y with a known arbitrary distribution into a Gaussian variable y_{trans}. It can be achieved through the connection between the two cumulative distribution functions (CDFs) of y and y_{trans}:

$$y_{\mathrm{trans}} = G^{-1}[F(y)] \, , \tag{1}$$

where $F(y)$ stands for the CDF of y (by definition having values from 0 to 1), and G^{-1} is the inverse CDF of a normal distribution with zero mean and unit standard deviation such as y_{trans} is designed to be. Here,

$$G^{-1}(x) = \sqrt{2}\, \mathrm{erf}^{-1}(2x - 1) \, , \tag{2}$$

where erf^{-1} is the inverse error function. The CDF of y can be determined empirically. In this study, we first run the SPEEDY model for 10 years and in order to compute the CDF of precipitation variables (previous 6-hour accumulated precipitation) at each grid point and at each season based on this 10-year model climatology. Accordingly, transformations of both observation and model precipitation variables are thus made in terms of their spatial location and season during the assimilation process. This technique is sometimes called "Gaussian anamorphosis" and has been also used by Schöniger et al. (2012) in hydrology, providing a more comprehensive theoretical explanation. Note that this method transforms the climatological distribution of the original variable into a Gaussian distribution as a whole, but not its error distribution at every measurement and model background. Nevertheless, we assume that the error distribution in a more Gaussian variable would also have more Gaussian error statistics, and test whether this method is really beneficial in the experiment results.

The transformation ensures a simple one-to-one relationship between the original variable and the transformed variable if their CDFs are *continuous*. Figure 1 illustrates how the transformation works for the precipitation distribution at a grid point near Maryland in the winter season. The probability density function (PDF) and CDF [i.e., $F(y)$] of the original precipitation variable are shown in Figures 1a and 1c, respectively. Using the inverse CDF of normal distribution G^{-1}, the $F(y)$ is converted back to the transformed variable y_{trans}, with the CDF shown in Figure 1d and the PDF in Figure 1b. It is apparent that the precipitation is not a continuous variable since it contains a large portion of zero values[1] so that the CDF is discontinuous at zero. The dashed parts of lines in Figures 1b, 1c, and 1d are associated with

[1] In this study, we define 6-hour accumulated precipitation less than 0.1 mm as "zero precipitation."

those zero precipitation values. This issue will be addressed and the figure will be further discussed in the next subsection.

In addition, G^{-1} will transform zero and one to $-\infty$ and $+\infty$ respectively, suggesting that the outliers of precipitation values will cause problems. In order to avoid problems transforming those outliers, we set all precipitation values with cumulative distribution less than 0.001 and greater than 0.999 to the values 0.001 and 0.999, respectively. This only affects very few values that are close to the tails of the model climatological distribution or that may even fall outside the distribution. Consequently, they will be transformed into -3.09 and 3.09.

(b) Handling zero precipitation

As mentioned in the last subsection, precipitation variables contain a large portion of zero values, which is manifested as a delta function in the PDF (Figure 1a). Since any deterministic transformation of a delta function will still result in a delta function, it would be impossible to obtain a transformed precipitation variable with perfect normal distribution if all precipitation values are considered. A naïve approach would be to only transform the non-zero part of precipitation data. However, this is not practical in data assimilation because even if all zero precipitation observations are discarded, in the background forecast there may be zero values at the corresponding observation location which still need to be transformed (via the observation operator) before they are passed into the assimilation calculation. In ensemble data assimilation framework, this problem is even more apparent than in variational data assimilation since it is very likely that a random ensemble member would have zero precipitation at an observation location. Therefore, a heuristic solution to the transform of zero precipitation values is necessary[2].

In our proposed algorithm, the CDF $F(y)$ is discontinuous at $y = 0$, thus the problem with zero precipitation in this algorithm is equivalent to assigning a value of cumulative probability F for zero precipitation ($y = 0$). In the absence of a better solution, a reasonable choice is to assign the middle value of zero-precipitation cumulative probability to $F(0)$. In the example shown in Figures 1c and 1d, the probability of zero precipitation is about 63.4% (CDF = 0.634 for minimum positive precipitation; open circles), thus $F(0)$ = 0.317 is assigned for all zero precipitation (solid circles) at that grid point. In this way, the zero precipitation in the transformed variable is still a delta function in its PDF (Figure 1b), but it is located at the *median* of the zero precipitation part of the normal distribution). Therefore, though not perfectly Gaussian, it is more reasonable than the original skewed distribution[3]. We tested other more sophisticated approaches, including one that assigned uniformly distributed random values to fill

[2] In the traditional logarithmic transformation, an arbitrary constant is usually added to the original precipitation value before the transformation [e.g., $y_{\text{trans}} = \log(y + 1)$] in order to avoid the singularity at zero precipitation.

[3] This approach to transforming zero precipitation does not maintain the properties of zero mean and unit standard deviation. However, this does not create problem in the data assimilation because such properties are essentially not required in the climatological distribution.

up the zero-precipitation cumulative probability so that a perfect Gaussian variable could be generated, but their experimental impact in the assimilation experiments were no better than that of the simple median approach.

We note that traditional precipitation assimilation systems in the variational framework usually discard the zero precipitation observations (e.g., Koizumi et al. 2005) because those observations are difficult to use. Nevertheless, zero precipitation observations should contain valuable (and accurate) information about the atmospheric state. With our current transformation algorithm handling the zero precipitation and an ensemble data assimilation system, zero precipitation observations are, indeed, assimilated. Instead of discarding zero observations, a different criterion is used in this study: assimilation is conducted at all grid points where at least some members of prior ensemble are precipitating (regardless of the observed values). This is because if the ensemble spread is zero (i.e., all forecasts have zero precipitation), it is not possible to assimilate precipitation using an EnKF. In section 4.a we will show that assimilating precipitation observations at locations with only a few precipitating members does not show improvements so that the criterion we have chosen is to require that at least half of the forecasts have positive precipitation, which controls the assimilation quality and saves computational time.

3. Experimental design

(a) The SPEEDY-LETKF system

The Simplified Parametrizations, primitivE-Equation DYnamics (SPEEDY) model (Molteni 2003) is a simple, computationally efficient, but realistic general circulation model widely used for data assimilation experiments. The version of SPEEDY model used in this study is run at a T30 resolution with 7 vertical sigma levels. It has five state variables: the zonal (U) and meridional (V) components of winds, temperature (T), specific humidity (Q), and surface pressure (Ps). The previous 6-hour accumulated precipitation (PP) is also outputted by the model.

The LETKF (Hunt et al. 2007) is an ensemble Kalman filter scheme that performs most of the analysis computations in ensemble space and in each local domain. As all other ensemble data assimilation schemes, the flow-dependent background error covariance $\mathbf{P}^b$ is inferred from the sample covariance among ensemble members. The background error covariance can be written as

$$\mathbf{P}^b = \frac{1}{K-1}\,\mathbf{X}^b(\mathbf{X}^b)^T\,, \tag{3}$$

where $\mathbf{X}^b = \left[\mathbf{x}^{b(1)} - \bar{\mathbf{x}}^b, \ldots, \mathbf{x}^{b(K)} - \bar{\mathbf{x}}^b\right]$ is the matrix whose columns are background ensemble perturbations (i.e., the departure of members from the ensemble mean) of state variables, and K is the ensemble size. The dimension of $\mathbf{P}^b$ is exceedingly large in modern NWP models, thus it is not computed explicitly. Instead, when performing the LETKF analysis, $\widetilde{\mathbf{P}}^a$, the analysis covariance *in ensemble space* is computed first (Hunt et al. 2007):

$$\widetilde{\mathbf{P}}^a = [(K-1)\,\mathbf{I} + (\mathbf{Y}^b)^T\mathbf{R}^{-1}\,\mathbf{Y}^b]^{-1}\,. \tag{4}$$

After that, the mean weight vector $\bar{\mathbf{w}}^a$ and the weight matrix for the ensemble perturbation $\mathbf{W}^a$ are computed from:

$$\bar{\mathbf{w}}^a = \widetilde{\mathbf{P}}^a\,(\mathbf{Y}^b)^T\mathbf{R}^{-1}(\mathbf{y}^o - \bar{\mathbf{y}}^b)\ , \tag{5}$$

$$\mathbf{W}^a = \left[(K-1)\,\widetilde{\mathbf{P}}^a\right]^{1/2}\ , \tag{6}$$

where $\mathbf{Y}^b = \left[\mathbf{y}^{b(1)} - \bar{\mathbf{y}}^b, \dots, \mathbf{y}^{b(K)} - \bar{\mathbf{y}}^b\right]$ is the matrix that consists of columns of background observation perturbations, $\mathbf{R}$ is the observation error covariance, and $\mathbf{y}^o$ is the observation. The background (forecasted) observation values are calculated through the observation operator: $\mathbf{y}^{b(i)} = H\left(\mathbf{x}^{b(i)}\right)$. Finally, the analysis ensemble mean and perturbations can be computed by applying the weights to the background ensemble:

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{X}^b\,\bar{\mathbf{w}}^a\ , \tag{7}$$

$$\mathbf{X}^a = \mathbf{X}^b\,\mathbf{W}^a\ . \tag{8}$$

In the LETKF, Eqs. (4)-(8) are computed locally for every model grid point with its nearby observations, which allows easy implementation of covariance localization and parallelization (Hunt et al. 2007). A computationally efficient code for the LETKF is available at the public Google Code platform by Miyoshi (http://code.google.com/p/miyoshi/), including the SPEEDY-LETKF system that couples the SPEEDY model with the LETKF codes.

When applying the Gaussian transformation, the precipitation observations in $\mathbf{y}^o$ are replaced by the transformed observations, and the transformation algorithm is also included in the observation operator *H* to get the transformed precipitation values from the background. In addition, the observation errors associated with each observation also have to be transformed. Conceptually,

$$\epsilon^o_{\text{trans}} \cong (y^o + \epsilon^o)_{\text{trans}} - y^o_{\text{trans}} \cong y^o_{\text{trans}} - (y^o - \epsilon^o)_{\text{trans}}\ , \tag{9}$$

where ϵ^o is the observation error and $\epsilon^o_{\text{trans}}$ is the transformed observation error whose squares appear in the diagonal elements of $\mathbf{R}$. This means that the observation error is rescaled based on the differences between the transformed observation value and its adjacent values (i.e., plus/minus one observation error). In this study, we calculate both $\epsilon^{o+}_{\text{trans}} = (y^o + \epsilon^o)_{\text{trans}}\text{-}y^o_{\text{trans}}$ and $\epsilon^{o-}_{\text{trans}} = y^o_{\text{trans}}\text{-}\left(y^o\text{-}\epsilon^o\right)_{\text{trans}}$, requiring them be at least 0.1 (unitless in the transformed variable), and then regarding their average as the transformed observation error; namely,

$$\sigma^o_{\text{trans}} = [\max(\sigma^{o+}_{\text{trans}},\ 0.1) + \max(\sigma^{o-}_{\text{trans}},\ 0.1)]/2\ . \tag{10}$$

(b) The observing system simulation experiment

The SPEEDY model is first run for a one year spin-up, arbitrarily denoted year 1981, and then for 10 years, from January 1, 1982 to January 1, 1992 forced by the climatological sea surface temperature. These 10 years of simulation are used to compute the precipitation CDF at each grid point and at each season in preparation for the Gaussian transformation as introduced in the previous section. The same run in the period from January 1, 1982 to January 1, 1983 is also regarded as the nature run, or the "truth" in the OSSEs. Simulated observations are taken from this nature run by adding random noise corresponding to the designated observation errors. The basic observing system used in this study is just conventional rawinsonde observations that are assimilated in the control run ("Raobs" hereafter). The rawinsonde locations are distributed realistically as shown by open circles in Figure 2. Variables assimilated include u, v winds, temperature, specific humidity, and surface pressure, whose observation errors are listed in Table 1. Additional precipitation observations are assimilated in other experiments to estimate the impact of the precipitation assimilation. The 6-hour accumulated precipitation data are gathered uniformly every 2 by 2 (4) model grids points over the globe simulating satellite retrievals (indicated with plus signs in Figure 2). The observation errors of precipitation observations are set to be either 20% or 50% of the observed values for the non-zero precipitation and no error when zero precipitation is observed in the nature run. Covariance localization is computed adjusting the observation errors by their distance (the "R localization" in Greybush et al. 2011), with a horizontal length scale L = 500 km and a vertical length scale of 0.1 in natural logarithm of pressure for all observations with two exceptions: (1) No vertical localization is applied for precipitation observations because of the expected correlation between precipitation and model variables in deep layers. (2) Reduced horizontal localization lengths for precipitation observations are used in some experiments ("PP_GT_10mR_0.5L" and "PP_GT_10mR_0.3L") in order to discuss the sensitivity of localization. In addition, the adaptive inflation scheme of Miyoshi (2011) is used.

Twenty ensemble members are used in our assimilation experiments. Starting from January 1, 1982, all experiments are initialized with the same initial ensemble created by a random choice of model conditions at unrelated time in the nature run, so they are very different from the "truth." Observation data are then assimilated into the model with a 6-hour cycle. All experiments are run for 1 year until January 1, 1983. The differences among experiments are summarized in Table 2. First, in "Raobs", only the rawinsonde observations are assimilated. We denote the main experiment showing the effectiveness of precipitation assimilation as "PP_GT_10mR", indicating that precipitation (PP) is assimilated, that the Gaussian Transformation (GT) is performed, and that the criterion requiring at least 10 members of the ensemble (half of the total ensemble size) to rain in order to use a precipitation observation (10mR) is being applied. All prognostic variables in the SPEEDY model are updated during the assimilation as in the standard formula of LETKF. The observation error of precipitation observations in this experiment is rather accurate, 20%, and the localization length of precipitation observations is the same as rawinsonde observations (i.e., L = 500 km). In "PP_GT_10mR_Qonly", only the specific humidity Q is updated during the LETKF assimilation of precipitation observations, trying to resemble what conventional "nudging" methods do by arbitrarily modifying the moisture field in the model. For other experiments, "PP_noGT_10mR" does not use Gaussian transformation; "PP_GT_ObsR" uses the traditional criterion that precipitation is only assimilated when at least a trace of rain is observed (ObsR > 0.1 mm $6h^{-1}$). In addition, "PP_GT_10mR_50%err" and "PP_noGT_10mR_50%err" are conducted to test the impact of lower observation accuracy on the precipitation assimilation, with much higher precipitation observation errors of 50% rather than 20% are used. The minimum required number of precipitating members is also varied ("PP_GT_1mR", "PP_GT_5mR", "PP_GT_15mR" and compared with "PP_GT_10mR"). Finally, different localization length scales ("PP_GT_10mR_0.5L", "PP_GT_10mR_0.3L") are also used in several experiments to test the sensitivity of assimilation criteria and the localization lengths of precipitation observations. Furthermore, for some experiments ("Raobs", "PP_GT_10mR", and "PP_GT_10mR_Qonly"), we also conduct 5-day free forecasts based on each 6-hourly ensemble mean analysis over the year in order to test whether the assimilation of precipitation is "remembered" during the forecast.

4. Results

(a) Effect of global precipitation assimilation

Figure 3a shows the evolution of the global root-mean-square (RMS) analysis errors (verified against the nature run) of the u-winds over one year. We only show this variable because the impacts are remarkably similar for all the model variables indicating that the assimilation of precipitation approach is indeed able to influence the full dynamical evolution of the model and not just the moist thermodynamics. Different time scales are used to show the spin-up stage in the first month and for the remaining 11 months after the spin-up. The averaged values of RMS analysis errors in the last 11 months are also listed in Table 3. Note that the spin-up takes about one month because the ensemble initial states were chosen to be very different from the nature run at the initial time. In the LETKF (or any EnKF) a long spin-up is required in order to estimate not only the truth (with the ensemble mean), but also the "errors of the day" with the ensemble perturbations. This spin-up period can be substantially reduced by applying Running in Place (RIP) or Quasi-Outer Loop (QOL), where during spin-up the observations are used more than once in order to extract more information from them (Yang et al. 2012; Kalnay and Yang 2010).

It is clear that when all variables (and therefore the full potential vorticity) are modified (PP_GT_10mR; blue line in Figure 3a), the improvement introduced by precipitation assimilation is quite large (27.2% reduction in the averaged global analysis error) after the first month of spin-up. Not only is the long-term averaged RMS error reduced, but the temporal variation of analysis accuracy is also reduced (e.g., the error jump observed in the Raobs experiment during July is not seen in PP_GT_10mR). This result is very encouraging because it clearly shows that assimilating precipitation does bring significant benefits to the LETKF analysis. In contrast, when only the moisture field is modified (PP_GT_10mR_Qonly; orange line in Figure 3a), the improvement is much smaller (13.6% reduction in the averaged global analysis error after the spin-up), even though this approach did benefit from the Gaussian transform of precipitation.

In addition to the LETKF analysis, the impact of precipitation assimilation on model forecasts is also shown on Figure 3b. The global RMS forecast errors of u-wind are computed with respect to the forecast time and averaged over the last 11 months (i.e., after the spin-up). It is evident that the improvements last throughout the 5-day forecasts, so that the effect of precipitation assimilation is not "forgotten" by the model during the forecast, as experienced with nudging. It is interesting that the improvement by LETKF modifying only moisture (PP_GT_10mR_Qonly) also lasts throughout the forecast, which seems more effective than nudging presumably because of the Gaussian transform and the use of an EnKF. However, the improvement in PP_GT_10mR_Qonly is much smaller than that in PP_GT_10mR, and its error growth rate (i.e., the slope) is close to that in Raobs whereas the error growth rate in PP_GT_10mR is smaller than for the other two experiments. As indicated before, similar improvements in the analysis and 5-day forecast errors are also observed in all other model variables, including the very important precipitation forecasts. Figure 4 shows that the precipitation forecasts are improved as well by assimilating the precipitation observations. Starting from 12 forecast hours, the error growth rates become stable, and the forecast improvement on precipitation in PP_GT_10mR is larger than 2 days.

The effects of Gaussian transformation (GT) and the criterion requiring at least 10 members to rain in order to use an observation (10mR) are examined assuming accurate precipitation (20% errors) by comparing the results of PP_GT_10mR, PP_noGT_10mR, and PP_GT_ObsR (Figure 5). As shown in the figure, during the spin-up stage the LETKF analysis without transforming the precipitation variable (PP_noGT_10mR; red line in Figure 5) is worse than that applying Gaussian transformation. However, with very accurate observations, the Gaussian transformation does not make a significant difference after the spin-up period (Table 4; 26.1% vs. 27.2% reduction in averaged global analysis errors). It is possible that the proposed Gaussian transformation is especially useful to the LETKF assimilation when the model background is less accurate and the difference between model background and the precipitation observations is large. Therefore, when the analysis is accurate enough after the first month of spin-up, the Gaussian transformation does not offer additional advantages. This point will be tested with experiments with less accurate precipitation observations, discussed in a later subsection.

In addition, Figure 5 also shows the impact of the criteria for assimilation of precipitation. We compare the results with the traditional criterion of assimilating only positive rain observations (PP_GT_ObsR) and our newly proposed criterion of requiring at least half of the members to rain, but allowing the assimilation of zero precipitation (PP_GT_10mR). The 10mR criterion seems to be essential in order to have an effective precipitation assimilation. The analysis of PP_GT_ObsR (green line in Figure 5) is obviously degraded from PP_GT_10mR (Table 5; only 0.3% reduction in the averaged global analysis error). In particular, the degradation comes mainly from the tropical region (30S ~ 30N; Table 5; 18.7% increase in the averaged analysis error), which indicates that this observation-based criterion is not useful. Additional experiments with different minimum numbers (1, 5, and 15 out of 20) of the precipitating member in order to pass the assimilation were also conducted. As shown in Table 5, it is interesting that with a criterion that is too lenient (requiring only 1 or 5 precipitating members), the improvement by precipitation assimilation is also degraded. This indicates that assimilating precipitation observations at locations where precipitating members are rare can hurt the analysis. If stricter criteria (10mR or even 15mR) are used as we do in most experiments in this study, the results are better. Note that this type of criteria also automatically allows zero precipitation observations to be assimilated (provided that there are enough precipitating members at the observation location). In addition, requiring more precipitating members in the background also reduces the number of observations assimilated and thus reduces computational time.

(b) Regional dependence

After observing the reduction of globally averaged analysis and forecast errors, the regional dependence of the impact of precipitation assimilation is considered. The RMS errors are computed for three regions: the Northern Hemisphere extratropics (30 – 90N; NH), the tropics (30S – 30N; TR), and the Southern Hemisphere extratropics (30 – 90S; SH). Figure 6 shows the RMS errors of u-wind in 0 – 5 day forecasts averaged over the last 11 months for main experiments as Figure 3b, but for each region. Averaged RMS analysis errors in the same 11 months in terms of separate regions are also listed in Table 3-6 for all experiments.

It is clear that, as in operational forecasts, these three regions have distinct characteristics of analysis errors, error growth rate, and the impact of precipitation assimilation. With only rawinsonde observations (Raobs), the analysis (0 hour) in the NH region is already very accurate, while the TR analysis is less accurate and the SH analysis is the least accurate. As a result, the precipitation assimilation only has a small effect on the NH region (20.6% reduction in PP_GT_10mR) but a large effect on the SH region (55.2% reduction in PP_GT_10mR). The effect on the TR region is even smaller (11.2% reduction in PP_GT_10mR), which would be explained by different dynamical instabilities and precipitation mechanisms between the tropical and extratropical regions. The prevailing convective precipitation in the tropics tends to maintain small-scale features and thus would be more difficult to capture in this low-resolution global model and by low-resolution observations. During the 5-day forecasts, the RMS errors in both NH and SH regions grow with similar rate, faster than that in the TR region, as observed in operational forecasts, due to the stronger growth rates of mid-latitude baroclinic instabilities. The RMS errors in the NH region are then close to those in the TR region at the end of the 5-day forecasts. The improvement by precipitation assimilation in the SH region is so large that the RMS analysis and most forecast errors in the SH region in PP_GT_10mR is even better than those in the TR region although without precipitation assimilation it is very inaccurate. The difference between modifying all variables and only modifying moisture by LETKF is also emphasized in the SH region during the later forecasts. It is also noted that in spite of different dynamical nature of error growth in these three regions, precipitation assimilation does lead to positive impacts in all regions.

Global maps of (temporally averaged) RMS errors and error reduction of the mid-level vorticity ($\sigma = 0.51$) for the 72-hour forecasts during the last 11 months are shown in Figure 7. As expected, the error in Raobs (contours) is large in the Southern Hemisphere since the conventional rawinsonde network is quite sparse in that region. The Southern Ocean near the southern end of South America has the largest error in the world presumably because it is the least observed. By contrast, the Raobs forecast error is generally small in the Northern Hemisphere, especially over the Euro-Asian continent with the densest rawinsonde observations. By including the precipitation observations in LETKF assimilation, the error reduction (i.e., the RMS error of PP_GT_10mR – the RMS error of Raobs; shade) is large in the SH extratropical region, smaller in the NH extratropical region, and smallest in the tropical region. Once again, the dynamical impact of assimilation of precipitation on the evolution is shown by the fact that the largest error reduction is almost collocated with the regions with the largest error in Raobs, where the room for improvement is large, and yet the error is still reduced even in rawinsonde-rich Northern Hemisphere. The tropical region, instead, shows the smallest improvement, and the eastern equatorial Pacific and the central Africa are the only two areas that show slightly negative impacts. Precipitation assimilation in EnKF has a profound impact on vorticity through the dynamical impact of giving higher weights to the ensemble members that show better precipitation, independent of the details of the physical processes involved in condensation and precipitation.

(c) Sensitivity to accuracy of the precipitation observations

As mentioned in subsection 4.a, with accurate precipitation observations of 20%, the application of the Gaussian transformation to the precipitation variable has only a minor impact on the LETKF analysis accuracy after the spin-up (Figure 5). However, this is not the case with the probably more realistic precipitation observation errors of 50%. Figure 8 and Table 4 shows the impact of both larger observation errors as well as the use of the Gaussian transformation. When the observation error of precipitation observations are increased from 20% to 50%[4], and the Gaussian transformation is used (PP_GT_10mR_50%err vs. PP_GT_10mR which uses 20%err), the analysis becomes only slightly worse (shown as a green line in Figure 8). However, without the Gaussian transformation and with 50% errors (PP_noGT_10mR_50%err; red line in Figure 8), the precipitation assimilation fails. The LETKF analysis in PP_noGT_10mR_50%err is worse than not assimilating precipitation in the globe and all separate regions (Table 4). In other words, without the Gaussian transformation the precipitation assimilation hurts the analysis, whereas PP_GT_10mR_50%err with the Gaussian transformation is almost as good as that with

[4] The 50%-error precipitation observations are generated independently from the nature run, and the observation errors are also set to 50% during the LETKF assimilation.

the much smaller 20% errors. This sensitivity test demonstrates the importance of the Gaussian transformation. Less accurate observations will tend to have larger differences from the model background and may not be able to make the analysis accurate enough, so that the non-Gaussian effects become more important for large errors. Note that a 50% error in precipitation observations is within a realistic range if they are satellite or radar retrieval products. Therefore, the Gaussian transformation proposed in this study seems essential for their practical assimilation.

(d) Sensitivity to the localization lengths of precipitation observations

In all experiments so far we have used the same horizontal localization length scale for precipitation assimilation as for rawinsonde observations (500 km, denoted as 1L). Since dense global precipitation observations are assimilated in our OSSEs, and precipitation has more local characteristics than other dynamical variables, we speculate that the optimal horizontal localization length scale for precipitation observations might be smaller than that for rawinsonde observations. Two additional experiments, PP_GT_10mR_0.5L and PP_GT_10mR_0.3L, with 0.5 and 0.3 times localization lengths for precipitation observations, respectively, are conducted. It is observed in Table 6 that the smaller length scales are helpful to the LETKF analyses, and the 0.5L (250 km) length scale would be the optimal setting under our current experimental design. The averaged RMS analysis error after the spin-up can be reduced by 32.7% of Raobs when the 0.5L length scale is used, compared with 27.2% when using the original length scale. This suggests that the optimal localization length could vary with different observation datasets and experimental settings and should be tuned appropriately.

5. Conclusions and discussion

Past attempts to assimilate precipitation observations into NWP models have found difficult to improve model analyses and, especially, model forecasts. In the experience with nudging or variational methods, the forecasts starting from analyses with precipitation assimilation lose their extra skill after a day or less (e.g., Errico et al. 2007). The linear representation of moist physical processes required in the variational data assimilation and the non-Gaussianity of both precipitation observations and model perturbations are two major problems in precipitation assimilation (e.g., Bauer et al. 2011).

The EnKF does not require linearization of the model, thus addressing the first problem. Besides, it is more efficient in improving the potential vorticity field than nudging or variational approaches by giving higher weights to ensemble members that are precipitating closer to the observations. Since potential vorticity is the variable that primarily determines the evolution of the forecast in NWP models, it is not surprising that the analysis improvements in EnKF would not be so quickly "forgotten" in the forecasts as in nudging.

In this study we tested these ideas with OSSEs of global precipitation assimilation with the SPEEDY model and the LETKF. In addition, we introduced two important changes in the data assimilation procedure that contribute to improving the performance of precipitation assimilation.

Since EnKF (like Kalman filtering) assume that variables and observations have Gaussian distributions, we first introduce a general algorithm to transform the precipitation variable into Gaussian distribution based on its climatological distribution. To handle the problem that the CDF of precipitation is discontinuous at zero, the middle value (median) of the zero-precipitation cumulative probability is chosen to transform all zero precipitation values. Second, we do not follow the approach used in most other studies of only assimilating positive precipitation observations. Instead, we propose a model background-based criterion in the ensemble data assimilation: precipitation observations are assimilated only at grid points where at least some members of prior ensemble are precipitating. This automatically allows zero precipitation observations to be assimilated.

Results in our simple OSSEs are extremely encouraging. By assimilating global precipitation, the globally averaged RMS analysis errors of u-winds after the spin-up stage are reduced by 27.2% as compared to only assimilating rawinsonde observations. The improvement is not "forgotten" and remains throughout the entire 5-day forecasts. All model variables show similar impacts of the precipitation assimilation. The improvement is much reduced when only modifying the moisture field by precipitation observations as done with nudging. By separating the globe into three verification regions, i.e., the NH extratropics, the tropics, and the SH extratropics, it is shown that the effect of precipitation assimilation is larger in the SH region than that in the NH region because the NH analyses are already accurate by denser rawinsonde stations. The tropical region shows the least improvement probably because of the slower dynamical instabilities and the prevailing convective precipitation type with small-scale features.

In addition, a number of comparisons among experiments were made in order to assess the impact of the Gaussian transformation, the observation selection criteria, the sensitivities to the precipitation error levels and the localization length scales of the precipitation observations. Applying the Gaussian transformation does not large impact on the analysis errors when the observation error level of precipitation is at an accurate 20% level, but it is very beneficial when observation errors are at a much higher 50% level. The proposed 10mR criterion (assimilating precipitation at the location where at least half of the members are precipitating) allows using some zero precipitation observations, and gives much better results than the traditional observation-based criterion of only assimilating positive precipitation, and even better than assimilating more observations with a looser criteria (1mR and 5mR criteria). Assigning smaller horizontal localization length scales for precipitation observations also improves the LETKF analysis in the present OSSE design.

Although these results are promising, it is important to recognize that the SPEEDY model is simple and that the simulated observations might be too ideal. Using the same model in creating the nature run and in the assimilation experiments neglects model errors and could also lead to overoptimistic results (i.e., the "identical twin" issue). Nevertheless, this is an essential first step to understand the feasibility and potential of the precipitation assimilation using an ensemble data assimilation method. The results indicate that the EnKF provides advantages for precipitation assimilation beyond the traditional nudging or variational methods. In addition, the explicit weights available in the LETKF are particularly useful to implement the "no-cost smoother" (Kalnay et al. 2007) and "Running in Place" (Yang et al. 2012), making better use of the information from a time sequence of observations. Since the precipitation is a variable that is tightly related to the past history of the moist physics, it is worthwhile to try these techniques in future studies.

Acknowledgements

We gratefully acknowledge support from the Office of Naval Research (ONR) grant N000141010149 under the National Oceanographic Partnership Program (NOPP). This study was also partially supported by NASA grants NNX11AH39G, NNX11AL25G, and NOAA grants NA100OAR4310248 and CICS-PAEK-LETKF11. George Huffman provided valuable guidance on the characteristics of the satellite precipitation data base.

References

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.

Davolio, S., and A. Buzzi, 2004: A nudging scheme for the assimilation of precipitation data into a mesoscale model. *Wea. Fcsting.*, **19**, 855–871, doi:10.1175/1520-0434(2004)019<0855:ANSFTA>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.

Falkovich, A., E. Kalnay, S. Lord, and M. B. Mathur, 2000: A new method of observed rainfall assimilation in forecast models. *J. Appl. Meteorol.*, **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.

Hou, A. Y., S. Q. Zhang, and O. Reale, 2004: Variational continuous assimilation of TMI and SSM/I rain rates: Impact on GEOS-3 hurricane analyses and forecasts. *Mon. Wea. Rev.*, **132**, 2094–2109, doi:10.1175/1520-0493(2004)132<2094:VCAOTA>2.0.CO;2.

Hou, A. Y., G. Skofronick-Jackson, C. D. Kummerow, and J. M. Shepherd, 2008: Chapter 6: Global precipitation measurement. in *Precipitation: Advances in Measurement, Estimation and Prediction*, Springer, 131–164.

Huffman, G. J. and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeor.*, **8**, 38–55, doi:10.1175/JHM560.1.

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.

Jones, P. D., and R. S. Bradley, 1992: Climatic variations in the longest instrumental records. *Climate Since A.D. 1500*, Routledge, London, 246–268.

Kalnay, E., and S.-C. Yang, 2010: Accelerating the spin-up of Ensemble Kalman Filtering. *Quart. J. Roy. Meteor. Soc.*, **136**, 1644–1651, doi:10.1002/qj.652.

Kalnay, E., H. Li, T. Miyoshi, S.-C. Yang, and J. Ballabrera-Poy, 2007: Response to the discussion on "4-D-Var or EnKF?" by Nils Gustafsson. *Tellus A*, **59**, doi:10.3402/tellusa.v59i5.15171.

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. *SOLA*, **1**, 45–48, doi:10.2151/sola.2005-013.

Li, X., and J. R. Mecikalski, 2010: Assimilation of the dual-polarization Doppler radar data for a convective storm with a warm-rain radar forward operator. *J. Geophys. Res.*, **115**, D16208, doi:10.1029/2009JD013666.

Li, X., and J. R. Mecikalski, 2012: Impact of the dual-polarization Doppler radar data on two convective storms with a warm-rain radar forward operator. *Mon. Wea. Rev.*, **140**, 2147–2167, doi:10.1175/MWR-D-11-00090.1.

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.

Lopez, P., and E. Moreau, 2005: A convection scheme for data assimilation: Description and initial tests. *Quart. J. Roy. Meteor. Soc.*, **131**, 409–436, doi:10.1256/qj.04.69.

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., 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.

Miyoshi, T., 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.

Molteni, 2003: Atmospheric simulations using a GCM with simplified physical parametrizations. I: model climatology and variability in multi-decadal experiments. *Clim. Dyn.*, **20**, 175–191, doi:10.1007/s00382-002-0268-2.

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**, doi:10.1029/2011WR010462.

Treadon, R. E., H.-L. Pan, W.-S. Wu, Y. Lin, W. S. Olson, and R. J. Kuligowski, 2003: Global and regional moisture analyses at NCEP. *Proc. ECMWF/GEWEX Workshop on Humidity Analysis*, Reading, United Kingdom, ECMWF, 33–48. [Link]

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.

Tsuyuki, T., 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.

Tsuyuki, T., and T. Miyoshi, 2007: Recent progress of data assimilation methods in meteorology. *J. Meteorol. Soc. Jpn.*, **85B**, 331–361, doi:10.2151/jmsj.85B.331.

Yang, S.-C., E. Kalnay, and B. R. Hunt, 2012: Handling nonlinearity in an ensemble Kalman filter: Experiments with the three-variable Lorenz model. *Mon. Wea. Rev.*, **140**, 2628–2646, doi:10.1175/MWR-D-11-00313.1.

Zhang, S. Q., M. Zupanski, A. Y. Hou, X. Lin, and S. H. Cheung, 2012: Assimilation of precipitation-affected radiances in a cloud-resolving WRF ensemble data assimilation system. *Mon. Wea. Rev.* (in press), doi:10.1175/MWR-D-12-00055.1.

Zupanski, D., and F. Mesinger, 1995: Four-dimensional variational assimilation of precipitation data. *Mon. Wea. Rev.*, **123**, 1112–1127, doi:10.1175/1520-0493(1995)123<1112:FDVAOP>2.0.CO;2.

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.

Tables

Table 1: The observation errors for the simulated observations.

Variable	Observation error
U	1.0 m s^{-1}
V	1.0 m s^{-1}
T	1.0 K
Q (specific humidity)	1.0×10^{-3} kg kg^{-1}
Ps (surface pressure)	1.0 hPa
PP (previous 6-hour accumulated precipitation)	20% or 50% (in different experiments)

Table 2: Design of all experiments.

Experiment	Observations		Gaussian transf.	Criteria for prcp. assimilation	Obs. error of prcp. obs.	Loc. lengths of prcp. obs.
	Raws.	Prcp.				
Raobs	X					
PP_GT_10mR	X	X	X	Prcp. members >=10	20%	1L (= 500km)
PP_GT_10mR_Qonly	X	X (only updating Q)	X	Prcp. members >=10	20%	1L
PP_noGT_10mR	X	X		Prcp. members >=10	20%	1L
PP_GT_ObsR	X	X	X	Obs. prcp. > 0.1 mm h^{-1}	20%	1L
PP_GT_10mR_50%err	X	X	X	Prcp. members >=10	50%	1L
PP_noGT_10mR_50%err	X	X		Prcp. members >=10	50%	1L
PP_GT_1mR	X	X	X	Prcp. members >=1	20%	1L
PP_GT_5mR	X	X	X	Prcp. members >=5	20%	1L
PP_GT_15mR	X	X	X	Prcp. members >=15	20%	1L
PP_GT_10mR_0.5L	X	X	X	Prcp. members >=10	20%	0.5L
PP_GT_10mR_0.3L	X	X	X	Prcp. members >=10	20%	0.3L

Table 3: Impact of precipitation assimilation on the last 11-month averaged analysis errors of u-wind.

Experiments	Last 11-month averaged RMSE of U (m/s) (percentage changes relative to Raobs)			
	Globe	NH	TR	SH
Raobs	1.58	0.67	1.64	2.03
PP_GT_10mR	1.15 (-27.2%)	0.53 (-20.6%)	1.45 (-11.2%)	0.91 (-55.2%)
PP_GT_10mR_Qonly	1.37 (-13.6%)	0.58 (-13.1%)	1.51 (-7.4%)	1.59 (-21.8%)

Table 4: Impact of the Gaussian transformation and accuracy of precipitation observations

Experiments	Last 11-month averaged RMSE of U (m/s) (percentage changes relative to Raobs)			
	Globe	NH	TR	SH
Raobs	1.58	0.67	1.64	2.03
PP_GT_10mR	1.15 (-27.2%)	0.53 (-20.6%)	1.45 (-11.2%)	0.91 (-55.2%)
PP_noGT_10mR	1.17 (-26.1%)	0.52 (-22.0%)	1.47 (-10.3%)	0.95 (-53.0%)
PP_GT_10mR_50%err	1.28 (-19.2%)	0.59 (-12.5%)	1.52 (-6.9%)	1.26 (-38.1%)
PP_noGT_10mR_50%err	1.87 (+17.8%)	0.79 (+18.2%)	2.00 (+22.0%)	2.29 (+12.9%)

Table 5: Impact of assimilation criteria of precipitation observations.

Experiments	Last 11-month averaged RMSE of U (m/s) (percentage changes relative to Raobs)			
	Globe	NH	TR	SH
Raobs	1.58	0.67	1.64	2.03
PP_GT_ObsR	1.58 (-0.3%)	0.69 (+3.4%)	1.94 (+18.7%)	1.40 (-31.0%)
PP_GT_1mR	1.29 (-18.6%)	0.57 (-14.3%)	1.62 (-0.9%)	1.04 (-48.6%)
PP_GT_5mR	1.19 (-25.2%)	0.52 (-22.3%)	1.50 (-8.5%)	0.94 (-53.6%)
PP_GT_10mR	1.15 (-27.2%)	0.53 (-20.6%)	1.45 (-11.2%)	0.91 (-55.2%)
PP_GT_15mR	1.13 (-28.9%)	0.52 (-23.0%)	1.42 (-13.4%)	0.89 (-56.0%)

Table 6: Impact of horizontal localization lengths of precipitation observations.

Experiments	Last 11-month averaged RMSE of U (m/s) (percentage changes relative to Raobs)			
	Globe	NH	TR	SH
Raobs	1.58	0.67	1.64	2.03
PP_GT_10mR	1.15 (-27.2%)	0.53 (-20.6%)	1.45 (-11.2%)	0.91 (-55.2%)
PP_GT_10mR_0.5L	1.07 (-32.7%)	0.48 (-28.0%)	1.31 (-20.0%)	0.95 (-53.4%)
PP_GT_10mR_0.3L	1.14 (-27.8%)	0.53 (-20.2%)	1.37 (-16.2%)	1.08 (-46.6%)

Figures

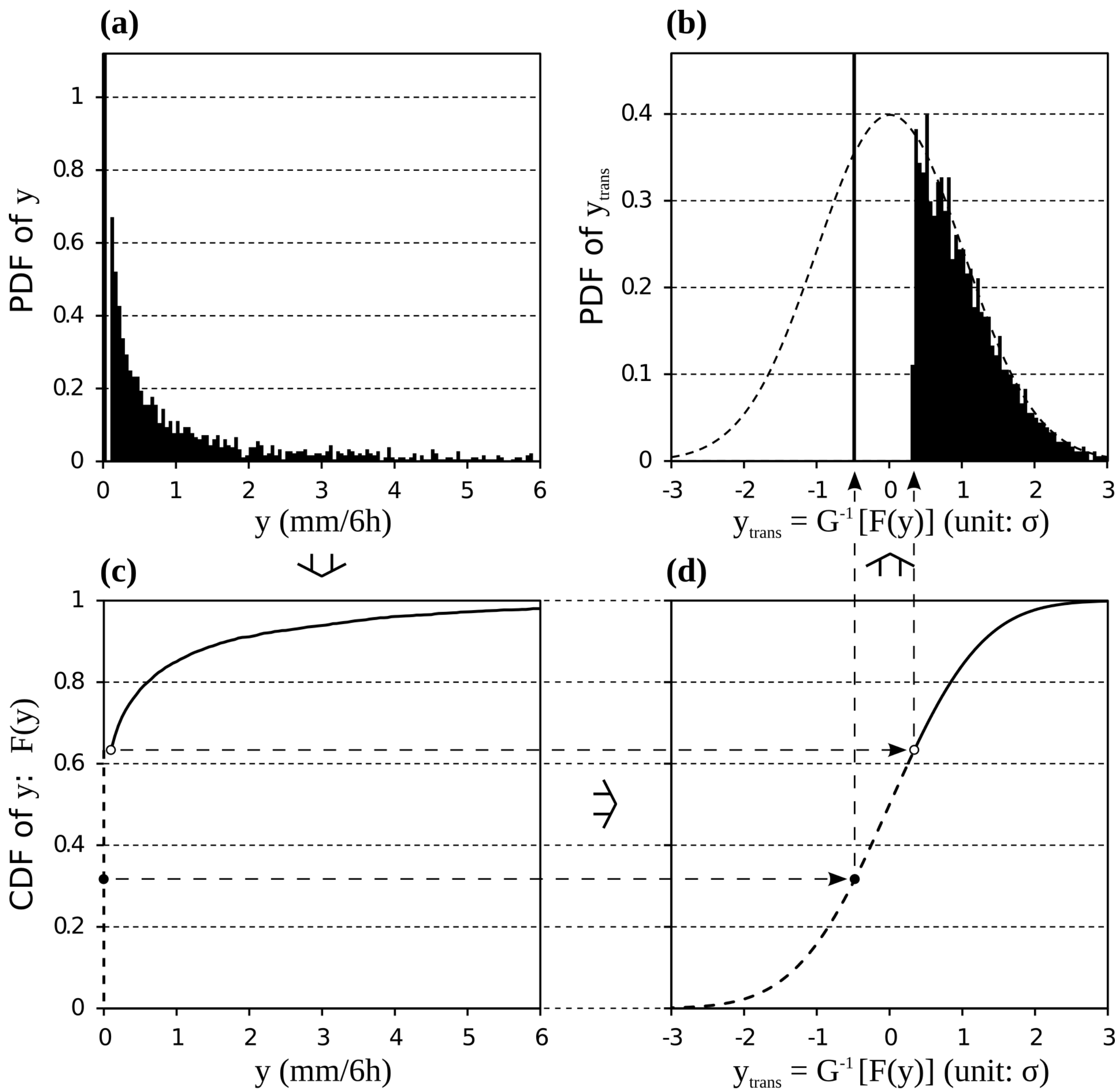

Figure 1: The probability density function and cumulative distribution function the of (a), (c) the original precipitation and (b), (d) the transformed precipitation at a grid point near Maryland (38.967N, 78.75W) in winter season (December – February) based on the 10-year nature run. The procedure of the Gaussian transformation is from (a) to (c), to (d), and to (b) as indicated by the arrows.

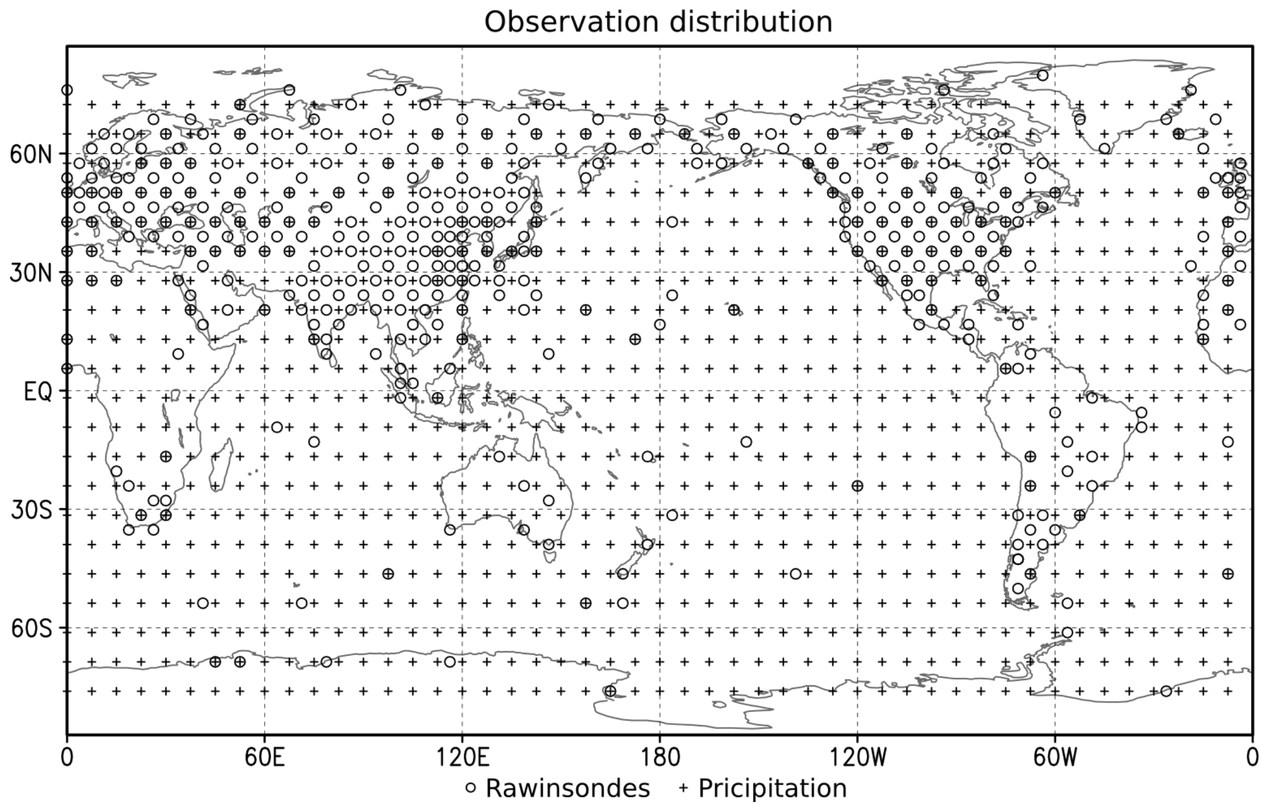

Figure 2: The spatial distribution of conventional rawinsonde observations (open circle) and global precipitation observations (plus sign) used in the OSSEs.

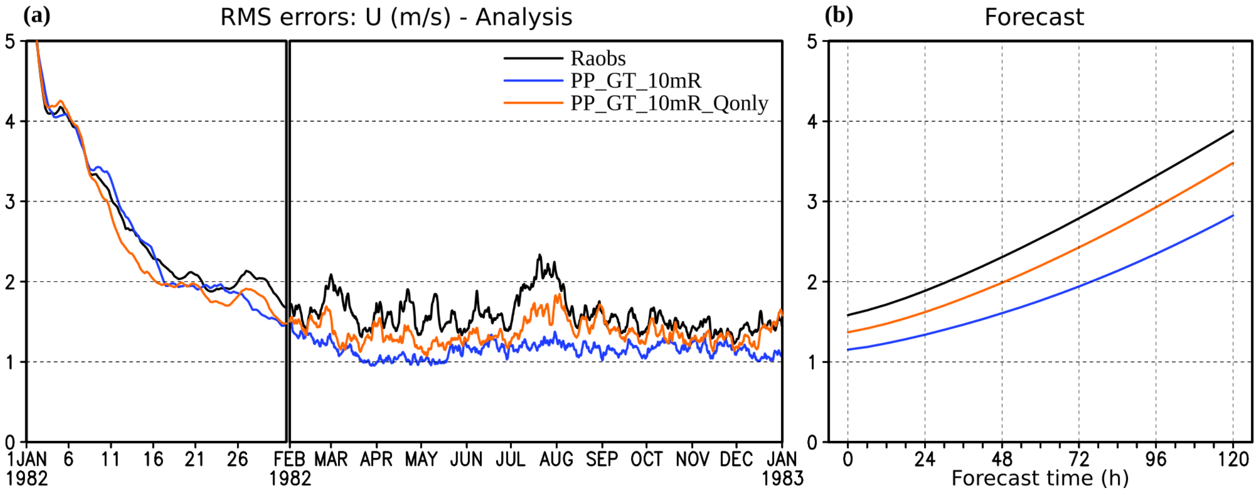

Figure 3: The global root-mean-square (a) analysis and (b) forecast errors (verified against the nature run) of u-winds in experiments Raobs, PP_GT_10nR, and PP_GT_10mR_Qonly. For the analysis errors, the evolution over one year is shown. Different time scales are used for the spin-up period (the first month) and the remaining 11 months. For the forecast errors, the 11-month (after the spin-up) averaged values are shown versus the forecast time.

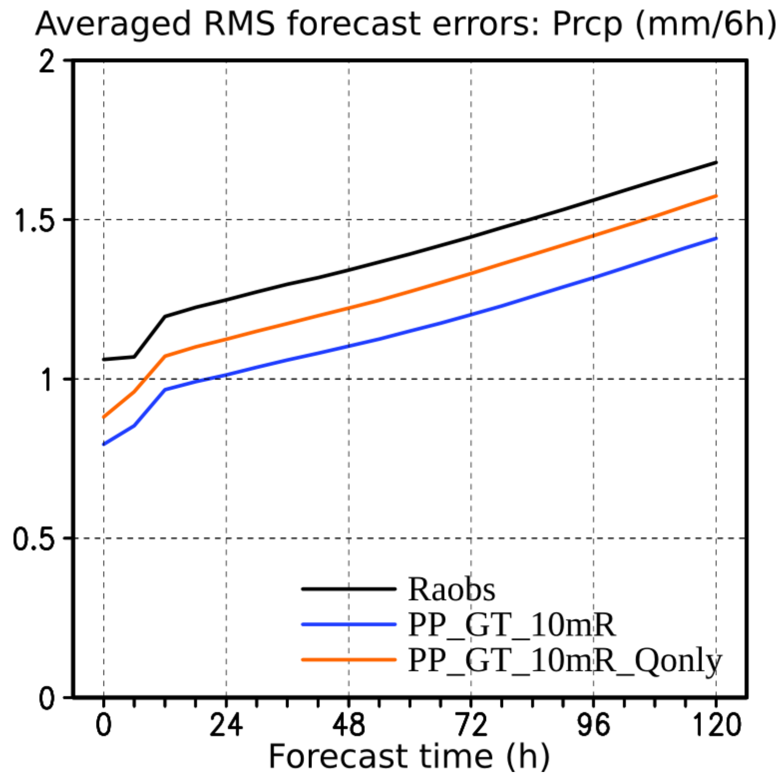

Figure 4: As in Fig. 3(b), but for precipitation forecast errors.

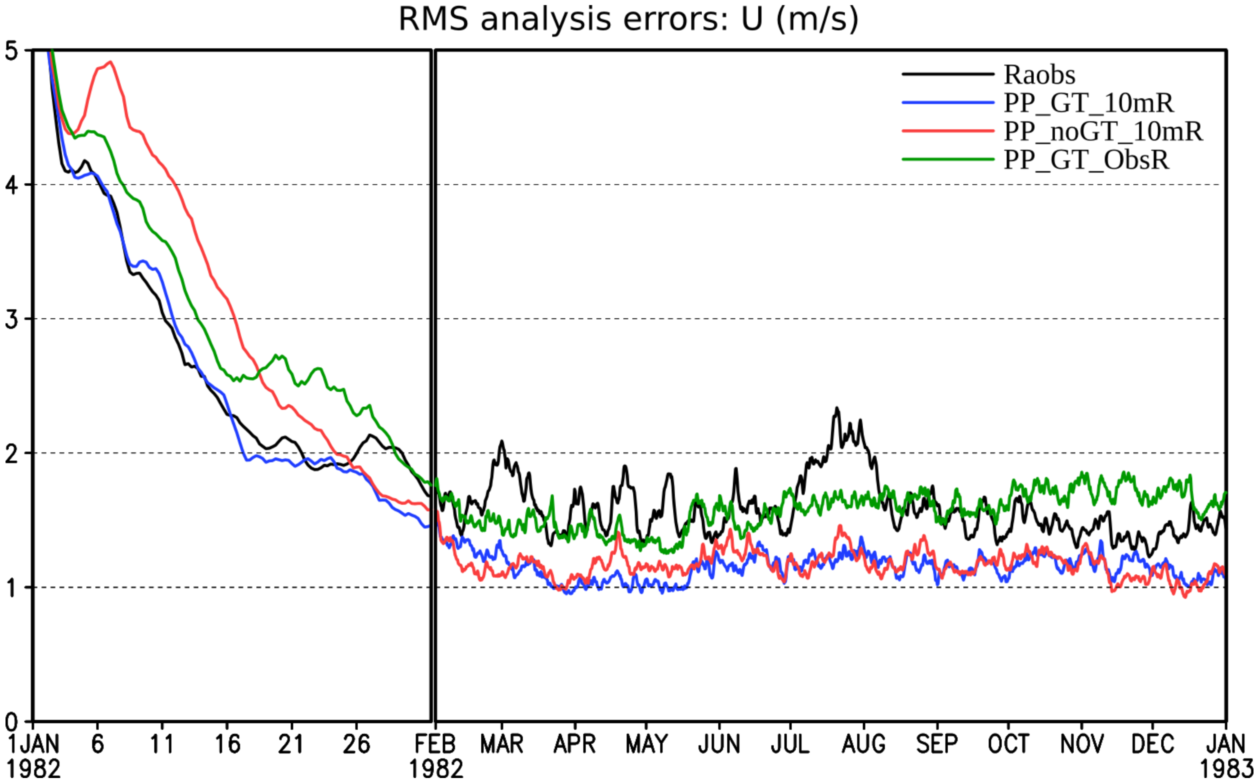

Figure 5: As in Fig. 3(a), but for experiments Raobs, PP_GT_10mR, PP_noGT_10mR, and PP_GT_ObsR.

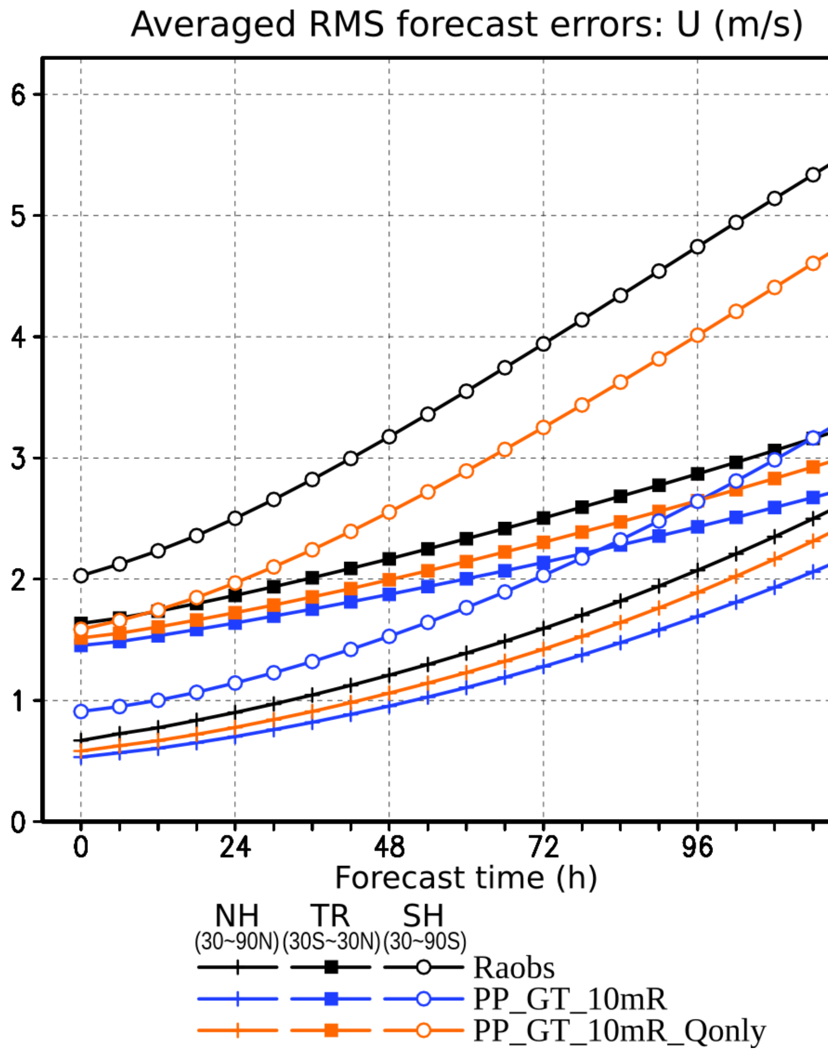

Figure 6: As in Fig. 3(b), but the RMS forecast errors are calculated separately for the Northern Hemisphere extratropics (30 – 90N; NH), the tropics (30S – 30N; TR), and the Southern Hemisphere extratropics (30 – 90S; SH), indicated by different marks on the lines.

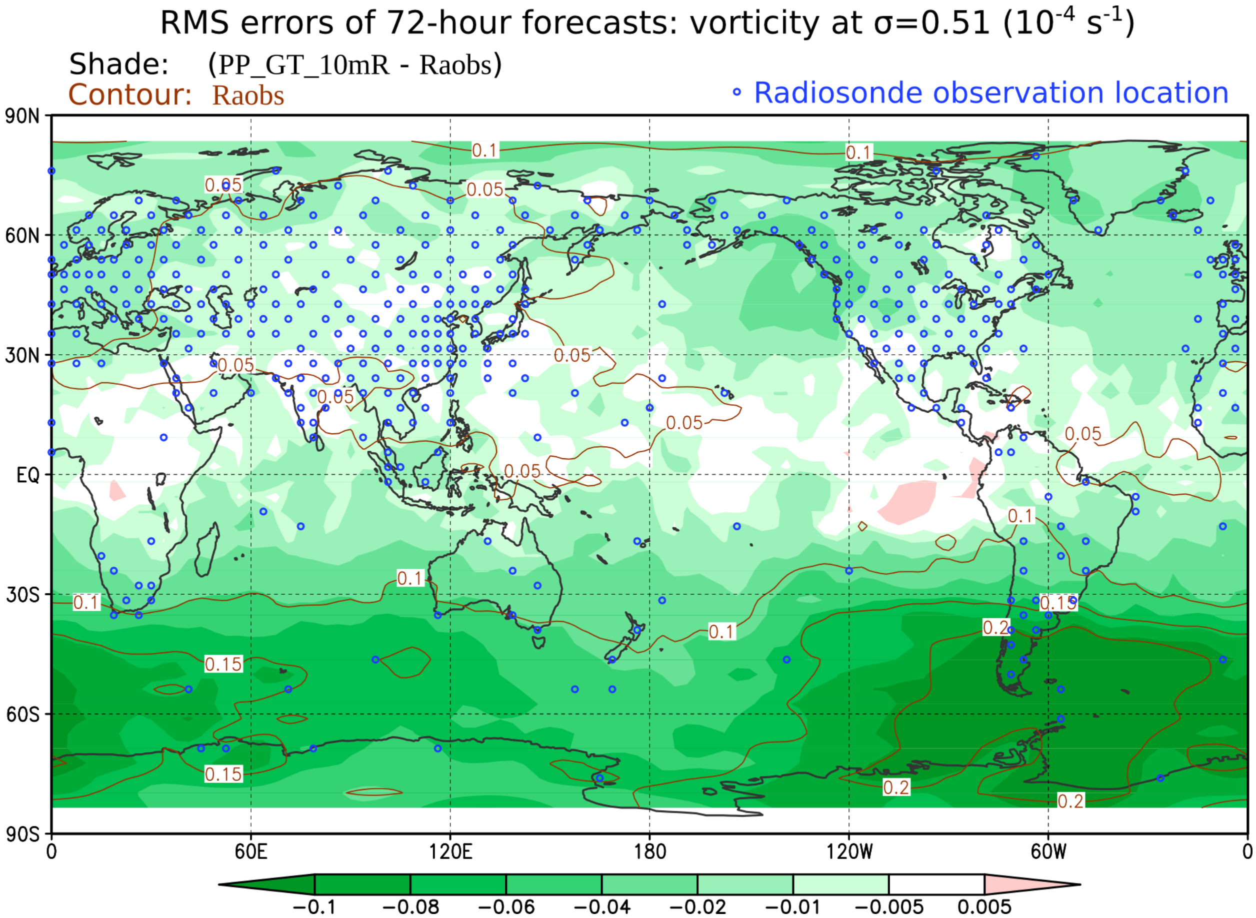

Figure 7: The global map of RMS 72-hour forecast errors of the vorticity at $\sigma = 0.51$ during the 11 months after the spin-up in Raobs (brown contour) and the corresponding error reduction from PP_GT_10mR to Raobs (shade). The rawinsonde observation locations are also shown in blue open circles.

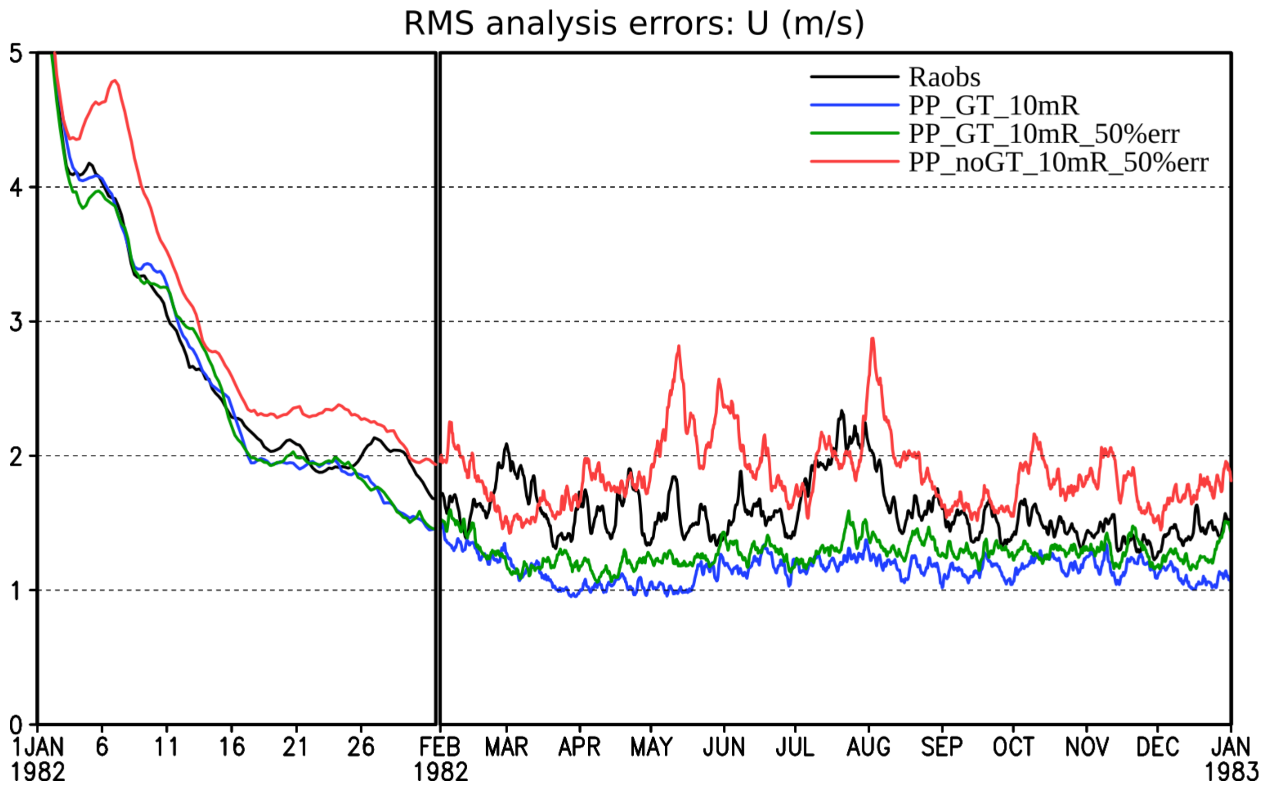

Figure 8: As in Fig. 3(a), but for experiments Raobs, PP_GT_10mR, PP_GT_10mR_50%err, and PP_noGT_10mR_50%err.