9.2 EFFECTIVE ASSIMILATION OF GLOBAL PRECIPITATION: SIMULATION EXPERIMENTS

Guo-Yuan Lien*, Eugenia Kalnay, and Takemasa Miyoshi
University of Maryland, College Park, Maryland

1. INTRODUCTION

Precipitation has long been one of the most important and useful meteorological observations. 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. Many efforts to assimilate precipitation observations have been made. Nudging or variational methods have been used to assimilate precipitation by modifying the model's moisture and sometimes temperature profiles as well, in order to obtain correct short-term precipitation according to the model parameterization of rain (e.g., Tsuyuki 1996; Davolio and Buzzi 2004; 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 the models revert to the regular forecasts soon after the assimilation of rain ceases, presumably because these methods are 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.

Precipitation processes parameterized by the model physics are usually very nonlinear. Therefore, it is very difficult and problematic to create the linearized version of the forward model which is required in the 4D-Var assimilation (Errico et al. 2007). In addition, the highly non-Gaussian distribution of the precipitation observations seriously violates the basic assumption of normal error statistics. Bauer et al. (2011) 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.

We propose to use the EnKF method to address these critical issues. First, the EnKF 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 to receive higher weights. Second, a general variable transformation is introduced to solve the problem that precipitation is highly non-Gaussian. In addition, we also allow assimilating zero precipitation observations by using a criterion that requires that at least several background ensemble members have positive precipitation. In this study, observing system simulation experiments (OSSEs) are carried out with a simplified atmospheric general circulation model in order to examine the effectiveness and feasibility of the proposed method.

2. METHODOLOGY

2.1 GAUSSION TRANSFORMATION

A general transformation algorithm to transform any variable y with a known arbitrary distribution into a Gaussian variable y_{trans} are defined through the connection between the two cumulative distribution functions (CDFs) of y and y_{trans}:

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

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,

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

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 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.

2.2 HANDLING ZERO PRECIPITATION

Figure 1 illustrates how the transformation works for the precipitation distribution at an example grid point. Using the inverse CDF of normal distribution G^{-1}, the CDF of the original precipitation variable [i.e., $F(y)$; Fig. 1c] is converted back to the transformed variable y_{trans}, with the CDF shown in Fig. 1d and the PDF in Fig. 1b. Since the precipitation data contain a large portion of zero values thus they are not continuous, a special treatment of the zero precipitation is needed. In the absence of a better solution, a reasonable choice is to

* *Corresponding author address:* Guo-Yuan Lien, Dept. of Atmospheric and Oceanic Science, Uni. of Maryland, College Park, MD 20742-2425; email: gylien@atmos.umd.edu

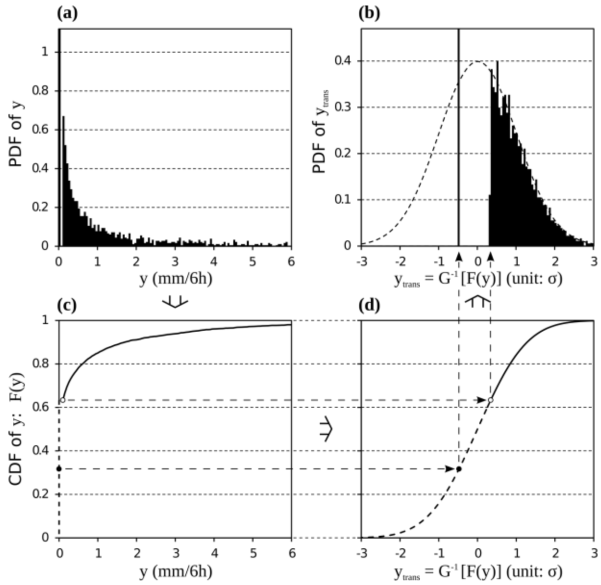

Figure 1: The probability density function and cumulative distribution function the of (a), (c) the original precipitation and (b), (d) the transformed precipitation at a grid point near Maryland (38.967N, 78.75W) in winter season (December – February) based on the 10-year nature run. The procedure of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

assign the middle value of zero-precipitation cumulative probability to $F(0)$. In this example, $F(0)$ = 0.317, which is equal to half of zero precipitation probability (63.4%), is assigned for all zero precipitation (solid circles). In this way, the zero precipitation in the transformed variable is still a delta function in its PDF (Fig. 1b), but it is located at the *median* of the zero precipitation part of the normal distribution). We also tested other more sophisticated approaches, but their experimental impact in the assimilation experiments was no better than that of the simple median approach.

With our current transformation algorithm handling the zero precipitation and an ensemble data assimilation system, zero precipitation observations can be naturally assimilated. Unlike the traditional precipitation assimilation that tends to discard 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, it is not possible to assimilate precipitation using an EnKF. We have chosen in this study is to require that at least half of the forecasts have positive precipitation, which can controls the assimilation quality (shown later) and saves computational time.

3. EXPERIMENTAL DESIGN

We conduct OSSEs using the Simplified Parametrizations, primitivE-Equation DYnamics (SPEEDY) model (Molteni 2003) coupled with the Local Ensemble Transform Kalman Filter (LETKF; Hunt et al. 2007). SPEEDY is a simple, computationally efficient, but realistic general circulation model. The version of SPEEDY model used in this study is run at a T30 resolution with 7 vertical sigma levels. LETKF is an ensemble Kalman filter scheme that performs most of the analysis computations in ensemble space and in each local domain. When applying the Gaussian transformation, the transformation algorithm is included in the observation operator of the precipitation variable. Besides, the observation errors associated with each observation are also transformed.

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. 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. 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. Observation data are then assimilated into the model with a 6-hour cycle. All experiments are run for 1 year.

Several experiments are summarized in Table 1. In the control run, only the realistically distributed conventional rawinsonde observations are assimilated (“Raobs” hereafter). For other experiments, global precipitation observations gathered uniformly every 2 by 2 model grids points are assimilated to estimate the impact of the precipitation assimilation. 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 to rain in order to use a precipitation observation (10mR) is applied. The observation error of precipitation observations in this experiment is 20%, which is rather accurate. In “PP_GT_10mR_Qonly”, only the specific humidity Q is updated during the LETKF assimilation of precipitation observations, resembling what conventional “nudging” methods do by arbitrarily modifying the moisture field in

Table 1: Design of experiments.

Experiment	Observations		Gaussian transf.	Criteria for prcp. assimilation	Obs. error of prcp. obs.
	Raws.	Prcp.			
Raobs	X				
PP_GT_10mR	X	X	X	Prcp. members >=10	20%
PP_GT_10mR_Qonly	X	X (only updating Q)	X	Prcp. members >=10	20%
PP_noGT_10mR	X	X		Prcp. members >=10	20%
PP_GT_ObsR	X	X	X	Obs. prcp. > 0.1 mm h^{-1}	20%
PP_GT_10mR_50%err	X	X	X	Prcp. members >=10	50%
PP_noGT_10mR_50%err	X	X		Prcp. members >=10	50%

the model. Experiment "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 observation accuracy on the precipitation assimilation, with much higher precipitation observation errors of 50% are used.

4. RESULTS

4.1 EFFECT OF PRECIPITATION ASSIMILATION

Figure 2a shows the evolution of the global root-mean-square (RMS) analysis errors (verified against the nature run) of the u-winds over one year. 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. It is clear that when all variables (and therefore the full potential vorticity) are modified (PP_GT_10mR; blue line in Fig. 2a), the improvement introduced by precipitation assimilation is quite large after the first month of spin-up. In addition to the LETKF analysis, the 0–5 day global RMS forecast errors of u-wind averaged over the last 11 months (i.e., after the spin-up) are also shown in Fig. 2b. 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. In contrast, when only the moisture field is modified (PP_GT_10mR_Qonly; orange line in Fig. 3a,b), the improvement in both analysis and forecasts is much smaller. Besides, the error growth rate (i.e., the slope) in PP_GT_10mR_Qonly is close to that in Raobs whereas the error growth rate in PP_GT_10mR is smaller as compared to the other two experiments. We only show the u-wind 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.

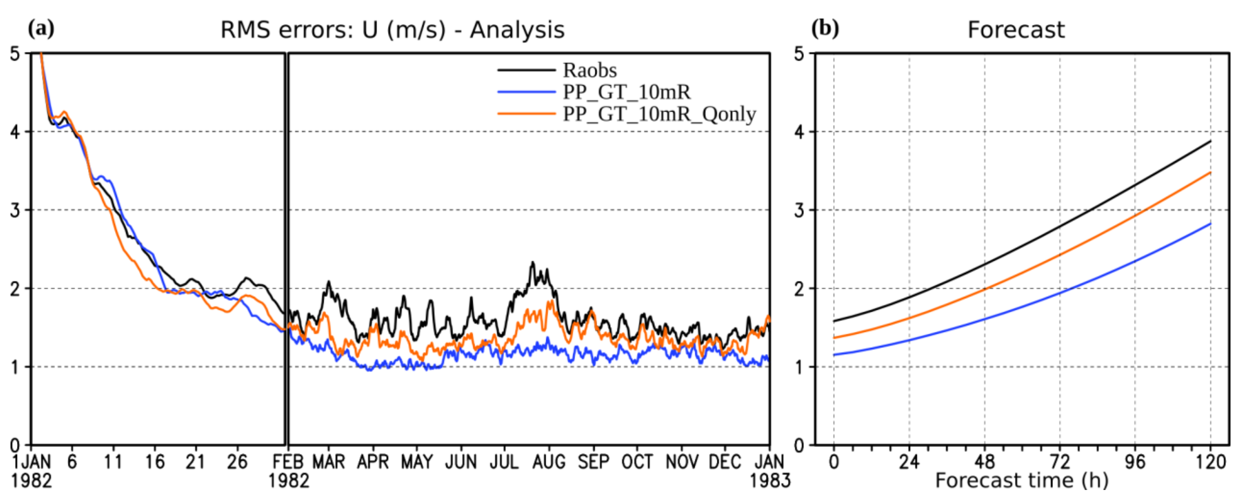

Figure 2: The global root-mean-square (a) analysis and (b) forecast errors (verified against the nature run) of u-winds in experiments Raobs, PP_GT_10nR, and PP_GT_10mR_Qonly. For the analysis errors, the evolution over one year is shown. Different time scales are used for the spin-up period (the first 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.

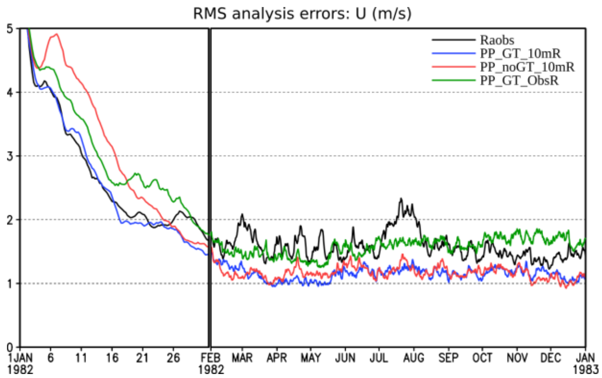

Figure 3: As in Fig. 2(a), but for experiments Raobs, PP_GT_10mR, PP_noGT_10mR, and PP_GT_ObsR.

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 by comparing the results of PP_GT_10mR, PP_noGT_10mR, and PP_GT_ObsR (Fig. 3). As shown in the figure, during the spin-up stage the LETKF analysis without transforming the precipitation variable (PP_noGT_10mR; red line in Fig. 3) is worse than that applying Gaussian transformation. However, with these accurate observations, the Gaussian transformation does not make a significant difference after the spin-up period. As to the observation selection criteria, the 10mR criterion seems to be essential in order to have an effective precipitation assimilation. The analysis of PP_GT_ObsR (green line in Fig. 3) is obviously degraded from PP_GT_10mR. Additional sensitivity experiments with different minimum numbers (1, 5, and 15 out of 20) of the precipitating member in order to pass to the assimilation were also conducted. It is concluded (not shown) that observations at locations where precipitating members are too rare can hurt the analysis, and requiring half (10) ensemble members are precipitating would be a proper criterion.

4.2 REGIONAL DEPENDENCE

To investigate the regional dependence of the impact of precipitation assimilation, 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 4 shows the RMS errors of u-wind in 0–5 day forecasts averaged over the last 11 months for main experiments as Fig. 2b, but for each region. It is clear that 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

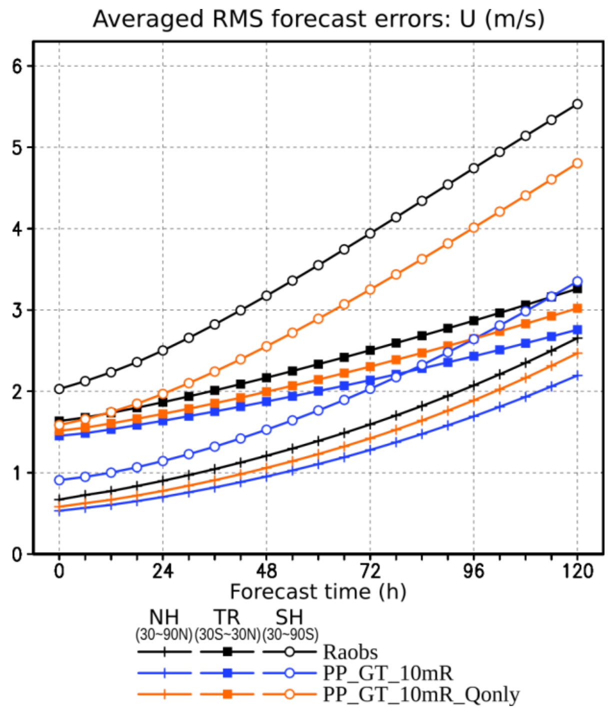

Figure 4: As in Fig. 2(b), but the RMS forecast errors are calculated separately for the Northern Hemisphere extratropics (30–90N; NH), the tropics (30S–30N; TR), and the Southern Hemisphere extratropics (30–90S; SH), indicated by different marks on the lines.

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 but a large effect on the SH region. The effect on the TR region is even smaller, which would be explained by different dynamical instabilities and precipitation mechanisms between the tropical and extratropical regions. 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 due to the stronger growth rates of mid-latitude baroclinic instabilities. It is noted that the improvement by precipitation assimilation in the SH region is quite large and the difference between modifying all variables and only modifying moisture by LETKF is emphasized in this region during the later forecasts.

4.3 SENSITIVITY TO ACCURACY OF THE PRECIPITATION OBSERVATIONS

As mentioned in subsection 4.1, 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 (Fig. 3). However, this is not the case with larger precipitation observation errors.

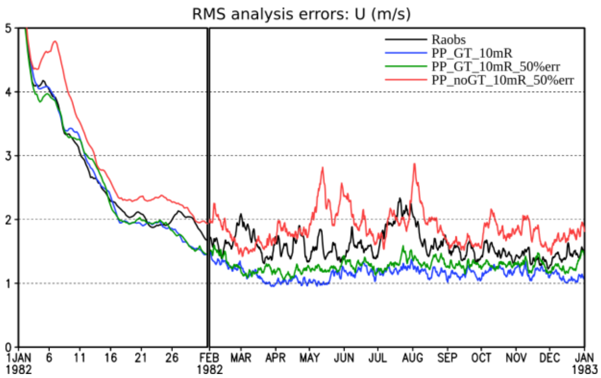

Figure 5: As in Fig. 2(a), but for experiments Raobs, PP_GT_10mR, PP_GT_10mR_50%err, and PP_noGT_10mR_50%err.

Figure 5 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% and the Gaussian transformation is used (PP_GT_10mR_50%err vs. PP_GT_10mR), the analysis becomes only slightly worse (shown as a green line in Fig. 5). However, without the Gaussian transformation and with 50% errors (PP_noGT_10mR_50%err; red line in Fig. 5), the precipitation assimilation fails and hurts the analysis. This sensitivity test demonstrates the importance of the Gaussian transformation to the practical assimilation of precipitation, since a 50% error in precipitation observations is within a realistic range if they are satellite or radar retrieval products.

5. CONCLUSIONS

Past attempts to assimilate precipitation observations into NWP models have found difficult to improve model analyses and, especially, model forecasts. 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).

An EnKF does not require linearization of the model, thus addressing the first problem. Besides, it is more efficient in improving the potential vorticity, which is the variable that primarily determines the evolution of the forecast in NWP models; therefore, the analysis improvements in EnKF would not be so quickly “forgotten” in the forecasts. In this study we test these ideas with OSSEs of global precipitation assimilation with the SPEEDY model and the LETKF. In addition, we introduce two important changes in the data assimilation procedure that contribute to improving the performance of precipitation assimilation. First, we 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 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 encouraging. By assimilating global precipitation, the globally averaged RMS analysis errors of u-winds after the spin-up stage are greatly reduced 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.

In addition, sensitivity tests show that 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 that only assimilate positive precipitation.

REREFENCES

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.

Davolio, S., and A. Buzzi, 2004: A nudging scheme for the assimilation of precipitation data into a mesoscale model. *Wea. Fcsting.*, **19**, 855–871.

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.

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.

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.

Mesinger, F. and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360.

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.

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.

Tsuyuki, T., 1996: Variational data assimilation in the tropics using precipitation data. Part II: 3D model. *Mon. Wea. Rev.*, **124**, 2545–2561.