A Heuristic Approach for Precipitation Data Assimilation: Effect of Forecast Errors and Assimilation of NCEP Stage IV Precipitation Analyses

ANDRÉS A. PÉREZ HORTAL, ISZTAR ZAWADZKI, AND M. K. YAU

Department of Atmospheric and Oceanic Sciences, McGill University, Montreal, Quebec, Canada

(Manuscript received 4 October 2019, in final form 23 January 2020)

ABSTRACT

Recently, Pérez Hortal et al. introduced a simple data assimilation (DA) technique named localized ensemble mosaic assimilation (LEMA) for the assimilation of radar-derived precipitation observations. The method constructs an analysis by assigning to each model grid point the information from the ensemble member that is locally closest to the precipitation observations. This study explores the effects of the forecasts errors in the performance of the method using a series of observing system simulation experiments (OSSEs) with different magnitudes of forecast errors employing a small ensemble of 20 members. The ideal experiments show that LEMA is able to produce forecasts with considerable and long-lived error reductions in the fields of precipitation, temperature, humidity, and wind. Nonetheless, the quality of the analysis deteriorates with increasing forecast errors beyond the spread of the ensemble. To overcome this limitation, we expand the spread of the ensemble used to construct the analysis mosaic by considering states at different times and states from forecasts initialized at different times (lagged forecasts). The ideal experiments show that the additional information in the expanded ensemble improves the performance of LEMA, producing larger and long-lived improvements in the state variables and in the precipitation forecast quality. Finally, the potential of LEMA is explored in real DA experiments using actual Stage IV precipitation observations. When LEMA uses only the background members, the quality of the precipitation forecast shows small or no improvements. However, the expanded ensemble improves the LEMA's effectiveness, producing larger and more persistent improvements in precipitation forecasts.

1. Introduction

Although the assimilation of precipitation into numerical weather prediction (NWP) models has a significant potential to improve the forecast quality, assimilating these observations is a particularly challenging task. One of the main reasons is that precipitation results from multivariate nonlinear processes that involve many state variables (e.g., temperature, humidity, pressure, winds). As such, obtaining precipitation from the model state is a deterministic forward problem with a unique solution, but the inverse problem is not: a given precipitation value can result from many different meteorological situations (states). Therefore, precipitation observations contain statistical information on each of the state variables related to the precipitation processes.

Conventional variational and ensemble Kalman filters (EnKF) data assimilation (DA) methods solve this inverse probabilistic problem assuming Gaussian distributions for the observation errors and prior estimates of the atmospheric states (Lorenc 1986; Hamill 2006). Several studies used these Gaussian DA approaches to assimilate precipitation observations over large domains with different degrees of success (Koizumi et al. 2005; Lopez and Bauer 2007; Lopez 2011; Kumar et al. 2014; Lien et al. 2016; Kotsuki et al. 2017). Nonetheless, the highly nonlinear moist physical processes and the non-Gaussian characteristics of precipitation errors complicate the assimilation of these observations using conventional approaches (Errico et al. 2007; Bauer et al. 2011; Lien et al. 2013).

In addition, simpler and more economical diabatic initialization (nudging) methods are frequently used for precipitation DA. They use theoretical or empirical relationships to modify the trajectory of a numerical model by forcing the model precipitation toward the observed values by adjusting the humidity or temperature profiles (e.g., Falkovich et al. 2000; Davolio and Buzzi 2004;

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

Corresponding author: Andrés A. Pérez Hortal, andres.perezhortal@mail.mcgill.ca

Davolio et al. 2017; Macpherson 2001; Bick et al. 2016; Jacques et al. 2018). These methods successfully force the precipitation toward the observed values during the assimilation windows, but most of the positive impacts in precipitation obtained by the DA are quickly forgotten ($\sim 3\text{--}6$ h).

Recent studies explore the potential of non-Gaussian approaches based on Monte Carlo methods, such as particle filters (PF, van Leeuwen 2009). Instead of assuming a Gaussian probability distribution for the model state, PF describe the model probability distribution by a discrete set of model states (ensemble members called particles). The posterior distribution properties are estimated using a weighted combination of all the members, where the members that are closer to the observations receive a higher weight. Poterjoy (2016) introduced a localized implementation of the PF (the local particle filter or LPF) that operates efficiently in high-dimensional systems and has potential for NWP applications. He obtained forecasts with an accuracy similar to those generated using an EnKF system (Poterjoy et al. 2017, 2019), although PF methods do not assume any particular distribution but rely only on the information contained in the ensemble forecast for the estimation of the atmospheric state.

In Pérez Hortal et al. (2019, hereafter PZY19), we introduced a heuristic approach for precipitation assimilation that directly uses the information in the ensemble (without relying on parametric assumptions) to estimate the atmospheric state. The method, called localized ensemble mosaic assimilation (LEMA), does not assume any particular shapes for the prior probability density (like Gaussianity). Instead, it is based on the intuitive idea that increasing the proximity in precipitation between model and truth will, on average, increase the proximity to the truth in state variables that caused the precipitation in the first place. This idea follows from the joint probability distributions of precipitation errors and state variables errors observed in the ensemble forecasts. In terms of forecast quality, the observing system simulating experiments (OSSEs) in PZY19 showed that LEMA produces a significant improvement in precipitation forecast that persists up for at least 12 h after the assimilation time. The improvements in precipitation are associated with a long-lasting reduction in the state variable errors (potential temperature, vapor mixing ratio, u wind, and v wind). However, the experiments in PZY19 were performed within ideal conditions: no model errors were considered (i.e., no shortcomings in representing the actual state of the atmosphere), and by construction, the truth was correctly covered by the background ensemble. This is not the case in real ensemble prediction systems where the ensemble spread does not represent the actual forecast uncertainty

(Fortin et al. 2014; Zhou et al. 2017). Since LEMA only relies on the information contained in the ensemble forecast to construct the analysis, in order to be effective, the spread of the ensemble members used by LEMA must sample all the possible forecast errors (timing, intensity, and model errors). Therefore, the underdispersivity of the ensemble forecasts will affect the performance of LEMA (as well as for any other methods that exclusively and strongly rely on the information in the ensemble).

The present study extends the OSSEs in PZY19 by considering situations that are likely to occur in real DA cases where the reality may be found partially or fully outside the spread of the background ensemble (underdispersive ensembles). As in PZY19, we extensively use OSSE experiments despite their limited usefulness for data assimilation in an operational context, where large discrepancy exists between forecasted and observed precipitation is typically beyond the differences in the context of OSSEs. However, important components of our objectives in this paper is the understanding of how information contained in ensemble forecasts propagates into state variables and the factors that may correct model trajectory over longer lead times, including the optimal scale of DA (PZY19) and ways to alleviate the underdispersivity of the ensemble. For these objectives OSSEs are essential.

Another objective of these experiments is to investigate how increasing forecast errors, that are not captured by the ensemble spread, affects LEMA's effectiveness. In particular, we will show that the quality of the analysis constructed by LEMA deteriorates with increasing forecast errors that are beyond the ensemble spread (underdispersive ensembles). A further objective is to present and characterize our first attempt to expand the ensemble spread by considering model states at different times and also states from forecasts initialized at different times (lagged forecasts).

Finally, we will explore the potential of LEMA with the expanded ensemble in DA experiments using actual Stage IV precipitation observations, considering situations where the forecast errors exceed the errors captured by the spread of background ensemble. For these unfavorable situations, when LEMA uses only the background members, the quality of the precipitation forecast shows small or no improvements. Nonetheless, we will show that the expanded ensemble provides a better representation of the actual forecast uncertainties. This improves LEMA's effectiveness in producing larger and more persistent improvements in precipitation forecasts.

The paper is organized as follows. The ensemble forecasts for the nine meteorological situations used in the OSSEs and the real DA experiments are described

TABLE 1. Ensemble forecasts for each precipitation event.

Case	Type	Analysis date	Name	Simulation start date
A	Synoptic front	0000 UTC 11 Apr 2013	En-6	1800 UTC 10 Apr
			En-12	1200 UTC 10 Apr
			En-18	0600 UTC 10 Apr
B	Cyclonic systems	1800 UTC 4 Apr 2013	En-6	1200 UTC 4 Apr
			En-12	0600 UTC 4 Apr
			En-18	0000 UTC 4 Apr
C	MCSs	0600 UTC 18 May 2013	En-6	0000 UTC 18 May
			En-12	1800 UTC 17 May
			En-18	1200 UTC 17 May
D	MCSs	0000 UTC 27 May 2013	En-6	1800 UTC 26 May
			En-12	1200 UTC 26 May
			En-18	0600 UTC 26 May
E	Synoptic front	0600 UTC 28 Dec 2015	En-6	0000 UTC 28 Dec
			En-12	1800 UTC 27 Dec
			En-18	1200 UTC 27 Dec
F	Synoptic front	0600 UTC 14 Oct 2012	En-6	0000 UTC 14 Oct
			En-12	1800 UTC 13 Oct
			En-18	1200 UTC 13 Oct
G	Synoptic front	1800 UTC 15 Aug 2016	En-6	1200 UTC 15 Aug
			En-12	0600 UTC 15 Aug
			En-18	0000 UTC 15 Aug
H	MCSs	0600 UTC 4 Jun 2014	En-6	0000 UTC 4 Jun
			En-12	1800 UTC 3 Jun
			En-18	1200 UTC 3 Jun
I	MCSs	1800 UTC 19 Apr 2012	En-6	1200 UTC 19 Apr
			En-12	0600 UTC 19 Apr
			En-18	0000 UTC 19 Apr

in section 2. In section 3 we recapitulate the LEMA method with additional details. In section 4 we evaluate the influence of forecast errors on the analysis quality using idealized OSSEs, showing the benefits of using the expanded ensemble in LEMA. In section 5 we present the results of the DA experiments with actual precipitation observations and present the benefits of using the expanded ensemble. Section 6 presents a discussion on the properties of the expanded ensemble and its relevance to other DA methods. Finally, section 7 contains a summary and conclusions.

2. Description of the experimental setup

a. Ensemble forecasts of the precipitation events

For this study, we consider nine precipitation events, named cases A to I. Five of the nine cases (A, B, E, F, G) are widespread precipitation events driven by cyclonic systems and with varied degrees of organization. In the rest of the cases (C, D, H, and I), precipitation was produced by several mesoscale convective systems (MCSs) scattered over the United States. A brief description of the precipitation events is shown in Table 1.

Case A is selected to characterize the effects of forecast errors on LEMA's performance in the context of OSSEs. Afterward, all the nine cases are used to explore

the potential of the method in a real DA application using actual Stage IV observations.

Our experiments employ the same NWP model and configuration as in PZY19. The numerical model is the Weather Research and Forecasting (WRF) Model with the Advanced Research WRF (ARW) dynamic solver (WRF-ARW), version 3.7.1 (Skamarock and Klemp 2008). The model domain covers the continental United States and southern Canada. The simulations are carried out using 41 vertical levels and 20 km horizontal grid spacing. The lateral boundary conditions (LBCs) and initial conditions (ICs) are downscaled from the Global Ensemble Forecast System (GEFS) data (1° resolution) to the WRF grid.

The main physics options used in the experiments are the WRF single moment microphysics scheme (WSM3; Hong and Lim 2006), the Yonsei University (YSU) boundary layer scheme (Hong et al. 2006), the Kain–Fritsch (KF) cumulus parameterization (Kain 2004), the Dudhia (1989) shortwave, and Rapid Radiative Transfer Model (RRTM) longwave radiation (Mlawer et al. 1997) schemes. Finally, the computational dynamic time step is 1 min. As for the other WRF parameters, we use the WRF default values.

For each case, we produce three ensemble forecasts of 21 members initialized at 6, 12, and 18 h before the

FIG. 1. Lagged-forecast scheme used to produce the ensemble forecasts. The red circles indicate the ICs downscaled from GEFS data while the numbers inside the circles denote the GEFS member. The blue lines indicate the spinup period while the red ones show the forecast period. Data assimilation is performed at $t = 0$ h and a 12 h forecast is run for each member.

DA is performed. These three ensembles will be denoted by En-6, En-12, and En-18, respectively. The ICs/LBCs for each of the ensemble forecast are downscaled from GEFS forecast data initialized at the same time as the WRF forecast initialization (Fig. 1). A detailed description of the ensemble forecasts for each case is shown in Table 1.

These ensemble forecasts are used to construct different idealized OSSEs as well as DA experiments with actual observations.

b. Observing system simulating experiments

In PZY19, the experimental setup used was optimal: the truth was covered by the background ensemble (symmetrical spread around the truth), and no model or observations errors were considered. Before considering assimilating actual precipitation observations, we should assess the consequence of forecast errors. Here, we will extend the previous study by considering more significant forecast errors. In these OSSEs, one model run is considered the “true” atmosphere and a different set of runs is considered as the background ensemble. The different forecast errors will be simulated by using different time lags between the initialization time of the truth and the background runs.

The main objective of the OSSEs is to understand how information contained in ensemble forecasts propagates into state variables for different situations with increasing forecast errors that are not captured by the ensemble

spread. A second objective is to present and characterize our first attempt to expand the ensemble spread by considering model states at different times and also states from forecasts initialized at different times (lagged forecasts). To that end, we will select case A for our analysis using an idealized setup. We carry out two OSSEs using the same background ensemble, but each experiment considers a different member as the truth. For easier comparison with the real DA experiments in section 5, for all the idealized experiments, a complete set of hourly precipitation accumulation observations are simulated from the member considered as the truth. The synthetic hourly precipitation observations are available every hour in the same domain as the Stage IV observations used in the real DA experiments (observations located east of 105°W). For simplicity, no model or observations errors are considered.

In the first experiment, we consider the En-6 runs with the member 0 as the truth run and the rest of the members as background. This represents an optimal situation where the truth is located at the center of the background ensemble. For this experiment, there is no initialization lag between the truth and the background runs. We denote this experiment by $\tau = 0$ h, where τ indicates the initialization lag.

In a second experiment, to simulate large forecast errors, we use the same background as in $\tau = 0$ h, but the “truth” run is taken to be the member 0 from the ensemble initialized 18 h before the analysis time (En-18/M0). For this experiment, named $\tau = 12$ h, there is a 12 h initialization lag between the truth and the background.

As we will show in section 4a, the background errors for each experiment increase with the increasing initialization lag between the truth and the background runs. In this section, we also explore the benefits of expanding the background ensemble for the analysis construction. For this, we perform a third experiment, $\tau = 12$ h/Xpd (Xpd indicates the expanded ensemble), similar to the $\tau = 12$ h experiment but constructing the analysis using the members from the background ensemble (member 1–20 from En-6) as well as all the members from the En-12 ensemble forecast. In addition, we further increase the ensemble by considering states at different times surrounding the analysis time 0, ± 1 , ± 2 h. It is important to note that the expanded ensemble is used only during the analysis construction. The new ensemble forecast is still produced by relaxing each background member toward the analysis.

A summary of all the OSSEs for a given case is shown in Table 2a.

c. Stage IV DA experiments

For the real DA, we assimilate the Stage IV hourly precipitation, a precipitation product that combines

TABLE 2. Experiments type summary. M# denotes the “member #” while M#₁–#₂ denotes members #₁ to #₂; τ indicate the initialization lag between the background and the “truth” runs; and “Xpd” denotes the experiments using the expanded ensemble in LEMA.

Observing system simulating experiments			
Name	Truth	Background	Expanded ensemble
$\tau = 0$ h	En-6/M0	En-6/M1–20	No
$\tau = 12$ h	En-18/M0	En-6/M1–20	No
$\tau = 12$ h/Xpd	En-18/M0	En-6/M1–20	Background + En-12
Stage IV DA experiments			
Name	Truth	Background	Expanded ensemble
StageIV/En-6	StageIV (east)	En-6	No
StageIV/Xpd	StageIV (east)	En-6	En-6 + En-12 + En-18

precipitation estimates from about 150 radars and about 5500 hourly rain gauge measurements over the continental United States (Baldwin and Mitchell 1997; Lin and Mitchell 2005). Over the Rocky Mountains area, the precipitation estimates are less reliable due to ground clutter, radar beam blockage, sparser radar coverage, and poorer rain gauge representativity. Therefore, all the observations located west of 105°W are discarded as in Lopez (2011).

The original precipitation data are available in a polar-stereographic grid with a 4 km grid spacing. To interpolate the observations to the model grid (20 km grid spacing), we first smooth the Stage IV data by running a 5×5 grid points moving-average window. This averaging matches the effective observations resolution with the model resolution and removes the small scales that are not represented by the model. Finally, the observations are interpolated onto the model grid using bilinear interpolation.

For the real DA with actual Stage IV precipitation observations, we use the nine precipitation events described in Table 2. For each precipitation event, we conduct two experiments: one using only the background to construct the Frankenstate, and the other using an expanded ensemble. For both experiments, we use as background ensemble the 21 members in the En-6 runs described in section 2a. The analysis is constructed 6 h after the background runs are initialized using the Stage IV precipitation observations.

In the first experiment (StageIV/En-6), we construct the Frankenstate using only the information in the background ensemble at the analysis times. In the second experiment (StageIV/Xpd), we construct the analysis using an expanded ensemble that includes the background ensemble En-6, and two other ensemble forecasts initialized at 12 and 18 h before the analysis time (En-12 and En-18 runs, respectively) described in section 2a. The ensemble is expanded even further by considering states at different times t surrounding the

analysis time (t_a), where $t = t_a + \Delta t$, with $\Delta t = 0, \pm 1, \pm 2$ h as described in section 3. This procedure expands the ensemble used to construct the analysis from the 21 members used in the first set of experiments to 315 members (21 members $\times$ 3 ensemble forecasts $\times$ 5 times).

A summary of all the real DA experiments for a given case is shown in Table 2b.

3. Revisiting LEMA

The localized ensemble mosaic assimilation (LEMA) creates an analysis using the information in the background ensemble by assigning to each vertical column in the model, the vertical profile of the state variables (u and v wind, temperature, humidity) from the ensemble member that is locally the closest to the precipitation observations (here, hourly accumulation). This analysis mosaic constructed using the selected members will be referred to as “Frankenstate.”

The local error in precipitation (local proximity) is measured by the mean absolute difference (MAD) between the forecasted hourly precipitation and the observed values, computed over a large square windows of $\Delta x = 820$ km width centered in the column. In PZY19, the MAD values are computed using precipitation intensity (in mm h^{-1}) expressed in dBZ values, transformed using the Marshall–Palmer relation. Although the same mathematical transformation can be used for hourly precipitation (in mm), the resulting units are not equivalent to dBZ, although they can be considered as “pseudo-reflectivity.” However, to avoid confusion with the reflectivity units, in this study we express the hourly precipitation in dBR units ($\text{dBR} = 10\log_{10}R$, with R in mm).

Therefore, for a given member “ m ,” the distance to the observations around the (i, j) horizontal grid point is

$$\text{MAD}_m(i, j) = \frac{1}{N_x N_y} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} |\text{dBR}_{x,y}^{\text{obs}} - \text{dBR}_{x,y}^m|, \quad (1)$$

where dBR^{obs} and dBR^m indicates the precipitation observations and the precipitation values corresponding to the m th member, respectively. The subindex x and y denotes the x index and y index of the horizontal grid point inside the observation windows. The summation limits, N_x and N_y , denotes the total number of horizontal grid points in the square window in each direction.

For each horizontal grid point, the ensemble member with the lowest MAD value is considered the “locally closest” one to the observations. To ensure that only members with MAD values strictly greater than zero are used in the closest member selection, during the selection process only members with a minimum precipitation coverage of $n_{\min} = 35$ grid points over the localization window are used. If over the observations window, the observed precipitation or no background member exceeds the minimum coverage $n_{\min}$, no information is assigned to that analysis column.

The new ICs for the ensemble forecast is produced by gradually forcing each member of the background ensemble toward the Frankenstate over a 30 min period preceding the analysis (nudging). With this procedure, the possible effects of sharp boundaries in the mosaic of closest members is attenuated.¹ The relaxation toward the Frankenstate is done by adding artificial terms to the model’s prognostic equations:

$$\left[\frac{\partial \phi(t)}{\partial t} \right]_{\text{new}} = \left[\frac{\partial \phi(t)}{\partial t} \right]_{\text{model}} + G[\phi^F - \phi(t)], \quad (2)$$

where $\phi(t)$ indicates a model variable at time t , ϕ^F the Frankenstate, and G is the nudging factor controlling the relative magnitude of the nudging term respect to other model processes. During the initialization, the nudging is only applied over the columns where a local closest member is found. On all the other grid points, we let the background to evolve without any artificial forcing. The forecasts initialized from the resulting ICs will be referred to as “Frankencasts.”

Since LEMA relies on the information contained in the ensemble forecast only, in order to be effective, the spread of the ensemble members used in the Frankenstate construction must be large enough to cover the reality. In other words, the ensemble used in the analysis construction should sample all the possible forecast errors (timing, intensity, and model errors). Since these forecast errors may not be captured by the background ensemble,

we modify the original LEMA approach (PZY19) by expanding background ensemble used to construct the analysis. The ensemble expanded by considering model states from forecasts initialized at previous times (time-lagged ensemble forecasts, see Hoffman and Kalnay 1983; Dalcher et al. 1988; Van Den Dool and Rukhovets 1994; Lu et al. 2007). Additionally, to further increase the sample of possible forecast errors, the ensemble is also expanded by considering, for each available member, states at different times t around the analysis time (t_a), where $t = t_a + \Delta t$, with $\Delta t = 0, \pm 1, \pm 2$ h. The final goal of this expansion is increasing the spread of the background ensemble to cover the “truth.”

Let us first revise the fundamental elements of LEMA, considering the experiment $\tau = 0$ h with the truth located at the center of the background ensemble for case A. We will consider the joint probability of the errors in potential temperature (ε_θ) for a model column and the local precipitation error [MAD, Eq. (1)] around that column. Only errors in potential temperature are shown, but similar results hold for the other variables. The joint probability is estimated from the background ensemble by the bidimensional histogram of the MAD, ε_θ pairs (columns), sampled over all the members and the entire domain, considering points with MAD > 0.

Figure 2a shows the joint probability of error in potential temperature and precipitation $p(\varepsilon_\theta, \text{MAD})$ when MAD is computed using an 820 km localization windows. The solid curve is the expected value of the conditional probability $p(\varepsilon_\theta | \text{MAD})$, representing the average relationship $\varepsilon_\theta = f(\text{MAD})$ per grid point. We can see that, on the average, when MAD decreases, the error in state variables steadily decreases as well. The joint probability can also be interpreted as the potential for improvement in potential temperature as MAD is reduced. This is the conceptual basis for LEMA: we assign to each grid point the vertical profile of the state variables (θ, q_v, U, V) of the ensemble member with the smallest MAD. This creates a mosaic of states (called Frankenstate) that, on the average, has smaller errors in the state variables.

Figure 2b shows the joint probability of the errors decrease² with respect to the Frankenstate (members with lowest MAD). The average gain in the ε_θ is approximately zero for small MAD gains, and increases as the reduction in MAD increases. However, there are grid points where the errors of these variables increase even when the MAD decreases. This a result of the stochasticity in the $\varepsilon_\theta = f(\text{MAD})$ relationship.

¹ Strictly speaking, nudging is not necessary. A filtering out of high frequencies can also be used or simply allowing gravity waves generated by the discontinuities to dissipate. However, the nudging procedure sets the stage for a future exploration in combining LEMA with latent heat nudging.

² The decrease in the errors is formulated in such way that positive values represent a positive impact in the analysis or forecast quality.

FIG. 2. Case A, experiment $\tau = 0$ h. (a) Joint probability $p(\varepsilon_\theta, \text{MAD})$ of given RMSE in potential temperature ε_θ and an error in precipitation MAD, and (b) joint probability $p(\Delta\varepsilon_\theta, \Delta\text{MAD})$ of decrease in errors when the member with lowest MAD is selected. The black curves indicate the expectation values with respect to the conditional probability. $\langle \varepsilon_\theta \rangle = \sum_{\varepsilon_\theta} \varepsilon_\theta p(\varepsilon_\theta | \text{MAD}) = \sum_{\varepsilon_\theta} \varepsilon_\theta p(\varepsilon_\theta, \text{MAD}) / p(\text{MAD})$, $\langle \Delta\varepsilon_\theta \rangle = \sum_{\Delta\varepsilon_\theta} \Delta\varepsilon_\theta p(\Delta\varepsilon_\theta | \Delta\text{MAD})$.

Given this degree of stochasticity, we may ask ourselves whether the member closest to the true precipitation is actually the best at maximizing the gain in state variables. Figure 3 shows the gain in state variables as a function of the proximity to true precipitation of the n th closest member. The results are shown for θ and q_v , but similar results are for U and V winds (not shown). Depending on the variable considered, 3 (for U , V , and θ) to 6 (for q_v) members close to the truth in precipitation are as effective as the closest one to reduce the error in state variables. Hence, this stochasticity in the joint probability of errors (Fig. 2a) has a beneficial aspect on LEMA: it provides a degree of tolerance to observations errors (appendix B, PZY19).

FIG. 3. Case A, experiment $\tau = 0$ h. Decrease in error (gain) in potential temperature ($\Delta\varepsilon_\theta$, green line) and vapor mixing ratio ($\Delta\varepsilon_{q_v}$, blue line) as a function of the proximity to the true precipitation of the n th closest member. The red lines indicate the decrease in the precipitation error ΔMAD when the n th closest member is selected.

Additionally, the adjustment of the atmospheric state will depend on the measure of the distance between model and observations used by LEMA and the optimization criteria used. In our study, we use the local MAD as a distance metric because it does not emphasize localized high values of errors. Using MAD as a distance measure, LEMA is optimized (by selecting the localization windows size) to maximize the transfer of information from precipitation to the state variables and for correcting the model trajectory in a way that remains closer to the observations for a longer lead time. In PZY19, we showed that the large observation window has a double benefit. It extends the area of influence of the observations as a larger portion of the domain is affected by LEMA and it also improves the transmission of information from precipitation to the state variables. The physical reason for this is simple: small-scale errors are determined by small-scale features (precipitation cells). As has been shown by Germann and Zawadzki (2002) and Surcel et al. (2015) and others, small scales are short lived (in agreement with the general correspondence between time and space scales of atmospheric phenomena, Orlanski 1975). This suggests that, on the average, these precipitation scales do not contain information relevant for a long-lasting correction of model trajectory. In other words, for data assimilation aiming at improvements lasting over 10 h or more at regional scales, the relevant information is not where it is intensely raining now but in what general region significant precipitation now occurs.

To further emphasize the last point, let us consider the scale dependence of forecast in the $\tau = 0$ h experiment

FIG. 4. Spatial-scale decomposition of DA impacts for case A's $\tau = 0$ h experiment using (left) 20 km localization windows and (right) 820 km windows. The DA impacts are measured by the relative improvements in the RMSE (ϵ) for potential temperature θ and vapor mixing ratio q_v , computed considering different spatial-scale intervals. The lines denote the decrease in RMSE averaged over all the ensemble members.

described in section 2, after assimilating synthetic observations using a 20 km and an 820 km localization windows. Figure 4 shows 12 h forecasts after precipitation was assimilated using MAD at 20 and 820 km. The diagram shows the relative decrease in error in θ and q_v at various scale intervals (similar results hold for U and V winds but not shown).³ When the Frankenstate is constructed using MAD at 20 km resolution, the analysis is constructed over the precipitation area getting closer to the detailed structure of the observed precipitation. On the other hand, using the large-scale component of MAD to correct model trajectory leads to an improvement in small scales ($\lambda < 120$ km) as well as in the large scales, with the gain increasing with scale.

³ The scale decomposition by intervals is done using the discrete cosine transform (DCT, Denis et al. 2002) and a 10th-order Butterworth bandpass filter. The DCT is equivalent to the fast Fourier transform (FFT), but it eliminates the problems associated with discontinuities at the boundaries of the domain.

4. Influence of forecast errors in the quality of the analysis

a. Effect of forecast errors

Hitherto, we considered a situation with a perfect model and the truth located at the center of the background ensemble. Figure 5a compares, at the time when DA takes place, the precipitation probability in the background ensemble with the truth precipitation for the $\tau = 0$ h experiment (case A). In this optimistic situation, the background ensemble captures remarkably well the “truth” precipitation pattern, with the regions with a high probability of precipitation (purple colors) being well collocated with the “truth” precipitation (black contours). This corresponds to a situation where the forecast errors are small, or in other terms, when observations fall within the ensemble forecast generated by perturbations of initial conditions. We know that in general this is not the case in real situations: ensemble forecasts are underdispersive. Hence, before considering assimilating actual precipitation observation, in this section we will assess the consequences of larger

FIG. 5. Case A OSSEs overview at 0000 UTC 11 Apr 2013, when DA is performed (a) the $\tau = 0$ h experiment, and (b) the $\tau = 12$ h experiment. The color plots depict the precipitation probability in the background ensemble (hourly accumulation values greater than 0.3 mm). The black contours show the synthetic precipitation observations while the blue contour indicates the area where the observations are available for the DA.

discrepancies between the forecast and the “truth” in the context OSSEs. Although the OSSEs may be greatly optimistic compared to real situations, they allow us to understand how LEMA works under different scenarios. In particular, we want to address how the potential for improvements in the state variables is affected by the magnitude of the forecast errors. A similar analysis is a challenging task in real DA cases where we only have a partial knowledge of the actual atmospheric state.

For this, let us consider the same background ensemble as before but the “truth” is taken from a run of the same meteorological situation but initialized 12 h before the background initialization. This corresponds to the $\tau = 12$ h experiment described in section 2b. The larger initialization lag (12 h) results in more substantial differences between the background and the “truth” precipitation patterns, with a significant mismatch between the areas with a high probability of precipitation in the background and the truth precipitation pattern (Fig. 5b). This indicates a less optimistic scenario where the truth is only partially covered by the background ensemble.

Figure 6a shows the joint probability of the decrease in the errors for this experiment. For comparison, the curve of the expected values from the $\tau = 0$ h experiment is shown here in gray. We notice that the errors in potential temperature have increased appreciably (by ~ 0.7 K). For small values of MAD, the expected value of ε_θ lost its dependence with MAD. This indicates that in regions with small MAD values, LEMA will not decrease the errors in state variables. For larger values of MAD, the average relationship between errors is maintained, and for the extreme large MAD the increase in ε_θ with increase of MAD is even stronger than the $\tau = 0$ h experiment. In this way, $p(\varepsilon_\theta, \text{MAD})$ is a diagnostic tool for assessing the effect of forecast errors on the analysis. Figure 6b indicates that the improvements in the error of potential temperature is very close to the one with small forecast error ($\tau = 0$ h experiment). In addition, large MAD reductions (ΔMAD) are limited to values smaller than 2 dBR (Fig. 6b). Since the initial error was larger (Fig. 6a) the relative gain is quite small. However, given the decrease in error in Fig. 6b and considering that we adjust four state variables, in the OSSE context we should expect LEMA to still be effective over the 12 h forecast but with lower skill. We will confirm this toward the end of this section.

Figure 7 shows the impacts of DA for case A on the root-mean-square error (RMSE) in potential temperature θ and vapor mixing ratio q_v , for the $\tau = 0$ h and $\tau = 12$ h experiments. The RMSE is computed over the entire domain for all the levels located in the troposphere. Similar results are obtained for U and V wind (not shown). As a result of the larger initialization lag between the background and the truth, the background forecast errors are larger in the $\tau = 12$ h than the $\tau = 0$ h runs (black and red lines in Figs. 7a,c). When the synthetic precipitation observations are assimilated, we obtain a persistent reduction in the errors for the state variables for both experiments (green and blue lines). Nevertheless, the error reduction is more important in both, absolute and in relative terms, for the $\tau = 0$ h experiment (smaller forecast errors). For the $\tau = 0$ h experiments, the ensemble-averaged decrease in RMSE for θ , q_v , U , and V are $\sim 12\%$ while for the $\tau = 12$ h runs, the ensemble-averaged error reduction is 3%–6%, remaining approximately constant over the entire 12 h forecast period.

The hourly accumulated precipitation errors are computed against the assimilated observations available only over the eastern United States (blue contour in Fig. 5). As for the improvements in the state variables, Fig. 8 shows that for the experiment where the background ensemble correctly samples the “truth” ($\tau = 0$ h), the DA produces considerable and long-lived improvements. For the $\tau = 12$ h experiment, where the members in the

FIG. 6. As in Fig. 2, but for experiment $\tau = 12h$.

background ensemble are asymmetrically distributed around the “truth,” the DA still produces persistent improvements, but of much smaller magnitude.

b. Expanded ensemble

In the preceding section, we showed that increasing forecast errors in the background reduce the performance of LEMA. This is not a surprising result since LEMA relies only on the information contained in the ensemble forecast, and for LEMA to be effective, the spread of the ensemble forecast must be enough to reflect its uncertainties and include model states close to the “truth.” On the contrary, if all the ensemble members in the background have large errors, the constructed Frankenstate will also have large errors.

The background ensembles used so far have been generated by the perturbations to initial conditions inherited from the GEFS data used to create the ICs/LBCs. In this section, we will use OSSEs to explore the benefits of

expanding the ensembles by including different forecast of the same meteorological situation. Concretely, we expand the background ensemble by considering states at different times and also other ensemble forecasts initialized at different times (lagged forecasts).

To that end, we perform a similar experiment to $\tau = 12h$ but constructing the Frankenstate using an expanded ensemble consisting of the background ensemble (En-6) and all members from the ensemble En-12. In addition, we further increase the ensemble by considering states at different times surrounding the analysis time 0, $\pm 1h$, $\pm 2h$. This results in an expanded ensemble with 315 members (instead of the original 20 members). This corresponds to the $\tau = 12h$ /Xpd experiment described in section 2b.

For the $\tau = 12h$ /Xpd experiment, the additional members in the expanded ensemble produce a small improvement in the proportionality between MAD and ϵ_θ for small MAD values (Fig. 9a). However, in the expanded ensemble, the pool of candidates to be selected for the Frankenstate is much larger. Hence, there is a greater probability of choosing a members with low MAD values, allowing bigger ΔMAD gains, that leads to larger improvements in the potential temperature errors ($\Delta \epsilon_\theta$, Fig. 9b).

Figure 10 shows the DA impacts on the RMSE for the state variables when we use the expanded ensemble to construct the Frankenstate. Results are only shown for potential temperature and vapor mixing ratio, but the results are similar for U and V winds. When the Frankenstate is constructed using only the background members En-12, LEMA produces only small improvements that persist over the entire 12h forecast period. However, better results are obtained when the Frankenstate is constructed with the expanded background ensemble ($\tau = 12h$ /Xpd). The state variables show a more significant reduction in the RMSE that persist over the entire forecast period (Fig. 10). Similar results are obtained for the precipitation forecast quality, computed over the observations domain (Fig. 11). The use of the expanded ensemble to construct the analysis results in a substantial improvement in precipitation forecast quality during the entire forecast period.

Since this procedure allows additional information to be available to LEMA, which translates into better forecast quality, we will consider the expanded ensemble as an integral component of the LEMA technique. We will see in the next section that the concept of expanded ensemble is crucial in the assimilation of actual observations.

5. Stage IV data assimilation experiments

In this section we present the results of the DA experiments with actual Stage IV observations described

FIG. 7. Impacts of DA for case A OSSEs on the entire domain, measured by the RMSE (ε) for potential temperature θ and vapor mixing ratio q_v . Similar results are for U and V winds. (left) The background and Frankencast RMSE for each variable, and (right) the relative decrease in RMSE for each Frankencast member. The thick lines denote the RMSE, and the decrease in the RMSE values averaged over all the ensemble members. The shading indicates the corresponding error variability over the ensemble members. Black and red colors indicate the background errors for the $\tau = 0$ and $\tau = 12$ h experiments while green and blue colors show (right) Frankencast errors and (left) improvements for the same experiments. The RMSE is computed over the entire domain for all the levels that are located in the troposphere.

in section 2c. Similar to the analysis of the DA impacts in the OSSEs, we evaluate LEMA using the RMSE of the hourly precipitation (in dBR units) computed against the Stage IV observations. Results for the ETS are not shown since LEMA only produced marginal improvements (<0.05). However, we will show next that considerable improvements are observed in RMSE, indicating that LEMA is more efficient in correcting the intensity of precipitation. This is expected, since LEMA uses MAD to measure the proximity to the observations that is sensitive to the precipitation intensity while ETS is only sensitive to the precipitation presence.

Let us first show the DA impacts in the forecast quality for the nine DA experiments carried out. Figures 12 and 13 show the DA impact in the precipitation RMSE when the expanded ensemble is used. To illustrate the benefits of using the expanded ensemble in real DA situations, Fig. 12 also shows the DA impacts for cases A, B, C, and D when only the background ensemble is used to construct the analysis.

For case A, when the Frankenstate is constructed using only the background ensemble (StageIV/En-6),

small improvements in the precipitation RMSE persist for the entire 12 h forecast (Figs. 12b,c). Nonetheless, higher gains are obtained when the analysis is constructed using the expanded ensemble (StageIV/Xpd).

For case B, using only the background (StageIV/En-6) produces small improvements in RMSE that persist over the entire forecast period (blue line in Figs. 12e,f). When LEMA uses the expanded ensemble to construct the analysis (StageIV/Xpd), a sharp decrease in the precipitation error is present during the first 2 forecast hours (green line). These improvements are almost lost from 2 to 6 h, and reappear again over the last 6 h.

For cases C and D precipitation was produced by several mesoscale convective systems (MCSs) scattered over the United States and is more localized than in cases A and B (widespread precipitation). For case C, in the StageIV/En-6 experiment, the assimilation of the observations does not improve the forecast quality. Only when the expanded ensemble is used (StageIV/Xpd) does LEMA produce long-lived improvements on the precipitation RMSE (Figs. 12h,i). Case D shows similar results as case B, with the DA produces small

FIG. 8. Impacts of DA on precipitation forecasts for the case A OSSEs computed over the observations domain. (left) The background and Frankencasts errors measured by (a) RMSE in dBR units, ϵ_R , and (c) ETS. (right) The error improvement achieved by each Frankencast member, with respect to the corresponding background member. The increase in ETS for the m th member is computed as $ETS_m^F - ETS_m^B$, where “F” and “B” indicate the error for the “Frankenstate” or “background,” respectively. The ETS is computed using a 0.3 mm detection threshold. Black and red colors indicate the background errors for the $\tau = 0$ and $\tau = 12$ h experiments while green and blue colors show (right) Frankencast errors and (left) improvements for the same experiments. The thick lines denote the ensemble-averaged values, while the shading shows the corresponding error variability over the ensemble members.

improvements in StageIV/En-6 experiment and a sharp decrease in the precipitation error during the first 2 forecast hours (green line in Figs. 12k,l) in the StageIV/Xpd experiment. As in case B, in the StageIV/Xpd experiment there is a loss of skill during 3–6 h of lead time that returns toward the end of the forecast period.

The previous results show that in real DA experiments, if only the background ensemble is used by LEMA, the quality of the precipitation forecast shows small or no improvements. However, when the expanded ensemble is used, LEMA produces larger improvements in precipitation forecasts. To confirm the potential of LEMA over a larger sample of cases, in Fig. 13 we show the DA improvements for five additional situations. Cases E and F are widespread precipitation events (Figs. 13a,d). Although in case G precipitation is also widespread, it is mostly produced by a large number of small-scale convective systems (Fig. 13g). Last, case H and I correspond to two mesoscale convective systems (MCS) precipitation events over central United States (Figs. 13j,m).

In four of the five cases, the improvements in the precipitation forecast persist up to 9 to 12 h (cases E, F, H, and I). However, case G (third row) shows the worst performance of LEMA where the improvements are quickly lost 3 h after the DA takes place. Let us examine this case in more detail. Figure 14 shows the precipitation field at 12 h forecast time without DA (upper left) and with DA (upper right, Frankencast forecast). The lower panels indicate the Stage IV fields and the reduction in the precipitation error by the DA. Although the impact of the DA as measured by RMSE is nil (see Fig. 13i), we point out that within the red ellipse in Fig. 14a, the background forecast indicates a well-organized line of precipitation that does not resemble the observed cellular structure (Fig. 14c). However, the forecast with DA correctly indicates a cellular structure. In addition, the most intense precipitation region (magenta circle) shows a marked DA quantitative improvement. The precipitation within the black circle in the northwest boundary of the observations domain indicates that the change is rather random, probably as a result of the inflow of information

FIG. 9. As in Fig. 6a, but for experiment $\tau = 12 \text{ h/Xpd}$.

from regions without DA. Thus, some improvements are present but not captured fully by the measure we use. The poor performance can perhaps be attributed to the very fragmented precipitation pattern, and hence very limited long-term information on the structure of small scales, and our emphasis on large scale and long duration of the effect of assimilation. In cases like this quantitative improvement by LEMA should not be expected over long lead times.

However, if we consider the average DA impact on precipitation over the nine cases, the assimilation results in long-lived positive impacts that persist up to 12 h of lead time (Fig. 15). These are encouraging results since the persistent improvements are obtained only with a single DA cycle and using an unfavorable setup, consisting of 3 small-sized and underdispersive ensembles (discussed in section 6).

Let us now investigate in more details the impacts of LEMA on the precipitation patterns. Hence, for the rest

FIG. 10. DA impacts on state variables RMSE (ε) for $\tau = 12 \text{ h}$ (green) and $\tau = 12 \text{ h/Xpd}$ (blue) experiments (case A OSSE). The RMSE is computed over the entire domain. The thick lines denote the ensemble-averaged values, while the shading shows the corresponding error variability over the ensemble members.

of this section, we will carry out our analysis only in a subset of four of the nine cases, namely cases A, B, C, and D. Figure 16 shows the observed and the forecasted precipitation patterns at 1 h of lead time for the four StageIV/Xpd experiments. Only member 0 is shown but the results are similar for other members. For case A (first row), the background forecast (second column) shows a good general agreement to the observed precipitation. When the observations are assimilated, the precipitation pattern becomes more similar to the observed one (third column). To better identify the positive (or negative) impacts of LEMA in the precipitation distribution, the fourth column in Fig. 16 shows the reduction in the precipitation errors obtained by the DA (positive gains: red and purple colors, negative gains: blue and green colors). For case A, LEMA is able to produce considerable improvements in the precipitation intensity over most of the frontal precipitation region.

FIG. 11. DA impacts on precipitation forecast for $\tau = 12h$ (green) and $\tau = 12h/Xpd$ (blue) experiments (case A OSSE), computed over the eastern Stage IV domain. The same precipitation metrics as in Fig. 8 are shown.

Similarly to case A, case B (second row) also shows a good general agreement with the observed precipitation, except in the southeastern part of the observation domain where precipitation is poorly represented. The DA is able to correct part the overestimation in the southeastern part of the domain (see red and purple colors in Fig. 16g).

For cases C and D (third and fourth row), the precipitation patterns are more localized than the other 2 cases and the position errors are important. For case C, the improvements are not as easy to identify as the other cases. However, Fig. 12 shows that there is a 10% improvement in the RMSE. Since the precipitation distance metric used by LEMA was optimized to reduce the large-scale precipitation errors, the improvements in precipitation occur only on average over a particular region. We will discuss this in more detail shortly. For case D, a large part of the improvements are due to the suppression of the overestimated precipitation in the western part of the domain (indicated by a red arrow in Figs. 16n,p).

Figure 16 also shows that LEMA is more effective in correcting large-scale errors in the precipitation than small-scale errors. This result is consistent with the analysis being constructed by selecting the members with the lower large-scale errors in precipitation (large MAD window). The use of large observation windows was motivated by the idea that decreasing errors at the large scales may have benefit at smaller scales (Durrant et al. 2013), leading to long-lived improvements in precipitation. Additionally, the large observation window is more effective to transfer information from precipitation to the state variables over a large region (PZY19). Furthermore, the reduction in large-scale errors is also a result of the stochastic relation between precipitation and the state variables. In LEMA, the transmission of the information from observations to the state variables is represented by the joint probability of a reduction in the precipitation and state variables errors. Figure 9b shows that, albeit there is an overall net reduction in the state errors, there is large variability around those mean values.

However, if we would want to put emphasis on short-term forecast, the distance metric used in LEMA and the observations windows can be optimized to effectively correct the small-scale precipitation errors. For example, one can use ETS over a smaller localization window to measure the local distance, or a weighted average of small-scale and large-scale MAD. Another alternative that may be promising is to combine LEMA, that focuses on large-scale precipitation errors, with Latent Heat Nudging, that takes into account the precipitation errors at the model resolution.

In our DA experiments, the extended ensemble was constructed by considering model states from different time-lagged forecasts, as well as model states at different times. This results in an ensemble with 315 members ($21 \text{ member} \times 3 \text{ ensemble forecasts} \times 5 \text{ times}$). All those members are candidates that may be used to construct the analysis. From that pool of candidates, for the analysis construction, LEMA only uses the members that better adjust to the observations. Additionally, the selected members can be used to assess the relative importance of the additional sources of information (time-lagged forecasts or models times) to the Frankenstate construction.

Figure 17a shows the number of grid points chosen from each lagged ensemble-forecast for the subset of four cases analyzed. Most the 315 members and all the lagged-forecasts ensembles are used in the Frankenstate construction, with contributions from the background ensemble (En-6) that are similar in magnitude to the additional lagged-forecasts ensembles En-12, En-18. In addition to the lagged forecasts, the expanded ensemble

FIG. 12. Impacts of DA on precipitation forecasts quality for (first row) case A, (second row) case B, (third row) case C, and (fourth row) case D for Stage IV DA experiments. (left) Observed precipitation (Stage IV) when the DA takes place. (middle) Hourly accumulation forecast error measured by RMSE (in dBR units). (right) Decrease in the RMSE ($\Delta\epsilon_R$) when the DA is applied (improvements). The background errors are shown in red while blue and green colors display the StageIV/En-6 and StageIV/Xpd errors and improvements. The lines denote the ensemble-averaged values while the shaded area indicates the variability around the mean.

includes different model forecast times around the analysis time to further expand the ensemble. Figure 17b shows the models times (time lag with respect to the analysis time) in the expanded ensemble. For the cases shown, all the model times included in the expanded ensemble contribute to the analysis. That is, LEMA using the expanded ensemble corrects errors originally present in the background ensemble (0 h time lag).

6. Discussion on the expanded ensemble properties

LEMA exclusively relies upon the information contained in the ensemble forecast, as represented by the joint probabilities of errors such as in Fig. 2. However,

in practice, ensembles are notoriously underdispersive (e.g., Zhou et al. 2017), which means that the ensemble does not represent all the forecast uncertainties. Possible causes include the underestimation of the errors in the initial conditions, nonoptimal perturbations that do not capture the growth of forecast errors, and model errors. Hence, forecast errors can be seen as a manifestation of the model and initial conditions errors not taken into account in the generation of the ensembles.

In section 4 we used the joint probabilities $p(\Delta\text{MAD}, \Delta\epsilon_\theta)$ to evaluate the effectiveness of LEMA for a particular situation (e.g., Figs. 2b and 6b). When the ensemble members used for the analysis construction capture the forecast errors, a decrease in the precipitation error

FIG. 13. As in Fig. 12, but for (first row) case E, (second row) case F, (third row) case G, (fourth row) case H, and (last row) case I.

MAD (by construction) results in a considerable decrease in the state variables errors. However, the latter approach is only suitable for idealized situations where the truth state is fully known. In a real case, the only information available is the forecast errors with respect to the observations (MAD). In this case, to measure the extent to which the ensemble used in LEMA capture the actual forecast errors, we can use the ensemble-averaged distributions of precipitation errors, $p(\text{MAD})$.

Figure 18 compares the spectrum of errors captured by the background spread (in green and red) to the $p(\text{MAD})$ of the actual errors with respect to the observations (orange curves) at 20 km resolution (grid scale). As shown, the forecast errors exceed the errors captured by the spread of ensemble forecasts En-6 (green), indicating that the background ensemble does not capture the actual uncertainties given by the orange curves, StageIV/En-6. With the expanded ensemble (red curves)

Case G at 12h of forecast time

FIG. 14. DA impacts of precipitation fields for case G StageIV/Xpd experiment at 12h of lead time. (a),(b) Ensemble-averaged precipitation for the runs without DA (background) and with DA (Frankencast), respectively. (c) Stage IV fields and (d) ensemble-averaged reduction in the precipitation error by the DA with respect to the background run. The black contour shows the Stage IV precipitation ($R = 0.3$ mm) observations while the blue contour indicates the area where the observations are available.

the forecast errors captured by the new ensemble spread, Spread/Xpd (considering all possible members as “truth”) the distribution of errors approaches the actual forecast errors blue dashed curves, StageIV/Xpd. Clearly, the expanded ensemble provides a better representation of the actual uncertainties, improving the performance of LEMA (see Fig. 12).

In our experiments, the perturbations in the ICs are inherited from the GEFS data used to create the ICs and LBCs. The GEFS data used to create the ICs/LBCs for the WRF runs never exceeded the 30h of lead time, where the GEFS dispersivity remains close to 1 for the 500 hPa geopotential height field (Zhou et al. 2017). However, for precipitation the forecast error probabilities, $p(\text{MAD})$, with respect to the precipitation observations (Stage IV) are quantitatively and qualitatively different from $p(\text{MAD})$ captured by the ensemble spread (Spread/En-6). These differences between OSSE and Stage IV experiments in the $p(\text{MAD})$ hold for different cases, representing different meteorological situations

and initial conditions (ICs). This difference could be partly (or mostly) due to large model errors (particularly errors in the convective parameterization), a common element to all the runs. Since the model errors are not explicitly taken into account, LEMA will try to compensate for those errors through the new ICs (analysis) by choosing the models state variables generating precipitation closer to the observations and an improvement in the precipitation forecast after the DA takes place. If this is the case, the forecast may diverge from the observations. This could be the cause of the temporary loss of skill observed for case B and D in the Stage IV DA experiments (see second and forth row in Fig. 12).

Perturbing the model's initial and boundary conditions, choosing different parameterizations included in the model formulation, are the usual methods to generate the spread of ensemble forecasts. In the OSSEs as well as in the assimilation of Stage IV data we used some forecast errors as a proxy for model and initial condition

FIG. 15. Decrease of the precipitation RMSE for the nine cases for the StageIV/Xpd experiment. The green line indicates the average improvements while the dark-green shaded region indicates the 95% confidence intervals. The full variability over the mean is indicated by the light-green shaded area.

errors to justify the expansion of the ensemble. A similar idea was used in a different context (Berner et al. 2015), where forecast errors are used as reflecting model errors. Hence, they can measure how different model-error schemes represent the actual uncertainties due to model errors. We show here that the initialization time of a limited area model is an aspect of uncertainty that can also be considered as an element of model perturbation that helps to expand the ensemble spread. Thus, the concept of the expanded ensemble, which we now consider as an integral part of LEMA, can be seen as a heuristic expansion of model perturbations. In an operational setup, where the ensemble forecasts are run periodically (e.g., every 6 h), this expansion of the model perturbations do not require additional model runs, since often the lagged forecasts are already available. Hence, the expanded ensemble is computationally inexpensive alternative to generate additional model forecast perturbations.

In section 5 we showed that when the expanded ensemble is used, all the lagged-forecast initialized at different times and different model times contribute to the Frankenstate (Fig. 17). This suggests that for precipitation forecasts, the different initialization and model times have a similar probability to represent the actual atmospheric state. That is, the members of the expanded ensemble appear to be as equiprobable with respect to precipitation errors as the members of the original ensemble generated by perturbations of the initial conditions. We have shown that the expanded ensemble

improves the coverage of the reality (dispersivity). To further emphasize this, Fig. 19 shows that the dispersivity of background (En-6) and the expanded ensemble (Xpd), with respect to the Stage IV observations. The correct dispersivity in the ensemble results when the RMSE of the ensemble mean matches the ensemble spread (Fortin et al. 2014). For the cases shown, the background ensemble (red curve) are underdispersive (dispersivity ~ 0.5 – 0.6) and the positive impacts of LEMA are limited (Fig. 12). The expanded ensemble considering all the candidate members (Xpd/Prior) have a much better dispersivity (closer to 1) as a result of a reduction in the RMSE of the ensemble mean (not shown). When we only keep the members selected by LEMA (Xpd/Posterior), the dispersivity is improved even further. This is an expected result since the members with larger errors are discarded, reducing the distance of the ensemble mean to the “truth.”

7. Summary and conclusions

In PZY19 we introduced a heuristic DA method for precipitation assimilation. The method, called localized ensemble mosaic assimilation (LEMA), uses the information from an ensemble of model runs in a direct manner to construct a mosaic analysis by selecting the ensemble members that locally have the smallest precipitation error (MAD) measured at large scales. The present study extends the idealized experiments in PZY19 by considering less optimistic situations that are likely to occur in real DA situations with significant forecast errors and underdispersive ensembles.

The OSSEs with different characteristics of forecasts errors show that, when a small number of ensemble members is used, the impacts of DA in the forecast quality deteriorate rapidly with increasing forecast errors beyond the spread of the ensemble. This is an expected result since LEMA relies on the information contained in the ensemble forecast only. If the ensemble does not cover the true state and all the ensemble members in the background have large errors, the constructed Frankenstate will also have large errors. To partially remedy this problem, we present a method to expand the spread of the background ensemble used by LEMA with the goal of improving the representation of the actual forecast uncertainties by considering additional lagged-forecasts ensembles and model states at different times. With OSSEs, we found that the additional information in the expanded ensemble improves the efficacy of LEMA, produces larger and long-lived improvements in the state variables and in the precipitation forecast quality. Although the OSSEs allow us to understand how the forecast errors affect the

FIG. 16. Observed and forecasted hourly precipitation for the StageIV/Xpd DA experiments at $t = 1$ h of lead time. (second column) The precipitation forecasts without DA (background) and (third column) with DA are shown for (first row) case A, (second row) case B, (third row) case C, and (fourth row) case D. (fourth column) Decrease in the precipitation error ($\Delta\epsilon_R$) by the DA at each grid point (difference between background and LEMA forecast). The black lines show the Stage IV precipitation ($R = 0.3$ mm) observations while the blue contour indicates the area where the observations are available.

FIG. 17. Histogram of selected members for the Stage IV/Xpdt experiments, for cases A to D. (a) Number of members selected from each of the lagged-forecasts runs. (b) Model times (time lag respect to the analysis time) selected from the expanded ensemble. The numbers in each block in the bar plots indicate the numbers of members selected for the corresponding source (lagged-ensemble or model time).

information transfer from precipitation to the state variables, they are overoptimistic compared to real situations.

Therefore, we also explore the potential of a LEMA in real DA experiments using Stage IV precipitation observations. As a proof of concept, we consider unfavorable situations where the forecast errors were not captured by the spread of background ensemble. For these unfavorable situations, when LEMA uses only the background members, the quality of the precipitation forecast shows small or no improvements. However, when the Frankenstate is constructed with the expanded ensemble, LEMA produces larger improvements in precipitation

forecasts. We consider these results encouraging since the expanded ensemble was constructed using 3 small-sized and underdispersive ensembles of 21 members each. However, this setup is only a proof-of-concept and more initialization lags are not harmful to LEMA, even if it produces an overdispersive ensemble. The member selection in LEMA will filter out the members that are far away from the truth and only use the information available in the members with the smallest large-scale MAD, improving the dispersivity of the expanded ensemble.

To understand the origins of the better effectiveness in LEMA, in section 6 we explored the properties of the expanded ensemble. In our real DA experiments, the forecast errors exceed the errors captured by the spread of the background ensembles, indicating that they do not represent the actual uncertainties. We showed that the expansion of the ensemble spread captures part of the uncertainties not taken into account in the original background ensemble, which results in a better performance of LEMA. However, despite the larger spread and corrected dispersivity of the expanded ensemble, not all of the forecast errors are taken into account. Figure 18 shows that the expanded ensemble does not fully correct the distribution of precipitation errors, especially for large MAD values. This difference could be partly explained by significant model errors not being explicitly taken into account in the expansion of the ensemble. Hence, to better capture the errors, the background ensemble can be further expanded by taking into account model errors. This can be done, for example, by adding perturbations to the Kain–Fritsch convective parameterization.

Additionally, the expanded ensemble concept is not only relevant to LEMA. It can be applied directly to any assimilation method that relies on the information contained in the ensemble to improve their performance when the background ensemble does not represent the actual forecast uncertainties or the size of the ensemble is small. For example, the expanded ensemble can be used in particle filters (PF) to improve the representation of the model probability density function (pdf). Although its potential is yet to be determined, the expanded ensemble can also be used to increase the ensemble size in EnKF, in particular when small-sized ensembles are used.

The main findings in this study are relevant to other approaches based on Monte Carlo methods that are similar to LEMA, such as the local particle filter (LPF) introduced in Poterjoy (2016). The analysis mosaic in LEMA has similarity to analysis mean obtained by LPF algorithm when the posterior distribution is strongly narrowed down (collapsing of the weights in the algorithm

FIG. 18. Case A. Probability of precipitations errors $p(\text{MAD})$ for the background ensemble (En-6) and the expanded ensemble (Xpd). The MAD is computed with respect to the Stage IV observations (Stage IV/prefix, blue and orange lines) and with respect to every possible members in the ensemble (spread/prefix, red and green lines). The probability is computed from the histogram (using 100 bins) of MAD values over every grid point where $\text{MAD} > 0$ and over all the ensemble members. The thick line denotes the average values over all the truths considered, while the shaded area the variability around the mean.

into a single member). When this collapse occurs, the mosaic of selected members resembles the one obtained in LEMA by selecting the member closest to the observations. In this particular case, the analysis mean produced by the narrow posterior distribution appears

to perform better in a deterministic forecast (Mark Buehner, private communication).

In this work and PZY19, the OSSEs are used to understand how the information contained in ensemble forecasts propagates into state variables and to determine

FIG. 19. Dispersivity computed with respect to the Stage IV observations for cases A to D. For each case the dispersivity is shown for the background ensemble used in the real DA experiments (En-6, red curve), the expanded ensemble considering all the possible candidates (Xpd/Prior, blue curve), the expanded ensemble considering only the members selected by LEMA (Xpd/Posterior, green curve). The dispersivity for the hourly accumulation (in dBR units) is computed as the ratio of the ensemble spread (spr_R) and the RMSE of the ensemble mean (ϵ_R).

the optimal parameters for LEMA. However, the OSSEs can also be used as part of the DA procedure for a flow-dependent optimization of the algorithm's parameters. For example, idealized experiments can be used before the analysis construction to determine the size of the window most adequate for each meteorological situation. This is part of our future work.

Finally, one limitation of LEMA not considered as yet is that all the background ensemble members are relaxed toward the single analysis (the Frankenstate). As we showed in PZY19, the ensemble spread is reduced, resulting in a very underdispersive ensemble at the beginning of the forecast. This limits the cycling of LEMA to long time intervals, of the order of 9 h, at which time the ensemble spread in the forecast is restored. A too short cycling period, Δt , of precipitation assimilation is not necessarily desirable because precipitation slowly decorrelates in time and hence information update is slow (particularly at large scales). The cycling period of 9 h is probably too long. However, it is possible to shorten the cycling period by adding to LEMA another heuristic step: nudging the ensemble toward the Frankenstate while maintaining the original (or adjusted) background's spread. This is part of next steps.

Acknowledgments. The research reported here has been supported by the NSERC/Hydro-Quebec Industrial Research Chair program. We acknowledge the comments and suggestions of three anonymous reviewers for their help in improving the paper.

REFERENCES

- Baldwin, M., and K. Mitchell, 1997: The NCEP hourly multisensor U.S. precipitation analysis. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc.
- 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, <https://doi.org/10.1175/2011BAMS3182.1>.
- Berner, J., K. R. Fossell, S.-Y. Ha, J. P. Hacker, and C. Snyder, 2015: Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. *Mon. Wea. Rev.*, **143**, 1295–1320, <https://doi.org/10.1175/MWR-D-14-00091.1>.
- Bick, T., and Coauthors, 2016: Assimilation of 3D radar reflectivities with an ensemble Kalman filter on the convective scale. *Quart. J. Roy. Meteor. Soc.*, **142**, 1490–1504, <https://doi.org/10.1002/qj.2751>.
- Dalcher, A., E. Kalnay, and R. N. Hoffman, 1988: Medium range lagged average forecasts. *Mon. Wea. Rev.*, **116**, 402–416, [https://doi.org/10.1175/1520-0493\(1988\)116<0402:MRLAF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1988)116<0402:MRLAF>2.0.CO;2).
- Davolio, S., and A. Buzzi, 2004: A nudging scheme for the assimilation of precipitation data into a mesoscale model. *Wea. Forecasting*, **19**, 855–871, [https://doi.org/10.1175/1520-0434\(2004\)019<0855:ANSFTA>2.0.CO;2](https://doi.org/10.1175/1520-0434(2004)019<0855:ANSFTA>2.0.CO;2).
- , F. Silvestro, and T. Gastaldo, 2017: Impact of rainfall assimilation on high-resolution hydrometeorological forecasts over Liguria, Italy. *J. Hydrometeorol.*, **18**, 2659–2680, <https://doi.org/10.1175/JHM-D-17-0073.1>.
- Denis, B., J. Côté, and R. Laprise, 2002: Spectral decomposition of two-dimensional atmospheric fields on limited-area domains using the discrete cosine transform (DCT). *Mon. Wea. Rev.*, **130**, 1812–1829, [https://doi.org/10.1175/1520-0493\(2002\)130<1812:SDOTDA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<1812:SDOTDA>2.0.CO;2).
- Dudhia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107, [https://doi.org/10.1175/1520-0469\(1989\)046<3077:NSOCOD>2.0.CO;2](https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2).
- Durran, D. R., P. A. Reinecke, and J. D. Doyle, 2013: Large-scale errors and mesoscale predictability in Pacific Northwest snowstorms. *J. Atmos. Sci.*, **70**, 1470–1487, <https://doi.org/10.1175/JAS-D-12-0202.1>.
- 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, <https://doi.org/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. Meteor.*, **39**, 1282–1298, [https://doi.org/10.1175/1520-0450\(2000\)039<1282:ANMOOR>2.0.CO;2](https://doi.org/10.1175/1520-0450(2000)039<1282:ANMOOR>2.0.CO;2).
- Fortin, V., M. Abaza, F. Anctil, and R. Turcotte, 2014: Why should ensemble spread match the RMSE of the ensemble mean? *J. Hydrometeorol.*, **15**, 1708–1713, <https://doi.org/10.1175/JHM-D-14-0008.1>.
- Germann, U., and I. Zawadzki, 2002: Scale-dependence of the predictability of precipitation from continental radar images. Part I: Description of the methodology. *Mon. Wea. Rev.*, **130**, 2859–2873, [https://doi.org/10.1175/1520-0493\(2002\)130<2859:SDOTPO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2002)130<2859:SDOTPO>2.0.CO;2).
- Hamill, T. M., 2006: Ensemble-based atmospheric data assimilation. *Predictability of Weather and Climate*, T. Palmer and R. Hagedorn, Eds., Cambridge University Press, 124–156.
- Hoffman, R. N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus*, **35A**, 100–118, <https://doi.org/10.1111/j.1600-0870.1983.tb00189.x>.
- Hong, S.-Y., and J.-O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42**, 129–151.
- , Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, <https://doi.org/10.1175/MWR3199.1>.
- Jacques, D., D. Michelson, J.-F. Caron, and L. Fillion, 2018: Latent heat nudging in the Canadian regional deterministic prediction system. *Mon. Wea. Rev.*, **146**, 3995–4014, <https://doi.org/10.1175/MWR-D-18-0118.1>.
- Kain, J. S., 2004: The Kain–Fritsch convective parameterization: An update. *J. Appl. Meteor.*, **43**, 170–181, [https://doi.org/10.1175/1520-0450\(2004\)043<0170:TKCPAU>2.0.CO;2](https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2).
- 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, <https://doi.org/10.2151/SOLA.2005-013>.
- Kotsuki, S., T. Miyoshi, K. Terasaki, G.-Y. Lien, and E. Kalnay, 2017: Assimilating the global satellite mapping of precipitation data with the nonhydrostatic icosahedral atmospheric model (NICAM). *J. Geophys. Res. Atmos.*, **122**, 631–650, <https://doi.org/10.1002/2016JD025355>.

- Kumar, P., C. M. Kishtawal, and P. K. Pal, 2014: Impact of satellite rainfall assimilation on weather research and forecasting model predictions over the Indian region. *J. Geophys. Res. Atmos.*, **119**, 2017–2031, <https://doi.org/10.1002/2013JD020005>.
- Lien, G.-Y., E. Kalnay, and T. Miyoshi, 2013: Effective assimilation of global precipitation: Simulation experiments. *Tellus*, **65A**, 19915, <https://doi.org/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, <https://doi.org/10.1175/MWR-D-15-0149.1>.
- Lin, Y., and K. E. Mitchell, 2005: The NCEP stage II/IV hourly precipitation analyses: Development and applications. *19th Conf. on Hydrology*, San Diego, CA, Amer. Meteor. Soc., 1.2, https://ams.confex.com/ams/Annual2005/techprogram/paper_83847.htm.
- Lopez, P., 2011: Direct 4D-Var assimilation of NCEP stage IV radar and gauge precipitation data at ECMWF. *Mon. Wea. Rev.*, **139**, 2098–2116, <https://doi.org/10.1175/2010MWR3565.1>.
- , and P. Bauer, 2007: “1D+4DVAR” assimilation of NCEP stage-IV radar and gauge hourly precipitation data at ECMWF. *Mon. Wea. Rev.*, **135**, 2506–2524, <https://doi.org/10.1175/MWR3409.1>.
- Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177–1194, <https://doi.org/10.1002/qj.49711247414>.
- Lu, C., H. Yuan, B. E. Schwartz, and S. G. Benjamin, 2007: Short-range numerical weather prediction using time-lagged ensembles. *Wea. Forecasting*, **22**, 580–595, <https://doi.org/10.1175/WAF999.1>.
- Macpherson, B., 2001: Operational experience with assimilation of rainfall data in the met office mesoscale model. *Meteor. Atmos. Phys.*, **76**, 3–8, <https://doi.org/10.1007/s007030170035>.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-K model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, <https://doi.org/10.1029/97JD00237>.
- Orlanski, I., 1975: A rational subdivision of scales for atmospheric processes. *Bull. Amer. Meteor. Soc.*, **56**, 527–530.
- Pérez Hortal, A. A., I. Zawadzki, and M. K. Yau, 2019: A heuristic approach for precipitation data assimilation: Characterization using OSSEs. *Mon. Wea. Rev.*, **147**, 3445–3466, <https://doi.org/10.1175/MWR-D-19-0034.1>.
- Poterjoy, J., 2016: A localized particle filter for high-dimensional nonlinear systems. *Mon. Wea. Rev.*, **144**, 59–76, <https://doi.org/10.1175/MWR-D-15-0163.1>.
- , R. A. Sobash, and J. L. Anderson, 2017: Convective-scale data assimilation for the weather research and forecasting model using the local particle filter. *Mon. Wea. Rev.*, **145**, 1897–1918, <https://doi.org/10.1175/MWR-D-16-0298.1>.
- , L. Wicker, and M. Buehner, 2019: Progress toward the application of a localized particle filter for numerical weather prediction. *Mon. Wea. Rev.*, **147**, 1107–1126, <https://doi.org/10.1175/MWR-D-17-0344.1>.
- Skamarock, W. C., and J. B. Klemp, 2008: A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *J. Comput. Phys.*, **227**, 3465–3485, <https://doi.org/10.1016/j.jcp.2007.01.037>.
- Surcel, M., I. Zawadzki, and M. K. Yau, 2015: A study on the scale dependence of the predictability of precipitation patterns. *J. Atmos. Sci.*, **72**, 216–235, <https://doi.org/10.1175/JAS-D-14-0071.1>.
- Van Den Dool, H. M., and L. Rukhovets, 1994: On the weights for an ensemble-averaged 6–10-day forecast. *Wea. Forecasting*, **9**, 457–465, [https://doi.org/10.1175/1520-0434\(1994\)009<0457:OTWFAE>2.0.CO;2](https://doi.org/10.1175/1520-0434(1994)009<0457:OTWFAE>2.0.CO;2).
- van Leeuwen, P. J., 2009: Particle filtering in geophysical systems. *Mon. Wea. Rev.*, **137**, 4089–4114, <https://doi.org/10.1175/2009MWR2835.1>.
- Zhou, X., Y. Zhu, D. Hou, Y. Luo, J. Peng, and R. Wobus, 2017: Performance of the new NCEP global ensemble forecast system in a parallel experiment. *Wea. Forecasting*, **32**, 1989–2004, <https://doi.org/10.1175/WAF-D-17-0023.1>.