Enhanced Adaptive Inflation Algorithm for Ensemble Filters

MOHAMAD EL GHARAMTI

National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 27 June 2017, in final form 2 January 2018)

ABSTRACT

Spatially and temporally varying adaptive inflation algorithms have been developed to combat the loss of variance during the forecast due to various model and sampling errors. The adaptive Bayesian scheme of Anderson uses available observations to update the Gaussian inflation distribution assigned for every state variable. The likelihood function of the inflation is computed using model-minus-data innovation statistics. A number of enhancements for this inflation scheme are proposed. To prevent excessive deflation, an inverse gamma distribution for the prior inflation is considered. A non-Gaussian distribution offers a flexible framework for the inflation variance to evolve during the update. The innovations are assumed random variables, and a correction term is added to the mode of the likelihood distribution such that the observed inflation is slightly larger. This modification improves the stability of the adaptive scheme by limiting the occurrence of negative and physically intolerable inflations. The enhanced scheme is compared to the original one in twin experiments using the Lorenz-63 model, the Lorenz-96 model, and an idealized, high-dimensional atmospheric model. Results show that the proposed enhancements are capable of generating accurate and consistent state estimates. Allowing moderate deflation is shown to be useful.

1. Introduction

The ensemble Kalman filter (EnKF) is an efficient estimation tool that has been used extensively in many geophysical applications during the last two decades (Evensen 2009). It operates sequentially by combining information from a prior state of a system and its associated variance with noisy observations to produce an analyzed posterior state with an updated uncertainty (Evensen 2003). Inflation has been introduced into stochastic (e.g., EnKF) and deterministic (e.g., Anderson 2001; Hoteit et al. 2002; Whitaker and Hamill 2002; Hunt et al. 2007; Sakov and Oke 2008) ensemble filters as a way to counteract the underestimation of the true variance. Large portions of the variance are often lost during the forecast step, but the inherent bias of the EnKF, in the form of sampling errors, is in the analysis step (Bocquet 2011; van Leeuwen 1999). A separate issue is whether there are sufficient members to track all the directions of uncertainty growth (Bocquet and Carrassi 2017). Other reasons for variance underestimation can be related to the simplified model dynamics and wrong parameterizations that may introduce large biases in addition to neglected model errors. Systematic observational and representativeness errors within the analysis step are also expected to lead to underestimates of the ensemble variance (Furrer and Bengtsson 2007).

Studies within the data assimilation (DA) literature have proposed a number of techniques that deal with all kinds of system errors. The majority of these techniques, apart from localization, which removes spurious correlations from the prior covariance, can be seen as a form of inflation. These can be split into four distinct categories. The first category is background covariance inflation, which itself is divided into additive and multiplicative inflation. Additive inflation (Mitchell and Houtekamer 2000) adds a random perturbation drawn from a specific error distribution to each ensemble member during the forecast step. Additive inflation can be applied to the prior ensemble in different forms (Raanes et al. 2015) and also within the update step (Hoteit et al. 2015). Multiplicative inflation (Anderson and Anderson 1999; Zheng 2009), on the other hand, increases the prior ensemble spread by a specific factor while pushing all members away from the mean. It has been reported that multiplicative inflation can help deal with observation network-related sampling errors, while additive inflation can help mitigate the modeling errors (Whitaker and Hamill 2012). In the current study, we only look into multiplicative inflation and its

Corresponding author: Mohamad El Gharamti, gharamti@ucar.edu

DOI: 10.1175/MWR-D-17-0187.1

spatial and temporal adaptive formulations. Observation error variance inflation falls within the second category. The idea behind inflating the observation error variance is to compensate for a wrongly specified error distribution on the observations (Liang et al. 2012) and, in other scenarios, to limit the impact of the Kalman update. Such an inflation procedure, which is often referred to as observation error moderation, has been applied in a number of ocean (e.g., Sakov et al. 2012; Karspeck 2016) and atmosphere (e.g., Li et al. 2009; Minamide and Zhang 2017) DA studies. The third inflation category includes the relaxation-to-prior perturbations (RTPP) (Zhang et al. 2004), and spread (RTPS) (Whitaker and Hamill 2012; Ying and Zhang 2015). The core of the relaxation method can be seen as a multiplicative inflation; however, it is applied only to the posterior ensemble, given the amount of ensemble spread reduction by the observations. In a recent study, Kotsuki et al. (2017) found that the adaptive versions of the RTPP and RTPS often lead to overdispersive (underdispersive) ensemble in sparse (dense) observation networks when compared to multiplicative prior inflation. The sampling error correction by Anderson (2012) can also be seen as a relaxation, in some sense. The fourth inflation category includes techniques that modify the ensemble by changing its size (Uzunoglu et al. 2007) or its physical and model-based nature (e.g., Meng and Zhang 2007; Berner et al. 2009). In addition to these four inflation categories, a few studies have focused on developing ensemble-based filters that do not require an explicit use of inflation. These methods, such as the ensemble time-local H_∞ filter (EnTLHF; Luo and Hoteit 2011) and the finite-size EnKF (EnKF-N; Bocquet 2011), have some sort of a built-in inflation by design.

Anderson (2009, hereafter A09) proposed an adaptive (multiplicative) inflation algorithm that is fully based on a Bayesian approach. The algorithm deals with inflation as a spatially heterogeneous random variable following a multivariate Gaussian distribution. As we will see in the next sections, the inflation likelihood—that is not Gaussian in terms of the inflation—often peaks at values that are smaller than 1 (sometimes negative). This may cause a deflation of the ensemble, producing poor estimates, and may possibly yield filter divergence over a long assimilation run. A common strategy for countering excessive deflation (e.g., Wang et al. 2007) is to bound the inflation by rejecting values that are smaller than 1. Here, we derive a new form for the likelihood function. The new form is similar to the one proposed in A09, with an additional correction term. The correction term helps reduce the probability for deflation, particularly for small ensemble sizes. Moreover, a Gaussian inflation prior does not respect the positivity constraint, such that drawing a negative inflation value is possible. As a more suitable alternative, Raanes (2016) recently proposed the use of an inverse chi-square distribution for the inflation prior. In this study, we use an inverse gamma distribution for the prior, and we present an efficient computational procedure to switch back and forth between Gaussian and inverse gamma probability density functions (pdfs).

Section 2 briefly introduces the adaptive inflation algorithm and also provides a useful geometrical interpretation for the problem. Section 3 discusses the enhancements to the prior and likelihood distributions. Section 4 presents results from two low-order models and an idealized 3D atmospheric system. Results from A09 are compared to the ones using the proposed enhanced inflation algorithm. Section 5 provides the discussion and conclusions.

2. Adaptive inflation review

a. Definitions, background, and innovation statistics

Consider a state vector $\mathbf{x}$ of size N_x, representing a 3D geophysical field such as temperature. In the ensemble framework, the probability density function of the model's state at any forecast time is approximated by an ensemble of realizations. For any kth element x^k, in the state vector, the first and second moments are approximated as

$$\overline{x}_b = \frac{1}{N_e}\sum_{i=1}^{N_e} x_i^k, \tag{1}$$

$$\widehat{\sigma_b}^2 = \frac{1}{N_e-1}\sum_{i=1}^{N_e}(x_i^k-\overline{x}_b)^2 = \frac{1}{N_e-1}\sum_{i=1}^{N_e}\widetilde{x}_i^2. \tag{2}$$

Here, $\overline{x}_b$ denotes the background ensemble mean, $\widetilde{x}_i$ is the ith background ensemble perturbation, N_e is the ensemble size, and $\widehat{\sigma_b}^2$ is the associated sample variance. Suppose that the true value of this state element is x_t and that an observation y is available through

$$y = x_t + \varepsilon_o, \tag{3}$$

where ε_o is the observation error, assumed Gaussian with zero mean and variance σ_o^2. From the definition of the innovation (Desroziers et al. 2005), the following relations can be derived:

$$d = y - \overline{x}_b = \varepsilon_o + (x_t - \overline{x}_b) = \varepsilon_o + \varepsilon_b, \tag{4}$$

$$\mathbb{E}(d) = \mathbb{E}(\varepsilon_o) + \mathbb{E}(\varepsilon_b) = 0, \tag{5}$$

$$\mathbb{E}(d^2) = \mathbb{E}(\varepsilon_o^2) + \mathbb{E}(\varepsilon_b^2) + 2\mathbb{E}(\varepsilon_o\varepsilon_b), \tag{6}$$

$$= \sigma_o^2 + \sigma_b^2, \tag{7}$$

where ε_b is the background error following $N(0, \sigma_b^2)$, and σ_b^2 is the true background variance; d is an

observation-minus-background-type innovation. Equation (7) assumes that the observation and background errors are uncorrelated (Lorenc 1986; Daley 1993). We assume that the ensemble mean $\overline{x}_b$ coincides with the true mean that one may obtain using a very large ensemble size.

The idea is to match the true variance by modifying the sample variance from the ensemble. One possible way is to impose the linear relation $\sigma_b^2 = \lambda_o \widehat{\sigma_b}^2$, and, assuming a correctly specified observational error variance, the following result can be obtained from Eq. (7):

$$\lambda_o = \frac{\mathbb{E}(d^2) - \sigma_o^2}{\widehat{\sigma_b}^2}, \tag{8}$$

where λ_o is a variance scaling factor, or inflation factor (Anderson 2007). This is a very well-known result that has been used in many atmospheric and oceanographic DA studies (e.g., Wang and Bishop 2003; Wang et al. 2007; Li et al. 2009; Miyoshi 2011; Karspeck et al. 2013). Given the limited ensemble size, $\widehat{\sigma_b}^2$ is often underestimated, and, thus, $\lambda_o > 1$. In cases where $\mathbb{E}(d^2) < \sigma_o^2$, the scaling becomes negative, and Eq. (8) is no longer useful. This often happens when the observation is highly uncertain or when it is very close to ensemble mean. Furthermore, if the numerator in Eq. (8) is smaller than the background sample variance, λ_o becomes smaller than 1, and, thus, a shrinkage or deflation of the ensemble spread is expected.

b. Geometrical interpretation

We have seen that in an error-free (ideal) system, Eq. (7) must be fulfilled. Given that the background and the observational errors are uncorrelated, their projections on a 2D plane are orthogonal. The orthogonality of ε_o and ε_b is only true from a statistical point view; that is, $\langle [\varepsilon_b], [\varepsilon_o] \rangle = 0 = \mathbb{E}(\varepsilon_b \varepsilon_o)$, in which $[\varepsilon_b]$ and $[\varepsilon_o]$ are two random background and observation errors, respectively. Assuming that the background error is projected on the positive x axis and that the observation error is projected on the positive y axis, there exists a point, say A, with coordinates (σ_b, σ_o) that lies on the circumference of a circle with radius $\sqrt{\mathbb{E}(d^2)}$ and origin 0 [reader may refer to Fig. 1 in Desroziers et al. (2005)]. When prior inflation is applied, Eq. (8) can be rearranged to obtain

$$\frac{\sigma_o^2}{\left[\sqrt{\mathbb{E}(d^2)}\right]^2} + \frac{\widehat{\sigma_b}^2}{\left[\sqrt{\mathbb{E}(d^2)}\Big/\sqrt{\lambda_o}\right]^2} = 1, \tag{9}$$

such that a new point, say B, with coordinates $(\widehat{\sigma_b}, \sigma_o)$ can be shown to live on an ellipse with origin 0, semimajor axis $\sqrt{\mathbb{E}(d^2)}$, and semiminor axis $\sqrt{\mathbb{E}(d^2)}/\sqrt{\lambda_o}$ (if $\lambda_o > 1$). Recall here that point A represents the truth, or the ideal system statistics, and point B is what we have from the ensemble and the imperfect observations. The eccentricity $0 < e < 1$ in Eq. (10) can be used as a measure of how much our elliptic shape deviates from being a circle[1]:

$$e = \sqrt{1 - \lambda_o^{-1}}. \tag{10}$$

Thus, in the worst case scenario, that is, an ensemble collapse with $\widehat{\sigma_b} \to 0$ ($\lambda_o \to \infty$), the eccentricity becomes 1. On the contrary, as e approaches 0, we are close to a circle situation and, hence, almost recovering the true background variance. In practice, e can be utilized as an elegant bounded parameter to comment on the status of the background ensemble, whether it is slightly ($e \to 0$) or severely ($e \to 1$) underestimated. In Fig. 1, a geometrical illustration of the inflation is shown. As can be seen, the background ensemble poorly delineates the true model forecast variance, with an eccentricity of 0.87. It is further interesting to note that in case of

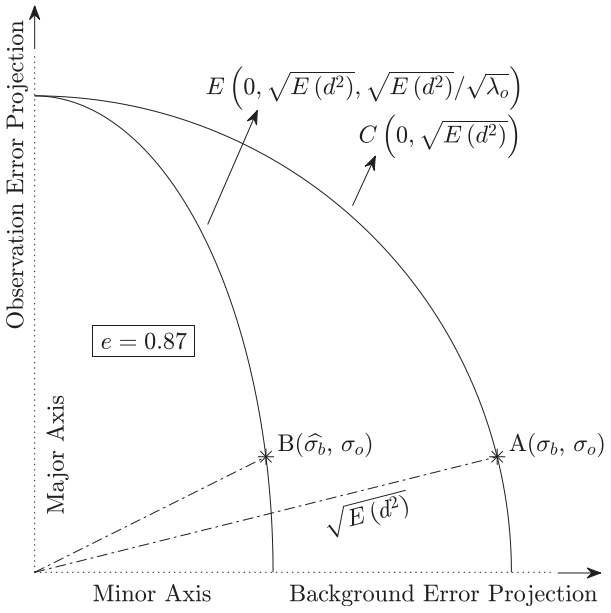

FIG. 1. Geometrical representation of the prior multiplicative inflation. The circle and point A describe the true background and observation error variances. Point B and its associated ellipse describe the statistics given by the ensemble. The parameters for the ellipse and the circle are displayed at the top of the graph. The horizontal axis represents the projections of the background errors, while the vertical axis represents those of the observation errors. The eccentricity e of the ellipse is computed based on Eq. (10).

[1] The eccentricity of an ellipse is measured as the ratio of the distance between the center of the ellipse and each focus to the length of the semimajor axis.

deflation, the major and minor axes are switched, and the eccentricity equation becomes $e = \sqrt{1-\lambda_o}$.

c. Anderson's adaptive inflation

The spatially and temporally adaptive inflation algorithm outlined by A09 is based on a Bayesian formulation, such that at any analysis time,

$$p(\lambda^k|d) \propto p(d|\lambda^k)\, p(\lambda^k), \tag{11}$$

where $p(\lambda^k)$ is the inflation prior for the kth inflation element, $p(d|\lambda^k)$ is the inflation likelihood given the discrepancy between the observation and the background, and $p(\lambda^k|d)$ is the posterior inflation density. For brevity, the time indexing has been removed. Here, the inflation is considered a random variable, and a different inflation value is assigned for every grid point in the domain, that is,

$$\boldsymbol{\lambda} = [\lambda^1, \ldots, \lambda^k, \ldots, \lambda^{N_x}]. \tag{12}$$

Equation (11) and the majority of the derivations in this paper are written, without loss of generality, in a scalar form for simplicity and to allow for an efficient serial processing of the observations (Anderson 2003). The prior marginal distribution for this kth inflation element is assumed Gaussian, with mean λ_b^k and variance $\sigma^2_{\lambda_b,k}$. For the likelihood to be unbiased (Dee 1995), the innovation is assumed to be drawn from a normal distribution with zero mean and variance $\theta^2 = \lambda_o^k \widehat{\sigma_b}^2 + \sigma_o^2$. As a way to spread the information spatially across all variables, A09 proposed equal correlation structure between the state variables and the inflation, such that

$$r = \mathrm{corr}(x^o, x^k) \quad k = 1, 2, \ldots, N_x, \tag{13}$$

$$\lambda_o^k = \left[\gamma(\sqrt{\lambda^k} - 1) + 1\right]^2. \tag{14}$$

Here, x^o refers to the observed variable. The notation x^o assumes an identity observational operator to simplify the equations. Parameter $\gamma = \kappa|r|$ and κ is a localization factor applied to limit the impact of observation on the update of state variable x^k. Accordingly, the posterior distribution in Eq. (11) becomes

$$p(\lambda^k|d) \propto \frac{1}{2\pi\sigma_{\lambda_b,k}\theta} \exp\left[-\frac{(\lambda^k - \lambda_b^k)^2}{2\sigma^2_{\lambda_b,k}} - \frac{d^2}{2\theta^2}\right]. \tag{15}$$

[This is Eq. (12) in A09.] The updated inflation value (i.e., λ_u^k) is obtained as the mode of $p(\lambda^k|d)$. This is done by first approximating the likelihood with a Taylor expansion and keeping only on the linear terms. The derivative of the resulting approximate posterior is then set to zero, leading to a quadratic equation in λ^k. Two solutions are obtained, and the one that is closest to the prior is selected. The reader may refer to appendix A in A09 for a detailed computation of the posterior inflation and its associated variance.

3. Enhanced adaptive inflation algorithm

a. The likelihood

In practice, studies have shown that the result of Eq. (8) is not always positive. This often happens because of a misspecification of the observational error variance [note that Eq. (8) is derived based on the assumption that σ_o^2 is true]. In fact, when $\lambda_o < 0$, this is an indication that the components $\sqrt{\mathbb{E}(d^2)}$, $\sqrt{\lambda_o}\widehat{\sigma_b}$, and σ_o cannot combine to form a triangle. Recall that for these sides to form a triangle, the magnitude of the sum of any two sides should be larger than the remaining third side. One way to satisfy such a triangle relation is to scale the observation error variance in addition to that of the background. For instance, Li et al. (2009) used different innovation statistics (Desroziers et al. 2005) to estimate both σ_b^2 and σ_o^2. On the other hand, Wang et al. (2007) approximated $\mathbb{E}(d^2)$ by averaging the innovations obtained over short periods of time. Such a time average often yields values that are closer to the expectation.

In the following, we propose a different form for Eq. (8). The derivation takes into account the entire prior ensemble and not just the mean when calculating the innovation statistics. A sample distance (innovation) can be expanded in terms of the underlying observation and background errors and ensemble perturbations as follows:

$$d_i = y - x_i^k = \varepsilon_o + x^t - (\overline{x}_b + x_i^k - \overline{x}_b), \tag{16}$$

$$= \varepsilon_o + (x^t - \overline{x}_b) - (x_i^k - \overline{x}_b), \tag{17}$$

$$= \varepsilon_o + \varepsilon_b - \widetilde{x}_i. \tag{18}$$

Squaring both sides leads to

$$d_i^2 = (\varepsilon_o)^2 + (\varepsilon_b)^2 + \widetilde{x}_i^2 + 2\varepsilon_o\varepsilon_b - 2\varepsilon_o\widetilde{x}_i - 2\varepsilon_b\widetilde{x}_i. \tag{19}$$

Now, average over all ensemble members, and further simplify:

$$\frac{1}{N_e}\sum_{i=1}^{N_e} d_i^2 = \frac{1}{N_e}\sum_{i=1}^{N_e}[(\varepsilon_o)^2 + (\varepsilon_b)^2 + \widetilde{x}_i^2 + 2\varepsilon_o\varepsilon_b - 2\varepsilon_o\widetilde{x}_i - 2\varepsilon_b\widetilde{x}_i], \tag{20}$$

$$= (\varepsilon_o)^2 + (\varepsilon_b)^2 + \frac{N_e - 1}{N_e}\widehat{\sigma_b}^2 + 2\varepsilon_o\varepsilon_b - \frac{2\varepsilon_o}{N_e}\sum_{i=1}^{N_e}\widetilde{x}_i - \frac{2\varepsilon_b}{N_e}\sum_{i=1}^{N_e}\widetilde{x}_i, \tag{21}$$

$$=(\varepsilon_o)^2+(\varepsilon_b)^2+\frac{N_e-1}{N_e}\widehat{\sigma_b}^2+2\varepsilon_o\varepsilon_b, \tag{22}$$

noting that the average of the ensemble perturbations is zero. Further, we take the expectation of both sides and impose a biased assumption of the ensemble innovations.[2] The equation form can then be simplified as follows:

$$\mathbb{E}(d^2)+\widehat{\sigma_b}^2=\sigma_o^2+\sigma_b^2+\frac{N_e-1}{N_e}\widehat{\sigma_b}^2, \tag{23}$$

where the variance of the innovations ensemble is equal to that of the prior. Introducing a new inflation factor λ_o^* and substituting $\sigma_b^2=\lambda_o^*\widehat{\sigma_b}^2$, we obtain

$$\lambda_o^*\widehat{\sigma_b}^2-\frac{1}{N_e}\widehat{\sigma_b}^2=\mathbb{E}(d^2)-\sigma_o^2, \tag{24}$$

$$\lambda_o^*=\frac{\mathbb{E}(d^2)-\sigma_o^2}{\widehat{\sigma_b}^2}+\frac{1}{N_e}=\lambda_o+\frac{1}{N_e}. \tag{25}$$

The constant $1/N_e$ can be thought of as a correction term to the original scaling factor in Eq. (8). Using Eq. (25), the peak for the likelihood function is shifted slightly to the right, as in Eq. (26). This behavior is pronounced when a very small ensemble size is used. Note that this term does not always guarantee a positive value; however, intuitively, it does reduce the occurrence of deflation and negative values:

$$\theta^2=\left\{\left[\gamma\left(\sqrt{\lambda^k}-1\right)+1\right]^2-\frac{1}{N_e}\right\}\widehat{\sigma_b}^2+\sigma_o^2. \tag{26}$$

b. The prior

Instead of a Gaussian, we will explore representing the inflation prior by an inverse gamma (IG) distribution (Stroud and Bengtsson 2007). Such a choice is important and can be useful in two ways. First, it would prevent negative and very tiny inflation values that often deteriorate the quality of the forecasts. Second, from a coding and computational perspective, this ensures a clean algorithm without lengthy checks on the sign and magnitude of the inflation factor.

The IG prior distribution for the kth inflation element (for brevity, we have dropped the superscript k) can be written as follows:

$$p(\lambda)=\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{-\alpha-1}\exp\left(-\frac{\beta}{\lambda}\right), \tag{27}$$

where Γ is the gamma function. The constants α and β are shape and rate parameters of the distribution, respectively. To apply Bayes's rule, as in Eq. (11), we first need to specify both α and β. This is not a straightforward exercise, considering that these two parameters are quite different from the Gaussian mean and variance. To make things easier, we assume that a Gaussian inflation prior is available, and all we have to do is to use the parameters of the Gaussian to try and find an appropriate α and β. To this end, we first use the mean (mode) of the Gaussian and equate it with the mode of the IG as follows:

$$\lambda_b=\frac{\beta}{\alpha+1}. \tag{28}$$

Next, the variance of the Gaussian is assumed equal to that of the IG. Then, by using Eq. (28), we get the following relation:

$$\sigma_{\lambda_b}^2=\frac{\beta^2}{(\alpha-1)^2(\alpha-2)},\quad \alpha>2, \tag{29}$$

$$=\frac{\beta^2}{(\beta/\lambda_b-2)^2(\beta/\lambda_b-3)}, \tag{30}$$

where Eq. (29) is the IG variance formula. After rearranging and simplifying Eq. (30), a cubic equation in β is obtained:

$$\frac{1}{\lambda_b^3}\beta^3-\left(\frac{7}{\lambda_b}+\frac{1}{\sigma_{\lambda_b}^2}\right)\beta^2+\frac{16}{\lambda_b}\beta-12=0. \tag{31}$$

To find its roots, we first compute the discriminant Δ, as shown below:

$$\Delta=-\frac{16(\sigma_{\lambda_b}^4+47\sigma_{\lambda_b}^2\lambda_b^2+3\lambda_b^4)}{\sigma_{\lambda_b}^6\lambda_b^4}. \tag{32}$$

Because $\Delta<0$, Eq. (31) has one real solution and two nonreal, complex conjugate roots. The real solution is selected and assigned to the rate parameter:

$$\overline{a}=\frac{\lambda_b^4}{\sigma_{\lambda_b}^3}\sqrt{\sigma_{\lambda_b}^4+47\sigma_{\lambda_b}^2\lambda_b^2+3\lambda_b^4}, \tag{33}$$

$$\overline{b}=\frac{1}{\sigma_{\lambda_b}^2}\sqrt[3]{\lambda_b^9+6\sqrt{3}\sigma_{\lambda_b}^6\overline{a}+75\sigma_{\lambda_b}^4\lambda_b^5+\sigma_{\lambda_b}^6\lambda_b^3+21\sigma_{\lambda_b}^2\lambda_b^7}, \tag{34}$$

$$\beta=\frac{\overline{b}}{3}+\frac{\lambda_b^3+7\sigma_{\lambda_b}^2\lambda_b}{3\sigma_{\lambda_b}^2}+\frac{\lambda_b^2(\sigma_{\lambda_b}^4+14\sigma_{\lambda_b}^2\lambda_b^2+\lambda_b^4)}{3\sigma_{\lambda_b}^4\overline{b}}. \tag{35}$$

[2] The left-hand side of Eq. (22) is approximated as $1/N_e\sum_{i=1}^{N_e}d_i^2\approx\left(1/N_e\sum_{i=1}^{N_e}d_i\right)^2+\mathbb{V}(d)=\mathbb{E}(d^2)+\widehat{\sigma_b}^2$, where $\mathbb{V}(d)$ denotes the variance of the innovation sample. In other words, the constant $N_e/(N_e-1)$ in front of $\widehat{\sigma_b}^2$ is omitted. This, generally, becomes a valid assumption when the ensemble size is large.

Consequently, the shape parameter can then be deduced using Eqs. (35) and (28). As an alternative, and instead of equating the mean of the Gaussian prior with the model of the IG distribution, one may choose to set the means of both distributions to be equal. This yields significantly simpler relations; however, because the posterior is approximated through the mode, as we will see next, this might, therefore, be unsuitable.

c. The posterior

Using the likelihood and the prior from the previous sections, the new posterior distribution $p(\lambda|d)$ takes the form

$$\frac{\beta^{\alpha}\lambda^{-\alpha-1}}{\sqrt{2\pi}\theta\Gamma(\alpha)}\exp\left(-\frac{d^2}{2\theta^2}-\frac{\beta}{\lambda}\right), \tag{36}$$

where both α and β are given as functions of λ_b and $\sigma^2_{\lambda_b}$. The variance of the likelihood θ^2 is given by Eq. (26) as a function of N_e, λ, γ, $\widehat{\sigma_b}^2$, and σ_o^2. To find the updated inflation or the mode of this distribution (i.e., λ_u), we follow A09 and first expand the likelihood density around the prior inflation λ_b using a Taylor expansion:

$$p(d|\lambda) \cong \underbrace{p(d|\lambda_b)}_{\ell} + \underbrace{\left.\frac{\partial p(d|\lambda)}{\partial\lambda}\right|_{\lambda_b}}_{\ell'}(\lambda-\lambda_b) + O(\lambda-\lambda_b)^2. \tag{37}$$

This approximate likelihood is then multiplied with the IG prior in Eq. (27). We take the derivative of the resulting function with respect to λ and obtain the following quadratic equation:

$$\left(1-\frac{\lambda_b}{\beta}\right)\lambda^2+\left(\frac{\overline{\ell}}{\ell'}-2\lambda_b\right)\lambda+\left(\lambda_b^2-\frac{\overline{\ell}}{\ell'}\lambda_b\right)=0. \tag{38}$$

Equation (38) has two real roots, and the one closest to λ_b is selected.

To find the updated rate and shape parameters (i.e., β_u and α_u), we first evaluate the ratio R of the posterior in Eq. (36) at point $\xi_1=\lambda_u+\sigma_{\lambda_b}$ to that at point $\xi_2=\lambda_u$ as follows:

$$R=\frac{p(\lambda|d)|_{\lambda=\xi_1}}{p(\lambda|d)|_{\lambda=\xi_2}}. \tag{39}$$

We note that the posterior in Eq. (36) is not exactly IG; however, from a series of experiments, we found out that it does behave like an IG distribution. Based on this, if one imposes an IG assumption of the posterior, R, which has already been obtained in Eq. (39), would be equivalent to

$$R\equiv\frac{\beta_u\Gamma(\alpha_u)^{-1}\xi_1^{-\alpha_u-1}\exp(-\beta_u/\xi_1)}{\beta_u\Gamma(\alpha_u)^{-1}\xi_2^{-\alpha_u-1}\exp(-\beta_u/\xi_2)}=\frac{\xi_2^{\alpha_u+1}\exp(\beta_u/\xi_2)}{\xi_1^{\alpha_u+1}\exp(\beta_u/\xi_2)}. \tag{40}$$

Using Eq. (28) for the mode, we substitute $\alpha+1=\beta_u/\lambda_u$, and after taking the logarithm of both sides, we get

$$\beta_u=\omega^{-1}\log(R), \tag{41}$$

$$\alpha_u=\frac{\beta_u}{\lambda_u}-1, \tag{42}$$

where $\omega=\xi_2^{-1}\log(\xi_2/\xi_1)+\xi_2^{-1}-\xi_1^{-1}$.

We can also find the associated Gaussian distribution, given the mean λ_u and a variance $\sigma^2_{\lambda_u}$, as in Eq. (43) below. This provides a simple framework to switch between the Gaussian and the IG distributions for the inflation:

$$\sigma^2_{\lambda_u}=\frac{\beta_u^2}{(\alpha_u-1)^2(\alpha_u-2)}. \tag{43}$$

Under the non-Gaussian assumption, the variance of the IG posterior is self-adapting, meaning that it can both increase and decrease after the update. This is empirically shown to prevent collapse in the spread of the inflation pdf.

d. Implementation

The enhanced adaptive inflation algorithm is implemented in the Data Assimilation Research Testbed (DART) at NCAR (www.image.ucar.edu/DAReS/DART/; Anderson et al. 2009). Listed below are some characteristics about the algorithm and its implementation procedure.

- The algorithm is coded in such a way that the user only deals with Gaussian inflation distributions. The motivation for this is to facilitate the choice of the initial inflation parameters and impose only marginal changes to old versions of DART. The algorithm operates by first transforming the Gaussian input into an inverse gamma, and then it applies the Bayesian update. The outputs are the corresponding Gaussian mean and variance.
- For computational reasons, and as outlined in A09, one may choose to keep the inflation variance unchanged during the update. One can also refrain from letting the variance increase in time and instead only accept updates that decrease the inflation variance.
- The algorithm, by construction, does not require enforcing the inflation to be nonnegative, given the IG assumption. This further simplifies the code and the list of the input parameters.
- The algorithm also features a damped inflation version to avoid overdispersion. The damping (Anderson et al. 2009),

which takes place right before applying $\sqrt{\lambda}$ to the state, is done according to $\sqrt{\lambda_b} = 1 + \rho(\sqrt{\lambda_u} - 1)$, where ρ is a user-defined parameter, often close to 1.

4. Numerical experiments

The enhanced spatially and temporally varying adaptive inflation scheme is compared to the algorithm of A09. The comparison is performed using (i) the Lorenz-63 (L63; Lorenz 1963), (ii) the 40-variable model of Lorenz (L96; Lorenz 1996), and (iii) a low-resolution atmospheric GCM with more than 10^5 variables. All experiments use the ensemble adjustment Kalman filter (EAKF; Anderson 2003). The Kalman update is done serially using DART. All experiments use damped adaptive inflation versions of A09 and the proposed scheme ($\rho = 0.9$). The root-mean-square (RMS) errors and average ensemble spread (AES) metrics are used to comment on the accuracy of the estimates. For the L63 model, the ensemble size is set to 5. The experiments with the two other models use 20 members.

a. Lorenz 63

For this model (see the appendix), we assume perfect modeling conditions, and we refrain from using localization. The goal is to study how well the inflation schemes can mitigate sampling errors in a strongly nonlinear regime and using a very small ensemble size. To this end, we also include results from the EnKF-N (Raanes 2016) as a benchmark. The EnKF-N accounts for errors in the prior ensemble mean and its associated covariance and is expected to outperform an EnKF with optimally tuned inflation (Bocquet 2011). Given its computational efficiency, only the dual formulation of the EnKF-N is implemented. The initial state of the model, consisting of three variables, $\{x, y, z\}$, is set to [20, 28, 8/3], with an observational error variance of $\sigma_o^2 = 2$ [following Harlim and Hunt (2007)]. A reference (truth) run is generated for 10 000 time steps, from which pseudo-observations are sampled every 10 model steps (i.e., $\Delta t = 0.1$). We conduct three assimilation experiments by changing the observation network. For the adaptive inflation schemes, the inflation lower bound is set to 1, and the inflation standard deviation is kept fixed in time and equal to 0.1. We plot in Fig. 2 the time-averaged forecast RMSE values obtained using the three schemes, evaluated only over the last 5000 steps. As can be seen, the EnKF-N outperforms the adaptive inflation schemes for all three observational scenarios. When only one variable is observed (here, x), the difference between the performances is only marginal. When two or three model variables are observed, the enhanced scheme demonstrates a better mitigation of sampling errors and suggests an average of 12% improvement in prediction accuracy over A09. With increasing nonlinearity (i.e., using observation frequencies such as $\Delta t = 0.25$ and $\Delta t = 0.50$), the performance (not shown) of the enhanced scheme was found closer to the EnKF-N, compared to A09. We note here that in the presence of model errors, the EnKF-N performance deteriorates strongly.

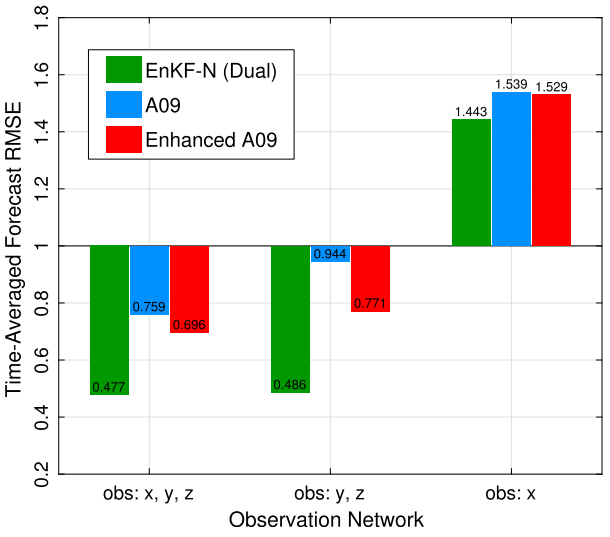

FIG. 2. Time average of forecast RMS errors and spread obtained using the L63 model for three observational network scenarios. The adaptive inflation of A09 is compared to the proposed enhanced scheme, in addition to the EnKF-N.

b. Lorenz 96

Forty state variables are defined as equally spaced on a [0, 1] periodic domain. A truth integration run of the L96 model is performed, starting from a point on the attractor with a time step of 0.05. We assume that every other variable is observed, for a total of 20 observations. A random draw from $N(0, 1)$ is added to the observations as a way to simulate the observation error. The observations are available every time step for a total of 100 000 steps. The results are reported using the last 90 000 steps to avoid any transient behavior. The initial ensemble members are random draws from the model climatology created by taking the 5000th step in a long free integration of the model. The true forcing of the model is set to 8. For localization, a 5th-order polynomial function of Gaspari and Cohn (1999) is used. The inflation is initialized by setting $\lambda_b = 1$ (i.e., no inflation) and $\sigma_{\lambda_b} = 0.6$ for all 40 variables. The inflation variance here is kept fixed at 0.6 for the entire DA cycle.

1) INFLATION CAP

In the following tests, perfect model conditions are assumed, and the performance is assessed based on a changing inflation lower bound (LB): either 0 or 1. The half-width c for localization is set to 0.2. In Fig. 3, the spatial average of the prior and posterior RMS and spread are plotted. Because every other variable is observed, a

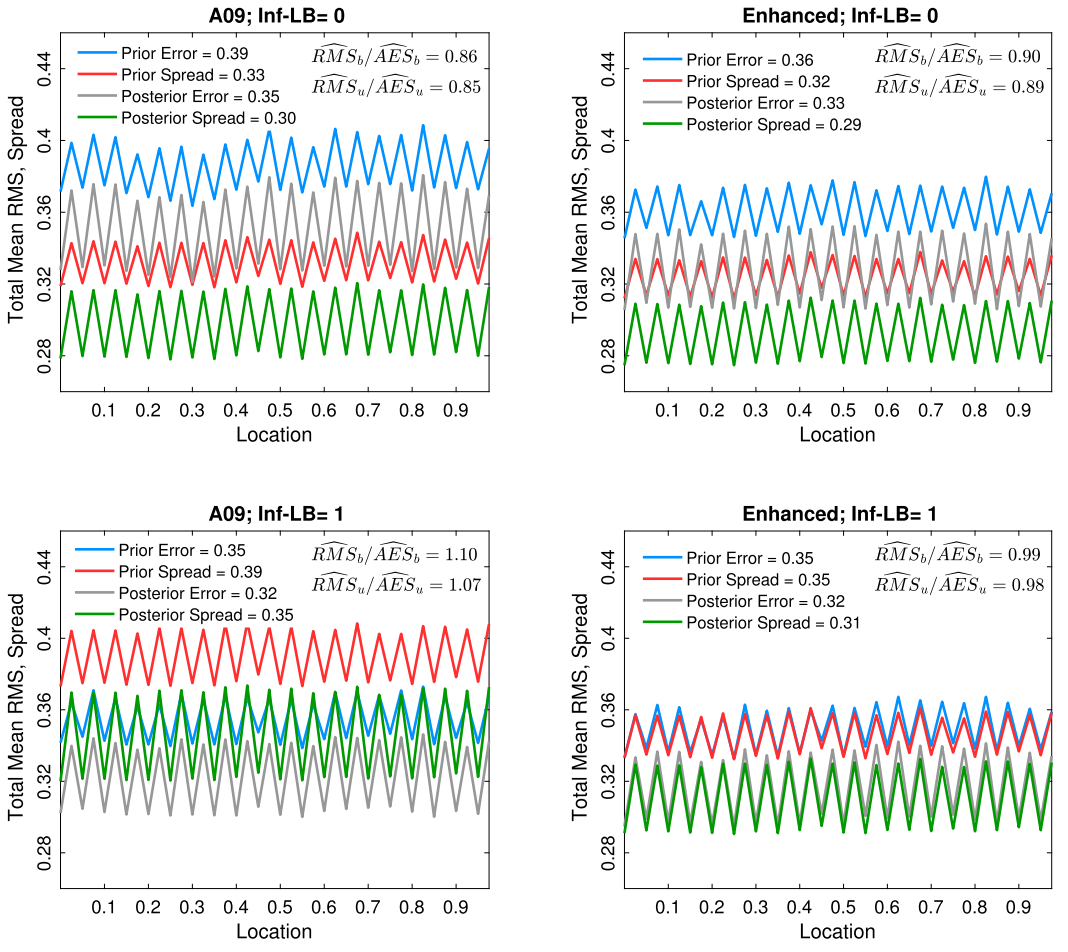

FIG. 3. The temporal average of total RMS errors and AES obtained using (left) A09 and (right) enhanced inflation algorithms. The top and bottom panels use an inflation lower bound (i.e., Inf-LB) of 0 and 1, respectively. Reported in the legends are the prior and posterior overall RMS and spread average. The ratio between the overall average RMS to spread is given for both the prior and the posterior estimates in the top-right corner of the panels.

saw-shaped behavior is obtained. As can be seen, when LB = 1, both algorithms suggest equivalent prior and posterior accuracy. The enhanced algorithm is around 8% more accurate than A09 when LB is set to 0. The overall prior and posterior RMS errors for both schemes are smaller when the LB is 1. In terms of the spread, the enhanced algorithm yields better consistency[3] in the estimates. Apparently, when the LB is 1, A09 displays a slight overdispersion in the prior estimates, with $\widehat{\text{RMS}}_b/\widehat{\text{AES}}_b = 1.10$. On the other hand, the consistency ratio for the priors and the posteriors using the enhanced algorithm is 0.99 and 0.98, respectively. When the LB is 0, the prior and posterior RMS obtained using A09 are larger than the spread.

In Fig. 4, we plot the temporal evolution of the inflation obtained using both algorithms during the last 1000 steps. The adaptive procedure in A09 (blue crosses) yields a wider inflation range, compared to the enhanced algorithm (red triangles). This explains the over- and underdispersed ensemble behavior suggested by A09 in Fig. 3. In space, A09's inflation structure was found more heterogeneous, in the sense that the difference between inflation for observed and unobserved cells is quite big. The enhanced scheme produced a smoother inflation field. Note that large ensemble variations between neighboring cells might create instabilities in some cases.[4] Related to inflation, we assessed the behavior of the eccentricity in Eq. (10) and

[3] The consistency in the estimates is measured here using the ratio between the overall RMS to spread averages.

[4] We have observed similar issues when assimilating into large ocean models. Big differences in the ensemble variance can lead to CFL violations, eventually causing failure of the barotropic solver.

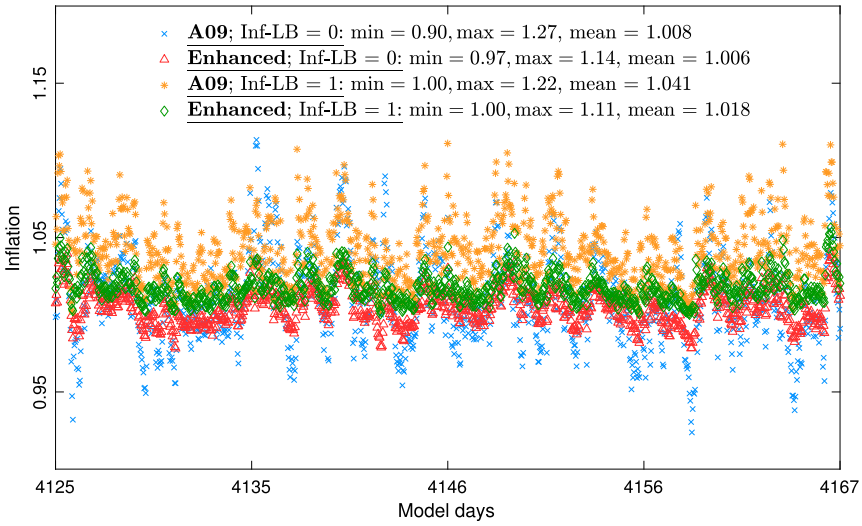

FIG. 4. Temporal evolution of the inflation (all 40 variables of L96) using A09 and the proposed enhanced algorithms over the last 1000 model time steps. Results are displayed for both LBs (0 and 1). The maximum value, minimum value, and the spatiotemporal average of inflation are given in the legend.

found that the value for e suggested by the enhanced scheme is closer to 0. On average, it is around 32% smaller than that of A09.

2) LOCALIZATION AND MODEL ERROR SENSITIVITY

The goal of the following experiments is to test the robustness of the schemes under challenging modeling and assimilation scenarios. For the localization sensitivity, the forcing is kept at 8, and the half-width parameter is varied between 0.1 and 0.8 with a step of 0.1. We tabulate the overall RMSE mean obtained using both algorithms in Table 1. As shown, the enhanced scheme suggests smaller RMSE values for all tested half-widths. For instance, when $c = 0.4$ and LB $= 0$, the enhanced inflation scheme yields estimates that are 9% more accurate than those of A09. The largest differences between the schemes happen for $c = 0.2$, 0.3, and 0.4.

Next, 14 experiments are conducted with different forcing values in the forecast model, $\{F = 1, 2, \ldots, 15\}$ (excluding the true value; i.e., $F = 8$). Without using inflation, a 20-member ensemble diverges quickly after few updates. Both A09 and the enhanced scheme suggest inflation values as high as 1.60 in order to compensate for strong modeling errors. Comparing the average RMSE values in Table 2, the enhanced scheme is found more accurate for all 14 tests. On top of the accuracy, few rank histograms (not shown) for randomly selected variables were assessed. We found that both algorithms generate relatively flat-shaped histograms for $3 < F < 13$. When the model error is very large (i.e., $F = \{1, 2, 3, 13, 14, 15\}$), U-shaped histograms start to become more dominant.

Overall, the proposed enhanced algorithm performed well against A09's scheme, suggesting high robustness to changes in model error and localization length scale. When the LB $= 0$, A09 performed poorly. Under perfect modeling conditions, the enhanced scheme generated small inflation values that were slightly larger than 1. In the presence of modeling errors, the scheme adequately responded and assigned necessary large inflation. We noticed that the occurrence of deflation, as compared to A09, has decreased by around 6%. These features can be attributed to the small right shift introduced to the likelihood probability density function. The IG enhancement to the prior leads to a decreased deflation and, consequently, less inflation, when compared to A09. In general, the performance with no deflation was found more accurate. This, however, cannot be generalized because it could only be a characteristic of the L96 model. As we will see in the next model results, we experience a different behavior when capping the inflation.

c. *Idealized global atmospheric model*

The model is a modified version of the dynamical core used in the Geophysical Fluid Dynamics Laboratory (GFDL) climate model (Anderson et al. 2004). The model (B grid) has no moisture and topography and no parameterizations for radiation, turbulence, and convection. To represent these, the model uses Held and Suarez (1994) forcing instead, which consists of Newtonian relaxation of the temperature field to a prescribed radiative

TABLE 1. The average forecast RMSE resulting from A09 and the enhanced inflation algorithms for two scenarios on the lower bound for inflation, namely, LB = 0 and LB = 1. The results are shown for eight different localization length scales.

	A09		Enhanced	
Half-width: c	LB = 0	LB = 1	LB = 0	LB = 1
0.10	0.4283	0.4032	0.4164	0.4015
0.20	0.3681	0.3436	0.3408	0.3315
0.30	0.3644	0.3267	0.3329	0.3191
0.40	0.3703	0.3260	0.3372	0.3200
0.50	0.3728	0.3211	0.3415	0.3210
0.60	0.3881	0.3302	0.3642	0.3288
0.70	0.3887	0.3336	0.3750	0.3315
0.80	0.3951	0.3411	0.3687	0.3361

equilibrium solution and Rayleigh damping of winds near the surface for dissipation. The prognostic variables in the model are surface pressure (PS), temperature T, zonal U, and meridional V wind components.

The B-grid model is configured with the minimum horizontal and vertical resolution required to generate baroclinic instabilities. The grid domain consists of 30 latitude and 60 longitude grid points and five levels spaced equally in pressure (level 1 being the highest), resulting in a total of 28 200 state variables. The time step in the model is 1 h. For more details, the reader may refer to Anderson et al. (2005). This model is available in the DART package under the name "bgrid-solo." In the following set of DA experiments, we test the performance of the inflation algorithms for constant and varying inflation variance scenarios. We use a predefined climatological state as an initial starting point to perform the truth run over a period of 400 days; however, we only use the last 200 days for assessment of the results. Three hundred PS observations, randomly distributed in the domain, are sampled every day with a Gaussian observation error variance of 1 hPa. All results reported here are based on the prior state (and inflation) estimates. The localization half-width is set to $c = 0.2$ rad.

1) FIXED σ_λ

In this section, the inflation is initialized with $\lambda_b = 1$. The standard deviation $\sigma_{\lambda_b} = 0.6$ is kept fixed for all DA cycles. The RMS errors for all variables obtained using A09's and the enhanced inflation algorithms are displayed in Fig. 5. The RMS errors for T, U, and V are averaged over all model levels. The best performance is obtained using the enhanced scheme with LB = 0. A09's scheme fails to generate accurate estimates when deflation is allowed in the adaptive procedure. On average, the enhanced scheme (LB = 0) outperforms A09 (LB = 1), suggesting an increased accuracy for PS, T, U, and V by 29.4%, 36.4%, 34.4%, and 34.3%, respectively. When

TABLE 2. As in Table 1, but, here, the sensitivity is given in terms of 14 different forcing parameters.

	A09		Enhanced	
Forcing: F	LB = 0	LB = 1	LB = 0	LB = 1
1	2.2170	2.2080	2.1522	2.1318
2	1.9676	1.9560	1.8934	1.8900
3	1.7435	1.7463	1.6486	1.6503
4	1.5295	1.5200	1.4343	1.4301
5	1.3179	1.3129	1.2234	1.2168
6	1.0904	1.0888	0.9989	0.9868
7	0.8078	0.8030	0.7414	0.7111
9	0.8486	0.8319	0.7900	0.7437
10	1.2083	1.2085	1.1121	1.0993
11	1.5322	1.5244	1.4113	1.3978
12	1.8553	1.8527	1.6916	1.7217
13	2.1493	2.1666	1.9809	2.0519
14	2.5317	2.4765	2.3138	2.2953
15	2.8118	2.7636	2.6068	2.5739

the LB = 1, the improvements over A09 are 11.8%, 9.1%, 12.5%, and 14.3%. To make sense of these statistics, the temporal average of the bias (LB = 0) for all variables is displayed in Fig. 6. The maps suggest that the largest biases are concentrated in "no observation" locations (e.g., central and north Asia) and dense observation areas (e.g., west of Australia). The absolute bias resulting from A09 grows up to 0.2 hPa, 0.04 K, and 0.3 m s^{-1}, for PS, T, and $U + V$, respectively. The estimates generated by the enhanced scheme are less biased. Given the instabilities in the B-grid model, most of the wind biases appear in midlatitudes (top-right panel).

Next, we plot in Fig. 7 PS inflation snapshots after 200 days (top panels), in addition to the temporal average (bottom panels) for LB = 0. At day 200, strong deflation areas in the high upper latitudes dominate A09's map. It is surprising to see that deflation happens mostly in places where observations are available (e.g., south of Australia). The enhanced scheme, on the contrary, yields reasonable inflation patterns, such that no-observation locations are assigned $\lambda \approx 1$, and locations with dense data points have inflation values as high as 2 (e.g., North Pacific). Deflation, using the enhanced scheme, is rarely observed and almost nonexistent. This could be attributed to the inclusion of the inverse gamma density in the prior. Averaging over 400 days, the enhanced scheme yields inflation patterns that spatially match the observation clusters in the domain. To illustrate, areas dominated with observations are generally bluish in color (i.e., $\lambda > 1$), while the rest of the domain is more reddish with almost no inflation. A09's inflation map is more heterogeneous, and no clear data-matching patterns are detected. Furthermore, inflation values in locations such as South Asia and the Indian Ocean are almost 25% larger than those suggested by

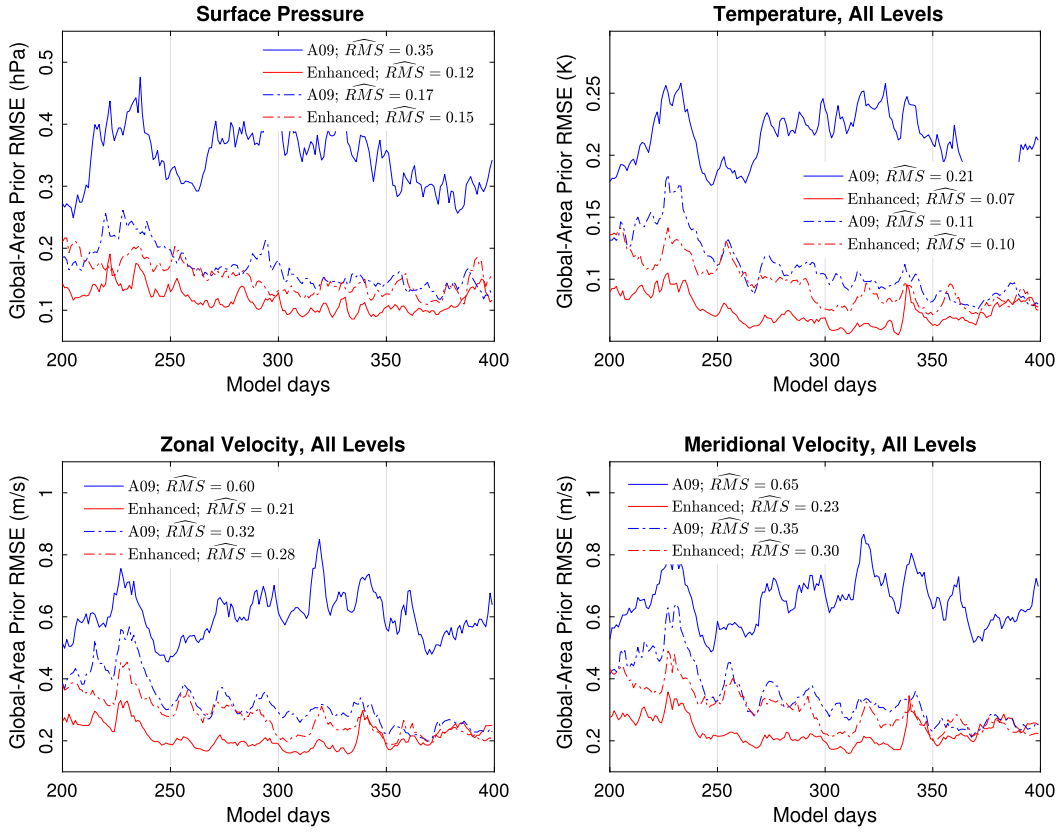

FIG. 5. Evolution of spatially averaged prior RMS in time for all variables of the dry dynamical core. Except for PS, the averaging is done for all model levels. The solid lines are results from experiments with inflation LB = 0. The dash–dotted lines, on the other hand, use LB = 1. The overall RMS averages (both in space and time) are reported in the legends of each panel.

the enhanced scheme. Consistent with the stability analysis described before, a number of no-inflation spots (red in color) in the domain are observed within large inflation zones (blue in color), as in the North and South Pacific. This strong inflation heterogeneity is one factor causing the degradation in the performance of A09 when LB = 0.

In addition to PS, we also analyze the behavior of the inflation along the vertical direction in the model. Figure 8 plots the average inflation maps for T at levels 1 and 5. Similar to PS, the enhanced scheme is shown to generate more reliable maps with consistent data-matching structures. Closer to the surface at L5, the inflation values are large, given the higher impact from PS observations. The maps produced by A09 consist of very large inflation regions, especially near the surface. Within these regions, few tiny no-inflation spots are observed. To better understand this behavior, we study in Fig. 9 the temporal change in the inflation's overall minimum and maximum values, in addition to the spatial average. Both schemes yield comparable inflation average with $\lambda \approx 1.2$. The results are different for the minimum and maximum inflation values. For instance, the minimum values obtained using the proposed algorithm tend to vary between 0.5 and 0.8, whereas most of the inflation minima obtained using A09 are equal to 0.1.[5] This explains the behavior observed in Figs. 7 and 8, in which A09 assigns very large inflation to combat such a strong deflation. Accordingly, the inflation range suggested by A09 is around 30% wider than that of the proposed scheme. Small inflation values are artifacts resulting from (i) the use of Gaussian priors and (ii) the inflation likelihoods peaking at very small values. Both of these issues seem to be partially, if not fully, resolved in the new inflation scheme.

2) DISCUSSION

When the inflation variance is fixed in time, the enhanced scheme with LB = 0 was shown to be the most reliable, providing accurate estimates. The results of the enhanced scheme with LB = 1 are, on average, 25% less accurate. One might argue that the main reason for using

[5] Note that the smallest inflation values are 0.1 and not 0 because of damping. With no damping, these become 0, and the assimilation fails.

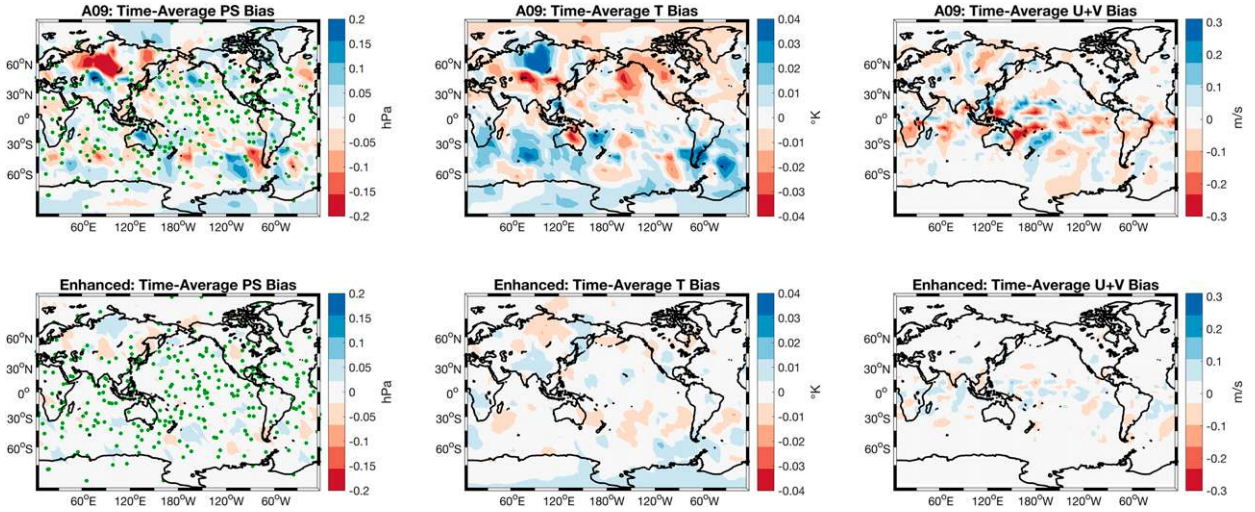

FIG. 6. Maps of time-averaged bias resulting from (top) A09 and (bottom) the enhanced adaptive inflation schemes. Shown on the maps with green circles are the location of the PS observations.

inflation is to avoid prior variance underestimation, and, thus, deflation should not be part of the scaling procedure. This is true to a certain extent; however, a more detailed and in-depth analysis is needed here.

(i) Preventing deflation means that Eq. (8), or Eq. (25) in our enhanced context, is no longer satisfied. As such, the innovation statistics in the following assimilation cycles, whether indicating inflation or deflation, may be incorrectly interpreted. Ignoring the likelihood information at one cycle may have consequences on future assimilations.

(ii) The process of inflation inserts an artificial spread in the ensemble that the model is unaware of. For a biased model, this can be extremely large. This may

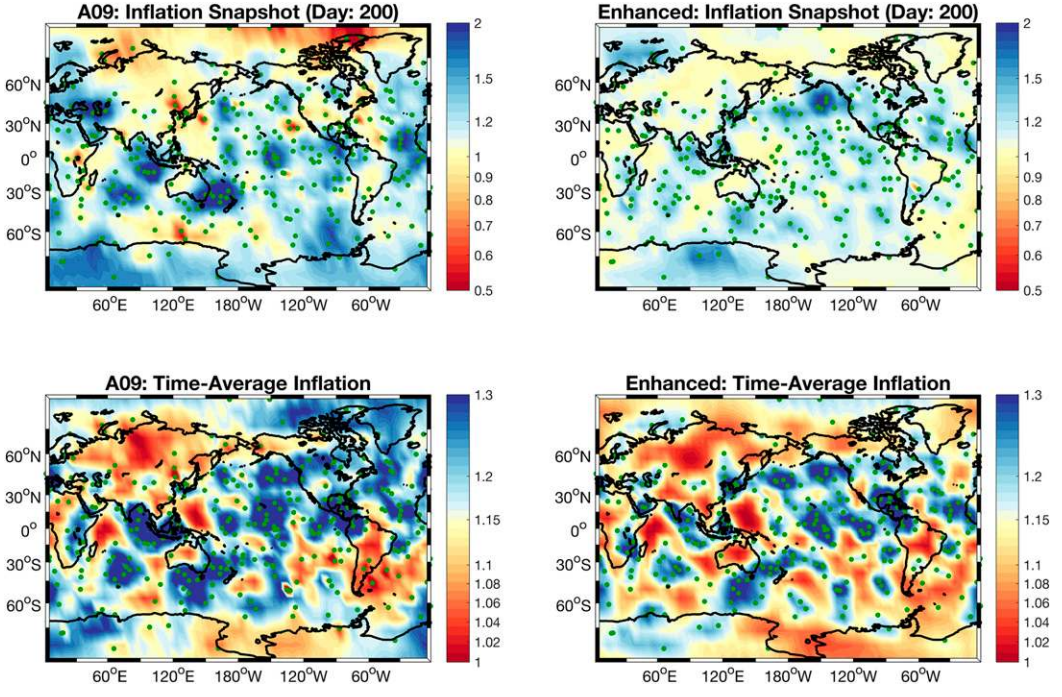

FIG. 7. (top) PS inflation maps after 200 days of assimilation resulting from (left) A09 and (right) the enhanced schemes. (bottom) As in (top), but displaying the overall time average (entire 400 days) instead. The green circles denote the PS observations locations.

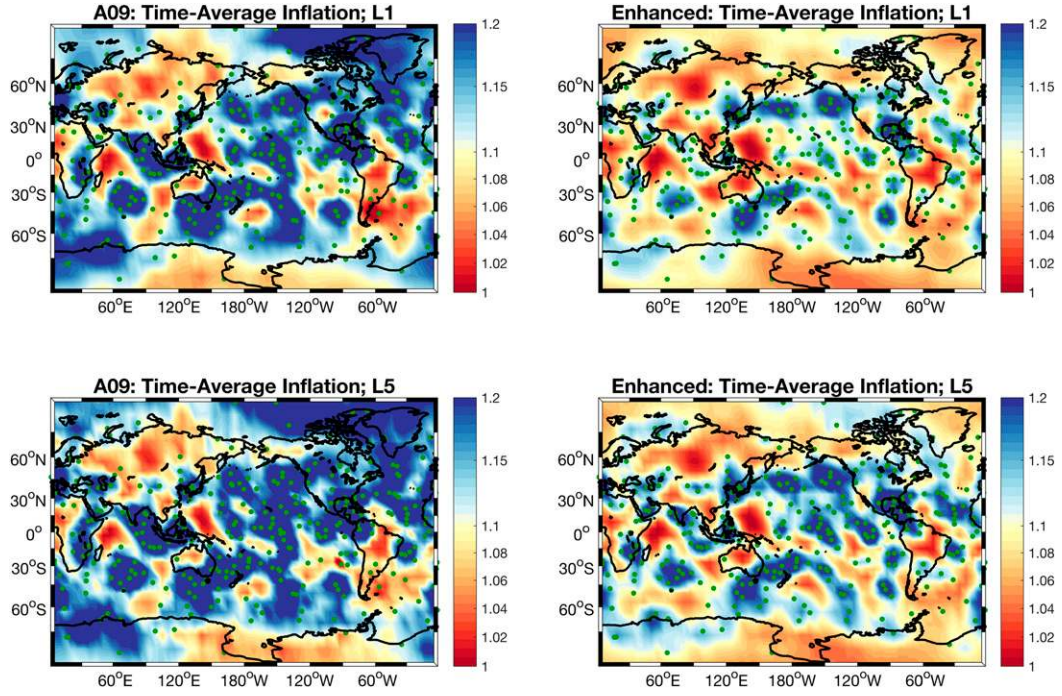

FIG. 8. (top) Time-average inflation maps for temperature from model level 1, resulting from (left) A09 and (right) the enhanced schemes. (bottom) As in (top), but for level 5. The green circles denote the PS observations locations.

require long cycles and successive Kalman updates before it can be removed, during which the accuracy of the forecast might be degraded. To better fit the data, deflation can be applied to get rid of the extra artificial spread. If the spread was to grow naturally in the ensemble due to the dynamics of the model alone, then deflation should not be utilized, and the Bayesian update can minimize the variance.

(iii) A strong deflation, as we experienced with A09, can be very unhealthy, producing poor estimates. The use of deflation should be moderate—hence the introduction of the IG instead of a Gaussian. Moreover, deflation was applied using A09's adaptive scheme 7489 times. This number was greatly reduced to 1144 times using the enhanced scheme, given the modified likelihood variance [Eq. (26)].

3) VARYING σ_λ

To test the functionality of the inflation variance, we now start the assimilation by setting $\sigma_\lambda(t_0) = 0.9$ and allow it to change in time. The minimum allowable variance is set to 0.01, such that after the update, if σ_{λ_b} becomes smaller than 0.1, it is kept unchanged. Here, we only test with LB = 1 for inflation, given the poor performance of A09 in the presence of deflation. The time evolution of σ_{λ_b} for PS and T is plotted in Fig. 10. The variance reduction for T, as shown, is smaller than that of PS. This is because PS is directly observed, and T is averaged over model levels that are far from the surface, and, thus, the impact of the updates is weaker. Given the Gaussian assumption, the inflation variance obtained by A09 monotonically decreases, eventually reaching 0.01 after 300 days. The enhanced scheme takes on a decreasing trend but with obviously a slower rate; σ_{λ_b} is seen to equilibrate at around 0.35 and 0.50 for PS and T, respectively. Furthermore, the variance suggested by the enhanced algorithm is shown to slightly increase at different times due to the IG assumption.

The effects of this behavior from both schemes are illustrated in Fig. 11, in which the RMS and AES of the estimates are displayed over time. Given the rapid decrease in the inflation variance, A09's performance tends to degrade, as shown in the last 50 days. When the inflation variance becomes very small, the effects of the Bayesian update decrease. For a longer simulation period (i.e., >400 days), the estimates are expected to get worse as the inflation algorithm becomes technically ineffective. For all variables, the RMSE suggested by A09 is almost double that of the spread, probably indicating an ensemble collapse (divergence). On the other hand, the enhanced scheme estimates are more accurate, further suggesting a consistency of ~1 over the entire assimilation period. Finally, we show the spatial maps for inflation

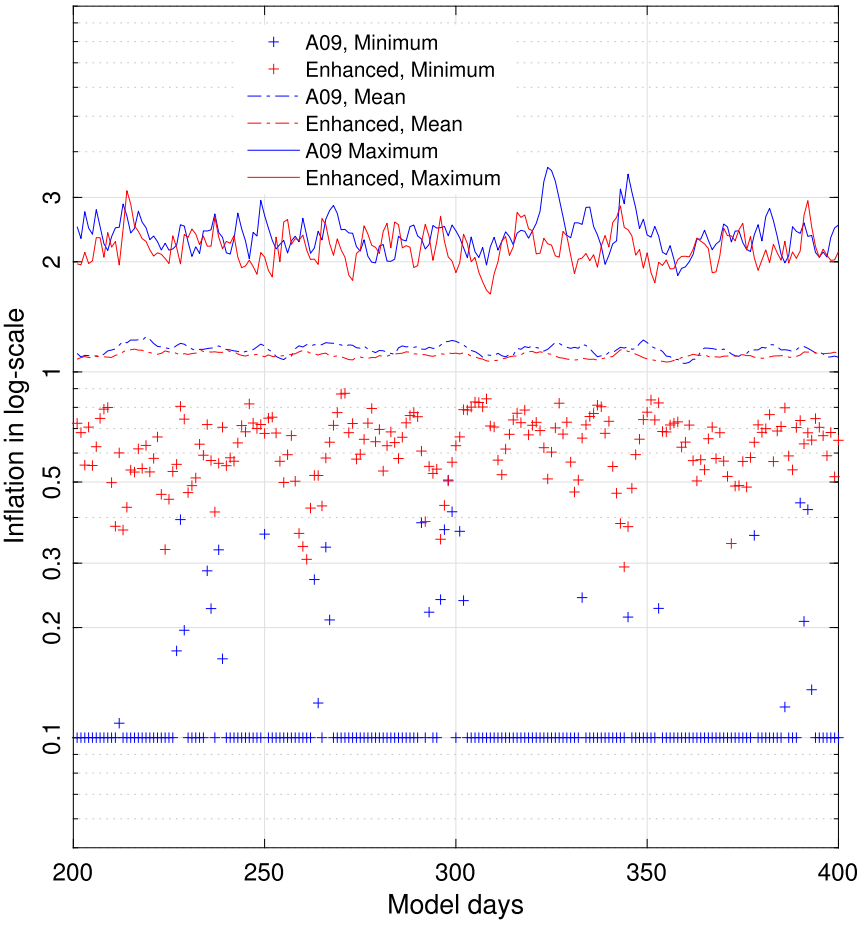

FIG. 9. Inflation values averaged in space, in the vertical, and over all prognostic variables obtained using A09 and the enhanced inflation schemes. The minimum, mean, and maximum values for these inflation values are plotted in time. The LB here is 0. The y axis is shown in log scale.

standard deviation after 50, 200, and 400 days in Fig. 12. A09 quickly decreases σ_{λ_b} in the vicinity of the observation points from 0.9 to 0.2 after 50 days. At day 400, the standard deviation becomes almost homogenous across the domain and equal to 0.1. The behavior of σ_{λ_b} suggested by the enhanced scheme is quite different. As can be seen, the major updates happen around the observation locations. Moreover, the inflation variance is reduced in time at a slower pace. The spatial patterns of σ_{λ_b} are comparable to the inflation structure observed in Fig. 7 (left panels), unlike what we obtained from A09. After 400 days, $\sigma_{\lambda_b} = 0.2$ in the vicinity of the observations and 0.9 for unobserved locations. This feature can be quite useful in realistic scenarios when the observation network changes constantly in time.

5. Summary and discussion

This paper studies the spatially and temporally varying adaptive prior covariance inflation for ensemble filters. Several enhancements are introduced to the previously documented algorithm of A09. The distance between the model forecast and the data points is assumed to be a random variable. This modifies the variance of the inflation likelihood with a factor of $1/N_e$, such that the distribution is slightly shifted to the right. Furthermore, the inflation prior is assumed inverse gamma instead of Gaussian. This restricts the sampling of the inflation to positive values only. The benefits of these enhancements are twofold. The modified likelihood is expected to behave better in the case of a very sparse observational network and/or assimilating observations with large uncertainties. ''Better'' here means that the occurrence of negative or small inflation values is reduced. The use of an inverse gamma prior limits the magnitude of deflation, should it occur. The enhanced scheme is compared to the original algorithm of A09 in twin experiments, using a low-dimensional Lorenz-96 and a high-dimensional idealized atmospheric model.

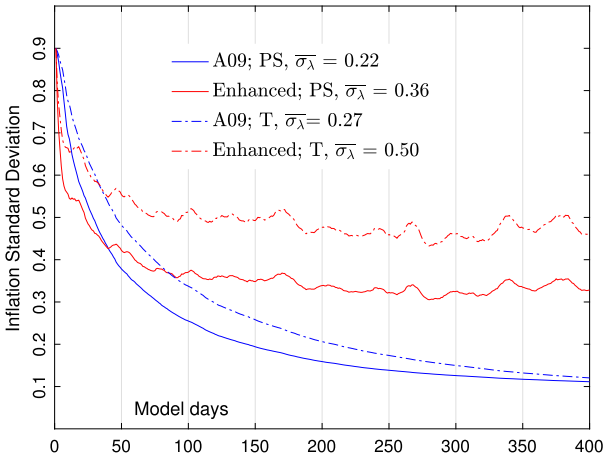

FIG. 10. Time evolution of inflation standard deviation σ_{λ_b} for PS and T (averaged over the entire five levels). Reported in the legend are the overall averages of σ_{λ_b}.

Under perfect modeling conditions, the enhanced scheme proved its efficiency in mitigating sampling errors, as compared to A09. This was verified by comparing the assimilation results to those of the EnKF-N benchmark. When the inflation is bounded below by 1, the accuracy suggested by the enhanced scheme is only marginally higher than that of A09. Moving the cap to 0, the enhanced scheme efficiently decreases both the magnitude and the frequency of deflation. This produced the most accurate and most consistent estimates in the atmospheric model. The performance of A09's algorithm, on the other hand, degraded strongly because the Bayesian update leads to extremely tiny and large inflation values. We note here that A09's algorithm was almost always used (at least by DART users) with an LB of 1. Allowing the inflation variance to change, the enhanced scheme offered more room for the inflation to evolve in time, and, thus, a better fit to the data was obtained. For A09, the inflation variance experienced a rapid decrease that forced the adaptive scheme to fix the prior, and this, in turn, reduced the accuracy of the estimates.

The presented results indicate that a moderate and not-so-frequent use of deflation can be quite useful. This is mainly because the mechanism by which inflation evolves in time is only based on the Bayesian update. Because the inflation distribution is kept unchanged during the forecast, an alternative way of dissipating an overinflated variance in the absence of observations is required. One example of such a scenario would be Doppler radar. Large inflation values are needed

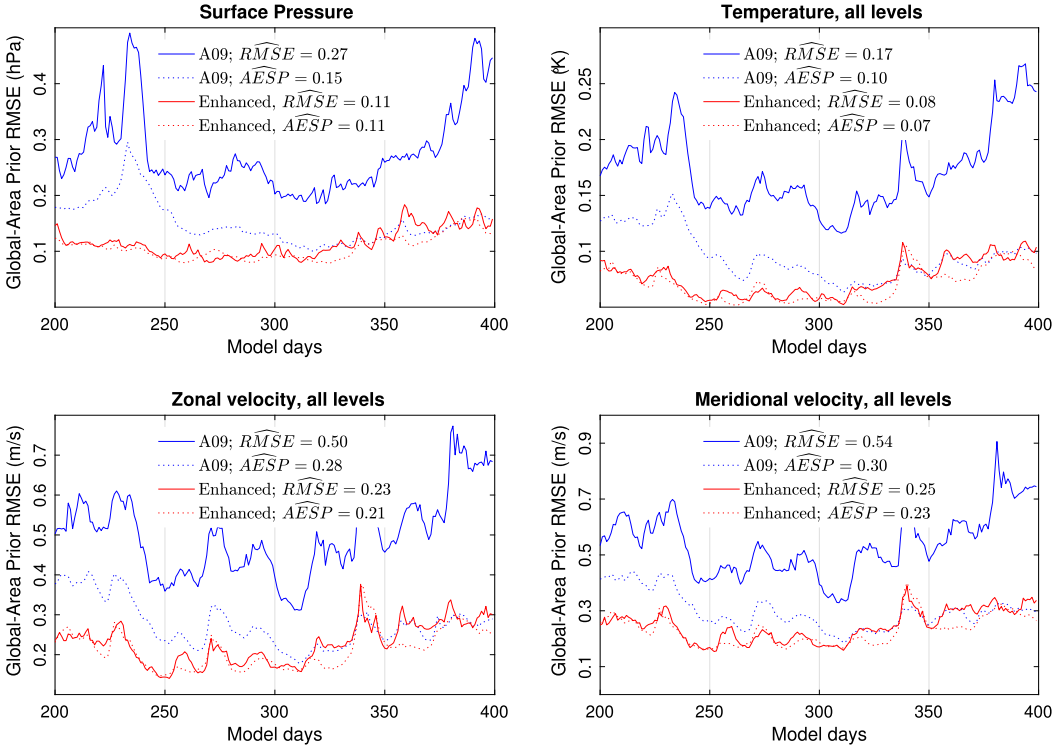

FIG. 11. Evolution of spatially averaged prior RMS/AES in time for all variables. Except for PS, the averaging is done for all model levels. The overall RMS/AES averages (both in space and time) are reported in the legends of each panel. Here, LB = 1.

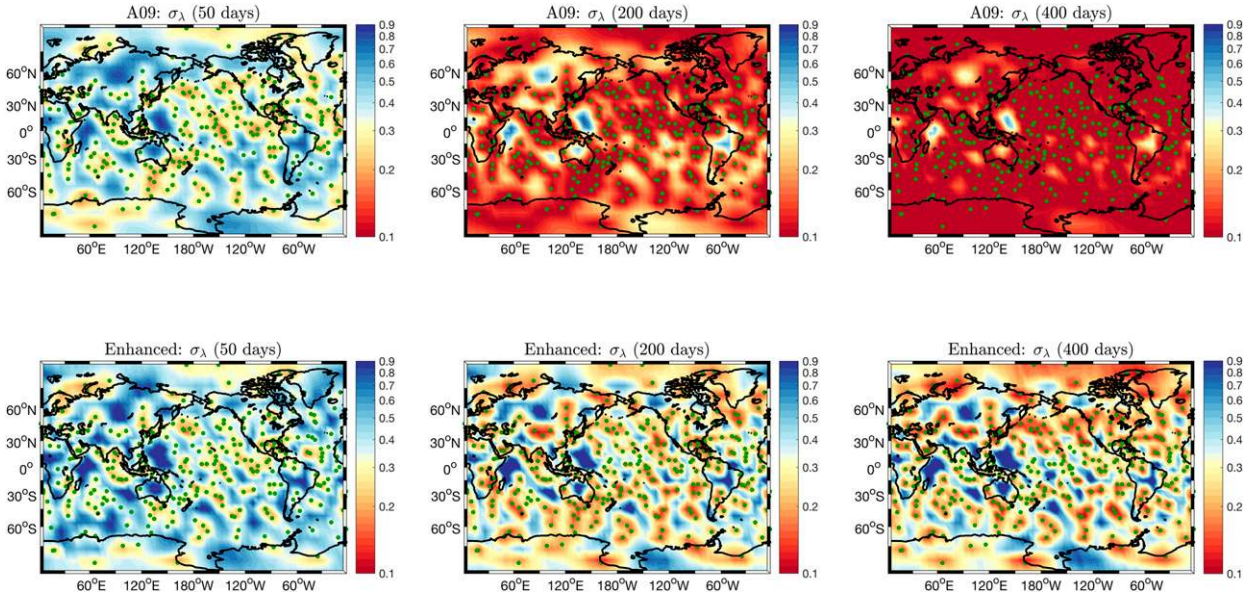

FIG. 12. Maps of the PS inflation standard deviation resulting from (top) A09 and (bottom) the enhanced adaptive inflation schemes. Shown on the maps with green circles are the location of the PS observations. Time (days) increases from left to right panels.

during a convective event. When the convection ends and Doppler observations vanish, the added artificial variance may be removed by other observations; however, the process can be very slow, leading to strong biases. The proposed scheme with no lower bound on the inflation can efficiently remove such variance through deflation only if there are some observations once the radar is gone. Another example is a tropical storm. To fit the data, the underestimated variance of the ensemble is heavily inflated. Once the hurricane passes, deflating the ensemble can help the model quickly forget about the hurricane and increase its prediction accuracy.

Despite the success of the presented inflation scheme, some limitations still exist.

- In terms of the inflation likelihood, the added correction term, which is based on a biased estimator for the expectation of the squared innovation sample, does not fully solve the problem. The observed inflation can still be negative when assimilating observations with large uncertainties. The correction term, however, partially mitigates the negative and deflation issues. It is important to note that the imposed right shift on the likelihood peak does not always yield larger posterior inflation values.
- An appropriate forecast model for the inflation is needed. Allowing the inflation first and second moments to evolve in time could improve the assimilation performance. Such a model should have strong correlations with the bias characteristics of the physical model, the observation network spatial structure, and how often the data become available. Ideas like deflation and damping were introduced to compensate for the absence of inflation time evolution.
- The presented inflation algorithm, unlike other schemes, such as the EnKF-N (Bocquet et al. 2015), does not account for errors in the ensemble mean and its covariance, given a finite ensemble size. Such confidence in the ensemble mean may further generate sampling errors that the adaptive scheme may not be aware of.
- When the sampling error is very large, the posterior variance may turn out to be small. One possible remediation is to inflate the posterior variance. Adapting the posterior variance inflation is not as simple as that of the prior because the observations would have been already used during the update. One could remove the impact of an observation to compute innovation and then apply the appropriate Bayesian update. This will be explored in detail in a future study. In fact, DART has supported this capability for a long time, and it has been used quite a bit, but it was not carefully documented.

The enhanced algorithm produced a robust scheme that is yet to be tested in realistic large-scale weather models. This task has been initiated, and experiments using the National Center for Atmospheric Research Community Atmosphere Model (CAM) are currently being conducted. Wind and temperature observations from radiosondes, ACARS, and aircraft, along with GPS radio occultation observations, are assimilated. Different 6-h forecasts of the atmospheric state are generated using the enhanced inflation scheme. The results of these tests will be analyzed in a follow-up publication.

Acknowledgments. The author is grateful for Patrick N. Raanes and another anonymous reviewer for their comments and suggestions. The author thanks Tim Hoar, Nancy Collins, Kevin Raeder, and Jonathan Hendricks for their immense technical support with DART. The author also thanks Jeffrey Anderson, Chris Snyder, and Alicia Karspeck for insightful discussions. The author sincerely appreciates the fruitful chats with Moustafa Gharamti. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation.

APPENDIX

Lorenz-63 System

The Lorenz-63 model (Lorenz 1963) is a model with three variables, defined by the equations

$$\frac{dx}{dt} = \sigma(y - x), \tag{A1}$$

$$\frac{dy}{dt} = \rho x - y - xz, \tag{A2}$$

$$\frac{dz}{dt} = xy - \beta z. \tag{A3}$$

The parameters are chosen such that the model produces chaotic dynamics; [$\sigma = 10$, $\rho = 28$, $\beta = 8/3$]. The time step in the mode is set to 0.01.

REFERENCES

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

——, 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634–642, https://doi.org/10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2.

——, 2007: An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus*, **59A**, 210–224, https://doi.org/10.1111/j.1600-0870.2006.00216.x.

——, 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **61A**, 72–83, https://doi.org/10.1111/j.1600-0870.2008.00361.x.

——, 2012: Localization and sampling error correction in ensemble Kalman filter data assimilation. *Mon. Wea. Rev.*, **140**, 2359–2371, https://doi.org/10.1175/MWR-D-11-00013.1.

——, and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758, https://doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2.

——, and Coauthors, 2004: The new GFDL global atmosphere and land model AM2–LM2: Evaluation with prescribed SST simulations. *J. Climate*, **17**, 4641–4673, https://doi.org/10.1175/JCLI-3223.1.

——, B. Wyman, S. Zhang, and T. Hoar, 2005: Assimilation of surface pressure observations using an ensemble filter in an idealized global atmospheric prediction system. *J. Atmos. Sci.*, **62**, 2925–2938, https://doi.org/10.1175/JAS3510.1.

——, T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano, 2009: The data assimilation research testbed: A community facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296, https://doi.org/10.1175/2009BAMS2618.1.

Berner, J., G. Shutts, M. Leutbecher, and T. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, **66**, 603–626, https://doi.org/10.1175/2008JAS2677.1.

Bocquet, M., 2011: Ensemble Kalman filtering without the intrinsic need for inflation. *Nonlinear Processes Geophys.*, **18**, 735–750, https://doi.org/10.5194/npg-18-735-2011.

——, and A. Carrassi, 2017: Four-dimensional ensemble variational data assimilation and the unstable subspace. *Tellus*, **69A**, 1304504, https://doi.org/10.1080/16000870.2017.1304504.

——, P. N. Raanes, and A. Hannart, 2015: Expanding the validity of the ensemble Kalman filter without the intrinsic need for inflation. *Nonlinear Processes Geophys.*, **22**, 645–662, https://doi.org/10.5194/npg-22-645-2015.

Daley, R., 1993: *Atmospheric Data Analysis.* Cambridge Atmospheric and Space Science Series, Vol. 2, Cambridge University Press, 457 pp.

Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1145, https://doi.org/10.1175/1520-0493(1995)123<1128:OLEOEC>2.0.CO;2.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396, https://doi.org/10.1256/qj.05.108.

Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–367, https://doi.org/10.1007/s10236-003-0036-9.

——, 2009: *Data Assimilation: The Ensemble Kalman Filter.* 2nd ed. Springer Science & Business Media, 307 pp.

Furrer, R., and T. Bengtsson, 2007: Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *J. Multivar. Anal.*, **98**, 227–255, https://doi.org/10.1016/j.jmva.2006.08.003.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, https://doi.org/10.1002/qj.49712555417.

Harlim, J., and B. R. Hunt, 2007: A non-Gaussian ensemble filter for assimilating infrequent noisy observations. *Tellus*, **59A**, 225–237, https://doi.org/10.1111/j.1600-0870.2007.00225.x.

Held, I. M., and M. J. Suarez, 1994: A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models. *Bull. Amer. Meteor. Soc.*, **75**, 1825–1830, https://doi.org/10.1175/1520-0477(1994)075<1825:APFTIO>2.0.CO;2.

Hoteit, I., D.-T. Pham, and J. Blum, 2002: A simplified reduced order Kalman filtering and application to altimetric data assimilation in tropical Pacific. *J. Mar. Syst.*, **36**, 101–127, https://doi.org/10.1016/S0924-7963(02)00129-X.

——, ——, M. Gharamti, and X. Luo, 2015: Mitigating observation perturbation sampling errors in the stochastic EnKF. *Mon. Wea. Rev.*, **143**, 2918–2936, https://doi.org/10.1175/MWR-D-14-00088.1.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112–126, https://doi.org/10.1016/j.physd.2006.11.008.

Karspeck, A. R., 2016: An ensemble approach for the estimation of observational error illustrated for a nominal 1° global ocean

model. *Mon. Wea. Rev.*, **144**, 1713–1728, https://doi.org/10.1175/MWR-D-14-00336.1.

——, S. Yeager, G. Danabasoglu, T. Hoar, N. Collins, K. Raeder, J. Anderson, and J. Tribbia, 2013: An ensemble adjustment Kalman filter for the CCSM4 ocean component. *J. Climate*, **26**, 7392–7413, https://doi.org/10.1175/JCLI-D-12-00402.1.

Kotsuki, S., Y. Ota, and T. Miyoshi, 2017: Adaptive covariance relaxation methods for ensemble data assimilation: Experiments in the real atmosphere. *Quart. J. Roy. Meteor. Soc.*, **143**, 2001–2015, https://doi.org/10.1002/qj.3060.

Li, H., E. Kalnay, and T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **135**, 523–533, https://doi.org/10.1002/qj.371.

Liang, X., X. Zheng, S. Zhang, G. Wu, Y. Dai, and Y. Li, 2012: Maximum likelihood estimation of inflation factors on error covariance matrices for ensemble Kalman filter assimilation. *Quart. J. Roy. Meteor. Soc.*, **138**, 263–273, https://doi.org/10.1002/qj.912.

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.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

——, 1996: Predictability: A problem partly solved. *Proc. Seminar on Predictability*, Reading, United Kingdom, ECMWF, 1–19.

Luo, X., and I. Hoteit, 2011: Robust ensemble filtering and its relation to covariance inflation in the ensemble Kalman filter. *Mon. Wea. Rev.*, **139**, 3938–3953, https://doi.org/10.1175/MWR-D-10-05068.1.

Meng, Z., and F. Zhang, 2007: Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect model experiments. *Mon. Wea. Rev.*, **135**, 1403–1423, https://doi.org/10.1175/MWR3352.1.

Minamide, M., and F. Zhang, 2017: Adaptive observation error inflation for assimilating all-sky satellite radiance. *Mon. Wea. Rev.*, **145**, 1063–1081, https://doi.org/10.1175/MWR-D-16-0257.1.

Mitchell, H. L., and P. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 416–433, https://doi.org/10.1175/1520-0493(2000)128<0416:AAEKF>2.0.CO;2.

Miyoshi, T., 2011: The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Mon. Wea. Rev.*, **139**, 1519–1535, https://doi.org/10.1175/2010MWR3570.1.

Raanes, P. N., 2016: Improvements to ensemble methods for data assimilation in the geosciences. Ph.D. thesis, Oxford University, 182 pp.

——, A. Carrassi, and L. Bertino, 2015: Extending the square root method to account for additive forecast noise in ensemble methods. *Mon. Wea. Rev.*, **143**, 3857–3873, https://doi.org/10.1175/MWR-D-14-00375.1.

Sakov, P., and P. R. Oke, 2008: A deterministic formulation of the ensemble Kalman filter: An alternative to ensemble square root filters. *Tellus*, **60A**, 361–371, https://doi.org/10.1111/j.1600-0870.2007.00299.x.

——, F. Counillon, L. Bertino, K. Lisæter, P. Oke, and A. Korablev, 2012: TOPAZ4: An ocean-sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Sci.*, **8**, 633–656, https://doi.org/10.5194/os-8-633-2012.

Stroud, J. R., and T. Bengtsson, 2007: Sequential state and variance estimation within the ensemble Kalman filter. *Mon. Wea. Rev.*, **135**, 3194–3208, https://doi.org/10.1175/MWR3460.1.

Uzunoglu, B., S. Fletcher, M. Zupanski, and I. Navon, 2007: Adaptive ensemble reduction and inflation. *Quart. J. Roy. Meteor. Soc.*, **133**, 1281–1294, https://doi.org/10.1002/qj.96.

van Leeuwen, P. J., 1999: Comment on ''Data assimilation using an ensemble Kalman filter technique.'' *Mon. Wea. Rev.*, **127**, 1374–1377, https://doi.org/10.1175/1520-0493(1999)127<1374:CODAUA>2.0.CO;2.

Wang, X., and C. H. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140–1158, https://doi.org/10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2.

——, T. M. Hamill, J. S. Whitaker, and C. H. Bishop, 2007: A comparison of hybrid ensemble transform Kalman filter–optimum interpolation and ensemble square root filter analysis schemes. *Mon. Wea. Rev.*, **135**, 1055–1076, https://doi.org/10.1175/MWR3307.1.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

——, and ——, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089, https://doi.org/10.1175/MWR-D-11-00276.1.

Ying, Y., and F. Zhang, 2015: An adaptive covariance relaxation method for ensemble data assimilation. *Quart. J. Roy. Meteor. Soc.*, **141**, 2898–2906, https://doi.org/10.1002/qj.2576.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2.

Zheng, X., 2009: An adaptive estimation of forecast error covariance parameters for Kalman filtering data assimilation. *Adv. Atmos. Sci.*, **26**, 154–160, https://doi.org/10.1007/s00376-009-0154-5.