Data assimilation challenges posed by nonlinear operators: A comparative study of ensemble and variational filters and smoothers

Kenta Kurosawa *

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

Jonathan Poterjoy

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

NOAA Atlantic Oceanographic and Meteorological Laboratory, Miami, Florida

**Corresponding author*: Kenta Kurosawa, kkurosaw@umd.edu

ABSTRACT

The ensemble Kalman Filter (EnKF) and the 4D variational method (4DVar) are the most commonly used filters and smoothers in atmospheric science. These methods typically approximate prior densities using a Gaussian and solve a linear system of equations for the posterior mean and covariance. Therefore, strongly nonlinear model dynamics and measurement operators can lead to bias in posterior estimates. To improve the performance in nonlinear regimes, minimization of the 4DVar cost function typically follows multiple sets of iterations, known as an “outer loop”, which helps reduce bias caused by linear assumptions. Alternatively, "iterative ensemble methods" follow a similar strategy of periodically re-linearizing model and measurement operators. These methods come with different, possibly more appropriate, assumptions for drawing samples from the posterior density, but have seen little attention in numerical weather prediction (NWP) communities. Lastly, particle filters (PFs) present a purely Bayesian filtering approach for state estimation, which avoids many of the assumptions made by the above methods. Several strategies for applying localized PFs for NWP have been proposed very recently. The current study investigates intrinsic limitations of current data assimilation methodology for applications that require nonlinear measurement operators. In doing so, it targets a specific problem that is relevant to the assimilation of remotely-sensed measurements, such as radar reflectivity and all-sky radiances, which pose challenges for Gaussian-based data assimilation systems. This comparison includes multiple data assimilation approaches designed recently for nonlinear/non-Gaussian applications, as well as those currently used for NWP.

1. Introduction

The ensemble Kalman Filter (EnKF; Evensen 1994; Houtekamer and Mitchell 1998; Evensen and van Leeuwen 2000) and the 4D variational method (4DVar; Thepáut and Courtier 1991) are the most commonly used filters and smoothers in atmospheric science. Ensemble/variational hybrid approaches (e.g., Hamill and Snyder 2000; Lorenc 2003; Buehner 2005) combine the flow-dependent ensemble covariance from an EnKF with climate-based covariance from variational methods. The methods have also become well-established and widely accepted for global weather prediction at major environmental prediction centers, such as the European Centre for Medium-Range Weather Forecasts (ECMWF) , UK Met Office, Environment and Climate Change Canada (ECCC), and National Centers for Environmental Prediction (NCEP). One strategy of the hybrid methods, denoted as ensemble-4DVar (E4DVar; Zhang et al. 2009) in this manuscript, typically uses tangent linear and adjoint model operators to minimize a cost function in the same manner as the traditional 4DVar data assimilation system. A second strategy is 4D-ensemble-Var (4DEnVar; Liu et al. 2008), in which the cost function minimization is computed based on an ensemble forecast instead of using tangent linear and adjoint models. In the 4DEnVar, temporal covariances are estimated from an ensemble of model trajectories that pass through the observation time window. In either case, both methods approximate prior densities using a Gaussian and perform linearizations to relax these assumptions. Therefore, strongly nonlinear model dynamics or measurement operators cause these methods to be biased, which leads to the suboptimal use of major Earth observing systems, such as satellite radiometers. For example, the combined impact of highly nonlinear model dynamics and measurement operators introduces major data assimilation challenges in weather regimes containing clouds or precipitation. As a result, most infrared satellite assimilation studies mainly focus on clear-sky observations (e.g., Errico et al. 2007; Fabry and Sun 2010; Geer and Bauer

2011; Zou et al. 2013; Okamoto et al. 2014; Minamide and Zhang 2017; Honda and Coauthors 2018). This follows despite the known benefits of assimilating cloudy radiances for weather forecasting (e.g., Vukicevic et al. 2004; Stengel et al. 2009; Privé et al. 2013). Some operational centers are making efforts to cope with these issues and assimilate cloudy and precipitating microwave radiances (e.g., Zhu et al. 2016; Geer et al. 2017, 2019). For further details on significant advances and current plans of operational centers that are close to implementing assimilation, we encourage readers to review the summary presented in Geer et al. (2018).

Several procedures have been proposed to improve the performance of these methods in nonlinear regimes. For example, in order to deal with issues within the 4DVar system (e.g., Bonavita et al. 2018), minimization of the 4DVar cost function typically follows multiple sets of iterations to re-linearize tangent linear and adjoints for the model, measurement operators, or both around an improved background solution. This step, known as an "outer loop," helps reduce bias caused by linear assumptions, thus making Gaussian error approximations more appropriate. The minimization strategy follows the Gauss–Newton method, which is guaranteed to approximate the posterior mode for local minima.

Alternatively, a number of methods fall under the generic category of "iterative ensemble methods", which follow a similar strategy of periodic re-linearization. Note that here “iterations” refers to multiple adjustments at a single time. Both 4DVar and the iterative ensemble methods re-linearize the observation operator. The only difference is that in 4DVar, the observation operator contains the nonlinear model. Gu and Oliver (2007) introduced the ensemble randomized maximal likelihood filter (EnRML) to handle nonlinearity by means of iterations of the EnKF. Sakov et al. (2012) proposed the iterative ensemble Kalman filter (IEnKF), which uses a deterministic update form, ensemble square root filter, while EnRML uses a stochastic update form, perturbed observations method. Following the introduction of ensemble Kalman smoother (EnKS;

van Leeuwen and Evensen 1996; Evensen and van Leeuwen 2000) for use in history matching by Kjervheim et al. (2011), the iterative forms of smoothers have developed into useful tools by the reservoir-engineering community for history matching reservoir models. Chen and Oliver (2012) proposed an iterative form of EnRML targeted for oil-reservoir modeling, and Bocquet and Sakov (2014) developed the iterative ensemble Kalman smoother (IEnKS), which extends IEnKF using a fixed-lag smoother with an ensemble variational method.

Emerick and Reynolds (2012) introduced the multiple data assimilation scheme (MDA) to improve EnKF estimates for nonlinear cases by assimilating the same data multiple times with the covariance matrix of the measurement errors multiplied by the number of data assimilation. We note that the name "MDA" is somewhat deceiving, as it is simply an application of tempering (Neal 1996). The process of the EnKF with MDA (EnKF-MDA) is based on the idea that a "large jump" between the forecast and analysis states could be reduced by assimilating the same data multiple times with increased measurement errors. MDA yields the same updated mean and covariance as would be obtained from assimilating the same data with the original measurement error covariance and no iterations when errors are Gaussian, and all operators are linear (Emerick and Reynolds 2012). For the nonlinear case, EnKF-MDA partly resolves issues with nonlinearity and leads to smaller bias than a conventional EnKF. Emerick and Reynolds (2013) developed the EnKS with MDA (EnKS-MDA) for reservoir simulations, and Bocquet and Sakov (2014) showed IEnKS with MDA significantly outperforms standard EnKF and EnKS in strongly nonlinear regimes with a simplified model. However, these methods have seen little attention in numerical weather prediction (NWP) communities. While the convergence properties of these methods are unknown, numerical experiments performed by Evensen (2018) suggest they can provide accurate solutions for mildly nonlinear problems.

Lastly, particle filters (PFs) present a purely Bayesian filtering approach for state estimation, which avoids many of the linear/Gaussian assumptions of the above methods. PFs provide a much more general, non-parametric estimate of the model probability density function (PDF), which is advantageous for non-Gaussian problems as long as a sufficient number of ensemble members exist. Nevertheless, these methods can easily diverge when a relatively small number of particles (ensemble members) are adopted for data assimilation; see Bengtsson et al. (2008), Bickel et al. (2008), and Snyder et al. (2008) for discussions on ensemble size requirements for PFs. Several strategies are proposed to overcome this filter collapse and apply PFs to data assimilation problems for operational NWP models very recently. One common effort to avoid filter divergence is to use localization, which restricts the influence of observations to nearby state variables. For example, Poterjoy (2016) introduced the localized PF, which assimilates observations with independent errors sequentially to combine sampled particles from a standard bootstrap PF with prior particles in a manner that satisfies a set of local constraints. Following this work, Poterjoy and Anderson (2016) and Poterjoy et al. (2017, 2019) demonstrate that the local PF works well for high-dimensional systems. For these studies, the authors compare the local PF with EnKFs for a simplified general circulation model and both idealized and real mesoscale convective systems in the Weather Research and Forecasting (WRF) model, respectively. Even more recently, Potthast et al. (2019) applied an alternative localized PF for global weather prediction using the Icosahedral Nonhydrostic Weather and Climate (ICON) model, which marks the first successful test of a PF in an operational framework. These studies provide an incentive to further explore the potential of localized PFs for weather prediction, especially considering the theoretical benefits they pose for assimilating remotely sensed measurements, such as satellite radiance and radar reflectivity, which require nonlinear measurement operators.

In addition to the methods described above, there are some notable developments related to treatment of nonlinearity and non-Gaussianity. For example, Bishop (2016) introduces the GIGG-EnKF algorithm, which retains the accuracy of the EnKF in the Gaussian case while lending it a high degree of accuracy when the forecast and observation uncertainty are gamma or inverse-gamma distributions. When conditions are not suitable for EnKF, such as the distribution of the prior and observation are not Gaussian distribution, and the observation operator is non-linear, Amezcua and Leeuwen (2014) apply a pre-processing step known as Gaussian anamorphosis to obtain state variables and observations that better fulfill the Gaussianity conditions. Fletcher (2010) and Fletcher and Jones (2014) present variants of variational solvers for issues with lognormal and mixed lognormal Gaussian distributed background and observation errors. While many methods have been proposed to deal with such difficult conditions, this study mainly focuses on the tempered iteration approach, which is relatively easy to implement in current NWP systems and can deal with these problems well.

In this study, we discuss EnKF-MDA, EnKS-MDA, E4DVar, 4DEnVar, and the local PF data assimilation methods and their use in applications that require nonlinear measurement operators. We also examine the sensitivity of each method to user-specified parameters, which include ensemble size, covariance localization radius of influence (ROI), inflation coefficients, data assimilation window length (DAW), and the number of iterations and outer loops. The comparisons are conducted with the 40-variable dynamical system introduced in Lorenz (1996, hereafter L96), using numerical experiments performed with conventional EnKF and EnKS techniques as benchmarks. This study provides a necessary first step in understanding the complexity of assimilating remotely-sensed measurements in weather models, which will require appropriate choices for data assimilation methodology going forward.

Three main goals of these experiments are as follows: 1) investigate intrinsic limitations of current data assimilation methodology for applications that require nonlinear measurement operators; 2) compare recently developed methods designed for nonlinear/non-Gaussian applications with those currently used for operational NWP; 3) inform ongoing efforts to design future geophysical modeling systems (e.g., NWP with Hurricane Analysis and Forecast System; HAFS), which will inevitably need to exploit remotely-sensed measurements.

The manuscript is organized in the following manner. In Section 2, we present algorithmic descriptions of each data assimilation method. Section 3, describes settings for data assimilation experiments and results from the cycling experiments. The last section summarizes the main findings of this study and discusses the potential of the methods for real numerical weather prediction.

2. Data Assimilation Methods

In this section, we present the mathematical framework for each method, along with the dynamical system adopted for performing numerical experiments. We use lowercase boldface font to indicate vectors, uppercase boldface font to indicate matrices, and italic font to indicate scalars and nonlinear operators.

In this study, let $\mathbf{x}^f$ be an N_x-dimensional background model forecast; let $\mathbf{y}$ be an N_y-dimensional set of observations; let $\mathbf{H}$ be the tangent linear operator that converts the model state to the observation space; let $\mathbf{R}$ be the $N_y \times N_y$ dimensional observation error covariance matrix; and let $\mathbf{P}$ be the $N_x \times N_x$ dimensional error covariance matrix. Superscript f and a denote forecast and analysis, respectively.

a. EnKF

The EnKF is an approximate but efficient application of the Kalman Filter (Kalman 1960) and explicitly includes the time evolution of error statistics, which operates effectively for moderately nonlinear dynamical systems. In EnKF, $\mathbf{P}$ is represented by ensemble members statistically. There is no need to consider the tangent linear model operator used in KF, so EnKF has many advantages for nonlinear dynamics. The analyzed state $\mathbf{x}^a$ is given by the following Kalman filter equations (e.g., Jazwinski 1970; Gelb et al. 1974)

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^f) \tag{1}$$

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^f\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1} \tag{2}$$

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f(\mathbf{I} - \mathbf{K}\mathbf{H})^{\mathrm{T}} + \mathbf{K}\mathbf{R}\mathbf{K}^{\mathrm{T}} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f. \tag{3}$$

For the ensemble formulation, the covariance matrix $\mathbf{P}$ can be defined as

$$\mathbf{P} = \mathbf{E}\mathbf{E}^{\mathrm{T}}, \tag{4}$$

$$\mathbf{E} = \frac{1}{\sqrt{N_e - 1}}[\delta\mathbf{x}^{(1)} \mid \cdots \mid \delta\mathbf{x}^{(N_e)}], \tag{5}$$

where $\delta\mathbf{x}^{(l)}$ is considered as a perturbation around $\mathbf{x}^{(l)}$, which is the l^{th} member from an ensemble of N_e model states.

The Kalman filtering algorithm requires the computation of $\mathbf{P}^a$ in (3). This process is equivalent to producing an appropriate analysis ensemble or “ensemble update," which has a sample covariance of $\mathbf{P}^a$. For this study, all algorithms requiring an EnKF to update ensemble members use the serial ensemble square-root filter (serial EnSRF; Whitaker and Hamill 2002). In general, this method provides a deterministic update of the ensemble mean and perturbations about the ensemble mean separately in a manner that satisfies the analysis mean and error covariance given by Kalman filter theory. The serial EnSRF assumes an ensemble update of the form

$$\mathbf{E}^a = (\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H})\mathbf{E}^f. \tag{6}$$

Andrews (1968) provides one solution, which involves Kalman gain matrix for perturbations of the form

$$\tilde{\mathbf{K}} = \mathbf{P}^f\mathbf{H}^\mathbf{T}[(\mathbf{H}\mathbf{P}^f\mathbf{H}^\mathrm{T} + \mathbf{R})^{-1/2}]^\mathbf{T}[(\mathbf{H}\mathbf{P}^f\mathbf{H}^\mathrm{T} + \mathbf{R})^{1/2} + \mathbf{R}^{1/2}]^{-1}. \tag{7}$$

If observations are uncorrelated ($\mathbf{R}$ is diagonal), each observation is treated serially, which makes the terms $\mathbf{H}\mathbf{P}^f\mathbf{H}^\mathrm{T}$ and $\mathbf{R}$ scalar. In this case, (3) can be simplified by assuming $\tilde{\mathbf{K}} = \alpha\mathbf{K}$ where α is a scalar value. The α was first derived by Potter (1964) as

$$\alpha = \left(1 + \sqrt{\frac{\mathbf{R}}{\mathbf{H}\mathbf{P}^f\mathbf{H}^\mathrm{T} + \mathbf{R}}}\right)^{-1}. \tag{8}$$

Thus, the serial version requires only the computation of a scalar factor to weight the traditional Kalman gain, and therefore is no more computationally expensive than the EnKF. In this study, observations are assumed to be independent of each other, which makes only the computation of (8) necessary. When assimilating a single observation through this formulation, $\mathbf{K}$ and $\mathbf{H}$ are vectors with N_x dimensions, and $\mathbf{R}$ is scalar. Therefore, for an individual observation, the terms $\mathbf{P}^f\mathbf{H}^\mathbf{T}$ and $\mathbf{H}\mathbf{P}^f\mathbf{H}^\mathrm{T}$ reduce to scalars and can be computed even if the measurement operator is fully nonlinear, which is done by applying this operator on each ensemble member before calculating sample statistics.

b. EnKS

The EnKS operates by storing ensemble members at past times and then modifying them by a gain matrix that considers observations at the current time. Whitaker and Compo (2002) introduced a serial ensemble square-root smoother (serial EnSRS), which uses Monte-Carlo estimates of forecast-analysis error cross-covariances needed to compute the Kalman smoother gain matrix.

While they applied the serial EnSRS to the fixed-lag Kalman smoother proposed by Cohn et al. (1994), in this study, we apply it as a fixed-interval Kalman smoother.

Here, define a subscript notation $m|n$ to indicate a quantity at observation time m, which incorporates knowledge of all observations up to and including time n. In this notation, (1) can be expressed as

$$\bar{\mathbf{x}}^{a}_{k|k} = \bar{\mathbf{x}}^{f}_{k|k-1} + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^{f}_{k|k-1}). \tag{9}$$

In the serial square-root smoother, we use $\mathbf{P}^{f}_{(m,n)}$ to denote a cross-covariance matrix between variables at times m and n. The gain matrix $\mathbf{K}$ involves the forecast error cross-covariance matrix $\mathbf{P}^{f}_{(k,k-l)}$ between $\mathbf{x}^{f}_{k|k-1}$ and $\mathbf{x}^{f}_{k-l|k-1}$.

$$\mathbf{K} = \mathbf{P}^{f}_{(k,k-l)}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^{f}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}, \tag{10}$$

where

$$\mathbf{P}^{f} = \mathbf{E}^{f}_{k|k-1}\mathbf{E}^{f\mathrm{T}}_{k|k-1} \tag{11}$$

$$\mathbf{P}^{f}_{(k,k-l)} = \mathbf{E}^{f}_{k|k-1}\mathbf{E}^{f\mathrm{T}}_{k-l|k-1}. \tag{12}$$

In the formulation of Cohn et al. (1994), this quantity is computed directly using the dynamical model because they developed the fixed-lag smoother without ensembles. On the other hand, the fixed-lag smoother with ensembles uses the dynamical model only when creating the background model forecast (Whitaker and Compo 2002). This idea can be directly implemented to the fixed-interval smoother. Note that the basic equations for the lag-0 implementation are identical to those of the serial EnSRF.

c. Multiple data assimilation (MDA)

Emerick and Reynolds (2012) introduced the MDA scheme, which assimilates the same data multiple times using an inflated covariance matrix of the measurement errors. They proved the equivalence between single and multiple data assimilations for the linear-Gaussian case. Although MDA contains approximations for the fully nonlinear case and the equivalence does not hold for the nonlinear case, MDA benefits from the inclusion of smaller incremental ensemble corrections.

When the same set of observations are assimilated N_a times, the inflated measurement error covariance matrix is used in (2),

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^{\mathrm{T}} (\mathbf{H}\mathbf{P}^f \mathbf{H}^{\mathrm{T}} + \alpha_i \mathbf{R})^{-1}, \tag{13}$$

where

$$\sum_{i=1}^{N_a} \frac{1}{\alpha_i} = 1. \tag{14}$$

Note that in this paper, we use $\alpha_i = N_a$ for $i = 1, \ldots, N_a$ for all experiments with MDA. Rommelse (2009) and Emerick and Reynolds (2012) suggest that when the assimilation of accurate data in non-Gaussian regimes requires a "large jump" between the forecast and analysis state, the magnitude of the jump can be overestimated by linear updates. This limitation of Gaussian data assimilation techniques is observed frequently for the assimilation of all-sky radiance measurements in weather models, which is one of the reasons to motivate the use of observation error inflation (e.g., Minamide and Zhang 2017) and other ingeneous approaches as descibed in Section 1. By using an inflated error covariance, a potentially large spurious update in the state vector is avoided. Going a step further, iterative techniques like MDA replace single updates with a series of smaller updates, which can correct filter or smoother updates that are too large.

In summary, the ensemble formulation of a fixed-interval serial EnSRS, with and without MDA, are realized by the following procedures. For DAW length $l = 0$, the serial EnSRS reduces to the serial EnSRF, and for $N_a = 1$, each iterative data assimilation cycle with MDA reduces to a single-step data assimilation scheme, such as standard EnKF and EnKS.

Algorithm 1: EnKS with MDA cycle

Function `EnKS-MDA_cycle`:

  for t = 1:time **do**

    if t is at the end of DAW **then**

      $t_0 \leftarrow t - l$

      for i = 1:iteration N_a **do**

        for k = 0:DAW length l **do**

          $\mathbf{x}^a_{t_0|t_0+k} \leftarrow$ Serial_EnSRS($\mathbf{x}^f_{t_0|t_0+k-1}$, $\mathbf{x}^f_{t_0+k|t_0+k-1}$, $\mathbf{y}_{t_0+k}$, $\alpha_i \mathbf{R}$)

          $\mathbf{x}^f_{t_0|t_0+k} \leftarrow \mathbf{x}^a_{t_0|t_0+k}$

        $\mathbf{x}^f_{t_0|t_0-1} \leftarrow \mathbf{x}^a_{t_0|t_0+l}$

      for $m = 1{:}N_e$ **do**

        $\mathbf{x}^{f(m)}_{t+1|t} \leftarrow M\mathbf{x}^{a(m)}_{t_0|t}$

    else

      for i = 1:iteration N_a **do**

        $\mathbf{x}^a_{t|t} \leftarrow$ Serial_EnSRS($\mathbf{x}^f_{t|t-1}$, $\mathbf{x}^f_{t|t-1}$, $\mathbf{y}_t$, $\alpha_i \mathbf{R}$)

        $\mathbf{x}^f_{t|t-1} \leftarrow \mathbf{x}^a_{t|t}$

      for $m = 1{:}N_e$ **do**

        $\mathbf{x}^{f(m)}_{t+1|t} \leftarrow M\mathbf{x}^{a(m)}_{t|t}$

  return

Algorithm 2: Serial EnSRS

Function `Serial_EnSRS(`$\mathbf{x}^f_{t-k|t-1}$, $\mathbf{x}^f_{t|t-1}$, $\mathbf{y}$, $\mathbf{R}$`):`

  for $j = 1{:}N_y$ **do**

    $\mathbf{E}^f_{t-k|t-1} = \frac{1}{\sqrt{N_e-1}}[\delta\mathbf{x}^{f(1)}_{t-k|t-1} \mid \cdots \mid \delta\mathbf{x}^{f(N_e)}_{t-k|t-1}]$

    $\mathbf{E}^f_{t|t-1} = \frac{1}{\sqrt{N_e-1}}[\delta\mathbf{x}^{f(1)}_{t|t-1} \mid \cdots \mid \delta\mathbf{x}^{f(N_e)}_{t|t-1}]$

    $\mathbf{P}^f = \mathbf{E}^f_{t|t-1}\mathbf{E}^{f\mathrm{T}}_{t|t-1}$

    $\mathbf{P}^f_{(t-k,t)} = \mathbf{E}^f_{t-k|t-1}\mathbf{E}^{f\mathrm{T}}_{t|t-1}$

    $\mathbf{K} = \mathbf{P}^f_{(t-k,t)}\mathbf{H}^{(j)\mathrm{T}}[\mathbf{H}^{(j)}\mathbf{P}^f\mathbf{H}^{(j)\mathrm{T}} + \mathbf{R}^{(j)}]^{-1}$

    $\bar{\mathbf{x}}^a_{t-k|t} = \bar{\mathbf{x}}^f_{t-k|t-1} + \mathbf{K}(\mathbf{y}^{(j)} - \mathbf{H}^{(j)}\mathbf{x}^f_{t|t-1})$

    $\alpha = (1 + \sqrt{\frac{\mathbf{R}^{(j)}}{\mathbf{H}^{(j)}\mathbf{P}^f\mathbf{H}^{(j)\mathrm{T}} + \mathbf{R}^{(j)}}})^{-1}$

    $\tilde{\mathbf{K}} = \alpha\mathbf{K}$

    $\mathbf{E}^a_{t-k|t} = \mathbf{E}^f_{t-k|t-1} - \tilde{\mathbf{K}}\mathbf{H}^{(j)}\mathbf{E}^f_{t|t-1}$

    $\mathbf{x}^a_{t-k|t} = \bar{\mathbf{x}}^a_{t-k|t} + \mathbf{E}^a_{t-k|t}$

    $\mathbf{x}^f_{t-k|t-1} \leftarrow \mathbf{x}^a_{t-k|t}$

  return $\mathbf{x}^a_{t-k|t}$

d. E4DVar and 4DEnVar

In this section, the equations of 4DVar, E4DVar, 4DEnVar are introduced briefly. For further details on these methods, we encourage readers to review the mathematical descriptions in Liu et al. (2009), Poterjoy and Zhang (2015), and Bannister (2017). The 4DVar method seeks a solution that minimizes the misfit of a control variable to the background state $\mathbf{x}^f_0$ at $t = 0$ and observations $\mathbf{y}_t$ at times $t = 0, 1, 2, \ldots, \tau$. The minimization is carried out with respect to increments $\delta\mathbf{x}_0$ from $\mathbf{x}^f_0$ (Courtier et al. 1994). The cost function is expressed as the sum of background (J_b) and observation

(J_o) terms:

$$J(\delta\mathbf{x}_0) = J_b(\delta\mathbf{x}_0) + J_o(\delta\mathbf{x}_0)$$
$$= \frac{1}{2}\delta\mathbf{x}_0^{\mathrm{T}}\mathbf{B}^{-1}\delta\mathbf{x}_0 + \frac{1}{2}\sum_{t=0}^{\tau}(\mathbf{H}_t\mathbf{M}_t\delta\mathbf{x}_0 - \mathbf{d}_t)^{\mathrm{T}}\mathbf{R}_t^{-1}(\mathbf{H}_t\mathbf{M}_t\delta\mathbf{x}_0 - \mathbf{d}_t), \tag{15}$$

where $\mathbf{B}$ is the background error covariance and $\mathbf{M}_t$ is the tangent linear model operator. The vector $\mathbf{d}_t$ contains the innovations at each time along a model trajectory from $\mathbf{x}_0^f$ and is given by

$$\mathbf{d}_t = \mathbf{y}_t - H_t[M_t(\mathbf{x}_0^f)], \tag{16}$$

where M_t and H_t are the nonlinear forecast model and observation operators, respectively. In practice, $\delta\mathbf{x}_0$ is replaced with $\mathbf{U}\mathbf{v}$, where $\mathbf{v}$ is the new control variable, and $\mathbf{U}$ is a square root of the background error covariance matrix ($\mathbf{B} = \mathbf{U}\mathbf{U}^{\mathrm{T}}$) (Lorenc 2003). The cost function in the control variable space and the gradient of the cost function with respect to the control variables become:

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{v} + \frac{1}{2}\sum_{t=0}^{\tau}(\mathbf{H}_t\mathbf{M}_t\mathbf{U}\mathbf{v} - \mathbf{d}_t)^{\mathrm{T}}\mathbf{R}_t^{-1}(\mathbf{H}_t\mathbf{M}_t\mathbf{U}\mathbf{v} - \mathbf{d}_t) \tag{17}$$
$$\nabla_{\mathbf{v}}J = \mathbf{v} + \sum_{t=0}^{\tau}\mathbf{U}^{\mathrm{T}}\mathbf{M}_t^{\mathrm{T}}\mathbf{H}_t^{\mathrm{T}}\mathbf{R}_t^{-1}(\mathbf{H}_t\mathbf{M}_t\mathbf{U}\mathbf{v} - \mathbf{d}_t) \tag{18}$$

For E4DVar and 4DEnVar, using a similar substitution described above, $\delta\mathbf{x}_0$ is separated into two terms to include a hybrid covariance in the variational cost function. For NWP applications, the ensemble contribution of the hybrid covariance is often much greater than the static covariance (Kleist and Ide 2015), however, such a choice is directly dependent upon the quality of ensemble, ensemble size, and model error. For the L96 model, Poterjoy and Zhang (2015) found the static error covariance to have a major impact only when an imperfect model is used for data assimilation, which is not explored in the current study. Therefore, we omit the use of a static error covariance to reduce the number of parameters to examine for this study. As a result, we have

$$\delta\mathbf{x}_0 = \delta\mathbf{x}_0^e = \mathbf{U}^e\mathbf{v}^e, \tag{19}$$

where $\delta\mathbf{x}_0^e$ is the increment resulting from the ensemble-estimated covariance. As described in Buehner (2005), $\mathbf{U}^e$ can then be written

$$\mathbf{U}^e = [\mathbf{e}^{(1)} \mid \cdots \mid \mathbf{e}^{(N_e)}] \tag{20}$$

$$\mathbf{P} \circ \mathbf{C} = \mathbf{U}^e \mathbf{U}^{e\mathbf{T}}, \tag{21}$$

$$\mathbf{e}^{(n)} = \sqrt{\frac{1}{N_e - 1}} \times \mathrm{diag}(\mathbf{x}_0^{f(n)} - \bar{\mathbf{x}}_0^f)\mathbf{C}^{\frac{1}{2}}, \qquad (n = 1, 2, \ldots, N_e), \tag{22}$$

where $\circ$ indicates element wise multiplication, and $\mathbf{C}$ is the correlation matrix used for localizing the ensemble covariance. From these equations, the cost function and the gradient of E4DVar are found by substituting $\mathbf{U}^e$ for $\mathbf{U}$ and $\mathbf{v}^e$ for $\mathbf{v}$ in (17) and (18). Using an ensemble forecast stored at each observation time in DAW, $\mathbf{M}_t\mathbf{U}^e$ can be rewritten as

$$\begin{aligned} \mathbf{M}_t\mathbf{U}^e &= [\mathbf{M}_t\mathbf{e}^{(1)} \mid \cdots \mid \mathbf{M}_t\mathbf{e}^{(N_e)}] \\ &= [\hat{\mathbf{e}}_t^{(1)} \mid \cdots \mid \hat{\mathbf{e}}_t^{(N_e)}] \end{aligned} \tag{23}$$

$$\begin{aligned} \hat{\mathbf{e}}_t^{(n)} &= \sqrt{\frac{1}{N_e - 1}} \times \mathrm{diag}(\mathbf{x}_t^{f(n)} - \bar{\mathbf{x}}_t^f)\mathbf{C}^{\frac{1}{2}} \\ &= \sqrt{\frac{1}{N_e - 1}} \times \mathrm{diag}(M_t(\mathbf{x}_0^{f(n)}) - \overline{M_t(\mathbf{x}_0^f)})\mathbf{C}^{\frac{1}{2}} \end{aligned} \tag{24}$$

By substituting (23) into (17) and (18), the 4DEnVar cost function and the gradient can be expressed without the tangent and adjoint model.

Note that while E4DVar uses tangent linear and adjoint models to propagate a localized error covariance through the DAW, 4DEnVar requires the localization of time covariances. Most previous studies use the same correlation matrix at each time thus ignoring the complexity of introducing a localization of time-dependent covariance (LTC) (Liu et al. 2009; Buehner et al. 2010; Liu and Xiao 2013; Fairbairn et al. 2014; Poterjoy and Zhang 2015).

The method also allows for the use of either the nonlinear operator H_t or the tangent linear operator $\mathbf{H}_t$ in its place. This study explores both approaches in 4DEnVar experiments to identify which option presents the largest advantage for nonlinear operators. To perform the localization, we calculate the tangent linear operator $\mathbf{H}_t$ at each time and use it to propagate a localized error covariance through the DAW. Moreover, this study re-runs the ensemble in outer loops for 4DEnVar, despite the fact that it is prohibitively costly for weather applications. This step is done to allow for a more direct comparison with incremental E4DVar with outer loops.

To form a hybrid analysis, the variational solution is typically taken as the posterior mean and posterior perturbations from an EnKF are recentered about this solution at the middle of the time window (Zhang et al. 2009; Poterjoy et al. 2014). This approach is more consistent with the methodology adopted at major NWP modeling centers (Bannister 2017). For the current study, we instead add posterior perturbations to the mean analysis at the end of each DAW. This option has a number of advantages, namely, the EnKF assimilates measurements at the appropriate times over an assimilation window, thus providing an EnKF posterior mean that is theoretically equivalent to the 4DVar posterior mean in the absence of sampling error and nonlinearity. It also permits a more direct comparison of smoothers and filters explored in this study.

In summary, the ensemble formulation of E4DVar and 4DEnVar are realized by the following procedures.

Algorithm 3: Ensemble/variational hybrid data assimilation without static error covariance

Function `ensemble_variational_hybrid(`$\mathbf{U}^e$, $\mathbf{x}_0^f$, $\mathbf{y}$, $\mathbf{R}$`):`

  if *4DEnVar w/o LTC* **then**

    $\mathbf{H}_t \leftarrow H_t$

  while *Outer Loop* **do**

    $\mathbf{d}_t = \mathbf{y}_t - H_t[M_t(\mathbf{x}_0^f)]$

    while *Inner Loop* **do**

      switch *Hybrid* **do**

        case *E4DVar*

          $\mathbf{D}_t \leftarrow \mathbf{M}_t \mathbf{U}^e$

        case *4DEnVar w/ LTC .or. 4DEnVar w/o LTC*

          $\mathbf{D}_t \leftarrow [\hat{\mathbf{e}}_t^{(1)} \mid \cdots \mid \hat{\mathbf{e}}_t^{(N_e)}]$

      $J(\mathbf{v}^e) = \frac{1}{2}\mathbf{v}^{e\mathrm{T}}\mathbf{v}^e + \frac{1}{2}\sum_{t=0}^{\tau}(\mathbf{H}_t\mathbf{D}_t\mathbf{v}^e - \mathbf{d}_t)^{\mathrm{T}}\mathbf{R}_t^{-1}(\mathbf{H}_t\mathbf{D}_t\mathbf{v}^e - \mathbf{d}_t)$

      $\nabla_{\mathbf{v}^e} J = \mathbf{v}^e + \sum_{t=0}^{\tau}\mathbf{D}_t^{\mathrm{T}}\mathbf{H}_t^{\mathrm{T}}\mathbf{R}_t^{-1}(\mathbf{H}_t\mathbf{D}_t\mathbf{v}^e - \mathbf{d}_t)$

      $\mathbf{v}^e = \arg\min(J(\mathbf{v}^e))$

    $\mathbf{x}_0^f \leftarrow \mathbf{x}_0^f + \mathbf{U}^e\mathbf{v}^e$

  $\mathbf{x}_0^a \leftarrow \mathbf{x}_0^f$

  return $\mathbf{x}_0^a$

e. The local PF

The current study uses the local PF proposed by Poterjoy et al. (2019). For simplicity, this section highlights important aspects of the local PF that are relevant to the comparisons performed in this study. Our experiments take advantage of additional regularization, tempering, and hybrid strategies that are unique to the local PF, which are briefly discussed in this section. For full details on this methodology, we refer readers to Poterjoy (2021).

The local PF assimilates observations serially, performing a bootstrap PF update for particles projected onto the current observation in the sequence, followed by a model-space update. For a given observation y, the model-space update replaces the standard bootstrap re-sampling step with one that merges sampled particles and prior particles:

$$\mathbf{x}_y^n = \overline{\mathbf{x}}_y + \mathbf{r}_1 \circ (\mathbf{x}^{k_n} - \overline{\mathbf{x}}_y) + \mathbf{r}_2 \circ (\mathbf{x}^{k_n} - \overline{\mathbf{x}}_y), \tag{25}$$

where $\mathbf{x}_y^n$ is an updated particle, $\mathbf{x}^n$ is the n^{th} prior particle, $\mathbf{x}^{k_n}$ is the n^{th} sampled particle, $\overline{\mathbf{x}}_y$ is the localized posterior mean based on importance weights that consider all observations up to y, and $\mathbf{r}_1$ and $\mathbf{r}_2$ are derived to satisfy the posterior mean and variance of marginals. The sampled particles are selected from a bootstrap re-sampling of past updated particles using a cumulative distribution formed by weights calculated from particle likelihoods for y. In general, the posterior particles formed from linear combinations of the sampled and prior particles are localized, because $\mathbf{r_1}$ and $\mathbf{r_2}$ are calculated based on localized moments.

Poterjoy et al. (2019) provide several improvements to the Poterjoy (2016) local PF, which are aimed at preventing particle weight collapse. In addition, Poterjoy (2021) introduces regularization and tempering methodology to further improve filter performance when sampling error is large. In short, regularization raises particle weights to a power β, which is pre-determined to yield marginal particle weights that have a specified "effective sample size," similar to the methodology described in Poterjoy et al. (2019). Regularization acts as a heuristic means of preventing weight collapse, similar to observation error inflation. It provides a strategy for assimilating observations through tempered iterations (Neal 1996), each with a unique set of β coefficients. Unlike regularization, tempering does not introduce bias in the posterior estimate.

The method also benefits from the use of a mixing parameter, γ, to increase particle diversity in the vicinity of observations. As described in Poterjoy (2021), $\mathbf{r_1}$ in (25) is multiplied by γ, which introduces a smooth "jittering" of particles. The coefficients in $\mathbf{r_2}$ are then modified so that the first two posterior moments are still maintained.

3. Cycling data assimilation experiments

We perform separate sets of data assimilation experiments to investigate limitations for nonlinear applications and examine the sensitivity of the methods to user-specified parameters. These parameters include the number of iterations, DAW, ensemble size, ROI, inflation, and measurement operators. The first two sets of experiments focus primarily on key parameters for smoothers, which are known to be sensitive to nonlinearity in model dynamics and measurement operators. These parameters are the number of iterations and DAW length. The third set of experiments focuses more broadly on the comparison between filters and smoothers. For this purpose, we select three types of observation networks, each differing primarily in choice of measurement operator. The system parameters for each of these cases are summarized in Table 1.

a. Experimental design

1) Model

We examine several aspects of the data assimilation methods by performing idealized numerical experiments with the L96 model (Lorenz 1996; Lorenz and Emanuel 1998). The model consists of variables x_i for $i = 1,2,\ldots,N_x$, which are equally spaced on a periodic domain. The variables are evolved in time using the set of differential equations,

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \tag{26}$$

with cyclic boundaries: $x_{i+N_x} = x_i$ and $x_{i-N_x} = x_i$. We integrate (26) forward numerically using the fourth-order Runge-Kutta method with a time step of 0.05 [units defined arbitrarily as 6 h; see Lorenz (1996)]. For this study, we fix N_x at 40 and use $F = 8.0$, which causes the model to behave chaotically.

2) Observations

In this study, we create observation networks of $N_y = 10$, $N_y = 15$, and $N_y = 20$ observations that are evenly spaced on model grid points. Note that for the case $N_y = 15$, we line up the observation points so that they were evenly distributed (i.e., 1, 4, 6, 9, 12, 14, 17, 20, 22, 25, 28, 30, 33, 36, 39). We simulate measurements every time step (6 h) by selecting values from a truth simulation, applying one of the operators discussed below, then adding uncorrelated Gaussian errors selected from $N(0,\sigma_y^2 I)$, where σ^2 is the measurement error variance.

Experiments include three forms of measurement operator. The "Linear Case" uses an H that selects model variables to be directly observed; i.e., $H(\mathbf{x}) = \hat{\mathbf{x}}$, where $\hat{\mathbf{x}}$ is a subset N_y variables in $\mathbf{x}$ chosen by H. The "Nonlinear Case 1" extends H to be quadratic: $H(\mathbf{x}) = \hat{\mathbf{x}} \circ \hat{\mathbf{x}}$. The "Nonlinear Case 2" introduces log and absolute value operators to the interpolated values: $H(\mathbf{x}) = \log[ABS(\hat{\mathbf{x}})]$, where ABS indicates the absolute value of each element. The second and third operators produce weak and strong nonlinearities, respectively. Note that we apply a simple gross error check for the third measurement operator to prevent observations from being assimilated if the value of $ABS(\hat{\mathbf{x}})$ is extremely small. Observation error standard deviations are set to $\sigma_y = 1.0$ for the first two experiments, but reduced to $\sigma_y = 0.1$ for the third case to compensate for the smaller information content provided by this observation network.

3) Observation timeline and verification

Observations are assimilated over a 3650-day period, and root-mean-square errors (RMSEs) from the last 3550 days are used to quantify the accuracy of the posterior analyses. The first 100 days of data assimilation act as a spinup period to allow members time to reach quasi-steady posterior solutions for the given setup of the model and observation network.

In the first sets of experiments described below, we perform direct comparisons of the different smoothers used for this study. For these experiments, we calculate RMSEs at the beginning of the DAW (smoother solution), because it more directly indicates how much information is being extracted from observations at future times. For experiments shown later in this section, which compare different forms of smoothers and filters, we calculate RMSEs at the end of the DAW (filter solution).

4) Treatment of sampling errors

Potential sources of bias in the estimation of the posterior include small ensemble sizes relative to the state dimensions, model errors, nonlinearities, and assumptions used to form data assimilation algorithms. Therefore, heuristic covariance localization strategies are needed to reduce noise introduced from ensemble error approximations by performing a Schur product between this matrix and an empirically defined correlation matrix with a tunable length scale parameter, or ROI. For this purpose, we use the fifth-order correlation function given by Eq. (4.10) of Gaspari and Cohn (1999).

The posterior covariance is inflated by replacing ensemble perturbations with linear combinations of posterior and prior perturbations, which is known as a covariance relaxation method (Zhang et al. 2004):

$$\mathbf{x}_n^{\prime a} \leftarrow (1-\alpha)\mathbf{x}_n^{\prime a} + \alpha \mathbf{x}_n^{\prime f}. \tag{27}$$

The α in (27) is called the “relaxation coefficient” and ranges from 0 to 1, where $\alpha = 0$ implies no inflation. We adopt this inflation strategy to remain consistent with Poterjoy and Zhang (2015), who perform a similar comparison of ensemble data assimilation algorithms, including hybrid covariance forms of E4DVar and 4DEnVar.

As previously stated, the local PF uses a mixing parameter to maintain particle diversity during updates. While this approach is effective at preventing filter divergence with small ensembles, it does not directly increase prior or posterior error variance in the same manner as relaxation. Similar to the α used in the relaxation method the coefficient γ is a scalar between 0 and 1. It further mixes prior particles and resampled particles everywhere particles are updated in state space, including in the vicinity of measurements.

b. Results

1) Sensitivity to the number of outer iterations

The variational and MDA techniques present different iterative strategies for coping with nonlinearity in model dynamics and measurement operators. For the first set of experiments, we explore the sensitivity of these methods to the number of iterations. In addition to providing a direct comparison of different smoothers for a nonlinear application, these experiments help motivate choices for iteration number in the filter/smoother comparisons that follow. As previously stated, we also explore the advantage of LTC, which is a localization of the ensemble covariance at each observation time in the window calculated with the tangent linear operator $\mathbf{H}_t$ at each time for nonlinear operators.

Figure 1 shows mean RMSEs of EnKS-MDA, E4DVar, 4DEnVar with LTC, and 4DEnVar without LTC from experiments with Nonlinear Case 1. Ensemble size N_e, relaxation coefficient α, and DAW are fixed at 10, 0.3, and 24 h, respectively. We find this window length to be sufficient for exploring sensitivity to outer loops without adding computational cost. We do not show results using Nonlinear Case 2 because all methods tested in this study (other than the PF) experience filter divergence when measurements are simulated with this operator. These results are discussed in the filter/smoother comparisons below.

For the observation networks tested in this study, we find that increasing the number of iterations has little impact on mean error for EnKS-MDA. For E4DVar and 4DEnVar, however, we confirm that multiple outer loops are required for optimal performance. Under various circumstances, outer loops are also needed to prevent filter divergence with the nonlinear measurement operator. For example, E4DVar with ROI fixed at 1 and a single outer loop shows a worse score than with multiple iterations. We also find that the minimum number of outer loops required to prevent filter divergence is sensitive to ROI. E4DVar experiments using an ROI of 3 and 5 require 2 and 3 outer iterations, respectively. Nevertheless, the improvements of multiple iterations beyond these numbers becomes negligible once a sufficient number is reached.

We also find E4DVar to be more stable than 4DEnVar for the tested observation networks. Recall, this method uses the tangent linear model to propagate increments along a nonlinear trajectory to future times, and its adjoint to propagate sensitivity gradients backward from observation times to the beginning of the DAW. The trajectory is updated between outer iterations to ensure that values propagated by the tangent linear and adjoint remain small enough for linear approximations to remain valid. In addition, the input of ensemble error covariance at a single time in this process (at the beginning of the DAW) greatly simplifies the removal of spurious error correlations through localization (Fairbairn et al. 2014; Poterjoy and Zhang 2015). For this reason, we find configurations of 4DEnVar that use LTC to be more stable than configurations without LTC. Based on this finding, we use this strategy for all remaining 4DEnVar experiments.

2) Smoother performance as a function of data assimilation window length

Several of the methods examined in this study are smoothers, which are sensitive to the choice of DAW. For the next set of experiments, we compare mean RMSEs of EnKS, EnKS-MDA, E4DVar, and 4DEnVar as a function of DAW (Fig. 2). As stated above, the verification for these experiments focuses on the posterior smoothing density; i.e., the analysis at the beginning of the DAW. For these experiments, we fix the ensemble size N_e, relaxation coefficient α, and ROI at 10, 0.3, and 3, respectively. The number of iterations (MDA) and outer loops (Var) are both set to 3. These decisions are based on results from the previous set of experiments, showing little benefit beyond 3 iterations for chosen model and observation networks. As we revisit later, in experiments with the Nonlinear Case 1, the observation value is closer to the truth all the time, making an order of RMSEs magnitude smaller than with the Linear Case.

We start by examining the impact of MDA on the EnKS. Our experiments show that MDA provides slight benefits over non-iterative configurations, even at DAW length $l = 0$ h and linear H (Fig. 2a). Note that EnKS is identical EnKF for this DAW length, so no benefits are expected from the iterations. One possible reason for the difference in skill between EnKS and EnKS-MDA at DAW length $l = 0$ h is due to small differences in how ensemble perturbations are adjusted through iterative steps. For linear cases with Gaussian prior, MDA yields the same posterior mean and covariance as would be obtained without iterations. As suggested by Rommelse (2009), the extra uncertainty included in measurements during each iteration ensures that adjustments from prior to posterior values are dampened, which is beneficial when linear updates overestimate the true impact of measurements that relate nonlinearly to model variables. Therefore, MDA provides an opportunity for the EnKF to remove over-adjustments that may occur during previous iterations. We suspect that a combination of serial processing of observations and iterative updates of members leads to slight improvements in how the EnKF samples from the posterior density, which is assumed to be non-Gaussian because of the nonlinear model. This finding explains why the MDA approach yields small improvements in posterior estimates over successive data assimilation steps, which is also explored later.

The advantage of the EnKS-MDA over the EnKS with the DAW length $l > 0$h is shown in both the linear and nonlinear cases. For both experiments, the MDA scheme resolves issues with the nonlinearlity of the model and observation measurement operators in DAW. EnKS is stable even with the longer DAW, but the quality of the analysis starts to degrade as the DAW length is increased beyond a certain point, because sampling error increases as the DAW become longer. Compared to 4DEnVar, EnKS is more stable with longer DAW. This indicates that the forecast error covariance matrix used for smoother is approximated more accurately by cross-covariance matrix ($\mathbf{P}^f_{(k,k-l)}$) in EnKS than by ensemble-based error covariance in 4DEnVar. Unlike the variational methods, the EnKS samples directly from the smoothing density rather than using a hybrid strategy of re-centering EnKF perturbations about a variational solution. Furthermore, the 4DEnVar experiment contains higher RMSEs than E4DVar because of the difficulty required in removing sampling errors from temporal error covariances when N_e is small (Fairbairn et al. 2014; Poterjoy and Zhang 2015).

3) Filter performance

In this section, we present results from experiments that examine the sensitivity and limitations of EnKF, EnKF-MDA, EnKS, EnKS-MDA, E4DVar, 4DEnVar, and the local PF to ROI, relaxation coefficient α, PF mixing coefficient γ, and the observation measurement operators. For all experiments, DAW for EnKS, EnKS-MDA, E4DVar, and 4DEnVar is set to 24 h, and the number of iterations and outer loops are set to 3. For the local PF, the regularization operates only when the effective ensemble size N_{eff} falls below a target value of N_{eff}. The target N_{eff} is fixed at $N^t_{\text{eff}} = 0.5 \times N_e$ for all experiments. We define filter divergence objectively by flagging configurations that produced 100-cycle average RMSEs larger than 2 with NA for "not available" in the figures.

Figure 3 shows mean RMSEs from the experiment with the Linear Case. Results from all methods, which use a fixed ensemble size N_e of 10, are displayed in charts that show RMSE as a function of tunable variables used to reduce the impact of sampling error. For example, Fig. 3 demonstrates that the optimal ROI and α are comparable for EnKF, EnKF-MDA, EnKS, EnKS-MDA, E4DVar, and 4DEnVar. In most cases, the optimal scores are typically found near values that lead to filter divergence. RMSEs from the local PF are slightly worse due to the small number of particles used in these experiments. Figure 4 shows results from experiments with the same settings except N_e is increased to 40. As expected, all methods become more stable and require less localization (larger ROI) and less inflation (smaller α and γ) as N_e increases. Comparing the results of the local PF from Fig. 3 and 4, it is clear that the larger ensemble size is required for the local PF to outperform the methods with a Gaussian prior with the tested observation network. EnKS shows clearly better performances than EnKF, and MDA makes EnKF and EnKS slightly improved, even with a linear measurement operator because of the reason mentioned in section 3.b.1.

Results from Nonlinear Case 1 experiments using $N_e = 10$ are shown in Fig. 5. Unlike experiments with the Linear operator, filter divergence occurs without setting strict limits on ROI and inflation coefficients for all methods. Despite the nonlinear measurement operator in these experiments, we find no benefits from the assimilation methods designed specifically for non-Gaussian applications, namely EnKF-MDA and the local PF. We believe this result occurs because of the accuracy and frequency at which these measurements are collected. For model variables that can reach magnitudes of $O(10)$, measuring the square of these variables with an error variance of 1 yields highly accurate information for characterizing the posterior. This factor, combined with the frequency of these measurements lead to prior and posterior members that remain close to the truth at all times, thus making Gaussian assumptions more valid. We revisit this property of the Nonlinear Case 1 measurement operator in the next section.

These experiments also continue to show clear benefits of E4DVar and 4DEnVar over EnKF, both in terms of stability and accuracy. We hypothesize that the 4D data assimilation methods are less sensitive to sampling noise, which becomes the dominant source of bias in mildly nonlinear regimes. Likewise, we find E4DVar to be more stable than 4DEnVar when N_e is small, owing mostly to the localization strategy adopted by this method. We note that all algorithms approach similar RMSEs as ensemble size increases; i.e., Fig. 6 shows results with $N_e = 40$ for the same observation network. The reason why E4DVar and 4DEnVar are more stable than EnKS is due to the small number of ensembles and the nonlinear observations that prevent from accurately estimating of the cross-covariance matrix in the Serial EnSRS.

Figure 7a shows the mean RMSEs from experiments of the local PF that use measurements simulated with Nonlinear Case 2 and $N_e = 40$. For this configuration, filter divergence occurs in all methods except the local PF, owing to the strong nonlinearity in the measurement operator. This observation network presents a case where nonlinearity in the application becomes a much larger factor than sampling error in ensemble-estimated prior and posterior distributions. Even with $N_e = 100$, the Gaussian-based methods fail to provide stable solutions despite the potentially large amount of information contained in these measurements, as indicated by the low RMSEs in the local PF posterior (Fig. 7b). Since the local PF makes no parametric assumptions about prior densities, non-Gaussian observation-space priors, which are produced by nonlinear measurement operators, do not have a negative impact on the filter. Therefore, it can continue to extract information from the observation network regardless of nonlinearity in H. These results confirm past studies, showing that local PF provides benefits when N_e is sufficiently large or when the observation operator is strongly nonlinear. It also demonstrates limitations in iterative techniques for cases where the observation function is quadratic and the posterior may be bimodal.

4) Filter performance for sparse observation networks

Using the mildly nonlinear observation operator (Nonlinear Case 1), we investigate the behavior of each method for increasingly sparse observation networks. These experiments use an observation frequency of 24 h, which is increased from 6 h in previous experiments, and N_y = 20, 15, and 10 for equally-spaced measurements at each observation time. We also fix the DAW for smoothers at 48 h; see Table 1 for full summary. These results are summarized in Figs. 8 – 10 using the same graphics adopted in the previous section comparing filter performance.

Compared to EnKF, the performance of EnKS becomes slightly worse for these observation networks. As discussed in Evensen and van Leeuwen (2000), the EnKS differs from the EnKF by computing updates of the model parameters using all the observations in DAW simultaneously rather than using recursive updates in time. Therefore, with these settings, the recursive updates of EnKF keep the model solutions close to the truth at any given time during the experiment, and operate on marginal densities that are relatively close to Gaussian at any given time. While posterior marginals of the smoothing density are expected to be close to Gaussian at the beginning of the DAW (Morzfeld and Hodyss 2019), marginals near the end of the DAW can evolve non-Gaussian characteristics because of nonlinearity in the model.

The benefits of MDA for EnKF are clearly shown in Fig. 8 and Fig. 9. For suboptimal configurations of the EnKF, prior members exhibit a larger variance thus allowing nonlinearity in H to become a significant source of bias for Gaussian methods. Therefore, the optimal EnKF configuration remains almost the same with MDA, but the set of parameters over which the filter remains stable becomes larger than that of the standard EnKF. For these observation networks, careful choices of ROI and α are sufficient for mitigating bias caused by Gaussian assumptions, but MDA helps prevent filter divergence when these parameters are improperly chosen.

For a long DAW (48 h) E4DVar becomes more stable than 4DEnVar with $N_y = 20$ (Fig. 8), but both methods diverge when observation density is decreased further (Fig.9–10). For these experiments, we find EnKS-MDA to be more accurate than the EnKS and much more stable than the variational methods. This result is anticipated in nonlinear regimes, since incremental updates reduce potential over-adjustments by the ensemble smoother over the time window. As previously stated, the improved performance over E4DVar and 4DEnVar for sparse observation networks (Fig. 9) must follow from the ability of EnKS-MDA to sample directly from the smoothing density, rather than relying on a hybrid approach, which is a clear advantage of this method. Algorithmically, the EnKS operates in a manner that is very similar to 4DEnVar, but with the added benefit of updating ensemble perturbations about the posterior mean, rather than re-centering EnKF perturbations about the posterior mode.

For the experiment with $N_y = 15$, we also verify the second moment of the posterior to examine potential shortcomings in uncertainty estimates. The observation network and ensemble size used in these simulations poses challenges for several data assimilation method used here, in that filter divergence is prevented for a narrower range of parameters than previous experiments. Figure 11 shows the ratio of spread to RMSEs, indicating whether the ensemble spread is overestimated or underestimated with respect to the RMSE. The results of all methods are presented except E4DVar and 4DEnVar, which do not estimate posterior variance—recall that ensemble perturbations are updated using an EnKF instead. Ideally, the spread and RMSE should be equivalent, but sampling error and assumptions made during data assimilation may lead to inconsistent results. Likewise, heuristic techniques for treating sampling errors, such as localization and covariance relaxation can also introduce suboptimal uncertainty estimates. For all filters and smoothers examined in this study, the best match between spread and RMSE tends to occur when RMSE is at a minima (Fig. 11a-d). The further away from the optimal parameter settings, the larger the mismatch between spread and RMSE. As such, filter divergence occurs when the spread begins to become overestimated or underestimated for all methods (Fig. 11a-d).

Despite the difficulty posed by these observation networks, we find that the local PF can be configured to produce stable results, even for data-sparse regimes, which was expected for this method (Poterjoy 2021). This property of the local PF is illustrated for the $N_y = 10$ case, where it is the only method that does not diverge for all parameter value (Fig. 10). These results demonstrate challenges that exist for the mildly nonlinear observation operator as the spatial and temporal density of measurements decreases to yield larger prior uncertainty.

5) Local PF performance as a function of ensemble size

Figure 12 shows the mean RMSEs of the local PF as a function of ensemble size. These experiments use a fixed PF mixing coefficient of $\gamma = 0.3$ and two N^t_{eff} values of $0.2 \times N_e$ and $0.8 \times N_e$. The results are similar for the cases with the linear and mildly nonlinear measurement operators (Fig. 12a and b) in that optimal ROI increases with ensemble size.

This is because the large ensemble size yields fewer sampling errors thus needing less localization. For the strongly nonlinear measurement operator, however, the difference in RMSEs for the range of ROI choices become small as the ensemble size increases (Fig. 12c). This result may reflect either the limited information contained in these measurements. That is, because they only observe the log of the absolute value of variables, distant multivariate updates from these measurements are truly very small, thus requiring very large ensemble sizes to estimate accurately. They may also suggests that sampling errors and other factors, such as assumptions made by local PF update equations, become less dominant for nonlinear applications of this type.

Furthermore, the experiments demonstrate a dependence of optimal N^t_{eff} on ensemble size. When the ensemble size is small, experiments with higher N^t_{eff} show more accurate results. As the ensemble size increases, the lower N^t_{eff} shows smaller posterior RMSEs. This suggests that the larger N^t_{eff} can result in over-inflation when the ensemble size is large.

4. Conclusions

In geophysical models, such as those used for numerical weather prediction, strongly nonlinear model dynamics and measurement operators can cause data assimilation methods to be biased. This study examines several procedures that are developed to overcome challenges posed by nonlinear operators, such as periodic re-linearization of tangent linear and adjoints in variational schemes, likelihood factorizations adopted by iterative ensemble filters and smoothers, and localized particle filters. These methods—some of which were originally designed for applications outside the weather community—are compared with methods currently used for operational NWP, namely EnKFs and hybrid variational methods with and without model adjoints.

This study adopts the 40-variable model of Lorenz (1996) to examine the selected data assimilation approaches. The small dimension of this model allow for extensive testing of each technique using a large variety of observation networks, each varying in density and the type of observations provided. For several observation networks used in this study, re-linearization of the model and measurement operators between outer iterations are required to prevent filter divergence. Once a sufficient number of outer iterations are reached to achieve stable results, the improvements are negligible.

The wide range of observation networks examined in this study yields a diverse set of results, which are summarized using posterior RMSEs. We acknowledge that this metric is not ideal for non-Gaussian regimes, particularly those characterized as multimodal. Nevertheless, the sharp failure of various techniques for non-Gaussian problems are easily identified by large values of RMSEs.

Each method examined in this study has clear advantages for specific regimes—which are identified to be a function of sampling error, nonlinearity in measurement operators, and observation density. This finding motivates the use of different choices of data assimilation methodology, depending on application.

The ensemble-variational smoother with an adjoint model, E4DVar, produces smaller RMSEs than 4DEnVar for all observation networks tested in this study. It also outperforms all other methods in regimes where sampling error is high, but the model solution is well-constrained by numerous accurate measurements; i.e., in weakly nonlinear regimes. This study also compares variational methods to an ensemble smoother, which is adapted from the fixed-lag EnSRS of Whitaker and Compo (2002). For regimes where sampling error is a more dominant sources of posterior bias than nonlinearity, the EnKS performs better than its filter counterpart. Adding iterations to EnKF and EnKS updates through MDA results in improved results for all nonlinear regimes, particularly for sparse observation networks and long DAW lengths. The EnKS with MDA is also found to outperform all methods for data assimilation problems characterized by high sampling error and weak nonlinearity. Likewise, it provides stable results in nonlinear regimes that cause E4DVar and 4DEnVar to experience filter divergence. For applications of this type, EnKS-MDA benefits from its ability to sample directly from the posterior smoothing density, rather than relying on a separate EnKF to update perturbations about a maximum likelihood solution.

Furthermore, ensemble filters outperform smoothers when nonlinearity in measurement operators or model dynamics have a dominant role in the data assimilation applications. This finding is consistent with past studies that compare filters and smoothers for problems of this type (Evensen and van Leeuwen 2000). For highly nonlinear regimes, the local PF is the only method that produces accurate results. The benefit of PF-based methodology, however, comes with the tradeoff of being more sensitive to sampling error. Therefore, it requires large ensemble sizes to produce RMSEs as low as ensemble and variational smoothers for quasi-linear regimes.

Owing to the nature of this study, all comparisons are performed in an idealized framework. These findings will ultimately help guide future data assimilation decisions for real geophysical problems, where the computational cost of exploring the sensitivity of data assimilation methodology and parameters is prohibitive. The major findings of this study demonstrate when to expect Gaussian filters and smoothers to be suboptimal and under what conditions iterative techniques provide added value over conventional methods. Choices of nonlinear measurement operators in this study are motivated by challenges faced by high-impact weather events, such as severe convective storms and tropical cyclones. In particular, all-sky satellite radiance measurements provide extensive, near-continuous data coverage for tropical cyclones over open oceans. These measurements are often difficult to use, owing to the highly non-Gaussian (often multi-modal) observation-space priors produced by nonlinear measurement operators. New operational weather prediction systems, such as NOAA's Hurricane Analysis and Forecast System (HAFS), will ultimately need to overcome barriers that currently exist in Gaussian-based data assimilation methodology to fully leverage measurements of this type, as several operational centers have made significant advancements to cope with the difficult conditions in the past years. Experiments performed in this study motivate applications of iterative ensemble approaches and the local PF for problems of this type.

Acknowledgments. Funding for this work was provided by NOAA grant #NA20OAR4600281 and NSF/CAREER Award #AGS1848363.

Data availability statement. All software used to generate results for this study is available upon request from the corresponding author.

References

Amezcua, J., and P. J. V. Leeuwen, 2014: Gaussian anamorphosis in the analysis step of the enkf: a joint state-variable/observation approach. *Tellus A: Dynamic Meteorology and Oceanography*, **66**, 23 493.

Andrews, A., 1968: A square root formulation of the Kalman covariance equations. *AIAA J.*, **6**, 1165–1168.

Bannister, R. N., 2017: A review of operational methods of variational and ensemble-variational data assimilation. *Quart. J. Roy. Meteor.*, **143**, 607–633.

Bengtsson, T., P. Bickel, and B. Li, 2008: Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. *Probability and Statistics: Essays in Honor of David A. Freedman, D. Nolan and T. Speed, Eds.*, **2**, 316–334.

Bickel, P., B. Li, and T. Bengtsson, 2008: Sharp failure rates for the bootstrap particle filter in high dimensions. *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, **3**, 318–329.

Bishop, C. H., 2016: The gigg-enkf: ensemble kalman filtering for highly skewed non-negative uncertainty distributions. *Quarterly Journal of the Royal Meteorological Society*, **142**, 1395–1412.

Bocquet, M., and P. Sakov, 2014: An iterative ensemble Kalman smoother. *Q. J. R. Meteorol. Soc.*, **140**, 1521–1535.

Bonavita, M., P. Lean, and E. Hólm, 2018: Nonlinear effects in 4d-var. *ECMWF*.

Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.*, **131**, 1013–1043.

Buehner, M., P. L. Houtekamer, C. Charette, H. Mitchell, and B. He, 2010: Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II. *Mon. Wea. Rev.*, **138**, 1567–1586.

Chen, Y., and D. S. Oliver, 2012: Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Math. Geosci.*, **44**, 1–26.

Cohn, S. E., N. Sivakumaran, and R. Todling, 1994: A fixed-lag Kalman smoother for retrospective data assimilation. *Mon. Wea. Rev.*, **122**, 2838–2867.

Courtier, P., J.-N. Thepáut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367–1387.

Emerick, A. A., and A. C. Reynolds, 2012: History matching time-lapse seismic data using the ensemble Kalman filter with multiple data assimilations. *Comput. Geosci.*, **16(3)**, 639–659.

Emerick, A. A., and A. C. Reynolds, 2013: Ensemble smoother with multiple data assimilation. *Comput. Geosci.*, **55**, 3–15.

Errico, R., P. Bauer, and J.-F. Mahfouf, 2007: Issues regarding the assimilation of cloud and precipitation data. *J. Atmos. Sci.*, **64**, 3785–3798.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99C**, 10 143–10 162.

Evensen, G., 2018: Analysis of iterative ensemble smoothers for solving inverse problems. *Computational Geosciences*, **22(3)**, 885–908.

Evensen, G., and P. J. van Leeuwen, 2000: An ensemble Kalman smoother for non-linear dynamics. *Mon. Wea. Rev.*, **128**, 1852–1867.

Fabry, F., and J. Sun, 2010: For how long should what data be assimilated for the mesoscale forecasting of convection and why? Part I: On the propagation of initial condition errors and their implications for data assimilation. *Mon. Wea. Rev.*, **138**, 242–255.

Fairbairn, D., S. R. Pring, A. C. Lorenc, and I. Roulstone, 2014: A comparison of 4DVar with ensemble data assimilation methods. *Quart. J. Roy. Meteor. Soc.*, **140**, 281–294.

Fletcher, S., and A. S. Jones, 2014: Multiplicative and additive incremental variational data assimilation for mixed lognorma gaussian errors. *Monthly Weather Review*, **142**, 2521–2544.

Fletcher, S. J., 2010: Mixed gaussian-lognormal four-dimensional data assimilation. *Tellus A*, **62**, 266–287.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor.*, **125**, 723–757.

Geer, A. J., and P. Bauer, 2011: Observation errors in all-sky data assimilation. *Quart. J. Roy. Meteor.*, **137**, 2024–2037.

Geer, A. J., S. Migliorini, and M. Matricardi, 2019: All-sky assimilation of infrared radiances sensitive to mid- and upper-tropospheric moisture and cloud. *Atmospheric Measurement Techniques*, **12**, 4903–4929.

Geer, A. J., and Coauthors, 2017: The growing impact of satellite observations sensitive to humidity, cloud and precipitation. *Quarterly Journal of the Royal Meteorological Society*, **143**, 3189–3206.

Geer, A. J., and Coauthors, 2018: All-sky satellite data assimilation at operational weather forecasting centres. *Quarterly Journal of the Royal Meteorological Society*, **144**, 1191–1217.

Gelb, A., J. F. Kasper, R. A. Nash, C. F. Price, and A. A. Sutherland, 1974: Applied optimal estimation. *The M.I.T. Press*, 374pp.

Gu, Y., and D. S. Oliver, 2007: An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *SPEJ*, **12**, 438–446.

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905–2919.

Honda, T., and Coauthors, 2018: Assimilating all-sky himawari-8 satellite infrared radiances: A case of typhoon soudelor (2015). *Mon. Wea. Rev.*, **146**, 213–229.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.

Jazwinski, A. H., 1970: Stochastic processes and filtering theory. *Academic Press*, 376pp.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, **82 (Series D)**, 35–45.

Kjervheim, J.-A., G. Evensen., J. Hove, and J. G. Vabø, 2011: An ensemble smoother for assisted history matching. *SPE*, 141929.

Kleist, D. T., and K. Ide, 2015: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part II: 4DEnVar and hybrid variants. *Mon. Wea. Rev.*, **143**, 452–470.

Liu, C., and Q. Xiao, 2013: Antarctic applications with Advanced Research WRF using real data. *Mon. Wea. Rev.*, **141**, 2721–2739.

Liu, C., Q. Xiao, and B. Wang, 2009: An ensemble-based four-dimensional variational data assimilation scheme. Part II. *Mon. Wea. Rev.*, **137**, 1687–1704.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP: A comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183–3203.

Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. Seminar on Predictability*, **1**, ECMWF, Reading, UK.

Lorenz, E. N., and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399–414.

Minamide, M., and F. Zhang, 2017: Adaptive observation error inflation for assimilating all-sky satellite radiance. *Mon Wea. Rev.*, **145 (3)**, 1063–1081.

Morzfeld, M., and D. Hodyss, 2019: Gaussian approximations in filters and smoothers for data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, **71:1**, 1–27.

Neal, R. M., 1996: Sampling from multimodal distributions using tempered transitions. *Stat. Comput.*, **6(4)**, 353–366.

Okamoto, K., A. P. McNally, and W. Bell, 2014: Progress towards the assimilation of all-sky infrared radiances: an evaluation of cloud effects. *Quart. J. Roy. Meteor.*, **140**, 1603–1614.

Poterjoy, J., 2016: A localized particle filter for high-dimensional nonlinear systems. *Mon. Wea. Rev.*, **144**, 59–76.

Poterjoy, J., 2021: Regularization and tempering for a moment-matching localized particle filter. *Quart. J. Roy. Meteor. Soc.*, under revision.

Poterjoy, J., and J. L. Anderson, 2016: Efficient assimilation of simulated observations in a high-dimensional geophysical system using a localized particle filter. *Mon. Wea. Rev.*, **144**, 2007–2020.

Poterjoy, J., 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.

Poterjoy, J., L. J. Wicker, and M. Buehner, 2019: Progress in the development of a localized particle filter for data assimilation in high-dimensional geophysical systems. *Mon. Wea. Rev.*, **147**, 1107–1126.

Poterjoy, J., and F. Zhang, 2015: Systematic comparison of four-dimensional data assimilation methods with and without a tangent linear model using hybrid background error covariance: E4DVar versus 4DEnVar. *Mon. Wea. Rev.*, **143**, 1601–1621.

Poterjoy, J., F. Zhang, and Y. Weng, 2014: The effects of sampling errors on the EnKF assimilation of inner-core hurricane observations. *J. Atmos. Sci.*, **142**, 1609–1630.

Potter, J., 1964: W matrix augmentation. M.I.T. Instrumentation Laboratory Memo SGA. *Massachusetts Institute of Technology, Cambridge, MA*, 5–64.

Potthast, R., A. Walter, and A. Rhodin, 2019: A localized adaptive particle filter within an operational NWP framework. *Mon. Wea. Rev.*, **147**, 345–362.

Privé, N. C., Y. Xie, J. Woollen, S. E. Koch, R. Atlas, and R. Hood, 2013: Evaluation of the Earth Systems Research Laboratory's global Observing System Simulation Experiment system. *Tellus*, **65A**, 19 011.

Rommelse, J., 2009: Data Assimilation in Reservoir Management. *Ph.D. thesis, Technical University of Delft, Delft, The Netherlands*.

Sakov, P., D. S. Oliver, and L. Bertino, 2012: An iterative EnKF for strongly nonlinear systems. *Mon. Wea. Rev.*, **140**, 1988–2004.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. *Mon. Wea. Rev.*, **136**, 4629–4640.

Stengel, M., P. Unden, M. Lindskog, P. Dahlgren, N. Gustafsson, and R. Bennartz, 2009: Assimilation of SEVIRI infrared radiances with HIRLAM 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **135**, 2100–2109.

Thepáut, J.-N., and P. Courtier, 1991: Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model. *Quart. J. Roy. Meteor. Soc.*, **117**, 1225–1254.

van Leeuwen, P. J., and G. Evensen, 1996: Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Wea. Rev.*, **124**, 2898–2913.

Vukicevic, T., T. Greenwald, M. Zupanski, D. Zupanski, T. V. Haar, and A. S. Jones, 2004: Mesoscale cloud state estimation from visible and infrared satellite radiances. *Mon. Wea. Rev.*, **132**, 3066–3077.

Whitaker, J., and G. Compo, 2002: An ensemble Kalman smoother for reanalysis. *Proc. Symp. on Observations, Data Assimilation and Probabilistic Prediction, Orlando, FL, Amer. Meteor. Soc.*, 144–147.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.

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.

Zhang, F., M. Zhang, and J. A. Hansen, 2009: Coupling ensemble Kalman filter with four-dimensional variational data assimilation. *Adv. Atmos. Sci.*, **26**, 1–8.

Zhu, Y., and Coauthors, 2016: All-sky microwave radiance assimilation in NCEP's GSI analysis system. *Mon. Wea. Rev.*, **144**, 4709–4735.

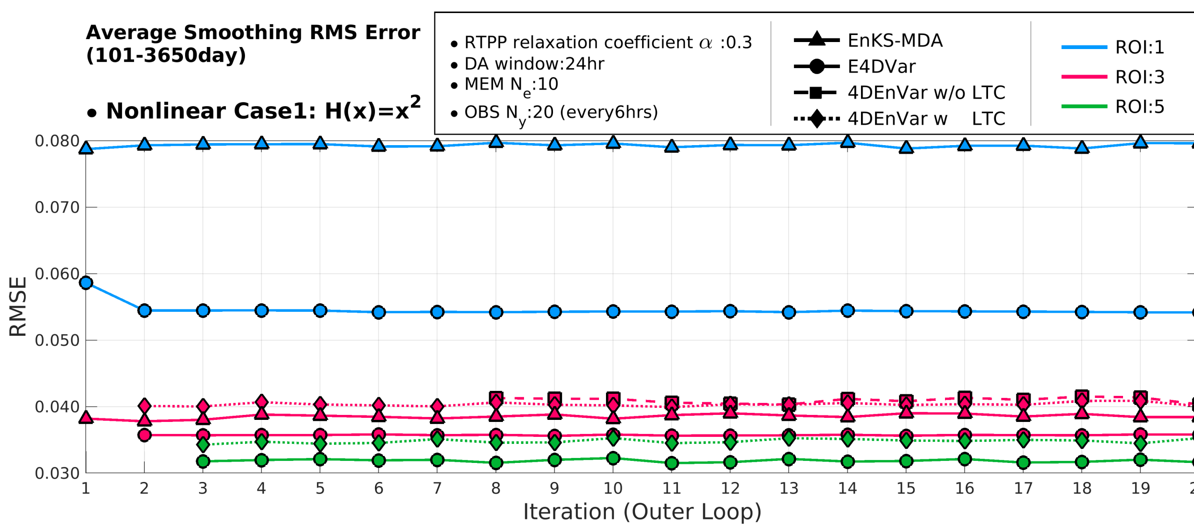
Zou, X., F. Weng, B. Zhang, L. Lin, Z. Qin, and V. Tallapragada, 2013: Impacts of assimilation of ATMS data in HWRF on track and intensity forecasts of 2012 four landfall hurricanes. *J. Geophys. Res. Atmos.*, **118**, 11 558–11 576.

LIST OF TABLES

TABLE 1. Configuration of cycling data assimilation experiments.

Expt	$H(x)$	σ_y	N_e	ROI	α,γ	N_y	Δt (h)
Linear Case	x	1.0	10,40	Variable	Variable	20	6
Nonlinear Case 1	x^2	1.0	10,40	Variable	Variable	20,15,10	6,24
Nonlinear Case 2	$\log(\|x\|)$	0.1	40,100	Variable	Variable	20	6

LIST OF FIGURES



867 FIG. 1. Mean analysis RMSEs as a function of the number of iteration or outer loop. Results are shown for
868 the Nonlinear Case1. Values are from the experiment with EnKS-MDA (triangle), E4DVar (circle), 4DEnVar
869 without LTC (square), and 4DEnVar with LTC (diamond), and ROI set to 1 (blue), 3 (red), and 5 (green). The
870 RMSEs are calculated at the start of the DAW (smoother solution).

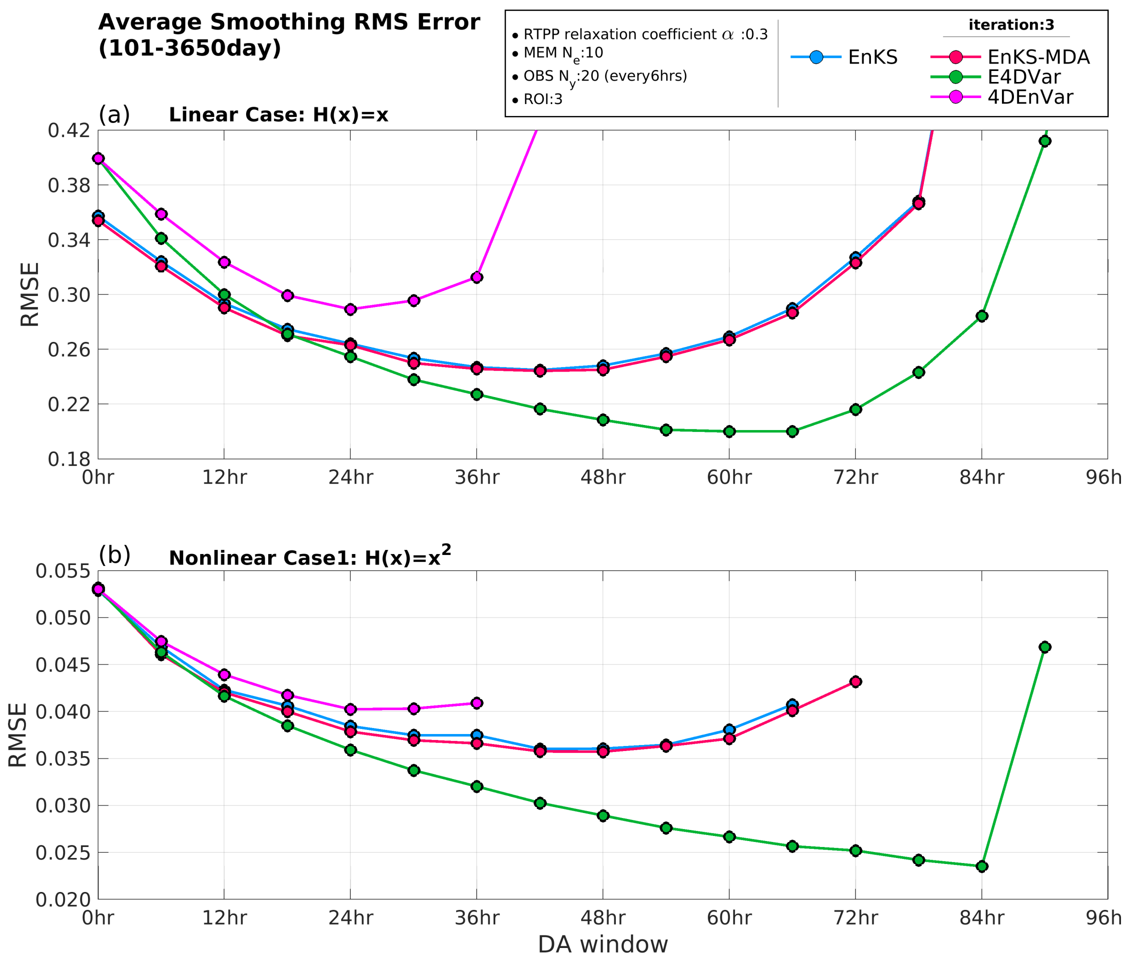
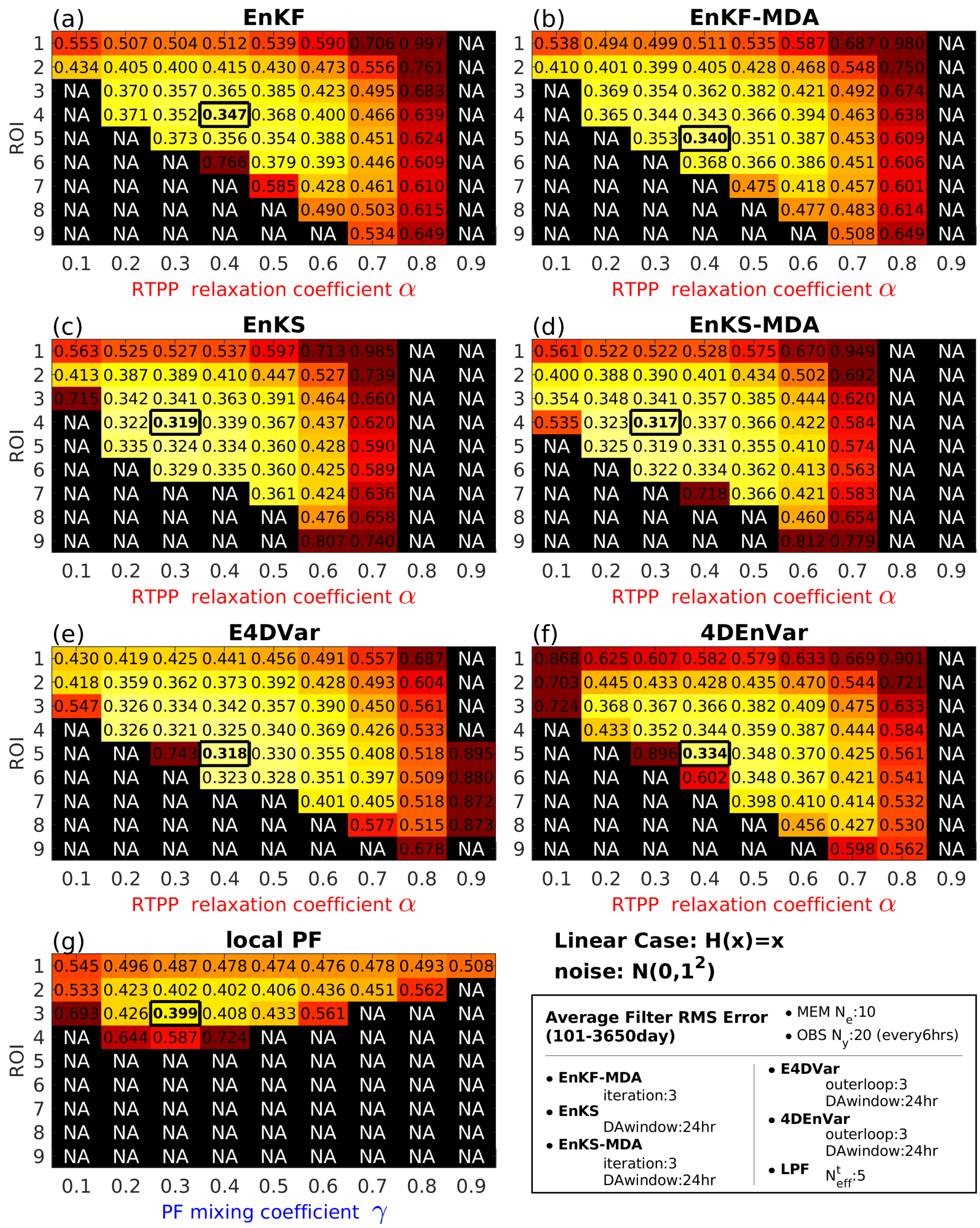


FIG. 2. Mean analysis RMSEs as a function of smoother lag. Results are shown for (a) Linear Case and (b) Nonlinear Case 1, with the EnKS (blue), the EnKS-MDA (red), the E4DVar (green), and the 4DEnVar (magenta). The number of iterations and outer loops is fixed at 3 for both cases. The RMSEs are calculated at the start of the DAW (smoother solution).



875 FIG. 3. Mean analysis RMSEs estimated for a range of relaxation coefficient α (a-f) and PF mixing coefficient γ (g) and ROI. Results are shown for experiments with the Linear Case and ensemble size is fixed at 10. Black shading indicates higher RMSEs, NA indicates that filter divergence occurs during the experiment, and the smallest errors are indicated by the black box. The RMSEs are calculated at the end of the DAW (filter solution).

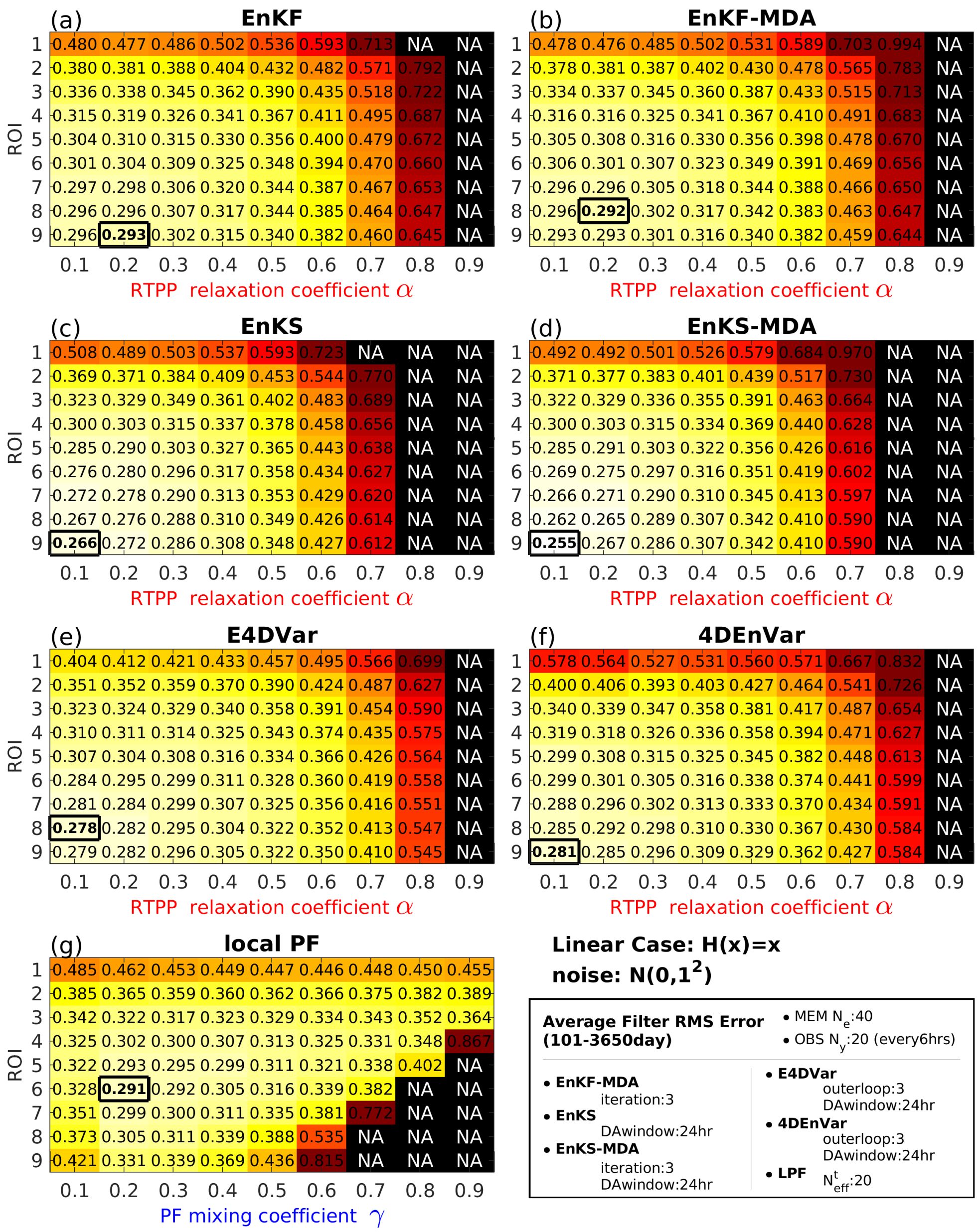


FIG. 4. As in Fig.3, but for ensemble size fixed at 40.

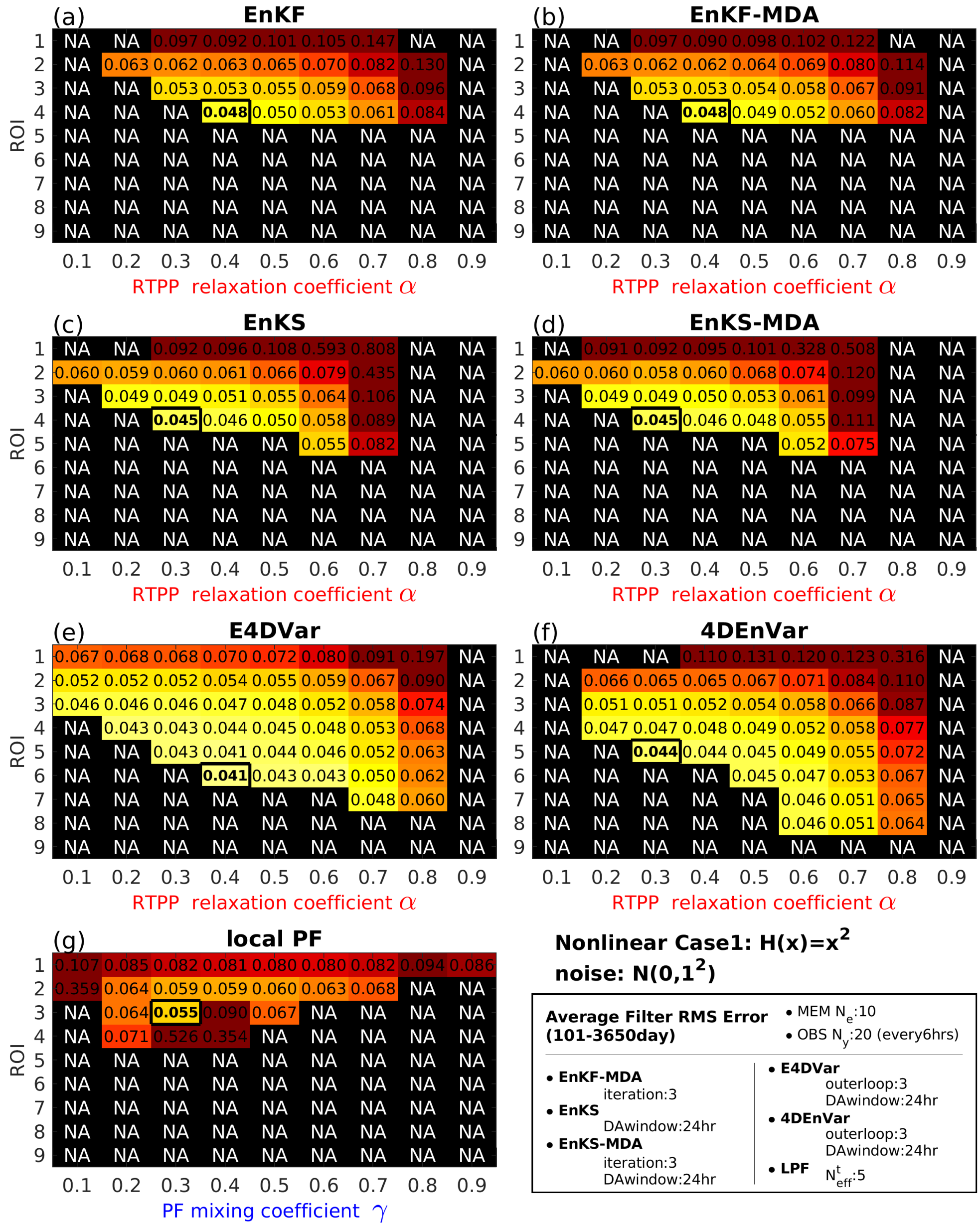


FIG. 5. Mean analysis RMSEs estimated for a range of relaxation coefficient α (a-f) and PF mixing coefficient γ (g) and ROI. Results are shown for experiments with the Nonlinear Case 1 and ensemble size is fixed at 10. Black shading indicates higher RMSEs, NA indicates that filter divergence occurs during the experiment, and the smallest errors are indicated by the black box. The RMSEs are calculated at the end of the DAW (filter solution).

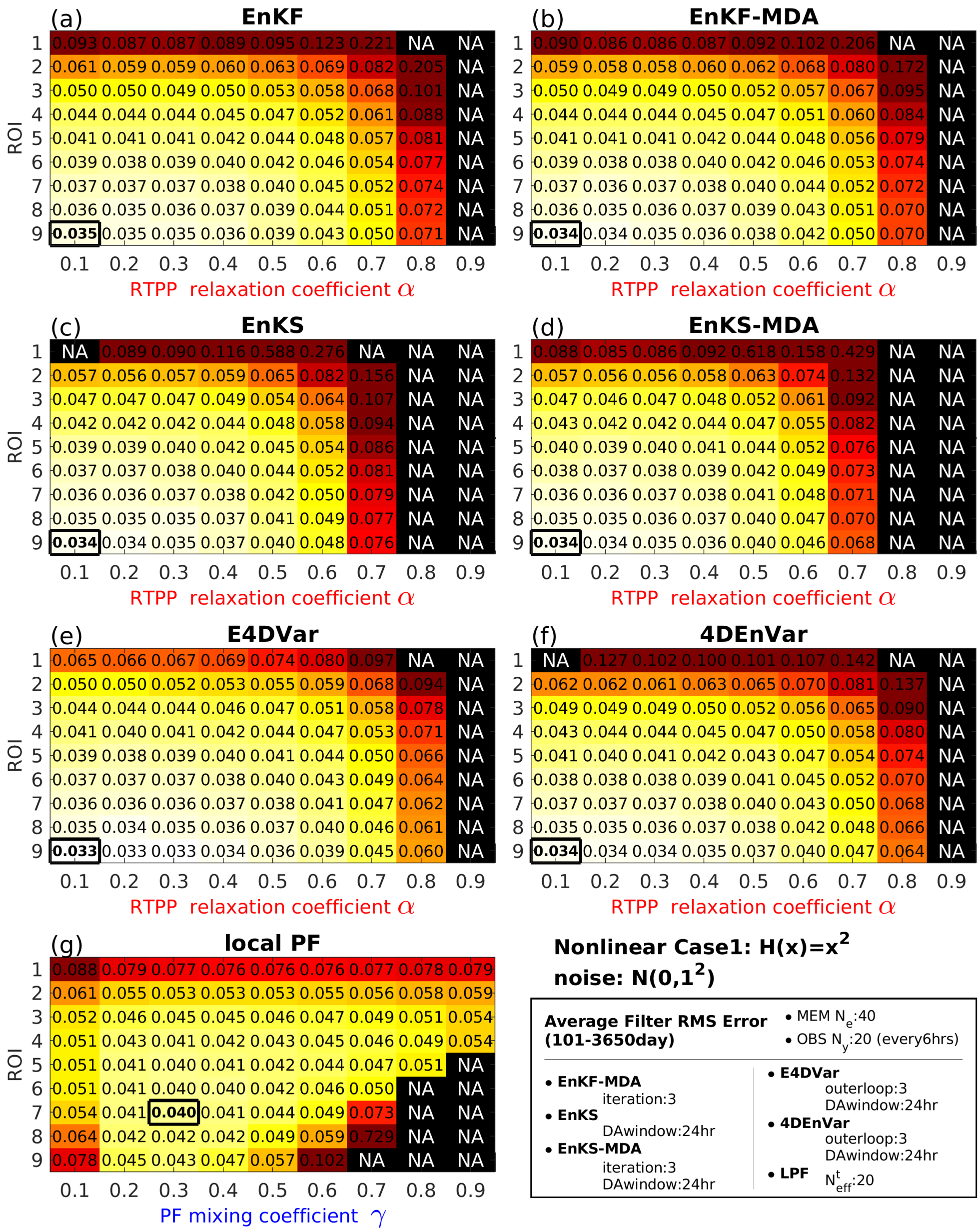


FIG. 6. As in Fig.5, but for ensemble size fixed at 40.

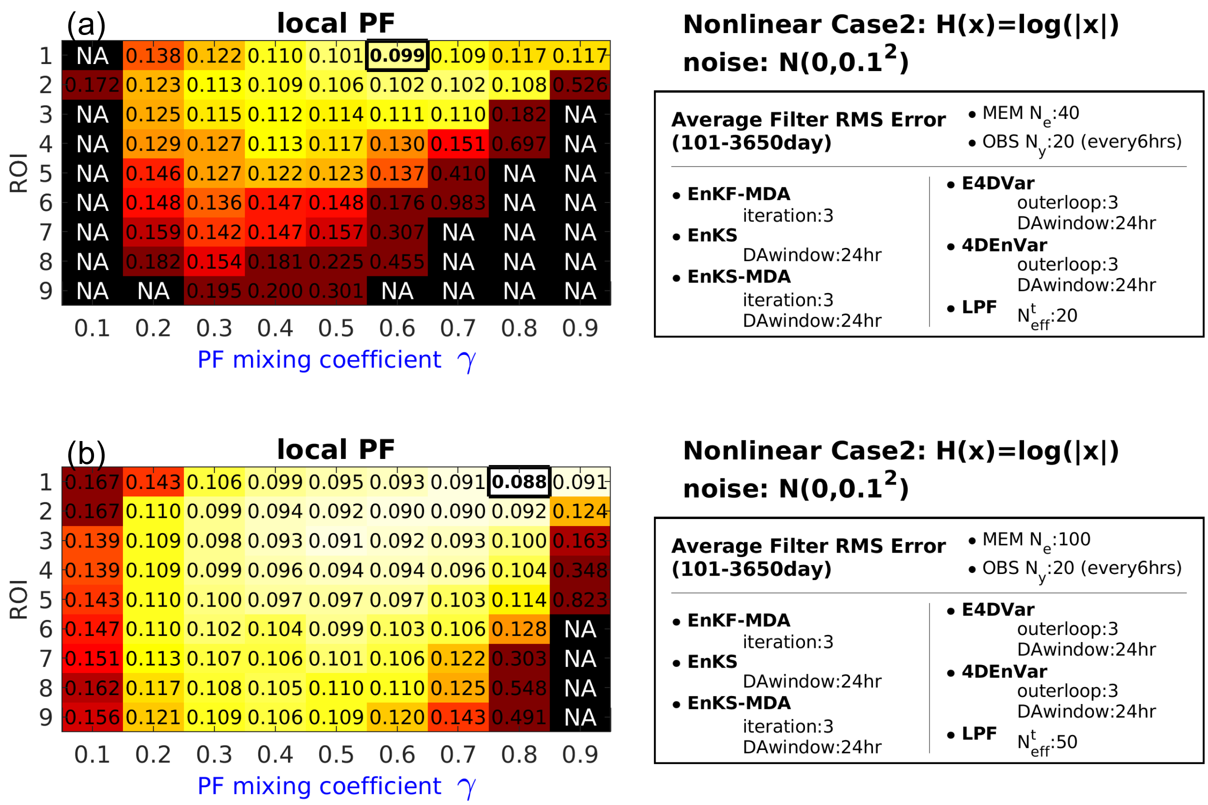


FIG. 7. Mean analysis RMSEs estimated for a range of PF mixing coefficient γ and ROI. Filter divergence occurs in all methods except the local PF, so only results of the local PF are shown for experiments with the Nonlinear Case 2 and ensemble size is fixed at 40 (a) and 100 (b). Black shading indicates higher RMSEs, NA indicates that filter divergence occurs during the experiment, and the smallest errors are indicated by the black box. The RMSEs are calculated at the end of the DAW (filter solution).

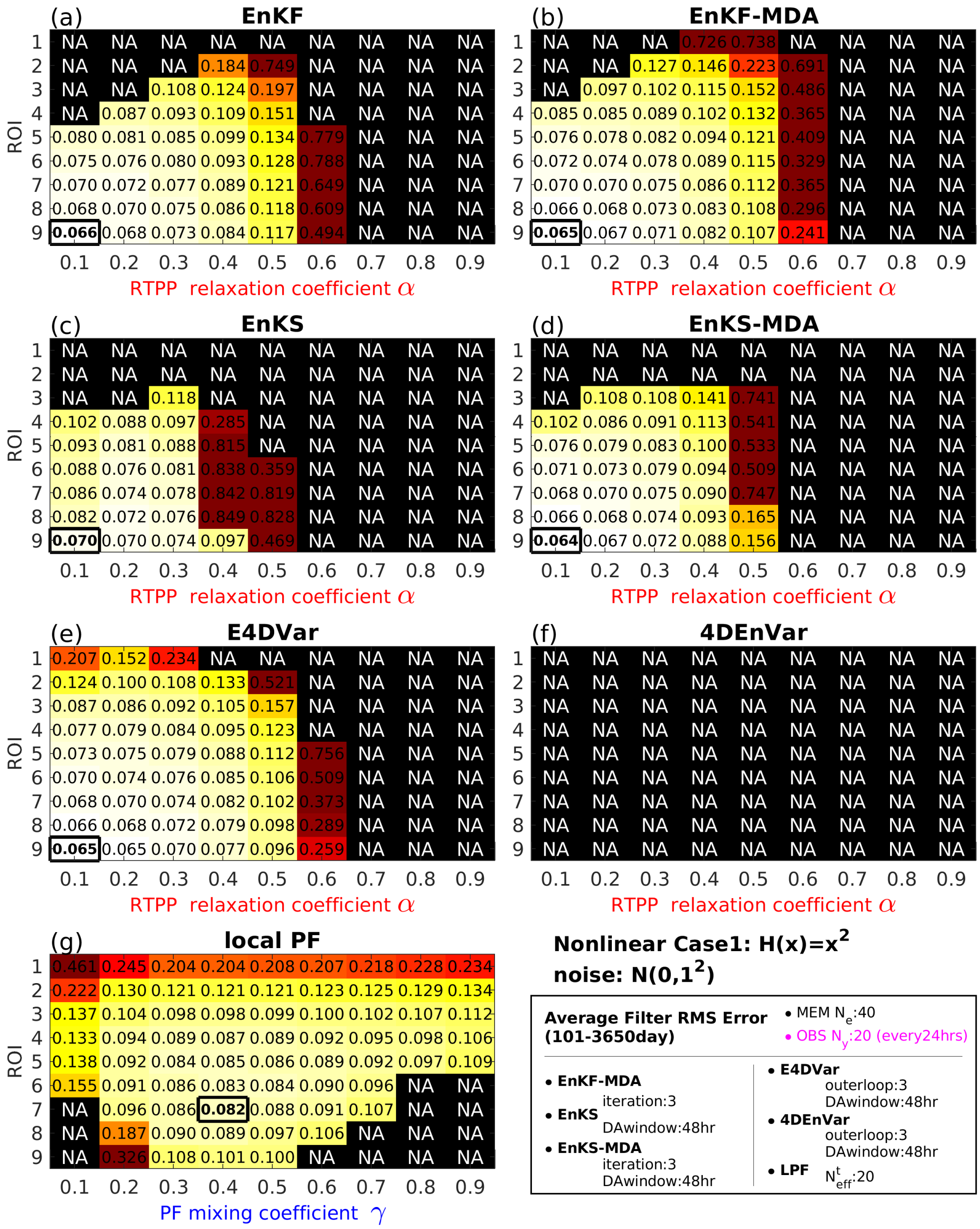


FIG. 8. As in Fig.6, but for the frequency of observations and DAW fixed at 24 h and 48 h, respectively.

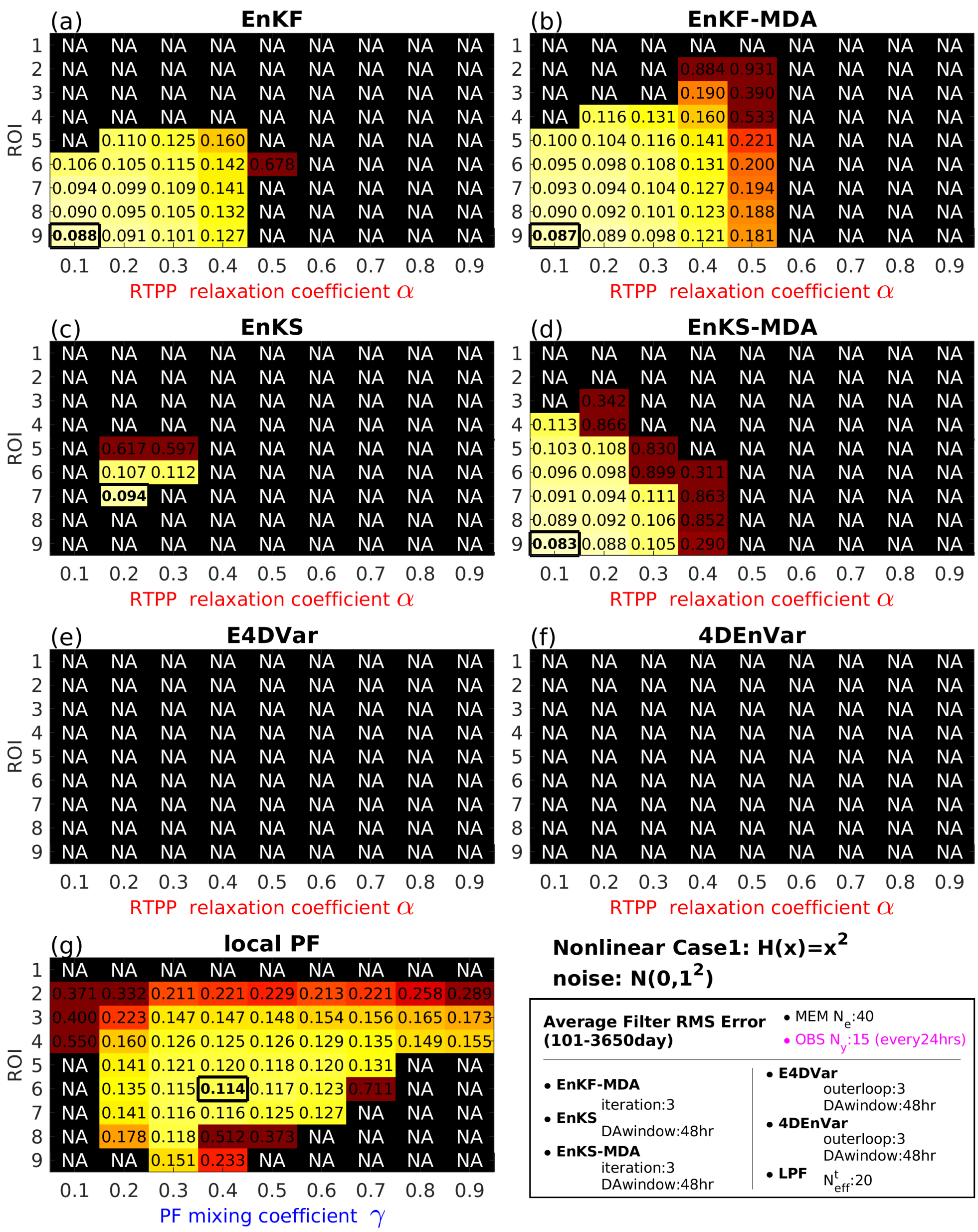


FIG. 9. As in Fig.8, but for the number of observations fixed at 15.

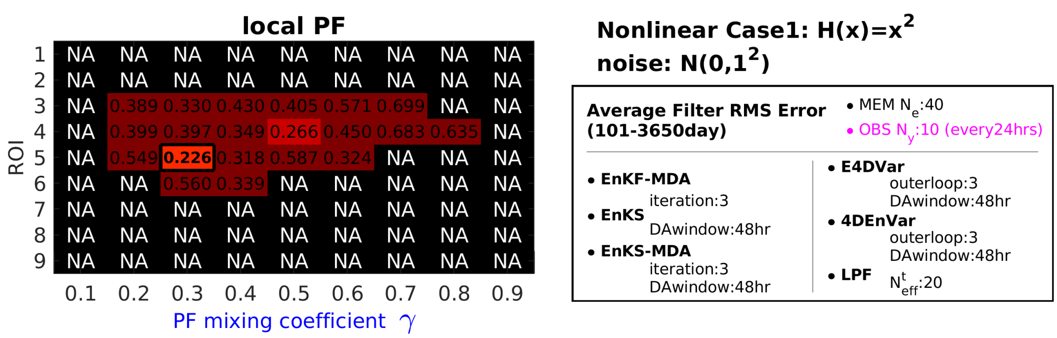


FIG. 10. As in Fig.8, but for the number of observations fixed at 10. Filter divergence occurs in all methods except the local PF, so only results of the local PF are shown.

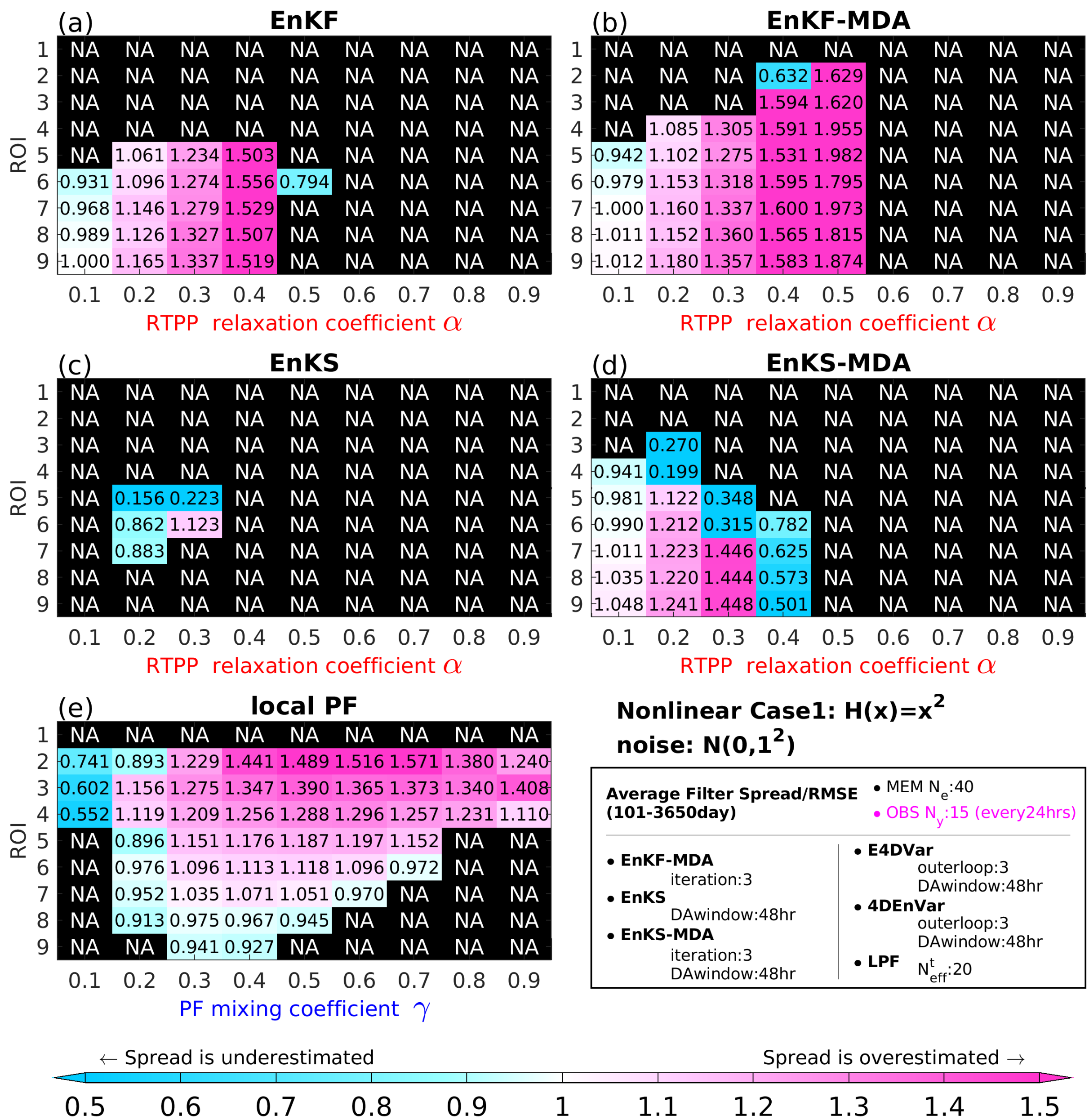
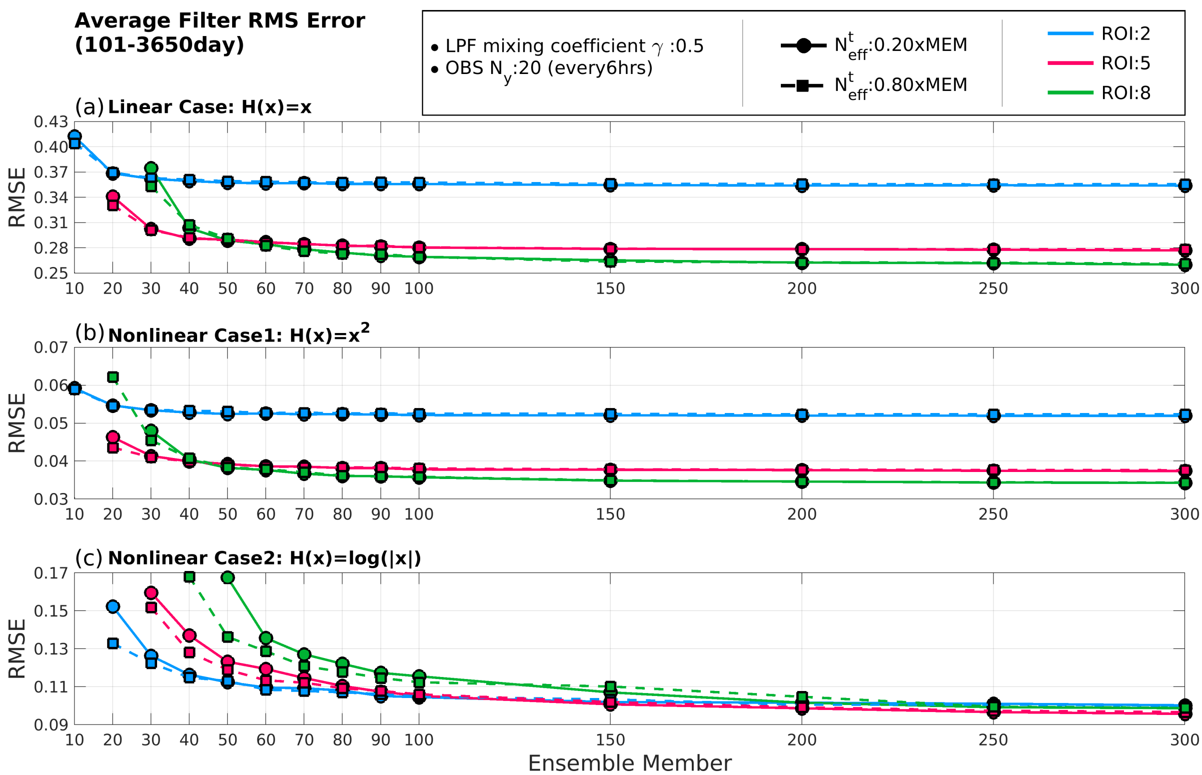


FIG. 11. Ratio of ensemble spread to mean analysis RMSEs estimated for a range of relaxation coefficient α (a-d) and PF mixing coefficient γ (e) and ROI. The experimental setting is the same as in Fig.9. NA indicates that filter divergence occurs during the experiment. The RMSEs and spread are calculated at the end of the DAW (filter solution).



894 FIG. 12. Mean analysis RMSEs of the local PF as a function of ensemble size. Results are shown for (a) Linear
895 Case, (b) Nonlinear Case 1, and (c) Nonlinear Case 2. Values are from the experiment with N_{eff} fixed at 0.20
896 $\times N_e$ (solid lines) and 0.80 $\times N_e$ (dashed lines), and ROI fixed at 2 (blue), 5 (red), and 8 (green).