An Efficient Bi-Gaussian Ensemble Kalman Filter for Satellite Infrared Radiance Data Assimilation

MAN-YAU CHAN,a,b JEFFREY L. ANDERSON,c AND XINGCHAO CHENA,b

a Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, Pennsylvania
b Center for Advanced Data Assimilation and Predictability Techniques, The Pennsylvania State University, University Park, Pennsylvania
c Data Assimilation Research Section, Computational Information Systems Laboratory, National Center for Atmospheric Research, Boulder, Colorado

ABSTRACT: The introduction of infrared water vapor channel radiance ensemble data assimilation (DA) has improved numerical weather forecasting at operational centers. Further improvements might be possible through extending ensemble data assimilation methods to better assimilate infrared satellite radiances. Here, we will illustrate that ensemble statistics under clear-sky conditions are different from cloudy conditions. This difference suggests that extending the ensemble Kalman filter (EnKF) to handle bi-Gaussian prior distributions may yield better results than the standard EnKF. In this study, we propose a computationally efficient bi-Gaussian ensemble Kalman filter (BGEnKF) to handle bi-Gaussian prior distributions. As a proof-of-concept, we used the 40-variable Lorenz 1996 model as a proxy to examine the impacts of assimilating infrared radiances with the BGEnKF and EnKF. A nonlinear observation operator that constructs radiance-like bimodal ensemble statistics was used to generate and assimilate pseudoradiances. Inflation was required for both methods to effectively assimilate pseudoradiances. In both 800- and 20-member experiments, the BGEnKF generally outperformed the EnKF. The relative performance of the BGEnKF with respect to the EnKF improved when the observation spacing and time between DA cycles (cycling interval) become sufficiently large. The BGEnKF generated less noise than the EnKF, suggesting that the BGEnKF produces more balanced analysis states than the EnKF. This proof-of-concept study motivates future investigation into using the BGEnKF to assimilate infrared observations into high-order numerical weather models.

KEYWORDS: Satellite observations; Data assimilation

1. Introduction

The advent of modern geostationary satellites ushered in an era of global mesoscale-resolving infrared observations of the atmosphere. Nowadays, operational numerical weather prediction centers include clear-sky infrared radiance observations into their data assimilation (DA). The inclusion of clear-sky observations has improved the analyzed temperature and humidity fields that are used to initiate numerical forecasts (Köpken et al. 2004; Munro et al. 2004; Yang et al. 2017; Burrows 2018). The inclusion of cloudy water vapor channel observations into DA has demonstrated the potential to further improve numerical weather prediction (Vukicevic et al. 2006; Errico et al. 2007; Otkin 2010; Bauer et al. 2018; Honda et al. 2018; Geer et al. 2019; Zhang et al. 2019; Otkin and Potthast 2019; Sawada et al. 2019; Chan et al. 2020). Improvements in the DA methods used might amplify the positive impacts of assimilating these observations.

Ensemble DA methods are often used to assimilate geostationary infrared radiances. These methods combine observations and their uncertainties with an ensemble of short-term forecasts (prior ensemble) to produce a more accurate and precise ensemble (posterior or analysis ensemble). More specifically, ensemble DA methods use the difference between observations and ensemble-derived simulated observations to update the ensemble. At the time of writing, the most popular ensemble DA methods used for satellite data assimilation assume that all ensemble members are drawn from a single multivariate Gaussian distribution (Gaussian DA methods; e.g., Bauer et al. 2018).

However, the ensemble members are not necessarily drawn from a single Gaussian distribution. Ensembles often possess mixtures of members that are cloudy at some forecasted location (cloudy members), and members that are cloud free at the same location (clear-sky members). The column-integrated condensate mass of clear-sky members is drawn from narrow distributions near zero, whereas the same quantity for cloudy members is drawn from distributions that are much wider and further from zero. Furthermore, because hydrometeors attenuate upwelling infrared and microwave radiation, satellite-sensed brightness temperatures vary depending on whether clouds are present (Geer and Bauer 2011; Harnisch et al. 2016; Minamide and Zhang 2019, 2017; Honda et al. 2018). In other words, the ensemble is drawn from multiple distributions, as opposed to a single Gaussian distribution. The prior ensemble distribution is thus a weighted sum of multiple distributions. As such, research into DA methods that handle weighted sum prior distributions can potentially improve the effectiveness of assimilating all-sky infrared observations.

One possibility is to extend Gaussian DA methods to handle Gaussian mixture model (GMM) prior probability distributions.
A GMM distribution models each of the component distributions in the prior as a Gaussian kernel. For instance, a kernel can be assigned to clear-sky members, and another kernel can be assigned to cloudy members. Furthermore, GMM distributions can also handle situations where some, or even all, of the variables are drawn from a single Gaussian distribution—the kernels would have identical means and covariances for these variables. Hence, extending Gaussian DA methods to handle GMM prior distributions is a reasonable approach (Anderson and Anderson 1999).

The idea of using GMMs to approximate non-Gaussian prior distributions for Bayesian inferences was first introduced by Als McNair and Sorenson (1972). In their GMM DA, four sets of parameters are used. The first parameter is the number of Gaussian kernels, which determines the maximum number of modes exhibited by the GMM distribution. This first parameter is not updated in the process of DA. The remaining three sets of parameters are the weights, means and covariances for each Gaussian kernel. These three parameters are updated in the process of DA. In GMM ensemble DA methods, these parameters are first estimated from the prior ensemble (prior parameters) and the ensemble is then updated to reflect the posterior parameters. Computationally affordable GMM ensemble DA methods for numerical weather prediction thus require computationally affordable methods to estimate the parameters and update the ensemble.

Various studies have proposed different methods to estimate the prior GMM PDF parameters and extend the Gaussian ensemble Kalman filter (EnKF; Evensen 1994; Houtekamer and Mitchell 1998; Burgers et al. 1998) to handle GMM prior PDFs (GMM EnKF). Anderson and Anderson (1999) efficiently estimated the parameters by assuming that the number of Gaussian kernels is the same as the number of ensemble members and that the center of each kernel corresponds to an ensemble member. Furthermore, the covariance of each kernel is a rescaled version of the entire ensemble’s covariance. However, these assumptions are inappropriate for infrared DA because only two kernels should suffice to approximate the clear-sky and cloudy members, and each kernel likely has a different covariance. Bengtsson et al. (2003) then proposed a GMM EnKF that constructed $G$ kernels by grouping ensemble members in the vicinity of $G$ randomly selected members, in high-dimensional state space. This method is very susceptible to sampling noise (Sondegaard and Lermusiaux 2013a). A more accurate method of estimating the prior parameters is to use an expectation maximization (EM) algorithm to fit Gaussian kernels to the ensemble (Dovera and Della Rossa 2011). However, in spaces with more than 100 dimensions, the EM algorithm can frequently run into values that are smaller than the computer’s floating point precision. To mitigate this issue, Sondergaard and Lermusiaux (2013a,b) performed the EM fitting in a stochastic subspace based on the forecast model’s governing equations (Sapsis and Lermusiaux 2009, 2012; Ueckermann et al. 2013). Unfortunately, utilizing this stochastic subspace requires deriving and coding a set of dynamically orthogonal field equations from the model equations. This procedure is harder and more labor-intensive than constructing model adjoint operators for four-dimensional variational data assimilation. Simpler, more efficient, and more appropriate methods of estimating the prior GMM PDF parameters are thus necessary in order to use GMM ensemble DA to practically assimilate satellite infrared observations.

Another challenge of GMM ensemble DA is the difficulty in updating the ensemble in a way consistent with the posterior parameters. The traditional approach is to directly compute the posterior parameters first, and then draw random samples based on these posterior parameters (Dovera and Della Rossa 2011; Anderson and Anderson 1999; Sondergaard and Lermusiaux 2013a). Standard drawing methods require the square-roots of the posterior covariance matrices (Rasmussen and Williams 2005), which are computationally expensive to determine in a high-dimensional space. While Sondergaard and Lermusiaux (2013a,b) performed the sampling in a low-dimensional subspace, their method is difficult to employ for complex models (see previous paragraph). A second approach is to first adjust the number of members in each kernel’s cluster until the number of members in each cluster is consistent with the posterior weights, and then perform perturbed observation EnKF updates on the ensemble members in each cluster (Burgers et al. 1998; Bengtsson et al. 2003). The cluster sizes are adjusted by duplicating (deleting) random members in clusters where the weight is increased (decreased) by DA. Since this second approach relies on perturbing the observations to prevent the emergence of identical posterior ensemble members, a deterministic EnKF cannot be employed with this approach.

In this study, we will use an ensemble of Weather Research and Forecasting (WRF) Model runs over the Maritime Continent to illustrate that a forecast ensemble can have members drawn from different distributions. Following that, we will introduce a simple heuristic method to fit a two-kernel GMM (bi-Gaussian) distribution to the ensemble. We then introduce an efficient bi-Gaussian ensemble Kalman filter (BGEEnKF) that updates the ensemble without the pitfalls listed in the previous paragraphs.

Tests of new ensemble DA methods with high-order models (e.g., WRF) typically require large amounts of computational resources and manpower. These expenditures are further exacerbated by the large volume of infrared observations typically assimilated. To motivate these expenditures, proof-of-concept tests of this BGEEnKF were done using the 40-variable Lorenz 1996 model (L96; Lorenz 2006) and will be presented in this article.

The setup and examination of the WRF illustration are presented in section 2. A heuristic approach to fitting a bi-Gaussian distribution to the forecast ensemble will be proposed in section 3. Section 3 then introduces the BGEEnKF algorithm. To mimic assimilating infrared brightness temperature observations in the L96 model, a nonlinear pseudoradiance observation operator will be introduced in section 4. The performance of the BGEEnKF in assimilating pseudoradiance observations against that of the EnKF in L96 is then examined in section 4. In addition, the performance will be examined for various observation spacing, ensemble sizes and cycling frequencies. Conclusions are presented in section 5.
2. Bi-Gaussian ensemble forecast statistics in WRF

a. Setup of WRF ensemble

The Advanced Research WRF (Skamarock et al. 2008) Model, version 3.8.1, was used to construct a realistic forecast ensemble over the Maritime Continent. The domain has 560 × 450 horizontal grid points with 9-km grid spacing and 45 model levels. The bottom 9 levels are within the lowest 1 km. The top of the domain is set to 20 hPa, and the domain covers a rectangular area that extends latitudinally from 20°S to 15°N and longitudinally from 85° to 130°E.

The WRF configuration employed is similar to that of Ying and Zhang (2017, 2018). Radiative processes were simulated using the updated Goddard shortwave scheme (Chou and Suarez 1999) and the global circulation model version of the Rapid Radiative Transfer Model (RRTMG) longwave radiation scheme (Iacono et al. 2008). The unified Noah land surface physics scheme (Chen and Dudhia 2001) was used to simulate surface processes, except that the surface skin temperature is diagnosed separately using the scheme of Zeng and Beljaars (2005). Subgrid-scale turbulent mixing is handled by the Yonsei University (YSU) boundary layer scheme (Hong et al. 2006). The WRF double-moment 6-class scheme (Lim and Hong 2010) was used to represent cloud microphysical processes.

Because the 9-km horizontal grid spacing used in this study is fine enough to resolve mesoscale convective processes, no cumulus parameterizations were employed. It should be pointed out, however, that the 9-km horizontal grid spacing is too coarse to resolve convective-scale updrafts. Nonetheless, the work of Wang et al. (2015) indicates that this resolution is sufficient to replicate the general precipitation, circulation, thermodynamic, and radiation features of the October 2011 and November 2011 Madden Julian Oscillation events observed over the Indian Ocean. Similar model setups have been used to study atmospheric overturning and the physical mechanisms of tropical convection over the Indian Ocean (Zhang et al. 2017; Chen et al. 2018a, b, c; Chen and Zhang 2019).

The initial conditions of the 50-member WRF were generated from the European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis 5 (ERA5) at 0000 UTC 30 May 2017. Since ERA5 only uses 10 ensemble members, the 50 ensemble perturbations needed were constructed from five groups of 10 ERA5 members at different times. The first 10 perturbations were generated by taking the difference between the 10 ERA5 members at 0000 UTC 26 May with the ERA5 ensemble mean at 0000 UTC 26 May. The second, third, fourth, and fifth groups of perturbations were constructed the same way based on the ERA5 ensemble at 0000 UTC 27, 28, 29, and 30 May. These 50 perturbations were added to the ERA5 ensemble mean at 0000 UTC 30 May 2017. The boundary conditions for all ensemble members are based on the ERA5 reanalysis. The WRF ensemble was then integrated forward for 24 h to generate flow-dependent ensemble statistics at 0000 UTC 31 May. Note that while the spinup process is longer than typical forecast cycles, it is common to use long spinups when initiating a new DA application (e.g., Ying and Zhang 2018; Chan et al. 2020) to prevent filter divergence.

b. Ensemble forecast statistics

Before delving into the forecast statistics of infrared brightness temperatures, we will briefly describe the case we are using as an illustration. Figure 1a shows the ensemble mean simulated Himawari-8 Advanced Himawari Imager (AHI) infrared window channel 14 brightness temperature field (Window-BT) at 0000 UTC 31 May. This field was generated by running the Community Radiative Transfer Model (release 2.3.0) on the 50 ensemble members, and then computing the ensemble mean. Prior to this time, the observed and simulated Window-BT indicate that a mesoscale convective system (MCS) entered the domain from the west. In the ensemble, this MCS evolved into a linear MCS to the west of Sumatra (Fig. 1a) that is consistent with Window-BT observations.

We will illustrate the point that the forecast ensemble is drawn from multiple distributions with significantly different prior ensemble statistics in the vicinity of this linear MCS. The simplest way is to show that statistics under clear-sky conditions are different from statistics under cloudy conditions. To do so, at each location in the WRF domain, we grouped ensemble members with nonnegligible total precipitate mixing ratio (>0.01 g kg⁻¹) at any grid point above 900 hPa into one cluster (cloudy cluster) and formed a second cluster with the remaining members (clear cluster). Three locations in the vicinity of the linear MCS with varying proportions of cloudy members (Fig. 1b) were selected and the pointwise ensemble histograms of Window-BT and 640 hPa water vapor mixing ratio (QVAPOR) are shown in Fig. 2. The ensemble data points over the area marked in Fig. 1 were aggregated to form a similar histogram in Fig. 3. Figures 2 and 3 indicate that clear-sky members have lower mean QVAPOR values than those of cloudy members. Furthermore, clear-sky Window-BT values have higher means and smaller spreads than those of cloudy members, which is consistent with earlier studies (Harnisch et al. 2016; Minamide and Zhang 2019, 2017; Honda et al. 2018). These differences between clear-sky and cloudy members illustrate that clear-sky and cloudy members might be drawn from different distributions.

The fact that multiple distributions can exist in the same location suggests that both the state and observation space prior distributions are the sum of multiple multivariate distributions. This summing property motivates the use of mixture distributions. Given the success of using Gaussian distributions in ensemble DA, it seems natural to extend existing DA methods to handle GMM distributions.

3. BGEnKF theory and algorithm for infrared satellite DA

a. The bi-Gaussian prior approximation and the resulting posterior PDF

While our WRF illustration and the work of several others (Geer and Bauer 2011; Harnisch et al. 2016; Minamide and Zhang 2017; Honda et al. 2018) support the idea that there are multiple kernels, it is difficult to estimate the kernel parameters (see section 1). To simplify the estimation process, we will
assumed that the prior distribution in the space updated by a single infrared observation has two kernels (i.e., the prior distribution is bi-Gaussian). Since members that are cloudy at the observation site are drawn from a distribution that is different from members that have no clouds at the observation site, this assumption implies that cloudy members belong to one kernel, and clear-sky members belong to the other kernel. The parameters of each kernel can then be easily estimated using members that belong to said kernel.

We will use a notation similar to Ide et al. (1997) to derive an algorithm to assimilate observations into this bi-Gaussian prior distribution. This filter will assimilate one observation at a time. Suppose that we are trying to constrain an ensemble of infrared radiance is constructed via a nonlinear forward operator \( h(x) \). Passing an ensemble of \( N_E \) forecasted states \( \{x_N^i, x_N^j, \ldots, x_N^{N_{k-1}}, x_N^k\} \) into \( h(x) \) results in an ensemble of \( N_E \) prior infrared radiances \( \{y_N^i, y_N^j, \ldots, y_N^{N_{k-1}}, y_N^k\} \). The prior PDF of the extended space vector \( \psi = [x^T, y^T] \) can then be written as

\[
p(\psi) = \sum_{\varepsilon=1}^{2} \frac{w_{\varepsilon}^f}{\sqrt{(2\pi)^N \det(P_{\varepsilon}^f)}} \exp \left\{ -\frac{1}{2} (\psi - \mu_{\varepsilon}^f)^T P_{\varepsilon}^{-1} (\psi - \mu_{\varepsilon}^f) \right\},
\]

where \( w_{\varepsilon}^f, \mu_{\varepsilon}^f, \) and \( P_{\varepsilon}^f \) are the prior scalar weight, ensemble mean vector, and ensemble covariance matrix of the \( g \)th Gaussian, where \( g \) can take on values 1 or 2, representing the clear-sky and cloudy clusters, respectively. Furthermore,

\[
\sum_{\varepsilon=1}^{2} w_{\varepsilon}^f = 1 \quad \text{and} \quad w_{\varepsilon}^f \geq 0 \quad \forall \quad g = 1, 2.
\]

In the limit where one of the prior weights is zero, the prior PDF in Eq. (1) collapses to a Gaussian prior PDF.

Assuming that \( y^o \) has Gaussian error statistics, the posterior PDF that results from assimilating \( y^o \) is bi-Gaussian (Dovera and Della Rossa 2011; Anderson and Anderson 1999):

\[
p(\psi|y^o) = \frac{p(\psi)P(y^o|\psi)}{P(y^o)} = \sum_{\varepsilon=1}^{2} \frac{w_{\varepsilon}^o}{\sqrt{(2\pi)^N \det(P_{\varepsilon}^o)}} \exp \left\{ -\frac{1}{2} (\psi - \mu_{\varepsilon}^o)^T P_{\varepsilon}^{-1} (\psi - \mu_{\varepsilon}^o) \right\},
\]

where \( w_{\varepsilon}^o, \mu_{\varepsilon}^o, \) and \( P_{\varepsilon}^o \) are the posterior scalar weight, ensemble mean vector and covariance matrix of the \( g \)th Gaussian kernel in the posterior PDF.

It should be mentioned here that the infrared observation errors do not strictly follow a Gaussian distribution (Harnisch et al. 2016; Minamide and Zhang 2017). Nonetheless, many studies have successfully assimilated infrared observations despite assuming Gaussian-distributed observation errors (e.g., Vukicevic et al. 2004; Otkin 2010; Zhang et al. 2016). Thus, this imperfect assumption is at least functional.
The posterior weight can be computed via (Dovera and Della Rossa 2011; Anderson and Anderson 1999)

\[ w_g^\omega = \frac{1}{Z} \frac{w_f^g}{\sqrt{2\pi(\sigma_{f,g}^2 + \sigma_o^2)}} \exp \left\{ -\frac{1}{2} \frac{(y^\omega - \bar{y}_f^g)^2}{\sigma_{f,g}^2 + \sigma_o^2} \right\}, \tag{3} \]

where \( \sigma_o^2 \) is the observation error variance, \( \sigma_{f,g}^2 \) is the variance of the prior simulated observations in cluster \( g \), and \( \bar{y}_f^g \) is the mean of prior simulated observations in cluster \( g \). Also, the normalization constant \( Z \) is

\[ Z = \sum_{g=1}^{G} \frac{w_f^g}{\sqrt{2\pi(\sigma_{f,g}^2 + \sigma_o^2)}} \exp \left\{ -\frac{1}{2} \frac{(y^\omega - \bar{y}_f^g)^2}{\sigma_{f,g}^2 + \sigma_o^2} \right\}. \tag{4} \]

To see how Eq. (3) increases or decreases the weight of a cluster, suppose that the Bayesian inference increased the weight of cluster 1. In other words, \( w_1^\omega > w_1^f \). If we substitute Eqs. (2), (3), and (4) into this inequality, take the natural logarithm of the result, and then rearrange the terms, we obtain

\[ \left[ \frac{y^\omega - \bar{y}_f}{\sigma_{f,1}^2 + \sigma_o^2} \right]^2 + \ln(\sigma_{f,1}^2 + \sigma_o^2) < \left[ \frac{y^\omega - \bar{y}_f^2}{\sigma_{f,2}^2 + \sigma_o^2} \right]^2 + \ln(\sigma_{f,2}^2 + \sigma_o^2). \tag{5} \]

We can view \( \left( \frac{y^\omega - \bar{y}_f^1}{\sqrt{\sigma_{f,1}^2 + \sigma_o^2}} \right) / \sqrt{\sigma_{f,1}^2 + \sigma_o^2} \) and \( \left( \frac{y^\omega - \bar{y}_f^2}{\sqrt{\sigma_{f,2}^2 + \sigma_o^2}} \right) / \sqrt{\sigma_{f,2}^2 + \sigma_o^2} \) as normalized distances. The above inequality indicates that if cluster 1’s normalized distance plus an offset \( \ln(\sigma_{f,1}^2 + \sigma_o^2) \) is smaller than cluster 2’s normalized distance plus another offset

---

**Fig. 2.** (top) Window-BT and (bottom) 640 hPa QVAPOR ensemble histograms at the locations (left) A, (center) B, and (right) C labeled in Fig. 1. In each histogram, the black lines outline the histogram formed by sorting the 50 members into 11 bins. Similarly, the blue lines outline the histograms formed by sorting the cloudy members into the same bins, and the red lines outline the histograms formed by sorting the clear members into the same bins. Note that location A has 16 cloudy members and 34 clear members (32% probability of cloudiness), location B has 29 cloudy members and 21 clear members (58% probability of cloudiness), and location C has 44 cloudy members and 6 clear members (88% probability of cloudiness).
[ln(σ₁² + σ₂²)], then cluster 1’s weight increases and cluster 2’s weight decreases. More generally, Bayesian inference increases the weight of the cluster with a smaller offset normalized distance.

It can be shown that \( \overline{w}_g \) and \( P_g \) are equivalent to performing Kalman filter updates on their corresponding \( \overline{w}_g \) and \( P_g \) (Anderson and Anderson 1999; Dovera and Della Rossa 2011):

\[
\overline{w}_g = \overline{w}_g + K_g \left( y - \overline{y}_g \right),
\]

\[
P_g = (I - K_g H)P_g,
\]

where \( K_g \) is the Kalman gain matrix:

\[
K_g = P_g^f H^T (σ_f² + σ_o²)^{-1}.
\]

In summary, applying Bayes’s rule to combine a bi-Gaussian prior PDF with Gaussian observations yields a bi-Gaussian posterior PDF. The Bayesian inference process increases the weight of the Gaussian kernel that is closer to the observation, while decreasing the weight of the Gaussian kernel further from the observation. Besides the weight update, the Bayesian inference process also modifies the mean and covariance matrix of each Gaussian via the Kalman filter equations.

b. Ensemble representation of the bi-Gaussian Bayesian inference

A BGEnKF requires a representation of \( w_g^f, \overline{w}_g, \) and \( P_g \) and an ensemble update scheme. As explained in section 3a, the ensemble can be broken into two clusters by the presence or
absence of clouds at the observation location. Supposing that \( N_{E,1}^{g} \) and \( N_{E,2}^{g} \) are the number of members in the clear and cloudy clusters, we can then calculate the prior weights via

\[
w_{g}^{f} = \frac{N_{E,1}^{g}}{N_{E,1}^{g} + N_{E,2}^{g}} \forall g = 1, 2.
\]

(8)

Furthermore,

\[
\overline{\psi}_{g}^{f} = \frac{1}{N_{E, g}^{f}} \sum_{n=1}^{N_{E, g}^{f}} \psi_{n}^{f} \forall g = 1, 2,
\]

where \( \psi_{n}^{f} \) is the \( n \)th member residing in cluster \( g \).

Since it is usually impractical to explicitly compute sample covariance matrices in high-dimensional spaces, the covariance matrix for a Gaussian is approximated by the ensemble perturbations in the corresponding cluster. The \( n \)th prior perturbation in the \( g \)th cluster is defined as

\[
\psi_{n}^{f} = \psi_{n}^{f} - \overline{\psi}_{g}^{f} \forall n = 1, 2, \ldots, N_{E, g}^{f}.
\]

(9)

After representing \( w_{g}^{f}, \overline{\psi}_{g}^{f}, \) and \( \psi_{n}^{f} \), the prior ensemble will be updated to produce a posterior ensemble that is consistent with Eqs. (3), (6), and (7). The values of \( \overline{\psi}_{g}^{f} \) and the posterior perturbations that represent \( \psi_{g}^{f} \) can be computed using the ensemble Kalman filter update equations on each cluster individually. To compute \( \overline{\psi}_{g}^{f} \), \( K_{g} \) is first computed directly from the \( g \)th cluster’s ensemble perturbations. In the case of assimilating a single observation,

\[
K_{g} = \frac{1}{\sigma_{g}^{2} + \sigma_{o}^{2}} \left[ \frac{1}{N_{E, g}^{f} - 1} \sum_{n=1}^{N_{E, g}^{f}} \left( \psi_{n}^{f} - \overline{\psi}_{g}^{f} \right) \right] \forall g = 1, 2,
\]

where \( y_{n}^{f} \) is the observed simulation in cluster \( g \). Then, \( \overline{\psi}_{g}^{f} \) is computed via Eq. (6).

The deterministic ensemble square root filter (EnSRF) proposed by Whitaker and Hamill (2002) is used in this study to update the prior perturbations in each cluster:

\[
\psi_{n}^{g} = \psi_{n}^{g} - \sqrt{\frac{\sigma_{o}^{2}}{\sigma_{g}^{2} + \sigma_{o}^{2}}} K_{g} \left( y_{n}^{g} - \overline{\psi}_{g}^{f} \right) \forall n = 1, 2, \ldots, N_{E, g}^{f},
\]

and \( g = 1, 2 \).

Note that the ensemble adjustment Kalman filter (EAKF) proposed by Anderson (2001) should yield similar results since the EAKF is a variant of the square root filter.

Since the \( g \)th prior weight is represented by the number of members in cluster \( g \), to represent the \( g \)th posterior weight, the number of members in the \( g \)th cluster will be changed to

\[
N_{E, g}^{f} = \text{round}(w_{g}^{f} N_{E}^{f}) \forall g = 1, 2,
\]

where \( \text{round}(\cdot) \) rounds a scalar argument to the nearest integer. We round \( w_{g}^{f} N_{E}^{f} \) to the nearest integer because a noninteger number of ensemble members cannot exist. Note that this rounding can introduce errors in the ensemble’s representation of the posterior weights, and the expected error is inversely proportional to the ensemble size. For practical ensemble sizes (10–100 members), this rounding error would be between 0.1 and 0.01.

If \( N_{E, g}^{f} \) is different from \( N_{E, g}^{f} \), we need to transfer members from one cluster to another. The number of members transferred is \( n_{\text{shift}} \). Ideally, the means and covariances of both clusters are preserved by this transfer process. However, in practical situations (\( N_{E} < N_{e} \)), the rank of the shrinking cluster covariance falls when \( n_{\text{shift}} \) members are removed, making it impossible to preserve the shrinking cluster covariance. It is nonetheless possible to preserve the shrinking cluster’s mean while minimizing the modifications to the perturbations: simply delete \( n_{\text{shift}} \) members with the smallest perturbations, and then subtract the mean of the remaining perturbations from the remaining perturbations. Finally, as we will demonstrate later, the ensemble mean and covariances of the expanding cluster can be preserved by the cluster transfer process.

We can construct the post-transfer expanding cluster perturbations through a linear combination of the pre-transfer perturbations. Let \( \Psi_{e, \text{old}} \) be an \( N_{e} \times N_{E, e} \) matrix that contains the pre-transfer perturbations of the expanding cluster in its columns, where \( N_{E, e}^{f} \) is the number of members in the expanding cluster before the transfer. Also, let \( \Psi_{e, \text{new}} \) be an \( N_{e} \times (N_{E, e} + n_{\text{shift}}) \) matrix that holds the post-transfer perturbations of the expanding cluster in its columns. The idea that the post-transfer perturbations are a linear combination of the pre-transfer perturbations can be expressed as

\[
\Psi_{e, \text{new}} = \Psi_{e, \text{old}} T,
\]

(10)

where \( T \) is an \( N_{E, e} \times (N_{E, e} + n_{\text{shift}}) \) matrix that needs to be determined.

We want the sample covariance of the post-transfer perturbations to be the same as that of the pre-transfer perturbations. In other words,

\[
\frac{1}{N_{E, e} + n_{\text{shift}} - 1} \Psi_{e, \text{new}}^{T} \Psi_{e, \text{new}} = \frac{1}{N_{E, e} - 1} \Psi_{e, \text{old}}^{T} \Psi_{e, \text{old}}.
\]

(11)

If we substitute Eq. (11) into Eq. (12), we obtain

\[
\frac{1}{N_{E, e} + n_{\text{shift}} - 1} \Psi_{e, \text{old}}^{T} \Psi_{e, \text{old}}^{T} = \frac{1}{N_{E, e} - 1} \Psi_{e, \text{old}}^{T} \Psi_{e, \text{old}}^{T}.
\]

(12)

This suggests that if

\[
T T^{T} = k I_{N_{E, e}}^{f},
\]

(13)

where \( I_{N_{E, e}}^{f} \) is an \( N_{E, e} \times N_{E, e} \) identity matrix and \( k = \sqrt{(N_{\text{shift}} + N_{E, e}^{f} - 1)(N_{E, e}^{f} - 1)} \), the covariance of the post-transfer perturbations will be the same as that of the pre-transfer perturbations.

Another desirable property of the post-transfer perturbations is that their sum yields a zero vector. This property ensures that the post-transfer and pre-transfer expanding clusters...
have the same ensemble means. The zero-sum property can be expressed as

$$
\Psi'_{c,\text{new}} = \begin{bmatrix} 1 \\ \vdots \\ 1\ N_f + n_{\text{shift}} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0\ N_c \end{bmatrix},
$$

(14)

where

$$
\begin{bmatrix} 1 \\ \vdots \\ 1\ N_f + n_{\text{shift}} \end{bmatrix}
$$

is a vector of ones with $N_f + n_{\text{shift}}$ rows, and

$$
\begin{bmatrix} 0 \\ \vdots \\ 0\ N_c \end{bmatrix}
$$

is a vector of zeros with $N_c$ rows.

Substituting Eq. (11) into Eq. (14) yields

$$
\Psi'_{c,\text{old}} T = \begin{bmatrix} 1 \\ \vdots \\ 1\ N_f + n_{\text{shift}} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0\ N_c \end{bmatrix}.
$$

(15)

Since the sum of the pre-transfer perturbations is zero,

$$
\Psi'_{c,\text{old}} c = \begin{bmatrix} 1 \\ \vdots \\ 1\ N_f + n_{\text{shift}} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0\ N_c \end{bmatrix},
$$

(16)

where $c$ can be any real number.

By comparing Eqs. (15) and (16), we can infer that

$$
T = \begin{bmatrix} 1 \\ \vdots \\ 1\ N_f + n_{\text{shift}} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1\ N_f \end{bmatrix}.
$$

In other words, if the elements in every row of $T$ sum to $c$, the zero-sum property is fulfilled. For convenience, we have set $c = k$. Thus,

$$
T = \begin{bmatrix} 1 \\ \vdots \\ 1\ N_f + n_{\text{shift}} \end{bmatrix} = k \begin{bmatrix} 1 \\ \vdots \\ 1\ N_f \end{bmatrix}.
$$

(17)

While Eqs. (13) and (17) provide guidance on the properties of $T$, it can be shown that there are an infinite number of solutions to Eqs. (13) and (17) for realistic values of $N_f$. Heuristic arguments can be used to constrain our choice of $T$. First, it is desired that the pre-transfer expanding cluster perturbations are preserved as much as possible. However, as shown in appendix A, it is impossible to preserve all the perturbations for all possible values of $n_{\text{shift}}$. Thus, a second property is considered: if there are some perturbations that are difficult to preserve, we should at least preserve the directions of the perturbations.

One possible formulation of $T$ that satisfies these properties is

$$
T = \begin{bmatrix} kI_{(N_f - m)} & 0 \\ 0 & I_m \end{bmatrix}
$$

where

$$
m = \begin{cases} n_{\text{shift}} - 1 & \forall n_{\text{shift}} \leq N_f, \\ N_f & \forall n_{\text{shift}} > N_f. 
\end{cases}
$$

(18)

and $E$ is an $m \times n_{\text{shift}}$ matrix that is described in appendix B. When $n_{\text{shift}} = 1$, the matrices $I_m$ and $E$ vanish from $T$ because $I_m$ and $E$ have zero rows. Furthermore, in the case where $n_{\text{shift}} > N_f$, the matrix $kI_{(N_f - m)}$ vanishes for similar reasons.

This formulation can be shown to satisfy the two desired properties. Suppose that $\Psi'_{c,\text{old}} = [\psi'_1, \psi'_2, \ldots, \psi'_{N_f - 1}, \psi'_{N_f}]$, where $\psi'_n$ is the $n$th perturbation (i.e., column) in $\Psi'_{c,\text{old}}$. Substituting Eq. (18) into Eq. (11) yields

$$
\Psi'_{c,\text{new}} = \begin{bmatrix} \Psi_{\text{infl}} & \Psi_{\text{copy}} & \Psi_{\text{comb}} \end{bmatrix},
$$

where

$$
\Psi_{\text{infl}} = \begin{bmatrix} k\psi'_1, k\psi'_2, \ldots, k\psi'_{N_f - m - 1}, k\psi'_{N_f} \end{bmatrix},
$$

$$
\Psi_{\text{copy}} = \begin{bmatrix} \psi'_{N_f - m - 1}, \psi'_{N_f - m + 1}, \ldots, \psi'_{N_f - 1}, \psi'_{N_f} \end{bmatrix},
$$

$$
\Psi_{\text{comb}} = \begin{bmatrix} \psi'_{N_f - m - 1}, \psi'_{N_f - m + 1}, \ldots, \psi'_{N_f - 1}, \psi'_{N_f} \end{bmatrix}E.
$$

The appearance of $\Psi_{\text{infl}}$ and $\Psi_{\text{copy}}$ in $\Psi'_{c,\text{new}}$ indicates that all of the pre-transfer expanding cluster perturbation directions are preserved in the post-transfer expanding cluster. Out of $N_f$, the post-transfer expanding cluster perturbations, $N_f - m$ of them are inflated versions of pre-transfer expanding cluster perturbations ($\Psi_{\text{infl}}$), and the remaining $m$ perturbations are preserved exactly by the transfer ($\Psi_{\text{copy}}$). The $\Psi_{\text{comb}}$ indicates that $n_{\text{shift}}$ perturbations were constructed by a linear combination of the $m$ preserved perturbations mentioned earlier.

The choice of which perturbations to inflate and preserve is arbitrary. In this study, the smallest $N_f - m$ perturbations are chosen to be inflated, and the rest of the perturbations are preserved. This is because the inflation of the larger perturbations is more likely to produce numerical instability during integration than that of the smallest perturbations. While this choice would introduce non-Gaussianity by decreasing the PDF near the mode and increasing the PDF near the tails, this is preferable to numerical integration failure. Future studies can investigate other viable choices of perturbations to inflate.

Using the method of constructing $E$ in appendix B, we can transfer members between clusters, update the kernel means and represent the update to the kernel covariances by applying the EnSRF within each kernel. A BGEEnKF algorithm to serially assimilate infrared observations is proposed below. The figures that illustrate each step, as well as the estimated computational complexity of each step, are indicated in the
brackets at the end of each step’s description. Note that $N_{E_{i}}^{c}$ denotes the pre-transfer size of the shrinking cluster, and $C_{v}$ denotes the computational cost of running the nonlinear observation operator on an ensemble member to generate a single simulated observation.

1) Select an observation that has not been assimilated and run the nonlinear observation operator on the ensemble to generate simulations of said observation $[\sim N_{E} \times C_{v}]$.
2) Partition the ensemble into two clusters, depending on whether the members have clouds at the observed column [Figs. 4a,b; $\sim N_{E}$].
3) Compute the prior weight, mean and ensemble perturbations of each cluster [Fig. 4c; $\sim N_{c} \times N_{E}$].
4) In each cluster, compute $K_{c}$ and perform the EnSRF on the cluster mean and perturbations [Fig. 4d; $\sim N_{c} \times N_{E}$].
5) Compute the posterior weights via Eq. (3) [Fig. 4e; $\sim 1$].
6) Estimate the number of posterior members in each cluster via Eq. (10), identify the expanding and shrinking clusters, and determine the number of members to transfer from the shrinking to expanding cluster $(n_{\text{shift}}) [\sim 1]$. 
7) Compute the Euclidean norm of the perturbations in the shrinking cluster, delete $n_{\text{shift}}$ shrinking cluster perturbation with the smallest Euclidean norms, and then subtract the mean of the remaining shrinking cluster perturbations from the remaining shrinking cluster perturbations [Fig. 4g; $\sim N_{c} \times N_{E}$].
8) In the expanding cluster, inflate the $N_{E_{i}}^{c} - m$ smallest perturbations by multiplying with factor $k$ [Fig. 4h; $\sim (N_{E_{i}}^{c} - m) \times N_{c}$].
9) Do nothing to the $m$ uninflected perturbations in the expanding cluster [Fig. 4h; $\sim 0$].
10) Construct matrix $E$ using the method in appendix B [$\sim m^2$].
11) Construct new perturbations in the expanding cluster by constructing a matrix containing the uninflected perturbations in its columns, and then right-multiplying said matrix with $E$. This is essentially evaluating the definition of $\Psi_{\text{comb}}$ [Fig. 4i; $\sim N_{c} \times m^2$].
12) Apply the expanding cluster perturbations onto the expanding cluster mean to construct the expanding cluster’s members, and likewise for the shrinking cluster $[\sim N_{c} \times N_{E_{i}}^{c}]$.
13) Return to step 1 and repeat until all infrared observations have been assimilated $[\sim 1]$. 

Note that steps 8 to 11 are equivalent to evaluating the right-hand side of Eq. (11) using matrix $T$ defined in Eq. (18). It should also be noted that in step 4, for each cluster, the Kalman gain matrix is computed using the subset of prior members that belong to said cluster.

In the case of operational DA with localization, we would typically expect $N_{c} \sim 10^4$ and $N_{E} \sim 10$. We can also suppose that $N_{E_{i}}^{c}$, $n_{\text{shift}}$, and $m$ to have the same order of magnitude as $N_{E}$ [i.e., $\sim 10$]. Plugging in these numbers reveals that step 11 ($N_{c} \times m^2$) has the highest order of magnitude estimate of $\sim 10^6$. Thus, step 11 is expected to be the dominant source of the BGEnKF’s computational complexity. The BGEnKF’s and standard EnKF’s computational complexities are displayed in Table 1.

While the BGEnKF has a higher computational cost than that of the EnKF (Table 1), the BGEnKF has a reasonable cost compared to earlier GMM methods that allow covariances to vary among kernels. Earlier methods that require the square root of the posterior covariance matrices in state space have a computational complexity that is untenable for high-order models $[\sim N_{E}^3 (= 10^{12})]$ (Rasmussen and Williams 2005). Furthermore, the BGEnKF does not depend on the forecast model being used, which makes it easier to implement and upkeep when compared against methods that utilize stochastic subspaces (Sondergaard and Lermusiaux 2013a,b; Lolla and Lermusiaux 2017a,b).

As a side note, the BGEnKF algorithm bears some resemblance to particle filters (e.g., van Leeuwen 2010; Poterjoy 2016; Vetra-Carvalho et al. 2018) in the sense that members that are not consistent with the observations are replaced by resampling from members that are consistent with the observations.

4. 40-variable L96 tests
a. Experiment setup

To examine the potential effects of assimilating infrared radiance observations using the BGEnKF and EnKF, the 40-variable L96 model was used as a proxy for an atmospheric model. The forcing term $F$ was set to 8 to produce chaotic model behavior. The Runge–Kutta fourth-order scheme was used to integrate the model. The model time step is set to 0.025 L96 time units. All ensembles used in this study are initiated by drawing from $\mathcal{N}(0,1)$, and then integrated for 25 L96 time units before the start of DA. This procedure resulted in ensembles that are effectively drawn from climatology. All results that will be shown in this paper are the statistics of running 40 independent trials with the ensemble and truth drawn from white noise and spun up.

The L96 model can be viewed as a simplified model of geopotential on a latitude ring (Lorenz 2006). This means that the spatial gradient of the L96 model variable can be treated as an indicator of upward or downward motion. Locations with upward motions indicated can be treated as cloudy locations, and locations with downward motions can be treated as clear-sky locations. For this project, we have considered locations with negative gradient $(x_{i+1} - x_{i} < 0)$ as cloudy locations, and locations with zero to positive gradient $(x_{i+1} - x_{i} \geq 0)$ as clear-sky locations. As seen in Fig. 3c and similar to the WRF example earlier, the L96 model variables for clear-sky and cloudy members follow different distributions.

A nonlinear observation operator was necessary to construct observations with ensemble statistics that mimic those of Window-BT ensemble statistics. The pseudoradiance observation at the $ith$ grid point in the Lorenz 1996 model $(y_{i})$ is defined as

$$y_{i} = \begin{cases} 
280 - 3 \left( \frac{\partial x}{\partial t} \right)^{2/3} & \forall x_{i+1} - x_{i} < 0 \\
280 + x_{i} + \left( \frac{\partial x}{\partial t} \right)^{2/3} & \forall x_{i+1} - x_{i} \geq 0
\end{cases}, \quad (19)
$$

where $x_{i}$ is the value of the L96 model variable at $ith$ grid point. The time tendency term $(\partial x/\partial t)$ in Eq. (19) is the L96 tendency equation. The $x_{i+1} - x_{i} < 0$ part of Eq. (19) corresponds to radiances under cloudy skies, and the $x_{i+1} - x_{i} \geq 0$ part corresponds to clear-sky radiances. This observation
FIG. 4. Illustration of the BGEnKF procedure for assimilating one observation in a two-variable scenario, using 10 ensemble members. First, (a) the prior ensemble members are broken into (b) two clusters, each cluster representing a Gaussian kernel. In this illustration, 6 members go to cluster 1, and 4 members go to cluster 2. (c) The prior weight, mean, and covariance of each Gaussian kernel is then computed. (d) Following that, the ensemble Kalman filter procedure is performed on each cluster to generate the posterior mean and covariance of each Gaussian kernel. (e) Using the prior weights, means, and covariances of both clusters, the posterior weights were then computed. (f) Simply a close-up version of (e). (g) In the shrinking cluster (cluster 2), members are randomly chosen and deleted until this cluster has an appropriate number of members to reflect its posterior weight. In this illustration, all 4 members in cluster 2 are deleted. (h) After the deletion, selected perturbations (three in this illustration) in the expanding cluster are inflated. (i) Finally, linear combinations of the uninflated perturbations are used to construct new perturbations in the expanding cluster to reflect the expanding cluster’s posterior weight. In this illustration, 4 perturbations are constructed.
The optimal ROI for 20-member scenario is determined by the average computation time to assimilate one observation into the ensemble. The computational complexity for this scenario is estimated to be 0.965 ms per observation. The leading-order term for computational complexity is given by $N_{m}N_{E}$, where $N_{m}$ is the number of members being transferred from the shrinking cluster to the expanding cluster, minus one.

Inflation of the prior is necessary for the various DA filters to ensure that the ensemble is clear or cloudy. In this study, the threshold is set to 20% of the ensemble size when the size of either cluster is below a threshold. This ensures that the potentially unphysical members in the post-transfer shrinking cluster are eliminated when the Bayesian inference is certain that the ensemble should be clear or cloudy.

When a clear-sky observation is assimilated, the Kalman update to the cloudy cluster is performed. On the other hand, if an observation is too far from either cluster (i.e., if $|y_{o} - y_{g}| > 3 \sqrt{\sigma_{g}^{2} + \sigma_{n}^{2}}$ for both $g = 1$ and 2), an EnKF is performed instead of the BGEnKF. An observation is rejected if it shows a large departure from the mean of all ensemble members.

Observations with large departures from the prior ensemble mean ($|y_{o} - y_{g}| > 3 \sqrt{\sigma_{g}^{2} + \sigma_{n}^{2}}$) are typically rejected in Gaussian data assimilation (Fletcher 2017). In this study, if an observation is too far from either cluster (i.e., if $|y_{o} - y_{g}| > 3 \sqrt{\sigma_{g}^{2} + \sigma_{n}^{2}}$ for both $g = 1$ and 2), an EnKF is performed instead of the BGEnKF. An observation is rejected if it shows a large departure from the mean of all ensemble members.

When a clear-sky observation is assimilated, the Kalman update to the cloudy cluster’s members can be unphysical. Likewise, there can be unphysical Kalman updates to the clear-sky cluster’s members when a cloudy observation is assimilated. Since the cluster that disagrees with the observation on cloudiness is typically the shrinking cluster, we have opted to set the shrinking cluster’s post-transfer size to zero and the expanding cluster’s post-transfer size to $N_{E}$ when the post-transfer size predicted by Eq. (10) is less than 20% of the ensemble. This ensures that the potentially unphysical members in the post-transfer shrinking cluster are eliminated when the Bayesian inference is certain that the ensemble should be clear or cloudy.

Finally, there can be instances where assimilating a clear-sky observation expands the cloudy cluster, or assimilating a...
cloudy observation expands the clear-sky cluster. This can happen because the BGEnKF does not explicitly consider whether the observation indicates clear or cloudy conditions; the BGEnKF simply expands the cluster with the smaller offset normalized distance [Eq. (5)]. To prevent potentially unphysical posterior states, when a clear-sky observation ($y_o > 280$ K) expands the cloudy cluster, or when a cloudy observation ($y_o < 270$ K) expands the clear-sky cluster, the EnKF is employed instead of the BGEnKF.

Two sets of experiments were carried out. The first set examined the effectiveness of the BGEnKF and EnKF with 800-member ensembles and applied no localization. The second set used 20-member ensembles and applied localization.

b. 800-member ensemble results

Figure 5 shows the root-mean-square (RMS) error (RMSE) and the ratio of RMS spread (RMSS) to RMSE (RMSS/RMSE). When no inflation is applied, both the EnKF and BGEnKF generally had performance similar to or worse than if no DA was used (NoDA). The poor performance of the two filters is generally accompanied by RMSS/RMSE values under 0.8 (Figs. 5b,e,h). These low RMSS/RMSE values suggest that severe underdispersion is responsible.

To mitigate the underdispersion, adaptive multiplicative inflation was introduced [Eq. (20)]. According to Fig. 5, this inflation method mitigated the underdispersion (RMSS/RMSE $\geq 1$) and generally improved the RMSE performance of both filters. The experiments using either filter now perform better than NoDA. This improved performance confirms that inflation is necessary for effective assimilation of pseudoradiance observations. Henceforth, we will only examine the results of using the filters with inflation.

According to Fig. 5 and Table 1, the BGEnKF experiments have RMSE that is either better than or comparable to that of the EnKF experiment. This comparison confirms the expectation that a bi-Gaussian filter is more appropriate.
at assimilating pseudoradiances with bi-Gaussian statistics (Fig. 3d) than a Gaussian filtering method.

Aside from that, the BGEnKF’s RMSE advantage appears to be small in the regime of dense and frequent observations (e.g., assimilating observations everywhere, every 0.1$t$), and in the regime of sparse and infrequent observations (e.g., assimilating observations every four grid points, every 0.2$t$). The weak advantage in the regime of dense and frequent observations is due to the small ensemble spread resulting from assimilating dense and frequent observations. With small spread, most of the ensemble almost always congregates within one of the clusters, causing the BGEnKF to almost always default to the EnKF (Fig. 5c), which explains the similarity in the BGEnKF’s and EnKF’s RMSE. In the other regime of weak advantage, the sparse and infrequent observations can only weakly constrain the BGEnKF’s and EnKF’s errors. This means that the BGEnKF’s weak advantage over the EnKF is because both experiments’ RMSEs are approaching the saturation value indicated by NoDA’s RMSE.

In summary, when tested with 800 ensemble members, the BGEnKF generally performs better than the EnKF. Furthermore, the BGEnKF’s advantage over the EnKF is weak in the regime of dense and frequent observations due to small ensemble spread, and weak in the regime of sparse and infrequent observations due to error saturation.

c. 20-member ensemble results

In practice, the ensemble size is usually smaller than the dimensions of the state space, meaning that there can be significant sampling errors in the ensemble statistics. To mimic this scenario, we have tested the BGEnKF and EnKF with 20 ensemble members. To mitigate sampling errors, the localization described in section 4a was applied. For each pair of observation spacing and cycling interval, ROI ranging from 1 to 20 grid spacings, in steps of 1 grid spacing, were tested 40 times each. The results plotted in Fig. 6 are for the ROI that yielded the smallest average prior RMSE for the 40 tests (optimal ROI).

Several features identified in the 800-member experiments are evident in the 20-member experiments. First, the BGEnKF’s RMSE is generally better or comparable to the EnKF’s. With adaptive inflation, both the BGEnKF and EnKF did not suffer from underdispersion. Finally,
similar nonmonotonic relationships between the BGEnKF’s RMSE advantage with respect to spatiotemporal density are seen in both the 800-member and 20-member experiments.

However, the RMSE’s of the 20-member experiments exhibit several differences with respect to that of 800-member experiments. While the RMSEs of both DA methods are smaller than the climatological value, their RMSEs are generally larger than when 800-member ensembles were employed (Fig. 6). Furthermore, the RMSE advantage of the 20-member BGEnKF over the 20-member EnKF is noticeably smaller than for their 800-member counterparts. These can be viewed as the impacts of increased sampling errors due to reduced ensemble size.

The optimal ROI for each pair of observation spacing and cycling interval are also shown in Fig. 6. For both the EnKF and BGEnKF, the optimal ROI generally falls with decreasing spatiotemporal density of assimilated observations. Furthermore, the optimal ROI of the BGEnKF is generally smaller than that of the EnKF. The optimal ROI of the BGEnKF also appears to be less sensitive to decreasing the spatiotemporal density of assimilated observations than that of the EnKF. Finally, the optimal ROI of the BGEnKF is generally at least as large as the grid spacing of assimilated observations.

The behavior of the BGEnKF’s optimal ROI has several implications. First, since the optimal ROI of the BGEnKF is smaller than that of the EnKF, the number of variables constrained by the BGEnKF is smaller than that of the EnKF. This suggests that the computational complexity estimate of the BGEnKF made earlier (section 3b or Table 1) might be excessive. In other words, if the BGEnKF’s optimal ROI is sufficiently smaller than the EnKF’s in practice, the BGEnKF’s computational cost might be comparable to the EnKF’s. Second, because the BGEnKF’s optimal ROI can be similar to that of the observation spacing, care must be taken when thinning infrared radiances. If the observations are thinned too much, the ROI can result in pockets between observations where DA has no direct impact. Finally, the fact that the BGEnKF’s optimal ROI varies with spatiotemporal assimilation density implies that an adaptive localization method might be a boon for practical applications of the BGEnKF.

**d. Integration stability of BGEnKF versus EnKF**

It is important to check if the BGEnKF will trigger severe numerical instabilities during forecasts. However, the Lorenz 1996 model is fairly resistant to initial shocks: numerical instabilities do not usually occur unless analyzed model values have very large absolute values. In light of this resistance, we will use the RMS of the 40-variable tendency of the model as a proxy to qualitatively estimate the BGEnKF’s potential to produce noise in practical situations. This RMS tendency is computed by first evaluating the model’s prognostic equation on an analyzed ensemble member, and then taking the RMS of the resulting 40-element rate-of-change vector.

Figure 7 shows the ensemble-averaged RMS tendencies of the NoDA and the various DA experiments, for various observation spacing, cycling intervals and ensemble sizes. Since continuous model runs without DA are numerically stable, NoDA experiments’ RMS tendencies will be taken
as the typical RMS tendency values. RMS tendency values that are increasingly greater than that of the NoDA indicate increasing potential for numerical noise growth. The EnKF experiments have the largest RMS tendencies among all experiments, implying that the EnKF has the largest potential to generate numerical instabilities when assimilating infrared observation. While the BGEnKF experiments generally have RMS tendencies greater than that of the NoDA, their RMS tendencies fall between that of EnKF and NoDA. This feature suggests that the BGEnKF has a smaller potential to generate numerical instabilities than the EnKF when assimilating infrared observations.

5. Summary and conclusions

We have highlighted the possibility that clear-sky and cloudy members are drawn from distinctly different distributions using an ensemble of WRF simulations of a linear MCS over the Maritime Continent. This result is consistent with the findings of Harnisch et al. (2016), Honda et al. (2018), and Minamide and Zhang (2017). Motivated by these results and studies, we proposed a formulation of the BGEnKF that can be more effective at assimilating infrared radiance observations than the EnKF. This new BGEnKF does not utilize the EM methods traditionally used to find the parameters of the GMM distribution, hence avoiding the difficulties in using EM methods. Furthermore, this new BGEnKF is computationally efficient because it does not require the costly square root decompositions of covariance matrices used in earlier GMM DA methods. Finally, the proposed BGEnKF does not require the difficult and labor-intensive derivation of dynamically orthogonal field equations.

To motivate resource-intensive studies to examine the effectiveness of our BGEnKF with high-order models, proof-of-concept tests were conducted using the 40-variable L96 model as a proxy. A nonlinear pseudoradiance operator was used to generate observations from the L96 model. These observations emulated the statistics of CRTM-derived WRF ensemble infrared brightness temperatures.

We found that inflation of the prior perturbations was necessary for both the EnKF and BGEnKF to effectively assimilate the pseudoradiance observations. With adaptive inflation, for tests with 800 and 20 ensemble members, the BGEnKF has RMSEs that are smaller than or similar to that of the EnKF. Finally, the BGEnKF is less likely to produce numerical noise than the EnKF when assimilating pseudoradiance observations. These results with the Lorenz 1996 model suggest that the BGEnKF can potentially be more effective at assimilating infrared observations than the EnKF.

A caveat to note here is that the BGEnKF is more susceptible to sampling noise than the EnKF. In our L96 tests with just 8 ensemble members (not shown), the BGEnKF’s performance is slightly weaker than that of the EnKF. This is likely due to two reasons. First, the two pairs of means and covariances in the BGEnKF are estimated using subsets of the ensemble whereas the EnKF uses the full ensemble to estimate a single pair of mean and covariances. Second, when small ensembles are used, the prior weight estimation becomes imprecise. Thus, the BGEnKF is unlikely to be more effective than the EnKF when the ensemble size is small (ensemble sizes under 10).

Future studies will examine the effectiveness of the BGEnKF algorithm at assimilating window channel infrared radiances in the context of mesoscale-resolving observing system simulation experiments (OSSEs). The impact of the BGEnKF on analysis and forecast model fields should be quantified and compared against that of the EnKF.

It should be noted that the most expensive step in the BGEnKF arises from a matrix multiplication and the computational complexities listed in Table 1 are for single-processor implementation. Since matrix multiplication can be easily parallelized, it is likely that computational cost of executing the BGEnKF in parallel might not be substantially more than that of the EnKF. Subsequent studies using parallelized BGEnKF and EnKF with high-order models can confirm this prediction.

Another avenue for future research concerns the assumption that the prior distribution is drawn from a bi-Gaussian distribution. While the bi-Gaussian assumption is an improvement over the traditional Gaussian assumption, the prior distribution might be more accurately modeled with more than two Gaussian kernels. The challenge with using more kernels lies in finding a way to separate members into more groups, without using EM algorithms in high-dimensional space or a stochastic subspace. Furthermore, to estimate the kernel parameters with sufficient accuracy, a larger ensemble is needed when more kernels are employed.

The advent of infrared radiance DA has proven to be a boon for numerical weather analyses and forecasts. Advances in infrared radiance data assimilation methods thus far have enhanced the effectiveness of these observations. The development and implementation of an efficient BGEnKF for infrared DA might further improve numerical weather analyses and forecasts.

Acknowledgments. This study is supported by the Water Cycle and Climate Extremes Modelling (WACCEM) project, which is funded by the U.S. Department of Energy Office of Science Biological and Environmental Research, as part of the Regional and Global Climate Modeling program. This study is also supported by the Office of Naval Research (Grant N000141812517), as well as the Graduate Visitor Program under the Advanced Study Program at the National Center for Atmospheric Research. The National Center for Atmospheric Research is sponsored by the National Science Foundation. 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. The initial and boundary conditions used to set up the WRF ensemble were based off the ERA5 reanalysis and ensemble (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=form). All experiments were done on the Stampede2 supercomputer of the Extreme Science and Engineering Discovery Environment (XSEDE), deployed at the Texas Advanced Computing Center (TACC). We would also like to thank Christopher M. Hartman and Zhu
Yao for proofreading the manuscript, as well as Yue Ying for his helpful discussions.

Data availability statement. The data and code used in this study are available on the Pennsylvania State University's Data Commons (https://doi.org/10.26208/3snv-s428).

APPENDIX A

Proof: Not all Expanding Cluster Perturbations can be Preserved When $n_{\text{shift}} < N_{E,E}^f$

For all expanding cluster perturbations to be preserved by the cluster transfer process, we require

$$T = [I_{N_{E,E}^f}, F], \quad (A1)$$

where $F$ is an $N_{E,E}^f \times n_{\text{shift}}$ matrix. If we substitute Eq. (A1) into Eq. (10), we obtain

$$I_{N_{E,E}^f} + FF^T = k^2 I_{N_{E,E}^f}, \quad (A2)$$

where $FF^T$ is an $N_{E,E}^f \times N_{E,E}^f$ matrix. For $n_{\text{shift}} < N_{E,E}^f$, $F$ has a maximum rank of $n_{\text{shift}}$. This means the maximum rank of $FF^T$ is $n_{\text{shift}}$. However, for Eq. (A2) to be satisfied, $FF^T$ must be $(k^2 - 1)I_{N_{E,E}^f}$, which has rank $N_{E,E}^f$. This rank contradiction indicates that it is impossible to preserve all perturbations in the case where $n_{\text{shift}} < N_{E,E}^f$.

APPENDIX B

Procedure to Generate Matrix $E$

For Eq. (15) to satisfy the pair of constraints imposed by Eqs. (10) and (14), matrix $E$ needs to obey corresponding constraints. Substituting Eq. (15) into Eqs. (10) and (14) yields:

$$EE^T = \left( \frac{n_{\text{shift}}}{N_{E,E}^f - 1} \right) I_m, \quad (B1)$$

$$E = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n_{\text{shift}}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n_{\text{shift}}} \quad (B2)$$

It can be shown that there is at least one solution to Eqs. (B1) and (B2) for $n_{\text{shift}} \geq 2$. The case of $n_{\text{shift}} = 1$ does not need to be considered because $E$ vanishes from Eq. (15) in that case. All relevant values of $n_{\text{shift}}$ are thus accounted for.

If each column of $E$ is a vector drawn from some Gaussian distribution, Eq. (B2) indicates that the sum of all the columns of $E$ is greater than zero ($k > 1$, see definition earlier). In other words, the sample mean vector is positive. Equation (B1) thus cannot be a rescaled definition of the covariance matrix of the samples because the mean sample vector of $E$ was not removed before computing the outer product. To construct $E$, we need some rescaled definition of the covariance matrix.

The rescaled covariance matrix of $E$ can be derived by first separating $E$ into two $m \times n_{\text{shift}}$ matrices

$$E = E + E', \quad (B3)$$

where

$$E = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n_{\text{shift}}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}_{n_{\text{shift}}} \quad (B3)$$

If Eq. (B3) is substituted into Eq. (B2), we obtain

$$E = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n_{\text{shift}}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n_{\text{shift}}}$$

which indicates that

$$E' = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n_{\text{shift}}} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{n_{\text{shift}}} \quad (B4)$$

In other words, the sum of the column vectors of $E'$ are the sample perturbations of $E$.

To construct a rescaled covariance matrix of $E$, substitute Eq. (B3) into the left side of Eq. (B1):

$$EE^T = (E + E')(E^T + E'E)^T = EE^T + EE^T + E'E^T + EE'E^T$$

Considering that

$$EE^T = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n_{\text{shift}}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n_{\text{shift}}} \quad (B1)$$

and

$$E' = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n_{\text{shift}}} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{n_{\text{shift}}} \quad (B4)$$

it can be shown that $E'E^T = EE'E = 0$. Then.

$$EE^T = EE^T + E'E^T = \begin{bmatrix} n_{\text{shift}} \\ \vdots \\ n_{\text{shift}} \end{bmatrix}_{n_{\text{shift}}} \quad (B5)$$

Thus, a rescaled covariance matrix of $E$ ($C_E$) can be obtained:

$$C_E = EE^T = \begin{bmatrix} n_{\text{shift}} \\ \vdots \\ n_{\text{shift}} \end{bmatrix}_{n_{\text{shift}}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{m \times n_{\text{shift}}} \quad (B4)$$
Eqs. (B3) and (B4) suggest the following method of constructing $E$ using a random number generator for the case where $n_{\text{shift}} \geq 2$ (Rasmussen and Williams 2005). We have indicated the estimated computational complexity for each step in the brackets.

1) Draw an $m \times n_{\text{shift}}$ matrix $W$ of white noise using the one-variable standard normal distribution $[\sim O(m \times n_{\text{shift}})]$.

2) Ensure that the mean of the column vectors in $W$ is zero $[\sim O(m \times n_{\text{shift}})]$.

3) Use Cholesky factorization to compute $L_W$ such that $L_W(L_W^T = WW^T \sim O(m^2))$.

4) Invert $L_W$ $[\sim O(m^3)]$.

5) Use Cholesky factorization to compute $L_E$ such that $L_E(L_E^T = C_E \sim O(m^3))$.

6) Construct $E'$ via $E' = L_E(L_W)^{-1}W \sim O(m^3)]$.

7) Construct $E$ via $E = E' + E \sim O(m^3)]$.

For the case where $n_{\text{shift}} = 1$, just apply Eq. (18) directly.

REFERENCES


Köpken, C., G. Kelly, and J.-N. Thépaut, 2004: Assimilation of Meteosat radiance data within the 4D-Var system at


