The potential benefits of handling mixture statistics via a bi-Gaussian EnKF: tests with all-sky satellite infrared radiances

Man-Yau Chan[1], Xingchao Chen[2], and Jeffrey L. Anderson[3]

[1]The Pennsylvania State University
[2]Pennsylvania State University
[3]National Center for Atmospheric Research (UCAR)

November 22, 2022

Abstract

The meteorological characteristics of cloudy atmospheric columns can be very different from their clear counterparts. Thus, when a forecast ensemble is uncertain about the presence/absence of clouds at a specific atmospheric column (*i.e.*, some members are clear while others are cloudy), that column's ensemble statistics will contain a mixture of clear and cloudy statistics. Such mixtures are inconsistent with the ensemble data assimilation algorithms currently used in numerical weather prediction. Hence, ensemble data assimilation algorithms that can handle such mixtures can potentially outperform currently used algorithms.
In this study, we demonstrate the potential benefits of addressing such mixtures through a bi-Gaussian extension of the ensemble Kalman filter (BGEnKF). The BGEnKF is compared against the commonly used ensemble Kalman filter (EnKF) using perfect model observing system simulated experiments (OSSEs) with a realistic weather model (the Weather Research and Forecast model). Synthetic all-sky infrared radiance observations are assimilated in this study. In these OSSEs, the BGEnKF outperforms the EnKF in terms of the horizontal wind components, temperature, specific humidity, and simulated upper tropospheric water vapor channel infrared brightness temperatures.
This study is one of the first to demonstrate the potential of a Gaussian mixture model EnKF with a realistic weather model. Our results thus motivate future research towards improving numerical Earth system predictions though explicitly handling mixture statistics.

The potential benefits of handling mixture statistics via a bi-Gaussian EnKF: tests with all-sky satellite infrared radiances

Man-Yau Chan[1,2], Xingchao Chen[1,2], Jeffrey L. Anderson[3]

[1]Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, Pennsylvania, USA

[2]Center for Advanced Data Assimilation and Predictability Techniques, The Pennsylvania State University, University Park, Pennsylvania, USA

[3]Data Assimilation Research Section, Computational Information Systems Laboratory, National Center for Atmospheric Research, Boulder, Colorado, USA

Key Points:

- Current ensemble DA methods assume that forecasts follow a normal distribution. This assumption is often invalid.
- In this study, we propose a computationally efficient ensemble DA method that handles clear and cloudy forecasts separately.
- This study uses a realistic weather model (WRF) to show that this method can outperform the EnKF.

Corresponding author: Man-Yau Chan, chanmanyau@gmail.com

Abstract

The meteorological characteristics of cloudy atmospheric columns can be very different from their clear counterparts. Thus, when a forecast ensemble is uncertain about the presence/absence of clouds at a specific atmospheric column (*i.e.*, some members are clear while others are cloudy), that column's ensemble statistics will contain a mixture of clear and cloudy statistics. Such mixtures are inconsistent with the ensemble data assimilation algorithms currently used in numerical weather prediction. Hence, ensemble data assimilation algorithms that can handle such mixtures can potentially outperform currently used algorithms.

In this study, we demonstrate the potential benefits of addressing such mixtures through a bi-Gaussian extension of the ensemble Kalman filter (BGEnKF). The BGEnKF is compared against the commonly used ensemble Kalman filter (EnKF) using perfect model observing system simulated experiments (OSSEs) with a realistic weather model (the Weather Research and Forecast model). Synthetic all-sky infrared radiance observations are assimilated in this study. In these OSSEs, the BGEnKF outperforms the EnKF in terms of the horizontal wind components, temperature, specific humidity, and simulated upper tropospheric water vapor channel infrared brightness temperatures.

This study is one of the first to demonstrate the potential of a Gaussian mixture model EnKF with a realistic weather model. Our results thus motivate future research towards improving numerical Earth system predictions though explicitly handling mixture statistics.

Plain Language Summary

The accuracy of a computer weather forecast often depends on the accuracy of the information inputted into the computer forecast system. The accuracy of the input in turn depends on the accuracy of the input-constructing algorithm. Such algorithms often use probabilistic forecasts from an earlier point in time and current atmospheric measurements to construct the inputs.

A common assumption in input-constructing algorithms is that the probabilistic forecasts follow multivariate normal distributions (henceforth called the normality assumption). However, in the frequent situation where the probabilistic forecasts are uncertain about the presence/absence of clouds, the normality assumption is violated. This is because clear atmospheric columns and cloudy atmospheric columns have distinctly different thermodynamic and dynamic characteristics. Such probabilistic forecasts thus have mixed statistics (henceforth termed mixed probabilistic forecasts). Addressing these mixed statistics can potentially improve forecasts.

In this study, we propose a new input-constructing algorithm that can explicitly handle mixed probabilistic forecasts. Compared to an existing popular algorithm, our algorithm is nearly as fast and can produced more accurate forecast inputs. Our work thus suggests that weather forecasts can be improved by upgrading input-constructing algorithms to treat a common situation where the normality assumption is violated.

1 Introduction

Earth system analysis and forecasting systems rely on ensemble data assimilation (ensemble DA, or EDA) methods to convert observations into corrections for Earth system model variables (Keppenne et al., 2005; Reichle et al., 2009; Edwards et al., 2015; Stammer et al., 2016; Park & Xu, 2016; ECMWF, 2016; Helmert et al., 2018; Hersbach et al., 2020). Current operational EDA methods typically assume that every member in an input forecast ensemble is drawn from a distribution only containing a single Gaussian kernel [*i.e.*, a Gaussian distribution; henceforth termed the unmixed ensemble assumption; *e.g.*, Geer et al. (2018) and Dowell et al. (2022)]. The effectiveness of such methods can thus can be limited by the validity of this assumption.

The unmixed ensemble assumption is violated for ensembles that are uncertain about the presence or absence of clouds at any model grid point. This is because clear atmospheric columns and cloudy atmospheric columns often have different dynamic, thermodynamic, and radiative properties [*e.g.*, Emanuel (1994), Markowski and Richardson (2010)]. Cloudy statistics are thus often different from clear statistics [*e.g.*, Grimes and Pardo-Igúzquiza (2010); Geer and Bauer (2011)]. If some ensemble members are cloudy at a location, and other members are clear at this location, the ensemble can exhibit mixed statistics (Harnisch et al., 2016; Minamide & Zhang, 2017; Honda et al., 2018; Chan, Anderson, & Chen, 2020). More evidence of mixed statistics can be found in the supporting information. The effectiveness of current operational EDA methods is likely limited in such situations.

This limitation can be mitigated by extending current operational EDA methods to handle mixed statistics. One possibility is to extend the commonly used ensemble Kalman filter, or the EnKF (Evensen, 1994; P. L. Houtekamer & Mitchell, 1998; Burgers et al., 1998; Tippett et al., 2003; Anderson, 2003; Whitaker & Hamill, 2002; Keppenne et al., 2005; Hunt et al., 2007; Reichle et al., 2009; Stammer et al., 2016; Edwards et al., 2015; Park & Xu, 2016; Helmert et al., 2018), to handle members drawn from forecast distributions with two Gaussian kernels. Specifically, we assume that forecast members that are clear at an observation site (henceforth, clear members) are drawn from one Gaussian kernel, and forecast members that are cloudy at this site (henceforth, cloudy members) are drawn from a different Gaussian kernel. This bi-Gaussian extension of the EnKF (henceforth, the BGEnKF) allows the clear ensemble statistics to be handled separately from the cloudy ensemble statistics (Chan, Anderson, & Chen, 2020), thus addressing the issue of mixed statistics.

We recently proposed a computationally efficient BGEnKF to handle mixtures of clear and cloudy members [Chan, Anderson, and Chen (2020); henceforth, the CAC20 BGEnKF]. Unlike similar methods proposed in the past (Dovera & Della Rossa, 2011; Reich, 2012; Sondergaard & Lermusiaux, 2013a, 2013b), the CAC20 BGEnKF does not use an expectation maximization (EM) algorithm to estimate the mean and covariances of the two Gaussian kernels. Instead, the CAC20 BGEnKF assigns the the sample mean and covariances of the cloudy members to one Gaussian kernel, and those of the clear members to the other Gaussian kernel. This assignment circumvents the computational cost and issues associated with using the EM algorithm in high dimensional spaces [see Chan, Anderson, and Chen (2020) for more information]. Furthermore, the CAC20 BGEnKF converts clear members into cloudy members, or *vice versa*, without involving the costly square-root computations or Cholesky decompositions of high-dimensional forecast covariance matrices.

The purpose of this study is to demonstrate that a variant of the CAC20 BGEnKF can outperform the EnKF using a realistic high-order weather model (the Weather Research and Forecast model; WRF). To do so, this new BGEnKF is implemented into the state-of-the-art Pennsylvania State University EnKF system [henceforth, the PSU-EnKF system; Meng and Zhang (2007, 2008), Chan, Zhang, et al. (2020)]. This demonstration is done using perfect model observing system simulation experiments (OSSEs) of a case of tropical convection over the equatorial Indian Ocean. This case occurred during the onset of the active phase of the October 2011 Madden-Julian Oscillation event [MJO; Madden and Julian (1971, 1972), and S. Wang et al. (2015)].

The structure of this paper is as follows. In section 2, we will give an overview of the BGEnKF algorithm, discuss how clear and cloudy members are identified, and modifications made to the CAC20 BGEnKF algorithm. A detailed description of the current BGEnKF, along with suggestions on handling more than two Gaussian kernels, can be found in the supporting information. Following that, we will discuss the setup of our OSSEs in section 3 and the results in section 4. We will then conclude in section 5.

2 On the BGEnKF algorithm

2.1 On the identification of clear and cloudy members

The BGEnKF requires identifying clear and cloudy members at each iteration of the serial data assimilation loop. A simple identification method is to check if the members' column-integrated liquid and/or frozen water mass contents exceed a threshold.

The choice of which phase of water to include in the column integration depends on the specifics of the forecast model. As will be discussed in section 3.3, this study used a WRF model setup with a 9-km horizontal grid spacing and without convective parameterization. This WRF model setup cannot realistically resolve trade cumuli since the typical width of trade cumuli is ∼1-km. As such, we consider columns with trade cumuli and entirely cloud-free columns as clear member columns, and the remaining members as cloudy member columns. Since trade cumuli do not typically grow above the melting layer (Johnson et al., 1999), clear members do not possess frozen water. It thus seems appropriate to use column-integrated ice mass content (ξ) to distinguish between clear and cloudy member columns. To be precise, we compute ξ at a given model column via

$$\xi \equiv \int_0^{z_{top}} \rho(q_i + q_s + q_g)dz \tag{1}$$

where z_{top} is the model top altitude and ρ represents air density. Furthermore, q_i, q_s and q_g are the mass mixing ratios of ice, snow and graupel, respectively.

In this study, we will consider model columns with $\xi \geq 1$ g/m^2 as cloudy, and model columns with $\xi < 1$ g/m^2 as clear. The cloudy and clear infrared window channel simulated brightness temperature statistics (Window-BT; central wavelength of 10.5 μm) do not vary noticeably for model column ξ thresholds between 0.8-1.2 g/m^2. Future studies can refine the threshold value or seek better ways to separate clear and cloudy column members.

2.2 Overview of the BGEnKF algorithm

This study's BGEnKF (and the CAC20 BGEnKF) assimilates observations with Gaussian observation likelihoods under the assumption that clear members are drawn from one Gaussian kernel and cloudy members are drawn from another Gaussian kernel. Suppose we seek to constrain the following extended state vector $\boldsymbol{\psi}$

$$\boldsymbol{\psi} \equiv \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{h}(\boldsymbol{x}) \\ \boldsymbol{\xi}(\boldsymbol{x}) \end{bmatrix} \tag{2}$$

where $\boldsymbol{x}$ represents the model state, $\boldsymbol{h}(\boldsymbol{x})$ represents applying the observation operator $\boldsymbol{h}$ on $\boldsymbol{x}$, and $\boldsymbol{\xi}(\boldsymbol{x})$ represents computing ξ at all observation sites [Eq. (1)]. Note that observation sites here refers to the latitude and longitude of the observation (*i.e.*, the vertical position is not considered for now). Supposing there are N_x elements in $\boldsymbol{x}$ and N_y observations, then $\boldsymbol{\psi}$ has $N_x + 2N_y$ elements.

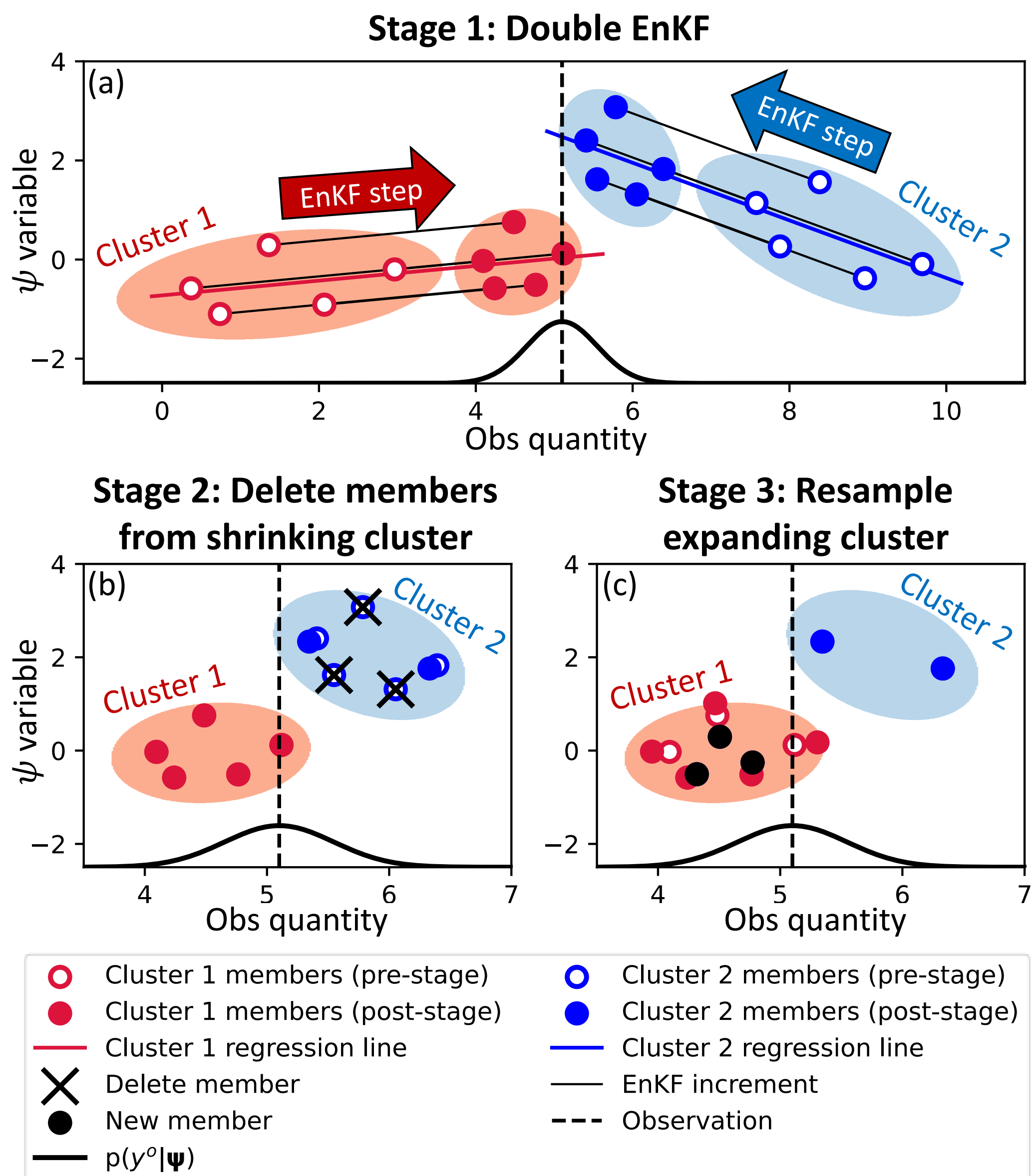


Figure 1. A bivariate demonstration of the three-stage process of the BGEnKF algorithm. The light red ovals highlight cluster 1 members and the light blue ovals highlight cluster 2 members. Prior to running the BGEnKF update, the prior members have already been separated into two clusters. The BGEnKF's first stage is to employ the EnKF update equations on the two clusters separately (panel a). In the second stage (panel b), the BGEnKF identifies the shrinking cluster (the blue cluster 2 in this case), deletes an appropriate number of members from this cluster, and adjusts the remaining members to prevent the deletion from changing this cluster's mean. The BGEnKF's final stage (panel c) is to recreate the deleted members by resampling from the expanding cluster (cluster 1).

The BGEnKF assumes that the prior probability density function [pdf; $p(\boldsymbol{\psi})$] can be represented by the bi-Gaussian pdf

$$p(\boldsymbol{\psi}) = w^f_{\text{clr}}\,\mathcal{G}\left(\boldsymbol{\psi};\ \overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{clr}}}, \boldsymbol{P}^{\boldsymbol{f}}_{\text{clr}}\right) + w^f_{\text{cld}}\,\mathcal{G}\left(\boldsymbol{\psi};\ \overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{cld}}}, \boldsymbol{P}^{\boldsymbol{f}}_{\text{cld}}\right). \tag{3}$$

The subscript "clr" denotes clear cluster quantities, and the subscript "cld" denotes cloudy cluster quantities. $\mathcal{G}\left(\boldsymbol{\psi};\ \overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{clr}}}, \boldsymbol{P}^{\boldsymbol{f}}_{\text{clr}}\right)$ denotes the clear cluster's Gaussian kernel with mean state $\overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{clr}}}$ and covariance matrix $\boldsymbol{P}^{\boldsymbol{f}}_{\text{clr}}$. Similarly, $\mathcal{G}\left(\boldsymbol{\psi};\ \overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{cld}}}, \boldsymbol{P}^{\boldsymbol{f}}_{\text{cld}}\right)$ denotes the cloudy cluster's Gaussian kernel with mean state $\overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{cld}}}$ and covariance matrix $\boldsymbol{P}^{\boldsymbol{f}}_{\text{cld}}$. The scalar quantities w^f_{clr} and w^f_{cld} are the respective weights of the clear and cloudy Gaussian kernels. Note that

$$w^f_{\text{clr}} + w^f_{\text{cld}} = 1, \qquad w^f_{\text{clr}} \geq 0, \quad \text{and,} \quad w^f_{\text{cld}} \geq 0.$$

The various parameters in Eq. (3) can be estimated by the procedure described in CAC20 or in the supporting information.

Upon assimilating an observation y^o with Gaussian observation error, the BGEnKF produces an ensemble that is consistent with the analysis pdf

$$p(\boldsymbol{\psi}|y^o) = w^a_{\text{clr}}\,\mathcal{G}\left(\boldsymbol{\psi};\ \overline{\boldsymbol{\psi}^{\boldsymbol{a}}_{\text{clr}}}, \boldsymbol{P}^{\boldsymbol{a}}_{\text{clr}}\right) + w^a_{\text{cld}}\,\mathcal{G}\left(\boldsymbol{\psi};\ \overline{\boldsymbol{\psi}^{\boldsymbol{a}}_{\text{cld}}}, \boldsymbol{P}^{\boldsymbol{a}}_{\text{cld}}\right). \tag{4}$$

Here, w^a_{clr} and w^a_{cld} are the respective analysis weights of clear and cloudy Gaussian kernels, $\overline{\boldsymbol{\psi}^{\boldsymbol{a}}_{\text{clr}}}$ and $\overline{\boldsymbol{\psi}^{\boldsymbol{a}}_{\text{cld}}}$ are the respective analysis means of the clear and cloudy Gaussian kernels, and $\boldsymbol{P}^{\boldsymbol{a}}_{\text{clr}}$ and $\boldsymbol{P}^{\boldsymbol{a}}_{\text{cld}}$ are the respective analysis covariances of the clear and cloudy Gaussian kernels. See CAC20 [or the supporting information] for the equations relating the analysis pdf's parameters to the forecast pdf's parameters.

The BGEnKF converts a forecast ensemble into an analysis ensemble through a three-stage process [illustrated in Figure 1]. First, two EnKF procedures are executed [Figure 1(a)]: once for clear members using clear forecast statistics $\left(\overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{clr}}}, \boldsymbol{P}^{\boldsymbol{f}}_{\text{clr}}\right)$, and a second time for cloudy members using cloudy forecast statistics $\left(\overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{cld}}}, \boldsymbol{P}^{\boldsymbol{f}}_{\text{cld}}\right)$. Afterwards, to reflect the update to the bi-Gaussian pdf weights, clear members will be replaced with cloudy members, or *vice versa*. For example, if the BGEnKF increased the weight on the clear Gaussian distribution (*i.e.*, $w^a_{\text{clr}} > w^f_{\text{clr}}$ and $w^a_{\text{cld}} < w^f_{\text{cld}}$), some cloudy members will be replaced with clear members. This is achieved by deleting some cloudy members [Figure 1(b)] and replacing the deleted members with resampled clear members [Figure 1(c)]. Once these three stages are completed, the ensemble obeys Eq. (4). See the supporting information for a detailed description of these three stages.

2.3 Revised extended state formulation for better scalable parallelism

The most important modification to the original CAC20 BGEnKF lies in the definition of $\boldsymbol{\psi}$. The CAC20 BGEnKF's $\boldsymbol{\psi}$ only contains $\boldsymbol{x}$ and a single observation. As such, the CAC20 BGEnKF algorithm is a sequential algorithm that scales inefficiently with parallelization on high latency clusters (Anderson & Collins, 2007). For more efficient scaling with parallelization, this study's $\boldsymbol{\psi}$ contains all of the information necessary to assimilate all observations [*i.e.*, Eq. (2); Anderson and Collins (2007)].

Since the definition of $\boldsymbol{\psi}$ has been modified, we will redefine our forecast ensemble. Supposing an ensemble size of N_E, the forecast $\boldsymbol{\psi}$ ensemble is constructed by evaluating

$$\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{n}} \equiv \begin{bmatrix} \boldsymbol{x}^{\boldsymbol{f}}_{\boldsymbol{n}} \\ \boldsymbol{h}(\boldsymbol{x}^{\boldsymbol{f}}_{\boldsymbol{n}}) \\ \boldsymbol{\xi}\left(\boldsymbol{x}^{\boldsymbol{f}}_{\boldsymbol{n}}\right) \end{bmatrix} \quad \forall\ n = 1, 2, \ldots, N_E \tag{5}$$

where $\boldsymbol{\psi}_n^f$ is the $\boldsymbol{\psi}$ of the $n-$th forecast member, and $\boldsymbol{x}_n^f$ is the $\boldsymbol{x}$ of the same forecast member.

The revised formulation enhances the scalability of the BGEnKF by avoiding evaluations of $\boldsymbol{h}(\boldsymbol{x})$ and $\boldsymbol{\xi}(\boldsymbol{x})$ at each iteration of the serial assimilation loop. This is because such evaluations may require costly inter-process communications. The removal of such evaluations is achieved through two modifications to the CAC20 BGEnKF. First, the assimilation of an observation uses the BGEnKF update equations (see CAC20 or the supporting information) to update all model state elements, all simulated observation state elements and all ξ elements in the forecast ensemble. The CAC20 BGEnKF, in contrast, updates all model state elements and only a single simulated observation state element. This difference in updates leads to a second modification: to assimilate the $m-$th observation, instead of evaluating $\boldsymbol{h}(\boldsymbol{x})$ and $\boldsymbol{\xi}(\boldsymbol{x})$, this study's BGEnKF only needs to read the corresponding simulated observation and the ξ values from $\boldsymbol{\psi}$.

2.4 Revised expanding cluster resampling procedure

The other major change to the CAC20 BGEnKF lies in the resampling matrix $\boldsymbol{T}$. $\boldsymbol{T}$ is used to resample the Gaussian kernel that better agrees with the assimilated observation, thus representing the increase in the weight of this kernel. The CAC20 BGEnKF uses a stochastic procedure to construct $\boldsymbol{T}$ [see Eq. (18) and Appendix B of CAC20]. Unfortunately, because random number generators are involved, the analysis ensemble generated on one computing cluster may not be easily replicated on another computing cluster.

To ensure the replicability of the BGEnKF's analysis ensembles, we replaced the stochastic component of the CAC20 BGEnKF's $\boldsymbol{T}$ [$\boldsymbol{W}$ in the Appendix B of Chan, Anderson, and Chen (2020)] with a deterministic one. Supposing that we want to add N_{new} cloudy members to the ensemble to represent an increased weight of the cloudy Gaussian distribution, the new deterministic $\boldsymbol{W}$ is defined as

$$\boldsymbol{W} \equiv \left[\boldsymbol{I}_{\boldsymbol{N}^*_{\text{new}}} \quad \boldsymbol{0}_{\boldsymbol{N}^*_{\text{new}} \times \left(\boldsymbol{N}_{\text{new}} - \boldsymbol{N}^*_{\text{new}}\right)} \right] - \frac{1}{N_{\text{new}}} \boldsymbol{1}_{\boldsymbol{N}^*_{\text{new}} \times \boldsymbol{N}_{\text{new}}} \tag{6}$$

where

$$N^*_{\text{new}} \equiv \begin{cases} N_{\text{new}} - 1 & \forall\ N_{\text{new}} \leq N_{\text{pre}} \\ \\ N_{\text{pre}} & \text{otherwise} \end{cases},$$

and N_{pre} is the number of cloudy members at the start of the resampling procedure. Furthermore, $\boldsymbol{I}_{\boldsymbol{N}^*_{\text{new}}}$ is an $N^*_{\text{new}} \times N^*_{\text{new}}$ identity matrix, $\boldsymbol{0}_{\boldsymbol{N}^*_{\text{new}} \times \left(\boldsymbol{N}_{\text{new}} - \boldsymbol{N}^*_{\text{new}}\right)}$ is an $N^*_{\text{new}} \times \left(N_{\text{new}} - N^*_{\text{new}}\right)$ matrix of zeros, and $\boldsymbol{1}_{\boldsymbol{N}^*_{\text{new}} \times \boldsymbol{N}_{\text{new}}}$ is an $N^*_{\text{new}} \times N_{\text{new}}$ matrix of ones. Note that Eq. (6) is also applied in the situation where N_{new} clear members are being added to the ensemble. A detailed description of the revised resampling procedure is provided in the supporting information.

Note that an interesting property of Eq. (6) is that the resulting $\mathbf{T}$ is a mostly diagonal matrix. Specifically, nearly all of the off-diagonal elements in $\mathbf{T}$ are either zero or much smaller than the diagonal elements (not shown). As a result, the resampled members are essentially copies of the pre-resampling members, plus some small perturbation. The CAC20 stochastic $\mathbf{W}$ formulation does not have this property. Future work can investigate how the BGEnKF's behavior changes with different $\mathbf{W}$ formulations.

2.5 Heuristic measures

2.5.1 Localization

The BGEnKF is likely more susceptible to sampling noise than the EnKF because the sample size used to estimate each cluster's mean state and Kalman gain are smaller than the sample size used to estimate the mean state and covariance matrix of the entire ensemble. As such, we employ two heuristic measures that are similar to those of CAC20. First, we spatially localize the BGEnKF analysis increment using the Gaspari-Cohn fifth order polynomial [GC99; Gaspari and Cohn (1999)]. If $\boldsymbol{\rho}$ represents a vector of GC99 localization factors, we construct the localized updated extended state vector for member n via

$$\boldsymbol{\psi}_n^a \leftarrow \boldsymbol{\rho} \circ \left(\boldsymbol{\psi}_n^a - \boldsymbol{\psi}_n^f \right) + \boldsymbol{\psi}_n^f \tag{7}$$

where $\circ$ represents element-wise multiplication. In the cases where either $w_{\text{clr}}^f = 1$ or $w_{\text{cld}}^f = 1$ (*i.e.,* the bi-Gaussian prior p.d.f. turns Gaussian), this localization method is identical to Kalman gain localization [*e.g.*, Anderson et al. (2009), Meng and Zhang (2008), Whitaker et al. (2008), P. L. Houtekamer and Zhang (2016)].

Note that this localization method [Eq. (7)] localizes the impacts of replacing clear members with cloudy members (or *vice versa*). As an example, suppose the BGEnKF replaces a cloudy forecast member with a clear analysis member. The localization process [Eq. (7)] first computes the difference between the cloudy forecast member and the clear analysis member (*i.e.*, the member's change due to the BGEnKF). This difference is then localized and applied to the cloudy forecast member. The resulting member follows the clear analysis member at the observation site and becomes increasingly like the cloudy forecast member with increasing distance from the observation site. Future work can examine other approaches to localize the impacts of deleting and replacing ensemble members.

2.5.2 Handling overly small clusters

The second heuristic sampling error mitigation measure is to switch from using the BGEnKF to using the EnKF whenever the pre-resampling expanding cluster is too small ($N_{\text{pre}} < 0.8N_E$), or whenever any cluster is too small (less than $0.1N_E$). A similar heuristic measure is used in CAC20.

2.5.3 Mitigating unphysical weight updates

Another issue specific to the BGEnKF is its occasional tendency to generate unphysical weight updates. Specifically, the BGEnKF occasionally expands the clear cluster when a cloudy observation is assimilated, and *vice versa*. This is because the BGEnKF does not explicitly consider whether an observation is clear or cloudy when assimilating it.

The BGEnKF is automatically switched to the EnKF whenever an unphysical weight update is detected. To do so, we first identify the whether the observation to be assimilated is definitively clear or cloudy. In the case of Window-BT values over tropical ocean, observation values warmer than 290 K are definitively clear, and observation values cooler than 280 K are definitively cloudy. If the observation is definitively clear, but the cloudy cluster is expanded by the BGEnKF, or *vice versa*, the BGEnKF will switch over to the EnKF.

3 Materials and methods

3.1 Description of October 2011 tropical convection case

The BGEnKF was tested against the EnKF using a case of tropical convection over the equatorial Indian Ocean during the October 2011 MJO. This case is chosen because it can be reasonably replicated by regional WRF models (S. Wang et al., 2015; F. Zhang et al., 2017; Ying & Zhang, 2017; Fu et al., 2017; X. Chen, Pauluis, & Zhang, 2018; X. Chen & Zhang, 2019; Ying & Zhang, 2018; Chan, Zhang, et al., 2020).

Our experiments are conducted over a three day period during the onset of this MJO event (15 October 2011 to 18 October 2011). Two persistent regions of enhanced convection (henceforth, "convective regions") are observed in the 4-km Global IR Dataset of Janowiak et al. (2001) [henceforth, the MERG dataset]. The first convective region (blue rectangle) occurs between 60 °E and 75 °E and persists beyond the three-day period. Westward propa-

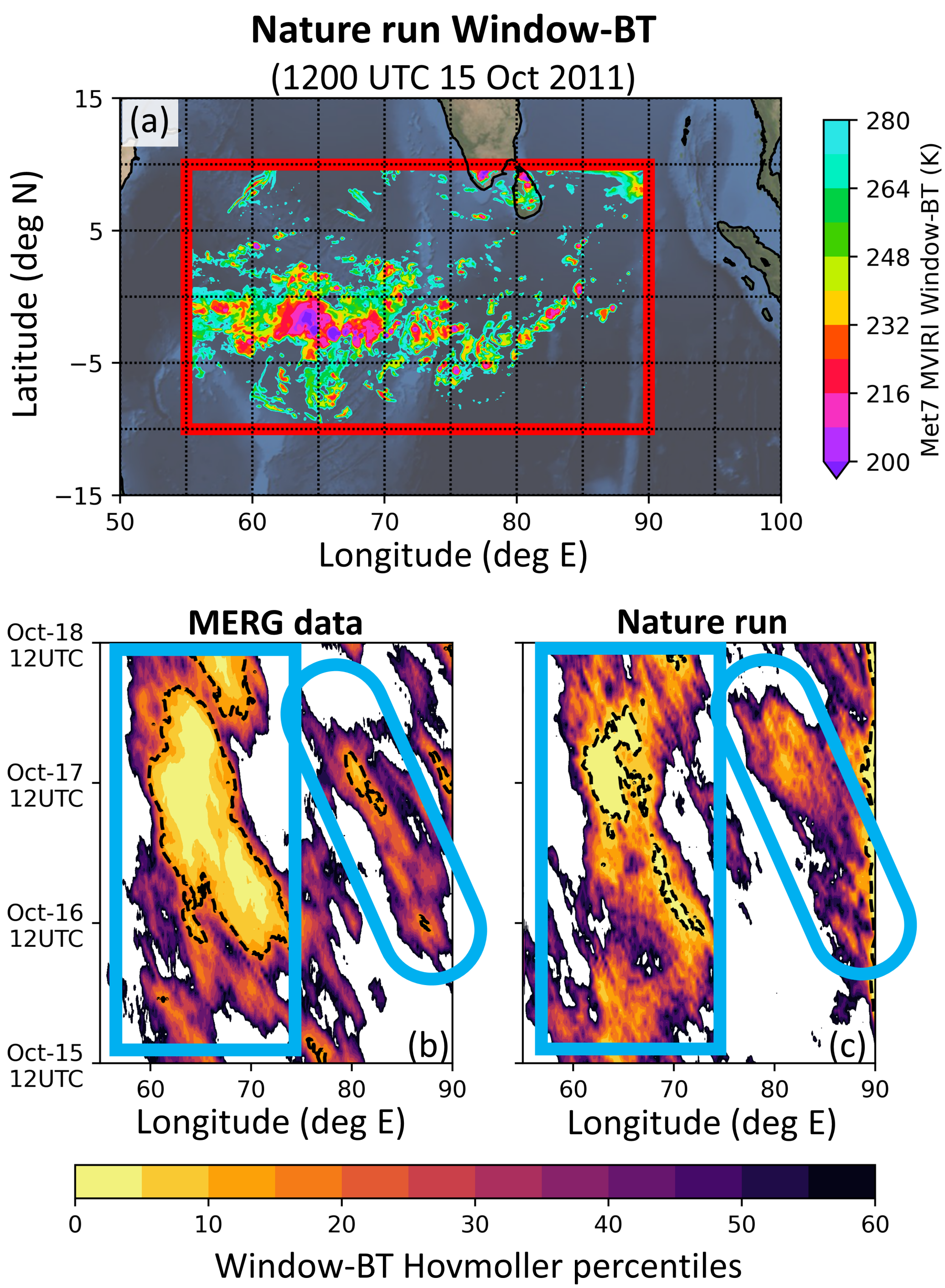


Figure 2. (a) Plot of our OSSE domain overlaid with the nature run's simulated Window-BT field at 1200 UTC on 15th October 2011. The red box in panel (a) indicates our study domain. Also shown are longitude-time diagrams for the MERG dataset (b) and nature run (c). In panels (b) and (c), the shadings indicate Window-BT Hovmoller percentile values. These Window-BT Hovmoller percentile values are constructed by first averaging Window-BT values between between 10°S and 10°N at every hour to produce a time-longitude array of latitudinally-averaged Window-BT values. These arrays are then converted into percentiles before producing the longitude-time percentile values. Note that the dashed black contours in (b) and (c) indicate areas where the time-longitude arrays of latitudinally-averaged Window-BT values are below 260 K.

gation is observed in some of the clouds in this region, most notably between 1200 UTC on 16 October and 0000 UTC on 18 October. The second convective region (blue oval) appears on the eastern edge of the study domain at 1200 UTC on 16th October and exhibits a westward propagation that is similar to that of the first system. We will later assess our OSSE's nature run simulation by checking the nature run against these two convective regions.

3.2 Setup of WRF model

The Advanced Research version of the WRF model (WRF-ARW) version 3.8 (Skamarock et al., 2008) is used in this study. Following Chan, Zhang, et al. (2020), we construct a 432×243 WRF domain over the study domain [red box in Figure 2(a)] with 9-km horizontal grid spacing and 45 model levels. The bottommost 9 levels are within the lowest 1-km of the atmosphere and the pressure level at the top of the domain is set to 20 hPa. The WRF integration time step is set to 20 seconds.

Our WRF model setup uses the following parameterization schemes. Cloud microphysical processes are handled by the WRF double-moment 6-class scheme (WDM6) proposed by Lim and Hong (2010). The updated Goddard shortwave scheme of Chou and Suarez (1999) and the Rapid Radiative Transfer Model (Global Circulation Model version; RRMTG) longwave scheme of Iacono et al. (2008) are used to parameterize radiative processes. The unified Noah land surface physics scheme (F. Chen & Dudhia, 2001) handles surface process and the Yonsei University (YSU) boundary layer scheme (Hong et al., 2006) is employed. No cumulus parameterization is employed because many studies have demonstrated that the 9-km grid spacing is sufficient to resolve tropical mesoscale convective systems (MCS) over the region (S. Wang et al., 2015; Ying & Zhang, 2017, 2018; F. Zhang et al., 2017; X. Chen, Pauluis, & Zhang, 2018; X. Chen, Pauluis, Leung, & Zhang, 2018; X. Chen & Zhang, 2019; X. Chen et al., 2020; Chan, Zhang, et al., 2020; Chan & Chen, 2021; X. Chen, Leung, Feng, & Song, 2021; X. Chen, Leung, Feng, Song, & Yang, 2021; X. Chen et al., 2022).

3.3 Setup of WRF ensemble and nature run

This study's WRF ensemble and nature run are constructed by combining two datasets from the European Center for Medium-Range Forecasts (ECMWF): the ECMWF Reanalysis Version 5 [ERA5; Hersbach et al. (2020)] and the ECMWF's 50-member perturbed forecasts (Swinbank et al., 2016). The ERA5 dataset is downloaded for every hour between 0000 UTC on 15 October to 1800 UTC on 18 October from the ECMWF's Climate Data Store (CDS). The ECMWF's perturbed forecasts are produced as part of The Observing System Research and Predictability Experiment (THORPEX) Interactive Grand Global Ensemble [TIGGE; Swinbank et al. (2016)] and is downloaded for 0000 UTC on 15 October from the ECMWF's Meteorological Archival and Retrieval System (MARS).

The ERA5 and ECMWF's 50-member perturbed forecasts (TIGGE ensemble) are processed using the WRF Preprocessing System and WRF's real data processor (`real.exe`) to produce a set of 51 WRF initial conditions files. Note that the ERA5 is used to fill in the data missing from the TIGGE ensemble above 200 hPa. The 50 WRF initial conditions from the TIGGE ensemble are then recentered on the ERA5 WRF initial condition file. The end result is a 51-member ensemble of WRF initial conditions, where member 51 is based entirely on the ERA5 (*i.e.*, the 51-st ensemble perturbation is zero). Note that this 51-st member is not used to initialize the nature run. One of the other initial conditions is used to initialize the nature run.

The lower and lateral boundary conditions used in this study are based entirely on the hourly ERA5 dataset (*i.e.*, the boundary conditions are unperturbed). While perturbed boundary conditions can increase the ensemble spread, the ensemble spread is usually reasonable even with unperturbed boundary conditions (not shown). Furthermore, as a first approach to studying the potential impacts of the BGEnKF in a high-order weather model setting, we want the differences between the nature run (described later) and the OSSE ensemble to be entirely due to differences in the initial conditions. Future work can extend this study to situations with perturbed boundary conditions.

We desire a nature run that is roughly one ensemble standard deviation from our experiments' ensembles. To select an appropriate initial condition file for such a nature run, we first integrate the 51 members forward for 12 hours (from 0000 UTC to 1200 UTC on 15 October 2011). This integration is performed to generate flow-dependent ensemble statistics that are consistent with the WRF model. After the 12-hour integration, we compute the following perturbation length metric (D^2) for each of the 51 ensemble members

$$D^2(n) \equiv \frac{1}{N_S\ N_i\ N_j} \sum_{v \in S} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \left(\frac{\Lambda\left(i,j,v,n\right) - \left\langle \Lambda\left(i,j,v\right) \right\rangle_n}{\sigma_{i,j,v}} \right)^2 . \tag{8}$$

$\Lambda\left(i,j,v,n\right)$ here is the value of a WRF-derived field v at horizontal index location (i,j) for ensemble member n. Furthermore, $\left\langle \Lambda\left(i,j,v\right) \right\rangle_n$ is the ensemble average of $\Lambda\left(i,j,v,n\right)$, and $\sigma_{i,j,v}$ is the ensemble standard deviation of $\Lambda\left(i,j,v,n\right)$. This means that the expression in the parentheses of Eq. (8) is the spread-normalized displacement of ensemble member n from the ensemble mean at location (i,j) for variable field v. The set S contains three 2D variables (precipitable water, column mass, and mass-integrated kinetic energy) and N_s is the size of the set S (*i.e.,* $N_S = 3$). Furthermore, N_i ($\equiv$ 432) is the number of east-west grid points and N_j ($\equiv$ 243) is the number of north-south grid points. The metric in Eq. (8) can thus be interpreted as being proportional to the spread-normalized Euclidean length of the n-th ensemble perturbation. As such, a D^2 value of unity means that the ensemble member is generally displaced from the ensemble mean by 1 standard deviation.

We define our nature run member to be the member whose D^2 value is closest to unity at 1200 UTC on 15 October. As a result, the nature run is based on member 10 of the TIGGE ensemble. The remaining 50 WRF members will be used for our cycling OSSE DA experiments.

3.4 Sanity check of nature run

Before proceeding, the nature run is checked by comparing it against the MERG dataset. Figure 2(b & c) shows longitude-time diagrams of the Window-BT percentiles from the MERG dataset and our nature run. The construction of these percentiles is explained in section 3.1 and in the caption of Figure 2.

We have opted to display the Window-BT percentiles instead of the Window-BT values because the WRF model tends to under produce clouds (*i.e.*, when compared to satellite observations, the nature run Window-BTs are warm biased). This is illustrated by the dashed contours in Figure 2(b & c), which highlights areas where the latitudinally-averaged values of Window-BT were cooler than 260 K. These areas are substantially larger in the MERG data than in the nature run, meaning that the nature run under produced clouds. Since converting the Window-BT values to percentile values weakens the visual interference from the cloud biases, we have opted to display the Window-BT percentiles over the Window-BT values.

Figure 2(c) indicates that the nature run also exhibits the two persistent convective regions observed in the MERG dataset (see section 3.1). These persistent convective regions are indicated by the blue rectangle and blue oval in Figure 2(c). Not only did the nature run's two persistent convective regions occur in locations and times similar to those of the MERG dataset (Figure 2(b)), these nature run regions also exhibit westward propagation patterns similar to those of the MERG dataset. As such, the nature run simulation reasonably replicates the anomalous convective behavior of the real atmosphere between 15 October to 18 October 2011.

3.5 Setup of DA experiments to test the BGEnKF

To test the BGEnKF, three 50-member ensemble experiments are conducted. All three experiments start at 1200 UTC on 15 October and terminate at 1200 UTC on 18 October, with hourly DA cycling (73 cycles in total). The construction and spin-up of these 50 members are described in section 3.3.

In the first experiment, no observations are assimilated (henceforth, NoDA experiment). The NoDA experiment serves as a baseline for comparing the performance of the EnKF and BGEnKF, and to measure imbalances induced by DA.

The other two experiments are the EnKF and BGEnKF experiments. The only difference between the EnKF and BGEnKF experiments is in the DA algorithm employed. The EnKF experiment will assimilate observations using the PSU-EnKF's (Meng & Zhang, 2007, 2008) default EnKF algorithm, and the BGEnKF experiment will assimilate observations using a new implementation of the BGEnKF into the PSU-EnKF. Note that both the EnKF and the BGEnKF are implemented into the PSU-EnKF using the high-latency strategy proposed by Anderson and Collins (2007).

As a first approach to testing the BGEnKF, only synthetic *Meteorological Satellite 7* Meteosat Visible Infra-Red Imager (MVIRI) Window-BT observations will be assimilated. Future work can investigate if our findings can be extended to situations where an entire suite of operationally-assimilated observations and observations from different infrared channels are assimilated.

The synthetic Window-BT observations are constructed by first running the Community Radiative Transfer Model (CRTM) release 2.3.0 on the nature run (see sections 3.3 and 3.4). The nature run's Window-BT values are then thinned to a horizontal spacing of 27-km (~11,500 observations per DA cycle). White noise with a standard deviation of 3 K is then added to the thinned nature run Window-BT values to simulate instrument noise, thus constructing the synthetic observations. Note that the observation errors are likely to be correlated in reality. This means our use of white noise is an imperfect approximation to actual observation errors. Future work can investigate if our results can be extended to situations with correlated Window-BT observation errors.

Common heuristic strategies are employed to assimilate the Window-BT observations. To limit the impact of sampling errors, horizontal localization is applied using the Gaspari-Cohn fifth-order polynomial (Gaspari & Cohn, 1999) with a 100-km radius of influence (P. L. Houtekamer & Mitchell, 2001; Greybush et al., 2011; P. L. Houtekamer & Zhang, 2016). No vertical localization is employed. We also employ the Adaptive Observation Error Inflation scheme (AOEI) of Minamide and Zhang (2017) to limit the deleterious increments that can result from clear-cloudy disagreements between the prior and observations (F. Zhang et al., 2016; Minamide & Zhang, 2017). To mitigate the tendency for ensemble under-dispersion to occur when the ensemble is clear and the observation is cloudy, the Adaptive Background Error Inflation scheme (ABEI) of Minamide and Zhang (2019) is applied. We also employ 80% relaxation to prior perturbations (RTPP) to maintain ensemble dispersion (F. Zhang et al., 2004). Similar combinations of heuristic strategies are commonly seen in the EnKF-based DA of infrared radiance observations (F. Zhang et al., 2016; Minamide & Zhang, 2018; Chan, Zhang, et al., 2020; Y. Zhang et al., 2019; Chan & Chen, 2021; Y. Zhang et al., 2021).

Aside from these common strategies, we also restrict the BGEnKF/EnKF from updating the domain-averaged specific humidity (QVAPOR) using Window-BT observations. Without this measure, both the BGEnKF and the EnKF experience filter divergence that is related to DA-induced dry biases within 48 hours of cycling. These dry biases are likely induced by the ensemble's tendency to be overly cloudy. The dry biases in the EnKF experiment are likely partly because of the EnKF's inability to handle clear and cloudy members separately (see section 4.3). As for the BGEnKF experiment, the dry bias can be explained by the fact that the BGEnKF algorithm frequently switches over to the EnKF algorithm (see section 4.2). Note that the BGEnKF generated smaller dry biases than the EnKF (not shown).

To prevent filter divergence due to DA-induced dry biases, we replace the 3D posterior mean QVAPOR field ($\overline{q_v^a}$) with the following modified mean QVAPOR field ($\overline{q_v^*}$):

$$\overline{q_v^*(i,j,k)} \equiv \overline{q^a(i,j,k)} - \frac{1}{N_i\, N_j}\sum_{i=1}^{N_i}\sum_{j=1}^{N_j}\left\{\overline{q_v^a(i,j,k)} - \overline{q_v^f(i,j,k)}\right\}. \tag{9}$$

Here, (i, j, k) refer to the west-east, south-north and bottom-top indices of the 3D QVAPOR fields and $\overline{q_v^f}$ refers to the 3D prior mean QVAPOR field.

3.6 Execution wall-time of the BGEnKF

Before proceeding, we should compare the execution wall-time of the BGEnKF and the EnKF. The BGEnKF algorithm took ~30 seconds to assimilate ~11,500 observations using 228 Intel Knight's Landing computer cores [distributed across 7 computational nodes on the National Energy Research Scientific Computing Center (NERSC) Cori supercomputer; each core has a clock rate of 1.4 GHz]. Assimilating the same observations via an EnKF algorithm took ~20 seconds of wall-time. For a fair comparison, this EnKF algorithm used the exact same code structure and computing resources, but with the cluster transfer and auxiliary variable update steps disabled. In other words, the BGEnKF used ~10 seconds more wall-time than the EnKF.

This ~10-second difference should be assessed in the context of the wall-time for the entire PSU-EnKF executable. The other components of the PSU-EnKF took ~100 seconds to execute. As such, the BGEnKF only added ~10% wall-time to the entire PSU-EnKF executable. The BGEnKF algorithm is thus likely affordable for research and operational groups that are already running serially-assimilating EnKFs [*e.g.*, Anderson et al. (2009)].

4 Perfect model WRF OSSE results

In the discussions to follow, we will be showing plots of normalized root-mean-square errors (nRMSEs) and normalized biases as functions of time and model level. The normalization is necessary for the ease of visualization, and uses the root-mean-square errors (RMSEs) of the NoDA experiment. The EnKF experiment's nRMSE at model level k and date t is defined as

$$\text{EnKF nRMSE}\,(k,t) \equiv \frac{\text{EnKF RMSE}\,(k,t)}{\text{NoDA RMSE}\,(k,t)} \tag{10}$$

and likewise for that of the BGEnKF and NoDA experiments (the NoDA's nRMSE values are always 1). Note that if a filter results in nRMSEs > 1.0, the assimilation of Window-BT via this filter degraded the ensemble with respect to the NoDA experiment. The reverse is true for nRMSEs < 1.0. We also define the normalized bias of the EnKF experiment to be

$$\text{EnKF normalized bias}\,(k,t) \equiv \frac{\text{EnKF bias}\,(k,t)}{\text{NoDA RMSE}\,(k,t)}, \tag{11}$$

and likewise for the BGEnKF and NoDA experiments. These biases are computed by subtracting the nature run fields from the forecast ensemble mean fields.

The nRMSEs and normalized biases are examined for six variable fields: the zonal wind velocity component field (U), the meridional wind velocity component field (V), the temperature field (T), the QVAPOR field (Q), the Window-BT field, and the upper tropospheric infrared water vapor channel brightness temperature field (WV-BT; central wavelength of 6.2 μm). The nRMSEs are plotted in Figures 3 and 5(a & b) and the normalized biases are plotted in Figures 4 and 5(c & d). All quantities are computed using forecast statistics.

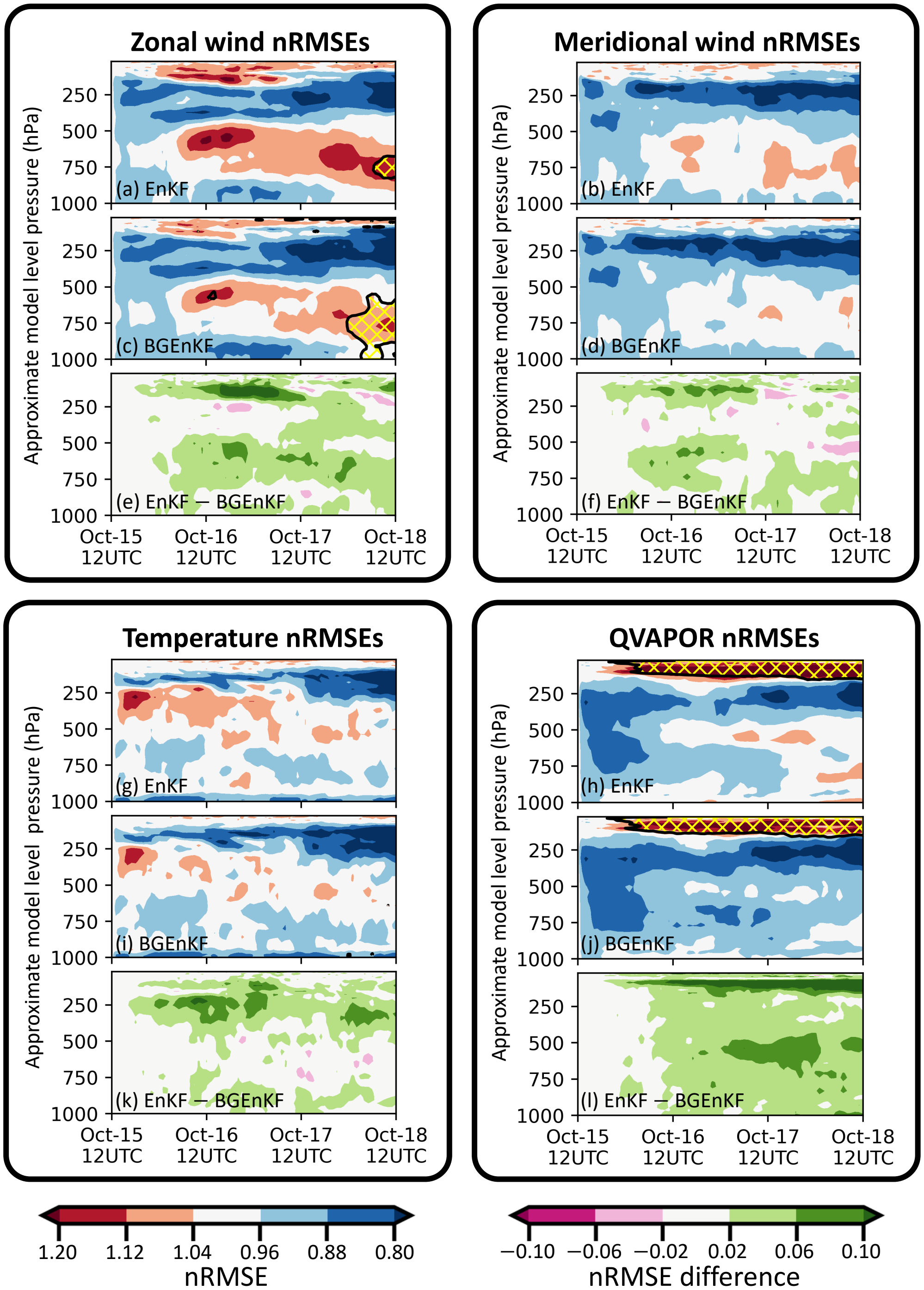


Figure 3. Plots of various prior ensemble statistics as functions of time and model level. For ease of interpretation, the model levels are displayed in terms of their approximate pressure levels (estimated using the definition of eta levels in WRF and assuming a surface pressure of 1000 hPa). The shadings indicate the NoDA-normalized RMSEs [nRMSEs; defined in Eq. (10)] for the EnKF (a, b, g & h) and BGEnKF (c, d, i & j) experiments, as well as the nRMSE differences between the EnKF and BGEnKF experiments (e, f, k & l). The nRMSEs and nRMSE differences are shown for the U field (a, c & e), V field (b, d & f), T field (g, i & k), and Q field (h, j & l). The areas outlined with a black contour and filled with yellow hatching have consistency ratios (spread/error) less than 0.75. Note that all displayed statistics are forecast statistics.

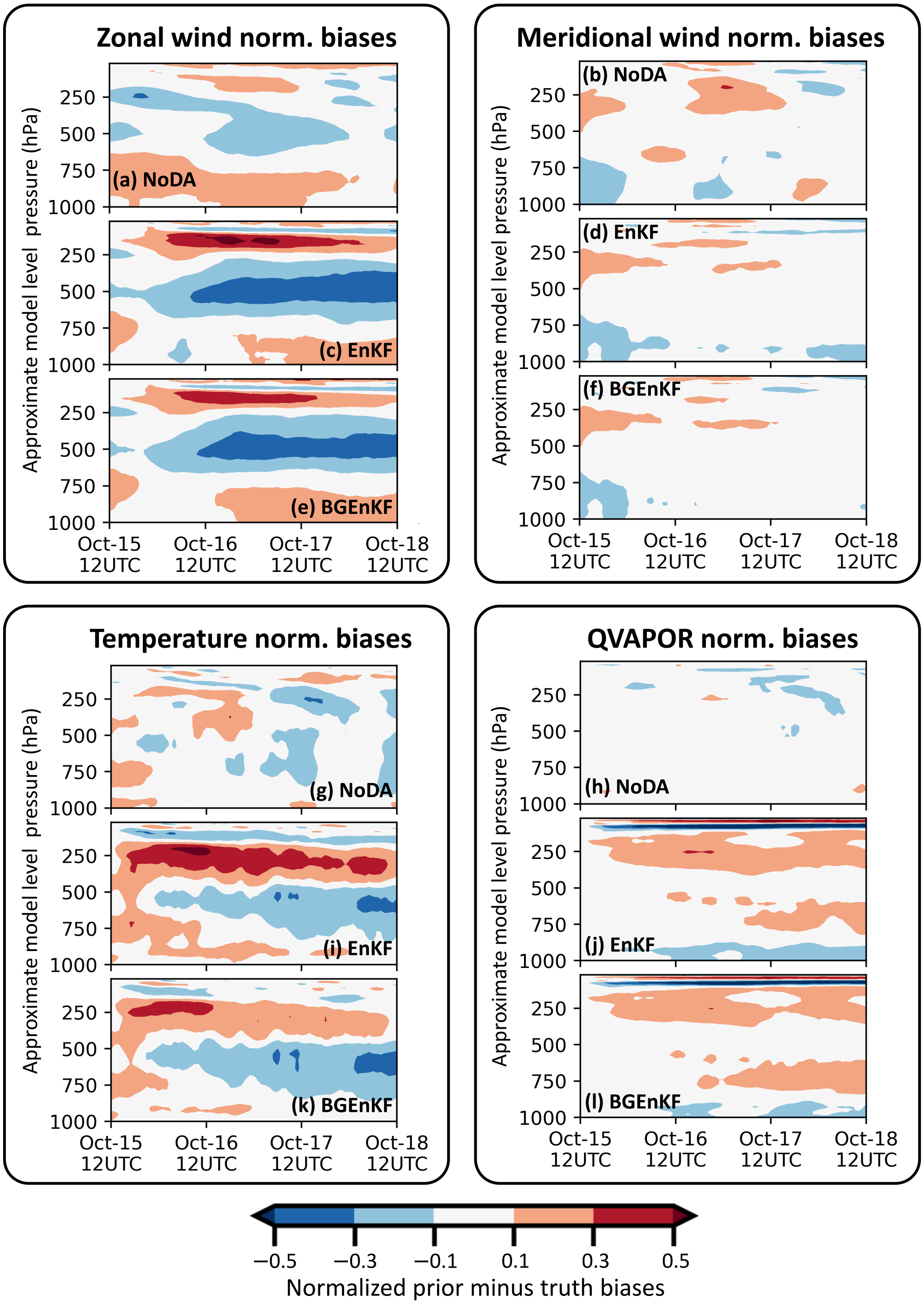


Figure 4. Plots of various prior ensemble normalized biases as functions of time and model level. These normalized biases are displayed for the U field (a, c & e), V field (b, d & f), T field (g, i & k), and Q field (h, j & l), for the NoDA (a, b, g & h), EnKF (c, d, i & j) and BGEnKF (e, f, k & l) experiments. Similar to Figure 3, the model levels are displayed in terms of approximate pressure levels. See the Eq. (11) for the definition of the normalized biases.

4.1 On differences in the BGEnKF's and the EnKF's performances during DA cycling

The nRMSEs and normalized biases of the BGEnKF experiment are generally better than or comparable to those of the EnKF experiment (Figures 3 to 5). For the U, V, T and Q fields, subtracting the BGEnKF's nRMSEs from the EnKF's nRMSEs generally results in positive values [Figure 3(e, f, k & l)]. The BGEnKF experiment also has better WV-BT nRMSEs than the EnKF experiment [Figure 5(b)]. The BGEnKF experiment also has smaller biases than the EnKF experiment in several places: the 100 hPa U field [Figure 4(c & e)], the 400–100 hPa T field [Figure 4(i & k)], the Window-BT field [Figure 5(e)], and WV-BT field [Figure 5(f)]. Otherwise, the BGEnKF and EnKF experiments have similar bias values. These results suggest that the BGEnKF is more suitable for assimilating all-sky Window-BT than the EnKF.

The BGEnKF's performance advantages over the EnKF can be separated into two types. In the first type, the BGEnKF generates larger improvements than the EnKF (*i.e.*, BGEnKF nRMSEs < EnKF nRMSEs < NoDA nRMSEs). This type of performance advantage occurs in multiple places (Figures 3 and 5): 1) the 800 hPa to 1000 hPa U field nRMSEs during the first 56 cycles, 2) the 100 hPa to 500 hPa U field nRMSEs during the last 36 DA cycles, 3) the near surface and ~250 hPa V field nRMSEs from 0000 UTC on 16th October to 0000 UTC on 17th October, 4) between 100 hPa to 300 hPa in the T field nRMSEs for most cycles, 5) between 250 to 600 hPa in the Q field nRMSEs for most cycles, and in the WV-BT nRMSEs for most DA cycles after 0000 UTC on 16th October. These differences are likely due to the BGEnKF's ability to handle mixture statistics, and suggest that the BGEnKF is more suitable for assimilating Window-BT than the EnKF.

The BGEnKF experiment's second type of performance advantage over the EnKF experiment is when the BGEnKF introduces milder degradations than the EnKF (*i.e.*, NoDA nRMSEs < BGEnKF nRMSEs < EnKF nRMSEs). In terms of nRMSEs (Figure 3), such situations are noticeable at the 100 hPa tropopause level and 500–700 hPa levels for the U and V fields, at the 200–500 hPa model levels for the T field, and at the 100 hPa level for the Q field. Such situations are also noticeable in the normalized biases of the ~100 hPa U field, the 100–400 hPa T field (Figure 4), and in the Window-BT and WV-BT fields (Figure 5). These are likely because 1) the BGEnKF can handle mixture statistics whereas the EnKF cannot, and 2) the BGEnKF experiment has smaller increments than the EnKF experiment because the BGEnKF experiment has smaller dispersion. Figure 3(a & c) shows an example of the latter: the BGEnKF U field has larger areas of low spread-to-error ratios (0.75) than the EnKF. The likely origins of the RMSE and bias degradations are discussed in section 4.2. Nonetheless, these results further support the notion that the BGEnKF is more appropriate for assimilating Window-BT observations than the EnKF.

The BGEnKF tends to result in smaller CRs than the EnKF because the BGEnKF can outright convert all clear member columns to cloudy member columns, or *vice versa*. Since clear and cloudy member columns are very different, having both types of columns present at the same time boosts the ensemble spread. If all clear member columns are converted to cloudy member columns, or *vice versa*, large perturbations relative to the ensemble mean are replaced with smaller perturbations. This replacement results in reduced ensemble dispersion. Since the EnKF lacks this mechanism of ensemble spread removal, the BGEnKF can remove more ensemble spread than the EnKF, thus resulting in smaller CRs than the EnKF. Future work can investigate if stronger inflation schemes are more appropriate for the BGEnKF.

Note that there are occasional situations where the EnKF outperforms the BGEnKF. For instance, at around 0000 UTC on 17th October the BGEnKF's U nRMSEs are slightly higher than the EnKF at 250 hPa (Figure 3(e)). Other examples include the T nRMSEs around 1200 UTC on 17th October (Figure 3). Nonetheless, if we integrate the forecast ensembles'

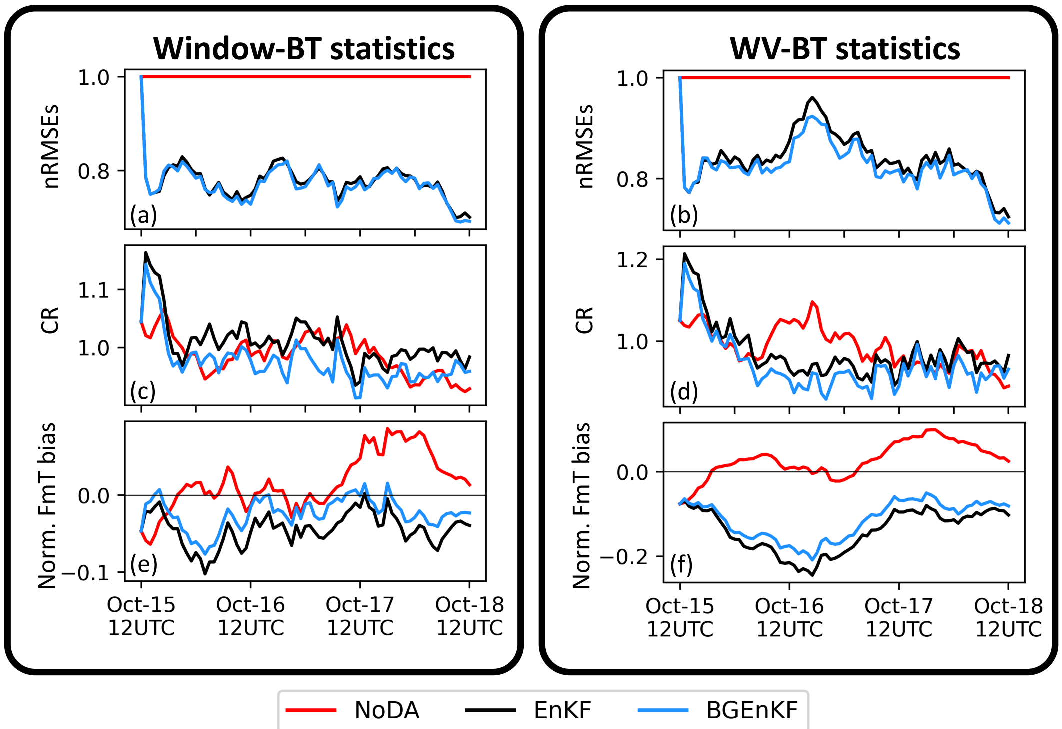


Figure 5. Time-series showing the performance statistics of the three experiments' prior ensembles in terms of Window-BT (a, c & e) and WV-BT (b, d & f). The definitions of nRMSEs (a & b) and normalized prior minus truth (Norm. FmT bias; e & f) are the same as in Figures 5 to 8. Like Figures 5 and 6, the consistency ratio (CR; c & d) here is defined as the ratio of spread to error.

nRMSEs with respect to pressure at every cycle, the resulting mass-weighted nRMSEs of the BGEnKF experiment will be lower than those of the EnKF experiment.

We have also examined day-long deterministic forecasts that are initialized from the analysis means of the EnKF and BGEnKF experiments (not shown). The BGEnKF experiment's RMSE performance advantage over the EnKF experiment persists for up to 9 hours of lead time in terms of the U, V and T fields. In terms of the 500–800 hPa Q field RMSEs, the BGEnKF experiment's RMSE advantage over the EnKF experiment persists throughout the 24 hours of integration. These results are as expected since the BGEnKF experiment has lower RMSEs than the EnKF experiment during DA cycling.

4.2 On the similar patterns observed in the performances of the BGEnKF and EnKF experiments

Though the BGEnKF experiment generally outperformed the EnKF experiment, there are common spatiotemporal patterns in their nRMSEs and normalized biases. For instance, Window-BT DA with either algorithm tends to degrade the 500–800 hPa U nRMSEs, and improve the 100–500 hPa U nRMSEs (Figure 3(a & c)). These similarities are likely because the BGEnKF frequently switches over to the EnKF. Figure 6(a) shows that the BGEnKF algorithm is only called to assimilate ∼10% of the Window-BT observations, meaning that the switching occurred for the remaining ∼90% of Window-BT observations. Future work can investigate if reducing the occurrence of such switches (*e.g.*, via weaker heuristic checks and larger ensembles) could improve the performance of the BGEnKF.

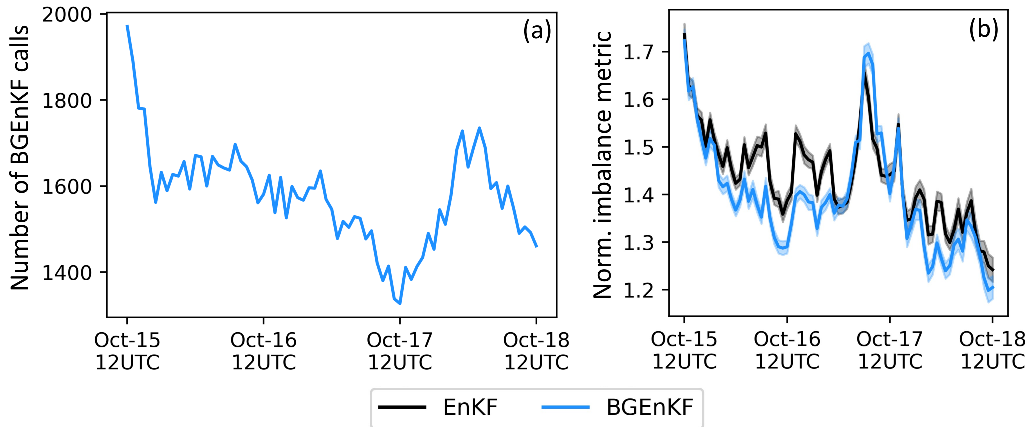


Figure 6. Plots showing the frequencies at which the two kernel BGEnKF update procedure is called in the BGEnKF experiment (a), and the normalized imbalance metric statistics for both the BGEnKF and EnKF experiments (b). For reference, 11502 IR observations are assimilated at each DA cycle. The normalized imbalance metric is defined in the text. The solid curves in (b) indicate the ensemble average of every member's normalized imbalance metric and the half-width of the shadings in (b) indicate twice the standard error of the members normalized imbalance metric.

It is notable that the BGEnKF outperforms the EnKF despite the high frequency of BGEnKF-to-EnKF switching. For instance, according to Figure 3(h, j & l), for the 24 cycles on 17th October and between 500 hPa to 700 hPa, the BGEnKF experiment has 0.06–0.1 less Q nRMSEs than the EnKF experiment. Since the EnKF experiment has Q nRMSEs of ∼1 then, the BGEnKF is able to introduce a ∼6–10% improvement over the EnKF. These are considerable improvements since the BGEnKF is only called on ∼10% of the Window-BT observations.

Given the frequent switching from the BGEnKF to the EnKF, the worse-than-NoDA RMSEs and biases in both the EnKF and BGEnKF experiments are likely caused by the EnKF algorithm. These degradations are likely caused by 1) non-Gaussian forecast statistics, 2) sampling errors, and 3) biases that are introduced by the assimilation of Window-BT. The first factor can originate from having mixtures of clear and cloudy members. Sampling errors can also introduce errors into the analysis, particularly over regions where the ensemble correlations are weak. This factor is likely present in our experiments because no vertical localization is used in this study. Future work can investigate if vertical localization can mitigate some of the RMSE and bias degradations (Lei & Anderson, 2014; Lei & Whitaker, 2015; Lei et al., 2016, 2020). Finally, since biases are a component of RMSEs [*e.g.*, Ying and Zhang (2017), Ying and Zhang (2018), and Chan, Zhang, et al. (2020)], biases that are introduced by Window-BT DA can contribute towards worse-than-NoDA RMSEs. While the contribution of biases to worse-than-NoDA RMSEs can be easily inferred (see the next paragraphs), the contributions from the first two factors cannot be easily teased apart.

To understand the contribution of biases to the occurrence of worse-than-NoDA RMSEs (*i.e.*, nRMSEs > 1), we computed the following fraction as a function of model level and time ($f_{\text{bias}}(k,t)$). For the EnKF experiment, we defined

$$\text{EnKF's } f_{\text{bias}}(k,t) \equiv \sqrt{\frac{[\text{EnKF's biases }(k,t)]^2 - [\text{NoDA's biases }(k,t)]^2}{[\text{EnKF's RMSEs }(k,t)]^2 - [\text{NoDA's RMSEs }(k,t)]^2}}$$

and likewise for the BGEnKF experiment. f_{bias} can be interpreted as the fractional contribution of biases to the worse-than-NoDA RMSE performance.

We found that for about 25–45% of the worse-than-NoDA situations (nRMSEs > 1) in the U and T fields, the majority of the nRMSE degradation (*i.e.*, $f_{\text{bias}} \geq 0.6$) can be explained by the the introduction of biases [*i.e.*, $p\ (f_{\text{bias}} > 0.6\ |\text{nRMSE} > 1) \ \in \ (0.25, 0.45)$]. This suggests that though DA-induced biases are important contributors towards the worse-than-NoDA RMSEs of either DA filters, the net contribution coming from other factors is also important. Future work can examine separating and quantifying the relative importance of these three factors towards the worse-than-NoDA RMSEs.

4.3 On the origin of biases in the EnKF and BGEnKF experiments

We now turn our attention to the U, T, Q, Window-BT and WV-BT biases that are introduced by Window-BT DA. Since the Q analysis increments are subject to bias removal (see last paragraph of section 3.5), the Q biases will be discussed later. The U, T and WV-BT biases are likely related to 1) a cold forecast minus truth (FmT) Window-BT bias at the start of all experiments, and 2) the persistence of these FmT Window-BT biases throughout all cycles (Figure 5(e)). Item 1 is essentially the result of drawing a single member from an ensemble – it is difficult to obtain a nature run whose domain-averaged Window-BT is always the same as that of the forecast ensemble. This is supported by the fact that the NoDA experiment's FmT Window-BT biases oscillate around zero (Figure 5(e)). More interestingly, item 2 indicates an over abundance of clouds in both DA experiments. Since WV-BT is cooler in the presence of clouds, the WV-BT bias is explained by the over abundance of clouds.

To understand the origin of the persistently cold FmT Window-BT biases, we examine the analysis ensembles' Window-BT biases. Running the CRTM on the analysis ensembles of the Window-BT DA experiments reveals analysis minus truth (AmT) Window-BT normalized biases that are typically around −0.25 (not shown). These bias values are a factor of 5 larger than the FmT normalized biases of around −0.05 (Figure 5(e)). The large AmT biases suggest that Window-BT DA resulted in overly cloudy analysis ensembles. Though the time-integration of these analysis ensembles dramatically reduces the over cloudiness (the normalized biases typically go from −0.25 to −0.05), some over cloudiness likely remain. As such, the U, T, Window-BT and WV-BT biases are likely caused by the EnKF and BGEnKF experiments introducing too many clouds into the analysis ensemble.

The over introduction of clouds is likely a result of the EnKF's inability to handle clear and cloudy members separately and the strong sensitivity of Window-BTs to hydrometeors. When both clear and cloudy members are present in the forecast ensemble, the EnKF's forecast mean state will contain some amount of clouds. Suppose that the correlations between Window-BT and hydrometeor mixing ratios are negative. If Window-BT observations with either small or negative innovations are assimilated, the clouds in the EnKF's mean state will either be unaffected (for small innovations) or be increased (for negative innovations). Since the EnKF will also reduce the size of the ensemble members' perturbations, the ensemble thus contracts around a cloudy mean state. The result is that clear column forecast members gain some amount of clouds, even in situations where the innovations are close to zero. Since Window-BTs are sensitive to the presence of clouds, running the CRTM on such members will generate cold cloudy Window-BT values. This mechanism of EnKF-induced over-cloudiness warrants future investigation.

Note that the BGEnKF experiment's over-cloudiness is likely caused by the mechanism in the previous paragraph. This is because the BGEnKF algorithm frequently switches over to the EnKF (for ~90% of assimilated observations). Since the BGEnKF can handle mixtures of clear and cloudy members, with less frequent switches, the BGEnKF is likely to have smaller biases. To test this possibility, smaller sampling errors are necessary to justify less frequent switches from the BGEnKF to the EnKF. Future work can thus investigate this possibility with larger ensembles.

With regards to the Q biases, since the analysis increment cannot modify the Q biases [see Eq. (9)], these biases are induced during the forecast step of the DA procedure. We can rule out the evaporation of DA-induced spurious clouds as an important source because the hydrometeor biases injected by the increment are an order of magnitude smaller than the Q bias growth during integration (not shown). Other processes are likely causing the Q biases. Some possibilities include enhancements to the upward transport of Q from the surface and/or the latent fluxes from the ocean surface. The exact origin of these Q biases can be investigated in future work.

4.4 On dynamical imbalances

Note that the BGEnKF introduces less dynamical imbalances into the ensemble than the EnKF. To measure dynamical imbalance, we compute the root-mean-square of the second time derivative of surface pressure during the time integration phase of each DA cycle (P. Houtekamer & Mitchell, 2005; Temperton & Williamson, 1981). These derivatives are computed via centered differencing (Press & Flannery, 2010) on three consecutive snapshots of the surface pressure field. These snapshots are spaced 30-minutes apart. The resulting imbalance metric is normalized using the NoDA experiment's imbalance metric. A normalized imbalance metric value of 1 indicates that a normal amount of fast-moving gravity waves is present. A value greater than 1 indicates that a higher than normal amount of fast-moving gravity waves is present, thus indicating DA-induced imbalances.

Figure 6(b) indicates that the BGEnKF experiment generally has either statistically indistinguishable or milder imbalances than the EnKF experiment. The only exception to this trend happens between 0000 UTC to 1200 UTC on 17th October. The BGEnKF is thus likely more appropriate than the EnKF at assimilating Window-BT observations.

5 Conclusions and future work

In this study, we compare the BGEnKF against the EnKF using perfect model OSSEs with a realistic weather model (WRF) for a case of tropical convection. These OSSEs are executed using the state-of-the-art PSU-EnKF system. Our results indicate that the BGEnKF outperforms the EnKF at assimilating synthetic Window-BT observations. We observe this performance advantage in terms of the RMSEs and biases of the U, V, T, Q, Window-BT and WV-BT fields. This performance advantage is likely due to the BGEnKF's ability to handle mixtures of clear and cloudy column members. These performance advantages are achieved even though the BGEnKF s only activated for $\sim$10% of the assimilated Window-BT observations. As such, these promising results motivate future work into the BGEnKF using real data.

There are several large areas of future research for the BGEnKF. The first large area concerns refining the BGEnKF algorithm. Future work can, for instance, seek less heuristic approaches to sort the ensemble into clusters in a computationally efficient manner. One option is to combine clustering algorithms [*e.g.*, k-means (Forgy, 1965; Lloyd, 1982), support-vector machines (Cortes & Vapnik, 1995) and expectation maximization (Sondergaard & Lermusiaux, 2013b)] with dimension reduction methods [*e.g.*, Sondergaard and Lermusiaux (2013b), Reddy et al. (2020), Albarakati et al. (2021)]. Since cluster sizes, and thus sampling errors, can vary in each iteration of the serial BGEnKF loop, future work can investigate using adaptive or empirical localization methods (Anderson, 2012; Anderson & Lei, 2013; Lei & Anderson, 2014) to improve the BGEnKF's performance. Future work can also examine more sophisticated methods to regulate when the BGEnKF switches over to the EnKF (*e.g.*, using the Shapiro-Wilk test for normality).

Another area of future work is to hybridize the BGEnKF with other DA algorithms. Hybridization with kernel filters (Anderson & Anderson, 1999; Hoteit et al., 2008; Stordal et al., 2011; Hoteit et al., 2012; Liu et al., 2016; Stordal & Karlsen, 2017; Kotsuki et al., 2022) can be achieved by assigning the clear cluster's covariance to clear member kernels and likewise for the cloudy member kernels. Existing ensemble-variational hybrid DA algorithms (Hamill & Snyder, 2000; Lorenc, 2003; Buehner, 2005; X. Wang et al., 2007) can also be hybridized with the BGEnKF. For instance, the BGEnKF can replace the EnKF component of such methods. Hybridization with DA methods that employ transport methods to update ensemble members (Reich, 2012; van Leeuwen, 2011; Marzouk et al., 2017; Hu & van Leeuwen, 2021; Evensen Geir et al., 2022) is also possible. This can provide a different method to shift members between clusters, as opposed to the current deletion-resampling method. Finally, the BGEnKF can be potentially hybridized with ensemble DA methods that allow non-parametric prior distributions. Such methods include particle filters (van Leeuwen, 2009; Poterjoy, 2016; Vetra-Carvalho et al., 2018; Poterjoy et al., 2019; van Leeuwen et al., 2019), the quantile conserving ensemble filter (Anderson, 2022), and the rank histogram filter (Anderson, 2010, 2019, 2020).

Since we have only tested the BGEnKF in a perfect model WRF OSSE using Window-BT observations, future work can test the BGEnKF in increasingly realistic scenarios, with other observation types, and/or in other Earth systems. For instance, since radar reflectivity observations are sensitive to the presence and absence of precipitation, the BGEnKF can potentially be better at assimilating such observations. The performance of the BGEnKF can also be compared with other popular DA algorithms in tests that assimilate the operational suite of atmospheric in-situ and remote observations. Imperfect model OSSEs and real data tests can also be done. The BGEnKF can also be tested in other Earth system components.

This study is among the first to demonstrate the potential of the BGEnKF with a high-order weather model. Our BGEnKF is computationally efficient, scalable with parallelization, and likely straightforward to implement in existing serial EnKF DA systems. These algorithmic properties and our promising results motivate future research into developing, testing and applying the BGEnKF, or similar GMM-EnKFs, for Earth systems DA.

6 Open Research

The data and software used in this study are either publicly available or available upon request. The WRF model software can be found on the National Center for Atmospheric Research's WRF website (https://www.mmm.ucar.edu/weather-research-and-forecasting-model). Our WRF ensemble is constructed using the ECMWF TIGGE data archived on the MARS system (https://apps.ecmwf.int/datasets/data/tigge) and the ERA5 data archived on the CDS system (https://cds.climate.copernicus.eu). The MERG data product is obtained from NASA's GES DISC (https://disc.gsfc.nasa.gov/datasets/GPM_MERGIR_1/summary). We have archived this study's experiments and a copy of the Fortran 90 BGEnKF module on the Pennsylvania State University's Data Commons (http://doi.org/10.26208/XV41-7N75). The Fortran 90 source code of the PSU-EnKF system, including the implemented BGEnKF, is available upon request.

Acknowledgments

This work is supported by the Office of Naval Research (ONR) Grant N00014-18-1-2517, the National Aeronautics and Space Administration (NASA) Grant 80NSSC22K0613, the National Center for Atmospheric Research (NCAR) Advanced Study Program Graduate Visitor Program (ASP GVP), and the Water Cycle and Climate Extremes Modeling (WACCEM) project. WACCEM 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, and NCAR is sponsored by the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation. Finally, the computations in this study are performed using the Texas Advanced Computing Center (TACC) Stampede2 supercomputer and the National Energy Research Scientific Computing Center (NERSC) Cori supercomputer. NERSC is a U.S. Department of Energy Office of Science User Facility operated under Contract DE-AC02-05CH11231.

References

Albarakati, A., Budišić, M., Crocker, R., Glass-Klaiber, J., Iams, S., Maclean, J., . . . Van Vleck, E. S. (2021). Model and data reduction for data assimilation: Particle filters employing projected forecasts and data with application to a shallow water model. *Computers and Mathematics with Applications*. doi: 10.1016/j.camwa.2021.05.026

Anderson, J. L. (2003, 4). A Local Least Squares Framework for Ensemble Filtering. *Monthly Weather Review*, *131*(4), 634–642. Retrieved from http://journals.ametsoc.org/doi/10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2 doi: 10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2

Anderson, J. L. (2010, 11). A Non-Gaussian Ensemble Filter Update for Data Assimilation. *Monthly Weather Review*, *138*(11), 4186–4198. Retrieved from http://journals.ametsoc.org/doi/10.1175/2010MWR3253.1 doi: 10.1175/2010MWR3253.1

Anderson, J. L. (2012). Localization and sampling error correction in ensemble Kalman filter data assimilation. *Monthly Weather Review*, *140*(7). doi: 10.1175/MWR-D-11-00013.1

Anderson, J. L. (2019). *A nonlinear rank regression method for ensemble Kalman filter data assimilation* (Vol. 147) (No. 8). doi: 10.1175/MWR-D-18-0448.1

Anderson, J. L. (2020). A marginal adjustment rank histogram filter for non-Gaussian ensemble data assimilation. *Monthly Weather Review*, *148*(8). doi: 10.1175/MWR-D-19-0307.1

Anderson, J. L. (2022). A Quantile-Conserving Ensemble Filter Framework. Part I: Updating an Observed Variable. *Monthly Weather Review*, *150*(5). doi: 10.1175/mwr-d-21-0229.1

Anderson, J. L., & Anderson, S. L. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*. doi: 10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2

Anderson, J. L., & Collins, N. (2007, 8). Scalable Implementations of Ensemble Filter Algorithms for Data Assimilation. *Journal of Atmospheric and Oceanic Technology*, *24*(8), 1452–1463. Retrieved from https://journals.ametsoc.org/view/journals/atot/24/8/jtech2049_1.xml doi: 10.1175/JTECH2049.1

Anderson, J. L., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., & Avellano, A. (2009, 9). The data assimilation research testbed a community facility. *Bulletin of the American Meteorological Society*, *90*(9), 1283–1296. Retrieved from https://journals.ametsoc.org/doi/10.1175/2009BAMS2618.1 doi: 10.1175/2009BAMS2618.1

Anderson, J. L., & Lei, L. (2013). Empirical localization of observation impact in ensemble Kalman filters. *Monthly Weather Review*, *141*(11). doi: 10.1175/MWR-D-12-00330.1

Buehner, M. (2005). Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quarterly Journal of the Royal Meteorological Society*, *131*(607). doi: 10.1256/qj.04.15

Burgers, G., Jan van Leeuwen, P., Evensen, G., Van Leeuwen, P. J., & Evensen, G. (1998, 6). Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, *126*(6), 1719–1724. Retrieved from http://journals.ametsoc.org/doi/10.1175/1520-0493(1998)126%3C1719:ASITEK%3E2.0.CO;2 doi: 10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2

Chan, M.-Y., Anderson, J. L., & Chen, X. (2020). An efficient bi-Gaussian ensemble Kalman filter for satellite infrared radiance data assimilation. *Monthly Weather Review*. doi: 10.1175/mwr-d-20-0142.1

Chan, M.-Y., & Chen, X. (2021). Improving Analyses and Forecasts of a Tropical Squall Line using Upper Tropospheric Infrared Satellite Observations. *Advances in Atmospheric Sciences, Accepted Manuscript*. Retrieved from http://www.iapjournals

.ac.cn/aas/en/article/doi/10.1007/s00376-021-0449-8http://www.iapjournals.ac.cn/aas/en/article/doi/10.1007/s00376-021-0449-8?viewType=HTML doi: 10.1007/S00376-021-0449-8

Chan, M.-Y., Zhang, F., Chen, X., & Leung, L. R. (2020). Potential Impacts of Assimilating All-sky Satellite Infrared Radiances on Convection-Permitting Analysis and Prediction of Tropical Convection. *Monthly Weather Review*. doi: 10.1175/mwr-d-19-0343.1

Chen, F., & Dudhia, J. (2001). Coupling and advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Monthly Weather Review*. doi: 10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2

Chen, X., Leung, L. R., Feng, Z., & Song, F. (2021, 10). Crucial Role of Mesoscale Convective Systems in the Vertical Mass, Water and Energy Transports of the South Asian Summer Monsoon. *Journal of Climate*, *-1*(aop), 1–46. Retrieved from https://journals.ametsoc.org/view/journals/clim/aop/JCLI-D-21-0124.1/JCLI-D-21-0124.1.xml doi: 10.1175/JCLI-D-21-0124.1

Chen, X., Leung, L. R., Feng, Z., Song, F., & Yang, Q. (2021, 9). Mesoscale Convective Systems Dominate the Energetics of the South Asian Summer Monsoon Onset. *Geophysical Research Letters*, *48*(17), e2021GL094873. Retrieved from https://onlinelibrary.wiley.com/doi/full/10.1029/2021GL094873https://onlinelibrary.wiley.com/doi/abs/10.1029/2021GL094873https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021GL094873 doi: 10.1029/2021GL094873

Chen, X., Leung, L. R., Feng, Z., & Yang, Q. (2022, 4). Precipitation-Moisture Coupling Over Tropical Oceans: Sequential Roles of Shallow, Deep, and Mesoscale Convective Systems. *Geophysical Research Letters*, *49*(7). Retrieved from https://onlinelibrary.wiley.com/doi/10.1029/2022GL097836 doi: 10.1029/2022GL097836

Chen, X., Nystrom, R. G., Davis, C. A., & Zarzycki, C. M. (2020, 12). Dynamical Structures of Cross-Domain Forecast Error Covariance of a Simulated Tropical Cyclone in a Convection-Permitting Coupled Atmosphere-Ocean Model. *Monthly Weather Review*, *149*(1), 41–63. Retrieved from https://journals.ametsoc.org/view/journals/mwre/149/1/mwr-d-20-0116.1.xml doi: 10.1175/mwr-d-20-0116.1

Chen, X., Pauluis, O. M., Leung, L. R., & Zhang, F. (2018). Multiscale atmospheric overturning of the Indian summer monsoon as seen through isentropic analysis. *Journal of the Atmospheric Sciences*. doi: 10.1175/JAS-D-18-0068.1

Chen, X., Pauluis, O. M., & Zhang, F. (2018). Atmospheric overturning across multiple scales of an MJO event during the CINDY/DYNAMO campaign. *Journal of the Atmospheric Sciences*. doi: 10.1175/JAS-D-17-0060.1

Chen, X., & Zhang, F. (2019). Relative Roles of Preconditioning Moistening and Global Circumnavigating Mode on the MJO Convective Initiation During DYNAMO. *Geophysical Research Letters*. doi: 10.1029/2018GL080987

Chou, M.-D., & Suarez, M. J. (1999). A Solar Radiation Parameterization Atmospheric Studies. *Technical Report Series on Global Modeling and Data Assimilation*.

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, *20*(3). doi: 10.1023/A:1022627411411

Dovera, L., & Della Rossa, E. (2011). Multimodal ensemble Kalman filtering using Gaussian mixture models. *Computational Geosciences*, *15*(2), 307–323. doi: 10.1007/s10596-010-9205-3

Dowell, D. C., Alexander, C. R., James, E. P., Weygandt, S. S., Benjamin, S. G., Manikin, G. S., . . . Alcott, T. I. (2022). The High-Resolution Rapid Refresh (HRRR): An Hourly Updating Convection-Allowing Forecast Model. Part 1: Motivation and System Description. *Weather and Forecasting*. Retrieved from https://journals.ametsoc.org/view/journals/wefo/aop/WAF-D-21-0151.1/WAF-D-21-0151.1.xml doi: 10.1175/WAF-D-21-0151.1

ECMWF. (2016). IFS Documentation CY41R2 - Part I: Observations | ECMWF. In *Ifs doc-*

umentation cy41r2 (chap. 1). ECMWF. Retrieved from https://www.ecmwf.int/en/elibrary/16646-ifs-documentation-cy41r2-part-i-observations

Edwards, C. A., Moore, A. M., Hoteit, I., & Cornuelle, B. D. (2015). Regional ocean data assimilation. *Annual Review of Marine Science*, *7*. doi: 10.1146/annurev-marine-010814-015821

Emanuel, K. A. (1994). *Atmospheric Convection*.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, *99*(C5), 10143–10162. Retrieved from https://doi.org/10.1029/94JC00572http://doi.wiley.com/10.1029/94JC00572 doi: 10.1029/94JC00572

Evensen Geir, Vossepoel Femke C., & van Leeuwen Peter Jan. (2022). Particle Flow for a Quasi-Geostrophic Model. In *Data assimilation fundamentals: A unified formulation of the state and parameter estimation problem* (pp. 199–206). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-030-96709-3_20 doi: 10.1007/978-3-030-96709-3{_}20

Forgy, E. W. (1965). Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, *21*(3).

Fu, J. X., Wang, W., Shinoda, T., Ren, H. L., & Jia, X. (2017). Toward Understanding the Diverse Impacts of Air-Sea Interactions on MJO Simulations. *Journal of Geophysical Research: Oceans*, *122*(11). doi: 10.1002/2017JC013187

Gaspari, G., & Cohn, S. E. (1999, 1). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 723–757. Retrieved from http://doi.wiley.com/10.1002/qj.49712555417 doi: 10.1256/smsqj.55416

Geer, A. J., & Bauer, P. (2011, 10). Observation errors in all-sky data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *137*(661), 2024–2037. Retrieved from http://doi.wiley.com/10.1002/qj.830 doi: 10.1002/qj.830

Geer, A. J., Lonitz, K., Weston, P., Kazumori, M., Okamoto, K., Zhu, Y., . . . Schraff, C. (2018). All-sky satellite data assimilation at operational weather forecasting centres. *Quarterly Journal of the Royal Meteorological Society*. doi: 10.1002/qj.3202

Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., & Hunt, B. R. (2011). Balance and ensemble Kalman filter localization techniques. *Monthly Weather Review*, *139*(2), 511–522. doi: 10.1175/2010MWR3328.1

Grimes, D. I., & Pardo-Igúzquiza, E. (2010). Geostatistical analysis of rainfall. *Geographical Analysis*, *42*(2). doi: 10.1111/j.1538-4632.2010.00787.x

Hamill, T. M., & Snyder, C. (2000). A hybrid ensemble Kalman filter-3D variational analysis scheme. *Monthly Weather Review*, *128*(8 II). doi: 10.1175/1520-0493(2000)128<2905:ahekfv>2.0.co;2

Harnisch, F., Weissmann, M., & Periáñez. (2016). Error model for the assimilation of cloud-affected infrared satellite observations in an ensemble data assimilation system. *Quarterly Journal of the Royal Meteorological Society*. doi: 10.1002/qj.2776

Helmert, J., Şorman, A. , Montero, R. A., De Michele, C., de Rosnay, P., Dumont, M., . . . Arslan, A. N. (2018). *Review of snow data assimilation methods for hydrological, land surface, meteorological and climate models: Results from a COST harmosnow survey* (Vol. 8) (No. 12). doi: 10.3390/geosciences8120489

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., . . . Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*. doi: 10.1002/qj.3803

Honda, T., Miyoshi, T., Lien, G. Y., Nishizawa, S., Yoshida, R., Adachi, S. A., . . . Bessho, K. (2018, 1). Assimilating all-sky Himawari-8 satellite infrared radiances: A case of Typhoon Soudelor (2015). *Monthly Weather Review*, *146*(1), 213–229. doi: 10.1175/MWR-D-16-0357.1

Hong, S. Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*. doi: 10.1175/

MWR3199.1

Hoteit, I., Luo, X., & Pham, D.-T. (2012, 2). Particle Kalman Filtering: A Nonlinear Bayesian Framework for Ensemble Kalman Filters. *Monthly Weather Review*, *140*(2), 528–542. Retrieved from https://journals.ametsoc.org/doi/10.1175/2011MWR3640.1 doi: 10.1175/2011MWR3640.1

Hoteit, I., Pham, D. T., Triantafyllou, G., & Korres, G. (2008). A new approximate solution of the optimal nonlinear filter for data assimilation in meteorology and oceanography. *Monthly Weather Review*, *136*(1). doi: 10.1175/2007MWR1927.1

Houtekamer, P., & Mitchell, H. L. (2005, 10). Ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3269–3289. doi: 10.1256/qj.05.135

Houtekamer, P. L., & Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, *126*(3), 796–811. doi: 10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2

Houtekamer, P. L., & Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *129*(1), 123–137. doi: 10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2

Houtekamer, P. L., & Zhang, F. (2016). Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *144*(12), 4489–4532. doi: 10.1175/MWR-D-15-0440.1

Hu, C. C., & van Leeuwen, P. J. (2021). A particle flow filter for high-dimensional system applications. *Quarterly Journal of the Royal Meteorological Society*, *147*(737). doi: 10.1002/qj.4028

Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007, 6). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, *230*(1-2), 112–126. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0167278906004647 doi: 10.1016/j.physd.2006.11.008

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research Atmospheres*. doi: 10.1029/2008JD009944

Janowiak, J. E., Joyce, R. J., & Yarosh, Y. (2001). A real-time global half-hourly pixel-resolution infrared dataset and its applications. *Bulletin of the American Meteorological Society*, *82*(2). doi: 10.1175/1520-0477(2001)082<0205:ARTGHH>2.3.CO;2

Johnson, R. H., Rickenbach, T. M., Rutledge, S. A., Ciesielski, P. E., & Schubert, W. H. (1999). Trimodal characteristics of Tropical convection. *Journal of Climate*, *12*(8 PART 1). doi: 10.1175/1520-0442(1999)012<2397:tcotc>2.0.co;2

Keppenne, C. L., Rienecker, M. M., Kurkowski, N. P., & Adamec, D. A. (2005). Ensemble Kalman filter assimilation of temperature and altimeter data with bias correction and application to seasonal prediction. *Nonlinear Processes in Geophysics*, *12*(4). doi: 10.5194/npg-12-491-2005

Kotsuki, S., Miyoshi, T., Kondo, K., & Potthast, R. (2022). A Local Particle Filter and Its Gaussian Mixture Extension Implemented with Minor Modifications to the LETKF. *Geoscientific Model Development Discussions*, *2022*, 1–38. Retrieved from https://gmd.copernicus.org/preprints/gmd-2022-69/ doi: 10.5194/gmd-2022-69

Lei, L., & Anderson, J. L. (2014). Comparisons of empirical localization techniques for serial ensemble kalman filters in a simple atmospheric general circulation model. *Monthly Weather Review*, *142*(2). doi: 10.1175/MWR-D-13-00152.1

Lei, L., Anderson, J. L., & Whitaker, J. S. (2016). Localizing the impact of satellite radiance observations using a global group ensemble filter. *Journal of Advances in Modeling Earth Systems*, *8*(2). doi: 10.1002/2016MS000627

Lei, L., & Whitaker, J. S. (2015). Model space localization is not always better than observation space localization for assimilation of satellite radiances. *Monthly Weather Review*, *143*(10). doi: 10.1175/MWR-D-14-00413.1

Lei, L., Whitaker, J. S., Anderson, J. L., & Tan, Z. (2020). Adaptive Localization for Satellite Radiance Observations in an Ensemble Kalman Filter. *Journal of Advances in*

Modeling Earth Systems, *12*(8). doi: 10.1029/2019MS001693

Lim, K. S. S., & Hong, S. Y. (2010). Development of an effective double-moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models. *Monthly Weather Review*. doi: 10.1175/2009MWR2968.1

Liu, B., Ait-El-Fquih, B., & Hoteit, I. (2016). Efficient kernel-based ensemble Gaussian mixture filtering. *Monthly Weather Review*, *144*(2). doi: 10.1175/MWR-D-14-00292.1

Lloyd, S. P. (1982). Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2). doi: 10.1109/TIT.1982.1056489

Lorenc, A. C. (2003). Modelling of error covariances by 4D-Var data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *129*(595 PART B), 3167–3182. doi: 10.1256/qj.02.131

Madden, R. A., & Julian, P. R. (1971). Detection of a 40–50 Day Oscillation in the Zonal Wind in the Tropical Pacific. *Journal of the Atmospheric Sciences*. doi: 10.1175/1520-0469(1971)028<0702:doadoi>2.0.co;2

Madden, R. A., & Julian, P. R. (1972). Description of Global-Scale Circulation Cells in the Tropics with a 40–50 Day Period. *Journal of the Atmospheric Sciences*. doi: 10.1175/1520-0469(1972)029<1109:dogscc>2.0.co;2

Markowski, P., & Richardson, Y. (2010). *Mesoscale Meteorology in Midlatitudes*. doi: 10.1002/9780470682104

Marzouk, Y., Moselhy, T., Parno, M., & Spantini, A. (2017). Sampling via measure transport: An introduction. In *Handbook of uncertainty quantification.* doi: 10.1007/978-3-319-12385-1{_}23

Meng, Z., & Zhang, F. (2007). Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect model experiments. *Monthly Weather Review*, *135*(4), 1403–1423. doi: 10.1175/MWR3352.1

Meng, Z., & Zhang, F. (2008). Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVAR in a real-data case study. *Monthly Weather Review*. doi: 10.1175/2007MWR2106.1

Minamide, M., & Zhang, F. (2017). Adaptive observation error inflation for assimilating all-Sky satellite radiance. *Monthly Weather Review*, *145*(3), 1063–1081. doi: 10.1175/MWR-D-16-0257.1

Minamide, M., & Zhang, F. (2018). Assimilation of all-sky infrared radiances from Himawari-8 and impacts of moisture and hydrometer initialization on convection-permitting tropical cyclone prediction. *Monthly Weather Review*, *146*(10), 3241–3258. doi: 10.1175/MWR-D-17-0367.1

Minamide, M., & Zhang, F. (2019). An adaptive background error inflation method for assimilating all-sky radiances. *Quarterly Journal of the Royal Meteorological Society*, *145*(719), 805–823. doi: 10.1002/qj.3466

Park, S. K., & Xu, L. (2016). *Data assimilation for atmospheric, oceanic and hydrologic applications (Vol. III)*. doi: 10.1007/978-3-319-43415-5

Poterjoy, J. (2016, 1). A localized particle filter for high-dimensional nonlinear systems. *Monthly Weather Review*, *144*(1), 59–76. Retrieved from `http://journals.ametsoc.org/doi/10.1175/MWR-D-15-0163.1` doi: 10.1175/MWR-D-15-0163.1

Poterjoy, J., Wicker, L., & Buehner, M. (2019). Progress toward the application of a localized particle filter for numerical weather prediction. *Monthly Weather Review*. doi: 10.1175/MWR-D-17-0344.1

Press, W., & Flannery, B. (2010). *Numerical Recipes in Fortran 90* (Vol. 35) (No. 6).

Reddy, G. T., Reddy, M. P. K., Lakshmanna, K., Kaluri, R., Rajput, D. S., Srivastava, G., & Baker, T. (2020). Analysis of Dimensionality Reduction Techniques on Big Data. *IEEE Access*, *8*. doi: 10.1109/ACCESS.2020.2980942

Reich, S. (2012). A Gaussian-mixture ensemble transform filter. *Quarterly Journal of the Royal Meteorological Society*, *138*(662). doi: 10.1002/qj.898

Reichle, R. H., Bosilovich, M. G., Crow, W. T., Koster, R. D., Kumar, S. V., Mahanama,

S. P. P., & Zaitchik, B. F. (2009). Recent Advances in Land Data Assimilation at the NASA Global Modeling and Assimilation Office. In *Data assimilation for atmospheric, oceanic and hydrologic applications.* doi: 10.1007/978-3-540-71056-1{_}21

Skamarock, W., Klemp, J., Dudhi, J., Gill, D., Barker, D., Duda, M., ... Powers, J. (2008). A Description of the Advanced Research WRF Version 3. NCAR Tech. Note NCAR/TN–468+STR, 113 pp. *NCAR TECHNICAL NOTE*. doi: 10.5065/D68S4MVH

Sondergaard, T., & Lermusiaux, P. F. (2013a). Data assimilation with gaussian mixture models using the dynamically orthogonal field equations. Part II: Applications. *Monthly Weather Review*, *141*(6), 1761–1785. doi: 10.1175/MWR-D-11-00296.1

Sondergaard, T., & Lermusiaux, P. F. (2013b, 6). Data assimilation with gaussian mixture models using the dynamically orthogonal field equations. Part I: Theory and scheme. *Monthly Weather Review*, *141*(6), 1737–1760. Retrieved from `https://journals.ametsoc.org/view/journals/mwre/141/6/mwr-d-11-00295.1.xml` doi: 10.1175/MWR-D-11-00295.1

Stammer, D., Balmaseda, M., Heimbach, P., Köhl, A., & Weaver, A. (2016). Ocean Data Assimilation in Support of Climate Applications: Status and Perspectives. *Annual Review of Marine Science*, *8*. doi: 10.1146/annurev-marine-122414-034113

Stordal, A. S., & Karlsen, H. A. (2017). Large sample properties of the adaptive gaussian mixture filter. *Monthly Weather Review*, *145*(7). doi: 10.1175/MWR-D-15-0372.1

Stordal, A. S., Karlsen, H. A., Nævdal, G., Skaug, H. J., & Vallès, B. (2011). Bridging the ensemble Kalman filter and particle filters: The adaptive Gaussian mixture filter. *Computational Geosciences*, *15*(2). doi: 10.1007/s10596-010-9207-1

Swinbank, R., Kyouda, M., Buchanan, P., Froude, L., Hamill, T. M., Hewson, T. D., ... Yamaguchi, M. (2016). The TIGGE project and its achievements. *Bulletin of the American Meteorological Society*. doi: 10.1175/BAMS-D-13-00191.1

Temperton, C., & Williamson, D. L. (1981). Normal mode initialization for a multilevel grid-point model. Part I: Linear aspects. *Monthly Weather Review*, *109*(4). doi: 10.1175/1520-0493(1981)109<0729:NMIFAM>2.0.CO;2

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., & Whitaker, J. S. (2003, 7). Ensemble Square Root Filters. *Monthly Weather Review*, *131*(7), 1485–1490. Retrieved from `http://journals.ametsoc.org/doi/10.1175/1520-0493(2003)131%3C1485:ESRF%3E2.0.CO;2` doi: 10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2

van Leeuwen, P. J. (2009). *Particle filtering in geophysical systems.* doi: 10.1175/2009MWR2835.1

van Leeuwen, P. J. (2011). Efficient nonlinear data assimilation for oceanic models of intermediate complexity. In *Ieee workshop on statistical signal processing proceedings.* doi: 10.1109/SSP.2011.5967700

van Leeuwen, P. J., Künsch, H. R., Nerger, L., Potthast, R., & Reich, S. (2019, 7). *Particle filters for high-dimensional geoscience applications: A review* (Vol. 145) (No. 723). John Wiley and Sons Ltd. doi: 10.1002/qj.3551

Vetra-Carvalho, S., van Leeuwen, P. J., Nerger, L., Barth, A., Altaf, M. U., Brasseur, P., ... Beckers, J. M. (2018). State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems. *Tellus, Series A: Dynamic Meteorology and Oceanography*. doi: 10.1080/16000870.2018.1445364

Wang, S., Sobel, A. H., Zhang, F., Qiang Sun, Y., Yue, Y., & Zhou, L. (2015). Regional simulation of the october and november MJO events observed during the CINDY/DYNAMO field campaign at gray zone resolution. *Journal of Climate*, *28*(6), 2097–2119. doi: 10.1175/JCLI-D-14-00294.1

Wang, X., Snyder, C., & Hamill, T. M. (2007). On the theoretical equivalence of differently proposed ensemble - 3DVAR hybrid analysis schemes. *Monthly Weather Review*, *135*(1). doi: 10.1175/MWR3282.1

Whitaker, J. S., & Hamill, T. M. (2002, 7). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, *130*(7), 1913–1924. Retrieved from `http://journals.ametsoc.org/doi/10.1175/1520-0493(2002)130%3C1913:EDAWPO%3E2.0.CO;2` doi: 10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2

Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y., & Toth, Z. (2008). Ensemble data assimilation with the NCEP global forecast system. *Monthly Weather Review*, *136*(2). doi: 10.1175/2007MWR2018.1

Ying, Y., & Zhang, F. (2017). Practical and intrinsic predictability of multiscale weather and convectively coupled equatorial waves during the active phase of an MJO. *Journal of the Atmospheric Sciences*, *74*(11), 3771–3785. doi: 10.1175/JAS-D-17-0157.1

Ying, Y., & Zhang, F. (2018). Potentials in improving predictability of multiscale tropical weather systems evaluated through ensemble assimilation of simulated satellite-based observations. *Journal of the Atmospheric Sciences*, *75*(5), 1675–1698. doi: 10.1175/JAS-D-17-0245.1

Zhang, F., Minamide, M., & Clothiaux, E. E. (2016). Potential impacts of assimilating all-sky infrared satellite radiances from GOES-R on convection-permitting analysis and prediction of tropical cyclones. *Geophysical Research Letters*, *43*(6), 2954–2963. doi: 10.1002/2016GL068468

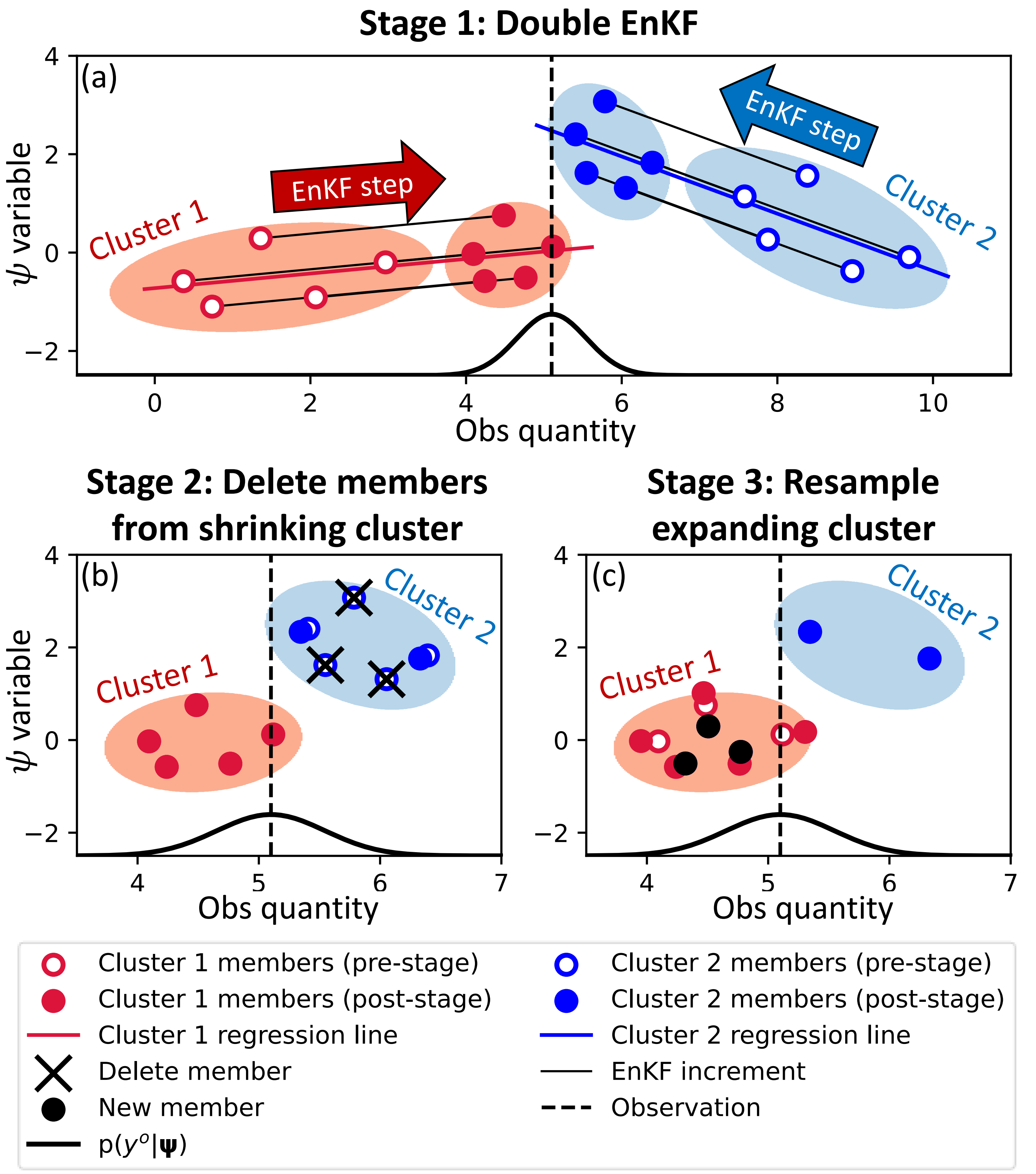
Zhang, F., Snyder, C., & Sun, J. (2004, 5). Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Monthly Weather Review*, *132*(5), 1238–1253. doi: 10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2

Zhang, F., Taraphdar, S., & Wang, S. (2017). The role of global circumnavigating mode in the MJO initiation and propagation. *Journal of Geophysical Research*, *122*(11), 5837–5856. doi: 10.1002/2016JD025665

Zhang, Y., Sieron, S. B., Lu, Y., Chen, X., Nystrom, R. G., Minamide, M., ... Zhang, F. (2021, 12). *Ensemble-Based Assimilation of Satellite All-Sky Microwave Radiances Improves Intensity and Rainfall Predictions for Hurricane Harvey (2017)* (Vol. 48) (No. 24). John Wiley and Sons Inc. doi: 10.1029/2021GL096410

Zhang, Y., Stensrud, D. J., & Zhang, F. (2019, 12). Simultaneous Assimilation of Radar and All-Sky Satellite Infrared Radiance Observations for Convection-Allowing Ensemble Analysis and Prediction of Severe Thunderstorms. *Monthly Weather Review*, *147*(12), 4389–4409. Retrieved from `https://journals.ametsoc.org/view/journals/mwre/147/12/mwr-d-19-0163.1.xml` doi: 10.1175/MWR-D-19-0163.1

Figure 1.



Stage 1: Double EnKF
(a)
EnKF step
EnKF step
Cluster 1
Cluster 2
ψ variable
Obs quantity
Stage 2: Delete members from shrinking cluster
(b)
Cluster 1
Cluster 2
ψ variable
Obs quantity
Stage 3: Resample expanding cluster
(c)
Cluster 1
Cluster 2
Obs quantity
Cluster 1 members (pre-stage)
Cluster 1 members (post-stage)
Cluster 1 regression line
Delete member
New member
p(y^o|ψ)
Cluster 2 members (pre-stage)
Cluster 2 members (post-stage)
Cluster 2 regression line
EnKF increment
Observation

Figure 6.

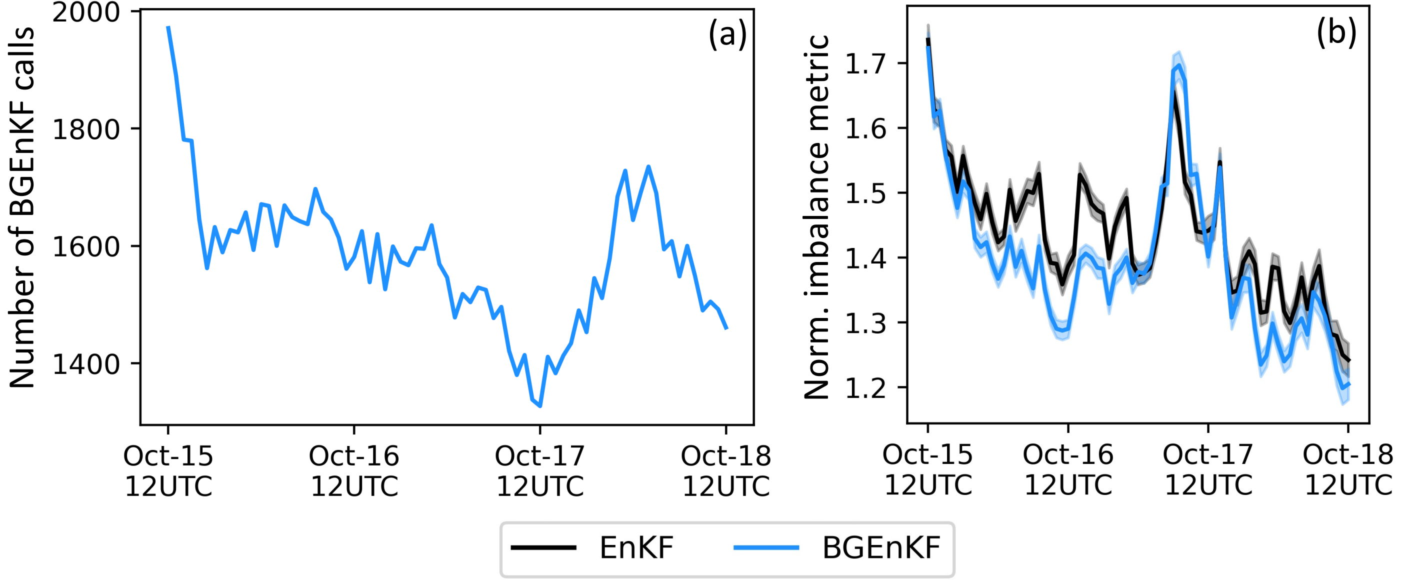

(a)
Number of BGEnKF calls
2000
1800
1600
1400
Oct-15 12UTC
Oct-16 12UTC
Oct-17 12UTC
Oct-18 12UTC
(b)
Norm. imbalance metric
1.7
1.6
1.5
1.4
1.3
1.2
Oct-15 12UTC
Oct-16 12UTC
Oct-17 12UTC
Oct-18 12UTC
EnKF
BGEnKF

Figure 2.

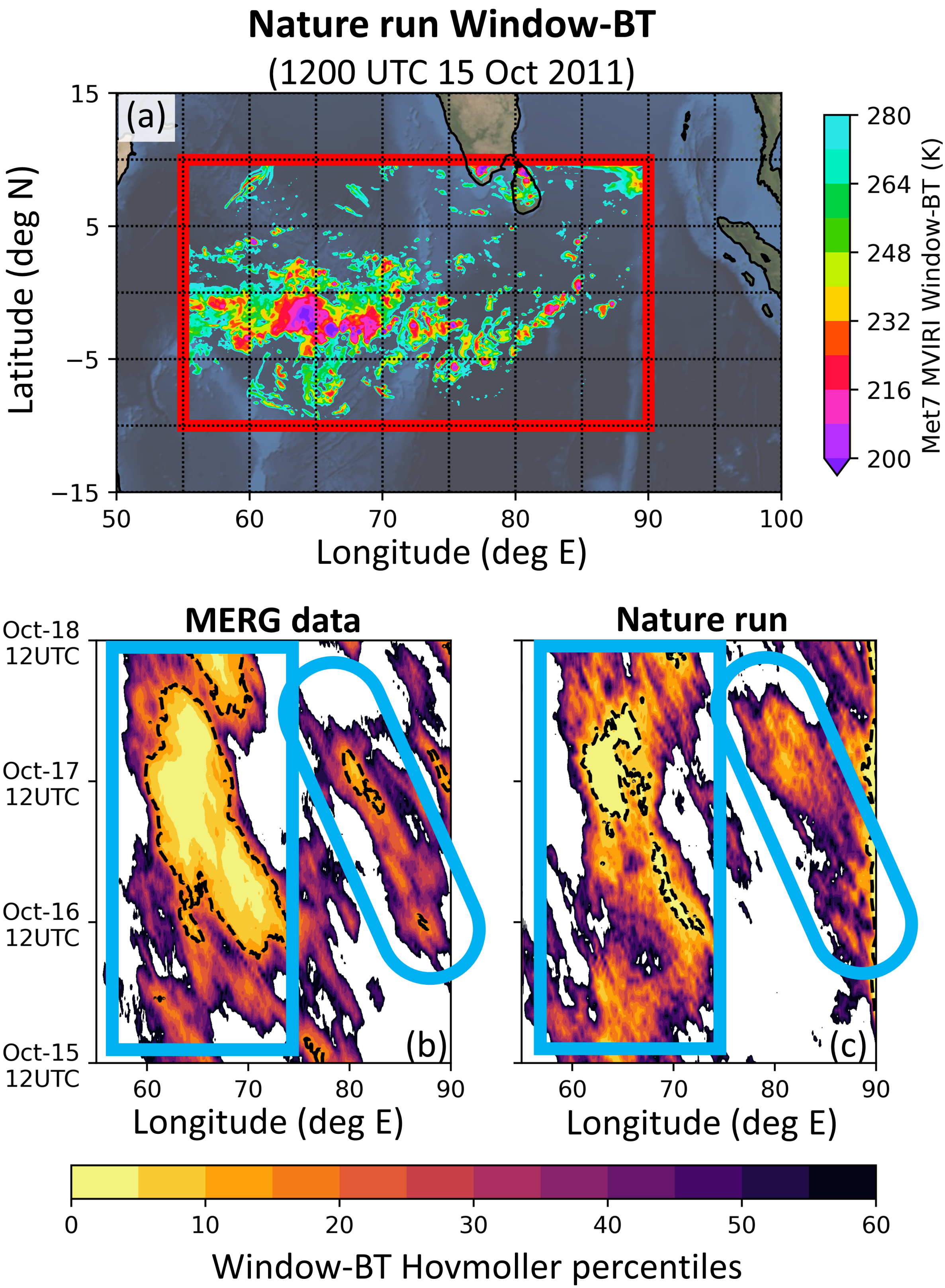

Nature run Window-BT
(1200 UTC 15 Oct 2011)
(a)
Latitude (deg N)
15
5
−5
−15
50
60
70
80
90
100
Longitude (deg E)
Met7 MVIRI Window-BT (K)
280
264
248
232
216
200
MERG data
Nature run
Oct-18 12UTC
Oct-17 12UTC
Oct-16 12UTC
Oct-15 12UTC
(b)
(c)
60
70
80
90
Longitude (deg E)
Longitude (deg E)
0
10
20
30
40
50
60
Window-BT Hovmoller percentiles

Figure 5.

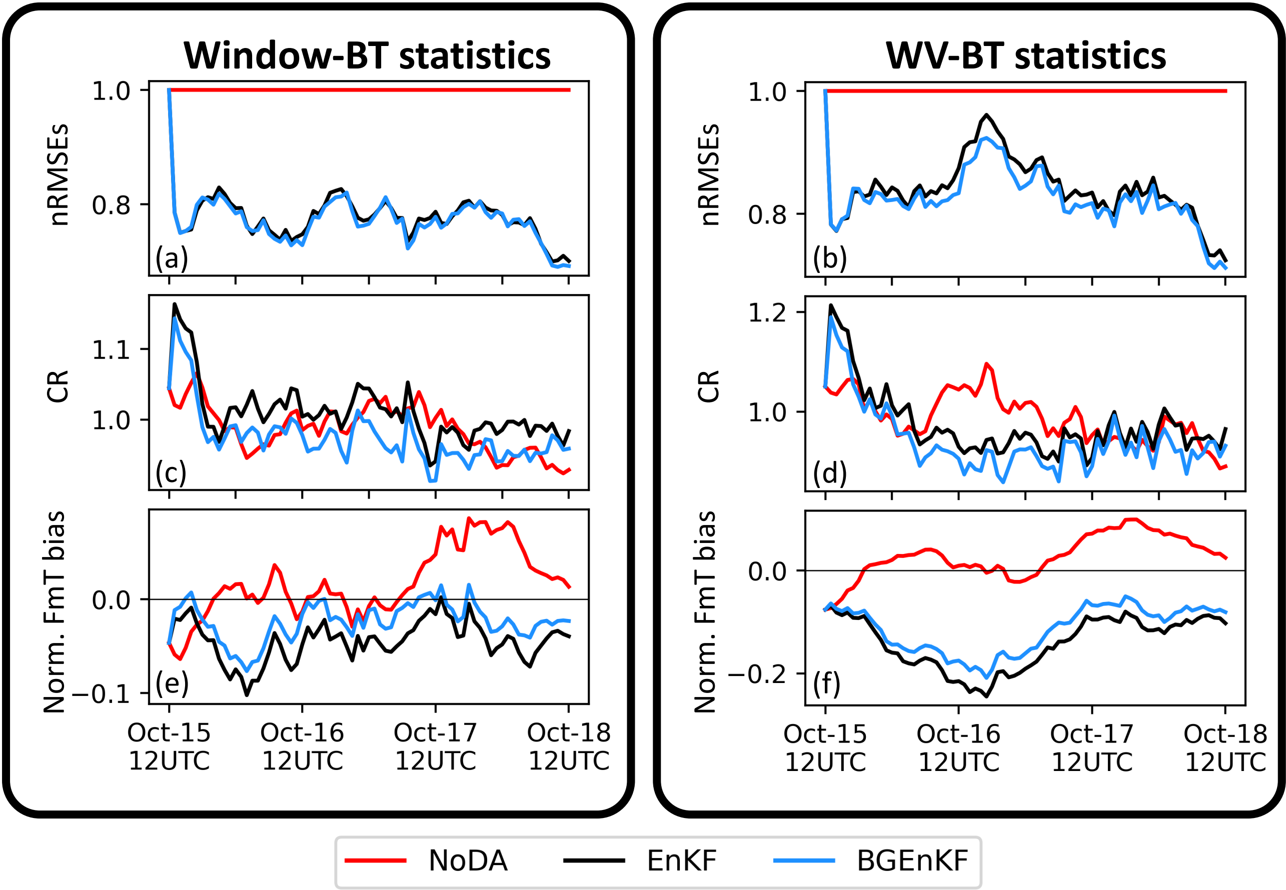
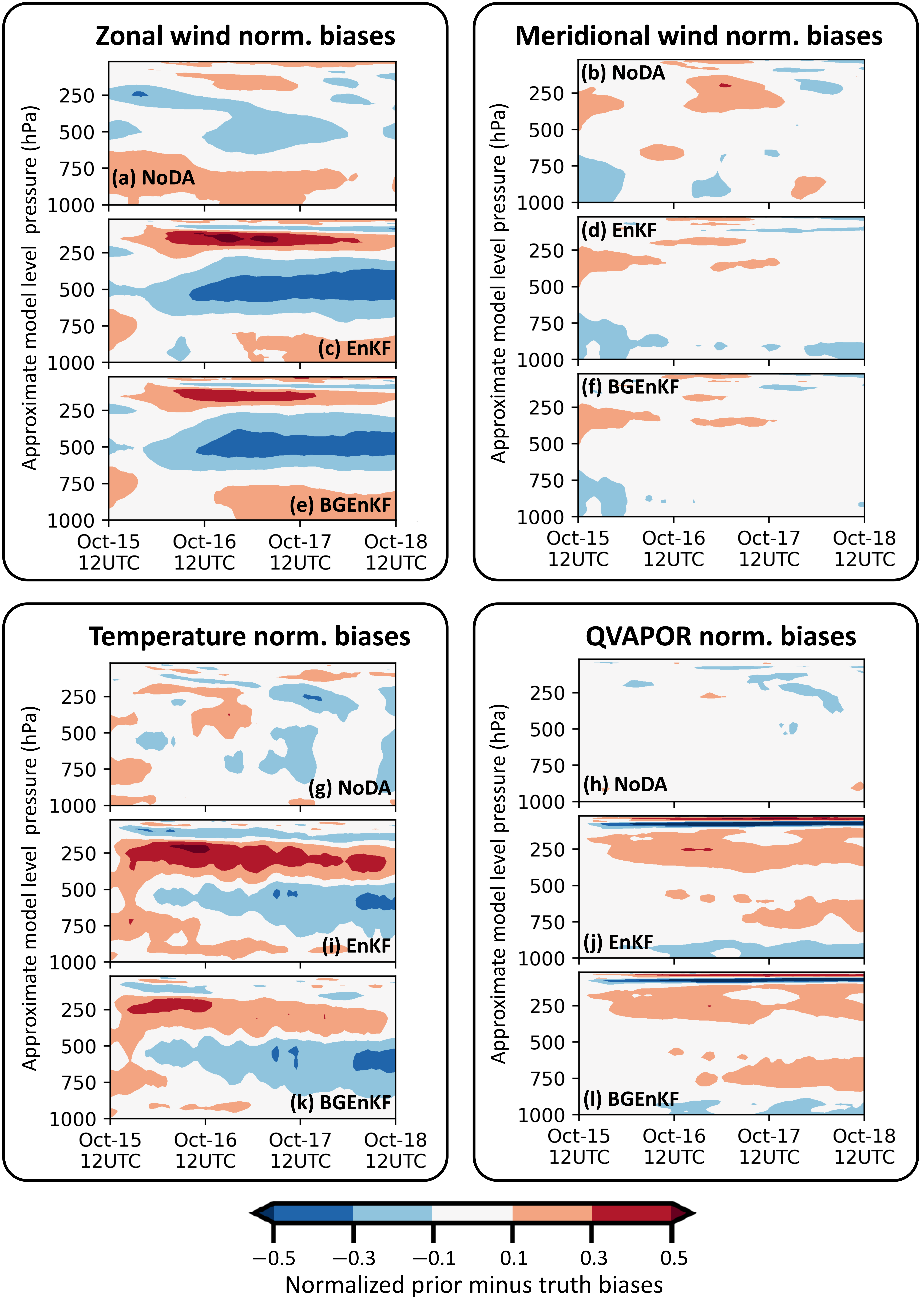
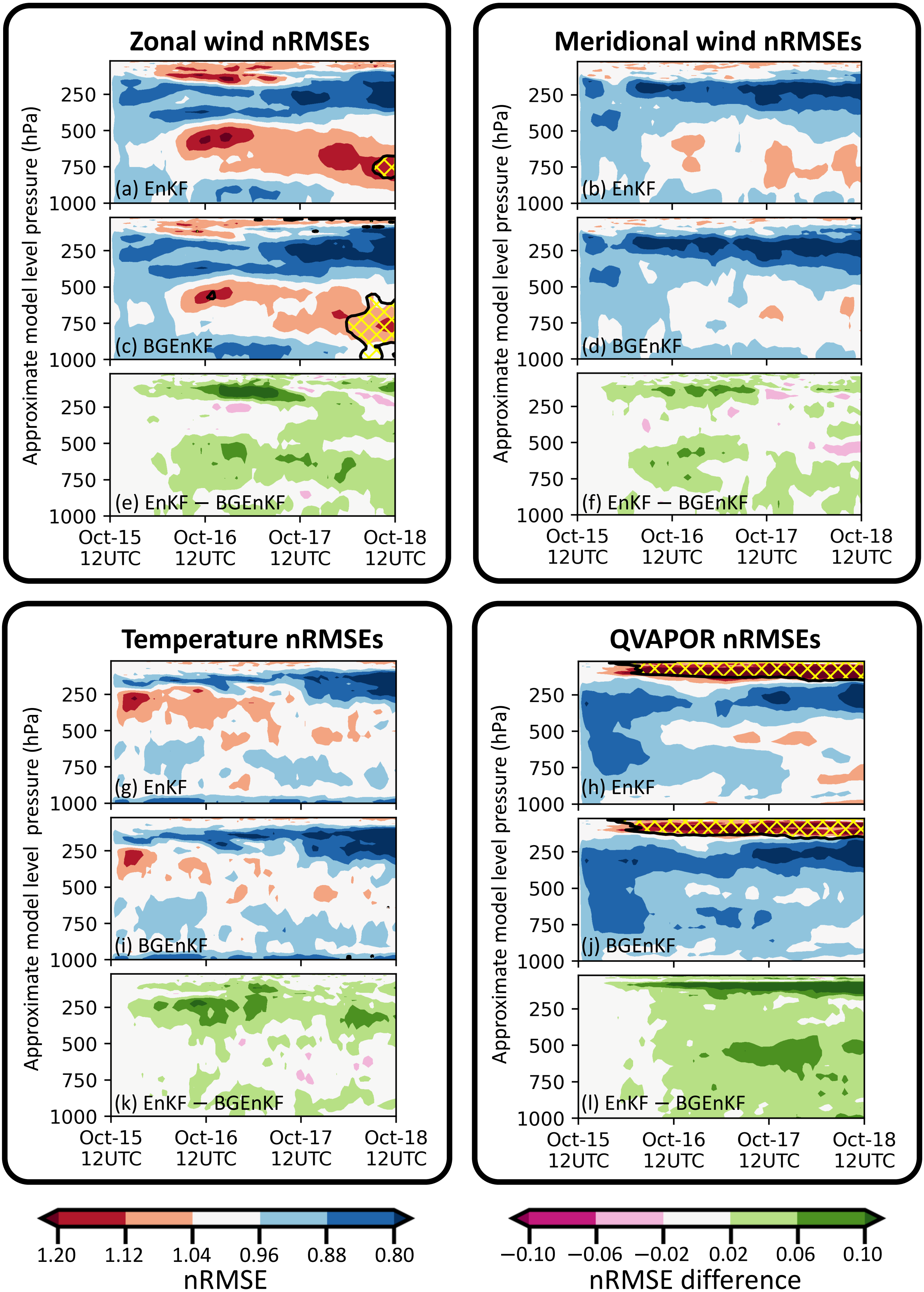

Window-BT statistics
WV-BT statistics
nRMSEs
CR
Norm. FmT bias
(a)
(b)
(c)
(d)
(e)
(f)
1.0
0.8
1.1
1.2
0.0
−0.1
−0.2
Oct-15 12UTC
Oct-16 12UTC
Oct-17 12UTC
Oct-18 12UTC
NoDA
EnKF
BGEnKF

Figure 4.



Zonal wind norm. biases
Meridional wind norm. biases
Temperature norm. biases
QVAPOR norm. biases
(a) NoDA
(b) NoDA
(c) EnKF
(d) EnKF
(e) BGEnKF
(f) BGEnKF
(g) NoDA
(h) NoDA
(i) EnKF
(j) EnKF
(k) BGEnKF
(l) BGEnKF
Approximate model level pressure (hPa)
250
500
750
1000
Oct-15 12UTC
Oct-16 12UTC
Oct-17 12UTC
Oct-18 12UTC
−0.5
−0.3
−0.1
0.1
0.3
0.5
Normalized prior minus truth biases

Figure 3.



Zonal wind nRMSEs
Meridional wind nRMSEs
Temperature nRMSEs
QVAPOR nRMSEs
Approximate model level pressure (hPa)
250
500
750
1000
(a) EnKF
(b) EnKF
(c) BGEnKF
(d) BGEnKF
(e) EnKF − BGEnKF
(f) EnKF − BGEnKF
(g) EnKF
(h) EnKF
(i) BGEnKF
(j) BGEnKF
(k) EnKF − BGEnKF
(l) EnKF − BGEnKF
Oct-15 12UTC
Oct-16 12UTC
Oct-17 12UTC
Oct-18 12UTC
1.20
1.12
1.04
0.96
0.88
0.80
nRMSE
−0.10
−0.06
−0.02
0.02
0.06
0.10
nRMSE difference

The potential benefits of handling mixture statistics via a bi-Gaussian EnKF: tests with all-sky satellite infrared radiances

Man-Yau Chan[1,2], Xingchao Chen[1,2], Jeffrey L. Anderson[3]

[1]Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, Pennsylvania, USA

[2]Center for Advanced Data Assimilation and Predictability Techniques, The Pennsylvania State University, University Park, Pennsylvania, USA

[3]Data Assimilation Research Section, Computational Information Systems Laboratory, National Center for Atmospheric Research, Boulder, Colorado, USA

Key Points:

- Current ensemble DA methods assume that forecasts follow a normal distribution. This assumption is often invalid.
- In this study, we propose a computationally efficient ensemble DA method that handles clear and cloudy forecasts separately.
- This study uses a realistic weather model (WRF) to show that this method can outperform the EnKF.

Corresponding author: Man-Yau Chan, `chanmanyau@gmail.com`

Abstract

The meteorological characteristics of cloudy atmospheric columns can be very different from their clear counterparts. Thus, when a forecast ensemble is uncertain about the presence/absence of clouds at a specific atmospheric column (*i.e.*, some members are clear while others are cloudy), that column's ensemble statistics will contain a mixture of clear and cloudy statistics. Such mixtures are inconsistent with the ensemble data assimilation algorithms currently used in numerical weather prediction. Hence, ensemble data assimilation algorithms that can handle such mixtures can potentially outperform currently used algorithms.

In this study, we demonstrate the potential benefits of addressing such mixtures through a bi-Gaussian extension of the ensemble Kalman filter (BGEnKF). The BGEnKF is compared against the commonly used ensemble Kalman filter (EnKF) using perfect model observing system simulated experiments (OSSEs) with a realistic weather model (the Weather Research and Forecast model). Synthetic all-sky infrared radiance observations are assimilated in this study. In these OSSEs, the BGEnKF outperforms the EnKF in terms of the horizontal wind components, temperature, specific humidity, and simulated upper tropospheric water vapor channel infrared brightness temperatures.

This study is one of the first to demonstrate the potential of a Gaussian mixture model EnKF with a realistic weather model. Our results thus motivate future research towards improving numerical Earth system predictions though explicitly handling mixture statistics.

Plain Language Summary

The accuracy of a computer weather forecast often depends on the accuracy of the information inputted into the computer forecast system. The accuracy of the input in turn depends on the accuracy of the input-constructing algorithm. Such algorithms often use probabilistic forecasts from an earlier point in time and current atmospheric measurements to construct the inputs.

A common assumption in input-constructing algorithms is that the probabilistic forecasts follow multivariate normal distributions (henceforth called the normality assumption). However, in the frequent situation where the probabilistic forecasts are uncertain about the presence/absence of clouds, the normality assumption is violated. This is because clear atmospheric columns and cloudy atmospheric columns have distinctly different thermodynamic and dynamic characteristics. Such probabilistic forecasts thus have mixed statistics (henceforth termed mixed probabilistic forecasts). Addressing these mixed statistics can potentially improve forecasts.

In this study, we propose a new input-constructing algorithm that can explicitly handle mixed probabilistic forecasts. Compared to an existing popular algorithm, our algorithm is nearly as fast and can produced more accurate forecast inputs. Our work thus suggests that weather forecasts can be improved by upgrading input-constructing algorithms to treat a common situation where the normality assumption is violated.

1 Introduction

Earth system analysis and forecasting systems rely on ensemble data assimilation (ensemble DA, or EDA) methods to convert observations into corrections for Earth system model variables (Keppenne et al., 2005; Reichle et al., 2009; Edwards et al., 2015; Stammer et al., 2016; Park & Xu, 2016; ECMWF, 2016; Helmert et al., 2018; Hersbach et al., 2020). Current operational EDA methods typically assume that every member in an input forecast ensemble is drawn from a distribution only containing a single Gaussian kernel [*i.e.*, a Gaussian distribution; henceforth termed the unmixed ensemble assumption; *e.g.*, Geer et al. (2018) and Dowell et al. (2022)]. The effectiveness of such methods can thus can be limited by the validity of this assumption.

The unmixed ensemble assumption is violated for ensembles that are uncertain about the presence or absence of clouds at any model grid point. This is because clear atmospheric columns and cloudy atmospheric columns often have different dynamic, thermodynamic, and radiative properties [*e.g.*, Emanuel (1994), Markowski and Richardson (2010)]. Cloudy statistics are thus often different from clear statistics [*e.g.*, Grimes and Pardo-Igúzquiza (2010); Geer and Bauer (2011)]. If some ensemble members are cloudy at a location, and other members are clear at this location, the ensemble can exhibit mixed statistics (Harnisch et al., 2016; Minamide & Zhang, 2017; Honda et al., 2018; Chan, Anderson, & Chen, 2020). More evidence of mixed statistics can be found in the supporting information. The effectiveness of current operational EDA methods is likely limited in such situations.

This limitation can be mitigated by extending current operational EDA methods to handle mixed statistics. One possibility is to extend the commonly used ensemble Kalman filter, or the EnKF (Evensen, 1994; P. L. Houtekamer & Mitchell, 1998; Burgers et al., 1998; Tippett et al., 2003; Anderson, 2003; Whitaker & Hamill, 2002; Keppenne et al., 2005; Hunt et al., 2007; Reichle et al., 2009; Stammer et al., 2016; Edwards et al., 2015; Park & Xu, 2016; Helmert et al., 2018), to handle members drawn from forecast distributions with two Gaussian kernels. Specifically, we assume that forecast members that are clear at an observation site (henceforth, clear members) are drawn from one Gaussian kernel, and forecast members that are cloudy at this site (henceforth, cloudy members) are drawn from a different Gaussian kernel. This bi-Gaussian extension of the EnKF (henceforth, the BGEnKF) allows the clear ensemble statistics to be handled separately from the cloudy ensemble statistics (Chan, Anderson, & Chen, 2020), thus addressing the issue of mixed statistics.

We recently proposed a computationally efficient BGEnKF to handle mixtures of clear and cloudy members [Chan, Anderson, and Chen (2020); henceforth, the CAC20 BGEnKF]. Unlike similar methods proposed in the past (Dovera & Della Rossa, 2011; Reich, 2012; Sondergaard & Lermusiaux, 2013a, 2013b), the CAC20 BGEnKF does not use an expectation maximization (EM) algorithm to estimate the mean and covariances of the two Gaussian kernels. Instead, the CAC20 BGEnKF assigns the the sample mean and covariances of the cloudy members to one Gaussian kernel, and those of the clear members to the other Gaussian kernel. This assignment circumvents the computational cost and issues associated with using the EM algorithm in high dimensional spaces [see Chan, Anderson, and Chen (2020) for more information]. Furthermore, the CAC20 BGEnKF converts clear members into cloudy members, or *vice versa*, without involving the costly square-root computations or Cholesky decompositions of high-dimensional forecast covariance matrices.

The purpose of this study is to demonstrate that a variant of the CAC20 BGEnKF can outperform the EnKF using a realistic high-order weather model (the Weather Research and Forecast model; WRF). To do so, this new BGEnKF is implemented into the state-of-the-art Pennsylvania State University EnKF system [henceforth, the PSU-EnKF system; Meng and Zhang (2007, 2008), Chan, Zhang, et al. (2020)]. This demonstration is done using perfect model observing system simulation experiments (OSSEs) of a case of tropical convection over the equatorial Indian Ocean. This case occurred during the onset of the active phase of the October 2011 Madden-Julian Oscillation event [MJO; Madden and Julian (1971, 1972), and S. Wang et al. (2015)].

The structure of this paper is as follows. In section 2, we will give an overview of the BGEnKF algorithm, discuss how clear and cloudy members are identified, and modifications made to the CAC20 BGEnKF algorithm. A detailed description of the current BGEnKF, along with suggestions on handling more than two Gaussian kernels, can be found in the supporting information. Following that, we will discuss the setup of our OSSEs in section 3 and the results in section 4. We will then conclude in section 5.

2 On the BGEnKF algorithm

2.1 On the identification of clear and cloudy members

The BGEnKF requires identifying clear and cloudy members at each iteration of the serial data assimilation loop. A simple identification method is to check if the members' column-integrated liquid and/or frozen water mass contents exceed a threshold.

The choice of which phase of water to include in the column integration depends on the specifics of the forecast model. As will be discussed in section 3.3, this study used a WRF model setup with a 9-km horizontal grid spacing and without convective parameterization. This WRF model setup cannot realistically resolve trade cumuli since the typical width of trade cumuli is ~1-km. As such, we consider columns with trade cumuli and entirely cloud-free columns as clear member columns, and the remaining members as cloudy member columns. Since trade cumuli do not typically grow above the melting layer (Johnson et al., 1999), clear members do not possess frozen water. It thus seems appropriate to use column-integrated ice mass content (ξ) to distinguish between clear and cloudy member columns. To be precise, we compute ξ at a given model column via

$$\xi \equiv \int_0^{z_{top}} \rho(q_i + q_s + q_g)dz \tag{1}$$

where z_{top} is the model top altitude and ρ represents air density. Furthermore, q_i, q_s and q_g are the mass mixing ratios of ice, snow and graupel, respectively.

In this study, we will consider model columns with $\xi \geq 1$ g/m^2 as cloudy, and model columns with $\xi < 1$ g/m^2 as clear. The cloudy and clear infrared window channel simulated brightness temperature statistics (Window-BT; central wavelength of 10.5 μm) do not vary noticeably for model column ξ thresholds between 0.8-1.2 g/m^2. Future studies can refine the threshold value or seek better ways to separate clear and cloudy column members.

2.2 Overview of the BGEnKF algorithm

This study's BGEnKF (and the CAC20 BGEnKF) assimilates observations with Gaussian observation likelihoods under the assumption that clear members are drawn from one Gaussian kernel and cloudy members are drawn from another Gaussian kernel. Suppose we seek to constrain the following extended state vector $\boldsymbol{\psi}$

$$\boldsymbol{\psi} \equiv \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{h}(\boldsymbol{x}) \\ \boldsymbol{\xi}(\boldsymbol{x}) \end{bmatrix} \tag{2}$$

where $\boldsymbol{x}$ represents the model state, $\boldsymbol{h}(\boldsymbol{x})$ represents applying the observation operator $\boldsymbol{h}$ on $\boldsymbol{x}$, and $\boldsymbol{\xi}(\boldsymbol{x})$ represents computing ξ at all observation sites [Eq. (1)]. Note that observation sites here refers to the latitude and longitude of the observation (*i.e.*, the vertical position is not considered for now). Supposing there are N_x elements in $\boldsymbol{x}$ and N_y observations, then $\boldsymbol{\psi}$ has $N_x + 2N_y$ elements.

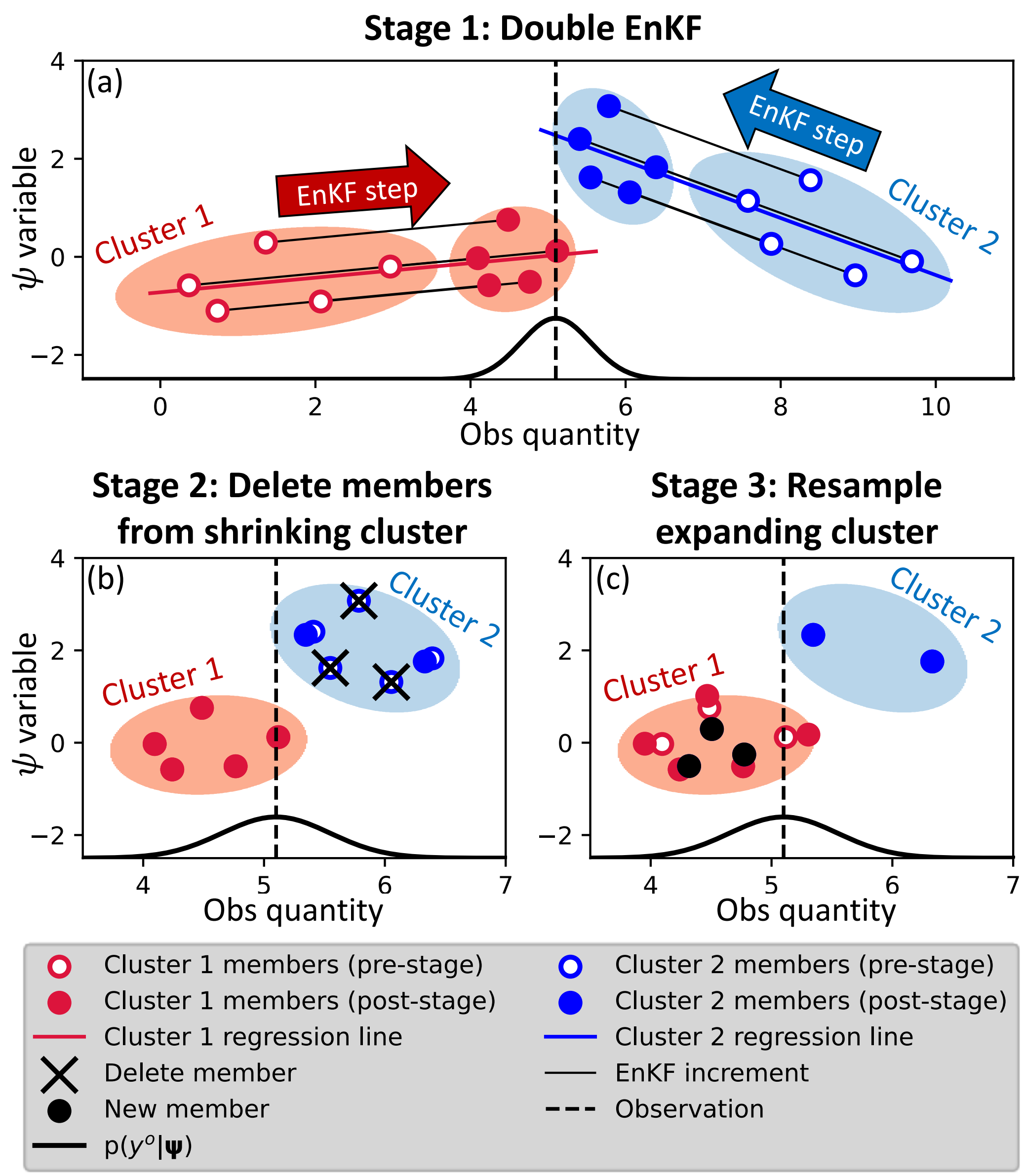


Figure 1. A bivariate demonstration of the three-stage process of the BGEnKF algorithm. The light red ovals highlight cluster 1 members and the light blue ovals highlight cluster 2 members. Prior to running the BGEnKF update, the prior members have already been separated into two clusters. The BGEnKF's first stage is to employ the EnKF update equations on the two clusters separately (panel a). In the second stage (panel b), the BGEnKF identifies the shrinking cluster (the blue cluster 2 in this case), deletes an appropriate number of members from this cluster, and adjusts the remaining members to prevent the deletion from changing this cluster's mean. The BGEnKF's final stage (panel c) is to recreate the deleted members by resampling from the expanding cluster (cluster 1).

The BGEnKF assumes that the prior probability density function [pdf; $p(\boldsymbol{\psi})$] can be represented by the bi-Gaussian pdf

$$p(\boldsymbol{\psi}) = w^f_{\text{clr}}\, \mathcal{G}\left(\boldsymbol{\psi};\ \overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{clr}}}, \boldsymbol{P}^{\boldsymbol{f}}_{\text{clr}}\right) + w^f_{\text{cld}}\, \mathcal{G}\left(\boldsymbol{\psi};\ \overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{cld}}}, \boldsymbol{P}^{\boldsymbol{f}}_{\text{cld}}\right). \tag{3}$$

The subscript "clr" denotes clear cluster quantities, and the subscript "cld" denotes cloudy cluster quantities. $\mathcal{G}\left(\boldsymbol{\psi};\ \overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{clr}}}, \boldsymbol{P}^{\boldsymbol{f}}_{\text{clr}}\right)$ denotes the clear cluster's Gaussian kernel with mean state $\overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{clr}}}$ and covariance matrix $\boldsymbol{P}^{\boldsymbol{f}}_{\text{clr}}$. Similarly, $\mathcal{G}\left(\boldsymbol{\psi};\ \overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{cld}}}, \boldsymbol{P}^{\boldsymbol{f}}_{\text{cld}}\right)$ denotes the cloudy cluster's Gaussian kernel with mean state $\overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{cld}}}$ and covariance matrix $\boldsymbol{P}^{\boldsymbol{f}}_{\text{cld}}$. The scalar quantities w^f_{clr} and w^f_{cld} are the respective weights of the clear and cloudy Gaussian kernels. Note that

$$w^f_{\text{clr}} + w^f_{\text{cld}} = 1, \qquad w^f_{\text{clr}} \geq 0, \quad \text{and,} \quad w^f_{\text{cld}} \geq 0.$$

The various parameters in Eq. (3) can be estimated by the procedure described in CAC20 or in the supporting information.

Upon assimilating an observation y^o with Gaussian observation error, the BGEnKF produces an ensemble that is consistent with the analysis pdf

$$p(\boldsymbol{\psi}|y^o) = w^a_{\text{clr}}\, \mathcal{G}\left(\boldsymbol{\psi};\ \overline{\boldsymbol{\psi}^{\boldsymbol{a}}_{\text{clr}}}, \boldsymbol{P}^{\boldsymbol{a}}_{\text{clr}}\right) + w^a_{\text{cld}}\, \mathcal{G}\left(\boldsymbol{\psi};\ \overline{\boldsymbol{\psi}^{\boldsymbol{a}}_{\text{cld}}}, \boldsymbol{P}^{\boldsymbol{a}}_{\text{cld}}\right). \tag{4}$$

Here, w^a_{clr} and w^a_{cld} are the respective analysis weights of clear and cloudy Gaussian kernels, $\overline{\boldsymbol{\psi}^{\boldsymbol{a}}_{\text{clr}}}$ and $\overline{\boldsymbol{\psi}^{\boldsymbol{a}}_{\text{cld}}}$ are the respective analysis means of the clear and cloudy Gaussian kernels, and $\boldsymbol{P}^{\boldsymbol{a}}_{\text{clr}}$ and $\boldsymbol{P}^{\boldsymbol{a}}_{\text{cld}}$ are the respective analysis covariances of the clear and cloudy Gaussian kernels. See CAC20 [or the supporting information] for the equations relating the analysis pdf's parameters to the forecast pdf's parameters.

The BGEnKF converts a forecast ensemble into an analysis ensemble through a three-stage process [illustrated in Figure 1]. First, two EnKF procedures are executed [Figure 1(a)]: once for clear members using clear forecast statistics $\left(\overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{clr}}}, \boldsymbol{P}^{\boldsymbol{f}}_{\text{clr}}\right)$, and a second time for cloudy members using cloudy forecast statistics $\left(\overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{cld}}}, \boldsymbol{P}^{\boldsymbol{f}}_{\text{cld}}\right)$. Afterwards, to reflect the update to the bi-Gaussian pdf weights, clear members will be replaced with cloudy members, or *vice versa*. For example, if the BGEnKF increased the weight on the clear Gaussian distribution (*i.e.*, $w^a_{\text{clr}} > w^f_{\text{clr}}$ and $w^a_{\text{cld}} < w^f_{\text{cld}}$), some cloudy members will be replaced with clear members. This is achieved by deleting some cloudy members [Figure 1(b)] and replacing the deleted members with resampled clear members [Figure 1(c)]. Once these three stages are completed, the ensemble obeys Eq. (4). See the supporting information for a detailed description of these three stages.

2.3 Revised extended state formulation for better scalable parallelism

The most important modification to the original CAC20 BGEnKF lies in the definition of $\boldsymbol{\psi}$. The CAC20 BGEnKF's $\boldsymbol{\psi}$ only contains $\boldsymbol{x}$ and a single observation. As such, the CAC20 BGEnKF algorithm is a sequential algorithm that scales inefficiently with parallelization on high latency clusters (Anderson & Collins, 2007). For more efficient scaling with parallelization, this study's $\boldsymbol{\psi}$ contains all of the information necessary to assimilate all observations [*i.e.*, Eq. (2); Anderson and Collins (2007)].

Since the definition of $\boldsymbol{\psi}$ has been modified, we will redefine our forecast ensemble. Supposing an ensemble size of N_E, the forecast $\boldsymbol{\psi}$ ensemble is constructed by evaluating

$$\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{n}} \equiv \begin{bmatrix} \boldsymbol{x}^{\boldsymbol{f}}_{\boldsymbol{n}} \\ \boldsymbol{h}(\boldsymbol{x}^{\boldsymbol{f}}_{\boldsymbol{n}}) \\ \boldsymbol{\xi}\left(\boldsymbol{x}^{\boldsymbol{f}}_{\boldsymbol{n}}\right) \end{bmatrix} \quad \forall \ n = 1, 2, \ldots, N_E \tag{5}$$

where $\boldsymbol{\psi}_n^f$ is the $\boldsymbol{\psi}$ of the $n-$th forecast member, and $\boldsymbol{x}_n^f$ is the $\boldsymbol{x}$ of the same forecast member.

The revised formulation enhances the scalability of the BGEnKF by avoiding evaluations of $\boldsymbol{h}(\boldsymbol{x})$ and $\boldsymbol{\xi}(\boldsymbol{x})$ at each iteration of the serial assimilation loop. This is because such evaluations may require costly inter-process communications. The removal of such evaluations is achieved through two modifications to the CAC20 BGEnKF. First, the assimilation of an observation uses the BGEnKF update equations (see CAC20 or the supporting information) to update all model state elements, all simulated observation state elements and all ξ elements in the forecast ensemble. The CAC20 BGEnKF, in contrast, updates all model state elements and only a single simulated observation state element. This difference in updates leads to a second modification: to assimilate the $m-$th observation, instead of evaluating $\boldsymbol{h}(\boldsymbol{x})$ and $\boldsymbol{\xi}(\boldsymbol{x})$, this study's BGEnKF only needs to read the corresponding simulated observation and the ξ values from $\boldsymbol{\psi}$.

2.4 Revised expanding cluster resampling procedure

The other major change to the CAC20 BGEnKF lies in the resampling matrix $\boldsymbol{T}$. $\boldsymbol{T}$ is used to resample the Gaussian kernel that better agrees with the assimilated observation, thus representing the increase in the weight of this kernel. The CAC20 BGEnKF uses a stochastic procedure to construct $\boldsymbol{T}$ [see Eq. (18) and Appendix B of CAC20]. Unfortunately, because random number generators are involved, the analysis ensemble generated on one computing cluster cannot be easily replicated on another computing cluster.

To ensure the replicability of the BGEnKF's analysis ensembles, we replaced the stochastic component of the CAC20 BGEnKF's $\boldsymbol{T}$ [$\boldsymbol{W}$ in the Appendix B of Chan, Anderson, and Chen (2020)] with a deterministic one. Supposing that we want to add N_{new} cloudy members to the ensemble to represent an increased weight of the cloudy Gaussian distribution, the new deterministic $\boldsymbol{W}$ is defined as

$$\boldsymbol{W} \equiv \begin{bmatrix} \boldsymbol{I}_{N^*_{\text{new}}} & \mathbf{0}_{N^*_{\text{new}} \times \left(N_{\text{new}} - N^*_{\text{new}}\right)} \end{bmatrix} - \frac{1}{N_{\text{new}}} \mathbf{1}_{N^*_{\text{new}} \times N_{\text{new}}} \tag{6}$$

where

$$N^*_{\text{new}} \equiv \begin{cases} N_{\text{new}} - 1 & \forall \; N_{\text{new}} \leq N_{\text{pre}} \\ \\ N_{\text{pre}} & \text{otherwise} \end{cases} ,$$

and N_{pre} is the number of cloudy members at the start of the resampling procedure. Furthermore, $\boldsymbol{I}_{N^*_{\text{new}}}$ is an $N^*_{\text{new}} \times N^*_{\text{new}}$ identity matrix, $\mathbf{0}_{N^*_{\text{new}} \times \left(N_{\text{new}} - N^*_{\text{new}}\right)}$ is an $N^*_{\text{new}} \times \left(N_{\text{new}} - N^*_{\text{new}}\right)$ matrix of zeros, and $\mathbf{1}_{N^*_{\text{new}} \times N_{\text{new}}}$ is an $N^*_{\text{new}} \times N_{\text{new}}$ matrix of ones. Note that Eq. (6) is also applied in the situation where N_{new} clear members are being added to the ensemble. A detailed description of the revised resampling procedure is provided in the supporting information.

Note that an interesting property of Eq. (6) is that the resulting $\mathbf{T}$ is a mostly diagonal matrix. Specifically, nearly all of the off-diagonal elements in $\mathbf{T}$ are either zero or much smaller than the diagonal elements (not shown). As a result, the resampled members are essentially copies of the pre-resampling members, plus some small perturbation. The CAC20 stochastic $\mathbf{W}$ formulation does not have this property. Future work can investigate how the BGEnKF's behavior changes with different $\mathbf{W}$ formulations.

2.5 Heuristic measures

2.5.1 Localization

The BGEnKF is likely more susceptible to sampling noise than the EnKF because the sample size used to estimate each cluster's mean state and Kalman gain are smaller than the sample size used to estimate the mean state and covariance matrix of the entire ensemble. As such, we employ two heuristic measures that are similar to those of CAC20. First, we spatially localize the BGEnKF analysis increment using the Gaspari-Cohn fifth order polynomial [GC99; Gaspari and Cohn (1999)]. If $\boldsymbol{\rho}$ represents a vector of GC99 localization factors, we construct the localized updated extended state vector for member n via

$$\boldsymbol{\psi}_n^a \leftarrow \boldsymbol{\rho} \circ \left(\boldsymbol{\psi}_n^a - \boldsymbol{\psi}_n^f \right) + \boldsymbol{\psi}_n^f \tag{7}$$

where $\circ$ represents element-wise multiplication. In the cases where either $w^f_{\mathrm{clr}} = 1$ or $w^f_{\mathrm{cld}} = 1$ (*i.e.,* the bi-Gaussian prior p.d.f. turns Gaussian), this localization method is identical to Kalman gain localization [*e.g.*, Anderson et al. (2009), Meng and Zhang (2008), Whitaker et al. (2008), P. L. Houtekamer and Zhang (2016)].

Note that this localization method [Eq. (7)] localizes the impacts of replacing clear members with cloudy members (or *vice versa*). As an example, suppose the BGEnKF replaces a cloudy forecast member with a clear analysis member. The localization process [Eq. (7)] first computes the difference between the cloudy forecast member and the clear analysis member (*i.e.*, the member's change due to the BGEnKF). This difference is then localized and applied to the cloudy forecast member. The resulting member follows the clear analysis member at the observation site and becomes increasingly like the cloudy forecast member with increasing distance from the observation site. Future work can examine other approaches to localize the impacts of deleting and replacing ensemble members.

2.5.2 Handling overly small clusters

The second heuristic sampling error mitigation measure is to switch from using the BGEnKF to using the EnKF whenever the pre-resampling expanding cluster is too small ($N_{\mathrm{pre}} < 0.8N_E$), or whenever any cluster is too small (less than $0.1N_E$). A similar heuristic measure is used in CAC20.

2.5.3 Mitigating unphysical weight updates

Another issue specific to the BGEnKF is its occasional tendency to generate unphysical weight updates. Specifically, the BGEnKF occasionally expands the clear cluster when a cloudy observation is assimilated, and *vice versa*. This is because the BGEnKF does not explicitly consider whether an observation is clear or cloudy when assimilating it.

The BGEnKF is automatically switched to the EnKF whenever an unphysical weight update is detected. To do so, we first identify the whether the observation to be assimilated is definitively clear or cloudy. In the case of Window-BT values over tropical ocean, observation values warmer than 290 K are definitively clear, and observation values cooler than 280 K are definitively cloudy. If the observation is definitively clear, but the cloudy cluster is expanded by the BGEnKF, or *vice versa*, the BGEnKF will switch over to the EnKF.

3 Materials and methods

3.1 Description of October 2011 tropical convection case

The BGEnKF was tested against the EnKF using a case of tropical convection over the equatorial Indian Ocean during the October 2011 MJO. This case is chosen because it can be reasonably replicated by regional WRF models (S. Wang et al., 2015; F. Zhang et al., 2017; Ying & Zhang, 2017; Fu et al., 2017; X. Chen, Pauluis, & Zhang, 2018; X. Chen & Zhang, 2019; Ying & Zhang, 2018; Chan, Zhang, et al., 2020).

Our experiments are conducted over a three day period during the onset of this MJO event (15 October 2011 to 18 October 2011). Two persistent regions of enhanced convection (henceforth, "convective regions") are observed in the 4-km Global IR Dataset of Janowiak et al. (2001) [henceforth, the MERG dataset]. The first convective region (blue rectangle) occurs between 60 °E and 75 °E and persists beyond the three-day period. Westward propa-

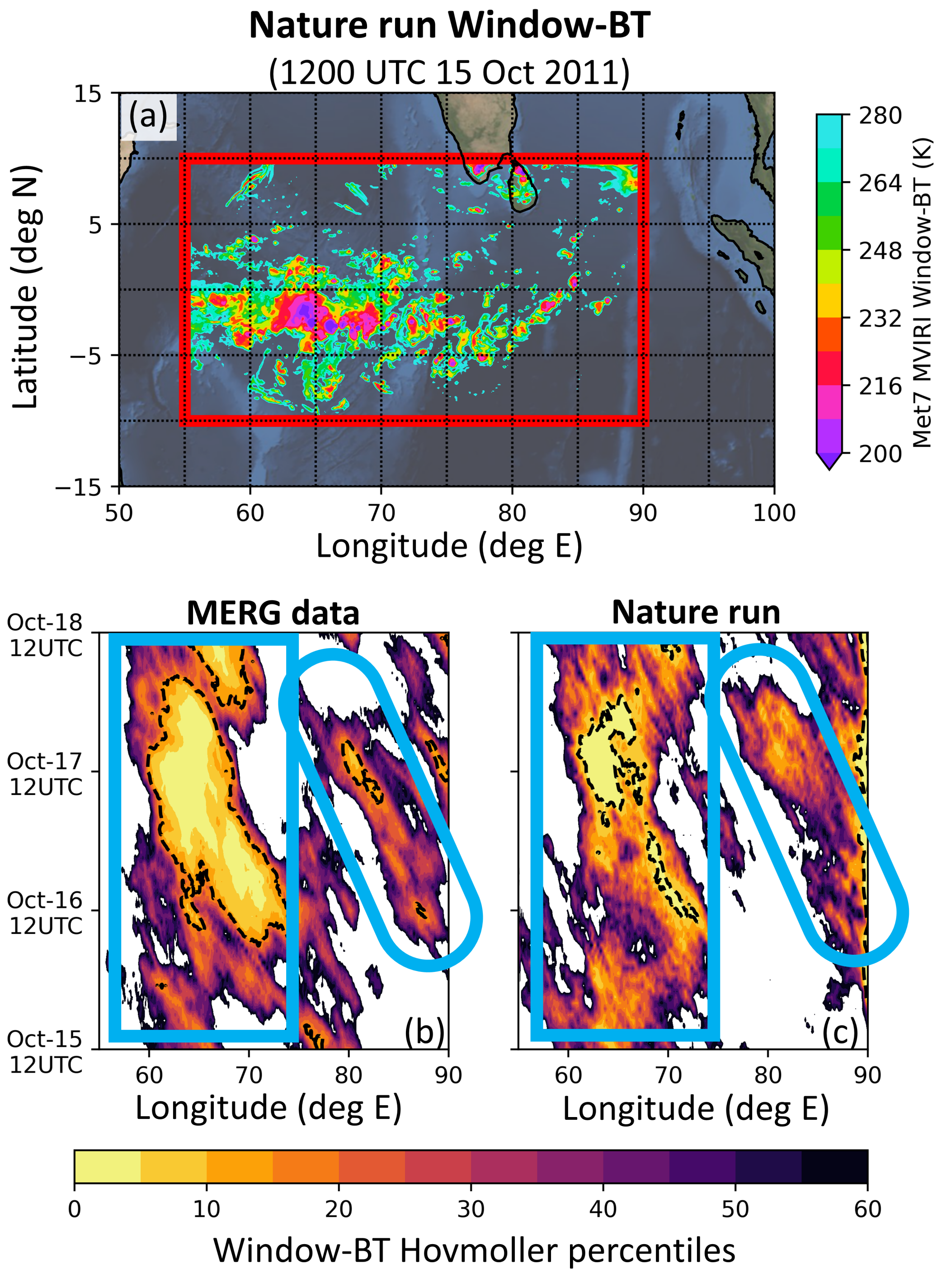


Figure 2. (a) Plot of our OSSE domain overlaid with the nature run's simulated Window-BT field at 1200 UTC on 15th October 2011. The red box in panel (a) indicates our study domain. Also shown are longitude-time diagrams for the MERG dataset (b) and nature run (c). In panels (b) and (c), the shadings indicate Window-BT Hovmoller percentile values. These Window-BT Hovmoller percentile values are constructed by first averaging Window-BT values between between 10°S and 10°N at every hour to produce a time-longitude array of latitudinally-averaged Window-BT values. These arrays are then converted into percentiles before producing the longitude-time percentile values. Note that the dashed black contours in (b) and (c) indicate areas where the time-longitude arrays of latitudinally-averaged Window-BT values are below 260 K.

gation is observed in some of the clouds in this region, most notably between 1200 UTC on 16 October and 0000 UTC on 18 October. The second convective region (blue oval) appears on the eastern edge of the study domain at 1200 UTC on 16th October and exhibits a westward propagation that is similar to that of the first system. We will later assess our OSSE's nature run simulation by checking the nature run against these two convective regions.

3.2 Setup of WRF model

The Advanced Research version of the WRF model (WRF-ARW) version 3.8 (Skamarock et al., 2008) is used in this study. Following Chan, Zhang, et al. (2020), we construct a 432×243 WRF domain over the study domain [red box in Figure 2(a)] with 9-km horizontal grid spacing and 45 model levels. The bottommost 9 levels are within the lowest 1-km of the atmosphere and the pressure level at the top of the domain is set to 20 hPa. The WRF integration time step is set to 20 seconds.

Our WRF model setup uses the following parameterization schemes. Cloud microphysical processes are handled by the WRF double-moment 6-class scheme (WDM6) proposed by Lim and Hong (2010). The updated Goddard shortwave scheme of Chou and Suarez (1999) and the Rapid Radiative Transfer Model (Global Circulation Model version; RRMTG) longwave scheme of Iacono et al. (2008) are used to parameterize radiative processes. The unified Noah land surface physics scheme (F. Chen & Dudhia, 2001) handles surface process and the Yonsei University (YSU) boundary layer scheme (Hong et al., 2006) is employed. No cumulus parameterization is employed because many studies have demonstrated that the 9-km grid spacing is sufficient to resolve tropical mesoscale convective systems (MCS) over the region (S. Wang et al., 2015; Ying & Zhang, 2017, 2018; F. Zhang et al., 2017; X. Chen, Pauluis, & Zhang, 2018; X. Chen, Pauluis, Leung, & Zhang, 2018; X. Chen & Zhang, 2019; X. Chen et al., 2020; Chan, Zhang, et al., 2020; Chan & Chen, 2021; X. Chen, Leung, Feng, & Song, 2021; X. Chen, Leung, Feng, Song, & Yang, 2021; X. Chen et al., 2022).

3.3 Setup of WRF ensemble and nature run

This study's WRF ensemble and nature run are constructed by combining two datasets from the European Center for Medium-Range Forecasts (ECMWF): the ECMWF Reanalysis Version 5 [ERA5; Hersbach et al. (2020)] and the ECMWF's 50-member perturbed forecasts (Swinbank et al., 2016). The ERA5 dataset is downloaded for every hour between 0000 UTC on 15 October to 1800 UTC on 18 October from the ECMWF's Climate Data Store (CDS). The ECMWF's perturbed forecasts are produced as part of The Observing System Research and Predictability Experiment (THORPEX) Interactive Grand Global Ensemble [TIGGE; Swinbank et al. (2016)] and is downloaded for 0000 UTC on 15 October from the ECMWF's Meteorological Archival and Retrieval System (MARS).

The ERA5 and ECMWF's 50-member perturbed forecasts (TIGGE ensemble) are processed using the WRF Preprocessing System and WRF's real data processor (`real.exe`) to produce a set of 51 WRF initial conditions files. Note that the ERA5 is used to fill in the data missing from the TIGGE ensemble above 200 hPa. The 50 WRF initial conditions from the TIGGE ensemble are then recentered on the ERA5 WRF initial condition file. The end result is a 51-member ensemble of WRF initial conditions, where member 51 is based entirely on the ERA5 (*i.e.*, the 51-st ensemble perturbation is zero). Note that this 51-st member is not used to initialize the nature run. One of the other initial conditions is used to initialize the nature run.

The lower and lateral boundary conditions used in this study are based entirely on the hourly ERA5 dataset (*i.e.*, the boundary conditions are unperturbed). While perturbed boundary conditions can increase the ensemble spread, the ensemble spread is usually reasonable even with unperturbed boundary conditions (not shown). Furthermore, as a first approach to studying the potential impacts of the BGEnKF in a high-order weather model setting, we want the differences between the nature run (described later) and the OSSE ensemble to be entirely due to differences in the initial conditions. Future work can extend this study to situations with perturbed boundary conditions.

We desire a nature run that is roughly one ensemble standard deviation from our experiments' ensembles. To select an appropriate initial condition file for such a nature run, we first integrate the 51 members forward for 12 hours (from 0000 UTC to 1200 UTC on 15 October 2011). This integration is performed to generate flow-dependent ensemble statistics that are consistent with the WRF model. After the 12-hour integration, we compute the following perturbation length metric (D^2) for each of the 51 ensemble members

$$D^2(n) \equiv \frac{1}{N_S\ N_i\ N_j} \sum_{v \in S} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \left(\frac{\Lambda\left(i,j,v,n\right) - \left\langle \Lambda\left(i,j,v\right)\right\rangle_n}{\sigma_{i,j,v}} \right)^2 . \tag{8}$$

$\Lambda\left(i,j,v,n\right)$ here is the value of a WRF-derived field v at horizontal index location (i,j) for ensemble member n. Furthermore, $\left\langle \Lambda\left(i,j,v\right)\right\rangle_n$ is the ensemble average of $\Lambda\left(i,j,v,n\right)$, and $\sigma_{i,j,v}$ is the ensemble standard deviation of $\Lambda\left(i,j,v,n\right)$. This means that the expression in the parentheses of Eq. (8) is the spread-normalized displacement of ensemble member n from the ensemble mean at location (i,j) for variable field v. The set S contains three 2D variables (precipitable water, column mass, and mass-integrated kinetic energy) and N_s is the size of the set S (*i.e.,* $N_S = 3$). Furthermore, N_i ($\equiv$ 432) is the number of east-west grid points and N_j ($\equiv$ 243) is the number of north-south grid points. The metric in Eq. (8) can thus be interpreted as being proportional to the spread-normalized Euclidean length of the n-th ensemble perturbation. As such, a D^2 value of unity means that the ensemble member is generally displaced from the ensemble mean by 1 standard deviation.

We define our nature run member to be the member whose D^2 value is closest to unity at 1200 UTC on 15 October. As a result, the nature run is based on member 10 of the TIGGE ensemble. The remaining 50 WRF members will be used for our cycling OSSE DA experiments.

3.4 Sanity check of nature run

Before proceeding, the nature run is checked by comparing it against the MERG dataset. Figure 2(b & c) shows longitude-time diagrams of the Window-BT percentiles from the MERG dataset and our nature run. The construction of these percentiles is explained in section 3.1 and in the caption of Figure 2.

We have opted to display the Window-BT percentiles instead of the Window-BT values because the WRF model tends to under produce clouds (*i.e.*, when compared to satellite observations, the nature run Window-BTs are warm biased). This is illustrated by the dashed contours in Figure 2(b & c), which highlights areas where the latitudinally-averaged values of Window-BT were cooler than 260 K. These areas are substantially larger in the MERG data than in the nature run, meaning that the nature run under produced clouds. Since converting the Window-BT values to percentile values weakens the visual interference from the cloud biases, we have opted to display the Window-BT percentiles over the Window-BT values.

Figure 2(c) indicates that the nature run also exhibits the two persistent convective regions observed in the MERG dataset (see section 3.1). These persistent convective regions are indicated by the blue rectangle and blue oval in Figure 2(c). Not only did the nature run's two persistent convective regions occur in locations and times similar to those of the MERG dataset (Figure 2(b)), these nature run regions also exhibit westward propagation patterns similar to those of the MERG dataset. As such, the nature run simulation reasonably replicates the anomalous convective behavior of the real atmosphere between 15 October to 18 October 2011.

3.5 Setup of DA experiments to test the BGEnKF

To test the BGEnKF, three 50-member ensemble experiments are conducted. All three experiments start at 1200 UTC on 15 October and terminate at 1200 UTC on 18 October, with hourly DA cycling (73 cycles in total). The construction and spin-up of these 50 members are described in section 3.3.

In the first experiment, no observations are assimilated (henceforth, NoDA experiment). The NoDA experiment serves as a baseline for comparing the performance of the EnKF and BGEnKF, and to measure imbalances induced by DA.

The other two experiments are the EnKF and BGEnKF experiments. The only difference between the EnKF and BGEnKF experiments is in the DA algorithm employed. The EnKF experiment will assimilate observations using the PSU-EnKF's (Meng & Zhang, 2007, 2008) default EnKF algorithm, and the BGEnKF experiment will assimilate observations using a new implementation of the BGEnKF into the PSU-EnKF. Note that both the EnKF and the BGEnKF are implemented into the PSU-EnKF using the high-latency strategy proposed by Anderson and Collins (2007).

As a first approach to testing the BGEnKF, only synthetic *Meteorological Satellite 7* Meteosat Visible Infra-Red Imager (MVIRI) Window-BT observations will be assimilated. Future work can investigate if our findings can be extended to situations where an entire suite of operationally-assimilated observations and observations from different infrared channels are assimilated.

The synthetic Window-BT observations are constructed by first running the Community Radiative Transfer Model (CRTM) release 2.3.0 on the nature run (see sections 3.3 and 3.4). The nature run's Window-BT values are then thinned to a horizontal spacing of 27-km (~11,500 observations per DA cycle). White noise with a standard deviation of 3 K is then added to the thinned nature run Window-BT values to simulate instrument noise, thus constructing the synthetic observations. Note that the observation errors are likely to be correlated in reality. This means our use of white noise is an imperfect approximation to actual observation errors. Future work can investigate if our results can be extended to situations with correlated Window-BT observation errors.

Common heuristic strategies are employed to assimilate the Window-BT observations. To limit the impact of sampling errors, horizontal localization is applied using the Gaspari-Cohn fifth-order polynomial (Gaspari & Cohn, 1999) with a 100-km radius of influence (P. L. Houtekamer & Mitchell, 2001; Greybush et al., 2011; P. L. Houtekamer & Zhang, 2016). No vertical localization is employed. We also employ the Adaptive Observation Error Inflation scheme (AOEI) of Minamide and Zhang (2017) to limit the deleterious increments that can result from clear-cloudy disagreements between the prior and observations (F. Zhang et al., 2016; Minamide & Zhang, 2017). To mitigate the tendency for ensemble under-dispersion to occur when the ensemble is clear and the observation is cloudy, the Adaptive Background Error Inflation scheme (ABEI) of Minamide and Zhang (2019) is applied. We also employ 80% relaxation to prior perturbations (RTPP) to maintain ensemble dispersion (F. Zhang et al., 2004). Similar combinations of heuristic strategies are commonly seen in the EnKF-based DA of infrared radiance observations (F. Zhang et al., 2016; Minamide & Zhang, 2018; Chan, Zhang, et al., 2020; Y. Zhang et al., 2019; Chan & Chen, 2021; Y. Zhang et al., 2021).

Aside from these common strategies, we also restrict the BGEnKF/EnKF from updating the domain-averaged specific humidity (QVAPOR) using Window-BT observations. Without this measure, both the BGEnKF and the EnKF experience filter divergence that is related to DA-induced dry biases within 48 hours of cycling. These dry biases are likely induced by the ensemble's tendency to be overly cloudy. The dry biases in the EnKF experiment are likely partly because of the EnKF's inability to handle clear and cloudy members separately (see section 4.3). As for the BGEnKF experiment, the dry bias can be explained by the fact that the BGEnKF algorithm frequently switches over to the EnKF algorithm (see section 4.2). Note that the BGEnKF generated smaller dry biases than the EnKF (not shown).

To prevent filter divergence due to DA-induced dry biases, we replace the 3D posterior mean QVAPOR field ($\overline{q_v^a}$) with the following modified mean QVAPOR field ($\overline{q_v^*}$):

$$\overline{q_v^*(i,j,k)} \equiv \overline{q^a(i,j,k)} - \frac{1}{N_i\, N_j} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \left\{ \overline{q_v^a(i,j,k)} - \overline{q_v^f(i,j,k)} \right\}. \tag{9}$$

Here, (i, j, k) refer to the west-east, south-north and bottom-top indices of the 3D QVAPOR fields and $\overline{q_v^f}$ refers to the 3D prior mean QVAPOR field.

3.6 Execution wall-time of the BGEnKF

Before proceeding, we should compare the execution wall-time of the BGEnKF and the EnKF. The BGEnKF algorithm took ~30 seconds to assimilate ~11,500 observations using 228 Intel Knight's Landing computer cores [distributed across 7 computational nodes on the National Energy Research Scientific Computing Center (NERSC) Cori supercomputer; each core has a clock rate of 1.4 GHz]. Assimilating the same observations via an EnKF algorithm took ~20 seconds of wall-time. For a fair comparison, this EnKF algorithm used the exact same code structure and computing resources, but with the cluster transfer and auxiliary variable update steps disabled. In other words, the BGEnKF used ~10 seconds more wall-time than the EnKF.

This ~10-second difference should be assessed in the context of the wall-time for the entire PSU-EnKF executable. The other components of the PSU-EnKF took ~100 seconds to execute. As such, the BGEnKF only added ~10% wall-time to the entire PSU-EnKF executable. The BGEnKF algorithm is thus likely affordable for research and operational groups that are already running serially-assimilating EnKFs [*e.g.*, Anderson et al. (2009)].

4 Perfect model WRF OSSE results

In the discussions to follow, we will be showing plots of normalized root-mean-square errors (nRMSEs) and normalized biases as functions of time and model level. The normalization is necessary for the ease of visualization, and uses the root-mean-square errors (RMSEs) of the NoDA experiment. The EnKF experiment's nRMSE at model level k and date t is defined as

$$\text{EnKF nRMSE}\,(k,t) \equiv \frac{\text{EnKF RMSE}\,(k,t)}{\text{NoDA RMSE}\,(k,t)} \tag{10}$$

and likewise for that of the BGEnKF and NoDA experiments (the NoDA's nRMSE values are always 1). Note that if a filter results in nRMSEs > 1.0, the assimilation of Window-BT via this filter degraded the ensemble with respect to the NoDA experiment. The reverse is true for nRMSEs < 1.0. We also define the normalized bias of the EnKF experiment to be

$$\text{EnKF normalized bias}\,(k,t) \equiv \frac{\text{EnKF bias}\,(k,t)}{\text{NoDA RMSE}\,(k,t)}, \tag{11}$$

and likewise for the BGEnKF and NoDA experiments. These biases are computed by subtracting the nature run fields from the forecast ensemble mean fields.

The nRMSEs and normalized biases are examined for six variable fields: the zonal wind velocity component field (U), the meridional wind velocity component field (V), the temperature field (T), the QVAPOR field (Q), the Window-BT field, and the upper tropospheric infrared water vapor channel brightness temperature field (WV-BT; central wavelength of 6.2 μm). The nRMSEs are plotted in Figures 3 and 5(a & b) and the normalized biases are plotted in Figures 4 and 5(c & d). All quantities are computed using forecast statistics.

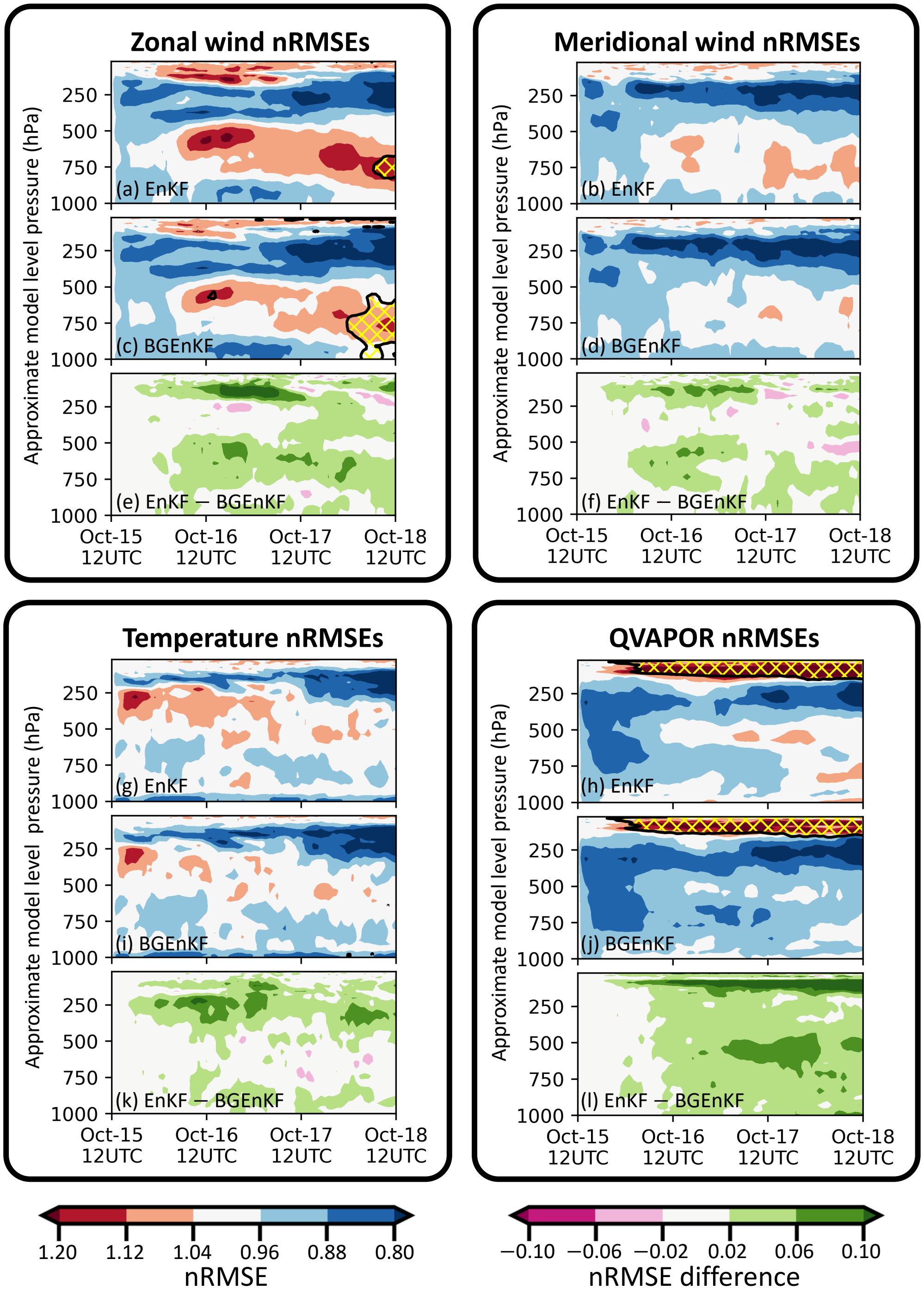


Figure 3. Plots of various prior ensemble statistics as functions of time and model level. For ease of interpretation, the model levels are displayed in terms of their approximate pressure levels (estimated using the definition of eta levels in WRF and assuming a surface pressure of 1000 hPa). The shadings indicate the NoDA-normalized RMSEs [nRMSEs; defined in Eq. (10)] for the EnKF (a, b, g & h) and BGEnKF (c, d, i & j) experiments, as well as the nRMSE differences between the EnKF and BGEnKF experiments (e, f, k & l). The nRMSEs and nRMSE differences are shown for the U field (a, c & e), V field (b, d & f), T field (g, i & k), and Q field (h, j & l). The areas outlined with a black contour and filled with yellow hatching have consistency ratios (spread/error) less than 0.75. Note that all displayed statistics are forecast statistics.

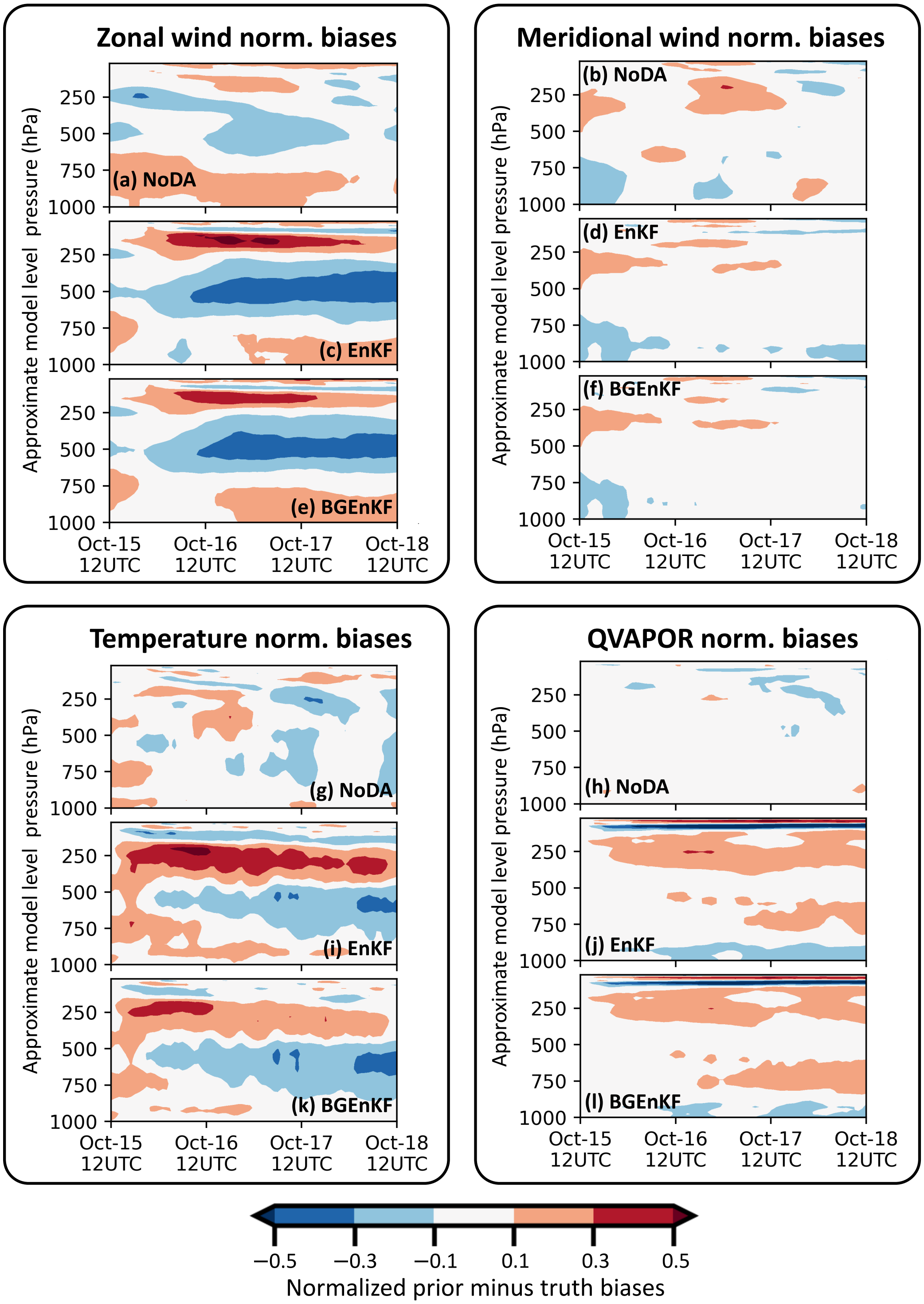


Figure 4. Plots of various prior ensemble normalized biases as functions of time and model level. These normalized biases are displayed for the U field (a, c & e), V field (b, d & f), T field (g, i & k), and Q field (h, j & l), for the NoDA (a, b, g & h), EnKF (c, d, i & j) and BGEnKF (e, f, k & l) experiments. Similar to Figure 3, the model levels are displayed in terms of approximate pressure levels. See the Eq. (11) for the definition of the normalized biases.

4.1 On differences in the BGEnKF's and the EnKF's performances during DA cycling

The nRMSEs and normalized biases of the BGEnKF experiment are generally better than or comparable to those of the EnKF experiment (Figures 3 to 5). For the U, V, T and Q fields, subtracting the BGEnKF's nRMSEs from the EnKF's nRMSEs generally results in positive values [Figure 3(e, f, k & l)]. The BGEnKF experiment also has better WV-BT nRMSEs than the EnKF experiment [Figure 5(b)]. The BGEnKF experiment also has smaller biases than the EnKF experiment in several places: the 100 hPa U field [Figure 4(c & e)], the 400–100 hPa T field [Figure 4(i & k)], the Window-BT field [Figure 5(e)], and WV-BT field [Figure 5(f)]. Otherwise, the BGEnKF and EnKF experiments have similar bias values. These results suggest that the BGEnKF is more suitable for assimilating all-sky Window-BT than the EnKF.

The BGEnKF's performance advantages over the EnKF can be separated into two types. In the first type, the BGEnKF generates larger improvements than the EnKF (*i.e.*, BGEnKF nRMSEs < EnKF nRMSEs < NoDA nRMSEs). This type of performance advantage occurs in multiple places (Figures 3 and 5): 1) the 800 hPa to 1000 hPa U field nRMSEs during the first 56 cycles, 2) the 100 hPa to 500 hPa U field nRMSEs during the last 36 DA cycles, 3) the near surface and ~250 hPa V field nRMSEs from 0000 UTC on 16th October to 0000 UTC on 17th October, 4) between 100 hPa to 300 hPa in the T field nRMSEs for most cycles, 5) between 250 to 600 hPa in the Q field nRMSEs for most cycles, and in the WV-BT nRMSEs for most DA cycles after 0000 UTC on 16th October. These differences are likely due to the BGEnKF's ability to handle mixture statistics, and suggest that the BGEnKF is more suitable for assimilating Window-BT than the EnKF.

The BGEnKF experiment's second type of performance advantage over the EnKF experiment is when the BGEnKF introduces milder degradations than the EnKF (*i.e.*, NoDA nRMSEs < BGEnKF nRMSEs < EnKF nRMSEs). In terms of nRMSEs (Figure 3), such situations are noticeable at the 100 hPa tropopause level and 500–700 hPa levels for the U and V fields, at the 200–500 hPa model levels for the T field, and at the 100 hPa level for the Q field. Such situations are also noticeable in the normalized biases of the ~100 hPa U field, the 100–400 hPa T field (Figure 4), and in the Window-BT and WV-BT fields (Figure 5). These are likely because 1) the BGEnKF can handle mixture statistics whereas the EnKF cannot, and 2) the BGEnKF experiment has smaller increments than the EnKF experiment because the BGEnKF experiment has smaller dispersion. Figure 3(a & c) shows an example of the latter: the BGEnKF U field has larger areas of low spread-to-error ratios (0.75) than the EnKF. The likely origins of the RMSE and bias degradations are discussed in section 4.2. Nonetheless, these results further support the notion that the BGEnKF is more appropriate for assimilating Window-BT observations than the EnKF.

The BGEnKF tends to result in smaller CRs than the EnKF because the BGEnKF can outright convert all clear member columns to cloudy member columns, or *vice versa*. Since clear and cloudy member columns are very different, having both types of columns present at the same time boosts the ensemble spread. If all clear member columns are converted to cloudy member columns, or *vice versa*, large perturbations relative to the ensemble mean are replaced with smaller perturbations. This replacement results in reduced ensemble dispersion. Since the EnKF lacks this mechanism of ensemble spread removal, the BGEnKF can remove more ensemble spread than the EnKF, thus resulting in smaller CRs than the EnKF. Future work can investigate if stronger inflation schemes are more appropriate for the BGEnKF.

Note that there are occasional situations where the EnKF outperforms the BGEnKF. For instance, at around 0000 UTC on 17th October the BGEnKF's U nRMSEs are slightly higher than the EnKF at 250 hPa (Figure 3(e)). Other examples include the T nRMSEs around 1200 UTC on 17th October (Figure 3). Nonetheless, if we integrate the forecast ensembles'

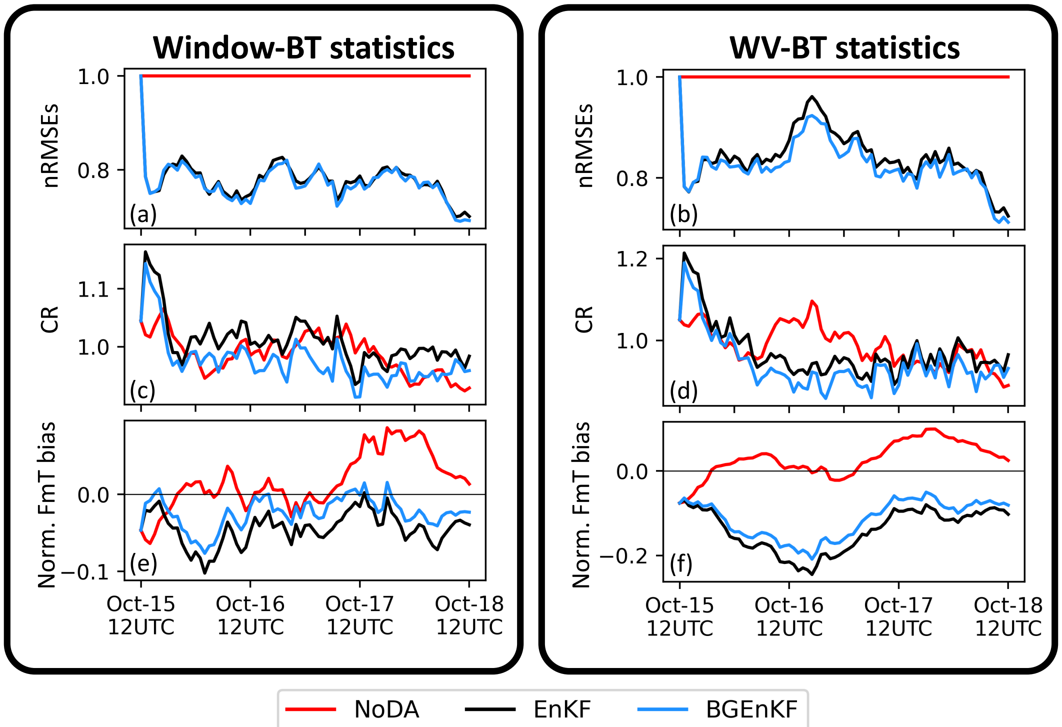


Figure 5. Time-series showing the performance statistics of the three experiments' prior ensembles in terms of Window-BT (a, c & e) and WV-BT (b, d & f). The definitions of nRMSEs (a & b) and normalized prior minus truth (Norm. FmT bias; e & f) are the same as in Figures 5 to 8. Like Figures 5 and 6, the consistency ratio (CR; c & d) here is defined as the ratio of spread to error.

nRMSEs with respect to pressure at every cycle, the resulting mass-weighted nRMSEs of the BGEnKF experiment will be lower than those of the EnKF experiment.

We have also examined day-long deterministic forecasts that are initialized from the analysis means of the EnKF and BGEnKF experiments (not shown). The BGEnKF experiment's RMSE performance advantage over the EnKF experiment persists for up to 9 hours of lead time in terms of the U, V and T fields. In terms of the 500–800 hPa Q field RMSEs, the BGEnKF experiment's RMSE advantage over the EnKF experiment persists throughout the 24 hours of integration. These results are as expected since the BGEnKF experiment has lower RMSEs than the EnKF experiment during DA cycling.

4.2 On the similar patterns observed in the performances of the BGEnKF and EnKF experiments

Though the BGEnKF experiment generally outperformed the EnKF experiment, there are common spatiotemporal patterns in their nRMSEs and normalized biases. For instance, Window-BT DA with either algorithm tends to degrade the 500–800 hPa U nRMSEs, and improve the 100–500 hPa U nRMSEs (Figure 3(a & c)). These similarities are likely because the BGEnKF frequently switches over to the EnKF. Figure 6(a) shows that the BGEnKF algorithm is only called to assimilate ∼10% of the Window-BT observations, meaning that the switching occurred for the remaining ∼90% of Window-BT observations. Future work can investigate if reducing the occurrence of such switches (*e.g.*, via weaker heuristic checks and larger ensembles) could improve the performance of the BGEnKF.

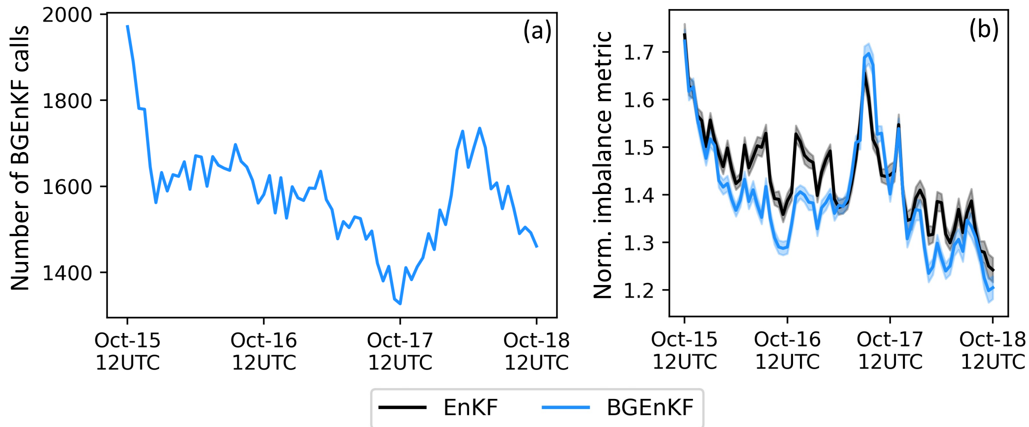


Figure 6. Plots showing the frequencies at which the two kernel BGEnKF update procedure is called in the BGEnKF experiment (a), and the normalized imbalance metric statistics for both the BGEnKF and EnKF experiments (b). For reference, 11502 IR observations are assimilated at each DA cycle. The normalized imbalance metric is defined in the text. The solid curves in (b) indicate the ensemble average of every member's normalized imbalance metric and the half-width of the shadings in (b) indicate twice the standard error of the members normalized imbalance metric.

It is notable that the BGEnKF outperforms the EnKF despite the high frequency of BGEnKF-to-EnKF switching. For instance, according to Figure 3(h, j & l), for the 24 cycles on 17th October and between 500 hPa to 700 hPa, the BGEnKF experiment has 0.06–0.1 less Q nRMSEs than the EnKF experiment. Since the EnKF experiment has Q nRMSEs of ~1 then, the BGEnKF is able to introduce a ~6–10% improvement over the EnKF. These are considerable improvements since the BGEnKF is only called on ~10% of the Window-BT observations.

Given the frequent switching from the BGEnKF to the EnKF, the worse-than-NoDA RMSEs and biases in both the EnKF and BGEnKF experiments are likely caused by the EnKF algorithm. These degradations are likely caused by 1) non-Gaussian forecast statistics, 2) sampling errors, and 3) biases that are introduced by the assimilation of Window-BT. The first factor can originate from having mixtures of clear and cloudy members. Sampling errors can also introduce errors into the analysis, particularly over regions where the ensemble correlations are weak. This factor is likely present in our experiments because no vertical localization is used in this study. Future work can investigate if vertical localization can mitigate some of the RMSE and bias degradations (Lei & Anderson, 2014; Lei & Whitaker, 2015; Lei et al., 2016, 2020). Finally, since biases are a component of RMSEs [*e.g.*, Ying and Zhang (2017), Ying and Zhang (2018), and Chan, Zhang, et al. (2020)], biases that are introduced by Window-BT DA can contribute towards worse-than-NoDA RMSEs. While the contribution of biases to worse-than-NoDA RMSEs can be easily inferred (see the next paragraphs), the contributions from the first two factors cannot be easily teased apart.

To understand the contribution of biases to the occurrence of worse-than-NoDA RMSEs (*i.e.*, nRMSEs > 1), we computed the following fraction as a function of model level and time ($f_{\text{bias}}(k,t)$). For the EnKF experiment, we defined

$$\text{EnKF's } f_{\text{bias}}(k,t) \equiv \sqrt{\frac{[\text{EnKF's biases }(k,t)]^2 - [\text{NoDA's biases }(k,t)]^2}{[\text{EnKF's RMSEs }(k,t)]^2 - [\text{NoDA's RMSEs }(k,t)]^2}}$$

and likewise for the BGEnKF experiment. f_{bias} can be interpreted as the fractional contribution of biases to the worse-than-NoDA RMSE performance.

We found that for about 25–45% of the worse-than-NoDA situations (nRMSEs > 1) in the U and T fields, the majority of the nRMSE degradation (*i.e.*, $f_{\text{bias}} \geq 0.6$) can be explained by the the introduction of biases [*i.e.*, $p\ (f_{\text{bias}} > 0.6\ |\text{nRMSE} > 1)\ \in\ (0.25, 0.45)$]. This suggests that though DA-induced biases are important contributors towards the worse-than-NoDA RMSEs of either DA filters, the net contribution coming from other factors is also important. Future work can examine separating and quantifying the relative importance of these three factors towards the worse-than-NoDA RMSEs.

4.3 On the origin of biases in the EnKF and BGEnKF experiments

We now turn our attention to the U, T, Q, Window-BT and WV-BT biases that are introduced by Window-BT DA. Since the Q analysis increments are subject to bias removal (see last paragraph of section 3.5), the Q biases will be discussed later. The U, T and WV-BT biases are likely related to 1) a cold forecast minus truth (FmT) Window-BT bias at the start of all experiments, and 2) the persistence of these FmT Window-BT biases throughout all cycles (Figure 5(e)). Item 1 is essentially the result of drawing a single member from an ensemble – it is difficult to obtain a nature run whose domain-averaged Window-BT is always the same as that of the forecast ensemble. This is supported by the fact that the NoDA experiment's FmT Window-BT biases oscillate around zero (Figure 5(e)). More interestingly, item 2 indicates an over abundance of clouds in both DA experiments. Since WV-BT is cooler in the presence of clouds, the WV-BT bias is explained by the over abundance of clouds.

To understand the origin of the persistently cold FmT Window-BT biases, we examine the analysis ensembles' Window-BT biases. Running the CRTM on the analysis ensembles of the Window-BT DA experiments reveals analysis minus truth (AmT) Window-BT normalized biases that are typically around −0.25 (not shown). These bias values are a factor of 5 larger than the FmT normalized biases of around −0.05 (Figure 5(e)). The large AmT biases suggest that Window-BT DA resulted in overly cloudy analysis ensembles. Though the time-integration of these analysis ensembles dramatically reduces the over cloudiness (the normalized biases typically go from −0.25 to −0.05), some over cloudiness likely remain. As such, the U, T, Window-BT and WV-BT biases are likely caused by the EnKF and BGEnKF experiments introducing too many clouds into the analysis ensemble.

The over introduction of clouds is likely a result of the EnKF's inability to handle clear and cloudy members separately and the strong sensitivity of Window-BTs to hydrometeors. When both clear and cloudy members are present in the forecast ensemble, the EnKF's forecast mean state will contain some amount of clouds. Suppose that the correlations between Window-BT and hydrometeor mixing ratios are negative. If Window-BT observations with either small or negative innovations are assimilated, the clouds in the EnKF's mean state will either be unaffected (for small innovations) or be increased (for negative innovations). Since the EnKF will also reduce the size of the ensemble members' perturbations, the ensemble thus contracts around a cloudy mean state. The result is that clear column forecast members gain some amount of clouds, even in situations where the innovations are close to zero. Since Window-BTs are sensitive to the presence of clouds, running the CRTM on such members will generate cold cloudy Window-BT values. This mechanism of EnKF-induced over-cloudiness warrants future investigation.

Note that the BGEnKF experiment's over-cloudiness is likely caused by the mechanism in the previous paragraph. This is because the BGEnKF algorithm frequently switches over to the EnKF (for ~90% of assimilated observations). Since the BGEnKF can handle mixtures of clear and cloudy members, with less frequent switches, the BGEnKF is likely to have smaller biases. To test this possibility, smaller sampling errors are necessary to justify less frequent switches from the BGEnKF to the EnKF. Future work can thus investigate this possibility with larger ensembles.

With regards to the Q biases, since the analysis increment cannot modify the Q biases [see Eq. (9)], these biases are induced during the forecast step of the DA procedure. We can rule out the evaporation of DA-induced spurious clouds as an important source because the hydrometeor biases injected by the increment are an order of magnitude smaller than the Q bias growth during integration (not shown). Other processes are likely causing the Q biases. Some possibilities include enhancements to the upward transport of Q from the surface and/or the latent fluxes from the ocean surface. The exact origin of these Q biases can be investigated in future work.

4.4 On dynamical imbalances

Note that the BGEnKF introduces less dynamical imbalances into the ensemble than the EnKF. To measure dynamical imbalance, we compute the root-mean-square of the second time derivative of surface pressure during the time integration phase of each DA cycle (P. Houtekamer & Mitchell, 2005; Temperton & Williamson, 1981). These derivatives are computed via centered differencing (Press & Flannery, 2010) on three consecutive snapshots of the surface pressure field. These snapshots are spaced 30-minutes apart. The resulting imbalance metric is normalized using the NoDA experiment's imbalance metric. A normalized imbalance metric value of 1 indicates that a normal amount of fast-moving gravity waves is present. A value greater than 1 indicates that a higher than normal amount of fast-moving gravity waves is present, thus indicating DA-induced imbalances.

Figure 6(b) indicates that the BGEnKF experiment generally has either statistically indistinguishable or milder imbalances than the EnKF experiment. The only exception to this trend happens between 0000 UTC to 1200 UTC on 17th October. The BGEnKF is thus likely more appropriate than the EnKF at assimilating Window-BT observations.

5 Conclusions and future work

In this study, we compare the BGEnKF against the EnKF using perfect model OSSEs with a realistic weather model (WRF) for a case of tropical convection. These OSSEs are executed using the state-of-the-art PSU-EnKF system. Our results indicate that the BGEnKF outperforms the EnKF at assimilating synthetic Window-BT observations. We observe this performance advantage in terms of the RMSEs and biases of the U, V, T, Q, Window-BT and WV-BT fields. This performance advantage is likely due to the BGEnKF's ability to handle mixtures of clear and cloudy column members. These performance advantages are achieved even though the BGEnKF s only activated for $\sim$10% of the assimilated Window-BT observations. As such, these promising results motivate future work into the BGEnKF using real data.

There are several large areas of future research for the BGEnKF. The first large area concerns refining the BGEnKF algorithm. Future work can, for instance, seek less heuristic approaches to sort the ensemble into clusters in a computationally efficient manner. One option is to combine clustering algorithms [*e.g.*, k-means (Forgy, 1965; Lloyd, 1982), support-vector machines (Cortes & Vapnik, 1995) and expectation maximization (Sondergaard & Lermusiaux, 2013b)] with dimension reduction methods [*e.g.*, Sondergaard and Lermusiaux (2013b), Reddy et al. (2020), Albarakati et al. (2021)]. Since cluster sizes, and thus sampling errors, can vary in each iteration of the serial BGEnKF loop, future work can investigate using adaptive or empirical localization methods (Anderson, 2012; Anderson & Lei, 2013; Lei & Anderson, 2014) to improve the BGEnKF's performance. Future work can also examine more sophisticated methods to regulate when the BGEnKF switches over to the EnKF (*e.g.*, using the Shapiro-Wilk test for normality).

Another area of future work is to hybridize the BGEnKF with other DA algorithms. Hybridization with kernel filters (Anderson & Anderson, 1999; Hoteit et al., 2008; Stordal et al., 2011; Hoteit et al., 2012; Liu et al., 2016; Stordal & Karlsen, 2017; Kotsuki et al., 2022) can be achieved by assigning the clear cluster's covariance to clear member kernels and likewise for the cloudy member kernels. Existing ensemble-variational hybrid DA algorithms (Hamill & Snyder, 2000; Lorenc, 2003; Buehner, 2005; X. Wang et al., 2007) can also be hybridized with the BGEnKF. For instance, the BGEnKF can replace the EnKF component of such methods. Hybridization with DA methods that employ transport methods to update ensemble members (Reich, 2012; van Leeuwen, 2011; Marzouk et al., 2017; Hu & van Leeuwen, 2021; Evensen Geir et al., 2022) is also possible. This can provide a different method to shift members between clusters, as opposed to the current deletion-resampling method. Finally, the BGEnKF can be potentially hybridized with ensemble DA methods that allow non-parametric prior distributions. Such methods include particle filters (van Leeuwen, 2009; Poterjoy, 2016; Vetra-Carvalho et al., 2018; Poterjoy et al., 2019; van Leeuwen et al., 2019), the quantile conserving ensemble filter (Anderson, 2022), and the rank histogram filter (Anderson, 2010, 2019, 2020).

Since we have only tested the BGEnKF in a perfect model WRF OSSE using Window-BT observations, future work can test the BGEnKF in increasingly realistic scenarios, with other observation types, and/or in other Earth systems. For instance, since radar reflectivity observations are sensitive to the presence and absence of precipitation, the BGEnKF can potentially be better at assimilating such observations. The performance of the BGEnKF can also be compared with other popular DA algorithms in tests that assimilate the operational suite of atmospheric in-situ and remote observations. Imperfect model OSSEs and real data tests can also be done. The BGEnKF can also be tested in other Earth system components.

This study is among the first to demonstrate the potential of the BGEnKF with a high-order weather model. Our BGEnKF is computationally efficient, scalable with parallelization, and likely straightforward to implement in existing serial EnKF DA systems. These algorithmic properties and our promising results motivate future research into developing, testing and applying the BGEnKF, or similar GMM-EnKFs, for Earth systems DA.

6 Open Research

The data and software used in this study are either publicly available or available upon request. The WRF model software can be found on the National Center for Atmospheric Research's WRF website (https://www.mmm.ucar.edu/weather-research-and-forecasting-model). Our WRF ensemble is constructed using the ECMWF TIGGE data archived on the MARS system (https://apps.ecmwf.int/datasets/data/tigge) and the ERA5 data archived on the CDS system (https://cds.climate.copernicus.eu). The MERG data product is obtained from NASA's GES DISC (https://disc.gsfc.nasa.gov/datasets/GPM_MERGIR_1/summary). We have archived this study's experiments and a copy of the Fortran 90 BGEnKF module on the Pennsylvania State University's Data Commons (http://doi.org/10.26208/XV41-7N75). The Fortran 90 source code of the PSU-EnKF system, including the implemented BGEnKF, is available upon request.

Acknowledgments

This work is supported by the Office of Naval Research (ONR) Grant N00014-18-1-2517, the National Aeronautics and Space Administration (NASA) Grant 80NSSC22K0613, the National Center for Atmospheric Research (NCAR) Advanced Study Program Graduate Visitor Program (ASP GVP), and the Water Cycle and Climate Extremes Modeling (WACCEM) project. WACCEM 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, and NCAR is sponsored by the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation. Finally, the computations in this study are performed using the Texas Advanced Computing Center (TACC) Stampede2 supercomputer and the National Energy Research Scientific Computing Center (NERSC) Cori supercomputer. NERSC is a U.S. Department of Energy Office of Science User Facility operated under Contract DE-AC02-05CH11231.

References

Albarakati, A., Budišić, M., Crocker, R., Glass-Klaiber, J., Iams, S., Maclean, J., . . . Van Vleck, E. S. (2021). Model and data reduction for data assimilation: Particle filters employing projected forecasts and data with application to a shallow water model. *Computers and Mathematics with Applications*. doi: 10.1016/j.camwa.2021.05.026

Anderson, J. L. (2003, 4). A Local Least Squares Framework for Ensemble Filtering. *Monthly Weather Review*, *131*(4), 634–642. Retrieved from http://journals.ametsoc.org/doi/10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2 doi: 10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2

Anderson, J. L. (2010, 11). A Non-Gaussian Ensemble Filter Update for Data Assimilation. *Monthly Weather Review*, *138*(11), 4186–4198. Retrieved from http://journals.ametsoc.org/doi/10.1175/2010MWR3253.1 doi: 10.1175/2010MWR3253.1

Anderson, J. L. (2012). Localization and sampling error correction in ensemble Kalman filter data assimilation. *Monthly Weather Review*, *140*(7). doi: 10.1175/MWR-D-11-00013.1

Anderson, J. L. (2019). *A nonlinear rank regression method for ensemble Kalman filter data assimilation* (Vol. 147) (No. 8). doi: 10.1175/MWR-D-18-0448.1

Anderson, J. L. (2020). A marginal adjustment rank histogram filter for non-Gaussian ensemble data assimilation. *Monthly Weather Review*, *148*(8). doi: 10.1175/MWR-D-19-0307.1

Anderson, J. L. (2022). A Quantile-Conserving Ensemble Filter Framework. Part I: Updating an Observed Variable. *Monthly Weather Review*, *150*(5). doi: 10.1175/mwr-d-21-0229.1

Anderson, J. L., & Anderson, S. L. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*. doi: 10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2

Anderson, J. L., & Collins, N. (2007, 8). Scalable Implementations of Ensemble Filter Algorithms for Data Assimilation. *Journal of Atmospheric and Oceanic Technology*, *24*(8), 1452–1463. Retrieved from https://journals.ametsoc.org/view/journals/atot/24/8/jtech2049_1.xml doi: 10.1175/JTECH2049.1

Anderson, J. L., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., & Avellano, A. (2009, 9). The data assimilation research testbed a community facility. *Bulletin of the American Meteorological Society*, *90*(9), 1283–1296. Retrieved from https://journals.ametsoc.org/doi/10.1175/2009BAMS2618.1 doi: 10.1175/2009BAMS2618.1

Anderson, J. L., & Lei, L. (2013). Empirical localization of observation impact in ensemble Kalman filters. *Monthly Weather Review*, *141*(11). doi: 10.1175/MWR-D-12-00330.1

Buehner, M. (2005). Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quarterly Journal of the Royal Meteorological Society*, *131*(607). doi: 10.1256/qj.04.15

Burgers, G., Jan van Leeuwen, P., Evensen, G., Van Leeuwen, P. J., & Evensen, G. (1998, 6). Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, *126*(6), 1719–1724. Retrieved from http://journals.ametsoc.org/doi/10.1175/1520-0493(1998)126%3C1719:ASITEK%3E2.0.CO;2 doi: 10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2

Chan, M.-Y., Anderson, J. L., & Chen, X. (2020). An efficient bi-Gaussian ensemble Kalman filter for satellite infrared radiance data assimilation. *Monthly Weather Review*. doi: 10.1175/mwr-d-20-0142.1

Chan, M.-Y., & Chen, X. (2021). Improving Analyses and Forecasts of a Tropical Squall Line using Upper Tropospheric Infrared Satellite Observations. *Advances in Atmospheric Sciences, Accepted Manuscript*. Retrieved from http://www.iapjournals

.ac.cn/aas/en/article/doi/10.1007/s00376-021-0449-8http://www.iapjournals.ac.cn/aas/en/article/doi/10.1007/s00376-021-0449-8?viewType=HTML doi: 10.1007/S00376-021-0449-8

Chan, M.-Y., Zhang, F., Chen, X., & Leung, L. R. (2020). Potential Impacts of Assimilating All-sky Satellite Infrared Radiances on Convection-Permitting Analysis and Prediction of Tropical Convection. *Monthly Weather Review*. doi: 10.1175/mwr-d-19-0343.1

Chen, F., & Dudhia, J. (2001). Coupling and advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Monthly Weather Review*. doi: 10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2

Chen, X., Leung, L. R., Feng, Z., & Song, F. (2021, 10). Crucial Role of Mesoscale Convective Systems in the Vertical Mass, Water and Energy Transports of the South Asian Summer Monsoon. *Journal of Climate*, *-1*(aop), 1–46. Retrieved from https://journals.ametsoc.org/view/journals/clim/aop/JCLI-D-21-0124.1/JCLI-D-21-0124.1.xml doi: 10.1175/JCLI-D-21-0124.1

Chen, X., Leung, L. R., Feng, Z., Song, F., & Yang, Q. (2021, 9). Mesoscale Convective Systems Dominate the Energetics of the South Asian Summer Monsoon Onset. *Geophysical Research Letters*, *48*(17), e2021GL094873. Retrieved from https://onlinelibrary.wiley.com/doi/full/10.1029/2021GL094873https://onlinelibrary.wiley.com/doi/abs/10.1029/2021GL094873https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021GL094873 doi: 10.1029/2021GL094873

Chen, X., Leung, L. R., Feng, Z., & Yang, Q. (2022, 4). Precipitation-Moisture Coupling Over Tropical Oceans: Sequential Roles of Shallow, Deep, and Mesoscale Convective Systems. *Geophysical Research Letters*, *49*(7). Retrieved from https://onlinelibrary.wiley.com/doi/10.1029/2022GL097836 doi: 10.1029/2022GL097836

Chen, X., Nystrom, R. G., Davis, C. A., & Zarzycki, C. M. (2020, 12). Dynamical Structures of Cross-Domain Forecast Error Covariance of a Simulated Tropical Cyclone in a Convection-Permitting Coupled Atmosphere-Ocean Model. *Monthly Weather Review*, *149*(1), 41–63. Retrieved from https://journals.ametsoc.org/view/journals/mwre/149/1/mwr-d-20-0116.1.xml doi: 10.1175/mwr-d-20-0116.1

Chen, X., Pauluis, O. M., Leung, L. R., & Zhang, F. (2018). Multiscale atmospheric overturning of the Indian summer monsoon as seen through isentropic analysis. *Journal of the Atmospheric Sciences*. doi: 10.1175/JAS-D-18-0068.1

Chen, X., Pauluis, O. M., & Zhang, F. (2018). Atmospheric overturning across multiple scales of an MJO event during the CINDY/DYNAMO campaign. *Journal of the Atmospheric Sciences*. doi: 10.1175/JAS-D-17-0060.1

Chen, X., & Zhang, F. (2019). Relative Roles of Preconditioning Moistening and Global Circumnavigating Mode on the MJO Convective Initiation During DYNAMO. *Geophysical Research Letters*. doi: 10.1029/2018GL080987

Chou, M.-D., & Suarez, M. J. (1999). A Solar Radiation Parameterization Atmospheric Studies. *Technical Report Series on Global Modeling and Data Assimilation*.

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, *20*(3). doi: 10.1023/A:1022627411411

Dovera, L., & Della Rossa, E. (2011). Multimodal ensemble Kalman filtering using Gaussian mixture models. *Computational Geosciences*, *15*(2), 307–323. doi: 10.1007/s10596-010-9205-3

Dowell, D. C., Alexander, C. R., James, E. P., Weygandt, S. S., Benjamin, S. G., Manikin, G. S., . . . Alcott, T. I. (2022). The High-Resolution Rapid Refresh (HRRR): An Hourly Updating Convection-Allowing Forecast Model. Part 1: Motivation and System Description. *Weather and Forecasting*. Retrieved from https://journals.ametsoc.org/view/journals/wefo/aop/WAF-D-21-0151.1/WAF-D-21-0151.1.xml doi: 10.1175/WAF-D-21-0151.1

ECMWF. (2016). IFS Documentation CY41R2 - Part I: Observations | ECMWF. In *Ifs doc-*

umentation cy41r2 (chap. 1). ECMWF. Retrieved from https://www.ecmwf.int/en/elibrary/16646-ifs-documentation-cy41r2-part-i-observations

Edwards, C. A., Moore, A. M., Hoteit, I., & Cornuelle, B. D. (2015). Regional ocean data assimilation. *Annual Review of Marine Science*, *7*. doi: 10.1146/annurev-marine-010814-015821

Emanuel, K. A. (1994). *Atmospheric Convection*.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, *99*(C5), 10143–10162. Retrieved from https://doi.org/10.1029/94JC00572http://doi.wiley.com/10.1029/94JC00572 doi: 10.1029/94JC00572

Evensen Geir, Vossepoel Femke C., & van Leeuwen Peter Jan. (2022). Particle Flow for a Quasi-Geostrophic Model. In *Data assimilation fundamentals: A unified formulation of the state and parameter estimation problem* (pp. 199–206). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-030-96709-3_20 doi: 10.1007/978-3-030-96709-3{_}20

Forgy, E. W. (1965). Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, *21*(3).

Fu, J. X., Wang, W., Shinoda, T., Ren, H. L., & Jia, X. (2017). Toward Understanding the Diverse Impacts of Air-Sea Interactions on MJO Simulations. *Journal of Geophysical Research: Oceans*, *122*(11). doi: 10.1002/2017JC013187

Gaspari, G., & Cohn, S. E. (1999, 1). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 723–757. Retrieved from http://doi.wiley.com/10.1002/qj.49712555417 doi: 10.1256/smsqj.55416

Geer, A. J., & Bauer, P. (2011, 10). Observation errors in all-sky data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *137*(661), 2024–2037. Retrieved from http://doi.wiley.com/10.1002/qj.830 doi: 10.1002/qj.830

Geer, A. J., Lonitz, K., Weston, P., Kazumori, M., Okamoto, K., Zhu, Y., ... Schraff, C. (2018). All-sky satellite data assimilation at operational weather forecasting centres. *Quarterly Journal of the Royal Meteorological Society*. doi: 10.1002/qj.3202

Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., & Hunt, B. R. (2011). Balance and ensemble Kalman filter localization techniques. *Monthly Weather Review*, *139*(2), 511–522. doi: 10.1175/2010MWR3328.1

Grimes, D. I., & Pardo-Igúzquiza, E. (2010). Geostatistical analysis of rainfall. *Geographical Analysis*, *42*(2). doi: 10.1111/j.1538-4632.2010.00787.x

Hamill, T. M., & Snyder, C. (2000). A hybrid ensemble Kalman filter-3D variational analysis scheme. *Monthly Weather Review*, *128*(8 II). doi: 10.1175/1520-0493(2000)128<2905:ahekfv>2.0.co;2

Harnisch, F., Weissmann, M., & Periáñez. (2016). Error model for the assimilation of cloud-affected infrared satellite observations in an ensemble data assimilation system. *Quarterly Journal of the Royal Meteorological Society*. doi: 10.1002/qj.2776

Helmert, J., Şorman, A. , Montero, R. A., De Michele, C., de Rosnay, P., Dumont, M., ... Arslan, A. N. (2018). *Review of snow data assimilation methods for hydrological, land surface, meteorological and climate models: Results from a COST harmosnow survey* (Vol. 8) (No. 12). doi: 10.3390/geosciences8120489

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*. doi: 10.1002/qj.3803

Honda, T., Miyoshi, T., Lien, G. Y., Nishizawa, S., Yoshida, R., Adachi, S. A., ... Bessho, K. (2018, 1). Assimilating all-sky Himawari-8 satellite infrared radiances: A case of Typhoon Soudelor (2015). *Monthly Weather Review*, *146*(1), 213–229. doi: 10.1175/MWR-D-16-0357.1

Hong, S. Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*. doi: 10.1175/

MWR3199.1

Hoteit, I., Luo, X., & Pham, D.-T. (2012, 2). Particle Kalman Filtering: A Nonlinear Bayesian Framework for Ensemble Kalman Filters. *Monthly Weather Review*, *140*(2), 528–542. Retrieved from `https://journals.ametsoc.org/doi/10.1175/2011MWR3640.1` doi: 10.1175/2011MWR3640.1

Hoteit, I., Pham, D. T., Triantafyllou, G., & Korres, G. (2008). A new approximate solution of the optimal nonlinear filter for data assimilation in meteorology and oceanography. *Monthly Weather Review*, *136*(1). doi: 10.1175/2007MWR1927.1

Houtekamer, P., & Mitchell, H. L. (2005, 10). Ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3269–3289. doi: 10.1256/qj.05.135

Houtekamer, P. L., & Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, *126*(3), 796–811. doi: 10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2

Houtekamer, P. L., & Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *129*(1), 123–137. doi: 10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2

Houtekamer, P. L., & Zhang, F. (2016). Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *144*(12), 4489–4532. doi: 10.1175/MWR-D-15-0440.1

Hu, C. C., & van Leeuwen, P. J. (2021). A particle flow filter for high-dimensional system applications. *Quarterly Journal of the Royal Meteorological Society*, *147*(737). doi: 10.1002/qj.4028

Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007, 6). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, *230*(1-2), 112–126. Retrieved from `https://linkinghub.elsevier.com/retrieve/pii/S0167278906004647` doi: 10.1016/j.physd.2006.11.008

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research Atmospheres*. doi: 10.1029/2008JD009944

Janowiak, J. E., Joyce, R. J., & Yarosh, Y. (2001). A real-time global half-hourly pixel-resolution infrared dataset and its applications. *Bulletin of the American Meteorological Society*, *82*(2). doi: 10.1175/1520-0477(2001)082<0205:ARTGHH>2.3.CO;2

Johnson, R. H., Rickenbach, T. M., Rutledge, S. A., Ciesielski, P. E., & Schubert, W. H. (1999). Trimodal characteristics of Tropical convection. *Journal of Climate*, *12*(8 PART 1). doi: 10.1175/1520-0442(1999)012<2397:tcotc>2.0.co;2

Keppenne, C. L., Rienecker, M. M., Kurkowski, N. P., & Adamec, D. A. (2005). Ensemble Kalman filter assimilation of temperature and altimeter data with bias correction and application to seasonal prediction. *Nonlinear Processes in Geophysics*, *12*(4). doi: 10.5194/npg-12-491-2005

Kotsuki, S., Miyoshi, T., Kondo, K., & Potthast, R. (2022). A Local Particle Filter and Its Gaussian Mixture Extension Implemented with Minor Modifications to the LETKF. *Geoscientific Model Development Discussions*, *2022*, 1–38. Retrieved from `https://gmd.copernicus.org/preprints/gmd-2022-69/` doi: 10.5194/gmd-2022-69

Lei, L., & Anderson, J. L. (2014). Comparisons of empirical localization techniques for serial ensemble kalman filters in a simple atmospheric general circulation model. *Monthly Weather Review*, *142*(2). doi: 10.1175/MWR-D-13-00152.1

Lei, L., Anderson, J. L., & Whitaker, J. S. (2016). Localizing the impact of satellite radiance observations using a global group ensemble filter. *Journal of Advances in Modeling Earth Systems*, *8*(2). doi: 10.1002/2016MS000627

Lei, L., & Whitaker, J. S. (2015). Model space localization is not always better than observation space localization for assimilation of satellite radiances. *Monthly Weather Review*, *143*(10). doi: 10.1175/MWR-D-14-00413.1

Lei, L., Whitaker, J. S., Anderson, J. L., & Tan, Z. (2020). Adaptive Localization for Satellite Radiance Observations in an Ensemble Kalman Filter. *Journal of Advances in*

Modeling Earth Systems, *12*(8). doi: 10.1029/2019MS001693

Lim, K. S. S., & Hong, S. Y. (2010). Development of an effective double-moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models. *Monthly Weather Review*. doi: 10.1175/2009MWR2968.1

Liu, B., Ait-El-Fquih, B., & Hoteit, I. (2016). Efficient kernel-based ensemble Gaussian mixture filtering. *Monthly Weather Review*, *144*(2). doi: 10.1175/MWR-D-14-00292.1

Lloyd, S. P. (1982). Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2). doi: 10.1109/TIT.1982.1056489

Lorenc, A. C. (2003). Modelling of error covariances by 4D-Var data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *129*(595 PART B), 3167–3182. doi: 10.1256/qj.02.131

Madden, R. A., & Julian, P. R. (1971). Detection of a 40–50 Day Oscillation in the Zonal Wind in the Tropical Pacific. *Journal of the Atmospheric Sciences*. doi: 10.1175/1520-0469(1971)028<0702:doadoi>2.0.co;2

Madden, R. A., & Julian, P. R. (1972). Description of Global-Scale Circulation Cells in the Tropics with a 40–50 Day Period. *Journal of the Atmospheric Sciences*. doi: 10.1175/1520-0469(1972)029<1109:dogscc>2.0.co;2

Markowski, P., & Richardson, Y. (2010). *Mesoscale Meteorology in Midlatitudes*. doi: 10.1002/9780470682104

Marzouk, Y., Moselhy, T., Parno, M., & Spantini, A. (2017). Sampling via measure transport: An introduction. In *Handbook of uncertainty quantification.* doi: 10.1007/978-3-319-12385-1{_}23

Meng, Z., & Zhang, F. (2007). Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect model experiments. *Monthly Weather Review*, *135*(4), 1403–1423. doi: 10.1175/MWR3352.1

Meng, Z., & Zhang, F. (2008). Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part III: Comparison with 3DVAR in a real-data case study. *Monthly Weather Review*. doi: 10.1175/2007MWR2106.1

Minamide, M., & Zhang, F. (2017). Adaptive observation error inflation for assimilating all-Sky satellite radiance. *Monthly Weather Review*, *145*(3), 1063–1081. doi: 10.1175/MWR-D-16-0257.1

Minamide, M., & Zhang, F. (2018). Assimilation of all-sky infrared radiances from Himawari-8 and impacts of moisture and hydrometer initialization on convection-permitting tropical cyclone prediction. *Monthly Weather Review*, *146*(10), 3241–3258. doi: 10.1175/MWR-D-17-0367.1

Minamide, M., & Zhang, F. (2019). An adaptive background error inflation method for assimilating all-sky radiances. *Quarterly Journal of the Royal Meteorological Society*, *145*(719), 805–823. doi: 10.1002/qj.3466

Park, S. K., & Xu, L. (2016). *Data assimilation for atmospheric, oceanic and hydrologic applications (Vol. III)*. doi: 10.1007/978-3-319-43415-5

Poterjoy, J. (2016, 1). A localized particle filter for high-dimensional nonlinear systems. *Monthly Weather Review*, *144*(1), 59–76. Retrieved from `http://journals.ametsoc.org/doi/10.1175/MWR-D-15-0163.1` doi: 10.1175/MWR-D-15-0163.1

Poterjoy, J., Wicker, L., & Buehner, M. (2019). Progress toward the application of a localized particle filter for numerical weather prediction. *Monthly Weather Review*. doi: 10.1175/MWR-D-17-0344.1

Press, W., & Flannery, B. (2010). *Numerical Recipes in Fortran 90* (Vol. 35) (No. 6).

Reddy, G. T., Reddy, M. P. K., Lakshmanna, K., Kaluri, R., Rajput, D. S., Srivastava, G., & Baker, T. (2020). Analysis of Dimensionality Reduction Techniques on Big Data. *IEEE Access*, *8*. doi: 10.1109/ACCESS.2020.2980942

Reich, S. (2012). A Gaussian-mixture ensemble transform filter. *Quarterly Journal of the Royal Meteorological Society*, *138*(662). doi: 10.1002/qj.898

Reichle, R. H., Bosilovich, M. G., Crow, W. T., Koster, R. D., Kumar, S. V., Mahanama,

S. P. P., & Zaitchik, B. F. (2009). Recent Advances in Land Data Assimilation at the NASA Global Modeling and Assimilation Office. In *Data assimilation for atmospheric, oceanic and hydrologic applications.* doi: 10.1007/978-3-540-71056-1{_}21

Skamarock, W., Klemp, J., Dudhi, J., Gill, D., Barker, D., Duda, M., . . . Powers, J. (2008). A Description of the Advanced Research WRF Version 3. NCAR Tech. Note NCAR/TN–468+STR, 113 pp. *NCAR TECHNICAL NOTE*. doi: 10.5065/D68S4MVH

Sondergaard, T., & Lermusiaux, P. F. (2013a). Data assimilation with gaussian mixture models using the dynamically orthogonal field equations. Part II: Applications. *Monthly Weather Review*, *141*(6), 1761–1785. doi: 10.1175/MWR-D-11-00296.1

Sondergaard, T., & Lermusiaux, P. F. (2013b, 6). Data assimilation with gaussian mixture models using the dynamically orthogonal field equations. Part I: Theory and scheme. *Monthly Weather Review*, *141*(6), 1737–1760. Retrieved from `https://journals.ametsoc.org/view/journals/mwre/141/6/mwr-d-11-00295.1.xml` doi: 10.1175/MWR-D-11-00295.1

Stammer, D., Balmaseda, M., Heimbach, P., Köhl, A., & Weaver, A. (2016). Ocean Data Assimilation in Support of Climate Applications: Status and Perspectives. *Annual Review of Marine Science*, *8*. doi: 10.1146/annurev-marine-122414-034113

Stordal, A. S., & Karlsen, H. A. (2017). Large sample properties of the adaptive gaussian mixture filter. *Monthly Weather Review*, *145*(7). doi: 10.1175/MWR-D-15-0372.1

Stordal, A. S., Karlsen, H. A., Nævdal, G., Skaug, H. J., & Vallès, B. (2011). Bridging the ensemble Kalman filter and particle filters: The adaptive Gaussian mixture filter. *Computational Geosciences*, *15*(2). doi: 10.1007/s10596-010-9207-1

Swinbank, R., Kyouda, M., Buchanan, P., Froude, L., Hamill, T. M., Hewson, T. D., . . . Yamaguchi, M. (2016). The TIGGE project and its achievements. *Bulletin of the American Meteorological Society*. doi: 10.1175/BAMS-D-13-00191.1

Temperton, C., & Williamson, D. L. (1981). Normal mode initialization for a multilevel grid-point model. Part I: Linear aspects. *Monthly Weather Review*, *109*(4). doi: 10.1175/1520-0493(1981)109<0729:NMIFAM>2.0.CO;2

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., & Whitaker, J. S. (2003, 7). Ensemble Square Root Filters. *Monthly Weather Review*, *131*(7), 1485–1490. Retrieved from `http://journals.ametsoc.org/doi/10.1175/1520-0493(2003)131%3C1485:ESRF%3E2.0.CO;2` doi: 10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2

van Leeuwen, P. J. (2009). *Particle filtering in geophysical systems.* doi: 10.1175/2009MWR2835.1

van Leeuwen, P. J. (2011). Efficient nonlinear data assimilation for oceanic models of intermediate complexity. In *Ieee workshop on statistical signal processing proceedings.* doi: 10.1109/SSP.2011.5967700

van Leeuwen, P. J., Künsch, H. R., Nerger, L., Potthast, R., & Reich, S. (2019, 7). *Particle filters for high-dimensional geoscience applications: A review* (Vol. 145) (No. 723). John Wiley and Sons Ltd. doi: 10.1002/qj.3551

Vetra-Carvalho, S., van Leeuwen, P. J., Nerger, L., Barth, A., Altaf, M. U., Brasseur, P., . . . Beckers, J. M. (2018). State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems. *Tellus, Series A: Dynamic Meteorology and Oceanography*. doi: 10.1080/16000870.2018.1445364

Wang, S., Sobel, A. H., Zhang, F., Qiang Sun, Y., Yue, Y., & Zhou, L. (2015). Regional simulation of the october and november MJO events observed during the CINDY/DYNAMO field campaign at gray zone resolution. *Journal of Climate*, *28*(6), 2097–2119. doi: 10.1175/JCLI-D-14-00294.1

Wang, X., Snyder, C., & Hamill, T. M. (2007). On the theoretical equivalence of differently proposed ensemble - 3DVAR hybrid analysis schemes. *Monthly Weather Review*, *135*(1). doi: 10.1175/MWR3282.1

Whitaker, J. S., & Hamill, T. M. (2002, 7). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, *130*(7), 1913–1924. Retrieved from http://journals.ametsoc.org/doi/10.1175/1520-0493(2002)130%3C1913:EDAWPO%3E2.0.CO;2 doi: 10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2

Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y., & Toth, Z. (2008). Ensemble data assimilation with the NCEP global forecast system. *Monthly Weather Review*, *136*(2). doi: 10.1175/2007MWR2018.1

Ying, Y., & Zhang, F. (2017). Practical and intrinsic predictability of multiscale weather and convectively coupled equatorial waves during the active phase of an MJO. *Journal of the Atmospheric Sciences*, *74*(11), 3771–3785. doi: 10.1175/JAS-D-17-0157.1

Ying, Y., & Zhang, F. (2018). Potentials in improving predictability of multiscale tropical weather systems evaluated through ensemble assimilation of simulated satellite-based observations. *Journal of the Atmospheric Sciences*, *75*(5), 1675–1698. doi: 10.1175/JAS-D-17-0245.1

Zhang, F., Minamide, M., & Clothiaux, E. E. (2016). Potential impacts of assimilating all-sky infrared satellite radiances from GOES-R on convection-permitting analysis and prediction of tropical cyclones. *Geophysical Research Letters*, *43*(6), 2954–2963. doi: 10.1002/2016GL068468

Zhang, F., Snyder, C., & Sun, J. (2004, 5). Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Monthly Weather Review*, *132*(5), 1238–1253. doi: 10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2

Zhang, F., Taraphdar, S., & Wang, S. (2017). The role of global circumnavigating mode in the MJO initiation and propagation. *Journal of Geophysical Research*, *122*(11), 5837–5856. doi: 10.1002/2016JD025665

Zhang, Y., Sieron, S. B., Lu, Y., Chen, X., Nystrom, R. G., Minamide, M., . . . Zhang, F. (2021, 12). *Ensemble-Based Assimilation of Satellite All-Sky Microwave Radiances Improves Intensity and Rainfall Predictions for Hurricane Harvey (2017)* (Vol. 48) (No. 24). John Wiley and Sons Inc. doi: 10.1029/2021GL096410

Zhang, Y., Stensrud, D. J., & Zhang, F. (2019, 12). Simultaneous Assimilation of Radar and All-Sky Satellite Infrared Radiance Observations for Convection-Allowing Ensemble Analysis and Prediction of Severe Thunderstorms. *Monthly Weather Review*, *147*(12), 4389–4409. Retrieved from https://journals.ametsoc.org/view/journals/mwre/147/12/mwr-d-19-0163.1.xml doi: 10.1175/MWR-D-19-0163.1

Supporting Information for "Potential benefits of handling mixture statistics via a bi-Gaussian EnKF: tests with all-sky satellite infrared radiances "

Man-Yau Chan[1,2], Xingchao Chen[1,2], Jeffrey L. Anderson[3]

[1]Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, Pennsylvania, USA

[2]Center for Advanced Data Assimilation and Predictability Techniques, The Pennsylvania State University, University Park, Pennsylvania, USA

[3]Data Assimilation Research Section, Computational Information Systems Laboratory, National Center for Atmospheric Research, Boulder, Colorado, USA

Contents of this file

Introduction

This document has several purposes. First, we will illustrate some differences between clear ensemble statistics and cloudy ensemble statistics. Differences like these motivate research into the BGEnKF and similar GMM-EnKFs. The second purpose is to provide a quick reference for other scientists to understand the BGEnKF, independently re-create our BGEnKF algorithm, and to support further development of the BGEnKF. To increase the accessibility of this area of research, we have written this document with graduate students in mind.

1. Text S1: Some differences between clear and cloudy member statistics

To set the stage, we plotted maps of the ensemble averaged Window-BT [Figure S1(b)] and the fraction of cloudy member columns in the ensemble [Figure S1(b)]. These ensemble quantities are constructed from the spun-up 50-member WRF ensemble described in the main text. Though the ensemble captured the general appearance of the organized convective features seen in the nature run [Figures 2(a) and S1(a)], the ensemble was uncertain about the presence/absence of clouds over much of the domain [Figure S1(b)]. This uncertainty is particularly noticeable over regions where the ensemble averaged Window-BT was between 248 K and 280 K.

Several differences between clear and cloudy member columns can be seen from Figure S1. First, the average Window-BT values of clear member columns are typically warmer than 280 K, whereas the average Window-BT values of cloudy member columns are cooler than 280 K [Figure S1(c & d)]. This difference is well known. As such, the Window-BT ensemble statistics of an ensemble of clear and cloudy member columns (henceforth, mixed ensemble) will exhibit mixed statistics.

The clear and cloudy member columns also differ noticeably in terms of their humidity fields and the Kalman gain linking Window-BT innovations to humidity increments. For the ease of visualization, we examined through a column-integrated measure of humidity that is a linear function of the WRF model state: the pseudo precipitable water (PPW). The PPW is defined as

$$\mathrm{PPW} \equiv \frac{g}{P_{\mathrm{sfc}} - P_{\mathrm{top}}} \int_0^1 q_v d\eta \tag{1}$$

where q_v refers to water vapor mass mixing ratio (QVAPOR), P_{sfc} and P_{top} refer to model surface pressure and model top pressure, and η refers to the WRF model's vertical coordinate. The PPW can be derived from the definition of precipitable water by applying the hydrostatic approximation, the definition of WRF η levels, and by assuming that P_{sfc} and P_{top} are constants ($P_{\mathrm{sfc}} \equiv 1000$ hPa, $P_{\mathrm{top}} \equiv 20$ hPa).

We opted to use the linear PPW over precipitable water (PW) because PW is a nonlinear function of the model state. Thus, the Kalman gain linking PW to Window-BT within the same model column is not mathematically equivalent to taking a column-integral of the Kalman gain linking QVAPOR to Window-BT. In contrast, said mathematical equivalence holds for PPW. Looking at PPW over PW thus allows us to get an accurate sense of what the EnKF would do to QVAPOR within a model column.

Figure S1(c & d) indicates that the PPW of cloudy member columns is higher than that of clear member columns. This is because clouds require nearly saturated humidity to materialize. As such, when the ensemble is mixed, mixture statistics in the humidity fields are likely.

We also examined the component of the Kalman gains responsible for propagating Window-BT innovations to QVAPOR: the least squares linear regression coefficient linking Window-BT to QVAPOR (Anderson, 2003). For the ease of visualization, we looked at the coefficient linking Window-BT to PPW within the same column. This coefficient (β) is defined as

$$\beta \equiv \frac{\mathrm{Cov}\left(\mathrm{PPW}, \mathrm{BT}\right)}{\mathrm{Var}\left(\mathrm{BT}\right)}. \tag{2}$$

$\mathrm{Cov}\left(\mathrm{PPW}, \mathrm{BT}\right)$ denotes the prior ensemble covariance between PPW and Window-BT within said model column, and $\mathrm{Var}\left(\mathrm{BT}\right)$ denotes the prior ensemble variance of Window-BT within the same column. In the limit where Var(BT) is much smaller than the observation error, the Kalman gain turns into β.

As can be seen from Figure S1(e & f), the clear member columns' statistically significant β values are generally an order of magnitude larger than those of the cloudy member columns. This difference suggests that the statistical relationship between Window-BT and humidity can vary dramatically depending on the absence/presence of clouds.

2. Text S2: Heuristic localized clustering of ensemble members

Since a mixture of clear and cloudy members results in a mixed prior distribution, it seems appropriate to explore an ensemble DA method that explicitly treat mixture distributions. Since the EnKF has been remarkably successful at assimilating infrared radiance observations (Otkin, 2012; F. Zhang et al., 2016; Honda et al., 2018; Minamide & Zhang, 2018; Y. Zhang et al., 2018; Otkin & Potthast, 2019; F. Zhang et al., 2019; Geer et al., 2019; Chan, Zhang, et al., 2020; Jones et al., 2020; Chan & Chen, 2021; Hartman et al., 2021; Y. Zhang et al., 2021), we will extend the EnKF to handle clear members and cloudy members separately.

A complication in handling clear members and cloudy members separately lies in the fact every member usually contains both clear model columns and cloudy model columns. Supposing we have $N_i * N_j$ model columns in the domain, there can exist up to $2^{N_i*N_j}$ possible spatial combinations of clear and cloudy columns in the domain. Sampling these $2^{N_i*N_j}$ combinations would require more than $2^{N_i*N_j}$ ensemble members – a likely impractical proposition. Dimensional reduction is necessary to reduce the required number of ensemble members.

A simple and natural dimensional reduction approach is to limit our clear/cloudy considerations to small regions of the domain. This dimensional reduction approach is effectively a type of spatial localization – a commonly employed heuristic method used to limit the effects of sampling errors on EnKFs (Houtekamer & Zhang, 2016). As a first attempt at employing this localization, suppose we are assimilating observations one-at-a-time (*i.e.,* serial assimilation). When assimilating the m-th observation, we will only consider model columns within 1 horizontal radius of influence (HROI) surrounding the observed column. If there are N_{loc} columns within 1 HROI of the observed column, the number of possible spatial combinations falls from $2^{N_i*N_j}$ to $2^{N_{loc}}$. For commonly used HROI values, $2^{N_{loc}} \ll 2^{N_i*N_j}$.

Though localization can dramatically reduce the number of spatial clear/cloudy combinations, $2^{N_{loc}}$ is likely greater than the number of ensemble members N_E. For instance, in the IR DA experiments of Chan, Zhang, et al. (2020) and Chan and Chen (2021), the HROI is approximately 11 model grid spacings (100-km HROI, 9-km grid spacing), meaning that there exist $\sim \pi * 11^2 \approx 363$ model columns within the localization zone. A typical ensemble size of ~ 50 is much less than the number of spatial combinations in this example ($\sim 2^{363}$). Another measure is necessary to further simplify the problem.

We opted to assume that there are at most two clear/cloudy spatial combinations within the localized zone. To understand the rationale, consider that localized serial EnKFs assume that all ensemble members within 1 radius of influence (ROI) of an observation to be drawn from a Gaussian distribution (Burgers et al., 1998; Whitaker & Hamill, 2002; Anderson, 2003). This is equivalent to assuming that there exists only one spatial combination within 1 ROI of said observation. Our two spatial combination assumption, though imperfect, is closer to the actual number of spatial combinations ($2^{N_{loc}}$) than the one spatial combination assumption.

We can now consider that the ensemble members are drawn from a mixture of two distributions within the localized region. The EnKF can be extended to handle this mixture distribution by replacing the EnKF's Gaussian prior assumption. Specifically, we consider that some prior members are drawn from one Gaussian distribution and the other members are drawn from a different Gaussian distribution. The prior ensemble is thus assumed to be drawn from a bi-Gaussian prior distribution. The resulting algorithm will be henceforth termed the bi-Gaussian EnKF (BGEnKF).

For the BGEnKF to work, it is necessary to separate the ensemble members into two groups (henceforth termed "clusters"). The sample statistics of each cluster will correspond to one of two Gaussian kernels. As a first approach, we will consider members that are clear at the observation site to be drawn from one Gaussian distribution (henceforth termed the "clear kernel" or "clear cluster"). The remaining members will be considered to be drawn from a different Gaussian distribution (henceforth, the "cloudy kernel" or "cloudy cluster"). More advanced clustering approaches, such as those involving machine learning (*e.g.*, support vector machines), can be considered at a later date.

3. Text S3: Bayes' rule for the BGEnKF

We will now formulate a serially assimilating BGEnKF (*i.e.*, the algorithm assimilates one observation at a time) starting from Bayes' rule and using a notation akin to that of Ide, Courtier, Ghil, and Lorenc (1997). In our earlier study (Chan, Anderson, & Chen, 2020), the BGEnKF was formulated as a model state space filter [or, in the terminology of Anderson and Collins (2007), a sequential filter]. However, multi-process implementations of sequential filters require inter-process communications at every iteration of the serial assimilation loop. The sequential filter formulation thus does not scale well with parallelization (Anderson & Collins, 2007).

To ensure that the BGEnKF algorithm scales well with parallelization, the BGEnKF is formulated to constrain an extended state vector $\boldsymbol{\psi}$ (Anderson & Collins, 2007). $\boldsymbol{\psi}$ will contain all of the variables used in the BGEnKF. Aside from containing the model state $\boldsymbol{x}$, $\boldsymbol{\psi}$ will also contain the simulated observation values $\boldsymbol{y}$ that correspond to said model state. Furthermore, since ξ [column-integrated frozen water mass content; see main text's Eq. (1)] can be used to discriminate clear column members from cloudy column members (see main text's section 2.2), we will include ξ at every observation site into $\boldsymbol{\psi}$. The vector $\boldsymbol{\xi}$ will be used to denote the ξ values at every observation site. We can thus define

$$\boldsymbol{\psi} \equiv \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \\ \boldsymbol{\xi} \end{bmatrix}. \tag{3}$$

Supposing N_x denotes the number of elements in $\boldsymbol{x}$ and N_y denotes the number of elements in $\boldsymbol{y}$ (and in $\boldsymbol{\xi}$), then $\boldsymbol{\psi}$ has $N_x + 2N_y$ elements. For the ease of writing, we will define

$$N_\psi \equiv N_x + 2N_y$$

With Eq. (3), we can construct an ensemble of forecasted $\boldsymbol{\psi}$ vectors. Supposing that we have a forecast ensemble of N_E model states $\left\{\boldsymbol{x_1^f}, \boldsymbol{x_2^f}, \ldots, \boldsymbol{x_{N_E}^f}\right\}$, we can define an ensemble of N_E forecasted extended state vectors via

$$\boldsymbol{\psi_n^f} \equiv \begin{bmatrix} \boldsymbol{x_n^f} \\ \boldsymbol{h}(\boldsymbol{x_n^f}) \\ \boldsymbol{\xi}\left(\boldsymbol{x_n^f}\right) \end{bmatrix} \quad \forall \;\; n = 1, 2, \ldots, N_E. \tag{4}$$

Here, $\boldsymbol{h}(\boldsymbol{x_n^f})$ represents calling the observation operator $\boldsymbol{h}$ on $\boldsymbol{x_n^f}$, and $\boldsymbol{\xi}\left(\boldsymbol{x_n^f}\right)$ represents evaluating ξ [Eq. (1) of main text] at every observation site using the using the information in $\boldsymbol{x_n^f}$.

Since the BGEnKF will be formulated as a serial assimilation algorithm, we can outline the essence of the algorithm by considering what happens when a single observation (y^o) is assimilated into an ensemble of forecasted $\boldsymbol{\psi}$ vectors. Like typical serially assimilating EnKF algorithms [*e.g.*, Whitaker, Hamill, Wei, Song, and Toth (2008), Anderson et al. (2009), and Meng and Zhang (2007)], the serially assimilating BGEnKF algorithm is of the form:

1. Construct an ensemble of forecasted $\boldsymbol{\psi}$ vectors (*i.e.*, $\left\{\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{1}}, \boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{2}}, \ldots, \boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{N_E}}\right\}$).
2. Select an unassimilated observation.
3. Divide the ensemble into the clear and cloudy clusters using the procedure described the main text's section 2.2.
4. Assimilate the selected observation using the BGEnKF to construct an ensemble of analyzed $\boldsymbol{\psi}$ vectors (*i.e.*, $\left\{\boldsymbol{\psi}^{\boldsymbol{a}}_{\boldsymbol{1}}, \boldsymbol{\psi}^{\boldsymbol{a}}_{\boldsymbol{2}}, \ldots, \boldsymbol{\psi}^{\boldsymbol{a}}_{\boldsymbol{N_E}}\right\}$)
5. If there are unassimilated observations remaining,
   (i) Overwrite the forecast ensemble with the posterior ensemble (*i.e.*, $\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{n}} \leftarrow \boldsymbol{\psi}^{\boldsymbol{a}}_{\boldsymbol{n}} \;\forall n = 1, 2, \ldots, N_E$).
   (ii) Return to step 2..
6. Exit.

We will thus formulate the BGEnKF equations by considering the assimilation of y^o into $\left\{\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{1}}, \boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{2}}, \ldots, \boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{N_E}}\right\}$. Supposing that the ensemble members have been sorted into the clear and cloudy clusters based on the ξ value at the observation site, the BGEnKF assumes that the prior probability density function [pdf; $p\left(\boldsymbol{\psi}\right)$] can be represented by the bi-Gaussian pdf

$$p\left(\boldsymbol{\psi}\right) = w^f_{\text{clr}}\, \mathcal{G}\left(\boldsymbol{\psi};\;\; \overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{clr}}}, \boldsymbol{P}^{\boldsymbol{f}}_{\text{clr}}\right) + w^f_{\text{cld}}\, \mathcal{G}\left(\boldsymbol{\psi};\;\; \overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{cld}}}, \boldsymbol{P}^{\boldsymbol{f}}_{\text{cld}}\right). \tag{5}$$

Throughout this document, we will use the subscript "clr" to denote clear cluster quantities, and the subscript "cld" to denote cloudy cluster quantities. $\mathcal{G}\left(\boldsymbol{\psi};\;\; \overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{clr}}}, \boldsymbol{P}^{\boldsymbol{f}}_{\text{clr}}\right)$ denotes the clear cluster's Gaussian kernel with mean state $\overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{clr}}}$ and covariance matrix $\boldsymbol{P}^{\boldsymbol{f}}_{\text{clr}}$. Similarly, $\mathcal{G}\left(\boldsymbol{\psi};\;\; \overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{cld}}}, \boldsymbol{P}^{\boldsymbol{f}}_{\text{cld}}\right)$ denotes the cloudy cluster's Gaussian kernel with mean state $\overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\text{cld}}}$ and covariance matrix $\boldsymbol{P}^{\boldsymbol{f}}_{\text{cld}}$. In general, the Gaussian pdf for a K-dimensional state $\boldsymbol{p}$ vector with some mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{C}$ is defined as

$$\mathcal{G}\left(\boldsymbol{p};\;\; \boldsymbol{\mu}, \boldsymbol{C}\right) \equiv \frac{1}{\sqrt{(2\pi)^K \det\left(\boldsymbol{C}\right)}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{p} - \boldsymbol{\mu}\right)^\top \boldsymbol{C}^{-1}\left(\boldsymbol{p} - \boldsymbol{\mu}\right)\right\}.$$

The scalar quantities w^f_{clr} and w^f_{cld} are the respective weights of the clear and cloudy Gaussian kernels. Note that

$$w^f_{\text{clr}} + w^f_{\text{cld}} = 1, \quad w^f_{\text{clr}} \geq 0, \quad \text{and,} \quad w^f_{\text{cld}} \geq 0.$$

The various parameters in the prior pdf [Eq. (5)] are estimated from the clustered forecast ensemble of $\boldsymbol{\psi}$ vectors. Suppose the set S_{clr} contains the ensemble member indices of clear cluster members [*i.e.*, the index n in Eq. (4)] and the set S_{cld} contains the ensemble member indices of cloudy cluster members. We first compute the number of members in the clear cluster (N^f_{clr}) and the number of members in the clear cluster (N^f_{cld}) via

$$N^f_{\text{clr}} \equiv \text{count}\left(S_{\text{clr}}\right), \quad \text{and,} \quad N^f_{\text{cld}} \equiv \text{count}\left(S_{\text{cld}}\right) \tag{6}$$

Supposing g is a placeholder that can be replaced with "clr" or "cld", $\text{count}\left(S_g\right)$ counts the number of elements in the set S_g. The parameters of Eq. (5) can then be estimated via

$$\overline{\boldsymbol{\psi}_{\boldsymbol{g}}^{\boldsymbol{f}}} \equiv \frac{1}{N_g^f} \sum_{n \in S_g} \boldsymbol{\psi}_{\boldsymbol{n}}^{\boldsymbol{f}}, \quad \overline{\boldsymbol{P}_{\boldsymbol{g}}^{\boldsymbol{f}}} \equiv \frac{1}{N_g^f - 1} \sum_{n \in S_g} \left(\boldsymbol{\psi}_{\boldsymbol{n}}^{\boldsymbol{f}} - \overline{\boldsymbol{\psi}_{\mathbf{g}}^{\boldsymbol{f}}}\right) \left(\boldsymbol{\psi}_{\boldsymbol{n}}^{\boldsymbol{f}} - \overline{\boldsymbol{\psi}_{\mathbf{g}}^{\boldsymbol{f}}}\right)^\top, \text{ and, } w_g^f \equiv \frac{N_g^f}{N_{\text{clr}}^f + N_{\text{cld}}^f}. \tag{7}$$

Note that the BGEnKF does not require any explicit estimate of the large matrices $\boldsymbol{P}_{\text{cld}}^{\boldsymbol{f}}$ and $\boldsymbol{P}_{\text{clr}}^{\boldsymbol{f}}$. Instead, like the typical serially assimilating EnKF, the BGEnKF only requires calculating a column of these matrices. This will be discussed in Text S4.

To assimilate y^o into $\left\{\boldsymbol{\psi}_{\boldsymbol{1}}^{\boldsymbol{f}}, \ldots, \boldsymbol{\psi}_{\boldsymbol{N_E}}^{\boldsymbol{f}}\right\}$, consider Bayes' rule:

$$p\left(\boldsymbol{\psi}|y^o\right) = \frac{p\left(\boldsymbol{\psi}\right)\ p\left(y^o|\boldsymbol{\psi}\right)}{p\left(y^o\right)} \tag{8}$$

where the marginal $p\left(y^o\right)$ normalizes the numerator of Eq. (8) [*e.g.*, Lorenc (1986)]. As we will show later, this normalization property is central to deriving the posterior weights of the clear and cloudy posterior kernels. Note that though the normalization property is used in the derivation, there is no need to explicit compute $p\left(y^o\right)$ at all in the BGEnKF algorithm.

If we assume Gaussian observation errors, the observation likelihood $p\left(y^o|\boldsymbol{\psi}\right)$ can be written as

$$p\left(y^o|\boldsymbol{\psi}\right) \equiv \mathcal{G}\left(\boldsymbol{H}\boldsymbol{\psi};\quad y^o, \sigma^{o2}\right) \tag{9}$$

where σ^{o2} is the observation error variance and $\boldsymbol{H}$ is a matrix that extracts the simulated observation from $\boldsymbol{\psi}$. Specifically, if y^o corresponds to the $(N_x + m)$-th element in $\boldsymbol{\psi}$, $\boldsymbol{H}$ is an $1 \times N_\psi$ matrix of the form

$$\boldsymbol{H} \equiv \begin{bmatrix} 0 & 0 & \ldots & 0 & 1 & 0 & \ldots & 0 & 0 \end{bmatrix}$$

where the only non-zero element (unity) is the $(N_x + m)$-th element.

Before proceeding further, note that the observation likelihoods for IR-BTs are not strictly Gaussian. The associated observation errors are known to be dependent on the presence/absence of clouds in the observed atmospheric columns (Geer & Bauer, 2011; Harnisch et al., 2016; Minamide & Zhang, 2017; Otkin et al., 2018). Furthermore, IR-BT values are bounded. Nonetheless, the successes seen in assimilating IR-BTs with EnKFs suggest that the imperfect Gaussian observation likelihood assumption is at least somewhat functional (Otkin, 2012; F. Zhang et al., 2016; Honda et al., 2018; Minamide & Zhang, 2018; Y. Zhang et al., 2018; Otkin & Potthast, 2019; F. Zhang et al., 2019; Geer et al., 2019; Chan, Zhang, et al., 2020; Jones et al., 2020; Chan & Chen, 2021; Hartman et al., 2021; Y. Zhang et al., 2021). We will thus proceed with the assumption that the observation likelihood is Gaussian.

For the ease of future reference, we will sketch out the main steps to derive the posterior pdf. Combining the bi-Gaussian forecast pdf [Eq. (5)] with the Gaussian observation likelihood [Eq. (9)] through Bayes rule [Eq. (8)] will result in

$$\begin{aligned} p\left(\boldsymbol{\psi}|y^o\right) = & w_{\text{clr}}^f\, \mathcal{G}\left(\boldsymbol{\psi};\quad \overline{\boldsymbol{\psi}_{\text{clr}}^{\boldsymbol{f}}}, \boldsymbol{P}_{\text{clr}}^{\boldsymbol{f}}\right) \mathcal{G}\left(\boldsymbol{H}\boldsymbol{\psi};\quad y^o, \sigma^{o2}\right) / p\left(y^o\right) \\ & + w_{\text{cld}}^f\, \mathcal{G}\left(\boldsymbol{\psi};\quad \overline{\boldsymbol{\psi}_{\text{cld}}^{\boldsymbol{f}}}, \boldsymbol{P}_{\text{cld}}^{\boldsymbol{f}}\right) \mathcal{G}\left(\boldsymbol{H}\boldsymbol{\psi};\quad y^o, \sigma^{o2}\right) / p\left(y^o\right) \end{aligned} \tag{10}$$

To proceed further, a well-known property is used: the multiplication of two Gaussian pdfs results in a scaled Gaussian pdf. This property is foundational to EnKFs (Evensen, 1994; Burgers et al., 1998; Houtekamer & Mitchell, 2001; Anderson, 2001; Bishop et al., 2001; Whitaker & Hamill, 2002; Tippett et al., 2003; Hunt et al., 2007). In this situation, for the term associated with cluster g [*e.g.*, Anderson and Anderson (1999)],

$$\mathcal{G}\left(\boldsymbol{\psi};\ \ \overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}}}, \boldsymbol{P}^{\boldsymbol{f}}_{\boldsymbol{g}}\right)\ \mathcal{G}\left(\boldsymbol{H}\boldsymbol{\psi};\ \ y^{o}, \sigma^{o2}\right) = \alpha_g\, \mathcal{G}\left(\boldsymbol{\psi};\ \ \overline{\boldsymbol{\psi}^{\boldsymbol{a}}_{\boldsymbol{g}}}, \boldsymbol{P}^{\boldsymbol{a}}_{\boldsymbol{g}}\right) \tag{11}$$

where $\overline{\boldsymbol{\psi}^{\boldsymbol{a}}_{\boldsymbol{g}}}$ represents the analyzed average state of cluster g, $\boldsymbol{P}^{\boldsymbol{a}}_{\boldsymbol{g}}$ represents the analyzed covariance matrices of said cluster, and α_g is a scaling factor. $\overline{\boldsymbol{\psi}^{\boldsymbol{a}}_{\boldsymbol{g}}}$ and $\boldsymbol{P}^{\boldsymbol{a}}_{\boldsymbol{g}}$ are related to $\overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}}}$ and $\boldsymbol{P}^{\boldsymbol{f}}_{\boldsymbol{g}}$ via the Kalman filter (KF) equations [*e.g.*, Lorenc (1986)]

$$\overline{\boldsymbol{\psi}^{\boldsymbol{a}}_{\boldsymbol{g}}} = \overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}}} + \boldsymbol{K}_{\boldsymbol{g}}\left(y^{o} - \boldsymbol{H}\overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}}}\right), \quad \text{and,} \quad \boldsymbol{P}^{\boldsymbol{a}}_{\boldsymbol{g}} = \left(\boldsymbol{I} - \boldsymbol{K}_{\boldsymbol{g}}\boldsymbol{H}\right)\boldsymbol{P}^{\boldsymbol{f}}_{\boldsymbol{g}}, \tag{12}$$

where $\boldsymbol{K}_{\boldsymbol{g}}$ is the Kalman gain matrix for cluster g. $\boldsymbol{K}_{\boldsymbol{g}}$ can be computed via

$$\boldsymbol{K}_{\boldsymbol{g}} \equiv \boldsymbol{P}^{\boldsymbol{f}}_{\boldsymbol{g}}\boldsymbol{H}^{\top}\left(\boldsymbol{H}\,\boldsymbol{P}^{\boldsymbol{f}}_{\boldsymbol{g}}\boldsymbol{H}^{\top} + \sigma^{o2}\right)^{-1} = \frac{\mathrm{Cov}\left(\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}},\, \boldsymbol{H}\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}}\right)}{\mathrm{Var}\left(\boldsymbol{H}\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}}\right) + \sigma^{o2}} \tag{13}$$

where

$$\begin{aligned}\mathrm{Cov}\left(\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}},\, \boldsymbol{H}\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}}\right) &\equiv \frac{1}{N^{f}_{g} - 1}\sum_{n_g \in S_g}\left(\boldsymbol{H}\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{n}} - \boldsymbol{H}\overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}}}\right)\left(\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{n}} - \overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}}}\right), \\ \mathrm{Var}\left(\boldsymbol{H}\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}}\right) &\equiv \frac{1}{N^{f}_{g} - 1}\sum_{n_g \in S_g}\left(\boldsymbol{H}\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{n}} - \boldsymbol{H}\overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}}}\right)^{2},\end{aligned}$$

and n_g is a dummy index that iterates over the member indices contained in S_g. The scaling factor α_g in Eq. (11) can be shown to be [*e.g.*, Anderson and Anderson (1999)]:

$$\alpha_g = \mathcal{G}\left(y^{o};\ \ \boldsymbol{H}\ \overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}}},\ \sigma^{o2} + \boldsymbol{H}\boldsymbol{P}^{\boldsymbol{f}}_{\boldsymbol{g}}\boldsymbol{H}^{\top}\right). \tag{14}$$

Note that $\boldsymbol{H}\,\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{n}}$, $\boldsymbol{H}\,\overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}}}$, and $\mathrm{Var}\left(\boldsymbol{H}\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}}\right)$ are scalars. Furthermore, if y^{o} corresponds to the $(N_x + m)$-th element of $\boldsymbol{\psi}$, then $\mathrm{Cov}\left(\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}},\, \boldsymbol{H}\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}}\right)$ is equal to the $(N_x + m)$-th column of $\boldsymbol{P}^{\boldsymbol{f}}_{\boldsymbol{g}}$.

Substituting Eq. (11) into Eq. (10) and results in

$$p\left(\boldsymbol{\psi}|y^{o}\right) = \frac{w^{f}_{\mathrm{clr}}\,\alpha_{\mathrm{clr}}\,\mathcal{G}\left(\boldsymbol{\psi};\ \ \overline{\boldsymbol{\psi}^{\boldsymbol{a}}_{\mathrm{clr}}}, \boldsymbol{P}^{\boldsymbol{a}}_{\mathrm{clr}}\right) + w^{f}_{\mathrm{cld}}\,\alpha_{\mathrm{cld}}\,\mathcal{G}\left(\boldsymbol{\psi};\ \ \overline{\boldsymbol{\psi}^{\boldsymbol{a}}_{\mathrm{cld}}}, \boldsymbol{P}^{\boldsymbol{a}}_{\mathrm{cld}}\right)}{p\left(y^{o}\right)}. \tag{15}$$

Since $p\left(y^{o}\right)$ normalizes Eq. (15), then,

$$\begin{aligned}p\left(y^{o}\right) &= \int_{R^{N_\psi}}\left\{w^{f}_{\mathrm{clr}}\,\alpha_{\mathrm{clr}}\,\mathcal{G}\left(\boldsymbol{\psi};\ \ \overline{\boldsymbol{\psi}^{\boldsymbol{a}}_{\mathrm{clr}}}, \boldsymbol{P}^{\boldsymbol{a}}_{\mathrm{clr}}\right) + w^{f}_{\mathrm{cld}}\,\alpha_{\mathrm{cld}}\,\mathcal{G}\left(\boldsymbol{\psi};\ \ \overline{\boldsymbol{\psi}^{\boldsymbol{a}}_{\mathrm{cld}}}, \boldsymbol{P}^{\boldsymbol{a}}_{\mathrm{cld}}\right)\right\}\mathrm{d}^{N_\psi}\boldsymbol{\psi} \\ &= w^{f}_{\mathrm{clr}}\,\alpha_{\mathrm{clr}} + w^{f}_{\mathrm{cld}}\,\alpha_{\mathrm{cld}}\end{aligned} \tag{16}$$

where $\int_{R^{N_\psi}}\{\cdot\}\,\mathrm{d}^{N_\psi}\boldsymbol{\psi}$ is an infinite N_ψ-dimensional volume integral of $\{\cdot\}$ over the N_ψ-dimensional space that $\boldsymbol{\psi}$ lives in [*i.e.*, an R^{N_ψ} space]. Substituting the marginal [Eq. (16)] back into Bayes' rule [Eq. (15)] gives us the following bi-Gaussian posterior pdf

$$p\left(\boldsymbol{\psi}|y^{o}\right) = w^{a}_{\mathrm{clr}}\,\mathcal{G}\left(\boldsymbol{\psi};\ \ \overline{\boldsymbol{\psi}^{\boldsymbol{a}}_{\mathrm{clr}}}, \boldsymbol{P}^{\boldsymbol{a}}_{\mathrm{clr}}\right) + w^{a}_{\mathrm{cld}}\,\mathcal{G}\left(\boldsymbol{\psi};\ \ \overline{\boldsymbol{\psi}^{\boldsymbol{a}}_{\mathrm{cld}}}, \boldsymbol{P}^{\boldsymbol{a}}_{\mathrm{cld}}\right) \tag{17}$$

where

$$w^a_{\mathrm{clr}} = \frac{w^f_{\mathrm{clr}}\,\alpha_{\mathrm{clr}}}{w^f_{\mathrm{clr}}\,\alpha_{\mathrm{clr}} + w^f_{\mathrm{cld}}\,\alpha_{\mathrm{cld}}}, \quad \text{and,} \quad w^a_{\mathrm{cld}} = \frac{w^f_{\mathrm{cld}}\,\alpha_{\mathrm{cld}}}{w^f_{\mathrm{clr}}\,\alpha_{\mathrm{clr}} + w^f_{\mathrm{cld}}\,\alpha_{\mathrm{cld}}}. \tag{18}$$

Like the EnKF, the BGEnKF will update the forecast ensemble to become consistent with the posterior bi-Gaussian pdf [Eq. (17)].

4. Text S4: Detailed description of the three-stage BGEnKF algorithm

The BGEnKF's updates to the ensemble is done through a three-stage update process (illustrated in the main text's Figure 1). In order of execution, these stages are: 1) the double EnKF stage, 2) the shrinking cluster member deletion stage, and 3) the expanding cluster member resampling stage. An outline of this three-stage BGEnKF update procedure can be found at the end of this section.

The double EnKF stage

The first stage [Figure 1(a)] is to represent the KF updates to the clusters' mean states and covariance matrices. We can thus use the ensemble square root filter of Whitaker and Hamill (2002) (EnSRF) to update each cluster's members. The EnSRF update equation (Whitaker & Hamill, 2002) for members in cluster g is

$$\boldsymbol{\psi}^{\boldsymbol{a}}_{\boldsymbol{n_g}} = \boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{n_g}} + \boldsymbol{K_g}\left(y^o - \boldsymbol{H}\overline{\boldsymbol{\psi_g}}\right) - \phi_g \boldsymbol{K_g}\left(\boldsymbol{H}\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{n_g}} - \boldsymbol{H}\overline{\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}}}\right) \quad \forall \;\; n_g \in S_g. \tag{19}$$

The Kalman gain matrix of cluster g ($\boldsymbol{K_g}$) can be computed via Eq. (13). ϕ_g is the EnSRF's square-root modification factor (Whitaker & Hamill, 2002), which can be computed via

$$\phi_g \equiv \left\{ 1 + \sqrt{\frac{\sigma^{o2}}{\sigma^{o2} + \mathrm{Var}\left(\boldsymbol{H}\boldsymbol{\psi}^{\boldsymbol{f}}_{\boldsymbol{g}}\right)}} \right\}^{-1}. \tag{20}$$

Note that the EnSRF-based cluster update equations can be replaced with those from the two-step ensemble adjustment Kalman filter (EAKF) of Anderson (2003). This is because the two filters have mathematically identical ensemble member update procedures.

The member deletion stage

In the second and third stages of the BGEnKF (Figure 1(b & c)), the number of ensemble members in each cluster (*i.e.*, cluster sizes) is updated to be consistent with the cluster's posterior weight [Eq. (18)]. The post-BGEnKF size of cluster g (N^a_g) can be determined by

$$N^a_g \equiv \mathrm{round}\left(N_E * W_g\right) \tag{21}$$

where $\mathrm{round}(\cdot)$ indicates rounding $\cdot$ to the nearest integer.

If the size of a cluster is reduced by the assimilation of y^o, we will delete members from said cluster (Figure 1(b)). The number of members to be deleted N_{del} is defined as

$$N_{\mathrm{del}} \equiv \begin{cases} N^f_{\mathrm{clr}} - N^a_{\mathrm{clr}} & \text{if } N^a_{\mathrm{clr}} < N^f_{\mathrm{clr}}, \\ \\ N^f_{\mathrm{cld}} - N^a_{\mathrm{cld}} & \text{if } N^a_{\mathrm{cld}} < N^f_{\mathrm{cld}}. \end{cases} \tag{22}$$

For simplicity, we will delete the members with the smallest N_{del} forecast-simulated observation perturbations. Since the deletion will cause the cluster's mean state to deviate from the theoretical mean state [Eq. (12)], we will recenter the remaining members around the theoretical value. Note that no heuristic adjustments were made to mitigate the changes in the cluster's sample covariance matrix due to the deletion process. This is because it is impossible to prevent such changes in practical situations [for $N_E < N_\psi$, the rank of the pre-deletion sample covariance matrix is guaranteed to be higher than the rank of the post-deletion sample covariance matrix; Chan, Anderson, and Chen (2020)].

The resampling stage

If the size of one cluster is reduced by the assimilation of y^o, the other cluster's size will increase to compensate for the reduction. This ensures that the total number of ensemble members is unchanged. To do so, the expanding cluster's ensemble members are resampled. The expanding cluster's sample mean state and sample covariance matrix should not be altered by resampling.

The computationally efficient resampling strategy proposed in Chan, Anderson, and Chen (2020) is to resample within the extended state subspace spanned by the expanding cluster's ensemble members (henceforth referred to as the subspace resampling strategy). This is the easiest to formulate in terms of the perturbations of the expanding cluster's members. Supposing that the subscript "pre" denotes expanding cluster quantities before resampling, we can compute the pre-resampling perturbations $\left\{\boldsymbol{\psi}_{\boldsymbol{n}}^{\boldsymbol{a}'} | n \in S_{\text{pre}}\right\}$ via

$$\boldsymbol{\psi}_{\boldsymbol{n}}^{\boldsymbol{a}'} \equiv \boldsymbol{\psi}_{\boldsymbol{n}}^{\boldsymbol{a}} - \overline{\boldsymbol{\psi}_{\mathbf{pre}}^{\boldsymbol{a}}} \quad \forall \;\; n \in S_{\text{pre}} \tag{23}$$

where $\overline{\boldsymbol{\psi}_{\mathbf{pre}}^{\boldsymbol{a}}}$ is the expanding cluster's mean state and S_{pre} is the set of member indices in the expanding cluster before resampling.

The central idea of the subspace resampling strategy is to construct a new set of perturbations via linear combinations of the pre-resampling perturbations. We will denote all post-resampling expanding cluster quantities with the subscript "post". Let S_{post} denote the set of member indices in the post-resampling expanding cluster. S_{post} thus includes the member indices in S_{pre} and the indices of the members deleted in the deletion stage. If we represent the set of post-resampling perturbation vectors as $\left\{\boldsymbol{\psi}_{\boldsymbol{n^*}}^{\boldsymbol{a^*}} | n^* \in S_{\text{post}}\right\}$, the strategy's central idea can then be mathematically expressed as

$$\boldsymbol{\psi}_{\boldsymbol{n^*}}^{\boldsymbol{a^*}} \equiv \sum_{n \in S_{\text{pre}}} \boldsymbol{\psi}_{\boldsymbol{n}}^{\boldsymbol{a}'} \; T_{n,n^*} \quad \forall \quad n^* \in S_{\text{post}}$$

where T_{n,n^*} is a to-be-determined scalar factor controlling how the n-th pre-resampling perturbation contributes to the n^*-th post-resampling perturbation. This linear combination idea can be more succinctly expressed as

$$\boldsymbol{\Psi}_{\mathbf{post}} \equiv \boldsymbol{\Psi}_{\mathbf{pre}} \boldsymbol{T}. \tag{24}$$

Here, $\boldsymbol{\Psi}_{\mathbf{pre}}$ is a matrix where each column contains a pre-resampling perturbation, and $\boldsymbol{\Psi}_{\mathbf{post}}$ is a matrix where each column contains a post-resampling perturbation. Supposing the pre-resampling cluster size is denoted by N_{pre} and the post-resampling cluster size is denoted by N_{post}, then $\boldsymbol{\Psi}_{\mathbf{pre}}$ is an $N_\psi \times N_{\text{pre}}$ matrix and $\boldsymbol{\Psi}_{\mathbf{post}}$ is an $N_\psi \times N_{\text{post}}$ matrix. If we denote the ℓ-th member index in S_{pre} as $n_{\text{pre},\ell}$, and likewise for the ℓ-th member index in S_{post}, we can explicitly write out $\boldsymbol{\Psi}_{\mathbf{pre}}$ and $\boldsymbol{\Psi}_{\mathbf{post}}$:

$$\begin{aligned}
\boldsymbol{\Psi}_{\mathbf{pre}} &\equiv \left[\boldsymbol{\psi}_{\boldsymbol{n}_{\mathbf{pre},1}}^{\boldsymbol{a}'} \;\; \boldsymbol{\psi}_{\boldsymbol{n}_{\mathbf{pre},2}}^{\boldsymbol{a}'} \;\; \cdots \;\; \boldsymbol{\psi}_{\boldsymbol{n}_{\mathbf{pre},\boldsymbol{N}_{\mathbf{pre}}}}^{\boldsymbol{a}'} \right], \\
\boldsymbol{\Psi}_{\mathbf{post}} &\equiv \left[\boldsymbol{\psi}_{\boldsymbol{n}_{\mathbf{post},1}}^{\boldsymbol{a^*}} \;\; \boldsymbol{\psi}_{\boldsymbol{n}_{\mathbf{post},2}}^{\boldsymbol{a^*}} \;\; \cdots \;\; \boldsymbol{\psi}_{\boldsymbol{n}_{\mathbf{post},\boldsymbol{N}_{\mathbf{post}}}}^{\boldsymbol{a^*}} \right].
\end{aligned} \tag{25}$$

Finally, $\boldsymbol{T}$ is an $N_{\text{pre}} \times N_{\text{post}}$ matrix containing all of the T_{n,n^*} values [*i.e.*, element (n, n^*) of $\boldsymbol{T}$ is equal to T_{n,n^*}].

$\boldsymbol{T}$ should be constructed such that the post-resampling perturbations have a mean of zero and have a covariance matrix equal to that of pre-resampling perturbations. As discussed in Chan, Anderson, and Chen (2020), there are an infinite number of possible $\boldsymbol{T}$'s that satisfy these two conditions. Following the discussions and heuristic arguments in Chan, Anderson, and Chen (2020), we chose to use

$$\boldsymbol{T} \equiv \begin{bmatrix} k\boldsymbol{I}_{\boldsymbol{N}_{\text{pre}}-\boldsymbol{N}^*_{\text{new}}} & \boldsymbol{0}_{(\boldsymbol{N}_{\text{pre}}-\boldsymbol{N}^*_{\text{new}})\times\boldsymbol{N}^*_{\text{new}}} & \boldsymbol{0}_{(\boldsymbol{N}_{\text{pre}}-\boldsymbol{N}^*_{\text{new}})\times\boldsymbol{N}_{\text{new}}} \\ \\ \boldsymbol{0}_{\boldsymbol{N}^*_{\text{new}}\times(\boldsymbol{N}_{\text{pre}}-\boldsymbol{N}^*_{\text{new}})} & \boldsymbol{I}_{\boldsymbol{N}^*_{\text{new}}} & \boldsymbol{E} \end{bmatrix} \tag{26}$$

where

$$N_{\text{new}} \equiv N_{\text{post}} - N_{\text{pre}}, \quad \text{and,} \quad N^*_{\text{new}} \equiv \begin{cases} N_{\text{new}} - 1 \;\; \forall \; N_{\text{new}} \leq N_{\text{pre}} \\ \\ N_{\text{pre}} \qquad \text{otherwise} \end{cases} . \tag{27}$$

Furthermore, for arbitrary integers η and μ, $\boldsymbol{I_\eta}$ is an $\eta \times \eta$ identity matrix, $\boldsymbol{0_{\eta\times\mu}}$ is an $\eta \times \mu$ matrix of zeros. k is the following scalar inflation factor

$$k \equiv \sqrt{\frac{N_{\text{new}} + N_{\text{pre}} - 1}{N_{\text{pre}} - 1}} \qquad \text{(note that } k \geq 1\text{)}. \tag{28}$$

The matrix $\boldsymbol{E}$ in Eq. (26) is an $N^*_{\text{new}} \times N_{\text{new}}$ matrix that will be defined shortly. Since $N^*_{\text{new}} < N_{\text{new}}$ [see Eq. (27)], $\boldsymbol{E}$ is a rectangular matrix with more columns than rows. Note that whenever $N_{\text{new}} > N_{\text{pre}}$, the $k\boldsymbol{I}_{\boldsymbol{N}_{\text{pre}}-\boldsymbol{N}^*_{\text{new}}}$ component vanishes from $\boldsymbol{T}$. Furthermore, whenever $N_{\text{new}} = 1$, the $\boldsymbol{I}_{\boldsymbol{N}^*_{\text{new}}}$ and $\boldsymbol{E}$ components vanish from $\boldsymbol{T}$.

Our choice of $\boldsymbol{E}$ is nearly identical to that of Chan, Anderson, and Chen (2020):

$$\boldsymbol{E} \equiv \frac{k-1}{N_{\text{new}}} \boldsymbol{1}_{\boldsymbol{N}^*_{\text{new}}\times\boldsymbol{N}_{\text{new}}} + \boldsymbol{L_E}(\boldsymbol{L_W})^{-1}\boldsymbol{W}. \tag{29}$$

Here, $\boldsymbol{1}_{\boldsymbol{N}^*_{\text{new}}\times\boldsymbol{N}_{\text{new}}}$ denotes an $N^*_{\text{new}} \times N_{\text{new}}$ matrix whose elements are all set to unity. Furthermore, $\boldsymbol{W}$ is an $N^*_{\text{new}} \times N_{\text{new}}$ matrix of the form

$$\boldsymbol{W} \equiv \begin{bmatrix} \boldsymbol{I}_{\boldsymbol{N}^*_{\text{new}}} & \boldsymbol{0}_{\boldsymbol{N}^*_{\text{new}}\times(\boldsymbol{N}_{\text{new}}-\boldsymbol{N}^*_{\text{new}})} \end{bmatrix} - \frac{1}{N_{\text{new}}} \boldsymbol{1}_{\boldsymbol{N}^*_{\text{new}}\times\boldsymbol{N}_{\text{new}}}. \tag{30}$$

Supposing that $\text{Chol}\,(\boldsymbol{S})$ denotes the Cholesky decomposition of an arbitrary symmetric matrix $\boldsymbol{S}$, following appendix B of Chan, Anderson, and Chen (2020), we define

$$\boldsymbol{L_W} \equiv \text{Chol}\left(\boldsymbol{W}\boldsymbol{W}^\top\right), \tag{31}$$

and

$$\boldsymbol{L_E} \equiv \text{Chol}\left(\frac{N_{\text{new}}}{N_{\text{pre}} - 1} \boldsymbol{I}_{\boldsymbol{N}^*_{\text{new}}} - \frac{(k-1)^2}{N_{\text{new}}} \boldsymbol{1}_{\boldsymbol{N}^*_{\text{new}}\times\boldsymbol{N}^*_{\text{new}}}\right). \tag{32}$$

The only difference between the current formulation of $\boldsymbol{E}$ and that of Chan, Anderson, and Chen (2020) lies in the $\boldsymbol{W}$ matrix. In Chan, Anderson, and Chen (2020), $\boldsymbol{W}$ is created from vectors of random white noise. For the ease of parallelization and to ensure replicability (*i.e.*, reruns of the BGEnKF should give the same result), we replaced that stochastic $\boldsymbol{W}$ generation procedure with a deterministic one [*i.e.*, Eq. (30)].

As discussed in Chan, Anderson, and Chen (2020), the resampled perturbations generated by the $\boldsymbol{T}$ defined in Eq. (26) has the property of preserving the pre-resampling perturbations (up to an inflation factor). More specifically, the first $N_{\text{pre}} - N^{*}_{\text{new}}$ resampled perturbations are inflated versions of the first $N_{\text{pre}} - N^{*}_{\text{new}}$ pre-resampling perturbations. The next N^{*}_{new} resampled perturbations are copies of N^{*}_{new} of the pre-resampling perturbations. Finally, the remaining N_{new} resampled perturbations are linear combinations of the copied perturbations.

Outline of three-stage BGEnKF update procedure to assimilate an observation

The outline of the three-stage BGEnKF procedure is as follows. Note that this outline assumes that the members have already been sorted into the clear and cloudy clusters (see the last paragraph of Text S2 for how members are sorted into the two clusters).

Stage 1: Double EnKF [illustrated in Figure 1(a)]

1. Do $g = \text{clr}, \text{cld}$
   (i) For cluster g, compute the Kalman gain [$\boldsymbol{K_g}$; Eq. (13)] and square-root modification factor [ϕ_g; Eq. (20)].
   (ii) Evaluate Eq. (19) for every ensemble member in cluster g.

Stage 2: Shrinking cluster member deletion [illustrated in Figure 1(b)]

1. Evaluate Eq. (21) to determine the targeted cluster sizes after assimilating the observation
2. If $N^{a}_{\text{clr}} < N^{f}_{\text{clr}}$, the clear cluster will be considered as the shrinking cluster.
3. If $N^{a}_{\text{cld}} < N^{f}_{\text{cld}}$, the cloudy cluster will be considered as the shrinking cluster.
4. If no shrinking cluster has been identified, terminate the current stage.
5. Compute N_{del} using Eq. (22).
6. Compute the current mean state of the shrinking cluster.
7. Delete the members with the smallest N_{del} forecast-simulated observation perturbations within the shrinking cluster.
8. Compute the mean state of the remaining members in the shrinking cluster.
9. Subtract the mean computed in step 8 from the mean computed in step 6.
10. Add the difference computed in step 9 to each of the remaining members in the shrinking cluster to recenter said members on the pre-deletion shrinking cluster mean state.

Stage 3: Resample expanding cluster members [illustrated in Figure 1(c)]

1. Evaluate Eq. (21) to determine the targeted cluster sizes after assimilating the observation
2. If $N^a_{\text{clr}} > N^f_{\text{clr}}$, the clear cluster will be considered as the expanding cluster.
3. If $N^a_{\text{cld}} > N^f_{\text{cld}}$, the cloudy cluster will be considered as the expanding cluster.
4. If no expanding cluster has been identified, terminate the current stage.
5. Compute N_{new} and N^*_{new} using Eq. (27).
6. Compute the expanding cluster's mean state vector.
7. Construct the expanding cluster's perturbation vectors via Eq. (23).
8. Construct matrix $\boldsymbol{W}$ by evaluating Eq. (30).
9. Construct $\boldsymbol{L_W}$ and $\boldsymbol{L_E}$ by evaluating Eqs. (31) and (32).
10. Construct $\boldsymbol{E}$ by evaluating Eq. (29).
11. Construct $\boldsymbol{T}$ by evaluating Eq. (26).
12. Evaluate Eq. (24) to resample the expanding cluster perturbations.
13. Add the expanding cluster's mean state (computed in step 6) to the resampled perturbations to construct the resampled expanding cluster ensemble members.

5. Text S5: Outline of the BGEnKF algorithm serial filtering workflow

We will now outline the workflow of the serially assimilating BGEnKF algorithm (illustrated in Figure S2). The serially assimilating BGEnKF algorithm executes the following list of steps.

1. Construct an ensemble of forecast $\boldsymbol{\psi}$ vectors from the prior ensemble using Eq. (4).
2. Select the first observation by setting $m = 1$.
3. Employ the adaptive observation error inflation (AOEI) of Minamide and Zhang (2017) to mitigate representation errors.
4. Extract an ensemble of ξ values from the ensemble of $\boldsymbol{\psi}$ vectors that corresponds to the m-th observation site. Members whose extracted ξ values are smaller than 1 g/m^2 are considered as clear members. The remaining members are considered as cloudy members.
5. Run through the heuristic checks in the main text's sections 2.5.2 and 2.5.3 to determine whether the BGEnKF or its single-kernel form (essentially an EnKF) should be used.
6. If any of the heuristic checks in step 5 fail, put all ensemble members into the clear cluster.
7. Apply the three-stage algorithm described in Text S4 to update the ensemble of $\boldsymbol{\psi}$ vectors.
8. Localize the $\boldsymbol{\psi}$ vector updates using the main text's Eq. (7).
9. Increment m (*i.e.*, $m \leftarrow m + 1$).
10. If there are unassimilated observations remaining, go back to step 3.
11. Extract the model states contained in the ensemble of $\boldsymbol{\psi}$ vectors, output said model states, and terminate the algorithm.

To implement this algorithm with parallelization on the PSU-EnKF system, we employed the low-latency computing cluster strategy proposed by Anderson and Collins (2007). Specifically, every process will receive a sub-domain's worth of model state variables, an entire domain of observation and simulated observation values, and an entire domain of ξ values. To assimilate an observation, each process will then update its sub-domain of model state variables, all of its simulated observations, and all of its ξ values. As such, no inter-process communications are needed within the serial assimilation loop.

6. Text S6: On generalizing the BGEnKF algorithm to handle more clusters

The BGEnKF algorithm can be generalized to handle an arbitrary number of ensemble clusters (*e.g.*, a three-cluster GMM-EnKF). We did not use more than two clusters in this study because this study is a first approach to testing a cluster GMM-EnKF with a realistic weather model. Furthermore, using more clusters means that each cluster will contain fewer members. With smaller cluster sizes, the deleterious impacts of sampling errors on each cluster's sample statistics are likely stronger. Considering the small ensemble size that will be used in this first-approach study (50 members), we opted to use two clusters for now.

To generalize the BGEnKF to handle N_c clusters, only a few modifications are needed: 1) the ensemble clustering method needs to be adjusted to sort the ensemble into the N_c clusters, and 2) a slightly different method would be needed to infer the posterior cluster sizes [Eq. (22)]. The latter modification is necessary because using Eq. (22) with more than 2 clusters can cause the total number of ensemble members to change. This change arises from the use of the rounding function. For instance, suppose we have 3 clusters with equal posterior weights (0.333333 each) and the ensemble size is 10. Using Eq. (22) will result in 3 members in each cluster, or 9 members in total. A different approach to convert the non-integer weights into integer cluster sizes is thus necessary for $N_c > 2$.

References

Anderson, J. L. (2001, 12). An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, *129*(12), 2884–2903. Retrieved from `http://journals.ametsoc.org/doi/10.1175/1520-0493(2001)129%3C2884:AEAKFF%3E2.0.CO;2` doi: 10.1175/1520-0493(2001)129⟨2884:AEAKFF⟩2.0.CO;2

Anderson, J. L. (2003, 4). A Local Least Squares Framework for Ensemble Filtering. *Monthly Weather Review*, *131*(4), 634–642. Retrieved from `http://journals.ametsoc.org/doi/10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2` doi: 10.1175/1520-0493(2003)131⟨0634:ALLSFF⟩2.0.CO;2

Anderson, J. L., & Anderson, S. L. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*. doi: 10.1175/1520-0493(1999)127⟨2741:AMCIOT⟩2.0.CO;2

Anderson, J. L., & Collins, N. (2007, 8). Scalable Implementations of Ensemble Filter Algorithms for Data Assimilation. *Journal of Atmospheric and Oceanic Technology*, *24*(8), 1452–1463. Retrieved from `https://journals.ametsoc.org/view/journals/atot/24/8/jtech2049_1.xml` doi: 10.1175/JTECH2049.1

Anderson, J. L., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., & Avellano, A. (2009, 9). The data assimilation research testbed a community facility. *Bulletin of the American Meteorological Society*, *90*(9), 1283–1296. Retrieved from `https://journals.ametsoc.org/doi/10.1175/2009BAMS2618.1` doi: 10.1175/2009BAMS2618.1

Bishop, C. H., Etherton, B. J., & Majumdar, S. J. (2001, 3). Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Monthly Weather Review*, *129*(3), 420–436. Retrieved from http://journals.ametsoc.org/doi/10.1175/1520-0493(2001)129%3C0420:ASWTET%3E2.0.CO;2https://journals.ametsoc.org/view/journals/mwre/129/3/1520-0493_2001_129_0420_aswtet_2.0.co_2.xml doi: 10.1175/1520-0493(2001)129⟨0420:ASWTET⟩2.0.CO;2

Burgers, G., Jan van Leeuwen, P., Evensen, G., Van Leeuwen, P. J., & Evensen, G. (1998, 6). Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, *126*(6), 1719–1724. Retrieved from http://journals.ametsoc.org/doi/10.1175/1520-0493(1998)126%3C1719:ASITEK%3E2.0.CO;2 doi: 10.1175/1520-0493(1998)126⟨1719:ASITEK⟩2.0.CO;2

Chan, M.-Y., Anderson, J. L., & Chen, X. (2020). An efficient bi-Gaussian ensemble Kalman filter for satellite infrared radiance data assimilation. *Monthly Weather Review*. doi: 10.1175/mwr-d-20-0142.1

Chan, M.-Y., & Chen, X. (2021). Improving Analyses and Forecasts of a Tropical Squall Line using Upper Tropospheric Infrared Satellite Observations. *Advances in Atmospheric Sciences, Accepted Manuscript*. Retrieved from http://www.iapjournals.ac.cn/aas/en/article/doi/10.1007/s00376-021-0449-8http://www.iapjournals.ac.cn/aas/en/article/doi/10.1007/s00376-021-0449-8?viewType=HTML doi: 10.1007/S00376-021-0449-8

Chan, M.-Y., Zhang, F., Chen, X., & Leung, L. R. (2020). Potential Impacts of Assimilating All-sky Satellite Infrared Radiances on Convection-Permitting Analysis and Prediction of Tropical Convection. *Monthly Weather Review*. doi: 10.1175/mwr-d-19-0343.1

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, *99*(C5), 10143–10162. Retrieved from https://doi.org/10.1029/94JC00572http://doi.wiley.com/10.1029/94JC00572 doi: 10.1029/94JC00572

Geer, A. J., & Bauer, P. (2011, 10). Observation errors in all-sky data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *137*(661), 2024–2037. Retrieved from http://doi.wiley.com/10.1002/qj.830 doi: 10.1002/qj.830

Geer, A. J., Migliorini, S., & Matricardi, M. (2019). All-sky assimilation of infrared radiances sensitive to mid- and upper-tropospheric moisture and cloud. *Atmospheric Measurement Techniques Discussions*. doi: 10.5194/amt-2019-9

Harnisch, F., Weissmann, M., & Periáñez. (2016). Error model for the assimilation of cloud-affected infrared satellite observations in an ensemble data assimilation system. *Quarterly Journal of the Royal Meteorological Society*. doi: 10.1002/qj.2776

Hartman, C. M., Chen, X., Clothiaux, E. E., & Chan, M.-Y. (2021, 7). Improving the Analysis and Forecast of Hurricane Dorian (2019) with Simultaneous Assimilation of GOES-16 All-Sky Infrared Brightness Temperatures and Tail Doppler Radar Radial Velocities. *Monthly Weather Review*, *149*(7), 2193–2212. Retrieved from https://journals.ametsoc.org/view/journals/mwre/149/7/MWR-D-20-0338.1.xml doi: 10.1175/MWR-D-20-0338.1

Honda, T., Miyoshi, T., Lien, G. Y., Nishizawa, S., Yoshida, R., Adachi, S. A., … Bessho, K. (2018, 1). Assimilating all-sky Himawari-8 satellite infrared radiances: A case of Typhoon Soudelor (2015). *Monthly Weather Review*, *146*(1), 213–229. doi: 10.1175/MWR-D-16-0357.1

Houtekamer, P. L., & Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *129*(1), 123–137. doi: 10.1175/1520-0493(2001)129⟨0123:ASEKFF⟩2.0.CO;2

Houtekamer, P. L., & Zhang, F. (2016). Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *144*(12), 4489–4532. doi: 10.1175/MWR-D-15-0440.1

Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007, 6). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, *230*(1-2), 112–126. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0167278906004647 doi: 10.1016/j.physd.2006.11.008

Ide, K., Courtier, P., Ghil, M., & Lorenc, A. C. (1997). Unified notation for data assimilation: Operational, sequential and variational. *Journal of the Meteorological Society of Japan*. doi: 10.2151/jmsj1965.75.1B{_}181

Jones, T. A., Skinner, P., Yussouf, N., Knopfmeier, K., Reinhart, A., Wang, X., ... Palikonda, R. (2020, 5). Assimilation of GOES-16 Radiances and Retrievals into the Warn-on-Forecast System. *Monthly Weather Review*, *148*(5), 1829–1859. Retrieved from https://journals.ametsoc.org/view/journals/mwre/148/5/mwr-d-19-0379.1.xml doi: 10.1175/MWR-D-19-0379.1

Lorenc, A. (1986). Analysis methods for numerical weather prediction. *Quart J. R. Mrt. Soc.*, *112*, 1177–1194. doi: 10.1256/smsqj.47413

Meng, Z., & Zhang, F. (2007). Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect model experiments. *Monthly Weather Review*, *135*(4), 1403–1423. doi: 10.1175/MWR3352.1

Minamide, M., & Zhang, F. (2017). Adaptive observation error inflation for assimilating all-Sky satellite radiance. *Monthly Weather Review*, *145*(3), 1063–1081. doi: 10.1175/MWR-D-16-0257.1

Minamide, M., & Zhang, F. (2018). Assimilation of all-sky infrared radiances from Himawari-8 and impacts of moisture and hydrometer initialization on convection-permitting tropical cyclone prediction. *Monthly Weather Review*, *146*(10), 3241–3258. doi: 10.1175/MWR-D-17-0367.1

Otkin, J. A. (2012). Assessing the impact of the covariance localization radius when assimilating infrared brightness temperature observations using an ensemble kalman filter. *Monthly Weather Review*. doi: 10.1175/MWR-D-11-00084.1

Otkin, J. A., & Potthast, R. (2019). Assimilation of All-Sky seviri infrared brightness temperatures in a regional-scale ensemble data assimilation system. *Monthly Weather Review*. doi: 10.1175/MWR-D-19-0133.1

Otkin, J. A., Potthast, R., & Lawless, A. S. (2018). Nonlinear bias correction for satellite data assimilation using taylor series polynomials. *Monthly Weather Review*. doi: 10.1175/MWR-D-17-0171.1

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., & Whitaker, J. S. (2003, 7). Ensemble Square Root Filters. *Monthly Weather Review*, *131*(7), 1485–1490. Retrieved from http://journals.ametsoc.org/doi/10.1175/1520-0493(2003)131%3C1485:ESRF%3E2.0.CO;2 doi: 10.1175/1520-0493(2003)131⟨1485:ESRF⟩2.0.CO;2

Whitaker, J. S., & Hamill, T. M. (2002, 7). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, *130*(7), 1913–1924. Retrieved from http://journals.ametsoc.org/doi/10.1175/1520-0493(2002)130%3C1913:EDAWPO%3E2.0.CO;2 doi: 10.1175/1520-0493(2002)130⟨1913:EDAWPO⟩2.0.CO;2

Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y., & Toth, Z. (2008). Ensemble data assimilation with the NCEP global forecast system. *Monthly Weather Review*, *136*(2). doi: 10.1175/2007MWR2018.1

Zhang, F., Minamide, M., & Clothiaux, E. E. (2016). Potential impacts of assimilating all-sky infrared satellite radiances from GOES-R on convection-permitting analysis and prediction of tropical cyclones. *Geophysical Research Letters*, *43*(6), 2954–2963. doi: 10.1002/2016GL068468

Zhang, F., Minamide, M., Nystrom, R. G., Chen, X., Lin, S.-J., & Harris, L. M. (2019, 7). Improving Harvey Forecasts with Next-Generation Weather Satellites: Advanced Hurricane Analysis and Prediction with Assimilation of GOES-R All-Sky Radiances. *Bulletin of the American Meteorological Society*, *100*(7), 1217–1222. Retrieved from https://journals.ametsoc.org/view/journals/bams/100/7/bams-d-18-0149.1.xml doi: 10.1175/BAMS-D-18-0149.1

Zhang, Y., Sieron, S. B., Lu, Y., Chen, X., Nystrom, R. G., Minamide, M., ... Zhang, F. (2021, 12). *Ensemble-*

Based Assimilation of Satellite All-Sky Microwave Radiances Improves Intensity and Rainfall Predictions for Hurricane Harvey (2017) (Vol. 48) (No. 24). John Wiley and Sons Inc. doi: 10.1029/2021GL096410

Zhang, Y., Zhang, F., & Stensrud, D. J. (2018). Assimilating all-sky infrared radiances from GOES-16 ABI using an ensemble Kalman filter for convection-allowing severe thunderstorms prediction. *Monthly Weather Review*. doi: 10.1175/MWR-D-18-0062.1

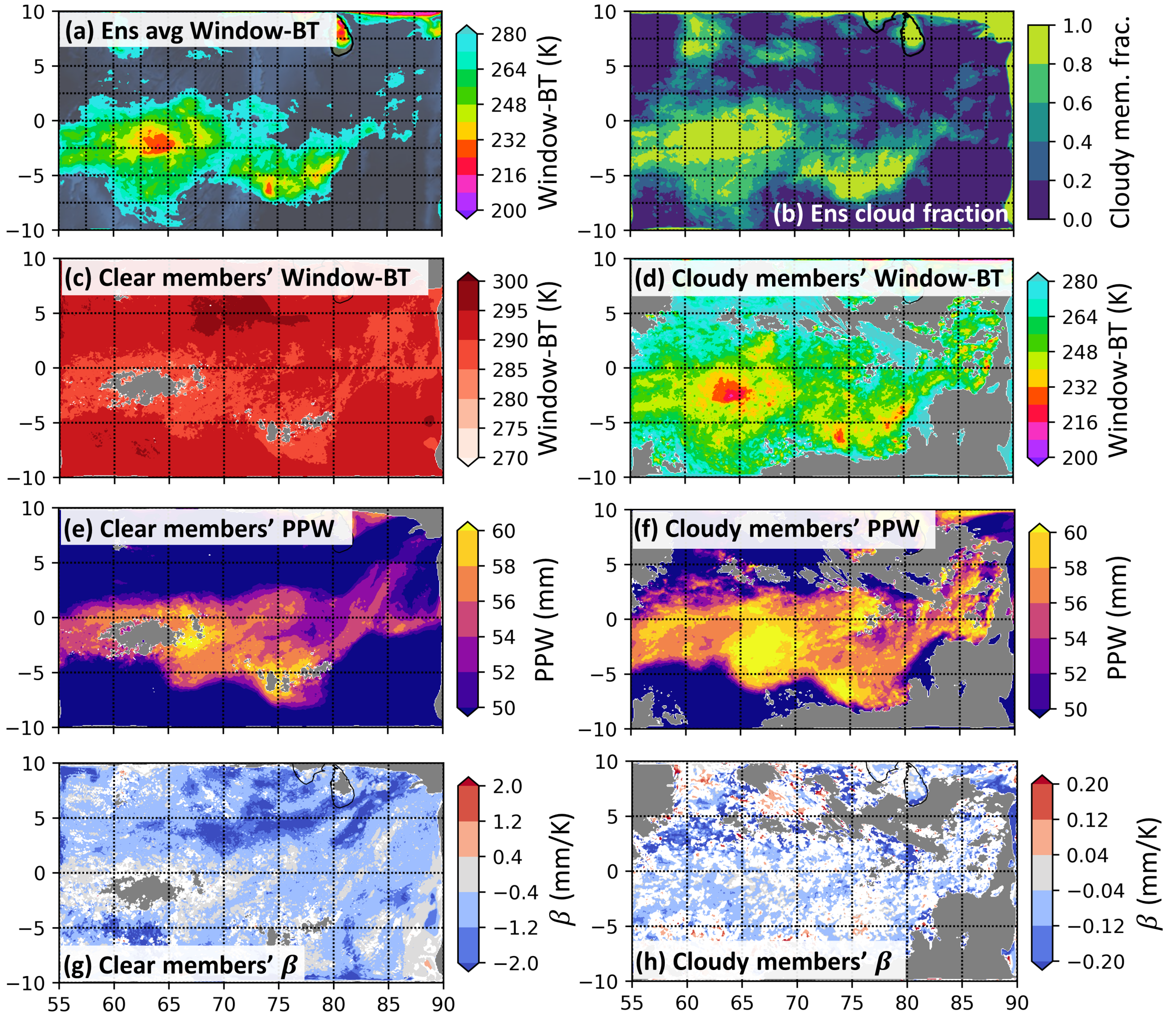


Figure S1: Latitude-longitude plots of various ensemble statistics at 1200 UTC on 15 October 2011 to illustrate the differences between clear and cloudy sky members at every model column. These quantities are generated using the 50-member ensemble described in the main text. The y-axes indicate latitude (degrees North), and the x-axes indicate longitude (degrees East). The plotted quantities are: the prior ensemble mean Window-BT (a), the fraction of cloudy member columns in the prior ensemble at every grid column (b), the mean Window-BTs of clear member columns (c), the mean Window-BT of cloudy member columns (d), the mean pseudo precipitable water (PPW) for clear member columns (e), the mean PPW for cloudy member columns (f), the linear regression coefficient between Window-BT and PPW (β) for clear member columns (g), and the β values for cloudy member columns (h). The gray shadings in panels c, e & g indicate locations where there are either less than 5 clear member columns, the clear member columns' Window-BT sample variance is zero, or the clear member columns' PPW sample variance is zero. The gray shadings in panels d, f & h indicate locations where there are either less than 5 cloudy member columns, the cloudy member columns' Window-BT sample variance is zero, or the cloudy member columns' PPW sample variance is zero. The white shadings in panels g indicate areas where the clear member columns' sample correlation between PPW and Window-BT is statistically insignificant, and likewise for the white shadings in panel h.

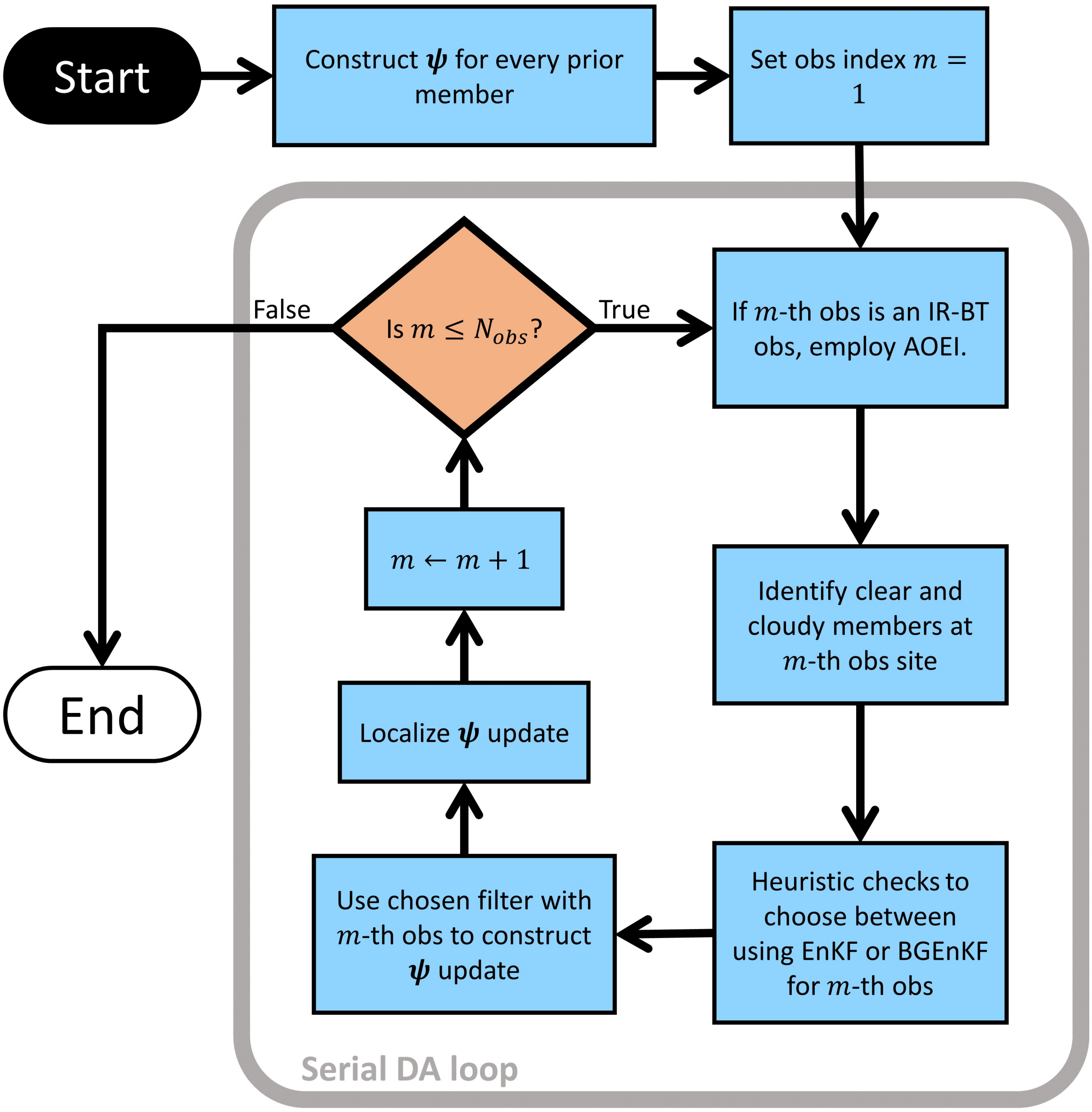


Figure S2: Workflow of the BGEnKF module implemented in the PSU DA system. "Obs" stands for "observations" and N_{obs} stands for the total number of observations. See the text for the definitions of the extended state vector $\boldsymbol{\psi}$ [Eq. (3)], the list of heuristic checks used to select between the EnKF and BGEnKF (main text section 2.5), and for a description of the BGEnKF update procedure (Text S4). The three-stage BGEnKF update procedure is illustrated in Figure 1 of the main text.