Augmented flow-dependent perturbations to mitigate sampling errors: Experiments for a regional application of the NOAA Unified Forecast System

Kenta Kurosawa[a,b] , Jonathan Poterjoy[b]

[a] *Center for Environmental Remote Sensing, Chiba University, Chiba, Japan*

[b] *Department of Atmospheric and Oceanic Science, University of Maryland, College Park, Maryland*

Corresponding author: Kenta Kurosawa, kurosawa@chiba-u.jp

ABSTRACT: Ensemble data assimilation in modern regional weather prediction models often faces challenges in managing sampling errors due to small ensemble size and model errors. Increasing the ensemble size is not often feasible because of the computational resources needed for implementing models with large, high-resolution domains. The current study introduces a strategy for mitigating issues of sampling error in operational data assimilation systems by supplementing ensemble-estimated error covariance needed for data assimilation with perturbations sourced from a global model. This approach resembles hybrid data assimilation methods that use a weighted sum of two background error covariances to mitigate sampling deficiency from ensembles. Specifically, we enhance the NOAA Hurricane Analysis and Forecast System (HAFS) by incorporating an ensemble Kalman filter (EnKF) with augmented perturbations that utilizes flow-dependent perturbations from the Global Data Assimilation System (GDAS) to reduce sampling errors. Additionally, we implement a localized particle filter (LPF) with augmented perturbations, which is not part of the original HAFS data assimilation system, and conduct a comparative analysis of the EnKF with augmented perturbations, the LPF with augmented perturbations, and a hybrid filter that combines the two methods. Experiments that rely on augmented perturbations from GDAS for updating 40-member ensembles are found to produce substantial improvements over benchmark experiments. The new approaches are evaluated over multi-week cycling data assimilation experiments focusing on Hurricanes Laura and Marco from August 2020.

1. Introduction

Sampling errors stemming from ensemble estimates of forecast uncertainty continue to be a substantial challenge for numerical weather prediction (NWP). The computational resources needed for handling large-scale and high-resolution datasets place large constraints on the ensemble afforded for operational prediction. As such, many strategies have been proposed for reducing bias in ensemble statistics caused by sampling errors and assumptions used to form data assimilation algorithms. Heuristic covariance localization and inflation strategies are frequently used methods and have also become widely accepted for large geophysical models (Anderson 2001; Houtekamer and Mitchell 2001; Hamill et al. 2001; Lorenc 2003; Bishop and Hodyss 2011).

One strategy, which is common for operational implementations of variational data assimilation, is the use of a linear combination of flow-dependent (from ensembles) and static (from climatology) background error covariance during data assimilation (Hamill and Snyder 2000; Lorenc 2003; Buehner 2005). The flow-dependent covariances from ensemble-based methods can describe physically consistent, time-variant errors that exhibit anisotropic spatial correlations, which are not easily parameterized from climatology. At the same time, the blending of ensemble and climatological statistics helps reduce sampling errors caused by the small ensemble size. This approach, referred to as hybrid methods, combines variational and ensemble approaches to leverage the advantages of both techniques. Various previous studies have indicated that combining variational and ensemble approaches provides better performance than either method alone; therefore, most major environmental prediction centers use hybrid methods (NCEP; Bannister 2017).

Hybrid data assimilation has been studied extensively over the past decades, and various alternative approaches have been proposed to improve its performance. One such approach is to supplement the dynamic ensemble members from the Ensemble Kalman filter (EnKF; Evensen 1994; Houtekamer and Mitchell 1998; Bishop et al. 2001; Anderson 2001; Whitaker and Hamill 2002) with additional ensemble members derived from climatological perturbations (Lei et al. 2021). This strategy, called the "integrated hybrid EnKF" method, utilizes climatological perturbations to approximate the static forecast error covariance, which allows for updating both the ensemble mean and perturbations with a hybrid error covariance within the EnKF framework.

In this study, our first objective is to assess a conceptually simple but effective strategy for reducing sampling bias in ensemble statistics derived from small regional ensembles, using hurricanes as the target application. As such, we extend the experimental framework in Lei et al. (2021) and employ a similar approach for treating sampling error that is motivated by the use of a hybrid error covariance matrix. For this strategy, flow-dependent error statistics derived from short-range forecasts are augmented by flow-dependent ensemble perturbations generated from a global model. This strategy is conceptually similar to past studies that introduce a hybrid error covariance in EnKFs by blending an ensemble error covariance with a low-rank, sample representation of a climatological error covariance matrix. However, our proposed strategy uses flow-dependent perturbations instead of climatological perturbations, which has the potential to provide more accurate assimilation results by reducing sampling error in pure ensemble estimates. Specifically, to improve the ensemble forecasts from the NOAA Hurricane Analysis and Forecast System (HAFS), we add ensemble perturbations from the Global Data Assimilation System (GDAS) and then estimate the Kalman gain in the EnKF update step with the augmented ensemble members.

Furthermore, ongoing research seeks to apply particle filters (PFs) for state estimation, which are inherently very sensitive to sampling errors (Snyder et al. 2008). Similar to EnKFs, current approaches for mitigating sampling error involve the use of localization and inflation (Poterjoy 2016; Penny and Miyoshi 2016; Potthast et al. 2019; Poterjoy et al. 2019). Similar filter methodologies have recently been extended with the advent of specific techniques, including the use of iterative strategies (Hu and van Leeuwen 2021; Poterjoy 2022b) and the incorporation of Gaussian mixtures (Kotsuki et al. 2022; Rojahn et al. 2023). Combining localized particle filters (LPFs) with EnKFs, which we will denote as "blending PF-EnKF", is another viable approach, as it also provides a bridge between PFs and robust Gaussian-based techniques like EnKFs (Frei and Künsch 2013; Robert et al. 2018; Grooms and Robinson 2021; Poterjoy 2022a; Kurosawa and Poterjoy 2021, 2023). This mixing methodology has already demonstrated promising results in idealized and real applications, particularly in high-dimensional data assimilation problems that are characterized by non-Gaussian errors, where the sole use of either PF or EnKF may have limitations.

In summary, the present study uses an experimental hurricane prediction system to examine new data assimilation strategies for regional weather models. The first component of this research is to explore the utility of augmented perturbations from a global model to reduce sampling deficiency in high-resolution ensembles produced within a limited-area model. The second component implements PFs with augmented perturbations, not originally part of the HAFS system, and compares its performance with the EnKF with augmented perturbations and a hybrid filter combining both methods.

The manuscript is organized in the following manner. Section 2 presents the data assimilation methods used in this study. Section 3 describes the experimental setup, including the model, and observation data used in the experiments. In Section 4, we present and discuss the results of the experiments. In the last section, we summarize our findings and suggest avenues for future research.

2. Background: data assimilation methods

In this section, we provide a mathematical description of each method and introduce the modeling system used for numerical experiments. Notational conventions are as follows: vectors are indicated by lowercase boldface font, matrices are indicated by uppercase boldface font, and scalars and nonlinear operators are indicated by italic font. The background model forecast (or prior) states are represented by an N_x-dimensional vector $\mathbf{x}^f$, while the observations are represented by a N_y-dimensional vector $\mathbf{y}$. The background (or prior) error covariance matrix is represented by the $N_x \times N_x$ dimensional $\mathbf{P}$, and the observation errors are assumed to have zero mean and covariance given by the $N_y \times N_y$ dimensional matrix $\mathbf{R}$. The superscripts f and a correspond to forecast and analysis, respectively.

a. EnKF

The EnKF is a computationally efficient method that is based on the Kalman filter (Kalman 1960) and is designed for moderately nonlinear dynamical systems. Unlike the extended Kalman filter (McElhoe 1966), which is a modification of the original Kalman filter that utilizes the tangent linear model operator to handle nonlinearities in the forecast model or measurement operators, the EnKF does not require the tangent linear model operator. Instead, the EnKF represents the error statistics of $\mathbf{P}$ using a statistical ensemble of model states. This approach bypasses the need for linearizing the forecast model, as in the extended Kalman filter. Consequently, the Kalman gain matrix in the EnKF is derived from the ensemble, facilitating its application to systems with moderate nonlinearity without the explicit use of a tangent linear model. The Kalman gain matrix is described as follows:

$$\mathbf{K} = \mathbf{E}^f \mathbf{D}^{f\top} \left(\mathbf{D}^f \mathbf{D}^{f\top} + \mathbf{R} \right)^{-1}, \tag{1}$$

where $\mathbf{E}^f$ consists of model-space forecast ensemble perturbations and $\mathbf{D}^f$ consists of observation-space forecast ensemble perturbations, with both matrices normalized by $\frac{1}{\sqrt{Ne-1}}$. For the ensemble formulation, the covariance matrix $\mathbf{P}$ can be defined as

$$\mathbf{P} = \mathbf{E}\mathbf{E}^{\mathrm{T}}, \tag{2}$$

$$\mathbf{E} = \frac{1}{\sqrt{N_e - 1}} [\delta \mathbf{x}^{(1)} \mid \cdots \mid \delta \mathbf{x}^{(N_e)}], \tag{3}$$

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

For this study, all algorithms requiring an EnKF to update ensemble members use the serial ensemble square-root filter (serial EnSRF; Whitaker and Hamill 2002). In general, this method provides a deterministic update of the ensemble mean and perturbations about the ensemble mean separately in a manner that satisfies the analysis mean and error covariance given by Kalman filter theory. To avoid large matrix inversions, observations are assumed to have independent errors and assimilated serially. When assimilating a single observation through this formulation, the measurement operators and $\mathbf{K}$ reduce to vectors of length N_x, and $\mathbf{R}$ is a scalar. Therefore, for an individual observation, the computation can be performed even if the measurement operator is fully nonlinear, which is done by applying this operator to each ensemble member before calculating sample statistics.

b. The local PF and mixed filter

In this section, we briefly outline important properties of the iterative local PF proposed by Poterjoy (2022b). This study takes advantage of the unique features of this filter, namely regularization, tempering, and mixing strategies. For those interested in a more detailed description of the methodology, we refer to Poterjoy (2022b) and Kurosawa and Poterjoy (2023).

The local PF assimilates observations with independent errors in a sequential manner and combines sampled particles and prior particles for each observation to introduce localization. The posterior particles can be adjusted in a way that is consistent with bootstrap sampling by updating the particles after each observation space sampling step. The n^{th} updated particle $\mathbf{x}_y^n$ is expressed as a linear combination of the re-sampled particle $\mathbf{x}^{k_n}$, and the prior particle $\mathbf{x}^n$, with $\overline{\mathbf{x}}_y$ representing the localized posterior mean accumulating the full weight of all observations up to y:

$$\mathbf{x}_y^n = \overline{\mathbf{x}}_y + \mathbf{r}_1 \circ (\mathbf{x}^{k_n} - \overline{\mathbf{x}}_y) + \mathbf{r}_2 \circ (\mathbf{x}^n - \overline{\mathbf{x}}_y), \tag{4}$$

where k_n is the index of each sampled particle. Here, the posterior mean is the best estimate of the state after incorporating observations, while the posterior variance reflects the uncertainty of this estimate. These are crucial for ensuring the particle updates align with the statistical properties of the posterior distribution. The coefficients, $\mathbf{r}_1$ and $\mathbf{r}_2$, are designed to maintain the posterior mean and variance, improving the accuracy and stability of the filter.

The iterative PF introduced by Poterjoy (2022b) allows for a "blended" PF-EnKF approach through the use of a mixing parameter, which determines when to switch from a PF to an alternative data assimilation technique that may be more appropriate for specific error distributions. Past research suggests that this mixing approach, especially when the blending coefficient is optimally tuned, can mitigate some of the sampling deficiencies associated with PFs for applications where the posterior distribution more closely resembles a Gaussian than the prior distribution (Kurosawa and Poterjoy 2023; McCurry et al. 2023). Although the mixing parameter can be adjusted independently for each grid point or variable, it is fixed at 0.5 in the current study, based on tests where the parameter was varied from 0 to 1 in increments of 0.25, with 0.5 yielding the best performance.

c. An augmented ensemble data assimilation scheme

The current study proposes a data assimilation system that uses additional ensemble forecast perturbations to reduce sampling deficiency in uncertainty estimates. This approach is largely inspired by the integrated hybrid ensemble Kalman filter with augmented perturbations (IHCEnKF; Lei et al. 2021), but uses samples generated from a global weather prediction system, rather than from a climatological error covariance. The IHCEnKF is a hybrid data assimilation method that combines the benefits of both ensemble and variational methods. It updates both the mean and perturbations in the EnKF framework using a hybrid background error covariance. This method can be implemented without making changes to existing codes, but requires larger ensembles compared to approaches that combine ensemble-based covariances with static climatological covariances, leading to increased computational cost. The scheme proposed in the current study extends the perturbations used to characterize prior uncertainty using flow-dependent data, which is more specific to the system being modeled. Flow-dependent data reflects the current state of the system, whereas climate ensemble data is based on balance constraints and historical weather patterns, which may not accurately reflect the current state of the system—especially for extreme events, such as tropical cyclones.

The current operational HAFS variational data assimilation already uses ensemble perturbations from GDAS to prescribe a flow-dependent background error covariance. Therefore, much of the computational expense for this step already exists in the operational system. This study's novelty is the introduction of GDAS perturbations in an update step, which can be used to cycle a high-resolution ensemble forward in time. In doing so, the GDAS perturbations are used to treat sampling deficiency during the update of ensemble perturbations for future implementations of HAFS that may use a self-cycled ensemble.

The dynamics governing hurricane track and intensity span a large number of scales, which require large domains and a convective-permitting resolution. These limitations restrict the ensemble size affordable for HAFS—which in this study, uses 40 members. Data assimilation steps performed in a prototype "basin-wide" version of HAFS use a background error covariance matrix $\mathbf{P}^f_{HAFS}$, which is solely derived from the 40-member HAFS ensemble. In the newly proposed system, $\mathbf{P}^f$ is replaced with a mixed covariance matrix $\mathbf{P}^f_{mix}$, which takes into account both the HAFS and GDAS ensembles to reduce sampling error with a small HAFS ensemble. $\mathbf{P}^f_{mix}$ is a weighted average of the perturbations from both ensembles:

$$\mathbf{P}^f_{mix} = \frac{N_{\text{HAFS}}}{N_{\text{HAFS}} + N_{\text{GDAS}}}\mathbf{P}^f_{\text{HAFS}} + \frac{N_{\text{GDAS}}}{N_{\text{HAFS}} + N_{\text{GDAS}}}\mathbf{P}^f_{\text{GDAS}}, \tag{5}$$

where N and the superscript f indicate the ensemble size and forecast, respectively.

This approach can be extended to PFs by introducing 80 additional particles that are generated by adding GDAS perturbations to the HAFS background mean. It is important to note that regardless of whether the augmented $\mathbf{P}^f$ is used, the prior ensemble mean at each data assimilation step is always the HAFS ensemble mean, which is crucial to maintain for regional models that operate at high resolution (Schwartz et al. 2022). Each ensemble perturbation from GDAS is produced by removing the original mean and centering the perturbations on the prior ensemble mean from HAFS at each analysis time. We also note that this approach ensures a balanced contribution from both ensemble systems, but may not be optimal for storm-scale data assimilation. GDAS perturbations are derived from a relatively low-resolution global model, and may not adequately represent storm-scale characteristics. If GDAS perturbations are given high weight, the high-resolution information provided by the HAFS ensemble may be unintentionally diluted, and the constructed background error covariance may limit its ability to capture local storm characteristics. Since the current study naively weights each set of perturbations based on the respective ensemble size, exploring alternative weighting strategies could potentially yield better results. Further information on HAFS and GDAS is provided in section 3a.

3. Configuration for numerical experiments

a. Models and experimental design

The GDAS uses a hybrid 4D-ensemble-Var (4DEnVar; Liu et al. 2008) configuration with 80 ensemble members to initialize the Global Forecast System (GFS) model. The GFS is used for medium-term numerical weather predictions in the US and it is built with the Geophysical Fluid Dynamics Laboratory (GFDL) Finite-Volume Cubed-Sphere (FV3) Dynamical Core and the Grid-Point Statistical Interpolation (GSI) data assimilation system. HAFS is also based on the FV3 and aims to provide operational analysis and forecasting with reliable and skillful guidance on tropical cyclone (TC) track and intensity, storm size, and weather hazards associated with TCs. Past configurations of HAFS have included a uniform global model with a high-resolution nest and a stand-alone high-resolution regional model (Dong et al. 2020; Hazelton et al. 2021; Gopalakrishnan et al. 2021). The HAFS modeling system can be initialized in two ways: (1) a "cold start," which is an initialization from GDAS, or (2) a warm start of the current forecast cycle, which sources data from its preceding forecast cycle to be used as a background, then performs its own data assimilation. The current operational HAFS system incorporates a 6-hourly four-dimensional ensemble variational (4DEnVar) method that employs the GDAS ensemble, relying exclusively on the background error covariance obtained from the 80-member GDAS without incorporating any static error covariance. This configuration only performs data assimilation over a high-resolution nest that covers TC vortices. Meanwhile, a prototype HAFS data assimilation system adopted for this research uses three-dimensional ensemble variational data assimilation (3DEnVar), with an ensemble updated using an EnKF, which derives the Kalman gain solely from the HAFS ensemble. Unlike the operational data assimilation system, the version used for this study performs data assimilation over a single fixed domain, which encompasses storms and their environments. In the current study, as highlighted in the previous subsection, we compare this configuration to one that uses both HAFS and GDAS ensembles to mitigate sampling errors inherent to the smaller HAFS ensemble.

The current study utilizes a fixed regional model configuration, with a model domain consisting of a single grid of 1440 $\times$ 1080 horizontal grid cells with a 6-km horizontal grid spacing, 81 vertical levels, and employing the FV3 dynamical core. The physics suite used in our configuration of HAFS resembles the GFS version 16 physics configuration but with modified boundary layer physics that is specific to TCs. The atmospheric model is also coupled with the Hybrid Coordinate Ocean Model (Chassignet et al. 2007, 2009). Identical to past operational configurations of the NOAA Hurricane Weather Research and Forecasting (HWRF) model, HAFS assimilates both conventional and satellite observations, in addition to hurricane-specific measurements collected from aircraft reconnaissance flights (Tong et al. 2018).

We conduct experiments using interpolated boundary conditions and initial conditions (for the first cycle) from the 2020 operational global FV3-based GFS, which has a horizontal spacing of approximately 13 km. Our configuration of HAFS uses a 6-km horizontal grid spacing for both deterministic and ensemble states. Each experiment covers a period of three weeks, from 00 UTC 11 August 2020 to 1800 UTC 29 August 2020, encompassing the entire life cycle of two landfalling hurricanes, AL13 (Laura) and AL14 (Marco), in the Gulf of Mexico (Fig. 1). We consider the first week as a spin-up period and use the results from the remaining weeks for validation.

b. Data assimilation configuration

The current study assimilates observations every 6-h to update a 40-member HAFS ensemble. Using results from experiments performed with conventional EnKF as a benchmark, or “control” (Fig. 2a), this study tests the local PF, the blending PF-EnKF, and their application with the augmented perturbations (Fig. 2b). Experiments that use augmented perturbations are denoted by a ”G” at the end of their name; i.e.,”EnKF-G,” ”PF-G,” and ”PF-EnKF-G.”

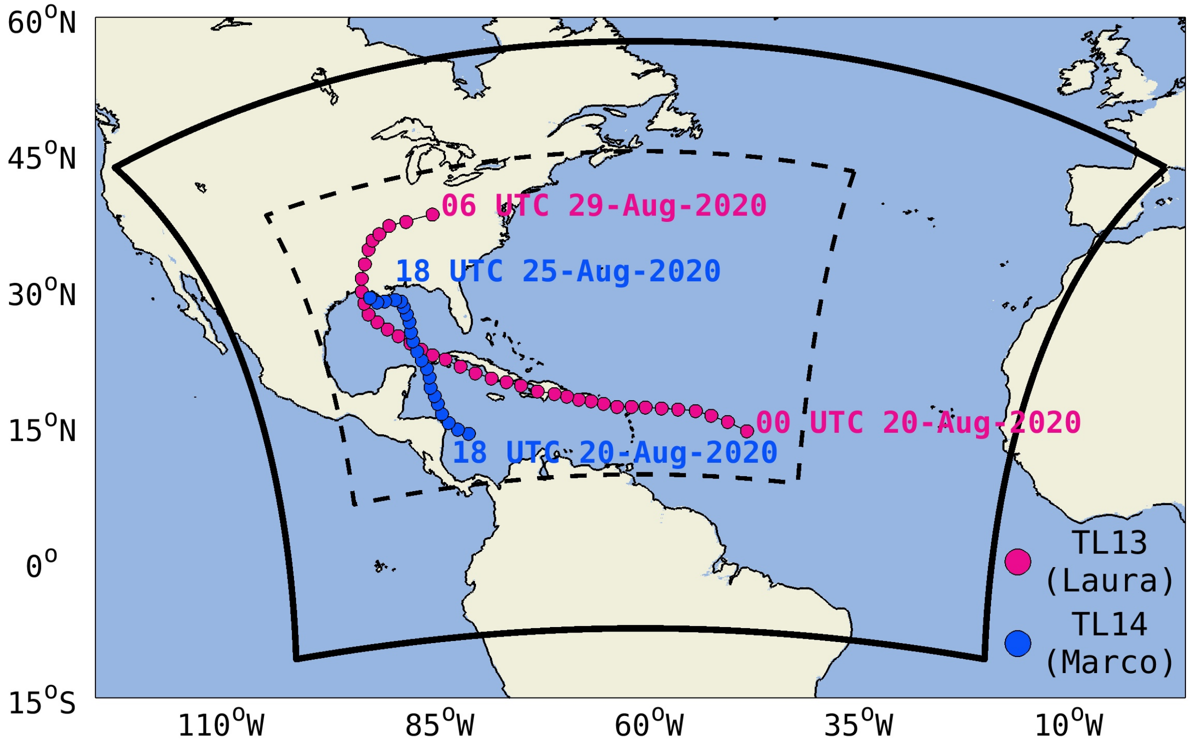


FIG. 1. HAFS domain with black bold lines outlining the boundaries, and the paths of hurricanes Laura (red) and Marco (blue). The black dashed lines indicate the area used for calculating the ERA5 scores in Fig. 7.

We implement augmenting perturbations from 80-member GDAS to improve sampling deficiency in 40-member HAFS ensemble data assimilation, i.e., $N_{\text{HAFS}} + N_{\text{GDAS}} = 120$. The ensemble perturbations from 6-h GDAS forecasts are introduced into the assimilation step by replacing the ensemble mean of the GDAS ensemble with the mean of the HAFS ensemble. This step occurs before data assimilation each analysis time. This design choice also permits the use of the local PF with augmented perturbations, which now uses an additional 80 members to estimate prior densities.

Localization is applied to reduce sampling noise introduced from ensemble error approximations. Horizontal and vertical localization length scales are set to $\sigma_h = 500$ km and $\sigma_v = 0.5$ natural log pressure, respectively. This study uses relaxation to prior spread (RTPS; Whitaker and Hamill 2012) to further help maintain ensemble spread for experiments that use the EnKF. The relaxation parameter for all variables is set to $\alpha = 0.95$, which was chosen based on tuning carried out through shorter experiments not shown in this paper. In Sec. 4b-d, the blending PF-EnKF uses a mixing coefficient (κ) of 0.5, implying an equal utilization of PF and EnKF updates.

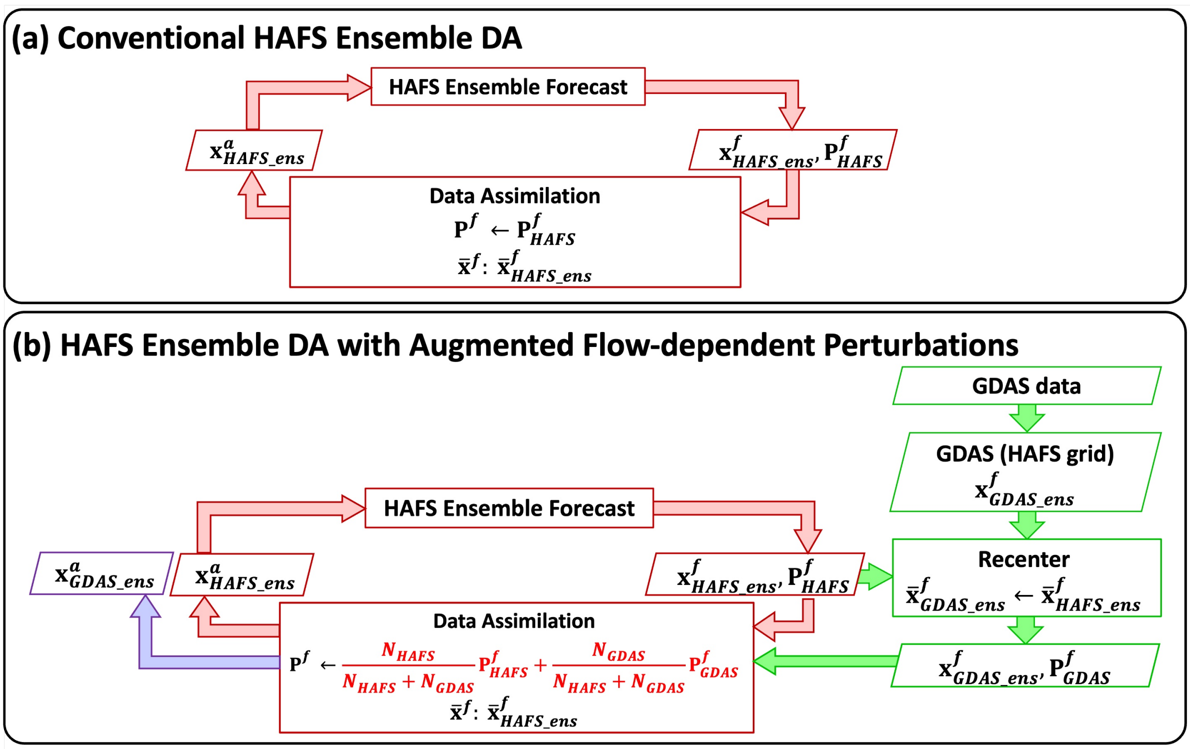


FIG. 2. Flowcharts of two experimental methods: (a) the conventional HAFS ensemble data assimilation experiment cycle and (b) the experiment cycle using the augmented flow-dependent perturbations. In (b), the forecast error covariance matrix $\mathbf{P}^f$ is replaced by a matrix of comprised of weighted HAFS and GDAS perturbations. A new step shown in green represents the re-centering of the ensemble mean using additional 80-member perturbations, and the purple process represents the ensemble analysis of GDAS, which is only used for data assimilation.

4. Experimental results from cycling experiments

a. Assessment of the augmented perturbations with EnKF

To assess the effectiveness of the proposed augmented ensemble data assimilation system, we conduct cycling experiments using EnKF with and without the augmented perturbations, denoted as EnKF-G and EnKF, respectively. For comparison, we examine prior ensemble spread in our cycled data assimilation experiments with values computed from the first 40 GDAS ensemble members. Therefore, the objective is to show how the ensemble spread of EnKF with augmented perturbations in previous data assimilation steps manifests as a change in prior spread over multiple weeks of data assimilation.

Figure 3 presents a comparison of the prior ensemble spread of temperature, u-wind, and specific humidity at 850 hPa for EnKF, EnKF-G, and GDAS averaged from 00 UTC on August 18 to 18 UTC on August 29, 2020. It is evident that GDAS exhibits the largest ensemble spread among the three, which is not surprising given that GDAS uses a larger ensemble size and stochastic physics to induce spread in the ensemble—with the latter not being a feature of the HAFS version adopted for this research. The EnKF experiment shows the smallest spread. The EnKF-G experiment falls between the two, indicating that the augmented perturbations lead to a moderate increase in ensemble spread.

To further investigate the impact of the augmented perturbations on the vertical structure of ensemble spread, we calculate the prior ensemble spread profiles of temperature, u-wind, and specific humidity from 0 hPa to 1000 hPa for EnKF, EnKF-G, and GDAS (Fig. 4). The results show that GDAS exhibits the largest ensemble spread throughout the entire vertical domain, while the EnKF experiment exhibits the smallest spread. The spread profiles for EnKF-G are consistently larger than those for EnKF across the vertical domain, which we will note in future sections to be crucial to obtaining more skillful analyses. Therefore, the results of our cycling experiments show that the proposed augmented ensemble data assimilation system increases the ensemble spread in a manner that is expected given our choices for weighting HAFS and GDAS perturbations in Kalman gain calculations, and the lack of stochastic physics in the HAFS ensemble.

b. Error verification in the observation space

To further examine the impact of using augmented perturbations to supplement prior uncertainty estimation during data assimilation (Figs. 3 and 4), we assess observation-space errors from our control EnKF experiment with those from EnKF-G, PF-G, and PF-EnKF-G. Figure 5 shows the vertical profiles of prior root-mean-square differences (RMSDs) and bias using all temperature and wind measurements assimilated between 1000 to 50 hPa, averaged from 00 UTC on August 18 to 18 UTC on August 29, 2020. Panels (a) and (b) indicate that EnKF has the largest RMSD, especially in the range from 600 hPa to 200 hPa, where many upper-air measurements (e.g., satellite-derived

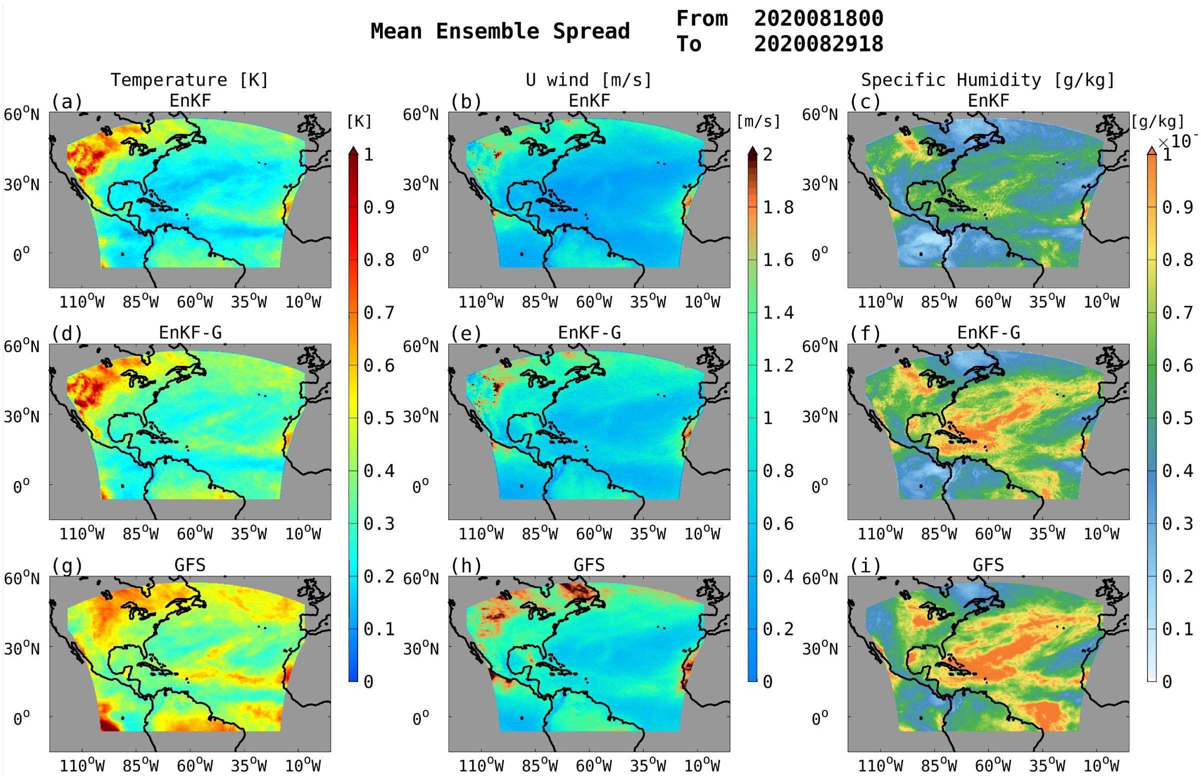


Fig. 3. Average prior ensemble spread of temperature, u-wind, and specific humidity at 850 hPa for (a-c) EnKF, (d-e) EnKF-G, and (g-i) GFS, averaged from 00 UTC on August 18 to 18 UTC on August 29, 2020.

atmospheric motion vectors and in situ data from aircraft) exist. Nevertheless, the benefits are much smaller in the lower troposphere, where HAFS exhibits a large increase in bias near the surface. Panels (c) and (d), representing wind RMSD and bias respectively, show that EnKF and PF-G display higher errors, with both methods indicating higher RMSDs and biases compared to the other techniques. In contrast, EnKF-G and PF-EnKF-G show improved scores, suggesting these methods are more effective in reducing both random errors and biases.

Figure 6 presents a time series of domain-average prior RMSDs and "total error" in observation space, calculated from prior members at each 6-hour data assimilation cycle. Total error is defined as the square root of the sum of the observation error variance and ensemble variance of the simulated observations (Houtekamer and Mitchell 2005). Ideally, total error should be equal to RMSDs, as it quantifies the expected standard deviation of the ensemble mean departures from noisy observations. First, focusing on the coefficient of total error to RMSDs, it is clear that the total

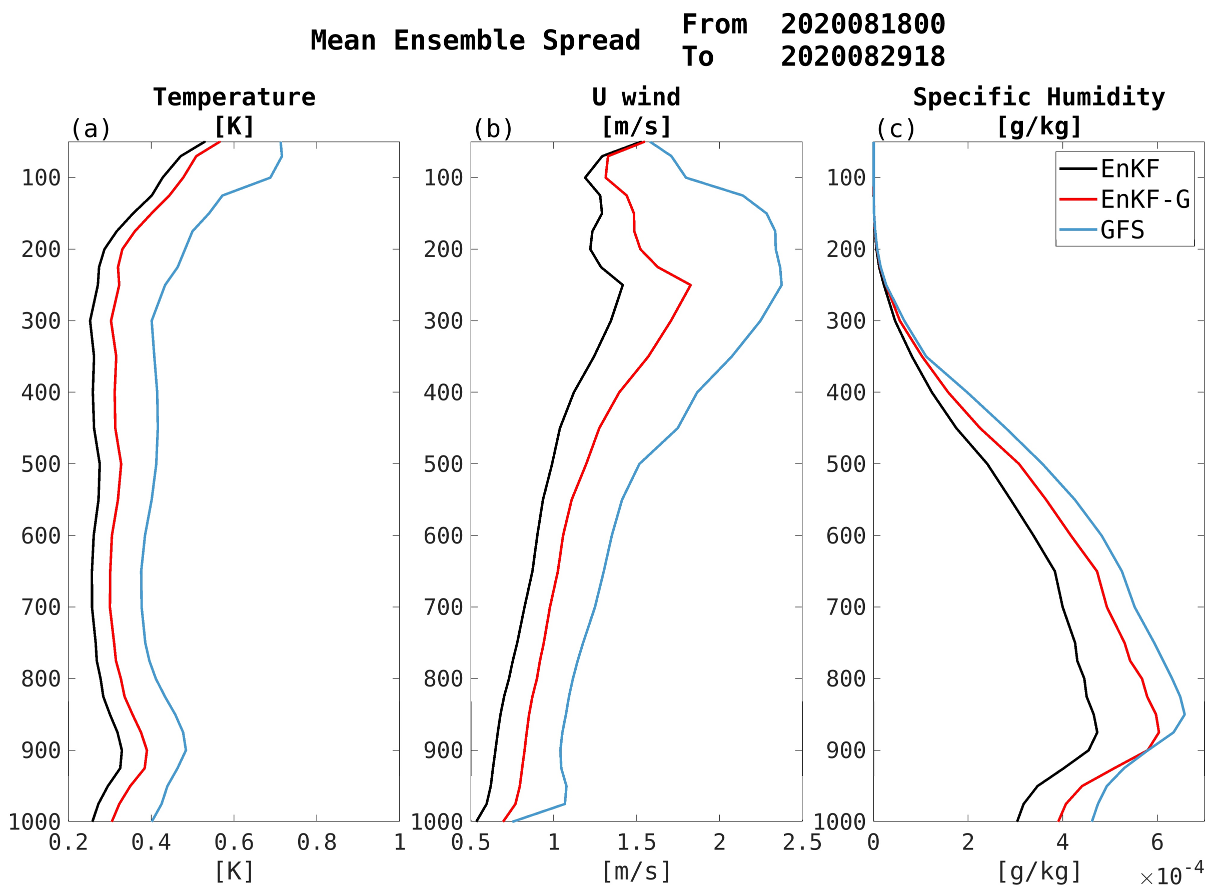


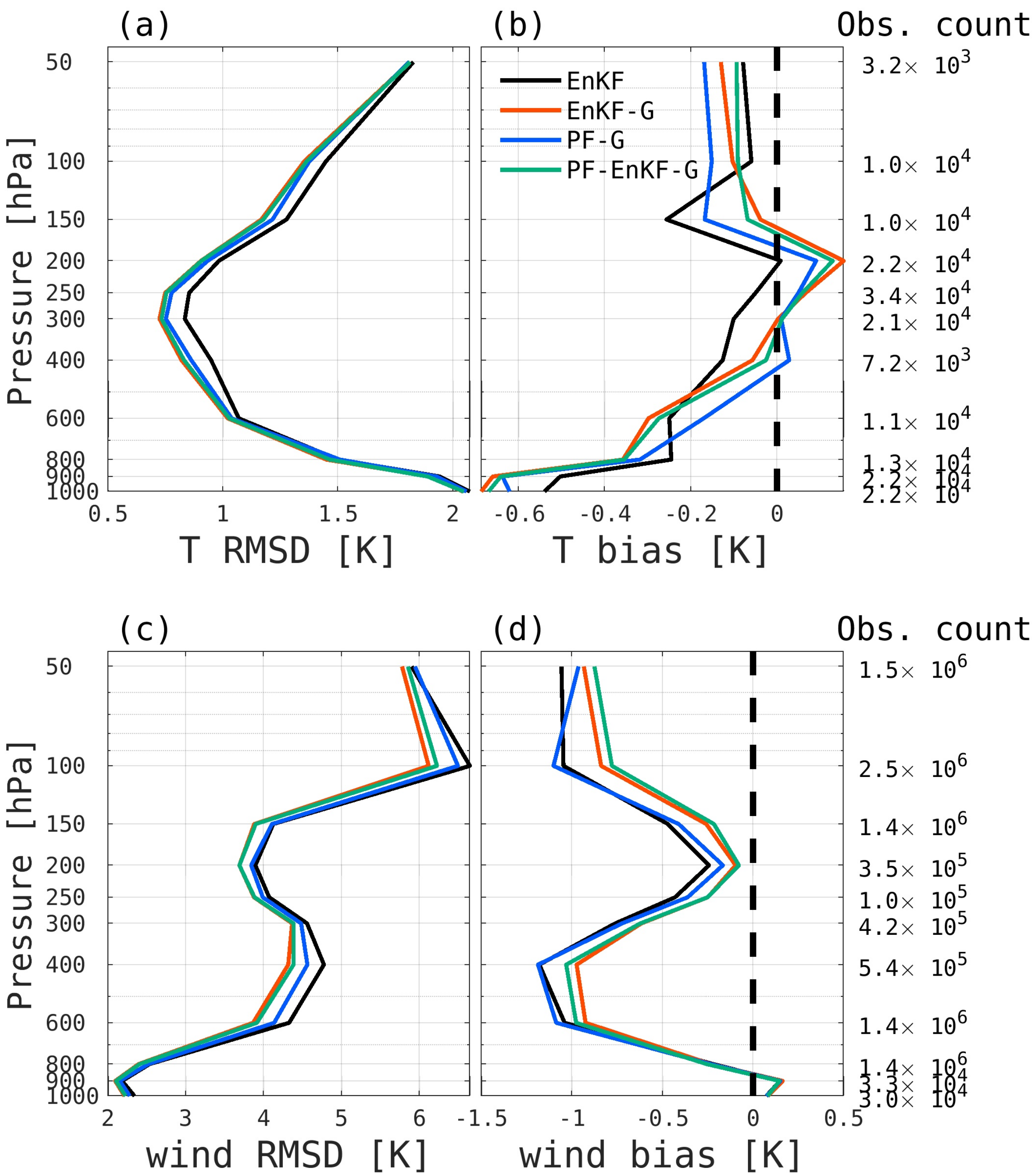
FIG. 4. Average prior ensemble spread profile of (a) temperature, (b) u-wind, and (c) specific humidity for EnKF (black), EnKF-G (red), and GFS (blue), respectively. These values were calculated throughout the cycle experiments conducted from 00 UTC on August 18 to 18 UTC on August 29, 2020.

error is underestimated for all variables except wind. In contrast, for wind, the total error is slightly overestimated in relation to the RMSDs, which we speculate is caused partly by an over-prescribed observation uncertainty for some of the verifying measurements. Further tuning the methods used to control ensemble spread, such as inflation and relaxation, could potentially mitigate this problem. Nevertheless, our attempts to increase spread without increasing RMSDs resulted in the current configuration, which also resembles specifications used in past studies for HWRF (Poterjoy et al. 2021). Furthermore, comparing RMSDs and total error across experiments shows that the augmented experiments generally have reduced RMSDs and a more appropriate representation of uncertainty for prognostic variables–with respect to the benchmark EnKF experiment. This verification confirms that extending ensemble perturbations to include samples from GDAS proves effective at reducing sampling deficiency in HAFS.

c. Domain-averaged error verification

This section compares prediction accuracy for the entire domain using the ECMWF Reanalysis v5 (ERA5) data, which incorporates 4DVar data assimilation and comprehensive observational data sets, including all-sky radiances (Hersbach et al. 2020). For this comparison, we look at temperature, wind, specific humidity, and absolute vorticity from August 18 to August 29, 2020. To mitigate the impact of boundary conditions, the domain over which the scores are calculated has been deliberately restricted to a smaller area, well within the interior of the larger domain, ensuring that the calculations reflect the dynamics less influenced by the boundaries. This area is represented by the black dashed lines in Fig. 1.

Figure 7 shows the average RMSDs, ensemble spread, and bias relative to ERA5 of the variables, with the forecast lead time on the horizontal axis from hour 0 to hour 102. An examination of RMSDs for temperature, winds, and specific humidity at short lead times (Fig. 7a–c) shows that PF-G has larger errors compared to the other methods that use augmented perturbations. This finding is not surprising, given that Kalman filter-based data assimilation techniques aim to find an analysis that minimizes mean squared errors, which favors a verification of RMSD near analysis times. However, as the forecast time increases, this difference diminishes, and the PF-G obtains comparable skill to the other experiments that use augmented perturbations. Likewise, the ensemble spread for PF-G is consistently the largest among the tested methods, though we note that all methods produce under-dispersed forecasts. As discussed in Poterjoy (2022a), posterior members produced by the LPF tend to undergo smaller geostrophic adjustment following initiation, which results in a more steady increase in ensemble spread at short lead times compared to EnKFs. Furthermore, each method that uses a full or partial EnKF update (with augmented perturbations) shows little difference in RMSD or ensemble spread across variables. Examining the bias in the forecasts, we find that the EnKF exhibits the largest bias overall. PF-G follows closely behind, even surpassing the EnKF bias in wind speed forecasts at lead times shorter than 48 hours. However, as the forecast lead time increases, the bias in PF-G becomes comparable to other methods using augmented perturbations (EnKF-G and EnKF-PF-G).



334 FIG. 5. Vertical profiles of RMSD and bias for temperature and wind in the observation space using different data assimilation methods: (a) temperature RMSD, (b) temperature bias, (c) wind RMSD, and (d) wind bias, evaluated with EnKF (black), EnKF-G (red), PF-G (blue), and PF-EnKF-G (green). These values were calculated throughout the cycle experiments conducted from 00 UTC on August 18 to 18 UTC on August 29, 2020.

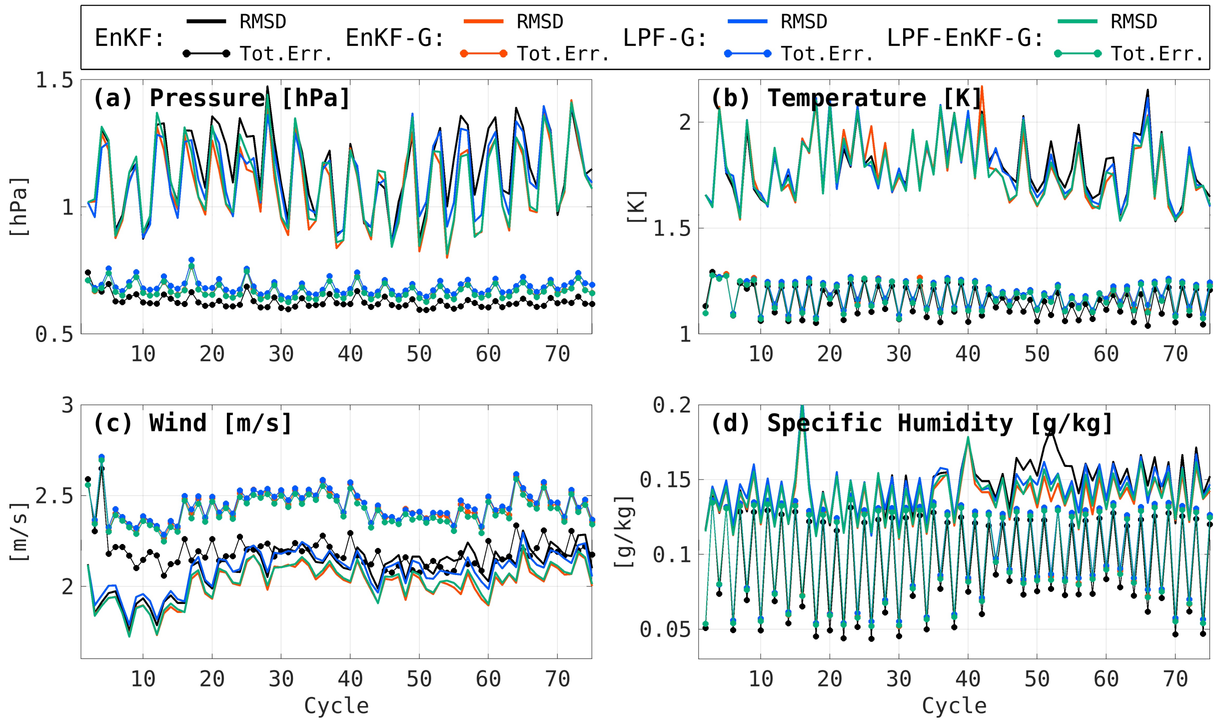


FIG. 6. Time series of domain-average prior RMSDs (solid lines) and total error (dotted lines) in the observation space of (a) pressure, (b) temperature, (c) wind, and (d) specific humidity, for EnKF (black), EnKF-G (red), PF-G (blue), and PF-EnKF-G (green).

We note that PF-G persistently produces the smallest errors in vorticity at all lead times. The markedly larger RMSEs in EnKF and mixed filter experiments at early lead times come from spuriously large wind gradients in EnKF analyses, which is consistent with coarse-resolution regional modeling experiments performed by Poterjoy (2022a). Likewise, these experiments result in a notable drop in error over the first 12 h as the model adjusts to wind analyses that are not supported by basic horizontal momentum balance. These effects are more notable for ensemble spread, which continues to drop over the first 48 h. Because EnKF adjustments to ensemble perturbations are also modified by posterior inflation, we suspect that this behavior is partly induced by the chosen inflation mechanism, namely, RTPS. Whitaker and Hamill (2012) note that RTPS tends to induce spread over a larger wave spectrum than alternative relaxation-based techniques—but at the expense of maintaining dynamical balance. As a result, many of the non-physical impacts of data assimilation on vorticity are removed during the model's integration leading to a decrease in ensemble spread at early forecast times.

In terms of bias, we observe an oscillating pattern. This behavior is likely due to the sensitivity of vorticity to spatial gradients in wind speed, making it susceptible to model instabilities and adjustments. Specifically, the model's resolution can contribute to this behavior. If the resolution is too coarse to accurately capture fine-scale variations in wind speed, it can introduce instabilities in the vorticity forecast, potentially causing the bias to oscillate as the forecast lead time increases.

Lastly, focusing further on the pure EnKF experiment that does not use the augmented approach, it exhibits the smallest ensemble spread and the largest RMSD values on average for all variables in Fig. 7. This suggests that the strategy aimed at addressing the sampling deficiency in ensemble statistics has the potential to enhance the effectiveness of various regional ensemble data assimilation systems.

d. Tropical Cyclone Forecast Verification

In this section, we verify model forecasts based on TC-specific metrics that are available from the National Hurricane Center (NHC) "best track" database, namely track, maximum 10-m surface winds, and minimum sea level pressure (MSLP). These comparisons examine 10-member ensemble forecasts that are initialized during times when AL13 (Laura) and AL14 (Marco) were of a tropical storm or greater intensity.

Figure 8 presents RMSDs and ensemble spread for the track and intensities of wind speed and MSLP throughout the experiments for AL13 and AL14. Experiments with the augmented perturbations consistently provide smaller RMSDs relative to pure EnKF, suggesting that the proposed approach is effective for both EnKFs and PFs. While PF-G is less skillful than EnKF-G and PF-EnKF-G, its performance is more sensitive to challenges related to the model's representation of rapid intensification and the use of the 6-km grid spacing, which is relatively coarse. By successfully integrating the advantages of EnKF-G, PF-EnKF-G achieves improved results over PF-G and qualitative benefits over EnKF-G that are described below. Consistent with the domain-wide verification, the EnKF without augmented perturbations produced the smallest spread while PF-G produced the largest–with the other two experiments falling between.

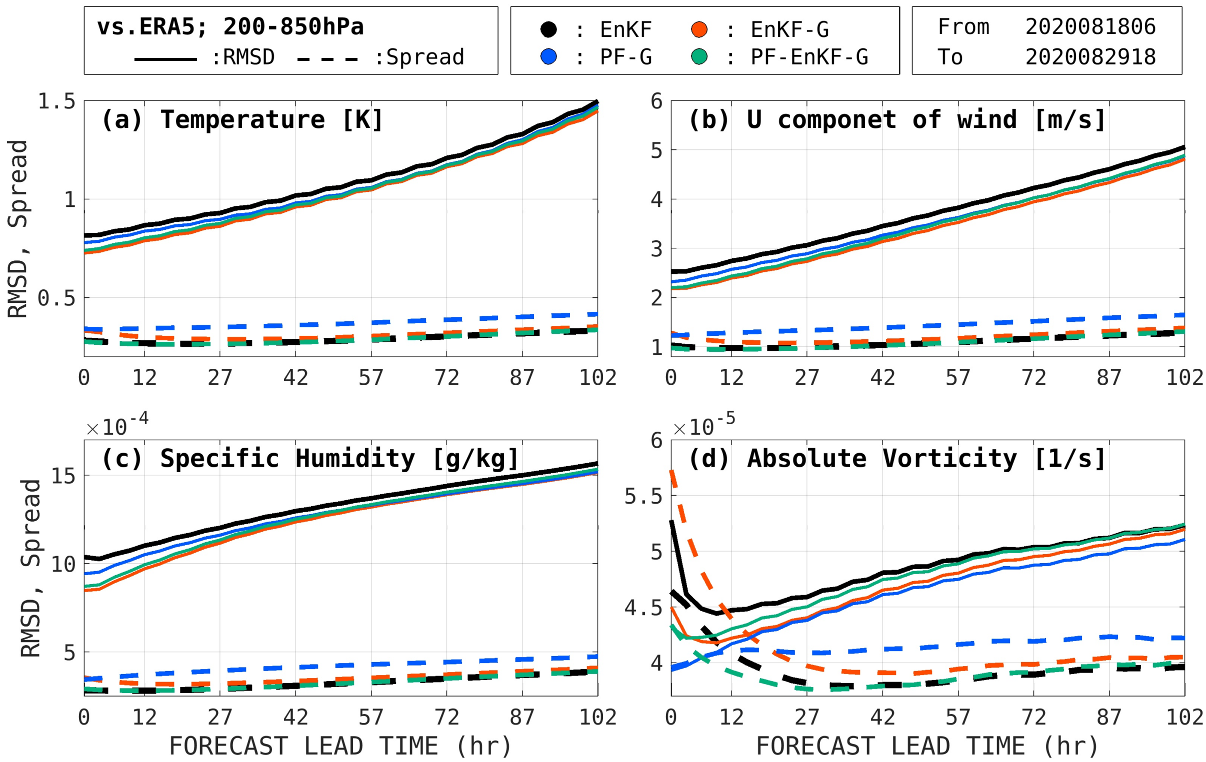


FIG. 7. Average RMSDs relative to ERA5 of (a) temperature, (b) u-wind, (c) specific humidity, and (d) absolute vorticity for EnKF (black), EnKF-G (red), PF-G (blue), and PF-EnKF-G (green), averaged from 00 UTC on August 18 to 18 UTC on August 29, 2020. The solid lines represent RMSD, and the dashed lines represent ensemble spread.

While track and error verifications highlight the overall performance of each method, a closer examination of AL13 and AL14 reveals further insight into the strengths and weaknesses of each data assimilation approach. First, we examine the case of Laura (Fig. 9). Ensemble forecasts generated during the control EnKF experiment predict the track of the TC accurately over most of the storm's life cycle but struggle in the early cycles to capture Laura's initial intensification. On the other hand, EnKF-G, which uses augmented perturbations, improves the accuracy of storm track and intensity during the early cycles and persistently produces an envelope of forecast solutions that capture the observed TC characteristics. This experiment, however, tends to produce spuriously large increases in winds during analyses, which rapidly decay in the first 6 h of forecasts; see sharp decrease in max winds in Fig. 8b. The spuriously large analysis winds are directly related to the large domain-wide vorticity errors highlighted in Fig. 7d, which tend to be restricted to smaller scales.

The PF-G experiment shows similar improvements over the control, primarily in early track forecasts for Laura, but tends to be less skillful than EnKF-G when the data assimilation needs to correct for major intensity errors (e.g., during Laura's rapid intensification period). This finding is a re-occurring challenge for PF-G in our experiments and is rather expected given that the LPF cannot easily shift members outside the span of the prior in the same manner as the EnKF. The issue is partially related to our use of a 6-km grid spacing for FV3, which leads to a low-intensity bias in our experiments. PF-EnKF-G, which combines both EnKF and PF equally, shows similar skill to EnKF-G, but avoids the spurious small-scale wind anomalies that dominate RMSD verifications at early lead times (Fig. 8b).

We observe similar advantages in the case of Marco as well (Fig. 10). Marco differed from Laura in that it was both a smaller and shorter-lived storm. These reasons are one factor that led to our control experiment missing its intensification into a tropical cyclone altogether (Figs. 10a–b). Using the augmented GDAS perturbations (EnKF-G and PF-G) significantly improves the ability of the filters to accurately adjust model states towards a realistic depiction of Marco, as is evident in the improved track forecasts. Following the same explanation provided for Laura in these experiments, we also note that PF-G (while improved over the EnKF) still shows difficulty matching the observed intensity of Marco. EnKF-G more rapidly spins up TC vortices, but shows signs of large spurious adjustments, as indicated by significant reductions in maximum wind speed following most forecast times (Fig. 10d), a problem that is again alleviated by PF-EnKF-G (Fig. 10h).

e. Spectral analysis of kinetic energy

To supplement our comparison of ensemble forecast skill, we scrutinize each data assimilation strategy's ability to produce analysis members that resemble model solutions. Figure 11 depicts the zonal kinetic energy (KE) at 250 hPa derived from a single analysis member of EnKF-G, PF-G, PF-EnKF-G. Results are averaged in the meridional direction and temporally before being plotted on a logarithmic scale with wavelengths decreasing to the right. The analysis KE spectra are compared to a climatological estimate for the FV3 model using 24 h forecasts from August 18 to August 29,

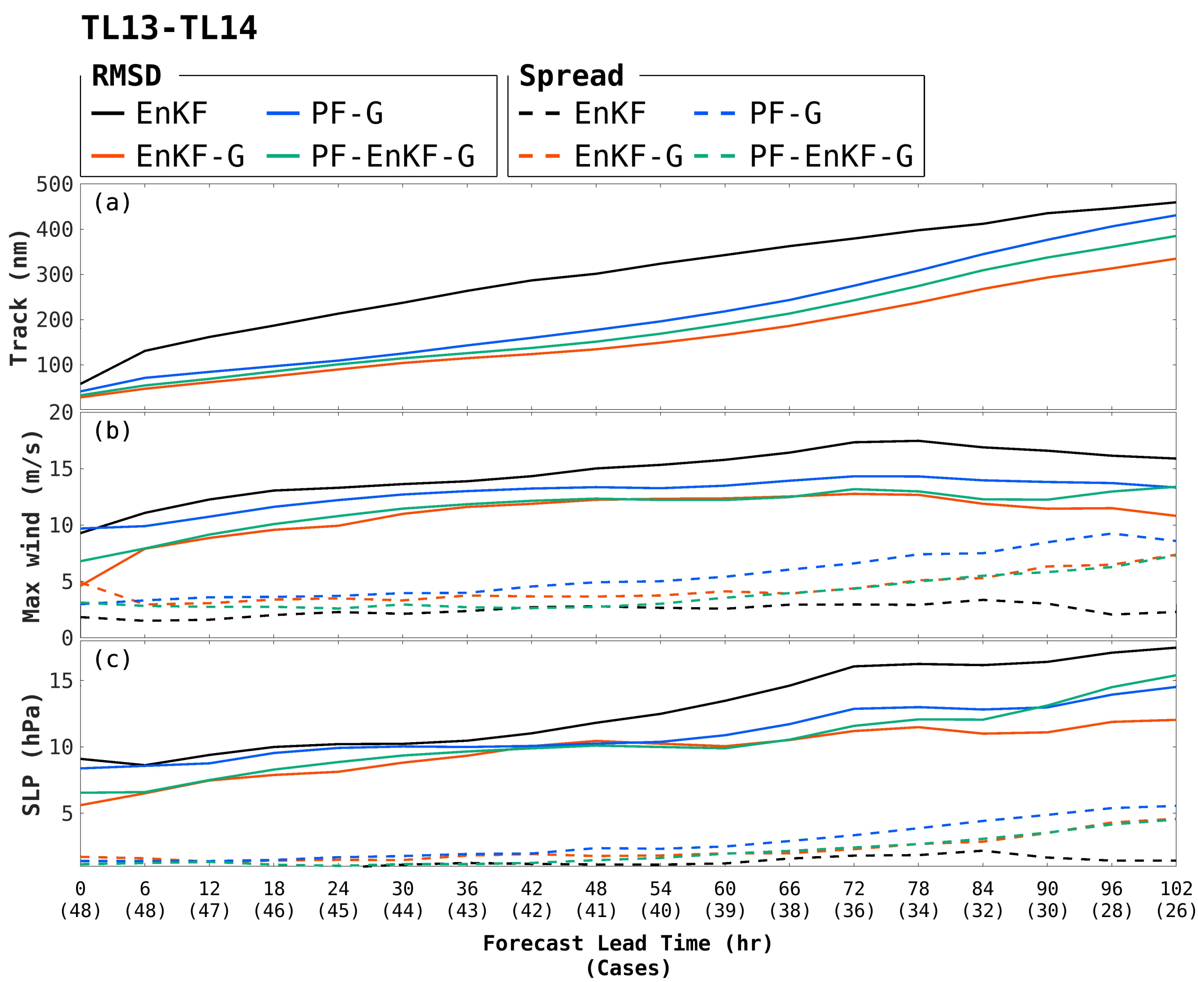


FIG. 8. Average RMSDs (solid) and ensemble spread (dashed) of (a) TC track, (b) max wind speed, and (c) sea level pressure for the periods during which AL13 (Laura) and AL14 (Marco) respectively occurred. These errors are verified against NHC best track and intensity data for EnKF (black), EnKF-G (red), PF-G (blue), and PF-EnKF filter (green). The horizontal axis shows the cumulative number of cases used for calculating the scores, along with the forecast lead time.

2020, which is verified to be identical for each experiment. For this comparison, deviations from the climatological estimate are assumed to stem from deficiencies in data assimilation, as discussed in Poterjoy (2022a).

Results indicate a consistent positive bias in KE for scales below ~400 km for the EnKF-G, demonstrating a significant deviation from model climatology; though not shown, we find a similar bias in the control EnKF experiment. This finding is consistent with verifications discussed

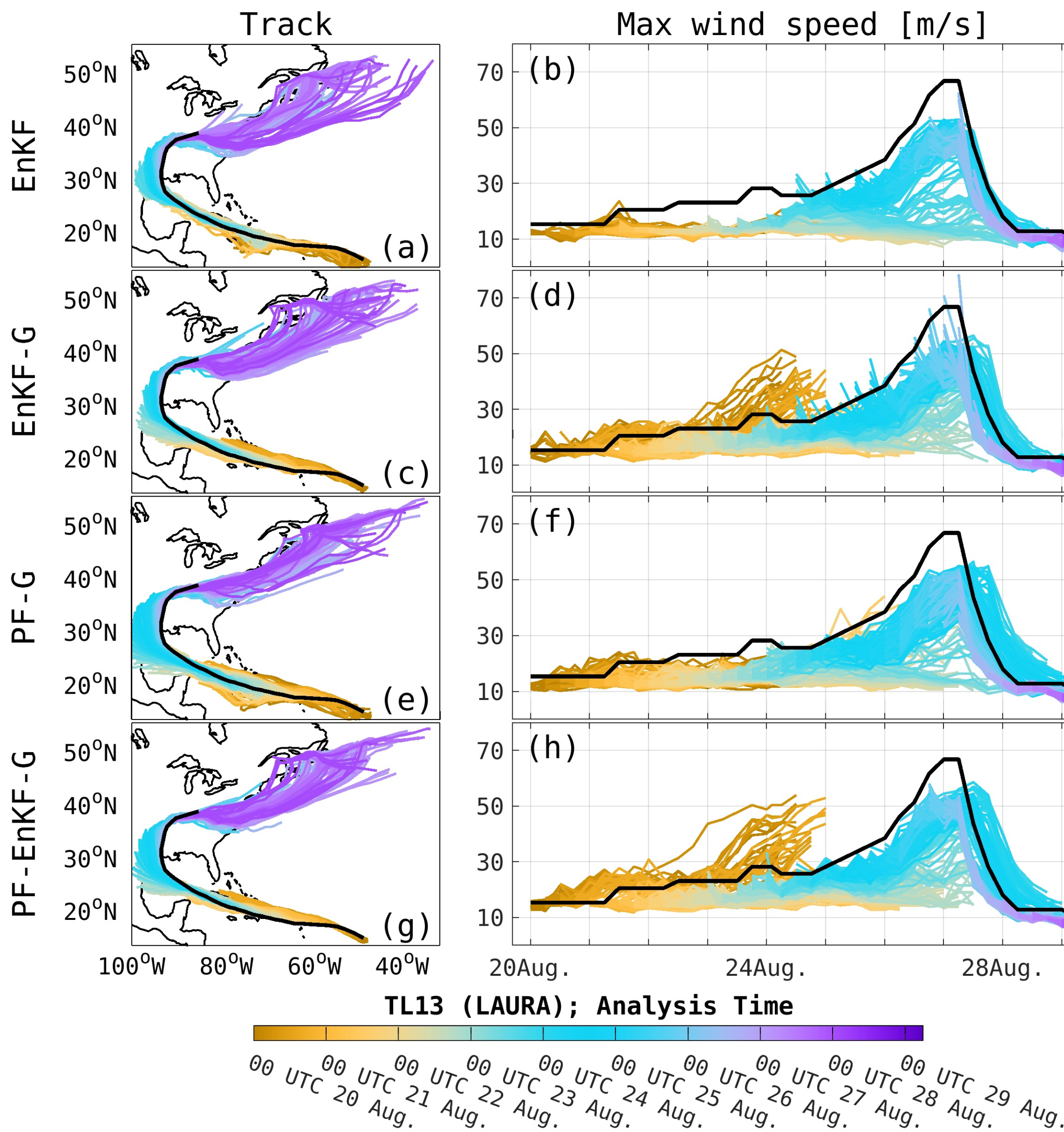


FIG. 9. Ensemble track (left) and intensity (left) forecasts from the (a) EnKF, (b) EnKF-G, (c) PF-G, and (d) PF-EnKF-G experiments between 0000 UTC 20 Aug. and 0600 UTC 29 Aug., focusing on AL13 (Laura). Forecasts are colored according to initialization time and NHC best-track data are plotted in black. The values are from 10 members of the 102-hour ensemble forecasts derived from the posterior state of each data assimilation method.

in previous sections, which identify a large spike in error for vorticity and max wind speeds with EnKF-G. Conversely, PF-G analyses and forecasts exhibit small but noteworthy, negative bias, which comes from populating the prior ensemble with a subset of coarse-resolution GDAS

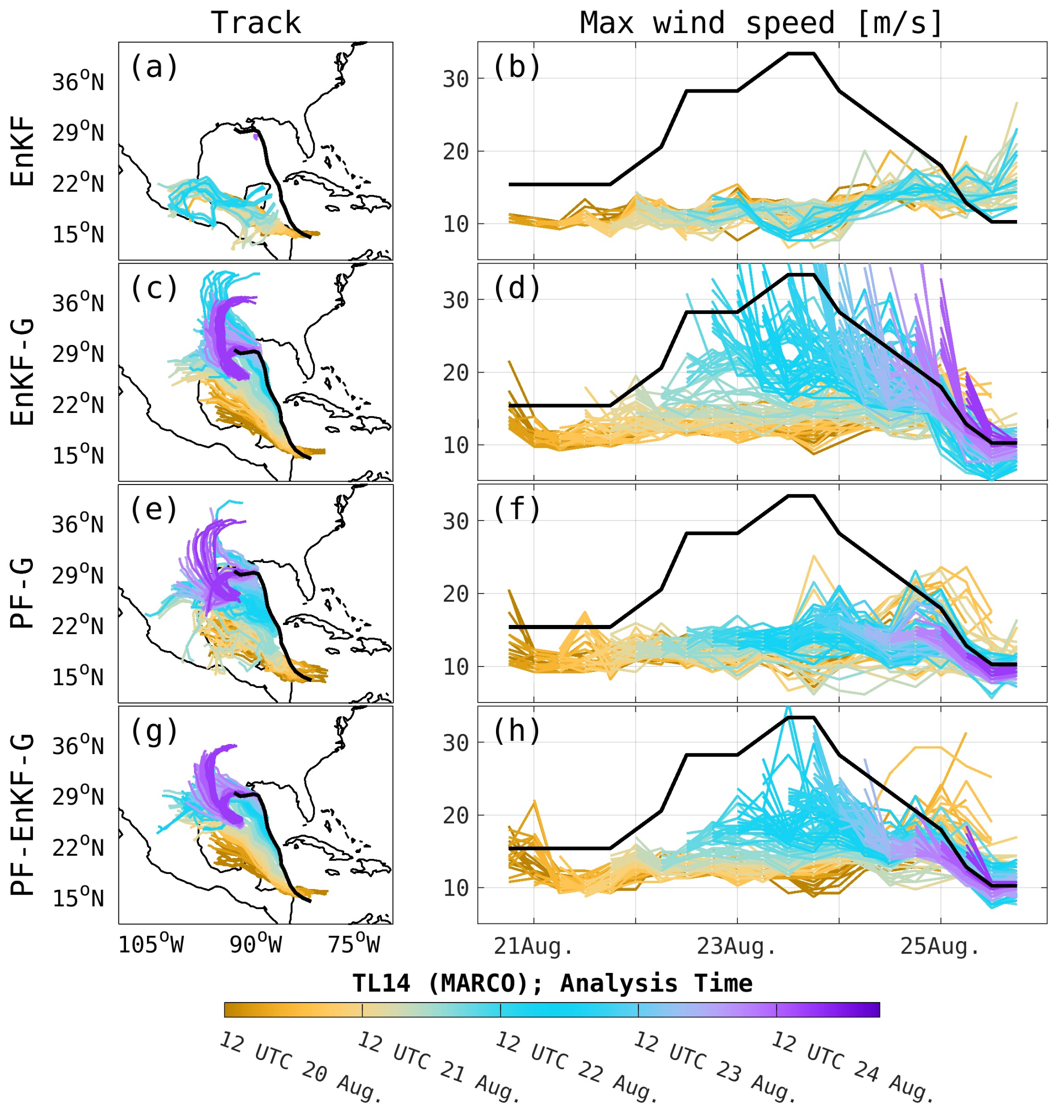


FIG. 10. As in Fig. 9, but for forecasts initialized between 1200 UTC 20 Aug. and 0600 UTC 25 Aug., focusing on AL14 (Marco).

perturbations—added to the HAFS mean. While EnKF-G ingests GDAS perturbations in the same manner as PF-G, the resulting KE bias is dominated by the assumptions used to adjust model states to reflect the analysis mean and error covariance determined from the Kalman filter update equations. Likewise, the PF-G KE bias is minimal compared to that of EnKF-G, suggesting that the LPF provides a closer alignment to plausible FV3 atmospheric states. Notably, FV3 requires ~12 h to dissipate the excess noise induced by the EnKF-G, which is visualized using lighter-shaded contours for longer lead times. The excess noise is typically removed in operational data assimilation systems by employing a digital filter (Lynch and Huang 1992) or through various other heuristic means. While the KE bias is often attributed to localization in ensemble filters, the positive bias is not found in PF-G which uses the same localization length scales as EnKF-G. As discussed in Poterjoy (2022a), a major source of KE bias stems from known limitations in data assimilation techniques that seek a minimum mean squared error estimate—as this estimate does not necessarily need to be a model solution when presented with non-Gaussian errors. We suspect that KE bias in EnKF-G members—and the cumulative impacts of this bias on 6-h forecasts—plays a role in the lack of spread found in this ensemble compared to PF-G. This aspect of the results will be the topic of a follow-up study.

Lastly, the PF-EnKF-G experiment contains substantially lower KE bias compared to EnKF-G. Consistent with past research using idealized models (Kurosawa and Poterjoy 2023) this method retains the positive benefits of Kalman filter-based data assimilation while mitigating some of the limitations associated with non-Gaussian priors.

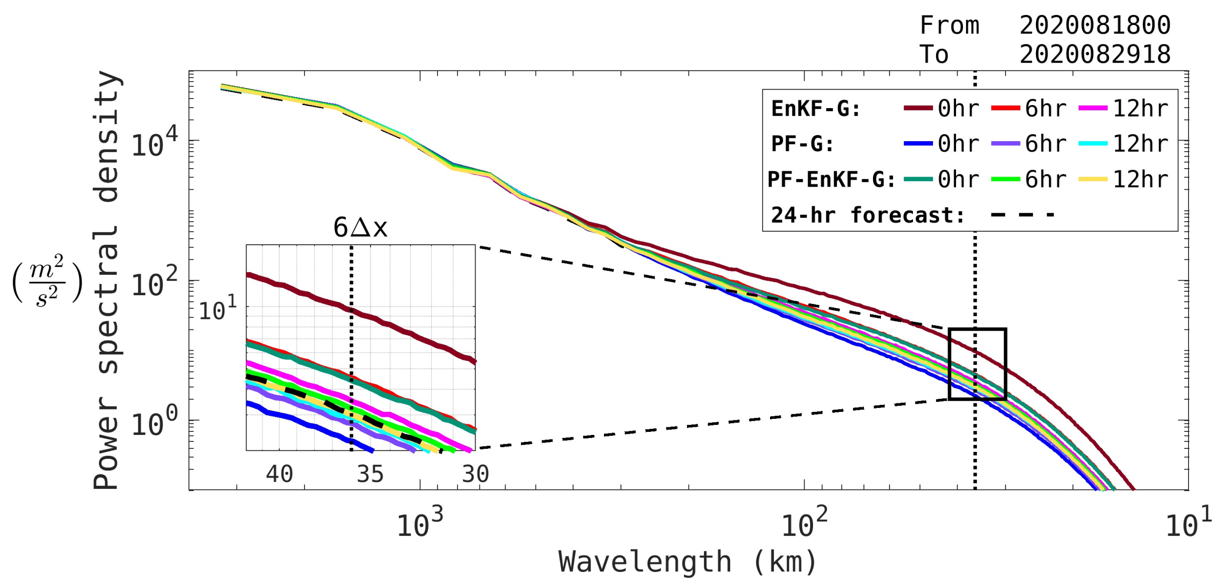


FIG. 11. 250-hPa kinetic energy spectrum averaged from August 18 to August 29, 2020 for single-member (red) EnKF-G, (blue) PF-G, and (green) PF-EnKF-G forecasts at 0-h, 6-h, and 12-h lead times. Darker colors represent earlier lead times, while lighter colors indicate later lead times. The dotted black line shows a climatological estimate for the FV3 model using 24-h forecasts. The vertical dashed line corresponds to the length scale that is 6 times the grid spacing of the model.

5. Conclusions

This study proposes an augmented ensemble data assimilation strategy that incorporates flow-dependent information generated from a global model into an ensemble data assimilation system for regional models. The primary goal is to reduce sampling deficiency when the high computational cost of generating high-resolution ensembles limits the number of members that can be used for data assimilation. We evaluate this approach through cycling experiments using the NOAA HAFS, focusing on the development and evolution of Hurricanes Laura and Marco in August 2020. The experiments demonstrate that the augmented ensemble data assimilation system successfully reduces sampling deficiencies in high-resolution 40-member HAFS analyses. Additionally, we implement PFs with augmented perturbations, which are not originally part of the HAFS system, and compare their performance with the EnKF and a hybrid filter that combines both methods.

We evaluate the performance of each data assimilation method in observation space and model space, using all available non-radiance measurements and ERA5 re-analyses, respectively. All data assimilation experiments that use augmented perturbations show reduced forecast errors, particularly from 600 hPa to 200 hPa, a region rich in upper-air measurements. This finding suggests a marked influence of the augmented approach for treating sampling deficiency in data-dense regions, which is an anticipated outcome given the theoretical limitations of ensemble data assimilation for well-observed high-dimensional dynamical systems (Hodyss and Morzfeld 2023). Additionally, when focusing on the atmospheric environment near TCs, the benchmark EnKF without augmented perturbations tends to underestimate storm intensity and completely miss the transition of one of our cases (Hurricane Marco) into a hurricane. We suspect that the small ensemble size, model resolution, and duration of the data assimilation experiments (3 weeks) were sufficient to cause the EnKF to experience filter divergence over portions of the model domain. The augmented approach, however, is found to mitigate this deficiency and produce much more skillful depictions of storm evolution. We further note that our choice of model grid spacing for these experiments (6 km) is less than the operational implementation of HAFS, which leads to additional intensity biases for storms. Likewise, subgrid-scale physical parameterization schemes and atmosphere-ocean coupling methodology in our version of HAFS predate operational versions of this model, which are additional sources of bias in our experiments.

We also investigate the implications of applying non-Gaussian data assimilation methods based on particle filters for HAFS by performing EnKF and PF experiments with augmented perturbations. We evaluate the prediction accuracy of each data assimilation technique over the entire model region using the ECMWF v5 reanalysis, and assess forecast accuracy for basic tropical cyclone metrics using NHC best track data. The EnKF performs best at the beginning of the forecast period—likely because the resulting analysis is derived to achieve a mean squared error estimate—but the advantages decrease as the forecast progresses to larger lead times. Nevertheless, the EnKF is found to induce a positive kinetic energy bias at shorter wavelengths in analysis members, owing to the use of error covariances alone when updating ensemble perturbations. The LPF, while showing a larger sensitivity to model bias, does not exhibit the same KE bias as the EnKF—a finding that is consistent with past studies. A mixed filter methodology that uses an intermediate LPF update before applying an EnKF update helps alleviate issues with the standalone (EnKF and LPF) data assimilation methods.

We emphasize that the operational data assimilation for HAFS is not ensemble based, and instead uses a variational scheme to update a single model state. Background error covariance for the variational analysis comes from GDAS instead of a self-cycled HAFS ensemble, which is a major difference between our methodology and the operational one. The current study deviates from the operational HAFS by using a serial ensemble square root filter and a local PF to perform the data assimilation. This means that observations are processed individually or in smaller subsets rather than simultaneously to update model states. This decision is mainly due to the design of the local particle filter and the blended PF-EnKF methods which require serial processing of observations to implement localization.

Moreover, the sensitivity of the proposed approach to the quality of GDAS forecast quality requires further investigation, including the development of flexible weighting schemes to prioritize higher-quality perturbations. By addressing these factors, the proposed approach could significantly enhance NOAA's regional forecasting capabilities while remaining practical for real-time applications.

We further note that our experiments do not apply specific approaches for reducing non-physical updates during data assimilation. This design choice differs from data assimilation systems that use digital filter initialization, normal mode initialization, incremental analysis update (IAU), or other means of reducing noise that may occur during data assimilation (Lynch and Huang 1992; Benjamin et al. 2004; Derber and Bouttier 1999; Bannister 2021). It is worth noting that these methods are often integrated into operational prediction systems to optimize their performances. While these techniques could potentially enhance the ability of the EnKF to produce large-scale quasi-balanced analyses in this study, the additional constraints may not be suitable for mesoscale weather systems, and would complicate the interpretation of results.

Lastly, the data assimilation experiments performed in this study adopt the same methodology for computing and removing bias for satellite radiance measurements that exists in operational implementations of HWRF and HAFS; i.e., coefficients for a state-dependent bias model are trained by GDAS and not updated within HAFS. This means of training a bias model is suboptimal, as shown in Knisely and Poterjoy (2023), and can have large implications for data assimilation methods that are more sensitive to bias, such as the LPF. The integration of new data assimilation techniques into a convective-permitting, basin-wide configuration of HAFS that performs its own online estimation of bias model coefficients will be the topic of a future study.

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

Data availability statement. All software, documentation, and methods used to support this study are available from the corresponding author at University of Maryland.

References

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129 (12)**, 2884 – 2903, https://doi.org/10.1175/1520-0493(2001)129⟨2884:AEAKFF⟩2.0.CO;2.

Bannister, R. N., 2017: A review of operational methods of variational and ensemble-variational data assimilation. *Q.J.R. Meteorol. Soc.*, **143 (703)**, 607–633, https://doi.org/10.1002/qj.2982.

Bannister, R. N., 2021: Balance conditions in variational data assimilation for a high-resolution forecast model. *Q.J.R. Meteorol. Soc.*, **147 (738)**, 2917–2934, https://doi.org/10.1002/qj.4106.

Benjamin, S. G., and Coauthors, 2004: An hourly assimilation–forecast cycle: The RUC. *Mon. Wea. Rev.*, **132 (2)**, 495 – 518, https://doi.org/10.1175/1520-0493(2004)132⟨0495:AHACTR⟩2.0.CO;2.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129 (3)**, 420 – 436, https://doi.org/10.1175/1520-0493(2001)129⟨0420:ASWTET⟩2.0.CO;2.

Bishop, C. H., and D. Hodyss, 2011: Adaptive ensemble covariance localization in ensemble 4D-VAR state estimation. *Mon. Wea. Rev.*, **139 (4)**, 1241 – 1255, https://doi.org/10.1175/2010MWR3403.1.

Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Q.J.R. Meteorol. Soc.*, **131 (607)**, 1013–1043, https://doi.org/10.1256/qj.04.15.

Chassignet, E. P., H. E. Hurlburt, O. M. Smedstad, G. R. Halliwell, P. J. Hogan, A. J. Wallcraft, R. Baraille, and R. Bleck, 2007: The HYCOM (HYbrid Coordinate Ocean Model) data assimilative system. *J. Mar. Syst.*, **65 (1)**, 60–83, https://doi.org/10.1016/j.jmarsys.2005.09.016.

Chassignet, E. P., and Coauthors, 2009: US GODAE: Global ocean prediction with the hybrid coordinate ocean model (hycom). *Oceanography*, **22 (2)**, 64–75, https://doi.org/10.5670/oceanog.2009.39.

Derber, J., and F. Bouttier, 1999: A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus A: Dynamic Meteorology and Oceanography*, **51 (2)**, 195–221, https://doi.org/10.3402/tellusa.v51i2.12316.

Dong, J., and Coauthors, 2020: The evaluation of real-time hurricane analysis and forecast system (HAFS) stand-alone regional (SAR) model performance for the 2019 atlantic hurricane season. *Atmosphere*, **11 (6)**, https://doi.org/10.3390/atmos11060617.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.: Oceans*, **99 (C5)**, 10 143–10 162, https://doi.org/10.1029/94JC00572.

Frei, M., and H. R. Künsch, 2013: Bridging the ensemble Kalman and particle filters. *Biometrika*, **100 (4)**, 781–800, https://doi.org/10.1093/biomet/ast020.

Gopalakrishnan, S., and Coauthors, 2021: 2020 HFIP R&D activities summary: Recent results and operational implementation. Tech. rep., NOAA/OAR and NWS. https://doi.org/10.25923/718e-6232.

Grooms, I., and G. Robinson, 2021: A hybrid particle-ensemble Kalman filter for problems with medium nonlinearity. *PLOS ONE*, **16 (3)**, 1–20, https://doi.org/10.1371/journal.pone.0248266.

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter–3D variational analysis scheme. *Mon. Wea. Rev.*, **128 (8)**, 2905 – 2919, https://doi.org/10.1175/1520-0493(2000)128⟨2905:AHEKFV⟩2.0.CO;2.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129 (11)**, 2776 – 2790, https://doi.org/10.1175/1520-0493(2001)129⟨2776:DDFOBE⟩2.0.CO;2.

Hazelton, A., and Coauthors, 2021: 2019 Atlantic hurricane forecasts from the global-nested hurricane analysis and forecast system: Composite statistics and key events. *Wea. Forecasting*, **36 (2)**, 519 – 538, https://doi.org/10.1175/WAF-D-20-0044.1.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Q.J.R. Meteorol. Soc.*, **146 (730)**, 1999–2049, https://doi.org/doi.org/10.1002/qj.3803.

Hodyss, D., and M. Morzfeld, 2023: How sampling errors in covariance estimates cause bias in the Kalman gain and impact ensemble data assimilation. *Mon. Wea. Rev.*, **151 (9)**, 2413 – 2426, https://doi.org/10.1175/MWR-D-23-0029.1.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126 (3)**, 796 – 811, https://doi.org/10.1175/1520-0493(1998)126⟨0796:DAUAEK⟩2.0.CO;2.

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

Houtekamer, P. L., and H. L. Mitchell, 2005: Ensemble Kalman filtering. *Q.J.R. Meteorol. Soc.*, **131 (613)**, 3269–3289, https://doi.org/10.1256/qj.05.135.

Hu, C.-C., and P. J. van Leeuwen, 2021: A particle flow filter for high-dimensional system applications. *Q.J.R. Meteorol. Soc.*, **147 (737)**, 2352–2374, https://doi.org/10.1002/qj.4028.

Kalman, R. E., 1960: A New Approach to Linear Filtering and Prediction Problems. *ASME. J. Basic Eng.*, **82 (1)**, 35–45, https://doi.org/10.1115/1.3662552.

Knisely, J., and J. Poterjoy, 2023: Implications of self-contained radiance bias correction for data assimilation within the hurricane analysis and forecasting system (HAFS). *Wea. Forecasting*, **38 (9)**, 1719 – 1738, https://doi.org/10.1175/WAF-D-23-0027.1.

Kotsuki, S., T. Miyoshi, K. Kondo, and R. Potthast, 2022: A local particle filter and its Gaussian mixture extension implemented with minor modifications to the LETKF. *Geosci. Model Dev.*, **15 (22)**, 8325–8348, https://doi.org/10.5194/gmd-15-8325-2022.

Kurosawa, K., and J. Poterjoy, 2021: Data assimilation challenges posed by nonlinear operators: A comparative study of ensemble and variational filters and smoothers. *Mon. Wea. Rev.*, **149 (7)**, 2369 – 2389, https://doi.org/10.1175/MWR-D-20-0368.1.

Kurosawa, K., and J. Poterjoy, 2023: A statistical hypothesis testing strategy for adaptively blending particle filters and ensemble Kalman filters for data assimilation. *Mon. Wea. Rev.*, **151 (1)**, 105 – 125, https://doi.org/10.1175/MWR-D-22-0108.1.

Lei, L., Z. Wang, and Z.-M. Tan, 2021: Integrated hybrid data assimilation for an ensemble Kalman filter. *Mon. Wea. Rev.*, **149 (12)**, 4091 – 4105, https://doi.org/10.1175/MWR-D-21-0002.1.

Liu, C., Q. Xiao, and B. Wang, 2008: An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Mon. Wea. Rev.*, **136 (9)**, 3363 – 3373, https://doi.org/10.1175/2008MWR2312.1.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP—a comparison with 4d-var. *Q.J.R. Meteorol. Soc.*, **129 (595)**, 3183–3203, https://doi.org/10.1256/qj.02.132.

Lynch, P., and X.-Y. Huang, 1992: Initialization of the HIRLAM model using a digital filter. *Mon. Wea. Rev.*, **120 (6)**, 1019 – 1034, https://doi.org/10.1175/1520-0493(1992)120⟨1019:IOTHMU⟩2.0.CO;2.

McCurry, J., J. Poterjoy, K. Knopfmeier, and L. Wicker, 2023: An evaluation of non-Gaussian data assimilation methods in moist convective regimes. *Mon. Wea. Rev.*, **151 (7)**, 1609 – 1629, https://doi.org/10.1175/MWR-D-22-0260.1.

McElhoe, B. A., 1966: An assessment of the navigation and course corrections for a manned flyby of Mars or Venus. *IEEE Transactions on Aerospace and Electronic Systems*, **AES-2 (4)**, 613–623, https://doi.org/10.1109/TAES.1966.4501892.

Penny, S. G., and T. Miyoshi, 2016: A local particle filter for high-dimensional geophysical systems. *Nonlin. Processes Geophys.*, **23 (6)**, 391–405, https://doi.org/10.5194/npg-23-391-2016.

Poterjoy, J., 2016: A localized particle filter for high-dimensional nonlinear systems. *Mon. Wea. Rev.*, **144 (1)**, 59 – 76, https://doi.org/10.1175/MWR-D-15-0163.1.

Poterjoy, J., 2022a: Implications of multivariate non-Gaussian data assimilation for multi-scale weather prediction. *Mon. Wea. Rev.*, **150 (6)**, 1475 – 1493, https://doi.org/10.1175/MWR-D-21-0228.1.

Poterjoy, J., 2022b: Regularization and tempering for a moment-matching localized particle filter. *Q.J.R. Meteorol. Soc.*, **148 (747)**, 2631–2651, https://doi.org/doi.org/10.1002/qj.4328.

Poterjoy, J., G. J. Alaka, and H. R. Winterbottom, 2021: The irreplaceable utility of sequential data assimilation for numerical weather prediction system development: Lessons learned from an experimental HWRF system. *Wea. Forecasting*, **36 (2)**, 661 – 677, https://doi.org/10.1175/WAF-D-20-0204.1.

Poterjoy, J., L. Wicker, and M. Buehner, 2019: Progress toward the application of a localized particle filter for numerical weather prediction. *Mon. Wea. Rev.*, **147 (4)**, 1107 – 1126, https://doi.org/10.1175/MWR-D-17-0344.1.

Potthast, R., A. Walter, and A. Rhodin, 2019: A localized adaptive particle filter within an operational NWP framework. *Mon. Wea. Rev.*, **147 (1)**, 345 – 362, https://doi.org/10.1175/MWR-D-18-0028.1.

Robert, S., D. Leuenberger, and H. R. Künsch, 2018: A local ensemble transform Kalman particle filter for convective-scale data assimilation. *Q.J.R. Meteorol. Soc.*, **144 (713)**, 1279–1296, https://doi.org/10.1002/qj.3116.

Rojahn, A., N. Schenk, P. J. van Leeuwen, and R. Potthast, 2023: Particle filtering and gaussian mixtures – on a localized mixture coefficients particle filter (LMCPF) for global NWP. *J. Meteor. Soc. Japan. Ser. II*, **101 (4)**, 233–253, https://doi.org/10.2151/jmsj.2023-015.

Schwartz, C. S., J. Poterjoy, G. S. Romine, D. C. Dowell, J. R. Carley, and J. Bresch, 2022: Short-term convection-allowing ensemble precipitation forecast sensitivity to resolution of initial condition perturbations and central initial states. *Wea. Forecasting*, **37 (7)**, 1259 – 1286, https://doi.org/10.1175/WAF-D-21-0165.1.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. *Mon. Wea. Rev.*, **136 (12)**, 4629 – 4640, https://doi.org/10.1175/2008MWR2529.1.

Tong, M., and Coauthors, 2018: Impact of assimilating aircraft reconnaissance observations on tropical cyclone initialization and prediction using operational HWRF and GSI ensemble–variational hybrid data assimilation. *Mon. Wea. Rev.*, **146 (12)**, 4155 – 4177, https://doi.org/10.1175/MWR-D-17-0380.1.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130 (7)**, 1913 – 1924, https://doi.org/10.1175/1520-0493(2002)130⟨1913:EDAWPO⟩2.0.CO;2.

Whitaker, J. S., and T. M. Hamill, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140 (9)**, 3078 – 3089, https://doi.org/10.1175/MWR-D-11-00276.1.