# The Irreplaceable Utility of Sequential Data Assimilation for Numerical Weather Prediction System Development: Lessons Learned from an Experimental HWRF System

JONATHAN POTERJOY,[a,b] GHASSAN J. ALAKA JR.,[b] AND HENRY R. WINTERBOTTOM[c,d]

[a] *University of Maryland, College Park, College Park, Maryland*
[b] *NOAA/Atlantic Oceanographic and Meteorological Laboratory, Miami, Florida*
[c] *Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado*
[d] *NOAA/Earth Systems Research Laboratories, Boulder, Colorado*

(Manuscript received 3 November 2020, in final form 13 February 2021)

ABSTRACT: Limited-area numerical weather prediction models currently run operationally in the United States and follow a ''partially cycled'' schedule, where sequential data assimilation is periodically interrupted by replacing model states with solutions interpolated from a global model. While this strategy helps overcome several practical challenges associated with real-time regional forecasting, it is no substitute for a robust sequential data assimilation approach for research-to-operations purposes. Partial cycling can mask systematic errors in weather models, data assimilation systems, and data preprocessing techniques, since it introduces information from a different prediction system. It also adds extra heuristics to the model initialization steps outside the general Bayesian filtering framework from which data assimilation methods are derived. This study uses a research-oriented modeling system, which is self-contained in the operational Hurricane Weather Research and Forecasting (HWRF) Model package, to illustrate why next-generation modeling systems should prioritize sequential data assimilation at early stages of development. This framework permits the rigorous examination of all model system components—in a manner that has never been done for the HWRF Model. Examples presented in this manuscript show how sequential data assimilation capabilities can accelerate model advancements and increase academic involvement in operational forecasting systems at a time when the United States is developing a new hurricane forecasting system.

SIGNIFICANCE STATEMENT: This study discusses a road map for designing numerical weather predictions systems that are more accessible to the research community. It is based on the premise that the statistical framework used for identifying initial conditions for dynamical models, such as weather prediction models, should play a larger role in model development, observation collection, and uncertainty quantification than currently exists for regional models. While this study uses examples motivated by one current operational weather model, the conclusions have broad implications. Ultimately, the long-term goals set forth by leaders in the atmospheric science community demand a more holistic evaluation of modeling systems than currently exists. This study is timely, considering the advancement of major modeling operational modeling efforts currently under way in the United States.

KEYWORDS: Kalman filters; Ensembles; Numerical weather prediction/forecasting; Mesoscale models; Regional models

## 1. Introduction

The physical processes governing the life cycle of tropical cyclones span a large spectrum of spatial and temporal scales. Weather phenomena of this type motivate the design of numerical modeling frameworks that consider multiscale interactions in initial conditions and forecasts. Ideally, such frameworks should also adhere to the probabilistic nature of the analysis and prediction problem posed by applications of this type. For example, changes in vortex structure, maximum wind speeds, and storm motion often depend on turbulent motions not represented well by observations and models, and likely exhibit intrinsic predictability limits (e.g., Lorenz 1969; Rotunno and Snyder 2008). Therefore, the most complete depiction of an observed tropical cyclone event is one that presents a number of plausible scenarios, which match prior physical knowledge contained in numerical models with measured evidence of the true atmosphere.

The above paradigm fits into two major National Oceanographic and Meteorological Agency (NOAA) initiatives, namely the Unified Forecasting System (UFS) and the Hurricane Forecast Improvement Project (HFIP). The UFS is a community framework for operational weather prediction and research that respects the multiscale nature of high-impact weather events, like tropical cyclones (Toepfer et al. 2018). Through this project, NOAA aims to advance a flagship modeling system based on the Finite Volume Cubed-Sphere (FV3) dynamical core (see NOAA/GFDL 2018). Likewise, HFIP sets ambitious goals for improving tropical cyclone forecasts in the United States, while emphasizing the importance of ensembles to communicate forecast uncertainty (Gall et al. 2013). As a part of HFIP, NOAA plans to advance the operational Hurricane Analysis and Forecast System (HAFS) to replace current forecast facilities as the hurricane application of the Unified Forecast System. The main objectives for HAFS include: 1) advancement of deterministic and ensemble prediction capabilities to seven days; 2) the fusion of modeling, data assimilation, and observations to produce a record of storm

*Corresponding author*: Dr. Jonathan Poterjoy, poterjoy@umd.edu

DOI: 10.1175/WAF-D-20-0204.1

analyses; and 3) improvements in statistical postprocessing methods for extracting information regarding forecast uncertainty (Marks et al. 2019). In the interim, opportunities exist for leveraging preexisting community software to explore science questions related to both UFS and HFIP/HAFS goals as unified modeling and data assimilation systems continue to mature.

Current global modeling efforts at NOAA already provide a starting point for addressing some of the challenges listed above for operational forecasting. Namely, the Global Ensemble Forecasting System (GEFS) seeks to provide a probabilistic depiction of the atmosphere conditioned on recent observations. A major deficiency in the GEFS, however, comes from the lack of a storm-resolving effective grid spacing, which limits its ability to capture important dynamics needed to model cloud-scale processes. Therefore, forecasters must rely on a suite of high-resolution limited-area models to supplement GEFS predictions with missing scales of motion. For the case of tropical cyclones, NOAA operational models include the Hurricane Weather Research and Forecasting (HWRF) Model and the Hurricanes in a Multiscale Ocean-coupled Nonhydrostatic (HMON) model. While a number of regional models in the United States contain their own data assimilation systems, they all adopt partial-cycling schedule, which increases their reliance on the Global Forecasting System (GFS) beyond the basic purpose of providing boundary conditions. Here, partial-cycling refers to the periodic recentering of the model on the GFS analysis. Because of this limitation, models such as HWRF—which is used in the current study—have never been run outside a partially cycled framework for extensive periods.

As the weather community begins to construct future operational weather predictions systems such as HAFS, it is important to acknowledge limitations in our past modeling systems and explore new strategies posed by research-oriented frameworks. The most straightforward way of demonstrating where scientific improvements—and in some cases, paradigm shifts—are needed is by concrete examples and the development of prototypes. As such, this manuscript explores findings from an experimental HWRF modeling system that is built around the philosophy outlined above. The purpose of this study is to use elementary data assimilation strategies to 1) uncover shortcomings of one partially cycled NOAA modeling system; 2) demonstrate outstanding challenges associated with NOAA's use of satellite measurements in hurricane models; and 3) highlight the utility of research-oriented modeling frameworks for data assimilation algorithm testing. These efforts exploit a dedicated probabilistic model framework, which will be denoted the "AOML-UMD ensemble system" in the current manuscript. The AOML-UMD ensemble system is a data assimilation and forecasting system built around the HWRF Model. Unlike past uses of the HWRF Model, including the HWRF prediction system run operationally by the Environmental Modeling Center (EMC) of the National Centers for Environmental Prediction (NCEP), the AOML-UMD ensemble system performs data assimilation continuously through large parts of hurricane seasons. This strategy is feasible because the experimental system is not limited by typical time constraints of operational modeling products.

By design, the AOML-UMD ensemble system is well-suited for performing holistic evaluations of numerical weather prediction systems. In particular, the authors recognize that probabilistic forecasts for tropical cyclones remain limited by the suboptimal quantification of initial condition uncertainty and the misrepresentation of physical processes in numerical weather prediction models, especially at subgrid scales. While these error sources are often difficult to distinguish from one another, sequential data assimilation techniques provide a means to rigorously test components of current forecast systems and validate changes proposed to model physics, observing systems, or data assimilation methods. As a starting point, one can examine observation-space diagnostics collected over long periods of time and space, which is done in the current study. This practice is common at many environmental modeling centers, where statistics acquired from large samples of model-observation comparisons help identify biases and justify upgrades. Short-term forecasts produced between data assimilation cycles can then be used to explore mechanisms leading to errors in weather and climate predictions (e.g., Rodwell and Jung 2008; Martin et al. 2010; Cavallo et al. 2016). Research of this type leverages the vast amount of "model state corrections" or analysis increments produced when performing data assimilation over long periods of time to determine why model trajectories systematically deviate from observations. This strategy is conceptually simple, but requires the dedicated support for a sequential data assimilation system—even if not used for real-time forecasting.

The framework discussed above contains a number of additional properties that distinguish it from regional modeling systems typically used for tropical cyclone research. A major difference is its use of an extensive regional domain, which covers an area large enough to permit the time-dependent calculation of bias correction coefficients used to assimilate radiances. This feature allows the AOML-UMD ensemble system to act as a standalone modeling framework, requiring a global model only for boundary conditions in the same manner as the HAFS system currently under development. The large domain size also enables interactions between tropical cyclones and the synoptic-scale environment represented across multiple hurricane basins, which motivated the development of "basin-scale" configurations of HWRF run at AOML (Zhang et al. 2016; Alaka et al. 2017). Users have full control over how multiscale weather features are represented in model initial conditions through choices of model domain and physics configurations, observations, observation preprocessing, and data assimilation. Consequently, the experimental prediction system contains a large parameter space for generating and calibrating ensemble forecasts. As will be described in the present manuscript, the above features make the AOML-UMD ensemble system a practical testbed for examining research questions posed by HFIP, while providing a cost-effective (regional) framework for exploring data assimilation and forecasting plans related to HAFS and future unified modeling efforts at NOAA.

The current manuscript introduces the AOML-UMD ensemble prediction system and summarizes findings that are relevant to the development of a next-generation hurricane modeling system. Given the methodology adopted for this work, most results will focus on observations-space verification

of analyses and short-term forecasts generated during the seasonal data assimilation experiments carried out over the 2017 and 2018 Atlantic and eastern North Pacific hurricane seasons. Section 2 presents the data assimilation and modeling framework. Section 3 describes the AOML-UMD ensemble system and its configuration for HWRF. Section 4 discusses results from experiments that explore multimonth biases in HWRF and satellite data assimilation practices used by the operational system. This section also explores strategies for testing new data assimilation algorithms, including a localized particle filter. The last section summarizes findings and discusses their relevance to the broader operational weather forecasting community.

## 2. Preexisting community infrastructure

The AOML-UMD ensemble system relies entirely on community software maintained by the Developmental Testbed Center (DTC), a distributed facility for research and operational model development managed by NOAA, the National Center for Atmospheric Research (NCAR), and the U.S. Air Force. This software includes multiple components of the operational HWRF forecasting system, such as the scripting infrastructure and workflow manager needed for its implementation, and various parts of the Gridpoint Statistical Interpolation (GSI) data assimilation package. In this section, we describe the operational HWRF forecasting system and discuss ongoing modeling efforts relevant to the AOML-UMD ensemble system.

### a. *The operational HWRF system*

HWRF is a regional model operated by EMC to provide real-time numerical guidance for tropical cyclone forecasting operations in the United States and its territories. It uses the Nonhydrostatic Mesoscale Model (NMM) dynamical core of the Weather Research and Forecasting (WRF) model and physics options tuned specifically for tropical cyclones and their environments (Gopalakrishnan et al. 2010; Atlas et al. 2015). All experiments described in this study use the 2017 version of HWRF. We refer to Biswas et al. (2017) for technical details of this model version, including a list of physics options.

EMC maintains the operational HWRF Model within a comprehensive data assimilation and forecasting platform, which includes annual changes to model physics, domain configuration, data assimilation, and observation processing. The most recent configurations all use a fixed outer domain with two storm-following inner nests centered on a designated tropical cyclone location. The model is initialized every 6 h following a rather complex workflow: 1) all domains are recentered on the most recent tropical cyclone location, determined by the National Hurricane Center (NHC), Central Pacific Hurricane Center (CPHC), or Joint Typhoon Warning Center (JTWC); 2) the parent domain inherits initial conditions interpolated from the GFS analysis; and 3) the inner nests are blended with the interpolated GFS solution near the periphery, but retain a solution coming from a cycled hybrid GSI-based ensemble 3D-variational update (3D-Var) in the vicinity of the storm (Zhang and Pu 2020). The hybrid data assimilation preceding the third step occurs after the tropical cyclone vortex is removed and relocated in prior model states. The relocated vortex is either modified from a previous forecast or replaced completely, depending on the storm intensity and whether a previous HWRF forecast is available at the time. We encourage readers to review Biswas et al. (2017) for a more complete description of procedures followed to run the HWRF Model.

In addition to the deterministic HWRF system, EMC operated a real-time HWRF ensemble to provide probabilistic guidance for tropical cyclones up until 2019 (Zhang et al. 2018). Like the deterministic implementation, the ensemble relied on the GFS for environmental initial conditions to avoid the computational cost of running an ensemble data assimilation system for the model's parent domain.

### b. *HWRF systems designed for data assimilation and ensemble forecasting research*

Given the narrow focus of the operational HWRF system described above, a number of data assimilation and prediction systems have emerged to fill the needs of the research community. This effort started at AOML with the Aksoy et al. (2012) HWRF Ensemble Data Assimilation System (HEDAS), which is based on the same sequential EnKF-based methodology used by the AOML-UMD ensemble system. Early applications of HEDAS focused on assimilating aircraft measurements from hurricane reconnaissance flights (Aksoy et al. 2013), but it is now used for a variety of data assimilation and observing system experiments at AOML (e.g., Steward et al. 2017; Christophersen et al. 2017). While HEDAS is a practical resource for examining inner-core tropical cyclone processes with HWRF, it currently has limited technical support outside AOML. For this reason, applications of HEDAS are restricted to modeling efforts smaller than the ones discussed in this manuscript.

Following HEDAS, Lu et al. (2017) introduced a research-oriented data assimilation and prediction system for HWRF that mirrors the developmental needs of the operational system. This framework uses hybrid three- and four-dimensional data assimilation methods to form high-resolution deterministic estimates of the model state over the life cycle of storms. The Lu et al. (2017) HWRF system has several added benefits over the operational version, such as four-dimensional data assimilation capabilities and some flexibility in choosing update frequency and domain configurations. Like the operational HWRF implementations, data assimilation procedures operate primarily on model variables in the vicinity of tropical cyclones, requiring the GFS model to provide environmental conditions.

### c. *The experimental "basin-scale" HWRF system*

When multiple storms exist simultaneously in the Atlantic hurricane basin, EMC operates multiple runs of the HWRF Model, providing independent numerical guidance for each storm. Motivated by the needs of future global weather prediction systems, scientists at AOML and EMC collaborated to form a parallel research version of HWRF called the "basin-scale" HWRF (Zhang et al. 2016), denoted hereafter as HWRF-b. The HWRF-b model uses a much larger fixed outer domain than the operational system, which spans multiple hurricane basins. It also runs multiple sets of movable nests,

which overlap when neighboring tropical cyclones are sufficiently close together. In addition to using the HWRF-b for research, AOML scientists run the model in a near-real-time forecast setting during hurricane seasons, an exercise that started in 2013.[1] The experimental model has demonstrated benefits in the form of superior tropical cyclone track forecasts over the operational system (Alaka et al. 2017, 2020), which is an anticipated outcome given the large domain and refined nesting strategy. The experimental model uses the same vortex relocation strategies as the operational model, and also relies mostly on the GFS analysis to provide its initial conditions; see Alaka et al. (2020) for details. Therefore, including data assimilation capabilities to the HWRF-b parent domain provides a natural progression from current modeling efforts at AOML. This feature brings the experimental model a step closer to achieving a framework that mimics how future global models with multiple nested domains will operate with sequential data assimilation.

## 3. The AOML-UMD ensemble forecasting system

### a. Motivation

The operational HWRF modeling system provides valuable numerical guidance for weather forecasters, but its reliance on GFS environmental conditions, limited use of sequential data assimilation, and heuristic vortex initialization procedures provide obstacles for scientists who wish to use the system for research. For example, empirical corrections made to prior tropical cyclone vortices before assimilating measurements can mask technical or fundamental issues with GSI data assimilation schemes or reduce the impact of potentially valuable new observing systems. Likewise, annual upgrades to the GFS model will ultimately impact the skill of HWRF in unknown ways. These factors could slow long-term progress toward improving forecast skill. In addition, the operational system does not permit the accumulation of observation-space model forecast diagnostics over long periods of time and space, owing to its intermittent use during hurricane seasons. This limitation reduces the amount of information available for examining biases in the model, which are needed to identify deficiencies in data assimilation or physical parameterization schemes. For similar reasons, the limited use of sequential data assimilation may also have large implications for how bias-correction procedures are implemented for satellite radiance measurements during their assimilation. Last, the HWRF initialization process, which is also adopted by the current HWRF-b system, decouples tropical cyclones from their environments. This feature, combined with the lack of data assimilation on the HWRF parent domain, makes it difficult to identify realistic modes of error growth in ensemble forecasts.

Past research-based HWRF configurations described in section 2b provide additional data assimilation and forecasting flexibility over the operational system, but contain the same limitations for probabilistic synoptic-scale weather prediction and bias estimation. Therefore, the construction of ensemble forecasts for research and operational prediction requires a standalone regional data assimilation system as described in this manuscript.

### b. A regional ensemble framework for HWRF

In this section, we introduce the AOML-UMD ensemble system, which includes sequential data assimilation and medium-range probabilistic forecasting components. The regional forecast system leverages the same Python-based scripting package as the operational HWRF Model, which means it uses a similar organization of directories and follows the same flow of observations, model states, and metadata. By construction, any advancement to software infrastructure, such as the addition of new observation types, can transition seamlessly into the operational HWRF system for testing—and vice versa. In doing so, this system carries forward with a strategy for research-to-operations activities that started with previous deterministic versions of the HWRF-b model (Zhang et al. 2016; Alaka et al. 2020).

Like the deterministic HWRF-b modeling system, the AOML-UMD ensemble uses an extensive static parent domain, which spans large portions of the Atlantic and eastern North Pacific hurricane basins; see Fig. 1 for 2017 domain configuration. This system uses the GSI ensemble Kalman filter (EnKF) to update a 60-member ensemble of HWRF Model states every 6 h to reflect the most recent conventional and satellite observations. The suite of measurements accessed by the data assimilation system is identical to the operational HWRF Model, thus allowing for real-time and retrospective forecasting experiments using both routinely collected measurements as well as those taken from aircraft reconnaissance missions. Following data assimilation, the system subsamples a specified number of ensemble members for producing probabilistic forecasts. For example, the 2018 version generates 20-member 5-day forecasts each day at 0000 and 1200 UTC using a single static domain for each member.

While the AOML-UMD ensemble prediction system represents a step toward a more autonomous modeling platform for HWRF, its basic structure resembles past efforts using other regional models. For example, the NCAR Advanced Research version of WRF (ARW) ensemble system (Schwartz et al. 2015) and the German Weather Service kilometer-scale ensemble data assimilation for the Consortium for Small-Scale Modeling (COSMO) or "KENDA" forecasting system (Schraff et al. 2016) both perform uninterrupted sequential data assimilation for extensive periods. Adopting this strategy greatly reduces the influence of global models on regional analysis and forecast verifications, thus allowing for a more transparent evaluation of model physics, data assimilation algorithms and parameters, observation quality control mechanisms, and observation impact, than all past implementations of the HWRF Model.

### c. Data assimilation configuration

The AOML-UMD ensemble system relies solely on the GSI EnKF data assimilation system for providing 6-h updates to HWRF Model states over time. The GSI EnKF is based on the

[1] An HFIP computing allocation on NOAA's research supercomputer, Jet, provides the computing resources for performing these experiments.

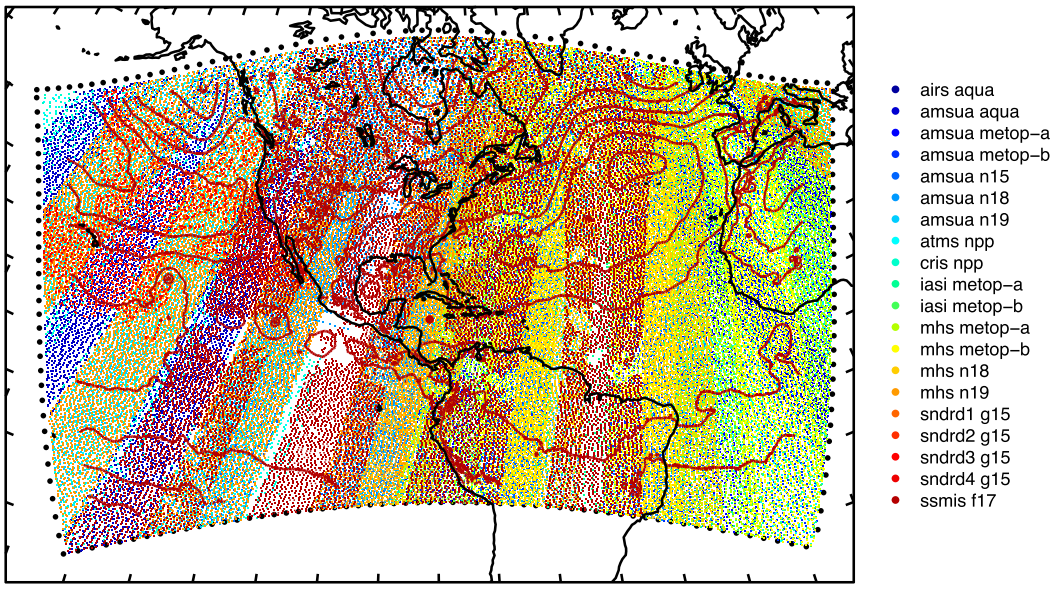


FIG. 1. The HWRF domain used for the 2017 HFIP demo is outlined on a map with land boundaries. Also plotted are contours of MSLP every 5 hPa for a single posterior member and locations of satellite radiance measurements at 1200 UTC 20 Jul 2017. The satellite data types are color coded according to the legend on the right.

Whitaker and Hamill (2002) ensemble square root filter and is the same ensemble data assimilation method used by multiple operational and experimental models in the United States, such as the GFS and HWRF Models, and the High-Resolution Rapid Refresh ensemble (HRRRE; Dowell 2020). In practice, ensemble data assimilation systems require modifications of sample-estimated error covariances to ensure they provide accurate and stable results for high-dimensional applications like numerical weather prediction. These mechanisms include covariance localization and variance inflation. For covariance localization, the GSI EnKF uses a continuous, smooth correlation function with compact support, given by Eq. (4.10) of Gaspari and Cohn (1999) to damp ensemble-estimated correlations to zero beyond a specified cutoff value as described in Hamill et al. (2001). To maintain variance in the ensemble following each data assimilation step, ensemble perturbations are inflated using the relaxation to prior spread (RTPS) technique described in Whitaker and Hamill (2012).

Both localization and RTPS methods require tunable parameters, which we examine through retrospective tests during the 2017 hurricane season (not shown). Using prior root-mean-square (RMS) fit to observations and prior observation-space ensemble spread, we arrived at the following EnKF configuration: 1200-km horizontal localization cutoff for Northern and Southern Hemisphere observations reduced to 900 km in the tropics; 1 scale-height vertical localization cutoff for conventional measurements; 2 scale-height vertical localization cutoff for satellite radiance measurements; 95% relaxation back to prior spread in the Northern and Southern Hemisphere reduced to 90% in the tropics.

Being a regional modeling system, the AOML-UMD ensemble requires boundary conditions from a global forecast model to operate, which we generate using posterior members and 6-h forecasts from the global data assimilation system (GDAS) ensemble. We interpolate each posterior HWRF member to posterior GFS ensemble members within a 360-km border on the domain boundary. The interpolation step helps reduce the mismatch between boundary conditions introduced from the GFS and posterior members produced during data assimilation. We also replace soil temperature and moisture and sea surface temperatures in posterior HWRF members with values from the GFS analysis, thus taking surface boundary conditions from the global model as well.

The data assimilation used for the AOML-UMD system performs its own bias correction for clear-air satellite radiance measurements, which is done through state augmentation during EnKF data assimilation (Miyoshi et al. 2010). Zhu et al. (2019) found this approach to perform comparably to the variational bias correction method used by the operational HWRF system. Assimilating radiance measurements requires the removal of biases coming from instruments, measurement operators, and the forecast model before they can be properly assimilated (e.g., Derber and Wu 1998). Because of the extensive domain size and continuous cycling of model states through time, estimating these biases adds value to the resulting prediction system; see section 4 for details. This feature provides another large distinction between the current HWRF data assimilation system and past strategies.

## 4. Seasonal HWRF experiments

During the 2017 and 2018 hurricane seasons, we used the NOAA Jet high-performance computing system to run the AOML-UMD ensemble as an HFIP demo project. The 2017 allocation provided resources for performing the first major testing of the experimental modeling system in near–real time,

following an initial trial period where the EnKF data assimilation system was tuned for the application. These tests examine whether the HWRF Model can run for long periods using uninterrupted sequential data assimilation to maintain consistency between the model and real atmosphere. They also provide a basis for retrospective modeling experiments to be performed after hurricane seasons, as demonstrated in sections 4b–c.

While near-real-time tests were performed in two seasons, results from 2018 were heavily biased, owing to a major issue uncovered in the 2018 operational HWRF modeling system during this study. This season was the first in which *GOES-16* provided operational geostationary satellite coverage over the Atlantic region, formally replacing *GOES-13* as GOES-East. Due to an error in naming convention for satellite measurements, atmospheric motion vectors (AMVs) estimated from GOES-East failed to be assimilated by our modeling system. Given that AMVs provide the only extensive wind information over open oceans, the sequential data assimilation strategy showed obvious differences from the GFS model, which ultimately led to this finding. This problem went undetected by the operational community because partial-cycling and vortex relocation strategies helped mask major deficiencies in data availability and assimilation strategy. The error ultimately nullifies any results obtained during the 2018 season, but provides another example for why developing sequential data assimilation capabilities are crucial for rigorously testing modeling systems outside of operations.

Despite the above limitations, the 2017 season provided ample data for evaluating various components of the HWRF modeling system and carrying forward with data assimilation experiments that require a standalone modeling system. This section discusses major findings from these numerical experiments.

### a. *Three-month sequential data assimilation experiments with HWRF*

The AOML-UMD ensemble system ran nonstop from 20 July 2017 to 31 October 2017 during the 2017 HFIP real-time demo, amounting to 415 sets of EnKF prior and posterior members. To the best of our knowledge, this experiment provides the longest period of uninterrupted data assimilation for the HWRF Model to date, thus providing a unique set of error statistics for evaluating its performance. These tests are carried out using a model grid spacing comparable to parent domains used by HWRF in past seasons, which parameterized convection. Therefore, error metrics used for this study are dominated mostly by errors in TC environments instead of errors within and near tropical cyclones.

For the 2017 experiment, we use the same model physics as the 2017 operational HWRF Model but with a modified domain configuration. The HWRF Model runs on a $598 \times 632 \times 61$ static grid with 18-km horizontal grid spacing and a model top at 2 hPa. The model levels follow the same vertical $\eta$ coordinates as the 2016 version of the operational HWRF Model, which also used a 2-hPa top; note the operational HWRF transitioned to a 5-hPa top for the 2017 version. Similar to past studies using sequential data assimilation to identify potential deficiencies in model physics, observations, or data assimilation systems, we use observation-space diagnostics for prior ensembles (6-h forecasts) to summarize errors in the experimental modeling system. We calculate root-mean-square differences (RMSDs) between prior ensemble means and conventional observations, alongside the "expected RMSDs." These quantities reflect domain-average errors for sets of observations assimilated during the same cycle, so that mean RMSD at a given time is given by $\sqrt{(1/N_y)\sum_{i=1}^{N_y}(y_i - \overline{x}_i)^2}$, where $y_i$ is the $i$th verifying observation and $\overline{x}_i$ is the prior ensemble mean projected onto $y_i$. Likewise, expected RMSDs are estimated as a function of error variance for each observation, $\sigma^2_{y_i}$, and the variance of prior ensemble members projected into observation space, $\sigma^2_{x_i}$; i.e., expected RMSD $= \sqrt{(1/N_y)\sum_{i=1}^{N_y}(\sigma^2_{y_i} + \sigma^2_{x_i})}$. Provided that the assigned observation error variance and the ensemble-estimated prior variance are accurate characterizations of the true errors, this metric should match the RMSDs when averaged over large numbers of independent samples. In this section, we discuss average RMSDs, expected RMSDs, mean error (bias), and observation counts accumulated over the test period. These "summary statistics" provide a quantitative measure of how effective the data assimilation system is at maintaining consistency between the HWRF Model and the observed atmosphere.

The first part of this study examines issues that may cause large-domain configurations of HWRF to "drift" from observations with time, despite the use of a continuously cycled data assimilation system. Figure 2 shows time series of summary statistics, calculated for prior zonal wind $u$ and meridional wind $v$. These values are calculated from atmospheric soundings, aircraft measurements, AMVs, as well as surface observing systems onboard ships, buoys, etc. In addition to the forecast metrics shown in the top panels of Fig. 2, the number of available observations at each cycle is plotted alongside the number of observations that pass thinning and quality control (QC) procedures. The observation counts indicate how many observations go into verification and help track the acceptance rate of measurements during data assimilation experiments. We also calculate time-average errors over the last 3 months of sequential data assimilation and bin the metrics according to pressure, i.e., in 50-hPa bins from the surface to 100 hPa (Fig. 3). In doing so, we omit the first 12 days of the experiment to allow ensemble-estimated errors and satellite bias correction coefficients to "spin up" from the GFS solution. This spin-up time is decided by the period needed for channel 10 (57 GHz) of the Advanced Microwave Sounding Unit (AMSU-A) brightness temperatures to yield quasi-steady innovations (see Fig. 8). As discussed in a future section, this channel is sensitive to stratospheric temperatures; therefore, the low model-top of HWRF provides a known bias difference from GFS.

The summary statistics in Figs. 2 and 3 show near-steady RMSDs over the test period, along with low domain-average bias and a relatively constant stream of observations getting passed through observation preprocessing steps. The pair of figures also show a close match between RMSDs and expected RMSDs, which indicates the data assimilation system is calibrated to provide an appropriate estimate of 6-h wind errors on average. Figures 4 and 5 show the same summary statistics described above, but for temperature $T$ and water vapor mixing ratio $q_v$. Unlike the wind variables, $T$ and $q_v$ exhibit a notable amount of bias throughout the cycling period, which follow from

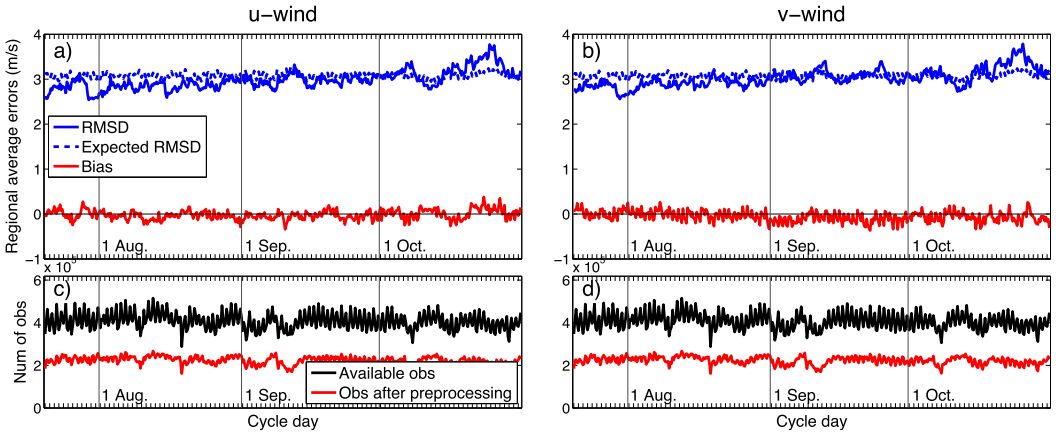


FIG. 2. (a),(b) Domain-average time series of observation-space statistics for (left) $u$ and (right) $v$. The statistics include RMSDs (solid blue lines), expected RMSDs (dashed blue lines), and mean error (red lines). (c),(d) Number of observations (black lines) and number of observations that pass thinning and quality control procedures (red lines).

potential deficiencies in the HWRF Model, data assimilation system, or error characterization for incoming measurements. The vertical extent of $q_v$ errors, for example, shows a negative moisture bias in low levels of the model domain which reaches a peak minimum value of $-1\,\mathrm{g\,kg^{-1}}$ near the surface. A time series of domain-mean RMSDs and bias show that both $q_v$ error metrics decrease steadily over the 3-month period (Fig. 4b). By comparing $q_v$ error profiles from the first and last months of the

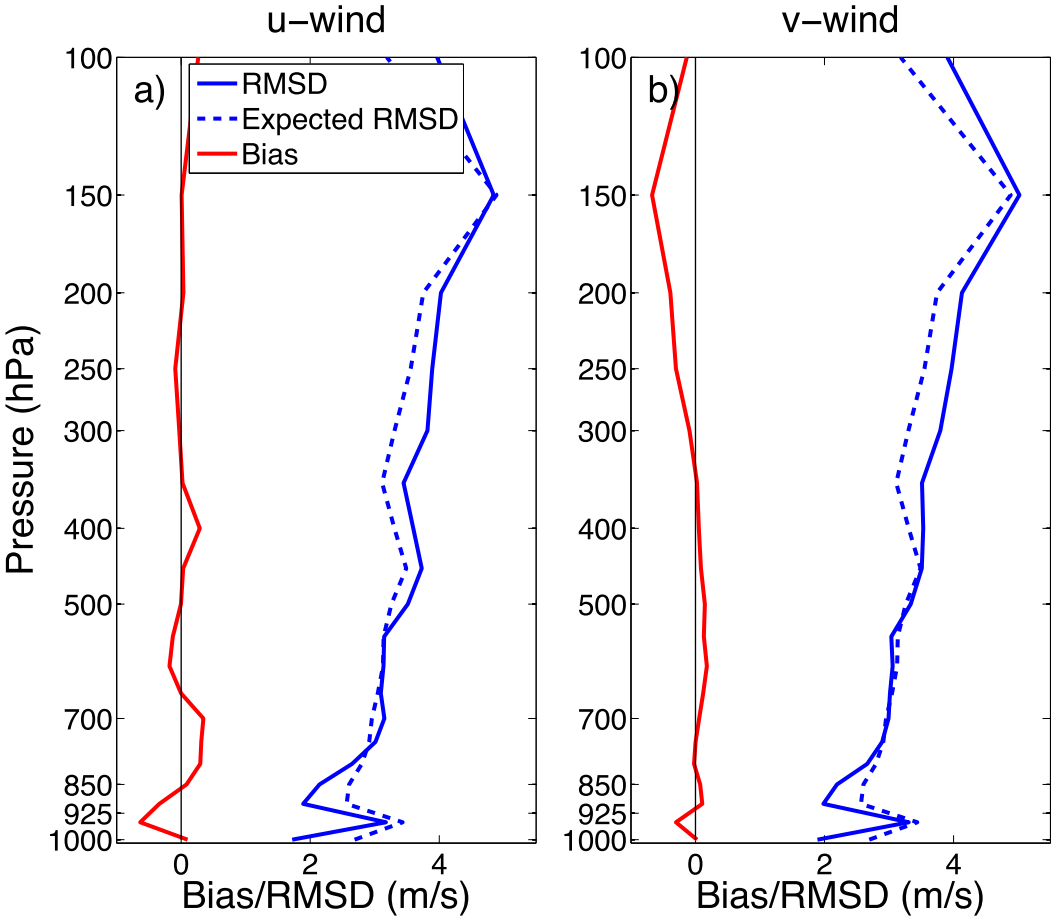


FIG. 3. Time-average RMSDs (solid blue lines), expected RMSDs (dashed blue lines), and mean error binned (red lines) by pressure level for (a) $u$ and (b) $v$.

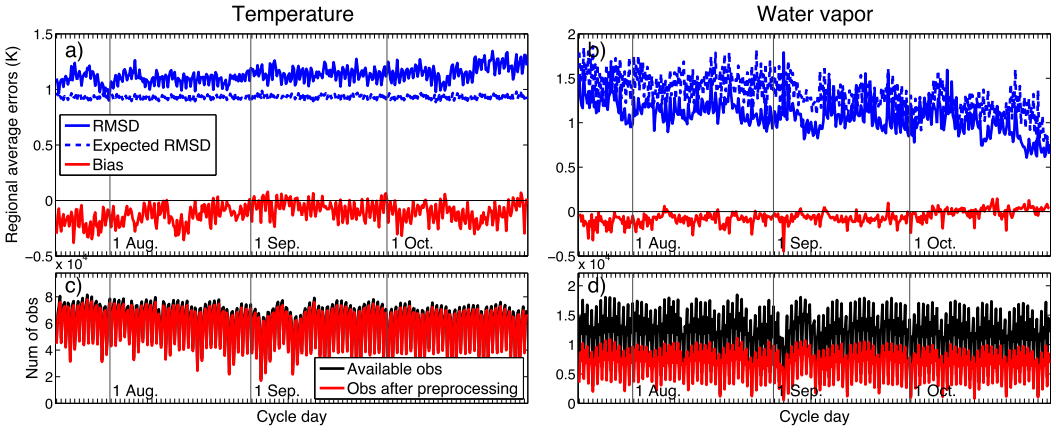


FIG. 4. As in Fig. 2, but for $T$ and $q_v$.

experiment (Fig. 6a) we find that the improvement occurs because of a decrease in magnitude for low-level moisture bias with time. The same figure also shows a reduction in $q_v$ RMSDs through all layers verified. A part of the $q_v$ bias comes from the GDAS initial conditions at the beginning of the experiment. Its reduction with time closely follows the adjustment of satellite bias correction coefficients for satellite radiances, which are taken to be the initial GDAS values over the experiments; we discuss this factor in section 4b. In addition, the verification shows overestimated prior ensemble spread in $q_v$ and underestimated prior ensemble spread in $T$ (Figs. 4 and 5). This result indicates possible deficiencies in error growth by the EnKF data assimilation, or a misspecification of observation error variances for the observing platforms providing data.

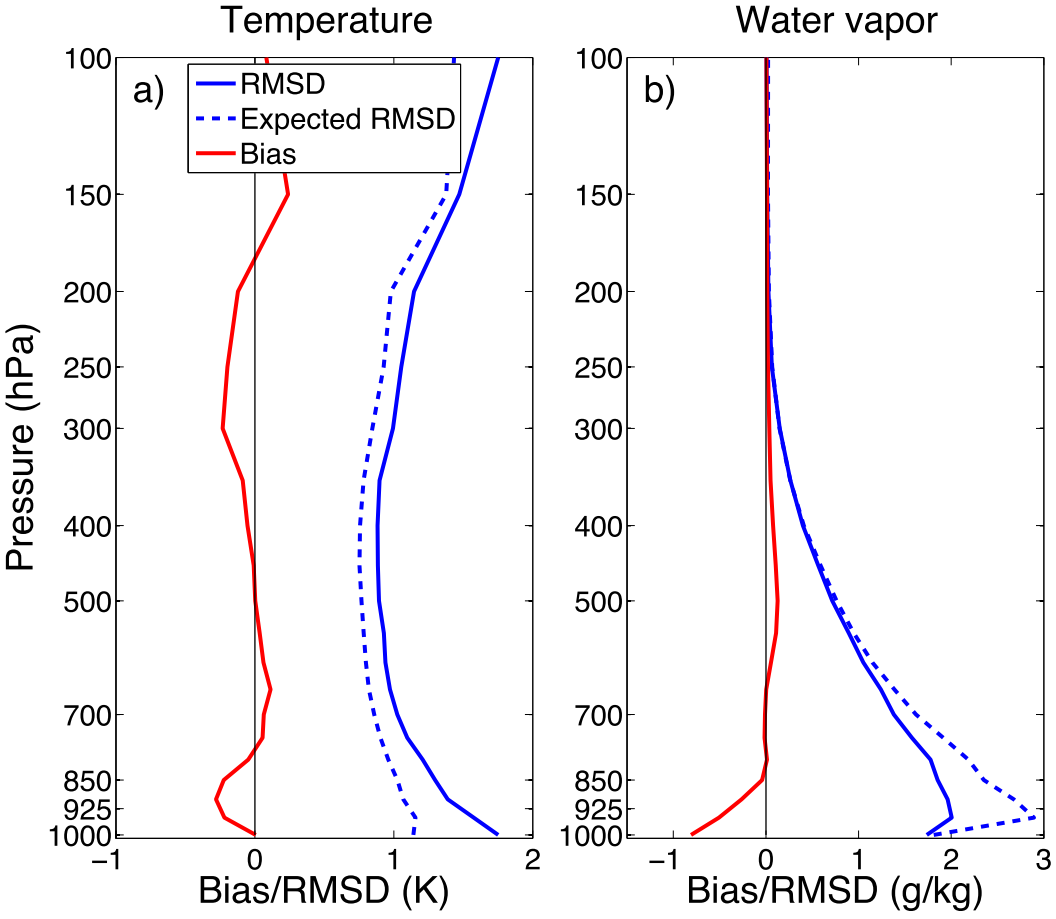


FIG. 5. As in Fig. 3, but for $T$ and $q_v$.

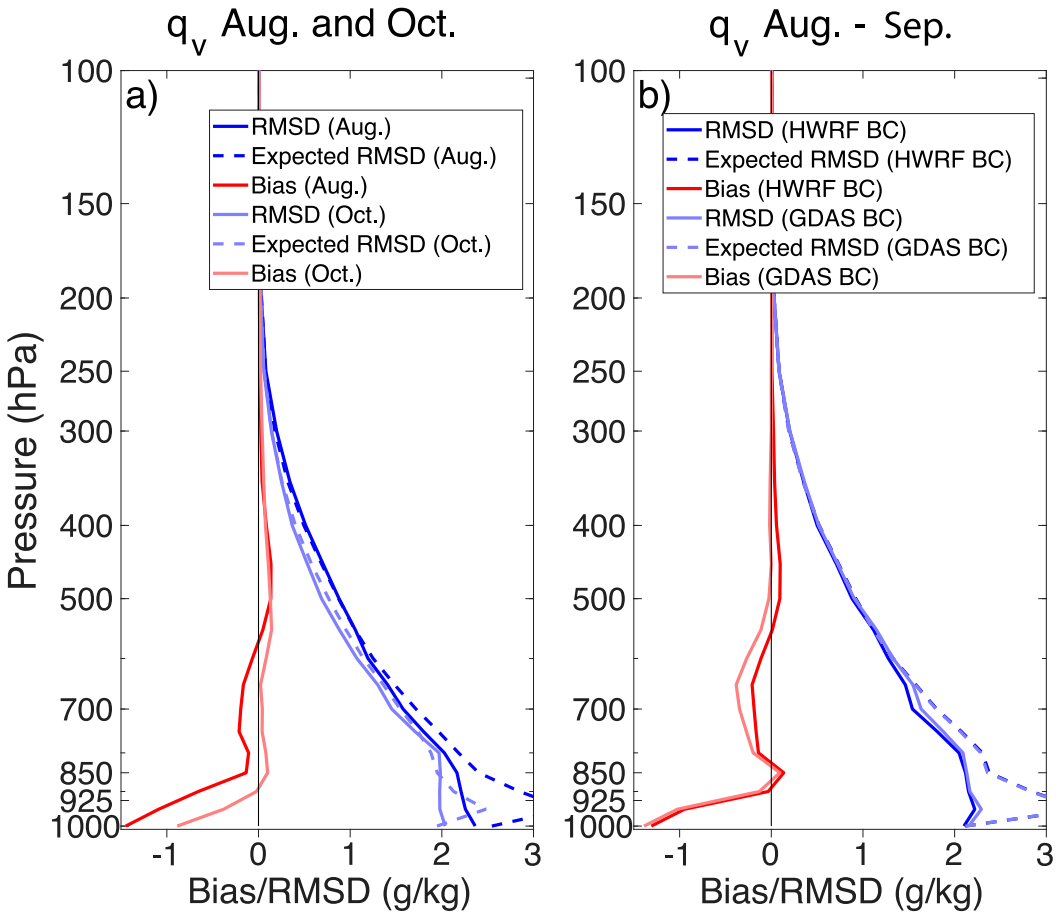


FIG. 6. Time-average errors binned by pressure level are stratified for $q_v$ over (a) the months of August (dark lines) and October (light lines). (b) The same metrics are averaged over the two months of August–September to compare HWRF BC (dark lines) and GDAS BC (light lines) experiments.

The sequential data assimilation still contains clear biases and suboptimal uncertainty estimates for $T$ and $q_v$ after calibration. Identifying and removing the source of these biases would require a more extensive examination of individual modeling system components than provided in the current study, but the evidence for these biases clearly exists in our experiments. Before arriving at these results, much more egregious problems in the HWRF modeling system produced a significant amount of bias or "model drift," which was diagnosed quickly from the observation-space statistics. For example, Fig. 7 shows a week-long time series of prior and posterior mean RMSDs, expected RMSDs, and bias for $q_v$ observations, starting from 20 July.[2] The errors are calculated over the Northern Hemisphere of our domain, where we identified the largest amount of bias in qualitative comparisons of prior fields and observations. In this figure, the $q_v$ statistics clearly show model drift for specific configurations of the HWRF Model. These experiments reveal that 2017 and earlier versions of the HWRF Model incorrectly specify the geographical locations of model grid points in the domain when nests are not present, which leads to large errors in shortwave radiation for single-domain configurations. The model error results in a large bias in surface thermodynamic variables that

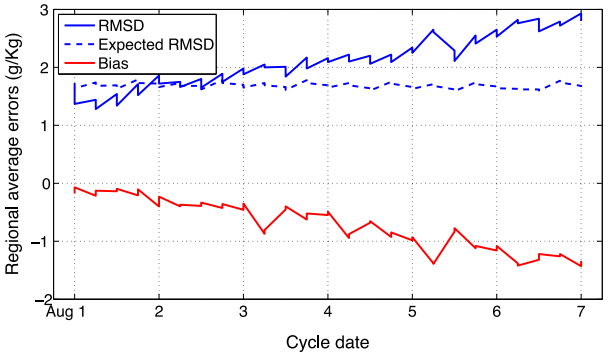


FIG. 7. Domain-average time series of prior and posterior RMSDs, expected RMSDs, and bias for $q_v$ during initial trial of AOML-UMD ensemble system.

[2] Prior and posterior errors are plotted together at the same update times, thus producing the observed "sawtooth" pattern.

carries through to subsequent data assimilation cycles. We suspect this issue may have impacted past decisions regarding data assimilation development for the HWRF Model. For example, section 1.1 of Tallapragada et al. (2014) note problems in attempting sequential data assimilation for large-scale fields in the HWRF Model, citing a notable amount of model drift in their experiments. Deficiencies of this type can be difficult to diagnose by validating model forecasts initialized from a different model (such as GFS) or using partial cycling, as they are often indistinguishable from the many other error sources present in free-running forecasts. Once identified by the cycling DA experiments, we implemented a correction to the model that allowed this study to continue.

### b. Satellite radiance data assimilation capability

A secondary objective for carrying out the 2017 real-time testing is to experiment with time-dependent satellite bias correction for the HWRF Model. Deficiencies in physical parameterization schemes, radiative transfer models, and sensors on satellite instruments can lead to systematic mismatches between prior model states and radiance measurements (Rizzi and Matricardi 1998). These errors often manifest themselves as a bias in radiance innovations, which must be removed to satisfy assumptions built into data assimilation algorithms. A common strategy for global modeling systems is to perform an adaptive estimation of biases (e.g., Derber and Wu 1998; Harris and Kelly 2001). Online bias calculations can capture temporal changes in error statistics, such as those caused by instrument sensor failure or seasonal shifts in model bias characteristics (Auligné et al. 2007).

In GSI, bias is included in measurement operator calculations for brightness temperatures $T_b$. This operator is given by

$$\tilde{h}(\mathbf{x},\boldsymbol{\beta}) = h(\mathbf{x}) + \sum_{i=1}^{N} \beta_i p_i(\mathbf{x}) + \sum_{j=1}^{K} \beta_{N+j} \phi^j, \qquad (1)$$

where $h(\mathbf{x})$ is the original measurement operator containing a radiative transfer model, and bias is calculated as a weighted sum of airmass predictors $p_i(\mathbf{x})$, and a $K$th-order polynomial function of viewing angle $\phi$. The airmass predictors include a global offset term (i.e., $p_0 = 1$), total column cloud liquid water, temperature lapse rate, and the square of lapse rate, while the viewing angle bias is represented by a fourth-order polynomial. The vector $\boldsymbol{\beta}$, which stores the weighting coefficient in the right two terms of (1), is updated alongside model state vectors within the EnKF using the strategy outlined in Miyoshi et al. (2010). This process differs from the approach used by GSI 3DVar and 4DEnVar, which contain an extra term for $\boldsymbol{\beta}$ in the variational cost function. We refer readers to Zhu et al. (2014) for a full description and motivation for the radiance bias model used by GSI.

Estimating $\boldsymbol{\beta}$ for statistical bias correction schemes requires large samples of independent model-observation comparisons, which is not often feasible for regional models with small domains (Lin et al. 2017). For example, the operational HWRF system must rely on $\boldsymbol{\beta}$ values estimated from the GDAS. Figure 1 shows the domain size and spatial distribution of satellite radiance locations at one time during the 2017 experiment period. Because our data assimilation system considers a much larger sample of radiance innovations than current HWRF configurations, we revisit current procedures used to assimilate satellite measurements. We assess the utility of performing our own bias correction by considering observation-space prior statistics and acceptance rates from conventional and radiance measurements. These statistics can identify potential problems with satellite data assimilation procedures in the AOML-UMD system and help guide future research directions.

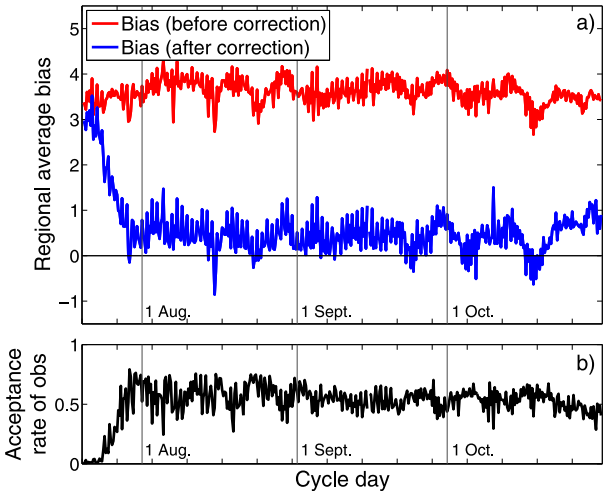


FIG. 8. (a) Time series of domain-average prior observation-space brightness temperature bias before bias correction (red lines) and after bias correction (blue lines) for channel 10 of AMSU-A. (b) Time series of acceptance rate for the measurements.

As shown in section 4a, prior RMSDs, biases, and acceptance rates for conventional observations show no decline in short-range forecast performance over the test period, which is one indication that the EnKF bias correction and assimilation steps are operating effectively. A close inspection of individual instrument channels provides a more direct examination of this feature. For example, Fig. 8a shows a time series of prior $T_b$ bias with and without bias correction for channel 10 of AMSU-A on the *NOAA-15* polar-orbiting satellite. Because this channel is sensitive to stratospheric temperature, the low model top of our HWRF domain (at 2 hPa) provides a constant source of $T_b$ bias in measurement operator calculations. After about 10 days of sequential data assimilation, the bias correction scheme is able to capture most of the bias and increase the observation acceptance rate from 0% to about 50% from a thinned subset of observations (Fig. 8b). This demonstration provides a sanity check on the bias model and bias correction coefficient estimation features of the GSI EnKF, since a major source of systematic error is known. In practice, the operational HWRF forecast system uses a blending of prior HWRF and GFS thermodynamic fields to fill-in upper levels of the atmosphere missing during radiance measurement operator calculations (Biswas et al. 2017). For simplicity, we turn off this option for the HWRF forecasting system and allow the bias correction procedures to account for the systematic error.

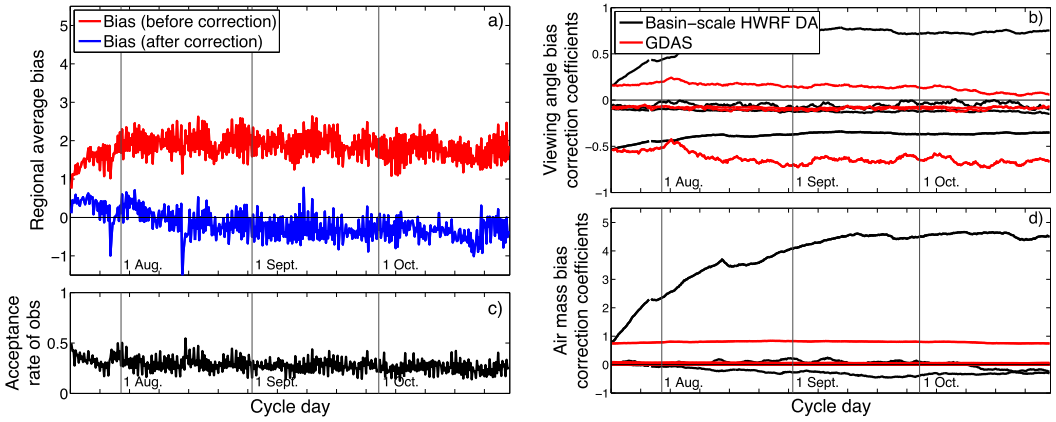


FIG. 9. (a) Time series of domain-average prior observation-space brightness temperature bias before bias correction (red lines) and after bias correction (blue lines) for channel 3 of MHS. (c) Time series of acceptance rate for the measurements. The time evolution of (b) airmass and (d) viewing angle bias correction coefficients for GDAS (red) and HWRF (black).

The AMSU-A example shown in Fig. 8 is a special case, where a truncated model top provides a large bias in observation-space priors, and the spatial and temporal coverage of measurements is sufficient for identifying the dominant error source. Though not shown, we find the bias correction algorithm to be less effective for certain observing systems, such as channels of the *GOES-15* sounder. We hypothesize this limitation follows from the lack of independent measurements passing through the data assimilation system; the number of unique GOES sounder observation locations over the HWRF domain is about an order of magnitude smaller than what is provided by other satellite platforms. This problem causes most *GOES-15* sounder measurements to be rejected by quality control.

We also identify situations where the bias correction operates effectively, while yielding a substantially different set of $\boldsymbol{\beta}$ coefficients than those estimated by GDAS for the GFS model. For this demonstration, we choose channel 3 (183 GHz) of the Microwave Humidity Sounder (MHS) on *NOAA-18*. The assimilation of these measurements is not directly affected by the 2-hPa model top specified in our HWRF configuration, as is the case for channel 10 of AMSU-A (see discussion above). It is, however, very sensitive to lower tropospheric moisture, which Fig. 5b suggests is vastly different between GFS and HWRF. Figure 9a shows the time series of domain-average prior $T_b$ bias for these measurements before and after applying the bias correction, and Figs. 9b and 9d show the corresponding change in angle and airmass $\beta_i$ terms relative to GDAS. The prior bias before performing a correction has values of about 1 K immediately after initializing the experiment from GFS, but bias values increase over the first few weeks before saturating to near 2 K. This finding suggests that, on average, HWRF yields a $\approx$1-K larger difference in MHS channel-3 $T_b$ for this satellite compared to GFS. During this period, the radiance bias correction scheme adjusts several coefficients in $\boldsymbol{\beta}$ to counteract the new bias sources. A major adjustment occurs in the global offset term for airmass coefficients, which can be easily distinguished as the largest term in Fig. 9d. Likewise, we observe notable changes in $\beta_i$ for the squared lapse rate predictor (coefficient with largest negative value in Fig. 9d) and angle terms (Fig. 9b). Though not shown, a similar adjustment is found for channel 3 of the MHS on the *Meteorological Operational satellite-A* (*Metop-A*).

The above findings motivate an additional data assimilation experiment to investigate the utility of online bias correction over the operational strategy for HWRF, which uses values directly from GDAS. Following a July spin up period, we repeat the sequential data assimilation over the months of August and September, but adopt time-dependent $\boldsymbol{\beta}$ from GDAS instead of estimating these coefficients online. For this test, measurement operators take advantage of the same strategy adopted by the operational HWRF Model, which is to blend prior HWRF and GFS thermodynamic fields to fill-in the atmosphere above the HWRF Model top (Biswas et al. 2017). In doing so, the additional experiment (denoted GDAS BC) provides a clean view of how GDAS biases can influence HWRF analyses through satellite data assimilation.

We find the choice of bias correction to have the largest impact on $q_v$, which is to be expected given that this variable exhibits the largest change in RMSD and bias over the experiment period (Fig. 6a). For this reason, we omit any discussion on wind and $T$ in this comparison. Figure 6b shows profiles of prior $q_v$ RMSD, bias, and spread for the online (HWRF BC) and GDAS BC experiments, averaged over the months of August and September. The figure indicates smaller RMSDs for the online experiment at all vertical levels. It also shows smaller bias in the HWRF BC experiment from the surface to 500 hPa, suggesting that GDAS bias coefficients explain a portion of the elevated low-level $q_v$ errors near the beginning of experiments. This result is somewhat counterintuitive, since the HWRF BC experiment yields the smallest prior fit to channel-3 MHS measurements at the beginning of our experiments (Fig. 9), where $\boldsymbol{\beta}$ was closer to GDAS values. Our findings underscore the complexity of interpreting comparisons between models and observations that are thought to be biased. In this case, we hypothesize that either persistent errors in the GFS partially cancel errors in the measurement

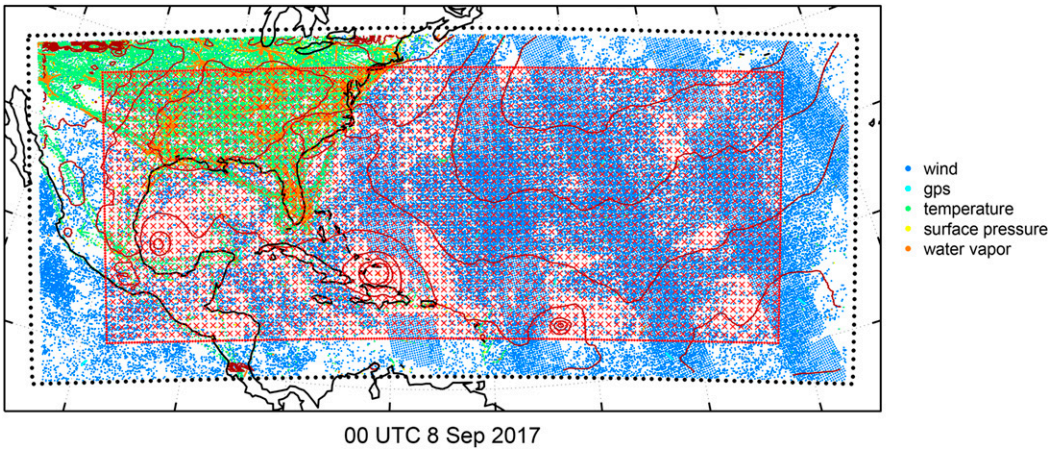


FIG. 10. The HWRF domain used for local PF experiments is outlined on a map with land boundaries. Also plotted are contours of MSLP every 5 hPa for a single posterior member, locations of conventional measurements at one time during the experiments (0000 UTC 8 Sep 2017), and the verification domain (red hatched region). Observation types are color coded according to the legend on the right.

operator or observed radiance measurements, or an inappropriate specification of $\boldsymbol{\beta}$ propagates bias to $q_v$ during early cycles. The experiment also reveals that small positive $q_v$ biases emerging above 500 hPa in the HWRF BC experiment are caused by a suboptimal treatment of upper-level biases. The result likely comes from the low HWRF Model top and decision to not adopt the operational GDAS-HWRF prior blending strategy in this experiment, thus providing further justification for this practice.

While online satellite bias correction may not be operationally feasible for HWRF, comparisons of this type provide valuable information regarding the HWRF Model—or how its bias characteristics differ from GFS. Online bias correction approaches also provide a valid benchmark for examining the impact of new satellite data assimilation and quality control strategies, which may be sensitive to biases and require a dedicated framework that is not heavily influenced by outside bias sources that cannot be accounted for.

### c. *New algorithm testing*

In addition to providing a platform for exploring choices of model and satellite data assimilation configuration, the AOML-UMD ensemble system also provides an appropriate framework for rigorously testing new data assimilation methodology. This section describes how previously run experiments are used for this purpose.

As discussed in previous sections, satellite radiance measurements require a dedicated sequential data assimilation system operating on either a global or large regional domain to accurately estimate and remove biases. While these biases can depend on assumptions made by data assimilation algorithms, they tend to be dominated by observations, measurement operators, and model errors (Dee and da Silva 1998; Eyre 2016). Therefore, modeling experiments constructed to examine new data assimilation algorithms can recycle previously estimated bias coefficients, which reduces the restriction on domain size typically needed for these measurements, so long as the model resolution and physical parameterization schemes are kept constant. This simplification greatly reduces the domain size needed to perform numerous experiments, thus making the rigorous evaluation of new data assimilation methodology significantly more affordable.

In this section, we leverage past experiments performed with the AOML-UMD ensemble system to test a new ensemble data assimilation system based on particle filters (PFs). Particle filters are sequential Monte Carlo methods that make no parametric assumptions for the underlying prior and posterior error distributions used during data assimilation. In doing so, they avoid Gaussian assumptions, which are built into the formulation of EnKFs, variational methods, and hybrid methods used within modern weather prediction systems (Bannister 2017). The theoretical benefits of PFs make them powerful tools for studying and predicting nonlinear dynamical systems, which has led to a considerable effort in the geoscience community to implement PF-based techniques for nonlinear problems such as weather prediction; see Van Leeuwen (2019) for a review.

Similar to EnKFs, high-dimensional applications of PFs require a careful treatment of sampling errors through strategies such as localization. One such method is the Poterjoy (2016) local PF, which has been tested extensively for low- and high-dimensional idealized models (Poterjoy and Anderson 2016; Poterjoy et al. 2017), and to a limited extent for real convective-scale applications with the WRF model (Poterjoy et al. 2019). These studies, however, have not demonstrated the feasibility of using the local PF for large multiscale applications, such as tropical cyclone weather forecasting. Given that much of the complexity in applying localized particle filter algorithms comes in the sampling step (i.e., the process of generating equally likely members from the posterior distribution), modeling systems that are restricted to partial cycling are not well suited for vetting algorithms of this type.

Using the AOML-UMD ensemble system, we add local PF capabilities to GSI and perform direct comparisons with the GSI EnKF over a month-long period. These experiments use a reduced domain size, shown in Fig. 10, and take place over the

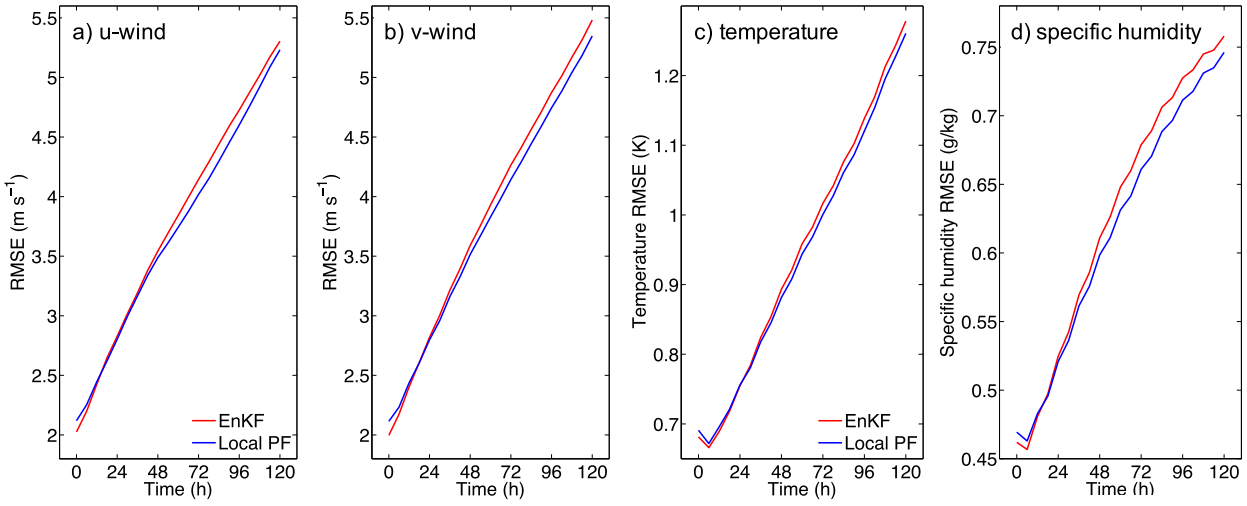


FIG. 11. Ensemble mean forecast RMSDs as a function of forecast lead time for (a) $u$, (b) $v$, (c) $T$, and (d) $q_v$ for the EnKF (red lines) and local PF (blue lines).

entire month of September 2017. Several major hurricanes from the 2017 season (i.e., Irma, Jose, Katia, Lee, and Maria) formed over the selected domain and period, thus providing a challenging test period for comparing data assimilation algorithms. The local PF and EnKF assimilate the same conventional and clear-sky satellite radiance measurements used for season-long testing, except satellite bias correction values come directly from the previously run large-domain experiments at each assimilation time. The domain configuration and ensemble size are also identical to past experiments, with the only exception being the lateral coverage of the domain.

Before performing comparisons with the EnKF, we thoroughly test various components of the new GSI local PF over a shortened (14-day) trial period to ensure the new algorithm works properly within the GSI framework. These tests include an examination of user-specified parameters needed for localization and inflation; see Poterjoy et al. (2019) for a full parameter list. Once properly configured, the performance of the local PF and EnKF are compared using 120-h ensemble forecasts generated every 12 h over the experiment period. To reduce computational cost, only the first 20 members from each ensemble are used in forecasts.

We compare the data assimilation strategies using error metrics that target overall performance of HWRF forecasts during the experiments—rather than traditional tropical cyclone-focused metrics often used for operational development. This decision is motivated by the small sample size available for metrics such as tropical cyclone track and intensity over the experiment, as well as their intrinsic limitations for measuring overall model and data assimilation performance. Ideally, model forecasts would be compared directly to point measurements during verification; e.g., (Poterjoy and Zhang 2014a,b). For the current application, the lack of upper-air in situ measurements over the open oceans limits the sample of verifying observation locations over the domain. We instead adopt GDAS analyses as our verifying dataset, which come from an ensemble four-dimensional variational smoother applied over a 6-h time windows (Kleist and Ide 2015). In addition to providing a verifying solution that contains more observations than our current experiments, there is no expectation for the GDAS solution to be biased more in the direction of either experiment performed in this study.

We use RMSDs between ensemble mean forecasts and GDAS analyses to compare the overall performance of the local PF and EnKF. These errors are averaged spatially over the verification domain shown by the red hatched region in Fig. 11. This region consists of grid points that are displaced at least 500 km laterally from HWRF domain boundaries, and spaced horizontally every 54 km and vertically every 500 m; this spacing was sufficient for reducing most of the large correlations between verifying grid points. By focusing on domain-wide performance of the different data assimilation techniques, forecast skill is dominated by synoptic-scale factors, but inevitably contains some signal from tropical cyclone position, intensity and structure. Given the coarse model grid spacing used for these experiments, we feel this metric provides the most appropriate indication of how well each data assimilation method performs.

Figure 11 shows time series plots of RMSDs for $u$, $v$, $T$, and $q_v$, averaged over all 52 sets of forecasts generated during the experiment. These results exclude HWRF forecasts from the first 5 days of sequential data assimilation to remove any memory of the 0000 UTC 1 September. GDAS ensemble used to initialize HWRF members on the beginning of the experiment. Ensemble mean forecasts are also scrutinized by analyzing the frequency at which the local PF provides lower forecast RMSDs than the EnKF. To summarize forecast skill in this manner, Fig. 12 shows the median and quartiles of EnKF minus local PF forecast RMSDs, with blue values indicating forecast times where the statistics indicate lower error values for the local PF. Unlike the temporally averaged metric in Fig. 11, the quantiles shown in Fig. 12 are not skewed by the presence of forecast outliers that may exist in the verifying sample, thus providing a secondary metric for forecast skill. Overall, both sets of statistics indicate marginal forecast

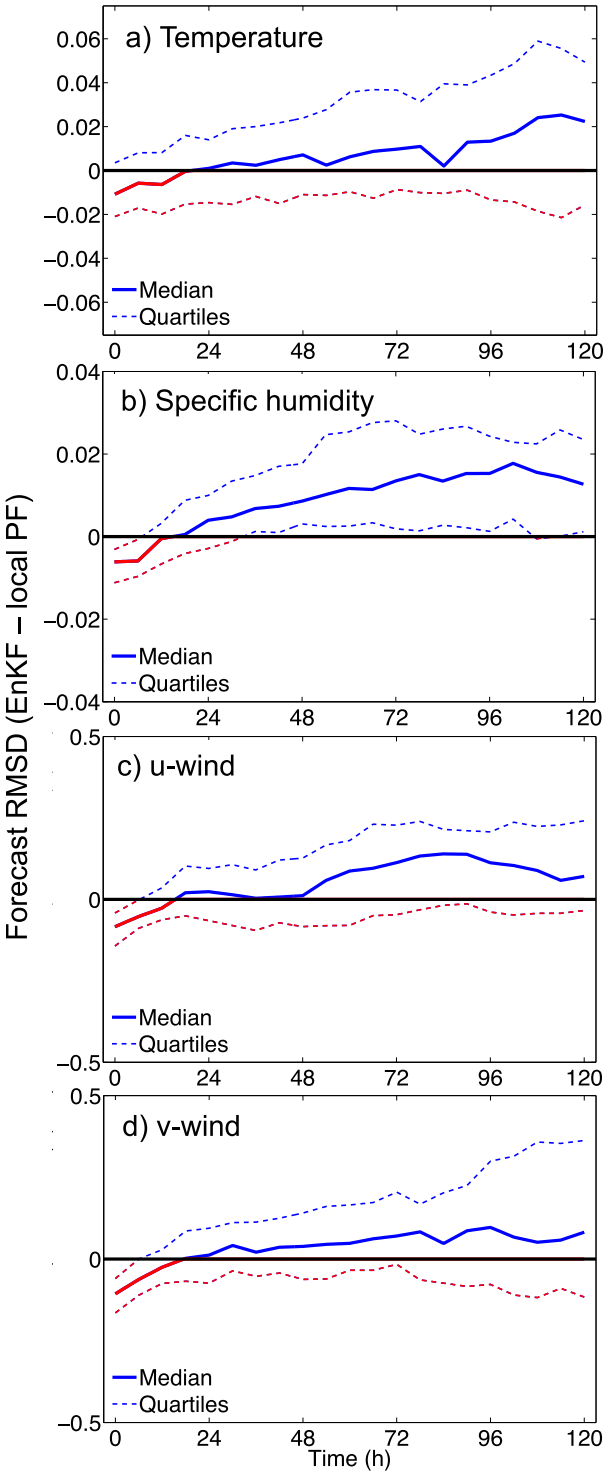


FIG. 12. Differences between ensemble mean forecast RMSDs for EnKF and local PF experiments as a function of forecast lead time for (a) $u$, (b) $v$, (c) $T$, and (d) $q_v$. Solid lines indicate median of differences and dashed lines indicate lower and upper quantiles. Red lines indicate metrics where the EnKF produces lower errors, while blue lines indicates metrics where the local PF produces lower errors.

improvements of the local PF over the benchmark EnKF, despite not matching the GDAS solution as closely at early forecast times. In addition to producing smaller average medium-range forecast errors than the EnKF, the local PF tends to produce lower RMSDs more frequently. The most persistent result occurs for $q_v$ forecasts, where the local PF produces a closer match to GDAS over 75% of the time for forecasts beyond 36 h.

Though not shown in the current manuscript, local PF forecasts benefit from a more dynamically consistent depiction of mesoscale features in the model domain, which were loosely constrained by observation networks present over the open oceans. A more detailed examination of the data assimilation algorithms and experiments, including an explanation for why the EnKF yields a closer match to GDAS at early forecast times, will be discussed in a future manuscript. The major findings from this comparison illustrate the value of a dedicated data assimilation framework for examining new methodology for numerical weather prediction. In this case, subtle differences in model forecast performance are found by comparing two different methods over month-long experiments using sequential data assimilation and satellite biases correction that is consistent with the HWRF Model.

## 5. Discussion and conclusions

This manuscript introduces the AOML-UMD ensemble prediction system, a standalone sequential data assimilation and forecasting framework for the HWRF Model. The methodology described in this study aims to fill a gap that currently exists between regional and global forecast systems that are used to provide numerical guidance for tropical cyclones. It also provides insight into how future modeling systems may evolve for tropical cyclone applications in the advent of greater computing resources and a unified modeling framework.

The experimental modeling system relies purely on ensemble data assimilation, owing to its simplicity and straightforward interpretation. Generating accurate deterministic forecasts for tropical cyclones is a challenging task for numerical weather prediction models because of initial condition uncertainty and the misrepresentation of physical processes during model integration. A well-calibrated ensemble can offer a much more complete depiction of how a given storm may evolve with time by delivering a set of equally likely model forecasts conditioned on past and present observations. Ensembles also provide the only affordable means of exploring statistical properties of non-Gaussian forecast probability densities, which evolve from nonlinear processes in weather models. While not discussed in the current manuscript, this factor is important for examining marginal distributions of the forecast probability density, such as tropical cyclone track and intensity metrics, and may be necessary for measuring incremental improvements in weather models. This approach differs from pure deterministic methods, which focus primarily on trying to identify the mode.

The above strategy is not always feasible for real-time use at operational centers, but it provides an invaluable framework for testing various components of a modeling system during

retrospective experiments. To demonstrate the utility of a dedicated sequential framework, we perform a series of experiments with the HWRF Model—an operational model that has never been examined extensively in this context. For these experiments, we focus primarily on multimonth observation-space error estimation for short-range forecasts, satellite data assimilation capabilities in HWRF, and month-long testing of a new data assimilation algorithm. In addition to uncovering multiple software errors in the operational HWRF system, these tests provide a number of important findings that would have been difficult to achieve without sequential data assimilation. We find that a persistent low-level dry bias develops early in our experiments when using an extensive coarse-resolution model configuration that is comparable to the parent domain used by the operational HWRF Model. This bias likely comes from the use of GDAS to initialize the ensemble and satellite bias correction coefficients at the first cycle of our experiments. The moisture bias approaches zero after multiple months of cycling, but dissipates less when GDAS bias correction coefficients are adopted in place of online estimates. While the use of GDAS bias correction values appears to be one deficiency in the operational strategy, the use of blended HWRF-GFS prior model solutions by the operational system removes an upper-level positive moisture bias that occurs in pure online experiments. Following the online estimate of satellite bias correction coefficients for HWRF, we reuse the coefficients in reduced-domain retrospective experiments to rigorously tune and test a data assimilation method based that has never been applied for synoptic-scale weather prediction. This method, called the local PF, is added to GSI so it can be compared directly with the EnKF. We find the local PF to provide stable performance over the month-long testing period and provide marginal improvements over the benchmark EnKF. These results are encouraging, owing to the theoretical advantages of the local PF for larger sample sizes and highly nonlinear applications, such as all-sky radiance data assimilation—neither of which are examined in the current study.

The forecasting system described here is designed to inform ongoing efforts to develop a model infrastructure that bridges between research and operations. In particular, it explores potential avenues for the future NOAA hurricane analysis and forecasting system (HAFS), aimed at meeting both short- and long-term operational forecast needs. In doing so, the current system targets improvements in probabilistic weather guidance through: 1) better preprocessing and assimilation of measurements, especially those coming from satellite sensors; 2) novel data assimilation strategies such as the local particle filter; and 3) a more rigorous estimation of model bias. Ultimately, our findings strongly suggest that sequential data assimilation efforts should precede advanced model testing for new weather prediction systems such as HAFS.

*Acknowledgments.* This study was supported by NOAA Grant NA19NES4320002 [Cooperative Institute for Satellite Earth System Studies (CISESS)] at the University of Maryland/ESSIC. Parts of this research were also performed while the first author held a National Research Council Research Associateship award at the NOAA Atlantic Oceanographic and Atmospheric Laboratory. The first author also thanks Ryan Torn for providing the initial motivation for this project.

*Data availability statement.* All metadata and ensemble analysis members produced over the course of this work are archived on the National Environmental Security Computing Center (NESCC) and available upon request.

## REFERENCES

Aksoy, A., S. Lorsolo, T. Vukicevic, K. J. Sellwood, S. D. Aberson, and F. Zhang, 2012: The HWRF Hurricane Ensemble Data Assimilation System (HEDAS) for high-resolution data: The impact of airport Doppler radar observations in an OSSE. *Mon. Wea. Rev.*, **140**, 1843–1862, https://doi.org/10.1175/MWR-D-11-00212.1.

——, S. D. Aberson, T. Vukicevic, K. J. Sellwood, S. Lorsolo, and X. Zhang, 2013: Assimilation of high-resolution tropical cyclone observations with an ensemble Kalman filter using NOAA/AOML/HRD's HEDAS: Evaluation of the 2008–11 vortex-scale analyses. *Mon. Wea. Rev.*, **141**, 1842–1865, https://doi.org/10.1175/MWR-D-12-00194.1.

Alaka, G. J. J., X. Zhang, S. G. Gopalakrishnan, S. B. Goldenberg, and F. D. Marks, 2017: Performance of basin-scale HWRF tropical cyclone track forecasts. *Wea. Forecasting*, **32**, 1253–1271, https://doi.org/10.1175/WAF-D-16-0150.1.

Alaka, J., D. Sheinin, B. Thomas, L. Gramer, Z. Zhang, B. Liu, H.-S. Kim and A. Mehra, 2020: A hydrodynamical atmosphere/ocean coupled modeling system for multiple tropical cyclones. *Atmosphere*, **11**, 869, https://doi.org/10.3390/atmos11080869.

Atlas, R., V. Tallapragada, and S. G. Gopalakrishnan, 2015: Advances in tropical cyclone intensity forecasts. *Mar. Technol. J.*, **49**, 149–160, https://doi.org/10.4031/MTSJ.49.6.2.

Auligné, T., A. P. McNally, and D. P. Dee, 2007: Adaptive bias correction for satellite data in a numerical weather prediction system. *Quart. J. Roy. Meteor. Soc.*, **133**, 631–642, https://doi.org/10.1002/qj.56.

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

Biswas, M. K., and Coauthors, 2017: Hurricane Weather Research and Forecasting (HWRF) Model: 2017 Scientific Documentation. NOAA/NCAR/Development Testbed Center Rep., 111 pp., https://dtcenter.org/HurrWRF/users/docs/scientific_documents/HWRFv3.9a_ScientificDoc.pdf.

Cavallo, S. M., J. Berner, and C. Snyder, 2016: Diagnosing model errors from time-averaged tendencies in the Weather Research and Forecasting (WRF) Model. *Mon. Wea. Rev.*, **144**, 759–779, https://doi.org/10.1175/MWR-D-15-0120.1.

Christophersen, H., A. Aksoy, J. Dunion, and K. Sellwood, 2017: The impact of NASA Global Hawk unmanned aircraft dropwindsonde observations on tropical cyclone track, intensity, and structure: Case studies. *Mon. Wea. Rev.*, **145**, 1817–1830, https://doi.org/10.1175/MWR-D-16-0332.1.

Dee, D. P., and A. M. da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.*, **124**, 269–295, https://doi.org/10.1002/qj.49712454512.

Derber, J. C., and W.-S. Wu, 1998: The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Mon. Wea. Rev.*, **126**, 2287–2299, https://doi.org/10.1175/1520-0493(1998)126<2287:TUOTCC>2.0.CO;2.

Dowell, D., 2020: HRRR Data-Assimilation System (HRRRDAS) and HRRRE Forecasts: Documentation. 8 pp., https://rapidrefresh.noaa.gov/internal/pdfs/2020_Spring_Experiment_HRRRE_Documentation.pdf.

Eyre, J. R., 2016: Observation bias correction schemes in data assimilation systems: A theoretical study of some of their properties. *Quart. J. Roy. Meteor. Soc.*, **142**, 2284–2291, https://doi.org/10.1002/qj.2819.

Gall, R., J. Franklin, F. Marks, E. Rappaport, and F. Toepfer, 2013: The Hurricane Forecast Improvement Project. *Bull. Amer. Meteor. Soc.*, **94**, 329–343, https://doi.org/10.1175/BAMS-D-12-00071.1.

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

Gopalakrishnan, S. G., Q. Liu, T. Marchok, D. Sheinin, N. Surgi, R. Tuleya, R. Yablonsky, and X. Zhang, 2010: Hurricane Weather Research and Forecasting (HWRF) Model scientific documentation. NOAA/NCAR/Development Testbed Center Rep., 75 pp.

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

Harris, B. A., and G. Kelly, 2001: A satellite radiance-bias correction scheme for data assimilation. *Quart. J. Roy. Meteor. Soc.*, **127**, 1453–1468, https://doi.org/10.1002/qj.49712757418.

Kleist, D. T., and K. Ide, 2015: An OSSE-based evaluation of hybrid variational-ensemble data assimilation for the NCEP GFS. Part II: 4DEnVar and hybrid variants. *Mon. Wea. Rev.*, **143**, 452–470, https://doi.org/10.1175/MWR-D-13-00350.1.

Lin, H. D., S. S. Weygandt, A. H. N. Lim, M. Hu, J. M. Brown, and S. G. Benjamin, 2017: Radiance preprocessing for assimilation in the hourly updating Rapid Refresh mesoscale model: A study using AIRS data. *Wea. Forecasting*, **32**, 1781–1800, https://doi.org/10.1175/WAF-D-17-0028.1.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307, https://doi.org/10.3402/tellusa.v21i3.10086.

Lu, X., X. Wang, Y. Li, M. Tong, and X. Ma, 2017: GSI-based ensemble-variational hybrid data assimilation for HWRF for hurricane initialization and prediction: Impact of various error covariances for airborne radar observation assimilation. *Quart. J. Roy. Meteor. Soc.*, **143**, 223–239, https://doi.org/10.1002/qj.2914.

Marks, F., N. Kurkowski, M. DeMaria, and M. Brennan, 2019: Hurricane Forecast Improvement Program Five-Year Plan: 2019–2024. NOAA, 83 pp., http://www.hfip.org/documents/HFIP_Strategic_Plan_20190625.pdf.

Martin, G. M., S. F. Milton, C. A. Senior, M. E. Brooks, S. Ineson, T. Reichler, and J. Kim, 2010: Analysis and reduction of systematic errors through a seamless approach to modeling weather and climate. *J. Climate*, **23**, 5933–5957, https://doi.org/10.1175/2010JCLI3541.1.

Miyoshi, T., Y. Sato, and T. Kadowaki, 2010: Ensemble Kalman filter and 4D-Var intercomparison with the Japanese operational global analysis and prediction system. *Mon. Wea. Rev.*, **138**, 2846–2866, https://doi.org/10.1175/2010MWR3209.1.

NOAA/GFDL, 2018: FV3 documentation and references. Accessed June 2018, https://www.gfdl.noaa.gov/fv3/fv3-documentation-and-references/.

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

——, and F. Zhang, 2014a: Intercomparison and coupling of ensemble and four-dimensional variational data assimilation methods for the analysis and forecasting of Hurricane Karl (2010). *Mon. Wea. Rev.*, **142**, 3347–3364, https://doi.org/10.1175/MWR-D-13-00394.1.

——, and ——, 2014b: Predictability and genesis of Hurricane Karl (2010) examined through the EnKF assimilation of field observations collected during PREDICT. *J. Atmos. Sci.*, **71**, 1260–1275, https://doi.org/10.1175/JAS-D-13-0291.1.

——, and J. L. Anderson, 2016: Efficient assimilation of simulated observations in a high-dimensional geophysical system using a localized particle filter. *Mon. Wea. Rev.*, **144**, 2007–2020, https://doi.org/10.1175/MWR-D-15-0322.1.

——, R. A. Sobash, and J. L. Anderson, 2017: Convective-scale data assimilation for the Weather Research and Forecasting Model using the local particle filter. *Mon. Wea. Rev.*, **145**, 1897–1918, https://doi.org/10.1175/MWR-D-16-0298.1.

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

Rizzi, R., and M. Matricardi, 1998: The use of TOVS clear radiances for numerical weather prediction using an updated forward model. *Quart. J. Roy. Meteor. Soc.*, **124**, 1293–1312, https://doi.org/10.1002/qj.49712454813.

Rodwell, M. J., and T. Jung, 2008: Understanding the local and global impacts of model physics changes: An aerosol example. *Quart. J. Roy. Meteor. Soc.*, **134**, 1479–1497, https://doi.org/10.1002/qj.298.

Rotunno, R., and C. Snyder, 2008: A generalization of Lorenz's model for the predictability of flows with many scales of motion. *J. Atmos. Sci.*, **65**, 1063–1076, https://doi.org/10.1175/2007JAS2449.1.

Schraff, C., H. Reich, A. Rhodin, A. Schomburg, K. Stephan, A. Perianez, and R. Potthast, 2016: Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). *Quart. J. Roy. Meteor. Soc.*, **142**, 1453–1472, https://doi.org/10.1002/qj.2748.

Schwartz, C. S., G. S. Romine, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2015: A real-time convection-allowing ensemble prediction system initialized by mesoscale ensemble Kalman filter analyses. *Wea. Forecasting*, **30**, 1158–1181, https://doi.org/10.1175/WAF-D-15-0013.1.

Steward, J. L., A. Aksoy, and Z. S. Haddad, 2017: Parallel direct solution of the ensemble square root Kalman filter equations with observation principal components. *J. Atmos. Oceanic Technol.*, **34**, 1867–1884, https://doi.org/10.1175/JTECH-D-16-0140.1.

Tallapragada, V., and Coauthors, 2014: Hurricane Weather Research and Forecasting (HWRF) Model: 2014 scientific documentation, HWRF V3.6a. NCAR Developmental Testbed Center Rep., 105 pp.

Toepfer, F., H. Tolman, T. L. Schneider, I. Stajner, and S. Warren, 2018: The Next Generation Global Prediction System (NGGPS) program update. *Eighth Conf. on Transition of Research to Operations*, Austin, TX, Amer. Meteor. Soc., 1.1, https://ams.confex.com/ams/98Annual/webprogram/Paper330985.html.

Van Leeuwen, P. J., H. R. Künsch, L. Nerger, R. Potthast, and S. Reich, 2019: Particle filters for high-dimensional geoscience applications: A review. *Quart. J. Roy. Meteor. Soc.*, **145**, 2335–2365, https://doi.org/10.1002/qj.3551.

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

1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

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

Zhang, S., and Z. Pu, 2020: Evaluation of the four-dimensional ensemble-variational hybrid data assimilation with self-consistent regional background error covariance for improved hurricane intensity forecasts. *Atmosphere*, **11**, 1007, https://doi.org/10.3390/atmos11091007.

Zhang, X., S. G. Gopalakrishnan, S. Trahan, T. S. Quirino, Q. Liu, Z. Zhang, G. Alaka, and V. Tallapragada, 2016: Representing multiple scales in the Hurricane Weather Research and Forecasting modeling system: Design of multiple sets of movable multilevel nesting and the basin-scale HWRF forecast application. *Wea. Forecasting*, **31**, 2019–2034, https://doi.org/10.1175/WAF-D-16-0087.1.

Zhang, Z., W. Wang, L. Zhu, B. Liu, A. Mehra, and V. Tallapragada, 2018: Performance of the 2017 Real-Time HWRF-Based Ensemble and Combined HWRF/HMON/COAMPS-TC Multimodel Ensemble. *25th Conf. on Probability and Statistics*, Austin, TX, Amer. Meteor. Soc., 2.5, https://ams.confex.com/ams/98Annual/webprogram/Paper332300.html.

Zhu, K. F., M. Xue, Y. J. Pan, M. Hu, S. G. Benjamin, S. S. Weygandt, and H. D. Lin, 2019: The impact of satellite radiance data assimilation within a frequently updated regional forecast system using a GSI-based ensemble Kalman filter. *Adv. Atmos. Sci.*, **36**, 1308–1326, https://doi.org/10.1007/s00376-019-9011-3.

Zhu, Y., J. Derber, A. Collard, D. Dee, R. Treadon, G. Gayno, and J. A. Jung, 2014: Enhanced radiance bias correction in the National Centers for Environmental Prediction's Gridpoint Statistical Interpolation data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **140**, 1479–1492, https://doi.org/10.1002/qj.2233.