Implications of Self-Contained Radiance Bias Correction for Data Assimilation within the Hurricane Analysis and Forecasting System (HAFS)

JOSEPH KNISELY[a] AND JONATHAN POTERJOY[a]

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

(Manuscript received 13 February 2023, in final form 3 July 2023, accepted 5 July 2023)

ABSTRACT: The Hurricane Analysis and Forecasting System (HAFS) is the next-generation, FV3-based tropical cyclone (TC) forecasting system. Unlike operational implementations of NOAA's Hurricane Weather Research and Forecast (HWRF) modeling system, current data assimilation (DA) capabilities in HAFS permit the uninterrupted basin-wide assimilation of measurements. This feature of HAFS opens a variety of new research directions for TC prediction, including new strategies for DA algorithm development and self-contained probabilistic forecasting. The present research focuses more narrowly on new opportunities HAFS brings for optimizing the use of satellite measurements for TC prediction. While satellite radiometers provide a wealth of information for characterizing temperature, moisture, and wind in TC environments, the provided measurements are often biased and contain unknown cross-channel error correlations. For mature global modeling systems, these statistics are estimated from information gathered during DA, namely, innovations collected over large spatial and temporal training periods. The estimated statistics, however, are imperfect owing to unknown error sources such as model process error, which are difficult to separate from observation error. As such, bias and uncertainty specifications that rely on information from external models are suboptimal—as is the current strategy for HWRF. In this paper, it will be demonstrated that bias estimation for satellite radiance observations is particularly sensitive to common design choices, such as using a bias model trained from the Global Data Assimilation System instead of within the native modeling system. Implications of this finding for TC prediction are examined over a 6-week period from 2020, which included the development and intensification of nine tropical cyclones.

SIGNIFICANCE STATEMENT: Tropical cyclone–focused numerical weather prediction is difficult due to complex nonlinear physical processes and a lack of in situ observations over open ocean. Prediction systems rely heavily on satellite radiance measurements, which have high spatial–temporal resolution over the entire domain but require bias correction. Estimation of observation bias requires long training periods and large spatial domain coverage, which is typically not permitted outside of global models. However, bias specification is strongly model dependent, as bias correction methods cannot easily separate model and observation bias. In this study, we perform satellite radiance bias specification internally for an experimental version of the NOAA Hurricane Analysis and Forecast System and demonstrate major implications for tropical cyclone prediction.

KEYWORDS: Tropical cyclones; Radiances; Satellite observations; Bias; Numerical weather prediction/forecasting; Data assimilation

1. Introduction

The accurate prediction of tropical cyclone (TC) track and intensity has always presented a unique problem. The extreme weather event, which lies somewhere in between the traditionally established spatial–temporal classifications of synoptic scale and mesoscale, develops predominately over the observation-sparse ocean and produces storm conditions too intense for common surface-level observation systems to properly measure. For modern numerical weather prediction (NWP) systems, these issues remain significant obstacles (Emanuel and Zhang 2016). To depict the TC inner-core, limited-area hurricane-focused NWP systems, such as the current operational Hurricane Weather Research and Forecasting (HWRF) Model, require a high-resolution grid and ample inner-core observations, such as those collected during aircraft reconnaissance missions (Zhang et al. 2009, 2013; Aksoy et al. 2012; Sippel et al. 2022). However, given the size of the storm and its interaction with broad synoptic-scale features, these limited-area models (LAMs) must also be able to accurately characterize synoptic-scale flow over a large domain, particularly for the purpose of predicting changes in TC structure as well as track and intensification (Carr and Elsberry 1997; Hanley et al. 2001; Zeng et al. 2007; Hendricks et al. 2010; Poterjoy and Zhang 2014, 2016; Jones et al. 2018).

Given the sparsity of in situ measurements over the open ocean, TC-focused NWP systems rely heavily on satellite radiance measurements, which have a high spatial and temporal resolution over the entire model domain (Prasad et al. 1998). These observations primarily provide information on atmospheric brightness temperature (BT), from which atmospheric properties such as temperature, moisture, and wind can be determined (Deng et al. 2009). Studies have shown improvements in TC prediction from assimilating satellite radiance both from infrared and microwave channels (McNally et al. 2000; Zapotocny et al. 2008; Li and Liu 2009; Liu et al. 2012; Schwartz et al. 2012; Xu et al. 2013; Zou et al. 2013). In this study, we concentrate on clear-sky radiances, meaning observations

Corresponding author: Joseph Knisely, jknisely@umd.edu

DOI: 10.1175/WAF-D-23-0027.1

over cloudless or thin cloud regions, versus cloudy- or all-sky radiance measurements, which describe observations over thick clouds or precipitating clouds. Although all-sky radiance observations represent the most comprehensive description of atmospheric state available, evaluating BT over cloudy regions is complicated by nonlinearity in the observation operator (Geer and Bauer 2011; Zhu et al. 2016; Minamide and Zhang 2017; Poterjoy et al. 2017). Furthermore, complex cloud microphysical processes are difficult to parameterize and modern radiative transfer models are imperfect, which both result in degraded model representations of cloudy BTs. For these reasons, cloudy-sky radiances are often removed or used sparingly for DA purposes (Reale et al. 2018). On the other hand, clear-sky radiances are much more easily evaluated and provide valuable temperature and moisture field data across large areas. For these reasons, the use of clear-air radiance measurements for numerical weather prediction has matured more rapidly than cloudy-sky measurements at operational modeling centers (McNally et al. 2007; Collard and McNally 2009).

Despite the known benefits, clear-sky radiance observations have been shown to be systematically biased, and thus require a bias correction (BC) scheme. Bias can arise from several different sources, namely, from the radiative transfer model, from the calibration of the satellite instrument, as well as errors in the background model state (Derber and Wu 1998; Zhu et al. 2014; Liu et al. 2018; Reale et al. 2018). With a method described in detail in the following section, the Hurricane Analysis and Forecast System (HAFS) currently uses the Gridpoint Statistical Interpolation (GSI) variational DA scheme, which performs bias correction twice during the analysis, once for each outer loop (Hu et al. 2016). The bias correction value is added to the observation space prior and consists of a sum of weighted bias predictors. The predictors are a function of atmospheric state, while the corresponding weights are determined from analysis innovation statistics during the minimization of the variational cost function (Benáček and Mile 2019). These statistics are collected continuously, meaning the DA system must be fully cycled, and cover a sufficiently large domain. For this reason, it is common for this technique to be employed in global systems but it is rarely used for LAMs given the restricted domain size as well as other common restrictions, such as the adoption of partially cycled DA (Lin et al. 2017; Poterjoy et al. 2021).

Instead, LAMs import predictor weight values from externally trained models. For example, HWRF is a partially cycled system that regularly adopts the Global Data Assimilation System (GDAS) analysis for the parent domain (Zou et al. 2013; Mehra et al. 2018), thus requires predictor weights from GDAS. This strategy is not ideal, given that the bias correction method does not differentiate between various sources of bias and thus would not be able to separate bias imposed by the externally trained model (Rizzi and Matricardi 1998; Poterjoy et al. 2021).

The HAFS is NOAA's next-generation FV3-based TC forecasting system. While early operational implementations of HAFS will strongly resemble the DA methodology used for HWRF, research versions of HAFS can operate within a flexible DA framework that allows users to avoid many of the heuristic design choices that are common for operational models. For example, HAFS has the ability to perform fully cycled DA across its entire domain, allowing for the continuous updating of BC predictor weights during each analysis cycle. This capability opens the door not just to online satellite radiance bias correction, but also many other experimental techniques that require a fully cycling system (Poterjoy et al. 2021). For this paper, we will focus more narrowly on the effects of the online BC technique, and compare it to model results wherein BC parameters are adopted from GDAS.

In section 2, we detail the radiance bias correction methodology and include relevant equations. Section 3 describes the data assimilation framework, model physics, and relevant observation details. Section 4 presents results in the form of domain- and time-averaged error verification, model field comparisons, as well as a case study of a notable TC that developed over the model time period. Section 5 gives a summary and the conclusions.

2. Radiance bias correction

To properly use satellite radiance BT measurements, we must correct biases between the observed radiances and first-guess model values. The one-step "enhanced" bias correction scheme is performed within GSI, and is described in detail by Zhu et al. (2014). For this study, we provide an abbreviated description of how this method operates within GSI.

The observation operator $\tilde{h}$ is described by

$$\tilde{h}(\mathbf{x}, \boldsymbol{\beta}) = h(\mathbf{x}) + b(\mathbf{x}, \boldsymbol{\beta}), \tag{1}$$

where $\mathbf{x}$ is the model state, $\boldsymbol{\beta}$ represents predictor coefficients, and $h(\mathbf{x})$ represents the radiative transfer model. HAFS incorporates the Joint Center for Satellite Data Assimilation (JCSDA) Community Radiative Transfer Model (CRTM) as a forward radiative transfer model to compute model-derived BT for various instruments and channels (Han et al. 2006).

The function $b(\mathbf{x}, \boldsymbol{\beta})$ is represented by a linear combination of N_p predictor variables $p_k(\mathbf{x})$, where $k = 0, 1, 2, \ldots, N_p$, with corresponding predictor coefficients $\boldsymbol{\beta}_k$. Thus, we have

$$b(\mathbf{x}, \boldsymbol{\beta}) = \boldsymbol{\beta}_0 + \sum_{k=1}^{N_p} \boldsymbol{\beta}_k p_k(\mathbf{x}). \tag{2}$$

Choices of predictors include four polynomials for scan angle; global offset; emissivity predictor; cloud liquid water (CLW) predictor; temperature lapse rate predictor; and square of the temperature lapse rate predictor. The emissivity predictor tends to be quite large over land for near-surface channels and was formulated to account for differences in emissivity over land versus sea. On the other hand, the CLW predictor is applied to microwave radiances over oceans only. This predictor tends to be small and only applies to clear-sky radiance measurements. Although the quality control procedure is intended to screen data affected by thick clouds and precipitation, some cloud contamination is missed, and this predictor is included to fix such biases. Finally, the lapse rate and lapse rate squared terms connect changes in transmittance between vertical layers with changes in temperature, and ultimately BT bias (Zhu et al. 2014). These coefficients are estimated during the minimization of the cost function:

$$J(\mathbf{x}, \boldsymbol{\beta}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^{\mathrm{T}}\mathbf{B}_{\mathbf{x}}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_b)^{\mathrm{T}}\mathbf{B}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_b) + \frac{1}{2}[\mathbf{y} - \tilde{h}(\mathbf{x}, \boldsymbol{\beta})]^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{y} - \tilde{h}(\mathbf{x}, \boldsymbol{\beta})], \quad (3)$$

where $\mathbf{y}$ represents observations, $\mathbf{x}_b$ and $\boldsymbol{\beta}_b$ are first-guess estimates for $\mathbf{x}$ and $\boldsymbol{\beta}$, respectively, and $\mathbf{B}_{\mathbf{x}}$ is the background-error covariance for $\mathbf{x}$. The matrix $\mathbf{B}_{\boldsymbol{\beta}}$ is the block-diagonal background-error covariance for predictor coefficients, given by

$$\mathbf{B}_{\boldsymbol{\beta}} = \begin{bmatrix} \mathbf{B}_{\beta}^{(1)} & & 0 \\ & \ddots & \\ 0 & & \mathbf{B}_{\beta}^{(J)} \end{bmatrix}, \quad (4)$$

where $\mathbf{B}_{\beta}^{(j)}$ is a diagonal matrix of error standard deviations for channel j ($j = 1, 2, \ldots, N_c$), specified as 10.0 for all coefficients of all channels (Zhu et al. 2014).

To minimize the cost function, we can combine the control vectors $\mathbf{x}$ and $\boldsymbol{\beta}$ into a single control vector $\mathbf{z}$, such that

$$\mathbf{z}^{\mathrm{T}} = [\mathbf{x}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}}], \quad (5)$$

with background of $\mathbf{z}^{\mathrm{T}} = [\mathbf{x}_b^{\mathrm{T}}, \boldsymbol{\beta}_b^{\mathrm{T}}]$. Assuming no cross covariance between $\mathbf{x}$ and $\boldsymbol{\beta}$, we write the error covariance matrix $\mathbf{B}$ as

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{\mathbf{x}} & 0 \\ 0 & \mathbf{B}_{\boldsymbol{\beta}} \end{bmatrix}. \quad (6)$$

The cost function J is then given by

$$J(\mathbf{z}) = \frac{1}{2}(\mathbf{z} - \mathbf{z}_b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{z} - \mathbf{z}_b) + \frac{1}{2}[\mathbf{y} - \tilde{h}(\mathbf{z})]^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{y} - \tilde{h}(\mathbf{z})]. \quad (7)$$

In addition, several channels are designated as "passive" channels, meaning these channels are read into GSI but do not impact the analysis. These channels may be monitored for research or quality control purposes, and are bias corrected in a manner similar to assimilated observations.

We differentiate our experiments by the technique through which we initialize the first-guess predictor coefficients at each cycle during data assimilation. Our control experiment retrieves GDAS-generated bias correction data for each cycle to serve as first-guess coefficients, while the "online" experiment uses the above variational (VarBC) method to generate the predictor coefficients during each analysis step. The coefficients are updated during each outer-loop calculation, thus twice for each analysis (Hu et al. 2016).

3. Model configuration

a. Forecast and analysis

HAFS is the next-generation operational analysis and prediction system for hurricane forecasting at NOAA. The current options for HAFS feature a uniform global model with a high-resolution nest (HAFS-global-nest) and a stand-alone high-resolution regional model (HAFS-SAR). For our research, we are utilizing the HAFS-SAR domain configuration, a version of which will become operational in summer 2023. Henceforth we will refer to this configuration of the model simply as HAFS. Like the operational HAFS, we use FV3-SAR and ensemble-based variational data assimilation with the ensemble pulled from GDAS for background uncertainty specification. Unlike the operational HAFS, our domain remains in a fixed location over multiple storms and does not make use of high-resolution nests or vortex relocation and uses 3D-Ensemble-Var (3DEnVar) in place of 4D-Ensemble-Var, which is adopted operationally. We choose to omit vortex relocation not only because we believe it will complicate the interpretation of our results, but also because we do not see it being used in the long term for TC data assimilation. Finally, we perform our own estimation of predictors for the bias correction model, as explained in more detail in section 3c (Dong et al. 2020; Hazelton et al. 2021; Gopalakishnan et al. 2021).

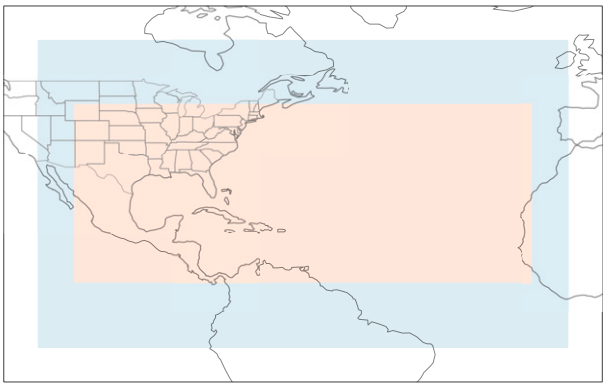


FIG. 1. In blue, the entire HAFS-SAR model domain. In orange, a smaller subdomain selected for error verification (see section 4).

The FV3 dynamical core for HAFS is based on a nonhydrostatic finite volume solver using a Lagrangian vertical coordinate (Hazelton et al. 2018). The model domain adopted for our study uses a single grid of 1440 × 1080 horizontal grid cells, with an approximate 6-km grid spacing, and 81 vertical levels in sigma-pressure hybrid coordinates. The lowest model level is about 25 m above the surface and the top isobaric level is 2 hPa. The domain is centered at 25°N, 60°W, with a 109.5° span for longitude and a 68.2° span for latitude, as seen in Fig. 1. For our experiments, boundary conditions and the first set of initial conditions are interpolated from the 2020 operational global FV3-based GFS (~13-km horizontal spacing) onto the 6-km HAFS domain. Boundary conditions are provided every 6 h from the same global GFS forecasts (Dong et al. 2020).

As previously mentioned, we use a GSI-based 3DEnVar data assimilation system. The data assimilation updates a single deterministic model solution every 6 h, using background error covariance estimated entirely from the 80-member 6-h GEFS forecasts used by the operational GDAS (Bannister 2017)—and no static error covariance. In addition, we employ the first guess at the appropriate time (FGAT) technique over

the 6-h DA window to account for differences between actual observation times and background times.

Consistent with Poterjoy et al. (2021), our configuration of HAFS does not make use of vortex relocation or finer-resolution nests, nor does it regularly recenter the model state onto GDAS solutions, which is a technique employed by operational LAMs. The partial-cycling strategy is often used to prevent model drift over time, which is often due to deficiencies in model physics, data assimilation, and boundary conditions. Nevertheless, we find our model to exhibit only a limited systematic change in background errors, which were determined from a careful inspection of observation-space statistics accumulated during the DA experiments.

b. Model physics

We use a model physics configuration that resembles the operational NCEP GFS version 16 (GFSv16), but with modifications that reflect boundary layer processes that are uniquely important for TCs. The physics suites employed include the following: the GFS Rapid Radiative Transfer Model for GCMs (RRTMG) shortwave/longwave radiation scheme (Mlawer et al. 1997); the GFS scale-aware simplified Arakawa–Schubert (SA-SAS) deep convection scheme (Han et al. 2017); the Noah-multiparameterization land surface model (Noah-MP LSM) (Niu et al. 2011); the GFS scale-aware TKE-based moist eddy-diffusion mass-flux (EDMF) PBL and free atmospheric turbulence scheme (Han and Bretherton 2019); the GFS surface layer scheme with HWRF exchange coefficients (Zheng et al. 2017); and the GFS orographic gravity wave drag scheme (Kim and Arakawa 1995).

The model configuration also features Community Mediator for Earth Prediction Systems (CMEPS)-based ocean coupling with the Hybrid Coordinate Ocean Model (HYCOM), which is described in detail in Chassignet et al. (2007, 2009). The HYCOM system features a 1/12° resolution North Atlantic basin domain with 41 vertical levels, initial conditions from NCEP's Real Time Ocean Forecast System (RTOFS), persistent oceanic lateral boundary conditions, and atmospheric forcing from GFSv16 for nonoverlapping areas.

c. Bias correction configurations

In this section, we outline a pair of experiments, hereby referred to as "control" or "GDAS BC" versus "online" or "HAFS BC," that examine the effectiveness of online VarBC for our LAM configuration of HAFS. The control experiment borrows bias correction coefficients from GDAS, importing them for each outer loop of the analysis. These GDAS-based coefficients are obtained directly from the real-time GDAS, which were updated via the same GSI-based VarBC method detailed in section 2 over an equivalent training period.

The online experiment initializes BC coefficients from GDAS values for the first analysis cycle, and then transitions to self-cycling these coefficients within HAFS, updating them during each outer loop and using the resulting output coefficients as the first guess for the subsequent cycle (Hu et al. 2016). In this way, the cycling of coefficients allows bias to accumulate during the data assimilation cycles, limiting potential system biases arising from the borrowing of GDAS coefficients. In other words, the configuration limits its potential influence from an outside model and allows for better identification and understanding of biases arising from the satellite instruments and measurement operator alone.

d. Observations and data assimilation schedule

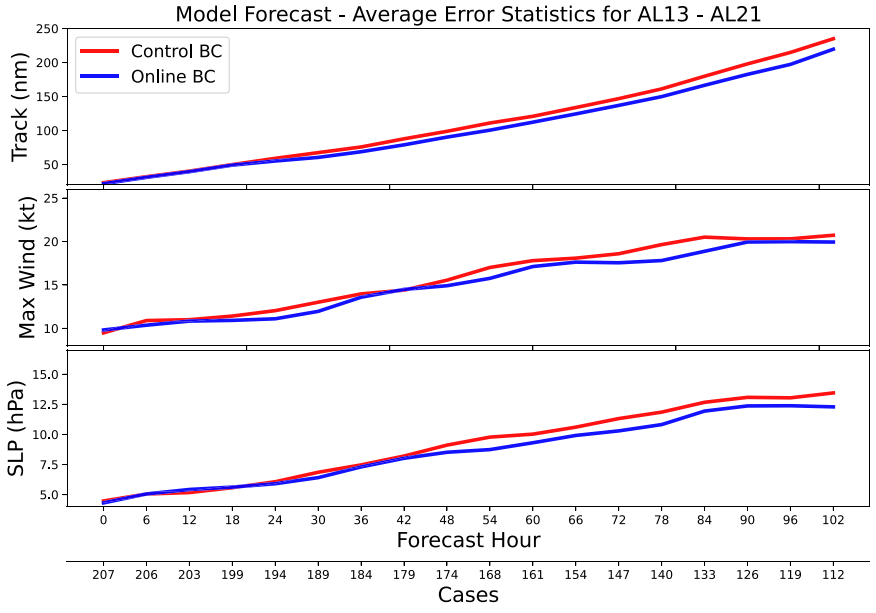
The experiments begin at 0000 UTC 18 August and end at 0000 UTC 22 September 2020, covering nine storms from TC Josephine to Hurricane Sally. In particular, this time span is chosen to include Hurricane Laura, a category-4 hurricane that was the strongest TC of the season by maximum sustained winds and among the most powerful hurricanes on record to make U.S. landfall (NOAA 2021). Our HAFS experiments assimilate measurements at 0000, 0600, 1200, and 1800 UTC each day, and generate medium-range (102-h) deterministic predictions on 0000 and 1200 UTC cycles. We also omit the first 7 days of data assimilation to allow HAFS to become independent from the GDAS initial conditions that are used to initialize experiments.

The suite of assimilated measurements includes all measurements assimilated by GFS except proprietary measurements, namely Aircraft Communications Addressing and Reporting System data and cloudy radiances. In addition, HAFS assimilates tail Doppler radar and TC inner core radar data as collected on aircraft reconnaissance flights, which are not assimilated by the GDAS. In addition, HAFS assimilates inner core data, including tail Doppler radar, dropsonde, and flight-level wind observations, which are not assimilated by the GDAS.

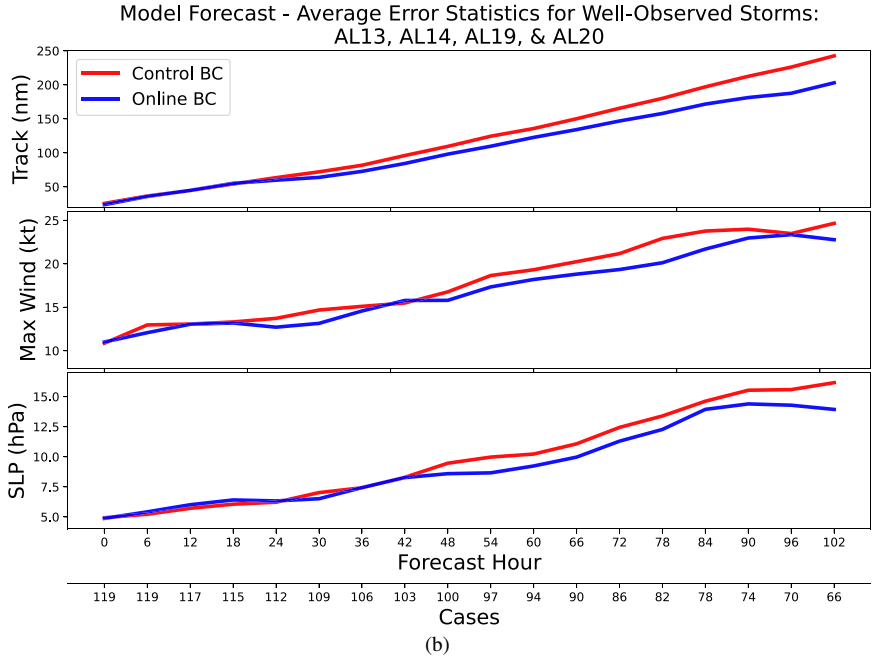
The Geophysical Fluid Dynamics Laboratory (GFDL) vortex tracker is used to generate the TC track information (Biswas et al. 2018). TC track, maximum 10-m surface wind, and minimum surface pressure from model forecasts are verified against "best track" data from the National Hurricane Center (Hazelton et al. 2021).

4. Results

The pair of experiments demonstrate substantial differences in domainwide error characteristics as well as forecast track and intensity errors, which we describe in this section. Moreover, we will show that many of the differences can be easily interpreted by analyzing the diverging evolution of predictor weights. More precisely, changes in weighted predictor fields over the domain amount to explainable differences in model variable fields, such as temperature and humidity, by changing the bias applied to the observation space background. For this purpose, we demonstrate how the online configuration affects weighted predictor fields and drives spatially coherent differences in thermodynamic variables between the experiments. From this analysis, we can determine how this configuration affects error characteristics and consequently improves TC forecasts. In general, we focus on variables that are directly related to design choices in our data assimilation system, choosing to omit variables that are more complex to attribute, such as storm surge and wave generation. Although these hazards are vitally important to consider



(a)



(b)

FIG. 2. Forecast error verification of control and online experiments. (a) Error for all TCs over the experimental period and (b) TCs without aircraft reconnaissance data are omitted. These errors are verified against NHC best track and intensity data.

when designing operational modeling systems, we limit the scope of this study to variables with simple interpretability.

a. Forecast error verification

Section 4b provides a verification of TC forecast error statistics, specifically: track, 10-m maximum winds, and minimum sea level pressure (SLP). These statistics are summarized for all storms in Fig. 2a, where forecast mean errors are calculated from a sample of storm forecasts over the 6-week experimental period and plotted as a function of forecast lead time. The number of cases for each time is also given on the secondary axis.

For each variable analyzed, error statistics up to 18–24 h are nearly identical across both experiments, after which differences begin to emerge. In regard to track performance, the online configuration provides consistently better results after

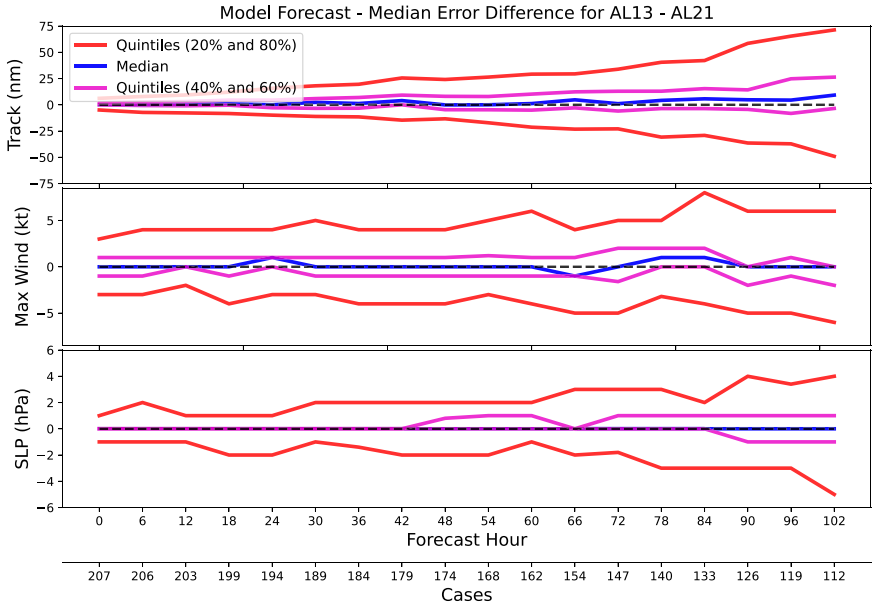


FIG. 3. Error verification of forecast absolute error differences between control and online experiments (online − control). The median difference is given in blue, while the inner and outer quintiles are given in purple and red, respectively. Errors are verified against NHC best track and intensity data.

18 h of lead time, while the intensity forecasts provide generally improved results over the same range.

Several previous studies have demonstrated that considerable forecast improvements can be obtained from the assimilation of flight level radar and dropsonde observations from reconnaissance flight missions (Zhang et al. 2009; Aksoy et al. 2012; Lu et al. 2017; Hartman et al. 2021; Sippel et al. 2022). Specifically, the assimilation of inner-core observations has been shown to significantly improve intensity forecasts even without the help of ad hoc vortex relocation techniques. For this reason, we further stratify the results to include only storms that had multiple reconnaissance flight missions requested by the National Hurricane Center, hereby referred to as "well-observed" storms (NHC; Fig. 2b). These storms are Hurricanes Laura, Marco, Sally, and Teddy. Since land-threatening storms are the most likely to receive NHC reconnaissance flight authorization, the accurate forecasting of these storms is particularly important for NWP centers.

Unsurprisingly, the forecast error statistics for well-observed storms are quite similar to the cumulative results for all storms, but a few key differences stand out. In regards to track forecasts, both experiments exhibit similar errors up to about 24 h, after which the online configuration exhibits markedly better performance. The intensity forecast results are mixed up until 42 h, after which the online experiment reports consistently lower errors. Note that the total intensity errors for these well-observed storms are generally higher than for the all-storms dataset; this is because these well-observed storms are, on average, more intense than the excluded storms, leading to proportionally larger errors.

It is likely that the similar performance and relatively low intensity errors of these experiments at early lead times is a reflection of the influence of reconnaissance observations. The divergence of these configurations and subsequent superiority of the online experiment is perhaps a sign that the HAFS BC configuration benefits significantly from the online clear-sky satellite bias correction when combined with valuable inner-core observations. Since the assimilated clear-sky radiances are unlikely to include measurements near the TC vortex due to extensive cloud cover, it instead seems likely the clear-sky radiance bias correction results in an improved characterization of the surrounding synoptic environment while these inner-core observations "fill in the gaps" near the storm center. Furthermore, if these inner-core observations can more precisely place the storm vortex location within the domain, and large-scale features are more accurate as a result of the online bias correction scheme, improved intensity forecasts will follow as a result of a more accurate environmental characterization (Aksoy et al. 2012; Lu et al. 2017; Hartman et al. 2021).

As a final verification of forecast performance and variability, we summarize model forecast differences between the control and online experiment with Fig. 3. We examine the same error statistics as above but take the difference in absolute errors between experiments at each lead time and plot the median as well as inner and outer quintiles. This test essentially shows us the frequency in which one experiment outperforms the other at each lead time. In regards to track performance, the online experiment generally provides improved error statistics more frequently than the control after 24-h lead time, as indicated by the median and inner quintiles. Conversely, the median and inner quintiles very consistently hover around zero for max wind and SLP. Comparing these results to Fig. 2a would suggest that while the frequency of

improved error statistics is relatively the same across experiments, the errors themselves tend to be smaller for the online BC experiment. We note that we performed a similar analysis for well-observed storms only (not shown) and the results were marginally improved in expected ways; i.e., all quintiles are slightly elevated in a manner that would be expected based on Fig. 2b.

b. Domain-averaged error verification

Next, we compare domainwide variables, such as temperature, humidity, and wind speeds. We verify the data versus ECMWF Reanalysis v5 (ERA5) data, which uses 4DVar data assimilation and comprehensive observation datasets including all-sky radiances; it is also relatively high-resolution compared to other global reanalysis datasets (Hersbach et al. 2020).

Figure 4 displays vertical profiles of horizontally averaged bias and root-mean-square error (RMSE) values for pertinent variables, i.e., temperature, specific humidity, and u and v winds, for both control and online experiments at 0000 UTC analysis time. Note that the investigated domain is slightly restricted compared to the model analysis domain (see Fig. 1) in order to minimize the effects of boundary conditions. In addition, the vertical profiles are capped at 200 mb (1 mb = 1 hPa), above which the effects of the model-top "sponge layer" produce unpredictable and difficult-to-characterize results.

The most prominent differences between control and online experiment error profiles can be found in temperature and humidity profiles, which is unsurprising given that these variables share a strong error dependence with BT. Both temperature bias profiles in Fig. 4a reflect a domainwide negative temperature bias present in this configuration of HAFS, particularly in the lower troposphere; however, there are marked differences in temperature bias and RMSE profiles in the mid and upper troposphere. The online experiment exhibits a slightly positive temperature bias from 950 to 750 mb, a more negative bias from 750 to 250 mb, and a significantly more positive bias above 250 mb. The specific humidity bias profile is largely similar between the two experiments; however, the control configuration exhibits slightly more negative biases at the surface and through the midtroposphere, from approximately 650 to 300 mb. Results are mixed for u and v winds, with only very slight differences in RMSE profiles as shown in Fig. 4b. We note that since these statistics are domain-averaged, regional biases of the opposing sign can cancel each other out, a fact that is explored further in the next section.

We also examine the temperature, specific humidity, u and v wind RMSE profiles across forecast lead times, where the online configuration tends to exhibit more substantial improvements in error profiles. Figure 5 shows time series of horizontal- and vertical column averages of RMSE for these variables taken from 0 to 102 h. Like the calculations for Fig. 4, the verification domain is restricted to a subset of points within the total HAFS domain (see Fig. 1). Each variable is column averaged from 900 to 250 mb so as to minimize the influence of surface and model top boundary conditions, except for specific humidity which was averaged up to 500 mb given the small mixing ratios in the upper troposphere. Temperature errors show notable improvements over early and late lead times, albeit with little difference between control and online from 6- to 84-h lead time. Conversely, across specific humidity and u and v winds, the difference between control and online is small at early lead times but consistently demonstrates improved error statistics at longer lead times, beyond 6–18 h. These improvements mostly trail off at lead times beyond 84 h. The improved domainwide error statistics for the online configuration at longer lead times shown in Fig. 5 mirror the improvements we see in storm track and intensity errors at longer lead times, as referenced in Fig. 2.

c. Temperature and humidity fields

In this section, we discuss major differences in the temperature and humidity fields between the online and control experiment. We highlight temperature and humidity in our discussion, as profiles of these variables are the most directly related to radiance. Furthermore, we emphasize discrepancies in experiments at different height layers and horizontally across the domain, particularly overland versus overocean. In doing so, we illustrate how the online bias correction strategy directly influences our state estimates, which in turn explain different forecast results between experiments.

In Fig. 6, we examine vertical profiles of domain- and time-averaged differences in temperature and specific humidity between online and control experiments. We also differentiate between total, overland, and oversea differences. Given that the domain is mostly ocean, it is unsurprising that total difference profiles for both temperature and humidity are quite close to their respective profiles over the ocean. However, stark differences exist over land versus ocean, which can be clearly distinguished in Figs. 7 and 8.

Relative to the control, we find that online configuration produces cooler surface and lower-tropospheric temperature fields in total and oversea locations, with a local minimum of around 850 mb, before becoming comparatively warmer into the midtroposphere, with local maxima around 650 and 350 mb. The warmest areas exist over the central Atlantic, particularly where most storms track during our experiments, as shown in Fig. 7. Conversely, this configuration produces slightly warmer temperature fields over land throughout the lower troposphere, with local maxima of around 900 mb. Throughout the midtroposphere, differences over land are slightly positive or near zero, although throughout western and northern North America these differences are strongly negative. Above 300 mb, all three profiles take a sharply negative turn.

Specific humidity exhibits less coherent layerwise patterns, as shown in Fig. 8, but examining its domain-averaged vertical profile in Fig. 6 reveals some notable differences overland versus overocean. For example, while the online experiment is uniformly drier at the surface, the total and oversea specific humidity difference is positive from 850 to 300 mb, with local maxima at 825 and 500 mb. Meanwhile, the overland profile remains dry throughout the lower troposphere, with a local minima at 775 mb, before switching signs at 650 mb and remaining positive throughout the rest of the midtroposphere, and reaching a local maximum at 525 mb. Above 300 mb, all three profiles approach zero, reflecting the low humidity in the upper troposphere.

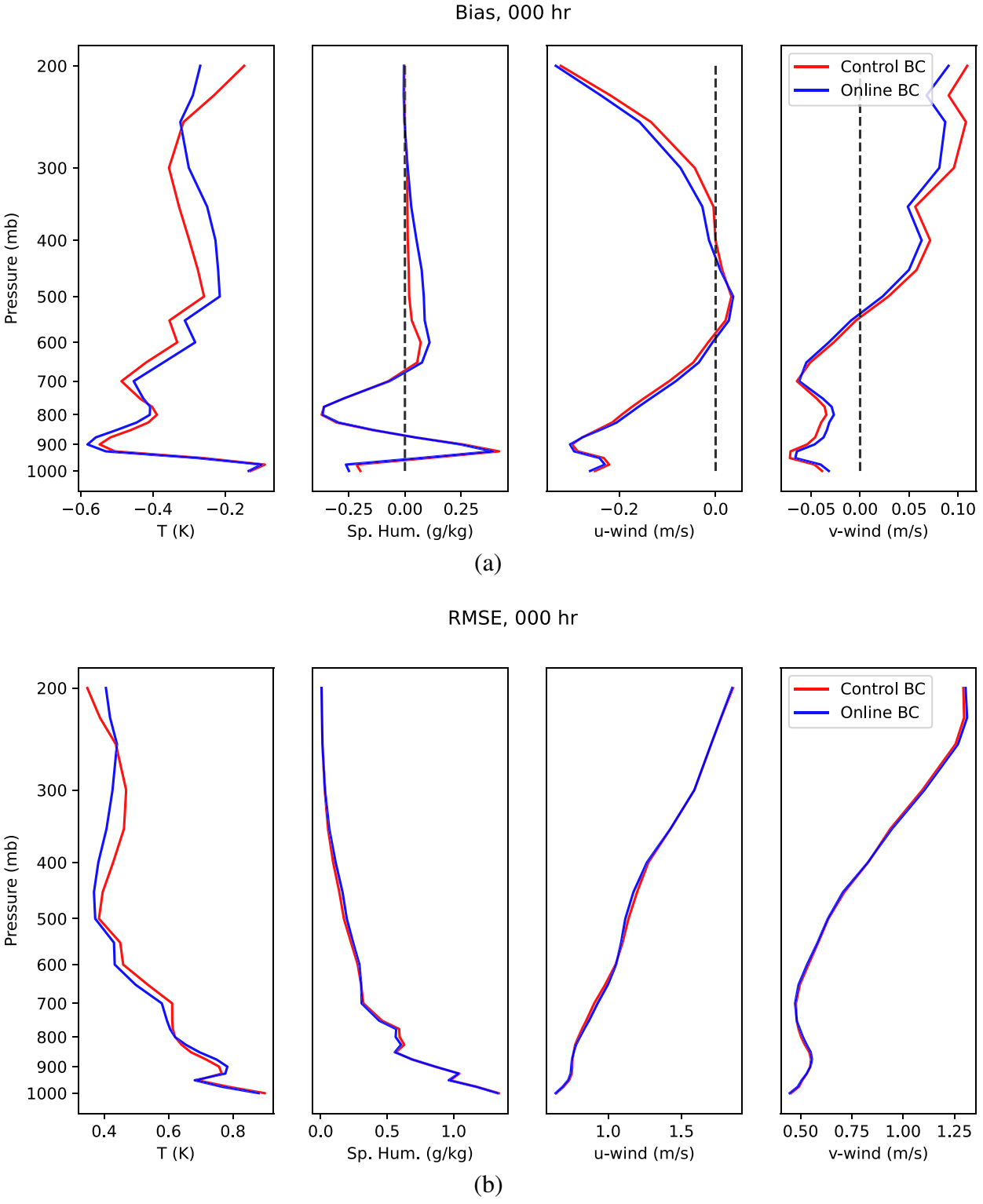


FIG. 4. (a) Bias and (b) RMSE verification of temperature, specific humidity, and u and v winds with ERA5 data, respectively. Data are time and domain averaged and displayed as a vertical profile.

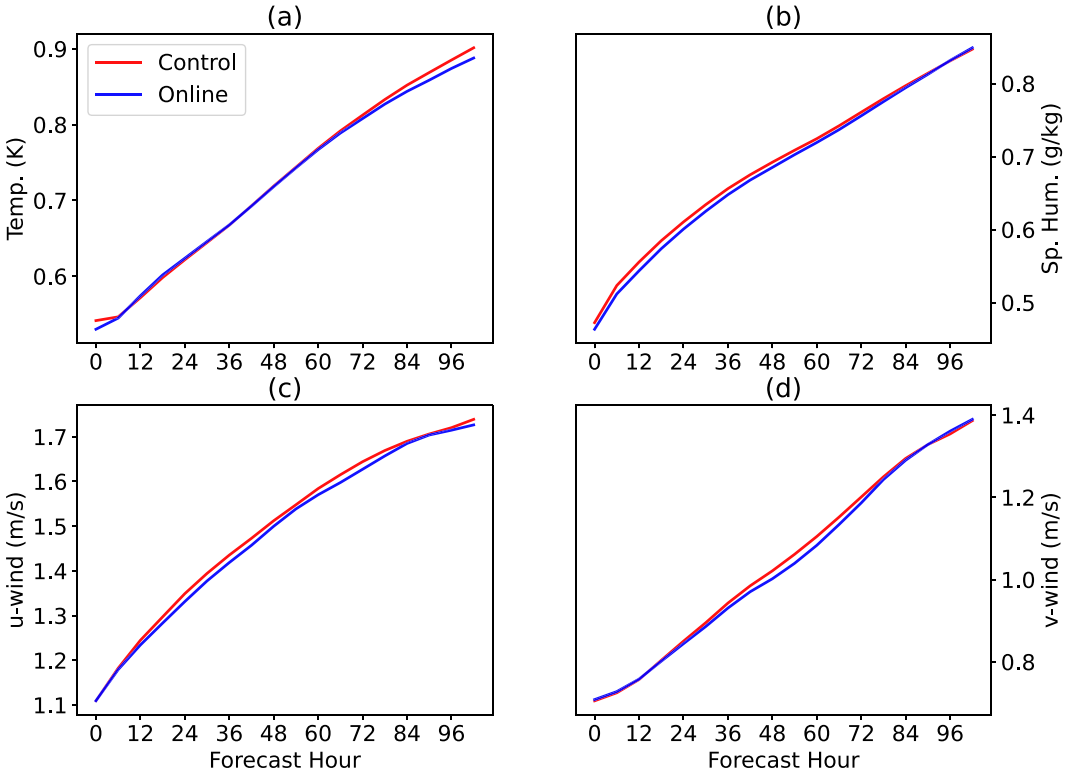


FIG. 5. (a) Temperature, (b) specific humidity, (c) u wind, and (d) v wind RMSE verification with ERA5 data. Data are domain and column averaged, and taken at 6-h lead-time increments up to 102 h. Columns span vertically from 900 to 250 mb for temperature and wind errors, and from 900 to 500 mb for specific humidity.

As we will demonstrate in section 4d, these differences over land and sea may arise from a number of factors, the most obvious of which are different radiative characteristics overland versus water. In addition, high-quality in situ observing systems are far more prevalent over land, which may lead to different bias characteristics. Alternatively, the analysis overland is less dependent on satellite radiances and thus fewer differences exist between the two configurations, explaining the overall less dramatic difference in temperature overland.

d. *Predictor evolution and verification*

To understand why the online configuration results in improved forecast errors and notable differences in model fields, it is first necessary to evaluate how the online configuration affects the evolution of bias correction predictor coefficients as well as the weighted predictors themselves. It would be difficult to evaluate the individual effects of each and every instrument- and channel-specific predictor field, and so instead we narrow our analysis to specific channels with qualities that are beneficial for our investigation.

With this in mind, we have chosen to focus on channel 4 of the Advanced Microwave Sounding Unit-A (AMSU-A) *NOAA-15* (*N15*) satellite instrument. This channel measures radiance over a frequency of 52.8 GHz, has a weighting function that peaks near the surface, and has a relatively wide spread of weighted predictor values, making it convenient for our analysis (Goldberg et al. 2001; Deng et al. 2009; Liu et al. 2012; Zhu et al. 2016). Furthermore, the AMSU-A instruments onboard NOAA polar-orbiting satellites are often indicated by ensemble and adjoint sensitivity analysis to yield large impacts on synoptic-scale forecasts (Ota et al. 2013). Moreover, the instrument has a high volume of observations ingested after quality control and relatively large differences in channel-specific weighted predictor values between control and online experiments.

Figures 9a and 9b show the evolution of channel-4 AMSU-A *N15* coefficients and weighted predictors over the 5 weeks of cycled DA analyzed for this study. Note that the magnitude of the unitless coefficients in Fig. 9a vary widely and that these coefficients are multiplied by bias predictors whose magnitude varies considerably themselves. Thus, the values of the coefficients are not as significant as the weighted predictors themselves in Fig. 9b which have units of brightness temperature and therefore are comparable.

Instead, Fig. 9a is notable because it shows the rate at which the online coefficients depart from values estimated by GDAS over the course of our experiments. While some coefficients seem to reach a quasi-steady estimate in 1–3 weeks, others, such as the global offset coefficient, appear to diverge up until the end of the experiment. Figure 9a also suggests that the HAFS online coefficients are considerably more noisy than their GDAS counterparts. Given that these HAFS

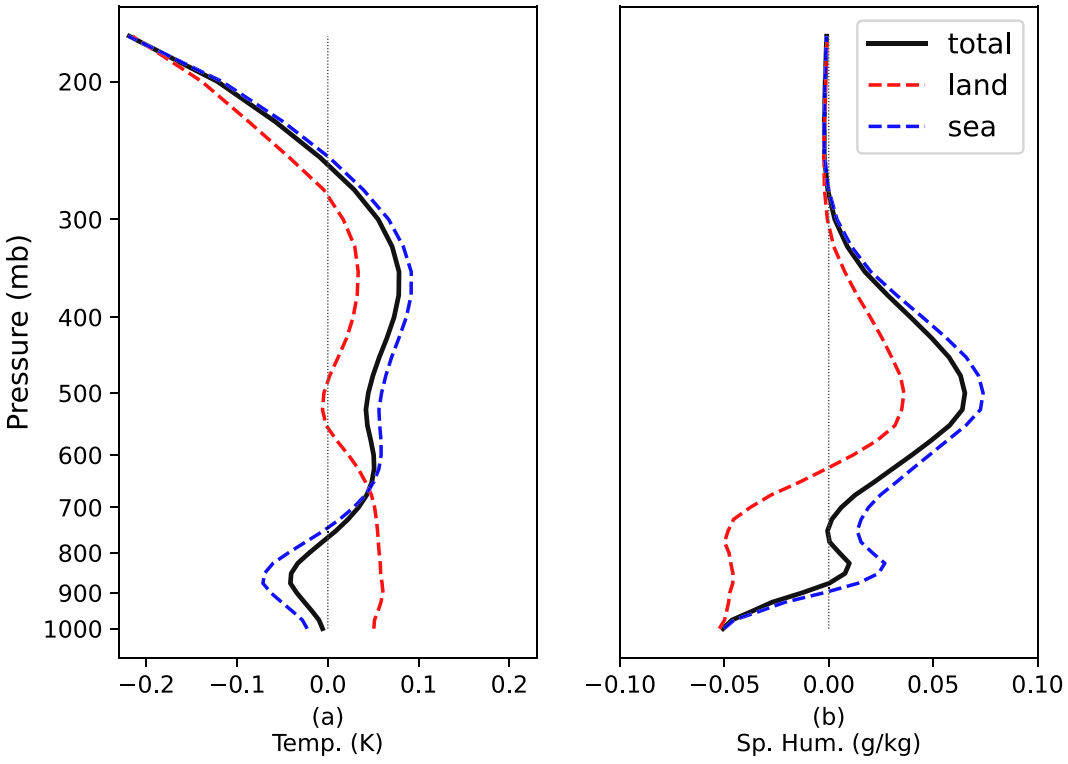


FIG. 6. Vertical profiles of domain- and time-averaged (a) temperature and (b) specific humidity differences between the online and control experiment (online − control). Differences overland, oversea, and total differences are all considered.

coefficients are derived over a limited-area domain and are influenced by boundary conditions, this behavior is not surprising. Also note that since the CLW predictor is not used by GDAS, the value of its predictor is exactly zero for the entirety of the control experiment.

Figure 9b displays the evolution of the weighted predictors themselves, that is, the product of domainwide predictor variables and the coefficient weights as featured in Fig. 9a. These values have units of BT (K), and combined constitute the total bias applied to the observation-space background as described in section 2. Note the strong diurnal pattern in weighted predictor value, and the relative similarity in values between experiments for the polynomial predictors. Instead, differences in the global offset, CLW, lapse rate, lapse rate squared, and emissivity predictors are predominated. Furthermore, much like in Fig. 9a, it can be seen that certain predictors begin to diverge 1–2 weeks into the experimental period, with some continuing to diverge even at the end of the 5-week period.

Although we have chosen to focus on channel 4 of the AMSU-A *N15* satellite instrument, there are many more significant channels across multiple instruments that fulfill similar criteria. Figure 10 details a number of these channels. Each bar plot displays a different instrument's time- and domain-average weighted predictor value for all nine predictors for the control and online experiment, as well as their difference.

Notice that for low-numbered channels, which tend to peak at or near the surface, the emissivity-weighted predictors tend to produce large disparities between the control and online configuration. Alternatively, for high numbered channels that peak in the upper troposphere and stratosphere, the predictor of interest tends to be the lapse rate and lapse rate squared predictors. It is also noteworthy that the distributions of weighted predictor values across instruments are very similar, e.g., channel 11 has similar predictor values and differences across multiple instruments. This suggests that these differences in bias are dominated by model bias. We speculate that model physics may be the major contributor to such biases, but ultimately emphasize this as an example of the importance of eliminating outside model bias, such as model bias from GDAS.

e. *Weighted predictor fields*

As an example of how the online VarBC method may affect model fields, it is useful to examine the difference in the weighted predictor fields between experiments for one specific channel and instrument, and then compare these differences to differences in model field variables. For the same reasons mentioned earlier, we examine the difference in weighted predictor fields between the online and control experiment for channel 4 of AMSU-A on the *N15* satellite.

Figure 11 shows the time-averaged online-control differences in weighted predictor field contributions for each predictor for

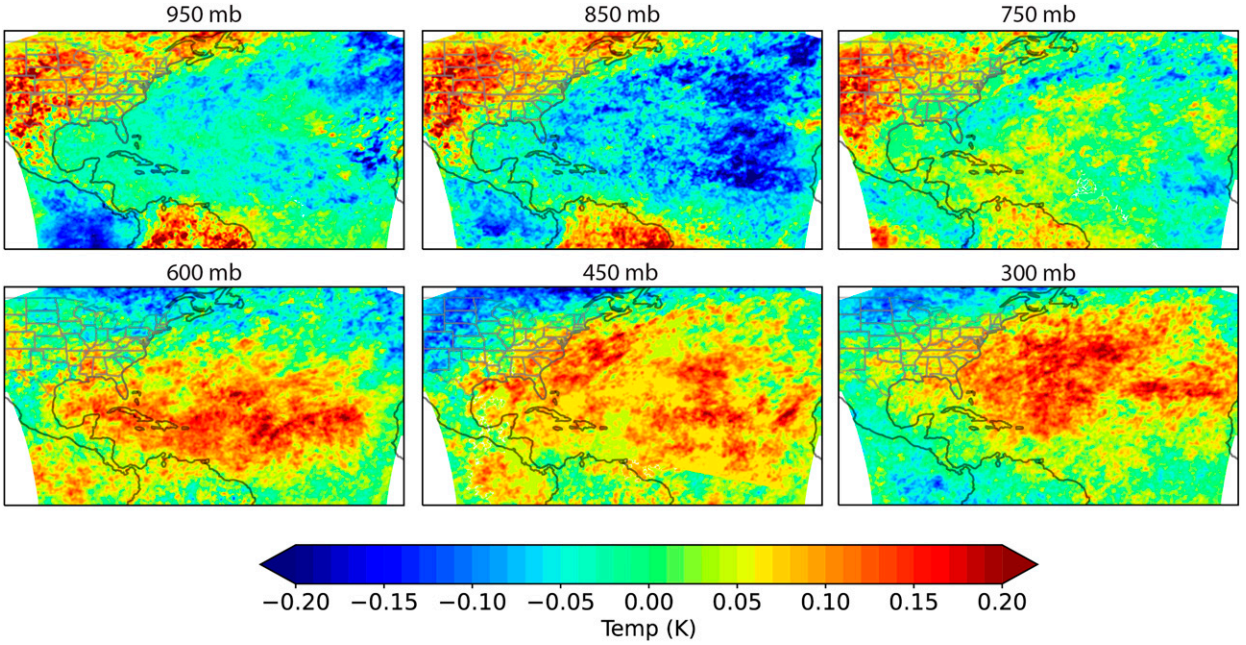


FIG. 7. Time-averaged temperature differences between the online and control experiment at 950, 850, 750, 600, 450, and 300 mb (online − control). Note that a Gaussian smoothing filter has been applied.

the AMSU-A channel 4. Although many of the weighted predictor fields show little coherence, a few display notable patterns. The global offset term tends to contribute to a uniform negative bias estimate across the entire domain, with a few exceptions that tend to congregate around coastlines, while the emissivity predictor is strongly positive across most of the North and South American landmass and mostly zero overocean. It is unsurprising that the emissivity predictor produces the starkest differences given that this channel peaks near the surface, and, as previously mentioned, this predictor tends to dominate for low-altitude peaking channels. The cloud liquid water predictor is conversely zero

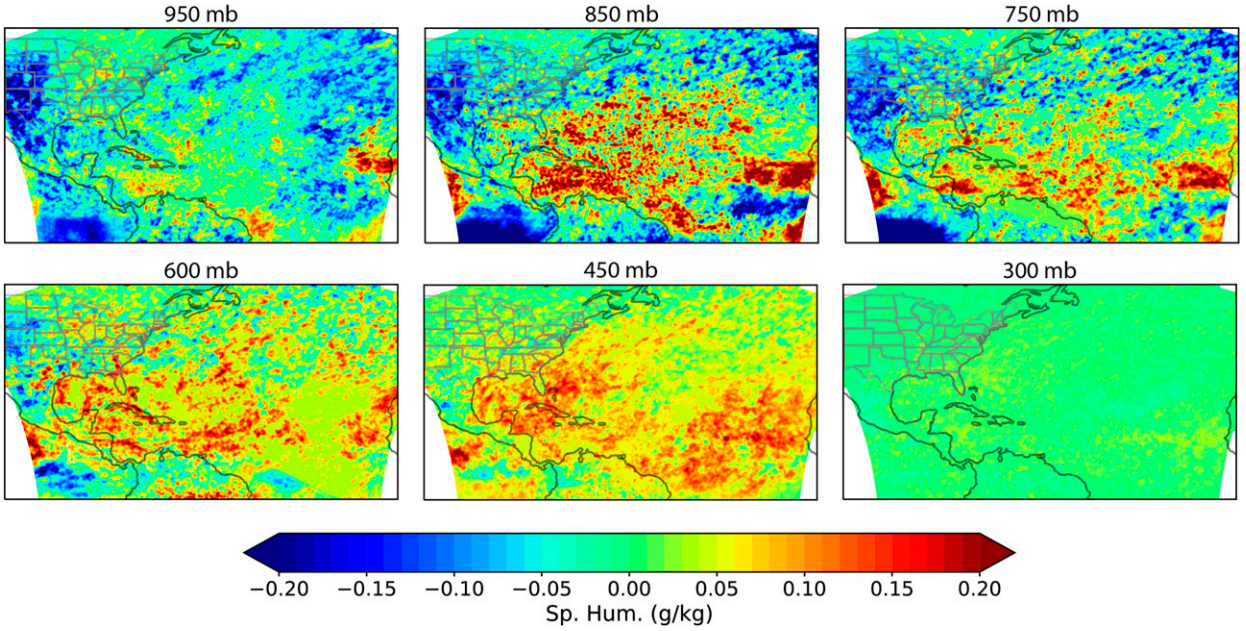


FIG. 8. Time-averaged specific humidity differences between the online and control experiment at 950, 850, 750, 600, 450, and 300 mb (online − control). Likewise, a Gaussian smoothing filter has been applied.

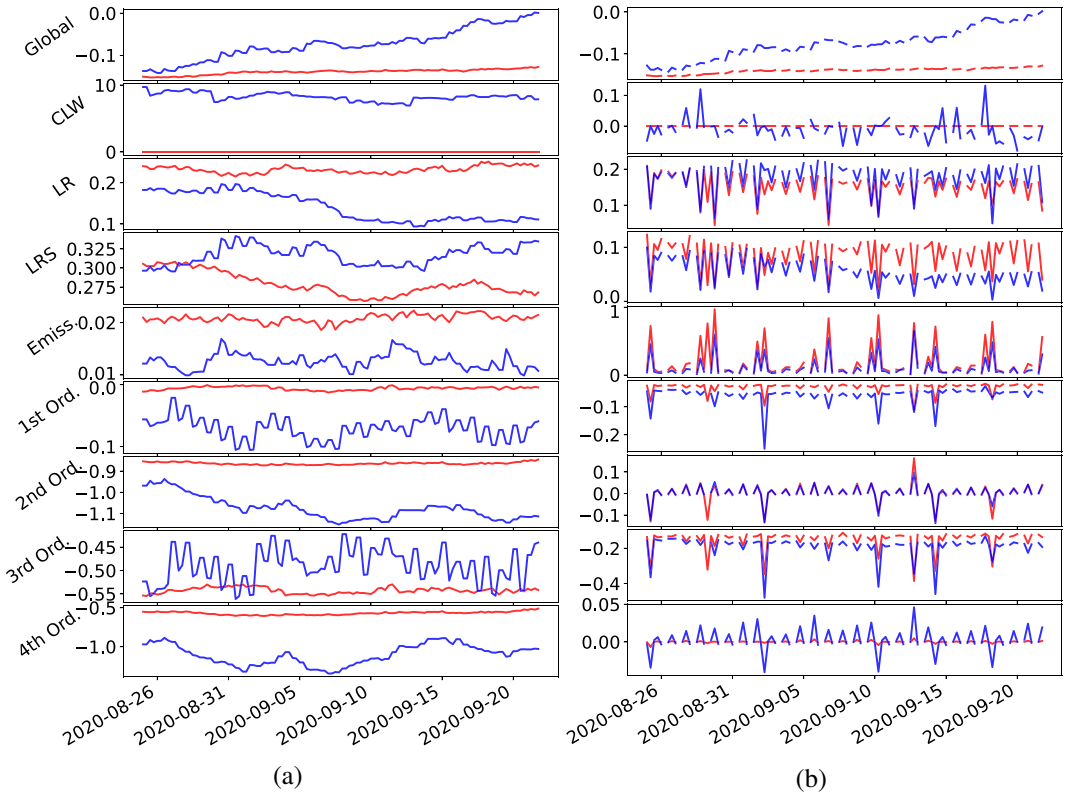


FIG. 9. (a) Bias predictor coefficient evolution over the course of the 5-week experiment period the GDAS BC (red) and HAFS BC (blue) for channel 4 of the AMSU-A *N15* satellite instrument. (b) Weighted bias predictor value evolution over the same period, in units of BT (K).

across most of the landmass, but exhibits varied patterns over water, having positive values over the Caribbean and mid-Atlantic basin versus negative values in the northern Atlantic region and in the eastern Pacific. The lapse rate and lapse rate squared predictors tend to produce a quasi-uniform pattern, albeit with the magnitude of the predictors increasing slightly toward the equator, while the polynomial predictors display less coherence across the domain.

In Fig. 12, we consider the influence of a single predictor by comparing time-averaged online − control temperature differences at 950 mb to weighted predictor field differences for the emissivity predictor. A correlation can be seen between the positive contribution of the emissivity predictor bias term and the warmer temperatures present overland in the online experiment's analysis. Since channel 4 peaks near the surface, it can be assumed the bias term has directly or indirectly affected the model field, but given the magnitude of the bias term it is likely this relationship falls under the former category. Of course, it is difficult to fully attribute these temperature field changes without fully considering the effects of each predictor over every channel and instrument, a task that is outside the scope of this paper. Instead, we focus on the macro effects of the self-cycling bias correction technique and its effects on TC track and intensity forecasting.

f. Case study

To give an example of how the online VarBC alters the track and intensity forecasts compared to the control experiment, we examine "both configuration" forecasts of Hurricane Teddy. Teddy, the 19th TC of the 2020 season, was a category-4 Cape Verde hurricane with a particularly long track, and multiple rapid changes in intensity during its lifetime. Teddy developed into a tropical cyclone on 14 September and a category-1 hurricane on 16 September, subsequently reaching its peak intensity on 17 September. The storm weakened over the next few days before it reintensified after interacting with a midlatitude upper-level trough. It later weakened and dissipated ather making landfall over Nova Scotia. Teddy's track was almost entirely over the observation-sparse ocean, and due to its significant land threat, featured multiple NHC reconnaissance mission datasets. It also featured some of the largest and strongest Atlantic hurricane wind fields on record (NOAA 2021). For these reasons, we determined Teddy would be an excellent choice to examine as a case study for our satellite radiance bias correction experiments.

Figure 13 shows Teddy's NHC best track estimate, along with the evolution of both the control and online experiments' track forecasts, and Fig. 14 displays similar information but for maximum wind speed error. The improved track forecasting skill of

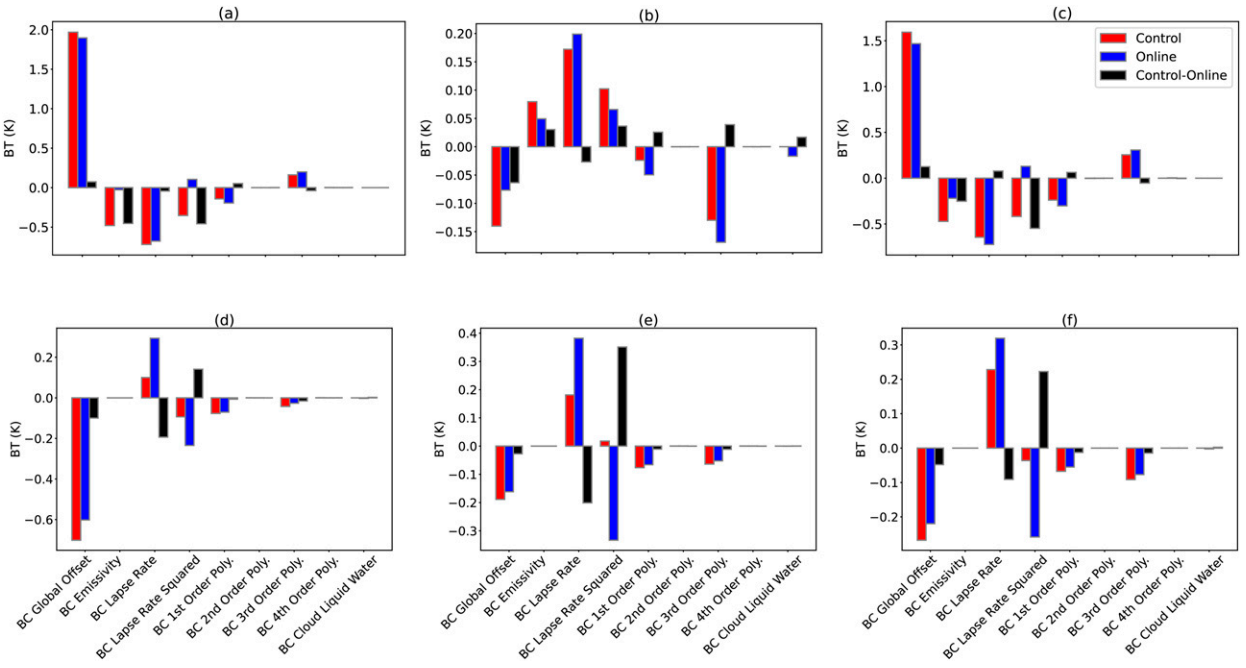


FIG. 10. These panels detail time- and domain-averaged weighted predictor values for select instruments and channels for each experiment. The *x* axis gives each weighted predictor, while the red, blue, and black colored bars give their time- and domain-averaged values for the control experiment, online experiment, and their difference, respectively. They are (a) MHS *MetOp-B*, channel 1; (b) AMSU-A *N15*, channel 4; (c) MHS *NOAA-19* (*N19*), channel 1; (d) AMSU-A *NOAA-18* (*N18*), channel 11; (e) AMSU-A *N18*, channel 11; (f) AMSU-A *MetOp-A*, channel 11.

the online experiment, especially at longer forecast lead times, is evident. Furthermore, we find that the online configuration results in noticeably improved track and intensity errors around the initial TC intensification time (~0000 UTC 14 September), and likewise for track error during Teddy's reintensification (~1200 UTC 22 September).

These conclusions are further supported in Fig. 15, where track and intensity error statistics are plotted as a function of lead time. Although the intensity forecast results are mixed at early lead times, the online configuration features substantially improved intensity forecasts after 42 h.

To explain how the online configuration produces more accurate track and intensity forecasts, we provide a more detailed analysis of model field variables at these two pivotal moments in Teddy's evolution: Initial intensification and reintensification. At initial intensification, we find that the online configuration produces a TC vortex that is deeper, tighter, and intensifies more quickly than for the control experiment. We examine a special verification domain that forms a ring or annulus around the vortex, extending from 200 to 2000 km from the predicted vortex center. In Fig. 16a, we display a time series of model field differences between online and control (online − control) for four pertinent model variables: temperature, specific humidity, relative humidity, and geopotential height. Consequently, we find that the online configuration produces environmental conditions that are moister at low altitudes surrounding the storm, leading to conditions that are more conducive to rapid intensification.

Following Teddy's initial intensification and subsequent decay, the storm drifted northward and interacted with an upper-level trough on 22 September, as seen in Fig. 17a. A secondary intensification occurred, resulting in a category-2 hurricane, which was accompanied by an altered north-northeastward track, as seen in Figs. 13 and 14. Although both configurations provided skillful predictions for intensity at this point, the control produced track forecasts that extended too far east initially, while the online experiment produced much more accurate track forecasts. Analysis of the model fields at extended lead times (48+ h) revealed that the control configuration predicted the upper-tropospheric trough to have propagated farther east than observed, as suggested in Fig. 13a, while the online configuration predicted a farther westward trough location, as can be seen in Fig. 17b. Due to this slightly more inaccurate environmental characterization from the control experiment, the forecasts initialized from these analyses brought the TC too far to the east.

To summarize, we find that online bias correction method produces improvements in track and intensity forecasting in this case, although we caution against the generalization of these results alone. Instead, we highlight that this case study demonstrates how a more accurate characterization of large-scale environment can lead to improved forecast results, particularly when coupled with inner-core observational datasets. Combined with the forecast error verification over a range of TCs presented in section a, particularly those with ample inner-core observations, we hypothesize that these results could extend to other TC forecasts as well.

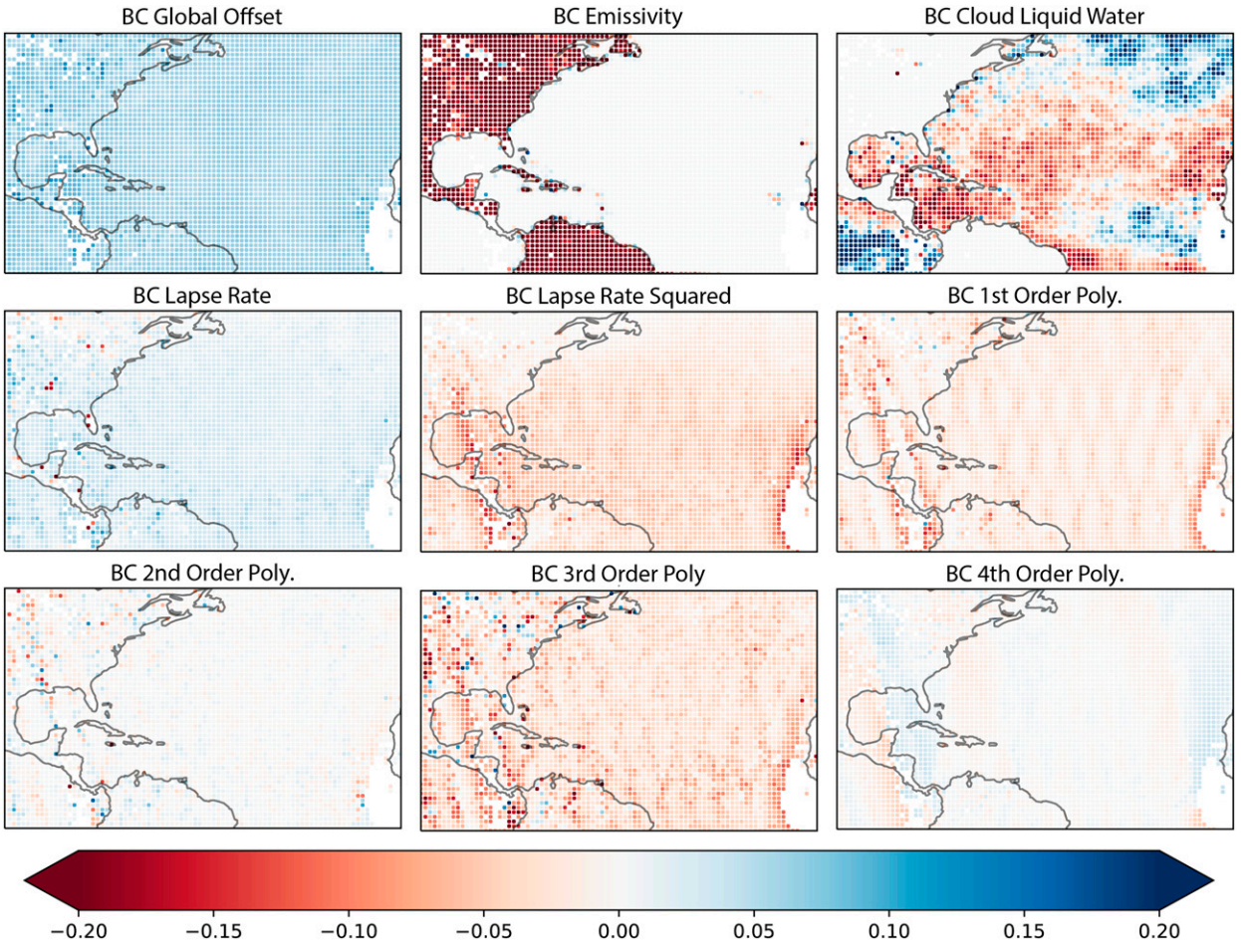


FIG. 11. Time-averaged weighted predictor differences between online and control experiments (online − control) for channel 4 of the AMSU-A *N15* instrument, in units of BT (K).

5. Discussion and conclusions

For many limited-area models, the decision to operate on a partially cycling basis often comes down to concerns of model drift and an insufficient sampling of innovations for estimating bias for satellite radiance measurements. For modeling systems that are able to avoid these challenges, the ability to run continuous, fully cycling, domainwide data assimilation can prove extremely valuable. One such benefit comes from the

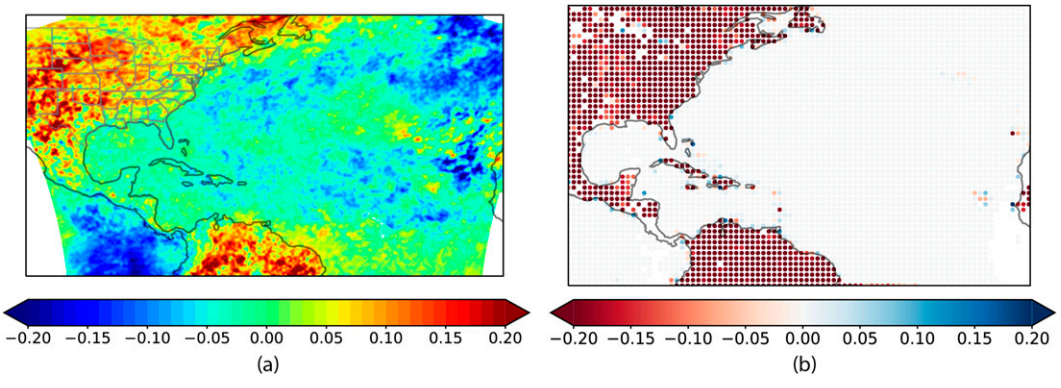


FIG. 12. (left) The time-average temperature field difference between control and online at 950 mb (online − control). (right) The difference in the emissivity predictor between control and online for channel 4 on the AMSU-A *N15* instrument. Note that the channel-4 weighting function peaks near the surface.

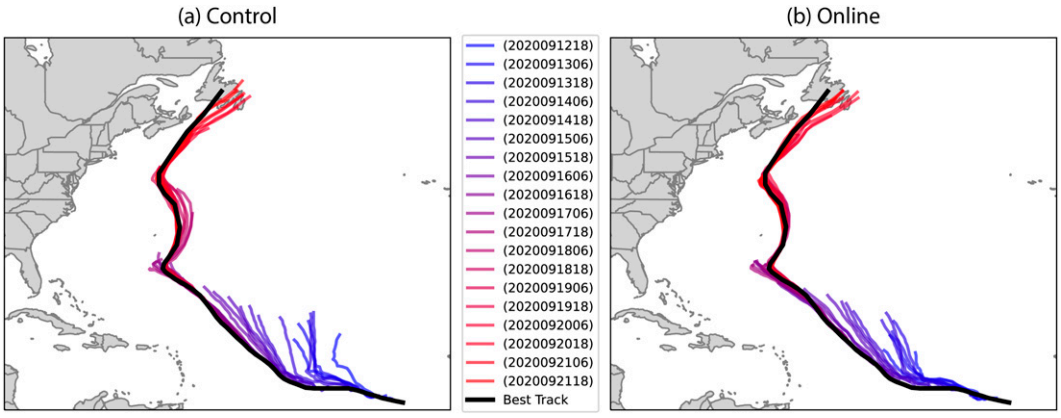


FIG. 13. Track error verification, (a) control vs (b) online, for AL20, Hurricane Teddy. The blue lines represent forecasts generated early in the model time span, transitioning to red as forecast time progresses. Black represents the best track estimate.

ability to self-cycle information such as innovations and model fields needed to estimate key parameters for specifying measurement bias, namely, predictor weights or "coefficients" that are used for statistical bias models. The decision for regional models to adopt bias correction coefficients from global models is usually one of practicality, given that partially cycling systems are unable to adopt an online configuration. However, even in cases where the system is fully cycling, these coefficients are still often adopted from lower-resolution global models. This decision is made because regional domain sizes are assumed to be too small, and global model data benefits from more thorough model tuning that could produce better bias correction configurations.

The current study investigates the potential of developing a fully cycled prediction system for the NOAA Hurricane Analysis and Forecast System (HAFS) that performs its own self-contained bias correction scheme for satellite radiance measurements. A number of key conclusions can be drawn from comparisons of a fully cycled HAFS system with and without this feature. First, introducing online estimates of bias correction parameters can produce improved TC forecasts, even in limited-area models such as HAFS. We postulate that the fully online bias correction leads to improved characterization of the surrounding synoptic environment during data assimilation, leading to more accurate steering flow associated with storms. In particular, we find the impact of online bias correction to be most noticeable for storms with extensive inner-core observations, as these observations can more precisely characterize TC vortex location within the domain by "filling in the gaps" near the storm center. We explore this finding in more detail for a case study of Hurricane Teddy, where the online configuration produced an environment that was more conducive to rapid intensification and consequently did a better job of forecasting Teddy's initial intensification and development. Likewise, Teddy's later interaction with

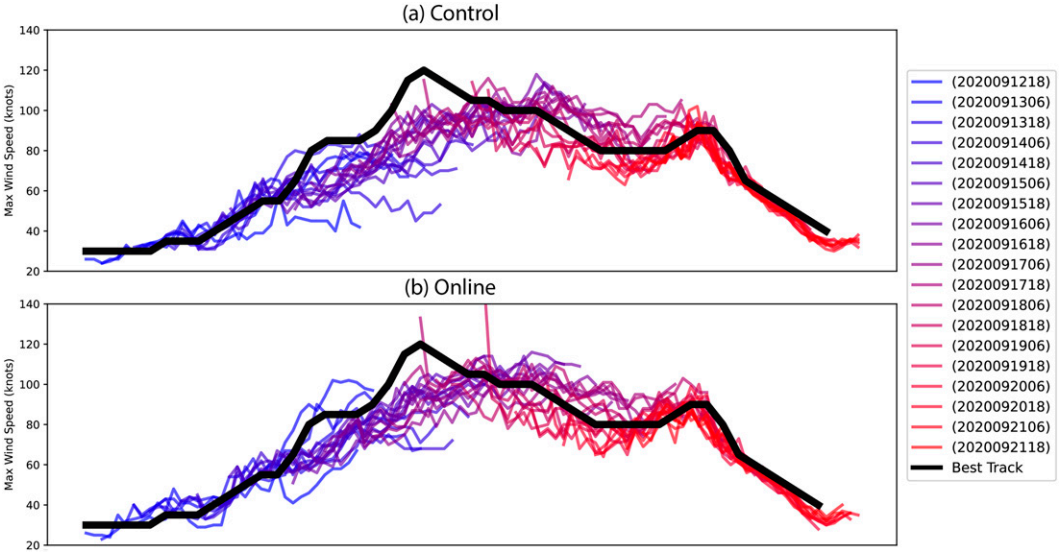


FIG. 14. As in Fig. 13, but for 10-m maximum wind speed.

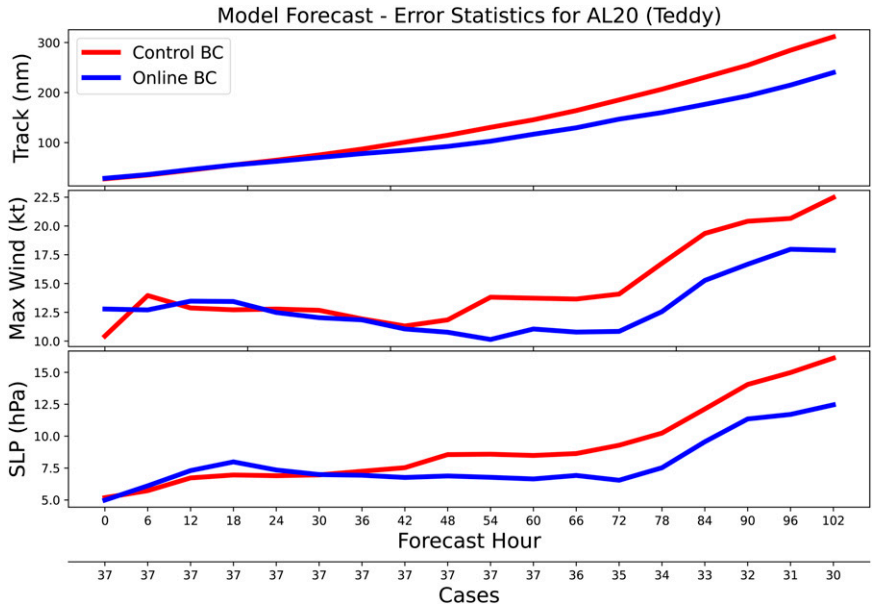


FIG. 15. Forecast error verification of control and online experiments for AL20, Teddy. These errors are verified against NHC best track and intensity data.

an upper-level trough and the resulting reintensification was more faithfully represented by the online experiment.

We also establish that fully online satellite bias correction can produce lower domain-averaged errors for key variables such as temperature and humidity, although these relationships are highly dependent on qualities such as lead time and height level and thus are arguably less robust than the TC-focused error statistics presented is section 4a.

Furthermore, we can draw some conclusions about major differences in mean bias over land versus over sea, and how

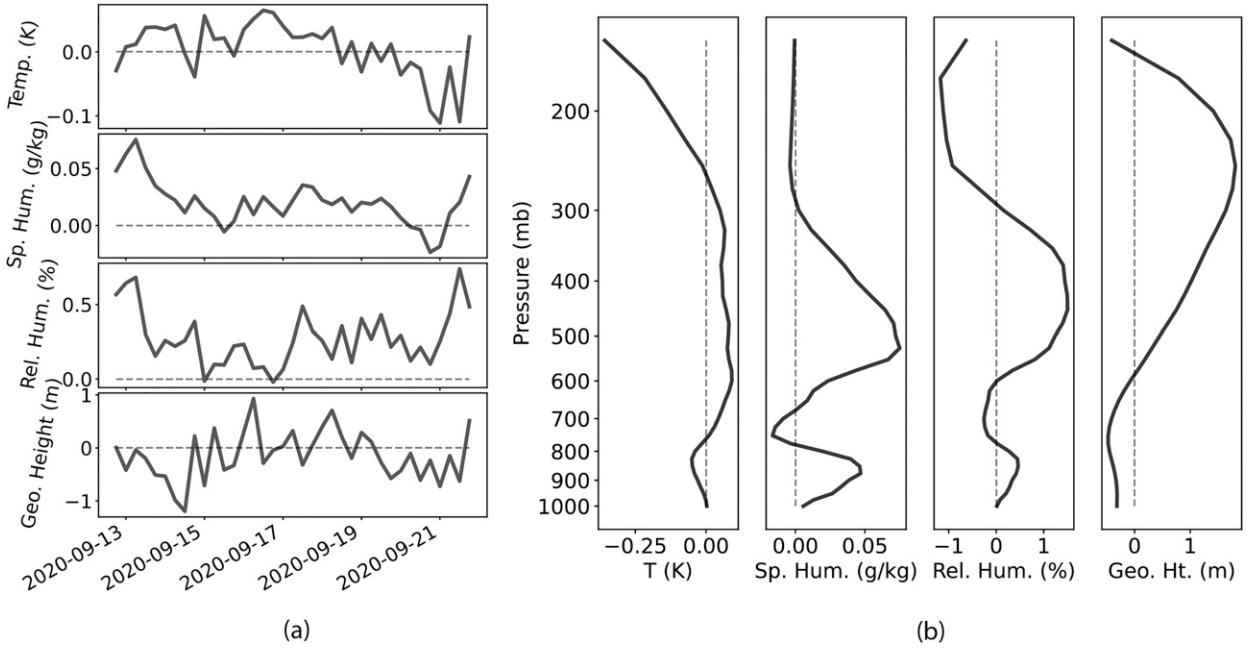


FIG. 16. (a) Displays a time series of model field differences between online and control (online − control) for four pertinent model variables: temperature, specific humidity, relative humidity, and geopotential height. The verification domain for these variables forms an annulus around the TC, extending from 200 to 2000 km around the TC. The values are column averaged from 950 to 500 mb. (b) Displays a time-averaged vertical profile of the same variables over the same verification domain.

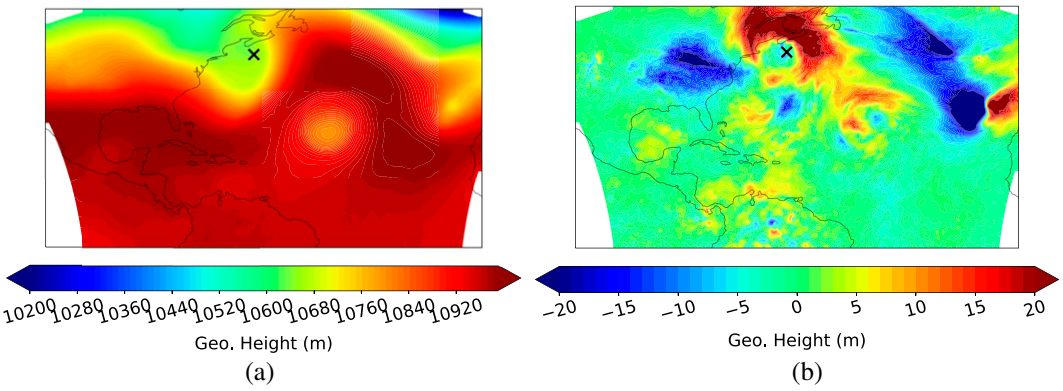


FIG. 17. (a) The 48-h lead-time 250-mb geopotential height model field generated at 0000 UTC 21 Sep for the control experiment. (b) The geopotential height difference between online and control at the same forecast time, lead time, and pressure level (online vs control). In both cases, a Gaussian smoothing filter has been applied. A black "X" marks the NHC best estimate location for Hurricane Teddy at 0000 UTC 23 Sep.

these differences can be tied to how biases are specified in satellite radiance measurements. For example, we find that the online method generally produces warmer and wetter model states throughout the midatmosphere both over land and over sea, but that the relationship is more complicated near the surface. Clearly this is a reflection of different radiative characteristics of land versus sea, and we demonstrate that these differences are a direct result of weighted predictor field specification. We also note that these differences in mean bias are largest over sea, where in situ measurements are sparse and accurate specification of satellite radiance measurements becomes particularly important.

This study also shows that modeling systems that ingest only clear-sky radiances should not adopt configurations that are tuned for ingesting all-sky radiances. In this case, the GDAS model has tuned its bias predictor coefficient values toward all-sky radiances, e.g., setting the cloud liquid water predictor to zero across the entire domain. This is problematic for the control HAFS system considering said system only assimilates clear-sky radiances, and thus would benefit from the cloud liquid water bias correction term. A similar conclusion can be drawn for models that borrow bias information, among other things, from global models with different model tops. HAFS has a maximum pressure level of 2 hPa, whereas GDAS extends well up into the upper stratosphere. This disparity results in a strong difference in model variables, such as temperature, at high altitudes. It also results in peculiar and unexplainable values for weighted predictors in channels that peak at high altitudes, particularly the global offset, lapse rate, and lapse rate squared predictors. These disparities make it difficult to assess the effectiveness of these bias predictors for these channels.

We also note that experiments performed in the current study adopt a single domain with 6-km grid spacing, which differs from the operational configuration of HAFS that uses a 2-km nest. While our results demonstrate that adopting a fully online bias correction strategy can produce notable differences in forecast skill, particularly mean errors, we speculate that these differences may be even larger if our study adopted the same convective-permitting 2-km grid spacing that is used for the operational HAFS. Doing so will further reduce any similarities in model process error that exist between resolved scales and choices of physical parameterization schemes between regional and global configurations of FV3 (HAFS versus GFS), and further exacerbate the issues raised in this study for quantifying bias in radiance measurements.

In concluding, we also note several directions in which this research could be continued or expanded. Given that several predictor coefficients and weighted predictor differences continue to diverge even at the end of the 6 weeks of cycling data assimilation, it may be useful to extend this study for entire hurricane seasons. It would also be beneficial to reproduce the online bias correction experiment starting with zero values for coefficients, as opposed to starting with coefficients adopted from GDAS. In this way, we could eliminate bias from the GDAS model more quickly, and it may further improve forecasts at earlier cycles in the experiment period. Eventually, this research may find value for the expansion of radiance assimilation from just clear-sky assimilation to all-sky assimilation; however, since all-sky radiance assimilation is still an active area of research, it is difficult to assess how this research may be relevant. Last, future research will focus on additional requirements needed to maintain a fully cycled ensemble data assimilation system for operational hurricane prediction. This effort parallels similar efforts underway for other regional modeling applications, such as severe convective storms (e.g., Schwartz et al. 2022).

Acknowledgments. This study was supported by NOAA Grant NA19NES4320002 [Cooperative Institute for Satellite Earth System Studies (CISESS)] at the University of Maryland/ESSIC, and NOAA Grants NA20OAR4600281 and NA22OAR4590184. Computing for this project was provided by the "Mississippi State University Orion" high-performing computing system.

Data availability statement. Data supporting the findings of this work are stored on the Mississippi State University

Orion computing facility and are available from the corresponding author upon request. Specifically, model fields are available in GRIB2 format and storm track and intensity metrics are stored in text files.

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 airborne Doppler radar observations in an OSSE. *Mon. Wea. Rev.*, **140**, 1843–1862, https://doi.org/10.1175/MWR-D-11-00212.1.

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.

Benáček, P., and M. Mile, 2019: Satellite bias correction in the regional model ALADIN/CZ: Comparison of different VarBC approaches. *Mon. Wea. Rev.*, **147**, 3223–3239, https://doi.org/10.1175/MWR-D-18-0359.1.

Biswas, M. K., D. Stark, and L. Carson, 2018: GFDL vortex tracker user's guide version 3.9a. NOAA/GFDL Doc., 35 pp., https://dtcenter.org/sites/default/files/community-code/gfdl/standalone_tracker_UG_v3.9a.pdf.

Carr, L. E., III, and R. L. Elsberry, 1997: Models of tropical cyclone wind distribution and beta-effect propagation for application to tropical cyclone track forecasting. *Mon. Wea. Rev.*, **125**, 3190–3209, https://doi.org/10.1175/1520-0493(1997)125<3190:MOTCWD>2.0.CO;2.

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**, 60–83, https://doi.org/10.1016/j.jmarsys.2005.09.016.

——, and Coauthors, 2009: US GODAE: Global ocean prediction with the HYbrid Coordinate Ocean Model (HYCOM). *Oceanography*, **22**, 64–75, https://doi.org/10.5670/oceanog.2009.39.

Collard, A. D., and A. P. McNally, 2009: The assimilation of infrared atmospheric sounding interferometer radiances at ECMWF. *Quart. J. Roy. Meteor. Soc.*, **135**, 1044–1058, https://doi.org/10.1002/qj.410.

Deng, G., D. Zhang, T. Zhu, and A. Wang, 2009: Use of the Advanced Microwave Sounding Unit data to improve typhoon prediction. *Prog. Natl. Sci.*, **19**, 369–376, https://doi.org/10.1016/j.pnsc.2008.08.001.

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.

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**, 617, https://doi.org/10.3390/atmos11060617.

Emanuel, K., and F. Zhang, 2016: On the predictability and error sources of tropical cyclone intensity forecasts. *J. Atmos. Sci.*, **73**, 3739–3747, https://doi.org/10.1175/JAS-D-16-0100.1.

Geer, A. J., and P. Bauer, 2011: Observation errors in all-sky data assimilation. *Quart. J. Roy. Meteor. Soc.*, **137**, 2024–2037, https://doi.org/10.1002/qj.830.

Goldberg, M. D., D. S. Crosby, and L. Zhou, 2001: The limb adjustment of AMSU-A observations: Methodology and validation. *J. Appl. Meteor.*, **40**, 70–83, https://doi.org/10.1175/1520-0450(2001)040<0070:TLAOAA>2.0.CO;2.

Gopalakishnan, S., and Coauthors, 2021: 2020 HFIP R&D activities summary: Recent results and operational implementation. NOAA HFIP Tech. Rep. HFIP2021-1, 49 pp., https://doi.org/10.25923/718e-6232.

Han, J., and C. S. Bretherton, 2019: TKE-based moist eddy-diffusivity mass-flux (EDMF) parameterization for vertical turbulent mixing. *Wea. Forecasting*, **34**, 869–886, https://doi.org/10.1175/WAF-D-18-0146.1.

——, W. Wang, Y. C. Kwon, S.-Y. Hong, V. Tallapragada, and F. Yang, 2017: Updates in the NCEP GFS cumulus convection schemes with scale and aerosol awareness. *Wea. Forecasting*, **32**, 2005–2017, https://doi.org/10.1175/WAF-D-17-0046.1.

Han, Y., P. van Delst, Q. Liu, F. Weng, B. Yan, R. Treadon, and J. Derber, 2006: JCSDA Community Radiative Transfer Model (CRTM): Version 1. NOAA Tech. Rep. NESDIS 122, 40 pp., https://repository.library.noaa.gov/view/noaa/1157.

Hanley, D., J. Molinari, and D. Keyser, 2001: A composite study of the interactions between tropical cyclones and upper-tropospheric troughs. *Mon. Wea. Rev.*, **129**, 2570–2584, https://doi.org/10.1175/1520-0493(2001)129<2570:ACSOTI>2.0.CO;2.

Hartman, C. M., X. Chen, E. E. Clothiaux, and M.-Y. Chan, 2021: Improving the analysis and forecast of Hurricane Dorian (2019) with simultaneous assimilation of *GOES-16* all-sky infrared brightness temperatures and tail Doppler radar radial velocities. *Mon. Wea. Rev.*, **149**, 2193–2212, https://doi.org/10.1175/MWR-D-20-0338.1.

Hazelton, A., L. Harris, and S.-J. Lin, 2018: Evaluation of tropical cyclone structure forecasts in a high-resolution version of the multiscale GFDL FvGFS model. *Wea. Forecasting*, **33**, 419–442, https://doi.org/10.1175/WAF-D-17-0140.1.

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

Hendricks, E. A., M. S. Peng, B. Fu, and T. Li, 2010: Quantifying environmental control on tropical cyclone intensity change. *Mon. Wea. Rev.*, **138**, 3243–3271, https://doi.org/10.1175/2010MWR3185.1.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hu, M., C. Zhou, H. Shao, D. Stark, and K. Newman, 2016: GSI advanced user's guide version 3.5. NOAA/GSI Doc., 124 pp., https://dtcenter.org/sites/default/files/community-code/gsi/docs/users-guide/AdvancedGSIUserGuide_v3.5.0.0.pdf.

Jones, T. A., X. Wang, P. Skinner, A. Johnson, and Y. Wang, 2018: Assimilation of *GOES-13* imager clear-sky water vapor (6.5 μm) radiances into a Warn-on-Forecast system. *Mon. Wea. Rev.*, **146**, 1077–1107, https://doi.org/10.1175/MWR-D-17-0280.1.

Kim, Y.-J., and A. Arakawa, 1995: Improvement of orographic gravity wave parameterization using a mesoscale gravity wave model. *J. Atmos. Sci.*, **52**, 1875–1902, https://doi.org/10.1175/1520-0469(1995)052<1875:IOOGWP>2.0.CO;2.

Li, J., and H. Liu, 2009: Improved hurricane track and intensity forecast using single field-of-view advanced IR sounding measurements. *Geophys. Res. Lett.*, **36**, L11813, https://doi.org/10.1029/2009GL038285.

Lin, H., 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.

Liu, Y.-A., H.-L. A. Huang, A. H. N. Lim, and W. Gao, 2018: Adaptive bias correction of advanced infrared sounding radiance assimilation in a regional model and its impact on typhoon forecast. *J. Appl. Remote Sens.*, **12**, 026012, https://doi.org/10.1117/1.JRS.12.026012.

Liu, Z., C. S. Schwartz, C. Snyder, and S.-Y. Ha, 2012: Impact of assimilating AMSU-A radiances on forecasts of 2008 Atlantic tropical cyclones initialized with a limited-area ensemble Kalman filter. *Mon. Wea. Rev.*, **140**, 4017–4034, https://doi.org/10.1175/MWR-D-12-00083.1.

Lu, X., X. Wang, M. Tong, and V. Tallapragada, 2017: GSI-based, continuously cycled, dual-resolution hybrid ensemble–variational data assimilation system for HWRF: System description and experiments with Edouard (2014). *Mon. Wea. Rev.*, **145**, 4877–4898, https://doi.org/10.1175/MWR-D-17-0068.1.

McNally, A. P., J. C. Derber, W. Wu, and B. B. Katz, 2000: The use of TOVS level-1b radiances in the NCEP SSI analysis system. *Quart. J. Roy. Meteor. Soc.*, **126**, 689–724, https://doi.org/10.1002/qj.49712656315.

——, P. D. Watts, J. A. Smith, R. Engelen, G. A. Kelly, J. N. Thépaut, and M. Matricardi, 2007: The assimilation of AIRS radiance data at ECMWF. *Quart. J. Roy. Meteor. Soc.*, **132**, 935–957, https://doi.org/10.1256/qj.04.171.

Mehra, A., V. Tallapragada, Z. Zhang, B. Liu, L. Zhu, W. Wang, and H.-S. Kim, 2018: Advancing the state of the art in operational tropical cyclone forecasting at NCEP. *Trop. Cyclone Res. Rev.*, **7**, 51–56, https://doi.org/10.6057/2018TCRR01.06.

Minamide, M., and F. Zhang, 2017: Adaptive observation error inflation for assimilating all-sky satellite radiance. *Mon. Wea. Rev.*, **145**, 1063–1081, https://doi.org/10.1175/MWR-D-16-0257.1.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, https://doi.org/10.1029/97JD00237.

Niu, G.-Y., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res.*, **116**, D12109, https://doi.org/10.1029/2010JD015139.

NOAA, 2021: Monthly tropical cyclones report for annual 2020. NOAA Tech. Rep., accessed 30 December 2022, https://www.ncei.noaa.gov/access/monitoring/monthly-report/tropical-cyclones/202013.

Ota, Y., J. C. Derber, E. Kalnay, and T. Miyoshi, 2013: Ensemble-based observation impact estimates using the NCEP GFS. *Tellus*, **65A**, 20038, https://doi.org/10.3402/tellusa.v65i0.20038.

Poterjoy, J., and F. Zhang, 2014: 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 ——, 2016: Comparison of hybrid four-dimensional data assimilation methods with and without the tangent linear and adjoint models for predicting the life cycle of Hurricane Karl (2010). *Mon. Wea. Rev.*, **144**, 1449–1468, https://doi.org/10.1175/MWR-D-15-0116.1.

——, 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.

——, G. J. Alaka Jr., 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**, 661–677, https://doi.org/10.1175/WAF-D-20-0204.1.

Prasad, V. S., K. J. Ramesh, A. K. Bohra, and R. C. Bhatia, 1998: Assessment of quality and impact of full resolution TOVS temperature profile data on the operational global data assimilation-forecast system of India. *Meteor. Atmos. Phys.*, **68**, 197–212, https://doi.org/10.1007/BF01030211.

Reale, O., E. L. McGrath-Spangler, W. McCarty, D. Holdaway, and R. Gelaro, 2018: Impact of adaptively thinned AIRS cloud-cleared radiances on tropical cyclone representation in a global data assimilation and forecast system. *Wea. Forecasting*, **33**, 909–931, https://doi.org/10.1175/WAF-D-17-0175.1.

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.

Schwartz, C. S., Z. Liu, Y. Chen, and X.-Y. Huang, 2012: Impact of assimilating microwave radiances with a limited-area ensemble data assimilation system on forecasts of Typhoon Morakot. *Wea. Forecasting*, **27**, 424–437, https://doi.org/10.1175/WAF-D-11-00033.1.

——, J. Poterjoy, J. R. Carley, D. C. Dowell, G. S. Romine, and K. Ide, 2022: Comparing partial and continuously cycling ensemble Kalman filter data assimilation systems for convection-allowing ensemble forecast initialization. *Wea. Forecasting*, **37**, 85–112, https://doi.org/10.1175/WAF-D-21-0069.1.

Sippel, J. A., X. Wu, S. D. Ditchek, V. Tallapragada, and D. T. Kleist, 2022: Impacts of assimilating additional reconnaissance data on operational GFS tropical cyclone forecasts. *Wea. Forecasting*, **37**, 1615–1639, https://doi.org/10.1175/WAF-D-22-0058.1.

Xu, D., Z. Liu, X.-Y. Huang, J. Min, and H. Wang, 2013: Impact of assimilating IASI radiance observations on forecasts of two tropical cyclones. *Meteor. Atmos. Phys.*, **122**, 1–18, https://doi.org/10.1007/s00703-013-0276-2.

Zapotocny, T. H., J. A. Jung, J. F. L. Marshall, and R. E. Treadon, 2008: A two-season impact study of four satellite data types and rawinsonde data in the NCEP Global Data Assimilation System. *Wea. Forecasting*, **23**, 80–100, https://doi.org/10.1175/2007WAF2007010.1.

Zeng, Z., Y. Wang, and C.-C. Wu, 2007: Environmental dynamical control of tropical cyclone intensity—An observational study. *Mon. Wea. Rev.*, **135**, 38–59, https://doi.org/10.1175/MWR3278.1.

Zhang, F., Y. Weng, J. A. Sippel, Z. Meng, and C. H. Bishop, 2009: Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **137**, 2105–2125, https://doi.org/10.1175/2009MWR2645.1.

Zhang, M., M. Zupanski, M.-J. Kim, and J. A. Knaff, 2013: Assimilating AMSU-A radiances in the TC core area with NOAA operational HWRF (2011) and a hybrid data assimilation system: Danielle (2010). *Mon. Wea. Rev.*, **141**, 3889–3907, https://doi.org/10.1175/MWR-D-12-00340.1.

Zheng, W., M. Ek, K. Mitchell, H. Wei, and J. Meng, 2017: Improving the stable surface layer in the NCEP Global Forecast

System. *Mon. Wea. Rev.*, **145**, 3969–3987, https://doi.org/10.1175/MWR-D-16-0438.1.

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.

——, and Coauthors, 2016: All-sky microwave radiance assimilation in NCEP's GSI analysis system. *Mon. Wea. Rev.*, **144**, 4709–4735, https://doi.org/10.1175/MWR-D-15-0445.1.

Zou, X., F. Weng, B. Zhang, L. Lin, Z. Qin, and V. Tallapragada, 2013: Impacts of assimilation of ATMS data in HWRF on track and intensity forecasts of 2012 four landfall hurricanes. *J. Geophys. Res. Atmos.*, **118**, 11 558–11 576, https://doi.org/10.1002/2013JD020405.