Received: 21 March 2024 | Revised: 31 July 2024 | Accepted: 12 August 2024

DOI: 10.1002/qj.4857

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Assimilating atmospheric motion vector winds using a feature track correction observation operator

Ross N. Hoffman[1,2] | **Hui Liu**[1,2] | **Katherine E. Lukens**[1,2] | **Kevin Garrett**[3] | **Kayo Ide**[4]

[1]NOAA/NESDIS/Center for Satellite Applications and Research (STAR), College Park, Maryland, USA

[2]Cooperative Institute for Satellite Earth System Studies (CISESS), University of Maryland, College Park, Maryland, USA

[3]NOAA/NWS/Office of Science and Technology Integration (OSTI), Silver Spring, Maryland, USA

[4]University of Maryland, College Park, Maryland, USA

Correspondence
Ross N. Hoffman, Cooperative Institute for Satellite Earth System Studies (CISESS), 5825 University Research Ct., College Park MD 20740, USA.
Email: rnh@umd.edu

Funding information
National Oceanic and Atmospheric Administration, Grant/Award Numbers: NA14NES4320003, NA19NES4320002

Abstract

Atmospheric motion vector (AMV) winds have positive impacts in operational numerical weather prediction (NWP) systems. These impacts might be improved with better treatment of the following error characteristics of AMVs. First, AMVs may have wind errors due to height assignment errors. Second, AMVs may have additional wind-speed biases in addition to those due to height assignment errors. Third, AMVs are representative of motion in a possibly thick atmospheric layer, not a single atmospheric level. Previous work proposed a variational feature track correction (FTC) method in which an observation operator is implemented that averages the NWP background winds optimally in the vertical. Here, a prototype feature track correction observation operator (FTC-OO) is implemented in the NOAA/NCEP data assimilation (DA) system. The parameters describing the vertical averaging are determined offline based on previous DA cycles. The FTC-OO reduces the observation minus background standard deviation by about 4%. Global observing-system experiments (OSEs) are performed comparing the FTC-OO with the operational observation operator. The forecast verification sample is 41 10-day forecasts. The OSEs show that the FTC-OO improves forecast skill, primarily for tropical geopotential height. Additional OSEs are performed that include *Aeolus* wind observations. The hypothesis that the *Aeolus* winds would enhance the impact of the FTC-OO was not borne out in these experiments—the *Aeolus* observations alone have a significant positive impact, but the impact of the FTC method in the presence of the *Aeolus* observations is neither enhanced nor degraded compared with the impact of the FTC method alone.

KEYWORDS

Aeolus HLOS winds, atmosphere, atmospheric motion vector (AMV) winds, data assimilation, feature track correction (FTC), NWP impact optimization, observation operator, winds

1 | INTRODUCTION

Atmospheric motion vectors (AMVs) are created in the millions each day by tracking clouds and water vapor (WV) features in satellite imagery at numerous wind-producing centers. AMVs include three-dimensional (3D) winds created by tracking features in retrieved WV imagery and cloud-track winds (CTWs) created by tracking infrared (IR) or visible imagery (e.g., Santek *et al.*, 2019, 2022). AMVs are assimilated in numerical weather prediction (NWP) systems and show positive impact (e.g., Lee *et al.*, 2022; Li *et al.*, 2020). However, they are often aggressively thinned, and their error characteristics are complex (e.g., Cordoba *et al.*, 2017; Lee & Song, 2017; Salonen *et al.*, 2015). Key among these error characteristics (Velden & Bedka, 2009) are the following:

- AMVs may have height assignment errors that induce wind errors when the observed heights are considered free of errors.
- AMVs may have additional wind-speed biases even after some bias correction.
- AMVs are representative of motion within a layer with depth up to a few hundred hPa, not a single atmospheric level.

Height assignments are made by comparing the cloud-top temperature with a "known" temperature profile. Errors in the temperature profile, as well as the methodology of matching the temperatures, can lead to height assignment errors (Velden & Bedka, 2009). Ultimately it must be recognized that AMVs have a somewhat complicated relationship with the true wind. Hoffman *et al.* (2022) propose a feature track correction (FTC) representation of AMVs to account for the above aspects of AMV errors.

When we began this work, a key motivation was the prospect of using Doppler wind lidar (DWL) observations to anchor AMVs in a variational FTC (VarFTC) system that would parallel using global navigation satellite system (GNSS) radio occultation (RO) observations to anchor radiance measurements in a variational bias correction (VarBC) system. VarBC has substantially improved the use of radiance observations (Zhu *et al.*, 2014). For VarBC, high-quality unbiased conventional and RO observations provide key anchoring information to account for biases in the NWP temperature background (i.e., the short-term forecast used as the background when assimilating observations, (e.g., Cucurull *et al.*, 2014; Healy, 2008)). Such observations are termed "anchoring" observations because they anchor the data assimilation (DA) to reality. VarFTC would do this for AMVs—that is, to account statistically for the above aspects of AMV errors using higher quality wind profiles as a reference.

The current study is a step towards developing VarFTC but falls short of demonstrating VarFTC in two important ways. First, the optimization of the feature track correction observation operator (FTC-OO) is performed offline using recent backgrounds. In a true VarFTC, the coefficients of the FTC-OO are determined inline as part of an augmented state vector. Appendix A.4 describes how this could be done. Second, the DWL observations used here are *Aeolus* wind profiles, which have larger than hoped for random errors and biases. Moreover, preprocessing of *Aeolus* winds uses European Centre for Medium-Range Weather Forecasts (ECMWF) background information. Ideally, the DWL observations would be assimilated as the Doppler shift (and amplitude) of the lidar signal backscattered by the sampled volume (Rennie & Isaksen, 2020, p. 16). This would parallel the way RO observations are assimilated as refractivities (or bending angles) and radiometer observations are assimilated as radiances instead of as retrieved temperature and humidity profiles. Assimilating observations close to the original measurement (refractivity not temperature, Doppler shift not wind) is critical in modern DA systems to eliminate the inevitable biases introduced by the assumptions and prior data used in retrieval methods. However, the feasibility of assimilating Doppler shift would require an efficient and fast forward operator to calculate Doppler shift from the NWP model state.

The present study assesses the utility of FTC in global NWP DA by comparing observing system experiments (OSEs) with and without the implementation of a prototype FTC-OO within the Global Statistical Interpolation (GSI) DA system, which is part of the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction (NCEP) Finite-Volume Cubed-Sphere Global Forecast System (FV3GFS). To test this hypothesis that DWL observations would anchor the FTC method, we conducted additional OSEs that added *Aeolus* observations of horizontal line-of-sight (HLOS) wind profiles. In what follows, *Aeolus* winds are *Aeolus* HLOS winds. *Aeolus* winds are observed by the first satellite DWL, which was hosted by the European Space Agency (ESA) Earth Explorer satellite named *Aeolus* (Rennie *et al.*, 2021; Stoffelen, 2005). The *Aeolus* wind observations were reprocessed with different software versions, referred to as baselines. This study made use of Baseline 10 (B10) *Aeolus* Level 2B (L2B) wind observations. The Atmospheric LAser Doppler INstrument (ALADIN) Processor Releases website lists additional baseline datasets that are available.[1] The L2B data make use of ECMWF temperature and humidity profiles in the Rayleigh–Brillouin correction and ECMWF wind profiles in the *Aeolus* DWL

main mirror temperature-dependent bias correction (Rennie *et al.*, 2021; Weiler *et al.*, 2021).

Aeolus measures the Doppler backscatter in two "channels." In the Mie channel, backscatter from clouds and aerosols is used to derive wind profiles, while in the Rayleigh channel molecular backscatter is used. Therefore, *Aeolus* Mie winds in clear conditions and *Aeolus* Rayleigh winds in cloudy conditions have large errors and are not used for NWP. In what follows, Rayleigh and Mie winds are the *Aeolus* Rayleigh-clear and Mie-cloudy winds, respectively. The characteristics of the Rayleigh and Mie winds are different, with the Mie winds having smaller mean and random errors. However, the Rayleigh winds are more numerous. The *Aeolus* HLOS L2B Rayleigh and Mie winds used in this study have been assimilated in operational DA systems at NWP centers worldwide and have demonstrated positive impact on global weather forecasts (e.g., Garrett *et al.*, 2022; Rennie *et al.*, 2021). The effectiveness of *Aeolus* winds to serve as anchoring observations should be increased because their height assignments are accurate, since they are based on range-gating the received signals, but might be decreased because they are HLOS winds and not 2D horizontal winds like the AMVs.

The main goals of this study are (1) to document our implementation of the FTC-OO in the GSI, (2) to examine the impact of the FTC-OO in the GSI, and (3) to determine whether these impacts are improved by the presence of *Aeolus* wind observations. Section 2 describes the FTC-OO, with details presented in the Appendix. Section 3 outlines the experimental setup and Section 4 presents the results. Section 5 summarizes the study, draws conclusions, and offers recommendations for future work.

2 | THE FEATURE TRACK CORRECTION OBSERVATION OPERATOR (FTC-OO)

DA systems minimize a cost function that "balances" the misfit to the observations and the misfit to the background (Kalnay, 2002). The minimization is with respect to the control vector, which includes the state vector describing dynamical variables of the NWP system. If the control vector includes additional parameters—related to physical parameterizations or observation operators—it is an augmented state vector. Within the DA, the misfits are combined in quadratic forms with weight matrices given by the inverse error covariance matrices. The error covariance matrices have a critical role in defining the DA cost function, but can be difficult to specify. Deriving this form of the cost function assumes no correlations between errors of the observation and background misfits. The misfit to the background is the difference between the background and the model state. The misfit to observations is the difference between the observation and the estimate of the observation determined from the model state, which is calculated by the observation operator. In this study we define a FTC-OO that estimates an AMV observation as a linear least-squares adjustment of a vertical average of the model horizontal wind profile. In practice, representations of the background and the data that have small or no correlations are preferred, to reduce the number of parameters that must be estimated to specify the necessary covariances. To that end, DA systems usually represent the wind field in terms of stream and velocity potential functions, which are assumed to have uncorrelated errors, rather than in terms of wind components which do have correlated errors. At the same time, wind observations are assimilated (here and in general) as wind components that are assumed to have uncorrelated errors. This might be improved upon by assimilating winds that are observed by feature tracking in terms of speed and direction, but we will not pursue that here, except to note that an FTC in terms of speed and direction could be implemented.

The FTC-OO is described for convenience as a two-step process for a single AMV observation, $\mathbf{V}^{o}$, at a reported observation pressure of p^{o}, where $\mathbf{V}$ is the horizontal vector wind, p is the pressure, and superscript o denotes an observation. Step 1 evaluates $\mathbf{V}(p)$, the model state horizontal vector wind profile at the AMV observation location and time. This is determined by applying the normal wind observation operator to the AMV observation latitude, longitude, and time at a number of pressure values including and surrounding p^{o}. Usually, and in the experiments reported here, that interpolation is linear in latitude, longitude, log of pressure, and time. Step 1 is symbolically written as

$$\mathbf{V}(p) = g(\mathbf{x}), \tag{1}$$

where $\mathbf{x}$ is the model state and g is the "horizontal wind profile observation operator." Step 2 estimates $\mathbf{V}^{o}$ as a function of $\mathbf{V}(p)$ as

$$\hat{\mathbf{V}} = f(\mathbf{V}(p)) = \gamma\overline{\mathbf{V}} + \delta\mathbf{V}, \tag{2}$$

where $\hat{\mathbf{V}}$ is the estimate of $\mathbf{V}^{o}$, $\overline{\mathbf{V}}$ is the vertical average of $\mathbf{V}(p)$, and γ and $\delta\mathbf{V}$ are multiplicative and additive bias correction parameters. The vertical average is over a pressure layer of thickness Δz that is offset from p^{o} by an amount h. The parameters γ, $\delta\mathbf{V}$, Δz, and h are determined as described below. Note that γ is a unitless scalar, $\delta\mathbf{V}$ has two components and the same units as $\mathbf{V}^{o}$, and Δz and h have units of pressure. The FTC-OO, $H^{\mathrm{FTC}}(\mathbf{x})$, combines Steps 1

and 2 to estimate a single AMV observation, $\mathbf{V}^{o}$, as

$$y = H^{\mathrm{FTC}}(\mathbf{x}) = f(g(\mathbf{x})) = \hat{\mathbf{V}}. \quad (3)$$

Appendix A.1 describes the implementation of the discretized FTC-OO.

The FTC optimization determines the quantities γ, $\delta\mathbf{V}$, Δz, and h to minimize the mean squared difference between $\mathbf{V}^{o}$ and $\hat{\mathbf{V}}$. The optimization is performed for various subsets of the AMVs. In the present study, the subsets are defined by all combinations of (1) the different AMV types defined in the GSI, (2) three geographic regions, and (3) three height bins. The AMV types used in this study and their acronyms are listed in the caption of Figure 1. The geographic regions are the Northern Hemisphere Extratropics (NHX, 20°N–90°N), the Tropics (20°S–20°N), and the Southern Hemisphere Extratropics (SHX, 20°S–90°S). The height bins, denoted low, middle, and high, are separated by the 800- and 450-hPa pressure levels. As described in Appendix A.1, interpolation near subset boundaries is made efficient by restating Equation (2) as a weighted sum.

Within each subset, for example, *Geostationary Operational Environmental Satellite* (*GOES*) IR tropical high AMVs, the calculation is as described by Hoffman *et al.* (2022), where the FTC parameters are determined by minimizing a cost function as two nested optimizations. The outer optimization determines the layer to be averaged over (Δz and h) and the inner optimization determines the multiplicative and additive bias corrections (γ and $\delta\mathbf{V}$) to be applied to the layer average horizontal vector wind, $\overline{\mathbf{V}}$. The inner optimization is a linear regression. The outer optimization is a search in the space of layer thickness (Δz) and layer offset (h) (see Appendix A.3). Because of the vertical discretization, Δz and h take on discrete values, thereby defining a 2D gridded cost function, which allows the use of a brute-force or directed search for the minimum on this grid. Discussions of the efficient implementation of the optimizations used in this study are given in Appendices A.2 and A.3.

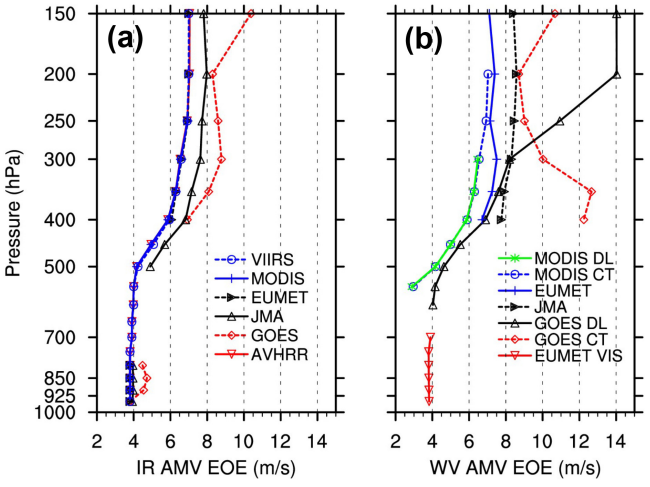

FIGURE 1 Estimated observation errors (EOEs, m/s) evaluated for all AMV types (different lines identified in the legend) assimilated in the BASE experiment during the study period, for all locations, for (a) IR AMV types and (b) WV and visible AMV types, displayed as a function of pressure (hPa). Figure acronyms: AVHRR=Advanced Very High Resolution Radiometer; CT=cloud top; DL=deep layer; EUMET=EUMETSAT=European Organisation for the Exploitation of Meteorological Satellites; *GOES*=Geostationary Operational Environmental Satellite; JMA=Japan Meteorological Agency; MODIS=Moderate Resolution Imaging Spectroradiometer; VIIRS=Visible Infrared Imaging Radiometer Suite; VIS=visible.

The FTC-OO parameters are calculated either a priori based on previous DA cycles or inline based on the current DA cycle. A priori calculations can be done before the current GSI cycle begins. The experiments presented here are all based on a priori calculations using the AMVs and backgrounds from the previous $N_{\mathrm{cycles}} = 28$ DA cycles (i.e., the previous 7 days; a plan for inline calculations is discussed in Appendix A.4). We tuned the parameter N_{cycles} based on offline calculations. The optimal value of N_{cycles} should be as small as possible to avoid using stale observations, but must be large enough that each sample of observations is sufficiently large. In choosing N_{cycles}, we preferred a choice that had a larger reduction in variance explained, a smaller value of N_{cycles}, mean differences closer to zero, and bias corrections ($\delta\mathbf{V}$) closer to zero. In retrospect, it should also be possible to use a fixed sample size of the previous AMV observations in each subset, going back in time only far enough as needed. Experience suggests that this sample size should be in the range of thousands of observations.

3 | OBSERVING-SYSTEM EXPERIMENT (OSE) SETUP

Four OSEs were performed using initial forecast times at 0000 UTC from August 2–September 16, 2019. The control experiment is named BASE. The other experiments replaced the operational AMV observation operator with the FTC-OO observation operator and/or added assimilation of *Aeolus* winds. These experiments are as follows:

- FTC: BASE + FTC-OO;
- DWL: BASE + *Aeolus* assimilation;
- FTCDWL: BASE + FTC-OO + *Aeolus* assimilation.

Our experiments assimilate all observation types that are assimilated operationally, including many types of AMVs. The AMV types fall into one of five categories—visible, short-wave IR, long-wave IR, WV

deep layer, and WV cloud-top AMVs. These AMVs are all single-level observations at the cloud top except that WV deep-layer AMVs are generated by tracking features in WV fields retrieved under clear conditions. Spatial coverage for the entire dataset is near-global, with partial gaps at around 60°S and 60°N. These gaps are present because at these latitudes geostationary Earth orbit (GEO) sensor zenith viewing angles are too high for reliable imaging and low-Earth orbit (LEO) collocations are not close enough in time. (LEO AMVs depend on comparing imagery from separate, but close in time, overpasses.) The majority of the AMVs in our experiments, ∼80% of the total, come from the NOAA *GOES* satellites. Note that in experiments FTC and FTCDWL, the FTC-OO is applied to AMV types with sufficient sample size (> 300 per cycle). The operational observation operator is applied to AMV types with smaller samples in experiments FTC and FTCDWL and to all AMVs in experiments BASE and DWL. Additional observations assimilated in these experiments include conventional observations (from surface stations, ships, radiosondes, etc.), radio occultation observations, scatterometer winds, and radiance observations from satellites in both LEO and GEO orbits.

The OSEs conducted for this study make use of the operational FV3GFS, a global 4D ensemble variational DA and forecast system (Wang & Lei, 2014). The DA component of the FV3GFS system is the GSI (, 2021; Kleist *et al.*, 2009). Specifically, we use the NCEP global workflow v15.3 at C384L64 research resolution, a lower horizontal resolution than that used in operations, where C384 indicates there are 384 grid points along each edge of the model's cubed sphere, giving an approximate ∼25 km horizontal grid, and L64 indicates there are 64 vertical levels. In addition to this deterministic component, the FV3GFS system also includes ensemble analyses and forecasts to estimate the background-error statistics. The ensemble resolution is C192 (∼50 km), about half that of the deterministic resolution. All forecasts and analyses make use of the same 64-level vertical grid.

In our experiments the FV3GFS system cycles every 6 hours, assimilating data observed from 3–9 hours after the start of the cycle to produce the analysis valid 6 hours after the start of the cycle, that is, valid at the start of the next cycle. 10-day forecasts are generated once a day from the analysis valid at 0000 UTC. The first five days are considered a spin-up period; all results presented are for the analysis valid at 0000 UTC from August 7–September 16, 2019 and forecasts from those initial conditions. We use ECMWF operational analyses for verification—analysis and forecast "errors" in this study are really differences with respect to the assumed ECMWF "truth."

ESA and ECMWF produced the operational *Aeolus* L2B B10 Rayleigh-clear and Mie-cloudy HLOS winds that are assimilated in experiments DWL and FTCDWL. These are not assimilated operationally by FV3GFS, but we have previously extended the GSI to do so (Garrett *et al.*, 2022) and developed a total least-squares (TLS) bias correction for the *Aeolus* winds (Liu *et al.*, 2022). Other details of how the *Aeolus* winds are quality-controlled and assimilated in our experiments are given by Garrett *et al.* (2022). In this study, we did not apply our own TLS bias correction to the *Aeolus* winds, to avoid complicating the interpretation of our results.

Of particular interest in the present study are the AMV estimated observation errors (EOEs), that is, assumed standard deviations (m/s), used in the DA. DA systems filter the observation minus background (OMB) difference to create the update, that is, the optimized analysis. Here, the background is the short-term forecast estimate of the observation, that is, the observation operator evaluated for that observation. (In the context of DA or time series analysis, these differences are known as innovations.) Within the DA, the OMB differences are scaled by the EOEs. The AMV EOEs in the GSI are based on estimates tabulated by pressure (in 50-hPa thick bins from 100 to 1000 hPa spaced by 50 hPa) and AMV type. The GSI adjusts the tabulated EOE observation by observation, based on quality-control information, observation density, and representativeness factors. Since the GSI does not currently thin AMVs available to it, and horizontal error correlations are ignored when assimilating these observations, the EOEs are inflated for those AMV types, notably the *GOES* AMVs, that are present in the database with high spatial density. Figure 1 presents the EOEs as used in experiment BASE for the AMVs assimilated in our experiments. EOEs are plotted in Figure 1 whenever there is at least an average of 800 AMVs for that subset per cycle. Note that the EOEs in Figure 1 vary substantially by type and vertical level. For example, the EOEs for *GOES* deep-layer water-vapor winds increase from 4 to 14 m/s from the mid to upper troposphere. Because the FTC-OO reduces representativeness errors, we adjust the EOEs for observations for which the OMB differences are calculated with the FTC-OO. (For details, see the discussion of AMV OMB wind differences in Section 4.)

4 | FTC-OO IMPACT ASSESSMENTS

In all statistical analyses presented here, the verification period includes all 0000 UTC initial times from August 7–September 16, 2019. The forecast verification sample is 41 10-day forecasts, and the analysis verification sample is 164 cycles. The boundary between tropical and extratropical verification regions is 20°. The results are verified with respect to the ECMWF operational analysis. ECMWF verification has the advantage that the ECMWF

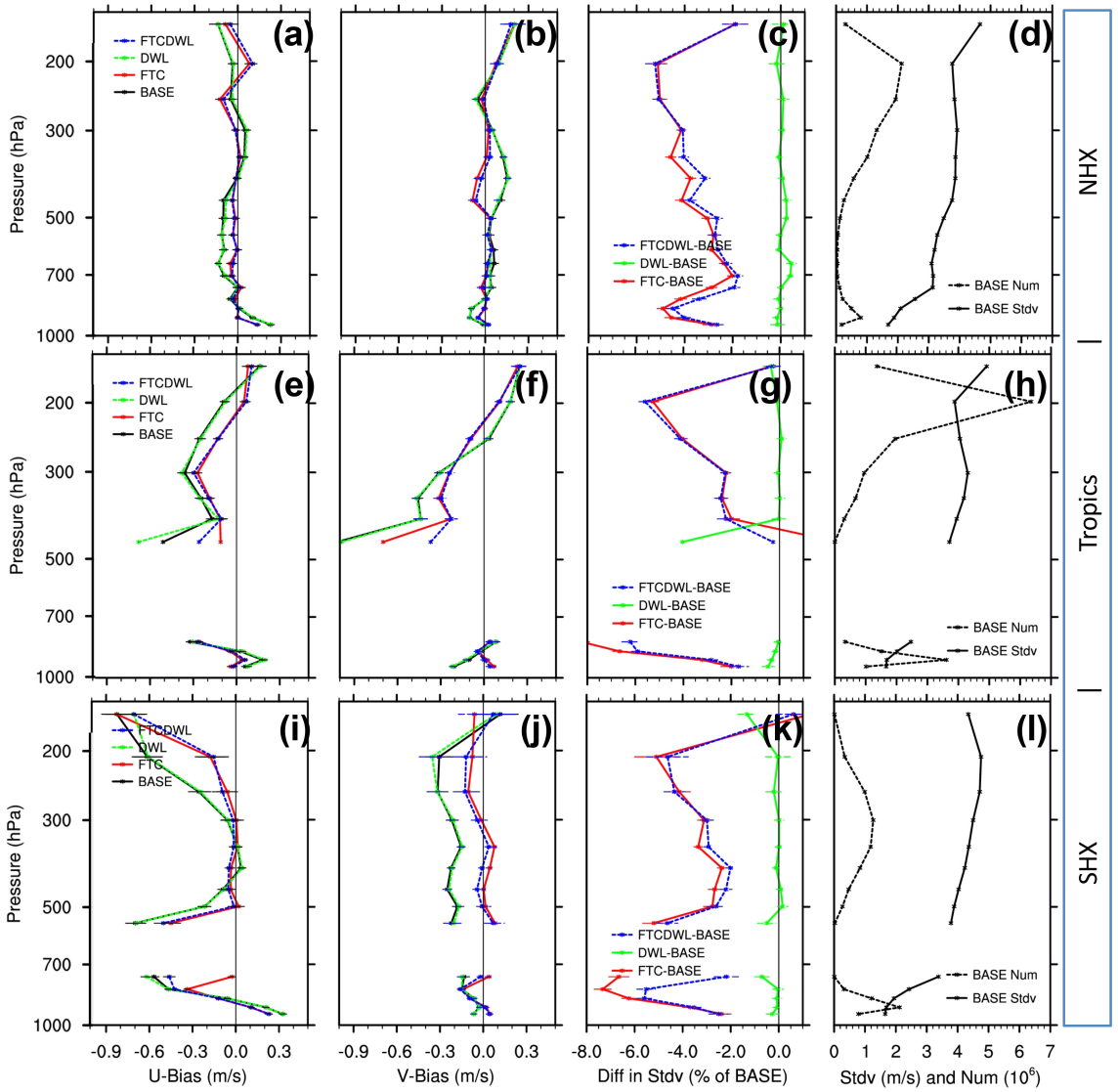

FIGURE 2 OMB statistics for all AMVs assimilated in the four experiments (different lines identified in the legend) during the study period, for all locations in the three geographic regions: (a–d) NHX, (e–h) Tropics, and (i–l) SHX, from top to bottom, for all AMV types. The statistics from left to right are (a,e,i) the mean *u*-wind OMB (m/s), (b,f,j) the mean *v*-wind OMB (m/s), (c,g,k) the vector standard deviation (Stdv in the figure labels) of the wind OMB (i.e., the mean speed of the horizontal vector wind OMB) expressed as a percentage with respect to BASE (%), and (d,h,l) the sample size (Num in the figure labels) in units of 10^6 counts for the experimental period and the vector standard deviation of the wind OMB for BASE (m/s), all displayed as a function of reported observation pressure (hPa). Estimated 95% confidence intervals (horizontal bars) are calculated from the sample standard deviation of the statistic.

analyses are independent of the NOAA analyses. However, it should be noted that the ECMWF system is most like the BASE configuration: unlike the ECMWF operational analysis in 2019, experiments FTC and FTCDWL use the FTC-OO and experiments DWL and FTCDWL assimilate *Aeolus* winds. In the results presented here, correlations between analyses and forecasts are not accounted for, but we discuss the estimated impact of such correlations at the end of this section.

OMB statistics for the AMVs should indicate reductions in both random errors and biases due to the FTC-OO. These reductions are seen in Figures 2 and 3 below and indicate where FTC-OO is having the biggest impact on adjusting the AMVs, but such reductions do not by themselves indicate that the FTC estimates of wind are closer to the truth than the original AMVs. For this we turn, later in this section, to OMB statistics for independent wind datasets and forecast-error statistics.

Figure 2 shows the variation of the AMV OMB statistics for all AMV types assimilated in our experiments. For experiments FTC and FTCDWL, the OMB values are calculated using the FTC-OO, while for experiments BASE and DWL the operational observation operator is used. Most AMVs are observed in the upper troposphere (at

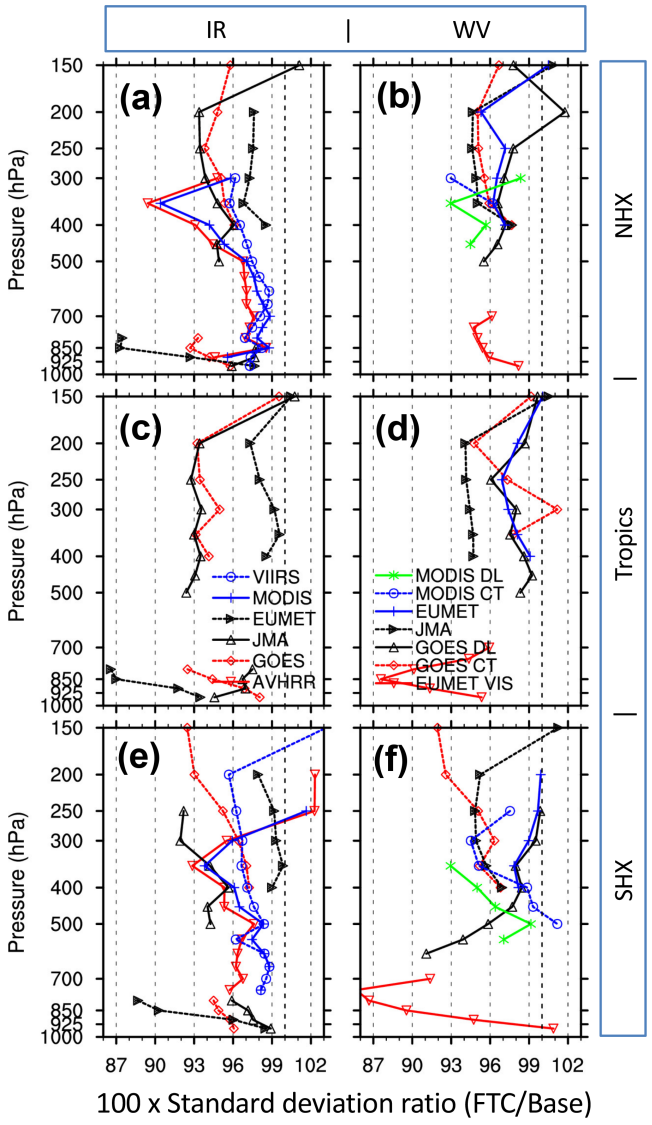

FIGURE 3 The ratio of the horizontal vector wind standard deviations of the wind difference (i.e., OMB) as a function of pressure (hPa) of the FTC-OO with respect to those of the operational observation operator, for the study period, for each AMV type (different lines identified in the legend), for all locations in the three geographic regions: (a,b) NHX, (c,d) Tropics, and (e,f) SHX, from top to bottom, for (a,c,e) the IR AMV types and (b,d,f) WV and visible AMV types. Figure acronyms: CT=cloud top; DL=deep layer; VIS=visible.

pressures less than 500 hPa) and near the top of the planetary boundary layer (panels d,h,l). Gaps in the figure panels (other than d,h,l) correspond to no observations, and statistics at the edges of these gaps may not be reliable. The u- and v-wind component (hereafter u- and v-wind) mean OMB differences are small (of the order of 0.1 m/s) in the NHX (panels a,b), but larger (of the order of 0.25–0.5 m/s) in the Tropics (panels e,f) and SHX (panels i,j). Mean OMB differences are small between experiments, but the mean OMB values for the FTC and FTCDWL experiments are consistently closer to zero than those of the BASE and DWL experiments. This is especially the case in the SHX: in Figure 2i,j, the u- and v-wind mean differences in the upper troposphere (at pressures less than 500 hPa) are closer to zero by ~0.25 m/s. Since the FTC parameters are a result of fitting samples in thick pressure layers (below 800 hPa, between 800 and 450 hPa, and above 450 hPa), we are not guaranteed improvements for sections of these thick layers. As a result, FTC-OO does not reduce the bias at a few levels in Figure 2. In all experiments and all regions, the vector standard deviation of the OMBs varies from about 1.5 m/s to 5 m/s as height increases (panels d,h,l). The FTC-OO reduces the vector standard deviation of the OMBs by roughly 4% (panels c,g,k).

Figure 3 shows the ratio of the vector standard deviation of the OMB wind difference between the FTC-OO and the operational observation operator for the different AMV types in the different regions. This ratio, φ, is expected to be less than one, indicating a positive impact. (Values larger than one indicating a negative impact occurs, but are very close to one.) In the figure, φ is always in the range [0.85, 1.02]. The smallest values (i.e., those corresponding to the greatest reduction in variance) are below 0.90 for low-level European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) AMVs in the SHX and Tropics. In experiments FTC and FTCDWL, operational EOEs (Figure 1) are multiplied by φ from Figure 3. Both the EOEs and φ values are tabulated by region, pressure level, and AMV type. Note that in Figure 3 the statistics are evaluated during the study period for the sample of all AMVs passing quality control in the BASE experiment—both in the BASE experiment and in a preliminary FTC experiment that did not adjust the EOEs.

Figures 4 and 5 show typical (mean) values of the AMV parameters Δz and h for the different AMV types that were assimilated in experiment FTC. First note the large variation in mean daily sample sizes, from about 0.75 million total *GOES* AMVs assimilated per day to only ~15,000 MODIS AMVs assimilated per day. The mean layer depth Δz is typically 100–200 hPa. Except for VIIRS IR AMVs, mean height corrections are all negative and for the most part act to lower the reported height by approximately half a vertical grid (i.e., $h \sim -12.5$ hPa); note that this correction is negligible for EUMETSAT and Japan Meteorological Agency (JMA) WV AMVs. The standard deviation across time for Δz is generally small (order of 20 hPa) except for the MODIS types, which have small samples. The standard deviation across time for h is also of order 20 hPa.

The mean u- and v-wind bias corrections (δu, δv) are mostly negative and typically of magnitude 0.5 m/s or less (Figure 6). These are components, so the mostly negative

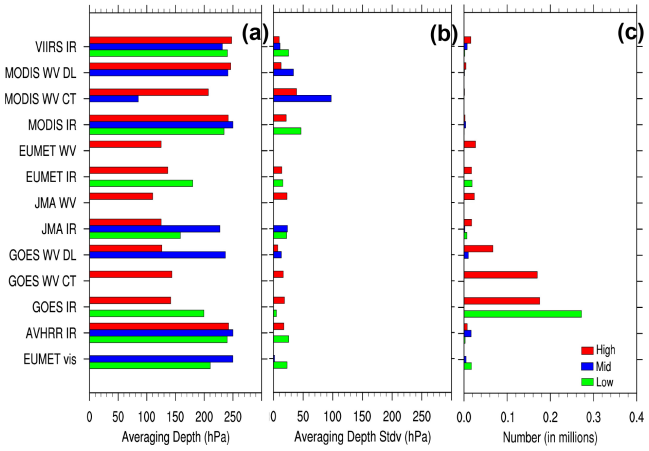

FIGURE 4 (a) The mean FTC parameter averaging depth (Δz, hPa), (b) its standard deviation across analyses, and (c) daily sample size (millions) for high, middle, and low winds (colors), evaluated for all AMVs of each type (vertical axis) assimilated in the FTC experiment during the study period, for all locations. Figure acronyms: CT=cloud top; DL=deep layer; vis=visible; mid=middle; Stdv=standard deviation.

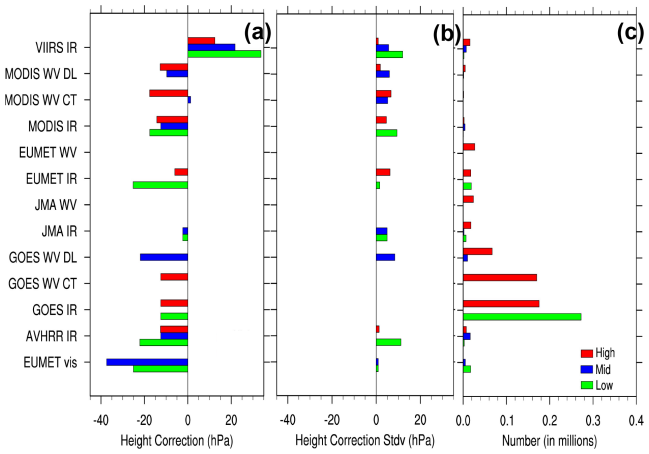

FIGURE 5 As in Figure 4, but for FTC parameter height correction (h, hPa).

values indicate that the FTC acts to make the estimated AMVs more northerly and easterly. Figure 6 shows an approximately 0.05 m/s positive shift (and reduced magnitudes) in the u-wind bias corrections in experiment FTCDWL compared with experiment FTC for each geostationary AMV type. This is small but systematic and only appears in the u-wind results. This is consistent with the fact that the *Aeolus* wind is essentially a u-wind in the Tropics due to the observing geometry, and may indicate some inconsistencies between the FV3GFS and *Aeolus* observations.

At very short forecast times, comparisons versus an analysis can obscure the true impact of observations due to errors in the verification. Therefore, we examined OMB statistics for an independent and reliable dataset, namely,

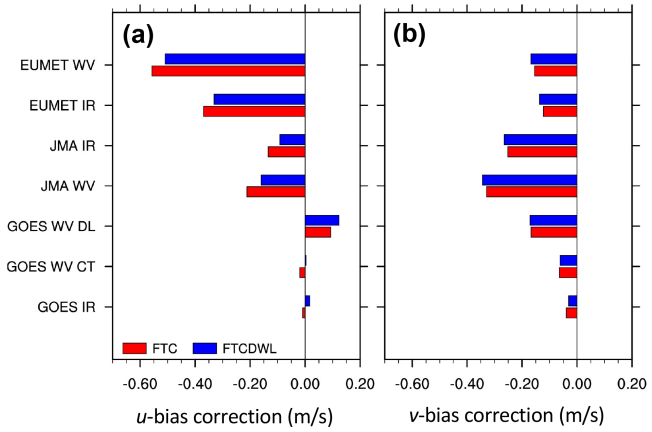

FIGURE 6 The u- and v-wind bias correction terms (δu, δv, m/s) in the Tropics for experiment FTC (red) and experiment FTCDWL (blue) calculated by the FTC optimization during the study period for vertical levels above 400 hPa for each geostationary AMV type.

conventional wind observations made by radiosondes and aircraft (Figure 7). In the NHX, the OMB standard deviation differences normalized by the BASE values are indistinguishable from each other and of magnitude less than 1%. In the Tropics and SHX, the experiments assimilating *Aeolus* winds reduce the OMB standard deviation differences by 1%–2%, except for u-wind in the tropical upper troposphere–lower stratosphere (UTLS), where reductions are larger, reaching 4% at 100 hPa. In the SHX, v-wind reductions are larger for experiment DWL than for experiment FTCDWL. In the Tropics and SHX, the results for u-wind are neutral with differences of magnitude less than 1% and the results for v-wind are negative with differences of order 1%.

While OMB statistics versus reliable assimilated observations, as in Figure 7, are not influenced by biases and random errors as can occur in an NWP verification of short-term forecasts, analyses like Figure 7 suffer from small or non-existent samples in parts of the spatial domain. Also, in the case of radiosondes, observations are mostly available only at the two main synoptic times of 0000 and 1200 UTC. Therefore, additional OMB analyses using satellite observations could be useful. For example, one could use other independent wind observations from lidars and scatterometers, or other observations, such as humidity-sensitive radiances, that are sensitive to the effects of wind observations on humidity transport.

Figure 8 shows the zonally averaged u-wind mean analysis for BASE (Figure 8a) and the differences comparing FTC, DWL, and FTCDWL with BASE. There are distinctive difference patterns of magnitudes up to 0.5 m/s in the tropical UTLS. In experiment FTC (Figure 8b), the mean u-wind increases by as much as

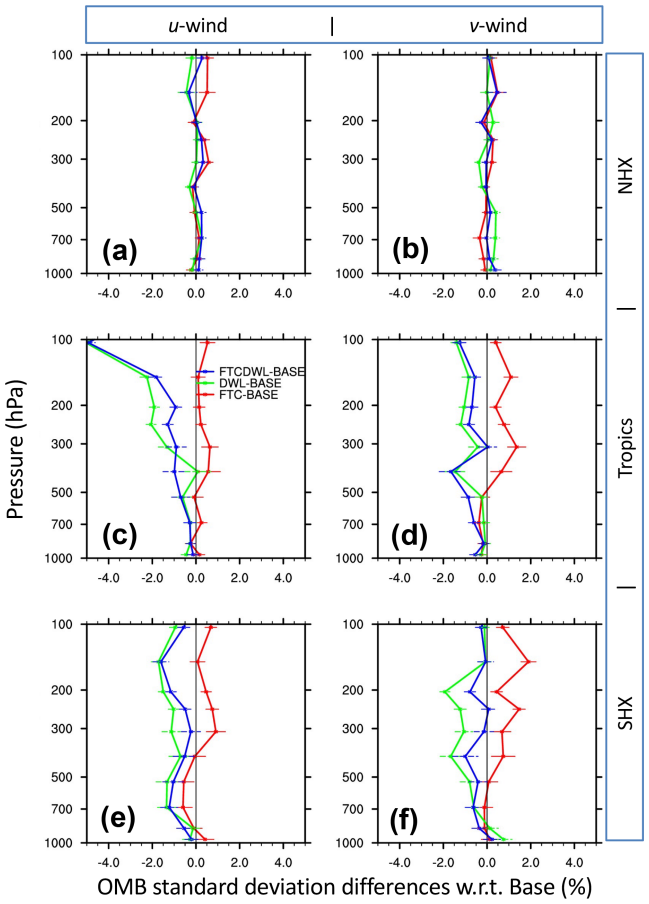

FIGURE 7 OMB impacts with respect to (w.r.t.) BASE for all conventional wind observations (radiosondes and aircraft). The panels show the standard deviation of the wind component OMB differences expressed as a percentage with respect to BASE (%) for the other three experiments (different lines identified in the legend) during the study period, for all locations in the three geographic regions: (a,b) NHX, (c,d) Tropics, and (e,f) SHX, from top to bottom, for the (a,c,e) *u*-wind and (b,d,f) *v*-wind components, displayed as a function of reported observation pressure (hPa). Estimated 95% confidence intervals (horizontal bars) are calculated from the sample standard deviation of the statistic.

0.25 m/s centered on the Equator at 200 hPa compared with BASE. In experiment DWL (Figure 8c), the mean *u*-wind decreases by as much as 0.5 m/s centered on the Equator at 100 hPa compared with BASE. Rennie *et al.* (2021) also reported large mean *u*-wind analysis differences in their OSE results at 150 and 100 hPa (their fig. 11). The patterns seen in Figure 8b,c are merged in Figure 8d for experiment FTCDWL.

Figure 9 shows the root-mean-square errors (RMSE, taking the ECMWF analysis as the truth) for heights, winds, and temperatures in the NHX, SHX, and Tropics for FTC, DWL, and FTCDWL compared with BASE in the form of forecast skill verification scorecards. The FTC/BASE comparison (Figure 9a) shows statistically significant improvement in forecast skill for geopotential height in the Tropics throughout the troposphere out to day 6. This is statistically significant, but not meteorologically important. Otherwise, this scorecard shows decreased skill in the SHX day-1 wind forecast in the UTLS and some increased skill in the day-6 wind forecast in the Tropics. The DWL/BASE comparison (Figure 9b) shows much greater improvements, especially for wind and temperature, for all levels in the Tropics at short forecast times, and for longer forecast times at upper levels. For example, the 100-hPa wind and temperature forecasts are significantly improved out to days 10 and 6, respectively. In the Tropics, the DWL experiment does not improve geopotential height forecasts except for the 100-hPa level at days 1 and 3. In the SHX there are very significant day-1 improvements and significant day-3 improvements for all variables except at lower levels. In the NHX there are minor improvements at day 1. The FTCDWL/BASE comparison (Figure 9c) combines all the improvements seen in the two other comparisons. Comparing Figure 9c with Figure 9b, we see improvements in the geopotential height and temperature forecasts in the troposphere in the Tropics and SHX. The scorecard for FTCDWL compared with DWL is not shown but is like the scorecard for FTC compared with BASE (Figure 9a).

Figure 10 visualizes the evolution of the differences (each experiment minus BASE) in the horizontal vector wind RMSE (ΔRMSE) in the SHX. The RMSE of the BASE experiment increases with forecast time to over 20 m/s at 300 hPa. The impacts in experiments DWL and FTCDWL are positive almost everywhere and increase slowly with forecast time, reaching over 0.5 m/s in the mid to upper troposphere ($400 < p < 200$). In experiment FTC, the ΔRMSE are initially negative until about day 4 and then increase rapidly, reaching over 0.3 m/s in a pattern like the DWL experiments. However, referring to Figure 9, the largest differences in Figure 10 are not statistically significant. Instead, Figure 9 shows that the most significant differences are at short forecast times, where the differences are small in magnitude but large as a percentage of the RMSE of the BASE experiment.

Plots of 500-hPa height forecast RMSE (Figure 11) for the Tropics and SHX show different patterns of positive impacts. In the Tropics, the impacts are marginal (Figure 11a), with significantly positive impacts for the DWL experiments out to day 8 (Figure 11b). In the SHX, improvements are greatest for FTCDWL at day 8 and beyond (Figure 11c), but the impacts are significantly positive for the FTC experiments out to day 7 (Figure 11d).

The difference summary assessment metrics (SAMs) combine a collection of forecast skill score differences (experiment minus BASE; Figure 12). Values above 0.0 represent an improvement of the forecast versus the BASE.

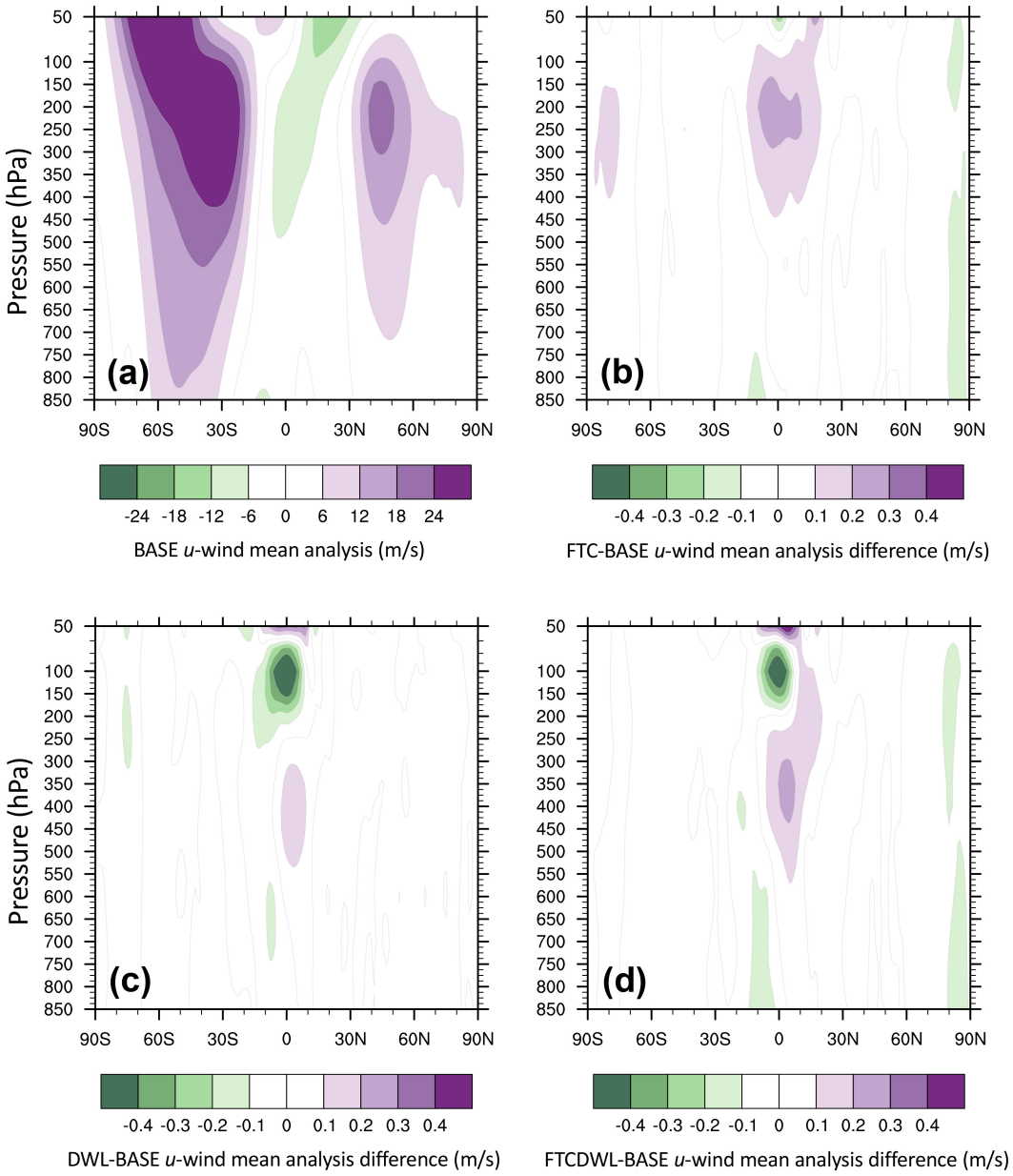

FIGURE 8 (a) The zonally and temporally averaged *u*-wind analysis (m/s) for BASE and differences with respect to BASE for (b) FTC, (c) DWL, and (d) FTCDWL.

A value of 0.02, for example, indicates the average normalized statistic is better (greater) than BASE by 0.02. Under the null hypothesis that there are no differences, all SAMs would be 0.5, so a 0.02 improvement can be considered a 4% improvement in normalized scores. Details are in Hoffman *et al.* (2018). The SAMs displayed in Figure 12 are consistent with the previous discussion of individual skill scores. Overall (Figure 12a), DWL and FTCDWL show an improvement of 4% and FTC shows a barely significant improvement of 1%. In the more granular difference SAMs, for different variables, forecast times, and regions (Figure 12b,c,d), we see that the DWL and FTCDWL impacts are similar, except that FTCDWL is noticeably better in the Tropics and for geopotential height. In these panels, FTC compared with BASE has a large improvement for geopotential height and minor improvements for the Tropics and SHX and for forecast times of days 6 and 7.

Estimating the impact of correlations on the error bars of Figures 11 and 12 is itself subject to sampling errors. As an approximation, we turn to results of Hoffman *et al.* (2018, their appendix d), which are based on three years of forecasts from three leading global operational NWP centers during 2015–2017. They found that "the combined [sample size] reduction factor for all dimensions is approximately 0.09 [appropriate for Figure 12a] and for

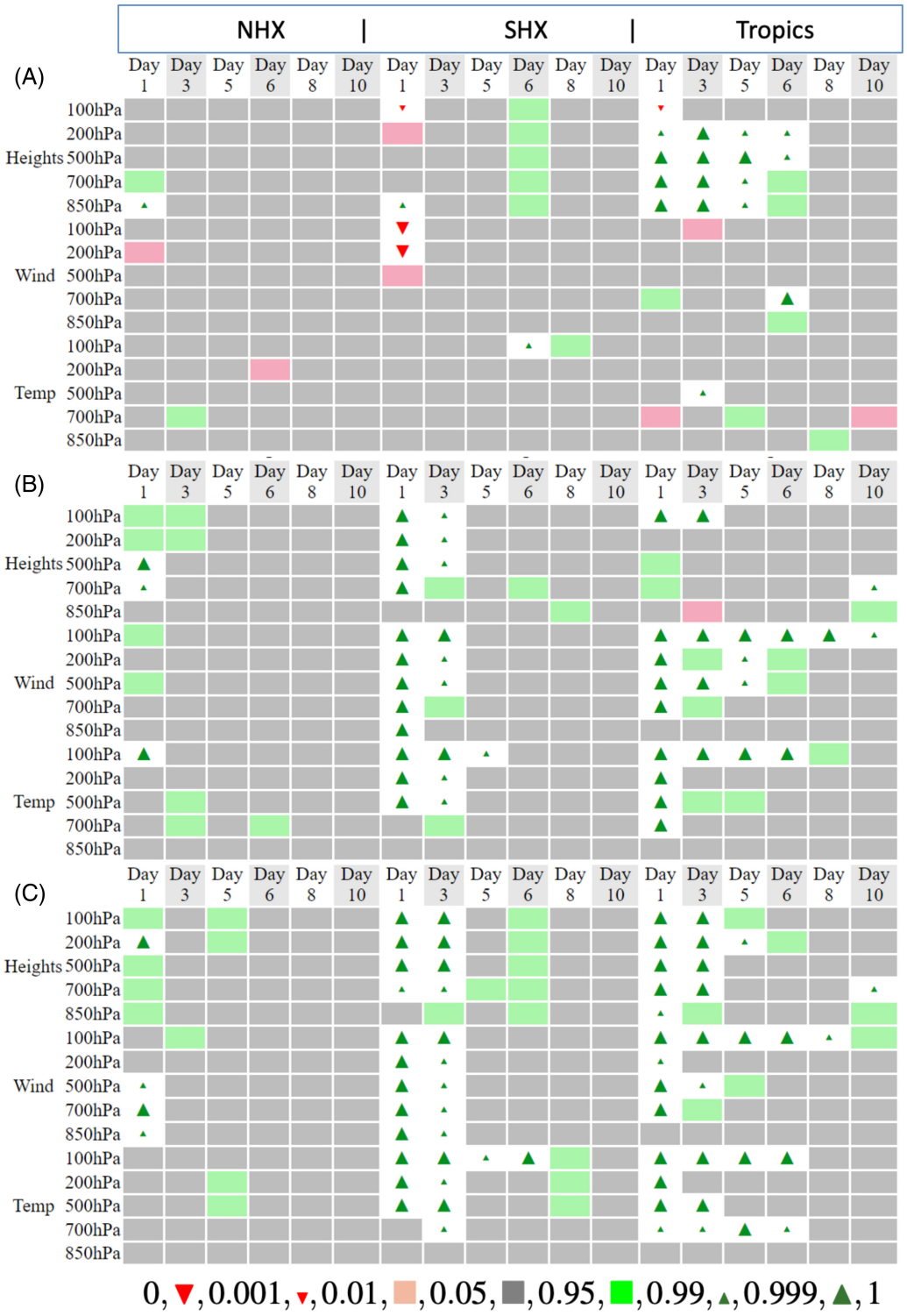

FIGURE 9 FV3GFS (day-1 to day-10) forecast skill verification scorecards comparing (from top to bottom) (a) FTC, (b) DWL, and (c) FTCDWL with BASE. Note that only RMSE heights, winds, and temperatures for NHX, SHX, and Tropics are extracted to create this figure. The symbols and colors indicate the probability that an experiment is better than BASE. As shown below the scorecard, the green symbols (from left to right) indicate that an experiment is better than BASE at the 95%, 99%, and 99.9% significance levels, respectively, while the red symbols indicate that an experiment is worse than BASE at the 99.9%, 99%, and 95% significance levels, respectively. Gray indicates no statistically significant differences.

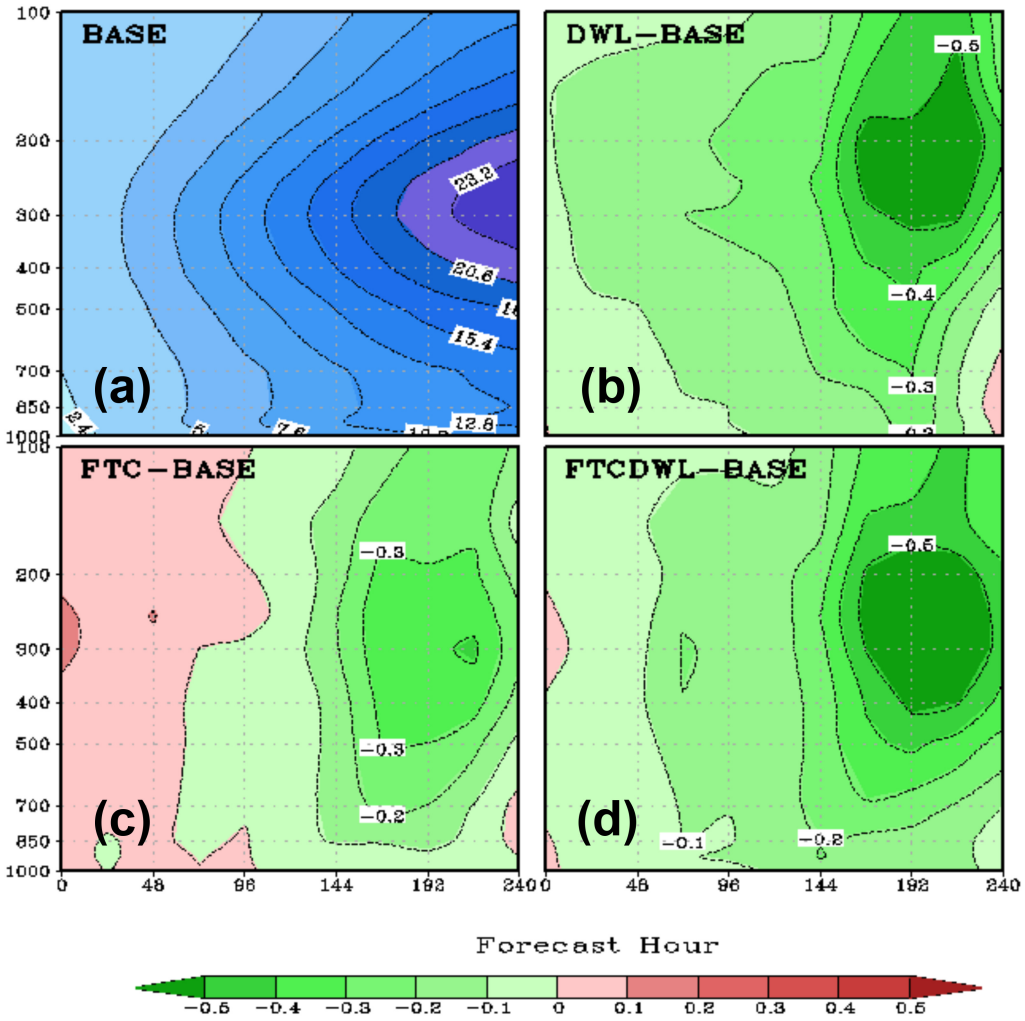

FIGURE 10 Southern Hemisphere extratropics forecast wind RMSE statistics (m/s) in time (h) versus pressure (hPa) cross-sections. The first panel (a) shows the forecast RMSE in the BASE experiment. The differences in the forecast RMSE with respect to BASE (ΔRMSE) are shown for (b) DWL, (c) FTC, and (d) FTCDWL. Note that increasing darker green shades correspond to increasing negative values and, hence, increasing positive impacts.

all verification time dimensions about 0.66 [appropriate for Figure 11]." The error bars should then be larger by a factor of one divided by the square root of this reduction factor, or 1.2 for Figure 11 and 3.3 for Figure 12a. Equivalently, the reported *p*-level of 95% could be adjusted. For figures like Figure 11 in which there are forecasts once daily, the *p*-level is actually approximately 90%.

5 | SUMMARY AND CONCLUSIONS

This study examined the OSE impact of a prototype implementation of the feature track correction observation operator (FTC-OO) in the FV3GFS. A goal of the FTC-OO is to reduce AMV error correlations, which are difficult to treat in DA systems. There are multiple potential sources of systematic errors (biases hereafter) in AMVs (as discussed in the Introduction) and in DA systems generally (Dee, 2005). Biases are normally assumed to be zero in operational DA systems. If present and unaccounted for, systematic errors induce error correlations. Eyre *et al.* (2022) point out that "Until recently, observation errors have all been treated as uncorrelated, with a diagonal R matrix. Where this has been known to be a poor approximation, the problem has been addressed empirically, either by inflating the error variances or thinning the observations or both. It is currently still the general practice to treat horizontally correlated errors by thinning." (However, in current operational use and in our experiments, data thinning of AMVs is disabled in the GSI.) While background-error correlations (the **B** matrix) are treated with varying degrees of sophistication, it is always assumed that the observation and background errors are uncorrelated. The FTC method and similar methods that are based on OMB values might seem to violate the cardinal assumption made in all DA systems of zero background minus observation-error covariances.

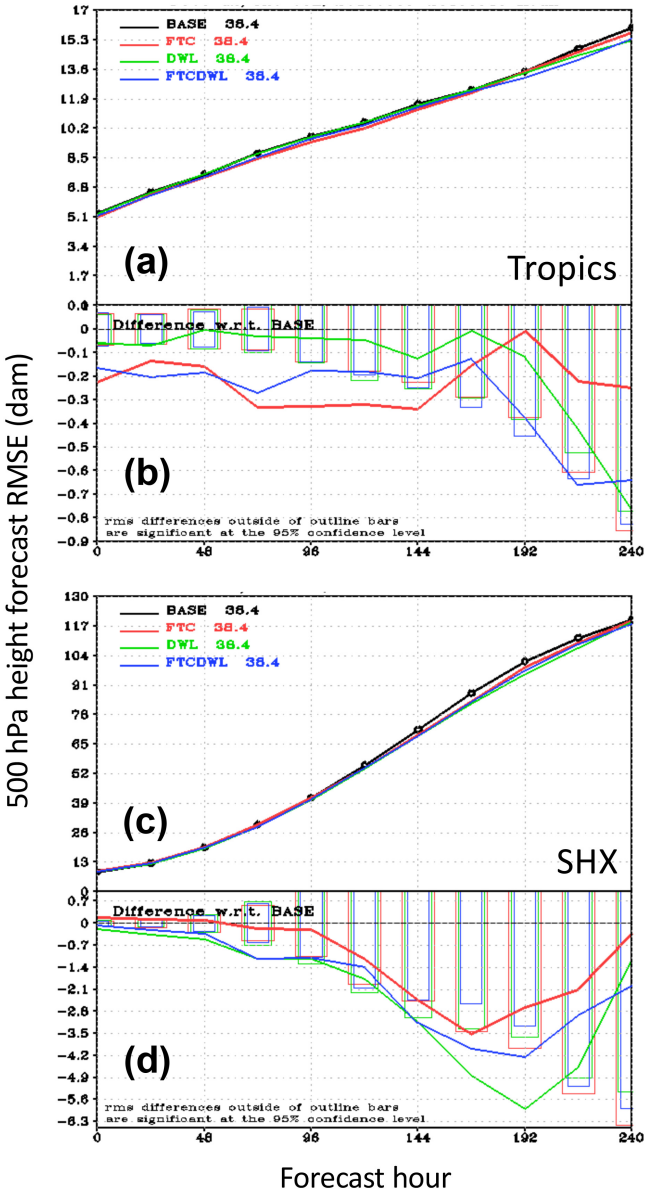

FIGURE 11 (a,c) RMSE of FV3GFS 0- to 10-day 500-hPa height forecasts (decameters) in the BASE, FTC, DWL, and FTCDWL experiments (colors) and RMSE differences with respect to (b, d) BASE in (a, b) the Tropics and (b, d) SHX. Note that (a,c) a decrease in RMSE or (b,d) a negative RMSE difference represents a positive impact. In the RMSE difference panels (b,d), the bars represent the 95% confidence interval.

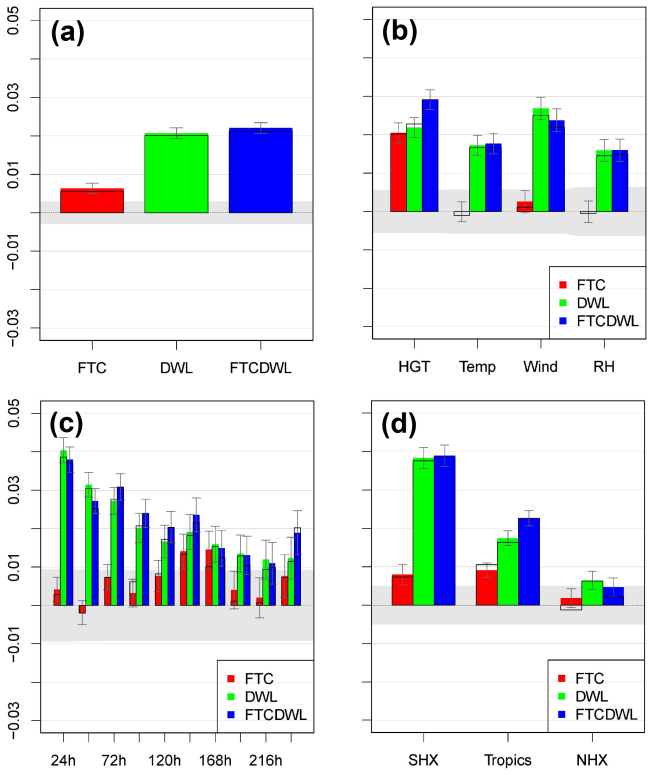

FIGURE 12 The difference Summary Assessment Metrics (SAMs) for FTC, DWL, and FTCDWL versus BASE experiments (colors). The SAMs are shown for (a) all metrics combined, for different (b) forecast variables (geopotential height (HGT), temperature (Temp), vector wind (Wind), and relative humidity (RH)), (c) forecast lead times, and (d) different geographic regions. The gray horizontal areas centered on the zero line indicate the 95% confidence interval under the null hypothesis that there is no difference between experiments for this metric. In addition, the sample-estimated 95% confidence interval is indicated by error bars at the ends of the color bars. Two normalizations are used, the empirical cumulative density function (ECDF, color-shaded) and rescaled-minmax normalization (black outline).

However, this assumption is not violated, because these "corrections" are based on past data or are formulated by augmenting the state vector, with the result that the bias corrections are part of the observation operator. At the same time, it must be recognized that, without enough observations that are highly accurate (i.e., without bias), bias-correction schemes in the context of DA run the risk of correcting relative biases between model and observations without reducing the true analysis biases.

We compared experiment FTC, which implemented the FTC-OO, with experiment BASE, our control experiment. In FTC, the FTC-OO is applied to all AMV types listed in Figure 1. The goal of using the FTC is to characterize AMV errors better and thereby improve their usefulness in DA systems. As seen in Equation (2), FTC attempts to match AMVs and model winds better by applying constant and speed-dependent corrections to a vertical average of the model winds in the FTC-OO. Compared with experiment BASE, experiment FTC reduces data misfit and improves the FV3GFS forecast skill, especially in the SHX and Tropics.

We also conducted OSEs with the addition of *Aeolus* wind observations, to examine the sensitivity of the FTC method to other wind observations. The experimental setups for experiments DWL and FTCDWL are

identical to experiments BASE and FTC, respectively, except for the addition of *Aeolus* wind observations. The comparisons of experiment FTC with BASE and experiment FTCDWL with DWL showed similar impacts on forecast skill. In both experiments FTC and FTCDWL, AMV OMBs show smaller mean and random differences in the upper troposphere of the SHX (Figure 2).

Overall, improvements in forecast skill due to the FTC-OO were small in our experiments and at best were just marginally significant. The main improvement of experiment FTC compared with experiment BASE (Figure 9a) is in forecast skill for geopotential height in the Tropics, which, while statistically significant, is not meteorologically important. However, the history of the evolution of NWP forecast skill has been incremental and small improvements have accumulated. In fact, the decision to implement a new method operationally often requires running impact experiments for much longer periods than we were able to use in our experiments. In the end, a new method with a good science basis may be implemented operationally so long as the impacts are neutral or better.

Unfortunately, the hypothesis that the *Aeolus* observations would enhance the impact of the FTC-OO was not borne out in our experiments; that is, the *Aeolus* observations alone have a significant positive impact, but the use of the FTC-OO method in the presence of the *Aeolus* observations only produces an improvement that parallels that seen without *Aeolus* winds, that is, improvements for tropical geopotential height forecasts. As described in the Introduction, we anticipated that *Aeolus* wind observations could anchor AMVs assimilated with the FTC method in a way similar to how RO observations anchor radiance observations assimilated with VarBC. However, currently available *Aeolus* wind observations, especially the Rayleigh wind observations, have larger than anticipated random errors and noticeable biases, which, as discussed at the end of Section 3, so far have only been corrected via comparison with an NWP system. First, the Rayleigh winds are subjected to two bias corrections based on ECMWF backgrounds, the Rayleigh–Brillouin correction and the *Aeolus* DWL main mirror temperature-dependent bias correction (Rennie *et al.*, 2021; Weiler *et al.*, 2021). Liu *et al.* (2023) showed the benefit of using FV3GFS backgrounds in these corrections when assimilating the Rayleigh winds in the FV3GFS system. Second, biases remain that vary with latitude, height, and wind speed, which have been removed with a TLS bias correction in other experiments (Liu *et al.*, 2022). It is expected that *Aeolus* follow-on missions will provide higher vertical resolution and higher accuracy wind profiles, which may then provide sufficient anchoring for the FTC method to realize the synergy of DWL and AMV observations as hypothesized.

The findings of this study suggest the potential positive impact of implementing the FTC-OO in the NOAA NWP system. Several possible research opportunities, including extensions of the implementation of the FTC-OO, are listed below.

- Subsets by AMV type could be replaced with finer subsets by AMV subtype: for example, stratify each *GOES* AMV type by satellite (e.g., *GOES* East and *GOES* West).
- Conduct OSEs for BASE+Mie and FTC+Mie, where "+Mie" indicates assimilation of *Aeolus* Mie-cloudy winds, without any bias correction and without *Aeolus* Rayleigh winds. These OSEs would test the synergy between *Aeolus* winds assimilated without bias corrections and AMVs assimilated with the FTC-OO. In fact, there is a bias correction, which accounts for the effect of fluctuations of temperature across the main mirror of the *Aeolus* DWL, that is applied to the Mie winds, but according to Weiler *et al.* (2021) it is an order of magnitude smaller than this correction is for the Rayleigh winds.
- Implement an "inline" version of the FTC method using AMV samples from the current DA cycle iteratively within the GSI (as described in Appendix A.4).
- Averaging weights with different shapes, for example, a triangle or a truncated Gaussian hill, could be used (as described in Appendix A.1).

AUTHOR CONTRIBUTIONS

Ross N. Hoffman: conceptualization; formal analysis; methodology; software; visualization; writing – original draft. **Hui Liu:** conceptualization; data curation; methodology; resources; software; visualization; writing – original draft. **Katherine E. Lukens:** conceptualization; data curation; investigation; writing – original draft. **Kevin Garrett:** conceptualization; funding acquisition; project administration; resources; supervision. **Kayo Ide:** conceptualization; formal analysis; investigation; project administration; supervision.

ACKNOWLEDGEMENTS

This research was funded by grants from the National Oceanic and Atmospheric Administration (NOAA) National Environmental Satellite, Data, and Information Service (NESDIS) to NOAA cooperative institutions (CIs) at the University of Maryland (UMD), College Park, Maryland, USA. At the UMD, authors R. N. Hoffman, Hui Liu, K. E. Lukens, and K. Ide received support from NOAA/NESDIS grants NA14NES4320003 to the Cooperative Institute for Climate and Satellites (CICS) and NA19NES4320002 to the Cooperative Institute for Satellite Earth System Studies (CISESS). CICS was and CISESS is administered by the Earth System Science

Interdisciplinary Center (ESSIC) at UMD. The authors acknowledge support and guidance from the NOAA/NESDIS Office of Projects, Planning, and Acquisition (OPPA) Technology Maturation Program (TMP). In 2023 OPPA became the Office of Space Weather Observations (SWO).

The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the authors and do not necessarily reflect those of NOAA or the U.S. Department of Commerce.

DATA AVAILABILITY STATEMENT

The *Aeolus* L2B Earth Explorer data used in this study are publicly available and can be accessed via the ESA *Aeolus* Online Dissemination System (https://aeolus-ds.eo.esa.int/oads/access/). The data related to the assimilation experiments are not publicly available due to the huge volume of data. We will try to provide access to the data upon request.

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ENDNOTE

[1]https://earth.esa.int/eogateway/instruments/aladin/processor-releases, accessed June 15, 2024.

ORCID

Ross N. Hoffman https://orcid.org/0000-0002-4962-9438
Hui Liu https://orcid.org/0000-0002-7959-0984
Katherine E. Lukens https://orcid.org/0000-0003-3469-8171
Kevin Garrett https://orcid.org/0000-0002-7444-4363
Kayo Ide https://orcid.org/0000-0001-5789-9326

REFERENCES

Cordoba, M., Dance, S.L., Kelly, G.A., Nichols, N.K. & Waller, J.A. (2017) Diagnosing atmospheric motion vector observation errors for an operational high resolution data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 143(702), 333–341. Available from: https://doi.org/10.1002/qj.2925

Cucurull, L., Anthes, R.A. & Tsao, L.-L. (2014) Radio occultation observations as anchor observations in numerical weather prediction models and associated reduction of bias corrections in microwave and infrared satellite observations. *Journal of Atmospheric and Oceanic Technology*, 31(1), 20–32. Available from: https://doi.org/10.1175/jtech-d-13-00059.1

Dee, D.P. (2005) Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613), 3323–3343. Available from: https://doi.org/10.1256/qj.05.137

Eyre, J.R., Bell, W., Cotton, J., English, S.J., Forsythe, M., Healy, S.B. et al. (2022) Assimilation of satellite data in numerical weather prediction. Part II: Recent years. *Quarterly Journal of the Royal Meteorological Society*, 148(743), 521–556. Available from: https://doi.org/10.1002/qj.4228

Garrett, K., Liu, H., Ide, K., Hoffman, R.N. & Lukens, K.E. (2022) Optimization and impact assessment of Aeolus HLOS wind assimilation in NOAA's global forecast system. *Quarterly Journal of the Royal Meteorological Society*, 148(747), 2703–2716. Available from: https://doi.org/10.1002/qj.4331

Healy, S.B. (2008) Forecast impact experiment with a constellation of GPS radio occultation receivers. *Atmospheric Science Letters*, 9(3), 111–118. Available from: https://doi.org/10.1002/asl.169

Hoffman, R.N., Kumar, V.K., Boukabara, S.-A., Ide, K., Yang, F. & Atlas, R. (2018) Progress in forecast skill at three leading global operational NWP centers during 2015-2017 as seen in Summary Assessment Metrics (SAMs). *Weather and Forecasting*, 33(6), 1661–1679. Available from: https://doi.org/10.1175/waf-d-18-0117.1

Hoffman, R.N., Lukens, K.E., Ide, K. & Garrett, K. (2022, 742) A collocation study of atmospheric motion vectors (AMVs) compared to Aeolus wind profiles with a feature track correction (FTC) observation operator. *Quarterly Journal of the Royal Meteorological Society*, 148, 321–337. Available from: https://doi.org/10.1002/qj.4207

Kalnay, E. (2002) *Atmospheric Modeling, Data Assimilation, and Predictability*. Cambridge: Cambridge University Press, 364 pp. Available from: https://doi.org/10.1017/CBO9780511802270

Kleist, D.T. & Coauthors. (2021) NCEP operational global data assimilation upgrades: from versions 15 through 16. In: *Special Symposium on Global and Mesoscale Models*. Boston, MA: Virtual, American Meteorological Society. Available from: https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/378554

Kleist, D.T., Parrish, D.F., Derber, J.C., Treadon, R., Wu, W.-S. & Lord, S. (2009) Introduction of the GSI into the NCEP global data assimilation system. *Weather and Forecasting*, 24(6), 1691–1705. Available from: https://doi.org/10.1175/2009waf2222201.1

Lee, E., Todling, R., Karpowicz, B.M., Jin, J., Sewnath, A. & Park, S.K. (2022) Assessment of Geo-Kompsat-2A atmospheric motion vector data and its assimilation impact in the GEOS atmospheric data assimilation system. *Remote Sensing*, 14(21), 5287. Available from: https://doi.org/10.3390/rs14215287

Lee, S. & Song, H.-J. (2017) Quantifying the inflation factors of observation error variance for COMS and MTSAT atmospheric motion vectors considering spatial observation error correlation. *Quarterly Journal of the Royal Meteorological Society*, 143(707), 2625–2635. Available from: https://doi.org/10.1002/qj.3113

Li, J., Li, J., Velden, C., Wang, P., Schmit, T.J. & Sippel, J. (2020) Impact of rapid-scan-based dynamical information from GOES-16 on HWRF hurricane forecasts. *Journal of Geophysical Research-Atmospheres*, 125(3), e2019JD031 647. Available from: https://doi.org/10.1029/2019jd031647

Liu, H., Garrett, K., Ide, K. & Hoffman, R.N. (2023) On the use of consistent bias corrections to enhance the impact of Aeolus Level-2B Rayleigh winds on National Oceanic and Atmospheric Administration global forecast skill. *Quarterly Journal of the Royal Meteorological Society*, 150(758), 355–372. Available from: https://doi.org/10.1002/qj.4600

Liu, H., Garrett, K., Ide, K., Hoffman, R.N. & Lukens, K.E. (2022) A statistically optimal analysis of systematic differences between Aeolus horizontal line-of-sight winds and NOAA's global forecast system. *Atmospheric Measurement Techniques*, 15, 3925–3940. Available from: https://doi.org/10.5194/amt-15-3925-2022

Rennie, M.P. & Isaksen, L. (2020) The NWP impact of Aeolus Level-2B winds at ECMWF. In: *Technical Memorandum 864*. Reading: ECMWF. Available from: https://doi.org/10.21957/alift7mhr

Rennie, M.P., Isaksen, L., Weiler, F., de Kloe, J., Kanitz, T. & Reitebuch, O. (2021) The impact of Aeolus wind retrievals on ECMWF global weather forecasts. *Quarterly Journal of the Royal Meteorological Society*, 147(740), 3555–3586. Available from: https://doi.org/10.1002/qj.4142

Salonen, K., Cotton, J., Bormann, N. & Forsythe, M. (2015) Characterizing AMV height-assignment error by comparing best-fit pressure statistics from the Met Office and ECMWF data assimilation systems. *Journal of Applied Meteorology and Climatology*, 54(1), 225–242. Available from: https://doi.org/10.1175/JAMC-D-14-0025.1

Santek, D., Dworak, R., Nebuda, S., Wanzong, S., Borde, R., Genkova, I. et al. (2019) 2018 Atmospheric motion vector (AMV) intercomparison study. *Remote Sensing*, 11(19), 1–27. Available from: https://doi.org/10.3390/rs11192240

Santek, D., Dworak, R., Wanzong, S., Rink, T., Lukens, K., Reiner, S. et al. (2022) *NWC SAF Winds Intercomparison Study Report: 2021*. Final report. Madison, Wisconsin, USA: University of Wisconsin-Madison. Available from: https://cimss.ssec.wisc.edu/iwwg/Docs/CIMSS.AMV.Comparison.2021.Report.02Nov2022.pdf

Stoffelen, A. (2005) The Atmospheric Dynamics Mission for global wind field measurement. *Bulletin of the American Meteorological Society*, 86(1), 73–87. Available from: https://doi.org/10.1175/BAMS-86-1-73

Velden, C.S. & Bedka, K.M. (2009) Identifying the uncertainty in determining satellite-derived atmospheric motion vector height attribution. *Journal of Applied Meteorology and Climatology*, 48(3), 450–463. Available from: https://doi.org/10.1175/2008jamc1957.1

Wang, X. & Lei, T. (2014) GSI-based Four–Dimensional Ensemble–Variational (4DEnsVar) data assimilation: Formulation and single-resolution experiments with real data for NCEP global forecast system. *Monthly Weather Review*, 142(9), 3303–3325. Available from: https://doi.org/10.1175/MWR-D-13-00303.1

Weiler, F., Rennie, M., Kanitz, T., Isaksen, L., Checa, E., de Kloe, J. et al. (2021) Correction of wind bias for the Lidar on-board Aeolus using telescope temperatures. *Atmospheric Measurement Techniques*, 14, 7167–7185. Available from: https://doi.org/10.5194/amt-14-7167-2021

Zhu, Y., Derber, J., Collard, A., Dee, D., Treadon, R., Gayno, G. et al. (2014) Enhanced radiance bias correction in the National Centers for Environmental Prediction's gridpoint statistical interpolation data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 140(682), 1479–1492. Available from: https://doi.org/10.1002/qj.2233

How to cite this article: Hoffman, R.N., Liu, H., Lukens, K.E., Garrett, K. & Ide, K. (2024) Assimilating atmospheric motion vector winds using a feature track correction observation operator. *Quarterly Journal of the Royal Meteorological Society*, 150(765), 5074–5093. Available from: https://doi.org/10.1002/qj.4857

APPENDIX

Four sections are presented below. The first three provide details for the implementation of the evaluation, inner optimization, and outer optimization of the feature track correction (FTC) observation operator (FTC-OO) and the fourth outlines a proposal to build variational FTC (VarFTC) systems based on the FTC-OO. The implementation descriptions include a number of efficiencies, including (1) interpolating the FTC-OO parameters across latitude and height boundaries rather than interpolating the FTC-OO values to reduce the burden on the adjoint calculation, (2) traversing the observation sample a single time while accumulating "universal sums" that can be used to solve all the least-squares optimizations within the outer optimization, and (3) using a directed search to solve the outer optimization. The VarFTC proposal includes a simple extension of the current study to implement the FTC-OO inline in the outer loop of the GSI immediately after the inline quality-control procedures.

A.1 FTC OBSERVATION OPERATOR EVALUATION

The initial step of the FTC-OO interpolates the model wind profile in the model native vertical coordinate system to a regular pressure grid $p_k = p^o + k\delta p$. The grid spacing is δp and the grid is indexed by k in $[k_{\min}, k_{\max}]$ centered on the observation level ($k = 0$). In the experiments reported here, $\delta p = 25$ hPa and $-k_{\min} = k_{\max} = 12$. Then, to evaluate Equation (3), the vertical average is set up as a weighted sum over the vertical grid, so that Equation (2) becomes

$$\widehat{\mathbf{V}} = f(\mathbf{V}(p)) = \gamma\overline{\mathbf{V}} + \delta\mathbf{V} = \sum_{k=k_{\min}}^{k_{\max}} w_k \mathbf{V}_k + \delta\mathbf{V}, \qquad \text{(A1)}$$

where $\mathbf{V}_k$ are the values of the model winds interpolated to the FTC-OO vertical grid and w_k are the weights. For a single subset, the weights w_k are identically equal to γ/n for $k \in [m, m+n-1]$ and zero otherwise. In terms of m and n, $\Delta z = n\delta p$ and $h = (m + (n-1)/2)\delta p$. In what follows, we will refer to the "FTC-OO null solution," which is identical to the ordinary AMV observation operator and occurs when $m = 0$, $n = 1$, $\gamma = 1$, and $\delta\mathbf{V} = 0$.

The definition of the w_k used here allows interpolation of w_k near latitude and height boundaries even when m and n are different for different geographic regions and/or height bins. This device is needed because the FTC solutions (i.e., m, n, γ, and $\delta\mathbf{V}$) are calculated separately for different subsets. Each subset is defined by an AMV type, a geographic region (NHX, Tropics, or SHX), and a height bin (low, middle, or high; Section 2). Near the boundaries (the 20° parallels and the 800- and 450-hPa levels) we interpolate the FTC estimates for the subsets on either

side of the boundaries. This interpolation is linear in latitude between 15° and 25°, and linear in the log of pressure between 850 and 750 hPa and between 500 and 400 hPa. As a simple example, consider an observation at 20°N and 447.21 hPa (i.e., the halfway point between 500 and 400 hPa in log space). The interpolated solution in this case is the average of four solutions—for the high-level Tropics, high-level NHX, mid-level Tropics, and mid-level NHX. For each specific AMV, the $\mathbf{V}_k$ are fixed, and Equation (A1) depends linearly on w_k and $\delta\mathbf{V}$. Instead of calculating and then interpolating several solutions of Equation (A1), we interpolate w_k and $\delta\mathbf{V}$ and then solve Equation (A1) once for the interpolated values of w_k and $\delta\mathbf{V}$. With this approach, changes to the GSI are minimal. In addition to implementing Equation (A1), the only new coding within the adjoint calculation is to calculate $\mathbf{V}_k$ on the regular pressure grid. Since the GSI already interpolates the model state to p^o, the pressure of the AMV, it is a simple matter to reuse that procedure to interpolate to the regular pressure grid.

Equation (A1) allows for alternative vertical averaging operators. In the experiments described here we always use the boxcar vertical average, that is, $w_k = \gamma/n$. However, any shape that can be described by Δz, h, and some measure of the amplitude of the shape (γ) can be accommodated. We experimented with an isosceles triangular weighting scheme and saw minor improvements.

A.2 FTC OBERVATION OPERATOR INNER OPTIMIZATION

We optimize the parameters of Equation (A1) by minimizing the squared difference between the observation, $\mathbf{V}^o$, and its estimate, $\widehat{\mathbf{V}}$. Since the optimization is subset by subset, the optimization is with respect to γ and not w_k. In the inner optimization, the parameters m and n are fixed and γ and $\delta\mathbf{V}$ are free. Thus, the inner optimization is a simple least-squares problem with the complication that the constant term, that is, one of the components of $\delta\mathbf{V}$, is different for the u- and v-winds. An additional issue is that, during the outer optimization, this problem must be solved multiple times for different values of m and n. To address the first issue, we do the following.

1. Center (i.e., remove the sample mean of) the data separately for the u- and v-wind components.
2. Determine γ by solving the least-squares through-the-origin problem for the centered data.
3. Evaluate $\delta\mathbf{V}$ making use of the mean values determined in step 1.

The statistics needed here involve summing over all observations in the sample. Rather than evaluate these statistics directly, we evaluate them by combining a set of universal statistics pre-calculated for the individual levels and subsamples. The statistics needed include the mean of $\mathbf{V}^o$ and $\overline{\mathbf{V}}$, the covariance of $\mathbf{V}^o$ and $\overline{\mathbf{V}}$, and the covariance of $\overline{\mathbf{V}}$ with itself. However, each of these is easily calculated in terms of the universal statistics—the mean of $\mathbf{V}^o$ and $\mathbf{V}_k$, the covariances of $\mathbf{V}^o$ and $\mathbf{V}_k$, and the covariances of $\mathbf{V}_k$ with itself. Note that the universal statistics must be maintained separately for the u- and v-winds. The advantage of this approach is that the universal statistics can be calculated once per subsample and saved and then used for any values of m and n and any combination of subsamples. As a result, this approach is much more efficient than the standard linear model used by Hoffman *et al.* (2022). The remainder of this section fills in the details.

For arbitrary Q and R, we use the notation $\langle Q\rangle$ to indicate the expectation of Q, which in practice is implemented as a sample average

$$\langle Q\rangle = S_Q/N, \tag{A2}$$

where S_Q is the sum of Q over the sample of size N. Some useful properties of the expectation operator are the following. Because the expectation of an expectation is the original expectation, that is, $\langle\langle Q\rangle\rangle = \langle Q\rangle$, we have, for $q = Q - \langle Q\rangle$ and $r = R - \langle R\rangle$, that

$$\langle q\rangle = 0, \tag{A3}$$

$$\langle q^2\rangle = \langle Q^2\rangle - \langle Q\rangle^2, \qquad \text{and} \tag{A4}$$

$$\langle qr\rangle = \langle QR\rangle - \langle Q\rangle\langle R\rangle. \tag{A5}$$

In what follows, to emphasize the parallel with the usual ordinary least-squares (OLS) notation of $y = ax + b + \varepsilon$, we will use Y and y for observation values and X and x for model values. For each observation location, let Y_u be the observed u-wind, $X_{u,k}$ be the model u-wind at level k, and $\overline{X}_u$ be the layer-averaged model u-wind, given by

$$\overline{X}_u = \sum w_k X_{u,k}. \tag{A6}$$

These definitions also hold with v replacing u.

To center the wind data, we remove the expected values for each component separately and stack them into vectors x and y, defined by

$$x = \begin{pmatrix} x_u \\ x_v \end{pmatrix} = \begin{pmatrix} \overline{X}_u \\ \overline{X}_v \end{pmatrix} - \begin{pmatrix} \langle\overline{X}_u\rangle \\ \langle\overline{X}_v\rangle \end{pmatrix},$$
$$y = \begin{pmatrix} y_u \\ y_v \end{pmatrix} = \begin{pmatrix} Y_u \\ Y_v \end{pmatrix} - \begin{pmatrix} \langle Y_u\rangle \\ \langle Y_v\rangle \end{pmatrix}. \tag{A7}$$

Because of Equation (A3), $\langle x\rangle = \langle y\rangle = 0$.

The OLS regression model through the origin is given by

$$y = ax + \varepsilon, \tag{A8}$$

where we have replaced γ with a. The OLS solution must satisfy $\langle x\varepsilon\rangle = 0$. If this does not hold, the predictor is correlated with the error and the predictor could be used to eliminate (predict) at least part of the error. Therefore, multiplying Equation (A8) by x and taking the expectation, the solution for a is given by

$$a = \frac{\langle xy\rangle}{\langle x^2\rangle}. \tag{A9}$$

Given a and ignoring ε, we substitute Equation (A7) into Equation (A8) to find

$$\begin{pmatrix} Y_u \\ Y_v \end{pmatrix} = a\begin{pmatrix} \overline{X}_u \\ \overline{X}_v \end{pmatrix} + \left[\begin{pmatrix} \langle Y_u\rangle \\ \langle Y_v\rangle \end{pmatrix} - a\begin{pmatrix} \langle\overline{X}_u\rangle \\ \langle\overline{X}_v\rangle \end{pmatrix}\right],$$
$$\begin{pmatrix} Y_u \\ Y_v \end{pmatrix} = a\begin{pmatrix} \overline{X}_u \\ \overline{X}_v \end{pmatrix} + \begin{pmatrix} b_u \\ b_v \end{pmatrix}, \tag{A10}$$

Thereby defining b_u and b_v. The last form in Equation (A10) is identical to the second form in Equation (A1) if we identify b_u and b_v as the components of $\delta\mathbf{V}$.

Given a, b_u, and b_v, we can evaluate the mean and mean squared difference for an independent sample in terms of expectations for the u- and v-winds separately. For the u-winds, this gives

$$\begin{aligned}
\langle\varepsilon_u\rangle &= \langle Y_u - (a\overline{X}_u + b_u)\rangle \\
&= \langle Y_u\rangle - a\langle\overline{X}_u\rangle - b_u, \qquad \text{and} \\
\langle\varepsilon_u^2\rangle &= \langle(Y_u - (a\overline{X}_u + b_u))^2\rangle \\
&= \langle Y_u^2 - 2a\overline{X}_uY_u - 2b_uY_u + a^2\overline{X}_u^2 + 2ab_u\overline{X}_u + b_u^2\rangle \\
&= \langle Y_u^2 - 2a\overline{X}_uY_u + a^2\overline{X}_u^2 - 2b_u\varepsilon_u - b_u^2\rangle \\
&= \langle Y_u^2\rangle - 2a\langle\overline{X}_uY_u\rangle + a^2\langle\overline{X}_u^2\rangle - 2b_u\langle\varepsilon_u\rangle - b_u^2.
\end{aligned} \tag{A11}$$

This also holds with v replacing u. For the dependent sample, $\langle\varepsilon_u\rangle = \langle\varepsilon_v\rangle = 0$. The u- and v-wind mean and mean squared difference are then averaged to get the corresponding values for the combined dataset,

$$\begin{aligned}
\langle\varepsilon\rangle &= (\langle\varepsilon_u\rangle + \langle\varepsilon_v\rangle)/2, \qquad \text{and} \\
\langle\varepsilon^2\rangle^{1/2} &= ((\langle\varepsilon_u^2\rangle + \langle\varepsilon_v^2\rangle)/2)^{1/2}.
\end{aligned} \tag{A12}$$

To solve Equation (A9) we need, for the u-wind components, $\langle x_uy_u\rangle$ and $\langle x_u^2\rangle$. Using Equations (A5) and (A4), these can be determined from

$$\begin{aligned}
\langle x_uy_u\rangle &= \langle\overline{X}_uY_u\rangle - \langle\overline{X}_u\rangle\langle Y_u\rangle, \\
\langle x_u^2\rangle &= \langle\overline{X}_u^2\rangle - \langle\overline{X}_u\rangle^2.
\end{aligned} \tag{A13}$$

This also holds for v replacing u. The statistics we need to solve Equation (A7), Equation (A11), and Equation (A13) are $\langle Y_u\rangle, \langle\overline{X}_u\rangle, \langle Y_u^2\rangle, \langle\overline{X}_uY_u\rangle, \langle\overline{X}_u^2\rangle$, and similar statistics with v replacing u. Using Equation (A6), expectations involving $\overline{X}_u$ are given by

$$\begin{aligned}
\langle\overline{X}_u\rangle &= \sum w_k\langle X_{u,k}\rangle, \\
\langle Y_u\overline{X}_u\rangle &= \sum w_k\langle Y_uX_{u,k}\rangle, \\
\langle\overline{X}_u^2\rangle &= \sum\sum w_jw_k\langle X_{u,j}X_{u,k}\rangle.
\end{aligned} \tag{A14}$$

This also holds with v replacing u. Finally, the univeral statistics needed are $\langle Y_u\rangle, \langle X_{u,k}\rangle, \langle Y_u^2\rangle, \langle X_{u,k}Y_u\rangle, \langle X_{u,j}X_{u,k}\rangle$, and similar statistics with v replacing u. To combine these universal statistics over multiple subsamples, the subsample sizes are also needed.

A.3 FTC OBSERVATION OPERATOR OUTER OPTIMIZATION

The outer optimization of the FTC-OO parameters finds the minimum of a 2D cost function in a $m - n$ grid corresponding to a discrete set of Δz and h values. This cost function has a single minimum, is well behaved, and a brute force search can be replaced with a directed search starting from a prior estimate. A reasonable prior would be a previous solution or the null FTC solution. Given the current estimate, the directed search finds the new estimate as the minimum within a 5×5 window centered on the current estimate in 2D space. If the current and new estimates are the same, the process terminates. (Note that a minimizing central point of a 3×3 window is not guaranteed to be a global minimum, but a minimizing central point of a 5×5 window is guaranteed to be a global minimum.) Every time the inner linear model is solved in a directed search, that cost-function value is saved in the 2D grid. As a practical matter, the $m - n$ grid is extended beyond the allowed search area and points outside the search area are initialized to infinity, while points within the search area are initialized to missing. Thus, solving the linear model is only required for unexplored (i.e., missing) points in the 5×5 window. The resulting 2D cost function at the end of such a directed search is shown in Figure A1. In Figure A1 the directed search began with a prior estimate of $\Delta z = 125$ hPa and $h = 0$ hPa and concluded at the global minimum at $\Delta z = 325$ hPa and $h = -25$ hPa. For the most part, the search jumps along odd values of n on the line $h = -25$ hPa. For n even, the smallest grid values are

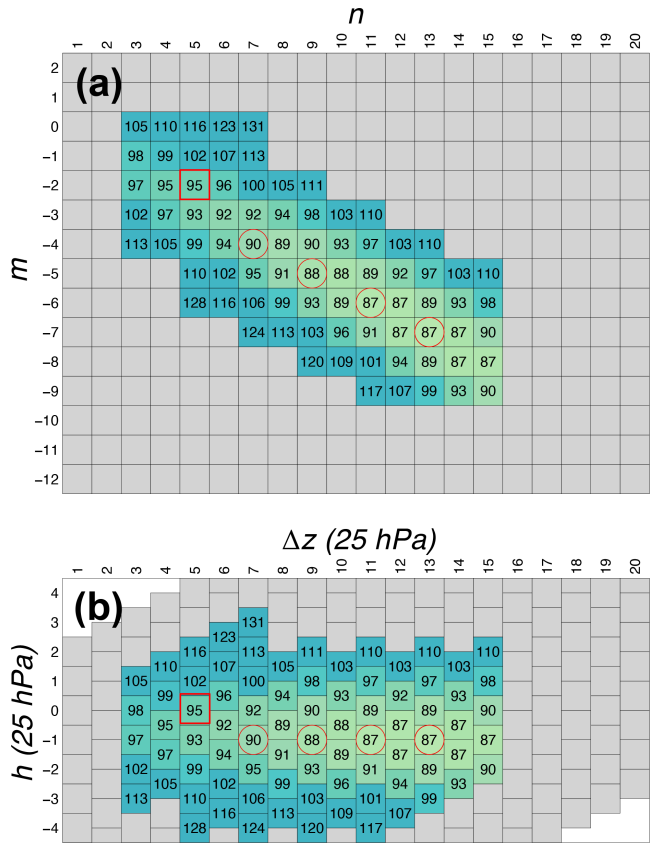

FIGURE A1 The value of the cost function that is minimized to determine the FTC parameters. The plotted values are normalized so that the normalized value is 100 for $\Delta z = 25$ hPa and $h = 0$ hPa, that is, the case when the layer averaged over is just the single level at the AMV pressure. (In this example, this cost-function value is 7073.) The cost function is plotted as a function of m and n in panel (a) and of z and h (in units of 25 hPa) in panel (b). The figure shows the state of the directed search when the minimum is found. The red square indicates the initial guess, and the red circles indicate the series of candidate minima found as the search progressed.

for $h/25 = -0.5$ and -1.5, but they are not as small as the neighboring grid values for odd n and $h/25 = -1$. This is the reason the directed search fails when using a 3×3 window.

A.4 PROPOSED VARIATIONAL FTC IMPLEMENTATION APPROACHES

Developing variational FTC (VarFTC) (or ensemble FTC) systems based on the FTC-OO would require embedding the evaluation of the FTC-OO parameters within the DA system. In this context the model state vector will be augmented by the FTC-OO parameters γ and $\delta\mathbf{V}$, as well as Δz and h. In the current study, Δz and h took on discrete values, but this limitation would be relaxed in a VarFTC system with suitable changes in the implementation of the vertical average in Equation (A1).

Usually, and in the current case, the augmented variables evolve according to persistence. That is, the background vector of the augmented variables is just the previous analysis value. An alternative to persistence is damped persistence towards some a priori values. In the current case, these a priori values might correspond to the null FTC solution, and e-folding times would have to be specified or estimated. In any case, the analyzed augmented variables would be constrained by a measure of the difference between the analysis and background augmented parameters. One approach is to treat the augmented state vector in the same way as the model state vector. That is, estimate the background-error covariance matrix for the augmented state vector. Conceptually, this is easy to do in any DA system that uses an ensemble of forecasts to estimate the background-error covariance matrix, but we anticipate there will be challenges in tuning such a system.

As a next step, we propose a simple extension of the current algorithm to an inline implementation in the GSI. In the GSI, the analysis minimization is done as a linearized inner-loop iterative calculation within an outer nonlinear loop with only a few (typically three) outer iterations. Before the start of each inner loop, quality controls are performed and the objective function is linearized. For inline calculations, the FTC-OO parameters would be calculated before the start of each inner loop immediately after the inline quality control, and the FTC-OO parameters would be held fixed within the inner loop. Note that, during the minimization, $\hat{\mathbf{V}}$ is always evaluated for the current estimate of the state vector, which is the analysis when the process terminates.

The inline calculations of the FTC-OO parameters must be constrained, since, in a single DA cycle, for a single subset, the subset sample may be too small. That is, the sample for a particular AMV type, geographic region, and height bin may be too small to calculate the FTC-OO parameters reliably. As an alternative to estimating an error covariance matrix of the FTC-OO parameters, we propose to choose the FTC-OO parameters to simultaneously minimize the mean squared difference between $\mathbf{V}^o$ and $\hat{\mathbf{V}}$ for the current cycle plus the mean squared difference between $\mathbf{V}^o$ and $\hat{\mathbf{V}}$ for recent previous cycles. The number of previous cycles would be selected so that the size of the prior sample exceeds a tunable constant. The second term (for the prior data) would be multiplied by a second tunable constant (β).

This proposed empirical data constraint for the inline implementation avoids estimating the parameters of some assumed distribution of the AMV errors. Only two parameters must be tuned, the multiplier (β) and the minimum size of the prior sample. Note that in the first term, that is, for the current cycle, $\hat{\mathbf{V}}$ is evaluated for the current

estimate of the state vector, but in the second term, that is, for the prior cycles, $\hat{\mathbf{V}}$ is evaluated for the prior analyses but with the current estimate of the FTC-OO parameters.

Finally, the two mean squared differences can be combined into one weighted-average squared difference, which can then be solved using the methods of Sections A.2 and A.3, except that the estimates of the expectations would be determined as weighted averages, with smaller weight given to the prior data. That is, instead of Equation (A2), we would use

$$\langle Q \rangle = \frac{N_{\mathrm{c}} S_{Q,\mathrm{c}} + \beta N_{\mathrm{p}} S_{Q,\mathrm{p}}}{N_{\mathrm{c}} + \beta N_{\mathrm{p}}}, \tag{A15}$$

where subscripts c and p denote current and prior cycles respectively. Note that the sums for the prior cycles do not change during the current cycle.