

# Overlapping Windows in a Global Hourly Data Assimilation System

LAURA C. SLIVINSKI,<sup>a,b</sup> DONALD E. LIPPI,<sup>c,d,e</sup> JEFFREY S. WHITAKER,<sup>b</sup> GUOQING GE,<sup>a,f</sup> JACOB R. CARLEY,<sup>d</sup>  
CURTIS R. ALEXANDER,<sup>f</sup> AND GILBERT P. COMPO<sup>a,b</sup>

<sup>a</sup> Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado

<sup>b</sup> NOAA/Physical Sciences Laboratory, Boulder, Colorado

<sup>c</sup> I.M. Systems Group, Inc., Rockville, Maryland

<sup>d</sup> NOAA/Environmental Modeling Center, College Park, Maryland

<sup>e</sup> University of Maryland, College Park, College Park, Maryland

<sup>f</sup> NOAA/Global Systems Laboratory, Boulder, Colorado

(Manuscript received 17 August 2021, in final form 16 February 2022)

**ABSTRACT:** The U.S. operational global data assimilation system provides updated analysis and forecast fields every 6 h, which is not frequent enough to handle the rapid error growth associated with hurricanes or other storms. This motivates development of an hourly updating global data assimilation system, but observational data latency can be a barrier. Two methods are presented to overcome this challenge: “catch-up cycles,” in which a 1-hourly system is reinitialized from a 6-hourly system that has assimilated high-latency observations; and “overlapping assimilation windows,” in which the system is updated hourly with new observations valid in the past 3 h. The performance of these methods is assessed in a near-operational setup using the Global Forecast System by comparing forecasts with in situ observations. At short forecast leads, the overlapping windows method performs comparably to the 6-hourly control in a simplified configuration and outperforms the control in a full-input configuration. In the full-input experiment, the catch-up cycle method performs similarly to the 6-hourly control; reinitializing from the 6-hourly control does not appear to provide a significant benefit. Results suggest that the overlapping windows method performs well in part because of the hourly update cadence, but also because hourly cycling systems can make better use of available observations. The impact of the hourly update relative to the 6-hourly update is most significant during the first forecast day, while impacts on longer-range forecasts were found to be mixed and mostly insignificant. Further effort toward an operational global hourly updating system should be pursued.

**KEYWORDS:** Kalman filters; Numerical weather prediction/forecasting; Data assimilation

## 1. Introduction

Currently, many global operational systems assimilate observations two to four times per day, including the National Oceanic and Atmospheric Administration (NOAA) Global Forecast System (GFS; [https://www.weather.gov/media/notification/pdf2/scn21-20\\_gfsv16.0\\_aac.pdf](https://www.weather.gov/media/notification/pdf2/scn21-20_gfsv16.0_aac.pdf)). In contrast, higher-resolution regional systems such as the High-Resolution Rapid Refresh system (HRRR; Alexander et al. 2020; Smith et al. 2008) assimilate observations hourly, or even more frequently for convective-scale models. Currently in the United States, the Rapid Refresh (RAP; Benjamin et al. 2016) takes 6-hourly fields from the GFS as lateral boundary conditions, assimilates observations, and provides hourly fields in real-time that can be used as boundary conditions to the HRRR. In contrast, if the global system assimilated observations hourly and provided real-time hourly analysis fields, these fields could be used as lateral boundary conditions for the HRRR and other high-resolution regional models.

Observations with high temporal frequency are also becoming more common (e.g., geostationary satellite observations from *GOES-16* and *Himawari-8/9* are available every 10–15 min), but it is likely that they are not being used as effectively as possible in a 6-hourly updating system. For example, clear-sky

geostationary radiances have a small impact on global forecast skill relative to radiances from polar orbiting satellites (Cardinali 2009). These high-frequency observations from geostationary satellites could conceivably be used more effectively in an hourly system: in particular, Kim and Kim (2019) found that shorter forecast lengths led to a greater impact of observations on forecast skill in a system with a hybrid ensemble/static background covariance. Global systems are also continually increasing in spatial resolution, to the point that they may be able to resolve convective systems in the near future. Accurate initialization of convective systems within a 6-hourly assimilation window would require development of new assimilation methods that can properly account for highly nonlinear, rapidly growing error in convective systems. Alternatively, an assimilation system with a shorter cadence could use existing methods because the error growth within a shorter assimilation window (of an hour or less) is more linear than in a longer window (e.g., Bocquet et al. 2010; Ruiz et al. 2021), though this would clearly not mitigate all non-Gaussianity (e.g., in bounded, moisture-related variables). In addition, current assimilation systems often struggle to accurately constrain the position of hurricanes, which can significantly degrade the forecast (Chen and Snyder 2007; Lu et al. 2017). The current operational Hurricane Weather Research and Forecasting (HWRF) Model uses “vortex relocation” to adjust the hurricane position in the background field prior to assimilating observations (e.g., Liu et al. 2006). However,

Corresponding author: Laura Slivinski, [laura.slivinski@noaa.gov](mailto:laura.slivinski@noaa.gov)

DOI: 10.1175/MWR-D-21-0214.1

© 2022 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

hurricane positions may also be better-estimated by more frequent assimilation, reducing non-Gaussian displacement errors since there is less time for the hurricane in the background field to drift away from the observations between analysis times. It is likely, then, that an hourly global assimilation system would not only provide more accurate lateral boundary conditions for a regional model, but could also be implemented in such a way that it would more broadly improve upon a 6-hourly global system.

There are several ways one could envision implementing hourly data assimilation. The simplest method would be identical to the 6-hourly system but with 1-h-long assimilation windows, symmetric about the nominal analysis time. The main challenge of this method is data latency: observations may not be available for assimilation until more than an hour after the observation is valid. This latency can be due to, for instance, data communication delays, preprocessing, and quality control (e.g., Lin et al. 2017). Thus, a symmetric hourly system that is run at, say, 30 min past the hour every hour would only be able to assimilate observations that are available within minutes after they are taken.

One method to overcome the data latency challenge is “catch-up cycles” in which the symmetric hourly system described above is run, but is reinitialized every 6–12 h from a separate global system that has a longer data receipt cutoff time, such as the current NOAA Global Data Assimilation System (GDAS) with a nearly 6-h data receipt cutoff time. This system provides 6-hourly fields that have assimilated observations with 3–9-h latency, so that the 1-hourly system is reinitialized from a system that has seen more high-latency observations. A version of catch-up cycling is currently used in the regional RAP system (Benjamin et al. 2016), not only to include information from latent observations but also because the global model can better represent long waves than the regional model.

However, this implementation of catch-up cycles has drawbacks. First, it requires two separate global assimilation systems (one 6-hourly and one 1-hourly), though this could be easier to maintain than one global system and two regional systems (as in the RAP/HRRR configuration.) Second, it is unsatisfying to have a partially cycled system, in which information potentially gained from hourly updates is lost every 6 h. Third, this method does not entirely avoid the issue of data latency, since the hourly cycles run between re-initialization cycles will still not see higher-latency observations, potentially resulting in system performance that depends on the time of day. Finally, Poterjoy et al. (2021) point out that partial cycling methods can obscure systematic errors in the model, observations, or assimilation techniques, which could hinder future developments and improvements (e.g., Wong et al. 2020).

Another method to overcome the challenge of data latency is so-called continuous data assimilation. This is currently implemented operationally at the European Centre for Medium-Range Weather Forecasts using outer-loops in a 4DVar system (Lean et al. 2021), but forecasts and analyses are not generated at an hourly cadence. More generally, many global data assimilation systems run (at least) two

analysis updates every cycle; one with an earlier data receipt cutoff time, for real-time forecasts, and one with a later data receipt cutoff time, for more accurate initialization of the next assimilation cycle (e.g., Haseler 2004; Kleist et al. 2009). The Naval Research Laboratory addressed the issue of data latency in their 3DVar ocean data assimilation system by assimilating all data received since the previous analysis time using First Guess at Appropriate Time (FGAT) (Cummings and Smedstad 2013). Along similar lines, Payne (2017) presented the idea of overlapping assimilation windows in a 4DVar system, in which assimilation updates are performed hourly but assimilate observations in backward-looking 3-h windows. Previous work (Fisher et al. 2011; Trémolet 2006) has also discussed overlapping assimilation windows using 4DVar, but not with the goal of producing real-time hourly analyses. To date, the feasibility of the overlapping windows method for hourly data assimilation has not been assessed in a near-operational setup.

To this end, we test the effectiveness of hourly data assimilation using overlapping windows as well as catch-up cycles in the GFS, assimilating observations with actual data latency. The present work is intended to provide a description of the overlapping windows method as well as initial promising results that motivate future work. Section 2 describes the methods and experiment setup; section 3 presents results from three experiments comparing 6-hourly assimilation with overlapping windows and catch-up cycles; section 4 provides a preliminary investigation into the causes of the improvements seen with overlapping windows; and section 5 provides conclusions and paths forward.

## 2. Methods

### a. Overlapping windows and catch-up cycles

Payne (2017) first introduced an implementation of overlapping windows to overcome the challenge of delayed observations in a rapidly cycling assimilation system. Figure 1 demonstrates the extent of the data latency problem in the global observing system, and illuminates the difficulty of performing real-time hourly data assimilation updates with non-overlapping windows: in such a system, the data receipt cutoff time would necessarily be no later than 1 h after the earliest possible observation in that window was taken. Figure 1 (right) shows that this cutoff would keep less than 40% of all observations valid in a single assimilation window. However, with overlapping windows, observations that have arrived in the past hour but that are valid in a longer window (say, 3 h) are assimilated; note that this means that each observation is assimilated no more than once. In this way, late-arriving observations can still impact a state estimate hours after they were measured. Figure 1 (right) shows that more than 95% of observations would be assimilated in such a system. We note that the aircraft observation counts in Fig. 1 are lower than normal due to COVID-19 (James et al. 2020) but are representative of the observation network for the experiments shown here.

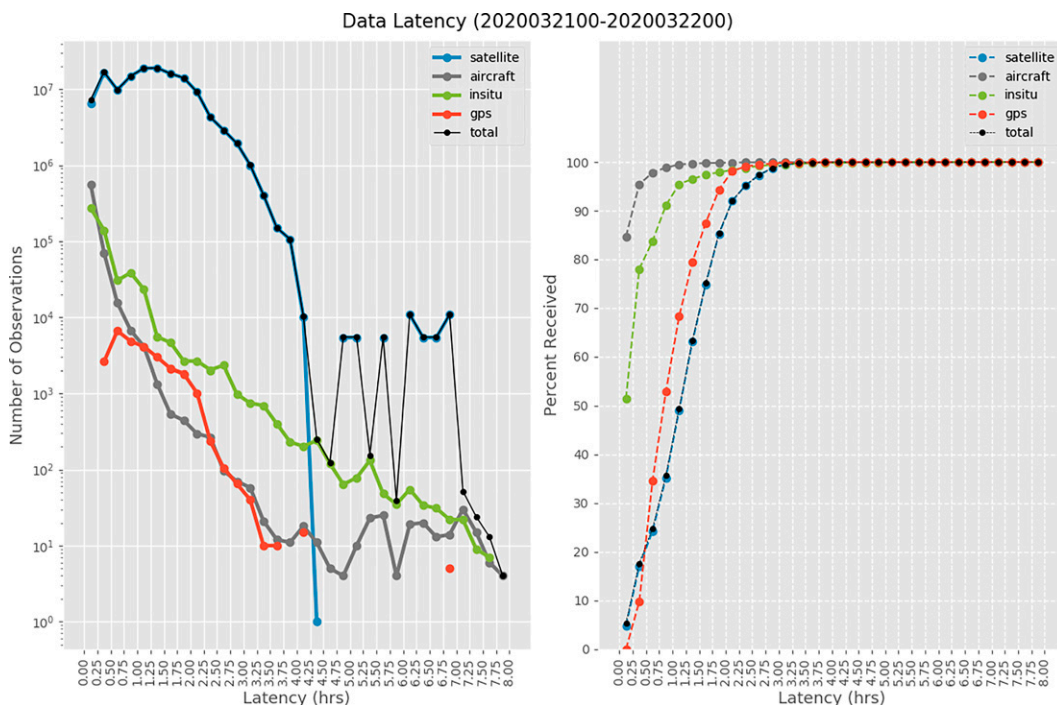


FIG. 1. (left) Number of observations as a function of latency (in 15-min bins) for different types of observations and (right) the cumulative distribution function of different types of observations as a function of latency, calculated using the GDAS observation dumps over the 24-h period of 0000–2300 UTC 22 Mar 2020. Latency is a positive quantity defined as the difference between the time an observation is received and the time for which it is valid.

The concept of overlapping windows in the presence of data latency was demonstrated in (Payne 2017) with a Lorenz-96 system and variational assimilation methods. Results showed promise, but further work was needed to evaluate this method in a more operationally relevant system. To that end, the tests shown in the present paper will use the Global Forecast Model (GFS), version 15,<sup>1</sup> which is based on the Finite Volume Cubed-Sphere dynamical core (FV3; Lin 2004; Putman and Lin 2007) with the hybrid-gain local ensemble transform Kalman filter (LETKF) (Penny 2014). The current NOAA global operational system uses the hybrid 4DVar assimilation method (Kleist and Ide 2015), but results (not shown) suggest that the hybrid-gain LETKF has comparable performance in a 6-hourly system and is computationally less expensive than hybrid 4DVar. Future work will investigate the hybrid 4DVar in a 1-hourly system.

Figure 2 (left panel) illustrates how the overlapping windows (OW) technique is implemented in this study. First, suppose initial conditions are available to cold-start the system at 0000 UTC. The forecast model is then run for 5 h with hourly output, providing first guess fields at 0100, 0200, ..., 0500 UTC (purple box). The analysis is then performed, assimilating observations in a 3.5-h window from 0100 to 0430 UTC (gray box). The nominal analysis time for this cycle is 0400 UTC; the hybrid-gain LETKF first calculates the 3DVar update

with FGAT valid at 0400 UTC, and then calculates LETKF updates at 0100, 0200, 0300, and 0400 UTC. Through the 4D LETKF covariances, every observation in the window can generate an increment to each hourly analysis. The 3DVar and LETKF updates are linearly combined based on a predetermined weighting parameter, resulting in hourly analysis fields from 0100 to 0400 UTC (note the 0100 UTC analysis is not used in the next cycle). For the second assimilation cycle, the aforementioned analyses valid at 0200–0400 UTC are used as the first three background fields; the model is then run from the 0400 UTC analysis for 2 h to generate a forecast that provides the last two background fields for the second cycle (see green box). In the second analysis cycle, observations valid from 0200 to 0530 UTC are assimilated, but only those observations that were not assimilated in the previous cycle are used in this cycle; no observations are assimilated more than once. The algorithm then repeats (blue box).

In the current 6-hourly operational system, the 4D Incremental Analysis Update (4DIAU) (Lei and Whitaker 2016) is used to prevent noise from gravity waves dominating the short-term forecast. Essentially, analysis increments are calculated throughout the assimilation window; the forecast model is then rerun for the assimilation window, adding these analysis increments to the model equations as a model forcing. Figure 3 illustrates how 4DIAU is implemented in a 6-hourly system with hourly background fields and 3-hourly increments. This method can be implemented in the context of overlapping windows (see Fig. 2, right panel), but we will

<sup>1</sup> GFSv15 was operational from June 2019 until March 2021 ([https://www.weather.gov/media/notification/scn19-40gfs\\_v15\\_1.pdf](https://www.weather.gov/media/notification/scn19-40gfs_v15_1.pdf)).

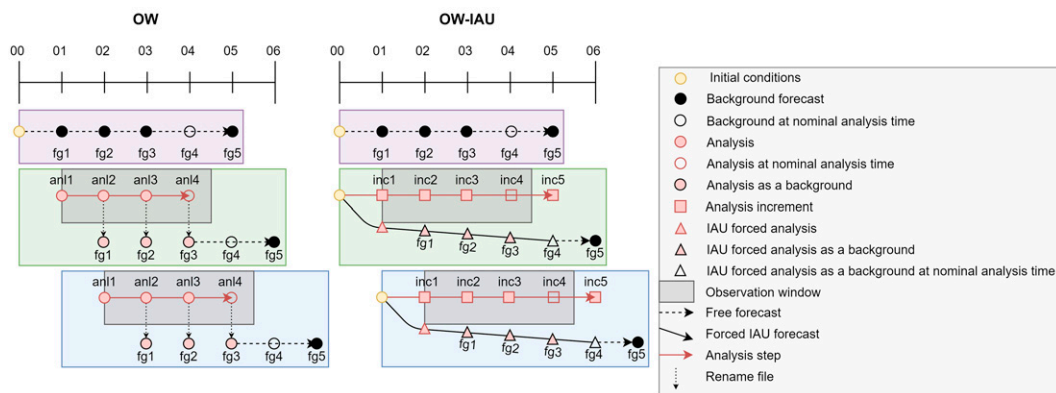


FIG. 2. Diagram of (left) overlapping windows without 4DIAU (OW) and (right) overlapping windows with 4DIAU (OW-IAU). The purple box demonstrates cycling from a cold start, the green box demonstrates the first full assimilation cycle, and the blue box shows the second full assimilation cycle. See the text for more details.

demonstrate below that it is unnecessary in an hourly setup and in fact, results in a loss of information across subsequent cycles.

The overlapping windows method will also be compared with a catch-up cycle algorithm inspired by the RAP. Figure 4 provides an overview of how catch-up cycles are implemented here. Analyses are generated hourly using 1-h symmetric assimilation windows. Continuously cycling such a system would only ever assimilate low-latency global observations, which compose less than 40% of all available observations (Fig. 1). To address the data latency issue, the catch-up cycle system is reinitialized every 6 h from a separate global 6-hourly system, which has a later data receipt cutoff time and therefore sees more high-latency global observations. In real-time, we anticipate that this system would be run using the NOAA Global Data Assimilation System (GDAS), which has a receipt cutoff time of about 3 h past the end of the assimilation window and lags real-time by approximately 6 h (i.e., the 0600 UTC analysis runs at about 1200 UTC).

Therefore, the catch-up cycle algorithm is designed such that it would be initialized from the end of a GDAS assimilation window (say 0900 UTC) with a run time 3 h later (1200 UTC), and then cycled hourly for 9 h. The first four cycles (blue filled circles in Fig. 4) are deemed the actual “catch-up” cycles before running the real-time hourly cycling (unfilled red circles in Fig. 4) beginning at 1300 until 1800 UTC. Note that the control experiment used for reinitialization includes 4DIAU, but the hourly cycling does not.

### b. Experiment setup

The hourly observation files in Binary Universal Form for the Representation of meteorological data (BUFR) used in this study were originally generated at the National Centers for Environmental Prediction/Environmental Modeling Center (NCEP/EMC) by gathering all available observations at 26 min past the hour except at 0000 and 1200 UTC, when observations are gathered at 58 min past the hour (consistent with current operational observation processing practices).

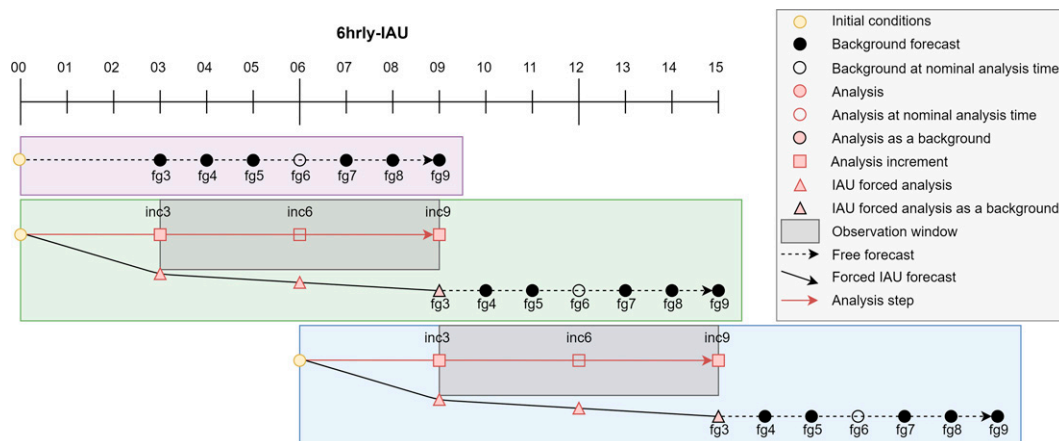


FIG. 3. Diagram of 6-hourly control configuration with 4DIAU and hourly background fields. The purple box demonstrates cycling from a cold start, the green box demonstrates the first full assimilation cycle, and the blue box shows the second full assimilation cycle. See the text for more details.



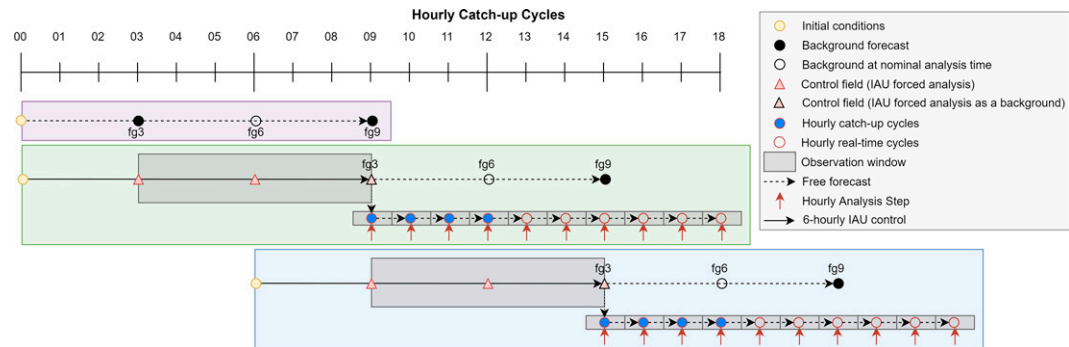


FIG. 4. Diagram of global hourly catch-up cycles. The red filled triangles denote the 3-hourly fields from a 6-hourly control experiment, after running 4DIAU for each cycle. The hourly cycling is initialized from the control field at the end of the 6-hourly assimilation window and then cycled hourly for 9 h (without 4DIAU). The first four hourly assimilation cycles (blue filled circles) would be the catch-up cycles, and the remaining six cycles (unfilled red circles) are the real-time hourly cycles.

Two different kinds of aggregated observation files, so-called “dump” files, were created: the “overlapping-window dumps” including all of the observations valid from 3 h prior to 30 min past the hour [−3 h, 30 min] and the “symmetric window dumps” including all observations valid from 30 min prior to 30 min past the hour [−30 min, 30 min]. Experiments with catch-up cycles assimilated these latter symmetric window dumps (see Table 1). For the overlapping-window dumps, consecutive BUFR files included overlapping data. To avoid assimilating observations more than once, the BUFR files were filtered to remove duplicate observations in the overlap period from subsequent cycles, using either observation receipt time information (when available) or a metadata comparison of consecutive files.

The overlapping windows method is tested in a hierarchy of retrospective experiments spanning four weeks in March–April 2020, and will be compared with the catch-up cycle method in the full-input experiment; see Table 1 for an overview. Both methods are studied using the GFS, version 15.

This version of the GFS uses three stochastic physics modules: Stochastically Perturbed Parameterization Tendencies (SPPT; Palmer et al. 2009; Shutts et al. 2011), stochastic humidity (SHUM; Tompkins and Berner 2008) and Stochastic Kinetic Energy Backscatter (SKEB; Berner et al. 2009). Each experiment will use a version of the hybrid-gain LETKF with 80 ensemble members at a resolution of C192 (i.e.,  $192 \times 192$  cells per face of the cubed sphere grid or approximately  $0.5^\circ$  grid spacing). Adaptive covariance inflation is implemented using both Relaxation to Prior Spread (RTPS; Whitaker and Hamill 2012) and Relaxation to Prior Perturbations (RTPP; Zhang et al. 2004). Unless otherwise noted, the RTPS and RTPP inflation parameters are 0.75 and 0.25, respectively; there is no high-resolution control member; and results shown use the ensemble mean. Results are shown for the overlapping windows method with and without 4DIAU. In each setup described below, the baseline will be a control experiment that uses 6-hourly non-overlapping assimilation windows with hourly background forecasts and 4DIAU with

TABLE 1. Summary of experiments shown in section 3.

Expt name	Inflation (RTPS/RTPP)	3DVar weight (%)	Obs types assimilated	Obs window (relative to nominal analysis time)	Obs receipt cutoff (relative to nominal analysis time)
NoSat, LETKF control	0.75/0.25	0	All except satellite radiances	[−3 h, 3 h]	2 h 45 min (GFS)
NoSat, LETKF OW	0.75/0.25	0	All except satellite radiances	[−3 h, 30 min]	58 min (0000 and 1200 UTC); 26 min (all other hours)
NoSat, LETKF OW-IAU	0.75/0.25	0	All except satellite radiances	[−3 h, 30 min]	58 min (0000 and 1200 UTC); 26 min (all other hours)
Full-input control	0.75/0.25	25	All	[−3 h, 3 h]	2 h 45 min (GFS)
Full-input OW	0.9/0.4	15	All	[−3 h, 30 min]	58 min (0000 and 1200 UTC); 26 min (all other hours)
Full-input catch-up cycles (hourly)	0.9/0.4	15	All	[−30 min, 30 min]	58 min (0000 and 1200 UTC); 26 min (all other hours)
Full-input catch-up cycles (control for reinitialization)	0.75/0.25	25	All	[−3 h, 3 h]	2 h 45min (GFS)

analysis increments calculated at the beginning, middle, and end of the window (Fig. 3). It will assimilate observations from the real-time GFS observation sets that are used in operations every 6 h. Note the GFS data receipt cutoff is about 15 min before the end of the observation window (2 h 45 min past the nominal analysis time at the center of the window) and therefore does not include high-latency data near the end of the window. This GFS-type control is used to evaluate relative performance of the hourly cycling because it is closest (though not identical) to what is currently used for operational forecasts.

First, overlapping windows will be demonstrated in a simple configuration that does not assimilate satellite radiances, using a pure LETKF (implemented by assigning zero weight to the 3DVar update in the hybrid-gain). In situ (including aircraft) and satellite wind observations are assimilated. Aircraft bias correction is not implemented here, to minimize the number of adaptive algorithms used in initial tests. The control in this set of experiments will also assimilate only non-radiance observations without aircraft bias correction using a pure LETKF.

Results will then be shown in a full-input system using the hybrid-gain LETKF. This system is allowed to assimilate satellite radiances, and adaptive aircraft bias correction and radiance bias correction are both active. The thinning of satellite observations, quality control methods, and bias correction systems used for online estimation of the bias correction coefficients in the hourly experiments have not been modified from the methods used in GFSv15 and in the control experiment. The static background covariance matrix used in the 3DVar portion of the algorithm has not been re-derived from the matrix used in GFSv15. Based on prior results (not shown), the hybrid-gain weighting parameter in the control experiment was tuned to assign 25% weight to the 3DVar gain. Because the static background covariance was not re-derived for an hourly cycling system, the overlapping windows and catch-up cycle experiments assigned 15% weight to the 3DVar gain; results with several weighting parameters ranging from 0% to 25% (not shown) suggest that this decrease from 25% provides a small but consistent improvement in background fits to observations.

### 3. Results

#### a. NoSat, pure LETKF

To determine how well the overlapping windows method works relative to the control, we first consider the root-mean-square (RMS) fit of each experiment's forecast fields to in situ observations (including radiosondes, dropsondes, and aircraft observations) for short forecast durations. Significance is evaluated using a Student's *t* test with inflation to account for temporal autocorrelations (Geer 2016). For a consistent comparison, the analysis fields from which the forecasts are initialized must have seen observations with (nearly) the same receipt cutoff time. For example, when comparing with observations in the 1200 UTC window (valid from 0900 to 1500 UTC), the control fields will be forecasts initialized from IAU-forced fields at 0900

UTC, which will have seen observations up to 0845 UTC (using the GFS data receipt cutoff; Fig. 3). The overlapping windows fields will be forecasts initialized from analyses at 0800 UTC, which will have seen observations valid up to 0826 UTC (using the overlapping windows data receipt cutoff). The first forecast hour (0800–0900 UTC) will be either a free forecast or forced by the IAU, depending on whether 4DIAU is used (Fig. 2). Thus, the overlapping windows forecasts will be initialized from analyses that have seen slightly fewer observations (a difference of ~15 min) than the control analysis and will be slightly longer forecasts (by ~1 h) than the forecasts from the control experiment. However, note that these are both the cycles that would be available to a forecaster at a given time. For instance, at 0900 UTC, the forecaster would have access to the 0800 UTC hourly update cycle, and the best available GFS analysis would be from the end of the 0600 UTC cycle (valid at 0900).

Figure 5 shows that the overlapping windows system, with and without 4DIAU, is comparable to the control system in the NoSat setup. Overlapping windows with 4DIAU (OW-IAU) has slightly but consistently larger RMS errors than the control when compared to wind and temperature observations at all levels. Conversely, overlapping windows without 4DIAU (OW) consistently, and often significantly, has smaller RMS errors relative to wind observations than the control. Relative to temperature and humidity observations, however, OW has slightly larger errors than the control. Overall, results in this simplified setup are promising. In the following section, the method will be examined in a more complex system.

First, however, we further investigate the differences in the overlapping windows method with and without 4DIAU. In the 6-hourly system, the 4DIAU is used to reduce gravity wave noise and prevent it from dominating in short-term forecasts. To assess the presence of imbalance, we examine 1-h global absolute surface pressure tendencies (not shown) which can act as a rough estimate for gravity wave noise (e.g., He et al. 2020; Lei and Whitaker 2016). In this metric, OW has about 20% higher noise than OW-IAU (with an absolute average tendency of about  $0.140 \text{ Pa s}^{-1}$ ) and about 35% higher noise than the control (with an absolute average tendency of about  $0.125 \text{ Pa s}^{-1}$ ) in the first 5–6 h of the forecast, but these differences decrease with longer forecast lengths (see section 3b.2). This suggests that gravity wave noise exists in the background fields of the hourly assimilation system, but does not dominate the short-term forecasts. In addition, removing the 4DIAU from the overlapping windows experiment does not degrade the performance, and seems to reduce RMS errors for wind forecasts (Fig. 5). The lack of degradation may be because the hourly increments in the overlapping windows method are smaller than those in the 6-hourly assimilation system (since fewer observations are assimilated each cycle and error growth between analysis times is likely more linear), and therefore do not initiate substantial gravity waves in the analysis updates.

Though 4DIAU is not needed to prevent gravity wave noise from dominating the forecasts in the hourly system (e.g., Fig. 5), it does contribute to increasing ensemble spread. Consequently, removing the IAU in the overlapping windows

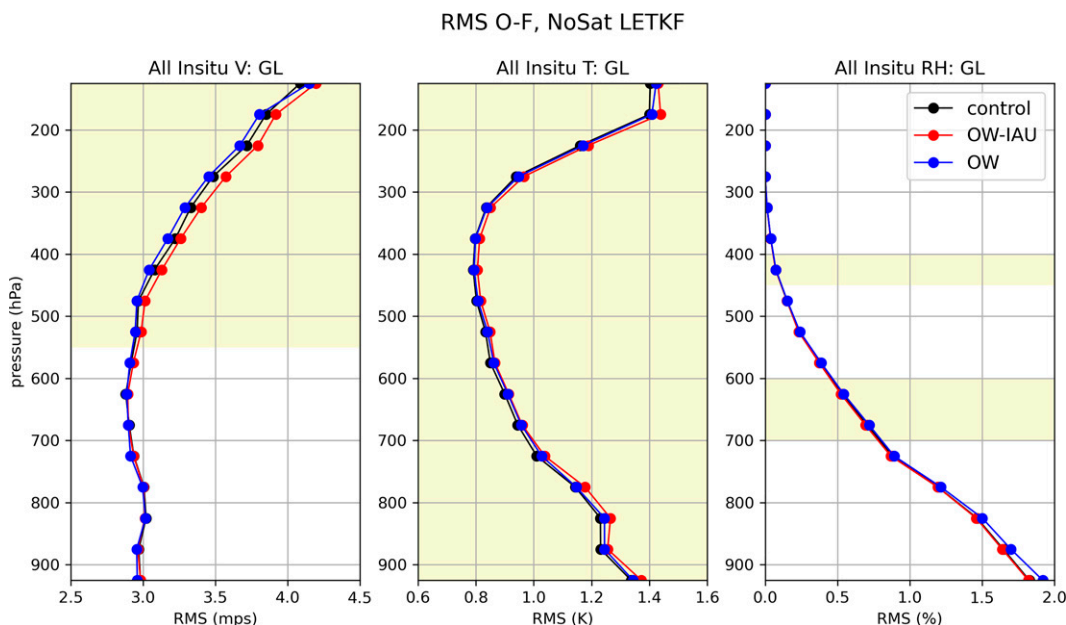


FIG. 5. Vertical profiles of short forecast RMS fits to global in situ observations of (left) vector wind ( $\text{m s}^{-1}$ ), (center) temperature (K), and (right) relative humidity (%) for the control (black), OW (blue), and OW-IAU (red) experiments in the simple NoSat, LETKF setup, averaged over the time period 1800 UTC 13 Mar–1800 UTC 13 Apr 2020. Yellow shading illustrates where the control and OW experiments differ significantly at the 95% level.

experiment (and necessarily using analyses as background fields in subsequent cycles; Fig. 2) decreases the ensemble spread of surface pressure by more than 50%, as shown for a representative date in Fig. 6 (cf. center and right panels). The OW ensemble spread is nearly 40% smaller than the control ensemble spread (cf. left and right panels). However, the spatial features are similar across all three experiments. In later experiments, we increase the inflation parameters in the OW experiments to compensate for removing the IAU while keeping the OW ensemble spread similar to the control ensemble spread (which has been previously tuned to minimize background errors).

Despite slightly larger gravity wave noise and smaller ensemble spread, OW consistently outperforms OW-IAU in wind and temperature at all levels because of the incremental nature of the 4DIAU (see appendix). In other words, since

the IAU forcing is added incrementally, only the fields at the end of the window have been fully informed by all of the observations (contrast Fig. 3 and right panel of Fig. 2). In OW-IAU, the earlier fields (that have only partially been informed by the observations) are suboptimal estimates but are still used as background fields for the next cycle. This effect likely compounds over many cycles, leading to an overall degradation in performance as well as larger ensemble spread which also indicates that the analyses are less constrained by observations (e.g., Fig. 6). Conversely, in OW, the background fields mostly consist of analyses that have each been informed by all observations (see appendix for more details). Note that this effect is specific to the overlapping windows method: in the control non-overlapping method with 4DIAU, the full IAU forcing has been included by the beginning of the next assimilation cycle (Fig. 3), guaranteeing that

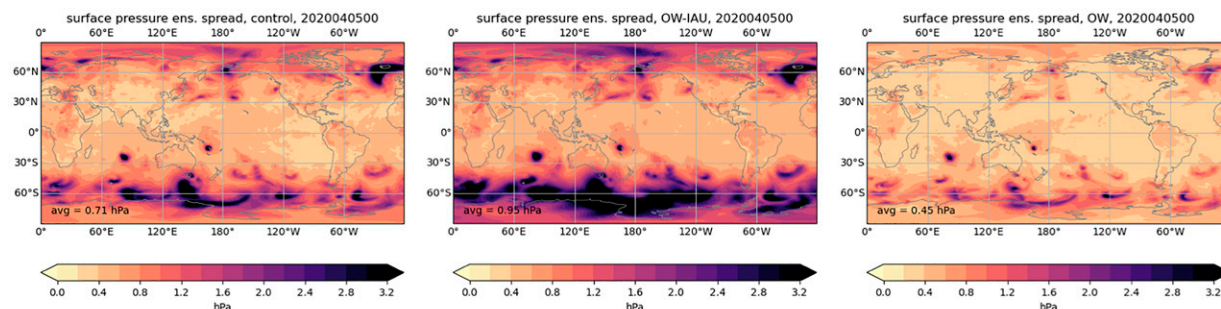


FIG. 6. Ensemble spread of 6-h background surface pressure (hPa) for the (left) control, (center) OW-IAU, and (right) OW at 0000 UTC 5 Apr 2020. Global average values are given in the bottom left of each panel.

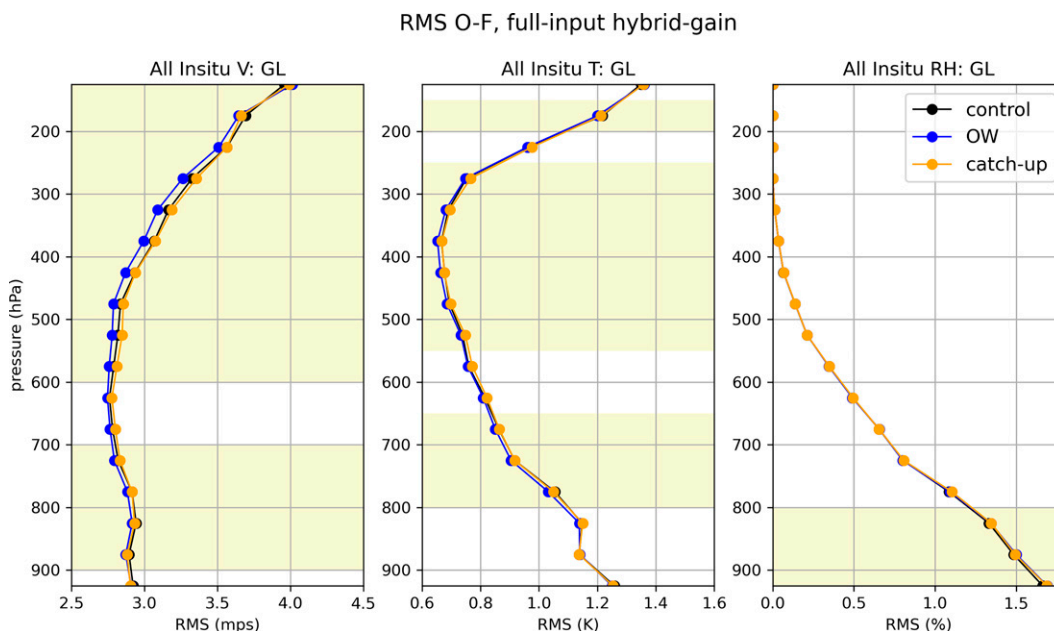


FIG. 7. Vertical profiles of short forecast RMS fits to global in situ observations of (left) vector wind, (middle) temperature, and (right) relative humidity for the control (black), OW (blue), and catch-up cycle (orange) experiments in the full-input hybrid-gain LETKF configuration, averaged over the time period 1800 UTC 13 Mar–1800 UTC 13 Apr 2020. Yellow shading demonstrates where the control and OW experiments differ significantly at the 95% level.

each of the background fields for the next cycle will include all available information from the current observations.

#### b. Full-input, hybrid-gain LETKF

The overlapping windows method is next evaluated in a more realistic setup and compared with the 6-hourly control as well as with catch-up cycles. Here, all available observations are assimilated, including radiances, and time-varying, adaptive bias corrections for radiances and aircraft are used (Zhu et al. 2014, 2015). The assimilation method is the hybrid-gain LETKF, which has the benefit of incorporating both an ensemble and a static background covariance, so that the effects of overlapping windows on these aspects can be investigated.

Based on results in section 3a, we increase the inflation parameters in OW and the hourly cycles within the catch-up cycle method to 0.9 and 0.4 for RTPS and RTPP, respectively. Note that the control used to reinitialize the catch-up cycles still uses 0.75 and 0.25 for RTPS and RTPP (see Table 1). As discussed earlier, the hybrid-gain parameter in the control experiment assigns 25% weight to the 3DVar gain, while the OW experiment assigns 15% weight to the 3DVar gain. On the basis of the results in the previous section, OW-IAU is not investigated in the full-input setup here.

##### 1) OVERLAPPING WINDOWS AND CATCH-UP CYCLES

In Fig. 7, short forecasts initialized from the full-input, hybrid-gain OW, catch-up cycles, and control analyses are compared to in situ observations of wind, temperature, and relative humidity. Similar to the results in the NoSat, pure

LETKF experiment, the wind forecasts from OW significantly outperform those from the control method at most levels (Fig. 7, left). In contrast to the NoSat, LETKF setup, OW has smaller temperature errors than the control at most levels (Fig. 7, center). The control experiment still has slightly smaller relative humidity errors than OW (Fig. 7, right); exploration of this is left for future work, but may be due to insufficient parameter tuning. The largest gain in performance is in the upper-tropospheric winds, consistent with previous findings using the regional hourly updated RAP system (Benjamin et al. 2016); this may be associated with the impact of aircraft observations (James and Benjamin 2017).

Despite increasing the inflation parameters, the ensemble spread of 6-h background surface pressure using OW is still about 25% smaller than the control ensemble spread (not shown). Modifying these parameters further could improve overall performance.

In this full-input setup, OW and the control are also compared with results from catch-up cycles (orange curves in Fig. 7). In general, the observation fit of the catch-up cycle backgrounds are very similar to those of the control. OW outperforms catch-up cycles in wind and temperature at nearly all levels. These differences are also significant at nearly all levels (not shown).

The behavior of the catch-up cycles is investigated in more detail in Fig. 8, which shows the 1h background fit to wind observations globally at all levels for catch-up cycles and OW, as well as the difference between the two. Note that the results in this figure are averaged over a shorter time period than earlier figures, but previous work (not shown) suggests that this does not impact the results. For OW, these are the



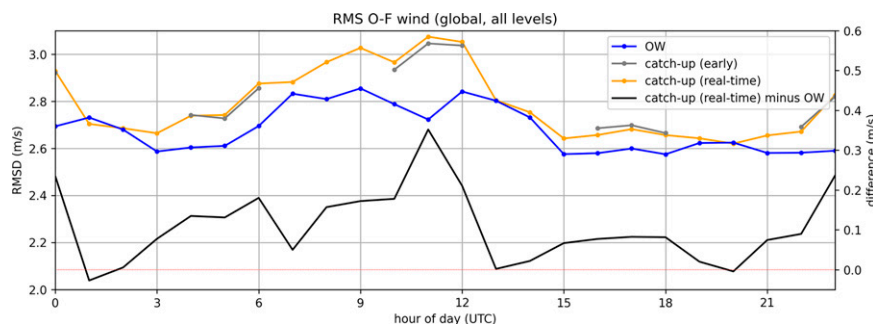


FIG. 8. RMS differences (RMSDs) between 1-h forecast fields and hourly in situ observations of wind averaged globally across all levels, as a function of time of day (left-hand axis). Results are averaged over 1600 UTC 13 Mar–1800 UTC 1 Apr 2020. The gray curve shows RMSDs of forecasts initialized from the “early” catch-up cycle analyses (blue filled circles in Fig. 4), and the orange curve shows RMSDs for forecasts initialized from real-time catch-up cycle analyses (red open circles in Fig. 4). The blue curve shows the OW performance. The black curve (right-hand axis) shows the difference between the real-time catch-up cycle performance and the OW performance. The dotted red line emphasizes where this difference is zero (right-hand axis).

backgrounds denoted by “fg4” (black open circles) in Fig. 2 (left). Here, the observations used for comparison are at hourly frequency ignoring latency: these observation sets were produced by separating the 6-hourly GFS dumps into hourly dumps, regardless of the observation’s time of arrival. The RMS differences of the first three catch-up cycles after reinitializing from the control (blue filled circles in Fig. 4) are shown in gray in Fig. 8, and the remaining cycles are shown in orange. The similarity between the orange and gray curves illustrates that reinitializing the catch-up cycles from the control provides negligible performance improvement, unlike results seen using the regional RAP model. This is likely because the main purpose of re-initialization in the RAP is to better represent the large-scale features from the global model that assimilating the regional observations alone cannot (e.g., Schwartz et al. 2020, 2021, 2022). In contrast, the purpose of re-initialization in the global catch-up cycles studied here is to include information from high-latency observations that would be assimilated in the control but not in a symmetric, non-overlapping hourly assimilation system. However, the control that we used to reinitialize the catch-up cycles assimilated the GFS observations with a receipt cutoff time of 2h, 45min after the nominal analysis time (i.e., the receipt cutoff time is nearly equal to the end of the valid observation window, rather than 3 h after the end of the window.) Thus, only low-latency observations are assimilated at the end of the control window. McNally (2019) suggests that the observations at the end of the window are the most important to forecast skill; we are therefore initializing the catch-up cycles from a suboptimal control experiment. Initializing the catch-up cycles from a control with a later receipt cutoff time (such as the GDAS observations with a cutoff of nearly 6 h after the nominal analysis time) could improve the performance of the “early” catch-up cycles, though results in section 4 (below) suggest that the later receipt cutoff time provides only a small improvement in the control performance.

The difference between the performance of real-time catch-up cycles and OW is also shown as the black curve in

Fig. 8. Catch-up cycles consistently have larger errors than OW, except at 0100 UTC, 1300 UTC, and 2000 UTC. We hypothesize that these differences are mainly attributable to the additional (high-latency) observations that OW assimilates. The differences between catch-up cycles and OW at 0100 and 1300 UTC may be smaller because these are 1h forecasts from analyses at 0000 and 1200 UTC, when the receipt cutoff time is 58 min (instead of 26 min; see Table 1) for both methods, allowing more time for observations to arrive. Additionally, most in situ observations (which are generally low-latency) are available at 0000 and 1200 UTC to be assimilated by catch-up cycles, thus reducing the benefit of backward-looking assimilation windows in OW. The small differences at 0700 and 1900–2000 UTC may also be the result of initializing from the 0600 and 1800 UTC analyses, when more observations are available.

## 2) 5-DAY FORECASTS

To investigate the skill of longer forecasts initialized from overlapping window analyses, vertical profiles of the improvement or degradation, relative to the control, of the RMS difference between in situ observations and OW forecasts of full-field wind, temperature, and specific humidity out to 5 days are examined (Fig. 9). These are deterministic forecasts at a resolution of C384 (approximately  $0.25^\circ$ ), initialized from the downscaled C192 ensemble mean analysis. Due to computational constraints, free forecasts were initialized once per day (at 0300 UTC for the control and 0200 UTC for OW). As in Figs. 5 and 7, the control and OW forecasts are each initialized from analyses that have seen approximately the same observations.

Benefits of hourly assimilation over the 6-hourly control are seen in the first 6–12 h of the forecast, as illustrated by the blue shading in wind and temperature. These benefits largely disappear at longer leads, when results are more mixed. Stippling on Fig. 9 denotes areas of significance at the 95% level

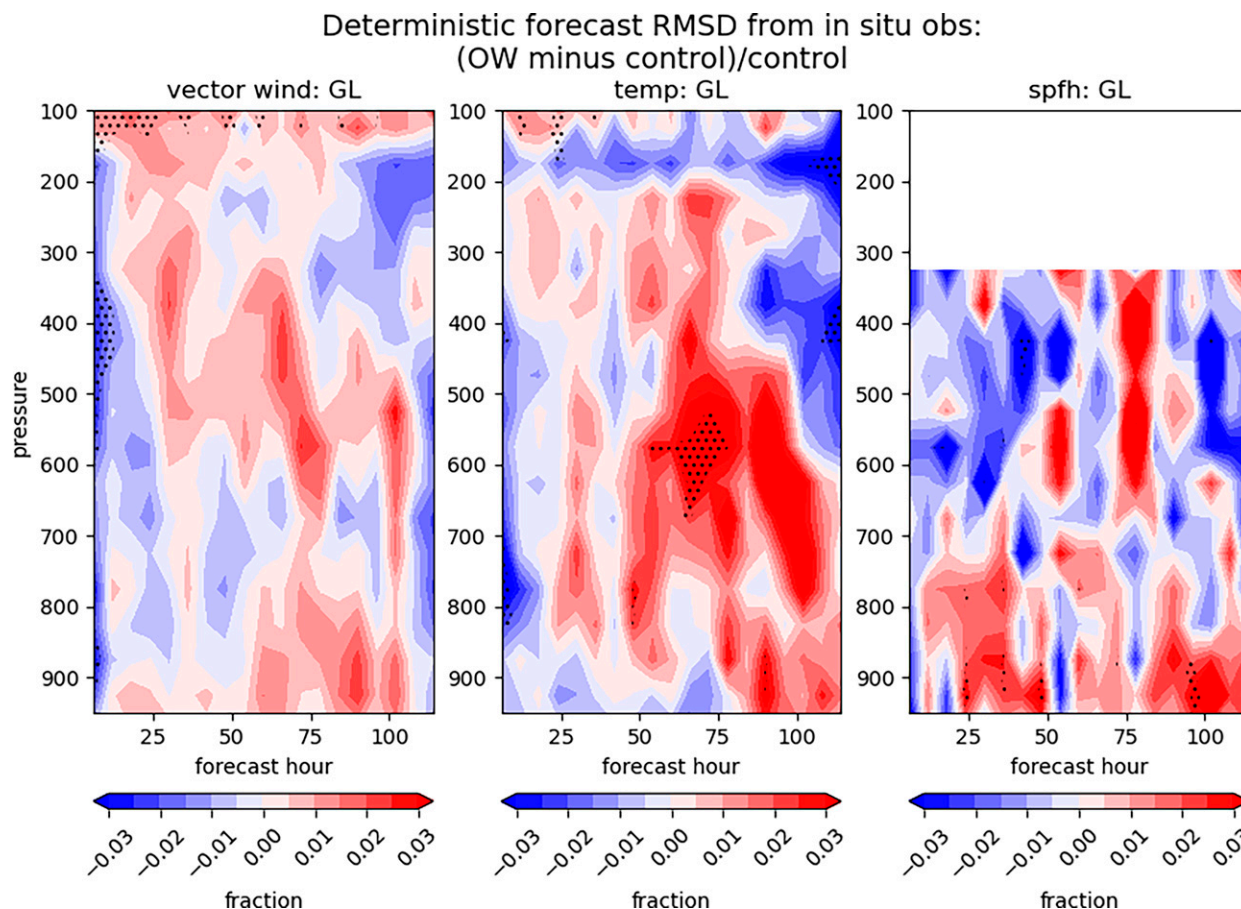


FIG. 9. Contour plot of global relative differences between the long forecast RMSDs of OW and of the control, relative to in situ observations, as a function of forecast hour and averaged over the period 0600 UTC 14 Mar 2020–0600 UTC 12 Apr, for (left) vector wind, (center) temperature, and (right) specific humidity. Blue colors denote that OW has smaller RMSDs than the control, and red colors denote that the control has smaller RMSDs than OW. Stippling denotes where the control and OW experiments differ significantly at the 95% level.

using Student's  $t$  test with inflation as described in (Geer 2016; Wilks 2011). Specifically, the statistical confidence interval is inflated with a factor determined by modeling the autocorrelation with a second-order autoregression model; see appendix C of (Geer 2016). While many of the short-term improvements in wind and temperature are statistically significant, there are also areas of significant degradation, particularly in midlevel temperature at hours 48 and beyond. However, since these results use only 30 forecasts, the estimate of the inflation itself is prone to sampling error (Geer 2016); future studies of forecast skill using this method will require longer experiment periods than the one month shown here. Regardless, given that the OW method has had limited parameter tuning relative to the control method, these results are promising, as the hourly assimilation does not consistently worsen long forecast performance. We anticipate that additional tuning to the OW method could improve results further.

Figure 10a illustrates the gravity wave noise (as measured by global absolute surface pressure tendencies, first discussed in section 3a) that results from removing the 4DIAU in overlapping windows relative to the control experiment, as a

function of forecast lead time. While OW has an increased surface pressure tendency of about  $0.01 \text{ Pa s}^{-1}$  (less than 10% of the control) in the first 12–15 h, this initial growth dissipates over longer forecast leads. Figure 10b shows the time series of the surface pressure tendency in the first three forecast hours for each experiment over the full month. The tendency of the OW experiment is larger than that of the control, but does not consistently increase in time, suggesting that the noise does not build up in a cycling system. These results further support the argument that the 4DIAU may not be necessary in an hourly cycling system with overlapping windows.

#### 4. Causes for improvement in an hourly system

There are likely several factors leading to the significant improvements of OW over the control and catch-up systems. First, OW cumulatively assimilates more observations than the control or catch-up experiments. This is because the 6-hourly systems used in the control and for reinitialization in catch-up cycles each use the operational GFS observation dataset, which has a 6-h valid observation window defined as

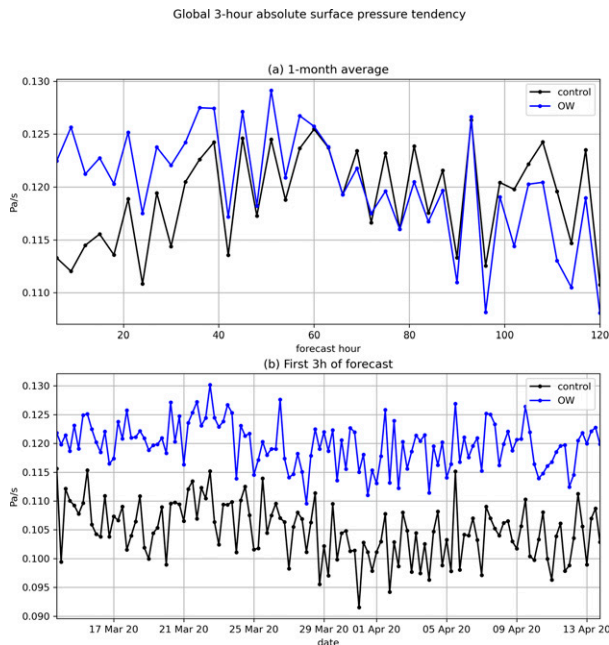


FIG. 10. Globally averaged 3-h absolute surface pressure tendency for the control experiment (black) and OW (blue) (a) as a function of forecast lead time, averaged over the 1-month period 14 Mar–13 Apr 2020, and (b) for the first 3 h of the forecast as a function of date–time for the cycled experiment.

$[-3 \text{ h}, 3 \text{ h}]$  symmetrically centered around the nominal analysis time and includes observations with receipt times of no later than approximately 2 h 45 min past the nominal analysis time. With this receipt cutoff time, high-latency observations at the end of the 6h window will not be assimilated into the 6-hourly system, but OW would assimilate those observations in the subsequent cycles.

The second factor is simply the hourly frequency of updating, which leads to shorter (and possibly more accurate) forecasts in the background fields. Updating the background

forecast hourly instead of every 6 h should result in a reduction of the detrimental effects of nonlinearity, both in the evolution of background-error covariance and in the linearization of the observation operator (e.g., Bonavita et al. 2018).

To investigate these two factors, a set of additional experiments is performed. Each experiment in this section uses the same full-input, hybrid-gain LETKF setup as in section 3b, but with RTPS and RTPP inflation parameters of 0.9 and 0.5, respectively, and a weighting factor of 20% is assigned to the 3DVar gain (Table 2). These experiments were run for a period of 9 days (13 March 2020–22 March 2020). The first experiment, “control\_gfs,” uses the 6-hourly system with 4DIAU and assimilates the GFS observations (as in section 3b). The second experiment, “control\_gdas,” uses the same 6-hourly system but assimilates the operational GDAS dataset, which has the same 6-hourly observation windows as the GFS but with a longer receipt cutoff time of 5h 50min past the center of the window (about 3h later than the GFS cutoff). A brief examination of the observation files suggests that the GDAS dataset includes about 15% more observations than the GFS dataset due to the later receipt cutoff time. Figure 11 (black curve and gray shading) shows that assimilating the additional observations from GDAS somewhat improves the 6-h RMS fit to in situ observations relative to the earlier GFS cutoff (as shown by the negative values of the difference between the control\_gdas and control\_gfs RMSDs).

The next experiments, “hourly\_gfs” and “hourly\_gdas,” assimilate the identical GFS or GDAS observations as the control experiments, but the background fields are updated each hour using a non-overlapping hourly cycling system by splitting the GFS or GDAS dumps into 6 subsets. For example, consider the 0600 UTC GFS dump window with observations valid from 0300 to 0900 UTC. The hourly\_gfs experiment will assimilate observations from this dump window that are valid from 0300 to 0430 UTC to generate an analysis at 0400 UTC. A 1-h forecast is initialized from this analysis to generate a background field at 0500 UTC. Observations from the same GFS dump window that are valid from

TABLE 2. Summary of experiments shown in section 4.

Expt name	Inflation (RTPS/ RTPP)	3DVar weight (%)	Obs types assimilated	Obs window (relative to nominal analysis time)	Obs receipt cutoff (relative to nominal analysis time)
Control_gfs	0.9/0.5	20	All	$[-3 \text{ h}, 3 \text{ h}]$	2 h 45 min (GFS)
Control_gdas	0.9/0.5	20	All	$[-3 \text{ h}, 3 \text{ h}]$	5 h 50 min (GDAS)
Hourly_gfs	0.9/0.5	20	All	$[-90 \text{ min}, 30 \text{ min}]$ for 0400, 1000, 1600, 2000 UTC; $[-30 \text{ min}, 0]$ for 0300, 0900, 1500, 2100 UTC; $[-30 \text{ min}, 30 \text{ min}]$ for all other times (see text for details)	2 h 45 min after 0000, 0600, 1200, 1800 UTC (GFS)
Hourly_gdas	0.9/0.5	20	All	$[-90 \text{ min}, 30 \text{ min}]$ for 0400, 1000, 1600, 2000 UTC; $[-30 \text{ min}, 0]$ for 0300, 0900, 1500, 2100 UTC; $[-30 \text{ min}, 30 \text{ min}]$ for all other times (see text for details)	5 h 50 min after 0000, 0600, 1200, 1800 UTC (GDAS)
OW	0.9/0.5	20	All	$[-3 \text{ h}, 30 \text{ min}]$	58 min (0000 and 1200 UTC); 26 min (all other hours)

## RMS O-F

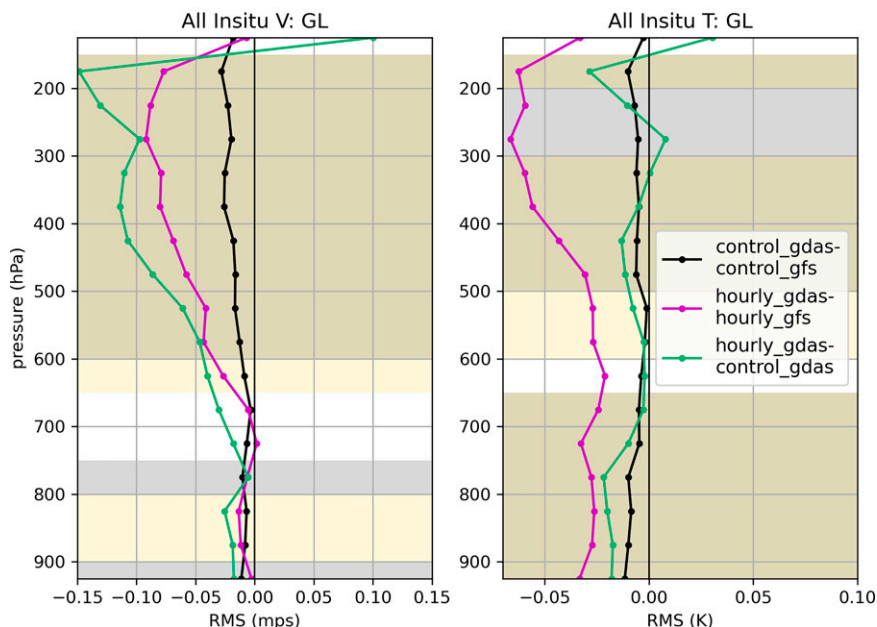


FIG. 11. Vertical profiles of differences of short forecast RMSDs with respect to global in situ observations of (left) vector wind ( $\text{m s}^{-1}$ ) and (right) temperature (K) for control\_gdas minus control\_gfs (black), hourly\_gdas minus hourly\_gfs (purple), and hourly\_gdas minus control\_gdas (green) in the full-input hybrid-gain LETKF setup, averaged over the time period 1200 UTC 13 Mar–1800 UTC 22 Mar 2020. Gray shading denotes where control\_gfs and control\_gdas differ significantly at the 95% level, and yellow shading denotes where hourly\_gfs and hourly\_gdas differ significantly at the 95% level.

0430 to 0530 UTC are assimilated to generate the analysis at 0500 UTC. The cycling continues in this way until 0900 UTC, which assimilates observations in the 0600 UTC GFS dump valid from 0830 to 0900 UTC to generate the analysis at 0900 UTC. The hourly\_gdas experiment proceeds in the same way, but assimilates observations from the GDAS dumps instead of the GFS dumps. These experiments were performed in this way to avoid any further processing of the GFS and GDAS dumps that could lead to slight differences in the observations assimilated.

The difference between RMSDs of the hourly\_gfs and hourly\_gdas experiments is shown in the pink curve in Fig. 11 (yellow shading emphasizes significant differences). Note that assimilating the larger GDAS observation dataset in the hourly system generally has a much greater effect than in the control system, particularly in mid- to upper-tropospheric winds and temperature at all levels (Fig. 11, compare pink and black curves).

By comparing hourly\_gdas with control\_gdas, the impact of updating with an hourly frequency can be isolated from the effect of assimilating extra observations. Figure 11 (green curve) shows that hourly\_gdas outperforms control\_gdas in wind and temperature at nearly all levels. This seems to suggest that the hourly updating frequency alone would provide significant improvement over a 6-hourly system; however, the hourly system using catch-up cycles

does not significantly improve over the 6-hourly control (Fig. 7). Instead, a comparison between the pink and black curves in Fig. 11 suggests that an hourly cycling system can better handle late-arriving observations than a 6-hourly system.

To determine how much improvement would be possible in an hourly system with reduced data latency, hourly\_gdas results are compared with those from OW (Fig. 12). Here, both experiments update with an hourly frequency, but hourly\_gdas essentially assimilates observations without data latency (recall that the GDAS data cutoff is 5 h 50 min after the nominal analysis time), while OW assimilates real-latency observations. Additionally, note that the end of the observation window for OW is about 30 min earlier than the end of the observation window used for hourly\_gdas (e.g., the 0000 UTC GDAS observation window would include observations valid until 0300 UTC, while the comparable OW observation window would only include observations valid until 0230 UTC). Comparisons between these two experiments (Fig. 12) therefore provide an upper-bound on the improvement that could be gained in an hourly system from reducing latency of observations. Results show that this potential improvement is smaller than the improvement gained by increasing the frequency of updates from 6-hourly to 1-hourly, suggesting that the overlapping-windows technique is largely successful in overcoming the issue of data



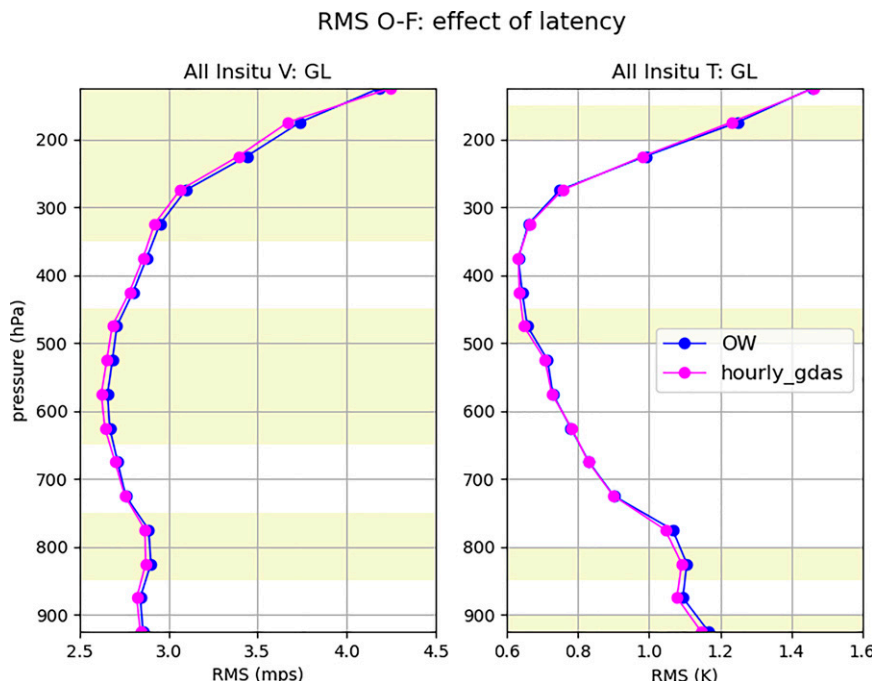


FIG. 12. Vertical profiles of short forecast comparisons with respect to global in situ observations of (left) vector wind (left) and (right) temperature for the OW (blue) and hourly\_gdas (purple) experiments in the full-input hybrid-gain LETKF setup, averaged over the time period 1200 UTC 13 Mar–1800 UTC 22 Mar 2020, to demonstrate the effects of data latency in an hourly system. Yellow shading illustrates where the two experiments differ significantly at the 95% level.

latency, though there is some room for improvement with decreased latency.

## 5. Discussion

A global hourly data assimilation system using an overlapping windows method to overcome the challenge of data latency was tested in a reduced resolution version of the NOAA operational forecast system using GFSv15 and a hybrid-gain 3DVar/LETKF assimilation algorithm. Despite minimal tuning of assimilation parameters, static background covariances, and bias correction methods, the overlapping windows method has comparable performance to a 6-hourly control system. In fact, the overlapping windows method improves upon a 6-hourly assimilation system in the context of short-term (less than 12 h) fits to observations. The improvements are greatest in winds and temperature in the upper troposphere, similar to what was found using the regional hourly updating RAP system. The benefits largely disappear at longer forecast leads, when results are more mixed. Longer experiments are needed to robustly estimate significance at those lead times.

Though the overlapping windows method had a slight increase in gravity wave noise without the 4DIAU relative to OW-IAU, this noise did not degrade the performance in short-term fits to observations, and the effects diminished at longer forecast leads. That the gravity wave noise did not

degrade performance is likely because the hourly system has smaller analysis increments than a 6-hourly system, in which some smoothing (from the 4DIAU or digital filtering) is required. As shown in section 3a, the 4DIAU cannot be straightforwardly applied in the overlapping windows framework without a loss of performance. While modifications to the 4DIAU procedure may be possible to minimize this loss of information, the 4DIAU adds computational burden and does not appear to be needed in an hourly OW framework.

Catch-up cycles were also investigated, in which an hourly cycling system is reinitialized from a control system every 6 h. This method had comparable results to the 6-hourly control, suggesting that any benefits gained from cycling hourly are lost when the system does not assimilate high-latency observations. Results from catch-up cycles also suggest that initializing from a control experiment does not significantly improve the results, though this may be because the control experiment assimilated only low-latency observations at the end of its assimilation windows.

An additional set of experiments suggests that the hourly frequency of updates provides most of the improvement in OW over the control system. However, this is not the whole story, since the catch-up cycle method does not outperform the 6-hourly control. The overlapping windows method likely outperforms both the control and catch-up cycles due to the interaction between additional observations and hourly cycling. Results comparing the effects of assimilating additional

observations in a 6-hourly system to those in a 1-hourly system suggest that an hourly cycling system can use higher-latency observations more effectively than a 6-hourly system by using data within the first hour that it becomes available. This particularly motivates future investigations of assimilating more high-temporal-frequency observations in an hourly system than are currently assimilated in the global system. Results from hourly experiments with different data latencies also suggest that an hourly global system would be beneficial regardless of whether data latency decreases in the future (with faster data communication links, for instance.)

This work has focused on the fairest possible comparisons between a real-time hourly system and the current 6-hourly GFS configuration, and therefore results are only shown from forecasts initialized one to four times per day, at nearly the same time. However, one significant benefit of a real-time hourly system over a 6-hourly system is that more accurate analyses will be available in real-time at every hour of the day. For example, at 1200 UTC, a forecaster would have access to the 1100 UTC hourly update cycle, which will have assimilated observations valid until 1130 UTC. In the current 6-hourly system, the most recent analysis available would be valid at 0900 UTC (the end of the 0600 UTC GDAS cycle), which would have assimilated observations valid until 0900 (including late-arriving observations). Regardless of data latency, the hourly updating system can provide more up-to-date, and therefore better-informed, fields than the 6-hourly system at most hours of the day. This could significantly improve weather situational awareness with particular impacts in the aviation sector, e.g., as described in (Benjamin et al. 2016).

These results suggest that further investigation of upgrading the GFS to an hourly system would be worthwhile, but they are also promising for the Rapid Refresh Forecast System (RRFS), NOAA's planned next-generation rapidly updated convection-allowing ensemble forecast system. While hourly updated lateral boundaries would benefit the limited area modeling approach (Black et al. 2021), the extension of overlapping windows to the RRFS data assimilation system may also obviate the longstanding need for complex partial cycling procedures.

Further work is needed to determine how viable overlapping windows would be for the operational GFS. More tuning of, for instance, the stochastic physics parameters, localization and inflation, static background error, and bias correction algorithms could be investigated and potentially provide further improvement. In particular, the radiance bias correction in the OW system may need modification due to the inhomogeneity in time of the in situ observations used for anchoring radiance observations (although there was no suspicious behavior in the time series of the bias correction coefficients used in the OW experiments; not shown). Additionally, the added computational burden of running an hourly global system is nontrivial, and would require extra attention to minimizing data transfer bottlenecks.

Future work will address the anticipated benefits of an operational global hourly assimilation system over a 6-hourly system in two particular contexts: initialization of regional model forecasts and hurricane forecasts. Experiments with regional model forecasts will determine whether a global

hourly system can provide better lateral boundary conditions to a high-resolution regional model than the current RAP, and case studies of significant hurricanes using the global OW system (at higher spatial resolution than implemented here) will demonstrate impacts of hourly cycling on fast-moving storms. Finally, the results shown here assimilated only the observations that are currently assimilated in the global system: we hypothesize larger improvements over the control system when assimilating additional high-frequency observations in an hourly system.

*Acknowledgments.* This work was supported by NOAA Grant NA19OAR0220182 and the NOAA Cooperative Agreement with CIRES, NA17OAR4320101. Support was also provided by the NOAA/Physical Sciences Laboratory and the NOAA/Global Systems Laboratory. Computational resources were provided by the NOAA Research and Development High Performance Computing System and the High Performance Computing Collaboratory at Mississippi State University. The authors gratefully acknowledge Shelley Melchior and the EMC ObsProc team for their work providing asymmetric observation dumps so that these experiments could test effects of real-time data latency. Chesley McColl (CIRES/NOAA PSL) is acknowledged for her help with filtering and post-processing these observations for assimilation. Sergey Frolov (NOAA PSL) is acknowledged for comments on an earlier version of this paper. We thank Stan Benjamin and Steven Weygandt (NOAA GSL) for helpful discussions. We also thank three anonymous reviewers for comments that improved this paper.

*Data availability statement.* The data used in this research are housed within the NOAA Research and Development High Performance Computing System and are available upon request.

## APPENDIX

### Performance of 4DIAU in Overlapping Windows

To delve into the causes of the differences between OW-IAU and OW, an experiment was run beginning from identical initial conditions for two sequential assimilation cycles, with and without the IAU, using the configuration described in section 3a (NoSat, pure LETKF). All stochastic forcing was turned off in this experiment. Background fields were generated for 0700–1100 UTC 13 March 2020 to span the first assimilation window (see Fig. A1). Each experiment was initialized from these (identical) background fields and then assimilated identical observations (the 3.5-h observation window from 0700 to 1030 UTC, as described in section 2 and Table 1), resulting in identical analysis fields at 0800, 0900, and 1000 UTC. For OW-IAU, the system was then initialized from the background field at 0700 UTC and run with IAU forcing until 1100 UTC, at which point the IAU forcing was turned off and the system ran in free-forecast mode for 6 h (through 1700 UTC). For OW, the analyses from 0800 to 1000 UTC were used as background fields for the subsequent cycle; the system was then initialized from the analysis

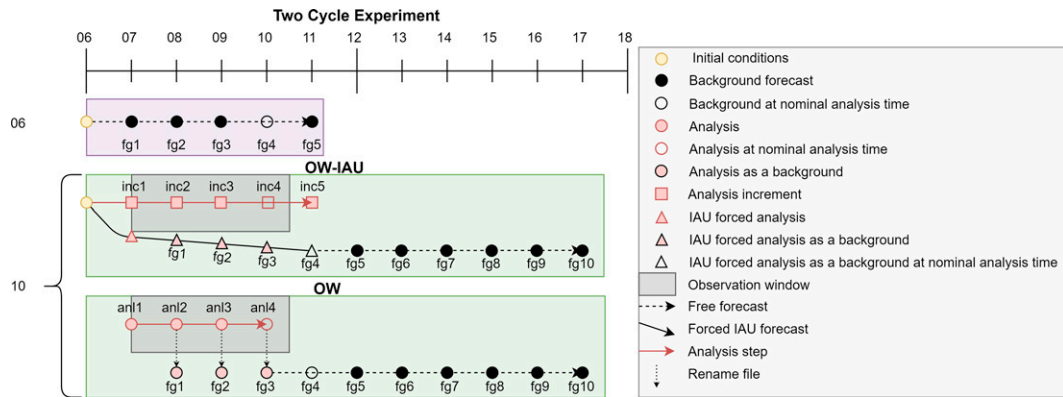


FIG. A1. Diagram for the two-cycle experiment comparing OW-IAU with OW.

field at 1000 UTC and ran as a free forecast from 1100 to 1700 UTC. During the second assimilation cycle, each experiment assimilates observations from 0800 to 1130 UTC; generates analyses at 0900, 1000, and 1100 UTC; and generates backgrounds at 1200 UTC with or without IAU forcing, respectively.

To compare OW-IAU with OW hour-by-hour, these fields were compared with a set of hourly observations with all latency removed: these observation sets were produced by separating the 6-hourly GFS dumps into hourly dumps, regardless of the observation's time of arrival. These are therefore not fits to *independent* observations, since many

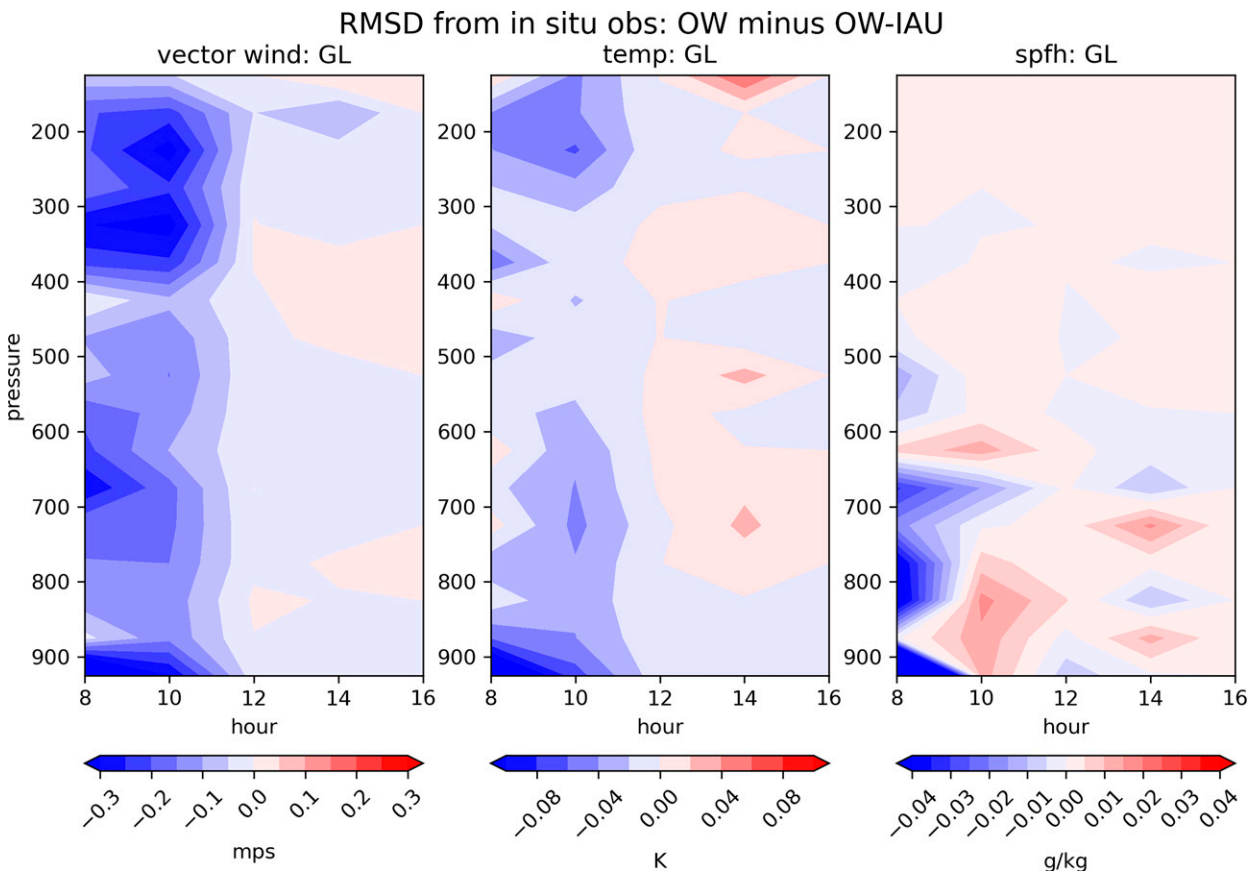


FIG. A2. Contour plot showing the difference between OW and OW-IAU RMSDs with global in situ observations of (left) vector wind, (center) temperature, and (right) specific humidity as function of hour in the first cycle of the two-cycle experiment described in the text. Blue colors denote that OW has smaller RMSDs, and red colors denote that OW-IAU has smaller RMSDs. Note that only even hours are shown; odd hours did not have enough data to calculate statistics.

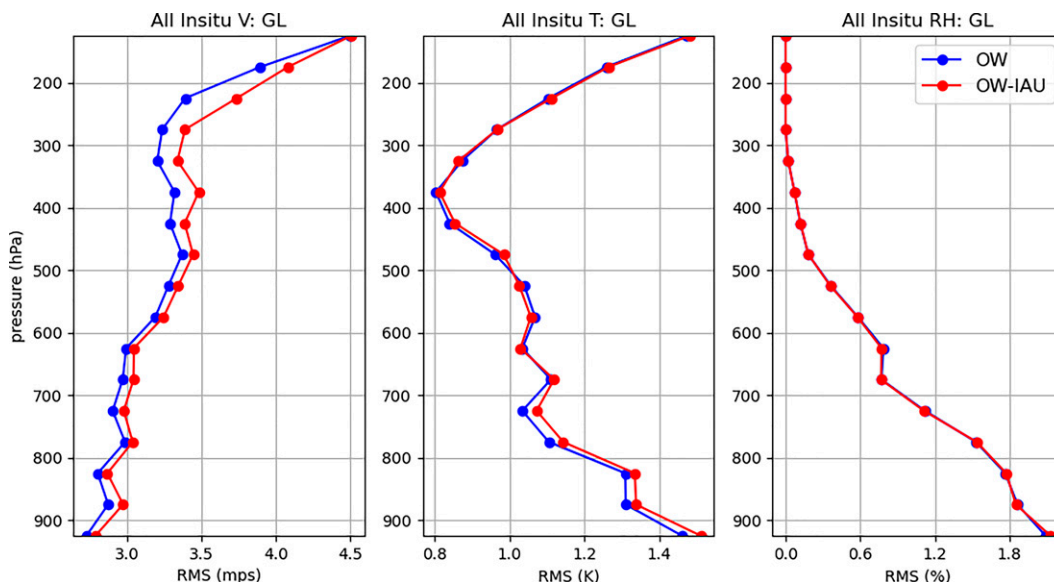


FIG. A3. Vertical profiles of RMS 1-h forecast differences with respect to global in situ observations of (left) vector wind, (center) temperature, and (right) relative humidity for OW (blue) and OW-IAU (red) for the second cycle of the two-cycle experiment described in the text.

of these observations were assimilated in the first cycle, so the errors themselves are not as informative as the difference between the OW-IAU errors and the OW errors; that difference will be entirely due to the effect of the IAU. Figure A2 shows the effect of the IAU: filled contours show the difference between the OW-IAU RMS fit to hourly observations and the OW RMS fit to hourly observations, for different forecast lead times. Blue colors signify that OW is closer to the observations than OW-IAU.

Figure A2 demonstrates that the largest differences between the OW-IAU and OW fields are at 0800 and 1000 UTC, when OW-IAU has larger errors than OW. This is because the OW fields at these times are the original analysis fields, while the OW-IAU fields are the IAU-forced analysis fields, which will not have been informed by all of the observations until the end of the window at 1100 UTC. The errors that correspond to the free forecasts (1200–1600 UTC) are very similar between OW-IAU and OW; this is because both experiments have been informed by all of the observations by 1200 UTC, and suggests that a free forecast from a raw analysis is close to a free forecast from an IAU-forced analysis.

However, the differences at 0800–1000 UTC affect the second cycle, since these fields are used as the backgrounds for the second cycle. Figure A3 compares the 1-h forecast fit with hourly observations at 1200 UTC, initialized from the first analysis cycle at 1100 UTC. At this point, all of the 1100 UTC observations have been assimilated in both experiments, but the OW errors are consistently smaller than the OW-IAU errors. This effect likely compounds over many cycles, leading to the differences seen in the previous section. Note that this effect is specific to the

overlapping windows method: in the control non-overlapping method with 4DIAU, the full IAU forcing has been included by the beginning of the next assimilation cycle, guaranteeing that each of the background fields for the next cycle will include all available information from the current observations.

## REFERENCES

- Alexander, C., and Coauthors, 2020: Rapid Refresh (RAP) and High-Resolution Rapid Refresh (HRRR) model development. *30th Conf. on Weather Analysis and Forecasting (WAF)/26th Conf. on Numerical Weather Prediction (NWP)*, Boston, MA, Amer. Meteor. Soc., 8A.1, <https://ams.confex.com/ams/2020Annual/webprogram/Paper370205.html>.
- Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, <https://doi.org/10.1175/MWR-D-15-0242.1>.
- Berner, J., G. J. Shutts, M. Leutbecher, and T. N. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, **66**, 603–626, <https://doi.org/10.1175/2008JAS2677.1>.
- Black, T. L., and Coauthors, 2021: A limited area modeling capability for the finite-volume cubed-sphere (FV3) dynamical core and comparison with a global two-way nest. *J. Adv. Model. Earth Syst.*, **13**, e2021MS002483, <https://doi.org/10.1029/2021MS002483>.
- Bocquet, M., C. A. Pires, and L. Wu, 2010: Beyond Gaussian statistical modeling in geophysical data assimilation. *Mon. Wea. Rev.*, **138**, 2997–3023, <https://doi.org/10.1175/2010MWR3164.1>.
- Bonavita, M., P. Lean, and E. Holm, 2018: Nonlinear effects in 4D-Var. *Nonlinear Processes Geophys.*, **25**, 713–729, <https://doi.org/10.5194/npg-25-713-2018>.



- Cardinali, C., 2009: Monitoring the observation impact on the short-range forecast. *Quart. J. Roy. Meteor. Soc.*, **135**, 239–250, <https://doi.org/10.1002/qj.366>.
- Chen, Y., and C. Snyder, 2007: Assimilating vortex position with an ensemble Kalman filter. *Mon. Wea. Rev.*, **135**, 1828–1845, <https://doi.org/10.1175/MWR3351.1>.
- Cummings, J. A., and O. M. Smedstad, 2013: Variational data assimilation for the global ocean. *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, Vol. II, S. K. Park and L. Xu, Eds., Springer, 303–343.
- Fisher, M., Y. Trémolet, H. Auvinen, D. Tan, and P. Poli, 2011: Weak-constraint and long-window 4D-Var. ECMWF Tech. Memo. 655, 49 pp., <https://www.ecmwf.int/sites/default/files/elibrary/2011/9414-weak-constraint-and-long-window-4dvar.pdf>.
- Geer, A. J., 2016: Significance of changes in medium-range forecast scores. *Tellus*, **68A**, 30229, <https://doi.org/10.3402/tellusa.v68.30229>.
- Haseler, J., 2004: Early-delivery suite. ECMWF Tech. Memo. 454, 37 pp., <https://www.ecmwf.int/sites/default/files/elibrary/2004/9793-early-delivery-suite.pdf>.
- He, H., L. Lei, J. S. Whitaker, and Z.-M. Tan, 2020: Impacts of assimilation frequency on ensemble Kalman filter data assimilation and imbalances. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002187, <https://doi.org/10.1029/2020MS002187>.
- James, E. P., and S. G. Benjamin, 2017: Observation system experiments with the hourly updating Rapid Refresh model using GSI hybrid ensemble–variational data assimilation. *Mon. Wea. Rev.*, **145**, 2897–2918, <https://doi.org/10.1175/MWR-D-16-0398.1>.
- , —, and B. D. Jamison, 2020: Commercial-aircraft-based observations for NWP: Global coverage, data impacts, and COVID-19. *J. Appl. Meteor. Climatol.*, **59**, 1809–1825, <https://doi.org/10.1175/JAMC-D-20-0010.1>.
- Kim, S.-M., and H. M. Kim, 2019: Forecast sensitivity observation impact in the 4DVAR and hybrid-4DVAR data assimilation systems. *J. Atmos. Oceanic Technol.*, **36**, 1563–1575, <https://doi.org/10.1175/JTECH-D-18-0240.1>.
- Kleist, D. T., and K. Ide, 2015: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part II: 4D-EnVar and hybrid variants. *Mon. Wea. Rev.*, **143**, 452–470, <https://doi.org/10.1175/MWR-D-13-00350.1>.
- , D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP Global Data Assimilation System. *Wea. Forecasting*, **24**, 1691–1705, <https://doi.org/10.1175/2009WAF2222201.1>.
- Lean, P., E. V. Hölm, M. Bonavita, N. Bormann, A. P. McNally, and H. Järvinen, 2021: Continuous data assimilation for global numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **147**, 273–288, <https://doi.org/10.1002/qj.3917>.
- Lei, L., and J. S. Whitaker, 2016: A four-dimensional incremental analysis update for the ensemble Kalman filter. *Mon. Wea. Rev.*, **144**, 2605–2621, <https://doi.org/10.1175/MWR-D-15-0246.1>.
- Lin, S.-J., 2004: A “vertically Lagrangian” finite-volume dynamical core for global models. *Mon. Wea. Rev.*, **132**, 2293–2307, [https://doi.org/10.1175/1520-0493\(2004\)132<2293:AVLFDC>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2).
- Lin, H., S. S. Weygandt, S. G. Benjamin, and M. Hu, 2017: Satellite radiance data assimilation within the hourly updated Rapid Refresh. *Wea. Forecasting*, **32**, 1273–1287, <https://doi.org/10.1175/WAF-D-16-0215.1>.
- Liu, Q., S. Lord, N. Surgi, Y. Zhu, R. Wobus, Z. Toth, and T. Marchok, 2006: Hurricane relocation in global ensemble forecast system. *27th Conf. on Hurricanes and Tropical Meteorology*, Monterey, CA, Amer. Meteor. Soc. P5.13, [https://ams.confex.com/ams/27Hurricanes/techprogram/paper\\_108503.htm](https://ams.confex.com/ams/27Hurricanes/techprogram/paper_108503.htm).
- 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., 2019: On the sensitivity of a 4D-Var analysis system to satellite observations located at different times within the assimilation window. *Quart. J. Roy. Meteor. Soc.*, **145**, 2806–2816, <https://doi.org/10.1002/qj.3596>.
- Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. ECMWF Tech. Memo. 598, 42 pp., <https://doi.org/10.21957/ps8gbwbdv>.
- Payne, T. J., 2017: Rapid update cycling with delayed observations. *Tellus*, **69A**, 1409061, <https://doi.org/10.1080/16000870.2017.1409061>.
- Penny, S. G., 2014: The hybrid local ensemble transform Kalman filter. *Mon. Wea. Rev.*, **142**, 2139–2149, <https://doi.org/10.1175/MWR-D-13-00131.1>.
- Poterjoy, J., 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>.
- Putman, W. M., and S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. *J. Comput. Phys.*, **227**, 55–78, <https://doi.org/10.1016/j.jcp.2007.07.022>.
- Ruiz, J., G.-Y. Lien, K. Kondo, S. Otsuka, and T. Miyoshi, 2021: Reduced non-Gaussianity by 30 s rapid update in convective-scale numerical weather prediction. *Nonlinear Processes Geophys.*, **28**, 615–626, <https://doi.org/10.5194/npg-28-615-2021>.
- Schwartz, C. S., M. Wong, G. S. Romine, R. A. Sobash, and K. R. Fossell, 2020: Initial conditions for convection-allowing ensembles over the conterminous United States. *Mon. Wea. Rev.*, **148**, 2645–2669, <https://doi.org/10.1175/MWR-D-19-0401.1>.
- , G. S. Romine, and D. C. Dowell, 2021: Toward unifying short-term and next-day convection-allowing ensemble forecast systems with a continuously cycling 3-km ensemble Kalman filter over the entire conterminous United States. *Wea. Forecasting*, **36**, 379–405, <https://doi.org/10.1175/WAF-D-20-0110.1>.
- , 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>.
- Shutts, G., M. Leutbecher, A. Weisheimer, T. Stockdale, L. Isaksen, and M. Bonavita, 2011: Representing model uncertainty: Stochastic parameterizations at ECMWF. *ECMWF Newsletter*, No. 129, ECMWF, Reading, United Kingdom, 19–24.
- Smith, T. L., S. G. Benjamin, J. M. Brown, S. Weygandt, T. Smirnova, and B. Schwartz, 2008: Convection forecasts from the hourly updated, 3-km High Resolution Rapid Refresh model. *24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 11.1, [https://ams.confex.com/ams/24SLS/techprogram/paper\\_142055.htm](https://ams.confex.com/ams/24SLS/techprogram/paper_142055.htm).

- Tompkins, A. M., and J. Berner, 2008: A stochastic convective approach to account for model uncertainty due to unresolved humidity variability. *J. Geophys. Res.*, **113**, D18101, <https://doi.org/10.1029/2007JD009284>.
- Trémolet, Y., 2006: Accounting for an imperfect model in 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **132**, 2483–2504, <https://doi.org/10.1256/qj.05.224>.
- Whitaker, J. S., and T. M. Hamill, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089, <https://doi.org/10.1175/MWR-D-11-00276.1>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.
- Wong, M., G. Romine, and C. Snyder, 2020: Model improvement via systematic investigation of physics tendencies. *Mon. Wea. Rev.*, **148**, 671–688, <https://doi.org/10.1175/MWR-D-19-0255.1>.
- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253, [https://doi.org/10.1175/1520-0493\(2004\)132<1238:IOIEAO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2).
- 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>.
- , —, R. J. Purser, B. A. Ballish, and J. Whiting, 2015: Variational correction of aircraft temperature bias in the NCEP's GSI analysis system. *Mon. Wea. Rev.*, **143**, 3774–3803, <https://doi.org/10.1175/MWR-D-14-00235.1>.