The Impact of Data Latency on Operational Global Weather Forecasting

SEAN P. F. CASEY[a,b] AND LIDIA CUCURULL[b]

[a] *Cooperative Institute for Marine and Atmospheric Studies, Miami, Florida*
[b] *NOAA/OAR/Atlantic Oceanographic and Meteorological Laboratory, Miami, Florida*

(Manuscript received 14 September 2021, in final form 8 April 2022)

ABSTRACT: The impact of low data latency is assessed using observations assimilated into the NCEP Finite-Volume Cubed-Sphere Global Forecast System (FV3GFS). Operationally, a full dataset is used to generate short-term (9-h) forecasts used as the background state for the next cycle, and a limited dataset with fewer observations is used for long-term (16-day) forecasts due to time constraints that exist in an operational setting. In this study, the sensitivity of the global weather forecast skill to the use of the full and limited datasets in both the short- and long-term forecasts (out to 10 days only) is evaluated. The results show that using the full dataset for long-term forecasts yields a slight improvement in forecast skill, while using the limited dataset for short-term forecasts yields a significant degradation. This degradation is primarily attributed to a decrease of in situ observations rather than remotely sensed observations, though no individual observation type captures the amount of degradation noted when all observations are limited. Furthermore, limiting individual types of in situ observations (aircraft, marine, rawinsonde) does not result in the level of degradation noted when limiting all in situ observations, demonstrating the importance of data redundancy in an operational observational system.

SIGNIFICANCE STATEMENT: Millions of observations are used in global models every day to understand the state of the atmosphere. These observations rely on quick transmission from observation source to weather centers for inclusion in operational models. For this study, we test how different groups of observations, which arrive at the model center at different times, impact the model forecast. We find that by not using the observations that take longer to arrive at the weather centers, the forecast is much worse, showing the importance of quick transmission of observations. Direct observations (those measured within the atmosphere) have a greater impact than remote observations (those viewed from afar, such as by satellites). However, no single observation type by itself causes a poor forecast by being limited, showing the importance of using different types of observations to capture the state of the atmosphere.

KEYWORDS: Numerical weather prediction/forecasting; Data assimilation

1. Introduction

In weather forecasting, an analysis can be defined as the best estimate of the true state of the atmosphere at a given time using all available appropriate data. Both observations and previous knowledge of the atmosphere are combined taking into account the corresponding error structures. Not only are analyses useful as a self-consistent diagnostic of the atmosphere, but they are also used as input (the initial state) for a forecast.

Observations of the actual state of the atmosphere are used to produce an analysis. In most cases, however, this is an undersampled system because the observations are sparse and not always directly related to model variables. To better resolve the current state of the atmosphere, the actual observations are usually supplemented with prior information of the atmospheric state, typically known as the "background field." Some of the ways this information can be determined are from climatology, a previously generated forecast, or a steady-state standard atmosphere. The observations are then used to adjust the background estimate to a more accurate analysis of the atmosphere, considering the uncertainties associated with the observations and the background. This process of using observations to numerically adjust a background state to create an analysis is called data assimilation (Kalnay 2002; Haupt et al. 2017).

Observations used in NWP can be split into two main categories: direct, or in situ measurements, and remotely sensed measurements. In situ ("situated in the original place") measurements are taken at the actual location within the atmosphere, and usually include surface pressure, wind, temperature, and moisture measurements. Rawinsondes have long been used to provide reliable in situ measurements of temperature, moisture, and winds at regular intervals at numerous locations (stations) over large swaths of Earth's land surfaces (and occasionally from ships as well). Marine surface observations, such as those measured from moving ships or buoys, provide vital coverage of various locations in the 67% of the globe covered by water, where few (if any) rawinsondes are available.

In recent decades, measurements of temperature, moisture, and winds provided by commercial aircraft have provided an additional important source of information used in data assimilation, helping to improve the NWP forecasts. This is particularly true during aircraft takeoffs and landings, when vertical profiles are measured as the plane climbs or descends through different levels of the troposphere. Petersen (2016) showed how the introduction of commercial aircraft observations a few decades ago has contributed to a near-50% reduction in forecast error in the NOAA GFS and Met Office (UKMO) models. James and Benjamin (2017, their Fig. 4) found that aircraft measurements provided nearly 3 times the error reduction for wind as the second-best observation source, with

Corresponding author: Sean P. F. Casey, Sean.Casey@noaa.gov

DOI: 10.1175/WAF-D-21-0149.1

similar reductions of error for relative humidity and temperature in the hourly updated Rapid Refresh assimilation-forecast system over the entire regional domain.

In contrast to in situ, i.e., directly measured observations, remotely sensed observations are measured from some distance away, typically from satellites, which employ a variety of observational instruments. Data from these satellites, such as from *Suomi-NPP* (Weng et al. 2012), and the MetOp series of satellites (Klaes et al. 2013), have become one of the dominant sources for improving NWP (English et al. 2013) forecasts. Cucurull and Anthes (2014) showed that IR, microwave (MW), and radio occultation (RO) observations had the largest impact in the Southern Hemisphere, with IR and MW observations having a small but significant impact in the Northern Hemisphere. Other studies, such as Garrett (2013), have quantified the benefits of satellite observations from new polar-orbiting satellites as mitigation strategies for the loss of observations due to the retirement of older satellite systems such as *NOAA-15, NOAA-18,* and *NOAA-19*.

A critical factor for all of these observations in impacting weather prediction is low data latency, or the length of time between the observation measurement and when the observation is received by the forecasting center. In operational NWP, data needs to be quickly downloaded from satellites to ground stations and ultimately transmitted to operational NWP centers. As an example, it is possible to obtain low data latency from polar orbit satellites through direct broadcast (DB) sites over CONUS and adjacent regions (Huang et al. 2016; Han and Jochum 2017). Lin et al. (2017) showed that radiances from polar-orbiting satellites carrying an AMSU-A instrument assimilated hourly in a "Rapid Refresh" model have significant impacts in terms of model error reduction.

Low data latency ensures that a larger percentage of observations can be assimilated into operational NWP models in an effort to improve weather forecast skill. On the regional scale, Wang et al. (2020) recently showed that the assimilation of low latency sounder observations into the WRF model (Powers et al. 2017) resulted in the best forecast skill (compared to other observation sources) when studying a mesoscale convective system over the central Great Plains. However, similar studies with a global-scale NWP model have not yet been conducted. Thus, the goal of this study is to investigate the sensitivity of the global-scale NWP forecast skill to the latency of the observations for different datasets in the operational data assimilation and forecast system at NOAA.

This paper is organized as follows. Section 2 describes the details of the NWP system and the different experiments conducted in this study. Sections 3 and 4 investigate the impacts of the different datasets in improving NWP forecast skill. Finally, section 5 summarizes the results and conclusions.

2. Methodology

All experiments in this study utilized the NCEP Finite Volume Cubed-Sphere Global Forecast System (FV3GFS) Version 15.3, accessed on 17 July 2020 (NOAA 2020; NOAA/NCEI 2020). FV3 refers to a six-sided (cubic) representation of Earth (Putman and Lin 2007). For research purposes, this version was run at a lower horizontal grid spacing, in accordance with guidelines for use of the Environmental Modeling Center's (EMC's) Hera supercomputer. In this setup at research resolution, analyses and ensemble runs used the six-sided cube with 192 grid cells on each side (C192), with an average grid width of ~50 km. Deterministic forecasts, on the other hand, were run at C384 (~25 km) resolution. For analyses and all runs, the operational configuration of 64 vertical levels was used. As in the operational setup, our experiments assimilated over 6 million observations per cycle every 6 h. A total of 20 ensemble members were used in each experiment, also generating 9-h forecasts for the calculation of background error covariances.

Operationally, the FV3GFS model produces daily extended ("long-term") 384-h (16 days) global forecasts initialized at 0000, 0600, 1200, and 1800 UTC, referred to as the GFS forecast. Due to limited computer resources, the forecast period is reduced to only 240 h (10 days) in our experiments, and only one forecast each day (at 0000 UTC) is produced. As in the operational configuration, 9-h forecasts were generated from the analysis to use as background for the next assimilation cycle and for generation of background error covariances. These short-term forecasts are referred to as the GDAS forecasts.

The two forecasts are visualized in Fig. 1, from NOAA (2015). Note the different start times for the GFS and GDAS forecasts. A 0000 UTC GDAS analysis/forecast starts running at 0600 UTC, after all observations from the 6-h time window centered upon 0000 UTC have been received by NCEP Central Operations. However, the 0000 UTC GFS analysis is generated at 0245 UTC to ensure that GFS products can be distributed to users in a timely manner. This "early" analysis (Wang and Lei 2014) limits the amount of observations available in generating the GFS analysis by approximately 20% compared to the GDAS "final" analysis. In practice, no observations near the end of the time window are assimilated in the early analysis. This means that the 4DEnsVar data assimilation system used by FV3GFS (Kleist and Ide 2015) is only able to use background error covariances from six times within the time window (-3, -2, -1, 0, $+1$, and $+2$ from the center of the time window). Thus, a 0600 UTC analysis calculates background error covariances from 0300 to 0900 UTC using the full dataset, while it only uses covariances from 0300 to 0800 UTC using the limited dataset.

A list describing all experiments for this study is given in Table 1. In experiment "operobs," the full observation dataset was used for short-term forecast steps (i.e., final analysis) and the limited observation dataset was used for long-term forecast steps (i.e., early analysis), as done operationally.

Experiment "fullobs" used the full observation dataset for both short- and long-term forecasts. This was done by altering the paths used in the setup of the data assimilation (analysis) step to pull from the full observation dataset regardless of whether short- or long-term forecasts were to be generated. Similarly, experiment "limobs" used the limited observation dataset for both short- and long-term forecasts, and was carried out in a similar manner to fullobs, but pulling from the limited dataset. It should be noted that fullobs was different from the other experiments in

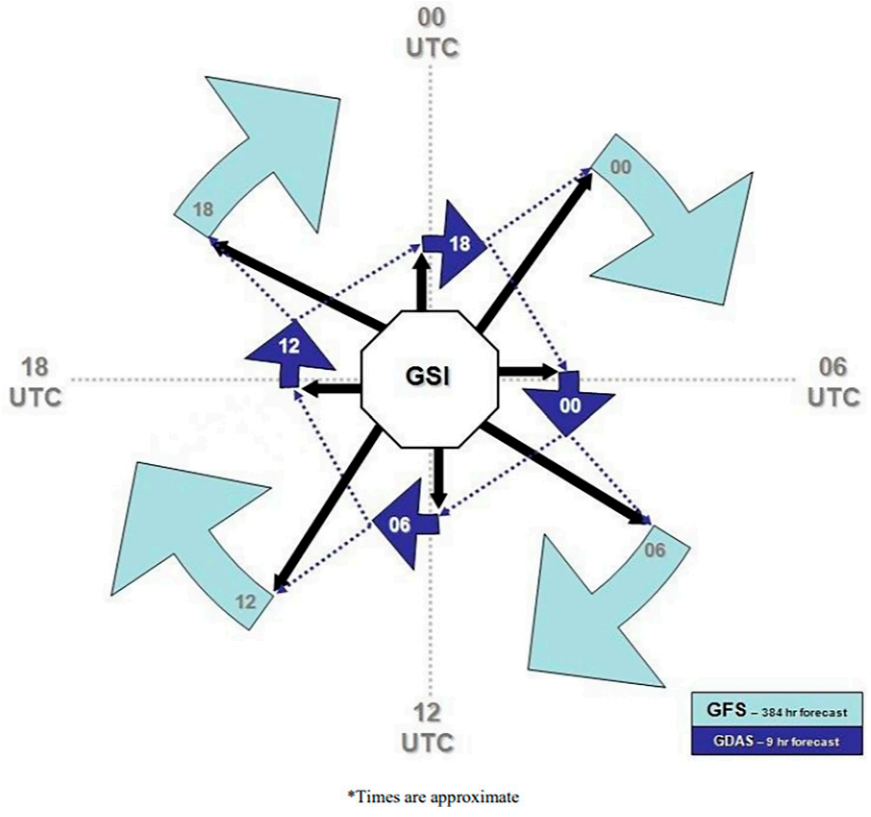

FIG. 1. GFS (light blue) and GDAS (dark blue) start times for each cycle. From NOAA (2015).

that its GDAS forecast initial conditions were the same as operobs, whereas the GDAS initial conditions differed for limobs and all other experiments. This difference in initial conditions can be amplified as the experiment moves forward through additional DA cycles.

The next three experiments, "limconv," "limgps," and "limrad," were similar to limobs, but only limiting conventional/aircraft observations, Global Navigation Satellite System Radio Occultation (GNSS-RO) profiles, and radiance soundings, respectively, in both short- and long-term forecasts. Experiments "limacft," "limam," and "limrds" were similar to limconv, but limit aircraft observations, both aircraft and marine observations, and rawinsondes, respectively, in both short- and long-term forecasts.

Forecasts for 0000 UTC were saved over the period 15 June–31 July 2020, with 1 June–14 June used as a model spinup period for each experiment. Forecast output was then compared to ECMWF (an independent data source with a full suite of assimilated observations, as opposed to treating one experiment here as the truth) Integrated Forecasting System (IFS; ECMWF 2022) 0000 and 1200 UTC operational model analyses for forecast verification using the Verification Statistical Database (VSDB; Brill and Iredell 1998). IFS version CY46R1 was used through 0000 UTC 13 July 2020, when ECWMF transitioned to version CY47R1. (Please see ECMWF 2022 for more information on the verification dataset.) These ECMWF IFS analyses are received in real time by NOAA/NCEP, and are at 1° resolution.

TABLE 1. List describing experiments in this study. Except for "operobs," all experiments assimilate data for long- and short-term forecasts.

Expt name	Expt description
operobs	Assimilates all operationally assimilated observations with the same configuration as in the operational system
fullobs	Assimilates full observation dataset
limobs	Assimilates limited observation dataset
limconv	Assimilates limited conventional/aircraft observation datasets
limgps	Assimilates limited GPS observation dataset
limrad	Assimilates limited radiance observation dataset
limacft	Assimilates limited aircraft observation dataset
limam	Assimilates limited aircraft/marine observation datasets
limrds	Assimilates limited rawinsonde observation datasets

While VSDB provides numerous statistical comparisons, only a few were used here for validation: Root-mean-square error (RMSE) of global/Northern Hemisphere extratropical (NHX; 20°–80°N)/Southern Hemisphere extratropical (SHX; 20°–80°S) 500-hPa temperature (T), NHX/SHX 500-hPa geopotential height (HGT) anomaly correlation (AC), and tropical (TRO; 20°S–20°N) 200-hPa vector wind RMS.

The 95% confidence intervals are included in many of the plots in the following two sections, and represent the range of values for which the difference between mean values of two given quantities can be assumed to be zero, assuming the difference between values follows a normal distribution. These are calculated using a paired two-sided t test at $p = 0.05$ (Hoffman et al. 2017) and follows the formula $\mathrm{CI} = F \times \mathrm{SD}/\sqrt{N-1}$, where CI denotes confidence interval, SD denotes the standard deviation of the difference between two values, and N denotes sample size. The term F represents the multiplication factor by which a 95% confidence interval is reached and is a function of sample size. For samples of size 46, the sample size for the following plots, an F value of 2.00 is used within VSDB (Brill and Iredell 1998).

3. Assimilation time windows with full and limited observation datasets

This section discusses the impacts of using the full and limited observation datasets in different scenarios. Experiment operobs was compared to experiments fullobs and limobs, as defined in Table 1.

First, thermodynamic differences were assessed using temperature. The top panel of Fig. 2 shows global 500-hPa temperature RMSE, using ECMWF analyses for verification out to 240 forecast hours for experiments operobs (black), fullobs (red), and limobs (green). RMSE for experiment operobs ranges from 0.6 K at analysis time up to 3.3 K at 240 h. When using the full observation dataset (fullobs), little difference is visible. The difference when using the limited observation dataset (limobs), however, is clear, ranging from 0.8 K at analysis time to 3.4 K at 240 h.

The bottom panel of Fig. 2 shows the RMS difference with respect to experiment operobs for experiments fullobs (red) and limobs (green). Rectangular bars denote the 95% confidence level, with corresponding colors for each experiment. It should be noted that VSDB does not take into account temporal correlation of errors, only the spatial correlation at the forecast hour in question. Despite the small numerical difference compared to operobs, the full observation dataset experiment shows statistically significant reductions in RMSE at 24–60 h, and again at 216 h. However, the degradation for the limited observation dataset experiment is far more evident, with statistically significant differences extending from the analysis time up to 216 h. Note also that the 0-h difference for limobs is 10–20 times the value of the corresponding difference for fullobs.

To investigate the impact based on the latitudinal range, Fig. 3 shows the 500-hPa temperature RMSE for NHX and SHX. NHX (Fig. 3a) exhibits similar behavior to the global temperature verification, with fullobs showing statistically significant improvement out to 60 h and limobs showing degradation

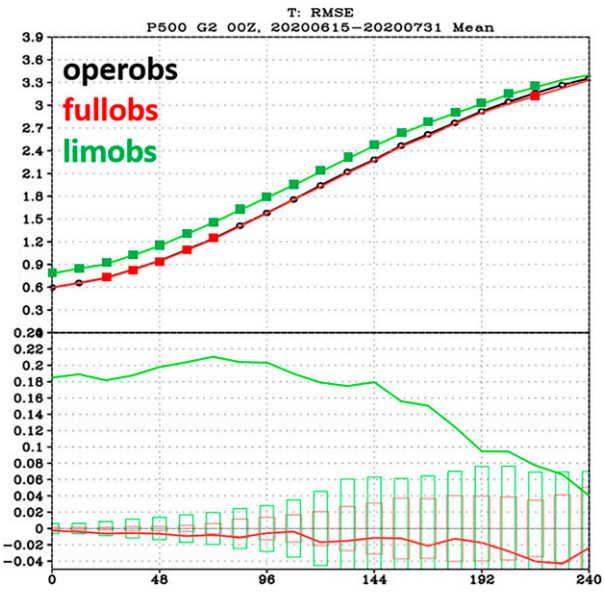

FIG. 2. (top) Global 500-hPa temperature RMSE, using ECMWF analyses for verification out to 240 forecast hours. Experiments operobs (black), fullobs (red), and limobs (green) are shown. (bottom) RMS difference with respect to experiment operobs for experiments fullobs (red) and limobs (green). Rectangular bars (in the bottom panel) denote 95% confidence level of spatial correlation of errors, and squares (in the top panel) denote those times where the difference for each experiment with respect to operobs is statistically significant.

out to 96 h. The magnitude of these differences, however, is smaller than when the verification is done globally. On the contrary, limobs shows a much larger degradation in the SHX (Fig. 3b), and this degradation is statistically significant out to 228 h. However, experiment fullobs results in a statistically significant improvement over operobs out to 60 h. Overall, Fig. 3 shows that improvement from fullobs is small and similar in both hemispheres, while a greater degradation from limobs is found in SHX. This greater impact in SHX may be due to the June–July time period chosen for this experiment, as this coincides with SHX winter, where the storm tracks and jet stream are stronger than in the summer.

For extratropical regions, midlevel geopotential height anomaly correlation is an important indicator of the NWP skill at synoptic scales. Figure 4 plots this quantity at 500 hPa for (Fig. 4a) NHX and (Fig. 4b) SHX. In both latitudinal ranges, experiment fullobs shows no statistically significant differences when compared to operobs. In contrast, experiment limobs shows statistically significant degradation in forecast skill. As in Fig. 3, this degradation is larger in the SHX, where the statistically significant degradation extends to 216 h (9 days) compared to the shorter 96 h (4 days) in the NHX.

Given the magnitude of the degradation in limobs, it was no surprise when a time series of forecast skill at a given time for each day in the experiment period (not shown) showed that this degradation was consistent among all experiment dates. Figure 5 plots the 500-hPa HGT difference between operobs and limobs for 0000 UTC 15 June 2016, the first cycle of the experimental

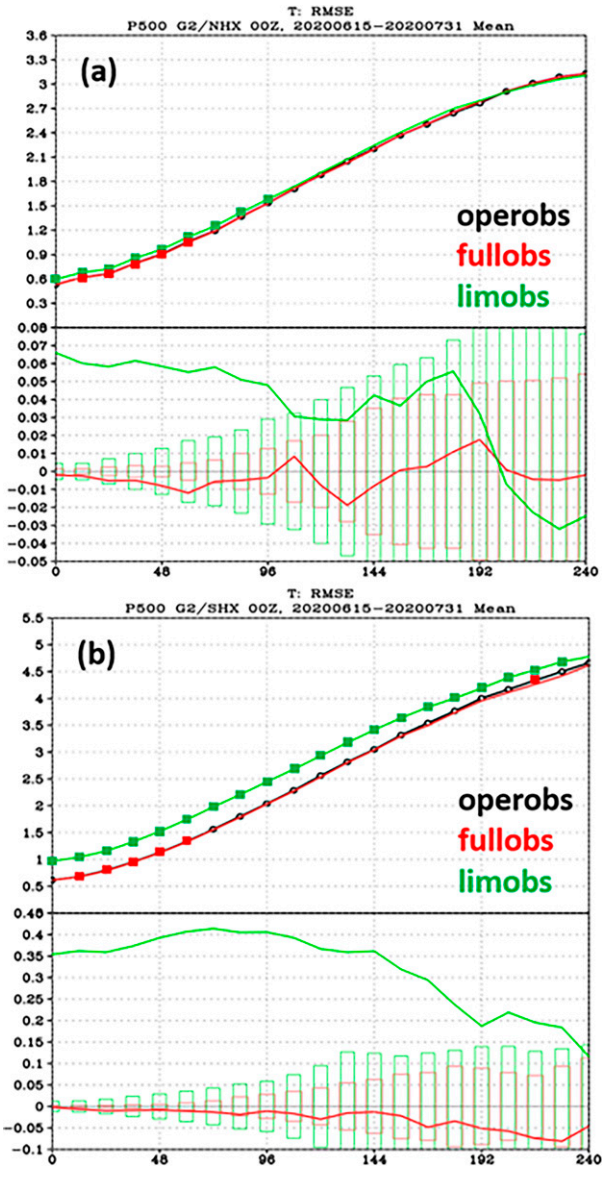

FIG. 3. As in Fig. 2, but for (a) Northern Hemisphere extratropics (NHX) and (b) Southern Hemisphere extratropics (SHX).

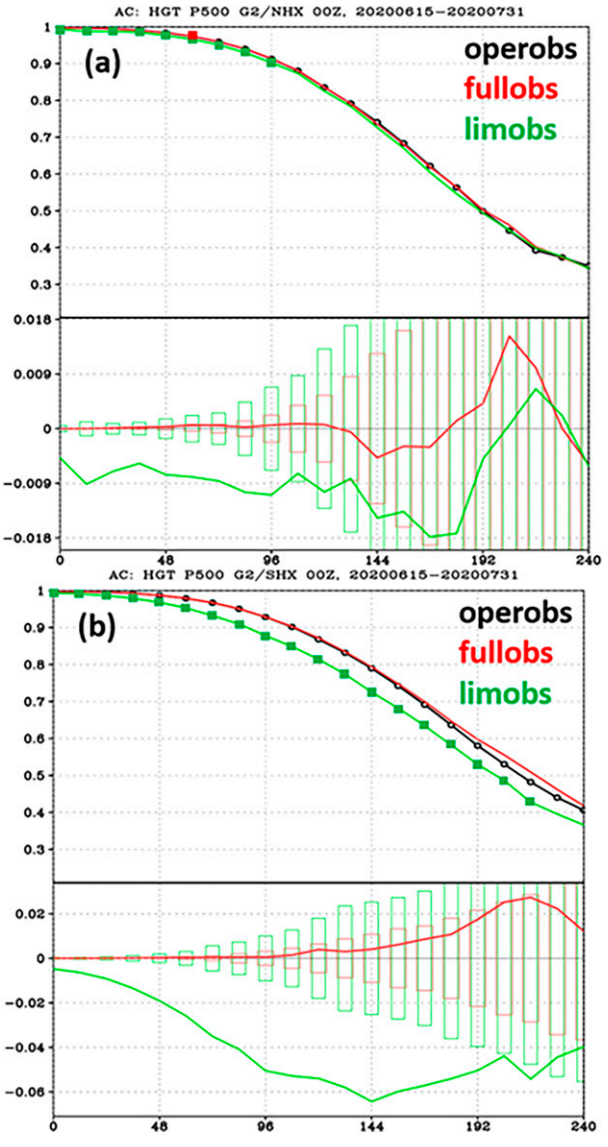

FIG. 4. As in Fig. 2, but for 500-hPa geopotential height anomaly correlation: (a) Northern Hemisphere (NH) and (b) Southern Hemisphere (SH).

period. After 24 h (top panel), the differences are on the order of 50 m over Antarctica and the surrounding ocean areas. As expected from Figs. 3 and 4, these differences are more prominent in SH than NH, though large differences are still noted in the NH arctic. After 120 h (bottom panel), these differences grow to where the forecasts for the two experiments look wildly different, with differences far exceeding 50 m covering nearly the entire area poleward of 30°.

Whereas thermodynamic variables such as temperature and geopotential height are effective in assessing extratropical synoptic-scale forecasts, tropical forecasts are also analyzed using kinematic observations such as wind. The impact of the full versus the limited dataset on the 200-hPa tropical vector wind RMSE is shown in Fig. 6. Experiment fullobs shows a reduction of the RMSE from 12 to 108 h, though, as in previous results, the magnitude of this difference is small compared to the larger impact of the limobs experiment. Here, limobs significantly increases the RMSE by roughly 20% at the analysis time, and shows statistically significant degradation at up to 72, 132–180, and 228–240 h.

Together, Figs. 2–6 indicate that using the full observation dataset for both short- and long-term forecasts results in small improvements in forecast skill, but these improvements are only statistically significant at certain (usually shorter) forecast lead times. On the other hand, the use of the limited observation dataset for both short- and long-term forecasts results in much larger and generally statistically significant errors of up to ~20% in the analysis (i.e., lead time = 0 h) and the forecasts

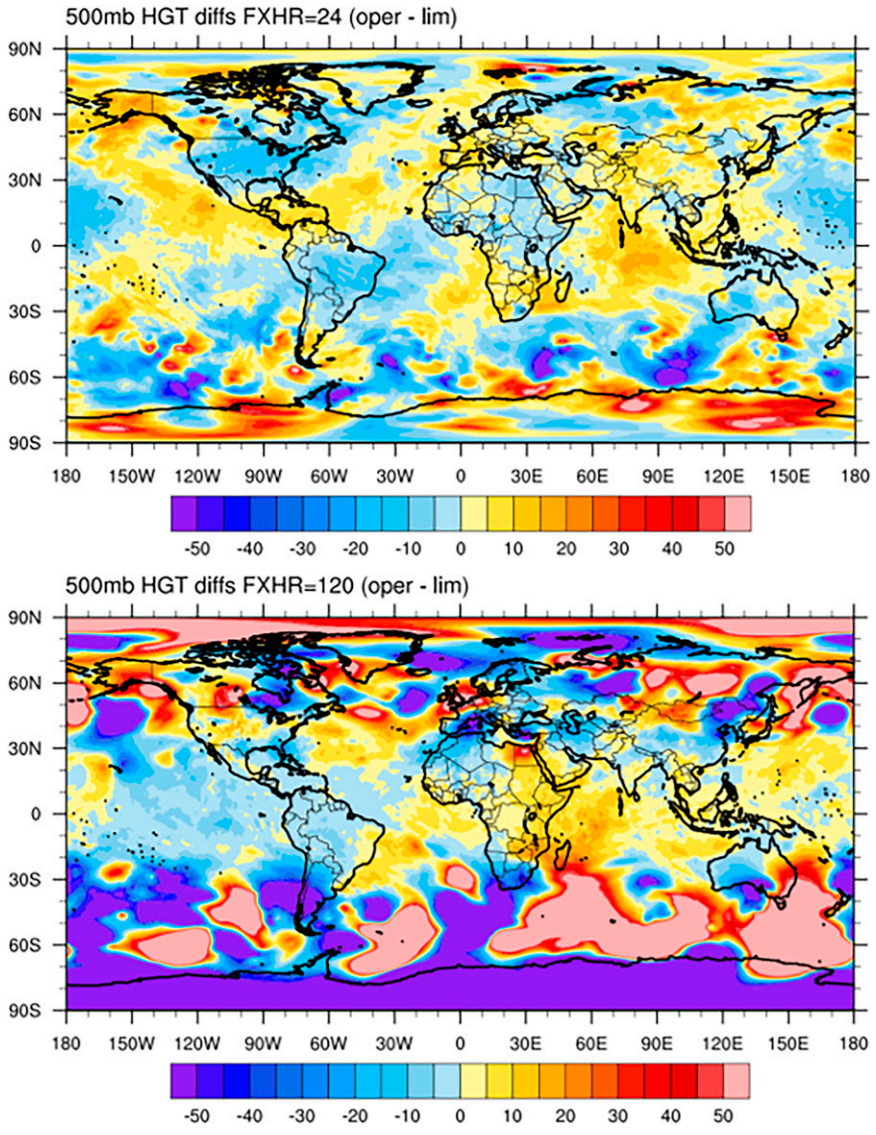

FIG. 5. Difference between 500-hPa HGT fields for operobs and limobs: (top) 24- and (bottom) 120-h forecast.

out to 10 days. To investigate the causes of this degradation in forecast skill, the next section looks more closely at the impacts of this limited-observation dataset in experiment limobs.

4. Detailed analysis of the impact of the limited-observation dataset

a. Impact of limiting different observation types

In this subsection, the impact of assimilating the limited observation dataset in experiment limobs described in section 3 will be broken down into individual observation sources. Table 2 shows how the assimilated observation count differs between the full and limited datasets, depending on observation type, for the GDAS forecast cycle at 0000 UTC 15 June 2020. Observation counts come from experiments operobs and limobs for the full and limited datasets, respectively. For in situ and atmospheric motion vector observations, captured in the surface pressure, wind, temperature, and moisture categories, less than 10% of the observations are missed in the limited dataset. A quarter of radiances, and nearly a fifth of GNSS radio occultations, however, are missing in the limited dataset assimilated observations. This suggests that satellite data latency is a key limiting factor in what observations are available in the limited versus full datasets.

The impacts from three different categories will be distinguished: conventional observations (i.e., any observation not taken from a satellite), GNSS radio occultation observations (bending angle profiles), and satellite radiances. These three experiments analyze the impact of limiting each of these observation types individually, by using the limited observation dataset for both short- and long-term forecasts. Conventional observations are limited in limconv, soundings of GNSS radio occultation in limgps, and satellite radiances in limrad.

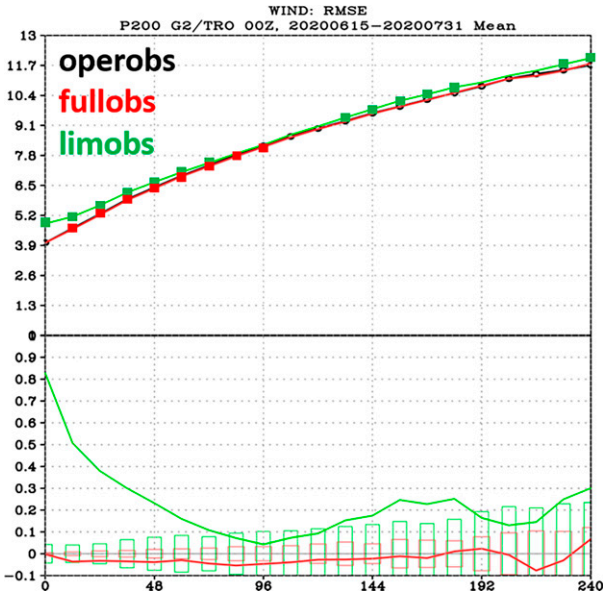

FIG. 6. As in Fig. 2, but for tropical 200-hPa vector wind RMSE.

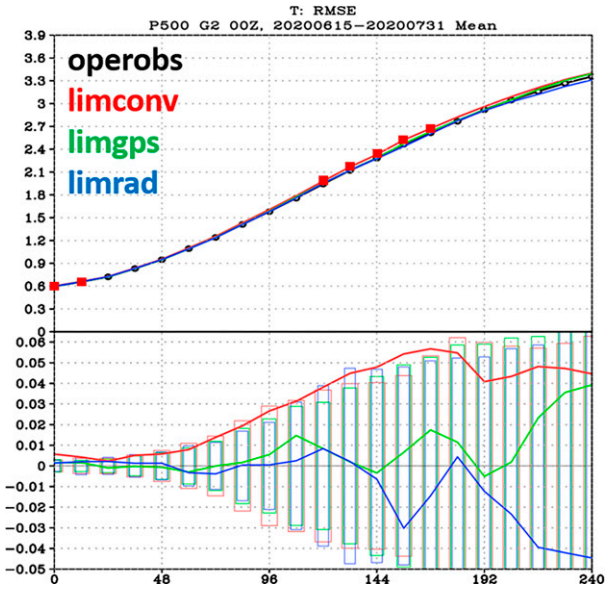

FIG. 7. As in Fig. 2, but for experiments operobs (black), limconv (red), limgps (green), and limrad (blue).

Similar to Fig. 2, Fig. 7 now shows the global 500-hPa temperature RMSE for experiments operobs (black), limconv (red), limgps (green), and limrad (blue). Experiment limconv results in a statistically significant increase in RMSE at the analysis time and at 120–168 h. On the other hand, experiments limgps and limrad show no statistically significant differences from operobs. This result seems to indicate that the in situ observations being limited in experiment limconv are the largest contributor to the degradation found in experiment limobs, despite Table 2 demonstrating that the percentage of missing observations for conventional observations being less than that for radiances and GNSS radio occultation. That being said, none of the individual limited datasets, limconv included, matches the magnitude of degradation obtained in limobs (Figs. 2 and 7). This suggests that observation redundancy, or having different observations over the same region, might partially compensate for data loss.

Figure 8a shows no statistically significant degradation in the NHX when one of the individual observation types is limited, despite the overall degradation found in Fig. 3a when all the observation types were limited. However, Fig. 8b shows that by limiting all conventional observations in the SHX, the forecast skill for the 500-hPa temperature RMSE significantly degrades at the analysis time and at 120–168 h. No significant degradation is found by limiting either GNSS-RO or satellite radiance observations. This result is surprising because the conventional observation network is sparse in the SHX when compared to the NHX. As such, forecasts in the SHX are typically reliant on satellite observations, such as GNSS-RO or radiances. This finding motivated further investigation, which is discussed in the next subsection.

TABLE 2. Number of observations assimilated for the full and limited datasets (for experiments operobs and limobs, respectively) for the GDAS forecast cycle at 0000 UTC 15 Jun 2016.

Observation type	Full dataset	Limited dataset
Surface pressure	72 122	68 041 (94%)
Wind	362 536	334 622 (92%)
Temperature	104 519	100 569 (96%)
Moisture	25 622	25 138 (98%)
Radiance	4 252 116	3 192 351 (75%)
Radio occultation	212 685	173 417 (82%)

b. *Impact of limiting different conventional observation types*

Most observations labeled "conventional" in this study can be grouped into three categories: rawinsondes, aircraft, and marine. Observation count differences between the full and limited datasets varies between these three groups. As an example, the location of temperature observations available at the first cycle of the experiment at 0000 UTC 15 June 2020 is shown in Fig. 9. As expected, most of the additional observations plotted in Fig. 9 are located in the Northern Hemisphere.

The only additional rawinsonde profile noted in the Southern Hemisphere in Fig. 9a is over central South America. Additional aircraft paths are visible in the Southern Hemisphere in Fig. 9b, specifically around Australia and New Zealand, the eastern coast of South America, and one vertical atmospheric profile over southern Africa. The 0000 UTC observations are shown in Fig. 9; greater variability in observation locations for aircraft is noted for 0600, 1200, and 1800 UTC (not shown) corresponding to differences in daytime periods for given locations. Additional surface marine observations are noted south of the equator in Fig. 9c, including some over the central Indian Ocean; off the coasts of Australia, Brazil, and South Africa; and one noted over the southern Pacific Ocean just west of 180° longitude.

Figure 9 indicates that most additional conventional observations available in the full dataset, but not in the limited

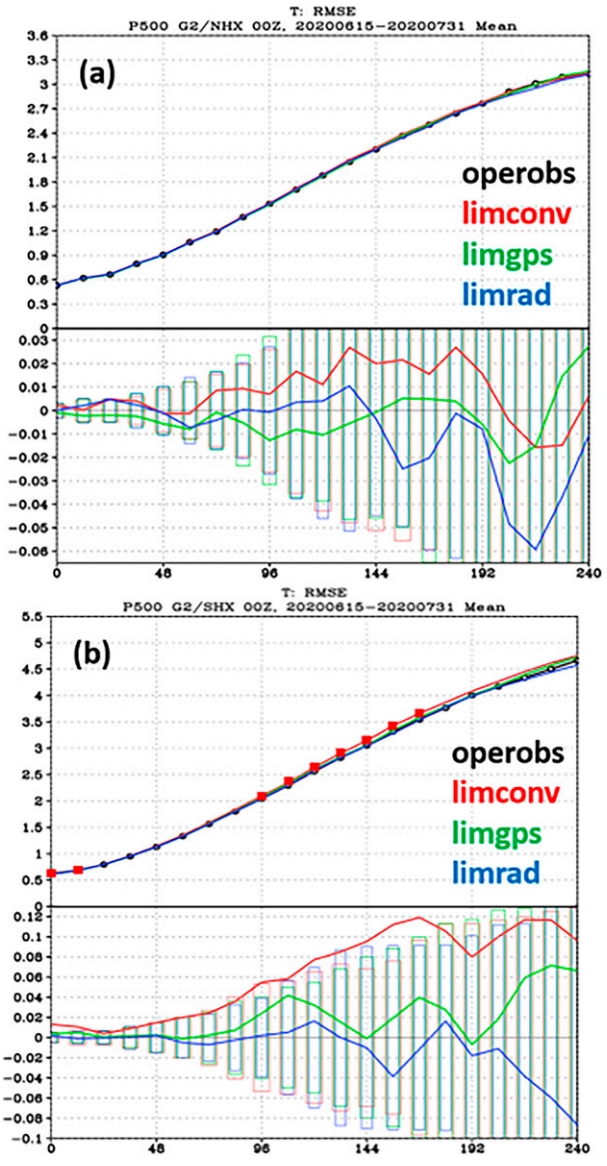

FIG. 8. As in Fig. 7, but for (a) NHX and (b) SHX.

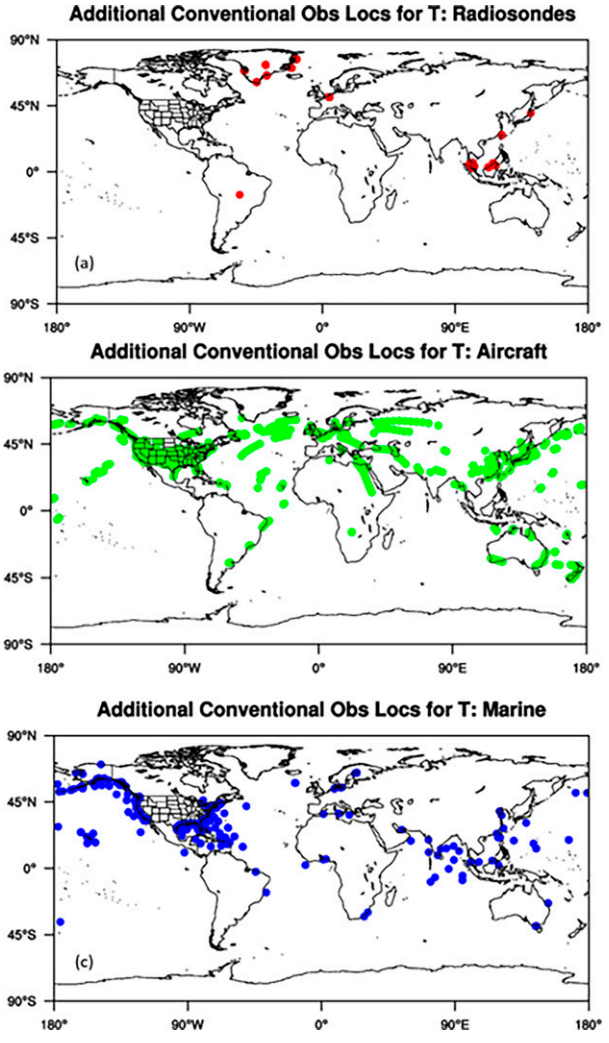

FIG. 9. Observation count differences between full- and limited-observation datasets for temperature measurements for (a) rawinsondes, (b) aircraft, and (c) marine observations at 0000 UTC 15 Jun 2020.

dataset, are aircraft observations, followed by surface marine and rawinsondes. This might also suggest that the degradation obtained when limiting conventional observations in Fig. 8b might be primarily caused by the removal of some aircraft observations.

Unlike aircraft and rawinsonde observations, where full tropospheric profiles are provided, marine observations are only available near the surface, suggesting that limiting the number of marine observations would have minimal atmospheric forecast impact compared to the impact of limiting the number of rawinsonde or aircraft observations. Rather than testing the impact of limiting marine observations alone, experiment limam (aircraft + marine) limits both aircraft and marine observations, and can be compared to experiment limacft (aircraft) directly to assess the impact of limiting the additional marine observations from the full dataset.

Figure 10 shows the forecast impacts on global 500-hPa T RMS of limiting the additional aircraft (limacft), aircraft + marine (limam), and rawinsonde (limrds) observations. When compared to limiting all the conventional observations, where a clear degradation compared to operobs was found (Fig. 7), none of the three individual datasets experiments show a comparable level of degradation. While both limacft and limam result in a slight degradation, experiment limrds actually shows a slight improvement. However, none of these differences are statistically significant at the 95% level.

Since the largest differences for the limconv experiment were in the SHX (Fig. 8), the impact of limacft, limam, and limrds are only evaluated over the SHX in Fig. 11. As in Fig. 10, no individual component of the conventional observations appears to have a significant impact, except for limacft experiment at 108 h. While limiting all the additional conventional observations at once results in degraded forecasts at 4–7 days, limiting individual

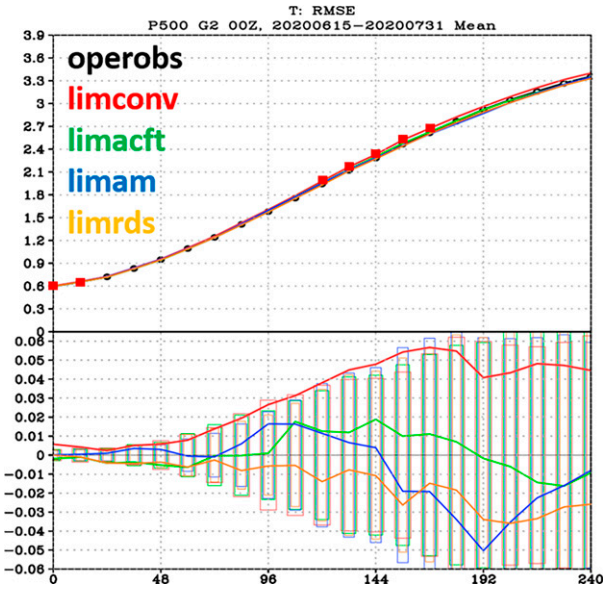

FIG. 10. As in Fig. 2, but for experiments operobs (black), limconv (red), limacft (green), limam (blue), and limrds (orange).

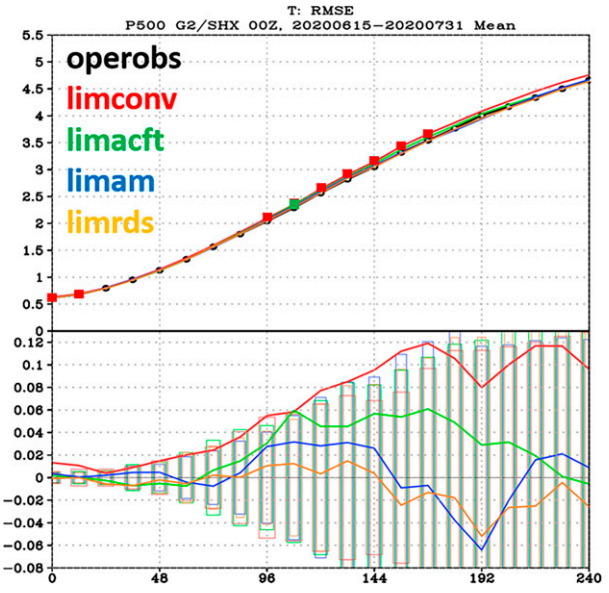

FIG. 11. As in Fig. 10, but over SHX.

datasets appears to have a minimal effect. This result again emphasizes how important redundancy in the observing system is, as similar information content might be available from different observation types and it is only when all the observations are removed at once that the system is negatively affected.

5. Conclusions

This study uses the NCEP FV3GFS model to investigate the impact of data latency in operational global weather forecasting. In this operational setting, a full dataset is used for short-term forecasts and a limited dataset with a reduced number of observations is used for long-term forecasts. The impact of data latency is investigated by using the two different datasets for the long- and short-term forecasts.

Initially, two experiments are run, the first using the full dataset for both short- and long-term forecasts and the second using the limited dataset for both short- and long-term forecasts. The experiment that uses the full dataset for both forecast lengths yields a small yet statistically significant improvement in forecast skill at a few lead times compared to the operational configuration. This small difference suggests that though some statistically significant improvements are noted when using the full dataset for long-term forecasts, the increase in forecast skill may not justify the lower data latency that would be required for all the observations to arrive in time for the early analysis.

On the other hand, using the limited dataset for both forecast lengths results in a substantial and statistically significant degradation in forecast skill at most lead times, an order of magnitude greater than that noted when using the full dataset. Focusing on this degradation, three additional experiments were run, one limiting all in situ observations, one limiting GNSS-RO observations, and one limiting satellite radiances. It is found that this degradation is largely attributed to decreased in situ observations, with negligible impact noted when GNSS-RO and radiance observations are limited. However, no individual observation type fully captures the degradation noted when all observations are limited.

Finally, three more experiments limit individual types of in situ observations (aircraft, marine, and rawinsonde). None of these three experiments show the degradation noted when limiting all in situ observations, demonstrating the importance of data redundancy in the operational observational system.

Four caveats should be kept in mind when considering these results. First, these experiments were not run at the full resolution of operational FV3GFS, and running on a coarser resolution could hide impacts from observations at scales between ~12 and 25 km. Second, this study only looked at one season, Northern Hemisphere summer, and as such could miss impacts from observations in other seasons. Third, the NCEP GFS/GDAS system was used, and while these results may be applicable to other NWP models using a hybrid 4DEnsVar system, it is unclear which results can be applied to other global NWP DA systems. Finally, the experiment period (June–July 2020) coincided with a decrease in aircraft observations due to the COVID-19 pandemic, and it is possible that with a fully operational commercial aircraft sector the impact in experiments such as limacft may show greater impact from aircraft observations. Addressing these caveats is left for further study.

Acknowledgments. This study was partially funded by the NESDIS OPPA Technology Maturation Program, and supported by NOAA/OAR/AOML and U.Miami/CIMAS Grant NA15OAR4320064. The authors wish to thank Kate Friedman for her assistance with the FV3GFS model; Sarah Ditchek and Michael Mueller for plotting assistance; Fanglin Yang for his assistance with the ECMWF IFS verification; and are grateful for early reviews of the manuscript by Stan Goldenberg and Michael Mueller. Three anonymous reviewers significantly contributed to improve the manuscript.

Data availability statement. All experiment results and configuration files are archived on the NCEP/Environmental Modeling Center High Performance Storage System, and available from the lead author upon request.

REFERENCES

Brill, K. F., and M. D. Iredell, 1998: EMC verification database. NCEP, 8 pp., http://www.emc.ncep.noaa.gov/mmb/papers/brill/VSDBformat.txt.

Cucurull, L., and R. A. Anthes, 2014: Impact of infrared, microwave and radio occultation satellite observations in operational numerical weather prediction. *Mon. Wea. Rev.*, **142**, 4164–4186, https://doi.org/10.1175/MWR-D-14-00101.1.

ECMWF, 2022: Changes in ECMWF model. Integrated Forecasting System, ECMWF, accessed 30 March 2022, https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model.

English, S., and Coauthors, 2013: Impact of satellite data. ECMWF Tech. Memo. 711, 46 pp., https://www.ecmwf.int/sites/default/files/elibrary/2013/9301-impact-satellite-data.pdf.

Garrett, K., 2013: Forecast impact assessments of SNPP ATMS. *JCSDA Quarterly*, No. 43, Joint Center for Satellite Data Assimilation, Boulder, CO, 3–4.

Han, W., and M. Jochum, 2017: Latency analysis of large volume satellite data transmissions. *2017 IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS 2017)*, Fort Worth, TX, IEEE, 384–387, https://doi.org/10.1109/IGARSS.2017.8126976.

Haupt, S. E., P. A. Jimenez, J. A. Lee, and B. Kosovic, 2017: Principles of meteorology and numerical weather prediction. *Renewable Energy Forecasting*, **2017**, 3–28, https://doi.org/10.1016/B978-0-08-100504-0.00001-9.

Hoffman, R. N., S.-A. Boukabara, V. K. Kumar, K. Garrett, S. P. F. Casey, and R. Atlas, 2017: An empirical cumulative density function approach to defining summary NWP forecast assessment metrics. *Mon. Wea. Rev.*, **145**, 1427–1435, https://doi.org/10.1175/MWR-D-16-0271.1.

Huang, A., and Coauthors, 2016: Community satellite processing package from direct broadcast: Providing real-time satellite data to every corner of the world. *2016 IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS 2016)*, Beijing, China, IEEE, 5532–5535, https://doi.org/10.1109/IGARSS.2016.7730443.

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.

Kalnay, E., 2002: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 368 pp., https://doi.org/10.1017/CBO9780511802270.

Klaes, K. D., F. Montagner, and C. Larigauderie, 2013: Metop-B, the second satellite of the EUMETSAT Polar System, in orbit. *Proc. SPIE*, **8866**, 886613, https://doi.org/10.1117/12.2022440.

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

Lin, H., 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.

NOAA, 2015: NOAA/NWS/NCEP/EMC Global Climate and Weather Modeling Branch: Running Global Model Parallel Experiments. NOAA, version 6.0, 18 February 2015.

——, 2020: FV3GFS Virtual Lab. NOAA, accessed 17 July 2020, https://vlab.noaa.gov/web/fv3gfs.

NOAA/NCEI, 2020: Global Forecast System (GFS) Atmospheric Model. NOAA/NCEI, https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast.

Petersen, R. A., 2016: On the impact and benefits of AMDAR observations in operational forecasting—Part I: A review of the impact of automated aircraft wind and temperature reports. *Bull. Amer. Meteor. Soc.*, **97**, 585–602, https://doi.org/10.1175/BAMS-D-14-00055.1.

Powers, J. G., and Coauthors, 2017: The Weather Research and Forecasting Model: Overview, system efforts, and future directions. *Bull. Amer. Meteor. Soc.*, **98**, 1717–1737, https://doi.org/10.1175/BAMS-D-15-00308.1.

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.

Wang, P., J. Li, and T. J. Schmit, 2020: The impact of low latency satellite sounder observations on local severe storm forecasts in regional NWP. *Sensors*, **20**, 650, https://doi.org/10.3390/s20030650.

Wang, X., and T. Lei, 2014: GSI-based four-dimensional ensemble-variational (4DEnsVar) data assimilation formulation and single-resolution experiments with real data for NCEP Global Forecast System. *Mon. Wea. Rev.*, **142**, 3303–3325, https://doi.org/10.1175/MWR-D-13-00303.1.

Weng, F., X. Zou, X. Wang, S. Yang, and M. D. Goldberg, 2012: Introduction to Suomi national polar-orbiting partnership advanced technology microwave sounder for numerical weather prediction and tropical cyclone applications. *J. Geophys. Res.*, **117**, D19112, https://doi.org/10.1029/2012JD018144.