ANALYSIS OF AN OBSERVING SYSTEM EXPERIMENT FOR THE JOINT POLAR SATELLITE SYSTEM

BY STEPHEN LORD, GEORGE GAYNO, AND FANGLIN YANG

An observing system experiment was conducted to measure the impact of withdrawing data from the afternoon orbit on global forecast skill.

For more than 40 years, satellite-based observations have contributed an increasing amount of information on atmospheric temperature and moisture structure, surface state, and cloud motion (as a proxy for winds). Satellite-based soundings are primarily from radiometric instruments measuring different parts of Earth's energy spectrum in the infrared (IR) and microwave (MW) regions. More recently, observations from the Global Navigation Satellite System (GNSS) provide accurate, nonbiased thermodynamic soundings in the stratosphere and much of the troposphere through a GNSS radio occultation (GNSS-RO) technique. These satellite observations are complementary to "conventional" observations from radiosondes, surface networks, aircraft, and radars—all globally distributed but confined primarily to land areas and occasional ships. Together, they comprise the Global Observing System (GOS), which is critical for operational numerical weather prediction (NWP).

While input from the GOS is critical, both the NWP forecast model and a global data assimilation system (DAS) are also critical for accurate prediction. The DAS extracts observed information on temperature, moisture, wind, and pressure, and combines it with information from the forecast model, usually a short-term (1–6 h) forecast valid at the analysis time (Kalnay 2003), to update the model initial conditions for the next forecast cycle. Importantly, the DAS and model can also be used to evaluate the impact of observations on forecast skill. Two currently used techniques are observing system experiments (OSEs) and the forecast sensitivity to observations (FSO) technique. In an OSE, a DAS and model forecast run is conducted using a baseline set of observations;

AFFILIATIONS: LORD—NOAA/National Weather Service, Silver Spring, and ESSIC, University of Maryland, College Park, College Park, Maryland; GAYNO AND YANG—I.M. Systems Group, and NOAA/NWS/NCEP/Environmental Modeling Center, College Park, Maryland
CORRESPONDING AUTHOR: Stephen J. Lord, NOAA/NWS/Office of Science and Technology Integration, 1325 East–West Highway, Silver Spring, MD 20910
E-mail: sjlord12345@gmail.com

The abstract for this article can be found in this issue, following the table of contents.
DOI:10.1175/BAMS-D-14-00207.1

In final form 8 October 2015

Publisher's Note: This article was revised on 17 October 2016 to insert footnote 3.

Unauthenticated | Downloaded 01/20/21 02:01 AM UTC

further runs are done but with denying or adding observations to measure forecast impact through a standard set of verification scores (e.g., Kelly et al. 2004; Zapotocny et al. 2008; and many others). This method can also be used on case studies to isolate the observational impact on specific, important meteorological events [e.g., McNally et al. (2014) for Hurricane Sandy (2012)]. More recently, the FSO technique was developed. Still using a DAS, a forecast model, and observations as tools, the FSO seeks to provide information on the reduction of (typically 24 hours) forecast error, made possible by each of the input observations. FSO can be executed using adjoints of the DAS and forecast model (e.g., Langland and Baker 2004) or through ensemble-based data assimilation techniques (e.g., Liu and Kalnay 2008; Ota et al. 2013). Compared to OSEs, FSO experiments require considerably reduced computational resources but do require additional system development and maintenance for the adjoints and/or the ensemble-based DAS.

Despite the relatively straightforward nature of OSEs and FSOs, it is important to make two further comments on factors that may influence OSE and/or FSO results from different NWP systems. First, current operational global 0–5-day predictions are very accurate and, consequently, small changes to the GOS, such as the loss of one satellite instrument, may not produce a large change in forecast skill as shown by traditional mean score differences. While results have clearly shown that forecast skill in the Southern Hemisphere (SH) is more dependent on satellite data (e.g., Kelly et al. 2004), it can be more difficult to see the impact in the Northern Hemisphere (NH), since conventional observations are far more plentiful and the impact of satellite data are correspondingly less. Second, observing system impacts depend on the DAS and the forecast model used in OSE. While most of GOS is commonly ingested by international NWP systems, data assimilation and model techniques differ, and these differences can be important in determining details of the OSE impacts.

The current GOS is changing, particularly for the satellite contribution. The Joint Polar Satellite System (JPSS) is introducing a pair of advanced sounders—the Advanced Technology Microwave Sounder (ATMS) and the Cross-Track Infrared Sounder (CrIS)—that will replace the legacy operational National Oceanic and Atmospheric Administration (NOAA) Polar-Orbiting Operational Environmental Satellite (POES) instruments in the afternoon (PM) orbit. The *Suomi National Polar-Orbiting Partnership* (*SNPP*) launched the first copies of these instruments on a National Aeronautics and Space Administration (NASA) research satellite in October 2011. Since the first operational JPSS satellite may not be launched until 2017, there is concern whether the POES and *SNPP* instruments will cease to function before the JPSS instruments are online, which can be up to one year after launch. Therefore, the JPSS program requested the National Centers for Environmental Prediction (NCEP) to provide an OSE to determine the forecast impact of losing the data from both the POES and JPSS instruments in the PM orbit.

This paper presents results from an OSE designed to demonstrate the impact of radiometric sounder data from the PM orbit in the JPSS era. Such OSE impacts for a possible JPSS data gap have previously been reported in the literature (e.g., McNally 2012; Garrett 2013; Cucurull and Anthes 2015), but each has used a different snapshot of GOS and a different experimental focus. Garrett (2013) focused on replacement of the current MW sounder constellation by ATMS, which is only part of the JPSS instrument impact. More extensive work by McNally (2012) focused on a two-orbit configuration, and Cucurull and Anthes (2015) added possible GNSS-RO impacts to the JPSS sounder issue. Each of these studies found minor impacts in standard verification scores to the loss of PM orbit data. Here, we present a similar OSE design to McNally (2012), but it is executed with the NCEP Global Data Assimilation System (GDAS) over a much longer period, thereby allowing a more robust analysis of statistical significance and possible seasonal changes in impact. Anticipating that the results will be qualitatively similar to those from both the McNally (2012) and Cucurull and Anthes (2015) work, we give an overview of all scores to place them in context of previous results. More importantly, however, we present a more detailed statistical analysis of the representative 500-hPa geopotential height anomaly correlation (Z500AC) skill score results that relates the observing system impacts to historical accuracy data. This analysis goes beyond the traditional comparison of mean skill scores by looking at the distribution of skill scores from each OSE run and how they change as a result of retaining or removing instruments in the PM orbit. We find that such an analysis can be used quantitatively to assess changes in risk associated with any OSE results or, more generally, any comparison between NWP systems.

The "Current and future polar-orbiting satellite system" section describes current and future polar-orbiting observing systems, and the "Measuring the impact of observing systems on numerical weather prediction" section describes the OSE and FSO techniques used to measure observing system impacts. The

Unauthenticated | Downloaded 01/20/21 02:01 AM UTC

OSE system used in this study, including the data assimilation and forecast system, is described in the "Description of the NCEP OSE system" section. The "OSE setup" section describes the control and experiment, and the next several sections present the evaluation procedures, overview of results, and a more detailed analysis of the Z500AC skill score distributions, respectively. The last section contains the summary and a short discussion. Appendix A summarizes the global observations available for this OSE, and appendix B provides some further background information and context for interpreting the forecast impacts presented in this paper.

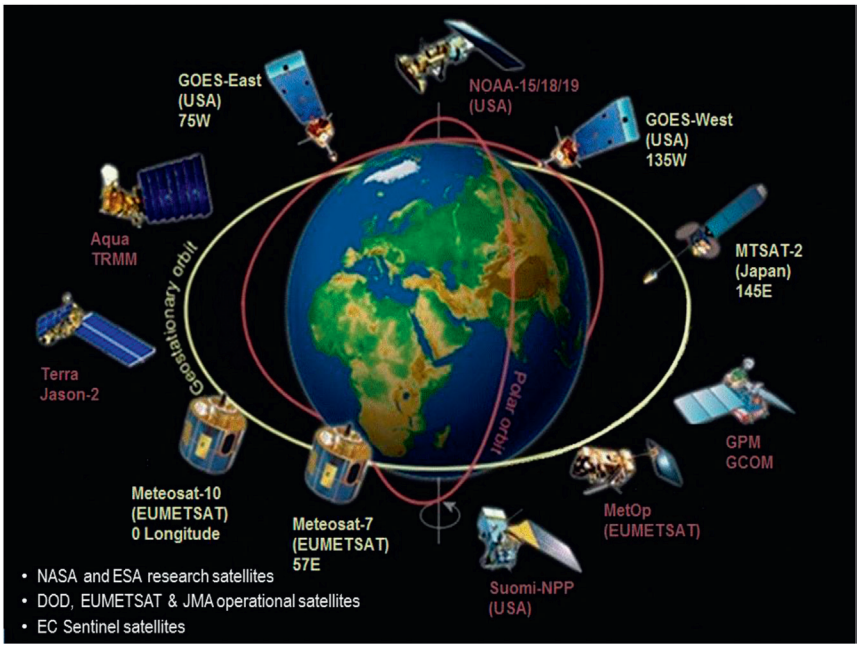


Fig. 1. The satellite-based observing system. Polar-orbiting satellites (red) are distinct from those in geostationary orbits (white) and are defined by their equatorial crossing time. Being in a sun-synchronous orbit, the satellite achieves global coverage as the Earth rotates beneath the orbit, which takes approximately two hours to complete. NOAA operational satellites generally occupy the PM orbit, while MetOp (EUMETSAT) satellites occupy the mid-AM orbit. Research satellites from NASA, the European Space Agency (ESA), and the European Commission (EC) may occupy either orbit. Operational U.S. Department of Defense (DOD) satellites have occupied a third (early AM) orbit (not shown). [Courtesy of NOAA/National Environmental Satellite, Data, and Information Service (NESDIS).]

THE CURRENT AND FUTURE POLAR-ORBITING SATELLITE SYSTEM. The satellite-based observing system (Fig. 1) is a critical component of GOS that supports routine operational weather prediction. For reference, a list of the primary GOS observations used by NCEP in 2012–13 is in appendix A. Because conventional observing systems, such as radiosondes, are mostly confined to continental areas and a few isolated islands and mobile platforms aboard ships, oceanic thermodynamic observations throughout the vertical atmospheric column are obtained almost exclusively from radiometric sounders on polar-orbiting satellites. Geostationary satellites, stationed over the equator at various longitudes, provide valuable imagery over the global domain and derived wind estimates from that imagery, but currently they carry only low-vertical-resolution infrared sounders that do not provide much information useful for NWP. Polar-orbiting satellites also host the GNSS-RO instruments that provide highly accurate and complementary atmospheric soundings.

Global coverage by atmospheric sounders is achieved through sun-synchronous (polar) orbits with different nominal equatorial crossing times (Fig. 1). Three polar orbits—PM, midmorning (mid-AM), and early morning (early AM)—provide complete global coverage every 6 h, while two orbits cover approximately 85% of Earth's surface (Fig. 2). Current international agreements have NOAA providing coverage from the PM orbit and the Meteorological Operational (MetOp) system, sponsored by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), occupying the mid-AM orbit.

In the United States, polar-orbiting instruments have been deployed for more than 40 years, but the current generation of operational NOAA POES instruments has reached the end of its life cycle. The last of the NOAA operational polar-orbiting satellites (*NOAA-19*) was launched in 2009 and hosts a pair of MW sounders [the Advanced Microwave Sounding Unit, instrument A (AMSU-A), and the Microwave Humidity Sounder (MHS)] and an IR sounder [the High Resolution Infrared Radiation Sounder (HIRS)]. Some POES instruments launched on previous NOAA satellites are still operating (Table A1) and their data are also used operationally. The POES instruments are being replaced by ATMS and CrIS, which have

Unauthenticated | Downloaded 01/20/21 02:01 AM UTC

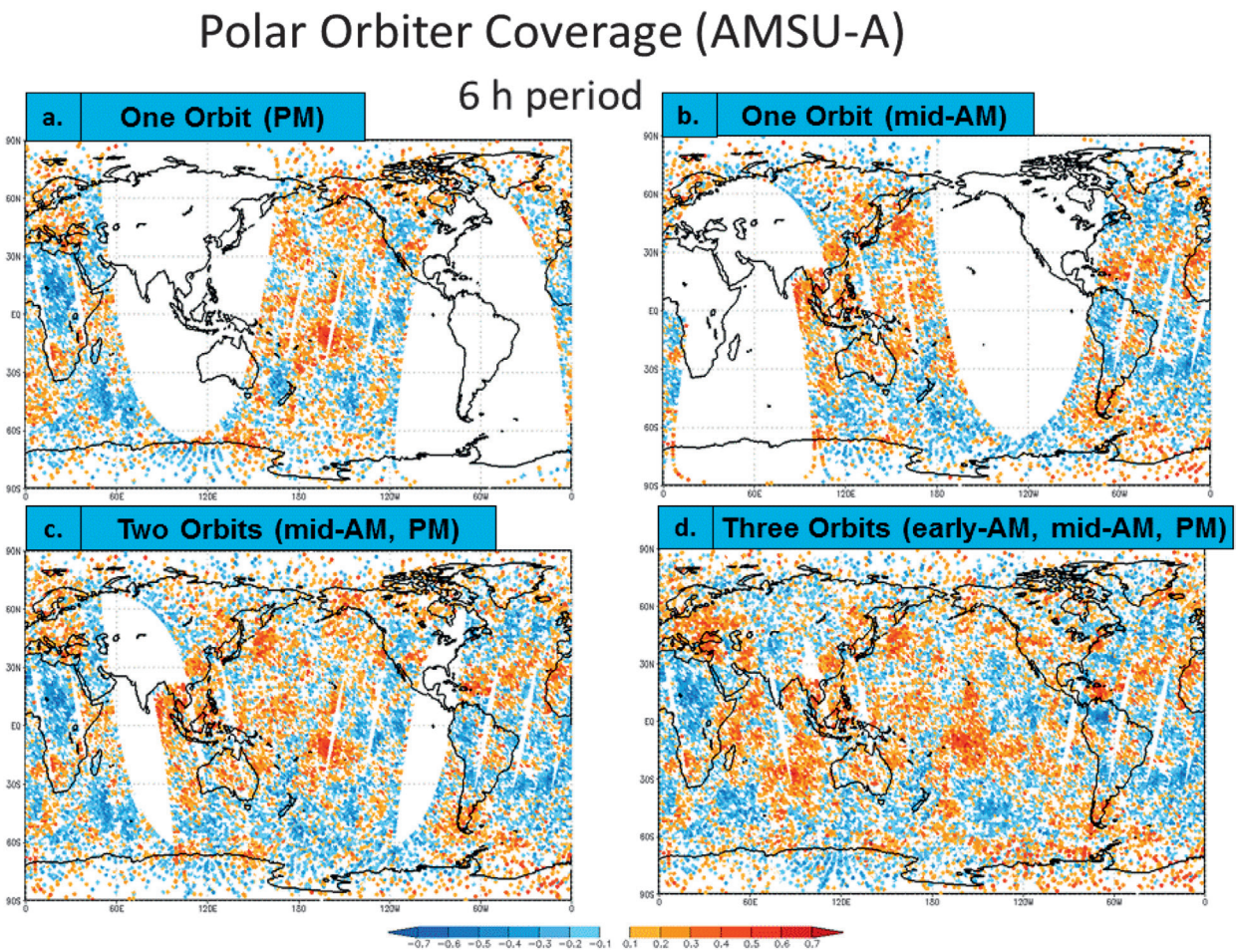


Fig. 2. Polar-orbiter observation coverage, illustrated using AMSU-A as an example, for a 6-h window centered on 1200 UTC 1 Jun 2011. Coverage shows the (a) PM, (b) mid-AM, (c) mid-AM and PM, and (d) mid-AM, PM, and *NOAA-15* orbital combinations. Because of its long-term drift from the PM orbit, *NOAA-15* is a proxy for the early-AM coverage. Colors depict differences between the observed channel-9 bias-corrected radiance value and the collocated GFS background (6-h forecast) value calculated with the operational CRTM.

improved instrument characteristics, including higher horizontal and vertical resolution and lower noise (e.g., Goldberg et al. 2013; Kim et al. 2014; Han et al. 2013; Zavyalov et al. 2013). In addition, the NASA research satellite *Aqua* provides observations from the hyperspectral Atmospheric Infrared Sounder (AIRS) and a partially operating AMSU-A in the PM orbit. In Europe, EUMETSAT launched its first polar-orbiting satellite *MetOp-A* in October 2006 with an AMSU-A, an MHS, and the hyperspectral Infrared Atmospheric Sounding Interferometer (IASI). Also on *MetOp-A* are the Advanced Scatterometer (ASCAT), the GNSS Receiver for Atmospheric Sounding (GRAS), and the Global Ozone Monitoring Experiment (GOME) for measuring surface winds, GNSS-RO, and ozone, respectively. *MetOp-B* was launched in September 2012 with the same sounding instruments. Currently, the Defense Meteorological Satellite Program (DMSP) satellites occupy the early-AM orbit and host the Special Sensor Microwave Imager/Sounder (SSMIS), which has some sounding channels similar to those on AMSU-A. However, the DMSP platforms are also nearing the end of their life cycles and the future of instrument(s) in the early-AM orbit is uncertain.

The future operational polar-orbiting satellite sounding system therefore will be primarily composed of JPSS and MetOp satellites in the PM and mid-AM orbits, respectively. Each satellite will have an MW and hyperspectral IR sounder, thereby forming a two-orbit, four-sounder (2O–4S) configuration. In the PM orbit, ATMS and CrIS have strong credentials, but nevertheless they are of approximately the same sounding capability as the current AMSU-A/MHS and AIRS. It is important to take note of these similarities in designing impact experiments for the future polar-orbiting satellite system.

Unauthenticated | Downloaded 01/20/21 02:01 AM UTC

MEASURING THE IMPACT OF OBSERVING SYSTEMS ON NUMERICAL WEATHER PREDICTION. There is considerable interest in the meteorological community and elsewhere about the impact of various GOS components on daily operational weather prediction skill, particularly in this period of rapid change in the satellite observing system. The World Meteorological Organization has sponsored international workshops every 4 years (e.g., Böttger et al. 2004; Pailleux et al. 2008; Andersson and Sato 2012) to review progress in observing system impacts for NWP. Testing the impact can be done in several ways by OSEs and the FSO technique, which differ in their approach but nevertheless use the power of modern data assimilation systems as their core software. In a typical OSE (e.g., Kelly et al. 2004; Zapotocny et al. 2008; McNally 2012; Cucurull and Anthes 2015), a control data assimilation and forecast experiment is conducted with all observations, and a second experiment is run without the observations of interest or with new observations added. Differences in performance skill are typically measured with standard scores such as the Z500AC (WMO 2010), root-mean-square (RMS) differences against both gridded analyses (GRD; RMS-GRD) and observations (OBS; RMS-OBS), and equitable threat scores (ETS) for precipitation (Wilks 1995).

FSO calculations measure the percentage contribution to the reduction of forecast error from each observation source (Langland and Baker 2004; Cardinali 2009; Gelaro and Zhu 2009; Ota et al. 2013; Lorenc and Marriott 2014). While OSEs and FSO studies have very different theoretical and algorithmic bases, they give consistent results on the relative importance of the most impactful observing systems (Gelaro and Zhu 2009) if the same data assimilation system is used. Nevertheless, OSEs and FSO studies from different data assimilation systems are not entirely consistent. For example, Joo et al. (2012) and Ota et al. (2013) report different rank orders of sensitivity using an FSO technique (Table 1). Some reasons for these discrepancies are discussed below.

While the design and execution of OSEs and FSOs are relatively straightforward, the results and their interpretation can be subject to many factors, including the representativeness of the analysis and forecast sample, and the overall quality of the analysis–forecast system (A-FS), including any forecast model bias. As also discussed by Cucurull and Anthes (2015), these factors cause forecast skill to depend on the season and meteorological conditions, so that details of observation impacts can also depend on the time period chosen for the experiment. To mitigate this dependency and to expose the NWP system to as many different weather regimes and observations as practical, experiments of at least 4–6 weeks for both winter and summer seasons are often conducted. Some OSEs are configured as case studies and can thereby directly illustrate forecast impacts on societally important meteorological events (e.g., McNally et al. 2014). However, case studies do not provide a statistically significant sample for overall impact and can often show no impact or even negative impact (the forecast is better without the observations in question; J. Yoe 2012, personal communication).

Table 1. Contribution to forecast model error reduction (from high to low) of selected major global observing systems listed by the Joo et al. (2012) and Ota et al. (2013) studies. These studies also considered satellite sounders but differences in their definitions precluded a side-by-side comparison.

Order	Joo et al.	Ota et al.
1	Radiosonde	Aircraft
2	Aircraft	Radiosonde
3	Land surface	AMVs
4	AMVs	GNSS-RO
5	Marine surface	Land surface
6	GNSS-RO	Marine surface

Accuracy of the A-FS and the details of observation processing are other important factors in determining the impact of observations. Some of these details include error assigned to the various observation types, quality control techniques and thresholds, and data thinning. Since the purpose of assimilating observations is to correct initial condition errors, a less accurate A-FS or larger assigned observation error may require more or different observations to achieve those corrections and, therefore, the overall observation impact can differ. Finally, the GOS information derives from different sources, some of which may add complementary information (as they have different observing techniques, and horizontal and vertical resolutions), but some may add resilience to the GOS by providing increased sampling over the globe. In the latter case, loss of one instrument of several similar ones can often be compensated by the DAS extracting additional information from the remaining instruments (Andersson and Sato 2012). For example, in 2012–13, five AMSU-A instruments provided operational data from (effectively) three different orbits (Table A1). In this case, withdrawal of one or more AMSU-A instruments may not impact the mean forecast skill in an OSE. In summary, quantitatively comparing OSEs should be done with caution, with an emphasis on a thorough understanding of the results.

Unauthenticated | Downloaded 01/20/21 02:01 AM UTC

Table 2. List of polar-orbiting observations selected for the CNTL and NOPM experiments.

Observing system	Orbit	CNTL	NOPM
AMVs (*Aqua*)	PM	Yes	No
AMVs (*Terra*)	AM	No	No
MetOp-A ASCAT	Mid-AM	Yes	Yes
MetOp-A IASI	Mid-AM	Yes	Yes
MetOp-A AMSU-A	Mid-AM	Yes	Yes
MetOp-A MHS	Mid-AM	Yes	Yes
Aqua AIRS	PM	Yes	No
NOAA-19 AMSU-A	PM	Yes	No
NOAA-19 MHS	PM	Yes	No

DESCRIPTION OF THE NCEP OSE SYSTEM.

Model and data assimilation. The NCEP operational global modeling system, as implemented on 22 May 2012, was used to execute the OSEs; its main components are the Global Forecast System (GFS), version 9.0.0, and the Gridpoint Statistical Interpolation analysis system (GSI), version 3.3. The GFS 9.0.0 is a global atmospheric spectral prediction model at 27-km (T574) resolution and 64 vertical levels (see www.emc.ncep.noaa.gov/GFS/doc.php for details). The GSI is a 3D hybrid (ensemble–variational) analysis system that provides the initial condition for the GFS from a blend of a first guess (a previous 9-h forecast) and both conventional and satellite observations within a 6-h data window, ±3 h from the analysis time (Parrish and Derber 1992; Derber and Wu 1998; Kleist et al. 2009). The background error is estimated by a GSI ensemble composed of 80 members executing at 55-km (T254) resolution (Kleist and Ide 2015; Wang et al. 2013). An ensemble Kalman filter (EnKF) generates flow-dependent, ensemble-based background error covariance estimates and a hybrid algorithm, using both static and ensemble-based background error estimates, is used to determine the analysis.

Satellite observations are assimilated as clear-sky radiances (Derber and Wu 1998; McNally et al. 2000), using the Community Radiative Transfer Model (CRTM) from the Joint Center for Satellite Data Assimilation (Chen et al. 2008, 2010). Quality control rejects cloud-contaminated observations detected in the infrared sensor data (Eyre and Menzel 1989). For thin clouds in the microwave, the retrieved cloud liquid water (Grody et al. 2001) is used as a bias correction predictor to remove the cloud radiative effect. GNSS-RO observations were assimilated as in Cucurull and Derber (2008) and later upgraded (Cucurull 2010; Cucurull et al. 2013).

The OSE was run using the same analysis–forecast ("cycled") configuration as NCEP's operations; a brief summary of that procedure follows. Four times per

Table 3. Standard verification scores for NH, SH, TR, CONUS, NA, and various hurricane basins. All verification is against ECMWF analyses, except for precipitation, hurricane track, and RMS vs observations.

Abbreviation	Score
NH-Z500-AC	500-hPa geopotential height anomaly correlation, NH and SH, for 24–240-h forecast period
SH-Z500-AC	
NH-MSLP-AC	Mean sea level pressure anomaly correlation, NH and SH, for 24–240-h forecast period
SH-MSLP-AC	
NH-T-RMS-GRD	RMS error for temperature, NH and SH, for 24–240-h forecast period vs gridded analysis every 50 hPa from 850 to 100 hPa
SH-T-RMS-GRD	
NH-W-RMS-GRD	RMS error for wind, NH, SH, and TR, for 24–240-h forecast period vs gridded analysis every 50 hPa from 850 to 100 hPa
SH-W-RMS-GRD	
TR-W-RMS-GRD	
CONUS PRC 24-48h	ETS for precipitation over CONUS for 24–48- and 60–84-h forecast periods
CONUS PRC 60-84h	
Hur-TRK-ATL	Hurricane track errors for Atlantic basin
Hur-TRK-EPAC	Hurricane track errors for east Pacific basin
NH-W-RMS-OBS 24-48h	RMS error of wind vs radiosonde observations for TR, NH, SH, and NA for 24–48-h forecast period and for every 25 hPa from 1000 to 100 hPa
SH-W-RMS-OBS 24-48h	
TR-W-RMS-OBS 24-48h	
NA-W-RMS-OBS 24-48h	

Unauthenticated | Downloaded 01/20/21 02:01 AM UTC

day (0000, 0600, 1200, and 1800 UTC), at approximately 3 h after cycle time, the GSI creates initial conditions for the GFS forecast model, which is run to 16 days. This is known as the "GFS" cycle. Then, at approximately 6 h after cycle time, the GDAS cycle begins with the GSI, creating another analysis using additional, late arriving data unavailable to the GFS cycle. The GDAS analysis is the initial condition for a 9-h forecast that serves as the first guess for the next GFS and GDAS cycles.

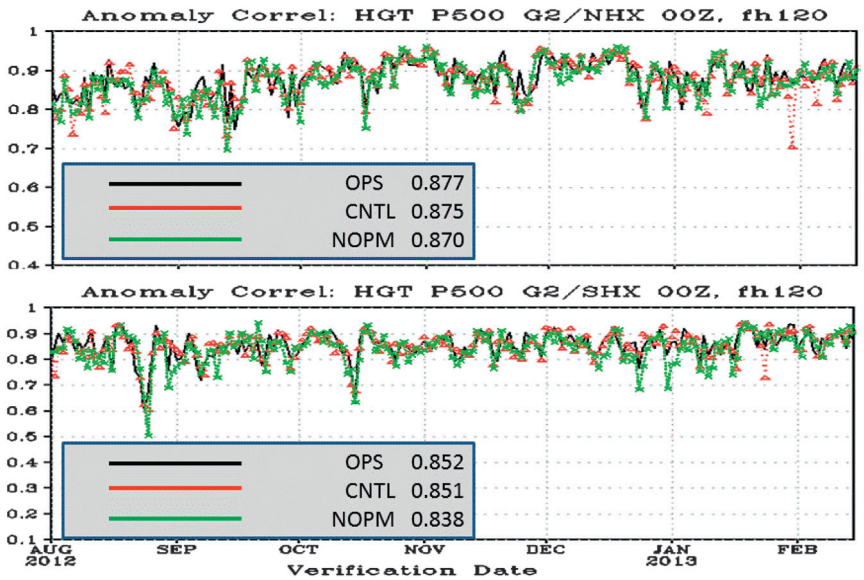


Fig. 3. Time series of 5-day Z500AC (top) NH and (bottom) SH scores for CNTL (red), NOPM (green), and OPS (black) forecasts at 0000 UTC for the period 1 Aug 2012–15 Feb 2013. Mean scores are also shown.

Observations selection. The observations selection for the OSE follows closely the choices of McNally (2012), bearing in mind that that study covered 3 months (December–February) of winter 2009/10. Conventional observations of all types (Table A1) are assimilated in all experiments, including globally distributed radiosondes, aircraft, and both overland and marine surface observations. Satellite data from geostationary and GNSS-RO sources are also assimilated in all experiments. Polar-orbiting instruments are selected as follows (see Table A1) and as summarized in Table 2. To simulate the 2O–4S configuration in the future JPSS era, MetOp (IASI, AMSU-A, and MHS) instruments are used from the mid-AM orbit and in the PM orbit *Aqua* (AIRS) and *NOAA-19* (AMSU-A and MHS) are selected as the PM sounding instruments for the OSE control (CNTL) run. Since the experiment was designed before the *SNPP* CrIS was assimilated operationally by NCEP, we used *Aqua* AIRS, in the PM orbit as proxy for the CrIS.[1] Note that AMSU-A and MHS combined have approximately the same spectral coverage as ATMS. Note also that MHS is not present on *Aqua* and that *Aqua* AMSU-A does not have a complete set of channels operating, thereby making *NOAA-19* instruments the preferred MW choice for the PM orbit. While we recognize that ATMS and CrIS are, in fact, superior instruments, we do not expect an OSE using these *NOAA-19* and *Aqua* proxies to yield substantially different results. In the no PM orbit (NOPM) OSE run, which simulates the absence of data in the PM orbit, all CNTL observations from *NOAA-19* and *Aqua* [AIRS and atmospheric motion vectors (AMVs)] are omitted from the data assimilation under the assumption that neither sounders nor an imager (for AMVs) will be in orbit for the NOPM scenario. AMVs from *Terra* were not available over the OSE time period. The MetOp scatterometer, ASCAT, was used in all runs.

For future reference, note that NCEP operations used the following additional satellite data (see Table A1): *Aqua* (AMSU-A), *NOAA-18* (AMSU-A, MHS), *NOAA-15* (AMSU-A), and HIRS on both *NOAA-17* and *NOAA-19*. *NOAA-18* observations are in the PM orbit (same as *Aqua* and *NOAA-19*), so they

Table 4. Statistical breakdown (%) of highest and lowest Z500AC scores among the CNTL, NOPM, and OPS runs for each verification date accumulated hemispherically. Each hemisphere had 293 cases.

	OPS	CNTL	NOPM
NH			
Highest	40.2	34.5	25.3
Lowest	32.4	30.4	37.2
SH			
Highest	39.2	34.1	26.6
Lowest	26.3	27.0	46.7

[1] For the MW instrument, we used *NOAA-19* AMSU-A and MHS instead of ATMS because it was originally planned to run an additional experiment to substitute ATMS for the *NOAA-19* MW instruments. However, this experiment was never executed due to the unavailability of computing and personnel resources.

Unauthenticated | Downloaded 01/20/21 02:01 AM UTC

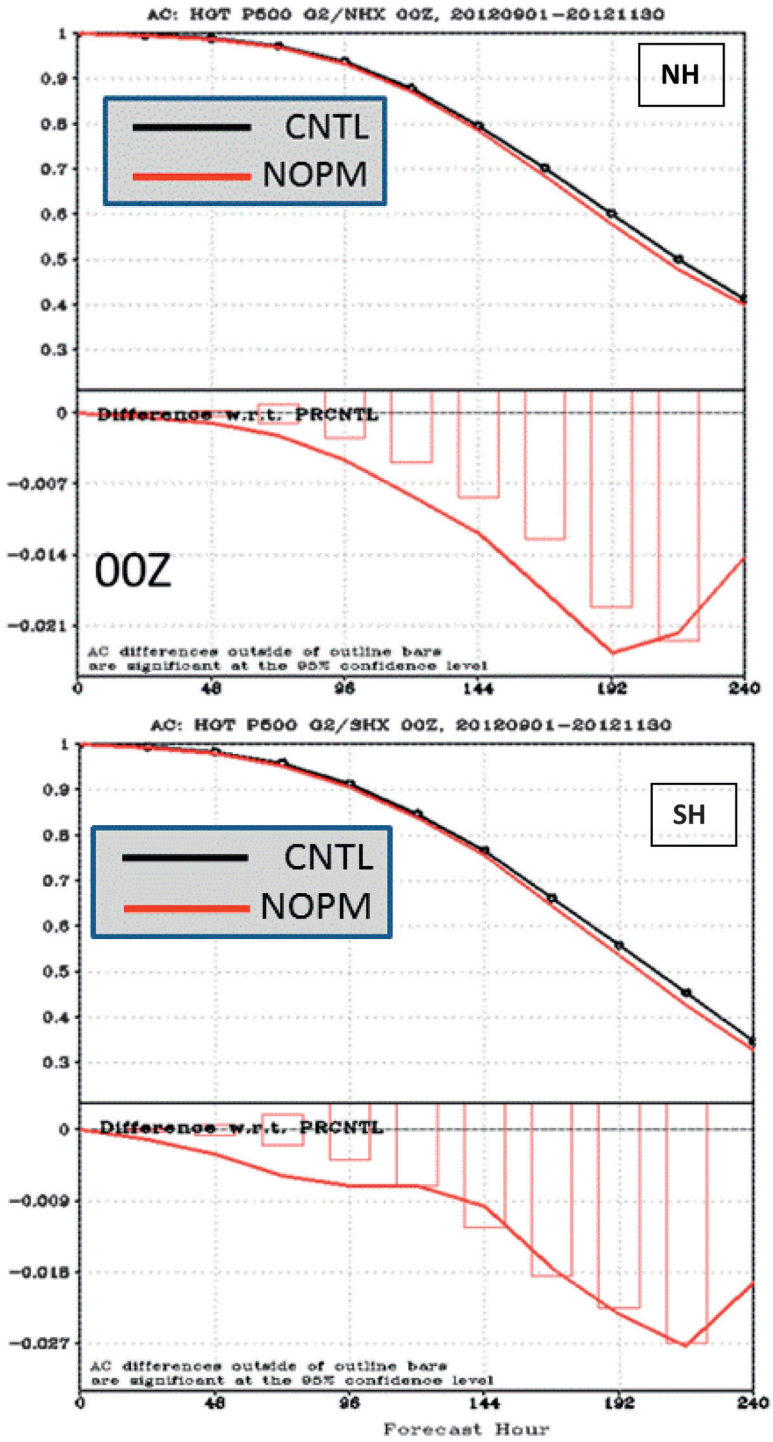


Fig. 4. Average (top) NH and (bottom) SH 0000 UTC Z500AC forecast scores for 1–10 days over SON for CNTL and NOPM. The lower panel of each plot shows the difference between CNTL and NOPM. Differences outside of the boxes are statistically significant at the 95% level.

are largely redundant and therefore add less additional information for data assimilation (EUMETSAT 2011). Compared to the hyperspectral information from AIRS and IASI, HIRS data provide relatively insignificant information. *NOAA-15* data coverage, on the other hand, does contribute by filling the uncovered area between the AM and PM orbits, so it effectively is an early-AM instrument as noted earlier (Fig. 2). NCEP did not use DMSP SSMIS data operationally until February 2015 and therefore the data were not included in this OSE.

OSE SETUP. The GDAS for both the CNTL and NOPM runs began on 0000 UTC 15 July 2012 and ended 7 months later on 0000 UTC 15 February 2013. Results in this paper cover the period 0000 UTC 1 August to the end of the runs in February; the July period is a spinup for all runs. The GFS forecast was run four times daily from the experiment's beginning until 3 November 2012 in order to generate the maximum number of global forecasts with hurricanes; thereafter, until 15 February 2013, the GFS forecast was run once per day at 0000 UTC. A total of 293 10-day forecasts[2] were made. While typical NCEP OSEs are run for 4–6 weeks for two seasons, this OSE extends for 7 months over three seasons and is one of the longest performed by NCEP. As such, it provides an opportunity to measure the impact of observations across different seasons with a continuous GDAS run and to assess the statistical significance with a very large number of cases. The CNTL and NOPM experiment differ only in the polar-orbiting observations omitted from the data assimilation in the NOPM experiment

[2] While the GFS forecast is run to 16 days in operations, the 10-day forecast for this OSE covers the most skillful part of the forecast that is most sensitive to, and appropriate for showing, the impact of observations and initial conditions on forecast accuracy.

Unauthenticated | Downloaded 01/20/21 02:01 AM UTC

(Table 2). Furthermore, we enhance the evaluation of the CNTL and NOPM runs with results from NCEP's operational (OPS) run, which was executed with the same model and data assimilation system and included all the observation sources listed in appendix A (see Table A1).

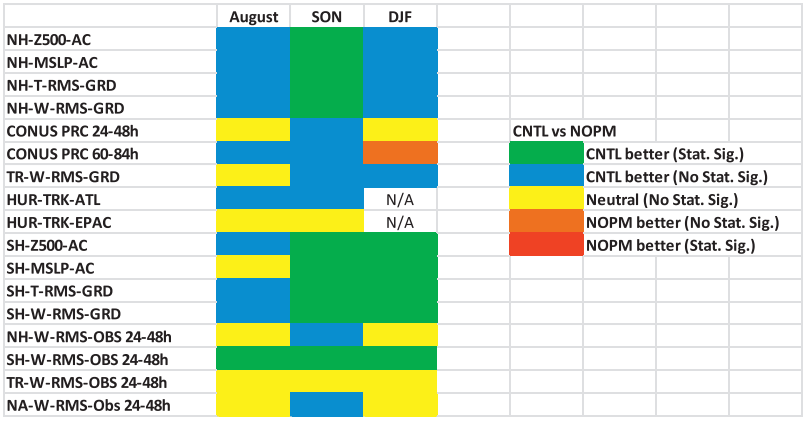


Fig. 5. Qualitative scorecard summary of OSE 1–10-day forecast performance for CNTL vs NOPM over the experimental period. Colors indicate an overall qualitative assessment of whether the CNTL or NOPM score was better for all forecast times and the associated statistical significance, if any. Results are divided into August, SON, and DJF periods. See Table 3 for definitions of each score.

EVALUATION PROCEDURES. A complete evaluation of all the OSE results is very complex and demanding, and is beyond the limited scope of this paper. As noted earlier, differences in skill are customarily measured by standard NWP scores, including the Z500AC, which is an overall measure of the skill (appendix B). Other standard statistical measures (Table 3) are summarized here and can provide supporting evidence by measuring different aspects of the model forecast output, for example, precipitation ETS and hurricane track errors. The evaluation covers almost three seasons and thereby captures seasonal variability, if any. Choosing representative case studies is a very subjective process, with its own challenges, and will not be attempted here since the focus is on the quantitative and objective information that can be gleaned from a more detailed analysis of the Z500AC score alone.

Standard verification scores (Table 3) were used to evaluate forecast skill for each run. Anomaly correlation and some RMS scores were verified against analyses. To provide an independent analysis estimate, the European Centre for Medium-Range Weather Forecasts (ECMWF) analysis was used as verification for geopotential height, temperature, and winds. Forecast verification for longer than 3 days using either the GDAS or a multicenter analysis does not change the results presented here simply because the forecast errors are much larger than any analysis differences. To verify precipitation, the NCEP Climate Prediction Center daily precipitation analysis, assembled from over 10,000 conterminous United States (CONUS) 24-h rain gauge reports, was used. Hurricane track scores, verified against National Hurricane Center best-track data for the Atlantic (ATL) and east Pacific (EPAC) basins, were accumulated from the four-per-day forecasts through 120 h. Short-term (24–48 h) temperature and wind forecasts were also verified against radiosonde observations in the NH, SH, tropics (TR), and North America (NA). All other statistics were generated from forecasts initiated at 0000 UTC 1 August 2012 to 0000 UTC 15 February 2013 for 0000 UTC and 1200 UTC 1 August to 1200 UTC 2 November 2012 for 1200 UTC. Statistical significance at the 95% confidence level was determined by a Student's *t* test (Hogg and Craig 1978) for all scores except for precipitation ETS listed in Table 3. For ETS, a Monte Carlo resampling method (Hammersley and Handscomb 1975) with 10,000 realizations was used to determine its confidence level. A qualitative scorecard was generated to provide an overview of all results.

For an additional perspective on the OSE impacts, these scores are compared with the annual distribution of 5-day Z500AC scores from the operational GFS (appendix B), the annual mean history of which is characterized by an increase of skill over time (Fig. B1). Furthermore, the distribution of Z500AC scores (Fig. B2) shows that, over the period 1996–2014, the percentage of low scores has decreased remarkably, while the percentage of high scores has increased. In this paper, we explore whether similar trends accompany changes to the satellite observing system, such as those tested with this OSE.

OVERALL RESULTS. Since a quantitative comparison of OSE results creates an enormous set of scores for different forecast variables, times, and model levels, the focus here is on the 5-day Z500AC score because it is an overall indicator of forecast quality in the extratropics and is commonly used for NWP model comparisons (Simmons and Hollingsworth 2002).

The NH and SH 5-day Z500AC time series (Fig. 3) shows that each of the CNTL, NOPM,

Unauthenticated | Downloaded 01/20/21 02:01 AM UTC

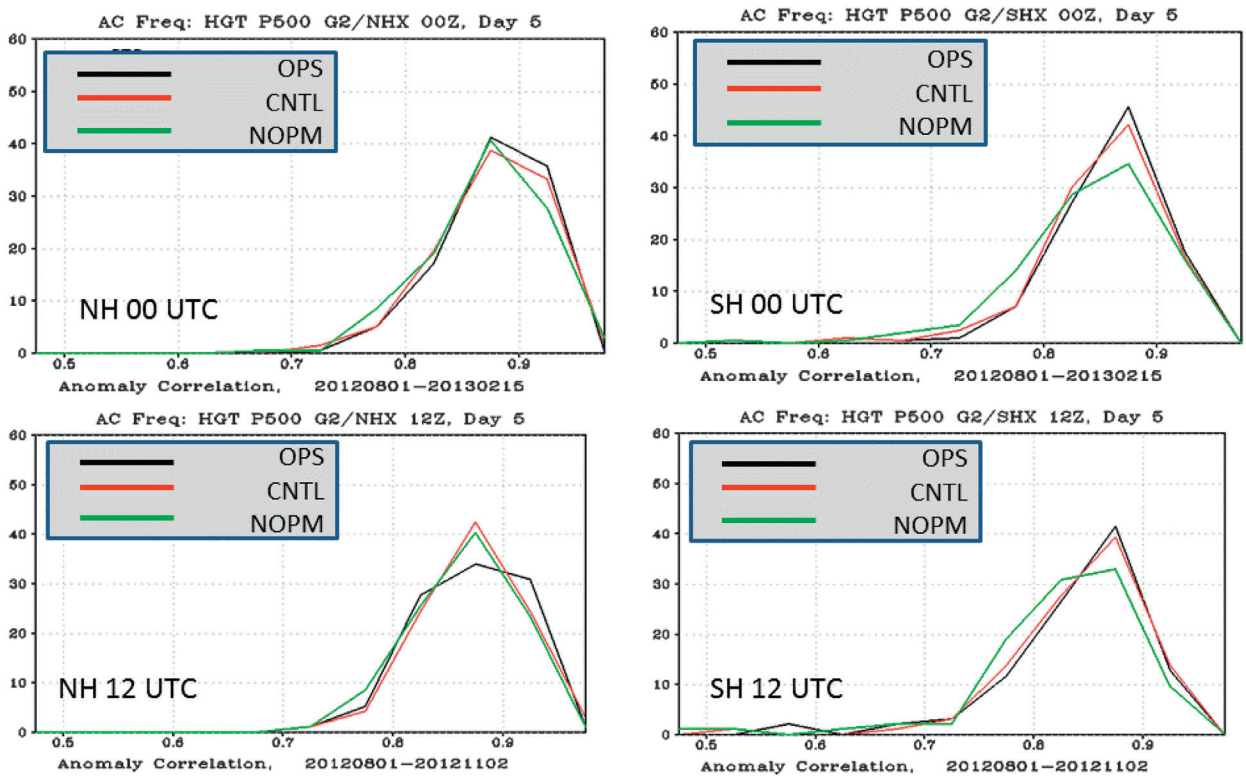


Fig. 6. Skill distribution of 5-day Z500AC scores for CNTL, NOPM, and OPS scores for (top) 0000 and (bottom) 1200 UTC in (left) NH and (right) SH. The 1200 UTC runs were produced only through 3 Nov 2012, so they do not include winter scores. Scores from each experiment are distributed into 20 bins of width 0.05 between 0.025 and 0.975. The abscissa is the Z500AC binned score, and the ordinate is the percentage in each bin.

and OPS runs have the highest or lowest scores at multiple times throughout the experiment period. Although there are episodic outlier results for the CNTL and NOPM runs—for example, late August in the SH (NOPM), from mid-September through mid-October in the NH (NOPM), and late January in the NH and SH (CNTL)—it appears visually that the performance of each of the three runs is almost indistinguishable from the others; that is, no run is superior throughout the entire time series. To confirm this fact, scores were compared head to head for each verification time. A frequency breakdown of the highest and lowest scores for all runs (Table 4) shows that the NOPM produces the smallest percentage of the highest scores in both hemispheres and the largest percentage of the lowest scores. The latter impact is stronger in the SH, where there are many fewer nonsatellite observations and the influence of satellite data is correspondingly larger as expected.

Table 5. Quintile cutoff values for 5-day Z500AC for CNTL based on the 293 cases, including 0000 and 1200 UTC; e.g., the lowest quintile for the NH has a maximum value of 0.835931.

Category (%)	Description	Upper cutoff value	
		NH	SH
<20	Worst	0.835931	0.814969
21–40	—	0.869417	0.844165
41–60	—	0.892238	0.869863
61–80	Good	0.915522	0.895398
81–100	Best	1.000000	1.000000

Over the entire experiment, the mean 5-day Z500AC score for 0000 UTC initial conditions (Fig. 3) is 0.005 and 0.013 larger for CNTL than for NOPM in the NH and SH, respectively—about equal to approximately one year of increase in the GFS annual mean Z500AC (appendix B). While the differences in CNTL and NOPM 5-day Z500AC scores appear to be small, if not underwhelming, they persist throughout the entire 1–10-day forecast period (Fig. 4) and are statistically significant at the 95% confidence level for 1–8 days in the NH and 1–5 days in the SH over

Unauthenticated | Downloaded 01/20/21 02:01 AM UTC

the September–November (SON) part of the experiment. From December to February (DJF; not shown), the CNTL Z500AC scores were significantly higher for 1–10 days only in the SH. In August (also not shown), the NH Z500AC scores were not significantly different, but the CNTL score was significantly higher for 1–3 days in the SH.

The comparison of overall CNTL and NOPM 1–10-day forecast performances is summarized by a scorecard (Fig. 5) showing superiority and any statistical significance for all scores (Table 3) over three subperiods: August only (AUG), SON, and DJF. In the NH, the CNTL is consistently superior at all forecast times for Z500AC, mean sea level pressure anomaly correction (MSLP-AC), and RMS-GRD scores, but it is significantly better for SON only. In the SH, where differences between the CNTL and the NOPM are larger, the CNTL is significantly better for both SON and DJF. CONUS precipitation scores for the CNTL are either neutral or insignificantly better for most subperiods, but they are worse for 60–84 h in DJF. Tropical RMS-GRD wind scores are not significantly better. Hurricane track errors for the CNTL are better in the ATL basin but neutral in the EPAC basin. A greater number of scores are statistically significant in the SH than in the NH, as might be expected due to the higher reliance on satellite data in the SH. Overall, statistically significant results are mostly found in SON and DJF but rarely in AUG, presumably because the number of AUG verifying times is much smaller. RMS-OBS scores for 24–48 h are mostly neutral (or not significant), the exception being SH winds. Interestingly, the statistical significance for NH Z500AC and other scores changes from significant in SON to insignificant in DJF; the reason for this change is not readily apparent.

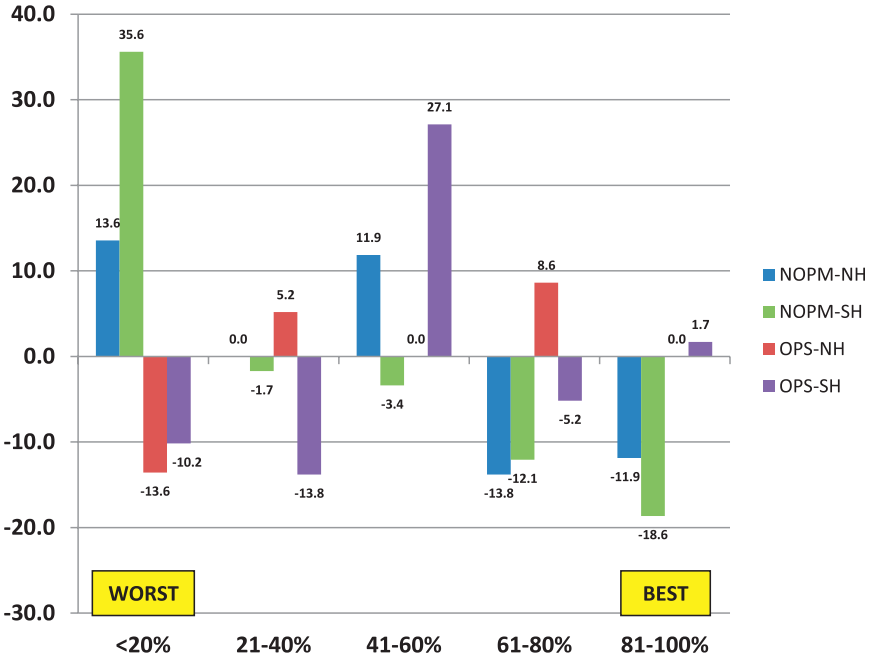


Fig. 7. Comparative fraction of 5-day NOPM and OPS Z500AC scores in each of the quintiles defined by Table 5 for the NH (blue and red) and SH (green and purple). Comparison measures the percentage increase or decrease of NOPM and OPS scores in each quintile relative to CNTL, which is (by construction) equally populated in each quintile.

The OPS mean 0000 UTC Z500AC score for the entire experimental period is 0.002 and 0.001 higher than CNTL in the NH and SH, respectively (Fig. 3). Compared to the CNTL experiment for all verification dates, OPS produces about 5% more instances in both hemispheres when it has the highest score (Table 4), but OPS also has about the same fraction of the lowest scores. Despite having about the same fraction of the lowest scores in a head-to-head comparison, it is interesting to note that the OPS time series also does not have any episodic low-score outliers. We speculate that this additional resilience of the OPS observing system (Andersson and Sato 2012), by having more AMSU-A instruments in orbit relative to the future 2O–4S configuration, has the potential to increase skill marginally in the time mean, but more importantly it may reduce the possibility of an episodic low-score forecast.

ANALYSIS OF SKILL DISTRIBUTION. After accounting for seasonal skill changes by collecting scores over a calendar year, the Z500AC skill distribution appears to be characteristic of a particular NWP system (appendix B). GFS improvements over the period 1996–2014 have resulted in a reduction of low scores and an increase in high scores. Since changing the observing system constitutes a change to the NWP system, we are motivated to apply the skill distribution analysis described in appendix B to the CNTL, NOPM, and OPS runs.

The OPS and CNTL skill distributions are very similar (Fig. 6), except at 1200 UTC in the NH, and markedly different from NOPM. The impact of removing the PM orbit is clearly seen as a shift in the distribution toward

Unauthenticated | Downloaded 01/20/21 02:01 AM UTC

Table A1. Observing systems used by the NCEP operational GDAS in 2012–13. Pibals = pilot balloons. MDCRS = Meteorological Data Collection and Reporting System. AMDAR = Aircraft Meteorological Data Relay. ASOS = Automated Surface Observing System. AWOS = Automated Weather Observing System. AVHRR = Advanced Very High Resolution Radiometer. SBUV/2 = Solar Backscatter Ultraviolet. MODIS = Moderate Resolution Imaging Spectroradiometer. *COSMIC-1 = Constellation Observing System for Meteorology, Ionosphere and Climate–1. GRACE-A = Gravity Recovery and Climate Experiment–A. SAC-C = Satélite de Aplicaciones Cientificas–C. C/NOFS = Communication/Navigation Outage Forecasting System. Earth Observation System–Aura = Aura.* OMI = Ozone Monitoring Instrument. GOES = Geostationary Operational Environmental Satellite. Meteosat = Meteorological Satellite. SEVIRI = Spinning Enhanced Visible and Infrared Imager. JMA = Japan Meteorological Agency. MTSAT = Multifunctional Transport Satellite.

	Platform	Instrument/measurement
Conventional		
Soundings	Balloon	Radiosonde
	Ship	Radiosonde
	Pibals	Wind
	Profiler	Wind
Commercial aircraft	MDCRS	Wind, temperature
	Canadian AMDAR	Wind, temperature
	European AMDAR	Wind, temperature
	Pilot reports	Wind, temperature
Land surface	Airport surface data	Surface meteorology
	ASOS and AWOS	Surface meteorology
	Mesonets	Surface meteorology
Marine	Ship	Surface meteorology, SST
	Moored buoy	Surface meteorology, SST
	Drifting buoy	Surface meteorology, SST
Hurricane	Dropsondes	Wind, temperature, moisture, pressure
	Military reconnaissance data	Wind, temperature, moisture, pressure
Satellite		
NOAA POES	*NOAA-19*	AMSU-A
		HIRS
		MHS
		AVHRR
		SBUV
	NOAA-18	AMSU-A
		MHS
		SBUV
	NOAA-17	HIRS[a]
		SBUV
	NOAA-15	AMSU-A
EUMETSAT (polar orbiting)	*MetOp-A*	IASI
		AMSU-A
		MHS
		HIRS
		GOME
		GRAS
		ASCAT

Unauthenticated | Downloaded 01/20/21 02:01 AM UTC

Table A1. Continued.		
	Platform	Instrument/measurement
NASA	*Aqua*	AIRS
		AMSU-A
		MODIS (AMVs)
	SNPP	ATMS
GNSS-RO	*COSMIC-1*	RO
	TerraSAR-X	RO
	GRACE-A	RO
	SAC-C	RO
	C/NOFS	RO
NASA	*Aura*	OMI
NOAA GOES	*GOES-13*	Sounder
	GOES-15	Sounder
		Imager (AMVs)
EUMETSAT	*Meteosat-10*	SEVIRI (Imager)
		SEVIRI (AMVs)
	Meteosat-7	Imager (AMVs)
JMA	*MTSAT-2*	Imager (AMVs)

[a] *NOAA-17* HIRS was assimilated until 13 Dec 2012.

increased frequency of low scores for both 0000 and 1200 UTC and in both hemispheres. It is more noticeable in the SH, where the influence of satellite data is larger and the time-mean loss of skill due to removing the PM orbit is correspondingly larger. Indeed, mean skill differences over the experimental period due to loss of the PM orbit (Fig. 4) show up more clearly in Fig. 6 as shifts in the NOPM skill distribution that are particularly evident at the tails of the distribution, that is, for the lowest and highest forecast scores. The shortened 1200 UTC time series does not include November–February cases, resulting in a smaller sample size by more than 50% and a broader NH skill distribution as shown (Fig. 6, bottom left) that is most likely a result of the smaller sample. The impact of additional instruments in OPS, relative to the CNTL, is present in the form of slightly more higher-than-average scores in the NH and fewer low scores in the NH and SH at 0000 UTC.

To quantify the abovementioned results, we divide the CNTL score distribution into quintiles, each with 20% of the total number of forecast scores (Table 5). The shape of this CNTL distribution is described by the quintile boundaries and is a reference against which the NOPM and OPS are then compared. For the sake of brevity, we label forecasts in the lowest quintile as the "worst"[3] of the CNTL distribution and the two highest quintiles as the "good" and "best" forecasts, respectively. Measuring the shifts of both NOPM and OPS skill distributions relative to the reference CNTL quintiles gives quantitative statements about the tails of the distributions, which (as noted previously) represent the probability of a worst or best GFS forecast.

The changes in skill distributions (Fig. 7) are calculated by determining the percentages of NOPM and OPS forecasts in each of the CNTL quintiles defined in Table 5. Relative to CNTL, NOPM is 13.6% more likely to produce a worst GFS forecast in the NH and 35.6% more likely to do so in the SH. For best GFS forecasts, NOPM is 11.9% less likely to populate the upper 20% of CNTL scores in the NH and 18.6% less likely in the SH. OPS changes relative to CNTL are less dramatic but nonetheless consistent with previous statements associated with Fig. 6. The additional observations in OPS reduce the likelihood of worst GFS forecasts in both the NH (13.6%) and SH (10.2%) but have little overall impact on improving the best GFS forecasts in either hemisphere. Instead, reductions in frequencies of the lowest 40% of CNTL SH scores of 10.2% and 13.8% show up as a large frequency increase (27.1%) in the middle quintile of the CNTL skill distribution.

SUMMARY AND DISCUSSION. An OSE using the NCEP GFS has been designed and executed to measure the potential impact of the loss of PM polar-orbiting observations in the future 2O–4S configuration of

[3] The appellation "worst" is relative and may, in fact, be a very good score when compared to other or past forecast systems.

Unauthenticated | Downloaded 01/20/21 02:01 AM UTC

the JPSS era. The control (CNTL) ingested observations from the operational GOS, including those from a polar-orbiting MW (temperature and moisture) sounder and a hyperspectral IR sounder in both the mid-AM and PM orbits. The PM instrument data were removed from the NOPM run. Data used were from the operational observations received by NCEP in the period 2012–13.

Removing PM orbit satellite observations results in generally inferior standard scores in the NH and SH, with the impact being greater in the SH. The NOPM experiment has inferior mean anomaly correlation and RMS scores, and these differences are statistically significant in SON in the NH and in both SON and DJF in the SH. Precipitation, tropical wind scores, and hurricane track errors are not significantly impacted although the trend is toward some degradation. These results, including the larger SH impact and more significant extratropical impact, are generally consistent with those from other OSEs over the last decade (Zapotocny et al. 2008; McNally 2012). Comparing the OSE CNTL with NCEP's OPS, it appears that adding three AMSU-A MW sounders increases the mean Z500AC score incrementally, but not significantly, in both hemispheres.

Analysis of the skill distributions for each of the CNTL, NOPM, and OPS runs is more revealing. Comparing CNTL and NOPM, removing the PM orbit data produces notable shifts toward increases in the number of low scores and clear decreases in the number of the highest scores. NOPM is 13.6% more likely to produce a worst GFS forecast in the NH and 35.6% more likely to do so in the SH. NOPM is 11.9% less likely to populate the upper 20% of the CNTL scores in the NH and 18.6% less likely in the SH. Comparing CNTL and OPS, there is a decrease in the likelihood of generating low scores in OPS of 13.6% and 10.2% in the NH and SH, respectively. These numbers suggest that the three additional AMSU-A instruments add resilience to the GOS, consistent with Andersson and Sato (2012). Furthermore, they suggest that an early-AM satellite, in particular, would add value and overall resilience to the GOS due to improved global data coverage over a 6-h period.

The skill distribution analysis demonstrates the well-known fact that the GFS (and any other operational forecast system) produces forecasts of variable skill from day to day. With an annual accumulation of scores to account for seasonal changes in forecast skill, it appears that the distributions are stable and well describe annual skill improvements due to scientific development and observing system changes. In particular, improvements to the GFS over almost two decades have dramatically changed the distribution of scores, and similar shifts in skill distribution are seen by removing the PM orbit observations in this OSE.

On any given day, even when the GDAS is cycled from its preceding instance, the resulting forecast may have more or less skill, depending on many factors. Some of these factors are the synoptic meteorology of that day; the accuracy of the initial analyses; the amount, type, and quality of observations; the ability of the quality control to remove erroneous observations and the ability of the observations to measure the key synoptic features; and the analysis and model accuracy for projecting the analysis forward in time. All in all, the diversity and complexity of these factors conspire to make the predictability of a single forecast from a single forecast system a challenging task, even when ensemble techniques are introduced (e.g., Wobus and Kalnay 1995; Tan and Xie 2003). The skill distribution analysis for NOPM can be quantitatively interpreted as an increased risk of producing more forecasts in the low end of the CNTL skill distribution and a reduced probability of producing forecasts at the high end.

We suggest that this skill distribution analysis could be useful for users, in particular for operational forecasters who desire and appreciate documentation on the performance of the numerical guidance used every day. Changes in the likelihood of making the worst or best forecasts (namely, on either end of the skill distribution) could be beneficial for forecaster services. In particular, quantifying a change in the risk of using guidance with an enhanced (or reduced) probability of making a comparatively better (or worse) forecast should provide decision-making information.

ACKNOWLEDGMENTS. The authors thank NCEP and the anonymous reviewers for their comments, R. Treadon for providing Fig. 2, and P. Caplan for originating the processing of GFS skill distribution. Partial support for this work was provided by the JPSS and Next Generation Global Prediction System (NGGPS) Programs via NOAA Grants 1312M41460 and NA14NES4320003, respectively.

APPENDIX A: LIST OF OBSERVING SYSTEMS USED IN NCEP OPERATIONAL GLOBAL DATA ASSIMILATION SYSTEM IN 2012–13.

The GOS is an ever-changing collection of instruments and systems that provides observations to international NWP centers and also serves local government, industry, and public needs. It is important to keep track of all GOS changes in instrument type, number, quality, etc., since it is clear that operational forecast quality depends on these factors. The failure of a particular satellite instrument, for example, is only predictable statistically, as any instrument can

Unauthenticated | Downloaded 01/20/21 02:01 AM UTC

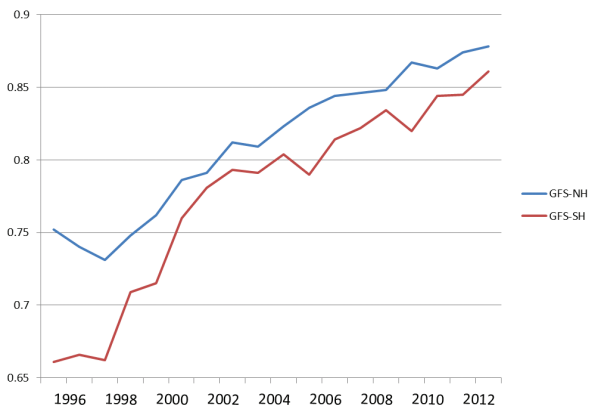


Fig. B1. Annual-mean NCEP GFS 5-day Z500AC scores from 1996 to 2013 in NH (blue) and SH (red).

exceed its designed lifetime or fail upon launch or soon thereafter, often with serious consequences due to the expense involved in its replacement.

Table A1 lists the types of observations, platforms, and instruments (or quantities and measurements) used operationally at NCEP during the period of this OSE.

APPENDIX B: CONTEXT FOR OBSERVING SYSTEM IMPACTS.

It is informative to place observing system impacts, as demonstrated by OSEs, in context with the long-term skill improvements to operational global forecast systems, such as the NCEP GFS, as documented by a standard and representative score such as Z500AC. The Z500AC is a representative score since it measures the skill of forecast of high and low pressure locations and the vertically averaged atmospheric state, and furthermore it has a long history as a performance metric. Other scores (such as root-mean-square error) tend to move in tandem with Z500AC, while scores for precipitation and hurricane track and intensity are more specialized and tend to measure less representative aspects of atmospheric behavior.

Operational NWP centers are constantly improving their analysis and forecast systems. System improvements can result from scientific development of their many complex components. Development areas include but are not limited to increased quantity and quality of ingested observations from the GOS; improvements to the data assimilation and quality control algorithms and procedures; and improvements to various aspects of the forecast model, such as the representation of physical processes, increasing horizontal and vertical resolution, and increasing computational efficiency. Increased computational efficiency is important because it enables more sophisticated science to be added while maintaining the same computational cost in operations.

The average improvement rate for operational global forecast systems is approximately one day of skill per decade (Simmons and Hollingsworth 2002); that is, the average skill of today's 5-day forecast is as good as that of a 4-day forecast produced a decade ago. The skill of the NCEP GFS has improved at the same rate, with average mean-annual increases in Z500AC of 0.007 (NH) and 0.010 (SH) as shown in Fig. B1. These increases in skill were due to the accumulated value of system improvements such as those noted above. For example, GFS horizontal resolution increases occurred in 1998 (100–70 km), 2002 (55 km),

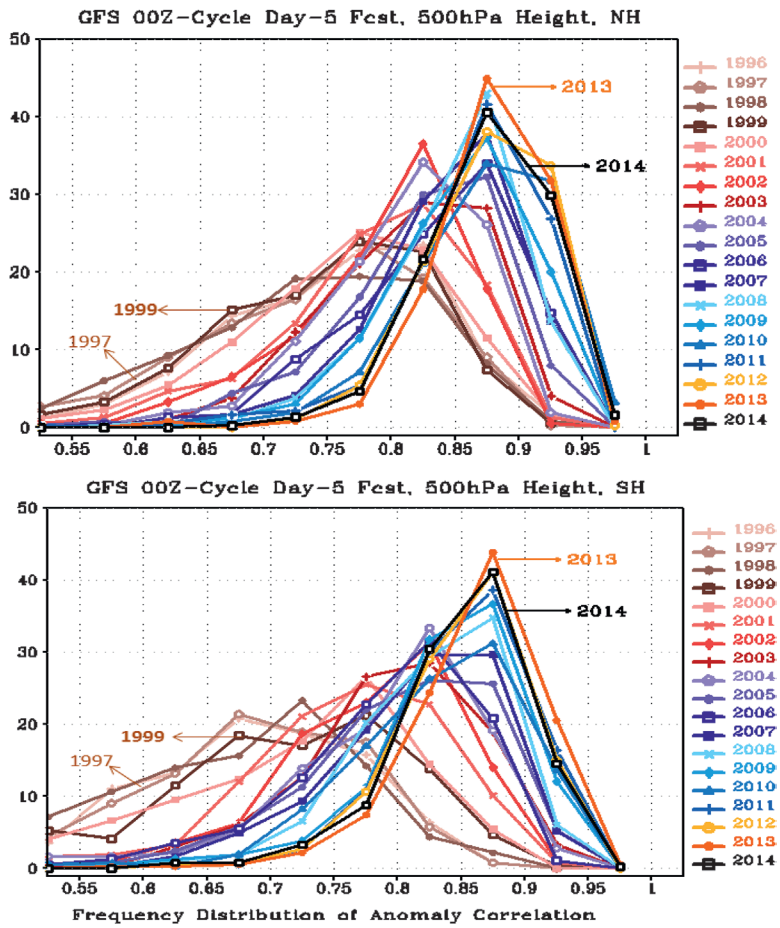


Fig. B2. As in Fig. 6, but for NCEP operational GFS 5-day forecasts at 0000 UTC, compiled annually from 1996 to 2014.

Unauthenticated | Downloaded 01/20/21 02:01 AM UTC

2005 (38 km), 2010 (23 km), and 2015 (13 km), and all were enabled by operational high-performance computing (HPC) increases and enhanced computational efficiency. Most of these horizontal-resolution changes resulted in higher annual scores the next year in one or both hemispheres (Fig. B1), even though other system changes undoubtedly contributed.

Distributions of GFS forecast skill for each year over the period 1996–2014 (Fig. B2) provide even more information on the impact of improvements. Despite some minor year-to-year variability in forecast skill due to different weather patterns and despite the fact that GFS upgrades occurred irregularly over this period, it is generally apparent that each annual skill distribution is unique to the GFS of that particular year. Notably, as annual-mean scores have increased, their skill distributions are characterized by a reduced frequency of low scores and an increased frequency of high scores. Contrast, for example, the distributions for NH over 1997–99 and 2012–14: Scores in the range 0.525–0.625 constituted 16%–18% of the total in the earlier period but 1% in the most recent years. From 1997 to 1999, the GFS scores did not reach 0.925 but, in each year of 2012–14, 30%–35% of the NH scores did so.

As a forecast system improves its ability to extract observational information through its DAS and increase its forecast skill through a better model, it becomes more resilient to changes in the observing system and less likely to produce forecasts in the lower range of scores.

REFERENCES

Andersson, E., and Y. Sato, Eds., 2012: WIGOS: WMO Integrated Global Observing System; Final report of the Fifth WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction, WMO Tech. Rep. 2012-1, 25 pp. [Available online at www.wmo.int/pages/prog/www/OSY/Reports/NWP-5_Sedona2012.html.]

Böttger, H., P. Menzel, and J. Pailleux, Eds., 2004: Proceedings of the Third WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction. WMO/TD-1228, 324 pp. [Available online at www.wmo.int/pages/prog/www/documents.html.]

Cardinali, C., 2009: Monitoring the observation impact on the short-range forecast. *Quart. J. Roy. Meteor. Soc.*, **135**, 239–250, doi:10.1002/qj.366.

Chen, Y., F. Weng, Y. Han, and Q. Liu, 2008: Validation of the Community Radiative Transfer Model by using CloudSat data. *J. Geophys. Res.*, **113**, D00A03, doi:10.1029/2007JD009561.

——, Y. Han, P. Van Delst, and F. Weng, 2010: On water vapor Jacobian in fast radiative transfer model. *J. Geophys. Res.*, **115**, D12303, doi:10.1029/2009JD013379.

Cucurull, L., 2010: Improvement in the use of an operational constellation of GPS radio occultation receivers in weather forecasting. *Wea. Forecasting*, **25**, 749–767, doi:10.1175/2009WAF2222302.1.

——, and J. C. Derber, 2008: Operational implementation of COSMIC observations into the NCEP's Global Data Assimilation System. *Wea. Forecasting*, **23**, 702–711, doi:10.1175/2008WAF2007070.1.

——, and R. A. Anthes, 2015: Impact of loss of microwave and radio occultation observations in operational numerical weather prediction in support of the U.S. data gap mitigation activities. *Wea. Forecasting*, **30**, 255–269, doi:10.1175/WAF-D-14-00077.1.

——, J. C. Derber, and R. J. Purser, 2013: A bending angle forward operator for global positioning system radio occultation measurements. *J. Geophys. Res. Atmos.*, **118**, 14–28, doi:10.1029/2012JD017782.

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

EUMETSAT, 2011: The case for EPS/MetOP Second Generation: Cost benefit analysis. 38 pp. [Available online at www.wmo.int/pages/prog/sat/meetings/documents/PSTG-3_Doc_11-04_MetOP-SG.pdf.]

Eyre, J. R., and W. P. Menzel, 1989: Retrieval of cloud parameters from satellite sounder data: A simulation study. *J. Appl. Meteor.*, **28**, 267–275, doi:10.1175/1520-0450(1989)028<0267:ROCPFS>2.0.CO;2.

Garrett, K., 2013: Forecast impact assessments of *SNPP* ATMS. *JCSCA Quarterly*, No. 43, Joint Center for Satellite Data Assimilation, College Park, MD, 3–4.

Gelaro, R., and Y. Zhu, 2009: Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models. *Tellus*, **61A**, 179–193, doi:10.1111/j.1600-0870.2008.00388.x.

Goldberg, M. D., H. Kilcoyne, H. Cikanek, and A. Mehta, 2013: Joint Polar Satellite System: The United States next generation civilian polar-orbiting environmental satellite system. *J. Geophys. Res. Atmos.*, **118**, 13 463–13 475, doi:10.1002/2013JD020389.

Grody, N., J. Zhao, R. Ferraro, F. Weng, and R. Boers, 2001: Determination of precipitable water and cloud liquid water over oceans from the NOAA 15 advanced microwave sounding unit. *J. Geophys. Res.*, **106**, 2943–2953, doi:10.1029/2000JD900616.

Hammersley, J. M., and D. C. Handscomb, 1975: *Monte Carlo Methods*. Methuen and Co., 178 pp.

Han, Y., and Coauthors, 2013: Suomi NPP CrIS measurements, sensor data record algorithm, calibration

Unauthenticated | Downloaded 01/20/21 02:01 AM UTC

and validation activities, and record data quality. *J. Geophys. Res. Atmos.*, **118**, 12 734–12 748, doi:10.1002/2013JD020344.

Hogg, R. V., and A. T. Craig, 1978: *Introduction to Mathematical Statistics*. 4th ed. Macmillan, 225 pp.

Joo, S., J. Eyre, and R. Marriott, 2012: The impact of MetOp and other satellite data within the Met Office global NWP system using an adjoint-based sensitivity method. Met Office Forecasting Research Tech. Rep. 562, 20 pp.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp.

Kelly, G., A. P. McNally, J.-N. Thepaut, and M. Szyndel, 2004: OSEs of all main data types in the ECMWF operation system. Proceedings of the Third WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction, WMO/TD-1228, 63–94. [Available online at www.wmo.int/pages/prog/www/documents.html.]

Kim, E., C.-H. J. Lyu, K. Anderson, R. V. Leslie, and W. J. Blackwell, 2014: S-NPP ATMS instrument prelaunch and on-orbit performance evaluation. *J. Geophys. Res. Atmos.*, **119**, 5653–5670, doi:10.1002/2013JD020483.

Kleist, D. T., and K. Ide, 2015: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part I: System description and 3D-hybrid results. *Mon. Wea. Rev.*, **143**, 433–451, doi:10.1175/MWR-D-13-00351.1.

——, 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, doi:10.1175/2009WAF2222201.1.

Langland, R. H., and N. L. Baker, 2004: Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus*, **56A**, 189–201, doi:10.1111/j.1600-0870.2004.00056.x.

Liu, J., and E. Kalnay, 2008: Estimating observation impact without adjoint model in an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **134**, 1327–1335, doi:10.1002/qj.280.

Lorenc, A. C., and R. Marriott, 2014: Forecast sensitivity to observations in the Met Office global numerical weather prediction system. *Quart. J. Roy. Meteor. Soc.*, **140**, 209–224, doi:10.1002/qj.2122.

McNally, A. P., 2012: Observing system experiments to assess the impact of possible future degradation of the Global Satellite Observing Network. ECMWF Tech. Memo. 672, 20 pp. [Available online at www.ecmwf.int/sites/default/files/elibrary/2012/11085-observing-system-experiments-assess-impact-possible-future-degradation-global-satellite.pdf.]

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

——, M. Bonavita, and J. Thepaut, 2014: The role of satellite data in the forecasting of Hurricane Sandy. *Mon. Wea. Rev.*, **142**, 634–646, doi:10.1175/MWR-D-13-00170.1.

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

Pailleux, J., E. Andersson, and M. Ondráš, Eds., 2008: Proceedings of the Fourth WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction. WMO/TD-1450, 214 pp. [Available online at www.wmo.int/pages/prog/www/documents.html.]

Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763, doi:10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2.

Simmons, A. J., and A. Hollingsworth, 2002: Some aspects of the improvement of skill of weather numerical prediction. *Quart. J. Roy. Meteor. Soc.*, **128**, 647–677, doi:10.1256/003590002321042135.

Tan, J., Z. Xie, and L. Ji, 2003: A new way to predict forecast skill. *Adv. Atmos. Sci.*, **20**, 837–841, doi:10.1007/BF02915409.

Wang, X., D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-based ensemble–variational hybrid data assimilation for NCEP Global Forecast System: Single-resolution experiments. *Mon. Wea. Rev.*, **141**, 4098–4117, doi:10.1175/MWR-D-12-00141.1.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press, 467 pp.

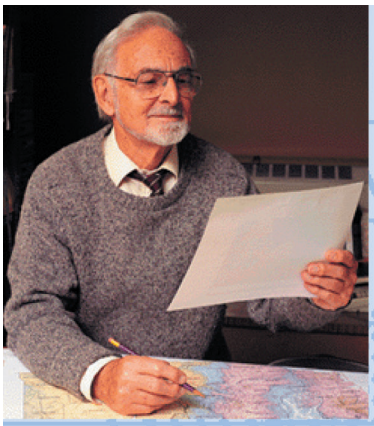
WMO, 2010: Global aspects. Vol. 1, Manual on the Global Data-Processing and Forecasting System, WMO-485, 196 pp. [Available online at www.wmo.int/pages/prog/www/DPFS/Manual/GDPFS-Manual.html.]

Wobus, R. L., and E. Kalnay, 1995: Three years of operational prediction of forecast skill at NMC. *Mon. Wea. Rev.*, **123**, 2132–2148, doi:10.1175/1520-0493(1995)123<2132:TYOOPO>2.0.CO;2.

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

Zavyalov, V., and Coauthors, 2013: Noise performance of the CrIS instrument. *J. Geophys. Res. Atmos.*, **118**, 13 108–13 120, doi:10.1002/2013JD020457.

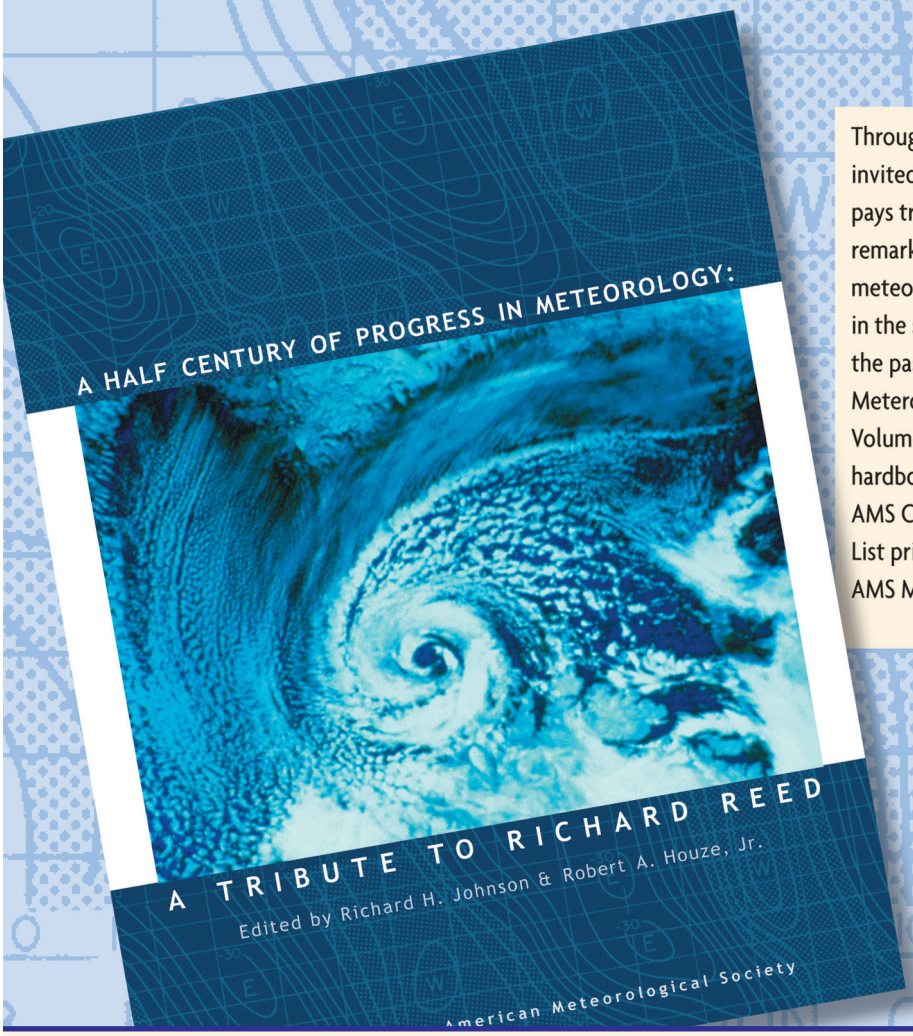
Unauthenticated | Downloaded 01/20/21 02:01 AM UTC



A Half Century of Progress in Meteorology: A Tribute to Richard Reed

edited by Richard H. Johnson and Robert A. Houze Jr.

with selections by: Lance F. Bosart Robert W. Burpee Anthony Hollingsworth James R. Holton Brian J. Hoskins Richard S. Lindzen John S. Perry Erik A. Rasmussen Adrian Simmons Pedro Viterbo



Through a series of reviews by invited experts, this monograph pays tribute to Richard Reed's remarkable contributions to meteorology and his leadership in the science community over the past 50 years. 2003. Meterological Monograph Series, Volume 31, Number 53; 139 pages, hardbound; ISBN 1-878220-58-6; AMS Code MM53.
List price: $80.00
AMS Member price: $60.00

ORDER ONLINE: bookstore.ametsoc.org or see the order form at the back of this issue

Unauthenticated | Downloaded 01/20/21 02:01 AM UTC