

The Reanalysis for the Global Ensemble Forecast System, Version 12[Ⓞ]

THOMAS M. HAMILL,^a JEFFREY S. WHITAKER,^a ANNA SHLYAEVA,^b GARY BATES,^c SHERRIE FREDRICK,^c
PHILIP PEGION,^c ERIC SINISKY,^{d,e} YUEJIAN ZHU,^e VIJAY TALLAPRAGADA,^e HONG GUAN,^{e,f} XIAQIONG ZHOU,^{d,e}
AND JACK WOOLLEN^{d,e}

^a NOAA/Physical Sciences Laboratory, Boulder, Colorado

^b Joint Center for Satellite Data Assimilation, University Corporation for Atmospheric Research, Boulder, Colorado

^c Cooperative Institute for Research in the Environmental Sciences, University of Colorado Boulder, Boulder, Colorado

^d I. M. Systems Group, Inc., College Park, Maryland

^e NOAA/NWS/NCEP/Environmental Modeling Center, College Park, Maryland

^f Systems Research Group, Inc., College Park, Maryland

(Manuscript received 10 February 2021, in final form 26 October 2021)

ABSTRACT: NOAA has created a global reanalysis dataset, intended primarily for initialization of reforecasts for its Global Ensemble Forecast System, version 12 (GEFSv12), which provides ensemble forecasts out to +35-days lead time. The reanalysis covers the period 2000–19. It assimilates most of the observations that were assimilated into the operational data assimilation system used for initializing global predictions. These include a variety of conventional data, infrared and microwave radiances, global positioning system radio occultations, and more. The reanalysis quality is generally superior to that from NOAA's previous-generation Climate Forecast System Reanalysis (CFSR), demonstrated in the fit of short-term forecasts to the observations and in the skill of 5-day deterministic forecasts initialized from CFSR versus GEFSv12. Skills of reforecasts initialized from the new reanalyses are similar but slightly lower than skills initialized from a preoperational version of the real-time data assimilation system conducted at the higher, operational resolution. Control member reanalysis data on vertical pressure levels are made publicly available.

KEYWORDS: Variational analysis; Ensembles; Data assimilation; Reanalysis data

1. Introduction

The Global Ensemble Forecast System (GEFS) is one of several prediction systems maintained by the U.S. National Weather Service (NWS). It supports medium-range weather and subseasonal to seasonal forecasting. The newest version 12 of the GEFS (GEFSv12 hereinafter) produces 31-member ensemble forecasts. The forecasts initialized at 0000 UTC are generated to +35-days lead time, and the forecasts initialized at 0600, 1200, and 1800 UTC are generated to +16-days lead time. The GEFSv12 system configuration and performance are documented in a companion article (Zhou et al. 2021, manuscript submitted to *Wea. Forecasting*). Like other prediction systems in the NWS, this system is typically updated every several years.

As current-generation ensemble guidance is limited by systematic errors including resolution limitations, errors in the mean state, and errors in the ensemble spread (standard deviation about the mean), products from the GEFS are commonly statistically postprocessed. A time series of past ensemble forecasts and coincident observations/analyses are used to estimate the systematic errors and adjust the real-time forecast, improving skill and reliability (Vannitsem

et al. 2018). Statistical postprocessing can also filter the predictable signal from the meteorological noise due to chaotic error growth and sampling variability due to finite ensemble size. When used in combination with higher-resolution analyses, statistical postprocessing of ensemble forecasts can also provide a downscaling related to the modulation of weather by physiographic features (Hamill and Whitaker 2006).

For infrequent events and longer-lead forecasts where skill is marginal, the statistical postprocessing is greatly aided from a long training dataset of reforecasts, i.e., retrospective forecasts using the same prediction system used to generate the real-time forecasts (Hamill et al. 2004, 2006, 2008, 2013; Hamill and Whitaker 2006; Hagedorn et al. 2008; Scheuerer and Hamill 2015). Similarity of the bias and error characteristics of the reforecasts and the real-time forecasts is highly desirable for statistical postprocessing; in this way the training data resemble the real-time forecast data. When this similarity is lost, statistically postprocessed product quality may suffer, especially the postprocessing of shorter-lead forecasts (Hamill 2017). Differences in the bias of short-term real-time forecasts versus reforecasts may occur in part because of the differing characteristics of the initial conditions in the reanalysis versus the real-time analysis. For this reason, initialization of reforecasts from an archive of past operational initial states from obsolete versions of modeling systems is unlikely to provide the desired statistical consistency; the initial conditions will reflect in part the systematic errors of the outdated prediction system versions used to provide the background forecasts in the data assimilation. Hence, if computationally feasible,

[Ⓞ]Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/MWR-D-21-0023.s1>.

Corresponding author: Thomas M. Hamill, tom.hamill@noaa.gov

production of a new set of retrospective analyses (reanalyses) consistent with the operational system is highly desirable for reforecast initialization.

Many organizations across the globe have created global, multidecadal reanalyses. These are retrospective gridded analyses of the state of the atmosphere (and sometimes other state components such as the land, ocean, and sea ice), commonly produced with a cycled data assimilation system leveraging a prior background forecast updated to the newly available observations. Global reanalyses have been provided by the Copernicus Climate Service/European Centre for Medium-Range Weather Forecasts (ECMWF; Uppala et al. 2005; Dee et al. 2011; Laloyaux et al. 2018; Hersbach et al. 2020), the Japan Meteorological Agency (Kobayashi et al. 2015; Harada et al. 2016), the National Aeronautics and Space Administration (NASA; Rienecker et al. 2011; Gelaro et al. 2017), the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR; Kalnay et al. 1996), and the NCEP–Department of Energy (Kanamitsu et al. 2002). Several have been generated by organizations within the National Oceanic and Atmospheric Administration, including the Climate Forecast System Reanalysis (CFSR; Saha et al. 2010) and several versions of the Twentieth Century Reanalysis (20CR; Compo et al. 2011; Slivinski et al. 2019). Most of these reanalyses, such as Copernicus/ECMWF’s “ERA5,” are intended primarily for weather and climate monitoring and coupled state estimation, and their design and choice of observations to assimilate emphasizes this. ERA5, for example, creates a near-surface temperature analysis using 2-m temperature observations alongside the full 3D atmospheric analysis. This temperature analysis has many applications, but it is used only indirectly in the initialization of ECMWF’s medium-range forecasts (to affect the soil-state estimate).

The main purpose of the reanalysis discussed in this paper, the reanalysis for the GEFSv12 is *reforecast initialization*. The GEFSv12 reanalysis was designed so that its initial conditions will be consistent with the operational data assimilation system [the data assimilation system for the Global Forecast System, version 16 (GFSv16)] used to initialize the real-time GEFSv12 ensemble predictions, to the extent practical. If a user seeks a reanalysis for other applications such as climate monitoring or evaluation over a longer time than the 2000–19 period covered by GEFSv12, they may be better served by other reanalyses such as ERA5 or 20CR or MERRA-2, depending on the application and the length of record needed. The GEFSv12 reanalysis is an atmosphere-only reanalysis, for reasons discussed later, including computational expense.

While designed to facilitate GEFSv12 reforecast initialization for statistical postprocessing, we envision this reanalysis having several other applications. For example, suppose in the coming years that developers will need to initialize experimental ensemble forecasts for high-impact events such as Hurricane Katrina (2005) or Sandy (2012). The GEFSv12 reanalysis will be a useful initialization dataset of choice for the next several years of retrospective uncoupled forecast

experiments for the GEFS and its deterministic counterpart, the GFS. The reanalysis will provide suitable initial conditions consistent with the underlying prediction system, and the reforecasts will provide a forecast performance baseline. Eventually the statistical character of the real-time initial conditions will change, be it due to more realistic couplings between land/ocean/atmosphere/ice, or prediction-system improvements that alter the background bias (Hamill 2017), or assimilation system improvements, or more observations. Further, reanalyses of the coupled state may be necessary to initialize next-generation coupled-model reforecasts. At this future point, another reanalysis will be necessary.

As will be discussed, producing acceptable statistical consistency between the GEFSv12 reanalyses and the GEFSv12 real-time analyses (produced through truncation of the GFSv16-based data assimilation system) was challenging. Generating a real-time analysis is itself computationally expensive when high analysis resolution and modern data assimilation techniques are used, techniques such as four-dimensional variational (4D-Var; Courtier et al. 1994; Rabier et al. 2000) methods or four-dimensional ensemble-variational analysis (4D-En-Var; Wang and Lei 2014; Kleist and Ide 2015). Even with using a reduced-resolution version of the data assimilation system and with splitting the reanalysis production into streams (parallel cycles processing different segments of multidecadal period), it took an extended period of time and large computational resources, $O(1)$ year and $O(1000)$ cores per stream, to generate these uncoupled reanalyses. This necessitated production of the reanalysis before the final configuration of the GEFSv12 and the operational data assimilation system were finalized, leading to some slight inconsistencies between the GEFSv12 reanalysis configuration and the real-time GEFSv12 configuration.

This article, then, describes the GEFSv12 reanalysis system. Diagnostics will focus on the general quality of the reanalysis in comparison with its NOAA predecessor, the CFSR. It will also provide some data to compare the reanalysis with contemporaneous real-time analyses produced for the GFS and GEFSv12 forecast initialization.

The paper is organized as follows. Section 2 will briefly describe the data that were assimilated to produce the reanalysis. Section 3 provides a brief description of the assimilation system, the forecast model, the ensemble prediction system, and the land, ocean, and sea ice initialization procedures. Section 4 describes the characteristics of the reanalysis through diagnostics such as the fit of short-term forecasts to observations and the accuracy of medium-range deterministic forecasts initialized with GEFSv12 versus the preceding CFSR. It will also provide some diagnostics of the character of initial conditions and forecasts initialized from GEFSv12 reanalyses versus preproduction parallel simulations of the GEFSv12, known as “retro” runs. Retro runs, like the eventual operational real-time GEFSv12 simulations, used initial conditions inherited from the operational analysis system, which used GFSv16. Section 5 reviews known issues with the reanalysis, and section 6 describes the data storage, before section 7 concludes.

2. Description of the data assimilated in the GEFSv12 reanalysis

As usual with modern-era reanalyses, a variety of observation types were assimilated. Figure 1 shows the assimilated observation counts across the major broad categories of observation types. In general, the observation count increases with time. To simplify the data processing, it was decided to leverage the observational data that had already been preprocessed to generate other previous reanalyses. Hence, before 1 March 2011, observations collected for input to the CFSR reanalysis (Saha et al. 2010) were used. The reader is referred to this publication for a review of these data. After this date, observations from the operational data stream were used. As can be seen in Fig. 1, there is no radical jump in observation count before versus after this date, nor would we expect any change in observation quality.

Figure 2 provides a more detailed description of the satellite observations that were assimilated from particular satellites, including general categories and the start and end dates of their assimilation. More details on particular channels that were assimilated are provided in Table S1 of the online supplemental material. In general, the microwave and infrared (IR) radiances were assimilated from a number of geostationary and polar-orbiting satellites from the United States and Europe.

Satellite radiance data now provide the majority of the observations assimilated and have the greatest impact on the reanalysis quality (Gelaro et al. 2010), and the GEFSv12 reanalysis used a large number of channels from a variety of satellites. For ozone observations, Solar Backscatter Ultraviolet version 2 (SBUV/2) observations from NOAA satellites were used for ozone profile and Ozone Monitoring Instrument (OMI) *Aura* for total ozone observations (see Table S2 of the online supplemental material for details). Variational quality control was also used. For this reanalysis, a newer variational bias correction was used (Zhu et al. 2014) that replaced the previous version (Derber and Wu 1998) used in CFSR production. All-sky microwave radiances (Zhu et al. 2016, 2019) were also assimilated in the GEFSv12 reanalysis but not in the CFSR.

One observational data source was problematic, namely, the velocity–azimuth display (VAD; Browning and Wexler 1968; Lee et al. 2014) winds. While the reanalyses were being computed, we discovered that there was insufficient quality control of these data in earlier streams, and sometimes the low-level winds were contaminated by bird migrations. We did not have the computational capacity or time to restart the computations from the beginning of the cycles. Accordingly, when discovered, assimilation of these wind observations was turned off. For the 1999 stream, VAD data were not assimilated after 0000 UTC 21 February 2001. For the 2003 stream, VAD data were not assimilated after 1800 UTC 6 January 2005. For the 2007 stream, VAD data were not assimilated after 1800 UTC 27 November 2008. For other streams, QC was deemed acceptable, and their assimilation was continued.

3. Data assimilation and prediction system

This section describes the reanalysis procedure used for reforecast initialization and a high-level description of the

underlying prediction system used to produce the background forecasts in the cycled reanalysis system. While the GEFSv12 reanalysis inherits many characteristics of the GFS real-time analysis system that initializes the real-time GEFSv12 ensemble forecasts, the reanalysis system differs somewhat from the real-time analysis system, and from the previous-generation reanalysis, CFSR, with changes described below.

Table 1 summarizes the major differences between the CFSR and the GEFSv12 reanalysis systems. The most significant changes included use of a new atmospheric dynamical core to provide background forecasts, discussed in section 3a below. Another major difference was that the GEFSv12 reanalyses excluded the generation of accompanying three-dimensional ocean reanalyses for weakly coupled forecast initialization. Instead, previously produced time series of optimal interpolation (OI) sea surface temperature (SST) analyses provided the ocean boundary condition for the reanalyses. This simplification was made in part because as the GEFSv12 real-time forecasts and reforecasts extend to +35-days lead time, changes in the ocean state were smaller in magnitude during this first month than they were for the seasonal forecasts produced by the previous Climate Forecast System version 2 (CFSv2). How SSTs evolve in the GEFSv12 reforecasts will be discussed in the accompanying article on the GEFS forecast and reforecast procedure (Zhou et al. 2021, manuscript submitted to *Wea. Forecasting*; Guan et al. 2021).

The GEFSv12 reanalyses were conducted in parallel streams of five or more years in length, starting in 1999, 2003, 2007, 2011, and 2015. As soil moisture is quite sensitive to its initialization and takes a long period to stabilize, the first year of each stream was discarded. Hence the 1999 stream provides the reanalysis initial conditions during 2000–03, and the 2003 stream provides them during 2004–08, and so forth (Rienecker et al. 2011). An examination of the time-mean soil moisture for three regions with strong land–atmosphere coupling (Koster et al. 2004) are shown in Fig. 3. These three regions are the U.S. Southern Great Plains, northern equatorial Africa, and India. The time- and domain-mean soil moistures in these three areas do not exhibit a temporal discontinuity between stream boundaries (left column). Similarly, scatterplots of beginning-of-stream versus end-of-stream soil moisture at each grid point in the three domains are shown in the right-hand column, with extremely high correlation and little sign of bias. From this we infer that the separation of the analysis procedure into streams appears to provide a legitimate way to parallelize the reanalysis productions without degrading near-surface analysis quality.

a. Atmospheric dynamical core and physical parameterization suite used for background forecasts

The “FV3” dynamical core (Lin 2004; Putman and Lin 2007) is used in both the deterministic GFS and in this version 12 of the GEFS that was used as the background forecast model in the cycled data assimilation. Predictions using this dynamical core were much less diffusive, containing more small-scale variability relative to the previous spectral dynamical core. The FV3 dynamical core is also capable of cloud-

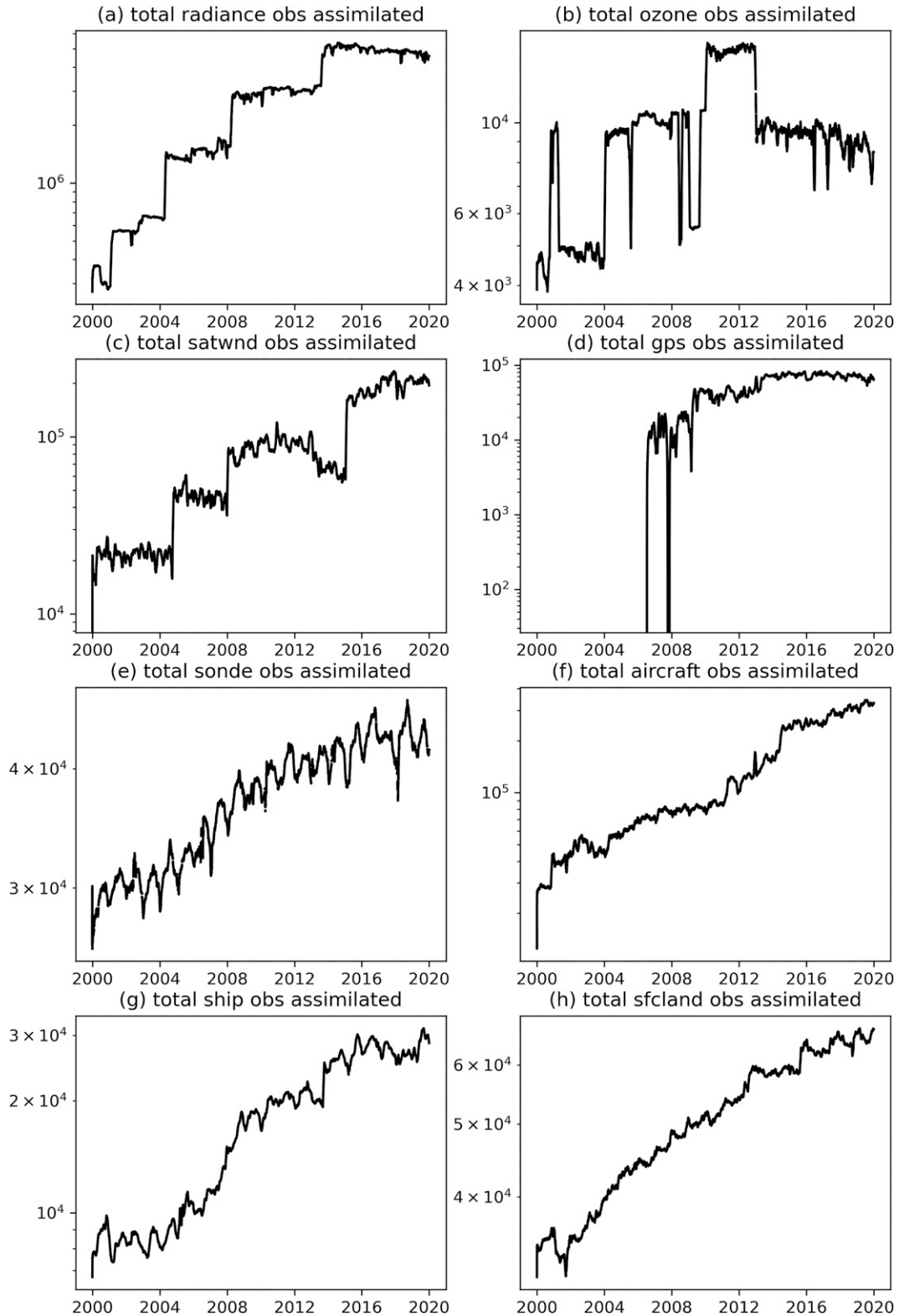


FIG. 1. Assimilated observation counts for major categories of observation types.

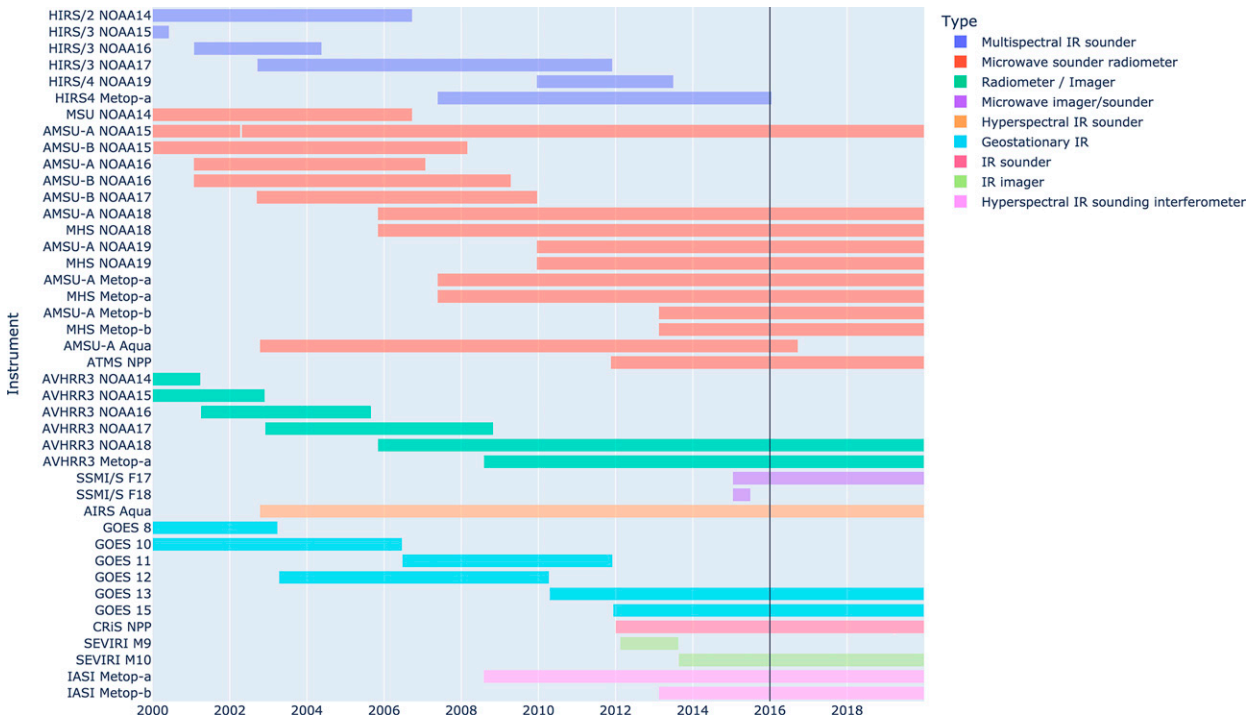


FIG. 2. Periods during which observations of various types were assimilated from satellite systems. The vertical black line indicates when the observation data switched from the data previously stored from the CFSR to the operational analysis data stream. Explanations of acronyms are provided in Table S1 of the online supplemental material.

resolving simulations when used at higher resolutions (Putman and Lin 2007). In the FV3-based GFS and GEFS, grid spacings are denoted with a “CZZZ” notation, denoting that there are $ZZZ \times ZZZ$ grid points across each of the six faces on the cube upon which Earth’s spheroidal shape is projected. For the C384 grid spacing chosen for the GEFSv12 reanalysis, this would indicate that there are 384×384 grid points around a latitude circle, or an effective grid spacing of approximately 0.23° . Vertical levels are shown in Fig. 4.

The parameterization suite used in conjunction with the FV3 dynamical core to produce the background forecasts was largely the same as that used in the GFS, version 15. Basic details on the parameterization suite, with associated references, can be found at https://dtcenter.ucar.edu/GMTB/v3.0/sci_doc/GFS_v15_page.html and https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gefs_v12.php.

b. Data assimilation procedure

We now consider the assimilation procedure, which generally reflects data assimilation system improvements incorporated into GFSv15.1 and GFSv15.2, as well as some of the improvements included in GFSv16. Ideally, the exact same assimilation procedure, types of observations, and prediction system would be used in the real-time and reanalysis systems. Because of the computational expense of producing multidecadal ensemble reanalyses, it was decided that the control reanalysis around which perturbed members are recentered would use a background forecast at half the resolution of the

real-time operational assimilation system for the GFS and GEFSv12, reanalysis C384 versus real-time C768. However, both the reanalysis’s and the real-time control’s forecast minus observation departures were computed at C384. Similarly, the ensemble providing background-error covariances was computed at C128 in the reanalysis versus C384 in the operational system. The impacts of resolution differences (and some other changes) will be examined in section 4, where diagnostics of the assimilation and forecast quality are examined.

Analyses were produced four times daily at 0000, 0600, 1200, at 1800 UTC, although the reforecasts were initialized at the end of the incremental analysis update window 3 h after the synoptic time. Figure 5 is a schematic illustrating the analysis workflow prior to application of the incremental analysis update. The data assimilation procedure used in the GEFSv12 reanalysis is a hybrid 4D-En-Var (Kleist and Ide 2015; Kleist et al. 2018). It reflects the improvements implemented for GFSv16 listed in https://www.weather.gov/media/notification/pdf2/scn21-20_gfsv16.0_aac.pdf, except for the inclusion of interchannel correlated observation errors for hyperspectral radiances and the use of updated land states from an offline run forced by observed precipitation. The 4D-En-Var was used to update the ensemble mean, while a local ensemble transform Kalman filter (LETKF) variant of the EnKF was used to update the 80 ensemble perturbations (Lei et al. 2018). The GEFSv12 reanalysis system also used a 4D incremental analysis update procedure (4D-IAU; Lei and

TABLE 1. A synthesis of the major differences between the CFSR and GFSv12 reanalyses.

Aspect changed	CFSR configuration	GEFSv12 configuration
Period of record	1978–current	2000–20
Atmospheric dynamical core and control forecast grid spacing	Spectral, T382L64 (~38-km grid)	FV3 (Lin 2004; Putman and Lin 2007); C384L64 (~25-km grid)
Microphysical parameterization	Zhao–Carr (Zhao and Carr 1997)	GFDL (Phillips and Donner 2006; Zhou et al. 2019)
Other parameterizations	Saha et al. (2010)	GFSv15 (2020)
Atmospheric data assimilation method	3D-Var (Parrish and Derber 1992; Kleist et al. 2009)	Hybrid 4D En-Var with 4D-IAU
Ensemble usage in data assimilation	None	80-member EnKF at C128L64 (~75 km) to provide background-error covariances
Ensemble stochastic physics	None (single control member for data assimilation)	Stochastically perturbed physical tendencies (SPPT), stochastic boundary layer relative humidity (SHUM), and stochastic kinetic energy backscatter (SKEB) (this paper)
Snow updates	“SNODEP” (Kopp and Kiess 1996) before 1997, NESDIS Interactive Multisensor Snow and Ice Mapping System (IMS; Helfrich et al. 2007) thereafter; updated four times daily	NESDIS IMS (Helfrich et al. 2007); updated only at 0000 UTC; otherwise, climatology for the other three cycles (a bug)
Land surface analysis	Separate land surface analysis with analyzed forcings (Saha et al. 2010)	Land surface forcings directly from short-term forecasts
Ocean analysis	SST via OI (Reynolds et al. 2002); rest of ocean state with 3D-Var using MOM4 ocean and weak coupling	SST via OI (Reynolds et al. 2002); no weak coupling in cycled DA and no full ocean analysis
Tropical cyclone processing	Vortex relocation to observed position (Liu et al. 1999)	Direct assimilation of central pressure; no relocation
Satellite bias correction	Variational (Derber and Wu 1998)	Variational (Zhu et al. 2014)
All-sky radiance assimilation	None	Zhu et al. (2016, 2019)
GPSRO assimilation	Refractivity	Bending angle
Variational correction of aircraft temperature bias	None	Zhu et al. (2015)

Whitaker 2016) to reduce gravity wave noise, and it used the linearized forward operator to compute observation prior ensemble perturbations (Shlyueva and Whitaker 2018).

After generating the reanalyses, we discovered a bug in the way snow was updated (see section 3d). To partially correct for this, we subsequently “replayed” (Orbe et al. 2017, section 2.1) the C384 deterministic solution to the C128 ensemble-mean analysis (after recentering the around the hybrid 4D-En-Var analysis) to downscale the C128 analysis to C384 resolution, while updating the land surface states using the correct snow analysis at each analysis time. The replay procedure utilized the same 4D-IAU process used in the data assimilation cycle. The only difference is that instead of recomputing the analysis increments by rerunning the hybrid 4D-En-Var data assimilation, it used the previously generated C128 analysis to compute the increments used to constrain the higher resolution model trajectory. This C384 “replay” analysis was used as the ensemble-mean state used to initialize the C384 reforecast ensemble (after superimposing the upscaled C128 ensemble perturbations).

For the 4D-IAU procedure, analyses were produced every 3 h within the 6-h assimilation window (at the beginning,

middle, and end of the window). The model was restarted from the beginning of the assimilation window and forced by analysis increments interpolated in time to the model time step from the 3-hourly fields.

Observation variational quality control and bias correction was performed by the operational variational Global Statistical Interpolation (GSI) code and was configured as in the operational GFS as of 2015. An aircraft temperature variational bias correction was employed in the reanalysis (Zhu et al. 2015) that was not used in the CFSR. The GEFSv12 reanalysis assimilated global positioning system radio occultations (GPSRO) bending angle, instead of refractivity as in CFSR (Cucurull et al. 2013).

c. Stochastic physics in the cycled ensemble for data assimilation

Model uncertainty in the background ensemble and in the 4D-IAU corrector segment was parameterized by a suite of schemes that consisted of stochastically perturbed physics tendencies (SPPT; Palmer et al. 2009), stochastic kinetic energy backscatter (SKEB; Shutts 2005; Berner et al. 2009) and stochastic specific humidity perturbations in the boundary layer

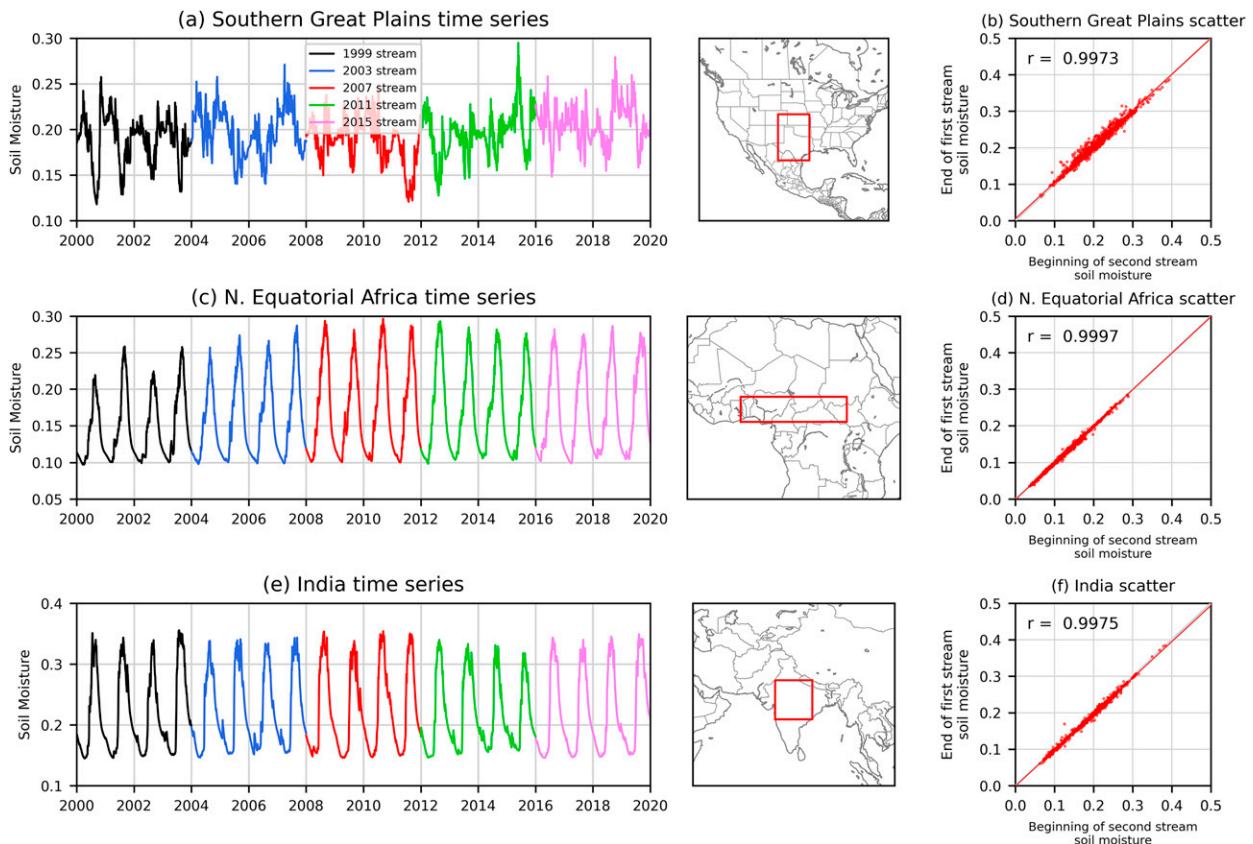


FIG. 3. Time series of mean volumetric soil moisture in overlapping streams for three areas with strong land–atmosphere coupling: (a) U.S. Great Plains, (c) northern equatorial Africa, and (e) India. (b),(d),(f) Also plotted are scatterplots of the soil moistures in each region between the last day of one stream and the first day of the next stream; each dot represents a separate grid point at one of the four stream boundaries.

(SHUM). The details of the GEFS implementation of SPPT and SKEB are outlined in Zhou et al. (2021, manuscript submitted to *Wea. Forecasting*).

There were several differences in the model uncertainty parameterizations used in the cycled data assimilation for reanalysis and real-time analysis relative to what were used in the GEFSv12 reforecasts and real-time forecasts. The first difference is that a single spatial and time scale for the random pattern was used in SPPT for the cycled ensemble data assimilation. The background ensemble used an e -folding horizontal length scale of 500 km, an e -folding time scale of 6 h, and an amplitude of 0.8, whereas the GEFSv12 predictions used multiple length and time scales similar to the ECMWF seasonal forecast system (Molteni et al. 2011; Zhou et al. 2021, manuscript submitted to *Wea. Forecasting*). The larger random patterns with their longer growth time scales are relevant for medium-range to S2S forecasts; however, these grow slowly and thus had negligible impact on the short-range background forecasts used in cycled data assimilation.

In several other respects, the application of the ECMWF SPPT method was changed. In early tests to integrate the SPPT to work with the FV3-based GFS, frequent model crashes occurred. These were traced to the interaction of the

planetary boundary layer (PBL) scheme and the mountain-blocking scheme. Both of these schemes produced very large momentum tendencies in regions of mountain blocking. When perturbed, they often produced unphysical oscillations that cause numerical stability issues. The solution chosen was to apply SPPT tendencies only above a dividing streamline, which was defined as the level in the atmosphere at which the flow below was blocked due to orography; the SPPT scheme was not active in the portion of the atmosphere that was considered blocked. Another issue with the implementation of SPPT was a wet precipitation bias globally. The perturbed moisture removed by the microphysics was not reflected in the precipitation at the surface. The solution was to perturb the surface precipitation with the same random number to ensure physical consistency with the perturbed tendencies in the atmosphere.

The SKEB for the cycled background ensemble was the same as GEFSv12, but here we used a single length-scale and time-scale amplitude of 500 km and 6 h, respectively.

In addition, another stochastic scheme was active for the cycled background ensemble that was not used in GEFSv12. Specific humidity tendency perturbations were applied to the lower layers of the model each physics time step. The perturbations rapidly decreased in amplitude above the surface,

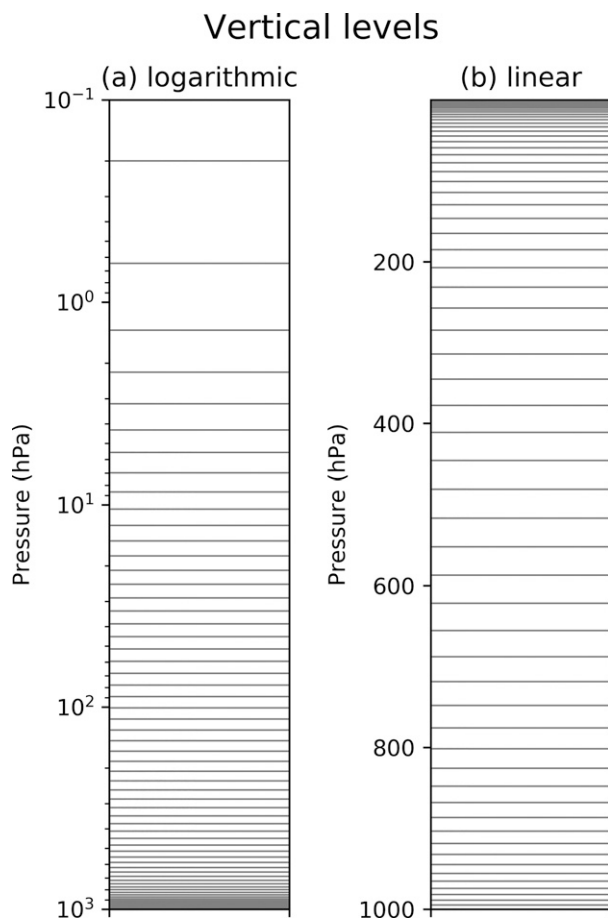


FIG. 4. Vertical levels for the cycled data assimilation and forecast model used in this reanalysis for a surface pressure of 1000 hPa, plotted as (a) logarithmic in pressure and (b) linear in pressure.

with an e -folding scale of 0.2 in the vertical coordinates (σ). Thus, if the surface pressure was 1000 hPa, the SHUM tendency perturbation at 800 hPa would be $\sim 36.7\%$ of that at the surface. The humidity perturbations were intended to account for unrepresented variability in small-scale convective triggering features such as gust fronts, cold pools, and subgrid humidity variability. This stochastic scheme was inspired by [Tompkins and Berner \(2008\)](#), which used random samples of specific humidity distributions

provided by the cloud scheme in the convection scheme. As opposed to the Tompkins and Berner method, which had ensembles of humidity profiles available, in the member-by-member processing of each background ensemble forecast only the single humidity column profile was available at a given grid point with this model's parameterization schemes. Hence, our approximation to their algorithm was to perturb the specific humidity with a random value, which is on the order of 0.1% per time step. Like SPPT and SKEB, the perturbations were correlated in space and time with an e -folding time scale of 6 h and spatial scale of 500 km.

d. Land, ocean, and sea ice initialization

We produced an atmosphere-only reanalysis and inherited analyses of other state components from other systems. Previously generated OI ([Reynolds et al. 2002](#)), version 2, SST analyses at $1/4^\circ$ provided the ocean state. This differs from the GEFSv12 real-time SST initialization procedure, known as near sea surface temperature (NSST; [Minnett et al. 2019](#)). As there can be strong diurnal variations in the ocean skin temperature, especially under sunny conditions and weak winds, the NSST includes some simplified dynamics of vertical mixing in the top ocean layers and its variation with atmospheric forcings. The background ocean SST state predicted by NSST is updated to available in situ and satellite observations. Use of the NSST algorithm was not included in the GEFSv12 reanalysis because its use resulted in excessive SST bias in climatologically cloudy regions during the early years of the reanalysis (not shown), when there were fewer in situ SST observations and infrequent IR views of the ocean. Note that NSST is turned on in the reforecasts and real-time GEFSv12 forecasts.

There were some systematic differences between SSTs from the OI and from NSST. [Figure 6](#) shows the mean skin temperature at 0000 UTC from the preproduction parallel SST analyses minus the skin temperature in the reanalysis during a period from late 2017 to late 2019. Over the ocean, the skin-temperature reanalysis was the OI SST analysis. Generally, mean differences were small over the oceans, though differences in excess of 1 C were found in higher-latitude oceans, with a cold bias of NSST relative to OI. The lack of major differences over the tropics suggests that surface-based convection should trigger similarly when initialized from reanalyses as in the real-time production system. This is probably of greater consequence to representation of the general circulation than the midlatitude differences.

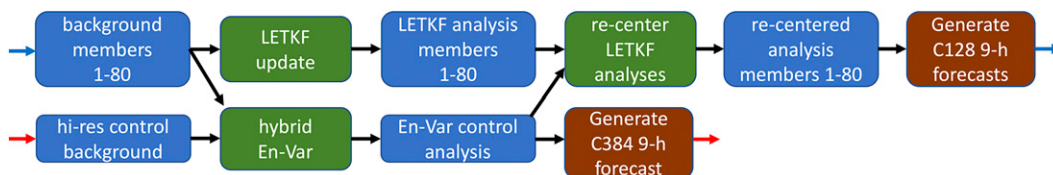


FIG. 5. Cycled data assimilation data and process flow used in the GEFSv12 reanalysis. Cycled lower-resolution LETKF analyses were maintained alongside the higher-resolution control En-Var analysis. Background cycled LETKF forecasts informed the background-error covariance model in the hybrid En-Var. LETKF analyses were recentered around the En-Var control. The LETKF analyses were computed at ensemble resolution C128, lower than operations by a factor of 3. The control analysis was computed at C384.

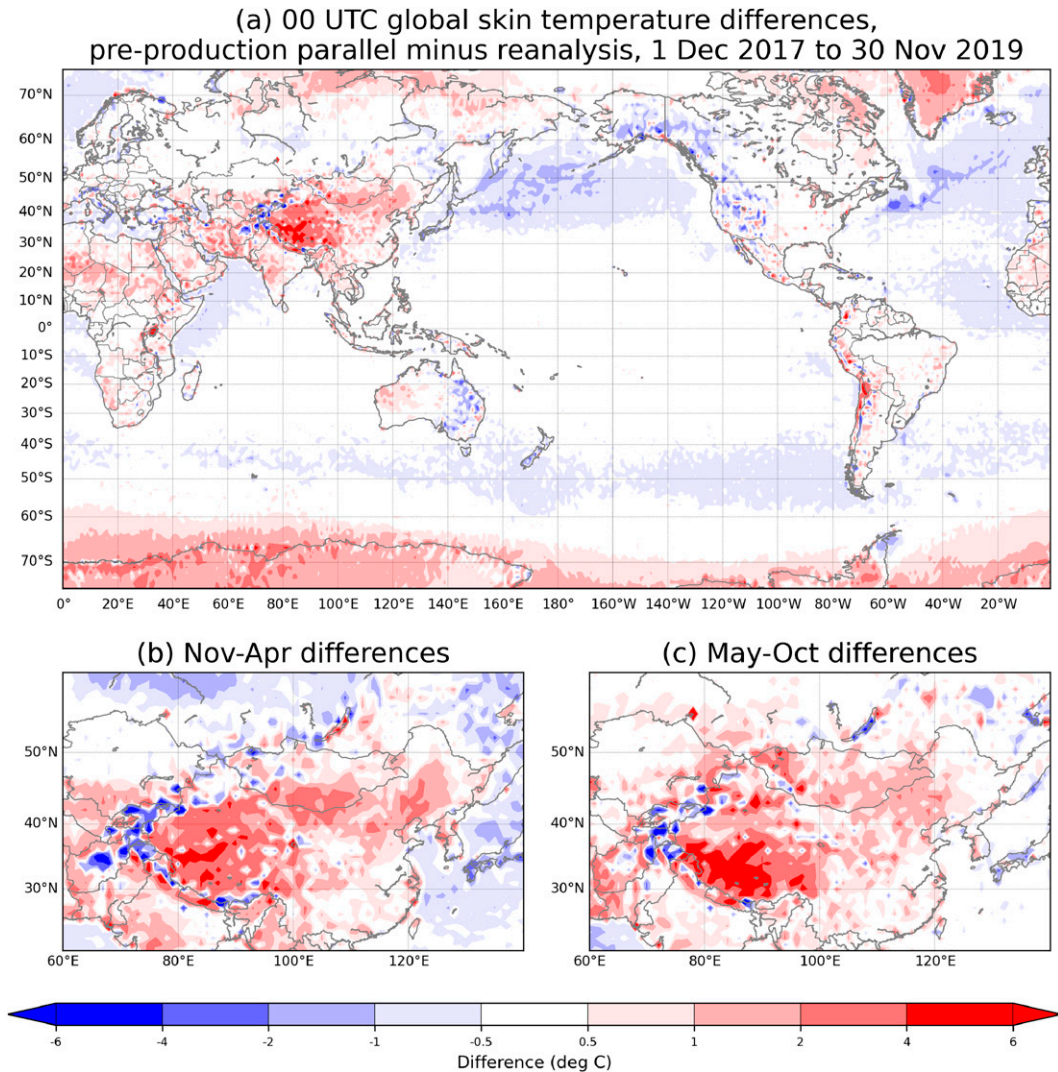


FIG. 6. Differences between 0000 UTC analyzed skin temperatures, preproduction parallel analysis minus reanalysis, (a) for the period 1 Dec 2017–30 Nov 2019 and for a subset of the domain for dates within this period and for the months (b) November–April and (c) May–October.

Initialization of snow and ice were designed to be as described in Saha et al. (2010). One known bug in this reanalysis was that the cycled DA erroneously inserted climatological snow amounts for three of the update cycles each day (0000, 0600, and 1200 UTC), whereas the actual snow analysis was inserted only at 1800 UTC. The previously described replay procedure that was used to regenerate the control C384 analysis was also applied ex post facto to adjust the reanalysis states to the correct snow analyses at all four synoptic times.

There was no direct data assimilation of top-level soil moisture observations or 2-m temperature/humidity in this reanalysis as there was in other operational prediction facilities (e.g., ECMWF 2019), although related techniques are in development within NOAA. Further, unlike the real-time GEFSv12, there was no insertion of a standalone soil-state

analysis from the Global Land Data Assimilation System (GLDAS; Rodell et al. 2004; Meng et al. 2012), as multidecadal GLDAS reanalyses were not available at the time of GEFSv12 reanalysis production. Instead, the background forecasts provided precipitation, temperature, and radiative forcings to the underlying cycled land model, but there was no explicit soil-state update based on atmospheric values. Deeper-soil moisture (layers 2–4, below 10 cm) were relaxed to externally specified climatology with a time scale of 60 days, with the climatology specified from an older version of the GLDAS system. The choice of 60 days was informed by experience (F. Yang 2021, personal communication).

There were some larger skin-temperature differences over land between the reanalysis and the preproduction parallel. In regions with more varied terrain, some of these may represent the differences in the resolution of the computational

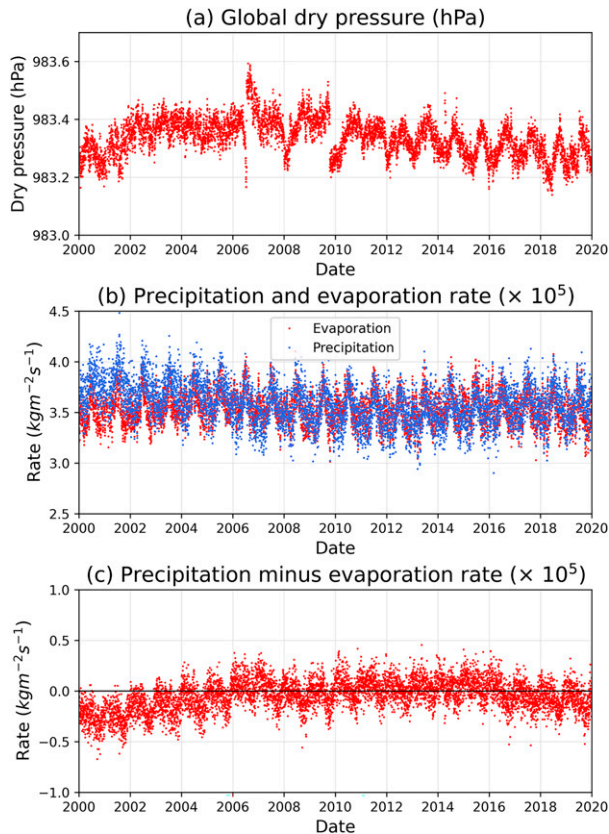


FIG. 7. Time series of (a) dry surface pressure, (b) precipitation and evaporation rate, and (c) precipitation minus evaporation rate.

grids, which was twice as high in the preproduction parallel (C768) than in the reanalysis (C384), with more terrain detail in the former. Still, areas such as the Himalayan plateau had mean temperatures warmer by several degrees C in the preproduction parallel analyses. Figures 6b and 6c show that the differences around China were persistent from cool to warm season and thus were unlikely due only to the previously mentioned bug related to the snow initialization once per day. Perhaps this region was particularly bias prone, and the insertion of GLDAS states in the preproduction parallel raised these surface temperatures relative to the reanalysis without GLDAS. Understanding the ultimate sources of these differences is left for possible future research.

4. Statistical characteristics of the reanalysis and scout runs

a. Conservation properties and the quasi-biennial oscillation

Figure 7a shows the time series of the analyzed global-mean dry surface pressure, which should be constant. There is an option for a weak constraint on dry mass during the data assimilation to enforce constancy, but it was inadvertently turned off for this reanalysis. Hence the dry surface pressure reflects random and systematic changes to the thermodynamic structure from the assimilation of observations. Analyzed

pressure exhibits two nonmeteorological jumps, the first around 13–15 July 2006 and the second around 11–13 October 2009. The first of the pressure jumps coincides with the advent of the assimilation of the more-dense GPSRO (Kursinski et al. 1997; Anthes et al. 2008). When these unbiased GPSRO observations began to be assimilated in large numbers, they provided an effective anchoring of the microwave radiance data (Cucurull et al. 2014), and these radiance bias corrections changed notably over a short period. The result was that the analyzed thermodynamic structure of the atmosphere changed, and with it the integrated surface pressure. An in-depth examination of the second pressure jump did not reveal any such conclusive cause.¹

The water cycle in the GEFSv12 reanalysis was not perfectly closed. Figures 7b and 7c show time series of globally averaged precipitation and evaporation and their difference. This makes the water-cycle balance comparable to the MERRA reanalysis but worse than in MERRA-2. The MERRA-2 reanalysis used extra algorithmic adjustments to ensure near closure of the water cycle (Bosilovich et al. 2017, Fig. 1).

Basic characteristics of the analysis of the quasi-biennial oscillation (QBO; Coy et al. 2016; Pascoe et al. 2005) are shown in Fig. 8. The QBO diagnostics are based on the monthly and zonally averaged reanalysis *u*-wind component from -10°S to 10°N . The annual cycle was first removed by subtraction of the 20-yr average of each month. Figure 8a provides the winds as a function of pressure (ordinate) and date (abscissa). Filtering was then performed with Fourier analysis, retaining only the first-nineteenth harmonics to retain variability longer than a year (periods of 12.6–240 months). The filtered data are presented in Fig. 8b. The time versus pressure series closely resembles those presented from other reanalyses (e.g., Coy et al. 2016, Fig. 5). The coarser stratospheric resolution in this reanalysis does not permit examination of characteristics above 10 hPa.

b. Fit of the background forecasts to observations

One way of monitoring the quality of the reanalysis is to display a time series of the statistics of differences between the short-term background forecasts (interpolated or converted to the observation type) and the observations. As more observations are added and as the forecast model and assimilation algorithms are improved from one reanalysis system to the next, one would expect the control background forecast to more closely fit the observations in the newer reanalysis system. As an example, the fit of background forecasts to “conventional” observations (primarily rawinsondes, aircraft, and surface and marine in situ observations) are shown in Fig. 9 for the 800–900-hPa layer for both this reanalysis and the previous generation reanalysis, the CFSR. Individual dots denote weekly averaged values, and fitted curves with annual

¹ We considered changes in the observation network as a cause, but there were no significant changes at this time. While SSM/I precipitation retrievals were turned off in the NASA reanalysis in September of 2009, these were not assimilated in the GEFSv12 reanalysis. Diagnostics that are not shown indicate that a similar jump in surface pressure also occurred at the same time in the CFSR.

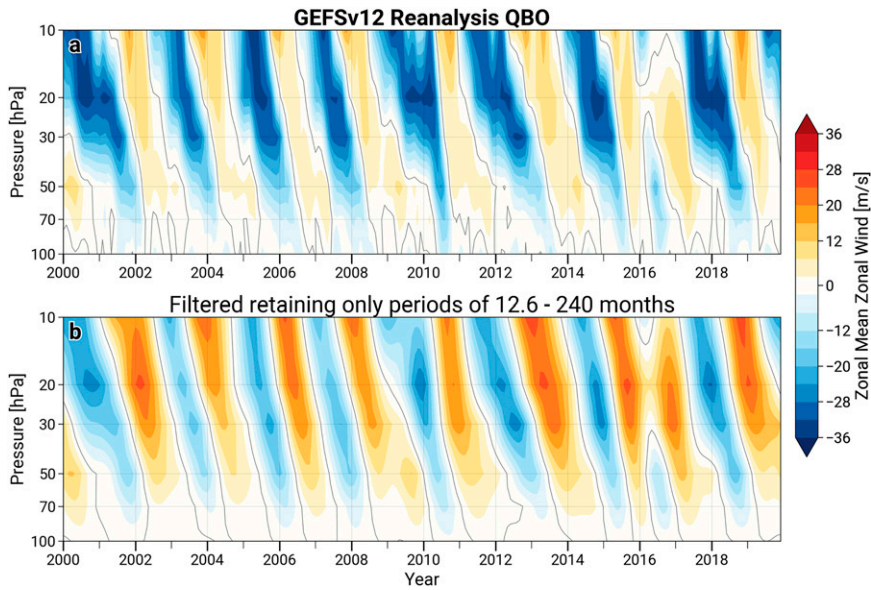


FIG. 8. Illustration of the monthly zonal-mean wind in the 10°S–10°N latitude band, illustrating the QBO in this reanalysis, (a) before and (b) after filtering.

cycles are also plotted. Root-mean-square errors (RMSE) in this reanalysis were generally decreased relative to CFSR throughout the reanalysis period, and biases were somewhat larger than CFSR in the first decade and smaller in the second

decade. Plots for other vertical layers are shown in the online supplemental material; for several tropospheric layers, GEFSv12 biases were somewhat larger than CFSR although RMSEs were generally smaller.

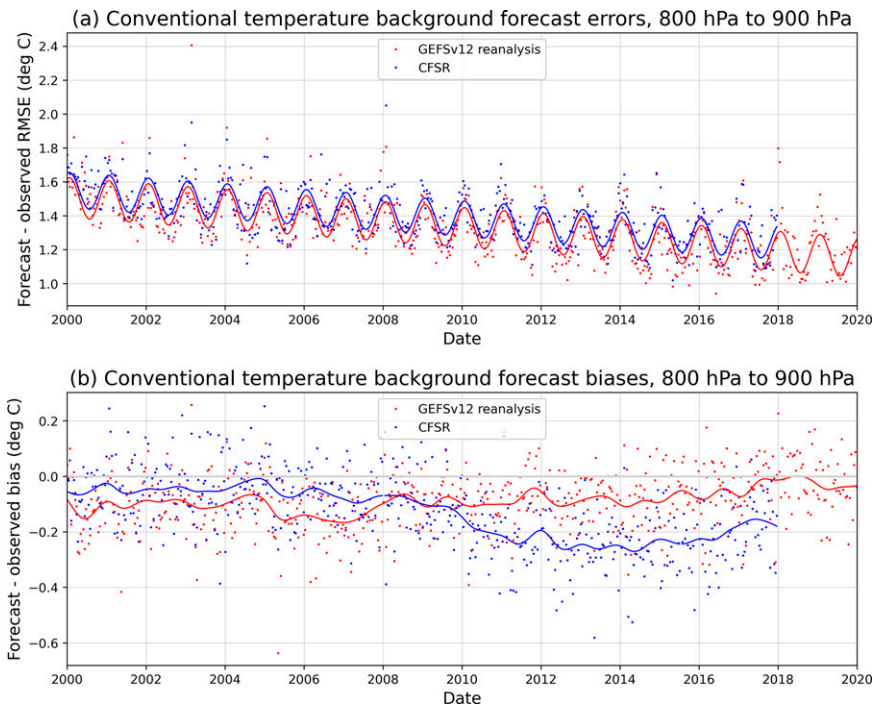


FIG. 9. (a) The 12.5-day average RMSE fit (dots) of control background forecast to conventional data, red for GEFSv12 analysis and blue for CFSR. (b) Averages of control background forecast bias. Colored lines in (a) are estimated with linear regression including an annual cycle, and the lines in (b) are estimated with a Gaussian kernel smoother with a 125-day *e*-folding time scale.

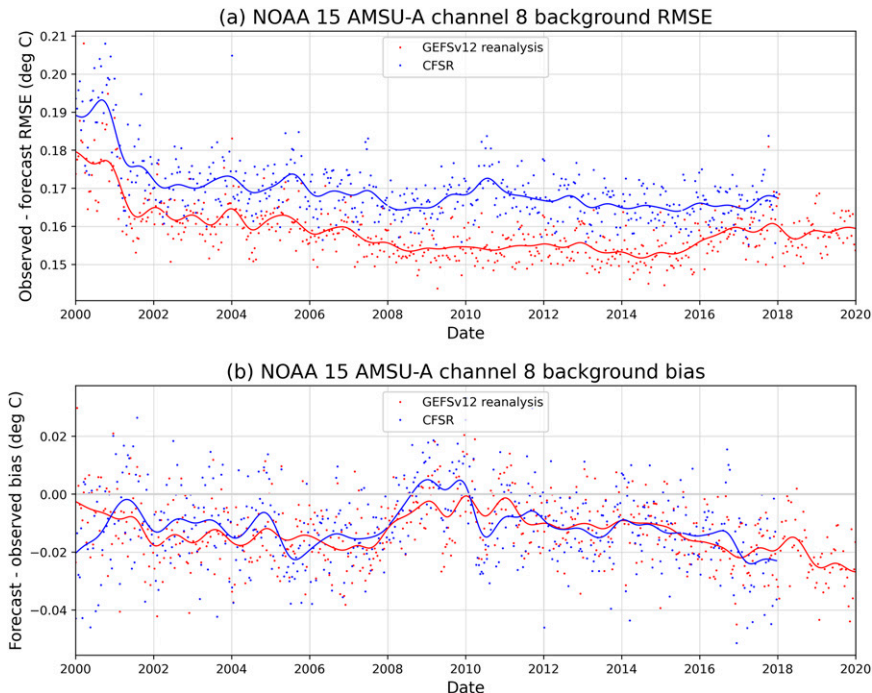


FIG. 10. Two-week average RMSE fit (dots) of control background forecast to the *NOAA-15* polar-orbiting satellite's AMSU-A channel 8 (peak weighting at ~ 150 hPa), red for GEFSv12 analysis and blue for CFSR. Overplotted are Gaussian kernel smoother RMSEs with an e -folding correlation time scale of 125 days. Also shown are (b) background biases.

Similarly, bias-corrected fits of the background forecasts to the *NOAA-15* satellite's Advanced Microwave Sounding Unit A (AMSU-A) channel 8 are shown in Fig. 10; other channels are shown in the online supplemental material. The different reanalyses used different bias-correction methods (Table 1). Channel 8 has a peak of its weighting function around 150 hPa. The RMSE fit to these observations was notably improved relative to CFSR. In most cases, other levels and other microwave channels provided qualitatively similar improvement of RMSE fits, but also generally larger biases in GEFSv12 versus CFSR. A few satellite channels had, at first glance, had much poorer background fits to the observations in the new reanalysis, (see Figs. S11 and S12 of the online supplemental material). However, these channels assimilated cloudy microwave radiances that the CFSR reanalysis did not use. The cloudy radiances typically have much larger differences between background and observed. In this way, the two reanalyses use different sets of observations, and direct comparisons cannot be easily performed.

The ultimate source of somewhat larger bias in many observational fits in GEFSv12 is still unknown. To meet production deadlines, we did begin production of the reanalysis before model physics were fully tuned, and hence it is possible that the (less optimally tuned) GEFSv12 reanalysis system produced background forecasts with larger biases. In the absence of many anchoring observations to help discriminate between background and observation bias, the GEFSv12 assimilation bias may then have been increased relative to CFSR.

c. *Scout-run prediction error characteristics*

Another way to evaluate the quality of the reanalysis was to conduct deterministic predictions initialized from the reanalysis and initialized from the reference standard, the previous generation CFSR. Accordingly, for each day during the reanalysis period, 5-day deterministic GFS predictions were generated from both reanalyses at C384 resolution. The two temperature predictions were then compared with an independent reference standard, in this case the ERA5 reanalysis (Hersbach et al. 2020). Figure 11 synthesizes the results as a function of the forecast lead time (abscissa) and pressure level (ordinate). The top rows show the time-averaged RMSE of the GEFSv12 predictions minus the RMSE from the CFSR-initialized predictions; red colors indicated where the CFSR-initialized forecasts had lower error relative to the ERA5 reference standard, and blue indicated where the GEFSv12 had lower errors. Forecasts from the GEFSv12 temperature reanalyses were generally lower in error, with the notable exception of tropical temperatures just below the tropopause. The GEFSv12 reanalysis used the new FV3 dynamical core as well as a new microphysics parameterization (Zhou et al. 2019). The reanalysis also fully applied humidity observation increments in the stratosphere. In similar testing of the preproduction cycle of GFS data assimilations and forecasts (the retro runs), similar humidity biases were noted, and the eventual operational GFS data assimilation configuration was changed to taper the data assimilation humidity increments in the stratosphere to zero. We assume that the upper-tropospheric and lower stratospheric temperature biases

2000-2019 temperature forecast RMSE difference and biases

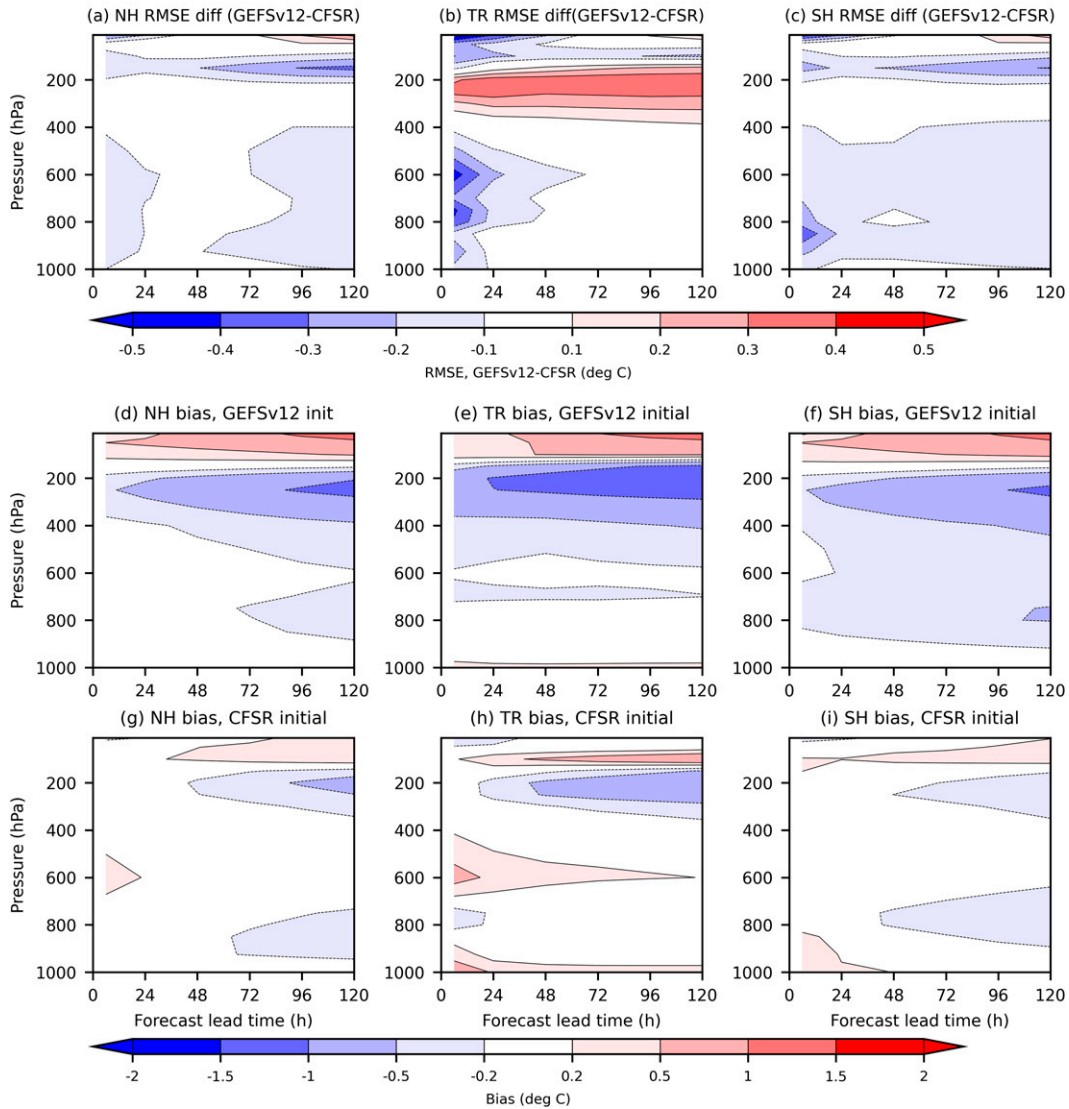


FIG. 11. RMSE and bias characteristics of deterministic FV3-based GFS temperature forecasts initialized from GEFSv12 reanalysis and CFSR initial conditions relative to ERA5 reanalyses. Average differences are shown as a function of forecast lead time (x axis) and pressure level (y axis). RMSE differences (GEFSv12 initialized minus CFSR initialized) are for the (a) Northern Hemisphere (20° – 90° N), (b) tropics (20° S– 20° N), and (c) Southern Hemisphere (90° – 20° S). Biases of GEFSv12 initialized forecasts are for the (d) Northern Hemisphere, (e) tropics, and (f) Southern Hemisphere. Also shown is bias of CFSR initialized forecasts for the (g) Northern Hemisphere, (h) tropics, and (i) Southern Hemisphere.

were related to this incomplete tuning of the reanalysis system prior to production.

The second row of Fig. 11 shows time-averaged prediction biases from the GEFSv12 reanalysis relative to the ERA5 analyses. The third row of Fig. 11 provides similar plots but for predictions initialized from the CFSR. Near-tropopause temperature biases were more pronounced in the forecasts initialized from the GEFSv12 reanalysis. Cold biases near 250 hPa increased with lead time and were especially prominent in the tropics. There were also more pronounced warm

stratospheric biases in predictions from the GEFSv12 reanalysis. While lower in magnitude, the increasing biases with lead time in CFSR-initialized forecasts indicates that the bias likely originated in the version of the FV3-based GFS forecast model used here, not in some aspect of the data assimilation.

Did the apparently increased upper-air temperature bias in the GEFSv12 result in forecasts with degraded accuracy? Figure 12 shows anomaly correlation (Wilks 2011, section 8.6.4) “die-off curves” for GFS deterministic predictions of temperature initialized from the GEFSv12 system (red) versus from

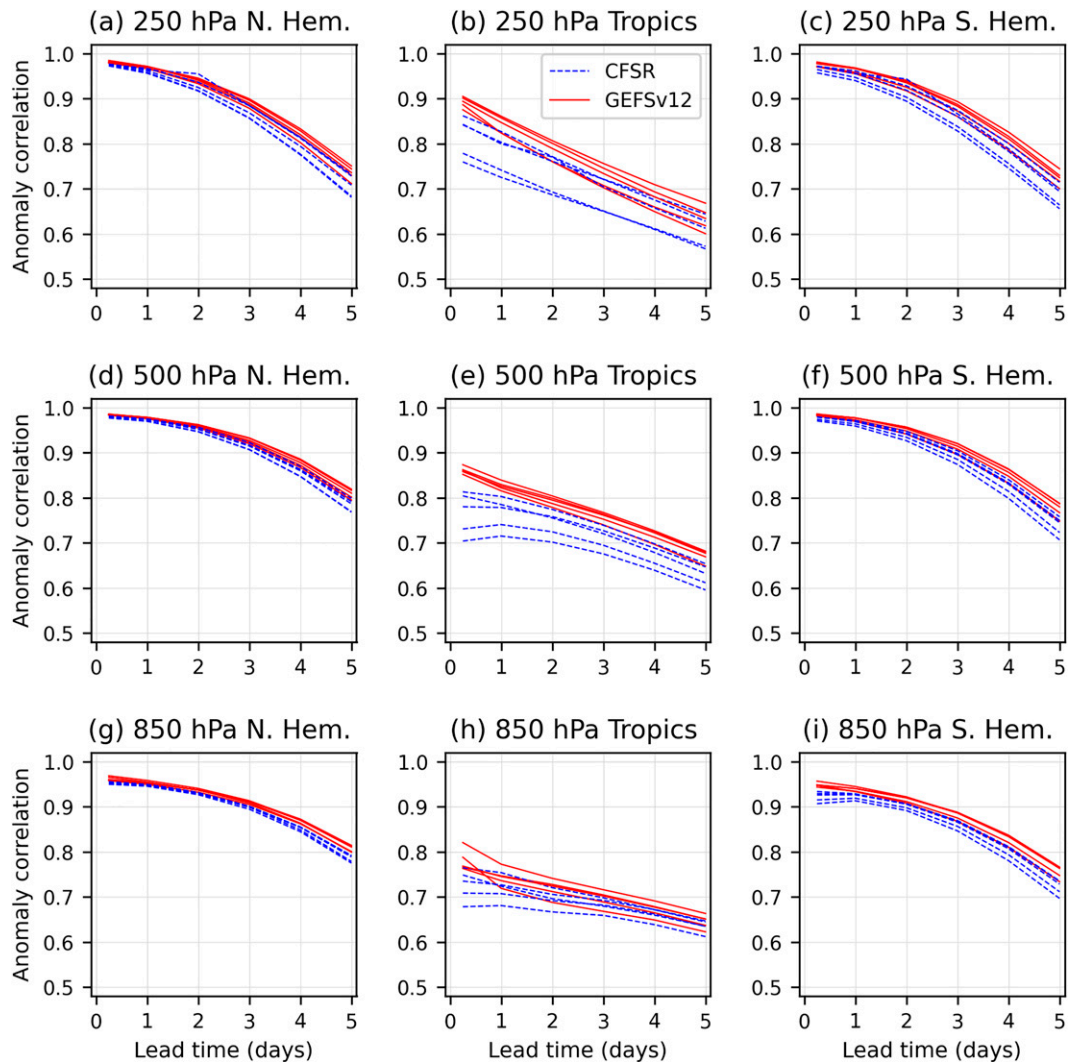


FIG. 12. Anomaly correlation (AC) “die off” curves of temperature computed with respect to ERA5 analyses for vertical levels (a)–(c) 250, (d)–(f) 500, and (g)–(i) 850 hPa for the (left) Northern Hemisphere (20° – 90° N), (center) tropics (20° S– 20° N), and (right) Southern Hemisphere (90° – 20° S). Red lines provide AC averages over each stream with GEFSv12 initialization of the FV3 GFS deterministic forecast, neglecting the first (spinup) year. Blue dashed lines provide averages over each stream with CFSR-initialized FV3 GFS deterministic forecasts.

the CFSR (dashed blue) using ERA5 reanalyses for verification. Five lines are plotted, one each representing the average over one of the five streams. While there was some variability in performance from one stream to the next (generally, the lower anomaly correlations correspond to earlier streams with sparser observational data), consistently across levels and regions, the predictions initialized from the GEFSv12 system were higher in anomaly correlation, especially in the tropics despite the temperature bias. This reinforces the inferences from fits to observations (Figs. 9, 10) that the GEFSv12 reanalysis had reduced errors relative to CFSR.

d. Comparisons with preproduction parallel forecasts

In addition to the multidecadal GEFSv12 reanalyses and the reforecasts generated from them, NOAA partners at the

Environmental Modeling Center also performed approximately two years of retrospective 31-member GEFSv12 forecasts initialized from the operational GFSv15 analyses. These November 2017–October 2019 retro forecast data were primarily used as a comparison with the then-operational GEFSv11 system to ensure that the new forecast system produced forecasts of equal or greater quality.

Perturbed initial conditions for the retro runs, as with the operational GEFSv12, used the background forecast perturbations from the previous analysis cycle. The GEFSv12 configuration was generally the same as that which was made operational, at the target operational C384 resolution. The retro forecasts form a useful baseline for comparison of the reforecasts, which, it was hoped, would be similar in error

characteristics. The cycled retro assimilations and forecasts did not use a strictly unchanging model and data assimilation configuration; as biases were discovered, parameter settings were changed. In this way, it proved difficult to isolate the specific parameter settings that may have been responsible for systematic differences between retro and reforecasts beyond the obvious differences in analysis resolution.

The reforecasts, as described more completely in Guan et al. (2021), used the perturbed forecast states starting from the end of the 9-h IAU window, that is, 3 h into the forecast. The five members were the control and first four from the 80-member ensemble.

Figure 13 illustrates differences in errors between global five-member ensemble-mean reforecasts and the mean of first five members of the retro ensemble forecasts over the ~2-yr period. The top row compares the mean prediction characteristics in the absence of any verifying reanalysis, the mean retro minus mean reforecast. As with the comparison with ERA5 in Figs. 11d–f and 13a indicates that the reanalysis-initialized predictions had mean temperatures that were lower than those from the retro runs near 250 hPa and that reforecast-initialized predictions of the stratosphere were warmer. Reforecast-initialized predictions of humidity above 200 hPa were lower than retro-run predictions. Temperature biases, through geostrophic adjustment, presumably affected the wind structure above this level. Possibly these systematic differences were attributable to differences in the cloud microphysical parameterization used in the cycled reanalysis and/or the application of a tapering of humidity increments in the stratosphere in the retro's cycled data assimilation. These tapering changes had not been made to the reanalysis system at the time of the production. Retro-run cycled assimilations and predictions were performed later, after subsequent tuning of the microphysics reduced biases and application of tapering.

RMS errors of near-surface temperature (Fig. 13g) and humidity (Fig. 13h) were very similar between retro and reforecast, suggesting that the reforecast data should provide acceptable similarity for postprocessing. Near-surface wind errors (Fig. 13i) were slightly larger in the reforecast, but only by a fraction of a meter per second.

A curious feature is shown in Fig. 13i, where u -component RMSE differences in the 200–400-hPa layer grow through day +7 but then get smaller. This can be explained by the differing lead times of forecast-error saturation. In general, reforecast-initialized winds at these levels had forecast errors corresponding to an ~12-h loss in prediction time relative to the retro; that is, reforecast-initialized forecasts at +7-days lead were as accurate as retro-run initialized forecasts at +7.5 days. The global saturation time scale was largely controlled by the (shorter) value in the 30°S–30°N band we call the “tropics” (not shown), given the large fraction of Earth's surface area in this latitude band. Thus, global reforecast prediction errors were more quickly reaching a saturation error value than retro predictions. But the retro predictions eventually began to saturate shortly thereafter as well. Despite different reforecast versus retro errors and biases in upper-air fields, the differences were smaller near Earth's surface,

which was of greater concern for most forecast applications such as postprocessing of sensible weather variables.

Clearly there are some differences in the statistical quality of the multidecadal reforecasts versus the 2017–19 retro runs. The impact of these differences on statistically postprocessed guidance was not evaluated here, for postprocessing is a diverse area of inquiry. Different methods are commonly used for different variables, different lead times (weather versus subseasonal), for instantaneous versus time-averaged quantities. Let us assume that biases are similar between reforecast and real-time guidance but that RMS errors are larger in the reforecast. Then one would expect that training a postprocessing method with reforecasts and then applying to a real-time system with lower errors will result in postprocessed guidance in real time that is higher in quality than were it applied to cross-validated reforecasts, with their larger errors. If biases are dissimilar between reforecast and real time, then the real-time product quality may be degraded because of the statistical inconsistency.

Another way of evaluating the relative quality of forecasts initialized from the reanalysis and from the retro runs is with anomaly correlation die-off curves (Fig. 14), as in Fig. 12. Again, the verifying analyses were provided by ERA5. There were differences in the anomaly correlations of the reforecast-initialized forecasts, which were slightly lower than for the retro runs. Consistent with Fig. 13, these were larger at higher levels and modest at lower levels. As most of the postprocessing is of surface-related variables, this suggests that the impact of differing skills at upper levels is of less importance. An understanding of the practical effect of reforecast skill differences is left for future research.

5. Known discrepancies between the reanalysis and the real-time analysis

Aside from the obvious resolution differences, Table 2 synthesizes what we believe to be the major discrepancies between the configuration of the reanalysis and the real-time forecast configuration.

6. Description of reanalysis data storage

Six-hourly control reanalyses for the 2000–19 period are publicly available on the “emcrzdm” server; 590 variables at 0.25° were created for the synoptic times 0000, 0600, 1200, and 1800 UTC in grib2 format. Because this server also provides operational forecast data, for some times of day, download speeds may be slow. The data can be found here:

7. Conclusions

This article has described the algorithms behind and some of the characteristics of a multidecadal global ensemble atmosphere-only reanalysis covering 2000–19. It was designed for initializing the atmospheric component of reforecasts for the new U.S. National Weather Service Global Ensemble Forecast System, version 12 (GEFSv12), described in a companion

GEFSv12 retro and reforecast differences and RMSE

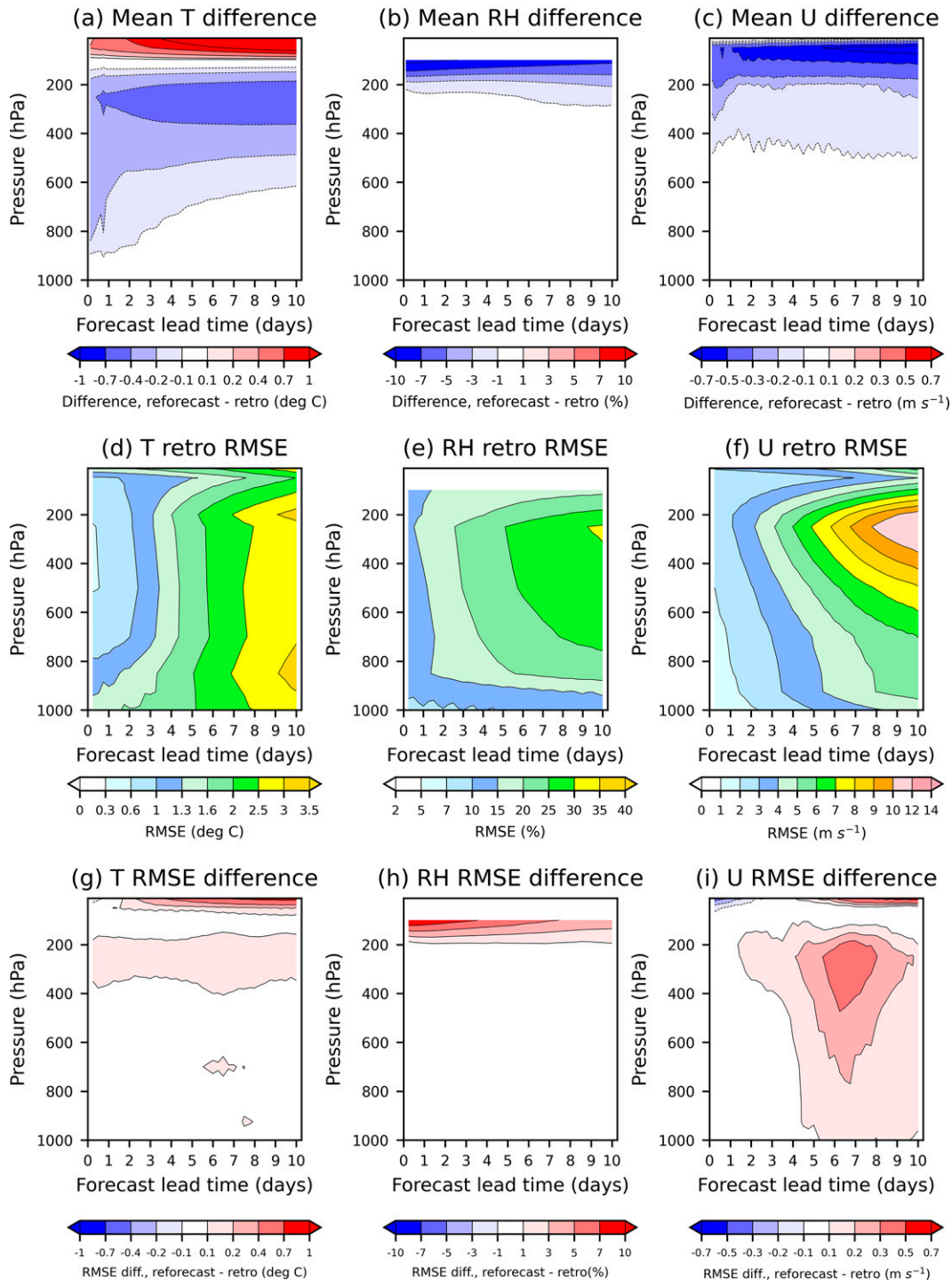


FIG. 13. Comparisons between mean of five-member C384 ensemble predictions initialized from the GEFSv12 reanalysis and the retro analysis for the period December 2017–November 2019. Data are plotted as a function of forecast lead time (abscissa) and vertical pressure level (ordinate); RH data were not available above 100 hPa. Shown are (a)–(c) the global-mean differences in reanalysis-initialized forecasts minus retro-initialized forecasts, (d)–(f) the retro-run prediction global and ensemble-mean RMSE, and (g)–(i) the difference in ensemble-mean RMSE of predictions from the reforecast minus those from the retro for (left) temperature, (center) relative humidity, and (right) u -wind component.

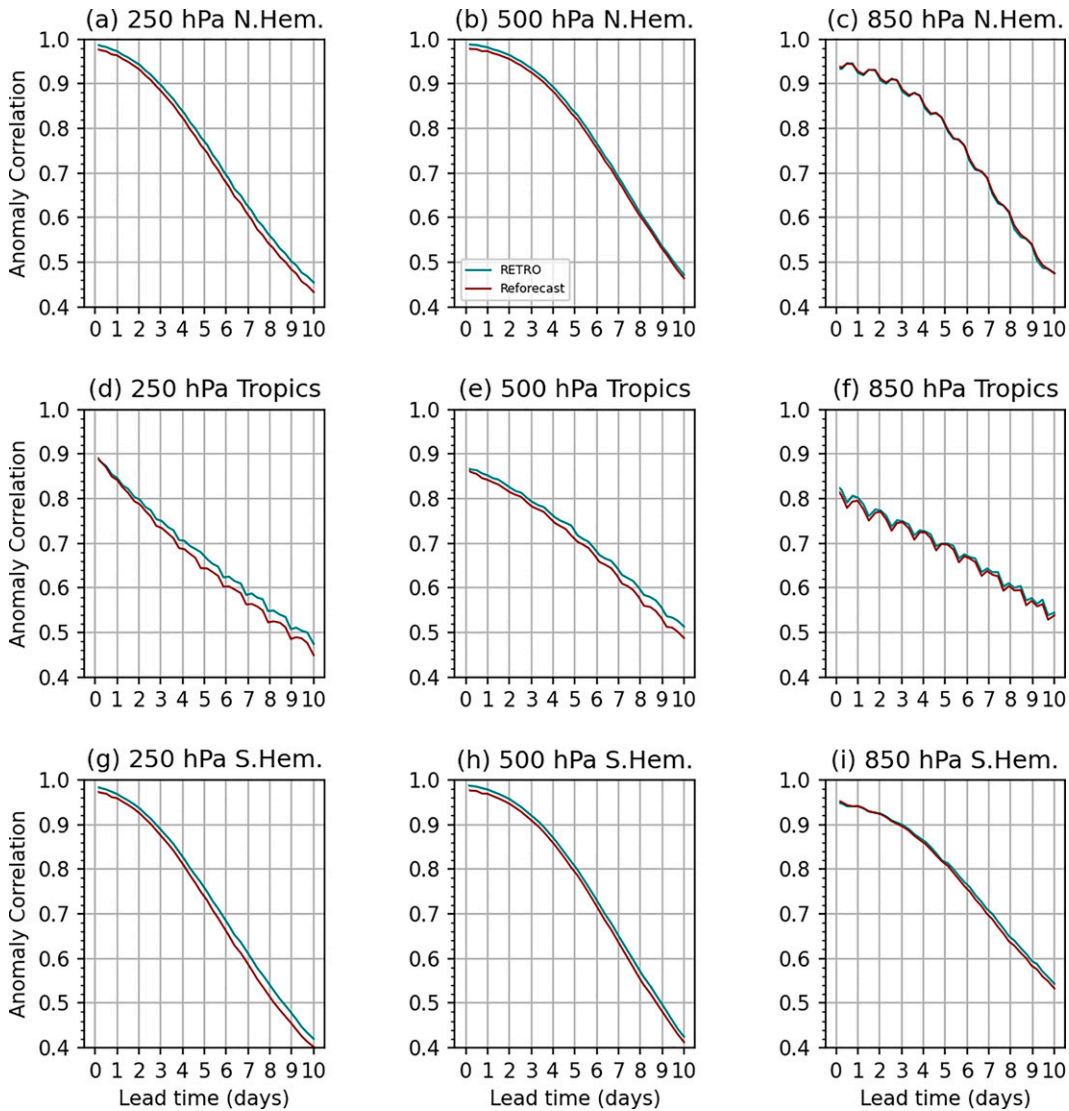


FIG. 14. Anomaly correlation die-off curves with respect to ERA5 analyses for forecast temperature in the (a)–(c) Northern Hemisphere, (d)–(f) tropics, and (g)–(i) Southern Hemisphere for (left) 250, (center) 500, and (right) 850 hPa. Reforecasts are presented in red, and retro simulations are in cyan.

article (Guan et al. 2021). It has approximate consistency with the operational data assimilation procedure used to initialize the real-time GEFSv12 forecasts. Considered over many metrics, the reanalysis quality generally appears to be an improvement relative to the previous-generation CFSR. However, initial-condition temperature bias was often larger in GEFSv12, especially near the tropical tropopause. We hypothesize that this was likely due to larger GEFSv12 background forecast bias as a consequence of starting the reanalysis production before the model was well tuned.

Partly because of computational expense that necessitated lower reanalysis resolution and partly because the reanalysis had to be created before the operational configuration was settled, there were inconsistencies with the real-time analysis. For one, reanalysis resolution was lower.

Also, upper-air thermodynamic variables were different between reanalyses and real-time analyses due to less tuning in the reanalysis with different parameterization configurations, and that affected the upper-air wind analyses as well. There were some near-surface differences, in one case due to a bug in the updating to snow analyses, as well as to different real-time versus reanalysis procedures for land and SST initialization.

We encountered significant challenges in the production of this reanalysis that are motivating some potential changes the next time a NOAA modern-era reanalysis is created. The most significant problem was that reanalysis computations had to be spread over a long period of time due to NOAA’s supercomputers being saturated. Consequently, the reanalysis configuration was decided roughly one year before final

TABLE 2. Consistency issues in the GEFSv12 reanalysis relative to the operational data assimilation scheme, and their consequences.

Issue	Consequence
OI SST in reanalysis vs NSST in operations	Reanalysis ocean initial SSTs are systematically warmer in some extratropical areas, with SST reforecasts inheriting this bias (see Fig. 4a)
Snow analyses only inserted for one cycle of four	Despite replay procedure to ameliorate differences, some land surface skin temperature bias (see Figs. 4b and 4c)
Soil temperature and moisture initialization are not via GLDAS in reanalysis	Freely cycling land surface state in reanalysis procedure can result in initial systematic differences of soil moisture and temperature
Microphysics and parameterization tuning	Biases in upper tropospheric and stratospheric variables in the reanalysis relative to real-time system (see Figs. 9, 11)
U.S. Doppler radar velocity–azimuth display winds turned off	Degraded lower-tropospheric wind analyses prior to their discovery (section 2 of this paper)
Velocity–azimuth display winds are not assimilated in much of the reanalysis because of insufficient quality control	Lower-tropospheric wind observations in reanalysis over United States are somewhat sparser than in real-time analysis

decisions were made on the real-time analysis configuration used for GEFSv12 initialization. Perhaps projects like reanalyses and the accompanying reforecasts are more appropriate for *cloud computing*, where their computations can be delayed until just before a model implementation and then performed on many processors over a short period of time. With the short production period, a consistent system configuration can be used for both reanalysis and the real-time system. The potential downside of this approach is the expense of cloud computing. However, when factoring in all the comparative costs of NOAA-owned versus cloud high-performance computing, the costs may be more comparable, especially if cloud computations are performed with less expensive “spot” instances, computational capacity that is unused by others.

Data preparation and quality control is always a challenge with reanalyses. While we leveraged archived NOAA observational data used in a previous reanalysis, the international sharing of reanalysis observational datasets would have many desirable consequences. With more people involved in data preparation and quality control, each organization would have greater confidence in the fidelity of these data. Also, were standardized observation datasets used by multiple organizations, then reanalysis comparisons will become more straightforward, and differences in quality can be attributed to model and assimilation system design rather than to observation dataset differences.

Inconsistencies between the reanalysis versus real-time analysis also occurred near Earth’s surface. Consistent Global Land Data Assimilation System (GLDAS) reanalysis states were not available at the time of reanalysis production, and the real-time sea surface temperature initialization procedure did not perform acceptably in the more data-sparse environment early in the reanalysis period. Hence, the development of consistent, coupled ocean–atmosphere data assimilation procedures shared between the reanalysis and the real-time system are a priority, if not a panacea for all issues. Similarly, improvement of land data assimilation procedures so that they can be implemented consistently between reanalysis and real-time operations are

also desirable. NOAA is currently working on procedures that improve on GLDAS procedures by using 2-m temperature and humidity observations and other data to make increments to the soil state (Draper 2021). When implemented across applications, this should both improve the consistency and make future reanalyses more widely useful, such as for climate monitoring applications in addition to reforecast initialization.

For brevity, this article did not include extensive comparisons with other reanalyses such as ERA5. Such a comparison is left as a subject for possible future research.

Acknowledgments. Jonathan Brannock of NOAA’s Big Data Project provided assistance in moving data to Amazon Web Services. Zac Lawrence of PSL and CIRES performed the analysis of the quasi-biennial oscillation in this paper. This project was supported by NOAA Weather Program Office Grant T8R1CRS-PCR, NOAA Climate Program Office Grants S8R1CWS-PRE and R8R1CWS-P01, and NWS OSTI Grants P8MWQNG-PTR, P8R1MP1-PPS, and M8M6H81-P10. We thank three anonymous reviewers for their detailed constructive criticism, which greatly improved the paper quality.

REFERENCES

- Anthes, R. A., and Coauthors, 2008: The COSMIC/FORMOSAT-3 mission: Early results. *Bull. Amer. Meteor. Soc.*, **89**, 313–334, <https://doi.org/10.1175/BAMS-89-3-313>.
- Berner, J., G. Shutts, M. Leutbecher, and T. N. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF Ensemble Prediction System. *J. Atmos. Sci.*, **66**, 603–626, <https://doi.org/10.1175/2008JAS2677.1>.
- Bosilovich, M. G., F. R. Robertson, L. Takacs, A. Molod, and D. Mocko, 2017: Atmospheric water balance and variability in the MERRA-2 reanalysis. *J. Climate*, **30**, 1177–1196, <https://doi.org/10.1175/JCLI-D-16-0338.1>.
- Browning, K. A., and R. Wexler, 1968: The determination of kinematic properties of a wind field using Doppler radar. *J. Appl. Meteor.*, **7**, 105–113, [https://doi.org/10.1175/1520-0450\(1968\)007<0105:TDOKPO>2.0.CO;2](https://doi.org/10.1175/1520-0450(1968)007<0105:TDOKPO>2.0.CO;2).

- Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28, <https://doi.org/10.1002/qj.776>.
- Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367–1387, <https://doi.org/10.1002/qj.49712051912>.
- Coy, L., K. Wargan, A. M. Molod, W. R. McCarty, and S. Pawson, 2016: Structure and dynamics of the quasi-biennial oscillation in MERRA-2. *J. Climate*, **29**, 5339–5354, <https://doi.org/10.1175/JCLI-D-15-0809.1>.
- Cucurull, L., 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, <https://doi.org/10.1029/2012JD017782>.
- , R. A. Anthes, and L. Tsao, 2014: Radio occultation observations as anchor observations in numerical weather prediction models and associated reduction of bias corrections in microwave and infrared satellite observations. *J. Atmos. Oceanic Technol.*, **31**, 20–32, <https://doi.org/10.1175/JTECH-D-13-00059.1>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Derber, J., and W.-S. Wu, 1998: The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Mon. Wea. Rev.*, **126**, 2287–2299, [https://doi.org/10.1175/1520-0493\(1998\)126<2287:TUOTCC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<2287:TUOTCC>2.0.CO;2).
- Draper, C., 2021: Accounting for land model uncertainty in numerical weather prediction ensemble systems: Toward ensemble-based coupled land/atmosphere data assimilation. *J. Hydrometeorol.*, **22**, 2089–2104, <https://doi.org/10.1175/JHM-D-21-0016.1>.
- ECMWF, 2019: IFS documentation—Cy46r1 operational implementation 6 June 2019. Part II: Data assimilation. ECMWF Doc., 103 pp., <https://www.ecmwf.int/en/eLibrary/19306-part-ii-data-assimilation>.
- Gelaro, R., R. H. Langland, S. Pellerin, and R. Todling, 2010: The THORPEX Observation Impact Intercomparison Experiment. *Mon. Wea. Rev.*, **138**, 4009–4025, <https://doi.org/10.1175/2010MWR3393.1>.
- , and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>.
- Guan, H., and Coauthors, 2021: GEFSv12 reforecast dataset for supporting subseasonal and hydrometeorological applications. *Mon. Wea. Rev.*, <https://doi.org/10.1175/MWR-D-21-0245.1>, in press.
- Hagedorn, R., T. M. Hamill, and J. S. Whitaker, 2008: Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part I: Two-meter temperature. *Mon. Wea. Rev.*, **136**, 2608–2619, <https://doi.org/10.1175/2007MWR2410.1>.
- Hamill, T. M., 2017: Changes in the systematic errors of global reforecasts due to an evolving data assimilation system. *Mon. Wea. Rev.*, **145**, 2479–2485, <https://doi.org/10.1175/MWR-D-17-0067.1>.
- , and J. S. Whitaker, 2006: Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. *Mon. Wea. Rev.*, **134**, 3209–3229, <https://doi.org/10.1175/MWR3237.1>.
- , —, and X. Wei, 2004: Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.*, **132**, 1434–1447, [https://doi.org/10.1175/1520-0493\(2004\)132<1434:ERIMFS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1434:ERIMFS>2.0.CO;2).
- , —, and S. L. Mullen, 2006: Reforecasts, an important dataset for improving weather predictions. *Bull. Amer. Meteor. Soc.*, **87**, 33–46, <https://doi.org/10.1175/BAMS-87-1-33>.
- , R. Hagedorn, and J. S. Whitaker, 2008: Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part II: Precipitation. *Mon. Wea. Rev.*, **136**, 2620–2632, <https://doi.org/10.1175/2007MWR2411.1>.
- , G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau Jr., Y. Zhu, and W. Lapenta, 2013: NOAA's second-generation global medium-range ensemble reforecast dataset. *Bull. Amer. Meteor. Soc.*, **94**, 1553–1565, <https://doi.org/10.1175/BAMS-D-12-00014.1>.
- Harada, Y., and Coauthors, 2016: The JRA-55 reanalysis: Representation of atmospheric circulation and climate variability. *J. Meteor. Soc. Japan*, **94**, 269–302, <https://doi.org/10.2151/jmsj.2016-015>.
- Helfrich, S. R., D. McNamara, B. H. Ramsay, T. Baldwin, and T. Kasheta, 2007: Enhancements to, and forthcoming developments in the Interactive Multisensor Snow and Ice Mapping System (IMS). *Hydrol. Processes*, **21**, 1576–1586, <https://doi.org/10.1002/hyp.6720>.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–472, [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, S. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP–DOE AMIP-II reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1644, <https://doi.org/10.1175/BAMS-83-11-1631>.
- Kleist, D. T., and K. Ide, 2015: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part II: 4D-EnVar and hybrid variants. *Mon. Wea. Rev.*, **143**, 452–470, <https://doi.org/10.1175/MWR-D-13-00350.1>.
- , D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP Global Data Assimilation System. *Wea. Forecasting*, **24**, 1691–1705, <https://doi.org/10.1175/2009WAF2222201.1>.
- , R. Mahajan, and C. Thomas, 2018: Data assimilation in the Next-Generation Global Prediction System (NGGPS) era: Initial implementation of FV3-based Global Forecast System (GFS). *JCSDA Quarterly*, No. 61, Joint Center for Satellite Data Assimilation, Boulder, CO, 1–9, <https://doi.org/10.25923/jw00-r987>.
- Kobayashi, S., and Coauthors, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, <https://doi.org/10.2151/jmsj.2015-001>.
- Kopp, T. J., and R. B. Kiess, 1996: The Air Force Global Weather Central Snow Analysis Model. Preprints, *15th Conf. on Weather Analysis and Forecasting*, Norfolk, VA, Amer. Meteor. Soc., 220–222.
- Koster, R. D., and Coauthors, 2004: Regions of strong coupling between soil moisture and precipitation. *Science*, **305**, 1138–1140, <https://doi.org/10.1126/science.1100217>.
- Kursinski, E. R., G. A. Hajj, J. T. Schofield, R. P. Linfield, and K. R. Hardy, 1997: Observing Earth's atmosphere with radio occultation measurements using the global positioning system. *J. Geophys. Res.*, **102**, 23 429–23 465, <https://doi.org/10.1029/97JD01569>.

- Lalouaux, P., E. de Boisseson, M. Balmaseda, J.-R. Bidlot, S. Broennimann, and R. Buizza, 2018: CERA-20C: A coupled reanalysis of the twentieth century. *J. Adv. Model. Earth Syst.*, **10**, 1172–1195, <https://doi.org/10.1029/2018MS001273>.
- Lee, W., X. Tang, and B. J. Jou, 2014: Distance velocity–azimuth display (DVAD)—New interpretation and analysis of Doppler velocity. *Mon. Wea. Rev.*, **142**, 573–589, <https://doi.org/10.1175/MWR-D-13-00196.1>.
- Lei, L., and J. S. Whitaker, 2016: A four-dimensional incremental analysis update for the ensemble Kalman filter. *Mon. Wea. Rev.*, **144**, 2605–2621, <https://doi.org/10.1175/MWR-D-15-0246.1>.
- , —, and C. Bishop, 2018: Improving assimilation of radiance observations by implementing model space localization in an ensemble Kalman filter. *J. Adv. Model. Earth Syst.*, **10**, 3221–3232, <https://doi.org/10.1029/2018MS001468>.
- Liu, Q., T. Marchok, H. Pan, M. Bender, and S. Lord, 1999: Improvements in hurricane initialization and forecasting at NCEP with global and regional (GFDL) models. NOAA Tech. Note 472, 7 pp.
- Lin, S.-J., 2004: A “vertically Lagrangian” finite-volume dynamical core for global models. *Mon. Wea. Rev.*, **132**, 2293–2307, [https://doi.org/10.1175/1520-0493\(2004\)132<2293:AVLFDC>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2).
- Meng, J., R. Yang, H. Wei, M. Ek, G. Gayno, P. Xie, and K. Mitchell, 2012: The land surface analysis in the NCEP Climate Forecast System Reanalysis. *J. Hydrometeorol.*, **13**, 1621–1630, <https://doi.org/10.1175/JHM-D-11-090.1>.
- Minnett, P. J., and Coauthors, 2019: Half a century of satellite remote sensing of sea-surface temperature. *Remote Sens. Environ.*, **233**, 111366, <https://doi.org/10.1016/j.rse.2019.111366>.
- Molteni, F., and Coauthors, 2011: The new ECMWF seasonal forecast system (system 4). ECMWF Tech. Memo. 656, 49 pp.
- Orbe, C., L. D. Oman, S. E. Strahan, D. W. Waugh, S. Pawson, L. L. Takacs, and A. M. Molod, 2017: Large-scale atmospheric transport in GEOS replay simulations. *J. Adv. Model. Earth Syst.*, **9**, 2545–2560, <https://doi.org/10.1002/2017MS001053>.
- Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weishaimer, 2009: Stochastic parametrization and model uncertainty. ECMWF Tech. Memo. 598, 42 pp.
- Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center’s spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763, [https://doi.org/10.1175/1520-0493\(1992\)120<1747:TNMCSS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2).
- Pascoe, C. L., L. J. Gray, S. A. Crooks, M. N. Juckes, and M. P. Baldwin, 2005: The quasi-biennial oscillation: Analysis using ERA-40 data. *J. Geophys. Res.*, **110**, D08105, <https://doi.org/10.1029/2004JD004941>.
- Phillips, V. T. J., and L. J. Donner, 2006: Cloud microphysics, radiation and vertical velocities in two- and three-dimensional simulations of deep convection. *Quart. J. Roy. Meteor. Soc.*, **132**, 3011–3033, <https://doi.org/10.1256/qj.05.171>.
- 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>.
- Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143–1170, <https://doi.org/10.1002/qj.49712656415>.
- Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609–1625, [https://doi.org/10.1175/1520-0442\(2002\)015<1609:AHSAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1609:AHSAS>2.0.CO;2).
- Rienecker, M. M., and Coauthors, 2011: MERRA: NASA’s Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, <https://doi.org/10.1175/JCLI-D-11-00015.1>.
- Rodell, M., and Coauthors, 2004: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.*, **85**, 381–394, <https://doi.org/10.1175/BAMS-85-3-381>.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, <https://doi.org/10.1175/2010BAMS3001.1>.
- Scheuerer, M., and T. M. Hamill, 2015: Statistical postprocessing of ensemble precipitation forecasts by fitting censored, shifted gamma distributions. *Mon. Wea. Rev.*, **143**, 4578–4596, <https://doi.org/10.1175/MWR-D-15-0061.1>.
- Shlyayeva, A., and J. S. Whitaker, 2018: Using the linearized observation operator to calculate observation space ensemble perturbations in ensemble filters. *J. Adv. Model. Earth Syst.*, **10**, 1414–1420, <https://doi.org/10.1029/2018MS001309>.
- Shutts, G. J., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102, <https://doi.org/10.1256/qj.04.106>.
- Slivinski, L. C., and Coauthors, 2019: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quart. J. Roy. Meteor. Soc.*, **145**, 2876–2908, <https://doi.org/10.1002/qj.3598>.
- Tompkins, A. M., and J. Berner, 2008: A stochastic convective approach to account for model uncertainty due to unresolved humidity variability. *J. Geophys. Res.*, **113**, D18101, <https://doi.org/10.1029/2007JD009284>.
- Uppala, S. M., and Coauthors, 2005: The ERA-40 re-analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, <https://doi.org/10.1256/qj.04.176>.
- Vannitsem, S., D. S. Wilks, and J. W. Messner, Eds., 2018: *Statistical Postprocessing of Ensemble Forecasts*. Elsevier, 347 pp.
- Wang, X., and T. Lei, 2014: GSI-based four-dimensional ensemble–variational (4DensVar) data assimilation: Formulation and single-resolution experiments with real data for the NCEP Global Forecast System. *Mon. Wea. Rev.*, **142**, 3303–3325, <https://doi.org/10.1175/MWR-D-13-00303.1>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Academic Press, 676 pp.
- Zhao, Q., and F. H. Carr, 1997: A prognostic cloud scheme for operational NWP models. *Mon. Wea. Rev.*, **125**, 1931–1953, [https://doi.org/10.1175/1520-0493\(1997\)125<1931:APCSFO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<1931:APCSFO>2.0.CO;2).
- Zhou, L., S. Lin, J. Chen, L. M. Harris, X. Chen, and S. L. Rees, 2019: Toward convective-scale prediction within the next generation global prediction system. *Bull. Amer. Meteor. Soc.*, **100**, 1225–1243, <https://doi.org/10.1175/BAMS-D-17-0246.1>.
- Zhu, Y., J. Derber, A. Collard, D. Dee, R. Treadon, G. Gayno, and J. A. Jung, 2014: Enhanced radiance bias correction in the National Centers for Environmental Prediction’s grid-point statistical interpolation data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **140**, 1479–1492, <https://doi.org/10.1002/qj.2233>.
- , —, R. J. Purser, B. A. Ballish, and J. Whiting, 2015: Variational correction of aircraft temperature bias in the NCEP’s GSI analysis system. *Mon. Wea. Rev.*, **143**, 3774–3803, <https://doi.org/10.1175/MWR-D-14-00235.1>.

- , and Coauthors, 2016: All-sky microwave radiance assimilation in NCEP's GSI analysis system. *Mon. Wea. Rev.*, **144**, 4709–4735, <https://doi.org/10.1175/MWR-D-15-0445.1>.
- , G. Gayno, R. J. Purser, X. Su, and R. Yang, 2019: Expansion of the all-sky radiance assimilation to ATMS at NCEP. *Mon. Wea. Rev.*, **147**, 2603–2620, <https://doi.org/10.1175/MWR-D-18-0228.1>.