An Evaluation of Precipitation Forecasts from Operational Models and Reanalyses Including Precipitation Variations Associated with MJO Activity

JOHN E. JANOWIAK

Cooperative Institute for Climate and Satellites, Earth Systems Science Interdisciplinary Center, University of Maryland, College Park, College Park, Maryland

PETER BAUER

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

WANQIU WANG

Climate Prediction Center, NOAA/NWS/National Centers for Environmental Prediction, Camp Springs, Maryland

PHILLIP A. ARKIN

Cooperative Institute for Climate and Satellites, Earth Systems Science Interdisciplinary Center, University of Maryland, College Park, College Park, Maryland

JON GOTTSCHALCK

Climate Prediction Center, NOAA/NWS/National Centers for Environmental Prediction, Camp Springs, Maryland

(Manuscript received 29 March 2010, in final form 29 June 2010)

ABSTRACT

In this paper, the results of an examination of precipitation forecasts for 1–30-day leads from global models run at the European Centre for Medium-Range Weather Forecasts (ECMWF) and the National Centers for Environmental Prediction (NCEP) during November 2007–February 2008 are presented. The performance of the model precipitation forecasts are examined in global and regional contexts, and results of a case study of precipitation variations that are associated with a moderate to strong Madden–Julian oscillation (MJO) event are presented. The precipitation forecasts from the ECMWF and NCEP operational prediction models have nearly identical temporal correlation with observed precipitation at forecast leads from 2 to 9 days over the Northern Hemisphere during the cool season, despite the higher resolution of the ECMWF operational model, while the ECMWF operational model forecasts are slightly better in the tropics and the Southern Hemisphere during the warm season. The ECMWF Re-Analysis Interim (ERA-Interim) precipitation forecasts perform only slightly worse than the NCEP operational model, while NCEP’s Climate Forecast System low-resolution coupled model forecasts perform the worst among the four models. In terms of bias, the ECMWF operational model performs the best among the four model forecasts that were examined, particularly with respect to the ITCZ regions in both the Atlantic and Pacific. Local temporal correlations that were computed on daily precipitation totals for day-2 forecasts against observations indicate that the operational models at ECMWF and NCEP perform the best during the 4-month study period, and that all of the models have low to insignificant correlations over land and over much of the tropics. They perform the best in subtropical and extratropical oceanic regions.

Also presented are results that show that striking improvements have been made over the past two decades in the ability of the models to represent precipitation variations that are associated with MJO. The model precipitation forecasts exhibit the ability to characterize the evolution of precipitation variations during a moderate–strong period of MJO conditions for forecast leads as long as 10 days.

Corresponding author address: John E. Janowiak, Cooperative Institute for Climate and Satellites, Earth Systems Science Interdisciplinary Center, University of Maryland, Suite 4001, Room 3031, 5825 University Research Court, College Park, MD 20740-3823.

E-mail: johnj@essic.umd.edu

DOI: 10.1175/2010MWR3436.1

© 2010 American Meteorological Society
1. Introduction

Numerical models evolve constantly due to the ever-present demand for more accurate weather and climate forecasts. Similarly, observational datasets improve with time largely due to advances in observing and communication systems, data storage capacity, and computing power. These circumstances warrant continued validation activities to assess potential model improvements with state-of-the-art observing systems.

Several investigations of model precipitation predictions have been conducted over the past few decades. Janowiak (1992) evaluated the performance of short-range forecasts (12–36 h) of precipitation that were generated by the operational numerical weather prediction models (ca. 1989) at the European Centre for Medium-Range Weather Forecasts (ECMWF) and at the National Centers for Environmental Prediction (NCEP) in the United States. One of the conclusions in that paper was that while the models did a very good job of representing the seasonal and climatological features of tropical rainfall, the temporal variations of these short-range rainfall predictions did a poor job of mimicking observed variations in the tropics. In fact, Janowiak et al. (1998) concluded that the correlation between observed precipitation (based in infrared satellite data) and the NCEP-1 reanalysis (representative of models around 1990) short-range precipitation forecasts were weak to nonexistent over much of the tropical Pacific.

More specifically, Hendon et al. (2000) concluded that the NCEP Medium-Range Forecast (MRF) model [early version of the present-day Global Forecast System (GFS) model] was not able to simulate Madden–Julian oscillation (MJO)-related variability very well. That result was unfortunate because the MJO is the dominant mode of intraseasonal variability in the tropics (Zhang and Hendon 1997) and has been linked to the onset of the Australian monsoon (Hendon and Liebmann 1990) and potentially to El Niño onset (Kessler et al. 1995). Furthermore, Ferranti et al. (1990) documented experiments that showed model forecast skill increased in the extratropics when the model tropical fields were forced to be influenced by observed MJO behavior, and thus demonstrated the potential for increased predictability if MJO variability can be mimicked in the models. In the past decade, Wheeler and Weickmann (2001) found that the NCEP model precipitation forecasts (ca. 1987) clearly depicted MJO behavior for predictions out to 5 days, but showed no semblance of MJO-related variations in precipitation forecasts out to 15 days.

Today, with the advantage of model improvements, advances in data assimilation systems, and ocean–atmosphere coupling in some models, the following question comes to mind: How do present-day model forecasts perform? After all, much better information about oceanic precipitation is available in the present-day compared to what was available when the studies referred to earlier were conducted. More accurate estimates of precipitation over the oceans are available at the time of this writing compared to during the era of the aforementioned studies. Furthermore, validating precipitation analyses are now available globally on daily and finer time scales so that more in-depth validations can be conducted on the model forecasts of precipitation than could be in the past when only monthly or pentad observed datasets were available. Furthermore, unlike most basic-state variables, observed precipitation is not directly assimilated into the models; precipitation is a model prognostic quantity. Therefore, assessments of model precipitation forecasts provide an avenue to assess model skill more independently than one can for most other state variables that are ingested during the assimilation process.

We present our assessment of the performance of present-day ECMWF and NCEP operational weather and climate prediction models in characterizing variations in precipitation, as part of The Observing System Research and Predictability Experiment (THORPEX) special collection in Monthly Weather Review. The primary focus is on the period November 2007–February 2008, which is a period when model forecast data were conveniently available and during which moderate-to-strong MJO activity was observed.

2. Data

Model predictions of precipitation were obtained from the ECMWF and NCEP on a 2.5° latitude–longitude grid and are compared with observations in this study. The “observations” are high spatial and temporal resolution satellite-derived estimates of precipitation that have been accumulated to daily totals and interpolated to 2.5° latitude–longitude grid form to match the model forecasts, along with a coarser heritage dataset that contains rain gauge data (over land only) as well as satellite-derived precipitation estimates.

In addition to operational model forecasts, we examine the precipitation predictions from NCEP’s Climate Forecast System (CFS) and ECMWF’s Re-Analysis Interim (ERA-Interim) to determine whether the intermittent upgrades of the operational models affect the forecast quality relative to the “frozen” model configurations of the CFS and ERA-Interim. The benefit of using reanalysis data rather than operational model output in the context of this paper is that the same model version and analysis system are run over the entire study period so that intermittent operational model upgrades
do not affect the forecast quality. The subsections below provide more detail about the models that are examined in this paper.

a. **NCEP Global Forecast System**

The GFS is an atmosphere-only model and is the primary global weather forecast model that is run operationally at NCEP. The forecasts analyzed in this study were from the system implemented in May 2007. The forecasts were initialized with the NCEP Global Data Assimilation System (GDAS) at the 0000 UTC model cycle. Both the model and initialization system for the GFS operational forecasts have been updated over time to include the latest improvements for the model physics and analysis algorithms and to assimilate new observations to produce the best initial conditions. (Changes to the GFS and GDAS during the past years are documented online at http://www.emc.ncep.noaa.gov/gmb/STATS/html/model_changes.html.) The GFS is integrated at T382 horizontal resolution (~35-km grid) for the first 180 h, after which a T190 resolution (~70-km grid) is used. Sea surface temperatures (SSTs) in the GFS are represented as climatology plus anomalies that decay from the initial observed values at an e-folding time scale of 90 days. The GFS uses a simplified Arakawa–Schubert (SAS) cumulus scheme with a saturated downdraft (Pan and Wu 1995; Grell 1993; Arakawa and Schubert 1974). The SAS scheme employs a condition for the onset of convection that the level of free convection must be within 150 mb from the original level of the air parcel (Hong and Pan 1998).

b. **ECMWF operational model**

The ECMWF operational forecasting system is an atmosphere-only model that comprises a hydrostatic (spectral) global atmospheric and wave modeling system that is initialized with analyses that are produced by a four-dimensional variational assimilation scheme. Since February 2006, the global operational model is run with a wavenumber truncation of 799 (T799) that corresponds to a spatial resolution of 25 km with 91 model levels in the vertical (model top at 0.01 hPa). During the period 2007–08, selected upgrades to the operational system (which are not in the ERA-Interim system) include the following: 1) revised convection and vertical diffusion, a new shortwave radiation scheme, and the inclusion of high-resolution sea surface temperature into the prediction model; 2) the addition of another loop at higher resolution in the minimization producing spatially better resolved increments in the data assimilation procedure; and 3) the addition of GPS radio occultation observations, and an increased number of rain-affected microwave imager radiances. In November 2007, a new model cycle was introduced that included substantial changes to convection and vertical diffusion, which caused the model to become dynamically more active and that resulted in improved representations of midlatitude synoptic activity, tropical Kelvin wave activity, and an increase of the amplitude of intraseasonal variability (Bechtold et al. 2008). The latter is expected to affect the representation of the MJO by the model and has been mainly modified by the improved convection scheme that interacts better with the environmental moisture fields, employs a more realistic convective adjustment time and vertical mass flux (Bechtold et al. 2008). More details of the ECMWF model physics parameterization developments and their impact on the model climate are given by Jung et al. (2010), who also cite the strong increase of rainfall variability introduced by the 2007 convection scheme updates. (Details of model and data assimilation formulation are available online at http://www.ecmwf.int/research/ifsdocs/CY33r1/index.html.)

c. **ECMWF ERA-Interim reanalysis**

Although ERA-Interim is a “reanalysis” system, the model that is used has a forecast capability, and we evaluate the precipitation forecasts from that model. The ERA-Interim uses the version of the ECMWF forecast model that became operational in September 2006 and is atmosphere only. The model physical parameterizations can therefore be expected to produce less precipitation variability than the operational model as of November 2007. ERA-Interim is run at 80-km horizontal resolution and 60 model levels in the vertical (model top at 0.1 hPa). ERA-Interim is considered an interim reanalysis because it was constructed based on the experience of the 40-yr ECMWF Re-Analysis (ERA-40; Uppala et al. 2005) and represents the preparation for a longer reanalysis that will extend back to the past 75–100 yr. The chosen period (1989–present) has permitted the use of a comparatively stable space-based observing constellation that provides radiances from infrared and passive microwave instruments together with derived near-surface scatterometer winds, cloud motion winds, and total column ozone retrievals.

With respect to ERA-40, numerous fundamental improvements to model, data assimilation, and observing system have been introduced. Apart from model resolution and vertical level upgrades, a 12-h four-dimensional variational assimilation system is operated that produces two analyses per day, and which initializes two 10-day forecasts per day. Largely improved background error statistics, humidity analysis, and model physics have been implemented since ERA-40. With respect to precipitation forecasts, the combination of better moist physics parameterization and humidity analysis has led to smaller humidity increments and reduced precipitation “spin down”
(i.e., a decline in global mean precipitation that persists after model initialization until the model reaches equilibrium) in ERA-Interim compared to ERA-40 (Andersson et al. 2007). ERA-Interim also has a more extensive use of satellite radiance data employing a data quality control system that draws on experience from ERA-40 and the 25-yr Japanese Re-Analysis (JRA-25) as well as a variational bias correction of satellite radiance data (Dec 2004). In particular the latter is expected to produce more consistent observational impact when, over longer time periods, new observation types are being implemented, instruments fail or are being replaced, and in the presence of radiative transfer model and interinstrument calibration biases. More details on ERA-Interim are available from Uppala et al. (2008).

d. NCEP Climate Forecast System

The CFS is a coupled atmosphere–ocean model. The atmospheric component of the CFS is the 2003 version of the NCEP operational atmospheric GFS model with a spectral truncation of 62 waves (T62) in the horizontal (approximately equivalent to a 200-km grid) and a finite differencing in the vertical with 64 sigma layers. The oceanic component of the CFS is the National Oceanic and Atmospheric Administration (NOAA) Geophysical Fluid Dynamics Laboratory Modular Ocean Model version 3 (MOM3; Pacanowski and Griffies 1998) with a zonal resolution of 1° and a meridional resolution of 1/3° between 10°S and 10°N, gradually increasing through the tropics until becoming fixed at 1° poleward of 30°S and 30°N. The vertical resolution of the oceanic component is 10 m from the surface to the 240-m depth, gradually increasing to about 511 m in the bottom layer. The atmospheric and oceanic components are coupled without any flux adjustment. The two components exchange daily averaged quantities once each day. Initial conditions for the CFS are taken from the NCEP Department of Energy (DOE) Reanalysis-2 (R2; Kanamitsu et al. 2002) for the atmosphere and from the NCEP Global Ocean Data Assimilation System (GODAS) for the ocean. More details can be found in Wang et al. (2005) and Saha et al. (2006). The GFS uses a simplified Arakawa–Schubert cumulus scheme with a saturated downdraft (Pan and Wu 1995; Grell 1993; Arakawa and Schubert 1974). The SAS scheme employs a condition for the onset of convection that the level of free convection must be within 150 mb from the original level of the air parcel (Hong and Pan 1998).

The CFS produces four forecast runs each day from the same oceanic initial condition but slightly different atmospheric initial conditions. The atmospheric initial condition for one forecast run is taken as the R2 0000 UTC analysis and that for the other three runs is from the same R2 initial state with small perturbations which are 3.3%, 6.7%, and 10% of the differences from the state 24 h earlier. In our analysis, an ensemble average of the four daily forecast runs is used for each initial day.

e. Observed precipitation data

One of the most widely used and familiar global precipitation datasets is the Global Precipitation Climatology Project analyses (GPCP; Huffman et al. 1997). GPCP has the desirable quality of rain gauge data incorporation that reduces bias in the satellite-derived precipitation estimates that are also used in those analyses. However, analyses are available only for pentad (5 day) and monthly accumulations, so these data are not ideal for examining features such as the evolution of MJO. The GPCP analyses use passive microwave data from a single platform (for climate stability concerns); thus, the accuracy (from a simple diurnal sampling point of view) of the precipitation estimates over oceanic regions is less than that of more recently developed high-resolution precipitation analyses such as the Climate Prediction Center (CPC) Morphing procedure (CMORPH; Joyce et al. 2004) or the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA; Huffman et al. 2007), among others, which incorporate passive microwave data from multiple sources. The GPCP does produce a daily product, but it is based on one to three passive microwave sensors (depending upon the observation date) and IR data and does not contain rain gauge data, although it is adjusted so that the monthly accumulations of the daily analyses equal the monthly GPCP product. Despite these shortcomings, we use GPCP data for global mean comparisons among the various datasets because of the wide use of that dataset and to provide perspective.

Over the past decade, several global (in longitude), high-resolution, satellite-derived precipitation datasets have become available that contain high-quality (but relatively sparse) passive microwave satellite data that are merged with relatively poor quality (but abundant) infrared data. The use of IR data limits the geographical coverage of these datasets to equatorward of about 60° latitude, and the period of record of these datasets is relatively short (generally only back to about 1998) because of the paucity of passive microwave data until the late 1990s. However, the fact that these analyses have high temporal resolution (½ hourly to 3 hourly) and use a relative abundance of passive microwave information (nine sensors during the period of this study) makes them attractive for use in studies in which precipitation variations over the course of hours to weeks and beyond are examined.

One such dataset is CMORPH, which is an analysis technique that blends together high-quality rainfall estimates from passive microwave sensors. Because those
sensors are housed on low earth orbit spacecraft, the estimates are neither spatially nor temporally complete. To remedy this problem, CMORPH uses IR data from geosynchronous satellites to integrate the precipitation features that have been identified by passive microwave data in time and space to produce spatially and temporally complete analyses of precipitation. The CMORPH estimates are produced at relatively high spatial (0.07° latitude–longitude) and temporal (30 min) resolution. For this study, the analyses were accumulated to daily totals and then interpolated to 2.5° latitude–longitude to match the model forecast data.

3. Large-scale evaluation

a. Global mean (60°N–60°S)

Time series of “global” mean precipitation during the November 2007–February 2008 period is shown in Fig. 1. Note that the global means are computed over the latitude range of 60°N–60°S because that is the latitudinal extent of CMORPH. The left panel shows time series of global mean daily precipitation for CMORPH, ERA-Interim, and CFS. Global mean time series of 2-day precipitation forecasts from ERA-Interim and CFS only are shown for clarity of the figure and were chosen because the global mean daily time series of ERA-Interim has the highest correlation (i.e., 0.69) with CMORPH and CFS has the lowest (i.e., 0.40). A summary of the global means for CMORPH and for each model appear in Table 1 along with the correlations among their time series.

The global mean of the models is about 3.3 mm day$^{-1}$ while the GPCP and CMORPH global means are considerably lower at 2.86 and 2.47 mm day$^{-1}$, respectively. Despite the large difference in the mean precipitation, the daily variability compares reasonably well. And although the correlations between CMORPH and the model forecasts are modest, the similarity among the time series is visually compelling.

The global mean precipitation averaged over the 120-day period for each model is plotted as a function of forecast lead in the right panel of Fig. 1. The range of mean precipitation among the four models and forecast leads is from about 3.24 to about 3.60 mm day$^{-1}$. The CFS global mean precipitation increases steadily, with forecast lead ranging from 3.28 mm day$^{-1}$ at 2-day forecast lead to 3.60 mm day$^{-1}$ at the 9-day forecast lead. The ECMWF global mean precipitation also increases with forecast lead but much less than CFS. In contrast, the GFS global mean precipitation decreases from the 2-day forecasts to day 5, then levels off and rises sharply between the day-8 and day-9 forecasts. The ERA-Interim global mean precipitation is quite constant near 3.22 mm day$^{-1}$ for all forecast leads.

b. Regional bias

The mean bias patterns among the two-day model forecasts are strikingly similar to both CMORPH (Fig. 2, Table 1. Global (60°N–60°S) mean precipitation (mm day$^{-1}$) and correlations among daily CMORPH and the model precipitation 2-day precipitation forecasts during 1 Nov 2007–28 Feb 2008. (Note: GPCP data are available only in pentad and monthly mean form, so correlations with daily data are not presented.)

<table>
<thead>
<tr>
<th></th>
<th>Global mean</th>
<th>ECMWF</th>
<th>ERA-Interim</th>
<th>GFS</th>
<th>CFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMORPH</td>
<td>2.47</td>
<td>0.66</td>
<td>0.69</td>
<td>0.53</td>
<td>0.40</td>
</tr>
<tr>
<td>ECMWF</td>
<td>3.24</td>
<td>—</td>
<td>0.74</td>
<td>0.74</td>
<td>0.56</td>
</tr>
<tr>
<td>ERA-Interim</td>
<td>3.24</td>
<td>—</td>
<td>0.70</td>
<td>0.70</td>
<td>0.52</td>
</tr>
<tr>
<td>GFS</td>
<td>3.35</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.66</td>
</tr>
<tr>
<td>CFS</td>
<td>3.28</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>GPCP</td>
<td>2.86</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

FIG. 1. (left) Global (60°N–60°S) daily mean precipitation for CMORPH, ERA-Interim, and CFS (2-day forecasts for the ERA-Interim and CFS). (right) Global (60°N–60°S) mean precipitation as a function of forecast lead averaged over November 2007–February 2008 for the various model forecasts (see legends). The global (60°N–60°S) mean for GPCP (version 2.1) is 2.86 mm day$^{-1}$ over the period.
FIG. 2. Mean bias of the model 2-day forecasts relative to (top) CMORPH and (middle) GPCP. The bias was computed from the simple difference at each grid location between the November 2007 and February 2008 mean precipitation of each model minus the observed mean over the same period. (bottom) The 4-month mean difference between the day-9 and day-2 forecasts. Units are mm day$^{-1}$. 
top) and GPCP (Fig. 2, middle). All four models have a positive bias over the storm track regions in the northwestern Atlantic and Pacific Oceans. This result is opposite of what Janowiak et al. (1998) discovered in an evaluation of the NCEP-1 reanalysis precipitation in which GPCP data were used as observations. Note that positive biases are also observed over most of Europe and over eastern and western North America with respect to CMORPH, but the bias over those regions is reduced substantially relative to GPCP. We place more confidence in GPCP over these regions since rain gauge data are included in the GPCP analyses. Similarly, all of the model forecasts have biases of 1–3 mm day$^{-1}$ in the zonal band between 50$^\circ$ and 60$^\circ$S latitude relative to CMORPH but again are near zero relative to GPCP. We have more trust in the CMORPH data in this region because the region is mostly ocean and CMORPH has much passive microwave information here than do the GPCP analyses. Bias is less in the Southern Hemisphere compared to the Northern Hemisphere, although that may because the study period occurs during the warm season in the Southern Hemisphere when the amplitude of energetic synoptic-scale systems is less than during the cold season.

The ECMWF forecasts have the least bias among the models over the tropical Atlantic and Pacific Oceans and over the ITCZ regions in particular, where the others have more intense rainfall in the central and eastern ITCZs in both oceans compared to observations. The biases are similar among all the models over Africa and Australia, but large differences are noted over northern and central South America where the bias is primarily positive in the ECMWF and GFS, while a preponderance of negative bias is seen there in the CFS and ERA-Interim forecasts. The implication from Fig. 1 (right panel) is that positive bias increases with forecast lead in the CFS and ECMWF (to a lesser degree) forecasts of precipitation, while the GFS forecasts produce slightly less precipitation with increasing forecast lead but ramps up sharply between the day-8 and day-9 forecasts, and the global mean ERA-Interim precipitation forecasts do not vary much with forecast lead. While this spatially aggregated bias information is interesting and useful, a more informative picture emerges when the spatial distribution of the time-mean differences between the day-9 and day-2 forecasts are plotted. In particular, the regions with bias in the day-2 forecasts do not necessarily maintain those biases in the longer lead forecasts and, in some cases, the bias is reduced (locally) in the longer lead forecasts. For example, the difference between the 9-day and 2-day CFS forecasts (Fig. 2, bottom) indicates that the bias is reduced locally in the longer lead forecasts (i.e., the sign of the differences in the bottom panel is opposite of the sign of the bias in the panels above) over the equatorial Atlantic east of Brazil, the eastern and central Pacific ITCZ, the South Pacific convergence zone (SPCZ), the central equatorial Indian Ocean, and central Brazil southeastward including the South Atlantic convergence zone (SACZ). In contrast, the longer lead CFS forecasts increase the biases that are observed in the day-2 forecasts over most of the Atlantic ITCZ, the western tropical and north-central Pacific, and over portions of equatorial and southern Africa.

Similar behavior is seen in the other models. The 9-day ECMWF forecasts reduce the bias observed in the day-2 forecasts over much of central South America and southern equatorial Africa, while increased bias is noted in the extreme western tropical Pacific, and over the northern fringe of the SPCZ. And while the area-averaged ERA-Interim precipitation amount is rather constant for all forecast leads, this is the result of drier forecasts over much of the Southern Hemisphere that are compensated by generally wetter forecasts in the Northern Hemisphere in the day-9 forecasts compared to day 2.

c. Temporal correlation

Correlations were performed between CMORPH and the 2-day forecasts at each grid point on daily data over the November 2007–February 2008 period for each model separately. The patterns and magnitudes of these local temporal correlations (Fig. 3) indicate that the models more strongly mimic the observed temporal precipitation variations over the subtropical and extratropical oceans compared to over land and in the tropics. The weak correlations over the Northern Hemisphere land surfaces may not be meaningful since the period of study (November–February) is the cool season in that hemisphere when the performance of satellite-derived precipitation estimates are relatively poor over extratropical continental regions. However, the correlations are weak over the warm season Southern Hemisphere and tropical land regions as well. Over the subtropical and extratropical oceans, the ECMWF and GFS model precipitation forecasts depict the temporal variability of the observed precipitation best among the models. It should also be noted that correlations are very weak over the equatorial Pacific, including the Pacific ITCZ, in all of the models and in the CFS in particular. This feature has been noted previously by Janowiak et al. (1998) with regard to the NCEP-1 reanalysis monthly precipitation anomalies.

As expected, the area-averaged local temporal correlations decrease with increasing forecast lead (Fig. 4). However, the differences among the model performance with regard to this statistic are enlightening. The ECMWF and GFS forecast correlations are clearly higher out to about 1 week than the others, and the coupled CFS model forecasts are the poorest of the four models. For the tropical and Southern Hemisphere zonal regions, the
ECMWF forecasts have higher average spatial correlation at each forecast lead, with GFS performing next best, followed by ERA-Interim, and CFS. Over the 30°–60°N zonal band, the averaged correlation for ECMWF and GFS is practically identical at all forecast leads. Because it is the cool season for that band during this November 2007–February 2008 study period, the implication is that the variations in the ECMWF and GFS daily precipitation forecasts are quite similar during extratropical winter when precipitation is primarily stratiform and synoptic scale in nature. The increased performance of the ECMWF in the other two zonal bands suggests that the ECMWF model handles the variations in convectively driven precipitation better than the other models, including the GFS, which may result from the recent improvements in the convective parameterization of the ECMWF model (see section 2b).

d. The distribution of precipitation rates (60°N–60°S)

There are a number of ways to characterize the distribution of precipitation rates among the models and observations, and we chose to do so as follows. For each model separately, we ordered the 2-day daily forecast nonzero precipitation amounts into a sorted ascending list and then computed the running sum of those individual daily values. Plots of the running totals (Fig. 5) provide the contribution of the total precipitation (i.e., the area-weighted sum of all daily amounts at all grid locations) in terms of the probability distribution of the daily precipitation amounts. The same procedure was applied to the daily CMORPH totals. Based on the plot in Fig. 5, it is apparent that the contribution to total areal and temporal accumulated precipitation is provided more by lighter amounts in the models compared to CMORPH. For example, about 40% of the total accumulated CMORPH precipitation (all grid points over the 120-day period) is contributed by daily amounts that are in the 95th percentile or higher compared to about 28%–35% for the GFS, CFS, ERA-Interim, and the ECMWF forecasts. Possible reasons for the ECMWF model forecast distribution mimicking CMORPH more closely than the other models may be that the ECMWF forecasts are calculated at higher spatial resolution (≈25 km) than the other models and because of the recent improvements to the convective parameterization.

This process was conducted for the 9-day forecasts for each model and the results are almost identical (not shown); thus, we conclude that the distribution of the intensity of daily totals is not a function of forecast lead time. This process was also performed separately for 30°–60°N, 30°N–30°S, and 30°–60°S, with essentially the same result.

4. Madden–Julian oscillation case study

a. Background

The Madden–Julian oscillation (MJO) phenomenon (Madden and Julian 1994) is a tropical feature that is
characterized by an eastward progression of alternating regions of enhanced and suppressed tropical convection that are observed over the Indian and Pacific Oceans primarily, although they can often be traced across the entire global tropics. The oscillating periodicity of these phenomena is approximately 30–60 days. Although the rainfall patterns associated with MJO often become disorganized and diffuse as they propagate over the relatively cool ocean waters of the eastern Pacific, variations in the nonrotational part of the atmospheric wind aloft, as represented in the velocity potential field, clearly depict the MJO propagation.

Despite the tropical origins of the MJO, the phenomenon can have impacts in the midlatitudes as well as in the tropics. For example, a number of studies have reported on tantalizing relationships between MJO and ENSO (Lau et al. 1989; Weickmann 1991; McPhaden 1999; Kessler and Kleeman 2000; Zhang and Gottschalck 2002). MJO events can contribute to very heavy precipitation over the Pacific Coast of the United States during wintertime (Higgins et al. 2000) via “atmospheric rivers” (Ralph et al. 2006), also known colloquially as “Pineapple Express” events, in which elongated bands of tropical moisture stream from the Hawaiian Island region toward the U.S. West Coast and result in flooding situations that may last a week or more. Furthermore, MJO can modulate tropical cyclone development (Liebmann et al. 1994; Maloney and Hartmann 2000; Hall et al. 2001; Bessafi and Wheeler 2006).

**b. Strategy to assess model performance**

The 4-month period examined in this paper (November 2007–February 2008) was chosen mainly because model forecast data were readily available for that period and because we knew, a priori, that it was an active MJO period. That period also coincided with moderate to strong La Niña conditions in the equatorial Pacific. And although we present results from just one period, we have performed cursory examinations on other MJO events of varying intensity and the same general behavior was observed, although that conclusion is based on the performance of just the GFS model. The results that we present in the following section are both qualitative and quantitative.

To compare the model performance in forecasting MJO activity, we use subjective visual interpretations of Hovmöller diagrams. Although subjective, we feel that the plots that we present are visually compelling and that our discussion of the observed and model behavior is presented in a broad fashion and without emphasizing subtleties that are not defensible given the subjectivity of the manner in which the model performance is evaluated. We attempt to quantify the comparisons by performing pattern correlations on the Hovmöller diagrams.
themselves, which we feel bolsters our qualitative assessments.

c. Model forecast performance

MJO activity is apparent in a Hovmöller diagram (Fig. 6) of CMORPH by the oscillatory behavior of the precipitation intensity over the period as well as the eastward propagation of the enhanced precipitation. Three distinct relatively wet events are highlighted by solid lines in Fig. 6, with the earliest event occurring near 80°E during early November and propagating eastward before subsiding near 160°W by the end of the month. The second event started near 70°E in early December 2007; its eastward propagation slowed late in the month until it ended near the date line in late-January 2008. A third event began near 70°E in mid-January 2008 and became a stationary feature in early February 2008. Sandwiched between these enhanced wet periods were two distinct relatively dry periods that range from the western Indian Ocean to near the date line from late November to late December and from early January 2008 through February 2008. Embedded within the envelope of eastward-propagating precipitation are westward-propagating elements.

Although not shown, the models depict this period of MJO activity extremely well in the 1-day forecasts of precipitation, and the forecasts continue to portray the MJO features quite well in amplitude and reasonably well in phase out to about 5 days compared to CMORPH (Fig. 7). Focusing on the strong episode in the middle of the plot (spanning 1 December–1 February), the ECMWF and GFS 5-day forecasts, in particular, portray the evolution of the events extremely well, while the CFS forecasts are less clear and have slower propagation speeds than the observations and the other models. Note that the solid lines that are drawn on the model forecast plots are from the observed (CMORPH) data so that differences among the model forecasts and the observations can be discerned easily. In addition, westward-moving elements within the larger-scale eastward-propagating envelope of anomalously strong convection are depicted fairly well by the models. All of the models predict the change from eastward propagation to a relatively stationary region of enhanced precipitation for the event that began in mid-January. The relatively dry period that is sandwiched between the first two enhanced wet periods (late November–late December) is more moist in the 5-day forecasts of all models compared to the CMORPH analyses (Fig. 6).

The enhanced wet periods depicted in the 10-day precipitation forecasts (Fig. 8) from the three models characterize the amplitude and pattern of the MJO, although the timing of the propagation appears to be a bit slower in the ECMWF and GFS forecasts compared to CMORPH. The westward-moving elements that are embedded within the eastward-propagating envelope are less apparent in the 10-day precipitation predictions than at shorter forecast leads. The changes introduced in the ECMWF model in November 2007 clearly enabled better MJO pattern propagation that had not been clearly identifiable with previous models at these forecast ranges (Bechtold et al. 2008).

Little difference in phase or magnitude is observed between the 10- (Fig. 8) and 15-day GFS and CFS forecasts (Fig. 9) while little semblance of MJO evolution is apparent in the 30-day CFS forecasts. In addition, the westward movement of elements embedded within the eastward-propagating envelope is essentially absent in the 15- and 30-day forecasts.

Very similar behavior is exhibited by the ERA-Interim precipitation forecasts (not shown). One notable difference in the ERA-Interim compared to the other models is that the precipitation intensity during the wet periods is reduced comparatively. This is also reflected in the convective intensity pattern comparisons between the ERA-Interim model cycle 31r1 and the operational model cycle (32r3; ca. November 2007) in reproducing the active phase of the MJO during December 1992–February 1993 (Bechtold et al. 2008, their Fig. 15, from seasonal model 15-day forecasts). In contrast, the
precipitation during the relatively dry periods is notably higher in the ERA-Interim forecasts compared to CMORPH and the other models at all forecast leads.

As a means to reduce the subjectivity that is inherent when making visual comparisons of Hovmöller diagrams, we computed the pattern correlation of the Hovmöller diagrams for CMORPH and for each model for 1–9-day forecast leads (Fig. 10). Several observations are clear from that figure. First, from the day-2 through day-9 forecasts, the ECMWF performs the best, followed in order by ERA-Interim, GFS, and CFS. Second, the CFS forecasts perform considerably worse than the other models. Finally, the GFS, ECMWF, and ERA-Interim day-1 forecasts have nearly identical pattern correlations with the CMORPH Hovmöller diagram.

While Hovmöller diagrams are excellent for depicting propagating features, spatial detail is degraded due to the latitudinal averaging that is inherent in them. Therefore, we show some time-averaged spatial depictions of this MJO activity in Fig. 11, which shows the evolution of precipitation “anomalies” (i.e., the difference from the November 2007–February 2008 average) for three separate periods for CMORPH and for the 10-day forecasts from the ECMWF and GFS models. These three periods show the progression of enhanced precipitation that is associated with an MJO event that began in the equatorial Indian Ocean during early December, was observed over the Maritime Continent from mid-December to early January, and moved to the western flank of the SPCZ during the first half of January. Note that the solid line on Fig. 11 depicts the approximate progression of the CMORPH positive precipitation anomaly with time.

During the formative stage of the event in the area south of the Bay of Bengal (Fig. 11, top row), for both the ECMWF and particularly the GFS, the 10-day precipitation model forecasts have less areal coverage of precipitation than CMORPH in that region. Also, while the pattern of anomalous precipitation straddles the equator in CMORPH, the model anomalies are mostly north of the equator during the 4–15 December period. Also, CMORPH and the GFS depict two separate anomalies—one in the southern Bay of Bengal and one in the southern Arabian Sea—whereas the ECMWF forecasts depict a contiguous band of positive anomalies that is centered south of the Indian peninsula.

The position of the precipitation anomalies during the 16 December–3 January period (Fig. 11, middle row) indicates that the ECMWF anomalies are farther west than CMORPH although the latitude of the enhanced precipitation is in good agreement with CMORPH. Meanwhile, the GFS 10-day forecasts have the center of enhanced precipitation even farther west than the ECMWF forecasts and farther north than the
Fig. 7. As in Fig. 6, but for 5-day forecasts of precipitation from the (left) ECMWF, (middle) GFS, and (right) CFS models. Note that for ease of comparison the lines on the plots indicate the propagation relative to the CMORPH precipitation.
FIG. 8. As in Fig. 7, but for 10-day model forecasts of precipitation.
Fig. 9. As in Fig. 7, but for 15-day forecasts from (left) GFS and (middle) CFS and (right) 30-day forecasts from CFS.
observations. In fact, the position of the area of enhanced precipitation in the GFS forecasts during this period is much closer to the pattern that is observed in CMORPH during the previous period.

During early January 2008 (Fig. 11, bottom row), the 10-day precipitation forecasts for both models agree remarkably well in the pattern, amplitude, and location of the enhanced convection. This region is near the western flank of the SPCZ, which is a region that may be more suited to the model strengths due to the extratropical influences in that area.

Time series of area-averaged precipitation for the same three time periods discussed above are shown in Fig. 12. The map insets in each time series plot indicate the area for which precipitation was averaged. These boxes were chosen so that the time evolution of the precipitation ramp up and die back could be more closely studied among CMORPH and the 10-day model forecasts. The top panel in Fig. 12 shows the time evolution of precipitation in the early stages of the enhanced precipitation that began in early December near 70°E (Fig. 6). The time series shows several peaks, but the peak that we shall focus upon is the largest amplitude peak that occurs near 10 December, which corresponds to the beginning of the event in the equatorial Indian Ocean south of the Bay of Bengal. Note that the observed precipitation begins to ramp up around 20 November, levels off briefly in early December, and
then ramps up sharply to about 22 mm day$^{-1}$. Thereafter the precipitation declines sharply. Both models clearly lag the observations at the initiation of the ramp up by a week or longer but the GFS peak intensity coincides well with CMORPH while the peak intensity in the ECMWF forecasts is several days later, with peak intensities of 15–18 mm day$^{-1}$. The middle panel in Fig. 12 shows the time evolution of the enhanced convection over Indonesia in late December to early January. In this case, the ECMWF 10-day forecasts appear to lead the observed intensity ramp up by several days with a more gradual rise of intensity over time than CMORPH or GFS, while the GFS ramp up lags by about 10 days. Both models reach peak intensity at about the same time (near 1 January), which is about a week later than CMORPH. Over the SPCZ region (Fig. 12, bottom panel), the GFS intensity ramp up is delayed by 5–7 days, but reaches a peak that is nearly coincident with CMORPH although the model maintains that peak for about 15 days, which is about twice as long as the observations indicate. Meanwhile, the ECMWF 10-day forecasts ramp up to peak intensity near the end of January, which is about 2 weeks past the observations. The intensity ramp up in the ECMWF forecasts is also very gradual compared to both GFS and CMORPH.

To this point, we have focused exclusively on the MJO in this section. To broaden our discussion, we now assess the ability of the model predictions to characterize precipitation variations in addition to the MJO over the Indian Ocean and western tropical Pacific. As a means to assess the skill of the model forecasts, we produced time series of pattern correlations between CMORPH and the model predictions over the domain of this study (15°N–15°S, 80°E–160°W) as a function of forecast lead (Fig. 13).

For the line labeled ‘‘CMORPH,’’ the correlations are conducted on CMORPH with itself, except lagged in time, so that a 1-day lag represents a ‘‘persistence’’ forecast (i.e., yesterday’s observation is used as the forecast for today). The horizontal line extending from a pattern correlation coefficient value of 0.51 represents the 1-day CMORPH persistence ‘‘forecast’’ that is used as a skill measure.

Fig. 13. Pattern correlation between CMORPH and model precipitation forecasts (see legend) over the domain defined by 15°N–15°S, 80°E–160°W for various forecast leads (x axis). Heavy solid line is for CMORPH that is lagged with itself, so that a 1-day lag represents a ‘‘persistence’’ forecast (i.e., yesterday’s observation is used as the forecast for today). The horizontal line extending from a pattern correlation coefficient value of 0.51 represents the 1-day CMORPH persistence ‘‘forecast’’ that is used as a skill measure.
CMORPH lag “2” (x-axis label), this means that the CMORPH was persisted from 2 days earlier, etc.

As expected, the pattern correlation decreases with increasing lead, but four particularly noteworthy features are evident. First, all of the model forecasts beat persistence for forecasts leads in the 2–9-day range; thus, all “add value” over persistence forecasts in which the observation “lag” is the same as the forecast “lead.” Interestingly, however, the CFS 1-day forecasts are considerably worse than persistence. Second, the ECMWF forecasts have as much skill as 1-day CMORPH persistence for forecasts out to 4.5 days based on this pattern correlation metric, while ERA-Interim has as much skill out to about 3.5 days, and GFS about 1.5 days. Third, the ECMWF forecasts perform better than the GFS at all forecasts leads by a margin of about 0.10–0.15, and the ERA-Interim is better than the GFS by about 0.05–0.10. Finally, the CFS forecasts are worse than persistence and all other model forecasts examined. We speculate that the relatively poor performance of the CFS compared to the other models that we examined is due primarily to differences in initial conditions, model resolution, and improvements in the other models. For example, CFS initial conditions are provided by the NCEP/DOE Climate Data Assimilation System Version 2 (CDAS2; Kanamitsu et al. 2002). That assimilation system assimilates only observations that were routinely available when the system was developed (ca. 2002). In contrast, the GFS uses initial conditions from the NCEP operational GDAS that have been upgraded substantially with time to enable the assimilation of all of the best available observations, as has the ECMWF models. In addition, the atmospheric component of the CFS is the frozen 2003 version of the GFS model while the other models have benefited from improvements that have been implemented over the years.

5. Conclusions

We have provided a qualitative and quantitative assessment on the ability of state-of-the-art operational forecast models to predict precipitation accurately over the period November 2007–February 2008. We have also provided information about the ability of the models to portray the temporal and spatial variability of precipitation observations. These objectives were accomplished by comparing the model forecasts of precipitation, at various forecast leads, to published and widely used state-of-the-art precipitation datasets.

The results of our study indicate that all four models that were examined exhibit common bias characteristics in magnitude but with some differences in the geographical distribution of them. The largest biases are in the storm-track regions in the northwestern extratropical Pacific and Atlantic and in the ITCZ regions of both oceans. Overall, the ECMWF operational model has the lowest bias. In terms of the variability, the highest local correlations on daily precipitation amounts are observed over the subtropical and extratropical oceans and lowest over the continental areas and the tropics, with the ECMWF and GFS performing the best among the four models. Interestingly, the temporal variability of the GFS and ECMWF precipitation forecasts are virtually identical during the Northern Hemisphere cool season, at 2–9-day forecast leads (Fig. 4), while the ECMWF forecasts are slightly better in the tropics and Southern Hemisphere extratropics during the warm season there. The ERA-Interim forecasts are slightly worse than the GFS in all three zones, while the CFS forecasts perform the poorest with respect to this statistic.

Regarding the distribution of daily rainfall amounts among the models and observations, the model contributions to the total rainfall from the most intense daily precipitation events are 5%–10% lower than is observed from satellite-derived precipitation estimates. Some of this difference may be a consequence of the limited spatial model resolution and therefore the insufficiently accurate representation of smaller-scale convective precipitation events. The differences may also be due to the fact that convective processes occur on spatial and temporal scales that currently cannot be resolved by the models and thus they employ parameterization schemes to determine the precipitation amounts that are generated by these processes (Ricciardulli and Sardeshmukh 2002).

Since the period of this study witnessed active MJO activity, we examined the model precipitation forecast ability to characterize MJO behavior. All models exhibit the capability to reproduce the MJO impressively in precipitation forecasts out to 10–15 days. This is in contrast to the results from the operational models about two decades earlier that produced MJO activity that was extremely weak compared to observations (Janowiak 1992; Hendon et al. 2000; Wheeler and Weickmann 2001). The improvement in the representation of the MJO likely results from the improved initialization systems as well as from improvements in the models. In particular, advances to cumulus convection schemes have made the initiation of convection more difficult, which enhances intraseasonal variability in the numerical models (Tokioka et al. 1988; Wang and Schlesinger 1999). Although the CFS coupled climate model exhibited the ability to characterize MJO propagation, the propagation speed is too slow compared to observations and the other models.

Using pattern correlations of daily precipitation amounts among the forecasts and observed data over the Indian Ocean and western tropical Pacific as a skill metric, the ECMWF and ERA-Interim clearly exhibit
more skill than the GFS and CFS and beat 1-day “persistence” CMORPH observations out to 3–4 days. The skill of the GFS forecasts is similar to ECMWF and ERA-Interim at 1-day forecast leads but consistently about 20% below the skill level of them at forecast leads of 2 days and longer. The CFS forecasts exhibit even less skill than the other models, and the speculation for the relatively poor performance is the poorer initial conditions used by the CFS and the considerably coarser resolution of that model.

A priori, we expected the CFS to perform better than the results that we have presented indicate because of the air–sea coupling in that model. We speculate that the relatively poor performance of the CFS compared to the other models that we examined is due primarily to differences in initial conditions, model resolution, and improvements in the other models. For example, CFS initial conditions are provided by the NCEP/DOE Climate Data Assimilation System Version 2 (CDAS2; Kanamitsu et al. 2002). That assimilation system assimilates only observations that were routinely available when the system was developed (ca. 2002). In contrast, the GFS uses initial conditions from the NCEP operational GDAS that have been upgraded substantially with time to enable the assimilation of all of the best available observations, as have the ECMWF models. In addition, the atmospheric component of the CFS is the year 2003 version of the GFS model while the other models have benefited from improvements that have been implemented over the years. Finally, the use of CDAS2 to initialize CFS may result in significant initial adjustments in the forecasts because of the differences between the forecast model and assimilation model, and this is less of a problem in the GFS forecasts because the same atmospheric model is used for GFS and GDAS. The CDAS2 is not only less accurate than GDAS, it also induces a substantial initial adjustment because of the inconsistency between the forecast model (GFS03) and initialization system (CDAS2). This initial adjustment also results in an even lower precipitation skill for day 1 than a persistence forecast does. In terms of spatial resolution, the CFS (T62; approximately 200-km grid) is considerably more coarse than the GFS, which is T382 (~35-km grid) for the first 180 h followed by the use a T190 (~70 km), ECMWF T799 (~25 km), and ERA-Interim T255 (~80 km).

It is clear that the capability of present-day weather prediction models to forecast precipitation variability has improved significantly over the past decade. We speculate that the reasons are probably due to advances in data assimilation, improved convective parameterization, finer model spatial resolution, and perhaps the availability of an unprecedented volume of passive microwave satellite data. It is beyond the scope of this paper and of our resources to determine the relative contribution to each of the above, but we hope that the results that we have presented in this paper will inspire those that do have the expertise and resources to embark on research to make that determination.

Acknowledgments. The ECMWF interim reanalysis was produced and kindly made available by Dick Dee (ECMWF) and his coworkers. Support for the first author (Janowiak) and one of the coauthors (Arkin) for this work was provided by the NOAA Climate Prediction Center through the Cooperative Institute for Climate and Satellites of the University of Maryland (Grant NA09NES4400006).

REFERENCES
Dee, D., 2004: Variational bias correction of radiance data in the ECMWF system. Proc. ECMWF Workshop: Assimilation of High Spectral Resolution Sounders in NWP, Reading, United Kingdom, ECMWF, 97–112.


——, D. Dee, S. Kobayashi, P. Berrisford, and A. Simmons, 2008: Towards a climate data assimilation system: Status update of ERA-Interim. *ECMWF Newsletter*, No. 115, ECMWF, Reading, United Kingdom, 12–18.

Wang, W., and M. E. Schlesinger, 1999: The dependence on convection parameterization of the tropical intraseasonal oscillation simulated by the UIUC 11-layer atmospheric GCM. *J. Climate*, 12, 1423–1457.


