Evaluating Tropical Cyclone Forecasts from the NCEP Global Ensemble Forecasting System (GEFS) Reforecast Version 2

WEIWEI LI AND ZHUO WANG
University of Illinois at Urbana–Champaign, Urbana, Illinois

MELINDA S. PENG
Marine Meteorology Division, Naval Research Laboratory, Monterey, California

(Manuscript received 26 December 2015, in final form 14 March 2016)

ABSTRACT
Tropical cyclone (TC) forecasts from the NCEP Global Ensemble Forecasting System (GEFS) Reforecast version 2 (1985–2012) were evaluated from the climate perspective, with a focus on tropical cyclogenesis. Although the GEFS captures the climatological seasonality of tropical cyclogenesis over different ocean basins reasonably well, large errors exist on the regional scale. As different genesis pathways are dominant over different ocean basins, genesis biases are related to biases in different aspects of the large-scale or synoptic-scale circulations over different basins. The negative genesis biases over the western North Pacific are associated with a weaker-than-observed monsoon trough in the GEFS, the erroneous genesis pattern over the eastern North Pacific is related to a southward displacement of the ITCZ, and the positive genesis biases near the Cape Verde islands and negative biases farther downstream over the Atlantic can be attributed to the hyperactive Africa easterly waves in the GEFS. The interannual and subseasonal variability of TC activity in the reforecasts was also examined to evaluate the potential skill of the GEFS in providing subseasonal and seasonal predictions. The GEFS skillfully captures the interannual variability of TC activity over the North Pacific and the North Atlantic, which can be attributed to the modulation of TCs by the El Niño–Southern Oscillation (ENSO) and the Atlantic meridional mode (AMM). The GEFS shows promising skill in predicting the active and inactive periods of TC activity over the Atlantic. The skill, however, has large fluctuations from year to year. The analysis presented herein suggests possible impacts of ENSO, the Madden–Julian oscillation (MJO), and the AMM on the TC subseasonal predictability.

1. Introduction
Tropical cyclones (TCs) are among the most destructive weather-related natural disasters. In the past two centuries, TCs have caused approximately 1.9 million deaths worldwide and killed more than 300,000 people in North America and the Caribbean alone (Tobin 1997; Shultz et al. 2005; Pielke et al. 2008). Meanwhile, property damage from TCs increases because of the growing coastal population and development. Skillful TC prediction is therefore of significant socioeconomic value. On the synoptic time scales, TC track forecasting has been improved significantly in the past decade (WMO 2007), and progress has also been made over the years toward improving the prediction skill of TCs on seasonal and longer time scales (e.g., Goldenberg et al. 2001; Camargo et al. 2007; Knutson et al. 2010; Vecchi et al. 2011). TC forecasts on the subseasonal time scales (7–60 days) provide useful information for early storm preparedness, especially in remote or large communities (Brunet et al. 2010). Statistical and dynamical models have been developed for TC subseasonal prediction. Leroy and Wheeler (2008) and Slade and Maloney (2013) applied logistic regression to predict TC frequency on the subseasonal time scales using the TC climatology, sea surface temperature (SST), El Niño–Southern Oscillation (ENSO), and Madden–Julian oscillation (MJO) indices as predictors, and they showed that skillful prediction can be achieved out to 3 weeks during strong MJO events. Roundy and Schreck (2009) further considered convectively coupled easterly waves and mixed Rossby–gravity waves to account for the

Corresponding author address: Zhuo Wang, Dept. of Atmospheric Sciences, University of Illinois at Urbana–Champaign, 105 S. Gregory St., Urbana, IL 61801.
E-mail: zhuowang@illinois.edu

DOI: 10.1175/WAF-D-15-0176.1

© 2016 American Meteorological Society
synoptic-scale variability associated with TC development (http://www.atmos.albany.edu/facstaff/rounby//tcforecast/tcforecast.html).

Dynamical prediction has improved tremendously in recent years as a result of improved model physics, initialization, and increased model resolutions (Klotzbach et al. 2012). Dynamical prediction employs numerical models that generate TC-like disturbances. Such disturbances can be tracked based on the dynamic and thermodynamic features of a TC [e.g., a warm-core structure and strong low-level cyclonic circulation; Gall et al. (2011) and Gopalakrishnan et al. (2012, 71–91)]. Vitart (2009) showed that the accurate representation of the MJO had a significant impact on the subseasonal prediction skill of landfalling TCs over North America and Australia in the European Centre’s Medium-Range Weather Forecasts (ECMWF) 46-days hindcasts. Elsberry et al. (2010) used an objective matching technique and showed that the ECMWF 32-day forecasts can provide useful TC track information on time scales of 10–30 days. Belanger et al. (2010, 2012) found that the TC genesis forecasts from the ECMWF have skill above climatology through 4 weeks over the Atlantic main development region [MDR; between 9° and 21.5°N spanning the Caribbean Sea and tropical Atlantic; Goldenberg et al. (2001)] and the Gulf of Mexico. They found that the regional forecast skill can be attributed to the ability of the model in realistically capturing the large-scale environment and the evolution of the MJO. Some dynamic models demonstrated comparable or better forecast skill than statistical models in some ocean basins (Vitart et al. 2010). In addition, multimodel ensemble techniques produce overall better forecasts than individual models and further improve the skill of dynamical prediction (Sivillo et al. 1997; Vitart 2006; Kirtman et al. 2014).

The present study will evaluate the dynamical prediction skill of TC activity in the National Centers for Environmental Prediction’s (NCEP) Global Ensemble Forecasting System (GEFS) using the GEFS Reforecast version 2 dataset (Hamill et al. 2013). The GEFS reforecasts have a forecast lead time up to 16 days and are available since 1 December 1984. Different from most previous studies, we will evaluate the skill of the GEFS from a “climate” perspective. The reasoning is that a skillful model will produce a realistic climatology of TC activity; otherwise, forecast errors may accumulate and manifest in the model climatology. Diagnosis of the long-term statistics will help to identify model error sources and can provide useful information on model improvement. In addition to TC climatology, we will evaluate the interannual and subseasonal variability of TC activity and assess the potential skill of the GEFS on the subseasonal to seasonal time scales, which is highly relevant to the goal of the National Oceanic and Atmospheric Administration (NOAA) to extend the lead time for skillful prediction of extreme events. Since a coarse-resolution global model generally has difficulty in predicting TC intensity, we will mainly focus on tropical cyclogenesis and briefly examine TC tracks.

The rest of the paper is organized as follows. The datasets and methods are described in section 2. The TC climatology in the GEFS is examined in section 3, and the related model biases and deficiencies in the model physics are investigated in section 4. The potential skill of the GEFS in TC subseasonal and seasonal prediction is assessed in section 5, followed by a summary and additional discussion in section 6.

2. Data and metrics

a. Global Ensemble Forecasting System Reforecast version 2

The GEFS reforecasts were developed by the NOAA’s Earth System Research Laboratory (ESRL) using version 9.0.1 of the operational GEFS (ftp://ftp.cdc.noaa.gov/Projects/Reforecast2). The reforecast is up to 16 days and consists of 11 ensemble members (one control run and 10 perturbed members). The reforecasts were initialized once per day at 0000 UTC with the Climate Forecast System Reanalysis (CFSR) through 20 February 2011 and with the analysis from the Gridpoint Statistical Interpolation (GSI) analysis system afterward (Hamill et al. 2013). We will refer to the initial conditions as the CFSR for brevity. The reforecasts have 42 vertical levels, and the horizontal resolution is T254 (approximately 40 km at 40° latitude) for the first week and is degraded to T190 (approximately 54 km at 40° latitude) after day 7.5. All reforecast data were saved at 1° × 1° horizontal resolution. Our analysis focuses on the forecast verification for the time period from January 1985 to December 2012. Such a long time period is critical for establishing robust statistics of weather events like TCs and for identifying possible systematic errors in the mean state.

We will evaluate TCs over different ocean basins for different forecast lead times. To exclude the forecast skill directly related to the model initialization, we omitted TC vortices during the first 18 forecast hours. The week-1 reforecasts include the forecast lead time of 24–162 h (days 2–7 with a 6-h interval), and the week-2 reforecasts include the forecast lead time of 168–306 h (days 8–13 with a 6-h interval). A TC year is defined as January–December for the Northern Hemisphere and from July to next June for the Southern Hemisphere (McBride and Keenan 1982; Vitart et al. 1997). Several TC indices were used in this study to facilitate quantitative evaluations. “TC counts” is defined
as the total number of TCs within a certain period over a certain basin; “TC days” is the sum of the lifetime of all TCs, measured in days; and “accumulated cyclone energy” (ACE; Bell et al. 2000) is calculated by integrating the squares of the 6-hourly maximum sustained surface wind speed (in \( \text{kt}^2 \), where 1 \( \text{kt} = 0.51 \text{ m s}^{-1} \)) over the lifetime of a TC for all TCs over a certain basin, which is a function of TC counts, lifetime, and intensity. To examine the mean state in the GEFS, long-term seasonal ensemble means were derived during 1985–2012, which were evaluated against the interim ECMWF reanalysis (ERA-Interim, ERAI; Dee et al. 2011). The ERAI data, with the original resolution of approximately 0.7°, were coarsened to 1° × 1° resolution to facilitate comparison.

In addition to the ERAI, the NOAA’s Optimum Interpolation SST version 2 (OISST; Reynolds et al. 2002) data and two satellite datasets were used. The Climate Prediction Center (CPC) morphing technique (CMORPH) derives precipitation estimates from the passive microwave data at a very high spatial and temporal resolution from December 2002 to the present (Joyce et al. 2004). The 3-hourly 0.25° × 0.25°CMORPH precipitation for the period 2003–2012 was mapped to 1.0° × 1.0°, and the daily average was taken to be consistent with the GEFS reforecasts. To evaluate the precipitation and column water vapor (CWV) relationship, we employed the version 7 Special Sensor Microwave Imager/Sounder (SSMIS) polar-orbiting dataset, which had simultaneous retrievals of precipitation and CWV (Horváth and Gentemann 2007; Wentz 2013). Both precipitation and CWV were coarsened to a 1.0° × 1.0° resolution grid mesh to be consistent with the GEFS reforecasts, and daily averages were taken over the tropical ocean areas for the same period as the CMORPH (2003–12).

### b. Detection of TCs

TCs in the GEFS reforecasts were tracked using the GFDL vortex tracker with an updated warm-core criterion (Gall et al. 2011; Gopalakrishnan et al. 2012, 71–91). This tracker has been adopted by NCEP for detecting and tracking TC vortices in their operational model since 1998. The scheme evaluates several key parameters, including relative vorticity, geopotential height, and wind speed at 850 and 700 hPa; sea level pressure; and 10-m wind speed. A threshold of 10-m maximum wind speed of 16.5 m s\(^{-1}\) or 32 kt was adopted for TC detection based on the data resolutions (Walsh et al. 2007), and it was also required that a TC had a warm core lasting at least 48 h cumulatively. The TC center was tracked by searching for the average location of the maximum or minimum of the key parameters, and a genesis was defined as the first record of a TC track. To avoid uncertainties due to changes in TC observing systems, short-lived TCs (lifetime less than 48 h) were excluded in both the observation and the GEFS reforecasts (e.g., Villarini et al. 2011; Chen and Lin 2013). Storms forming poleward of 40° latitude were regarded as extratropical cyclones and were excluded. TCs were tracked in individual ensemble members, and the ensemble mean of TCs based on the 11 ensemble members was evaluated against the International Best Track Archive for Climate Stewardship (IBTrACS v03r05; Knapp et al. 2010). Since all ensemble members have a good level of agreement in the TC climatological biases and mean state biases, only the ensemble means will be discussed in the following sections for brevity.

### 3. Climatology of TC activity in the GEFS

We first examined the long-term mean and seasonality of TCs during 1985–2012 in the GEFS. As shown in Fig. 1, the seasonal evolutions of TC genesis frequency over different ocean basins resemble the observed seasonality in both the GEFS week-1 and -2 reforecasts. The GEFS captures the peak storm season from June/July to October over the western and eastern North Pacific, as well as over the North Atlantic. The model also reproduces the bimodal distribution over the Indian Ocean, with a primary (postmonsoon) peak in October–November and a secondary (premonsoon) peak in May–June. In the Southern Hemisphere, TC activity in the GEFS peaks during January–February, same as in the observations. On the other hand, quantitative differences are evident over most ocean basins. For example, negative biases in TC frequency are found over the western and eastern North Pacific, and positive biases are present over the north Indian Ocean and the Southern Hemisphere. The seasonality curves over the Atlantic are nearly perfect, but as shown next, large biases exist at the regional scale.

To evaluate the spatial distribution of TC formations, the TC genesis density function (GDF) was derived. It is defined as the total number of TCs per year forming within a 10° × 10° box centered on each 1.0° grid point, and the long-term mean during 1985–2012 is shown in Fig. 2. The spatial pattern of the GDF in the GEFS week-1 reforecasts qualitatively agrees with that derived from the IBTrACS (Figs. 2a,b): most model TCs form equatorward of 20° latitude, the GEFS reproduces the GDF maxima over the western and eastern North Pacific, more storms form over the Bay of Bengal than the Arabian Sea, and a nearly zonal band of TC genesis is present along 15°S. However, large biases exist at the regional scale (Fig. 2c). Over the western North Pacific, the GEFS has a strong negative bias in GDF to the east of the Philippines and a weak positive bias over the East China Sea and the northern central Pacific near the date...
Over the eastern North Pacific, a dipole pattern indicates a southeastward shift of the TC genesis center. Over the Atlantic, almost no TCs form over the west Atlantic and the central MDR (west of 40°W). In contrast, hyperactive cyclogenesis occurs near the Cape Verde islands (off the coast of West Africa). Consistent with the seasonality shown in Fig. 1, positive biases are found over the Southern Hemisphere and the north Indian Ocean. The differences in the GDF between the GEFS week-1 and -2 reforecasts suggest that the negative biases in the western North Pacific and the dipole pattern biases in the eastern North Pacific increase with the forecast lead times (Fig. 2d).

We now examine the spatial distribution of the TC tracks. The track density function (TDF) is defined as the total number of TC days per year within a 10° × 10° box centered on each 1.0° grid point (Fig. 3). The TDF depends on both TC genesis frequency and the subsequent storm tracks, and the environmental factors affecting storm tracks and TC formations are not necessarily the same (Mei et al. 2014). The spatial pattern of the TDF in the GEFS week-1 reforecasts is broadly consistent with that in the observations (Figs. 3a,b), such as the high track density in the western and eastern North Pacific, the relatively low TDF in the North Atlantic, and the zonally elongated TC tracks in the Southern Hemisphere. The reforecasts also capture the large poleward extension of TC tracks in the western North Pacific and the more latitudinally confined TC activity in the eastern North Pacific. This implies that the GEFS is skillful in reproducing the large-scale steering flows. However, large regional biases exist over all ocean basins. Figure 3c shows that the TDF is low in the GEFS over the western North Pacific and the western North Atlantic, which can be at least partially attributed to the negative genesis biases in these regions. The positive biases in the TDF over the central and eastern MDR are consistent with the hyperactive TC formation near the
4. Impact of environmental condition biases on TC prediction

a. Biases in tropical cyclogenesis

The TC climatology in the GEFS shows some regional biases. Since TC genesis depends on large-scale environmental conditions, it is natural to ask which mean state errors in the model contribute to the genesis biases. We will use the genesis potential index (GPI) to address this.
question, and we focus on the western North Pacific, the eastern North Pacific, and the North Atlantic during July–October (JASO). GPI is a function of environmental variables and can serve as a proxy for genesis probability. Following Emanuel and Nolan (2004), it is defined as

$$\text{GPI} = 10^5 \eta^{3/2} \left(\frac{\text{RH}}{50}\right)^3 \left(\frac{\text{PI}}{70}\right)^3 \left(1 + 0.1 \text{VWS}\right)^{-2},$$

where $\eta$ is 850-hPa absolute vorticity, RH is 700-hPa relative humidity, and VWS is the 850–200-hPa vertical
wind shear (vector difference). PI is the potential intensity (Emanuel 1995), which is a function of the sea surface temperature and the pressure and vertical profiles of temperature and specific humidity, as well as the ratio of the exchange coefficient for enthalpy to the drag coefficient.

The monthly mean GPI in the GEFS was calculated from the monthly mean atmospheric fields and SST in the GEFS day-8 reforecast, and the long-term mean GPI was derived by averaging over 1985–2012. The GPI was then calibrated (or rescaled) so that the average TC frequency predicted by the GPI over the North Pacific and the North Atlantic (0°–40°N, 100°–355°E) was the same as observed. The calibrated long-term seasonal mean GPI derived from the ERAI and the OISST is shown in Fig. 4a. It captures the major centers of genesis activity over the western North Pacific, the eastern North Pacific, and the Atlantic, but overpredicts genesis over the central Pacific, which is a common issue with the GPI (Camargo et al. 2014).

Figure 4b displays the differences in the calibrated seasonal mean GPI between the GEFS day-8 reforecast and the ERAI. The genesis density function differences between the GEFS day-8 reforecast and the IBTrACS are shown in Fig. 4c to facilitate comparisons. The negative biases in GPI prevail over both the North Pacific and the North Atlantic (Fig. 4b). The broad consistency with the genesis density differences (Fig. 4c) suggests that the biases in environmental conditions can to a large extent explain the negative biases over the western and eastern North Pacific and the Atlantic. However, exceptions exist near Central America and the west coast of Africa, suggesting that the positive genesis biases in these two regions cannot be simply attributed to biases in the environmental conditions.

The relative contribution of an environmental variable to the genesis biases can be examined by replacing its monthly mean from the GEFS reforecast by the one from the ERAI to recalculate the GPI (with all the other variables from the GEFS reforecast). The GPI derived exclusively from the GEFS were then subtracted from the recalculated GPI, and the resultant differences can indicate how much the correction in a certain variable may affect the biases in GPI. The 850-hPa vorticity, 700-hPa RH, and 200–850-hPa VWS were examined in Figs. 4d–f, respectively. Readers, however, should be cautioned that these variables may be closely related to each other in some ocean basins.

In Fig. 4d, a positive difference over the western North Pacific suggests that the correction in the 850-hPa vorticity field reduces the negative genesis bias in this region. In other words, the bias in the GEFS vorticity

---

**Fig. 4.** (a) The calibrated long-term seasonal mean (JASO) GPI from the ERAI. (b) The calibrated long-term seasonal mean GPI differences between the GEFS day-8 reforecast and the ERAI. (c) As in Fig. 4c, but for JASO from the GEFS day-8 reforecast. (d) The contribution of the 850-hPa relative vorticity (Vort850) to the genesis bias (see the text for details). Results shown are the same as in (d) but for the (e) 700-hPa relative humidity (RH700) and (f) 200–850-hPa VWS.
field makes a large contribution to the negative genesis bias in that region. On the other hand, the correction in the 850-hPa vorticity field enhances the negative biases over the subtropical west Pacific and the subtropical west Atlantic, suggesting errors in the GEFS vorticity field may compensate for the biases in other variables in these regions.

Similar diagnoses were carried out by replacing the 700-hPa RH (Fig. 4e) and the 850–200-hPa vertical shear (Fig. 4f), respectively. Figure 4e suggests that the correction in the 700-hPa RH helps to reduce the negative genesis biases over the South China Sea, the Philippine Sea, and the subtropical west Atlantic (note the partial cancellation between the vorticity and the RH fields in this region). Figure 4f shows that the correction in the vertical shear helps to improve the GPI prediction over the subtropical west Pacific, the eastern North Pacific, the subtropical west Atlantic, and the Atlantic MDR.

Overall, Fig. 4 suggests that TC genesis in the GEFS is more sensitive to the dynamic variables over the western North Pacific and to the thermodynamic ones over the North Atlantic, consistent with the observational analysis by Peng et al. (2012) and Fu et al. (2012). Next, we will examine the environmental biases within the context of large-scale circulations, and also investigate the positive genesis biases near Central America and the west coast of Africa, which the GPI diagnosis fails to explain.

1) Western North Pacific

A third of global TCs form over the western North Pacific, 75% of which develop in the monsoon trough environment (Ritchie and Holland 1999). The dynamic and thermodynamic conditions in the monsoon trough zone are favorable to the growth and development of a TC, such as the low-level convergence and cyclonic vorticity, weak vertical wind shear, high midlevel relative humidity, and warm SST (e.g., Holland 1995; Chen et al. 1996; Gray 1998). The successful prediction of a monsoon trough may contribute to the skillful extended-range forecast of tropical cyclogenesis (Nakano et al. 2015).

The ERAI shows a northwest–southeast-oriented monsoon trough across 110°–150°E (Fig. 5a). The low-level monsoonal southwesterlies extend from the Indochina Peninsula to 150°E. The GEFS day-8 reforecast is broadly consistent with the ERAI results (Fig. 5b), but the monsoon trough is much weaker than observed and confined to the South China Sea, while the trade wind easterlies prevail east of the Philippines. The weaker monsoon trough is associated with negative biases in column water vapor and precipitation (Fig. 5c), and weaker low-level convergence and cyclonic vorticity (not shown) over the South China Sea and the Philippine Sea (not shown), all contributing to the underprediction of TC activity in that region (Fig. 2c). Also note that the extent (or strength) of the subtropical ridge over East Asia is underpredicted by the GEFS as well, which results in biases in the steering flow and the storm tracks (Fig. 3c), as well as overpredicted extratropical cyclogenesis over the East China Sea and Japan (not shown).

2) Eastern North Pacific

The eastern North Pacific is the second most active basin for TC activity (Peduzzi et al. 2012; Jin et al. 2014). TC formation over this basin is more confined in both the longitudinal and latitudinal extents, resulting in the highest genesis density over the globe (Davis et al. 2008; Camargo et al. 2008).

TC activity in this region is closely related to the regional Hadley circulation and the intertropical convergence zone (ITCZ; Zhang and Wang 2015). To evaluate the skill of the GEFS in capturing the ITCZ, precipitation is examined here. Figure 6a shows the difference of the long-term mean precipitation rate in JASO during 2003–12 between the GEFS day-8 reforecast and the CMORPH. A dipole pattern in the precipitation difference over the eastern North Pacific indicates a southward displacement (~2°) of the ITCZ in the GEFS, along with the overpredicted precipitation in the Central America. Previous studies have suggested that tropical cyclogenesis is sensitive to the meridional displacement of the ITCZ, which is associated with variations in the low-level vorticity, divergence, vertical wind shear, etc. (Molinari and Vollaro 2000; Merlis et al. 2013). In addition, the ITCZ breakdown is an important mechanism for tropical cyclogenesis over the east Pacific (e.g., Nieto Ferreira and Schubert 1997; Wang and Magnusdottir 2005). Given that the easternmost tip of the ITCZ is the most unstable region along the vorticity strip (Wang and Magnusdottir 2006), the southeastward shift of the TC genesis center is consistent with the southward displacement of the ITCZ in the GEFS (Fig. 2c). The biases in the 700-hPa RH also contribute to the positive genesis bias over Central America (Fig. 4e).

3) North Atlantic

The ITCZ is also displaced southward over the central MDR of the Atlantic (50°–15°W; Fig. 6b), which is consistent with the negative genesis biases in this region (Merlis et al. 2013). The GPI diagnosis suggests that the genesis biases over the Atlantic can be mainly attributed to the biases in relative humidity and vertical shear (Figs. 4e,f), while it fails to explain the positive genesis bias near the Cape Verde islands (Fig. 4c).
Most Atlantic TCs (~60%) develop from African easterly waves (AEWs), especially over the MDR (e.g., Landsea et al. 1998; McTaggart-Cowan et al. 2008). Given the role of the AEWs in tropical cyclogenesis, we next examine the representation of the AEWs in the GEFS. The seasonal variances of 2.5–9-day bandpass-filtered 850-hPa meridional wind along 15°N were calculated and then averaged over 28 yr (1985–2012) to represent the wave activity. Figure 7a shows the variances from the ERAI, the CFSR (i.e., day 0), and the GEFS day-8 reforecast. The ERAI shows a peak in variance (~12.5 m² s⁻²) near the West African coastline, and the wave activity over the African continent is much weaker. This is consistent with the observed AEW structure: the AEWs are mainly confined to 600–700 hPa south of the African easterly jet over the land, and
become stronger and attain a deeper structure offshore as a result of coastal convection and/or interaction with the northern wave track (Thorncroft and Hodges 2001; Hankes et al. 2015). Compared to the ERAI, the AEWs in the CFSR are much stronger. Results show a primary peak over the ocean (~18°W) similar to the ERAI but with larger amplitude, and a secondary peak over the land (~7.5°W). The synoptic-scale biases in the CFSR can be an error source for the subsequent TC forecasts through downscale cascading (Peters and Roebber 2014). The positive biases in the variance over the land are even stronger in the GEFS day-8 reforecast, and the wave activity over the land is comparable to or even stronger than that over the ocean, suggesting that the AEWs have an unrealistically deep structure over the land.

To further examine the structure of the AEWs, we regressed the 2.5–9-day bandpass-filtered meridional wind and the 2.5–9-day bandpass-filtered diabatic heating rate $Q_1$ (Yanai et al. 1973) at each grid point and each pressure level against the 2.5–9-day bandpass-filtered 850-hPa meridional wind at a reference point (15°N, 12°W) over West Africa. Before applying the bandpass filter and linear regression, the seasonal cycle was removed by subtracting the climatological mean daily data. Figures 7b–d show the anomalies in the meridional wind and $Q_1$ fields for one standard deviation of the 850-hPa meridional wind at the reference point. Consistent with previous studies (e.g., Kiladis et al. 2006), the ERAI shows an eastward tilt to the wave structure below 700 hPa and a westward tilt above. The weakening amplitude with height over the land indicates that the wave has a shallow structure in the lower troposphere. The $Q_1$ field shows that shallow heating occurs in the southerly flow west of the trough axis, and the heating near the wave trough is rather weak. Figure 7c shows that the CFSR is generally consistent with the ERAI, but the AEWs are slightly stronger at the lower troposphere (~850 hPa). In addition, the $Q_1$ field shows that the heating in the CFSR is stronger, deeper, and closer to the wave trough axis than that in the ERAI. The wave is even stronger in the GEFS day-8 reforecast, with stronger and deeper heating in the vicinity of the 700-hPa trough (Fig. 7d). Since a wave of a deeper structure and with active convection is more conducive to TC formation (Raymond and López Carrillo 2011; Wang 2012), the biases in the AEWs help to explain the positive TC genesis biases near the Cape Verde islands. Meanwhile, the stronger and deeper waves can be attributed to the stronger deep convective heating associated with the AEWs in the GEFS.

b. Possible deficiencies in the model physics

The biases in the RH and $Q_1$ fields hint at possible deficiencies in the model physics, especially the cumulus...
scheme. The GEFS (version 9.0.1) employed the simplified Arakawa–Schubert (SAS) deep convection scheme. By depleting the excessive instability in the atmospheric column, the SAS scheme in the GEFS was modified to suppress the grid-scale precipitation and to make the cumulus convection stronger and deeper (Han and Pan 2011). Figure 8a shows the average daily precipitation rate in 1-mm-wide bins of CWV over the tropical ocean areas from the SSMIS and the GEFS day-8 and -16 reforecasts. Consistent with the SSMIS, the GEFS captures the nonlinear relationship between precipitation and CWV and the exponential increase in precipitation for large CWV. However, the GEFS generates too much precipitation for a given value of CWV, implying that precipitation may be triggered too early in the model in terms of the CWV accumulation. A similar bias was found by Li et al. (2014) in the operational forecasts of the Navy Operational Global Atmospheric Prediction System (NOGAPS), which also employed the SAS scheme. The difference between day-8 and -16 reforecasts shows an increasing bias with the forecast lead time.

Figure 8b shows the probability distribution of CWV over the tropical oceans from the SSMIS, the CFSR (day 0), and the GEFS day-8 and -16 reforecasts. Consistent with the SSMIS, the GEFS depicts a bimodal distribution of CWV with a primary peak around 57 mm and a secondary peak around 34 mm. A dry bias is found in the GEFS reforecasts: the CWV of maximum frequency of occurrence shifts from 56 to 48 mm in the GEFS day-8 reforecast, and to 46 mm in the day-16 reforecast. Given the precipitation–CWV relationship in Fig. 8a, the dry

![Graph](https://example.com/graph.png)
bias in CWV indicates the underprediction (overprediction) of heavy (light) precipitation in the GEFS, which is confirmed in Fig. 8c. Compared to the CMORPH findings, the GEFS underpredicts moderate-to-heavy precipitation (>30 mm day$^{-1}$) by over 70%. Consistent with CWV, the negative bias in precipitation increases with the forecast lead times. The frequency of drizzle-to-light (15–30 mm) precipitation is slightly underpredicted in the day-8 reforecast and overpredicted in the day-16 reforecast.

The early initiation of precipitation with respect to the CWV is consistent with hyperactive convection over West Africa. This was also confirmed by Bombardi et al. (2015), who found deep convection was triggered too frequently in the SAS scheme. The underprediction of heavy precipitation is consistent with the weaker monsoon trough in the GEFS than observed (Fig. 5). A weaker monsoon trough along with negative genesis biases over the western North Pacific was also found in
the Model for Prediction Across Scales (MPAS) when using the SAS scheme (C. Davis 2015, personal communication). Overall, it suggests that an improved cumulus parameterization may help to reduce the errors in the model mean state and improve TC prediction skill on the regional scale. In addition to the cumulus scheme, the lack of air–sea interaction and the deficiencies in other model physics, such as the boundary layer and land surface parameterizations (Taylor and Clark 2001), may also lead to biases in the diabatic heating field and African easterly waves. Further study is warranted to improve our understanding and the model representation of these processes.

5. Potential prediction skill on the subseasonal and seasonal time scales

In this section, we will examine the interannual and subseasonal TC variability in the GEFS. The investigations will help to evaluate how well the model reproduces the impacts of some large-scale climate modes on TCs and provide insights into the potential skill of the GEFS in making real-time operational subseasonal and seasonal TC predictions.

a. Interannual TC variability

The time series of the annual TC counts and ACE from the GEFS week-1 and -2 reforecasts and the IBTrACS are shown in Figs. 9 and 10. Visual inspection suggests that the GEFS week-1 and -2 reforecasts have a good level of agreement with the IBTrACS over the western North Pacific, the eastern North Pacific, and the Atlantic (Figs. 9a–c and 10a–c), especially for ACE. The GEFS also captures the out-of-phase relationship of the TC activity between the eastern North Pacific and the Atlantic (Wang and Lee 2009; Zhang and Wang 2015). On the other hand, the GEFS fails to skillfully predict the interannual TC variability over the north Indian Ocean and the Southern Hemisphere (Figs. 9d,e and 10d,e). The Spearman rank correlations with IBTrACS are significant for the GEFS week-1 and -2 reforecasts over the eastern North Pacific and the Atlantic and for the week-2 reforecasts over the western North Pacific ($r \geq 0.48$; Table 1), while the correlations are much weaker over the north Indian Ocean and the Southern Hemisphere. The correlations of ACE are much higher than those of the TC counts over the North Pacific and the North Atlantic (Table 1), which can probably be attributed to the strong control of ACE by the SST. It is interesting to note that the rank correlations in the week-2 reforecasts, although still insignificant, are higher than those in the week-1 reforecasts over the north Indian Ocean and the Southern Hemisphere.

Previous studies have shown that ENSO impacts the TC activity in various ocean basins, and that the AMM strongly modulates TCs over the North Atlantic (e.g., Gray 1984; Landsea and Knaff 2000; Goldenberg et al. 2001; Jin et al. 2014). Such low-frequency climate modes provide sources of predictability for TC activity on the subseasonal and longer time scales, and it is interesting to examine how well the GEFS represents their impacts. Table 2 lists the Pearson correlations of the GEFS’ TC indices with the observed Niño-3.4 and the AMM indices (http://www.esrl.noaa.gov/psd) over the western North Pacific, the eastern North Pacific, and the North Atlantic. The correlations with the TC indices derived from IBTrACS are also listed for comparison. Consistent with previous studies (e.g., Lander 1994; Goldenberg and Shapiro 1996), IBTrACS shows that the AMM index is moderately correlated with both TC counts and ACE over the eastern North Pacific and the North Atlantic. Over the western North Pacific, ENSO strongly modulates the interannual variability of ACE ($r = 0.72$) but is weakly correlated with the annual TC counts. Over the North Atlantic, the AMM plays a prominent role in the interannual variability of TC counts ($r = 0.74$) and ACE ($r = 0.78$). The GEFS week-1 and -2 reforecasts capture all the observed significant correlations between ENSO and the TC indices, including a strong correlation with ACE and a weak correlation with the TC counts over the western North Pacific. This is consistent with the relatively low prediction skill of TC counts compared to that of ACE over the western North Pacific. It is also interesting to note that the GEFS overpredicts all the correlations with ENSO, especially for ACE in the week-2 reforecasts. The GEFS captures the positive correlations between the AMM and the Atlantic TC activity, but the correlations are weaker than those observed, which is probably associated with the erroneous location of the ITCZ over the Atlantic.

b. Subseasonal TC variability in the Atlantic

In this section, we will examine how well the inactive/active TC periods are captured by the GEFS. For brevity, we will focus on the Atlantic basin. As an example, the time series of the weekly TC days from the GEFS week-1 and -2 reforecasts were compared with that derived from IBTrACS for the hurricane season in 2000 in Fig. 11a. The annual cycle has been removed to highlight the subseasonal variability. The Atlantic basin underwent three active periods of TC activity, in mid-August, mid-September, and early to mid-October, respectively. The GEFS captures the peaks in TC days during August and early October but misses a primary peak in mid-September and a small peak in mid-October.
The Pearson correlations are 0.79 and 0.53 in the week-1 and -2 reforecasts, respectively. Similar time series were constructed for the weekly cyclone energy (WCE) being defined as the squares of the 6-hourly maximum sustained surface wind speed (in \( \text{kt}^2 \)) integrated over all active TCs for the time period of 1 week (Fig. 11b). The GEFS shows higher correlations with IBTrACS for WCE than for TC days, especially in the week-1 reforecasts. It is also notable that both the week-1 and -2 reforecasts underpredict the amplitude of the WCE peak in late September–early October. This can be partly attributed to the underpredicted peak in the TC days (Fig. 11a) as well as the underpredicted TC intensity in the global model.

Other years can be evaluated similarly, and the model skill is summarized by the time series of the Pearson correlation between the observed and predicted TC days or WCE (Fig. 12). The time series of the correlation coefficients for TC days show that the GEFS agrees well with the observation in the week-1 reforecasts, with the correlation coefficient above 0.6 in most years. The correlations for the week-2 reforecasts are weaker compared to the week-1 reforecasts but still above the 95% confidence level in most years. The mean correlations during...
1985–2012 are 0.65 for the week-1 reforecasts and 0.44 for the week-2 reforecasts. The correlations for WCE are generally higher than those for TC days, with the mean correlations of 0.75 for the week-1 reforecasts and 0.50 for the week-2 reforecasts. Figure 12 suggests that the model has reasonable skill in predicting the active and inactive TC periods with a lead time of 7–14 days.

A striking feature in Fig. 12 is the large year-to-year variability of the correlation coefficients, especially for the TC days. Since the GEFS reforecasts used a fixed-model version, this suggests that intrinsic predictability may vary from year to year. Previous studies have suggested that the low-frequency climate modes such as the ENSO and MJO can enhance the TC subseasonal prediction skill (Leroy and Wheeler 2008; Slade and Maloney 2013). It is natural to ask whether TCs are more predictable in some climate conditions, such as during strong ENSO or AMM events or in a year of active MJO.

To examine how the prediction skill of TC subseasonal variation is modulated by different climate modes, Fig. 13 shows the mean correlations of TC days and WCE with the observations in different climate regimes. The stratification was based on a ±0.5 standard deviation (SD) of the seasonal mean AMM or ENSO index during July–November in each year. Here, the MJO activity is represented by the seasonal mean amplitude of the daily velocity potential MJO indices (VPM; Ventrice et al. 2013), and a larger seasonal mean amplitude indicates stronger MJO activity. The VPM is similar to the all-season real-time multivariate MJO index (RMM; Wheeler and Hendon 2004), but using 200-hPa velocity
potential instead of outgoing longwave radiation. Ventrice et al. (2013) show that the VPM can better represent the MJO activity in the Western Hemisphere than the RMM index. Our analysis also shows a stronger contrast in the stratification using the VPM than using the RMM or a local MJO index defined by the space–time-filtered OLR (not shown). The two-tailed Student’s $t$ test was applied to explore the statistical significance of the difference in the mean correlations between the neutral group (from $-0.5$ to 0.5 SD) and an anomalous group ($>0.5$ or $<-0.5$ SD).

Figure 13 shows that prediction skill tends to be higher (weaker) in years of strong (weak) MJO activity. The model also tends to have better skill during strong ENSO years than in neutral years, especially in strong La Niña years. The AMM does not seem to impact the prediction skill, except that the correlation of WCE is slightly higher in the positive AMM phase. This may be due to the underpredicted correlation between the AMM and the TC counts in the GEFS (Table 2). In addition, most differences in Fig. 12 are statistically insignificant, which is probably due to the small sample size and the large fluctuations.

### 6. Summary and discussion

The tropical cyclone (TC) prediction skill of the NCEP Global Ensemble Forecasting System (GEFS) Reforecast version 2 was evaluated against IBTrACS, the interim ECMWF reanalysis (ERAi), and two satellite datasets with a special focus on tropical cyclogenesis. Different from most previous studies, we assessed the model’s skill from a “climate” perspective. The basic reasoning is that forecast errors may accumulate on longer time scales and the evaluation of climatology can help to identify model error sources. The evaluation of TC activity in the GEFS shows that the GEFS captures the seasonality of TC activity reasonably well and the spatial pattern of TC activity in the GEFS is broadly consistent with the observations, but quantitative errors are present over all ocean basins. Positive biases in genesis frequency are found over the Southern Hemisphere and the Indian Ocean, and negative biases over the western North Pacific and the eastern North Pacific. The genesis center also shifts southeastward over the eastern North Pacific. Although the long-term mean basin-wide TC counts are well captured by the GEFS over the Atlantic, genesis almost exclusively occurs near the Cape Verde islands and is substantially underpredicted over the rest of the Atlantic basin. The biases in TC genesis also lead to large biases in the TC track distribution.

Different genesis pathways are dominant over different ocean basins, and genesis biases over different ocean basins are thus associated with errors in different aspects of the large-scale and synoptic-scale circulation patterns. Over the western North Pacific, most TCs develop in the monsoon trough environment, and a weaker monsoon trough in the GEFS and the associated biases in the low-level vorticity and midlevel relative humidity lead to the negative bias in TC genesis. Over the eastern North Pacific, the ITCZ breakdown is an important mechanism for tropical cyclogenesis, and TC activity is modulated by the variability of the regional Hadley circulation. The genesis biases in the GEFS can be largely attributed to the erroneous location of the ITCZ over the eastern North Pacific. Over the Atlantic, most

### Table 2. The Pearson correlations of the Niño-3.4 and AMM (July–November mean) indices with the interannual variability of TC counts and ACE in the GEFS and IBTrACS over different basins. The numbers in boldface are the significant coefficients with 95% confidence.

<table>
<thead>
<tr>
<th>Basin</th>
<th>TC counts (Niño-3.4)</th>
<th>ACE (Niño-3.4)</th>
<th>TC counts (AMM)</th>
<th>ACE (AMM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP</td>
<td>0.32</td>
<td>0.36</td>
<td>0.35</td>
<td>0.38</td>
</tr>
<tr>
<td>EP</td>
<td>0.48</td>
<td>0.66</td>
<td>0.75</td>
<td>0.61</td>
</tr>
<tr>
<td>NA</td>
<td>−0.42</td>
<td>−0.66</td>
<td>−0.46</td>
<td>−0.52</td>
</tr>
<tr>
<td>NA</td>
<td>0.74</td>
<td>0.48</td>
<td>0.52</td>
<td>0.72</td>
</tr>
</tbody>
</table>
TCs originate from tropical easterly waves. The GEFS overpredicts the wave activity over West Africa. Compared to the ERAI, the waves in the GEFS are stronger and deeper, and are accompanied by larger diabatic heating around the wave trough axis, which lead to prevailing positive genesis biases (or early TC development) near the coast of West Africa and negative biases farther downstream. The diagnosis of precipitation and column water vapor (CWV) reveals a dry bias in CWV and indicates that precipitation is initiated too early with respect to CWV accumulation in the GEFS. It is suggested that improvement in the cumulus parameterization may reduce the model mean state errors and enhance the TC prediction skill on the regional scale. Although some of the biases presented in this study are likely tied to a certain model resolution and the version of the model physics and dynamics, the result can be used to guide future GFS and GEFS development for better TC forecasts.

The potential skill of GEFS in subseasonal-to-seasonal predictions was evaluated by examining the subseasonal and interannual variability of TC activity in the GEFS. The GEFS is able to depict the interannual variability of TC counts and ACE over the North Pacific and the Atlantic, but shows much poorer skill over the Indian Ocean and the Southern Hemisphere than the other basins. The model skill in ACE tends to be higher than that in TC counts, probably because of the control of TC intensity by the SST. The skill of the model over the North Pacific and the North Atlantic is linked to the impacts of the ENSO and/or the AMM. The correlation
analysis suggests that the GEFS overpredicts the impacts of ENSO over the North Pacific and the North Atlantic but underpredicts the impacts of the AMM over the North Atlantic.

The subseasonal variability of weekly TC days and cyclone energy was also examined over the Atlantic, and the model skill was evaluated based on the correlations between the observed time series and the forecasts in each year during 1985–2012. The mean correlation in the week-1 (week 2) reforecasts is 0.65 (0.44) for the TC days and 0.75 (0.50) for cyclone energy, suggesting that the GEFS captures the active and inactive periods of TC activity reasonably well. On the other hand, the large fluctuations in the correlation coefficients from year to year imply that the intrinsic predictability varies. It is found that the active (inactive) MJO events contribute to the higher (lower) prediction skill of the subseasonal TC variations. In addition, the model tends to have better skill during strong ENSO events, especially during the La Niña years. However, most differences are insignificant, possibly because of the small sample size and the unrealistic representation of these climate modes and their impacts in the GEFS. Further study is needed to examine the role of low-frequency modes in TC predictability on subseasonal and seasonal time scales.

Acknowledgments. The authors thank three anonymous reviewers for their constructive comments on
the manuscript. This work is supported by NOAA’s Research to Operations (R2O) Initiative and ONR’s Unified Parameterization Department Research Initiative (DRI). The SSM/I and SSMIS data are produced by the Remote Sensing Systems and sponsored by the NASA Earth Science MEaSUREs program. The ERA-Interim data were provided by the Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory.

FIG. 13. The mean correlation coefficients of the (left) TC days and (right) WCE with the observations stratified by different climate indices: (a),(b) the VPM indices, (c),(d) the Niño-3.4 index, and (e),(f) the AMM index in July–November during 1985–2012. The red and blue bars indicate the GEFS week-1 and -2 reforecasts, respectively. Plus signs indicate that the difference between an anomalous group and the corresponding neutral group exceeds the 90% confidence level.


