

Subseasonal Tropical Cyclone Genesis Prediction and MJO in the S2S Dataset

CHIA-YING LEE

International Research Institute for Climate and Society, Columbia University, Palisades, New York

SUZANA J. CAMARGO

Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York

FRÉDÉRIC VITART

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

ADAM H. SOBEL

Department of Applied Physics and Applied Mathematics, and Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York

MICHAEL K. TIPPETT

Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York, and Department of Meteorology, King Abdulaziz University, Jeddah, Saudi Arabia

(Manuscript received 9 November 2017, in final form 30 April 2018)

ABSTRACT

Subseasonal probabilistic prediction of tropical cyclone (TC) genesis is investigated here using models from the Seasonal to Subseasonal (S2S) Prediction dataset. Forecasts are produced for basin-wide TC occurrence at weekly temporal resolution. Forecast skill is measured using the Brier skill score relative to a seasonal climatology that varies monthly through the TC season. Skill depends on models' characteristics, lead time, and ensemble prediction design. Most models show skill for week 1 (days 1–7), the period when initialization is important. Among the six S2S models examined here, the European Centre for Medium-Range Weather Forecasts (ECMWF) model has the best performance, with skill in the Atlantic, western North Pacific, eastern North Pacific, and South Pacific at week 2. Similarly, the Australian Bureau of Meteorology (BoM) model is skillful in the western North Pacific, South Pacific, and across northern Australia at week 2. The Madden-Julian oscillation (MJO) modulates observed TC genesis, and there is a relationship, across models and lead times, between models' skill scores and their ability to accurately represent the MJO and the MJO–TC relation. Additionally, a model's TC climatology also influences its performance in subseasonal prediction. The dependence of the skill score on the simulated climatology, MJO, and MJO–TC relationship, however, varies from one basin to another. Skill scores increase with the ensemble size, as found in previous weather and seasonal prediction studies.

1. Introduction

The Madden–Julian oscillation (MJO; Madden and Julian 1972) modulates tropical cyclone (TC) activity globally. The probability of TC genesis is typically greater during or after a strong convective MJO phase than at other times (Camargo et al. 2009; Klotzbach 2014; Klotzbach and Oliver 2015a). In the Atlantic

(Mo 2000; Maloney and Hartmann 2000b; Klotzbach 2010; Klotzbach and Oliver 2015b), the enhanced storm genesis occurs when a strong MJO is active in the Indian Ocean. In contrast, in the eastern North Pacific, enhanced TC genesis occurs when a strong MJO is active in the central and eastern North Pacific (Molinari et al. 1997; Maloney and Hartmann 2000a, 2001; Aiyyer and Molinari 2008). Similarly, in the western North Pacific (Nakazawa 1988; Liebmann et al. 1994; Sobel and Maloney 2000; Kim et al. 2008; Li and Zhou 2013),

Corresponding author: Chia-Ying Lee, clee@iri.columbia.edu

DOI: 10.1175/WAF-D-17-0165.1

© 2018 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

North (Nakazawa 1988; Liebmann et al. 1994; Kikuchi and Wang 2010; Krishnamohan et al. 2012) and South (Bessafi and Wheeler 2006; Ho et al. 2006) Indian Ocean, and South Pacific (Hall et al. 2001), the number of storms increases when the MJO is active in these basins. Additionally, typhoon tracks shift eastward when the convective MJO is active in the Indian Ocean and shift westward in the western Pacific (Kim et al. 2008). Rapidly intensifying storms are more frequent in the Atlantic when the MJO is active in the Indian Ocean (Klotzbach 2012). A strong active MJO increases local values of an empirical TC genesis index (Camargo et al. 2009) through systematic enhanced low-level absolute vorticity and increased midlevel relative humidity.

In these observational studies, it is often stated that accurate predictions of the MJO and knowledge of the MJO–TC relationship offer the potential for forecasts of the probability of TC genesis with a few weeks lead time. Regional statistical models for subseasonal TC prediction have in fact been developed (Leroy and Wheeler 2008; Slade and Maloney 2013) using MJO indices, as well as other environmental parameters. When an MJO index is added as one of the predictors, there is a significant, albeit small, improvement of skill at leads up to 3 weeks. For longer leads, the forecast skill is thought to be primarily from the climatological seasonal cycle and interannual variability. Reforecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) also suggest that the accuracy of the MJO prediction has a significant impact on the predicted TC frequency (Vitart 2009). Compared to Southern Hemisphere TC statistical forecasts (Leroy and Wheeler 2008), the ECMWF model has greater skill in predicting TC occurrence at week 1, while the statistical model performs better for longer leads (Vitart et al. 2010). Furthermore, the ECMWF skill in predicting Atlantic hurricane activity is sensitive to the MJO phase and amplitude at the time of the model initialization (Belanger et al. 2010).

With the increasing demand for forecasts on the time scale between weather and seasonal–interannual—the so-called subseasonal time scale—an international effort was initiated to improve and develop various aspects of dynamical subseasonal predictions, including subseasonal TC forecasts. A key goal of these efforts is improved understanding of the factors that affect forecast prediction skill. The multimodel Seasonal to Subseasonal Prediction (S2S; Vitart et al. 2017) dataset, containing extensive reforecasts with lead times up to 60 days, is ideal for this task. In this study, we focus on the subseasonal prediction of TC genesis in the S2S reforecasts. While the ability of global models to simulate the MJO–TC modulation (Vitart 2009; Satoh et al. 2012;

Kim et al. 2014; Murakami et al. 2015; Xiang et al. 2015) and the prediction skill of TC genesis prediction on subseasonal time scales have been analyzed (Belanger et al. 2010; Elsberry et al. 2011; Tsai et al. 2013; Elsberry et al. 2014; Nakano et al. 2015; Barnston et al. 2015; Yamaguchi et al. 2015; Li et al. 2016; Camp et al. 2018) for various models, this is the first comprehensive multimodel, multiyear analysis of reforecasts of TC genesis prediction on subseasonal time scales.

Here, we use the S2S data to construct probabilistic and deterministic reforecasts of the of basin-wide TC occurrence with weekly temporal resolution. The prediction skill is evaluated using the mean square error skill score and the Heidke skill score for deterministic forecasts, as well as the Brier skill score for probabilistic forecasts. Reforecasts, observations, skill scores, and other analysis methods are described in section 2. We then discuss TC climatology in the reforecasts in section 3 to define the tropical storm thresholds and seasonality for prediction skill evaluation. The simulated and observed MJO modulation of TC genesis is examined in section 4. Then, we analyze the prediction skill as well as the potential predictability in section 5. Connections between the skill scores and the model characteristics, the initialization, and the ensemble prediction system design are examined in section 6. Results are then summarized in section 7.

2. Data and methods

a. S2S reforecasts

Table 1 shows some basic characteristics of the S2S reforecasts used here. They are obtained from coupled, global general circulation models run by six operational centers: the Australian Bureau of Meteorology (BoM), the China Meteorological Administration (CMA), the ECMWF, the Japan Meteorological Agency (JMA), the Météo-France/Centre National de Recherche Météorologiques (MetFr), and the National Centers for Environmental Prediction (NCEP). The first ensemble member is the unperturbed control run. Note that because the designs of the ensemble prediction systems (specifically, the frequency of forecasts and ensemble size) differ among these agencies, the reforecasts are heterogeneous. We treat such differences in system design as additional factors contributing to prediction skill. Another heterogeneous feature is that the reforecast periods differ. While this might affect the comparison, we do not think it is likely to qualitatively change the relative skill of the forecast systems. Further details of the S2S dataset are described by Vitart et al. (2017). All the S2S reforecasts are archived on a $1.5^\circ \times 1.5^\circ$ grid at daily resolution.

TABLE 1. Characteristics of the six S2S reforecasts used in this paper.

Model	Forecast time (days)	Resolution	Period	Ensemble size	Frequency and sample size
BoM	0–64	2°, L17	1981–2013	33	~5 days, 2160
CMA	0–61	1°, L40	1994–2014	4	Daily, 7665
ECMWF	0–46	0.25° for first 10 days; 0.5° after day 10, L91	1994–2014	11	~4 days, 2058
JMA	0–33	0.5°, L60	1981–2010	5	~10 days, 1079
MetFr	0–61	~0.7°, L91	1993–2014	15	~15 days, 528
NCEP	0–44	~1°, L64	1999–2010	4	Daily, 4380

b. TCs in the S2S models and observations

To track TCs in the S2S models, we employ the tracker from Vitart and Stockdale (2001). The tracker defines a storm center at a local minimum sea level pressure where 1) a local vorticity maximum ($>3.5 \times 10^{-5} \text{ s}^{-1}$) at 850 hPa is nearby, 2) a local maximum in the vertically averaged temperature (warm core, $>0.5^\circ\text{C}$) in between 250 and 500 hPa is within a distance (in any direction) equivalent to 2° latitude, 3) the two locations detected from criteria 1 and 2 above are within a distance equivalent to 8° latitude, and 4) a local maximum thickness between 1000 and 200 hPa can be identified within a distance equivalent to 2° latitude. Additionally, a detected storm must last at least 2 days to be included in our analysis.

In general, the criteria used in a tracker should vary with model resolution (Walsh et al. 2007; Camargo 2013). TC detection, however, is very sensitive to the input thresholds (Horn et al. 2014; Zarzycki and Ullrich 2017); changing criteria without a thorough investigation could potentially introduce artifacts into the results. Furthermore, all the S2S data are archived on a common grid. Therefore, in this study, we use the same criteria (as described above) for all models. A potential impact due to the interpolation of the atmospheric fields from a high-resolution model output to a low-resolution common grid is that it might reduce the strength of the warm core, the vorticity, and the pressure. The vorticity might be most strongly affected. Because the criteria used here were set for a low-resolution model, the impact on the weakening of these fields is not expected to affect the number of the detected TCs. The S2S TC tracks contain daily values of maximum sustained winds and storm locations.

Observations of tropical cyclone tracks are derived from the HURDAT2 dataset, produced by the National Hurricane Center (NHC; Landsea and Franklin 2013) and from the Joint Typhoon Warning Center (JTWC; Chu et al. 2002). Both best-track datasets include 1-min maximum sustained wind, minimum sea level pressure (not used in this study), and storm location every 6 h.

Following the conventional definitions, the TC basins are the Atlantic (ATL), northern Indian Ocean (NI), western North Pacific (WNP), eastern North Pacific (ENP), southern Indian Ocean (SIN, 0° – 90°E), northern Australia (AUS, 90° – 160°E), and southern Pacific (SPC, east of 160°E).

c. MJO definition

The real-time multivariate MJO indices (RMM1 and RMM2) are calculated using zonal winds at 200 and 850 hPa and outgoing longwave radiation (Gottschalck et al. 2010; Wheeler and Hendon 2004; Vitart 2017). Observational RMM indices are calculated using ERA-Interim reanalysis data.

d. Skill scores

A skill score is an index that measures the model prediction skill relative to a reference value. Three skill scores are used here: mean-square error (MSE) skill score (MSESS), Heidke skill score (HSS), and Brier skill score (BSS). MSESS and HSS are for evaluating deterministic forecasts while BSS is for probabilistic predictions.

MSESS is applied to the predicted storm numbers and is defined as

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2 \quad \text{and} \quad (1)$$

$$\text{MSESS} = 1 - \frac{\text{MSE}}{\text{MSE}_{\text{ref}}}, \quad (2)$$

where N is the total number of forecasts, P_i is the predicted genesis number for the i th forecast, and O_i is the i th observation. MSE_{ref} is the MSE of a reference based on observed climatology. MSESS larger than 0 means the model has higher skill than the climatological reference.

HSS compares the proportion of correct categorical forecasts to that which would be expected by random forecasts that are statistically independent of the observations. We use two categories here: 0 for no genesis and 1 for one or more storms forming during the forecast period. The

ratio of the number of correct forecasts to that of all forecasts, commonly called percent correct (PC), is defined as

$$\text{PC} = \frac{a + d}{a + b + c + d} = \frac{a + d}{n}, \quad (3)$$

where a represents the frequency of observed geneses that are correctly forecast, b is the frequency of false alarms, c represents the observed geneses that are not forecast, and d are cases that were neither forecast nor occurred.

The marginal probability of a 1 forecast is $(a + b)/n$ and that for a 1 observation is $(a + c)/n$. Thus, the probability of a correct 1 forecast by chance is $(a + b)/n \times (a + c)/n$. Similarly, the probability of having a correct 0 forecast by chance is $(b + d)/n \times (c + d)/n$. Thus, the probability E of a correct forecast due to chance is

$$E = \left(\frac{a + d}{n}\right) \left(\frac{a + c}{n}\right) + \left(\frac{b + d}{n}\right) \left(\frac{c + d}{n}\right). \quad (4)$$

HSS is therefore defined as

$$\text{HSS} = \frac{\text{PC} - E}{1 - E} = \frac{2(ad - bc)}{(a + c)(c + d) + (a + b)(b + d)}. \quad (5)$$

The HSS is 1 if all forecasts were correct (i.e., when PC equals 1) and is 0 if the model has no skill (i.e., PC equals E).

BSS is used to assess the skill of a probabilistic forecast of basin-wide TC occurrence relative to a climatological forecast. The Brier skill (BS) is defined as

$$\text{BS} = \frac{1}{N} \sum_{i=1}^N (p_i - o_i)^2 \quad \text{and} \quad (6)$$

$$\text{BSS} = 1 - \frac{\text{BS}}{\text{BS}_{\text{ref}}}, \quad (7)$$

where N is the total number of forecasts, o_i is the i th observation, and p_i is the predicted probability of TC occurrence for the i th forecast, defined as

$$p_i = \frac{1}{M} \sum_{j=1}^M P_{i,j}. \quad (8)$$

In Eq. (8), M is the number of ensembles, $P_{i,j}$ is the genesis prediction from the j th ensemble member for the i th forecast. Both $P_{i,j}$ and o_i are 0 for no genesis and are 1 for one or more occurrences of storm genesis during the forecast period. Thus, BS is the mean-square probability forecast error. The BS_{ref} is similar to BS but for a reference forecast based on the observed climatology. In

this study, two climatologies are used. One is the seasonally varying climatology at a monthly time resolution, while the other is a constant, annual mean climatology. When a model is skillful compared to the climatology, the BSS is positive.

e. Candy plot analysis

To analyze the dependence of TC genesis on MJO phases, the probability density function (PDF) of storm genesis is calculated in each TC basin and binned by MJO phase. To identify favorable and unfavorable MJO phases, values of the TC number in each week are randomly shuffled in time throughout the entire period to obtain PDFs independent of the MJO. The favorable MJO phases are then defined when the unshuffled PDF is larger than the 90th percentile of the 4000 PDFs obtained from randomly swapping the data, and the unfavorable MJO phases are defined when the unshuffled PDF is less than the 10th percentile of the randomized PDFs. The PDFs are then organized by the longitude along the Y axis from ATL to ENP and by the MJO phases along the X axis, like a sheet of candies.

The candy plot analysis is conducted for observations and the six S2S models. The comparison of the global pattern (consisting of PDFs from all TC basins) between each S2S model and the observed pattern is then quantified using r^2 (i.e., the fraction of the variance of the observed PDFs that is predicted by the model PDFs). In this analysis, we include only storms when the magnitude of the MJO index is larger than one standard deviation. The fractions of storms we used are 60% in the observations, and roughly 60%, 50%, 40%, 52%, 65%, and 52% in BoM, CMA, ECMWF, JMA, MetFr, and NCEP reforecasts, respectively.

3. Tropical cyclone climatology

a. Intensity and tropical storm threshold

Since their horizontal resolutions are inadequate to represent the TC inner-core structure, the global models used here are not able to simulate the highest observed TC intensities. Another factor that impacts the simulated intensities as represented in the S2S archive is that the model outputs are instantaneously archived on a $1.5^\circ \times 1.5^\circ$ grid in the S2S database every 24 h. As a result, the cumulative density distribution (CDF) of TCs' lifetime maximum intensity (LMI) shows that the median LMI for the observed storms is 50 kt ($1 \text{ kt} = 0.51 \text{ m s}^{-1}$) while it is in the range of 25–35 kt for the S2S models, except for BoM, which has a median LMI of 40 kt (Fig. 1). The BoM model is able to simulate stronger storms than other S2S models that have higher horizontal resolutions. This could be due to its physical

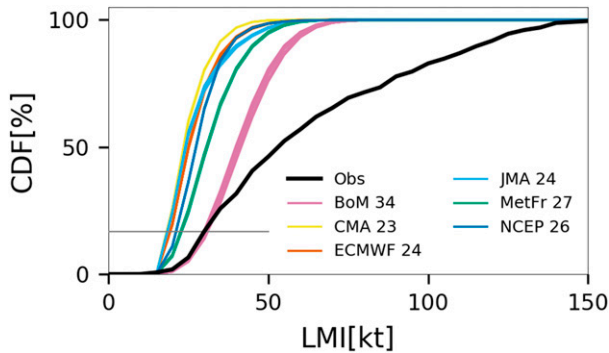


FIG. 1. CDFs (%) of the observed (black) and the simulated (colored) LMIs from each of the ensemble members from the six S2S models.

parameterizations, dynamical cores, or both. Multiple studies with other global climate models have noted that both factors are important, so that the maximum TC intensities simulated by a model are not a simple function of that model's horizontal resolution (Vitart et al. 2001; Murakami et al. 2012; Zhao et al. 2012; Reed et al. 2015; Duvel et al. 2017; Kim et al. 2018). Although there is a significant low bias in the S2S TC intensities, we can categorize storms using quantile analysis (Camargo and Barnston 2009). For example, the observed tropical storm (TS) wind speed threshold is 34 kt, which in the observed LMI distribution corresponds to the 18th percentile (gray line in Fig. 1). Thus, we define the tropical storm threshold as the 18th percentile of the LMI CDF in each model. These are 34, 23, 24, 24, 27, and 26 kt, respectively, for the BoM, CMA, ECMWF, JMA, MetFr, and NCEP models. In this study, we only consider TCs that reach the tropical storm threshold thus defined in observations and the model reforecasts.

b. Genesis and TC season definition

Genesis time is defined here as the time of the first point recorded on each forecast track. TCs that exist prior to the model initialization time have already undergone genesis. Nevertheless, for purposes of model evaluation, we refer to the first recording time (usually day 1, or $t = 24$ h) of the preexisting storms as their genesis time. The reason for including preexisting storms in our analysis will be discussed in section 5. With the exception of week 1—when there is a higher TC occurrence because of the preexisting storms—the forecast genesis climatology does not change much with lead time. Therefore, while we only show here the genesis climatology for week 2 forecasts (Fig. 2), our results are also valid at longer lead times.

Globally, the ECMWF model that is statistically significant generates 20% more TCs than observed, while CMA,

MetFr, and NCEP have genesis rates 140%, 65%, and 80% higher than observed, respectively. In contrast, the BoM and JMA models generate 35% and 45% fewer TCs than are present in the observed climatology. Low-resolution models often have unrealistically high TC genesis rates in the subtropics, when storms are detected and tracked using algorithms with model-dependent thresholds (Camargo 2013). However, the difference maps between simulated and observed genesis counts (Figs. 2b–g) suggest that this is not the case for the S2S models, since the errors in the subtropics are much smaller than those in the tropical belt (30°S – 30°N).

Regionally, the strongest observed local maxima of the TC genesis rate occur in the ENP and WNP (Fig. 2a). In the three Southern Hemisphere basins (SIN, AUS, and SPC), the observed storms form in an elongated area around 15°S . In general, the S2S models are able to capture these local maxima (not shown), and the ECMWF model has the smallest regional biases, followed by the BoM model. The JMA model underestimates the rate of TC genesis everywhere, while CMA, MetFr, and NCEP overestimate it. In the individual basins, the models that have the smallest mean bias (in number of storms per year) are the MetFr model (-0.01) for the ATL, the BoM (0.22) for the NI, the ECMWF (0.11) for the WNP, the MetFr (-0.24) for the ENP, the JMA (-0.3) for the SIN, the BoM (-0.6) for the AUS, and both the ECMWF (0.33) and BoM (0.36) for the SPC. The CMA model has the largest positive bias in the Pacific Ocean (Fig. 2c), with more than 1 storm per year per grid ($4^{\circ} \times 4^{\circ}$) between 4° and 12°N , compared to the observed climatological mean of less than 0.2 storms per year per grid (Fig. 2a).

Despite these biases in the total TC counts and genesis spatial distribution, the S2S models represent the annual cycle of TC genesis reasonably well (Fig. 3). We define regionally varying TC seasons that consist of the months with genesis rates higher than 5% of the annual genesis rate in each region. Using this definition, the TC seasons in some models are slightly different than in the observations. For example, the observed hurricane season in the ENP is defined as May–October, but it is from July to December in the BoM model. The simulated TC seasons in the ECMWF model best match those from the observations.

Although there are differences between simulated and observed TC seasons, our goal is to have skillful TC predictions during the observed TC seasons. Therefore, the observed TC seasons are used in our prediction skill evaluation below. Using this definition, the TC seasons are June–November in the ATL, May–October in the ENP, May–December in the WNP, April–June and September–December in the NI, October–April in the SIN, and November–April in the AUS and SPC.

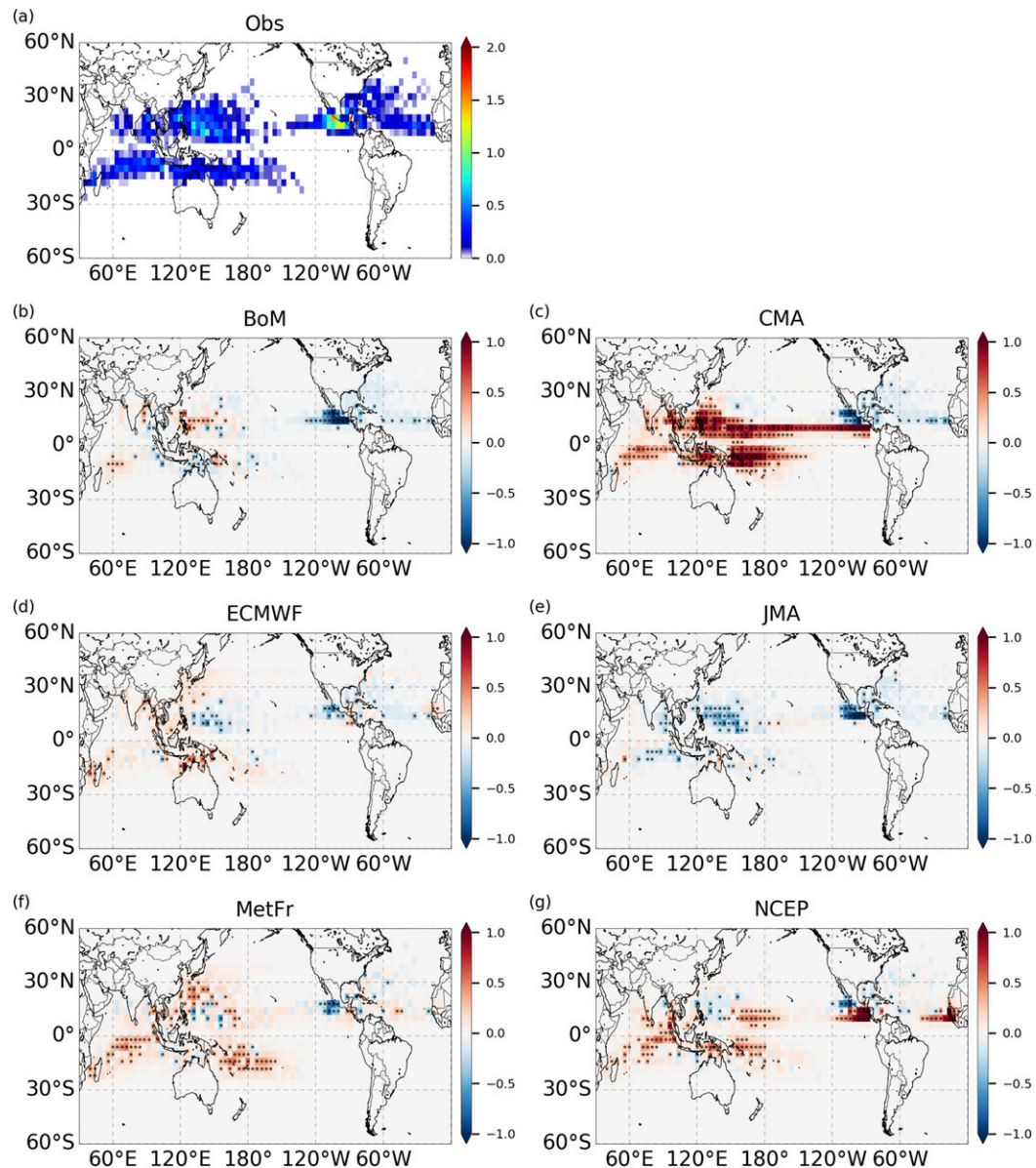


FIG. 2. (a) The observed genesis density in storm numbers per $4^\circ \times 4^\circ$ per year. (b)–(g) Week 2 ensemble mean genesis density biases in the BoM, CMA, ECMWF, JMA, MetFr, and NCEP models. The black dots indicate where the error is larger than one standard deviation of the natural variability.

4. MJO–TC modulation

Next, we examine whether the S2S models are capable of simulating the observed MJO–TC modulation. Specifically, we refer to the spatial distribution of genesis as a function of the MJO phase, such as the anomalous fields in Figs. 4 and 5, and the basin-wide PDFs in Fig. 6. Our focus is on the week 2 reforecasts, when all S2S models show skill in predicting the MJO (Vitart 2017).

Global climate models are able to simulate the observed dependence of TC genesis on the MJO

(Nakazawa 1988; Liebmann et al. 1994; Mo 2000; Maloney and Hartmann 2000b,a; Kim et al. 2008; Li and Zhou 2013; Krishnamohan et al. 2012; Bessafi and Wheeler 2006). High horizontal resolution is often cited as a necessary condition to capture the MJO–TC modulation (e.g., Zhang 2013; Camargo and Wing 2016). However, the necessary resolution is not precisely defined. For example, the high horizontal resolution Zhang (2013) and Camargo and Wing (2016) referred to varies from 50 km for the EMCWF (Vitart 2009) and the Geophysical Fluid Dynamics Laboratory (GFDL) High

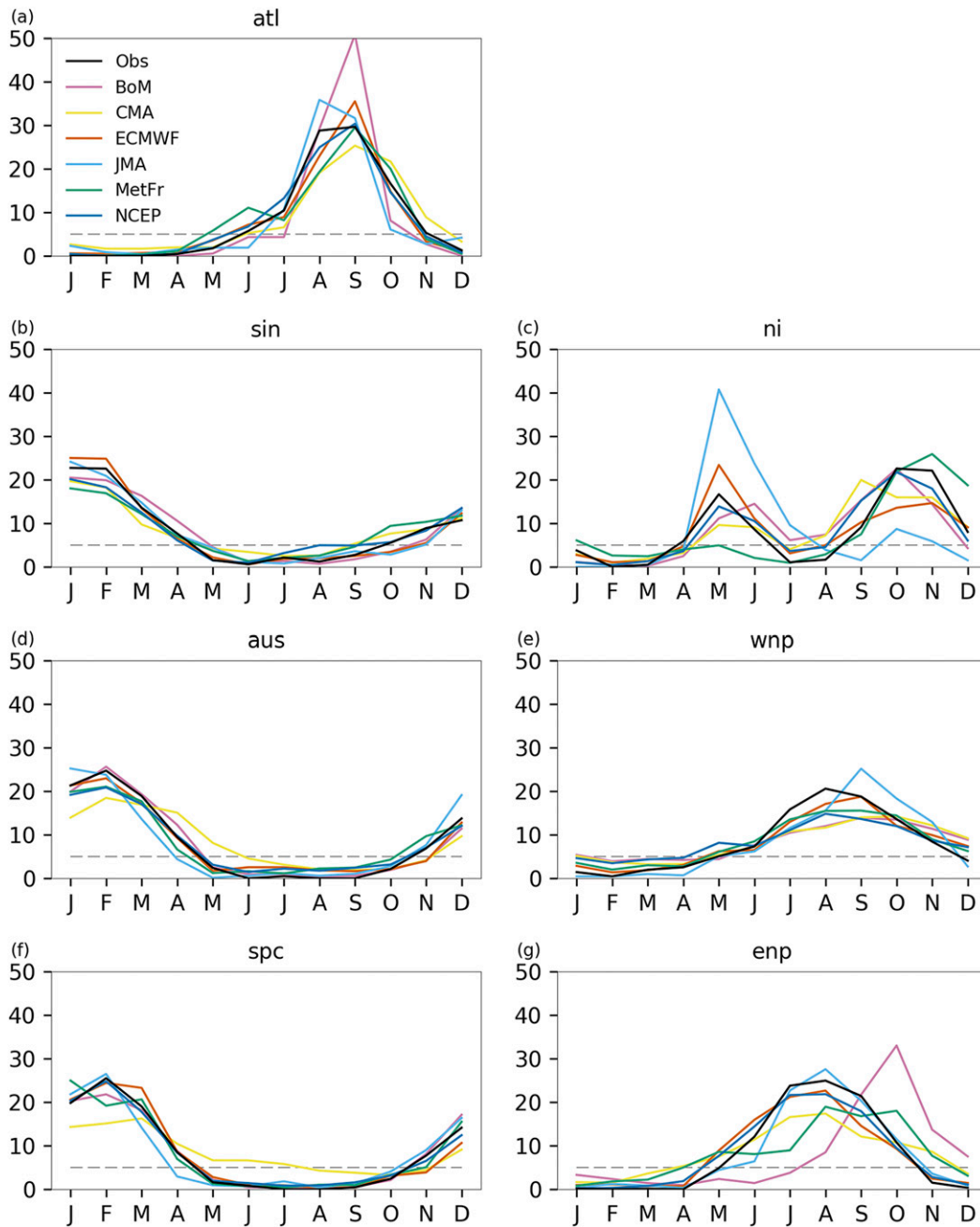


FIG. 3. Observed (black) and simulated (colored) ensemble mean seasonality (%). At each basin, the months when the genesis rate is larger than 5% (gray dashed lines) of the observed annual genesis rate are defined as the TC season.

Resolution Atmospheric Model (HiRAM; Jiang et al. 2012) to 14 km for the Japanese Nonhydrostatic Icosahedral Atmospheric Model (NICAM; Oouchi et al. 2009). Furthermore, good representations of convection and microphysics are key elements in order to simulate well the MJO in global models (Kim et al. 2012, 2014; Holloway et al. 2013; Kang et al. 2016). Changing the

horizontal resolution (with the same physical packages for the same model) does not necessarily improve the MJO simulation (Jia et al. 2008; Holloway et al. 2013; Hung et al. 2013). Horizontal resolution might be important primarily through its influence on the models' ability to simulate TCs and their interaction with the ambient environment (Kim et al. 2018).

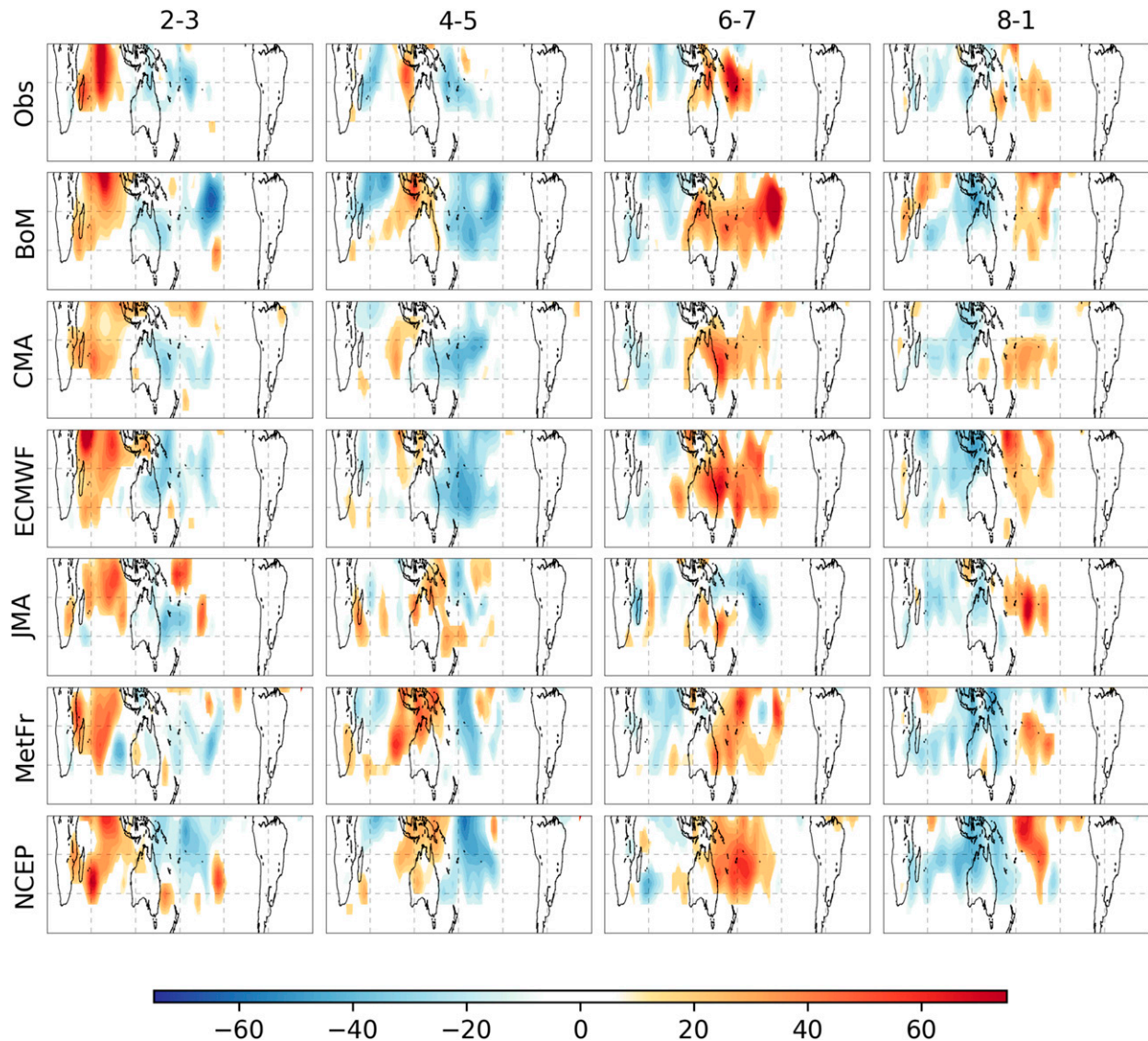


FIG. 4. Ensemble mean genesis anomalies (%) at every two MJO phases from 2–3, 4–5, 6–7, and 8–1 in the Southern Hemisphere from the observations and the week 2 forecasts from six S2S models.

The S2S models have horizontal grid spacings from 0.25° to 2° . All of them are able to capture, at least qualitatively, the observed eastward propagation of TC genesis anomalies in the Southern Hemisphere with increasing MJO phase (Fig. 4). While the observed eastward-propagating signal is weaker in the Northern Hemisphere (Fig. 5), we can still see positive TC genesis anomalies propagating from the NI to the ATL (i.e., from MJO phases 2–3 to 8–1). The Northern Hemisphere eastward propagation is stronger in most of the S2S models than in the observations. The observed positive anomalies in the ENP for MJO phases 6 and 7 are too strong and expand toward the WNP for the ECMWF and MetFr models. Similarly, the WNP

anomalies for MJO phases 8–1 are overpredicted and expand toward the ENP for the BoM, CMA, ECMWF, and NCEP models. The JMA model is not able to capture the MJO eastward propagation in the Northern Hemisphere.

To further identify the MJO phases that are favorable for TC genesis in individual basins, we perform a candy plot analysis (section 2), which shows the storm genesis rate binned by MJO phase in each TC basin. In the observations (Fig. 6a), 31% of observed ATL hurricanes form when the MJO convection center is in the Indian Ocean (MJO phases 2–3). Similarly, in the SIN, a higher rate of genesis (50% of all storms) occurs during the MJO phases 2–3. The favorable MJO phases

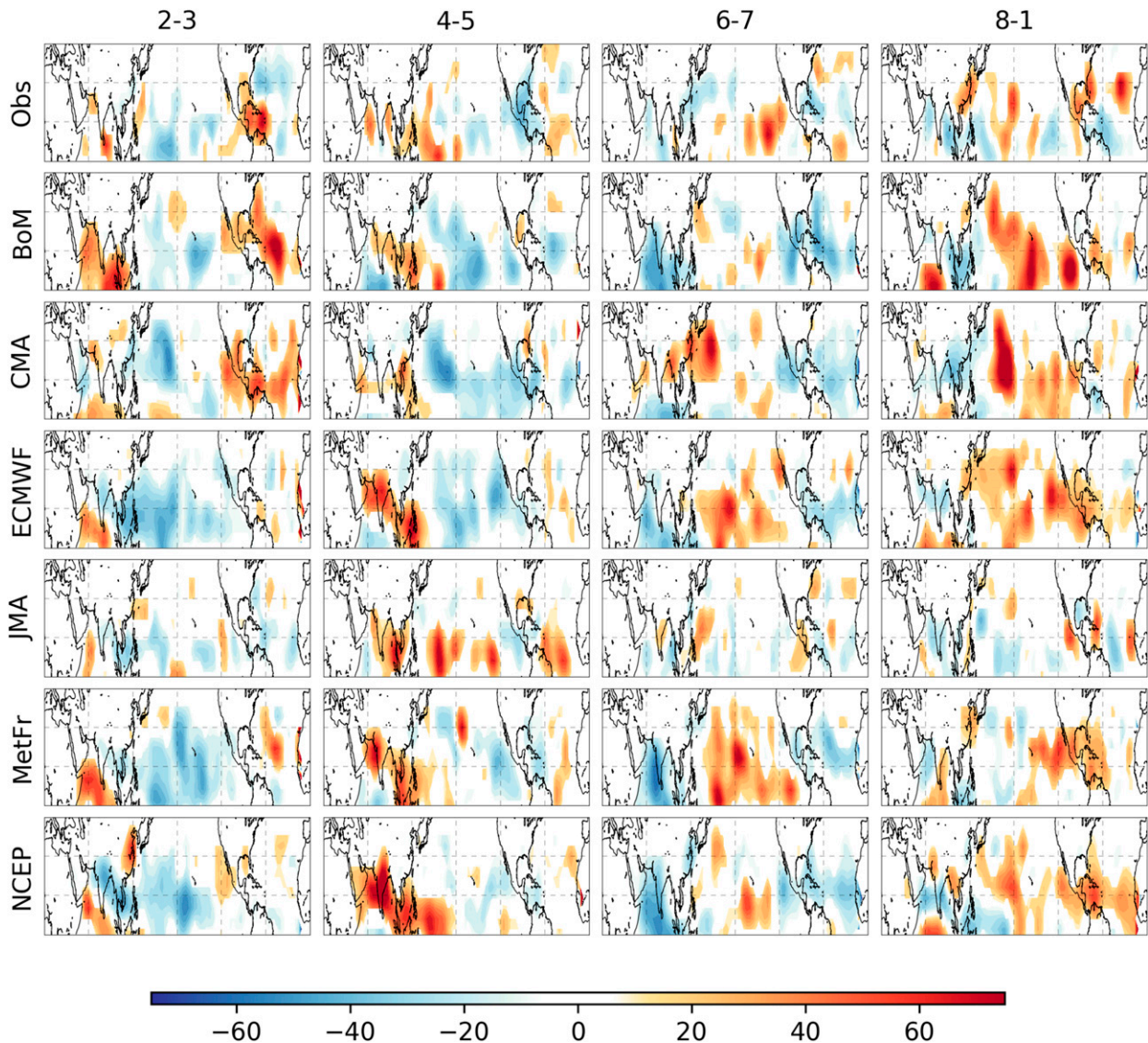


FIG. 5. As in Fig. 4, but for the Northern Hemisphere.

for TC occurrence are 3–5 for the NI, 3–4 for the AUS, 5–6, for the WNP, 7–8 for SPC, and 7 and 8–1 for the ENP. The favorable MJO phases are listed by basin with increasing longitudes so that they line up from the bottom-left corner to top-right corner in Fig. 6a.

The favorable MJO phases in the S2S models also show a bottom-left to top-right trend, with the exception of the JMA model. The ECMWF model (with horizontal resolutions of 0.25° – 0.5°) best simulates the observed MJO–TC pattern (Fig. 6b), and explains 64% (r^2) of the observed variance. The MetFr, NCEP, CMA, and BoM models (in order of model resolution from 0.7° to 1°) explain 42%, 50%, 41%, and 47% of the variance, respectively. The JMA model does not capture the upward trend of the MJO–TC pattern, because MJO

phases 5 and 6 occur much more frequently in the model than in the observations, especially during the Northern Hemisphere TC seasons (Fig. 6e). As a result, despite having 0.5° horizontal grid spacing, the JMA model explains only 23% of the observed MJO–TC relationship. Our results suggest that while model resolution plays an important role in simulating TCs, it is probably not the most important factor for simulating the observed modulation of TC genesis by the MJO.

5. Genesis forecast skill

On the scale of weather forecasting (2–5 days), deterministic prediction of TC occurrence is often used. Beyond 5 days, probabilistic forecasts from ensemble

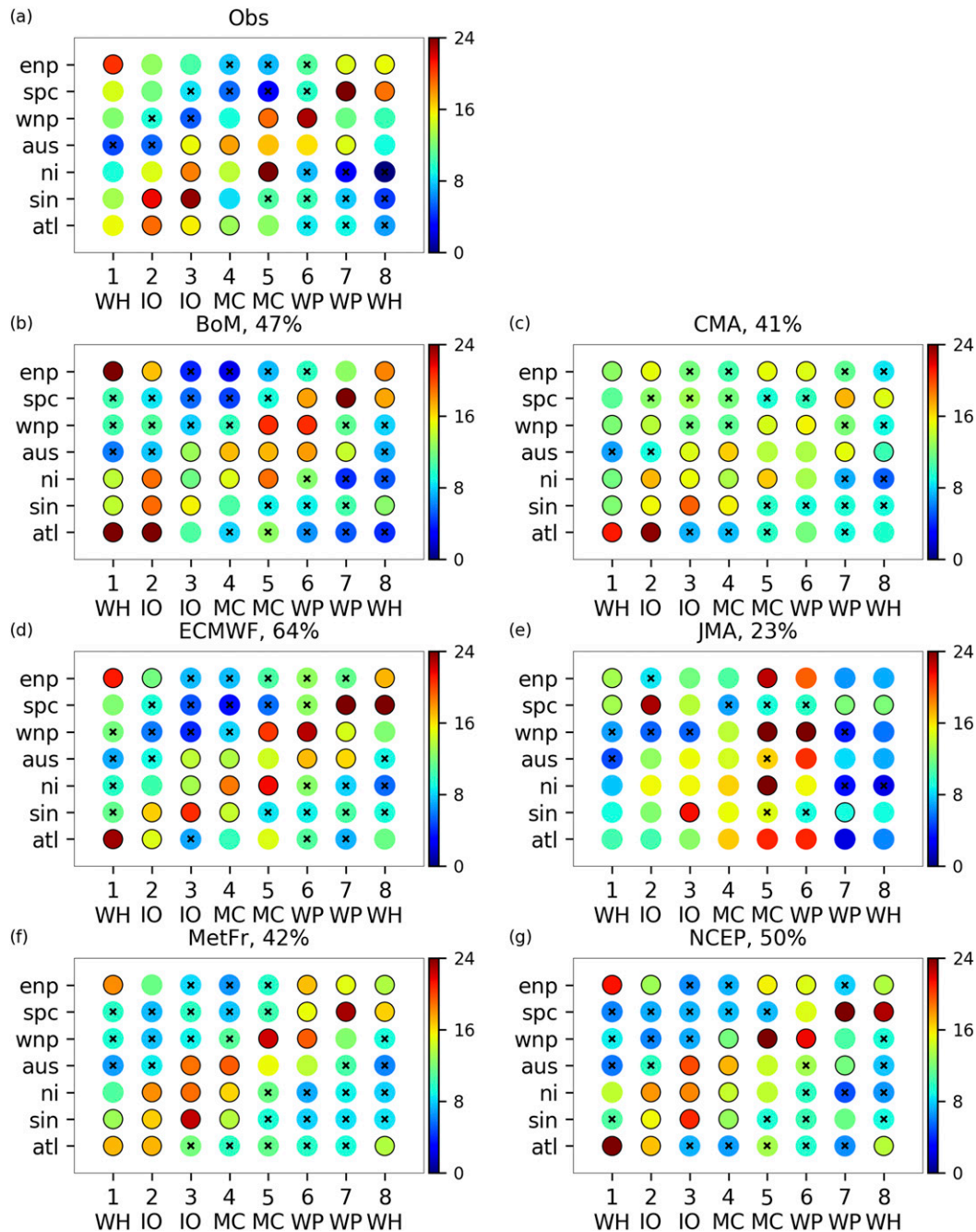


FIG. 6. Candy plot for the MJO–TC relationship in the observations and from six S2S models from week 2 forecasts. The color of each candy indicates the PDF (%) in the corresponding MJO phase in the basin. The sum of the circles across the MJO phases in each basin is 100%. The black circle at the edge indicates the value is above the 90th percentile while the cross symbol (×) at the center means the value is below the 10th percentile. In the title of each subplot from the simulations, we label the r^2 value, which represents the fraction of the observed pattern explained by the model simulation.

systems are used because they provide information on the uncertainty of the forecasts (Elsberry et al. 2014). Furthermore, the prediction skill from the mean of an ensemble might be higher than that of a single

“deterministic” run even if the latter has higher resolution. In this section, we will show the models’ skill in predicting TC genesis in both deterministic and probabilistic forecasts. Our focus is the probabilistic

prediction skill, and we will also discuss its potential predictability.

a. Deterministic prediction: MESS and HSS

The deterministic TC occurrence prediction is defined using the ensemble mean forecasts here. The deterministic prediction skill is quantified using MESS and HSS (section 2). HSS measures the S2S models' ability to forecast storm occurrence (regardless of number) while MESS evaluates not only occurrence but also the number of the storms.

Comparing to the random predictions, the S2S models are more skillful at predicting storm occurrence within a basin at most leads (Fig. 7). For week 1 occurrence prediction, the ECMWF model has the highest values of the HSS in all basins except in the ATL basin where the NCEP model has higher HSS. After week 1, the values of HSS in most of the models drop significantly, and it is hard to distinguish among them. Note that HSS compares the ratio of correct forecasts from the S2S models to that expected by chance, without taking into account the climatology. The values of HSS alone do not tell if the S2S models are more skillful than climatology.

Therefore, we create a no-skill reference forecast that shows the skill score from knowing only the seasonal climatology (dashed line in Fig. 7). The no-skill reference is calculated by verifying the S2S predictions against observations that are shuffled by year. For example, the S2S predictions from 2005 are evaluated using observations from a randomly selected year. Doing so, we keep the seasonality in the shuffled observations but remove the year-to-year dependence. With a few exceptions, the HSS values of the no-skill references are positive. This is because the S2S models simulate the observed genesis seasonality reasonably well, as we discussed earlier in the section 3. In the cases when the value of the model HSS is close to the no-skill reference, the climatology contributes to most of the prediction skill. In Fig. 7, the solid lines merge to the dashed lines for most of the models after week 2, except for the ECMWF model. For the BoM and CMA models in the Atlantic basin, even the week 1 prediction skill is largely due to knowing the seasonal climatology. In the ECMWF model, the HSS is larger in all basins than the corresponding no-skill reference, indicating that there are additional factors that contribute to the deterministic prediction skill.

Analyses using MESS suggest that the S2S models are not skillful at predicting TC frequency except for week 1 forecasts from the ECMWF model in the NI, AUS, WNP, and ENP and from the BoM model in the SPC and AUS (not shown).

b. Probabilistic prediction: BSS

Next, we investigate the performance of the S2S models in predicting the probability of weekly TC occurrence. Two different Brier skill scores are calculated here. The first (called BSS_c, where "c" stand for "constant"; dashed lines in Fig. 8) compares the Brier score of the forecast to that of a constant, the observed annual mean climatology. Positive values of BSS_c mean that the model is more skillful than a constant climatological prediction. We consider storms with geneses in all months, and with this score, forecasts receive credit for correctly matching the annual cycle in TC genesis frequency. This is consistent with standard practice in the verification of short-term weather forecasts, in which the total values of meteorological variables, as opposed to anomalies from a seasonal climatology, are verified against observations.

The second Brier skill score (called BSS without a subscript here; solid lines in Fig. 8) compares the Brier skill of the forecasts to that of an observed monthly varying climatology. In this case, only storms that formed during the observed TC seasons are considered (as defined in section 3). This is more typical in seasonal predictions and provides a stricter measure for evaluating TC genesis prediction skill. S2S models achieve positive BSS values when they capture deviations from the observed seasonality. A positive BSS in this case means that the model is more skillful than the monthly varying climatological prediction.

The values of BSS_c are often positive, indicating that S2S models are more skillful than an annually constant forecast in most of the TC basins. The BSS_c values are also often noticeably greater than those of BSS because the reference forecast used for BSS_c is less skillful than that of BSS. Forecasts from the CMA and those for the NI are exceptions. In the CMA forecasts, the values of BSS_c are much closer to those of BSS in SIN, AUS, and WNP. This is because the CMA model's climatology is poor, with too many storms forming during the observed off season (Fig. 3). The differences between BSS_c and BSS in the NI are much smaller than in other basins, and there are no positive BSS_c values after week 1. In other words, none of the S2S models has more skill in predicting the TC season than an annually constant forecast after week 1 in the NI. In contrast, JMA, ECMWF, NCEP, and MetFr all have positive BSS_c in the ATL up to week 5, although ECMWF is the only model that has a positive BSS after week 1.

The BSS values for all models drop significantly from week 1 to week 2 (Fig. 8), similar to results shown for HSS (Fig. 7). This large drop is connected to our genesis definition by accounting for preexisting storms at the

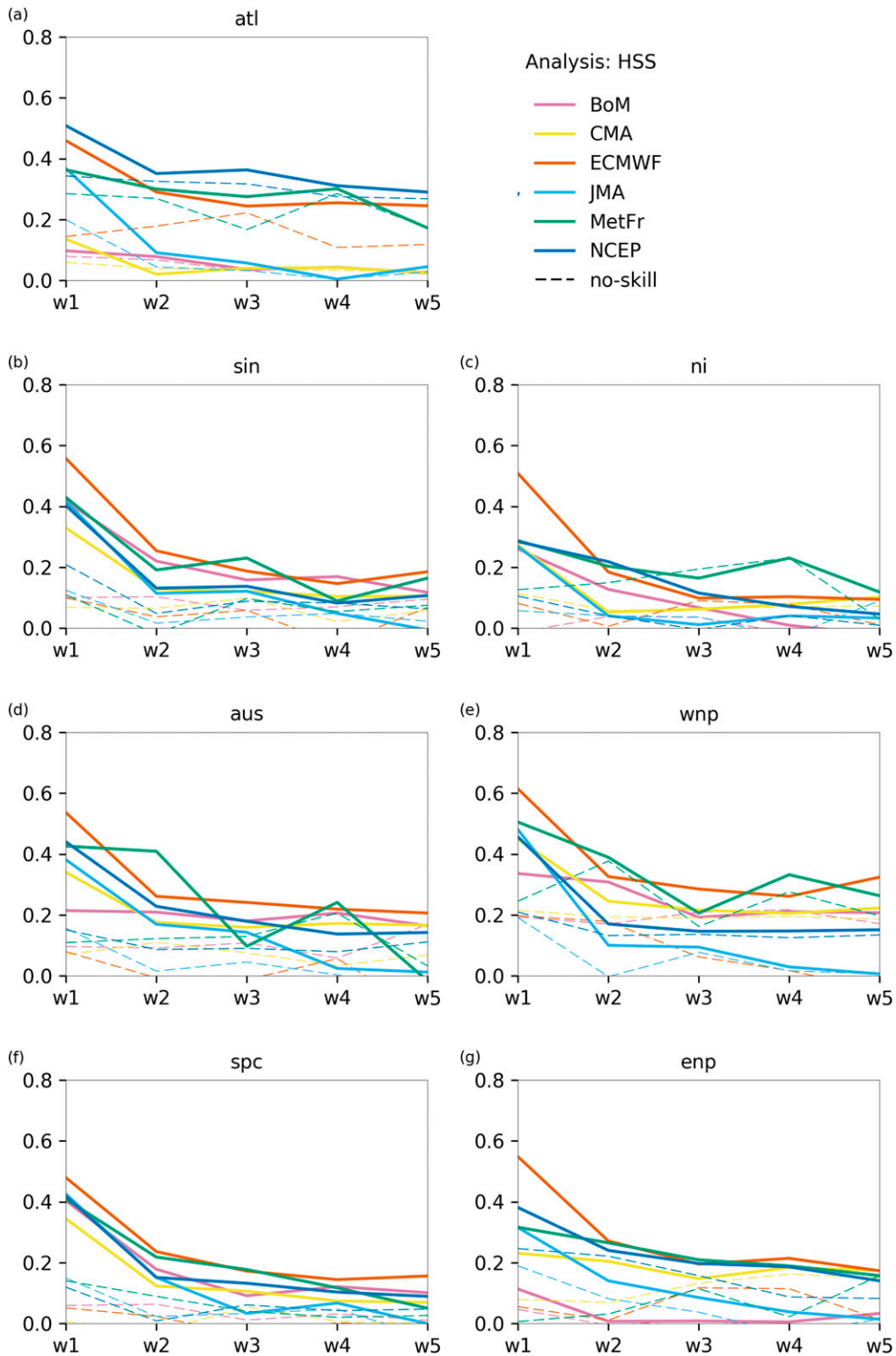


FIG. 7. Weekly, basin-wide HSS (solid lines) and the no-skill references (dashed lines) in the S2S model.

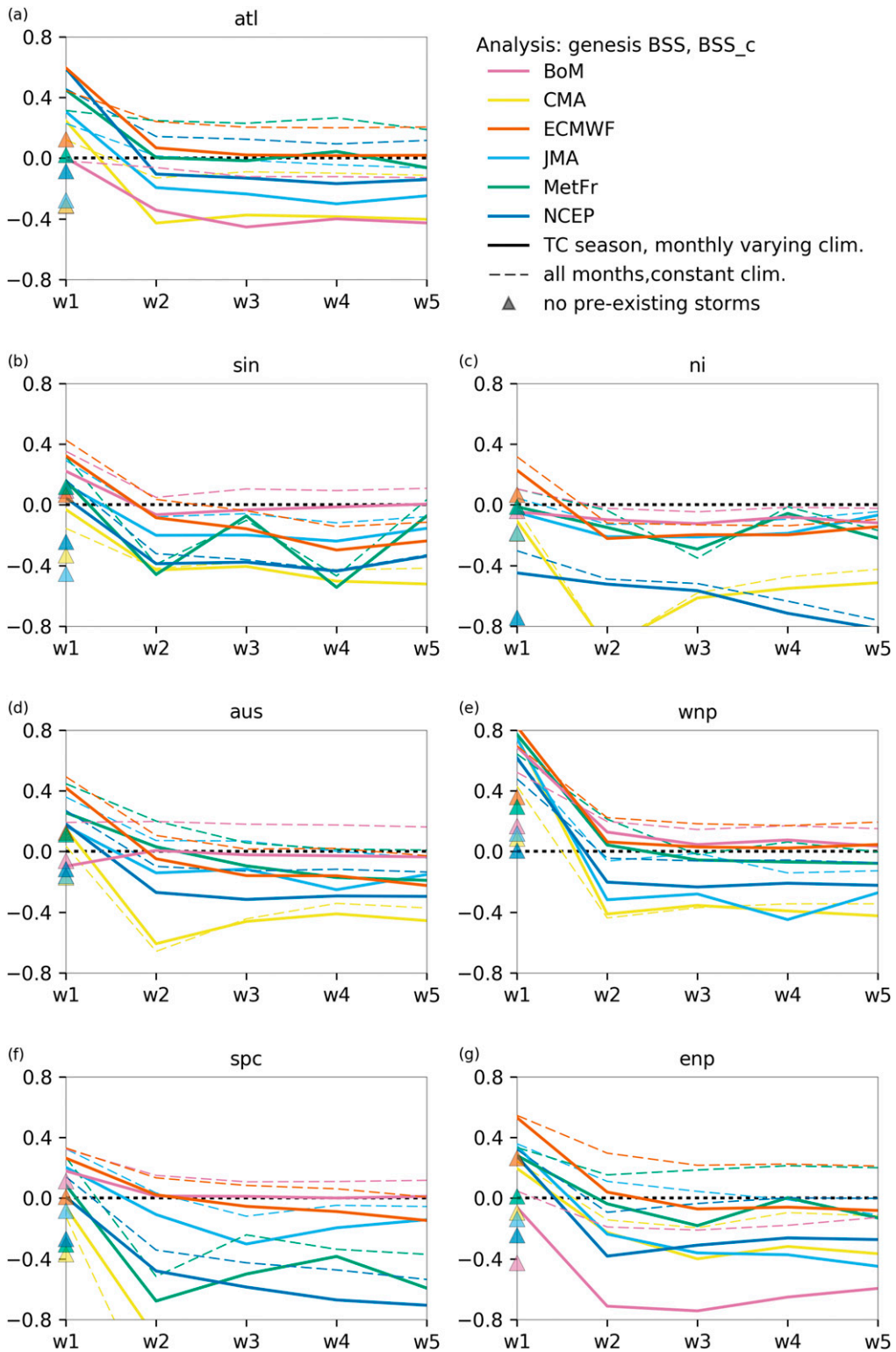


FIG. 8. Weekly, basin-wide BSSs in the S2S model relative to the constant (dashed lines) and to the monthly varying climatological (solid) predictions. The triangle markers indicate the BSS values without preexisting storms.

time of model initialization, which leads to a high association between forecasts and observations (therefore high BSS). Without including preexisting TCs,¹ the week 1 BSS values (the triangle markers in Fig. 8) are close to but (most of them) still slightly higher than those at week 2. By keeping the preexisting storms, we acknowledge that initialization is one of the factors contributing to a dynamical model's prediction skill. Most S2S models are skillful at week 1 in most basins, with the exception of the NI. Low or even negative week 1 BSS values in a basin are related to poor model initialization in those regions. From weeks 2 to 5, most models' BSS values level off, with the forecast errors saturating at week 2. This is consistent with the fact that the genesis climatology does not vary significantly after week 2. In some cases, such as for the ECMWF model in the SPC and the NCEP model in the NI, the model error continues to grow and therefore the BSS values decrease with increasing lead time. The MetFr's BSS values in the SIN and NI fluctuate, with relatively higher values during weeks 3 and 5 than in weeks 2 and 4. We speculate that these fluctuations are not meaningful but result from inadequate sample size.

Based on the BSS evaluation, the ECMWF model has skill up to week 5 in predicting TC occurrence in the ATL and WNP, up to week 2 in the SPC and ENP, but has no skill in the SIN, NI, and AUS after week 1. The BoM model has positive skill in the WNP up to week 5 and in the SPC up to week 2. The MetFr model is skillful up to week 2 in the WNP. The CMA, JMA, and NCEP models have no skill after week 1. Compared to the existing basin-wide statistical models (Leroy and Wheeler 2008; Slade and Maloney 2013),² the ECMWF, BoM, and MetFr models have comparable prediction skill. At week 1, the BSSs from these multiple logistic regression models are 0.13 and 0.17 in the ATL and ENP (Slade and Maloney 2013), 0.09 in the SIN,

0.06–0.08 near AUS, and 0.045 in the SPC (Leroy and Wheeler 2008). The highest S2S BSS values (from ECMWF) at week 1 are 0.7 in the ATL, WNP, and ENP; 0.35 in the AUS; 0.25 in the SPC; 0.25 in the SIN; and 0.06 in the NI. (Without considering the preexisting storms, they are 0.127, 0.36, and 0.27 in the ATL, WNP, and ENP, respectively; 0.126 in the AUS; 0.07 in the SIN and NI; and 0.01 in the SPC.) At week 2, the statistical models have BSS values of 0.11 and 0.16 in the ATL and ENP (Slade and Maloney 2013), 0.07 in the SIN, 0.05–0.07 near AUS, and 0.001 in the SPC (Leroy and Wheeler 2008). The highest S2S BSS values at the same lead time are 0.15 in the ATL, 0.19 in the WNP, 0.105 in the ENP (from ECMWF), 0.056 in the SPC, and 0.08 in the AUS (from BoM). None of the S2S models has positive skill in the SIN and NI at week 2. From week 3, the BSS values from the statistical models are overall better than those from the S2S models.

c. Potential predictability

The BSS values are low for weeks 2–5 in all global basins, even for those models that are skillful (with BSS above zero in Fig. 8). This raises the question of what are the upper limits of the subseasonal TC genesis prediction in these S2S models. We estimate these limits by computing potential predictability (Buizza 1997).³ For each of the S2S models, reforecasts from one of the ensemble members are treated as a fake “observation,” to which predictions from the rest of the members are verified against using the Brier Skill score (the same as was used for calculating the actual skill). This process is repeated for each ensemble member, and then we average the skill scores. Replacing the observations with a model forecast renders the model “perfect” in that the representation of the atmosphere in the forecast and in the target being forecast then are identical, without any systematic biases. The sources of the errors that remain are the uncertainties in the initial conditions and the unpredictable noise within the model.

Commonly, but not always (Kumar et al. 2014), the potential skill is larger than the actual skill. With the perfect observational data, sufficient ensemble spread, and unbiased models, the positive difference between the potential and actual skill levels can be interpreted as an indication of room for improvement in the models. Under the imperfect situation with possible errors in the observational data, insufficient ensemble spread, and models with systematic biases, the positive

¹ A preexisting TC is defined when the storm is identified by the tracker on day 1 with intensity greater than the respective TS threshold and when there is an observed storm greater than 34 kt (the TS threshold for observations) within 500-km distance of the simulated storm.

² While the mathematical formulas in the statistical models are similar from one basin to another, the predictors and how sensitive they are varies. Leroy and Wheeler (2008) focused on the southern oceans and used two MJO indices, the ENSO SST index, the Indo-Pacific SST, and the regional TC seasonal climatology. Slade and Maloney (2013), who focused on the Atlantic and east Pacific basins, used MJO and ENSO indices, as well as a regional genesis climatology. For the Atlantic basin, an additional predictor representing the variability of SST in the main development region is used. We do not distinguish between the different statistical models in our discussions but refer the interested reader to those studies.

³ The potential predictability discussed here is model dependent, not the intrinsic potential predictability of TC genesis.

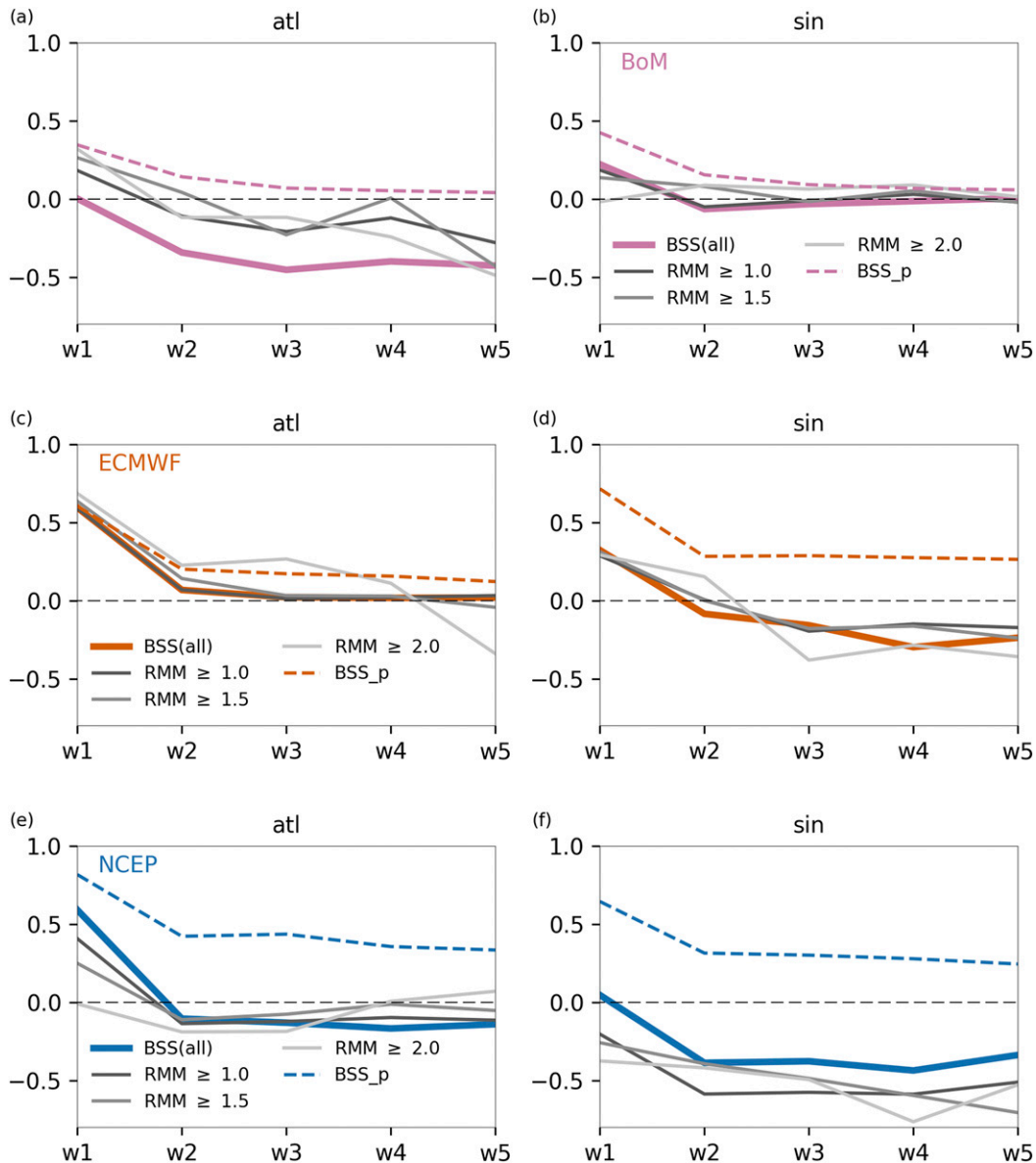


FIG. 9. Weekly, basin-wide BSS (solid) and the potential predictability (BSS_p; dashed) for the ATL and SIN basins from the BoM, ECMWF, and NCEP models. The grayish lines are the BSSs with data binned by the observed MJO magnitude at the time of initialization.

difference can be a consequence of these deficiencies. Observational errors lead to artificially low actual skill while insufficient ensemble spread and a biased model can result in artificially high potential skill (Wheeler et al. 2017).

The potential skill score used here is called BSS_p, where the “p” stands for “potential.” As an example, Fig. 9 shows the BSS_p results (dashed lines) of the BoM, ECMWF, and NCEP models in the ATL and SIN. The S2S models perform relatively better in the ATL than in other basins while the SIN is one of the more

challenging basins. Similar to BSS, the values of BSS_p (dashed lines) drop significantly after week 1 and level off from weeks 2 to 5. The BSS_p values are positive everywhere for all S2S models at all leads (not shown).

The differences between the BSS_p and the BSS (the gap between the dashed and the solid lines in Fig. 9) vary with lead time, basin, and model. They are usually smallest at week 1, indicating a positive contribution from the initialization. The differences are largest in the NCEP model when compared to those in the ECMWF and BoM. In addition to having low actual skill, the

NCEP model has only four members, which might result in insufficient ensemble spread, and therefore artificially high BSS_p. In the three Southern Hemisphere TC basins (figures for SPC and AUS are not shown), the BoM model's actual skill is close to the respective potential one. The values of BSS are closer to those of BSS_p in the ECMWF model in the ATL. Across models, the positive differences between BSS_p and BSS are largest in the CMA, and smallest in the ECMWF (not shown). Similar to the NCEP, CMA has both low actual skill and only four ensemble members. While there is no easy way to definitely explain the gap between potential and actual skill levels, large differences suggest deficiencies in the model, the ensemble system, or the data.

6. Discussion of the source of probabilistic prediction predictability

The ability of global models to represent TC genesis depends on the model characteristics, including the dynamical core (Reed et al. 2015; Vitart et al. 2001; Kim et al. 2018), the physical parameterizations (Reed and Jablonowski 2011), and the model resolution (Kajikawa et al. 2016). These characteristics are responsible for how well the genesis climatology, the MJO, and the interaction between the MJO and TCs (or interactions between two weather systems in general) are represented in the model. A good TC forecast (at least on weather scales) strongly relies on the model initialization (i.e., the data assimilation scheme). Additionally, the skill of the ensemble prediction system is sensitive to the design of the forecast, such as the ensemble size and the range of the model spread—topics that have been broadly studied within the context of weather (Wilson et al. 1999; Richardson 2001), seasonal (Brankovic and Palmer 1997; Deque 1997; Kumar et al. 2001; Kumar and Chen 2015), and decadal (Siens et al. 2016) predictions. In this section, we discuss the impact of the model characteristics, initialization, and the ensemble size on the sub-seasonal genesis probabilistic prediction skill.

a. BSS and climatology, MJO, and MJO–TC relationship

To examine how a model's TC genesis prediction skill is influenced by its representation of the observed TC climatology, the MJO, and the MJO–TC relationship, we compute the correlation, across lead times and models, between the BSS values (from Fig. 8) and the three verification indices, which are 1) the correlation of the basin-wide, monthly, genesis frequency between simulations and observations for TC climatology, 2) the bivariate correlation of RMM indices for rating the models' performance on the MJO [same as those shown

in Vitart (2017)], and 3) the fraction of variance r^2 of the observed MJO–TC relationship explained by the models from the candy plot analysis (Fig. 6). The correlations are calculated using data from weeks 1 to 5 (r_{w1-5}), as well as from weeks 2 to 5 (r_{w2-5}), with the latter excluding the influence of the initialization on these measures.

Figure 10 shows the scatterplots of BSS and these three verification indices in the ATL. The BSS values are positively correlated with the verification indices for the TC climatology, the MJO, and the MJO–TC relationship from weeks 1 to 5 (r_{w1-5}). The positive correlation can be partially attributed to the dependence of both quantities on the lead time; the week 1 BSS and the measure of week 1 genesis climatology (or of MJO and of the MJO–TC relationship) are both higher than those at week 2. Using data from weeks 2 to 5, the positive correlations remain, and the correlation coefficient (r_{w2-5}) becomes smaller, especially the r_{w2-5} between BSS and the index for the MJO simulation (Fig. 10b). Results from Fig. 10 suggest that the ATL BSS from weeks 2 to 5 is a consequence of the models' representation of the MJO–TC relationship more than it is a consequence of the relationship between the TC climatology and the MJO. The r_{w2-5} from BSS and the index for the MJO–TC relationship is above the 90% significance level.

Similar analyses are conducted for other basins, and there is no consistent dependence of BSS on the three indices examined (Fig. 11). In the SIN and NI, the BSS values are positively correlated to the models' performance in simulating the MJO. The BSSs in AUS and WNP are positively correlated to all three verification indices. The correlation with TC climatology and the MJO–TC relationship is equally strong in the WNP while the correlation with TC climatology is strongest in the ENP. The r_{w2-5} between BSS and the verification indices for the climatology and the MJO–TC relationship in the WNP and ENP are both above the 90% significance level. It is noted that the correlation does not necessarily represent causality. Furthermore, BSS and the three verification indices are not independent of each other, even for those that are above the 90% significance level. Results from Fig. 11 merely show their dependency. There might exist causality or both quantities might be affected by some factors that are not discussed here.

b. BSS and initial MJO magnitude

Belanger et al. (2010) found that the subseasonal Atlantic hurricane prediction in the ECMWF model depends on the MJO magnitude at the time of model initialization. Following their work, we bin the reforecasts

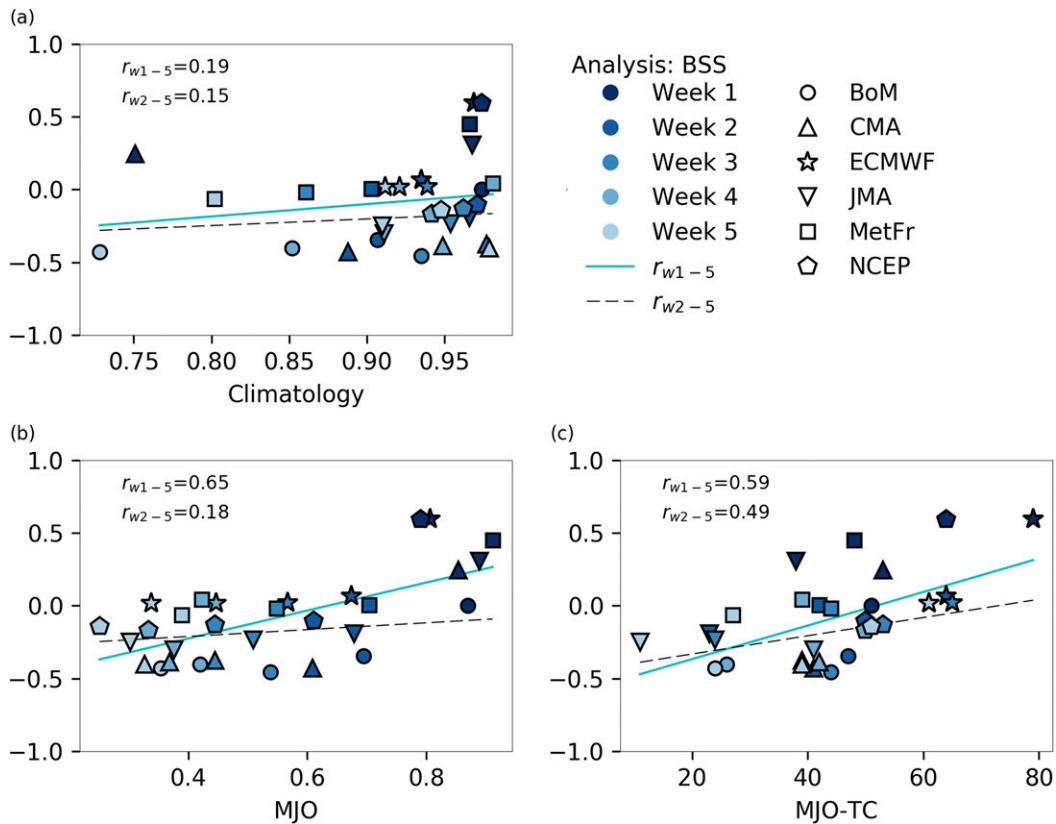


FIG. 10. Basin-wide BSSs (from Fig. 8) as a function of verification indices for (a) the genesis climatology, (b) the MJO, and (c) the MJO–TC relationship from weeks 1 to 5 (colors) in all six S2S models (symbols). The three indices are the correlation of the basin-wide genesis frequency between simulations and the observations, the bivariate correlation of RMM, and r^2 from the candy plot analysis (Fig. 6). The solid cyan lines show the best-fit line and the correlation coefficient between BSS and the indices with data from all leads (r_{w1-5}), while the dashed gray lines are for data from weeks 2 to 5 (r_{w2-5}).

by MJO magnitude at the initial time. Our results for the ECMWF model in the ATL (grayish lines in Fig. 9) are consistent with those of Belanger et al. (2010); higher BSSs with increasing MJO magnitude. We, however, do not find such a relationship in other basins. There is no consistent dependence between BSS and the initial MJO magnitude across basins for the BoM and NCEP models (Fig. 9) nor for other S2S models (not shown) either.

c. BSS and ensemble size

Next, we consider the impact of the ensemble size on the forecast skill of subseasonal TC predictions. In particular, we are interested in exploring whether the low skill scores of the NCEP model are due to its small ensemble size (four members) in the S2S reforecasts. (In the climatology and candy plot analyses, the NCEP system’s performance is as good as those of the BoM, and MetFr models, but its skill scores are negative after week 1.) To examine this question, we first reduce the

number of ensemble members for all S2S models to four in the BSS calculation. We focus our analysis in the BoM, ECMWF, and MetFr models at weeks 2 and 5, as the CMA and JMA models only have four and five ensemble members, respectively. As expected, the ECMWF, BoM, and MetFr BSS values drop to below zero when only four ensemble members are used in the calculation (Fig. 12). The ECMWF system is still more skillful than NCEP with four ensemble members, as is the BoM model in the basins, where it is skillful with all 33 ensemble members.

We further calculate the BSS values with increasing ensemble size and find that the BoM model reaches a saturation point with roughly 15 ensemble members; that is, further increases in the ensemble size do not benefit the genesis prediction skill. The ECMWF and MetFr models, with 11 and 15 ensemble members, respectively, seem to be close to their saturation points as well, although their BSS values are not flat yet. In other words, the forecast strategy of the

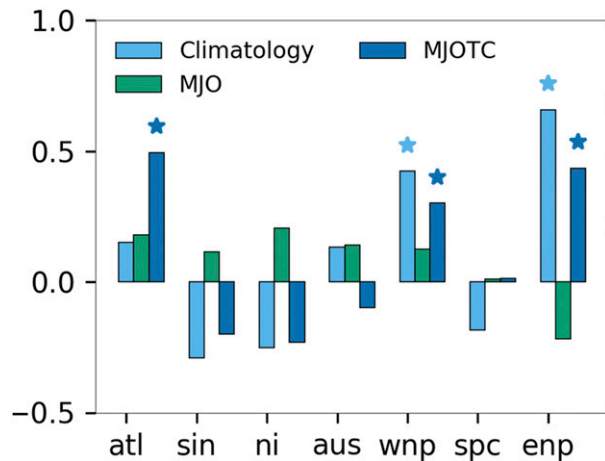


FIG. 11. The r_{w2-5} from Fig. 10, but for all global TC basins. Light blue bars show genesis climatology, green bars show the MJO, and the dark blue bars show the TC–MJO relationships. The correlations coefficients above the 90% significance level are marked with stars.

ECMWF and MetFr is probably more efficient than that chosen by BoM. Thus, we can expect that the NCEP, JMA, and CMA models will have better skill scores using larger ensemble sizes. This is particularly true in the case of the NCEP model, as the JMA and CMA models have larger biases in their simulated TC genesis climatologies.

7. Summary

The subseasonal prediction skill of TC genesis forecasts, at the basin level, is examined in this study using reforecasts from six different ensemble prediction systems. Skill scores are calculated for deterministic and probabilistic predictions. We compute the potential predictability as well. Most forecasts are skillful for week 1 (days 1–7), the period when initialization is important. The prediction skill drops significantly for weeks 2–5, when the models' performance is associated with the models' ability to simulate TC genesis climatology, the MJO, and the interaction between the MJO and TCs.

Deterministically, the S2S models are skillful at predicting basin-wide TC occurrence at all lead times (from weeks 1 to 5), but not skillful at predicting the genesis frequency. From weeks 2 to 5, the Brier skill scores (BSSs) of the probabilistic predictions from all models in the ATL, WNP, and ENP are found to be positively related to how well the global MJO–TC relationship is captured by the model. The BSS is positively related to the models' performance in simulating the MJO in the ATL, SIN, NI, AUS, WNP, and ENP, and to the

accuracy of the simulated TC climatology in the ATL, AUS, WNP, and ENP.

Among the six models, the ECMWF model delivers the best performance in reproducing the observed genesis TC climatology and is skillful in forecasting TC genesis up to week 5 in the ATL and WNP, week 2 in the SPC and ENP, and week 1 for the SIN, NI, and AUS. The BoM system has positive skill up to week 4 in the WNP, week 2 in the SPC and AUS, and week 1 in the ATL and SIN. The MetFr model has skill in the WNP up to week 2, and week 1 in the other basins, except the NI. The CMA, JMA, and NCEP models show no skill in predicting TC genesis from weeks 2 to 5. Among the TC basins, subseasonal TC predictions in the Indian Ocean and southern oceans show the least skill, as most of the S2S models have less skill than the monthly climatological probabilities after week 1. In contrast, more S2S models have positive skill in the North Atlantic and North Pacific basins after week 1.

From weeks 2 to 5, the BSSs in all basins are either close to zero (having little skill compared to the climatological forecast) or below zero (having no skill), indicating the difficulty of subseasonal-scale prediction with the current generation of models. The comparison between actual and potential skill suggests that the S2S models may not have yet reached their limits in predicting TC occurrence in all basins, though some models are close to that mark in some basins. The values of the BSSs are close to their respective potential skill levels in the ECMWF model in the ATL and in the three Southern Hemisphere TC basins in the BoM. The BSSs in NCEP and CMA are most distant from their potential values, though this might be due to their low actual skill as well as the insufficient number of ensemble members, which can result in artificially high potential skill.

While current skill scores are still low, the forecasts are produced directly without any bias correction. As noted by Vitart et al. (2010), the skill of a model can be extended by a few weeks by bias correction through postprocessing techniques based on the past hindcast performance. Furthermore, one can use derived parameters, such as the genesis potential index, rather than the direct output from TC detection, which might have useful results. Even with no bias correction, the most skillful models (ECMWF and BoM) have comparable (or slightly higher) skill to the existing regional statistical models at weeks 1 and 2. It is noted that our results may not reflect the maximum potential skill of some models, such as CMA, JMA, and NCEP, since their S2S reforecast ensembles are small (four or five members).

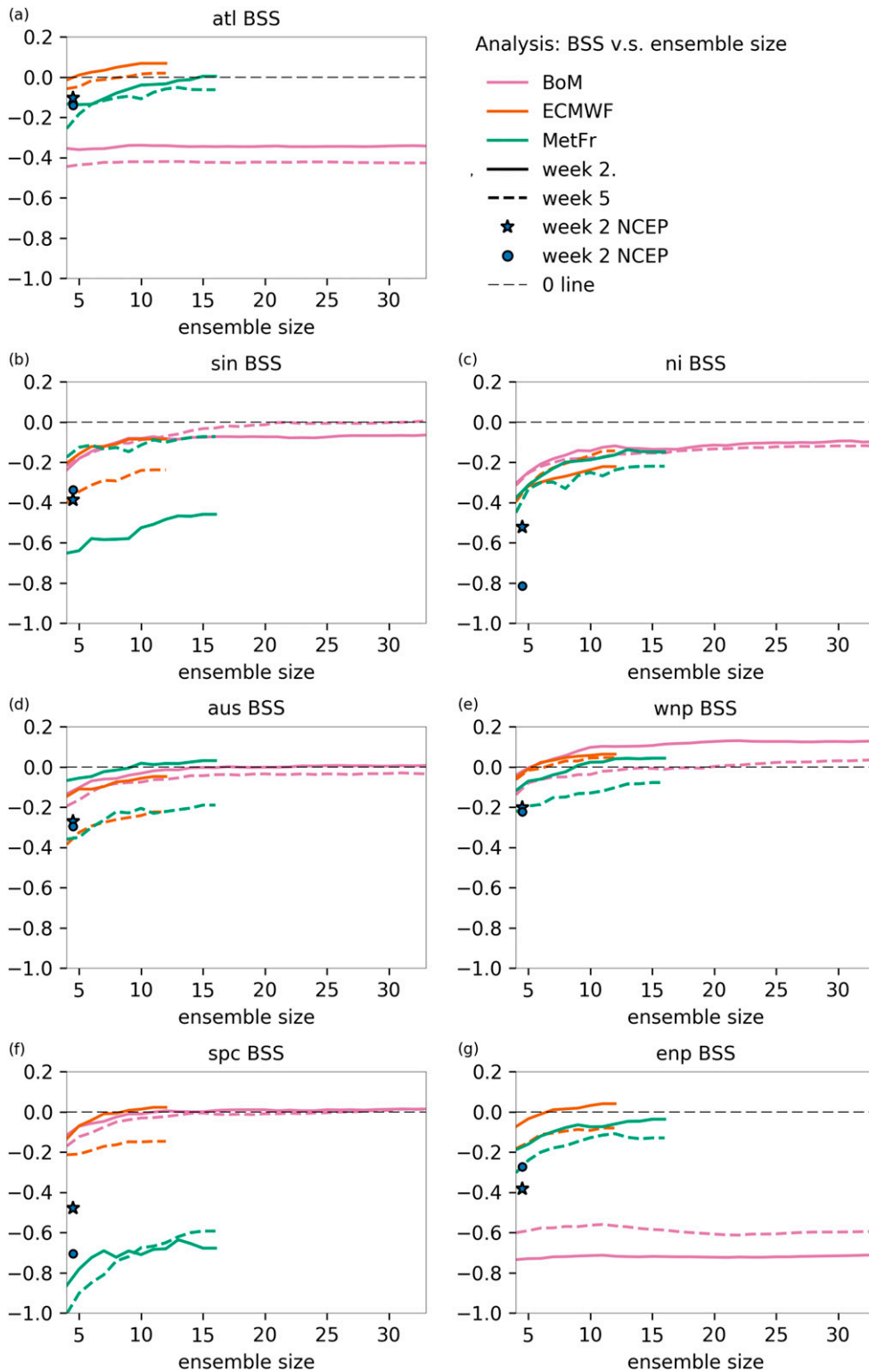


FIG. 12. Basin-wide BSSs at weeks 2 and 5 as a function of ensemble size in BoM, ECMWF, and MetFr. The blue star and circle indicate the BSSs from the NCEP model (four ensemble members) at weeks 2 and 5, respectively.

Acknowledgments. The research was supported by NOAA S2S Project NA16OAR4310079. Discussions with Drs. Shuguang Wang and Daniel Stern during the course of this study are appreciated. We thank Dr. Carl Schreck and an anonymous reviewer for their comments and constructive suggestions.

REFERENCES

- Ayyer, A., and J. Molinari, 2008: MJO and tropical cyclogenesis in the Gulf of Mexico and eastern Pacific: Case study and idealized numerical modeling. *J. Atmos. Sci.*, **65**, 2691–2704, <https://doi.org/10.1175/2007JAS2348.1>.
- Barnston, A. G., N. Vignaud, L. N. Long, M. K. Tippett, and J.-K. E. Schemm, 2015: Atlantic tropical cyclone activity in response to the MJO in NOAA's CFS model. *Mon. Wea. Rev.*, **143**, 4905–4927, <https://doi.org/10.1175/MWR-D-15-0127.1>.
- Belanger, J. I., J. A. Curry, and P. J. Webster, 2010: Predictability of North Atlantic tropical cyclone activity on intraseasonal time scales. *Mon. Wea. Rev.*, **138**, 4362–4374, <https://doi.org/10.1175/2010MWR3460.1>.
- Bessafi, M., and M. C. Wheeler, 2006: Modulation of South Indian Ocean tropical cyclones by the Madden–Julian oscillation and convectively coupled equatorial waves. *Mon. Wea. Rev.*, **134**, 638–656, <https://doi.org/10.1175/MWR3087.1>.
- Brankovic, C., and T. N. Palmer, 1997: Atmospheric seasonal predictability and estimates of ensemble size. *Mon. Wea. Rev.*, **125**, 859–874, [https://doi.org/10.1175/1520-0493\(1997\)125<0859:ASPAEO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<0859:ASPAEO>2.0.CO;2).
- Buizza, R., 1997: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF Ensemble Prediction System. *Mon. Wea. Rev.*, **125**, 99–119, [https://doi.org/10.1175/1520-0493\(1997\)125<0099:PFSOEP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<0099:PFSOEP>2.0.CO;2).
- Camargo, S. J., 2013: Global and regional aspects of tropical cyclone activity in the CMIP5 models. *J. Climate*, **26**, 9880–9902, <https://doi.org/10.1175/JCLI-D-12-00549.1>.
- , and A. G. Barnston, 2009: Experimental dynamical seasonal forecasts of tropical cyclone activity at IRI. *Wea. Forecasting*, **24**, 472–491, <https://doi.org/10.1175/2008WAF2007099.1>.
- , and A. A. Wing, 2016: Tropical cyclones in climate models. *Wiley Interdiscip. Rev.: Climate Change*, **7**, 211–237, <https://doi.org/10.1002/wcc.373>.
- , M. C. Wheeler, and A. H. Sobel, 2009: Diagnosis of the MJO modulation of tropical cyclogenesis using an empirical index. *J. Atmos. Sci.*, **66**, 3061–3074, <https://doi.org/10.1175/2009JAS3101.1>.
- Camp, J., and Coauthors, 2018: Skilful multi-week tropical cyclone prediction in ACCESS1 and the role of the MJO. *Quart. J. Roy. Meteor. Soc.*, <https://doi.org/10.1002/qj.3260>, in press.
- Chu, J.-H., C. R. Sampson, A. Lavine, and E. Fukada, 2002: The Joint Typhoon Warning Center tropical cyclone best-tracks, 1945–2000. Naval Research Laboratory Tech. Rep. NRL/MR/7540-02-16, 22 pp., http://www.usno.navy.mil/NOOC/nmfc-ph/RSS/jtwc/best_tracks/TC_bt_report.html.
- Deque, M., 1997: Ensemble size for numerical seasonal forecasts. *Tellus*, **49A**, 74–86, <https://doi.org/10.1034/j.1600-0870.1997.00005.x>.
- Duvel, J.-P., S. J. Camargo, and A. H. Sobel, 2017: Role of convection scheme in modeling initiation and intensification of tropical depressions over the North Atlantic. *Mon. Wea. Rev.*, **145**, 1495–1509, <https://doi.org/10.1175/MWR-D-16-0201.1>.
- Elsberry, R. L., M. S. Jordan, and F. Vitart, 2011: Evaluation of the ECMWF 32-day ensemble predictions during the 2009 season of the western North Pacific tropical cyclone events on intraseasonal timescales. *Asia-Pac. J. Atmos. Sci.*, **47**, 305–318, <https://doi.org/10.1007/s13143-011-0017-8>.
- , H.-C. Tsai, and M. S. Jordan, 2014: Extended-range forecasts of Atlantic tropical cyclone events during 2012 using the ECMWF 32-day ensemble predictions. *Wea. Forecasting*, **29**, 271–288, <https://doi.org/10.1175/WAF-D-13-00104.1>.
- Gottschalck, J., and Coauthors, 2010: A framework for assessing operational Madden–Julian oscillation forecasts: A CLIVAR MJO Working Group project. *Bull. Amer. Meteor. Soc.*, **91**, 1247–1258, <https://doi.org/10.1175/2010BAMS2816.1>.
- Hall, J. D., A. J. Matthews, and D. J. Karoly, 2001: The modulation of tropical cyclone activity in the Australian region by the Madden–Julian oscillation. *Mon. Wea. Rev.*, **129**, 2970–2982, [https://doi.org/10.1175/1520-0493\(2001\)129<2970:TMOTCA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2970:TMOTCA>2.0.CO;2).
- Ho, C.-H., J.-H. Kim, J.-H. Jeong, and H.-S. Kim, 2006: Variation of tropical cyclone activity in the South Indian Ocean: El Niño–Southern Oscillation and Madden–Julian oscillation effects. *J. Geophys. Res.*, **111**, D22191, <https://doi.org/10.1029/2006JD007289>.
- Holloway, C. E., S. J. Woolnough, and G. M. S. Lister, 2013: The effects of explicit versus parameterized convection on the MJO in a large-domain high-resolution tropical case study. Part I: Characterization of large-scale organization and propagation. *J. Atmos. Sci.*, **70**, 1342–1369, <https://doi.org/10.1175/JAS-D-12-0227.1>.
- Horn, M., and Coauthors, 2014: Tracking scheme dependence of simulated tropical cyclone response to idealized climate simulations. *J. Climate*, **27**, 9197–9213, <https://doi.org/10.1175/JCLI-D-14-00200.1>.
- Hung, M.-P., J.-L. Lin, W. Wang, D. Kim, T. Shinoda, and S. J. Weaver, 2013: MJO and convectively coupled equatorial waves simulated by CMIP5 climate models. *J. Climate*, **26**, 6185–6214, <https://doi.org/10.1175/JCLI-D-12-00541.1>.
- Jia, X., C. Li, J. Ling, and C. Zhang, 2008: Impacts of a GCM's resolution on MJO simulation. *Adv. Atmos. Sci.*, **25**, 139–156, <https://doi.org/10.1007/s00376-008-0139-9>.
- Jiang, X., M. Zhao, and D. E. Waliser, 2012: Modulation of tropical cyclones over the eastern Pacific by the intraseasonal variability simulated in an AGCM. *J. Climate*, **25**, 6524–6538, <https://doi.org/10.1175/JCLI-D-11-00531.1>.
- Kajikawa, Y., Y. Miyamoto, R. Yoshida, T. Yamaura, H. Yashiro, and H. Tomita, 2016: Resolution dependence of deep convections in a global simulation from over 10-kilometer to sub-kilometer grid spacing. *Prog. Earth Planet. Sci.*, **3**, 16, <https://doi.org/10.1186/s40645-016-0094-5>.
- Kang, I.-S., M.-S. Ahn, and Y.-M. Yang, 2016: A GCM with cloud microphysics and its MJO simulation. *Geosci. Lett.*, **3**, 16, <https://doi.org/10.1186/s40562-016-0048-1>.
- Kikuchi, K., and B. Wang, 2010: Formation of tropical cyclones in the northern Indian Ocean associated with two types of tropical intraseasonal oscillation modes. *J. Meteor. Soc. Japan*, **88**, 475–496, <https://doi.org/10.2151/jmsj.2010-313>.
- Kim, D., A. H. Sobel, A. D. Del Genio, Y. Chen, S. J. Camargo, M.-S. Yao, M. Kelley, and L. Nazarenko, 2012: The tropical subseasonal variability simulated in the NASA GISS general circulation model. *J. Climate*, **25**, 4641–4659, <https://doi.org/10.1175/JCLI-D-11-00447.1>.
- , and Coauthors, 2014: Process-oriented MJO simulation diagnostic: Moisture sensitivity of simulated convection. *J. Climate*, **27**, 5379–5395, <https://doi.org/10.1175/JCLI-D-13-00497.1>.

- , and Coauthors, 2018: Process-oriented diagnosis of tropical cyclones in high-resolution GCMs. *J. Climate*, **31**, 1685–1702, <https://doi.org/10.1175/JCLI-D-17-0269.1>.
- Kim, J.-H., C.-H. Ho, H.-S. Kim, C.-H. Sui, and S. K. Park, 2008: Systematic variation of summertime tropical cyclone activity in the western North Pacific in relation to the Madden–Julian oscillation. *J. Climate*, **21**, 1171–1191, <https://doi.org/10.1175/2007JCLI1493.1>.
- Klotzbach, P. J., 2010: On the Madden–Julian oscillation–Atlantic hurricane relationship. *J. Climate*, **23**, 282–293, <https://doi.org/10.1175/2009JCLI2978.1>.
- , 2012: El Niño–Southern Oscillation, the Madden–Julian oscillation and Atlantic basin tropical cyclone rapid intensification. *J. Geophys. Res.*, **117**, D14104, <https://doi.org/10.1029/2012JD017714>.
- , 2014: The Madden–Julian oscillation’s impacts on worldwide tropical cyclone activity. *J. Climate*, **27**, 2317–2330, <https://doi.org/10.1175/JCLI-D-13-00483.1>.
- , and E. C. Oliver, 2015a: Modulation of Atlantic basin tropical cyclone activity by the Madden–Julian oscillation (MJO) from 1905 to 2011. *J. Climate*, **28**, 204–217, <https://doi.org/10.1175/JCLI-D-14-00509.1>.
- , and —, 2015b: Variations in global tropical cyclone activity and the Madden–Julian oscillation (MJO) since the mid-twentieth century. *Geophys. Res. Lett.*, **42**, 4199–4207, <https://doi.org/10.1002/2015GL063966>.
- Krishnamohan, K. S., K. Mohanakumar, and P. V. Joseph, 2012: The influence of Madden–Julian oscillation in the genesis of north Indian Ocean tropical cyclones. *Theor. Appl. Climatol.*, **109**, 271–282, <https://doi.org/10.1007/s00704-011-0582-x>.
- Kumar, A., and M. Y. Chen, 2015: Inherent predictability, requirements on the ensemble size, and complementarity. *Mon. Wea. Rev.*, **143**, 3192–3203, <https://doi.org/10.1175/MWR-D-15-0022.1>.
- , A. G. Barnston, and M. P. Hoerling, 2001: Seasonal predictions, probabilistic verifications, and ensemble size. *J. Climate*, **14**, 1671–1676, [https://doi.org/10.1175/1520-0442\(2001\)014<1671:SPPEAE>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<1671:SPPEAE>2.0.CO;2).
- , P. Peng, and M. Chen, 2014: Is there a relationship between potential and actual skill? *Mon. Wea. Rev.*, **142**, 2220–2227, <https://doi.org/10.1175/MWR-D-13-00287.1>.
- Landsea, C. W., and J. L. Franklin, 2013: Atlantic hurricane database uncertainty and presentation of a new database format. *Mon. Wea. Rev.*, **141**, 3576–3592, <https://doi.org/10.1175/MWR-D-12-00254.1>.
- Leroy, A., and M. C. Wheeler, 2008: Statistical prediction of weekly tropical cyclone activity in the Southern Hemisphere. *Mon. Wea. Rev.*, **136**, 3637–3654, <https://doi.org/10.1175/2008MWR2426.1>.
- Li, R. C. Y., and W. Zhou, 2013: Modulation of western North Pacific tropical cyclone activity by the ISO. Part I: Genesis and intensity. *J. Climate*, **26**, 2904–2918, <https://doi.org/10.1175/JCLI-D-12-00210.1>.
- Li, W. W., Z. Wang, and M. S. Peng, 2016: Evaluating tropical cyclone forecasts from the NCEP Global Ensemble Forecasting System (GEFS) reforecast version 2. *Wea. Forecasting*, **31**, 895–916, <https://doi.org/10.1175/WAF-D-15-0176.1>.
- Liebmann, B., H. H. Hendon, and J. D. Glick, 1994: The relationship between tropical cyclones of the western Pacific and Indian Oceans and the Madden–Julian oscillation. *J. Meteor. Soc. Japan*, **72**, 401–411, https://doi.org/10.2151/jmsj1965.72.3_401.
- Madden, R. A., and P. R. Julian, 1972: Description of global circulation cells in the tropics with a 40–50 day period. *J. Atmos. Sci.*, **29**, 1109–1123, [https://doi.org/10.1175/1520-0469\(1972\)029<1109:DOGSCC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1972)029<1109:DOGSCC>2.0.CO;2).
- Maloney, E. D., and D. L. Hartmann, 2000a: Modulation of eastern North Pacific hurricanes by the Madden–Julian oscillation. *J. Climate*, **13**, 1451–1460, [https://doi.org/10.1175/1520-0442\(2000\)013<1451:MOENPH>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<1451:MOENPH>2.0.CO;2).
- , and —, 2000b: Modulation of hurricane activity in the Gulf of Mexico by the Madden–Julian oscillation. *Science*, **287**, 2002–2004, <https://doi.org/10.1126/science.287.5460.2002>.
- , and —, 2001: The Madden–Julian oscillation, barotropic dynamics, and North Pacific tropical cyclone formation. Part I: Observations. *J. Atmos. Sci.*, **58**, 2545–2558, [https://doi.org/10.1175/1520-0469\(2001\)058<2545:TMJOBOD>2.0.CO;2](https://doi.org/10.1175/1520-0469(2001)058<2545:TMJOBOD>2.0.CO;2).
- Mo, K. C., 2000: The association between intraseasonal oscillations and tropical storms in the Atlantic basin. *Mon. Wea. Rev.*, **128**, 4097–4107, [https://doi.org/10.1175/1520-0493\(2000\)129<4097:TABIOA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)129<4097:TABIOA>2.0.CO;2).
- Molinari, J., D. Knight, M. Dickinson, D. Vollaro, and S. Skubis, 1997: Potential vorticity, easterly waves, and eastern Pacific tropical cyclogenesis. *Mon. Wea. Rev.*, **125**, 2699–2708, [https://doi.org/10.1175/1520-0493\(1997\)125<2699:PVEWAE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<2699:PVEWAE>2.0.CO;2).
- Murakami, H., and Coauthors, 2012: Future changes in tropical cyclone activity projected by the new high-resolution MRI-AGCM. *J. Climate*, **25**, 3237–3260, <https://doi.org/10.1175/JCLI-D-11-00415.1>.
- , and Coauthors, 2015: Simulation and prediction of category 4 and 5 hurricanes in the high-resolution GFDL HiFLOR coupled climate model. *J. Climate*, **28**, 9058–9079, <https://doi.org/10.1175/JCLI-D-15-0216.1>.
- Nakano, M., M. Sawada, T. Nasuno, and M. Satoh, 2015: Intraseasonal variability and tropical cyclogenesis in the western North Pacific simulated by a global nonhydrostatic atmospheric model. *Geophys. Res. Lett.*, **42**, 565–571, <https://doi.org/10.1002/2014GL062479>.
- Nakazawa, T., 1988: Tropical super clusters within intraseasonal variations over the western Pacific. *J. Meteor. Soc. Japan*, **66**, 823–839, https://doi.org/10.2151/jmsj1965.66.6_823.
- Oouchi, K., A. T. Noda, M. Satoh, H. M. H. Tomita, T. Nasuno, and S. Iga, 2009: A simulated preconditioning of typhoon genesis controlled by a boreal summer Madden–Julian Oscillation event in a global cloud-system-resolving model. *SOLA*, **5**, 65–68, <https://doi.org/10.2151/sola.2009-017>.
- Reed, K. A., and C. Jablonowski, 2011: Impact of physical parameterizations on idealized tropical cyclones in the Community Atmosphere Model. *Geophys. Res. Lett.*, **38**, L04805, <https://doi.org/10.1029/2010GL046297>.
- , J. T. Bacmeister, N. A. Rosenbloom, M. F. Wehner, S. C. Bates, P. H. Lauritzen, J. E. Truesdale, and C. Hannay, 2015: Impact of the dynamical core on the direct simulation of tropical cyclones in a high-resolution global model. *Geophys. Res. Lett.*, **42**, 3603–3608, <https://doi.org/10.1002/2015GL063974>.
- Richardson, D. S., 2001: Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quart. J. Roy. Meteor. Soc.*, **127**, 2473–2489, <https://doi.org/10.1002/qj.49712757715>.
- Satoh, M., and Coauthors, 2012: The Intra-Seasonal Oscillation and its control of tropical cyclones simulated by high-resolution global atmospheric models. *Climate Dyn.*, **39**, 2185–2206, <https://doi.org/10.1007/s00382-011-1235-6>.
- Sienz, F., W. A. Mueller, and H. Pohlmann, 2016: Ensemble size impact on the decadal predictive skill assessment. *Meteor. Z.*, **25**, 645–655, <https://doi.org/10.1127/metz/2016/0670>.

- Slade, S. A., and E. D. Maloney, 2013: An intraseasonal prediction model of Atlantic and east Pacific tropical cyclone genesis. *Mon. Wea. Rev.*, **141**, 1925–1942, <https://doi.org/10.1175/MWR-D-12-00268.1>.
- Sobel, A. H., and E. D. Maloney, 2000: Effect of ENSO and MJO on the western North Pacific tropical cyclones. *Geophys. Res. Lett.*, **27**, 1739–1742, <https://doi.org/10.1029/1999GL011043>.
- Tsai, H.-C., R. L. Elsberry, M. S. Jordan, and F. Vitart, 2013: Objective verifications and false alarm analyses of western North Pacific tropical cyclone event forecasts by the ECMWF 32-day ensemble. *Asia-Pac. J. Atmos. Sci.*, **49**, 409–420, <https://doi.org/10.1007/s13143-013-0038-6>.
- Vitart, F., 2009: Impact of the Madden Julian oscillation on tropical storms and risk of landfall in the ECMWF forecast system. *Geophys. Res. Lett.*, **36**, L15802, <https://doi.org/10.1029/2009GL039089>.
- , 2017: Madden–Julian oscillation prediction and teleconnections in the S2S database. *Quart. J. Roy. Meteor. Soc.*, **143**, 2210–2220, <https://doi.org/10.1002/qj.3079>.
- , and T. N. Stockdale, 2001: Seasonal forecasting of tropical storms using coupled GCM integrations. *Mon. Wea. Rev.*, **129**, 2521–2537, [https://doi.org/10.1175/1520-0493\(2001\)129<2521:SFOTSU>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2521:SFOTSU>2.0.CO;2).
- , J. L. Anderson, J. Sirutis, and R. E. Tuleya, 2001: Sensitivity of tropical storms simulated by a general circulation model to cumulus parameterization. *Quart. J. Roy. Meteor. Soc.*, **127**, 25–51, <https://doi.org/10.1002/qj.49712757103>.
- , A. Leroy, and M. C. Wheeler, 2010: A comparison of dynamical and statistical predictions of weekly tropical cyclone activity in the Southern Hemisphere. *Mon. Wea. Rev.*, **138**, 3671–3682, <https://doi.org/10.1175/2010MWR3343.1>.
- , and Coauthors, 2017: The Subseasonal to Seasonal (S2S) Prediction project database. *Bull. Amer. Meteor. Soc.*, **98**, 163–173, <https://doi.org/10.1175/BAMS-D-16-0017.1>.
- Walsh, K. J. E., M. Fiorino, C. W. Landsea, and K. L. McInnes, 2007: Objectively determined resolution-dependent threshold criteria for the detection of tropical cyclones in climate models and reanalyses. *J. Climate*, **20**, 2307–2314, <https://doi.org/10.1175/JCLI4074.1>.
- Wheeler, M. C., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917–1932, [https://doi.org/10.1175/1520-0493\(2004\)132<1917:AARMMI>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2).
- , H. Zhu, A. H. Sobel, D. Hudson, and F. Vitart, 2017: Seamless precipitation prediction skill comparison between two global models. *Quart. J. Roy. Meteor. Soc.*, **143**, 374–383, <https://doi.org/10.1002/qj.2928>.
- Wilson, L. J., W. R. Burrows, and A. Lanzinger, 1999: A strategy for verification of weather element forecasts from an ensemble prediction system. *Mon. Wea. Rev.*, **127**, 956–970, [https://doi.org/10.1175/1520-0493\(1999\)127<0956:ASFVOW>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<0956:ASFVOW>2.0.CO;2).
- Xiang, B., and Coauthors, 2015: Beyond weather time-scale prediction for Hurricane Sandy and Super Typhoon Haiyan in a global climate model. *Mon. Wea. Rev.*, **143**, 524–535, <https://doi.org/10.1175/MWR-D-14-00227.1>.
- Yamaguchi, M., F. Vitart, S. T. K. Lang, L. Magnusson, R. L. Elsberry, G. Elliot, M. Kyouda, and T. Nakazawa, 2015: Global distribution of the skill of tropical cyclone activity forecasts on short- to medium-range time scales. *Wea. Forecasting*, **30**, 1695–1709, <https://doi.org/10.1175/WAF-D-14-00136.1>.
- Zarzycki, C. M., and P. A. Ullrich, 2017: Assessing sensitivities in algorithmic detection of tropical cyclones in climate data. *Geophys. Res. Lett.*, **44**, 1141–1149, <https://doi.org/10.1002/2016GL071606>.
- Zhang, C., 2013: Madden–Julian oscillation: Bridging weather and climate. *Bull. Amer. Meteor. Soc.*, **94**, 1849–1870, <https://doi.org/10.1175/BAMS-D-12-00026.1>.
- Zhao, M., I. M. Held, and S.-J. Lin, 2012: Some counterintuitive dependencies of tropical cyclone frequency on parameters in a GCM. *J. Atmos. Sci.*, **69**, 2272–2283, <https://doi.org/10.1175/JAS-D-11-0238.1>.