A Statistical Intraseasonal Prediction Model of Extended Boreal Summer Western North Pacific Tropical Cyclone Genesis

HAIKUN ZHAO,[a] YING LU,[b] XIANAN JIANG,[c] PHILIP J. KLOTZBACH,[d] LIGUANG WU,[e] AND JIAN CAO[b]

[a] *Key Laboratory of Meteorological Disaster, Ministry of Education, and Joint International Research Laboratory of Climate and Environment Change, and Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disaster, and Pacific Typhoon Research Center, Nanjing University of Information Science and Technology, Nanjing, China*
[b] *Key Laboratory of Meteorological Disaster, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing, China*
[c] *Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, Los Angeles, California/Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California*
[d] *Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado*
[e] *Department of Atmospheric and Oceanic Sciences and Institute of Atmospheric Sciences, Fudan University, Shanghai, China*

(Manuscript received 6 February 2021, in final form 11 December 2021)

ABSTRACT: An L2 regularized logistic regression model is developed in this study to predict weekly tropical cyclone (TC) genesis over the western North Pacific (WNP) Ocean and subregions of the WNP including the South China Sea (SCS), the western WNP (WWNP), and the eastern WNP (EWNP). The potential predictors for the TC genesis model include a time-varying TC genesis climatology, the Madden–Julian oscillation (MJO), the quasi-biweekly oscillation (QBWO), and ENSO. The relative importance of the predictors in a constructed L2 regression model is justified by a forward stepwise selection procedure for each region from a 0-week to a 7-week lead. Cross-validated hindcasts are then generated for the corresponding prediction schemes out to a 7-week lead. The TC genesis climatology generally improves the regional model skill, and the importance of intraseasonal oscillations and ENSO is regionally dependent. Over the WNP, there is increased model skill over the time-varying climatology in predicting weekly TC genesis out to a 4-week lead by including the MJO and QBWO, whereas ENSO has a limited impact. On a regional scale, ENSO and then either the MJO or QBWO are the two most important predictors over the EWNP and WWNP after the TC genesis climatology. The MJO is found to be the most important predictor over the SCS. The logistic regression model is shown to have comparable reliability and forecast skill scores to the ECMWF dynamical model on intraseasonal time scales.

SIGNIFICANCE STATEMENT: Skillful forecasts of tropical cyclone activity on time scales from short-range to seasonal are now issued operationally. Although there has been great progress in the understanding of physical mechanisms driving tropical cyclone (TC) activity, intraseasonal prediction of TCs remains a significant scientific challenge. This study develops a statistically based intraseasonal model to predict weekly TC genesis over the western North Pacific Ocean basin. The intraseasonal prediction model developed here for TC genesis over the western North Pacific basin shows skill extending out to four weeks. We discuss the regional dependence of the model skill on ENSO and other subseasonal climate oscillations. This approach provides skillful intraseasonal forecasting of TCs over the western North Pacific basin.

KEYWORDS: Tropical cyclones; ENSO; Forecasting techniques; Statistical forecasting; Intraseasonal variability

1. Introduction

Tropical cyclones (TCs) are one of the most destructive natural disasters worldwide, posing a major threat to life and property over coastal and adjacent inland regions (Kleinen 2007; Zhang et al. 2009; Bakkensen et al. 2018; Klotzbach et al. 2018). To provide improved guidance for TC-associated disaster prevention and mitigation, there has been recent increasing attention on the improvement of extended-range prediction of TC activity (Vitart and Robertson 2018; Vitart et al. 2019; Merryfield et al. 2020). Over the past few decades, there has been significant progress made in the understanding of the physical mechanisms driving TC activity (Bender et al. 2010; Vecchi et al. 2019; Zhang et al. 2020). Skillful forecasts of TC activity on time scales from short-range (i.e., day to day) to seasonal are now issued operationally (Bauer et al. 2015; Klotzbach et al. 2019).

The skill of short-range TC forecasts arises largely from the initial conditions and generally shows an increasing trend associated with increased model resolution and improved data assimilation systems (Bauer et al. 2015). On seasonal time scales, TC forecasts largely depend on slowly varying large-scale environmental conditions such as sea surface temperatures (SSTs) (Gray 1979, 1984; McBride and Zehr 1981; Goldenberg and Shapiro 1996). Statistical models (Gray et al.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-21-0110.s1.

Zhao's ORCID: 0000-0002-1771-1461.

Corresponding author: Haikun Zhao, zhk2004y@gmail.com

DOI: 10.1175/JCLI-D-21-0110.1

1992; Chan et al. 2001; Saunders and Lea 2005), hybrid statistical–dynamical models (Vecchi et al. 2011; Klotzbach et al. 2020), and dynamical models (Vitart and Stockdale 2001; Camargo et al. 2005) have been successfully implemented for seasonal outlooks of TC activity (Klotzbach et al. 2019). On intraseasonal time scales, forecasts of TC activity are generally based on relationships with intraseasonal oscillations (ISOs), convectively coupled equational waves, and SST (Leroy and Wheeler 2008; Lee et al. 2018; Vitart and Robertson 2018; Camargo et al. 2019).

Intraseasonal prediction of TCs remains a significant scientific challenge, although some statistical, dynamical, and hybrid statistical–dynamical techniques have been developed to predict TCs on intraseasonal time scales (Leroy and Wheeler 2008; Elsberry et al. 2010, 2014; Zhu et al. 2017; Jiang et al. 2012, 2018; Klotzbach et al. 2019; Gregory et al. 2020). In an effort to arrive at a seamless suite of forecast products, a subseasonal-to-seasonal (S2S) project was initiated by the World Weather Research Program and the World Climate Research Program (Vitart et al. 2017). One of the key focuses of this project was to assess the skill of 11 dynamical models from operational agencies for intraseasonal prediction of extreme climate and weather events including TC activity, thus improving extended-range forecasting (Vitart et al. 2010; Vitart and Robertson 2018).

During recent years, some dynamical models have shown remarkable improvement in forecasting the Madden–Julian oscillation (MJO; Madden and Julian 1971), which is thought to be a major predictability source at intraseasonal time scales (Vitart 2014; Lim et al. 2018; Kim et al. 2019). Statistically based models for the intraseasonal prediction of TCs have shown skill comparable to techniques using either dynamical models or hybrid statistical–dynamical models (Vitart et al. 2010; Kim et al. 2018; Wang et al. 2019). Dynamical models still have considerable room for improvement in their prediction of ISO and tropical waves (Serra et al. 2014; Jiang et al. 2020).

Over the western North Pacific (WNP) Ocean basin, several prior studies have investigated intraseasonal forecasts of TC activity using dynamical models or statistical–dynamical techniques (Elsberry et al. 2010, 2014; Camargo et al. 2019). Statistical models have also been developed for skillful intraseasonal prediction of TC genesis. These statistical forecasts have focused on the Southern Hemisphere (Leroy and Wheeler 2008), or on the eastern North Pacific and North Atlantic Ocean basins (Slade and Maloney 2013), while to our knowledge, no publications have outlined a statistical model for intraseasonal prediction of TC genesis in the WNP basin. The factors affecting TC genesis in the WNP basin are somewhat different than those driving TC activity in the Southern Hemisphere, eastern North Pacific, and North Atlantic basins. This study uses a method that is similar to that of Leroy and Wheeler (2008) and Slade and Maloney (2013) but with a focus on the WNP. We construct a statistical prediction model for intraseasonal TC genesis over the WNP basin and test the prediction skill of the model. The statistical regression model utilized here is different from those two papers, in that it uses a L2 regularized model to largely overcome potential overfitting issues.

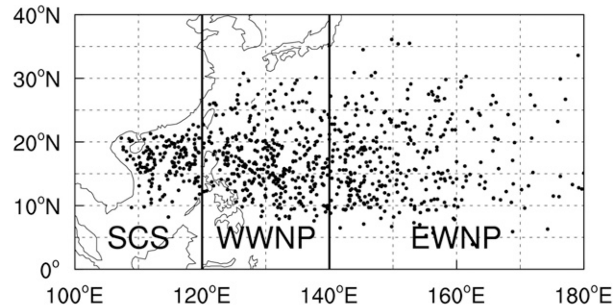

FIG. 1. Western North Pacific (WNP) study area and the three subregions as defined in this study: the South China Sea (SCS), the western WNP (WWNP), and the eastern WNP (EWNP). TC genesis locations from 1 May to 31 Oct from 1979 to 2019 are shown.

Section 2 describes the datasets used in this study and the logistic model development for predicting weekly TC genesis. Section 3 discusses the relative importance of potential predictors for the intraseasonal prediction model. Section 4 presents the details of the construction of the intraseasonal prediction model for WNP TC genesis. Section 5 assesses the skill of the logistic regression model for predicting weekly TC genesis, compares the model skill with the European Centre for Medium-Range Weather Forecasts (ECMWF) model, and develops a real-time forecast scheme for this model using a method that does not involve filtering. Section 6 summarizes the study.

2. Data and method

a. Data

TC data are obtained from the Joint Typhoon Warning Center (JTWC) best-track dataset (Chu et al. 2002), which includes latitude, longitude, minimum central pressure, and maximum sustained wind at 6-h intervals. Only TCs reaching tropical storm intensity (i.e., $\geq$34 kt; 1 kt $\approx$ 0.51 m s^{-1}) during the extended boreal summer (i.e., 1 May–31 October) from 1979 to 2019 over the WNP basin (0°–40°N, 100°E–180°) are considered in this study. The genesis time and location for each TC is defined as the first time and location at which the TC reached 34 kt during its lifetime. Since there are substantial differences in physical processes driving TC genesis over subregions of the WNP basin (Wang et al. 2000; Wang et al. 2007; Li and Zhou 2013a), optimal predictions for subregions may be achieved using different predictors than those for forecasts over the whole basin (Leroy and Wheeler 2008). In this study, the whole WNP is divided into three subregions including the South China Sea (SCS), the western WNP (WWNP), and the eastern WNP (EWNP) (Fig. 1). The selection of these three subregions is based on both physical and operational considerations (Rasmusson and Carpenter 1982; Wang et al. 2007; Wang and Chan 2002), with the boundary at 120°E providing a natural geographical separation between the waters of the western Philippine Islands and the eastern

Philippine Islands, and the boundary at 140°E roughly separating regions that are known to behave oppositely in response to El Niño–Southern Oscillation (ENSO) (Wang and Chan 2002; Camargo and Sobel 2005; Zhao et al. 2010).

Among the latest generation of WNP-based subseasonal forecast models, the ECMWF model has the best skill in predicting TC genesis anomalies at 2–3-week lead times in most basins (Lee et al. 2018, 2020). To compare the forecast skill of the dynamical and statistical predictions on intraseasonal time scales, hindcasts of TC genesis prediction by the ECMWF model are obtained from the S2S dataset. Further details of the S2S dataset are described in Vitart et al. (2017). The ECMWF model has a 46-day forecast lead time (~6-week lead) and 11 ensemble members for reforecasts. These reforecasts are run twice per week (Monday and Thursday) from 2000 to 2019. TCs are tracked in the dynamical model hindcasts using the method described in Vitart and Stockdale (2001). The TC tracks contain daily values of maximum sustained winds and storm locations. Note that hindcasts of the 1-week lead generated by the ECMWF model include preexisting storms (Lee et al. 2018). Since the ECMWF model used here is not able to simulate the highest observed TC intensities, only TCs that reach the tropical storm wind speed threshold in the model, which is defined as the 18th percentile of the lifetime maximum intensity cumulative density distribution (i.e., 24 kt), are considered in this study, following the method of Lee et al. (2018). For the dynamical model, the probability of TC genesis is computed from the fraction of the 11 ensemble members, defined as

$$p_i = \frac{1}{M}\sum_{j=1}^{M} P_{i,j}, \tag{1}$$

where M is the number of ensemble members and $P_{i,j}$ is the genesis prediction from the jth ensemble member for the ith forecast; $P_{i,j}$ is 0 for no genesis and is 1 for one or more storm genesis events during the forecast period.

To obtain the ENSO index, SST data are computed from the Hadley Centre sea ice and sea surface temperature (HadISST) dataset (Rayner et al. 2003). Outgoing longwave radiation (OLR) is used to quantify intraseasonal modes. Daily mean OLR data from 1979 to 2019 with a horizontal resolution of 2.5° latitude × 2.5° longitude are obtained from the National Oceanic and Atmospheric Administration (NOAA) (Liebmann and Smith 1996). Using empirical orthogonal function (EOF) analyses of physical variables (e.g., OLR, precipitation, and winds) as in previous studies (Wheeler and Hendon 2004; Lee et al. 2013; Li and Zhou 2013a; Zhao et al. 2015a,b, 2016), the two leading ISO modes with a prevailing period of 10–20 days [quasi-biweekly oscillation (QBWO)] and 30–60 days (MJO) are obtained using OLR. Both of these modes can significantly affect TC genesis over the WNP basin and are consequently selected as potential predictors for intraseasonal prediction of TCs.

b. *L2 regularized logistic regression model*

We now develop a statistical model for intraseasonal prediction of TC genesis over the WNP basin using multiple logistic regression. Logistic regression has been widely used in physical and social science (Leroy and Wheeler 2008; Ogutu et al. 2012; Zhou et al. 2020), and is useful for probabilistic forecasts as it allows the input predictand to be dichotomous (0 or 1) and forecasts a probability between 0 and 1. A dichotomous index is used to represent whether the event occurs or not, denoted by 1 (if the event occurs) or 0 (if it does not occur). The logistic regression hypothesis function $h_\beta(x)$ is defined as follows:

$$h_\beta(x) = \hat{P} = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_m x_m}}{1 + e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_m x_m}}, \tag{2}$$

where $\hat{P}$ is the predicted probability of TC genesis with a value between 0 and 1, $(x_1, x_2, \ldots, x_m)$ are the predictors, and $(\beta_1, \beta_2, \ldots, \beta_m)$ represent the regression coefficients of the predictand and the historical predictor values.

Regression models tend to construct a hypothesis function that fits all of the training samples if possible. If the model is trained for too long on a limited sample dataset, it begins to learn the noise and overfits the training data. That is, an overfitted model is useful only for the training data and not for independent data. If a model cannot generalize well to new data, then it will not be successful in real-time prediction. In addition, if overfitting occurs due to model complexity, it makes sense to reduce the number or the weight of certain predictors. However, it is hard to know which features to remove from a specific model in advance. Regularization methods can be particularly helpful to avoid overfitting. These methods work by adding a penalty term, such as the L1 norm and L2 norm, to the model's loss function (Ng 2004; Ogutu et al. 2012). A regression model that uses the L2 regularization technique is termed a ridge regression model (Hoerl and Kennard 1970). Ridge regression shrinks the regression coefficients by imposing a penalty on their size, rather than discarding a subset of the predictors, to reduce the prediction error of the model. In this study, a model with L2 regularized logistic regression (adding an L2 norm to the loss function of logistic regression) is developed to generate hindcasted probabilities. The loss function of L2 regularized logistic regression is expressed as

$$J(\beta) = -\frac{1}{n}\sum_{i=1}^{n}\{o_i \log h_\beta(x_i) + (1 - o_i)\log[1 - h_\beta(x_i)]\} + \frac{\lambda}{2n}\sum_{j=1}^{m}\beta_j^2, \tag{3}$$

where the term o_i is the observed probability of a cyclogenesis event and λ denotes the penalty parameter. A larger λ value will tend to shrink the regression coefficients toward zero. The first term on the right-hand side of Eq. (3) is the original logistic regression loss, and the second term represents the L2 penalty (i.e., ridge penalty). Notice that the intercept β_0 has been left out of the penalty term. The regression coefficients are calculated by minimizing the loss function $J(\beta)$. The logistic regression with L2 regularization can shrink the regression coefficients to avoid

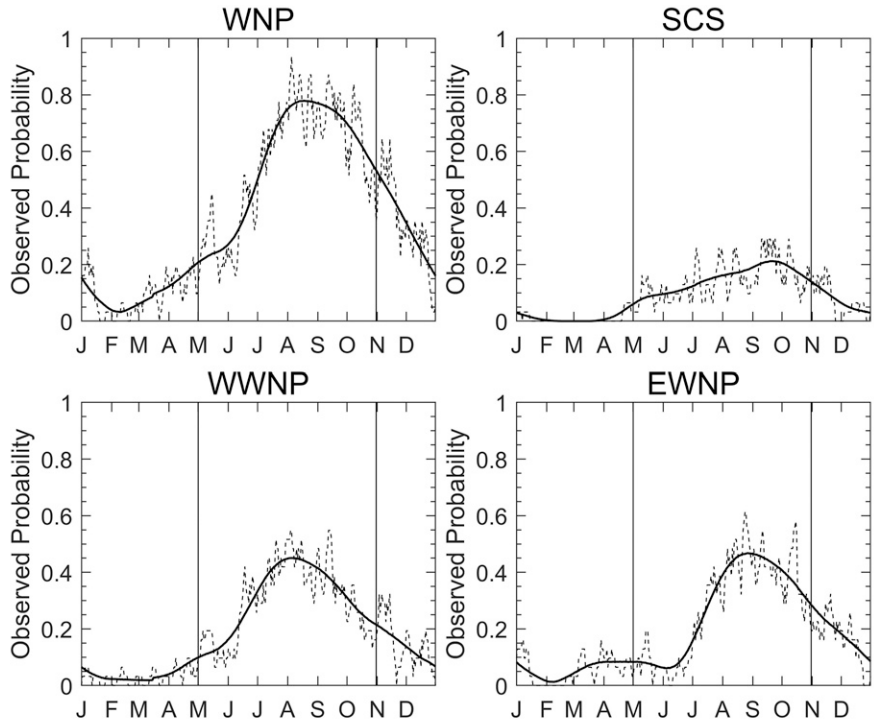

FIG. 2. Climatology of weekly tropical cyclogenesis probability for 1979–2009 in the full WNP and for the three subregions SCS, WWNP, and EWNP. The dashed curve is a raw climatology derived from observations, and the solid curve represents a smoothed climatology. Tick marks on the x axis show the first day of each month. Vertical lines denote the time range used in this study.

overfitting while retaining all of the variables in the model. The importance of the variable can be ignored when the regression coefficient is close to zero.

In this study, we use L2 regularized logistical regression to generate forecasts out to a 7-week lead (W0–W7) for TC genesis probability following Leroy and Wheeler (2008) and Slade and Maloney (2013). To eliminate potential bias in the model with TCs forming on a specific start day for each week (e.g., Sunday vs Monday), we develop the model using overlapping weeks, starting on every day. In addition, W0 is defined as the 7-day period centered on each day that the model is initialized, which provides no predictability but is used as a means of comparison. Individual models are developed for week leads from W0 to W7 for the whole WNP basin and for the three subregions (i.e., SCS, WWNP, and EWNP).

3. Predictor selection scheme

a. Climatology of TC genesis

A climatology of TC genesis is constructed based on observations during 1979–2009 for the whole WNP basin and for each of the three subregions, representing their corresponding average occurrence probability of weekly TC genesis. This climatology is selected as a predictor to characterize the seasonal cycle of weekly TC genesis (Fig. 2). Since sample sizes are relatively small, a raw climatological probability calculated by the observed weekly TC genesis probability (a binary probability) for every year is quite noisy (dashed curve), and the use of this raw climatology as a predictor can affect the performance of the regression model. Therefore, a smoothed climatology (solid curve) is adopted as a predictor instead of using the raw climatology for the whole WNP basin and for the three subregions. As was done by Leroy and Wheeler (2008) for the Southern Hemisphere, a smoothed climatology was similarly obtained by applying a low-pass filter to the raw climatology.

b. ENSO

In addition to the seasonal cycle of TC genesis, intraseasonal TC prediction also needs to include predictors on interannual time scales to represent the low-frequency large-scale environmental conditions affecting TC genesis. Due to its significant impacts on TC activity over the WNP basin (Lander 1994; Wang and Chan 2002; Camargo and Sobel 2005; Zhao et al. 2011; Patricola et al. 2018), ENSO is selected as a potential predictor for intraseasonal modulation of TC genesis. More TCs generally form over the southeastern region of the WNP basin, track westward and have an increased chance of intensification during El Niño years relative to La Niña years. As shown in Fig. 3, there is no

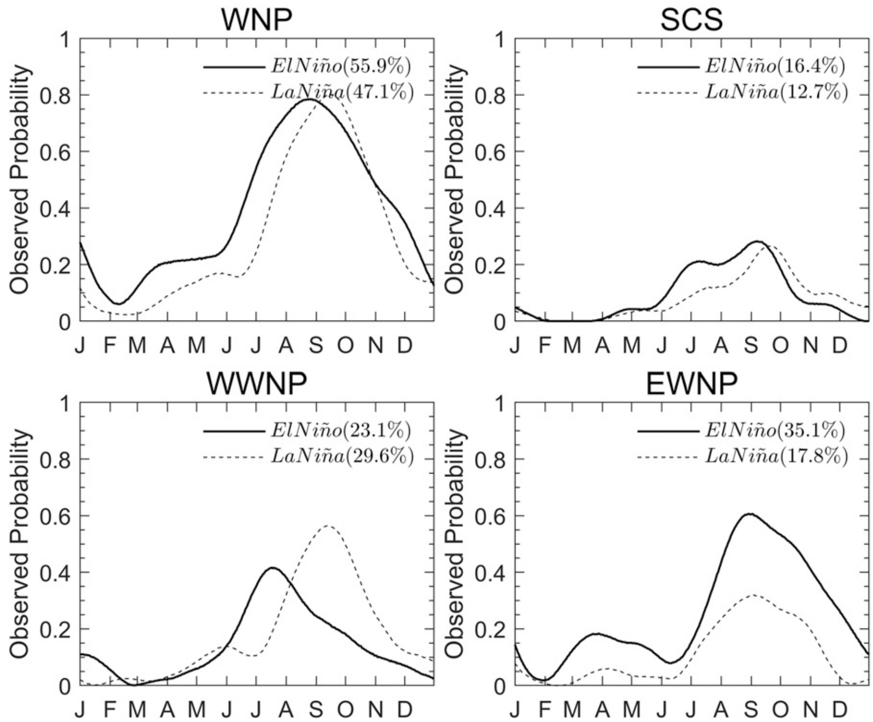

FIG. 3. TC genesis probability curves during strong El Niño (solid curve) and La Niña (dashed curve) years for the full WNP and for the three subregions SCS, WWNP, and EWNP. Strong El Niño years are 1982, 1987, 1991, 1994, 1997, 2002, 2009, and 2015. Strong La Niña years are 1988, 1995, 1998, 1999, 2007, 2010, and 2011. The average probability from 1 May to 31 Oct for El Niño and La Niña years is in parentheses.

significant difference in the climatological TC genesis probability over the whole WNP basin between El Niño and La Niña years. However, when examining the climatology of TC genesis probability for the three subregions, the peak in the WWNP subregion shows a lower or higher probability than in the EWNP subregion during El Niño or La Niña years, respectively, consistent with the previously reported significant impacts of ENSO on TC genesis location. In this study, ENSO is represented by the oceanic Niño index (ONI), defined as the 3-month running mean of SST anomalies over the Niño-3.4 region (5°S–5°N, 170°–120°W). Strong El Niño or La Niña years are defined as 5 consecutive overlapping 3-month periods when the ONI index anomaly is greater than +0.5°C or less than −0.5°C, respectively. This anomaly must also equal or exceed 1°C for at least 3 consecutive months during this time period. Following this definition, there are 15 strong ENSO years including eight strong El Niño years (1982, 1987, 1991, 1994, 1997, 2002, 2009, and 2015) and seven strong La Niña years (1988, 1995, 1998, 1999, 2007, 2010, and 2011) during 1979–2019.

c. MJO

Intraseasonal oscillations are regarded to be the main predictability source for intraseasonal prediction (Leroy and Wheeler 2008; Slade and Maloney 2013; Jiang et al. 2018; Vitart and Robertson 2018). Several studies have documented a strong modulation of TC activity by ISOs (Maloney and Hartmann 2000; Kim et al. 2008; Jiang et al. 2012; Li and Zhou 2013a; Zhao et al. 2015a,b), with more TCs during convectively active ISO phases and fewer TCs during convectively suppressed ISO phases. The Madden–Julian oscillation (Madden and Julian 1971) mode is a dominant ISO mode in the tropics. During the boreal summer, the MJO over the WNP basin shows a significant northward or northwestward propagation (Murakami 1984; Kemball-Cook and Wang 2001; Lee et al. 2013), which is distinct from the MJO during boreal winter that predominantly propagates eastward (Madden and Julian 1971). However, two studies have used the real-time multivariate MJO (RMM) index developed by Wheeler and Hendon (2004) as a main predictability source for the development of an intraseasonal prediction model for TC genesis (Leroy and Wheeler 2008; Slade and Maloney 2013). While the RMM index throughout the year describes MJO activity, it is not expected to fully represent seasonality and regionality of the MJO due to possible impacts from extratropical regions (Kikuchi et al. 2012).

To represent boreal summer MJO activity (Kikuchi et al. 2012; Lee et al. 2013), the MJO index in this study is constructed based on empirical orthogonal function (EOF) analysis of daily OLR anomalies over the region 0°–40°N, 100°E–180° from 1 May to 31 October for 1979–2019

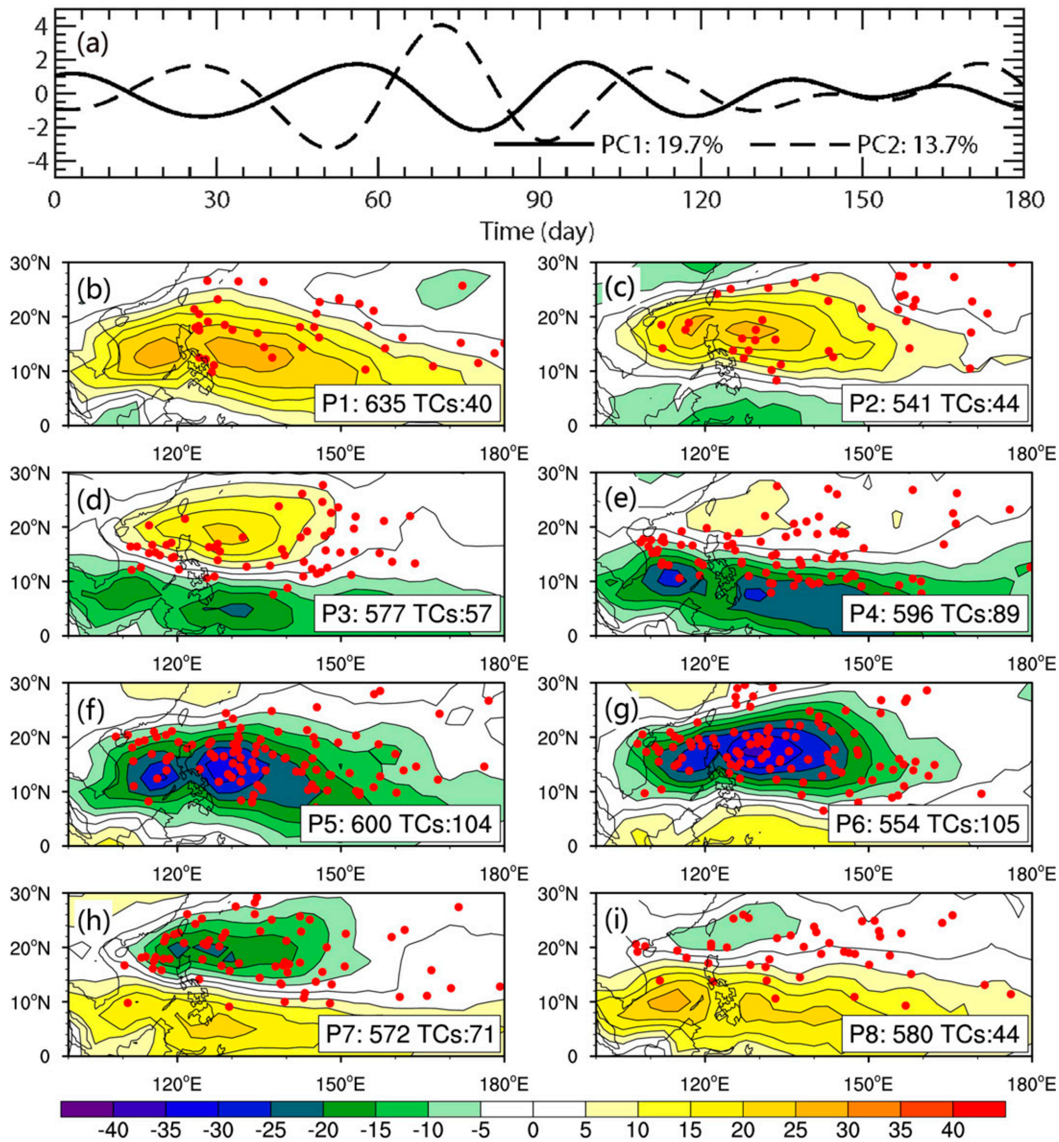

FIG. 4. (a) Boreal summer intraseasonal oscillation (MJO) activity depicted by the leading two principal components (PC1 and PC2; only a partial length of the series is displayed) of daily outgoing longwave radiation (OLR) anomalies. The OLR anomalies have been filtered by a Lanczos bandpass filter with cutoff periods of 30 and 60 days. Also shown are the variance contributions of the principal components. (b)–(i) TC genesis locations (red dots) and the number of TCs during MJO phases along with the 30–60-day-filtered OLR anomalies for the period of 1 May–31 Oct from 1979 to 2019. Also listed is the number of TCs counted (the second number) and the number of MJO days (the first number) with an amplitude ≥1 for phase 1–phase 8.

following previous studies (Jiang et al. 2012; Zhao et al. 2015a). The OLR anomalies are first filtered by a Lanczos bandpass filter (Duchon 1979) with 30–60-day cutoff periods before EOF analysis. The leading two EOF modes explain 19.8% and 13.8% of the variance of 30–60-day bandpass-filtered OLR anomalies, respectively. These two EOF modes can well represent boreal summer MJO activity over the WNP (Fig. 4) and are consistent with results from Kikuchi et al. (2012) and Lee et al. (2013). The corresponding principal components (PCs) of these EOF modes are hereinafter referred to as MJO-PC1 and MJO-PC2. We further examine the modulation of TCs over the WNP basin by MJO events when its amplitude, defined by the first two leading PCs (i.e., $\text{MJO-PC1}^2 + \text{MJO-PC2}^2$), is ≥ 1.0. As shown in Fig. 4, significantly more TCs occur during the convectively enhanced MJO phases (i.e., phases 4–7) whereas fewer TCs form during the convectively suppressed MJO phases (i.e., phase 8 and phases 1–3). Given the close relationship between TC genesis over the WNP basin and MJO activity, both indices (i.e., MJO-PC1 and MJO-PC2) are

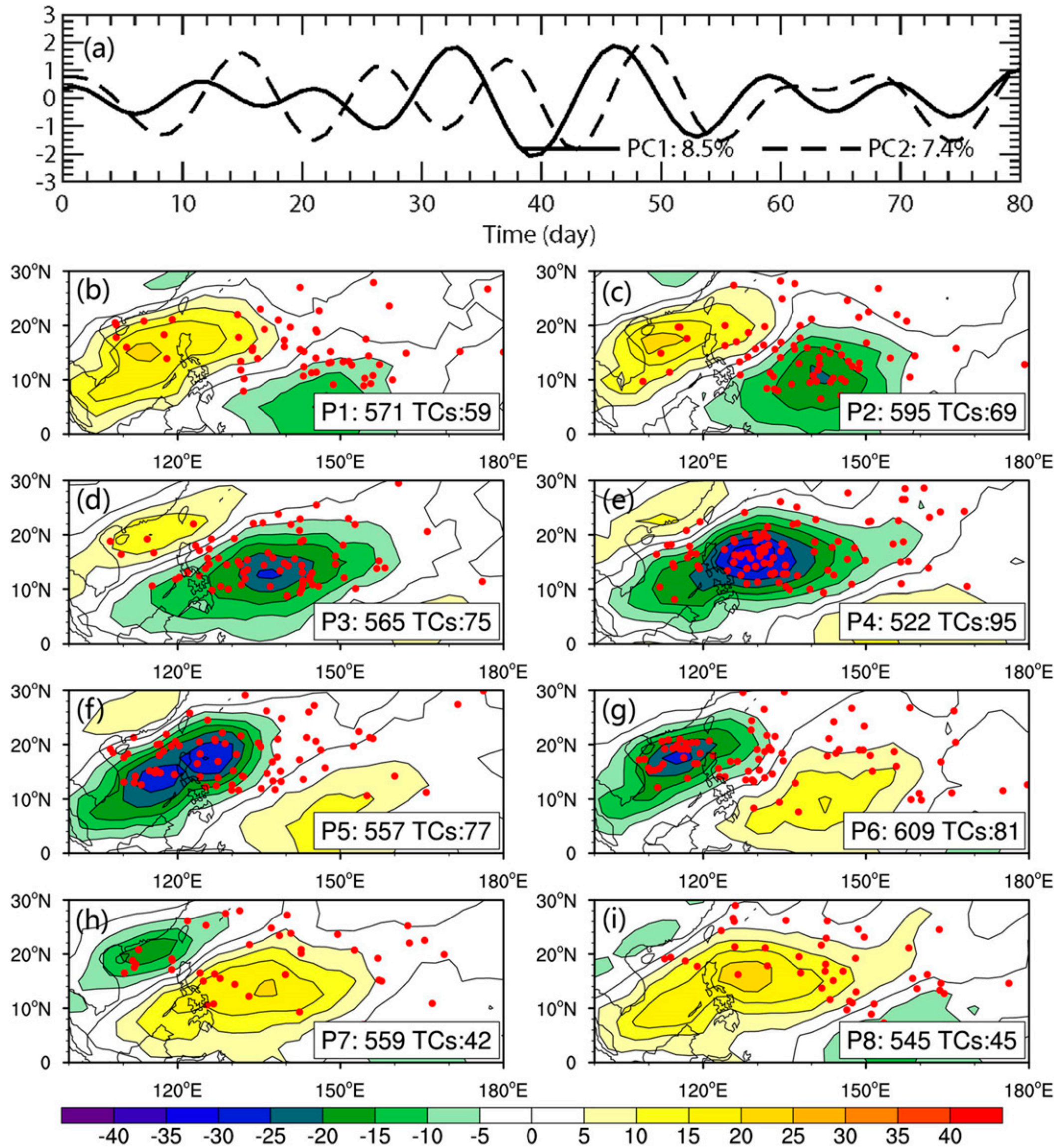

FIG. 5. As in Fig. 4, but for the quasi-biweekly oscillation (QBWO). The OLR anomalies have been filtered by a Lanczos bandpass filter with cutoff periods of 10 and 20 days. Also listed is the number of TCs counted and the number of QBWO days with an amplitude ≥1 for phase 1–phase 8.

selected as potential predictors for the multiple logistic regression model.

d. QBWO

The quasi-biweekly oscillation is regarded as another dominant ISO mode and is characterized by westward or northwestward propagation and a 10–20-day periodicity (Li 1996; Zhao et al. 2015a, 2016). Studies have suggested that the QBWO has stronger kinetic energy variance than the MJO and consequently modulates TCs over the WNP through changes in the large-scale environment (Li 1996; Li and Zhou 2013a,b; Zhao et al. 2016). The QBWO is also extracted using EOF analysis but is based on 10–20-day bandpass-filtered OLR anomalies. The first two leading modes explain 8.5% and 7.4% of the variance of 10–20-day bandpass-filtered OLR anomalies, respectively, representing the westward or northwestward propagation of the QBWO (Fig. 5). Similarly, a strong modulation by the QBWO on TC genesis is clearly evident in Fig. 5, with more TC genesis largely clustered over the regions with enhanced QBWO-associated convective activity (phases 2–5) in comparison with that with suppressed phases (phase 1 and phases 6–8). Zhou et al. (2018) also suggested a significant increase in TC growth rate during the convectively enhanced QBWO phases. The PCs of the first two leading EOF modes for the QBWO (i.e., QBWO-PC1 and

QBWO-PC2) are also included as potential predictors in the regression model.

e. Importance of the combined modulation of TCs by ENSO and intraseasonal oscillations

As highlighted earlier, ENSO is one of the most important interannual air–sea coupled modes, and the MJO and QBWO are two of the most important intraseasonal modes that can significantly modify the large-scale environment, thus affecting TC activity (Camargo et al. 2007, 2009; Li and Zhou 2013a,b; Zhao et al. 2015a,b, 2016; Zhao and Wang 2019; Hansen et al. 2020). TC genesis largely depends on TC-favorable large-scale environmental conditions, such as increased low-level vorticity, increased midlevel moisture, reduced vertical wind shear, and increased SST (Gray 1979; 1984; McBride and Zehr 1981; Goldenberg and Shapiro 1996). Using a genesis potential index (GPI) developed by Emanuel and Nolan (2004), Zhao et al. (2015a,b) suggested that relative humidity and low-level relative vorticity appeared to be the two most important environmental fields modulated by intraseasonal oscillations that then consequently impacted TC genesis over the WNP basin. Camargo et al. (2007) also investigated the relative importance of large-scale factors associated with ENSO in modulating TC genesis through a GPI analysis and found that both increased relative humidity and increased low-level relative vorticity were primarily responsible for the eastward shift in genesis location in the western North Pacific when El Niño conditions occurred.

However, the influence of ENSO and ISO on TC genesis is not independent or linear. Previous studies have emphasized the combined modulation of TC genesis over the WNP basin by the two leading ISO modes (i.e., QBWO and MJO) through changing large-scale conditions for TC genesis (Mao and Chan 2005; Li and Zhou 2013a,b; Zhao et al. 2015b). Li and Zhou (2013a) suggested that the WNP is controlled predominantly by MJO-induced positive OLR anomalies during convectively suppressed QBWO phases, leading to a significant suppression of TC genesis. However, TC suppression in the convectively suppressed MJO phases is generally weaker than during the convectively enhanced QBWO phases, mainly due to the counterbalance of the QBWO and MJO-associated large-scale conditions.

Moreover, a few studies have also highlighted the combined effect of ENSO and the ISO on WNP TC genesis (Li et al. 2012; Klotzbach and Oliver 2015; Han et al. 2019). For example, Li et al. (2012) found that ENSO can affect both genesis location and frequency of TCs by changing the location and intensity of MJO-associated convection, with stronger modulation of TC genesis by the MJO during El Niño years relative to that during neutral and La Niña years. Han et al. (2019) suggested that TC genesis modulation by the QBWO and ENSO is mainly related to spatial changes in the large-scale environment and synoptic-scale eddy kinetic energy that affect TC genesis associated with active or inactive phases of the QBWO under different ENSO phases.

The respective impact of ENSO and the ISO and their combined effect on TC genesis on intraseasonal time scales are considered in this study. We construct a statistical intraseasonal prediction model of TC genesis over the WNP basin. We find that a statistical prediction model has better skill in capturing TC genesis on intraseasonal time scales by incorporating both ENSO and ISO indices relative to a model using just ENSO or the ISO individually. Additionally, as mentioned in section 3b, ENSO significantly impacts TC genesis in the WWNP and EWNP but has a minimal impact on SCS TC genesis. When the convective centers of the MJO and QBWO are located over the SCS and WWNP, there is a stronger modulation of TC genesis in the SCS and WWNP than when convectively driven ISO modulation is located over the EWNP. We will show in section 5a that both ENSO and the ISO considerably impact the regional differences in TC genesis prediction skill.

4. Development of logistic regression model

a. Real-time prediction

The approach to construct the logistic regression model for intraseasonal prediction of TC genesis over the WNP basin in this study largely follows the approach employed by Leroy and Wheeler (2008) and Slade and Maloney (2013). To represent what would be available for a real-time forecast, all datasets utilizing the ONI are lagged by one month, and the four indices representing the MJO and QBWO are lagged by one day. The climatological values for the week being forecast (note that climatology is the only predictor not lagged at each forecast lead) and the five lagged predictors are used to calculate regression coefficients that are then input into Eq. (1) to generate forecasts out to a 7-week lead for the whole WNP basin and for the three subregions.

b. Relative importance of predictors

A forward stepwise selection scheme can be generally used to provide an order of the importance of predictors. Here, we use this scheme to quantify the relative importance of input variables for TC genesis prediction on intraseasonal time scales. Individual models are developed for each predictor k, and then chi-squared statistics are calculated. After including the predictor with the highest chi-squared statistic in the logistic regression model, the remaining $k - 1$ predictors are separately added into the model, and new chi-squared statistics are calculated. We repeat this step until the introduction of new predictors does not significantly reduce the sum of squared errors [i.e., under a 95% significance critical value following Cheng et al. (2006)]. In other words, the predictor that holds a lower p value (<0.05) is likely to be a more meaningful addition to the model. The relative importance of predictors in the stepwise regression scheme is usually consistent with results of individual predictors regressed on the probability of TC genesis in univariate analyses.

Table 1 summarizes the results of the forward selection scheme at each week lead for the full WNP and for the three subregions. Over the whole WNP basin, the climatology of TC genesis is the most important predictor, explaining the most variability in TC genesis. The MJO (including MJO-PC1

TABLE 1. Predictor selection order as determined by the forward selection scheme for the full WNP and for the three subregions, the South China Sea (SCS), the western WNP (WWNP), and the eastern WNP (EWNP), from 0-week to 7-week lead, using all available input data from 1979 to 2019. The number 1 denotes the first predictor chosen by the selection scheme, the number 2 denotes the second predictor selected, and so on. An asterisk indicates that the predictor failed the selection test and was not chosen.

Zone	Lead	Clim	MJO-PC1	MJO-PC2	QBWO-PC1	QBWO-PC2	ENSO
WNP	W0	1	2	3	4	5	6
	W1	1	2	6*	3	5	4
	W2	1	3	2	6*	5*	4
	W3	1	3	2	5	6*	4
	W4	1	2	5	4	6*	3
	W5	1	2	5*	6*	4*	3
	W6	1	6*	3	5*	4*	2
	W7	1	4*	3	6*	5*	2
SCS	W0	2	1	4	3	5	6*
	W1	2	1	6*	3	4	5*
	W2	1	2	3	4	6*	5*
	W3	1	3	2	4	5	6*
	W4	2	1	3	6*	4	5*
	W5	2	1	5*	3*	4*	6*
	W6	1	2	3	5*	6*	4*
	W7	1	3	2	4*	6*	5*
WWNP	W0	1	2	4	3	6	5
	W1	1	2	6	3	5	4
	W2	1	4	3	6	5	2
	W3	1	4	3	5	6*	2
	W4	1	3	4	5	6*	2
	W5	1	3	5*	4*	6*	2
	W6	1	5*	3	4*	6*	2
	W7	1	4	3	5*	6*	2
EWNP	W0	1	4	6*	5	3	2
	W1	1	4	5*	6*	3	2
	W2	1	5	3	6*	4	2
	W3	1	3	6*	4*	5*	2
	W4	1	3	4	5*	6*	2
	W5	1	3*	4*	5*	6*	2
	W6	1	3*	5*	6*	4*	2
	W7	1	3	6*	5*	4*	2

and MJO-PC2) is found to be the second most important predictor, especially for forecasts at W1–W5 leads. The QBWO shows less importance from a 1-week to 4-week lead, indicating that it provides relatively low predictability for TC genesis on intraseasonal time scales. ENSO has a limited impact on the prediction skill over the entire WNP, especially at W0–W3 leads, likely due to the opposite influences of ENSO on the WWNP and EWNP subregions. Over the SCS subregion, both ISO (i.e., MJO and QBWO) have a more important influence than ENSO. Over the WWNP subregion, the MJO and QBWO are the most important factors at W0–W1 leads, while ENSO is the second-most important predictor after the climatology of TC genesis for W2–W7 leads. Over the EWNP subregion, the inclusion of the MJO and QBWO has a limited impact on model skill, with ENSO being the second-most important predictor after the climatology of TC genesis at all forecast leads.

Note that the stepwise regression produces the "best" model by retaining a subset of the predictors and discarding the rest, resulting in a possibly lower prediction error than the full model (Thompson 1995; Whittingham et al. 2006; Smith 2018). However, the "best" model often exhibits high variance and may be overfit (Henderson and Denison 1989; Mundry and Nunn 2009). In this study, we just use forward stepwise regression to provide an order of predictor importance, but we do not use it to generate forecast probabilities. All potential predictors are used to develop the logistic regression model. The relative importance of potential predictors using the forward stepwise regression is found to be nearly consistent with the skill improvement in BSS. This consistency is useful for physical interpretation of selected predictors. More analyses will be shown in section 5a.

c. Cross-validation

A cross-validation method including both internal validation and external validation is used to generate hindcast probabilities and assess model skill. Data for the period from 2010 to 2019 are used to perform external validation. Hindcasts are generated for each year from 1979 to 2009 using a two-step cross-validation method. First, one year (internal validation) is successively left out of the period from 1979 to 2009. Second, the model is developed with the remaining 30 years (training dataset) to generate hindcasts for the one year removed (internal validation) and for 2010–19 (external validation). Note that the climatological probability of cyclogenesis (Clim) is recalculated for each different training period. These hindcast probabilities are independently generated from 1 May to 31 October, resulting in overlapping weekly probabilities as was done by Leroy and Wheeler (2008).

Examples of hindcast probabilities for a W1 lead by the regression model using all potential predictors are shown in Fig. 6. We examine hindcast probabilities for the strong El Niño year of 1997 and the strong La Niña year of 1999. As expected, the hindcast probability is lower than climatology over the WWNP subregion during the El Niño year and higher than climatology over the WWNP subregion during the La Niña year, with the opposite signal over the EWNP subregion. There is also stronger forecast TC genesis variability over the WWNP subregion than the EWNP subregion on intraseasonal time scales, due to a relatively stronger impact of ISO over the WWNP subregion. Moreover, we find that hindcast probabilities consistently follow climatology during these two ENSO events over the whole WNP basin and the SCS subregion. These results are largely explained by the opposing impact of ENSO on WWNP and EWNP TC genesis and the relatively weak effect of ENSO on TC genesis in the SCS subregion.

5. Model skill

a. Assessment of logistic model predictability

1) Performance of predictive model skill using the Brier score

The prediction skill of the regression model is quantitatively assessed by the Brier score. The Brier score is used to evaluate the accuracy of probabilistic forecasts using

$$\mathrm{BS} = \frac{1}{n}\sum_{i=1}^{n}(y_i - o_i)^2, \tag{4}$$

where n is the number of events, y_i represents the forecast probability of event i, and the term o_i is equal to 1 if the event i occurred and 0 if the event did not occur. The Brier score can take on any value between 0 and 1 since both the forecasts and observations are bounded by 0 and 1. The Brier score is the mean-square error of the forecast probability, and thus a lower Brier score means a better prediction and a larger Brier score means a worse prediction. The Brier skill score (BSS) is a metric that tells us how well the Brier score of a forecast model compares to a reference model. The BSS is calculated as

$$\mathrm{BSS} = 1 - \frac{\mathrm{BS}}{\mathrm{BS}_{\mathrm{ref}}}, \tag{5}$$

where BS represents the Brier score value of the hindcasts; $\mathrm{BS}_{\mathrm{ref}}$ is the Brier score calculated from a reference strategy that uses a seasonal mean climatology (as shown by the horizontal lines in Fig. 8, which varies with region only) to generate hindcast probabilities. Using this reference strategy, the same probability for each day can be predicted, indicating the mean seasonal probability of TC genesis in a week. Higher BSS indicates greater improvement of the hindcast model relative to the reference model.

In general, BSSs generated with all potential predictors are higher than with a subset of predictors, while occasionally higher BSSs are found when a subset of predictors are used. This usually occurs when the first few predictors have much higher predictive skill at TC genesis than the remaining predictors. To represent the skill improvement and the relative importance of predictors for W0–W7, the hindcast BSSs for models using all potential predictors and models using a subset of predictors are illustrated in Fig. 7.

Over the WNP basin, when all potential predictors are considered, the skill of the model using internal validation shows improvement over the reference model using a mean seasonal climatology from 16% to 24% during W0–W5. Similar results can be found for the skill of the model using external validation in comparison with the reference model using a mean seasonal climatology (~17%–24% for W0–W7). These results suggest that the capability of the L2 regularized regression model is strong and will likely show robust future skill at WNP TC genesis prediction on intraseasonal time scales. When using only the time-varying climatology (i.e., Clim) in the model using internal or external validation, the model respectively shows an ~16% or ~17% improvement over the reference model. By comparing BSSs for the model using a time-varying climatology and ENSO (i.e., Clim+ENSO) and the model using all predictors (All), we find almost no additional improvement. This finding is consistent with the limited impact of ENSO on all WNP TCs due to the opposite impact of ENSO on WWNP and EWNP TC genesis. The BSS of the model using climatology and ISO (i.e., Clim+MJO+QBWO) generates an improvement using an internal validation of greater than ~9% for W0 lead decreasing to ~1% for W4 lead over the Clim model and the Clim+ENSO model. We find similar improvement from incorporation of the ISO when using external validation. These results highlight the importance of both the TC genesis climatology and ISO for predictability of TC genesis over the WNP basin.

Over the SCS subregion, we find little improvement over the reference model when using the TC genesis climatology and either internal or external validation, likely caused by the low probability of TC genesis over the SCS region. The Clim+ENSO model adds little additional skill and even causes some degradation at certain lead times, coinciding with the limited impact of ENSO on TC genesis over this region. Additional substantial improvement is obtained by incorporating ISO (i.e., Clim+MJO+QBWO) at all forecast leads relative to the Clim model. In summary, ISO are of great importance for the prediction of TC genesis over the SCS subregion.

When using internal validation, the Clim model for WWNP and EWNP show improvements of ~6% and ~10% over the reference model respectively, and ENSO provides an additional ~1% improvement over the Clim model alone at all

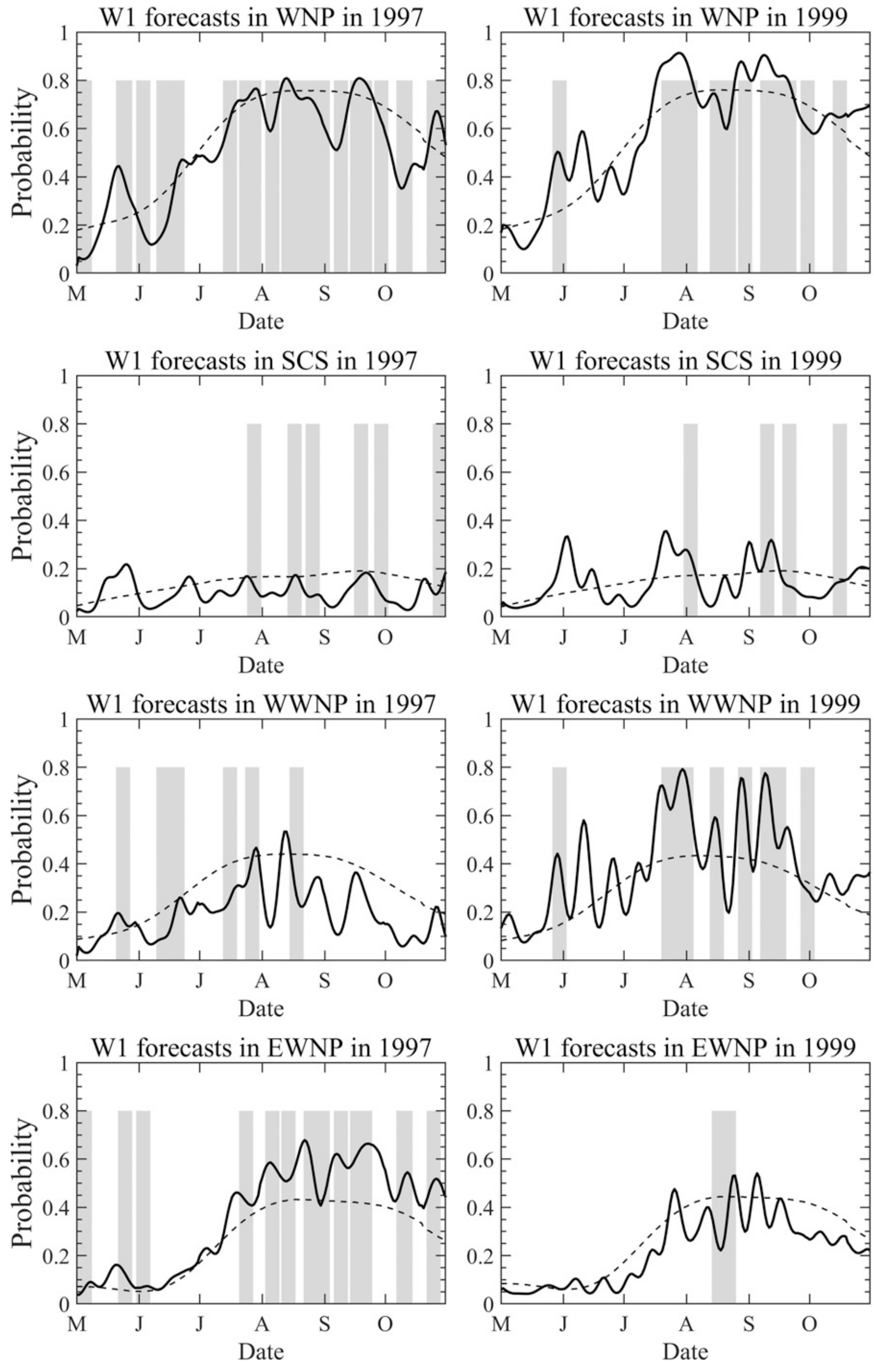

FIG. 6. Cross-validated hindcasts (solid curves) and cross-validated climatology (dashed curves) for W1 of (left) 1997 and (right) 1999 for (top) the full WNP and for the three subregions (top middle) SCS, (bottom middle) WWNP, and (bottom) EWNP. Gray bars represent a week in which TC genesis occurred. Each gray bar lasts at least 1 week because of overlapping weekly probabilities.

forecast leads, both in the WWNP and EWNP subregions. Inclusion of ISO in the model using internal validation generates an improvement greater than ~1%–3% relative to the Clim model during W0–W4 leads for the WWNP but generates almost no improvement at W0 and W1 for the EWNP. In the external validation, ISO provides more skill improvement than ENSO over the WWNP with limited improvement in the first two weeks over the EWNP. In summary, the climatology of TC genesis appears to be the most important predictor for the WWNP and EWNP. Over the WWNP, ISO provides

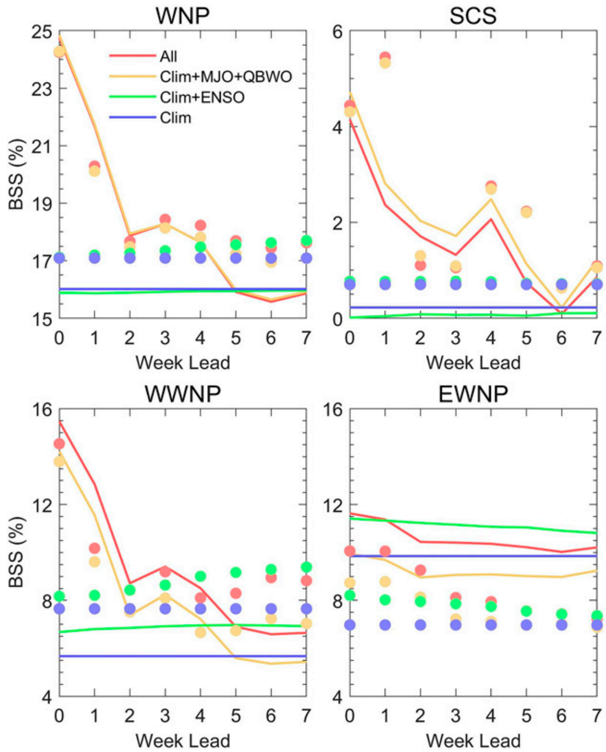

FIG. 7. Brier skill scores of models using both internal validation (lines) and external validation (dots) for the full WNP and for the three subregions SCS, WWNP, and EWNP. The statistical models are developed with a training dataset from 1979 to 2009 (where one year is successively left out of the complete record), and a hindcast for the year left out of the dataset is then produced (i.e., hindcasts for internal validation). We separately test over the period from 2010 to 2019 (i.e., hindcasts of external validation). Shown are the Brier skill scores for models using all potential predictors (ALL) and comparison models using only a subset of selections: intraseasonal oscillations and climatology (Clim + MJO + QBWO), climatology and ENSO (Clim + ENSO), and climatology only (label "Clim").

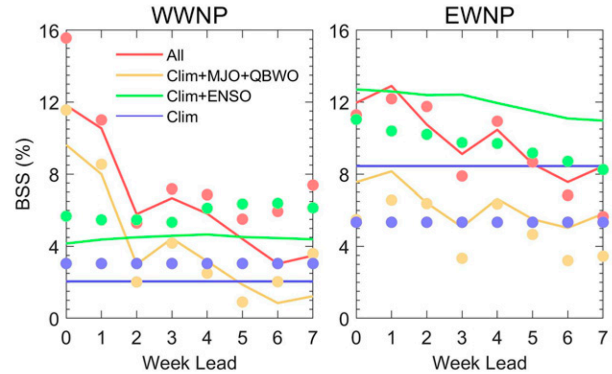

FIG. 8. As in Fig. 7, but for strong ENSO years over (left) WWNP and (right) EWNP.

more skill improvement than ENSO. By contrast, additional skill improvement over the EWNP is mainly from ENSO.

2) ROLE OF REGIONAL DEPENDENCE OF ENSO AND ISO

The inclusion of ENSO over the WWNP and EWNP generates considerable skill increases, with ~1% additional improvement in the models using both internal and external validation relative to the Clim model. As mentioned in section 3b, there are fewer or more TC genesis events over the WWNP during warm or cold ENSO phases, respectively, with the opposite relationship occurring in the EWNP. The analyses of BSSs also support the significant and opposite impact of ENSO on TC genesis in the WWNP and EWNP. We further compare the model skill for strong ENSO years and find that there is a substantial increase in skill by including ENSO in the TC genesis prediction model, with ~3% improvement over the WWNP and ~4% improvement over the EWNP using internal validation (Fig. 8). The relative importance of potential predictors is almost identical in the models using either external or internal validation. Additionally, the impact of ENSO on the prediction skill of the regression models shows an amplification during strong ENSO events relative to all years (Figs. 7 and 8).

As shown in Figs. 4 and 5, both the MJO and QBWO show a convective modulation maximum over the WWNP, and these ISOs have a more important modulation of TC genesis over this region when compared with either the SCS or EWNP. The WWNP also shows the most improvement in model skill by incorporating the ISO. To further clarify the improvements in skill, we now examine the model skill over the WWNP for selected strong MJO and QBWO events. We define strong events to be days when the respective ISOs have amplitudes greater than 2. During the strongly convective MJO phases, there is improved skill relative to the Clim model for the first four weeks except for the W2 lead. In the models using internal validation, there are significant improvements at W1 lead with 9% and 7% additional improvement in the convectively enhanced MJO phases (phases 4–7) and convectively suppressed MJO phases (phase 8 and phases 1–3) respectively (Fig. 9), while only a 6% increase is found for the model using all days as shown in Fig. 7. The stronger modulation of TC genesis during high-amplitude MJO events is also reflected by a large improvement in skill at W0, although W0 provides limited predictability. During active QBWO phases, similar skill improvement is found at the shortest lead times (Fig. 9). As expected, the improvement in skill does not exceed W3 for the QBWO, primarily due to a shorter periodicity for the QBWO. In summary, both the MJO and QBWO can substantially improve model skill, with an increase in skill extending to three weeks for the QBWO and to four weeks for the MJO.

3) TYPHOON-STRENGTH BSS

Over the WNP basin, ~65% of TCs reach typhoon intensity (>63 kt) at some point during their lifetime. In this section, we examine the difference in model skill for prediction of TC genesis for the whole WNP when the maximum lifetime intensity of the TC is a tropical storm (34–63 kt) or a typhoon. As

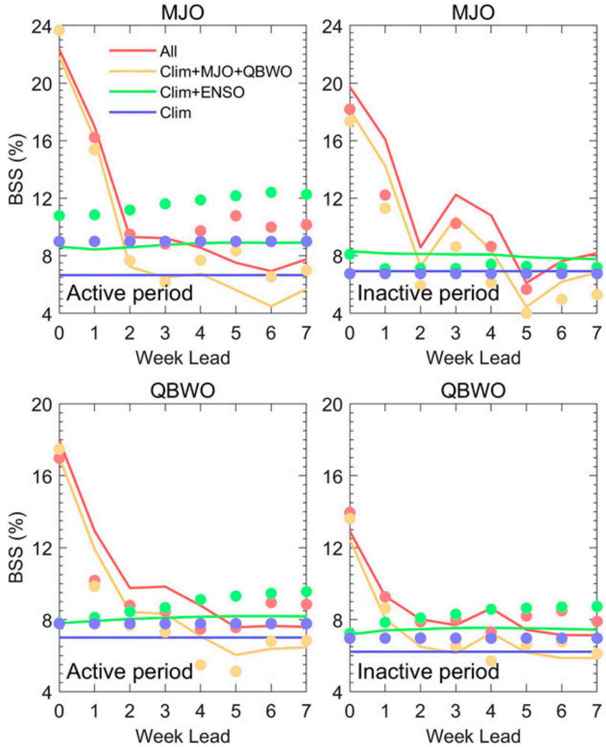

FIG. 9. As in Fig. 7, but for days with amplitude >2 during (top) MJO or (bottom) QBWO (left) active and (right) inactive phases over WWNP subregion.

shown in Fig. 10, the prediction model for typhoons using the "Clim" results in 11% and 13% skill improvement over the reference model using the internal and external validation, respectively. By contrast, there is less improvement for tropical storms with skill increases of 4% and 7% in the models using both internal and external validation. The enhancement provided by intraseasonal variations spans W0–W3 for typhoons and extends to W5 for tropical storms. This difference in skill prediction may be due to the regional dependence of ENSO and ISO. As shown in Fig. 11, most TCs reaching typhoon strength form over the region from 120° to 150°E, which coincides with the WWNP and EWNP, while TCs remaining below typhoon strength primarily form over the SCS and WWNP. Inclusion of ISO improves the skill at W0–W4 lead over the WWNP and at W0–W1 lead over the EWNP, consistent with the distribution of typhoons. Incorporating ISO in the TC genesis model over the SCS and WWNP can extend the skill for tropical storms to a 5-week lead. As was shown earlier, ENSO continues to have a limited impact on the prediction of TCs over the whole WNP on intraseasonal time scales for both typhoons and tropical storms.

b. *Comparison of dynamical and statistical predictions*

We next examine the skill in predicting TC genesis over the WNP between the ECMWF model and the logistic regression model outlined in this study. For verification purposes, hindcasts with the statistical model are generated for the same weeks as are available from the ECMWF model, during the TC season of WNP defined from 1 May to 31 October. Then, hindcasts of the dynamical and statistical models are verified against JTWC TC observations over the WNP.

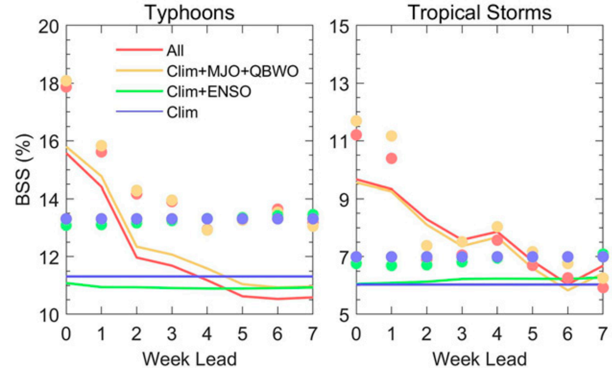

FIG. 10. As in Fig. 7, but for (left) typhoons with maximum wind speed >63 kt and (right) tropical storms with maximum wind speed between 34 and 63 kt over the full WNP basin.

1) Reliability diagrams

We use reliability diagrams (Wilks 1995) to examine the reliability of the dynamical and statistical models. The hindcast probabilities and corresponding observations are binned into 10 equally sized groups and then are averaged for each group for both hindcast and observed probabilities. As shown in Fig. 12, we focused on model skill on intraseasonal time

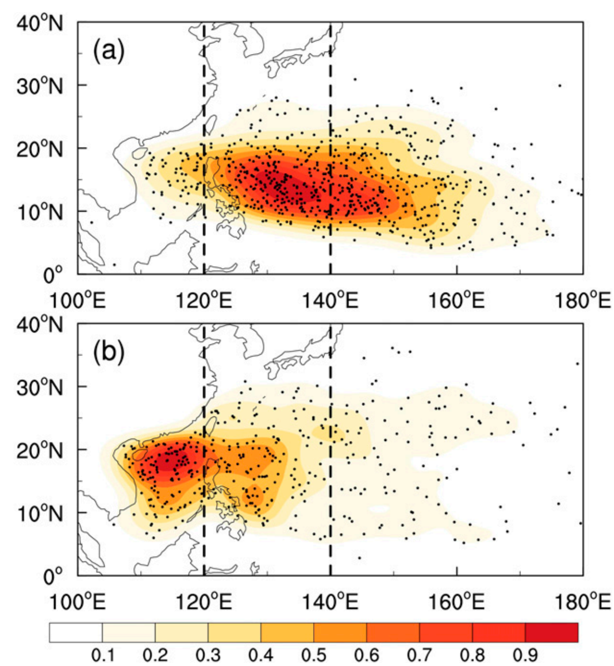

FIG. 11. Probability density distribution of TC genesis for (a) typhoons and (b) tropical storm over the full WNP (shaded areas) from 1 May to 31 Oct for 1979–2019, along with TC genesis locations.

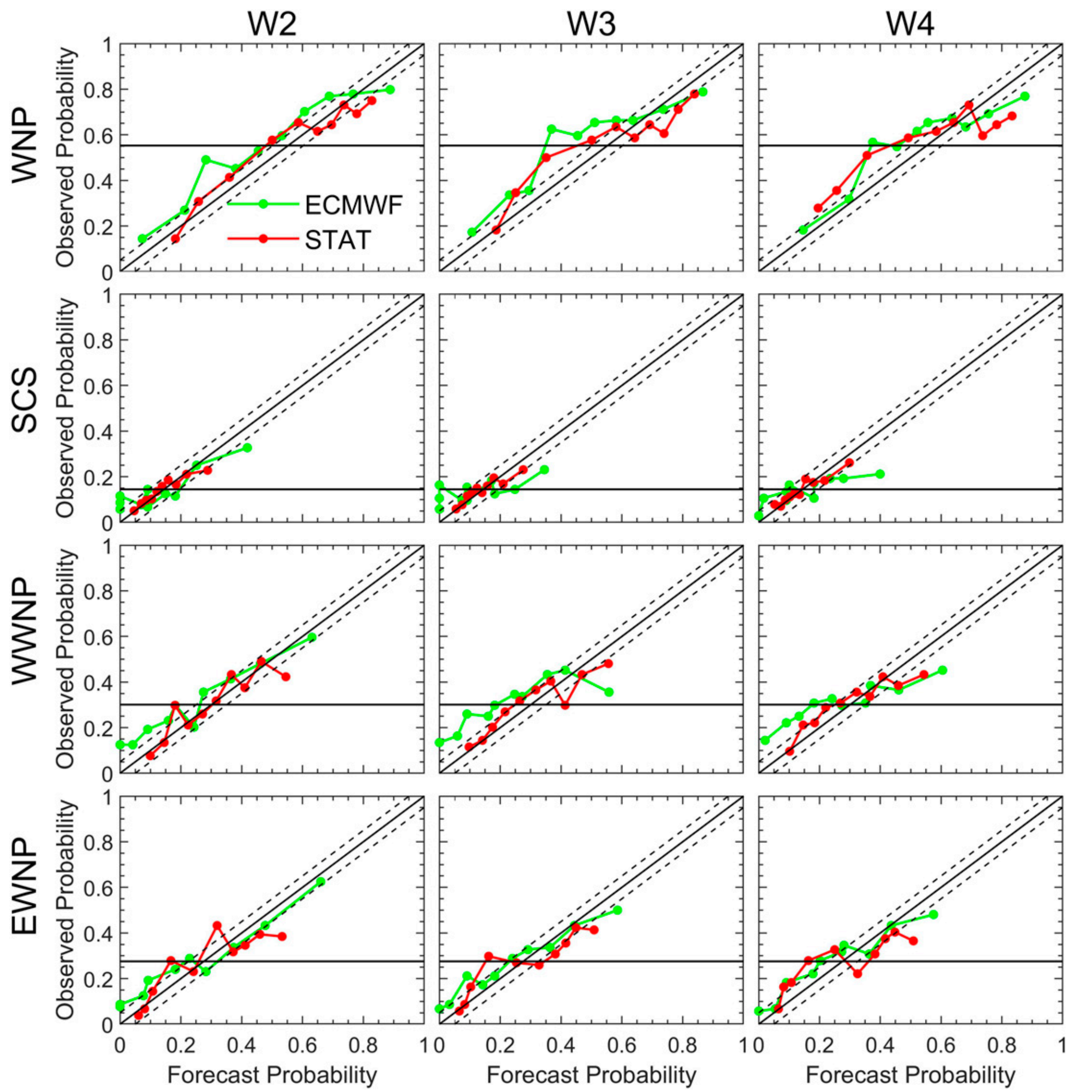

FIG. 12. Reliability diagrams from a week-2 to a week-4 forecast lead over the full WNP and for the three subregions SCS, WWNP, and EWNP. The green line corresponds to the ECMWF model, and the red line corresponds to the statistical model using logistic regression. All hindcasts and their corresponding observations are binned into 10 equal-sized groups according to forecast probabilities. Probabilities of each group are averaged and portrayed as a dot. Dots are connected to form a reliability curve. The horizontal solid lines indicate mean observed probabilities. A perfect forecast is shown by the solid diagonal line. The dashed diagonal lines represent a 10% interval centered on a perfect forecast.

scales (forecast ranges more than 2 weeks) and provide an example of reliability curves of dynamical and statistical models for W2–W4 leads. A 10% interval centered on the perfect forecast (solid diagonal lines) is denoted by the two dashed diagonal lines. The model is considered reliable when the reliability curve lies in the 10% error interval. When the reliability curve lies above the perfect forecast, the model underestimates the probability for TC genesis. In contrast, when the reliability curve lies below the perfect forecast, the model overestimates the probability for TC genesis.

Over the full WNP and the three subregions, TC genesis is underpredicted in the ECMWF model at W2–W4 leads. As the forecast lead increases, the ECMWF model shows a trend to underestimate the low and middle probability groups. Forecasts produced by the statistical models for W2 are more reliable, with the reliability curves of the statistical model closer to the diagonal than the dynamical model. Similarly, the statistical models are shown to be reliable over the SCS and WWNP. However, over the full WNP and the EWNP, the statistical model also underestimates the probability of TC genesis for the low and middle probability groups at W3–W4 leads. At longer forecast leads, underprediction of TC genesis is more common in both the dynamical and statistical models. Overall, the statistical prediction model appears

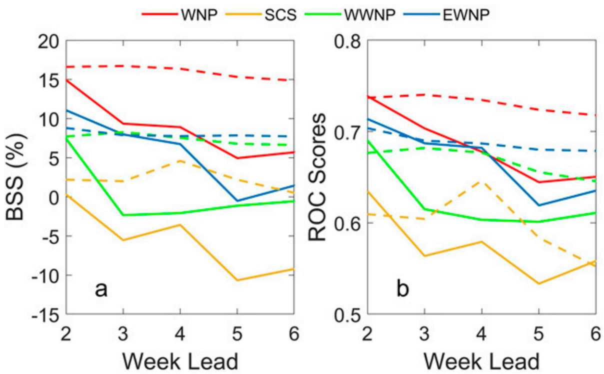

FIG. 13. (a) Brier skill scores and (b) ROC scores of the ECMWF model (solid lines) and the statistical model developed with logistic regression (dashed lines) for the full WNP (red) and the three subregions SCS (orange), the WWNP (green), and EWNP (blue).

more reliable than the ECMWF model during the W2–W4 forecast leads.

2) Probabilistic forecast skill

Probabilistic forecast skill for the ECMWF model is measured using the Brier skill score, using the reference forecast strategy as the statistical model. In addition, the relative operating characteristic (ROC) score (defined as the area under the ROC curve; Buizza and Palmer 1998; Mason and Graham 1999), which relates the hit rate to the corresponding false alarm rate, is also used to assess the skill of the probabilistic forecasts. The ROC score is equal to 1.0 for a perfect forecast and 0.5 for a no-skill forecast.

Previous studies found that the ECMWF model had greater skill in predicting TC occurrence at week 1 than statistical models (Vitart et al. 2010; Lee et al. 2018, 2020). Here we compare the dynamical and statistical models forecast skill at W2–W6 lead times, because it is difficult to eliminate the effect of preexisting storms that are included the hindcasts of the W1 lead generated by the ECMWF model. With forecast lead time increasing, the ECMWF model skill as measured by both BSSs and ROC scores decreases (Fig. 13). As compared with the reference model, the ECMWF model displays positive BSSs at W2–W6 lead over the full WNP and the EWNP, while the model has negative BSSs except for a week-2 lead over the SCS and WWNP. The forecast skill of the statistical model is nearly equal to the ECMWF model at W2 lead, and the statistical model performs better for longer leads than the dynamical model, as shown in both Brier skill scores and ROC scores.

c. *Real-time forecast method for the L2 regression model*

The regression model developed here can be directly applied to forecast weekly TC genesis over the WNP using the ISO indices from the climate model's forecast of the base state. More importantly, to ensure that the developed L2 regression model can be used for real-time forecasting, the real-time ISO indices were extracted by a nonfiltering method following previous studies (Kikuchi et al. 2012; Hsu et al. 2015; Kiladis et al. 2014). In this study, we adopt the nonfiltering method developed by Hsu et al. (2015) to extract the signals of the MJO and QBWO. Specifically, the real-time MJO index can be obtained by the following four steps. First, the climatological annual cycle is removed from the raw data by subtracting a climatological 90-day low-pass filtered component, based on the 1979–2009 period. Then, a 30-day mean of the previous 30 days is subtracted from the anomaly field above to remove lower-frequency variability. We next apply a 5-day running mean to remove higher-frequency variability for the MJO mode. Last, real-time PCs of the MJO mode can be obtained by projecting the OLR anomaly fields onto each mode derived from the intraseasonal time-filtered historical data. Likewise, the real-time QBWO index can be obtained by the following four steps. First, the climatological annual cycle is removed from the raw data by subtracting a climatological 90-day low-pass filtered component, based on the 1979–2009 period. Then, a 10-day mean of the previous 10 days is subtracted from the anomaly field above to remove lower-frequency variability. We next apply a 3-day running mean to remove higher-frequency variability for the QBWO mode. Last, real-time PCs of the QBWO mode can be obtained by projecting the OLR anomaly fields onto each mode derived from the intraseasonal time-filtered historical data. Details on the processing of real-time ISO indices can be found in Hsu et al. (2015).

An example of real-time monitoring of the MJO index and the QBWO index is shown in Fig. 14. The real-time monitoring time series of both the MJO and QBWO modes are able to capture 30–60-day and 10–20-day ISO signals respectively (Figs. 14a,b), with significant correlations of 0.82 and 0.78 for PC1 and PC2, respectively, of the MJO and 0.75 and 0.70 respectively for PC1 and PC2 of the QBWO. Similar consistency is also found for extended periods as shown in Table 2. During May–October of 2010–19, the first two leading PCs correlation coefficients of MJO and QBWO modes using the nonfiltering method and the bandpass-filtering method are ~0.8 and ~0.7, respectively, which is supported by a statistical analysis of 31 years of data during 1979–2009.

Moreover, we compare the model skill during 2010–19 for ISO indices with the bandpass-filtering and nonfiltering methods. Figure 14c shows the differences of BSS values for the regression models using climatology and the ISO (i.e., Clim+MJO+QBWO). Over the full WNP and its three subregions, the absolute differences in model skill based upon the ISO signal extraction using these two methods was small, with no more than a 1.6% difference between these two approaches. To further demonstrate real-time model skill, we compare the hindcasts, skill scores, and reliability diagrams for the whole basin and the three subregions using the real-time indices and bandpass-filtered indices in the regression model and found similar results with just slight differences in amplitude (see Figs. S1–S6 in the online supplemental material). In summary, the developed L2 regression model can be used for real-time forecasts of weekly TC genesis over the WNP on intraseasonal time scales, using a nonfiltering method to extract the ISO signal.

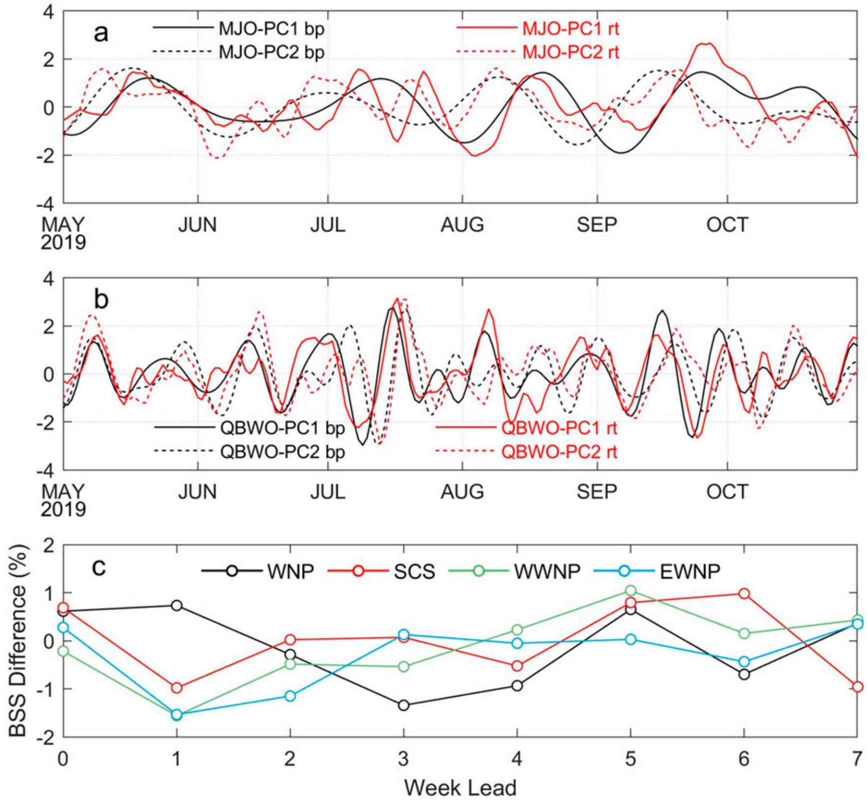

FIG. 14. The first two leading PCs of (a) the MJO mode and (b) the QBWO mode during May–October of 2019. Black lines represent the time series of two PCs obtained using a bandpass-filtered field, and red lines represent the time series of real-time PCs obtained using a nonfiltering method. (c) The difference of BSS values between the nonfiltering method and the bandpass-filtering method for statistical models using climatology and intraseasonal oscillations (i.e., Clim+MJO+QBWO) in May–October of 2010–19. A positive or negative value respectively means a relatively higher or lower skill of model using real-time ISO indices than the model with ISO indices using bandpass filtering.

6. Summary

In this study, we developed a statistical intraseasonal prediction model for WNP TC genesis using L2 regularized logistic regression. Since there are substantial regional differences in physical controlling mechanisms for TCs, we divided the whole WNP basin into three subregions (SCS, WWNP, and EWNP) and developed a statistical prediction model for each of these subregions on intraseasonal time scales using the same approach. For the climatology of TC genesis, two indices representing the MJO, two indices representing the QBWO, and the ONI index characterizing ENSO are chosen as potential predictors for the regression model in this study. Appropriate lags were introduced for each predictor according to their availability in real time. We assessed the relative importance of potential predictors over the whole WNP basin and for the three subregions for W0–W7 lead using a forward stepwise selection procedure. Independent models are developed for the whole WNP and for the three subregions.

In general, the time-varying climatology of TC genesis appears to be the most important predictor and shows increased model skill for predicting weekly TC genesis. By including ISO, the model skill improvement extends to a 4-week lead over the whole WNP basin. By contrast, ENSO has a limited impact on the skill improvement for the whole WNP basin, due to the opposite impact of ENSO on TC genesis between the WWNP and EWNP. In addition to the regional dependence of TC prediction skill on ENSO, the intraseasonal prediction of TC genesis also shows a strong regional dependence on ISO. There is a substantial increase in model skill by

TABLE 2. Correlations of the principal components (PCs) for the MJO mode and the QBWO mode for real-time monitoring using a nonfiltering method and a bandpass-filtering method for the training period (1979–2009) and for the hindcast period (2010–19).

	1979–2009		2010–19	
Period	PC1	PC2	PC1	PC2
MJO	0.82	0.80	0.78	0.77
QBWO	0.73	0.73	0.72	0.74

including ENSO over the WWNP and EWNP, while both the MJO and QBWO improve model skill over the SCS and WWNP but have limited impact over the EWNP. The regional dependence of ENSO and ISO can largely explain the difference in model prediction skill for typhoons and tropical storms over the whole WNP basin. The skill of the statistical prediction models for TC genesis over the WNP basin on intraseasonal time scales is then compared with the ECMWF dynamical model. We find comparable reliability and forecast skill scores to the ECMWF dynamical model at W2, with higher forecast skill scores from W3 to W6 weeks.

The prediction skills are assessed in the models using both internal and external validation, and we find similar skill between models using both internal (1979–2009) and external validation (2010–19). Given the potential impact of decadal climatological differences on model skill (Liu and Chan 2013; Zhao and Wang 2016, 2019; Murakami et al. 2020), we further compared the skills of the model with a climatology of 1979–97 and the model with a climatology of 1998–2009. We found very small differences in skill between the models developed over the two subperiods for the full WNP and for the three subregions (results are not shown). These analyses imply that the statistical intraseasonal prediction model of TC genesis in this study is not sensitive to the decadal background state. Last, we have implemented predictions through real-time MJO and QBWO indices using a nonfiltering method and found that the statistical model shows skill for real-time extended TC genesis over the WNP basin. The forecast skill is also comparable to the MJO and QBWO signals that are extracted using a bandpass filtering.

While the models presented here are promising, more potential predictability for intraseasonal prediction should be considered including other equatorial waves, extratropical teleconnection, and SST anomalies in other ocean basins (Yu et al. 2016; Camargo et al. 2019; Vitart et al. 2019). Key environmental factors and large-scale circulation systems affecting the prediction of TCs on intraseasonal time scales should be further explored by systematically analyzing and assessing the prediction skill of TCs on intraseasonal time scales in current dynamical models. More recently, increasing attention has been paid to forecasts using machine learning and artificial intelligence. Matsuoka et al. (2018) successfully detected TC genesis and its precursors in the WNP using a two-dimensional deep convolutional neural network (CNN) model. Ham et al. (2019) also used a CNN model to better predict ENSO events. The application of deep-learning approaches may yield improved skill at forecasting TC activity on intraseasonal time scales.

Acknowledgments. This research was jointly supported by the National Natural Science Foundation of China (Grants 42192551, 41922033, 41730961, and 42005017). Author Klotzbach acknowledges a grant from the G. Unger Vetlesen Foundation. The numerical calculations in this study have been done on the supercomputing system at the Supercomputing Center of Nanjing University of Information Science and Technology.

Data availability statement. The data used in this paper are available online from the following sources: Joint Typhoon Warning Center (JTWC) best-track data (http://www.metoc.navy.mil/jtwc/), NOAA Interpolated OLR (https://psl.noaa.gov/data/gridded/data.interp_OLR.html), HadISST (https://www.metoffice.gov.uk/hadobs/index.html), and ECMWF track predictions of S2S reforecasts (https://acquisition.ecmwf.int/ecpds/data/list/).

REFERENCES

Bakkensen, L. A., D.-S. R. Park, and R. S. R. Sarkar, 2018: Climate costs of tropical cyclone losses also depend on rain. *Environ. Res. Lett.*, **13**, 074034, https://doi.org/10.1088/1748-9326/aad056.

Bauer, P., A. J. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525**, 47–55, https://doi.org/10.1038/nature14956.

Bender, M. A., T. R. Knutson, R. E. Tuleya, J. J. Sirutis, G. A. Vecchi, S. T. Garner, and I. M. Held, 2010: Modeled impact of anthropogenic warming on the frequency of intense Atlantic hurricanes. *Science*, **327**, 454–458, https://doi.org/10.1126/science.1180568.

Buizza, R., and T. N. Palmer, 1998: Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.*, **126**, 2503–2518, https://doi.org/10.1175/1520-0493(1998)126<2503:IOESOE>2.0.CO;2.

Camargo, S. J., and A. H. Sobel, 2005: Western North Pacific tropical cyclone intensity and ENSO. *J. Climate*, **18**, 2996–3006, https://doi.org/10.1175/JCLI3457.1.

——, A. G. Barnston, and S. E. Zebiak, 2005: A statistical assessment of tropical cyclone activity in atmospheric general circulation models. *Tellus*, **57A**, 589–604, https://doi.org/10.3402/tellusa.v57i4.14705.

——, K. A. Emanuel, and A. H. Sobel, 2007: Use of a genesis potential index to diagnose ENSO effects on tropical cyclone genesis. *J. Climate*, **20**, 4819–4834, https://doi.org/10.1175/JCLI4282.1.

——, 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.

——, and Coauthors, 2019: Tropical cyclone prediction on subseasonal time-scales. *Trop. Cyclone Res. Rev.*, **8**, 150–165, https://doi.org/10.1016/j.tcrr.2019.10.004.

Chan, J. C. L., J. Shi, and K. S. Liu, 2001: Improvements in the seasonal forecasting of tropical cyclone activity over the western North Pacific. *Wea. Forecasting*, **16**, 491–498, https://doi.org/10.1175/1520-0434(2001)016<0491:IITSFO>2.0.CO;2.

Cheng, Q., P. K. Varshney, and M. K. Arora, 2006: Logistic regression for feature selection and soft classification of remote sensing data. *IEEE Geosci. Remote Sens. Lett.*, **3**, 491–494, https://doi.org/10.1109/LGRS.2006.877949.

Chu, J.-H., C. R. Sampson, A. S. Levine, and E. Fukada, 2002: The Joint Typhoon Warning Center tropical cyclone best-tracks, 1945–2000. Ref. NRL/MR/754002, 16 pp., http://www.usno.navy.mil/NOOC/nmfc-ph/RSS/jtwc/best_tracks/TC_bt_report.html.

Duchon, C. E., 1979: Lanczos filtering in one and two dimensions. *J. Appl. Meteor.*, **18**, 1016–1022, https://doi.org/10.1175/1520-0450(1979)018<1016:LFIOAT>2.0.CO;2.

Elsberry, R. L., M. S. Jordan, and F. Vitart, 2010: Predictability of tropical cyclone events on intraseasonal timescales with the

ECMWF monthly forecast model. *Asia-Pac. J. Atmos. Sci.*, **46**, 135–153, https://doi.org/10.1007/s13143-010-0013-4.

——, 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.

Emanuel, K. A., and D. S. Nolan, 2004: Tropical cyclone activity and global climate. Preprints, *26th Conf. on Hurricanes and Tropical Meteorology*, Miami, FL, Amer. Meteor. Soc., 240–241.

Goldenberg, S. B., and L. J. Shapiro, 1996: Physical mechanisms for the association of El Niño and West African rainfall with Atlantic major hurricane activity. *J. Climate*, **9**, 1169–1187, https://doi.org/10.1175/1520-0442(1996)009<1169:PMFTAO>2.0.CO;2.

Gray, W. M., 1979: Hurricanes: Their formation, structure and likely role in the general circulation. *Meteorology over the Tropical Oceans*, D. B. Shaw, Ed., Royal Meteorological Society, 155–218.

——, 1984: Atlantic seasonal hurricane frequency. Part II: Forecasting its variability. *Mon. Wea. Rev.*, **112**, 1669–1683, https://doi.org/10.1175/1520-0493(1984)112<1669:ASHFPI>2.0.CO;2.

——, C. W. Landsea, P. W. Mielke, and K. J. Berry, 1992: Predicting Atlantic seasonal hurricane activity 6–11 months in advance. *Wea. Forecasting*, **7**, 440–455, https://doi.org/10.1175/1520-0434(1992)007<0440:PASHAM>2.0.CO;2.

Gregory, P., F. Vitart, R. Rivett, A. Brown, and Y. Kuleshov, 2020: Subseasonal forecasts of tropical cyclones in the Southern Hemisphere using a dynamical multimodel ensemble. *Wea. Forecasting*, **35**, 1817–1829, https://doi.org/10.1175/WAF-D-20-0050.1.

Ham, Y.-G., J.-H. Kim, and J.-J. Luo, 2019: Deep learning for multi-year ENSO forecasts. *Nature*, **573**, 568–572, https://doi.org/10.1038/s41586-019-1559-7.

Han, X., H. Zhao, X. Li, G. B. Raga, C. Wang, and Q. Li, 2019: Modulation of boreal extended summer tropical cyclogenesis over the northwest Pacific by the quasi-biweekly oscillation under different El Niño–Southern Oscillation phases. *Int. J. Climatol.*, **40**, 858–873, https://doi.org/10.1002/joc.6244.

Hansen, K. A., S. J. Majumdar, and B. P. Kirtman, 2020: Identifying subseasonal variability relevant to Atlantic tropical cyclone activity. *Wea. Forecasting*, **35**, 2001–2024, https://doi.org/10.1175/WAF-D-19-0260.1.

Henderson, D. A., and D. R. Denison, 1989: Stepwise regression in social and psychological research. *Psychol. Rep.*, **64**, 251–257, https://doi.org/10.2466/pr0.1989.64.1.251.

Hoerl, A. E., and R. W. Kennard, 1970: Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55–67, https://doi.org/10.1080/00401706.1970.10488634.

Hsu, P.-C., T. Li, L. You, J. Gao, and H.-L. Ren, 2015: A spatial–temporal projection model for 10–30 day rainfall forecast in South China. *Climate Dyn.*, **44**, 1227–1244, https://doi.org/10.1007/s00382-014-2215-4.

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.

——, B. Xiang, M. Zhao, T. Li, S. J. Lin, Z. Wang, and J. H. Chen, 2018: Intraseasonal tropical cyclogenesis prediction in a global coupled model system. *J. Climate*, **31**, 6209–6227, https://doi.org/10.1175/JCLI-D-17-0454.1.

——, A. Adames, D. Kim, E. Maloney, H. Lin, H. Kim, C. Zhang, C. DeMott, and N. Klingaman, 2020: Fifty years of research on the Madden–Julian oscillation: Recent progress, challenges, and perspectives. *J. Geophys. Res. Atmos.*, **125**, e2019JD030911, https://doi.org/10.1029/2019JD030911.

Kemball-Cook, S., and B. Wang, 2001: Equatorial waves and air–sea interaction in the boreal summer intraseasonal oscillation. *J. Climate*, **14**, 2923–2942, https://doi.org/10.1175/1520-0442(2001)014<2923:EWAASI>2.0.CO;2.

Kikuchi, K., B. Wang, and Y. Kajikawa, 2012: Bimodal representation of the tropical intraseasonal oscillation. *Climate Dyn.*, **38**, 1989–2000, https://doi.org/10.1007/s00382-011-1159-1.

Kiladis, G. N., J. Dias, K. H. Straub, M. C. Wheeler, S. N. Tulich, K. Kikuchi, K. M. Weickmann, and M. J. Ventrice, 2014: A comparison of OLR and circulation-based indices for tracking the MJO. *Mon. Wea. Rev.*, **142**, 1697–1715, https://doi.org/10.1175/MWR-D-13-00301.1.

Kim, H., F. Vitart, and D. E. Waliser, 2018: Prediction of the Madden–Julian oscillation: A review. *J. Climate*, **31**, 9425–9443, https://doi.org/10.1175/JCLI-D-18-0210.1.

——, M. A. Janiga, and K. Pegion, 2019: MJO propagation processes and mean biases in the SubX and S2S reforecasts. *J. Geophys. Res. Atmos.*, **124**, 9314–9331, https://doi.org/10.1029/2019JD031139.

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.

Kleinen, J., 2007: Historical perspectives on typhoons and tropical storms in the natural and socio-economic system of Nam Dinh (Vietnam). *J. Asian Earth Sci.*, **29**, 523–531, https://doi.org/10.1016/j.jseaes.2006.05.012.

Klotzbach, P. J., and E. C. J. Oliver, 2015: Variations in global tropical cyclone activity and the Madden–Julian oscillation since the midtwentieth century. *Geophys. Res. Lett.*, **42**, 4199–4207, https://doi.org/10.1002/2015GL063966.

——, S. G. Bowen, R. Pielke, and M. Bell, 2018: Continental U.S. hurricane landfall frequency and associated damage: Observations and future risks. *Bull. Amer. Meteor. Soc.*, **99**, 1359–1376, https://doi.org/10.1175/BAMS-D-17-0184.1.

——, and Coauthors, 2019: Seasonal tropical cyclone forecasting. *Trop. Cyclone Res. Rev.*, **8**, 134–149, https://doi.org/10.1016/j.tcrr.2019.10.003.

——, L.-P. Caron, and M. M. Bell, 2020: A statistical/dynamical model for North Atlantic seasonal hurricane prediction. *Geophys. Res. Lett.*, **47**, e2020GL089357, https://doi.org/10.1029/2020GL089357.

Lander, M. A., 1994: An exploratory analysis of the relationship between tropical storm formation in the western North Pacific and ENSO. *Mon. Wea. Rev.*, **122**, 636–651, https://doi.org/10.1175/1520-0493(1994)122<0636:AEAOTR>2.0.CO;2.

Lee, C. Y., S. J. Camargo, F. Vitart, A. H. Sobel, and M. K. Tippett, 2018: Subseasonal tropical cyclone genesis prediction and MJO in the S2S dataset. *Wea. Forecasting*, **33**, 967–988, https://doi.org/10.1175/WAF-D-17-0165.1.

——, ——, ——, ——, J. Camp, S. Wang, M. K. Tippett, and Q. Yang, 2020: Subseasonal predictions of tropical cyclone occurrence and ACE in the S2S dataset. *Wea. Forecasting*, **35**, 921–938, https://doi.org/10.1175/WAF-D-19-0217.1.

Lee, J.-Y., B. Wang, M. C. Wheeler, X. Fu, D. E. Waliser, and I.-S. Kang, 2013: Real-time multivariate indices for the boreal summer intraseasonal oscillation over the Asian summer monsoon region. *Climate Dyn.*, **40**, 493–509, https://doi.org/10.1007/s00382-012-1544-4.

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, C.-Y., 1996: Quasi-two weeks oscillation in the tropical atmosphere. *Theor. Appl. Climatol.*, **55**, 121–127, https://doi.org/10.1007/BF00864707.

Li, R. C. Y., and W. Zhou, 2013a: 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.

——, and ——, 2013b: Modulation of western North Pacific tropical cyclone activity by the ISO. Part II: Tracks and landfalls. *J. Climate*, **26**, 2919–2930, https://doi.org/10.1175/JCLI-D-12-00211.1.

——, ——, J. Chan, and P. Huang, 2012: Asymmetric modulation of western North Pacific cyclogenesis by the Madden–Julian oscillation under ENSO conditions. *J. Climate*, **25**, 5374–5385, https://doi.org/10.1175/JCLI-D-11-00337.1.

Liebmann, B., and C. A. Smith, 1996: Description of a complete (interpolated) outgoing longwave radiation dataset. *Bull. Amer. Meteor. Soc.*, **77**, 1275–1277, https://doi.org/10.1175/1520-0477-77.6.1274.

Lim, Y., S.-W. Son, and D. Kim, 2018: MJO prediction skill of the subseasonal-to-seasonal prediction models. *J. Climate*, **31**, 4075–4094, https://doi.org/10.1175/JCLI-D-17-0545.1.

Liu, K. S., and J. C. L. Chan, 2013: Inactive period of western North Pacific tropical cyclone activity in 1998–2011. *J. Climate*, **26**, 2614–2630, https://doi.org/10.1175/JCLI-D-12-00053.1.

Madden, R. A., and P. R. Julian, 1971: Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **28**, 702–708, https://doi.org/10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2.

Maloney, E. D., and D. L. Hartmann, 2000: 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.

Mao, J. Y., and J. C. L. Chan, 2005: Intraseasonal variability of the South China Sea summer monsoon. *J. Climate*, **18**, 2388–2402, https://doi.org/10.1175/JCLI3395.1.

Mason, S. J., and N. E. Graham, 1999: Conditional probabilities relative operating characteristics, and relative operating levels. *Wea. Forecasting*, **14**, 713–725, https://doi.org/10.1175/1520-0434(1999)014<0713:CPROCA>2.0.CO;2.

Matsuoka, D., M. Nakano, D. Sugiyama, and S. Uchida, 2018: Deep learning approach for detecting tropical cyclones and their precursors in the simulation by a cloud-resolving global nonhydrostatic atmospheric model. *Prog. Earth Planet. Sci.*, **5**, 80, https://doi.org/10.1186/s40645-018-0245-y.

McBride, J. L., and R. Zehr, 1981: Observational analysis of tropical cyclone formation. Part II: Comparison of non-developing versus developing systems. *J. Atmos. Sci.*, **38**, 1132–1151, https://doi.org/10.1175/1520-0469(1981)038<1132:OAOTCF>2.0.CO;2.

Merryfield, W. J., and Coauthors, 2020: Current and emerging developments in subseasonal to decadal prediction. *Bull. Amer. Meteor. Soc.*, **101**, E869–E896, https://doi.org/10.1175/BAMS-D-19-0037.1.

Mundry, R., and C. L. Nunn, 2009: Stepwise model fitting and statistical inference: Turning noise into signal pollution. *Amer. Nat.*, **173**, 119–123, https://doi.org/10.1086/593303.

Murakami, H., T. L. Delworth, W. F. Cooke, M. Zhao, B. Xiang, and P. C. Hsu, 2020: Detected climatic change in global distribution of tropical cyclones. *Proc. Natl. Acad. Sci. USA*, **117**, 10 706–10 714, https://doi.org/10.1073/pnas.1922500117.

Murakami, M., 1984: Analysis of deep convective activity over the western Pacific and Southeast Asia. Part II: Seasonal and intraseasonal variations during northern summer. *J. Meteor. Soc. Japan*, **62**, 88–108, https://doi.org/10.2151/jmsj1965.62.1_88.

Ng, A. Y., 2004: Feature selection, L1 vs. L2 regularization, and rotational invariance. *Proc. 21st Int. Conf. on Machine Learning*, Banff, AB, Canada, ACM, https://doi.org/10.1145/1015330.1015435.

Ogutu, J. O., T. Schulz-Streeck, and H. P. Piepho, 2012: Genomic selection using regularized linear regression models: Ridge regression, lasso, elastic net and their extensions. *BMC Proc.*, **6**, S10, https://doi.org/10.1186/1753-6561-6-S2-S10.

Patricola, C. M., S. J. Camargo, P. J. Klotzbach, R. Saravanan, and P. Chang, 2018: The influence of ENSO flavors on western North Pacific tropical cyclone activity. *J. Climate*, **31**, 5395–5416, https://doi.org/10.1175/JCLI-D-17-0678.1.

Rasmusson, E. M., and T. H. Carpenter, 1982: Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño. *Mon. Wea. Rev.*, **110**, 354–384, https://doi.org/10.1175/1520-0493(1982)110<0354:VITSST>2.0.CO;2.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, https://doi.org/10.1029/2002JD002670.

Saunders, M. A., and A. S. Lea, 2005: Seasonal prediction of hurricane activity reaching the coast of the United States. *Nature*, **434**, 1005–1008, https://doi.org/10.1038/nature03454.

Serra, Y. L., X. Jiang, B. Tian, J. Astua, E. D. Maloney, and G. N. Kiladis, 2014: Tropical intra-seasonal oscillations and synoptic variability. *Annu. Rev. Environ. Resour.*, **39**, 189–215, https://doi.org/10.1146/annurev-environ-020413-134219.

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.

Smith, G., 2018: Step away from stepwise. *J. Big Data*, **5**, 32, https://doi.org/10.1186/s40537-018-0143-6.

Thompson, B., 1995: Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educ. Psychol. Meas.*, **55**, 525–534, https://doi.org/10.1177/0013164495055004001.

Vecchi, G. A., M. Zhao, H. Wang, G. Villarini, A. Rosati, A. Kumar, I. M. Held, and R. Gudgel, 2011: Statistical–dynamical predictions of seasonal North Atlantic hurricane activity. *Mon. Wea. Rev.*, **139**, 1070–1082, https://doi.org/10.1175/2010MWR3499.1.

——, and Coauthors, 2019: Tropical cyclone sensitivities to CO_2 doubling: Roles of atmospheric resolution, synoptic variability and background climate changes. *Climate Dyn.*, **53**, 5999–6033, https://doi.org/10.1007/s00382-019-04913-y.

Vitart, F., 2014: Evolution of ECMWF sub-seasonal forecast skill scores. *Quart. J. Roy. Meteor. Soc.*, **140**, 1889–1899, https://doi.org/10.1002/qj.2256.

——, 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.

——, and A. W. Robertson, 2018: The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *npj Climate Atmos. Sci.*, **1**, 3, https://doi.org/10.1038/s41612-018-0013-0.

——, 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.

——, and Coauthors, 2019: *Sub-seasonal to Seasonal Prediction of Weather Extremes.* Elsevier, 365–386, https://doi.org/10.1016/B978-0-12-811714-9.00017-6.

Wang, B., and J. C. L. Chan, 2002: How strong ENSO events affect tropical storm activity over the western North Pacific. *J. Climate*, **15**, 1643–1658, https://doi.org/10.1175/1520-0442(2002)015<1643:HSEEAT>2.0.CO;2.

——, R. Wu, and X. Fu, 2000: Pacific–East Asian teleconnection: How does ENSO affect East Asian climate? *J. Climate*, **13**, 1517–1536, https://doi.org/10.1175/1520-0442(2000)013<1517:PEATHD>2.0.CO;2.

Wang, G., J. Su, Y. Ding, and D. Chen, 2007: Tropical cyclone genesis over the South China Sea. *J. Mar. Syst.*, **68**, 318–326, https://doi.org/10.1016/j.jmarsys.2006.12.002.

Wang, S., A. H. Sobel, M. K. Tippett, and F. Vitart, 2019: Prediction and predictability of tropical intraseasonal convection: Seasonal dependence and the Maritime Continent prediction barrier. *Climate Dyn.*, **52**, 6015–6031, https://doi.org/10.1007/s00382-018-4492-9.

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.

Whittingham, M. J., P. A. Stephens, R. B. Bradbury, and R. P. Freckleton, 2006: Why do we still use stepwise modelling in ecology and behavior? *J. Anim. Ecol.*, **75**, 1182–1189, https://doi.org/10.1111/j.1365-2656.2006.01141.x.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction.* Academic Press, 464 pp.

Yu, J., T. Li, Z. Tan, and Z. Zhu, 2016: Effects of tropical North Atlantic SST on tropical cyclone genesis in the western North Pacific. *Climate Dyn.*, **46**, 865–877, https://doi.org/10.1007/s00382-015-2618-x.

Zhang, G., H. Murakami, T. R. Knutson, R. Mizuta, and K. Yoshida, 2020: Tropical cyclone motion in a changing climate. *Sci. Adv.*, **6**, eaaz7610, https://doi.org/10.1126/sciadv.aaz7610.

Zhang, Q., L. Wu, and Q. Liu, 2009: Tropical cyclone damages in China 1983–2006. *Bull. Amer. Meteor. Soc.*, **90**, 489–495, https://doi.org/10.1175/2008BAMS2631.1.

Zhao, H., and C. Wang, 2016: Interdecadal modulation on the relationship between ENSO and typhoon activity during the late season in the western North Pacific. *Climate Dyn.*, **47**, 315–328, https://doi.org/10.1007/s00382-015-2837-1.

——, and ——, 2019: On the relationship between ENSO and tropical cyclones in the western North Pacific during the boreal summer. *Climate Dyn.*, **52**, 275–288, https://doi.org/10.1007/s00382-018-4136-0.

——, L. Wu, and W. Zhou, 2010: Assessing the influence of the ENSO on tropical cyclone prevailing tracks in the western North Pacific. *Adv. Atmos. Sci.*, **27**, 1361–1371, https://doi.org/10.1007/s00376-010-9161-9.

——, ——, and ——, 2011: Interannual changes of tropical cyclone intensity in the western North Pacific. *J. Meteor. Soc. Japan*, **89**, 243–253, https://doi.org/10.2151/jmsj.2011-305.

——, X. Jiang, and L. Wu, 2015a: Modulation of Northwest Pacific tropical cyclone genesis by the intraseasonal variability. *J. Meteor. Soc. Japan*, **93**, 81–97, https://doi.org/10.2151/jmsj.2015-006.

——, R. Yoshida, and G. B. Raga, 2015b: Impact of the Madden–Julian oscillation on western North Pacific tropical cyclogenesis associated with large-scale patterns. *J. Appl. Meteor. Climatol.*, **54**, 1413–1429, https://doi.org/10.1175/JAMC-D-14-0254.1.

——, C. Wang, and R. Yoshida, 2016: Modulation of tropical cyclogenesis in the western North Pacific by the quasi-biweekly oscillation. *Adv. Atmos. Sci.*, **33**, 1361–1375, https://doi.org/10.1007/s00376-016-5267-z.

Zhou, F., and Coauthors, 2020: Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. *Lancet*, **395**, 1054–1062, https://doi.org/10.1016/S0140-6736(20)30566-3.

Zhou, H., P. Hsu, and Y. Qian, 2018: Close linkage between quasi-biweekly oscillation and tropical cyclone intensification over the western North Pacific. *Atmos. Sci. Lett.*, **19**, e826, https://doi.org/10.1002/asl.826.

Zhu, Z., T. Li, L. Bai, and J. Gao, 2017: Extended-range forecast for the temporal distribution of clustering tropical cyclogenesis over the western North Pacific. *Theor. Appl. Climatol.*, **130**, 865–877, https://doi.org/10.1007/s00704-016-1925-4.