Global Verification of Tropical Cyclone Genesis Forecasts from Five Global Numerical Models

DANIEL J. HALPERIN[a] AND ROBERT E. HART[b]

[a] *Embry-Riddle Aeronautical University, Daytona Beach, Florida*
[b] *Florida State University, Tallahassee, Florida*

(Manuscript received 15 July 2024, in final form 18 December 2024, accepted 28 February 2025)

ABSTRACT: Prior studies by the authors have documented the verification statistics of disturbance-based tropical cyclone (TC) genesis forecasts over the North Atlantic (AL) and eastern North Pacific (EP) basins, which led to the development of real-time probabilistic TC genesis guidance based on multiple logistic regression [the TC Logistic Guidance for Genesis (TCLOGG)]. This study provides a substantial update to that prior work by analyzing a more recent period (2017–22) with one additional model [Navy Global Environmental Model (NAVGEM)], expanding the forecast period temporally to 7 days, and expanding the study domain spatially to include all basins [except the central North Pacific (CP) basin, where the sample size of TC genesis events was too small to generate meaningful statistics]. TC genesis forecasts from five global models are verified against the NHC's and JTWC's best tracks. Verification statistics exhibit nontrivial interannual and model-to-model variability rendering it unfeasible to attempt to define repeatable performance rankings among the models. Nevertheless, results indicate that the ECMWF model exhibits the largest mean success ratio (SR) overall, while the UKMO and GFS models exhibit the greatest probability of detection (POD). All models exhibit a clear trade-off between SR and POD, yielding mean critical success index values < 0.35 for any individual model and basin. The ECMWF and UKMO models exhibit the greatest critical success index (CSI) values globally. This study provides additional evidence that some best track TC genesis events can be detected at least 1 week in advance, but maximum lead times are inconsistent. The resulting dataset of verified forecasts will serve as an updated training dataset for enhanced and updated TCLOGG products.

KEYWORDS: Tropical cyclones; Forecast verification/skill; Numerical weather prediction/forecasting; Model evaluation/performance

1. Introduction

Accurately predicting tropical cyclone (TC) genesis remains a priority for the research, operational forecasting, and numerical model development communities (e.g., Yang 2018; Dunion et al. 2023; Hon et al. 2023; Rajasree et al. 2023). Output from numerical models and forecast guidance products that rely on postprocessed numerical model output are often primary sources of TC genesis guidance for operational forecasters. Prior studies have analyzed TC genesis forecast verification statistics at various lead times (e.g., Briegel and Frank 1997; Chan and Kwok 1999; Cheung and Elsberry 2002; Elsberry et al. 2009, 2010, 2011; Tsai et al. 2011). Improvements over time in numerical models' ability to analyze and predict TC-like structures in the forecast fields have been well documented. While models struggled to predict TC genesis on the order of a few days during the 1990s and 2000s (e.g., Beven 1999; Pasch et al. 2006, 2008; Schumacher et al. 2009), there was increased ability to predict TC genesis out to a week or more during the 2010s (e.g., Halperin et al. 2013; Komaromi and Majumdar 2015; Halperin et al. 2016; Chen et al. 2019a) due to increased computational capabilities which supported improved model resolution, more sophisticated physics parameterizations, and advancements in data assimilation. While previous studies noted interannual variability with respect to TC genesis forecast skill, it was unclear whether such variability existed due to differences in model configuration, large-scale flow patterns, or genesis pathways (e.g., McTaggart-Cowan et al. 2008, 2013; Russell et al. 2017; Whitaker and Maloney 2018; Núñez Ocasio et al. 2020, 2021; Zhan et al. 2022) from which TC genesis events occurred, or a combination of such factors.

Recent studies have helped quantify these impacts (e.g., Wang et al. 2018; Chen et al. 2019a; Teng et al. 2020; Halperin et al. 2020; Li et al. 2020; Liang et al. 2021, 2022; Lin et al. 2023). Halperin et al. (2020) compared a homogeneous set of (re)forecasts from three operational configurations of the GFS model (2013–14, 2015 [v12.0.0], and 2016 [v.13.0.2]) and found TC genesis performance over the North Atlantic (AL) and eastern North Pacific (EP) basins generally degraded between 2013–14 and 2015, then improved in 2016 in terms of critical success index (CSI). Chen et al. (2019a) compared a homogeneous set of forecast initializations between the then-operational spectral GFS and two versions of the GFS using the finite-volume cubed-sphere (FV3) dynamical core. They found that the FV3-based GFS configurations produced greater TC genesis success ratios over most basins and longer TC genesis forecast lead times over all basins compared to the spectral GFS.

Additionally, following the classification of TC genesis events by pathway, recent studies have linked greater TC genesis predictably in model forecasts to nonbaroclinic, low-level baroclinic, and (monsoon) trough TC genesis pathways (Elsberry et al. 2014; Wang et al. 2018; Lin et al. 2023), the monsoon shear line (Liang

Corresponding author: Daniel J. Halperin, daniel.halperin@erau.edu

DOI: 10.1175/WAF-D-24-0130.1

TABLE 1. The output grid spacing for each model used in this study.

Model	Δx	Δy
CMC	1°	1°
ECM	0.5°	0.5°
GFS	0.5°	0.5°
NAV	0.5°	0.5°
UKM	0.140625°	0.09375°

et al. 2021), strong convectively enhanced periods of the MJO (Jiang et al. 2018), and the ability of models to predict MJO–TC interactions (Lee et al. 2018).

Additional studies investigated model biases that led to false alarm forecasts. For example, the GFS has a documented bias of producing false alarm forecasts over the eastern main development region (MDR) in the AL basin. Brammer et al. (2018) found that the GEFS overintensified an African easterly wave due to short-term precipitation errors, which led to increased vorticity and vertical alignment of the circulation. There was also evidence of strong convective feedbacks in some versions of the GFS based on composites of false alarm forecasts over the AL basin (Halperin et al. 2020).

The dataset of TC genesis forecasts previously verified by the authors (Halperin et al. 2013, 2016; hereafter H13 and H16, respectively) was used as the training dataset for the development of the TC Logistic Guidance for Genesis (TCLOGG; Halperin et al. 2017). The real-time disturbance-specific probabilistic TC genesis guidance from TCLOGG was initially implemented for the AL and EP basins to provide guidance for the NHC's Tropical Weather Outlook (TWO). The present study provides an update to H16 by examining TC genesis verification statistics during the period 2017–22, including the Navy Global Environmental Model (NAVGEM; Hogan et al. 2014) in the analysis, and expanding the analysis spatially to include the JTWC's area of responsibility and temporally to include TC genesis forecast events out to 7 days. The central North Pacific (CP) basin is excluded from this study due to the small sample size of TC genesis events during the study period ($N = 2$). The forecasts verified here serve as the training dataset for an enhanced and expanded TCLOGG product suite that provides real-time guidance globally.

As in numerous prior studies on forecast techniques improvement, the scope of the current study is focused on the background, method, statistics, and performance summary of the approach, since the statistical results and rigor are essential for demonstrating that contribution. Therefore, we do not focus on in-depth physical interpretation and theoretical implications of the results herein, although such interpretation will be the goal of a future study.

2. Data and methods

a. Data

A local archive of operational global model output was maintained during the study period for the following models: the Global Deterministic Prediction System from Environment and Climate Change Canada (CMC), the Integrated Forecasting System from the ECMWF (ECM), GFS, NAVGEM, and the global model from the UKMO (UKM). Table 1 provides the output grid spacing for each model. While the data were interpolated to 0.5° or 1° grids for most models, the UKM data were available at the native model resolution. Genesis was defined as the first instance of a tropical depression (TD) or tropical storm (TS) in the Automated TC Forecast (ATCF; Sampson and Schrader 2000) system b decks (i.e., best tracks; Landsea and Franklin 2013) from the NHC (AL and EP basins) and the JTWC [western North Pacific (WP), north Indian Ocean (IO), and Southern Hemisphere (SH) basins]. The period of study is 2017–22 in the ATCF b decks. For the SH basin, the ATCF year runs from 1 July through 30 June. For example, the verification statistics for the SH during the 2022 season include model forecasts initialized from 1 July 2021 through 30 June 2022. Storms in the best tracks that never attained TC status were excluded from the study. Information regarding model upgrades during the study period are available in the literature (Lin 2004; Harris and Lin 2013; Chen et al. 2019b; McTaggart-Cowan et al. 2019; Heming and Titley 2023) and/or from modeling centers (NOAA 2024; ECMWF 2024; Environment and Climate Change Canada 2024; HYCOM 2024).

TABLE 2. The threshold values for each component of the TC tracker for each model and basin.

Basin	Model	850-hPa ζ maximum ($\times 10^{-5}$ s^{-1})	250–850-hPa ΔZ maximum (m)	925-hPa wind speed maximum (m s^{-1})
AL	CMC	16.8	9446.4	18.5
AL	ECM	27.9	9470.5	17.1
AL	GFS	21.2	9462.6	17.7
AL	NAV	14.3	9487.4	17.3
AL	UKM	31.2	9458.1	18.7
EP	CMC	16.4	9506.4	17.3
EP	ECM	27.2	9498.0	16.8
EP	GFS	23.1	9499.6	16.5
EP	NAV	15.2	9501.8	16.2
EP	UKM	27.4	9501.6	16.5
IO	CMC	19.4	9527.1	18.3
IO	ECM	31.4	9518.6	18.0
IO	GFS	23.2	9522.3	16.6
IO	NAV	13.3	9525.8	16.9
IO	UKM	32.1	9540.0	18.9
SH	CMC	20.2	9538.4	19.4
SH	ECM	30.9	9531.9	19.4
SH	GFS	26.8	9532.6	19.0
SH	NAV	14.1	9538.3	17.8
SH	UKM	33.3	9540.0	20.1
WP	CMC	20.1	9551.0	19.6
WP	ECM	30.0	9541.2	19.6
WP	GFS	24.8	9540.0	18.6
WP	NAV	14.0	9545.5	17.8
WP	UKM	28.8	9542.0	18.6

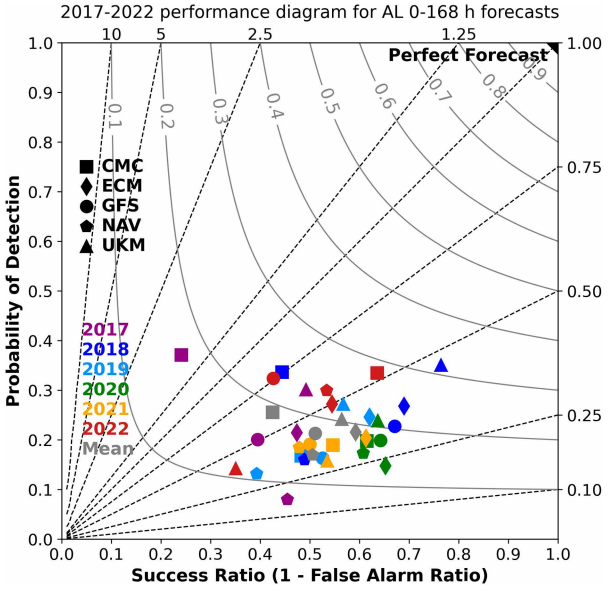


FIG. 1. Performance diagram showing annual average values of the following verification metrics for TC genesis forecasts over the AL basin for the 0–168-h period: SR (x axis), POD (y axis), BIAS (dashed lines), and CSI (gray solid lines). For a perfect forecast, all metrics would equal unity. Each model is represented by a different shape. Each year is color coded with cool (warm) colors representing early (late) years in the dataset. The gray markers indicate 2017–22 mean values.

b. *Methods*

The definition of TC genesis in the model forecast output is the same as in H13 and H16:

1) a relative minimum in mean sea level pressure (MSLP),
2) a relative maximum in 850-hPa relative vorticity that exceeds the model-specific threshold value (explained below) within 2° latitude and longitude of the MSLP minimum (hereafter, "latitude and longitude" will be omitted when referring to distance in degrees),
3) a relative maximum in 250–850-hPa thickness that exceeds the model-specific threshold value within 2° of the MSLP minimum,
4) the 925-hPa wind speed must exceed the model-specific wind threshold at any point within 5° of the MSLP minimum, and
5) criteria 1–4 must be met for at least 24 consecutive forecast hours.

The genesis time occurs at the beginning of the period where criterion 5 is analyzed. The threshold values for maximum 850-hPa relative vorticity, 250–850-hPa thickness, and 925-hPa wind speed used in H13 and H16 were initially calculated from model analyses during the period 2008–10. These threshold values have been updated/calculated for all basins using model analyses during the period 2015–20. Table 2 shows the threshold values for each model and basin, which are held constant for the entire period of this study. Intuitively, the higher (native) resolution models typically exhibit larger threshold values for 850-hPa relative vorticity.

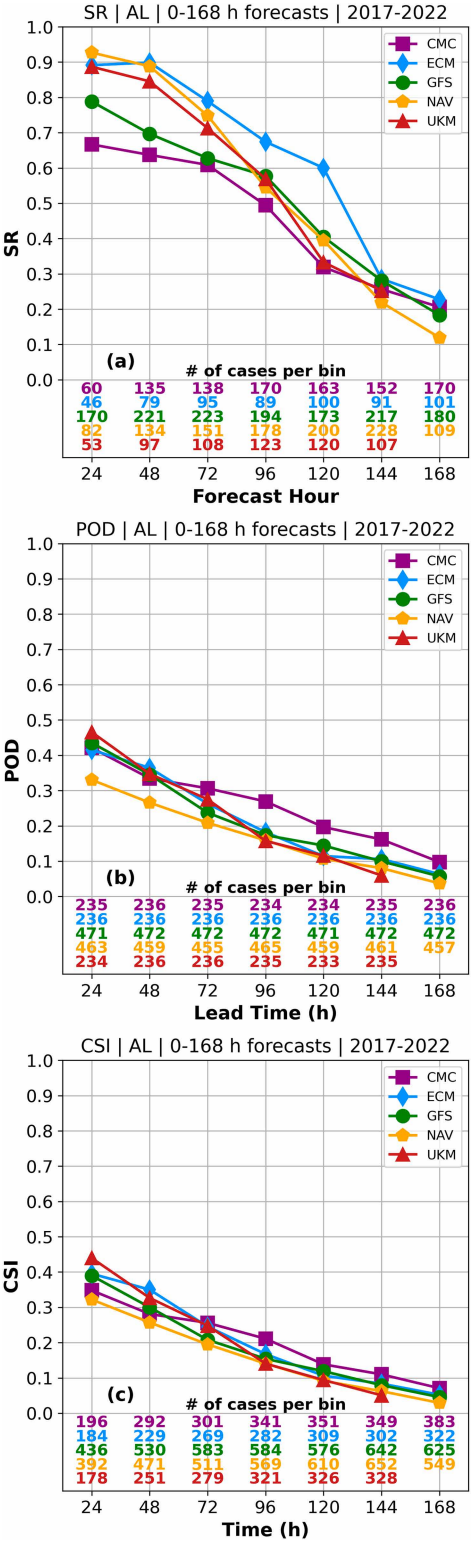
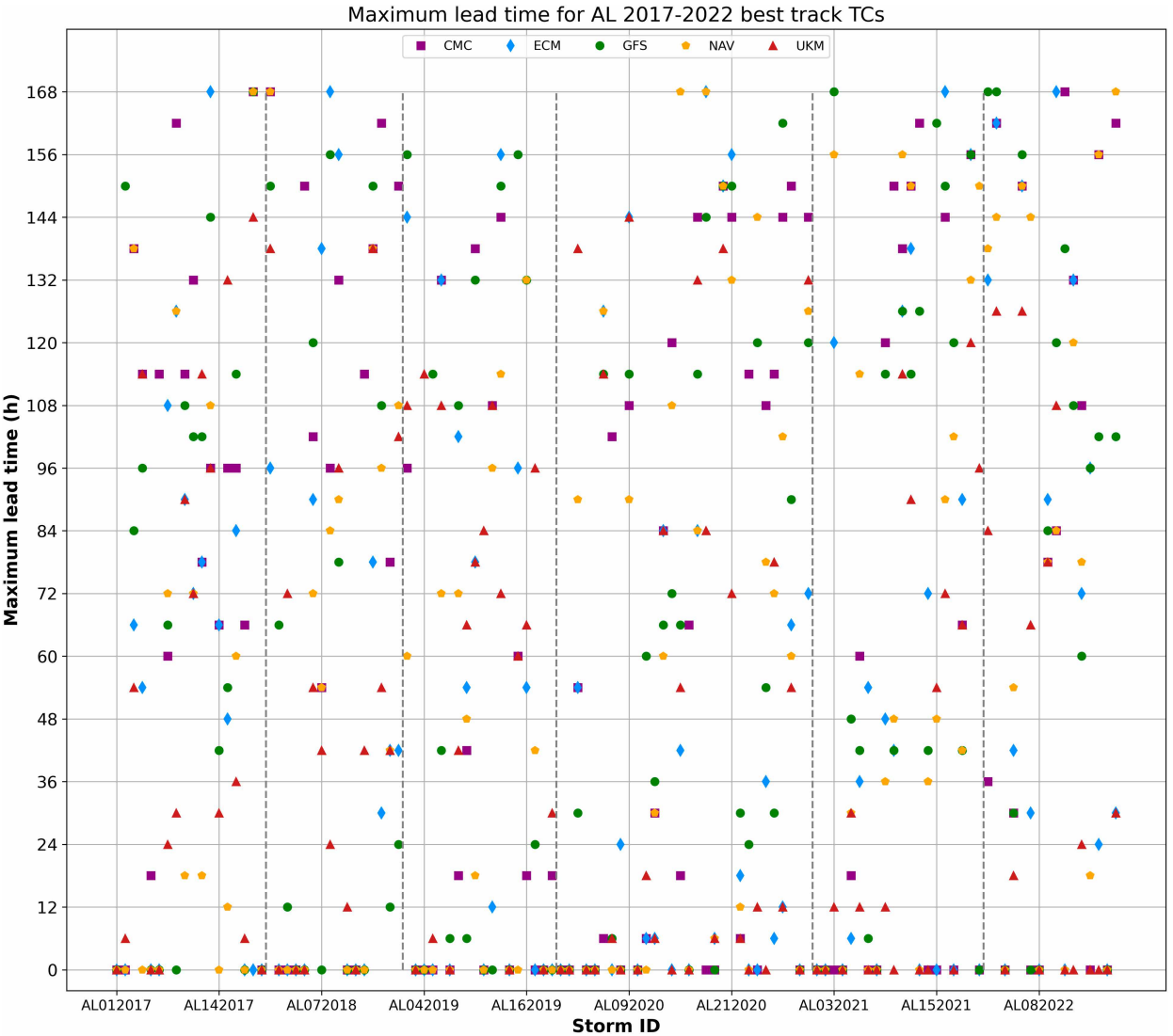


FIG. 2. Mean values of (a) SR, (b) POD, and (c) CSI as a function of hour for the AL basin during the period 2017–22 for forecasts verified during the 0–168-h period.



	CMC (63 h)	ECM (33 h)	GFS (54 h)	NAV (45 h)	UKM (30 h)
CMC (63 h)		0.1	0.989	0.247	0.032
ECM (33 h)			0.1	0.557	0.893
GFS (54 h)				0.271	0.034
NAV (45 h)					0.38
UKM (30 h)					

FIG. 3. The maximum lead time for detection of each best track TC in the AL basin. Best track TCs are ordered on the *x* axis chronologically by their ATCF ID number. Vertical dashed gray lines separate ATCF years. The maximum possible lead time is 168 h for all models, except 144 h for the UKM. A maximum lead time of 0 h indicates that the model entirely failed to predict genesis for a given best track TC. The corresponding table provides the median of the maximum lead time values for each model (rows and columns), and the values in each cell represent the *p* value of a Mann–Whitney *U* test/Wilcoxon rank-sum test comparing the median values between models. Cells shaded in cyan represent statistically significant differences between models with 95% confidence.

The TC tracking algorithm was run on every available model cycle from the CMC, ECM, GFS, and NAV (UKM) during the study period out to 192 h (168 h). Since the TC tracking algorithm's genesis definition requires that criteria 1–4 above are met for at least 24 consecutive hours in the forecast cycle, it can identify TC genesis events out to 168 h (except 144 h for the UKM) in the forecast cycle. Each model forecast TC genesis event is classified as a hit or false alarm using the "OPS"[1] scoring method

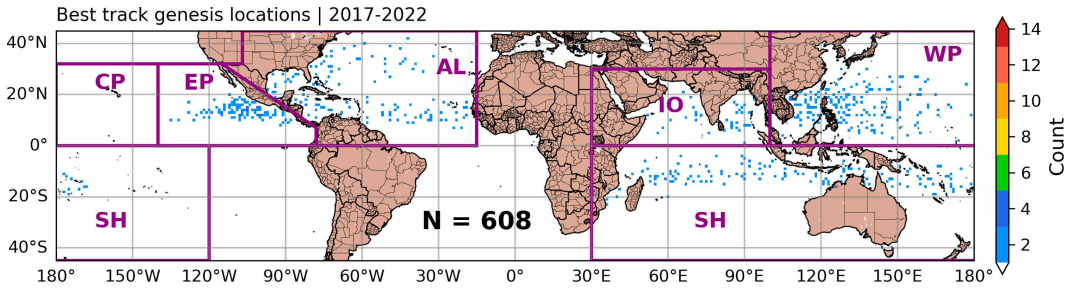


FIG. 4. Counts of TC genesis events in each 1° latitude/longitude grid box from the ATCF b decks during the 2017–22 study period. Purple polygons represent basin boundaries used in this study. CP basin TC genesis events are excluded.

in the Model Evaluation Tools (MET) TC-Gen tool (Jensen et al. 2023). The definitions for hit and false alarm events are identical to those in H16. However, the spatial threshold to match a model genesis forecast to an ATCF system is 500 km in this study instead of 5° in H16. Contingency table counts are calculated for each model and for each basin, with a focus on hit, false alarm, and miss events. A miss occurs when genesis is not forecast in a model for a best track TC when the model initialization time is within 168 h (except 144 h for the UKM) of best track TC genesis. Correct null events are not considered given the rare occurrence of TC genesis and to avoid false inflation of aggregate metrics that use it.

Bulk verification statistics, including probability of detection (POD), success ratio (SR), frequency bias (BIAS), and CSI, were calculated and displayed using performance diagrams (Roebber 2009) to visually compare model performance. These statistics are analyzed for TC genesis events occurring within 7 days of the model initialization time to match the current operational TWO maximum forecast period at NHC and CPHC. From Roebber (2009),

$$\mathrm{POD} = \frac{\mathrm{hits}}{\mathrm{hits} + \mathrm{misses}}, \tag{1}$$

$$\mathrm{SR} = \frac{\mathrm{hits}}{\mathrm{hits} + \mathrm{false\ alarms}}, \tag{2}$$

$$\mathrm{Bias} = \frac{\mathrm{POD}}{\mathrm{SR}}, \tag{3}$$

$$\mathrm{CSI} = \frac{1}{\frac{1}{\mathrm{SR}} + \frac{1}{\mathrm{POD}} - 1}. \tag{4}$$

The number of hits and misses is counted with respect to the forecast lead time (i.e., the difference between the model initialization time and the best track genesis time) for the calculation of POD. In contrast, the number of hits and false alarms is counted with respect to the forecast hour within the model cycle (i.e., the difference between the model initialization time and the model forecast genesis valid time) for the calculation of SR. For all calculations, the tracker runs out to 168 h (except 144 h for the UKM), but events are subset by different temporal ranges, such as when analyzing the SR, POD, and CSI as a function of hour in 24-h periods. This is in contrast to H13 and H16 where the number of hits was counted only with respect to the forecast hour because only SR was analyzed as a function of time in the model cycle in those studies. The maximum lead time for each best track TC (e.g., Chen et al. 2019a; Halperin et al. 2020) was compared, and geographic plots of hit, false alarm, and miss events were constructed in an attempt to subjectively identify any spatial biases within the models.

The results of this study are sensitive to the TC tracker used. The H13 tracker was developed based on tracker attributes from Cheung and Elsberry (2002) and recommendations from Walsh et al. (2007). It shares some similarities with operational and other heavily used cyclone trackers (e.g., Marchok 2002; Hart 2003; Biswas et al. 2018; Magnusson et al. 2021). While there are varying degrees of complexity, all trackers use some combination of MSLP and/or relative vorticity to identify the location of the cyclone. Additional criteria are included to identify warm-core systems (based on temperature and/or thickness or vortex alignment) and confirm that the cyclone has obtained some minimum level of intensity or organization (e.g., MSLP gradients and/or wind speed thresholds).

There is a known limitation with the H13 tracker's ability to detect TC genesis occurring poleward of 25°, often due to the thickness threshold not being met perhaps due to the background environment being less barotropic or the forming cyclone being closer to subtropical or the fact that TCs forming poleward of 25° simply do not have much time to spin up in the model due to a documented life cycle lag (Schenkel and Hart 2012). A tracker with a latitude-dependent thickness threshold was tested (not shown) but that led to a large false alarm ratio where too many nontropical cyclones were being detected at high latitudes. Even after considering this limitation, some of the best track TCs that were not detected at all occurred at latitudes equatorward of 25°, suggesting that other factors are influencing a model's inability to detect a given best track TC. Given the H13 tracker's use in previous studies and for generating the real-time TCLOGG guidance

[1] MET contains two scoring methods. The OPS method best aligns with the logic used to verify TWO forecasts from the NHC.

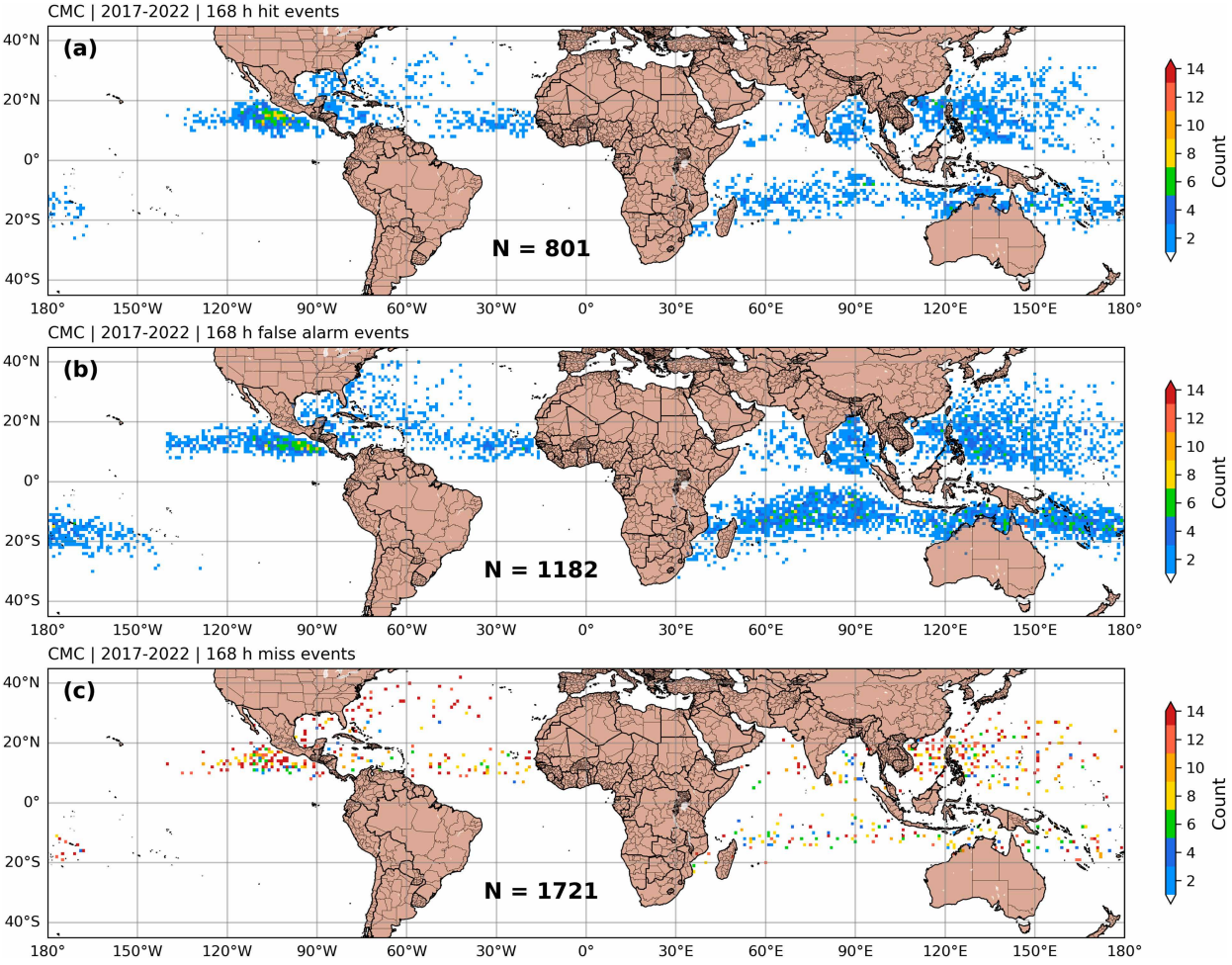


FIG. 5. Counts of TC genesis forecasts in each 1° latitude/longitude grid box from the CMC that were classified as (a) hits, (b) false alarms, and (c) misses during the 0–168-h verification period.

for over a decade, we opt to continue using it for the current study.

3. Results

a. *AL basin*

Performance diagrams for TC genesis forecasts over the AL basin reveal substantial interannual variability among a given model during the study period (Fig. 1). The relative performance among all five models was similar during 2020 and 2021, unlike in other years. The impact of model configuration changes is apparent in some cases. For example, a major update to the atmospheric physics parameterizations in the CMC in 2019 (McTaggart-Cowan et al. 2019) is evident when comparing the annual verification scores between the periods 2017–18 and 2019–22. The POD decreased and SR increased markedly in the latter period. Heming and Titley (2023) found that an update to the UKM in May 2022 that included ocean coupling and physics changes resulted in weaker TCs overall, which is consistent with the notably smaller POD from the UKM compared to the other four models during that year. The ECM exhibits the greatest mean SR value (0.59) but a midrange mean POD value (0.22). In contrast, the CMC exhibits the smallest mean SR value (0.43) but the largest POD value (0.26). These characteristics yield identical mean CSI values (0.19) among those two models, with all models between 0.15 (NAV) and 0.2 (UKM).

SR, POD, and CSI values generally decrease as a function of hour for all models (Fig. 2). For most forecast hour bins, the ECM (CMC) exhibits the greatest (smallest) SR values. The GFS, NAV, and UKM (ECM) exhibit mean SR values > 0.5 through forecast hour 96 (120). There is a fair amount of spread in SR values by model in the beginning of the forecast cycle ($\sim$0.25 for the first 2 days of the forecast cycle), with a convergence to SR values in the 0.22–0.29 range for forecast hours 126–144 and 0.12–0.23 range for forecast hours 150–168. Some of this convergence to small SR values after forecast hour 120 can be attributed to the definition of a hit event in this study: TC genesis has to occur in the ATCF b decks within

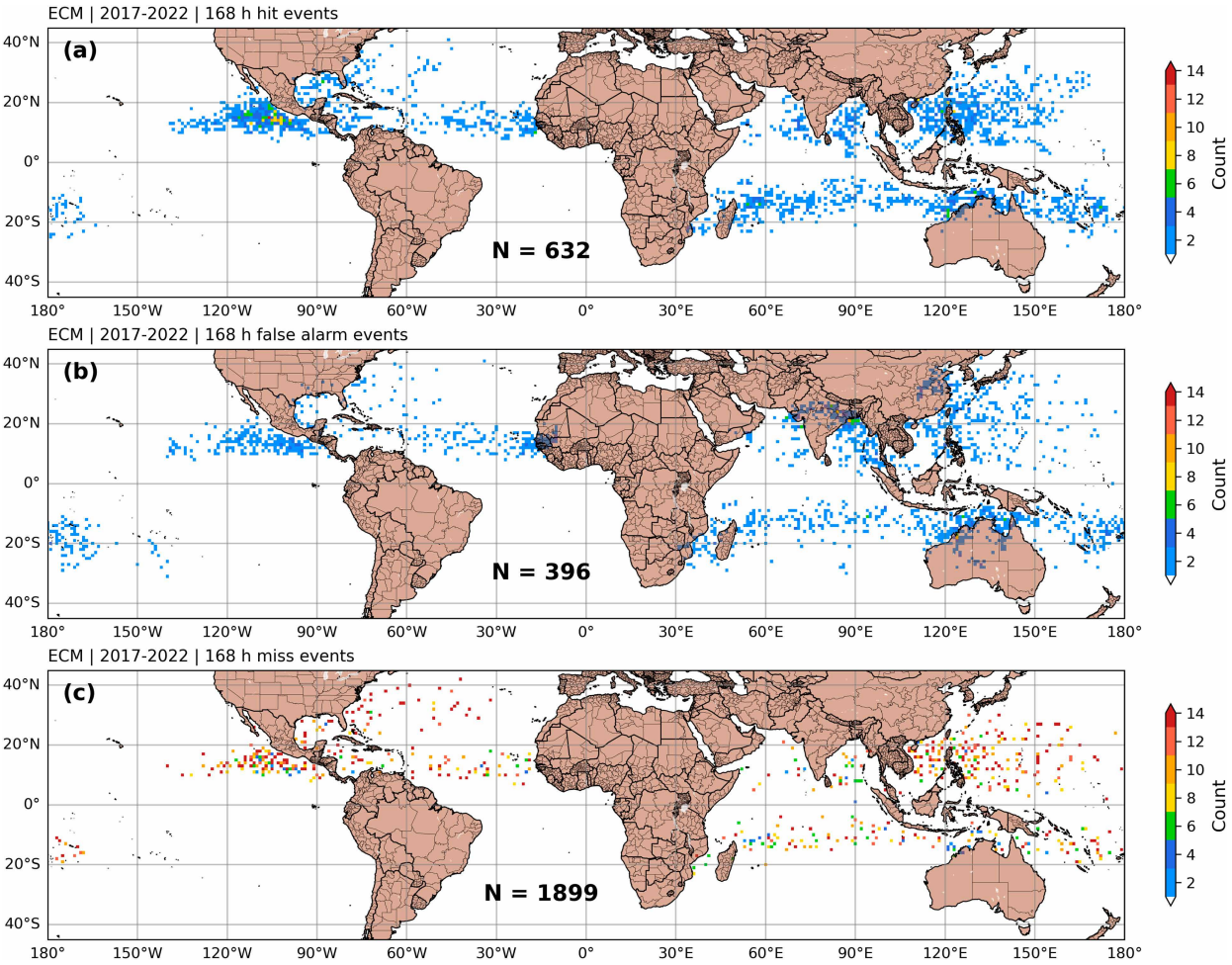


FIG. 6. As in Fig. 5, but for the ECM.

168 h of the model initialization time. Therefore, if a TC genesis event occurs in the model fields at forecast hour 168, but best track genesis occurs at forecast hour 174 or later, then the forecast is necessarily classified as a false alarm. Nevertheless, the decrease in SR values as a function of forecast hour is expected as model error grows over time (e.g., Lorenz 1982; Dalcher and Kalnay 1987; Royer et al. 1994).

Even at lead times within 24 h, all models exhibit POD values < 0.5. The NAV exhibits the smallest POD at most lead time bins. The other four models are tightly clustered for lead times ≤ 48 h, but the CMC has the largest POD at lead times ≥ 72 h. POD values for all models are ≤ 0.1 for lead times of 150–168 h. Note that the sample size for each bin is the number of best track TCs in the study period (118 for the AL) multiplied by the number of model cycles per day less the number of model cycles where data were missing or incomplete since a hit or miss event will be counted for each lead time where model data are available.

The CSI values exhibit the least amount of spread among models for a given time bin. The CMC, ECM, and UKM all exhibit the greatest CSI value for at least one time bin. CSI values decrease from the 0.21–0.4 range at 6–24 h to the 0.03–0.07 range at 150–168 h.

Despite the observed small SR, POD, and CSI values after forecast hour 120 for all models, the models can at times successfully predict genesis for a given best track TC at least 1 week in advance (Fig. 3). This result is encouraging, but overall, there is large variability in the maximum lead time for detection among models and best track TCs. In fact, Fig. 3 illustrates that many AL best track TCs are not detected at all (i.e., markers plotted at $y = 0$). This is likely partially attributed to the characteristics of the TC tracker used, which includes a 250–850-hPa thickness threshold value that must be exceeded for the cyclone to be considered tropical. However, this may also be due to the models struggling to predict TC genesis from tropical transition pathways (e.g., McTaggart-Cowan et al. 2013). Indeed, this result is consistent with Wang et al. (2018), as they found that the POD of TC genesis from tropical transition pathways was smaller than nonbaroclinic and low-level baroclinic pathways over the AL basin using GEFS forecasts. Other

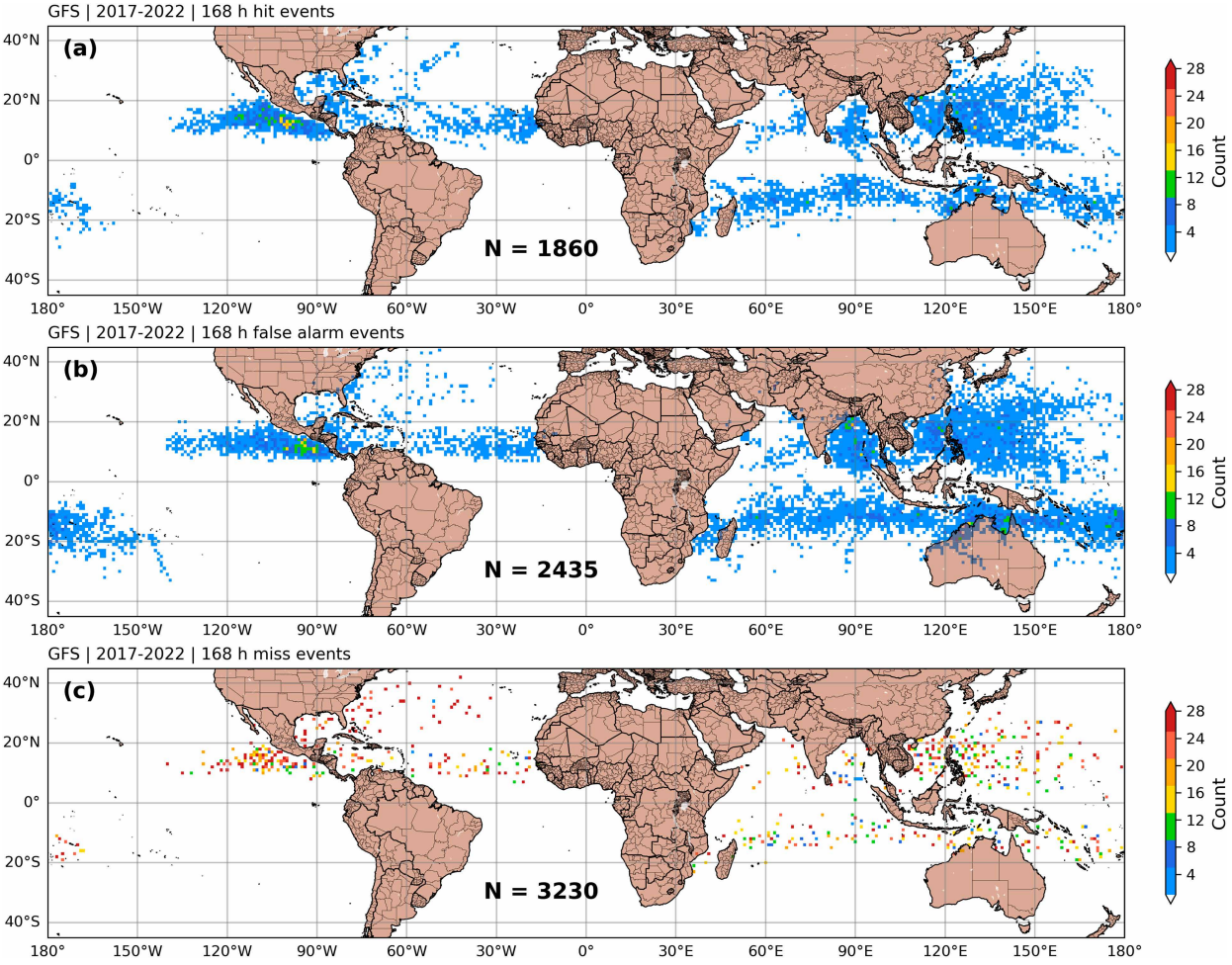


FIG. 7. As in Fig. 5, but for the GFS.

possible explanations include the models not correctly simulating the phasing of African easterly waves (AEWs) and mesoscale convective systems or the strength of the West African monsoon, which have been shown to be important predictors of TC genesis from AEWs in the AL basin (Núñez Ocasio et al. 2020, 2021, 2024; Núñez Ocasio and Rios-Berrios 2023).

Overall, the CMC (UKM) exhibited the greatest (smallest) median of the maximum lead time at 63 h (30 h). When intercomparing all five models, the CMC and GFS exhibited significantly longer medians of maximum lead time than the UKM with 95% confidence, according to a Mann–Whitney *U* test/Wilcoxon rank-sum test (e.g., Wilks 2011). Results in other basins that refer to the statistical significance of differences in the medians of maximum lead time will be conducted using the same test and confidence level.

The locations of the best track genesis events and the spatial extent of each basin are provided in Fig. 4. While it is interesting to view the forecasts geographically, the relatively small sample size makes it difficult to confidently assess potential spatial biases. For example, the CMC has hit and false alarm events distributed throughout the AL basin, particularly over the MDR and Caribbean Sea (Fig. 5). There is a cluster of false alarm events just offshore of the southeastern United States, but this is a result of multiple false alarm forecasts over time for potential tropical cyclone 10 in 2017 (i.e., a single nondeveloping disturbance) rather than an indication of a model bias. The greatest number of miss events occurs over the Gulf of Mexico and poleward of 30°N, with fewer miss events across the MDR and Caribbean Sea. The ECM, which has a smaller POD overall, tends to successfully predict TC genesis over the MDR, Caribbean Sea, and the Gulf of Mexico (Fig. 6). The model also produced false alarm cases over the MDR, but there is a dearth of false alarm forecasts over the Caribbean Sea and Gulf of Mexico, suggesting that a forecaster may have increased confidence in development predicted in those regions. Relatively few forecasts (hits or false alarms) occurred poleward of 20°N over the Atlantic Ocean, which is where the most miss events are identified, although there are TC genesis events that go undetected throughout the basin. The GFS produces numerous TC genesis forecasts over

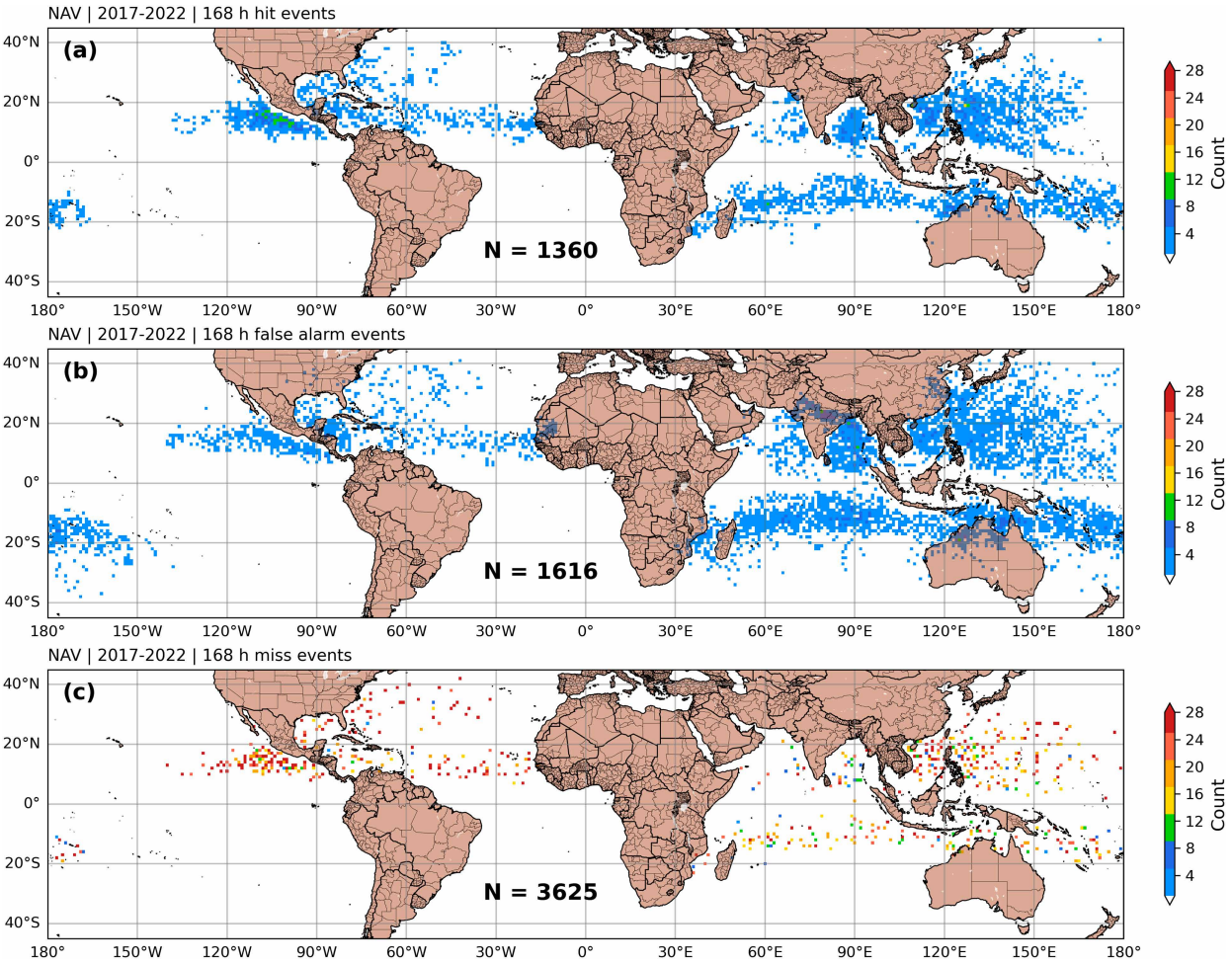


FIG. 8. As in Fig. 5, but for the NAV.

the MDR and western Caribbean Sea (Fig. 7). While most of the regions where genesis is predicted contain a mix of hit and false alarm events, the events in the southwestern Caribbean Sea near the coast of Central America from Nicaragua to Panama are mostly false alarms. These false alarms often occur during the early and late portions of the hurricane season (not shown) and appear to be associated with the Central American gyre (Papin et al. 2017). The NAV exhibits a pattern similar to the GFS but with fewer overall forecast events (Fig. 8). Most forecast events over the Gulf of Mexico are hits. The distribution of UKM forecast genesis events (Fig. 9) is similar to that of the ECM, with false alarm events mostly over the MDR. Forecast genesis events occurring over the western Caribbean Sea and Gulf of Mexico tend to develop into best track TCs. Miss events occur throughout the basin.

b. *EP basin*

The forecast verification statistics are better overall in the EP basin compared to the AL basin (Fig. 10). The mean POD values over the EP basin are greater than those over the AL basin for all models. In addition, the mean SR values over the EP basin are greater (less) than those over the AL basin for the CMC, ECM, and GFS (UKM). The resulting mean CSI values are between 0.2 (NAV) and 0.31 (ECM). All models except the GFS exhibited their highest SR values in 2022. Despite several different GFS configurations during the study period, including a new dynamical core in 2019 (Lin 2004; Harris and Lin 2013; Chen et al. 2019b), there is relatively little interannual variability in GFS SR values in the EP basin. Instead, the GFS interannual variability is largely in terms of POD. The CMC model configuration change in 2019 is also evident in the EP basin verification statistics, with recent years exhibiting larger SR values and smaller POD values. Overall, there is a fair amount of spread in the verification metrics among the five models for any given year.

As in the AL basin, SR, POD, and CSI decrease as a function of hour (Fig. 11), with SR values > 0.5 for the GFS and UKM through forecast hour 96. The ECM and NAV models exhibit SR values > 0.5 for longer in the forecast cycle than over the AL basin, through forecast hour 144 and 120, respectively. The observed convergence to small SR values among all models over the AL basin is not found over the EP basin.

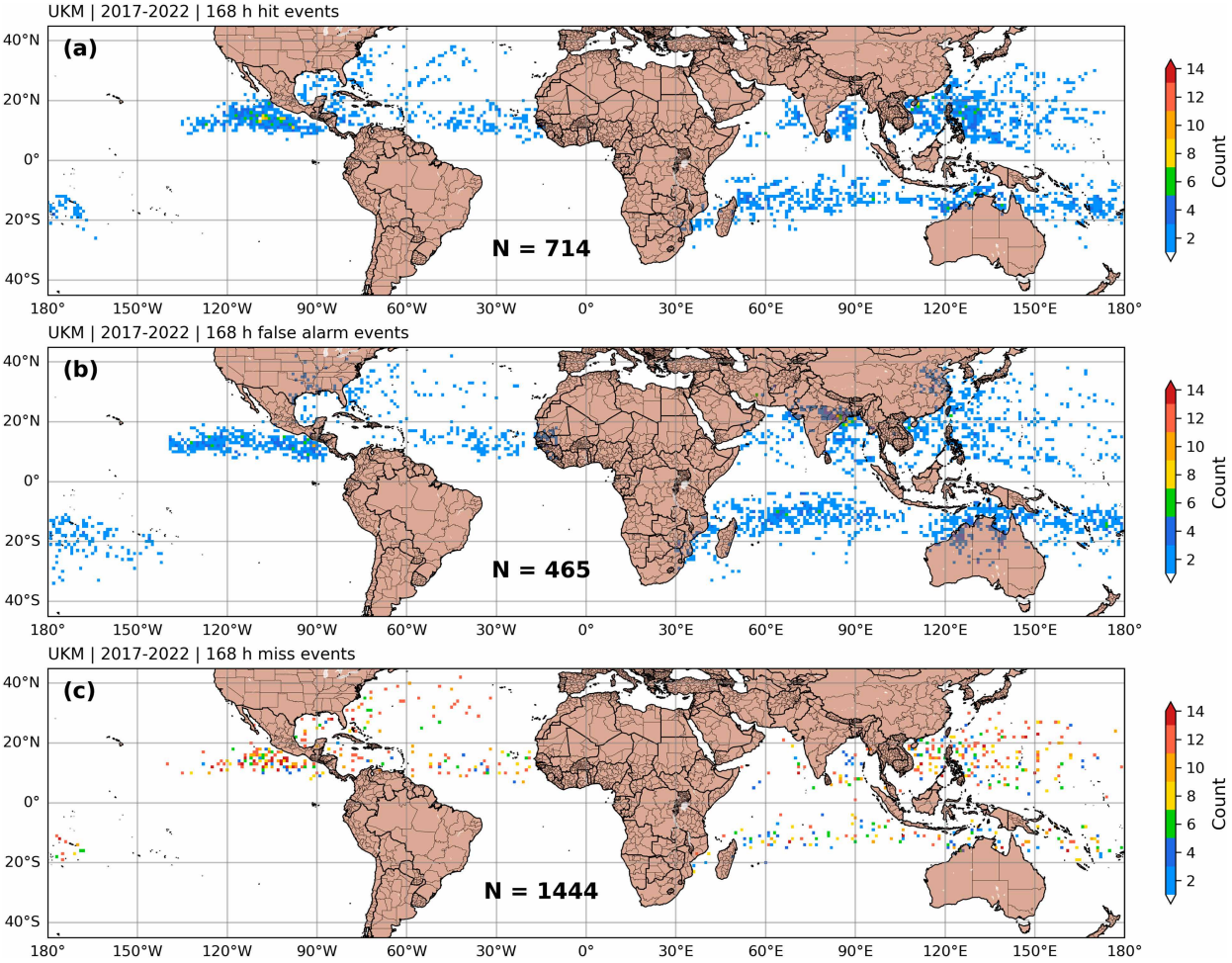


FIG. 9. As in Fig. 5, but for the UKM.

In fact, the ECM and NAV models exhibit SR values 0.18–0.25 larger at days 6 and 7 over the EP compared to the AL basin.

The POD is larger over the EP than the AL. Even so, POD values > 0.5 occur only for the CMC, ECM, and GFS for lead times out to 48 h. The GFS (NAV) exhibits the largest (smallest) POD at most lead-time bins. The ECM and GFS CSI values are >0.5 out to 24 h, with the ECM maintaining the greatest value relative to all models for all but the 126–144-h bin. The NAV CSI values are the smallest among the models, largely due to its small POD over the EP basin.

Similar to the results from the AL basin, the models were able to detect the formation of some best track TCs up to 1 week in advance, but there was a large degree of variability among models and TCs (Fig. 12). Consistent with the overall larger POD over the EP compared to the AL basin, all five models exhibited greater medians of maximum lead time. Similar to the AL basin, the CMC and GFS (UKM) exhibited the largest (smallest) median of maximum lead time over the EP basin. The CMC and GFS exhibited significantly longer medians of maximum lead time than the NAV and UKM, while the ECM also exhibited a significantly longer median of maximum lead time than the UKM.

All of the models produce numerous hit events in the area between 10° and 20°N and between 95° and 120°W (Figs. 5–9). However, the CMC, GFS, and UKM exhibit a relative maximum of false alarm events east of ~95°W, just south of the Gulf of Tehuantepec and west of the Gulf of Papagayo. We speculate that these models may overpredict the surface relative vorticity generation associated with gap wind events in this region (Zehnder 1991; Zehnder and Gall 1991; Mozer and Zehnder 1996; Zehnder et al. 1999; Molinari et al. 2000; Holbach and Bourassa 2014), leading to erroneous predictions of TC genesis. It is also possible that the models are not correctly simulating the local generation of easterly waves (e.g., Rydbeck et al. 2017; Whitaker and Maloney 2020) or the relationships between MJO and Caribbean low-level jet phase/index and large-scale conditions (un)favorable for the development and intensification of easterly waves into TCs in this region (e.g., Molinari and Vollaro 2000; Serra et al. 2010; Whitaker and Maloney 2018).

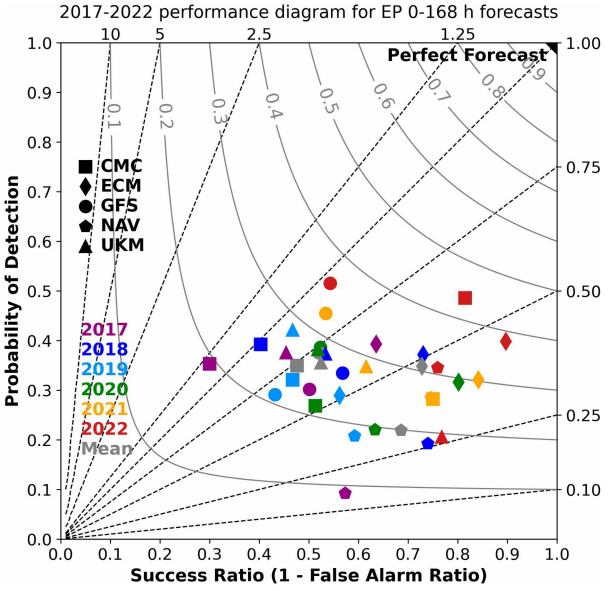


FIG. 10. As in Fig. 1, but for the EP basin.

c. *IO basin*

The forecast verification metrics over the IO basin are mixed, with some of the greatest mean POD values but the smallest mean SR values of any basin (Fig. 13). Mean POD ranges from 0.39 (GFS and UKM) to 0.43 (ECM), while mean SR ranges from 0.28 (NAV) to 0.36 (ECM). With the larger POD, the models exhibit larger CSI values over the IO than over the AL basin. The relative ranking of the CMC (second for all metrics) is better over the IO than over the AL and EP basins. Despite the considerable interannual variation and model-to-model differences, the mean scores are tightly clustered for all models with SR, POD, and CSI values only exhibiting a range of 0.08, 0.04, and 0.04, respectively. Sample sizes across the IO basin are considerably smaller than in the other basins, so results for the IO basin should be interpreted cautiously.

The SR values peak for all models in the forecast hour 30–48 period, perhaps indicating that there are issues with model initialization in this basin (Fig. 14). SR values are <0.5 for all models and time periods except the CMC and GFS at forecast hours 30–48. The SR values decrease to between 0.1 and 0.2 at forecast hours 150–168. The POD evolution over time is quite consistent among the models, with no single model performing best or worst at all lead times. While the POD values start notably higher over the IO basin (>0.6) compared to the AL and EP basins, the POD drops off to <0.2 for the longest lead times, which is consistent with the EP basin. The largest variance in CSI scores among the models occurs in the 6–24-h time bin with CSI peaking in the 30–48-h bin, following the evolution of the SR values.

The relatively large POD values over the IO basin result in relatively high medians of the maximum lead time for each best track TC, with some TCs detected 1 week in advance

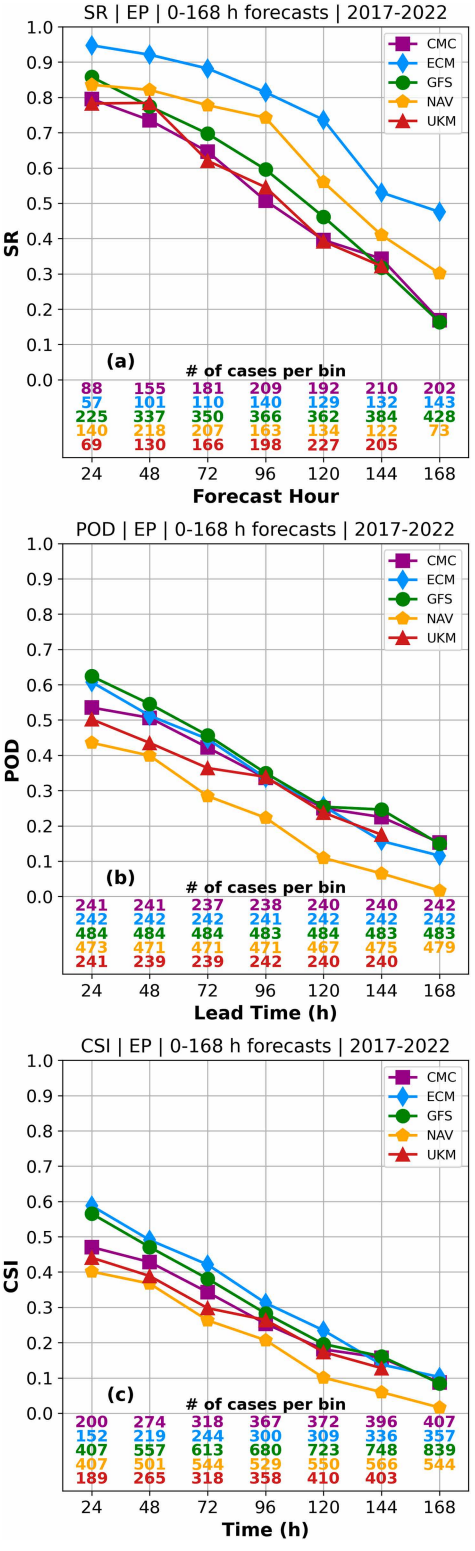
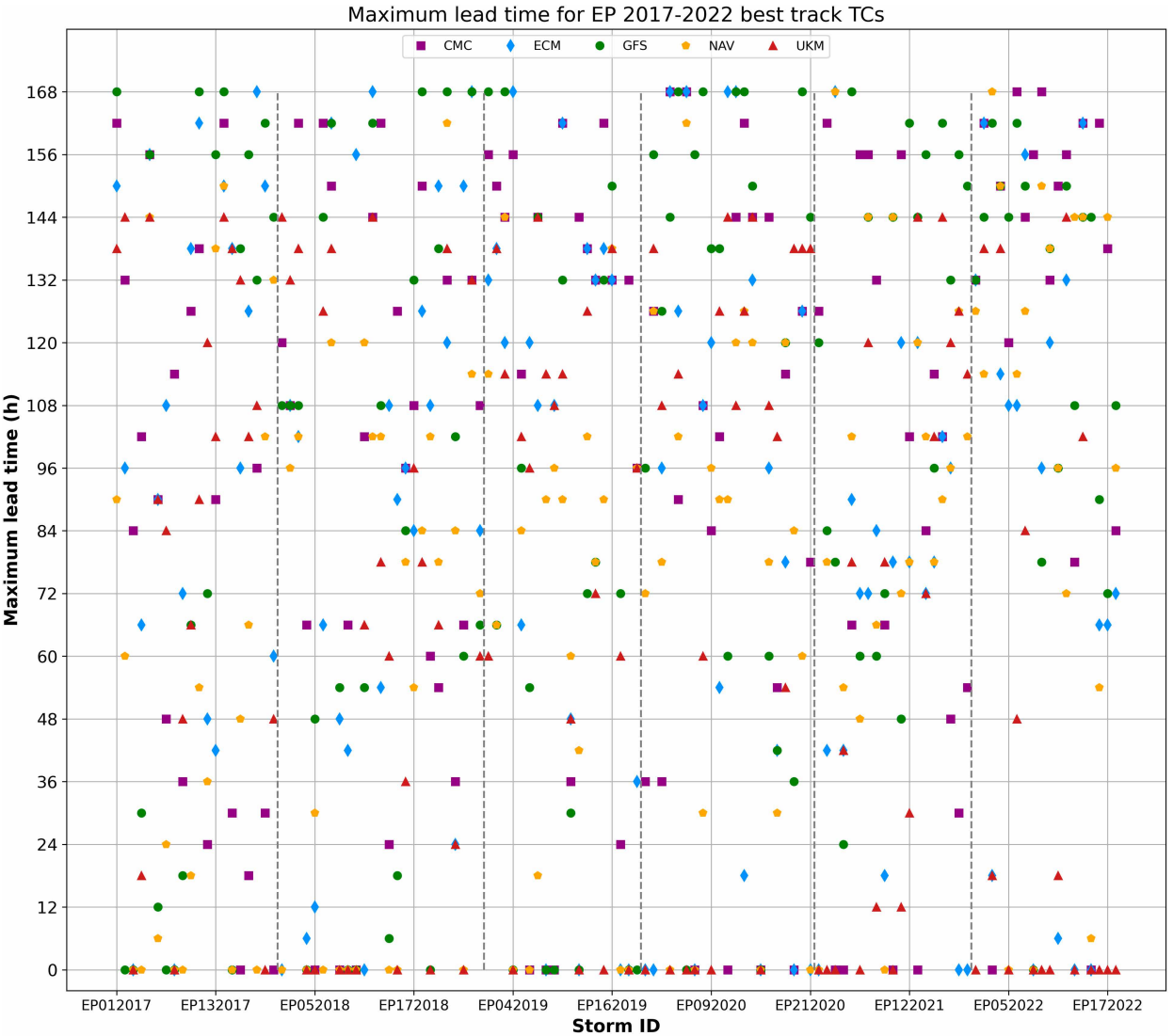


FIG. 11. As in Fig. 2, but for the EP basin.



	CMC (108 h)	ECM (96 h)	GFS (108 h)	NAV (81 h)	UKM (78 h)
CMC (108 h)		0.158	0.503	0.0	0.0
ECM (96 h)			0.06	0.056	0.033
GFS (108 h)				0.0	0.0
NAV (81 h)					0.76
UKM (78 h)					

FIG. 12. As in Fig. 3, but for the EP basin.

(Fig. 15). The UKM stands out with the relatively smallest median of maximum lead time (63 h), which is significantly different than the medians of the maximum lead time for the ECM, GFS, and NAV.

All of the models exhibit a relative maximum in false alarm forecasts just offshore of Bangladesh, centered near 20°N, 90°E (Figs. 5–9). The spatial distribution of these false alarm events is in good agreement with the climatology of monsoon low pressure and monsoon depression genesis from Hurley and Boos (2015, their Fig. 2), suggesting that some of these events may have verified as monsoon lows that did not intensify into TCs. The ECM, NAV, and UKM even produce TC genesis forecasts over northern India. Given the potential for the TC tracker criteria used here to be met over small landmasses (e.g., islands), we chose not to include a land/ocean mask in this study. The models generally exhibit fewer miss events over the Bay of Bengal compared to the Arabian Sea.

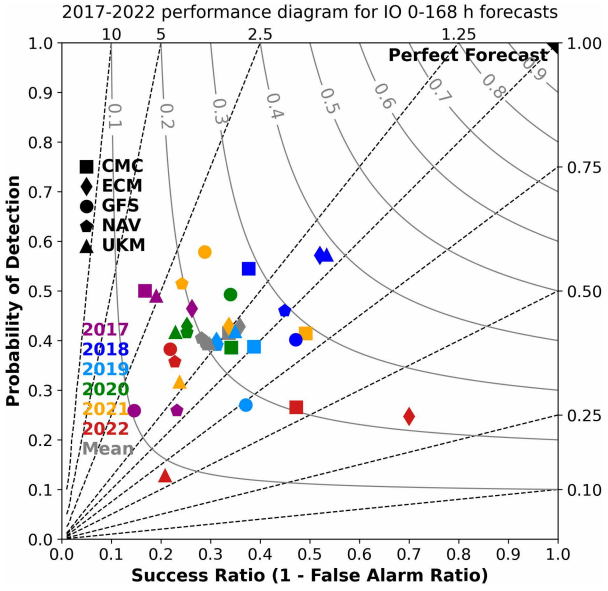


FIG. 13. As in Fig. 1, but for the IO basin.

d. *SH basin*

Probability of detection values over the SH basin were on average comparable to the IO basin, with values ≥ 0.4 for most models (Fig. 16). With some exceptions, POD values peaked during 2018–20 for all models. SR values were larger over the SH than the IO and in the 0.32–0.52 range (except for the CMC), which yields CSI values in the 0.19–0.29 range. The ECM (CMC) exhibits the largest (smallest) SR and CSI values. The relative ranking of the UKM is higher over the SH compared to the AL, EP, and IO basins.

Similar to the IO basin, the SR values peak at forecast hours 30–48 over the SH basin, except for the CMC which is nearly steady between forecast hours 6 and 48 (Fig. 17). The GFS and NAV SR values are quite similar throughout the period. The ECM exhibits SR values > 0.5 during forecast hours 30–120. Its SR value of 0.55 at forecast hours 102–120 is 0.22 larger than the next best model at that time (UKM). The POD values show a nearly constant rate of decline from values > 0.7 (except the NAV) at the 6–24-h lead times to the 0.08–0.15 range at the 150–168-h lead times. Similar to the IO basin, POD values remained ≥0.45 for all models at lead times out to 72 h. The CSI values peak at the 30–48-h time bin, following the SR evolution. The ECM exhibits the greatest CSI values for most time bins, mainly a result of its large SR scores compared to the other models.

Consistent with large POD values, all models exhibit relatively long medians of maximum lead time over the SH basin, with the CMC, GFS, and NAV all near 5 days and the ECM and UKM near 4 days (Fig. 18). More best track TCs went completely undetected during 2021–22 compared to prior years. In particular, the UKM missed several best track TCs. The CMC, GFS, and NAV exhibit significantly longer medians of maximum lead time than the ECM and UKM.

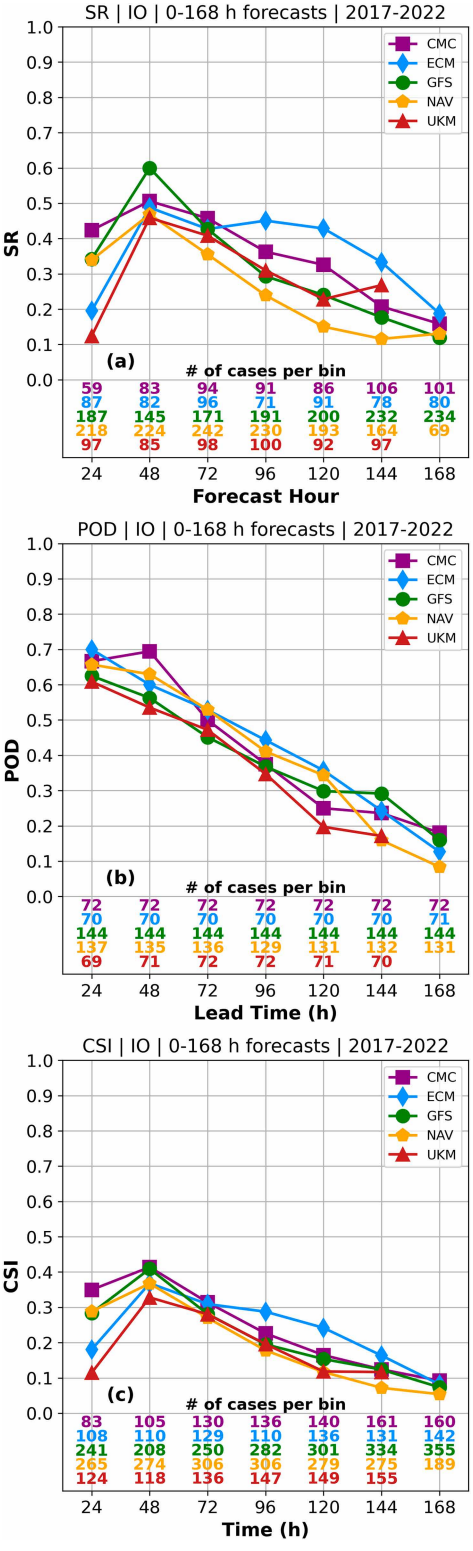
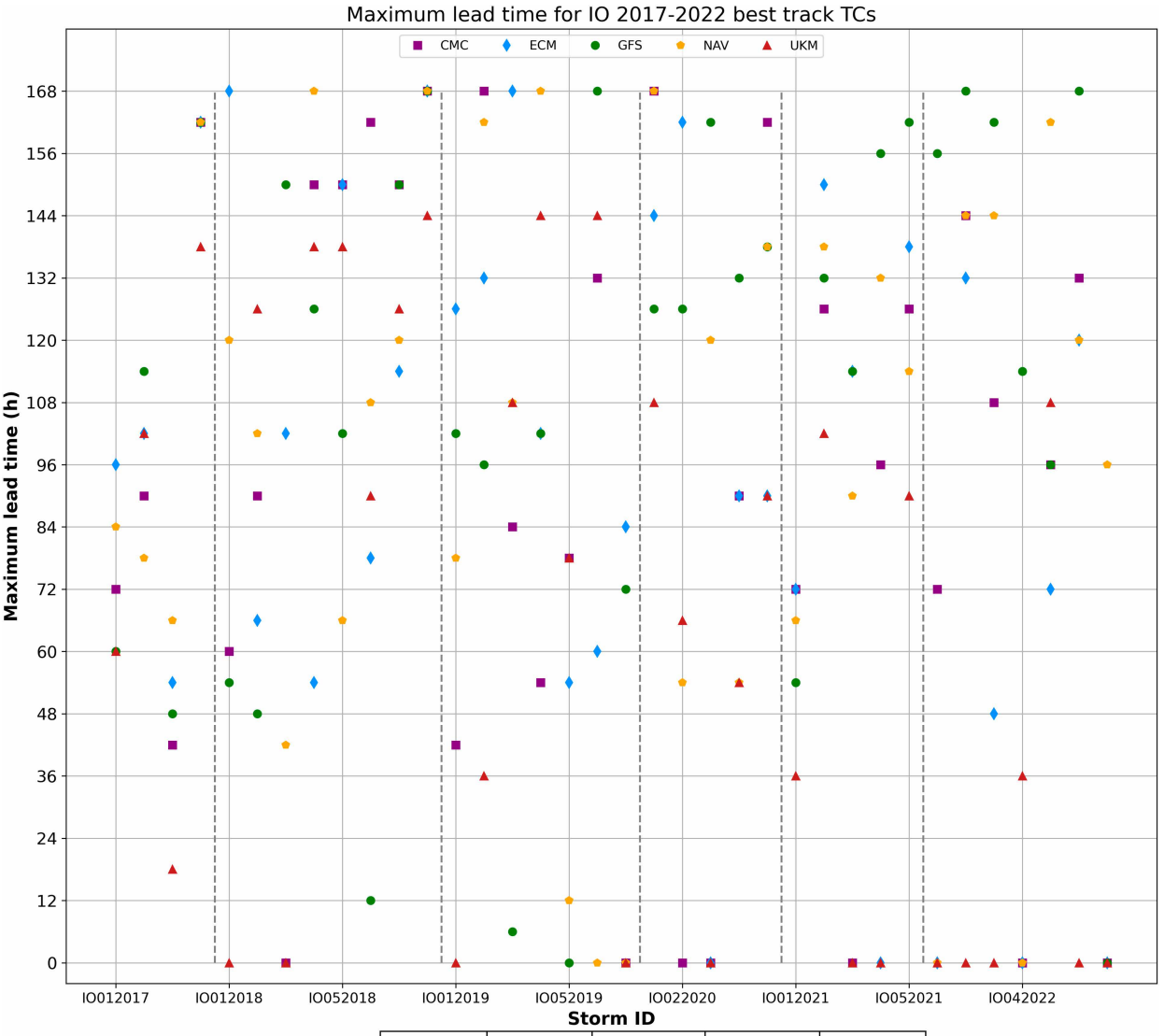


FIG. 14. As in Fig. 2, but for the IO basin.



	CMC (90 h)	ECM (99 h)	GFS (120 h)	NAV (108 h)	UKM (63 h)
CMC (90 h)		0.862	0.152	0.572	0.056
ECM (99 h)			0.198	0.634	0.032
GFS (120 h)				0.443	0.001
NAV (108 h)					0.01
UKM (63 h)					

FIG. 15. As in Fig. 3, but for the IO basin.

The CMC, GFS, and NAV produce a relative maximum of false alarms over the Gulf of Carpentaria, with the GFS also producing numerous false alarms over the Joseph Bonaparte Gulf (Figs. 5, 7, and 8). Another comparison with Hurley and Boos (2015, their Fig. 5) suggests that some of these false alarm events may be monsoon lows or depressions that do not intensify into TCs. Elsewhere, the CMC (GFS) overpredicts genesis over the southern Indian Ocean (South Pacific Ocean). The ECM and UKM, which have the largest mean SR values over the SH basin, do not have any discernible areas of persistent erroneous TC genesis, except for a few events that occurred over Australia (Figs. 6 and 9).

e. WP basin

TC genesis forecasts for the WP basin exhibit mean CSI values in the 0.21–0.27 range (Fig. 19). The same general comparison observed in other basins remains true for the SR, with

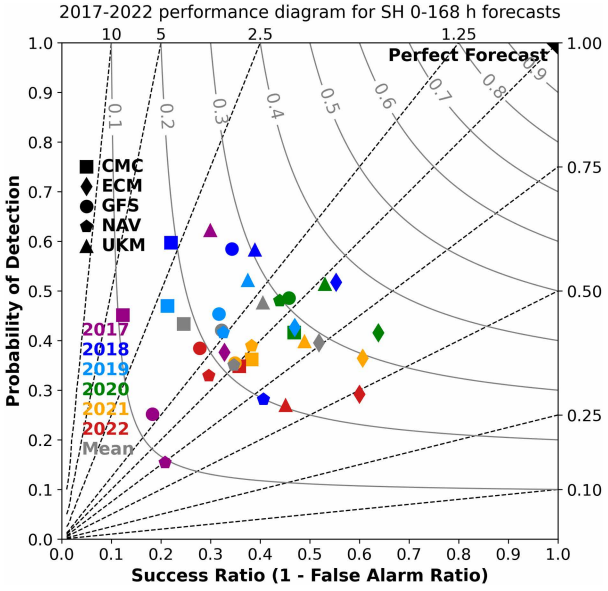


FIG. 16. As in Fig. 1, but for the SH basin.

the ECM (CMC) exhibiting the largest (smallest) mean values. Mean SR values are most comparable to the AL basin. Meanwhile, the GFS (ECM) exhibits the greatest (smallest) POD, with values most similar to the EP basin. Similar to the SH basin, POD appears to have decreased since 2021 but only for the CMC, ECM, and UKM models.

The expected pattern of SR generally decreasing with forecast hour is observed in the WP basin (Fig. 20). The models tend to fall into two groups, with a very similar evolution for each member in a group: the ECM and UKM SR values peak at forecast hour 30–48, with generally larger SR values overall, and the CMC, GFS, and NAV with generally smaller SR values. The CMC, GFS, and NAV (ECM and UKM) models exhibit SR values ≥ 0.5 through forecast hour 72 (120). As in other basins, the ECM generally exhibits the greatest SR values throughout the forecast period. There are no distinct groups in terms of POD and CSI values. The GFS exhibits the greatest POD at all lead times except 150–168 h, while the ECM exhibits the smallest POD at all lead times after 48 h. POD values > 0.45 occur out to lead times of 48 h, which is 24 h less than in the IO and SH basins. The CSI values are tightly clustered and decrease from >0.4 at the 6–24-h time bin to <0.1 at the 150–168-h time bin.

As in the other basins, some best track TCs were predicted up to 1 week in advance (Fig. 21). The medians of maximum lead time are between those of the AL (shortest lead times) and SH (largest lead times) basins. Consistent with its relatively larger POD values over the WP basin, the GFS exhibits the greatest median of maximum lead time, and that value (108 h) is significantly longer than those of the CMC, ECM, and UKM. The CMC and NAV also exhibit significantly longer medians of maximum lead time than the ECM and UKM.

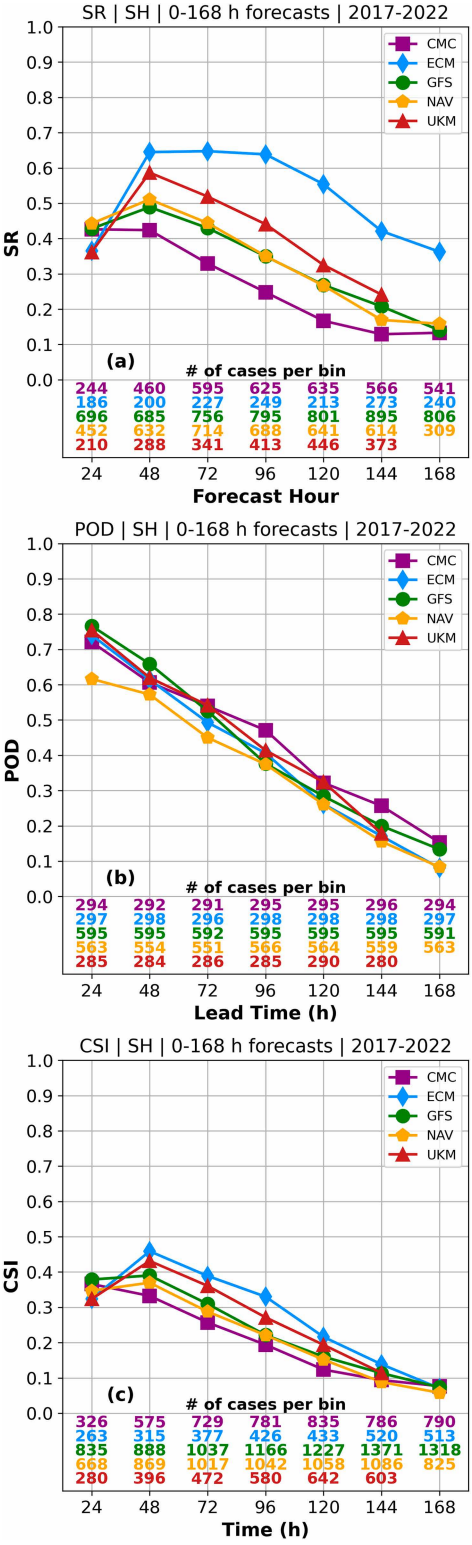
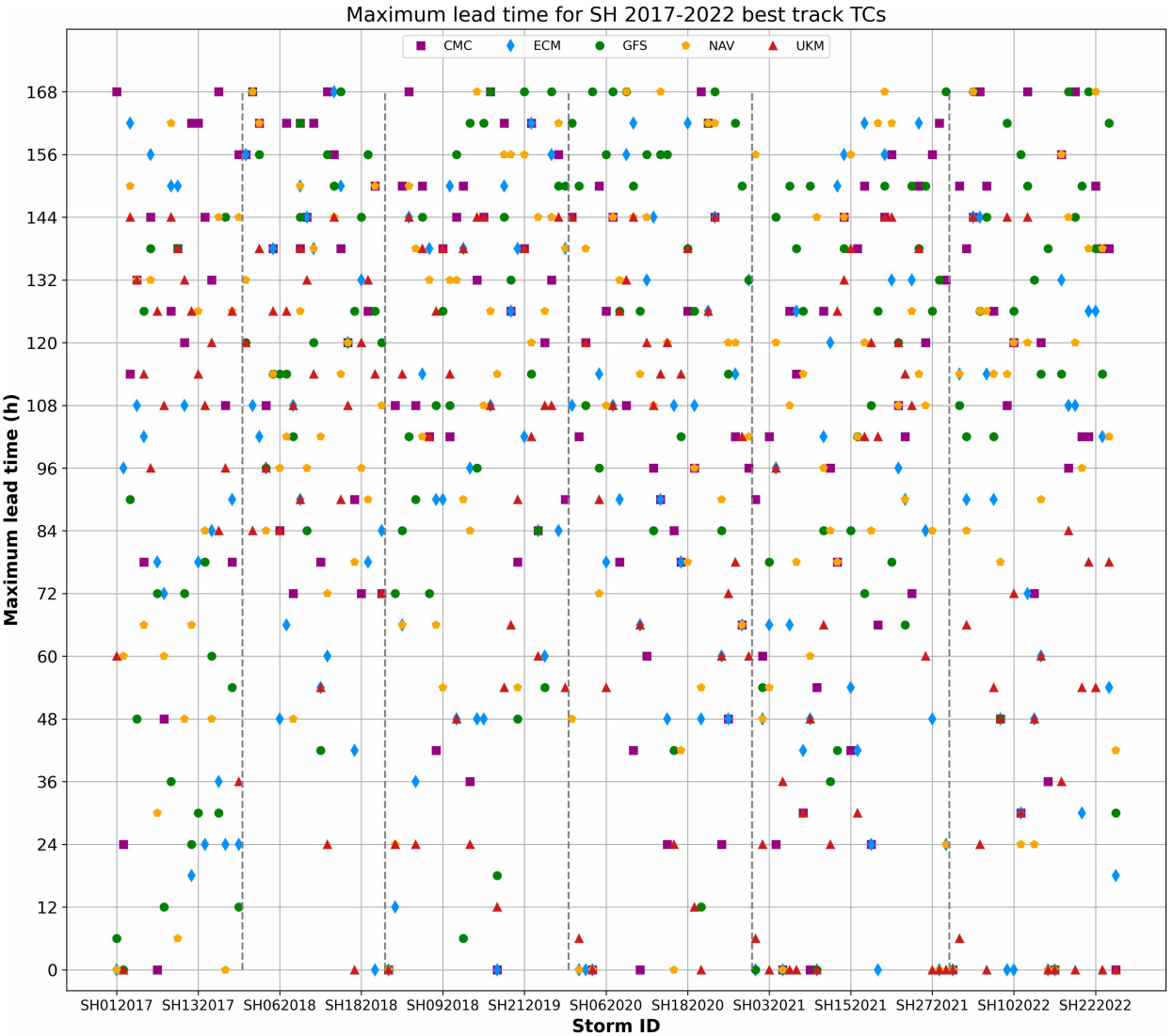


FIG. 17. As in Fig. 2, but for the SH basin.



	CMC (120 h)	ECM (90 h)	GFS (126 h)	NAV (114 h)	UKM (96 h)
CMC (120 h)		0.0	0.642	0.177	0.0
ECM (90 h)			0.0	0.006	0.57
GFS (126 h)				0.071	0.0
NAV (114 h)					0.001
UKM (96 h)					

FIG. 18. As in Fig. 3, but for the SH basin.

All models are generally most cyclogenetic over the South China Sea and Philippine Sea (Figs. 5–9), coincident with the climatological relative maximum in monsoon depression events (Hurley and Boos 2015). The ECM and UKM produce relatively few false alarm forecasts over these regions, which is unsurprising given their relatively larger SR values compared to the other models. In contrast, the GFS produces numerous false alarms over these regions, while the relative maximum of false alarms occurs over the Philippine Sea for the CMC and NAV.

4. Summary and conclusions

This study provided an update and expansion of the training dataset used to produce the Tropical Cyclone Logistic Guidance for Genesis (TCLOGG) disturbance-specific, probabilistic forecasts. Real-time guidance from TCLOGG was initially available in 2014 for the AL and EP basins. The guidance was expanded to the CP basin in 2017 and globally in 2023.

In this study, tropical cyclone (TC) genesis events were detected in model forecast fields from the CMC, ECM, GFS,

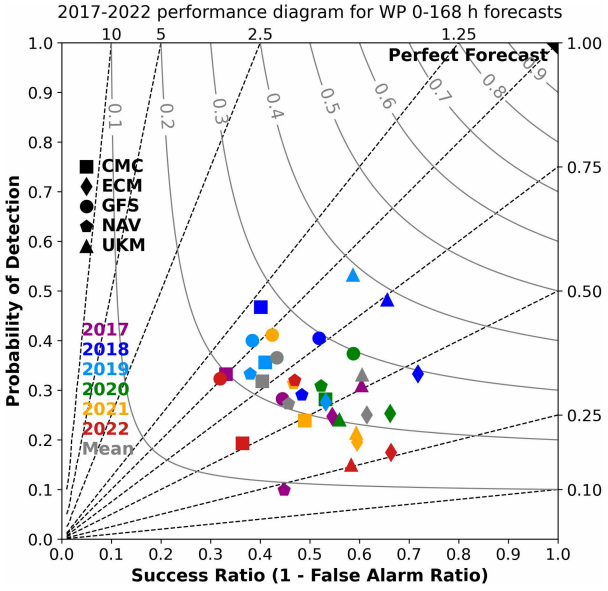


FIG. 19. As in Fig. 1, but for the WP basin.

NAV, and UKM global models and verified against observations from the NHC and JTWC during the period 2017–22. The TC tracker used here is the same as that used in H13 and H16, which has some similarities to operational TC and other heavily used trackers (e.g., Marchok 2002; Hart 2003; Biswas et al. 2018; Magnusson et al. 2021). Therefore, results found here are tracker dependent. The verification period of 7 days in this study was chosen to align with the NHC's and CPHC's maximum forecast period in the operational TWO product. Contingency table statistics were computed for each model, year, and basin.

Results show large interannual variability within a single model and model-to-model variability for a given year and basin. The model rankings are often dependent on year and basin, making a clear "winner" impossible to determine. Nevertheless, a few themes did emerge from this study: 1) the ECM exhibited the largest 2017–22 mean success ratio (SR) value globally and in each individual basin (Fig. 22); 2) a large SR value often comes at the expense of a smaller probability of detection (POD); 3) the SR–POD trade-off yields mean critical success index (CSI) values no higher than 0.35 for any individual model and basin and less than 0.3 when averaged globally for each model during 2017–22; 4) SR, POD, and CSI values unsurprisingly generally decrease as hour increases, but the rate of change as a function of hour can be quite different among the models for a given basin; and 5) while the models can detect some best track TCs 1 week in advance, there are also numerous best track TCs whose genesis goes completely undetected by the models, which may be due to the characteristics of the TC tracker used in this study and/or the (in)ability of the models to predict TC genesis from certain genesis pathways. Overall, while many of the verification statistics indicated that global models continue to be a useful source of TC genesis

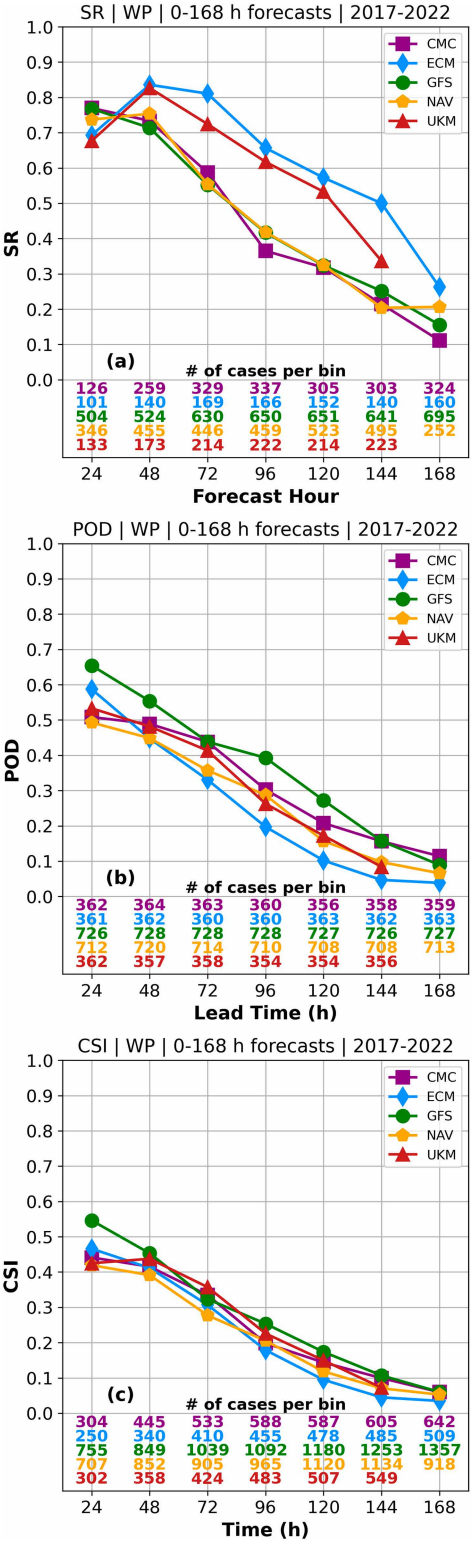
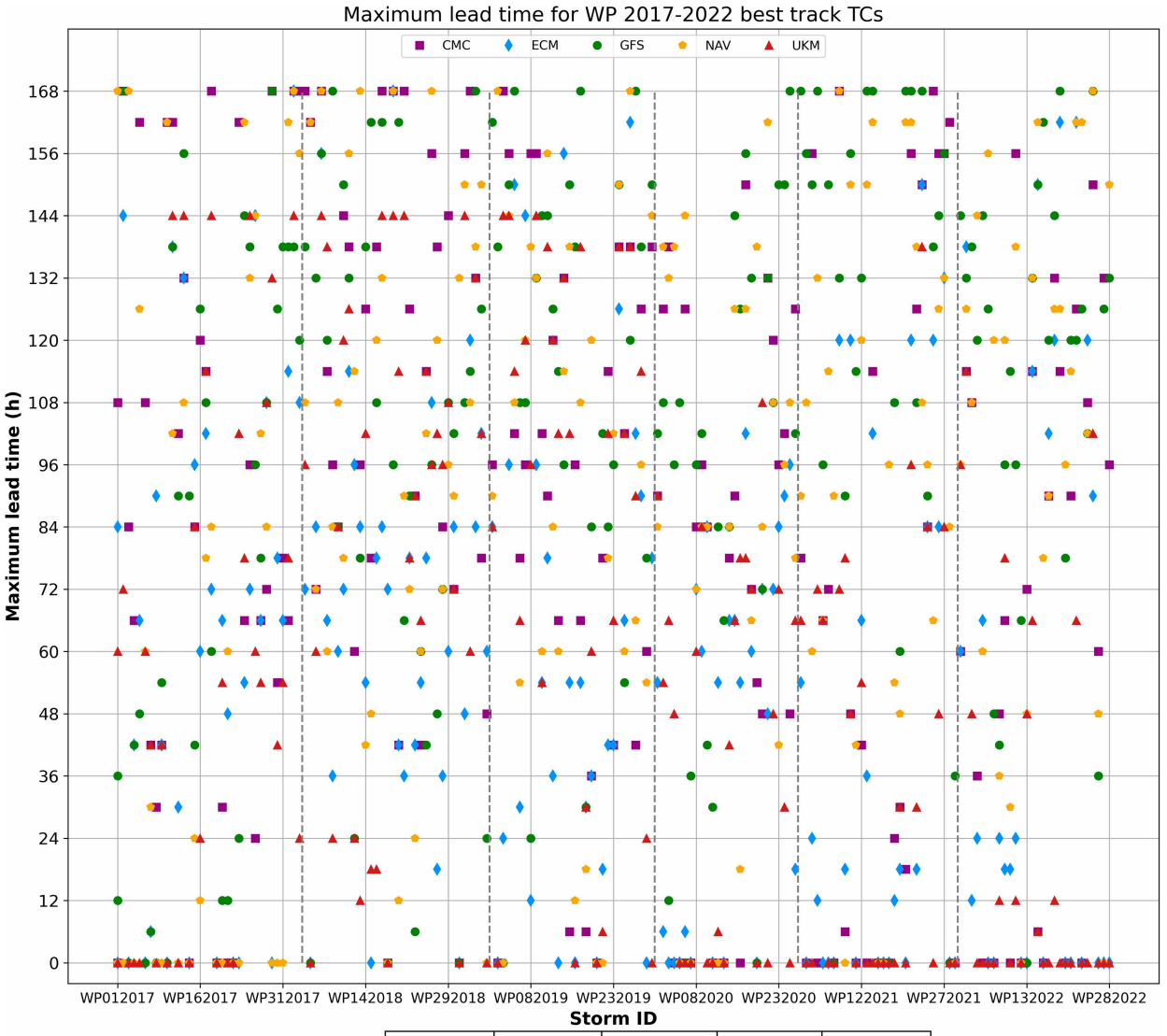


FIG. 20. As in Fig. 2, but for the WP basin.



	CMC (87 h)	ECM (54 h)	GFS (108 h)	NAV (96 h)	UKM (54 h)
CMC (87 h)		0.0	0.009	0.412	0.0
ECM (54 h)			0.0	0.0	0.701
GFS (108 h)				0.065	0.0
NAV (96 h)					0.0
UKM (54 h)					

FIG. 21. As in Fig. 3, but for the WP basin.

guidance, this study affirms the conclusion of Lin et al. (2023) that there remains room for additional improvements. Verifying TC genesis forecasts globally by genesis pathway may help provide insight regarding where such improvements are most needed. The forecasts verified here were used as the updated training dataset for post-processed guidance from TCLOGG. Future studies will describe the enhancements and extensions of the TCLOGG product suite since 2017.

Acknowledgments. The authors thank Julian Heming from the UKMO for providing the UKM model data used in this study. Expansion of TCLOGG guidance products is being developed by Ryan Remondelli at Florida State

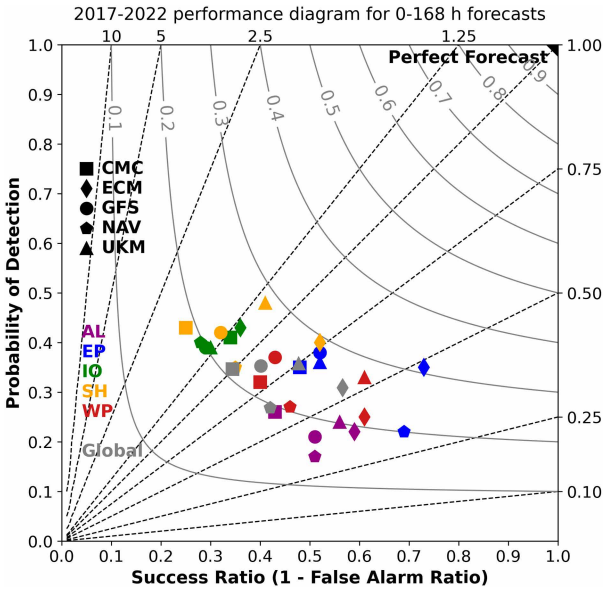


FIG. 22. As in Fig. 1, but for 2017–22 mean values for each basin.

University. The authors thank NOAA Joint Hurricane Testbed and ONR points of contact during the period of study Robert Ballard, Eric Blake, Alan Brammer, Josh Cossuth, Levi Cowan, Mark DeMaria, Wallace Hogsett, Chris Landsea, Matt Onderlinde, Richard Pasch, Matt Sardi, Brian Strahl, and Brian Zachry. This manuscript benefited from the constructive comments of three anonymous AMS reviewers. This study was supported by NOAA Awards NA18NWS4680066 and NA19OAR4590134/5 and ONR Award N00014-22-1-2704/5.

Data availability statement. Global model data are available from the TIGGE data archive (although potentially with different grid spacing than the data used in this study). Given the excessive volume of data required to complete this study and prior studies, it is impractical for the authors to fulfill individual data requests.

REFERENCES

Beven, J., 1999: The boguscane—A serious problem with the NCEP medium range forecast model in the tropics. Preprints, *23rd Conf. on Hurricanes and Tropical Meteorology*, Dallas, TX, Amer. Meteor. Soc., 845–848.

Biswas, M. K., D. Stark, and L. Carson, 2018: GFDL vortex tracker users guide version 3.9a. 35 pp., https://dtcenter.org/sites/default/files/community-code/gfdl/standalone_tracker_UG_v3.9a.pdf.

Brammer, A., C. D. Thorncroft, and J. P. Dunion, 2018: Observations and predictability of a nondeveloping tropical disturbance over the eastern Atlantic. *Mon. Wea. Rev.*, **146**, 3079–3096, https://doi.org/10.1175/MWR-D-18-0065.1.

Briegel, L. M., and W. M. Frank, 1997: Large-scale influences on tropical cyclogenesis in the western North Pacific. *Mon. Wea. Rev.*, **125**, 1397–1413, https://doi.org/10.1175/1520-0493(1997)125<1397:LSIOTC>2.0.CO;2.

Chan, J. C. L., and R. H. F. Kwok, 1999: Tropical cyclone genesis in a global numerical weather prediction model. *Mon. Wea. Rev.*, **127**, 611–624, https://doi.org/10.1175/1520-0493(1999)127<0611:TCGIAG>2.0.CO;2.

Chen, J.-H., S.-J. Lin, L. Zhou, X. Chen, S. Rees, M. Bender, and M. Morin, 2019a: Evaluation of tropical cyclone forecasts in the Next Generation Global Prediction System. *Mon. Wea. Rev.*, **147**, 3409–3428, https://doi.org/10.1175/MWR-D-18-0227.1.

——, and Coauthors, 2019b: Advancements in hurricane prediction with NOAA's next-generation forecast system. *Geophys. Res. Lett.*, **46**, 4495–4501, https://doi.org/10.1029/2019GL082410.

Cheung, K. K. W., and R. L. Elsberry, 2002: Tropical cyclone formations over the western North Pacific in the Navy Operational Global Atmospheric Prediction System forecasts. *Wea. Forecasting*, **17**, 800–820, https://doi.org/10.1175/1520-0434(2002)017<0800:TCFOTW>2.0.CO;2.

Dalcher, A., and E. Kalnay, 1987: Error growth and predictability in operational ECMWF forecasts. *Tellus*, **39A**, 474–491, https://doi.org/10.3402/tellusa.v39i5.11774.

Dunion, J. P., and Coauthors, 2023: Recommendations for improved tropical cyclone formation and position probabilistic forecast products. *Trop. Cyclone Res. Rev.*, **12**, 241–258, https://doi.org/10.1016/j.tcrr.2023.11.003.

ECMWF, 2024: Evolution of the IFS. https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model.

Elsberry, R., W. Clune, and P. Harr, 2009: Evaluation of global model early track and formation prediction in the western North Pacific. *Asia-Pac. J. Atmos. Sci.*, **45**, 357–374.

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.

——, ——, and ——, 2011: Evaluation of the ECMWF 32-day ensemble predictions during 2009 season of 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.

Environment and Climate Change Canada, 2024: Chronology of changes to the Global Deterministic Prediction System (GDPS). Accessed 15 June 2024, https://eccc-msc.github.io/open-data/msc-data/nwp_gdps/changelog_gdps_en/

Halperin, D. J., H. E. Fuelberg, R. E. Hart, J. H. Cossuth, P. Sura, and R. J. Pasch, 2013: An evaluation of tropical cyclone genesis forecasts from global numerical models. *Wea. Forecasting*, **28**, 1423–1445, https://doi.org/10.1175/WAF-D-13-00008.1.

——, ——, ——, and ——, 2016: Verification of tropical cyclone genesis forecasts from global numerical models: Comparisons between the North Atlantic and eastern North Pacific basins. *Wea. Forecasting*, **31**, 947–955, https://doi.org/10.1175/WAF-D-15-0157.1.

——, R. E. Hart, H. E. Fuelberg, and J. H. Cossuth, 2017: The development and evaluation of a statistical–dynamical tropical cyclone genesis guidance tool. *Wea. Forecasting*, **32**, 27–46, https://doi.org/10.1175/WAF-D-16-0072.1.

——, A. B. Penny, and R. E. Hart, 2020: A comparison of tropical cyclone genesis forecast verification from three Global Forecast System (GFS) operational configurations.

Wea. Forecasting, **35**, 1801–1815, https://doi.org/10.1175/WAF-D-20-0043.1.

Harris, L. M., and S.-J. Lin, 2013: A two-way nested global-regional dynamical core on the cubed-sphere grid. *Mon. Wea. Rev.*, **141**, 283–306, https://doi.org/10.1175/MWR-D-11-00201.1.

Hart, R. E., 2003: A cyclone phase space derived from thermal wind and thermal asymmetry. *Mon. Wea. Rev.*, **131**, 585–616, https://doi.org/10.1175/1520-0493(2003)131<0585:ACPSDF>2.0.CO;2.

Heming, J., and H. Titley, 2023: MET Office tropical cyclone modelling update. *77th Interdepartmental Hurricane Conf.*, Miami, FL, NOAA, 5.4, https://www.weather.gov/media/nws/IHC2023/Presentations/Heming_UKMET.pdf.

Hogan, T. F., and Coauthors, 2014: The Navy Global Environmental Model. *Oceanography*, **27** (**3**), 116–125, https://doi.org/10.5670/oceanog.2014.73.

Holbach, H. M., and M. A. Bourassa, 2014: The effects of gap-wind-induced vorticity, the monsoon trough, and the ITCZ on East Pacific tropical cyclogenesis. *Mon. Wea. Rev.*, **142**, 1312–1325, https://doi.org/10.1175/MWR-D-13-00218.1.

Hon, K. K., and Coauthors, 2023: Recent advances in operational tropical cyclone genesis forecast. *Trop. Cyclone Res. Rev.*, **12**, 323–340, https://doi.org/10.1016/j.tcrr.2023.12.001.

Hurley, J. V., and W. R. Boos, 2015: A global climatology of monsoon low-pressure systems. *Quart. J. Roy. Meteor. Soc.*, **141**, 1049–1064, https://doi.org/10.1002/qj.2447.

HYCOM, 2024: NAVGEM. Accessed 15 June 2024, https://www.hycom.org/dataserver/navgem.

Jensen, T., and Coauthors, 2023: The MET version 11.1.0 user's guide. Developmental Testbed Center, accessed 10 December 2023, https://metplus.readthedocs.io/projects/met/en/v11.1.0/Users_Guide/index.html.

Jiang, X., 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.

Komaromi, W. A., and S. J. Majumdar, 2015: Ensemble-based error and predictability metrics associated with tropical cyclogenesis. Part II: Wave-relative framework. *Mon. Wea. Rev.*, **143**, 1665–1686, https://doi.org/10.1175/MWR-D-14-00286.1.

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.

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.

Li, M., and Coauthors, 2020: An examination of the predictability of tropical cyclone genesis in high-resolution coupled models with dynamically downscaled coupled data assimilation initialization. *Adv. Atmos. Sci.*, **37**, 939–950, https://doi.org/10.1007/s00376-020-9220-9.

Liang, M., J. C. L. Chan, J. Xu, and M. Yamaguchi, 2021: Numerical prediction of tropical cyclogenesis Part I: Evaluation of model performance. *Quart. J. Roy. Meteor. Soc.*, **147**, 1626–1641, https://doi.org/10.1002/qj.3987.

——, ——, ——, and ——, 2022: Numerical prediction of tropical cyclogenesis. Part II: Identification of large-scale physical processes under the monsoon shear line synoptic pattern. *Quart. J. Roy. Meteor. Soc.*, **148**, 1965–1982, https://doi.org/10.1002/qj.4288.

Lin, S.-J., 2004: A "vertically Lagrangian" finite-volume dynamical core for global models. *Mon. Wea. Rev.*, **132**, 2293–2307, https://doi.org/10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2.

——, H.-H. Hsu, C.-Y. Tu, and C.-H. Chih, 2023: Assessment of an experimental version of fvGFS for TC genesis forecasting ability in the western North Pacific. *Wea. Forecasting*, **38**, 2271–2287, https://doi.org/10.1175/WAF-D-23-0056.1.

Lorenz, E. N., 1982: Atmospheric predictability experiments with a large numerical model. *Tellus*, **34A**, 505–513, https://doi.org/10.3402/tellusa.v34i6.10836.

Magnusson, L., and Coauthors, 2021: Tropical cyclone activities at ECMWF. ECMWF Tech. Memo. 888, 140 pp., https://www.ecmwf.int/en/elibrary/81277-tropical-cyclone-activities-ecmwf.

Marchok, T. P., 2002: How the NCEP tropical cyclone tracker works. Preprints, *25th Conf. on Hurricanes and Tropical Meteorology*, San Diego, CA, Amer. Meteor. Soc., Vol. **1**, P1.13, https://ams.confex.com/ams/25HURR/techprogram/paper_37628.htm.

McTaggart-Cowan, R., G. D. Deane, L. F. Bosart, C. A. Davis, and T. J. Galarneau Jr., 2008: Climatology of tropical cyclogenesis in the North Atlantic (1948–2004). *Mon. Wea. Rev.*, **136**, 1284–1304, https://doi.org/10.1175/2007MWR2245.1.

——, T. J. Galarneau Jr., L. F. Bosart, R. W. Moore, and O. Martius, 2013: A global climatology of baroclinically influenced tropical cyclogenesis. *Mon. Wea. Rev.*, **141**, 1963–1989, https://doi.org/10.1175/MWR-D-12-00186.1.

——, and Coauthors, 2019: Modernization of atmospheric physics parameterization in Canadian NWP. *J. Adv. Model. Earth Syst.*, **11**, 3593–3635, https://doi.org/10.1029/2019MS001781.

Molinari, J., and D. Vollaro, 2000: Planetary- and synoptic-scale influences on eastern Pacific tropical cyclogenesis. *Mon. Wea. Rev.*, **128**, 3296–3307, https://doi.org/10.1175/1520-0493(2000)128<3296:PASSIO>2.0.CO;2.

——, ——, S. Skubis, and M. Dickinson, 2000: Origins and mechanisms of eastern Pacific tropical cyclogenesis: A case study. *Mon. Wea. Rev.*, **128**, 125–139, https://doi.org/10.1175/1520-0493(2000)128<0125:OAMOEP>2.0.CO;2.

Mozer, J. B., and J. A. Zehnder, 1996: Lee vorticity production by large-scale tropical mountain ranges. Part I: Eastern North Pacific tropical cyclogenesis. *J. Atmos. Sci.*, **53**, 521–538, https://doi.org/10.1175/1520-0469(1996)053<0521:LVPBLS>2.0.CO;2.

NOAA, 2024: Changes to NCEP models and implementation dates to NOAAPORT. Accessed 15 June 2024, https://www.nco.ncep.noaa.gov/pmb/changes/.

Núñez Ocasio, K. M., and R. Rios-Berrios, 2023: African easterly wave evolution and tropical cyclogenesis in a pre-Helene (2006) hindcast using the Model for Prediction across Scales-Atmosphere (MPAS-A). *J. Adv. Model. Earth Syst.*, **15**, e2022MS003181, https://doi.org/10.1029/2022MS003181.

——, J. L. Evans, and G. S. Young, 2020: A wave-relative framework analysis of AEW–MCS interactions leading to tropical cyclogenesis. *Mon. Wea. Rev.*, **148**, 4657–4671, https://doi.org/10.1175/MWR-D-20-0152.1.

——, A. Brammer, J. L. Evans, G. S. Young, and Z. L. Moon, 2021: Favorable monsoon environment over eastern Africa for subsequent tropical cyclogenesis of African easterly waves. *J. Atmos. Sci.*, **78**, 2911–2925, https://doi.org/10.1175/JAS-D-20-0339.1.

——, C. A. Davis, Z. L. Moon, and Q. A. Lawton, 2024: Moisture dependence of an African easterly wave within the West

African monsoon system. *J. Adv. Model. Earth Syst.*, **16**, e2023MS004070, https://doi.org/10.1029/2023MS004070.

Papin, P. P., L. F. Bosart, and R. D. Torn, 2017: A climatology of Central American gyres. *Mon. Wea. Rev.*, **145**, 1983–2000, https://doi.org/10.1175/MWR-D-16-0411.1.

Pasch, R. J., P. A. Harr, A. Lixion, J. G. Jiing, and G. Elliot, 2006: An evaluation and comparison of predictions of tropical cyclogenesis by three global forecast models. Preprints, *27th Conf. on Hurricanes and Tropical Meteorology*, Monterey, CA, Amer. Meteor. Soc., 14A.5, https://ams.confex.com/ams/27Hurricanes/webprogram/Paper108725.html.

——, E. Blake, J. Jiing, M. Mainelli, and D. Roberts, 2008: Performance of the GFS in predicting tropical cyclone genesis during 2007. *28th Conf. on Hurricanes and Tropical Meteorology*, Orlando, FL, Amer. Meteor. Soc., 11A.7, https://ams.confex.com/ams/28Hurricanes/techprogram/paper_138218.htm.

Rajasree, V. P. M., and Coauthors, 2023: Tropical cyclogenesis: Controlling factors and physical mechanisms. *Trop. Cyclone Res. Rev.*, **12**, 165–181, https://doi.org/10.1016/j.tcrr.2023.09.004.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, https://doi.org/10.1175/2008WAF2222159.1.

Royer, J.-F., R. Stroe, M. Déqué, and S. Vannitsem, 1994: An improved formula to describe error growth in meteorological models. *Predictability and Nonlinear Modelling in Natural Sciences and Economics*, J. Grasman and G. van Straten, Eds., Springer, 45–56.

Russell, J. O., A. Aiyyer, J. D. White, and W. Hannah, 2017: Revisiting the connection between African easterly waves and Atlantic tropical cyclogenesis. *Geophys. Res. Lett.*, **44**, 587–595, https://doi.org/10.1002/2016GL071236.

Rydbeck, A. V., E. D. Maloney, and G. J. Alaka Jr., 2017: In situ initiation of East Pacific easterly waves in a regional model. *J. Atmos. Sci.*, **74**, 333–351, https://doi.org/10.1175/JAS-D-16-0124.1.

Sampson, C. R., and A. J. Schrader, 2000: The automated tropical cyclone forecasting system (version 3.2). *Bull. Amer. Meteor. Soc.*, **81**, 1231–1240, https://doi.org/10.1175/1520-0477(2000)081<1231:TATCFS>2.3.CO;2.

Schenkel, B. A., and R. E. Hart, 2012: An examination of tropical cyclone position, intensity, and intensity life cycle within atmospheric reanalysis datasets. *J. Climate*, **25**, 3453–3475, https://doi.org/10.1175/2011JCLI4208.1.

Schumacher, A. B., M. DeMaria, and J. A. Knaff, 2009: Objective estimation of the 24-h probability of tropical cyclone formation. *Wea. Forecasting*, **24**, 456–471, https://doi.org/10.1175/2008WAF2007109.1.

Serra, Y. L., G. N. Kiladis, and K. I. Hodges, 2010: Tracking and mean structure of easterly waves over the intra-Americas Sea. *J. Climate*, **23**, 4823–4840, https://doi.org/10.1175/2010JCLI3223.1.

Teng, H.-F., J. M. Done, C.-S. Lee, H.-H. Hsu, and Y.-H. Kuo, 2020: Large-scale environmental influences on tropical cyclone formation processes and development time. *J. Climate*, **33**, 9763–9782, https://doi.org/10.1175/JCLI-D-19-0709.1.

Tsai, H.-C., K.-C. Lu, R. L. Elsberry, M.-M. Lu, and C.-H. Sui, 2011: Tropical cyclone–like vortices detection in the NCEP 16-Day ensemble system over the western North Pacific in 2008: Application and forecast evaluation. *Wea. Forecasting*, **26**, 77–93, https://doi.org/10.1175/2010WAF2222415.1.

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.

Wang, Z., W. Li, M. S. Peng, X. Jiang, R. McTaggart-Cowan, and C. A. Davis, 2018: Predictive skill and predictability of North Atlantic tropical cyclogenesis in different synoptic flow regimes. *J. Atmos. Sci.*, **75**, 361–378, https://doi.org/10.1175/JAS-D-17-0094.1.

Whitaker, J. W., and E. D. Maloney, 2018: Influence of the Madden–Julian oscillation and Caribbean low-level jet on east Pacific easterly wave dynamics. *J. Atmos. Sci.*, **75**, 1121–1141, https://doi.org/10.1175/JAS-D-17-0250.1.

——, and ——, 2020: Genesis of an East Pacific easterly wave from a Panama bight MCS: A case study analysis from June 2012. *J. Atmos. Sci.*, **77**, 3567–3584, https://doi.org/10.1175/JAS-D-20-0032.1.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*, Vol. 100. 3rd ed. Academic Press, 704 pp.

Yang, F., 2018: GDAS/GFS V15.0.0 upgrades for Q2FY2019. 59 pp., https://www.emc.ncep.noaa.gov/users/Alicia.Bentley/fv3gfs/updates/EMC_CCB_FV3GFS_9-24-18.pdf.

Zehnder, J. A., 1991: The interaction of planetary-scale tropical easterly waves with topography: A mechanism for the initiation of tropical cyclones. *J. Atmos. Sci.*, **48**, 1217–1230, https://doi.org/10.1175/1520-0469(1991)048<1217:TIOPST>2.0.CO;2.

——, and R. L. Gall, 1991: On a mechanism for orographic triggering of tropical cyclones in the eastern North Pacific. *Tellus*, **43A**, 25–36, https://doi.org/10.3402/tellusa.v43i1.11913.

——, D. M. Powell, and D. L. Ropp, 1999: The interaction of easterly waves, orography, and the intertropical convergence zone in the genesis of eastern Pacific tropical cyclones. *Mon. Wea. Rev.*, **127**, 1566–1585, https://doi.org/10.1175/1520-0493(1999)127<1566:TIOEWO>2.0.CO;2.

Zhan, R., Y. Wang, and Y. Ding, 2022: Impact of the western Pacific tropical easterly jet on tropical cyclone genesis frequency over the western North Pacific. *Adv. Atmos. Sci.*, **39**, 235–248, https://doi.org/10.1007/s00376-021-1103-1.