Atlantic Tropical Easterly Wave and Cyclogenesis Forecasting in the Physical and New AI Forecast Systems at ECMWF

SHARANYA J. MAJUMDAR,[a] LINUS MAGNUSSON,[b] QUINTON A. LAWTON,[c] REBECCA EMERTON,[b] SIMON T. K. LANG,[b] MICHAEL MAIER-GERBER,[d] AND DAVID S. RICHARDSON[b]

[a] Department of Atmospheric Sciences, University of Miami, Miami, Florida
[b] ECMWF, Reading, United Kingdom
[c] NSF National Center for Atmospheric Research, Boulder, Colorado
[d] ECMWF, Bonn, Germany

(Manuscript received 23 September 2025, in final form 8 February 2026, accepted 13 April 2026)

ABSTRACT: ECMWF forecasts of African easterly waves and tropical cyclogenesis in the Atlantic basin are investigated during 2020–24, with a focus on Integrated Forecasting System (IFS) upgrades and the new Artificial Intelligence Forecasting System (AIFS). Ensemble-based probabilistic forecasts, valid at the time a tropical storm was named, exhibited high variability, with sensitivity to the maximum wind speed threshold. In many cases, the probabilities rose sharply when the lead time was reduced from 72 to 48 h. The average probabilities have generally increased in each year, including when the IFS ensemble grid spacing was reduced from 18 to 9 km in 2023. Across 18 developing tropical cyclones in 2024, the probabilities based on the artificial intelligence (AI)-based ensemble system ["AIFS-continuous ranked probability score (CRPS)"] often exceeded IFS probabilities for 84–120-h lead times, especially for stronger waves. In contrast, the AIFS-CRPS probabilities were lower for 36–48-h lead times, especially for the weakest waves. The average AIFS-CRPS ensemble-mean position forecast error was often lower than that of the AIFS-Single, IFS deterministic, and IFS ensemble-mean forecasts. The wave locations in the control (deterministic) IFS forecasts exhibited a greater southward bias than the single AIFS forecasts ("AIFS-Single"), whereas both models exhibited a slow bias in longitude. Mean absolute errors in AIFS-Single were mostly smaller for 3–5-day forecasts of 850–500-hPa wave-relative environmental vorticity and specific humidity, with lower biases in these fields. Overall, the AIFS-CRPS ensemble and AIFS-Single forecasts serve as a useful complement to the IFS.

SIGNIFICANCE STATEMENT: ECMWF has recently made upgrades to the Integrated Forecasting System (IFS), and in 2025, the Artificial Intelligence Forecasting System (AIFS) became operational. The IFS ensemble-based probabilities of tropical cyclone formation between 2020 and 2024 varied from storm to storm, with a frequent sharp jump in probability between 3- and 2-day forecasts of formation. In 2024, the AIFS had similarities and differences to the IFS, including higher probabilities for stronger tropical waves at 3.5–5 days and lower probabilities for weaker tropical waves at 1.5–2 days. The single IFS forecasts of tropical waves were on average further south, weaker, and moister than the AIFS counterparts. We expect the AIFS to add value to the IFS for forecasts of tropical cyclone formation.

KEYWORDS: North Atlantic Ocean; Tropical cyclones; Ensembles; Model comparison; Numerical weather prediction/forecasting; Machine learning

1. Introduction

Forecasting the genesis of tropical cyclones (TCs) remains an important challenge. In some instances, a newly named tropical storm has intensified rapidly into a major hurricane and made landfall within just 3 days (e.g., Ida 2021 and Helene 2024), accentuating the need for high-confidence probabilistic forecasts up to a week prior to its formation to enable initial preparations and the deployment of resources to assist with the potential aftermath. In the Atlantic basin, the National Hurricane Center (NHC) issues 2- and 7-day Tropical Weather Outlooks to illustrate areas of disturbed weather that have the potential to develop into a TC.

Genesis forecasts, especially in the 7-day range, rely on predictions from numerical weather prediction (NWP) models. The NHC uses a statistical–dynamical tool that utilizes logistic regression to compute genesis probabilities, based on several deterministic NWP models (Halperin and Hart 2025). Ensembles of global model forecasts are also used to generate genesis probabilities (Majumdar and Torn 2014; Yamaguchi and Koide 2017; Titley et al. 2019; Dunion et al. 2023; Hon et al. 2023; Zhang et al. 2023). A probabilistic evaluation of 1–10-day ECMWF Integrated Forecasting System (IFS) ensemble forecasts (Molteni et al. 1996) between 2019 and 2023 was recently performed by Richardson et al. (2025). They found that genesis is picked up at least 7 days ahead in about half of the cases, although the forecast confidence is often only strong within 3 days of genesis. They also indicated that genesis probabilities can be jumpy between successive runs.;

Tropical cyclones originate from precursor disturbances over the ocean. Forecasts are complicated due to these precursors being disorganized clusters of thunderstorms. Strong precursors with high relative vorticity and favorable environmental conditions are more likely to develop into TCs. In most cases, since

Corresponding author: Sharanya J. Majumdar, s.majumdar@miami.edu

DOI: 10.1175/WAF-D-25-0200.1

genesis usually occurs equatorward of 25°N, the ocean thermal structure is favorable. In contrast, the vertical wind shear can vary, as can the premoistening process, which is necessary to sustain deep convection (Nolan 2007; Wang 2012, 2014). Thermodynamic processes (e.g., Russell and Aiyyer 2020; Lawton and Majumdar 2023) have recently been emphasized, including the areal coverage of intense convection in developing waves (e.g., Brammer et al. 2018; Zawislak 2020) and locally increasing convective intensity (Nolan et al. 2007; Wang et al. 2010; Fritz et al. 2016). We also note that there is likely significant variability in the coverage, intensity, and duration of convection (Wang 2018), which could lead to variability in forecast skill. It is unclear whether global models and ensembles are able to capture all these processes adequately, although they have shown promise in genesis prediction (e.g., Halperin and Hart 2025; Richardson et al. 2025).

At ECMWF, several advances have recently taken place. In the IFS, the horizontal grid spacing in the 50-member ensemble was reduced from 18 to 9 km in 2023 (version 48r1; Lang et al. 2023), and a new stochastically perturbed parameterization (SPP) scheme (Leutbecher et al. 2017; Ollinaho et al. 2017; Lang et al. 2021a) was introduced in late 2024 (version 49r1). Additionally, a new Artificial Intelligence Forecasting System (AIFS) has been introduced, with separate single ("AIFS-Single"; Lang et al. 2024a) and 50-member ensemble forecasts ["AIFS-continuous ranked probability score (CRPS)"; Lang et al. 2024b] going operational in 2025, representing the first fully operational weather prediction model and ensemble that use machine learning. AIFS-Single and AIFS-CRPS are distinct models that are separately trained, although they share the same code and training data.

An assessment of early artificial intelligence (AI)-based forecasts demonstrated the potential for improved forecasts relative to the IFS, including for TC track (Ben Bouallègue et al. 2024). In the first documentation of skill in the single AIFS forecasts, Lang et al. (2024a) demonstrated that the AIFS possessed statistical superiority over the IFS for upper-air and surface weather variables, including TC track. However, similar to the findings of Ben Bouallègue et al. (2024), they found the AIFS to be inferior to the IFS for intensity, potentially due to the generally too low intensity of TCs in the training datasets, and the tendency of the AIFS to smooth the forecast fields. Haiden et al. (2025; Fig. 46) also cited similar results for July 2024–June 2025. Other AI-based models are also showing promise for TC track (e.g., Radford et al. 2025).

No studies have yet investigated the ECMWF AIFS for genesis prediction. Since the AIFS is new, the sample is small (2023–24 for AIFS-Single and 2024 for AIFS-CRPS), and genesis is verified probabilistically, a rigorous verification for the Atlantic basin is not yet feasible. Instead, we focus on insights into the AIFS relative to the IFS. Given that the AIFS is trained on reanalysis data with a ~30-km grid spacing and an IFS analysis scaled up from ~9 to ~30 km, and that genesis involves convective-scale processes, it is not clear whether the AIFS can provide the added value for genesis that has been demonstrated for synoptic-scale metrics. We note that several statistical schemes using other AI techniques have been developed for genesis (Zhang et al. 2022; Loi et al. 2024; Nguyen and Kieu 2024) with encouraging results and statistical connections with key physical variables.

Since about 60% of Atlantic TCs develop from African easterly waves (AEWs) (Russell et al. 2017), it is instructive to compare the basic characteristics of these waves in the IFS and AIFS and assess whether the ensemble-based genesis probabilities in these systems are related to the AEW properties. A useful study for context is Elless and Torn (2018), who found that ECMWF AEW position forecasts had a slow-motion bias and that the ensemble forecasts were underdispersive for the periods of 2007–09 and 2011–13. They found mixed connections between errors in the AEW intensity (area-averaged curvature vorticity) and errors in the convection. Lawton et al. (2025) identified relationships between interannual variability of TC activity and AEWs including their near-environmental characteristics. The object-oriented tropical easterly wave tracking software "QTrack" [introduced by Lawton et al. (2022) and available via Lawton (2024)] has been implemented at ECMWF to track AEWs in the Atlantic, facilitating our investigation of wave characteristics in the ECMWF framework.

In this paper, we provide an overview of the impact of recent upgrades to the IFS in the context of genesis, together with an early assessment of AIFS-Single and ensemble forecasts of genesis and tropical easterly waves. Our specific goals are to discern the effects of increasing the resolution, modifying the ensemble perturbation scheme, the use of AI-based versus physics-based frameworks, and how the predictability of genesis differs in tropical cyclones that form from easterly waves versus those that do not. The intent of the paper is to provide context for ongoing tropical cyclone modeling effects, tropical weather forecasting, and AI applications in research and forecasting. The roadmap for the paper is provided in Fig. 1, in which we will address the following sequence of questions:

- What are the characteristics of IFS ensemble-based forecast probabilities between 2020 and 2024? (section 3a)
- Do these probabilities vary based on (i) the region in the Atlantic basin and (ii) whether or not the precursor disturbance is a tropical easterly wave? (section 3a)
- Do the characteristics of the probabilistic forecasts change when (i) the grid spacing of the ensemble is reduced from 18 to 9 km and (ii) the SPP ensemble perturbation technique is introduced? (section 3b)
- How do the probabilities in the new AIFS-CRPS ensemble system compare with those in the AIFS ensemble? (section 3c)
- How do the position forecast errors in the deterministic and ensemble mean compare between IFS and AIFS? (section 3d)
- For the tropical easterly waves that developed into TCs, how do their forecasts compare for the IFS and AIFS? (section 4)

Following a description of the data and methods in section 2, the above questions are investigated for 1–10-day ensemble-based probabilistic forecasts and single forecasts in sections 3 and 4. Concluding remarks are provided in section 5.

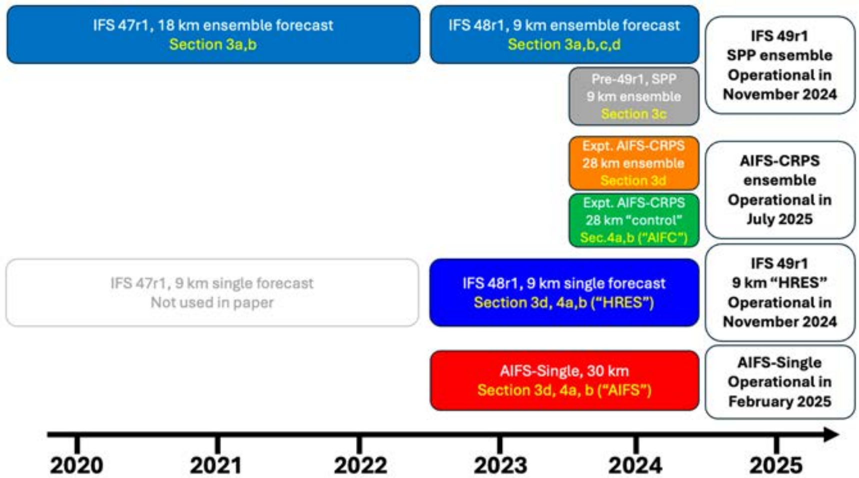

FIG. 1. Timeline of the version upgrades and new forecast systems at ECMWF, including the relevant sections in the paper where the various systems are used.

2. Data and methods

a. ECMWF IFS

For the period examined in this paper, the IFS comprised a single high-resolution ("HRES") forecast integrated from the ECMWF analysis, a control forecast integrated from the same analysis (not used in this study), and a 50-member ensemble of perturbed forecasts. The IFS is typically upgraded once per year with a new "cycle." Our sample of TC cases begins with cycle 47r1, which was launched on 30 June 2020, with an 18-km ensemble. An upgrade on 12 October 2021 (cycle 47r3) included the improved representation of moist physics and the increased assimilation of satellite observations in cloudy regions. From 27 June 2023 (cycle 48r1), a significant upgrade involved reducing the grid spacing in the ensemble to 9 km (Lang et al. 2023). Cycle 49r1 was launched on 12 November 2024, and the experimental run of this version before its launch will be compared here with cycle 48r1. Two significant changes to the ensemble in cycle 49r1 include an increased resolution and soft centering of the ensemble of data assimilations (EDAs) and the inclusion of SPP to represent model uncertainty.[1] In addition to these changes, most cycles include changes to the data assimilation and observation usage (Magnusson et al. 2025).

The initial conditions for the 9-km HRES forecasts are provided by the 4D-Var data assimilation scheme. A flow-dependent background error covariance matrix is estimated via the EDA (Isaksen et al. 2010), which is also used to perturb the ensemble initial conditions. The perturbation methodology is described in Lang et al. (2021b). We also note that the ensemble initial conditions are perturbed using dry singular vectors (SVs) in the 30°–90°N and 30°–90°S latitude bands, moist SVs for TCs, and moist SVs in the Caribbean.

[1] Additional details of changes to the ECMWF IFS configurations are provided in https://www.ecmwf.int/en/forecasts/documentation-and-support/changes-ecmwf-model.

b. ECMWF AIFS

AIFS-Single forecasts (Lang et al. 2024a), which became operational on 25 February 2025, are computed from the IFS initial conditions. In this paper, we use experimental AIFS-Single forecasts from 2023 to 2024, which have the same configuration with a ~28-km grid. The AIFS architecture is based on a graph neural network encoder and decoder, with a sliding window transformer processor. It is pretrained on 6-h forecasts from the fifth generation ECMWF atmospheric reanalysis (ERA5; Hersbach et al. 2020) dataset from 1979 to 2022. Following this, a rollout on operational IFS analyses between 2016 and 2022 is performed, with training on successive forecasts up to 72 h. Successive 6-h forecasts are initialized from previous forecasts using an autoregressive model. The loss function is mean-square error (MSE). Known limitations include blurred fields at longer forecast times, since the MSE loss function encourages the smoothing of fields to avoid a "double penalty" when weather features are misplaced and a lack of reproduction of energy spectra or dynamical relationships that are evident in physical models.

The AIFS-CRPS ensemble (Lang et al. 2024b) went operational at ECMWF on 1 July 2025 ["AIFS Ensemble, version 1 (ENS v1)"; ECMWF 2025]. The loss function is based on the "almost fair" CRPS. The probabilistic training involves the propagation of a small ensemble forward in time using separate model instances (with the same weights) within the encoder–processor–decoder framework (similar to that used for AIFS-Single), following which ensemble sharding is used to gather the ensemble forecasts. The CRPS loss is then calculated from the forecast ensemble, using the ERA5 analysis as the target and operational IFS analyses during fine-tuning. Each of the 50 AIFS-CRPS ensemble members is initialized from the perturbed initial condition of an IFS ensemble member. Additionally, a "control" AIFS-CRPS forecast is initialized from the unperturbed IFS analysis. Like for the 50 ensemble members, the forward model is stochastic due to the probabilistic training approach (and hence is not a control in the conventional

NWP sense), and it appears less skillful when verified as a deterministic model compared with AIFS-Single and the IFS control member. We use this AIFS-CRPS control forecast (referred to herein as "AIFC") as a representative member of the ensemble to connect the ensemble-based probabilities (section 3) with the corresponding easterly wave characteristics (section 4).

c. Tropical cyclone and wave tracking

To identify TCs in the ensemble forecasts, we use the operational ECMWF tracker (Magnusson et al. 2021), which identifies warm-core vortices at each time step and then estimates the surface position, minimum surface pressure, and maximum 10-m wind speed (Vmax) at the full model resolution (the grid spacings for each version are listed in Fig. 1). Given the negative bias in Vmax in the IFS (Majumdar et al. 2023) and ERA5, we expect that the threshold value of Vmax in the IFS (and AIFS) will be lower than the corresponding NHC best track value. While there is some arbitrariness in the selection of the threshold Vmax value, we choose 25 kt (1 kt $\approx$ 0.51 m s^{-1}) for both IFS and AIFS for simplicity, given that 20 kt is very weak, and the probabilities of 30-kt vortices seem abnormally low (Fig. 2 below).

To identify and track AEWs in the individual forecasts, we use version 0.0.1 of the QTrack methodology of Lawton et al. (2022), which follows Brammer and Thorncroft (2015) and Elless and Torn (2018) and has recently been implemented at ECMWF. Maxima of 700-hPa curvature vorticity equatorward of 25°N, averaged within a disk of radius 600 km with a minimum threshold value of 10^{-7} s^{-1}, are identified via a weighted centroid technique and tracked. As with all conventional trackers, this version of QTrack successfully captures wave propagation in the eastern and central Atlantic, although it can miss very weak waves, especially in the western Atlantic and Caribbean (Downs et al. 2025). For computational efficiency, we use the model fields on a $1^\circ \times 1^\circ$ grid, following Lawton et al. (2022). The wave-relative diagnostics are also averaged within disks of 600-km radius, with our emphasis on environmental variables. We note that after a TC has formed, the wave center is not necessarily the same as the TC center in the best track, since the wave center locates a broad region of vorticity rather than a small closed circulation.

d. Case selection

The sample of cases is based on the availability of forecast data, which is limited for the experimental forecast versions. For the ensemble-based probabilities in section 3, the IFS includes all Atlantic basin TCs between July 2020 and November 2024 (Table 1). A case is regarded as a "wave" if the main precursor disturbance was cited in the NHC Tropical Cyclone Report (NHC 2025) as being of wave origin. The minority of nonwave origin TCs formed from precursors such as gyres, troughs, broad low pressure systems, and mesoscale convective systems. For the comparison between IFS cycle 48r1 (which was operational through most of 2024) and the then-experimental parallel cycle that became IFS cycle 49r1 in November 2024, the sample is restricted to daily 0000 UTC initializations in 2024 for 14 TCs. The AIFS-CRPS ensemble was initiated twice per day in 2024 in experimental mode, and here, we compare the 0000 and 1200 UTC ensembles against their IFS counterparts for the 2024 season.

For the wave tracking in section 4, a homogeneous sample of HRES, AIFS-Single (labeled as "AIFS"), and AIFS-Control (labeled as AIFC) forecasts in 2024, initialized daily at 0000 UTC, is used. A homogeneous sample of HRES and AIFS forecasts initialized daily at 0000 UTC during 2023–24 is also examined.

e. Computation of ensemble-based TC genesis probabilities

We use 0–10-day ensembles initialized at 0000 and 1200 UTC, with forecast outputs in 12-hourly increments. If the NHC declared a named tropical storm at 0600 UTC (1800 UTC), we use the position valid at the next 1200 UTC (0000 UTC) time as the genesis location. We compute the percentage of the 50 perturbed members that were identified as a $\geq$25-kt TC within 1000 km of the actual position at the genesis time (when the tropical storm is named). This allows for realistic errors in position and timing and is generous in allowing a threshold that is lower than the NHC threshold for a tropical storm (34 kt). No time tolerance is included, and we consider the forecast a "miss" if no TC is identified in an ensemble member at the time that the tropical storm is named. We examined different space and time tolerances, and the results differed minimally. Given the small sample, we do not perform a probabilistic verification and do not formally calibrate the Vmax threshold value (as Majumdar and Torn 2014 had done). Instead, our focus is on comparing relative probabilities across forecasting systems.

f. Easterly wave forecast evaluations

The wave track locations for the single HRES, AIFS, and AIFC forecasts are computed using QTrack, as are the locations from the IFS analysis, which are used as the genesis locations. There are minor differences between the initial tracked location in the AIFS versus the IFS since QTrack postprocesses and smooths wave tracks after the complete track is generated, but these do not affect the results. A total of 18 waves from 2023 to 2024 are included, with 7 of the 25 waves in Table 1 omitted due to missing data (Bret) or the wave not being tracked continuously (Gert, Harold, Jose, Katia, Francine, Sean). The latitudinal and longitudinal errors of the forecasted wave locations are computed relative to the analysis locations.

For wave-relative environmental variables, we compute 850-, 700-, and 500-hPa vorticity and specific humidity, 200-hPa divergence, and 200–850-hPa wind shear, each averaged in a disk of radius 600 km centered on the wave location. The errors in these variables are computed relative to the corresponding values in the IFS analysis. Since vorticity and divergence are not available as AIFS outputs, we compute the vorticity via Stokes' theorem, as the closed line integral of the azimuthal wind around the circle of fixed radius (the circulation), divided by the area within the circle. Using the 2D divergence theorem, we compute the divergence as the net outward flow (radial wind) normal to the circle, divided by the area within the circle.

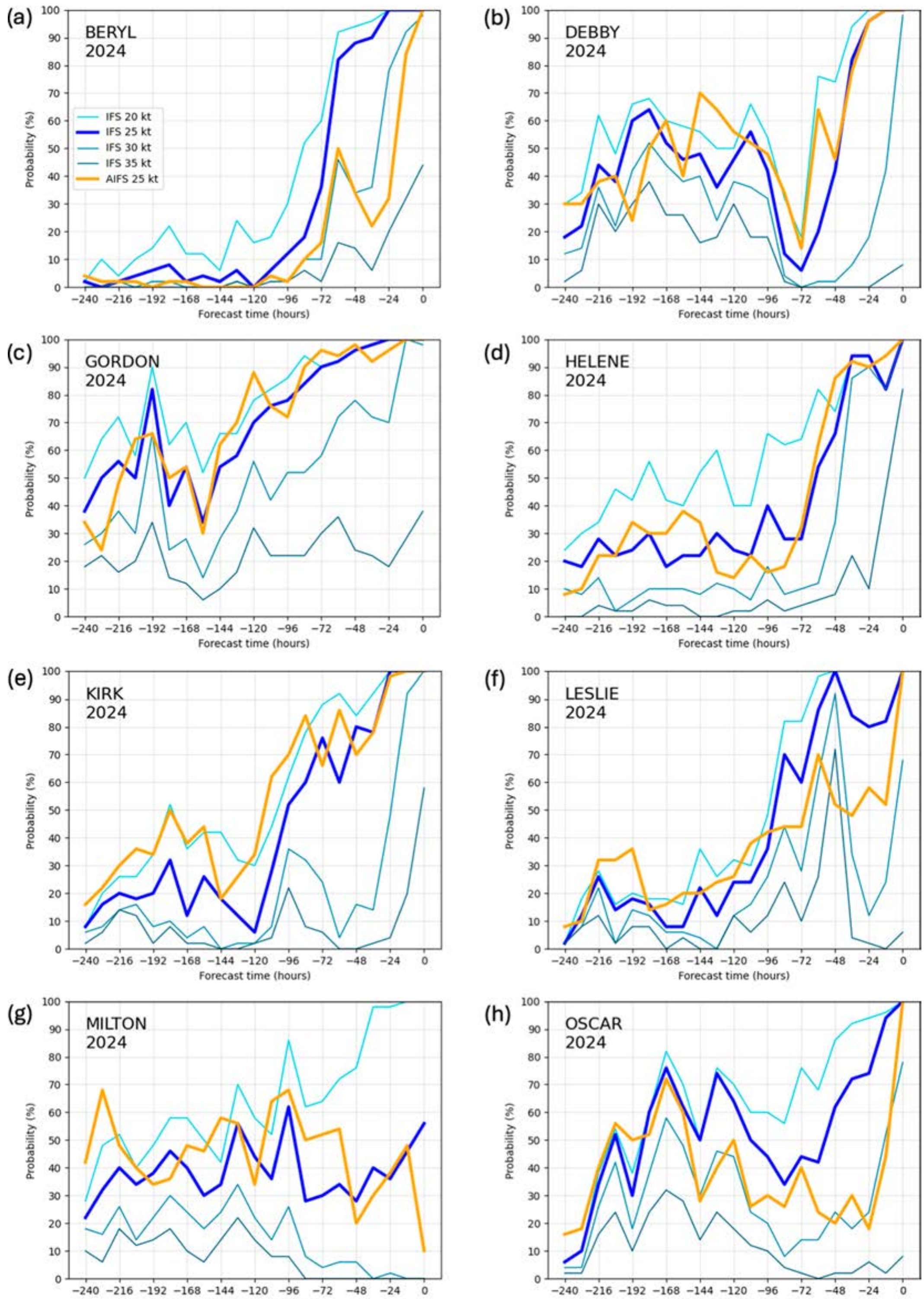

FIG. 2. Progression of 10–0-day ensemble-based probabilities for eight cases in 2024, valid at the time the tropical storms are named: (a) Beryl, (b) Debby, (c) Gordon, (d) Helene, (e) Kirk, (f) Leslie, (g) Milton, and (h) Oscar. The probabilities are computed for different Vmax threshold values, with the thick blue (orange) lines corresponding to IFS (AIFS-CRPS) probabilities when the Vmax threshold is set to 25 kt.

TABLE 1. List of cases, with those of wave origin bolded. Potential TCs and cases that did not reach tropical storm status were excluded, as was the short-lived Subtropical Storm Alpha (2020). The italicized storms in 2023–24 correspond to the 18 wave cases that are included in the wave-tracking study in section 4.

Year	No. of cases	TC name
2020	23	**Gonzalo**, **Hanna**, **Isaias**, **Josephine**, Kyle, **Laura**, **Marco**, **Nana**, Omar, **Paulette**, **Rene**, Sally, **Teddy**, **Vicky**, Wilfred, **Beta**, **Gamma**, **Delta**, Epsilon, **Zeta**, **Eta**, Theta, **Iota**
2021	20	Ana, Bill, **Claudette**, Danny, **Elsa**, **Fred**, **Grace**, Henri, **Ida**, **Julian**, **Kate**, **Larry**, **Mindy**, **Nicholas**, Odette, **Peter**, **Rose**, **Sam**, **Victor**, Wanda
2022	14	Alex, **Bonnie**, Colin, Danielle, **Earl**, **Fiona**, **Gaston**, **Hermine**, **Ian**, **Julia**, Karl, **Lisa**, Martin, Nicole
2023	19	Arlene, **Bret**, ***Cindy***, Don, ***Emily***, ***Franklin***, **Gert**, **Harold**, Idalia, **Jose**, **Katia**, ***Lee***, ***Margot***, ***Nigel***, Ophelia, ***Philippe***, ***Rina***, Sean, ***Tammy***
2024	18	Alberto, ***Beryl***, ***Chris***, ***Debby***, ***Ernesto***, **Francine**, ***Gordon***, Helene, Isaac, ***Joyce***, ***Kirk***, ***Leslie***, Milton, Nadine, ***Oscar***, Patty, Rafael, Sara

3. Results: Ensemble forecast probabilities of TC genesis

a. IFS ensemble-based probabilistic forecasts

We begin by illustrating 0–10-day IFS probabilities valid at the time the tropical storm was named, for cases from 2024 that were interesting and/or high-impact events. For Beryl, an early season TC, 52% of the 84-h ensemble forecast members predicted ≥ 20-kt vortices (thin light blue line) at the time Beryl was named. In contrast, for the 25-kt threshold (thick blue line), the 84-h forecast probability was only 18%, before a rapid increase to 82% in the 60-h forecast (Fig. 2a). For Debby, the IFS ensemble predicted substantial confidence (~40%–60%) of a TC forming in the 4–9-day forecasts, using the 25-kt threshold (Fig. 2b). However, the probability dropped dramatically to 8% at 72 h, before rising sharply to 100%. Debby was an interesting case in which the earlier ensemble members contained vortices east of Florida, whereas the 72-h predictions were likely influenced by land interactions, before the shorter-range predictions correctly shifted Debby's genesis over the Caribbean Sea near southern Cuba. In contrast to Debby, Gordon was a well-behaved robust wave over the open ocean. Following high IFS probabilities in its 5–10-day forecasts, the values for Gordon then increased steadily (Fig. 2c). Helene, which developed from a Central American Gyre and rapidly intensified into a destructive hurricane just 2 days after it was named, had only a 28% chance of formation 72 h before this time (Fig. 2d). While Helene's genesis source and mechanism were different to Beryl, the progression of probabilities was somewhat similar with a sharp increase in the 3-day range. For all time ranges from 60 to 204 h, the probability values were ~20% higher when the 20-kt threshold was used, indicating a high sensitivity to the Vmax threshold. Next, Kirk showed a similar progression using the 25-kt threshold, but for the stricter 30-kt threshold, the probabilities were only ~20% or even lower for the 36–60-h forecasts, again indicating high sensitivity to the Vmax threshold (Fig. 2e). The progression of IFS ensemble probabilities for Leslie was similar to those of Kirk, with a similar "dropout" at very short lead times when the 30-kt threshold was used, suggesting difficulty in spinning up an adequate vortex 1 day after initialization (Fig. 2f). Milton, which formed from a complex trough system, wavered between 30% and 60% for 0–9-day forecasts using the 25-kt threshold (Fig. 2g). With the 20-kt threshold, the progression steadily increased to 100%, whereas for 30 kt, the progression remarkably decreased to 0%. Finally, Oscar exhibited run-to-run inconsistency (or "jumpiness," Richardson et al. 2025) 5–9 days prior to its naming, following which confidence in formation decreased before increasing again (Fig. 2h).

These and other cases illustrate the storm-to-storm variability in the probabilities, some low probabilities just ~72 h prior to their naming, and some unexpected drops in probability as the lead time decreases. There is often a high sensitivity of the probabilities to the Vmax threshold (Zhang et al. 2023). Our probabilities are undesirably low for the 30- and 35-kt thresholds and are elevated for the 20-kt threshold. We keep our threshold at 25 kt for the rest of this paper.

Across all the cases and lead times, the AIFS-CRPS ensemble probabilities (orange lines in Fig. 2) using the 25-kt threshold do not show obvious systematic differences compared with the IFS, even though the AIFS-CRPS is of lower resolution (~28 vs 9 km in the IFS). Some cases (Beryl, Leslie, Oscar) show substantially lower AIFS-CRPS probabilities at ~2 days, whereas others (Kirk and others not shown) show substantially higher AIFS-CRPS probabilities at ~5 days. We will revisit this in section 3c. The Debby and Oscar cases show similar undesirable fluctuations in AIFS-CRPS as in the IFS, suggesting that the ensemble forecasts, regardless of the model, are strongly influenced by the initial conditions.

The chart of IFS ensemble probabilities for all 94 cases between 2020 and 2024 exhibits a wide range of probabilities at each time, indicating large case-to-case variability (Fig. 3). While the probabilities generally increase as the lead time approaches 0 h, a fairly abrupt transition from many low probabilities for ~72-h forecasts to many high probabilities for ~48-h forecasts is evident, as illustrated by Beryl. The number of cases with midrange probabilities (30%–60%) is surprisingly low within this time frame, suggesting that the ensemble members are consistent with each other in their predictions of whether a TC will form or not. It remains an open question whether this jump in probabilities is related to an abrupt change in predictability, which may be associated with the model evolution from the initial conditions, and/or intrinsic atmospheric processes in which the mid–lower-tropospheric vorticity only begins to consolidate within 48 h of genesis (e.g., Fig. 5d of Lawton and Majumdar 2023). On the other hand, robust waves with steady high probabilities such as Gordon are in a minority.

The geographical distributions of IFS ensemble probabilities at 48, 72, 120, and 168 h illustrate some regional variations,

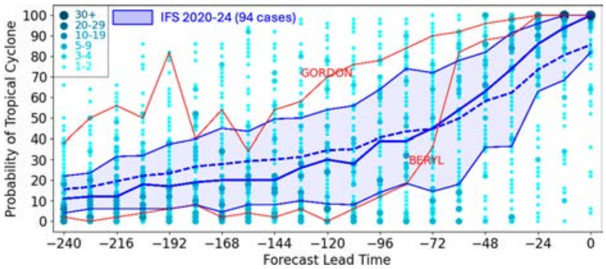

FIG. 3. Progression of IFS ensemble-based probabilities (with Vmax threshold = 25 kt) for all 94 TCs between 2020 and 2024, with 0 h corresponding to the time at which the cyclone became a named tropical storm. The larger, darker shaded dots represent more cases at a given probability. The dashed line illustrates the mean probability for each lead time among the 94 cases, the thin solid lines show the 25th- and 75th-percentile values, and the thick solid line shows the median value. Two cases, Beryl from Fig. 2a and Gordon from Fig. 2c, are shown for reference.

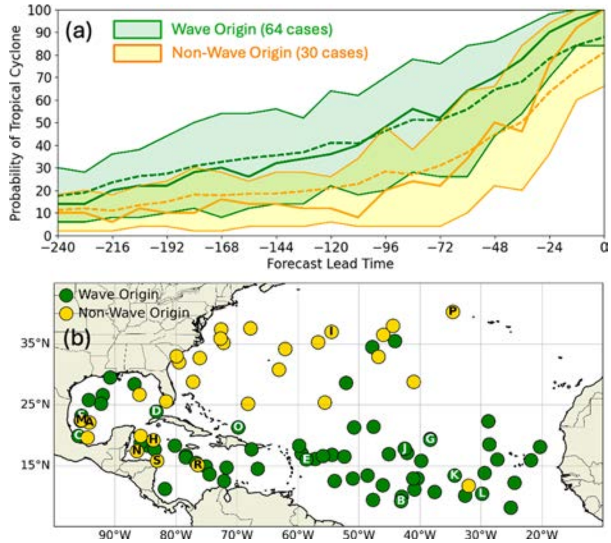

FIG. 5. (a) As in Fig. 3, divided into wave vs nonwave cases. (b) Locations of tropical storm declaration for wave and nonwave cases. The first letters of storms named in 2024 are indicated.

consistent with Richardson et al. (2025). At 48 h, most probabilities are high (≥60%) as expected (Fig. 4a). However, at 72 h, some probabilities have diminished, especially in the Caribbean and higher latitudes. At 120 h, the probabilities diminish a little further in these regions and noticeably equatorward of 15°N (including Beryl, Kirk, and Leslie). Many cases in these regions drop below 10% for 168-h forecasts, the range for which the NHC issues its Tropical Weather Outlooks. However, in the Bay of Campeche and main development region (especially the northern side), the probabilities often remain robust, even at 168 h. These results suggest that it is important to account for the regional climatology of the initial precursor disturbance when a genesis forecast is being made.

Next, we examine whether the probabilities are related to the type of precursor disturbance. The probabilities for cases of nonwave origin are generally lower than those of wave origin (Fig. 5a), with the 25th percentile of wave cases (lower thin green line) exhibiting similar probabilities to the 50th percentile of nonwave cases (thick orange line) at most lead times. Consistent with this, the range of probabilities for nonwave cases is narrower (with lower probabilities) at 5–10-day lead times than for wave cases. At 2-day lead times, the median probability of a TC for nonwaves is only 48%. In contrast, the median 2-day probability for waves is 70%, and the sharp increase in probabilities around this lead time is not so noticeable as for nonwaves. In the Atlantic Ocean south of 22°N, all but one case are of wave origin (Fig. 5b). In the Caribbean and Gulf regions, the origins are mixed, with several cases forming from waves and others from features such

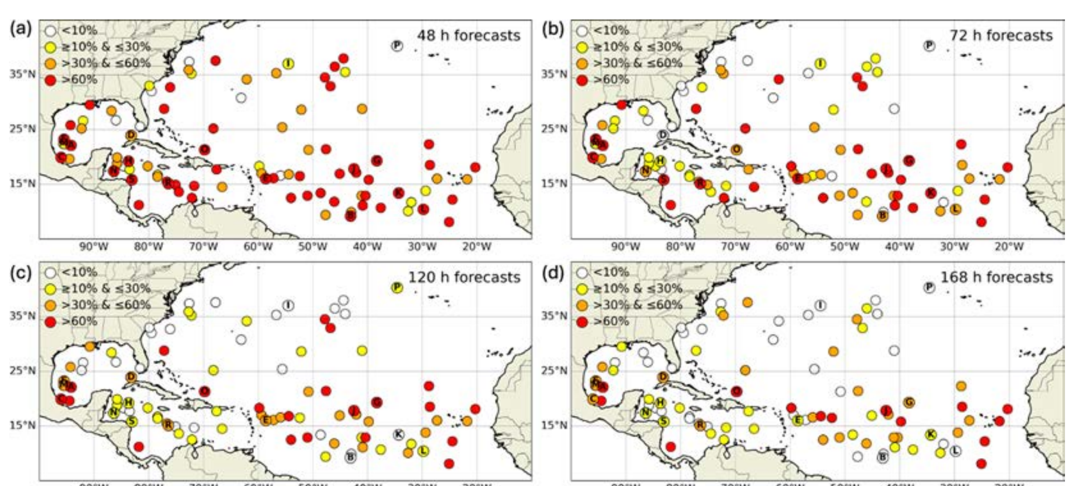

FIG. 4. IFS ensemble-based probabilities at (a) 48, (b) 72, (c) 120, and (d) 168 h for all 94 cases between 2020 and 2024, valid at the location and time at which each tropical storm was named. The first letters of storms named in 2024 are indicated (e.g., "B" for Beryl).

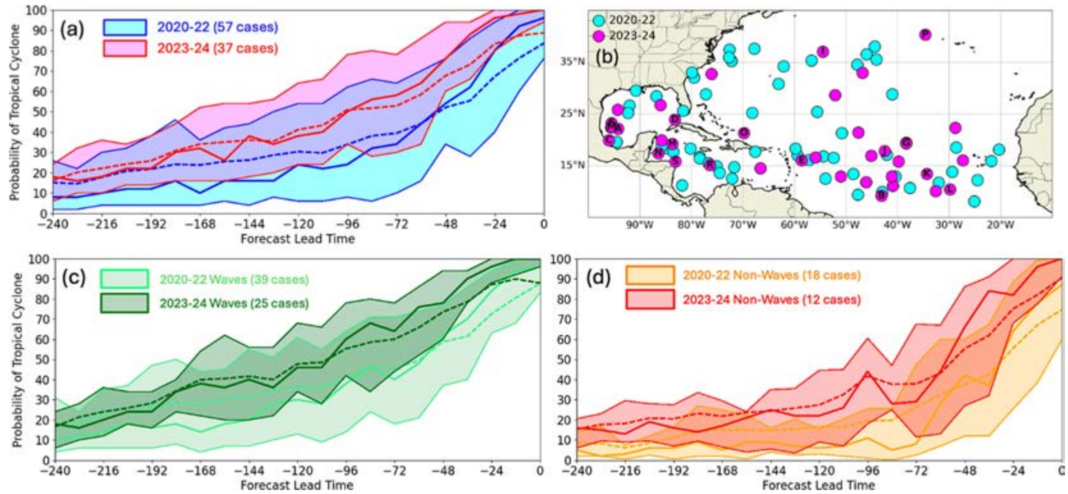

FIG. 6. (a) As in Fig. 3, divided into the 2020–22 and 2023–24 periods. (b) Locations of tropical storm declaration for 2020–22 (cyan) and 2023–24 (magenta). (c) As in (a), but for the subset of waves only. (d) As in (a), but for the subset of nonwaves only.

as the Central American Gyre. Figure 5b combined with Fig. 4 reveals that many low-probability cases of nonwave origin are situated just offshore of the eastern seaboard of the United States, over the north–central Atlantic, and in the extreme western Caribbean. We suggest that tropical cyclogenesis from troughs, gyres, and baroclinic systems in operational models requires additional investigation (following Wang et al. 2018).

b. Impact of updating cycle version

Prior to the first Atlantic TC in 2023, a major upgrade to the IFS ensemble (version 48r1) involved a reduction in the grid spacing from 18 to 9 km. While the 18-km (2020–22) and 9-km (2023–24) samples are distinct, so a direct evaluation of the resolution upgrade is not possible, the upgrade suggests a shift toward higher probabilities (Fig. 6a). The 25th percentile of the 9-km sample (lower thin red line) displays similar probabilities to the 50th percentile of the 18-km sample (thick blue line) at most times. There is also a decrease in the range of probabilities. Although the 2020–22 period had proportionally more cases of nonwave origin over the Atlantic Ocean north of 25°N (Fig. 6b), a similar increase in forecast probabilities is evident for subsamples comprising only waves (Fig. 6c) and only nonwaves (Fig. 6d).

We next examine the effects of the most recent IFS upgrade at the time of writing, from cycle 48r1 to 49r1. The experimental cycle that became 49r1 was run once per day (0000 UTC) for a limited period in 2024, covering 14 Atlantic TCs (Debby–Rafael). In contrast to the previous upgrade (from 47r1 to 48r1), the probabilities only increased marginally, although the 25th and 50th percentiles are slightly raised in 49r1 for the ~72-h (Fig. 7a) forecasts. To examine whether the SPP in 49r1 may have helped spin up more TCs for these cases, the differences between the probabilities are presented in Fig. 7b for 72–84-h forecasts. The three cases with the highest increase in probability are Nadine, Francine, and Ernesto. Each of these storms formed from complex broad areas of low pressure in which the NHC short-range genesis probabilities were low (NHC 2025), and the SPP may have contributed to greater convection. For the wave cases, the impact of the SPP was usually neutral (and negative for Gordon). Overall, we suggest that the introduction of SPP in 49r1 did not have as big an impact on raising the probabilities as the increase in resolution from 18 to 9 km in 48r1 (Fig. 6a).

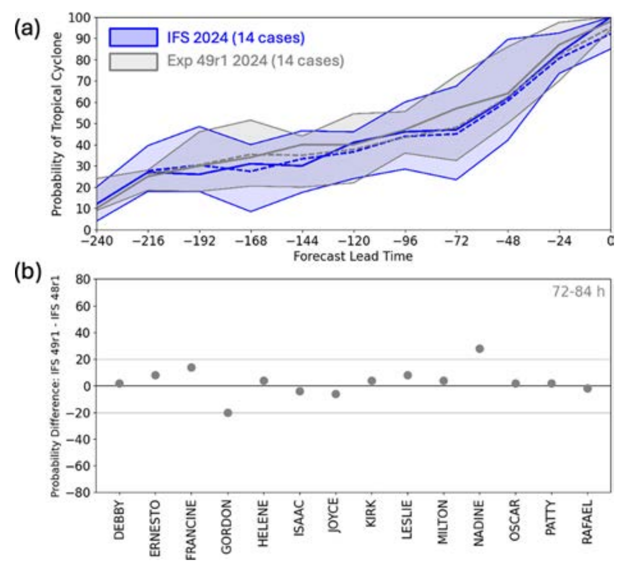

FIG. 7. (a) As in Fig. 3, but for the then-operational IFS cycle 48r1 and a daily parallel cycle that later became IFS cycle 49r1, for a homogeneous sample of 14 TCs in 2024. In this figure only, "0" refers to combined 0- and 12-h forecasts, "24" refers to 24- and 36-h forecasts, and so on. The genesis times are 0000 or 1200 UTC, consistent with Figs. 2–6. (b) Difference between experimental 49r1 and then-operational 48r1 probabilities for each individual TC in 2024, for 72–84-h ensemble forecasts.

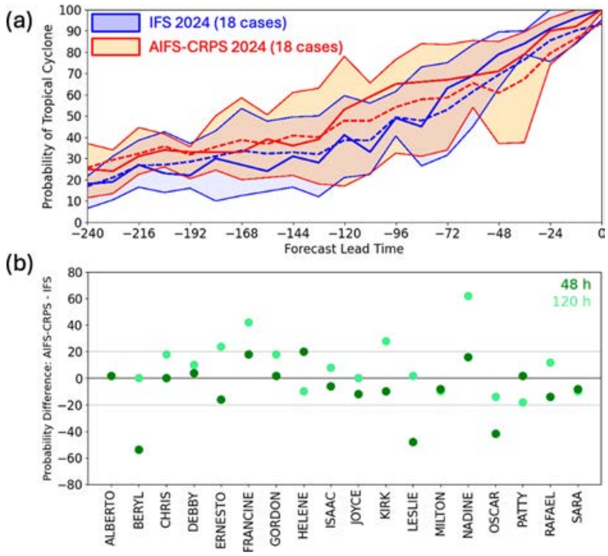

FIG. 8. (a) As in Fig. 3, but for the operational IFS ensemble and the parallel AIFS-CRPS ensemble, over a homogeneous sample of 18 TCs in 2024. (b) Difference between AIFS and IFS probabilities for each individual TC in 2024, for 48-h (dark green) and 120-h (light green) ensemble forecasts.

c. IFS and AIFS-CRPS ensemble-based probabilistic forecasts

The AIFS-CRPS ensemble was produced in experimental mode during 2024, covering all 18 Atlantic basin TCs. It exhibits a higher median and mean probability than the IFS ensemble for lead times exceeding 72 h and a lower median and mean probability for lead times shorter than 48 h (Fig. 8a). An examination of individual 120-h forecasts reveals four waves in which the AIFS-CRPS forecast probability exceeds its IFS counterpart by around 20%: Chris, Ernesto, Gordon, and Kirk, in addition to two cases that were not tracked waves (Francine and Nadine) (Fig. 8b). The 48-h probabilities indicate that AIFS-CRPS predicts a much lower probability (>40% difference) than the IFS ensemble for the Beryl, Leslie, and Oscar waves (Fig. 8b). These differences are also discernible in those cases included in Fig. 2, and we will revisit them in the context of wave strength in section 4.

To qualify these differences beyond the visual distinction, a nonparametric test for the paired differences is conducted using a standard Wilcoxon signed-rank test. The null hypothesis is that the differences between the IFS and AIFS-CRPS probabilities are zero for each lead time. The p values for 36- and 48-h forecasts are 0.14 and 0.18, respectively, indicating no clear significance in the differences, whereas they are 0.10 or below for 120–144-h forecasts, indicating that the AIFS-CRPS probabilities are significantly greater at the 90% level. We emphasize that the differences between the IFS and AIFS-CRPS probabilities need to be interpreted with caution, given the small sample size of genesis cases.

The geographical distributions of the 2024 cases in Fig. 9 reveal several similarities and differences between IFS and AIFS-CRPS. For the two higher-latitude cases (Isaac and Patty), both ensembles had low 48- and 120-h probabilities. In contrast, the 48-h probabilities for the cases west of 75°W were mostly high in both ensembles (except for Debby with its land interaction challenges and Milton which had high probabilities for the generous 20-kt Vmax threshold; Fig. 2g). The lowest-latitude waves, Beryl and Leslie, had only medium (30%–60%) 48-h probabilities in the AIFS ensemble, whereas all the waves in the IFS ensemble had high (>60%) 48-h probabilities. For the 120-h forecasts, all the cases where AIFS-CRPS is at least one probability category higher are equatorward of 25°N across the full tropical Atlantic basin.

We close this section by summarizing the average 2-, 5-, and 8-day probabilities for each of the years 2020–24 (Fig. 10). The corresponding lines for the intervening forecast lead times largely lie in sequence and are not shown. First, the IFS probabilities in 2020 were lower than in the later years. The average

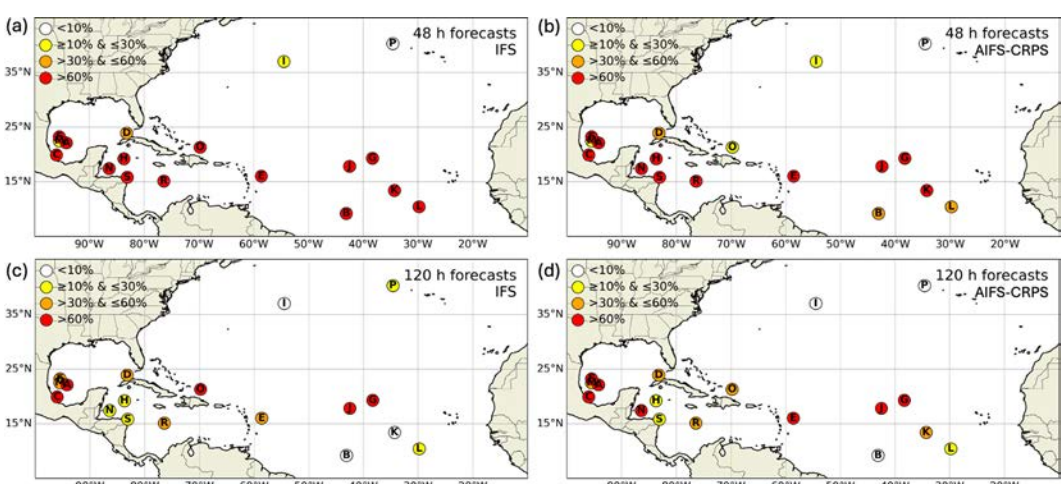

FIG. 9. The 48-h ensemble-based probabilities for (a) IFS and (b) AIFS-CRPS for each 2024 case, at the location at which each tropical storm was named. The 120-h probabilities for (c) IFS and (d) AIFS.

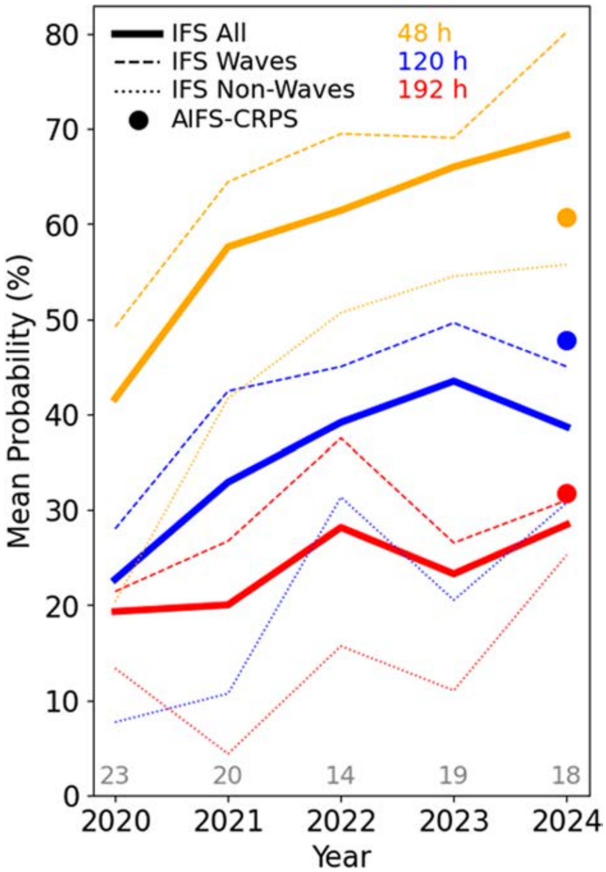

FIG. 10. Mean ensemble-based probabilities by year, for 48-, 120-, and 192-h forecasts. Thick lines: IFS average for all cases. Dashed lines: IFS average for wave cases. Dotted lines: IFS average for nonwave cases. Large dots: AIFS-CRPS average (2024 only). The sample size for each year is listed in gray.

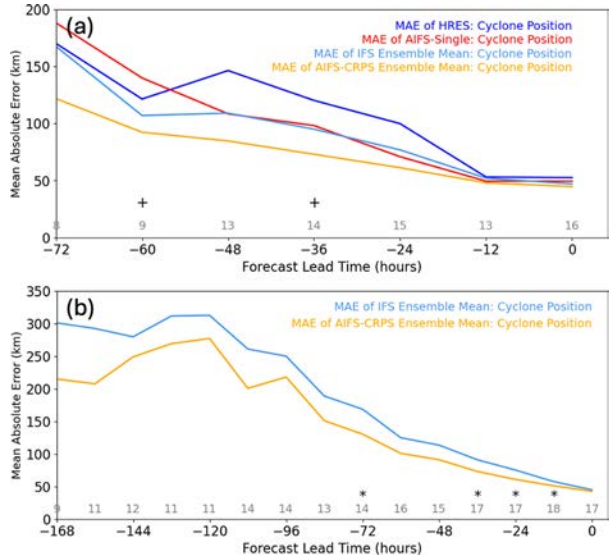

FIG. 11. MAE (km) of position valid at the time the tropical storm is named, for the 2024 season, for (a) a homogeneous sample of HRES, AIFS-Single, IFS ensemble mean, and AIFS-CRPS ensemble mean and (b) a homogeneous sample of the IFS and AIFS-CRPS ensemble means only. In (a), + symbols are marked at the times when the MAE of the AIFS-CRPS ensemble mean is statistically superior to the MAE of AIFS-Single. In (b), * symbols are marked at the times when the MAE of the AIFS-CRPS ensemble mean is statistically superior to the MAE of the IFS ensemble mean.

probabilities mostly increased progressively between 2020 and 2024. Although an increase was evident in Fig. 6a between 2022 and 2023, when the resolution was changed from 18 to 9 km, it was not markedly higher than between other consecutive years. The probabilities for the wave cases (dashed lines) are generally increasing year to year and are consistently much higher than the probabilities for the nonwave cases (dotted lines) in each year and for each lead time. Finally, the average AIFS-CRPS probabilities in 2024 (large dots) are lower than IFS for 2-day forecasts but higher for 5-day forecasts and beyond.

d. *Forecasts of genesis position in the IFS and AIFS*

The forecast errors of the cyclone position valid at the genesis time are computed for the HRES, IFS ensemble mean, AIFS-Single, and AIFS-CRPS ensemble-mean forecasts, for the 18 cases in 2024. The ensemble mean is only computed if at least 20% (10 of the 50 members) of the ensemble members contain a ≥25-kt TC within 1000 km of the best track position. For a homogeneous sample of all four forecasts, this criterion needs to be met for both ensemble means, and the TC also needs to have formed in the HRES and AIFS-Single forecasts for a position forecast to exist. The sample size drops to just 6 for forecasts beyond 72 h, so we only include forecasts out to 72 h before genesis (Fig. 11a). The mean absolute error (MAE) of the IFS ensemble mean is lower than that of the single IFS forecast (HRES), and the MAE of the AIFS-CRPS ensemble mean is lower than that of AIFS-Single. A one-sided *t* test reveals that the MAE of the AIFS-CRPS ensemble mean is statistically significantly lower than that of AIFS-Single at the 95% level for 36- and 60-h forecasts. Given that the sample is small largely due to HRES and/or AIFS-Single not predicting the existence of a TC, the same evaluation is conducted using a larger homogeneous sample containing only the ensemble means (Fig. 11b). For all lead times out to 168 h, the MAE of the AIFS-CRPS ensemble mean is lower than that of the IFS ensemble mean, with statistical significance at the 95% level for 12-, 24-, 36-, and 72-h forecasts. These results indicate the potential utility of the AIFS to predict genesis positions.

One caveat of this evaluation is that TCs at the time of genesis are difficult to track in models and in the NHC best track, given that the low pressure systems are weak and very broad, with substantial uncertainty in the central position at the time of genesis. Landsea and Franklin (2013) reflect this difficulty in the best track uncertainty estimates, with position uncertainty estimates of 68 km for tropical storms without aircraft data, and it is also evident in the MAE of the ECMWF analysis positions evaluated against the final best track.

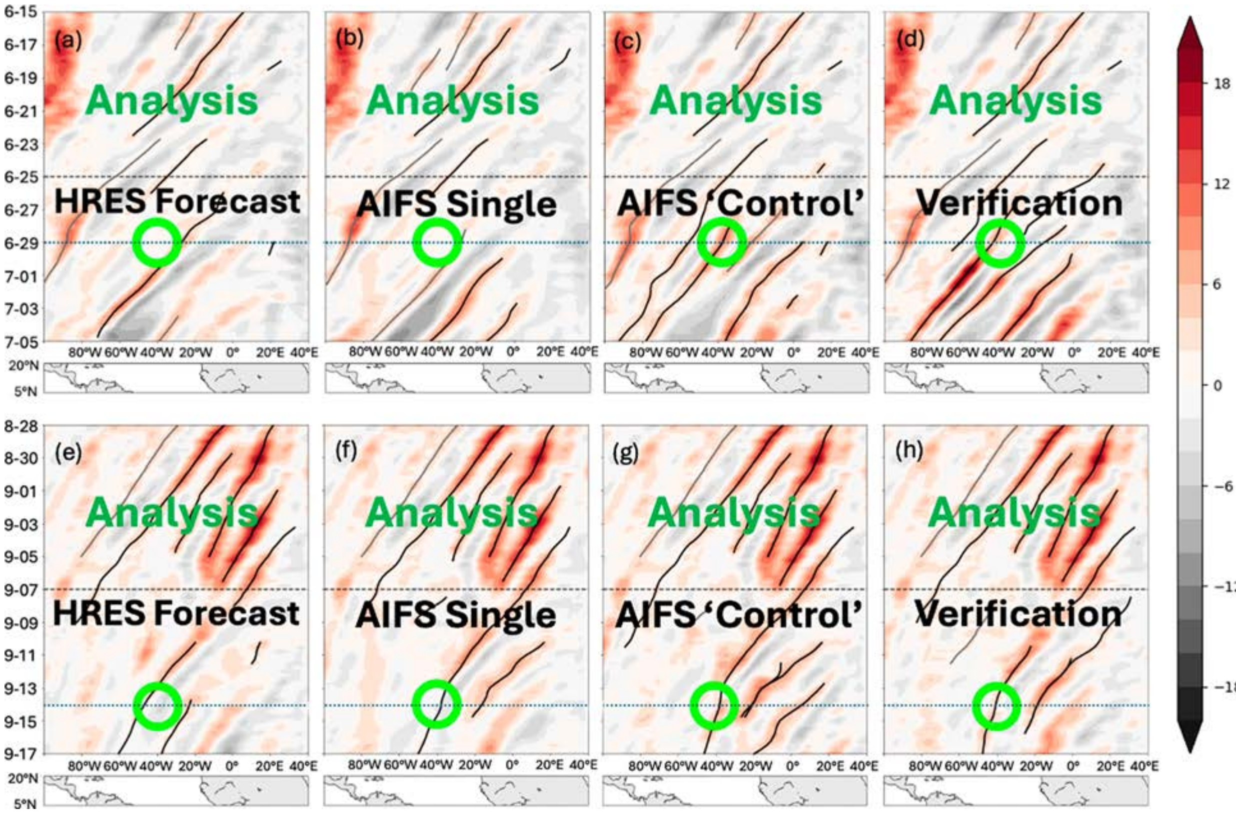

FIG. 12. Hovmöller diagrams using the Lawton et al. (2022) tracker. Pre-Beryl (2024), for (a) HRES, (b) AIFS-Single, (c) AIFS-CRPS control forecasts initialized at 0000 UTC 25 Jun 2024, and (d) the corresponding ECMWF analysis. Pre-Gordon (2024) for (e) HRES, (f) AIFS-Single, (g) AIFS-CRPS control forecasts initialized at 0000 UTC 7 Sep 2024, and (h) the corresponding ECMWF analysis. The month and day at 0000 UTC time are on the *y* axis. The initial time is indicated by the horizontal dashed line, and the time at which the cyclone is named is indicated by the horizontal dotted line below it. The time and longitude at which each storm is named are at the center within the green circle. The shading illustrates the average 700-hPa curvature vorticity between 5° and 20°N. The black lines represent tracked waves of African origin, and the gray lines indicate waves that could not be tracked back to Africa. The geography and relevant latitude band are included at the bottom for reference.

An investigation of the corresponding intensity (Vmax) forecasts reveals that the mean errors in Vmax are biased low by 5–15 kt (not shown). This is expected, due to the limited ability of the models, data assimilation schemes, and training datasets to resolve convective-scale processes (consistent with Lang et al. 2024a). Additionally, since genesis is focused on the minimum of the TC intensity range, our sample, which is restricted to TCs with Vmax ≥ 25 kt, is necessarily biased. This is in contrast to standard evaluations of intensity predictions of mature TCs, in which the predicted and verified range of intensities is far greater, and the sample is much less compromised by forecast cases in which the TC does not exist at all.

4. Results: Tracking African easterly waves in the IFS and AIFS

We now examine the characteristics of the tracked waves up to the time the tropical storm is named. One objective is to evaluate AIFS-Single (AIFS) in comparison with the unperturbed IFS single forecast (HRES), for a homogeneous sample of 18 cases during 2023–24. Another objective is to provide context for the probabilistic forecasts from the IFS and AIFS-CRPS ensembles shown in Figs. 2, 8, and 9, using the corresponding wave forecasts from 2024. We use AIFC, the control member of the AIFS-CRPS ensemble, to illustrate an ensemble member's behavior. Following an evaluation of the wave track forecasts, we assess the wave-relative environmental variables.

a. Wave tracks and coordinates

The time–longitude Hovmöller diagrams in the top row of Fig. 12 demonstrate challenges in the 4-day predictions of the wave that formed into Beryl (2024). Both HRES and AIFS missed the wave (Figs. 12a,b), although a wave was predicted behind it. In contrast, AIFC caught the wave (Fig. 12c) at the time Beryl formed, similar to the wave tracked in the ECMWF analyses (Fig. 12d). This case illustrates the complex scenarios that can occur with weak and/or interacting waves. For the 3-day forecasts initialized 1 day later, all of HRES, AIFS, and AIFC were able to capture the pre-Beryl wave (not shown). In contrast to Beryl, the HRES and/or AIFS forecasts provided trackable waves out to 8 days for 13 of the other 17 wave cases in 2023–24. An example is Gordon, whose 7-day forecast is captured in each of HRES, AIFS, and AIFC (Figs. 12e–h). Minor differences in the wave propagation speed are evident, with HRES being slightly faster than the analysis, whereas AIFS and AIFC are more accurate. AIFS and AIFC also caught the nondeveloping wave behind pre-Gordon. Overall, a qualitative investigation of Hovmöller diagrams for all cases does not reveal that any one of HRES, AIFS, or AIFC is obviously superior in capturing the wave propagation and genesis.

We use the information gained from the Hovmöller diagrams (and maps of the tracks, not shown) to evaluate the forecast errors for the 18 waves in 2023–24. We take a homogeneous sample of integer-day forecasts (up to 10 days) initialized at 0000 UTC each day, for which the correct wave is successfully tracked in HRES and AIFS (and AIFC in 2024). For each wave, we impose a longitudinal limit of ±6° and a latitudinal limit of ±5°, excluding wave predictions that are outside these limits. Based on an examination of each wave, there is no ambiguity within this limit, whereas separate waves are identified outside this limit in a few instances. Our imposed limitation restricts our ability to evaluate forecasts beyond 8 days, where there can be significant ambiguity and one case can contaminate the position error statistics. We also note that neither system consistently identifies and tracks the wave for longer than the other.

The HRES mean errors in latitude are consistently negative at all lead times, up to about 1°, indicating that HRES is too far to the south on average (Fig. 13a). In contrast, the AIFS mean errors in latitude are closer to zero for 3- and 5–8-day forecasts. For longitude, the HRES (AIFS) exhibits a clear positive, slow bias for 3–6 (4–7)-day forecasts, consistent with Elless and Torn (2018). The MAE of latitude is lower in AIFS than HRES by ~0.5° for 5–8-day forecasts, and the MAE of longitude is also lower by ~0.5° in AIFS for 4- and 6–8-day forecasts (Fig. 13b). Given the small sample and variability across cases, we are cautious about statistical significance. A metric that is less influenced by outliers in a small sample is the frequency of superior performance (FSP; Magnusson et al. 2025). The 5–8-day AIFS forecasts of latitude were superior to HRES in 60%–80% of the cases (Fig. 13c). An examination of the distributions of position errors reveals characteristics that are consistent with the mean errors, with an additional finding that about a quarter of the 5–8-day HRES forecasts have negative (fast) longitudinal errors of higher magnitude than 2° (e.g., Gordon in Fig. 12e), although the fraction of HRES (and AIFS) forecasts with positive (slow) errors exceeding 2° is greater (not shown). Restricting the sample to 2024, the MAEs of latitude and longitude in AIFC are comparable to those of HRES (not shown).

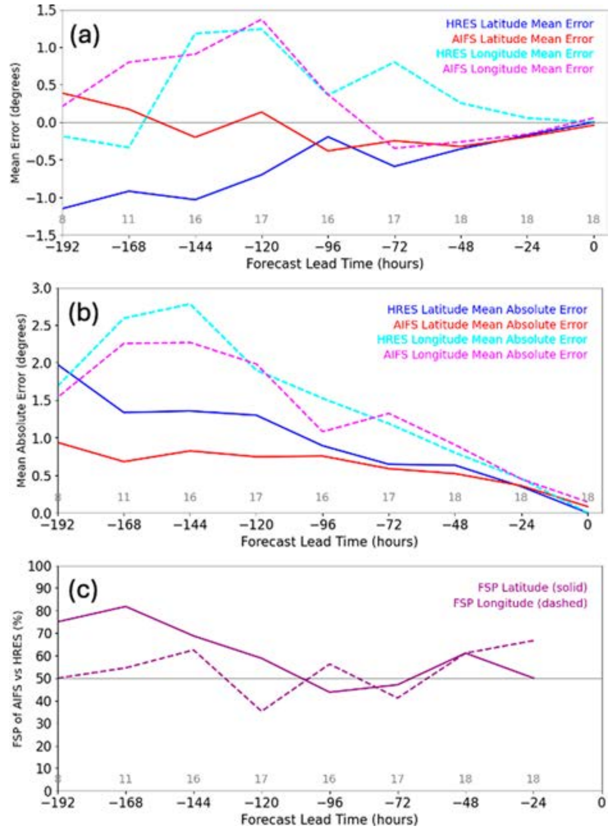

FIG. 13. (a) Mean forecast errors of wave latitude and longitude (°) for 0–8-day forecasts, for HRES and AIFS-Single in 2023–24. The mean error at a given lead time is defined as the average over all cases of forecast–analysis. A negative (positive) latitudinal error corresponds to the forecast wave being too far south (north). A negative (positive) longitudinal error corresponds to the forecast wave being too fast (slow). (b) MAE. (c) FSP for AIFS-Single over HRES, for wave latitude and longitude. The 0-h FSP is not useful and is omitted.

b. *Wave-relative variables*

Finally, we examine the forecast characteristics of the wave-relative environmental variables. To examine whether a relationship exists between the IFS and AIFS-CRPS ensemble probability differences (Figs. 2, 8, and 9) and the forecasted wave strength, we use HRES and AIFC forecasts of wave-relative 850-hPa vorticity averaged within a disk of radius of 600 km for the nine trackable waves in 2024 (Fig. 14). At 0 h, the range of vorticity analysis values at the time the storm was named is evident, with Oscar, Leslie, and Beryl being the three weakest waves. Each of these waves was a small storm at genesis (NHC 2025). These are also the three waves for which the 48-h AIFS-CRPS probabilities were substantially lower than the corresponding IFS probabilities (Fig. 8b). Given that the resolution is lower in AIFS-CRPS (28 km) than the IFS ensemble (9 km), AIFS-CRPS may have struggled to capture these small storms. Additionally, the IFS may be better able to physically spin up TCs from weak waves within 48 h. Oscar is an outlier case that had nearly zero environmental vorticity in the analysis and short-range forecasts. For the 48-h forecasts of Beryl and Leslie, the HRES values are higher than AIFC and closer to the verifying analysis values. Another time of interest in section 3 is 120 h, where the AIFS-CRPS probabilities were ~20% higher than those in the IFS for the waves that developed into Chris, Ernesto,

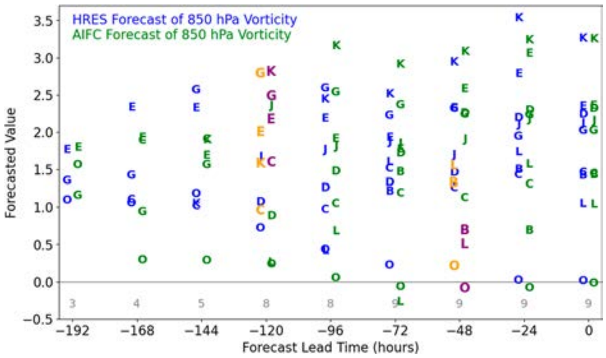

FIG. 14. Distributions of HRES (blue) and AIFC (green) forecast values of wave-relative 850-hPa vorticity (10^{-4} s^{-1}) for each of the nine trackable wave cases in 2024. The letter refers to the first letter of the storm name. The cases at 48 and 120 h that are discussed in the text are labeled in orange (HRES) and purple (AIFC) for easy reference. The number of cases at each lead time is listed in gray. A 120-h HRES forecast value of -1.3 for Leslie is off the chart.

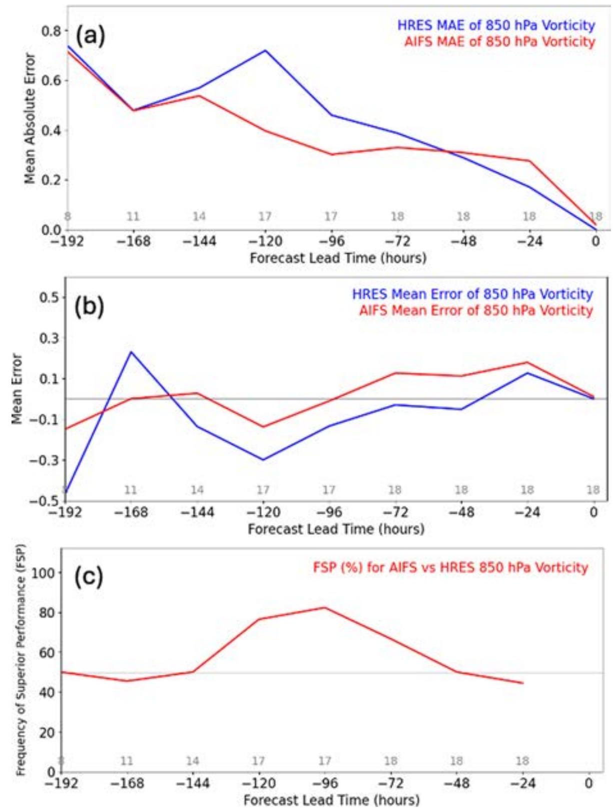

FIG. 15. (a) MAE of wave-relative 850-hPa environmental vorticity (10^{-4} s^{-1}) for 0–8-day HRES and AIFS forecasts of all 18 cases of wave origin during 2023–24. (b) Mean error. (c) FSP of AIFS over HRES. The sample size at each lead time is in gray.

Gordon, and Kirk. The 120-h forecast values of 850-hPa vorticity in AIFC are higher than HRES for Chris, Ernesto, and Kirk. The latter two storms also possessed the highest vorticity out of the nine cases at the times they were named. For Gordon, the 120-h HRES vorticity forecast was overly strong. While the AIFC forecast vorticity was weaker than HRES, it was still stronger than the analysis value. In all four cases, the 120-h AIFC forecast value was closer to the verifying analysis than the HRES value. Given these results, we suggest that the AIFS-CRPS ensemble members may struggle to develop tropical cyclones from weak waves at 48 h compared with the IFS, whereas the converse is true for stronger waves at 120 h. While AIFS-Single is not shown in Fig. 14 because it is not a member of the IFS or AIFS-CRPS ensembles, its distribution of 850-hPa vorticity values is comparable to HRES and AIFC, with no distinct or surprising differences.

Turning our attention to how AIFS-Single compares against HRES for wave-relative fields for the 18 wave cases in 2023–24, the MAE of 850-hPa vorticity is lower in AIFS between 3 and 5 days and higher in AIFS at 1 day (Fig. 15a). The 120-h HRES mean errors of up to -0.3×10^{-4} s^{-1} demonstrate that HRES underpredicts vorticity on average, whereas AIFS, by dint of its training on analysis and reanalysis fields that capture synoptic-scale fields well, exhibits a lower mean error within $\pm 0.1 \times 10^{-4}$ s^{-1} (Fig. 15b). The FSP shows superior AIFS forecasts for 60%–80% of the cases in the 3–5-day time range.

For the weakest (Beryl) and strongest (Kirk) waves, the model differences in the physical wind fields are apparent. For Beryl, the 48-h HRES forecast produces a weak closed vortex, similar to that in the verifying analysis, whereas AIFS and AIFC fail to do so (Figs. 16a–d). In contrast, AIFC, and to a lesser extent AIFS, develops a TC-like vortex for Kirk at 120 h, whereas HRES does not (Figs. 16e–h). We note that the AIFS can predict realistically complex flows in the wave-relative framework, such as a double vortex for Margot (2023) that was consistent with the analysis (not shown).

For wave-relative 850-hPa specific humidity, the MAE reveals lower errors in AIFS at all lead times, especially 2–4 days (Fig. 17a). The mean error in AIFS is again small and neutral relative to HRES (Fig. 17b), suggesting that the lower troposphere in HRES is too moist (by 0.4–0.6 g kg^{-1} for 2–4 days). AIFS performs better in 60%–80% of the cases across most lead times (Fig. 17c). This discrepancy is exemplified in the Leslie case, where the wave-relative moisture in HRES is higher than in AIFS and AIFC (Fig. 18). AIFS produces a smooth field, as expected from its training algorithm (Lang et al. 2024a), whereas the AIFS-CRPS ensemble (of which AIFC is a member) is designed to permit sharper gradients (Lang et al. 2024b). The specific humidity field in AIFC compares well to the analysis.

The question of whether the differences in the IFS and AIFS-CRPS ensemble-based probabilities exhibit a relationship with the environmental moisture values was also investigated. The equivalent of Fig. 14 for moisture (not shown) mostly exhibited larger forecast values for HRES than AIFC for the 2024 waves of interest, regardless of their probability or strength. Hence, unlike for vorticity, we do not suggest that the characteristics of the HRES versus AIFC moisture forecasts can be connected with low or high probabilities in their respective ensembles. One observation worth noting is that

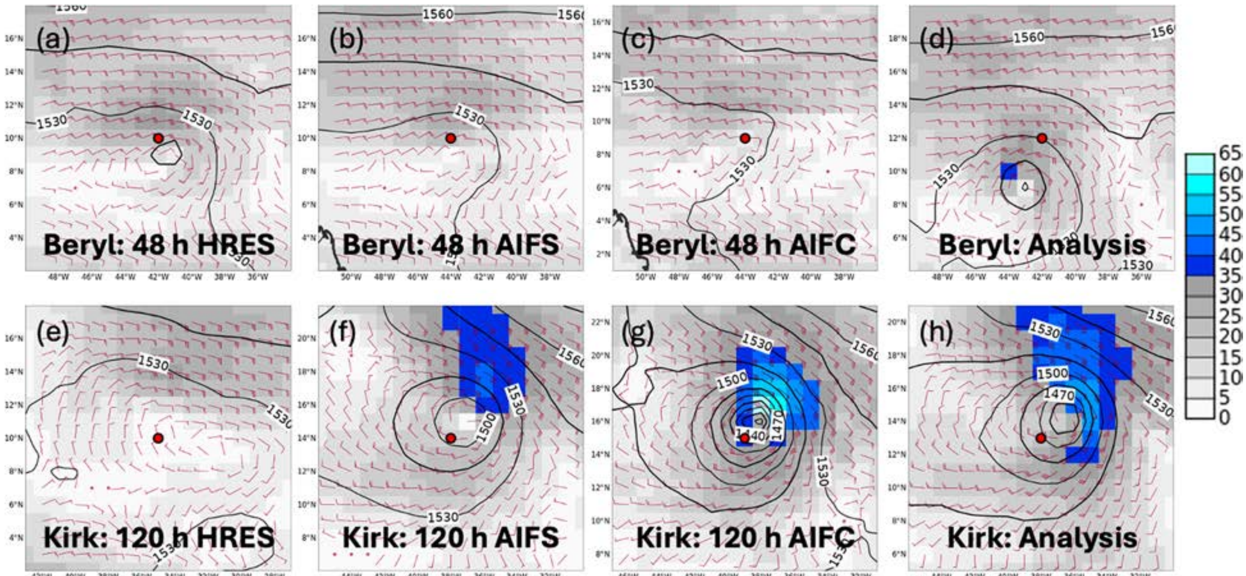

FIG. 16. The 850-hPa wind barbs (kt), wind speed (kt; shaded), and geopotential height (m; contour) for (a)–(d) 48-h forecasts and the corresponding analysis of pre-Beryl and (e)–(h) 120-h forecasts and the corresponding analysis of pre-Kirk. The red dot indicates the wave location (not the TC location).

the two driest environments at the analysis time were for Beryl and Leslie, and the moister environments in the analysis and both HRES and AIFC forecasts were for Chris, Ernesto, Gordon, and Kirk.

We investigated other wave-relative variables, and the comparisons generally displayed similar error and physical characteristics in the 700- and 500-hPa vorticity and specific humidity as in Figs. 15–18. The results were neutral for 200-hPa divergence and 250–850-hPa vertical wind shear, with similar results for the physics-based and AI-based systems.

5. Concluding remarks

Significant recent developments in the ECMWF forecasting suite require users to be aware of the characteristics and associated sensitivities for their applications of interest. The application in this paper is tropical cyclone formation, using ensemble-based probabilistic forecasts and tracked easterly waves in single forecasts. The introduction of the AIFS accentuates the need to examine its characteristics and benchmark it against the IFS.

In the IFS ensemble, we found considerable variability in the probabilities across 94 Atlantic basin TCs in 2020–24, which is expected given the variability in structures of precursor disturbances. The probability values exhibited high sensitivity to the threshold value of the maximum wind speed. A fairly abrupt transition from low probabilities at 72-h lead time (prior to the tropical storm being named) to high probabilities at 48 h was also identified. The cases with the lowest probabilities corresponded to disturbances that included broad lows and troughs, midlatitude systems, weak waves in the deep tropics, and occasionally waves that were interacting with another disturbance.

The average IFS probabilities have largely increased steadily each year during 2020–24, especially on the lower end (0%–20%). The change in IFS ensemble resolution from 18 to 9 km in cycle 48r1 exhibited a shift toward higher probabilities, although a direct attribution to the improved resolution is not proven. While the subsequent upgrade to IFS cycle 49r1 did not exhibit clear differences in the probabilities for the wave cases in 2024, the probabilities were increased in three cases of genesis from broad lows and troughs, possibly indicating that the SPP may serve to increase the convection in these systems.

The AIFS-CRPS ensemble-based probabilities of genesis were examined for the first time and were competitive with their IFS counterparts. AIFS-CRPS produced some higher probabilities than IFS, especially for 84–120-h forecasts, and were likely connected to higher AIFS-CRPS predictions of environmental vorticity. For 36–48-h forecasts, the opposite result was found. AIFS-CRPS produced lower probabilities in a few weak waves, with the IFS better able to spin up these waves. AIFS-CRPS also suffered from similar "dropouts" as the IFS ensemble in some short-range forecasts, suggesting a strong dependence on the initial conditions. Overall, AIFS-CRPS serves as a useful complement to the IFS ensemble, adding value in some cases, but showing deficiencies especially at short lead times for weak waves.

An examination of position errors for the genesis cases in 2024 revealed that the average AIFS-CRPS ensemble-mean position forecast error was often lower than that of AIFS-Single and the IFS ensemble mean, for forecasts valid at the genesis time. This initial result, while for a small sample, and recognizing that the position errors in the best track at the genesis time, opens new avenues for future evaluations of tropical cyclone track, intensity, and size parameters using single and ensemble forecasts initialized before the genesis time.

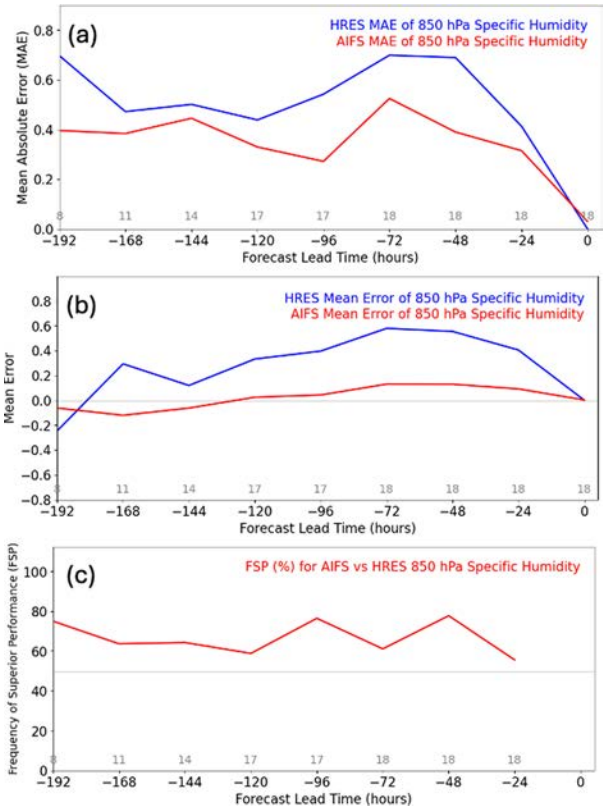

FIG. 17. (a) MAE of wave-relative 850-hPa specific humidity (g kg^{-1}) for 0–8-day HRES and AIFS forecasts of all 18 cases of wave origin during 2023–24. (b) Mean error. (c) FSP of AIFS over HRES. The sample size at each lead time is in gray.

Most easterly waves equatorward of 25°N that developed into TCs were trackable using QTrack (Lawton et al. 2022) for lead times out to ~8 days in the HRES and AIFS-Single systems. A few weak waves were difficult to track beyond a few days. A southward bias in the wave latitude was evident in the IFS forecasts, compared with the analysis values. AIFS-Single did not exhibit a latitudinal bias. A slow longitudinal bias was evident in both systems. Generally, the AIFS wave locations were more consistent with the verifying analysis than the IFS. For the wave-relative variables, AIFS-Single exhibited lower errors than the IFS for 3–5-day forecasts of 850–500-hPa environmental vorticity and specific humidity, whereas the IFS did not clearly demonstrate superiority over AIFS-Single for any variables. The AIFS-Single 3–5-day forecasts yielded stronger waves than the IFS. The IFS was also too moist, whereas the mean errors in the humidity and vorticity were generally small and neutral in AIFS-Single.

Challenges remain in tropical wave and TC genesis forecasting, stemming from the disorganized disturbances, multiscale processes, and the role of convection in the maintenance and propagation of the waves. For weaker waves, our results may depend on the choice of variable (area-averaged 700-hPa curvature vorticity) and tracking methodology. Since its implementation at ECMWF, QTrack has been upgraded to include waves that form west of ~40°W, allowing for weaker waves that form or reintensify in the Caribbean (Lawton 2024; Lawton et al. 2025). Additionally, a new deep learning method developed by Downs et al. (2025), which is now used by NHC's Tropical Analysis and Forecast Branch, shows promise for tracking weak developing and importantly nondeveloping waves westward of 50°W, where conventional trackers often lose the wave. Future applications of wave trackers to include ensemble forecasts will be a new avenue for advanced, probabilistic wave forecasting.

The physical properties of the TC genesis process will need to be examined further in the evolving IFS and AIFS, together with the roles of observations and data assimilation, following Magnusson et al. (2025), which was limited to mature tropical cyclones. Our identification of some cases of very low probability, even < 3 days prior to the tropical storm being named, motivates the need to understand physically what distinguishes these low-probability cases that develop into TCs from low-probability cases that do not develop. We expect that convective-scale processes may be more prevalent, which provokes new questions about stochasticity and the limits of predictability for these types of cases.

The software and diagnostics developed for this paper can easily be applied to further models and version upgrades. Additional priorities include the establishment of a probabilistic verification framework for developing and nondeveloping waves, following Majumdar and Torn (2014) and Richardson et al. (2025), and regular evaluations of jumpiness following the latter paper. Our paper also offers a starting point to diagnose and evaluate additional wave-relative fields including stratiform and convective precipitation and other process-oriented variables such as heating tendencies (e.g., Majumdar et al. 2023).

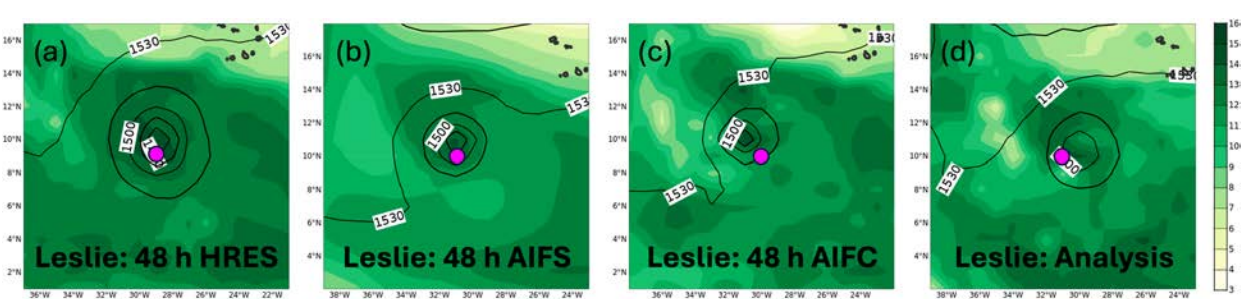

FIG. 18. The 850-hPa specific humidity (g kg^{-1}; shaded) and geopotential height (m; contour) for 48-h forecasts and the corresponding analysis of pre-Leslie. The magenta dot indicates the tracked wave location.

We close by suggesting several avenues for improving ensemble predictions of genesis. It is worth examining the reliability of the ensembles. If the ensembles are underdispersive, new perturbation methods may be beneficial to create a broader spread of solutions. The sensitivity to initial conditions also motivates the question of how to improve the initialization of disorganized clusters of thunderstorms, via conventional or AI-based techniques. The encouraging early performance of the AIFS for TC genesis sets a challenge for developers to yield additional improvements to machine learning–based systems. Potential advances include increasing the resolution, incorporating convective scales into the training, improving scale-aware loss objectives (e.g., Lang et al. 2025), and increasing the number of ensemble members. Finally, we suggest that each new version of any ensemble requires a recalibration of the threshold values before a quantitatively reliable probabilistic forecast can be issued.

Acknowledgments. Sharanya J. Majumdar gratefully acknowledges the University of Miami and ECMWF for a sabbatical year and subsequent visits during which this study was performed, support from NSF Grants AGS-1747781 and AGS-2438140, and discussions at NHC, Colorado State, and NCAR. Quinton A. Lawton acknowledges support from NSF Grant AGS-1747781, an NSF Graduate Research Fellowship (1938060), and an NSF NCAR Advanced Study Program Postdoctoral Fellowship. Part of this material is based upon work supported by the NSF NCAR, which is a major facility sponsored by the NSF under Cooperative Agreement 1852977. We thank two anonymous reviewers for their comments and suggestions.

Data availability statement. The easterly wave tracks were generated using the QTrack Python module v0.0.1, first developed by Lawton et al. (2022), which is available on GitHub (Lawton 2024; https://github.com/qlawton/QTrack). The ECMWF data are available from ECMWF, and software is available from the first author upon request. The tropical cyclone best track data are available in NOAA's International Best Track Archive for Climate Stewardship (IBTrACS) archive at https://doi.org/10.25921/82ty-9e16 (Knapp et al. 2010, 2018).

REFERENCES

Ben Bouallègue, Z., and Coauthors, 2024: The rise of data-driven weather forecasting: A first statistical assessment of machine learning–based weather forecasts in an operational-like context. *Bull. Amer. Meteor. Soc.*, **105**, E864–E883, https://doi.org/10.1175/BAMS-D-23-0162.1.

Brammer, A., and C. D. Thorncroft, 2015: Variability and evolution of African easterly wave structures and their relationship with tropical cyclogenesis over the eastern Atlantic. *Mon. Wea. Rev.*, **143**, 4975–4995, https://doi.org/10.1175/MWR-D-15-0106.1.

——, ——, 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.

Downs, W. B., S. J. Majumdar, and A. D. Mahoney, 2025: Using deep learning to identify tropical easterly waves, the intertropical convergence zone, and the monsoon trough. *Mon. Wea. Rev.*, **153**, 1581–1604, https://doi.org/10.1175/MWR-D-24-0195.1.

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, 2025: Implementation of AIFS ENS v1. ECMWF, accessed 11 May 2026, https://confluence.ecmwf.int/display/FCST/Implementation+of+AIFS+ENS+v1.

Elless, T. J., and R. D. Torn, 2018: African easterly wave forecast verification and its relation to convective errors within the ECMWF ensemble prediction system. *Wea. Forecasting*, **33**, 461–477, https://doi.org/10.1175/WAF-D-17-0130.1.

Fritz, C., Z. Wang, S. W. Nesbitt, and T. J. Dunkerton, 2016: Vertical structure and contribution of different types of precipitation during Atlantic tropical cyclone formation as revealed by TRMM PR. *Geophys. Res. Lett.*, **43**, 894–901, https://doi.org/10.1002/2015GL067122.

Haiden, T., and Coauthors, 2025: Evaluation of ECMWF forecasts. ECMWF Tech. Memo. 931, 68 pp., https://doi.org/10.21957/51e665e5c8.

Halperin, D. J., and R. E. Hart, 2025: Global verification of tropical cyclone genesis forecasts from five global numerical models. *Wea. Forecasting*, **40**, 1125–1145, https://doi.org/10.1175/WAF-D-24-0130.1.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

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.

Isaksen, L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud, 2010: Ensemble of data assimilations at ECMWF. ECMWF Tech. Memo. 636, 48 pp., https://doi.org/10.21957/obke4k60.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone data. *Bull. Amer. Meteor. Soc.*, **91**, 363–376, https://doi.org/10.1175/2009BAMS2755.1.

——, H. J. Diamond, M. C. Kossin, and C. J. Schreck, 2018: International Best Track Archive for Climate Stewardship (IBTrACS) project, version 4. NOAA National Centers for Environmental Information, accessed 11 May 2026, https://doi.org/10.25921/82ty-9e16.

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.

Lang, S. T. K., S.-J. Lock, M. Leutbecher, P. Bechtold, and R. M. Forbes, 2021a: Revision of the stochastically perturbed parametrisations model uncertainty scheme in the Integrated Forecasting System. *Quart. J. Roy. Meteor. Soc.*, **147**, 1364–1381, https://doi.org/10.1002/qj.3978.

——, and Coauthors, 2021b: More accuracy with less precision. *Quart. J. Roy. Meteor. Soc.*, **147**, 4358–4370, https://doi.org/10.1002/qj.4181.

——, M. Rodwell, and D. Schepers, 2023: IFS upgrade brings many improvements and unifies medium-range resolutions. *ECMWF Newsletter*, No. 176, ECMWF, Reading, United Kingdom, 21–28, https://doi.org/10.21957/slk503fs2i.

——, and Coauthors, 2024a: AIFS—ECMWF's data-driven forecasting system. arXiv, 2406.01465v2, https://doi.org/10.48550/arXiv.2406.01465.

——, and Coauthors, 2024b: AIFS-CRPS: Ensemble forecasting using a model trained with a loss function based on the continuous ranked probability score. arXiv, 2412.15832v1, https://doi.org/10.48550/arXiv.2412.15832.

——, M. Leutbecher, and P. Maciel, 2025: A multi-scale loss formulation for learning a probabilistic model with proper score optimisation. arXiv, 2506.10868v1, https://doi.org/10.48550/arXiv.2506.10868.

Lawton, Q. A., 2024: QTrack (version v0.0.1): An African easterly wave tracker. GitHub, accessed 11 May 2026, https://github.com/qlawton/QTrack.

——, and S. J. Majumdar, 2023: Convectively coupled Kelvin waves and tropical cyclogenesis: Connections through convection and moisture. *Mon. Wea. Rev.*, **151**, 1647–1666, https://doi.org/10.1175/MWR-D-23-0005.1.

——, ——, K. Dotterer, C. Thorncroft, and C. J. Schreck, 2022: The influence of convectively coupled Kelvin waves on African easterly waves in a wave-following framework. *Mon. Wea. Rev.*, **150**, 2055–2072, https://doi.org/10.1175/MWR-D-21-0321.1.

——, B. Weiser, and S. J. Majumdar, 2025: On the interannual variability of African easterly waves and its relationship with El Niño–Southern Oscillation. *J. Climate*, **38**, 5791–5808, https://doi.org/10.1175/JCLI-D-25-0113.1.

Leutbecher, M., and Coauthors, 2017: Stochastic representations of model uncertainties at ECMWF: State of the art and future vision. *Quart. J. Roy. Meteor. Soc.*, **143**, 2315–2339, https://doi.org/10.1002/qj.3094.

Loi, C. L., C.-C. Wu, and Y.-C. Liang, 2024: Prediction of tropical cyclogenesis based on machine learning methods and its SHAP interpretation. *J. Adv. Model. Earth Syst.*, **16**, e2023MS003637, https://doi.org/10.1029/2023MS003637.

Magnusson, L., and Coauthors, 2021: Tropical cyclone activities at ECMWF. ECMWF Tech. Memo. 888, 140 pp., https://doi.org/10.21957/zzxzzygwv.

——, and Coauthors, 2025: The role of observations in ECMWF tropical cyclone initialisation and forecasting. *Quart. J. Roy. Meteor. Soc.*, **151**, e4924, https://doi.org/10.1002/qj.4924.

Majumdar, S. J., and R. D. Torn, 2014: Probabilistic verification of global and mesoscale ensemble forecasts of tropical cyclogenesis. *Wea. Forecasting*, **29**, 1181–1198, https://doi.org/10.1175/WAF-D-14-00028.1.

——, L. Magnusson, P. Bechtold, J. R. Bidlot, and J. D. Doyle, 2023: Advanced tropical cyclone prediction using the experimental global ECMWF and operational regional COAMPS-TC systems. *Mon. Wea. Rev.*, **151**, 2029–2048, https://doi.org/10.1175/MWR-D-22-0236.1.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF Ensemble Prediction System: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119, https://doi.org/10.1002/qj.49712252905.

Nguyen, Q., and C. Kieu, 2024: Predicting tropical cyclone formation with deep learning. *Wea. Forecasting*, **39**, 241–258, https://doi.org/10.1175/WAF-D-23-0103.1.

NHC, 2025: Tropical cyclone reports. https://www.nhc.noaa.gov/data/tcr/.

Nolan, D. S., 2007: What is the trigger for tropical cyclogenesis? *Aust. Meteor. Mag.*, **56**, 241–266, https://doi.org/10.1071/ES06025.

——, E. D. Rappin, and K. A. Emanuel, 2007: Tropical cyclogenesis sensitivity to environmental parameters in radiative–convective equilibrium. *Quart. J. Roy. Meteor. Soc.*, **133**, 2085–2107, https://doi.org/10.1002/qj.170.

Ollinaho, P., and Coauthors, 2017: Towards process-level representation of model uncertainties: Stochastically perturbed parametrizations in the ECMWF ensemble. *Quart. J. Roy. Meteor. Soc.*, **143**, 408–422, https://doi.org/10.1002/qj.2931.

Radford, J. T., I. Ebert-Uphoff, J. Q. Stewart, K. D. Musgrave, R. DeMaria, N. Tourville, and K. Hilburn, 2025: Accelerating community-wide evaluation of AI models for global weather prediction by facilitating access to model output. *Bull. Amer. Meteor. Soc.*, **106**, E68–E76, https://doi.org/10.1175/BAMS-D-24-0057.1.

Richardson, D. S., H. L. Cloke, L. Magnusson, S. J. Majumdar, J. A. Methven, and F. Pappenberger, 2025: Skill and consistency of ECMWF forecasts of Atlantic tropical cyclone genesis. *Wea. Forecasting*, **40**, 703–717, https://doi.org/10.1175/WAF-D-24-0115.1.

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.

Russell, J. O. H., and A. Aiyyer, 2020: The potential vorticity structure and dynamics of African easterly waves. *J. Atmos. Sci.*, **77**, 871–890, https://doi.org/10.1175/JAS-D-19-0019.1.

Titley, H. A., M. Yamaguchi, and L. Magnusson, 2019: Current and potential use of ensemble forecasts in operational TC forecasting: Results from a global forecaster survey. *Trop. Cyclone Res. Rev.*, **8**, 166–180, https://doi.org/10.1016/j.tcrr.2019.10.005.

Wang, Z., 2012: Thermodynamic aspects of tropical cyclone formation. *J. Atmos. Sci.*, **69**, 2433–2451, https://doi.org/10.1175/JAS-D-11-0298.1.

——, 2014: Role of cumulus congestus in tropical cyclone formation in a high-resolution numerical model simulation. *J. Atmos. Sci.*, **71**, 1681–1700, https://doi.org/10.1175/JAS-D-13-0257.1.

——, 2018: What is the key feature of convection leading up to tropical cyclone formation? *J. Atmos. Sci.*, **75**, 1609–1629, https://doi.org/10.1175/JAS-D-17-0131.1.

——, M. T. Montgomery, and T. J. Dunkerton, 2010: Genesis of pre–Hurricane Felix (2007). Part II: Warm core formation, precipitation evolution, and predictability. *J. Atmos. Sci.*, **67**, 1730–1744, https://doi.org/10.1175/2010JAS3435.1.

——, 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.

Yamaguchi, M., and N. Koide, 2017: Tropical cyclone genesis guidance using the early stage Dvorak analysis and global ensembles. *Wea. Forecasting*, **32**, 2133–2141, https://doi.org/10.1175/WAF-D-17-0056.1.

Zawislak, J. A., 2020: Global survey of precipitation properties observed during tropical cyclogenesis and their differences compared to nondeveloping disturbances. *Mon. Wea. Rev.*, **148**, 1585–1606, https://doi.org/10.1175/MWR-D-18-0407.1.

Zhang, R., Q. Liu, R. Hang, and G. Liu, 2022: Predicting tropical cyclogenesis using a deep learning method from gridded satellite and ERA5 reanalysis data in the western North Pacific basin. *IEEE Trans. Geosci. Remote Sens.*, **60**, 4101810, https://doi.org/10.1109/TGRS.2021.3069217.

Zhang, X., J. Fang, and Z. Yu, 2023: The forecast skill of tropical cyclone genesis in two global ensembles. *Wea. Forecasting*, **38**, 83–97, https://doi.org/10.1175/WAF-D-22-0145.1.