Factors Influencing the Track of Hurricane Dorian (2019) in the West Atlantic: Analysis of a HAFS Ensemble

ANDREW HAZELTON,[a,b] GHASSAN J. ALAKA JR.,[b] MICHAEL S. FISCHER,[a,b] RYAN TORN,[c] AND SUNDARARAMAN GOPALAKRISHNAN[b]

[a] *Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, Florida*
[b] *NOAA/AOML/Hurricane Research Division, Miami, Florida*
[c] *University at Albany, State University of New York, Albany, New York*

(Manuscript received 19 April 2022, in final form 29 September 2022)

ABSTRACT: Hurricane Dorian (2019), a category-5 tropical cyclone (TC), was characterized by a large spread in track forecasts as it moved northwest. A set of 80 ensemble forecasts from the Hurricane Analysis and Forecast System (HAFS) was produced to evaluate Dorian's track spread and the factors that contributed to it. Track spread was particularly critical at long lead times (5–7 days after initialization near the Lesser Antilles), because of the uncertainty in the location of landfall and hazards. Four clusters of members were analyzed based on the 7-day track, characterized by Dorian moving: 1) slowly near the northern Bahamas (closest to reality), 2) across the Florida Peninsula, 3) slowly into Florida's east coast, and 4) quickly north of the Bahamas. Ensemble sensitivity techniques were applied to identify areas that were most critical for Dorian's track. Key differences were found in the strength of the subtropical ridge over the western Atlantic Ocean with a weaker ridge and slower easterly steering flow in the offshore groups. Subtle differences in the synoptic pattern over the United States also appeared to affect the timing of Dorian's northward turn, specifically the strength of a shortwave trough moving over the Ohio Valley. Despite some early track differences, the correlation between early and late track errors was not significant. An examination of four members further highlights the differences in steering and the strength of the subtropical ridge. This study demonstrates the utility of ensemble datasets for studying TC forecast uncertainty and the importance of medium-range modeling of synoptic-scale steering features to accurately predict the track of tropical cyclones.

SIGNIFICANCE STATEMENT: Hurricane Dorian was a catastrophic hurricane for the Bahamas and got very close to Florida without directly impacting the state. Some early forecasts showed the storm moving directly into or across Florida; others correctly showed the storm stalling over the Bahamas and then turning northward. This track forecast uncertainty made preparations in Florida challenging; therefore, we wanted to better understand why Dorian took the track that it did, to see what this tells us about the factors that affect hurricane tracks, and learn for future storms. We looked at an ensemble of 80 runs of a hurricane model, initiated at the same time. Some runs showed a Florida landfall; others showed Dorian stalling over the Bahamas. The strength of the subtropical ridge over the Atlantic north of Dorian and an upper-level trough of low pressure over the United States were key influences on storm path. These two large-scale features were better forecast in the ensemble members that correctly showed Dorian stalling and turning northward. This study shows how useful ensembles can be for understanding the processes driving hurricane motion and also shows that it is critical to forecast multiple synoptic-scale features correctly to accurately predict a hurricane's track 5–7 days in advance.

KEYWORDS: Synoptic-scale processes; Tropical cyclones; Ensembles; Numerical weather prediction/forecasting

1. Introduction

Although tropical cyclone (TC) track forecasts have improved significantly over the past several decades, progress in prediction of TC track has slowed (especially for shorter forecast lead times) more recently (Landsea and Cangialosi 2018), and there are still cases that prove very challenging for track predictions for both numerical models and human forecasters. Challenging track forecasts become especially important when the range of possible track outcomes can mean the difference between a landfall or no landfall for a populated area, due to the costs and hazards associated with large-scale evacuations (e.g., Wu et al. 2012). Hurricane Dorian in 2019 was one such case with uncertainty as to whether the TC would cross the Bahamas and make landfall on the east coast of Florida as a major hurricane, or stay over the Bahamas. In this study, we utilize an ensemble approach to understand uncertainty in the track of Dorian during this period, as well as the large-scale factors that caused the storm to ultimately stall over the Bahamas and turn north, rather than continuing west into Florida.

TC track forcing has been studied for decades. Chan (2005) provides a summary of some of the key physics that affect TC motion. One of the leading factors influencing TC track is the environmental steering flow. George and Gray (1976) used rawinsonde data to study TC tracks over the western North Pacific Ocean and found that the flow 1°–7° from the TC was correlated with the motion of the TC. They also found that the speed was

Corresponding author: A. Hazelton, Andrew.Hazelton@noaa.gov

DOI: 10.1175/MWR-D-22-0112.1

best predicted by the flow at 700 hPa, while the direction was best predicted by the flow at 500 hPa. Later studies highlighted the importance of steering flow in different layers of the atmosphere as well as the intensity and depth of the TC vortex, with stronger storms tending to be steered by deeper-layer flow (Velden and Leslie 1991). More recent studies (e.g., Galarneau and Davis 2013) have expanded on this concept and attempted to define an optimal TC steering layer that can be used to understand TC motion as well as errors in TC motion and environmental steering.

Another tool used to understand and forecast TC tracks, particularly for TCs with uncertain medium-to-long-range track forecasts, is ensemble methods (e.g., Zhang and Krishnamurti 1997). For example, Munsell and Zhang (2014) used Weather Research and Forecasting (WRF) Model ensembles to study the track uncertainty of Hurricane Sandy (2012) and found that the environmental steering and motion in the first 24–48 h of the forecast was most important in determining the TC's eventual track toward the United States. Torn et al. (2015) also examined Hurricane Sandy using a set of GFS ensemble forecasts, and showed the importance of steering from an upper-level ridge, which was modulated by convection and precipitation on the north side of the TC. Torn et al. (2018) used the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble to study the sensitivity of TC tracks in deformation steering zones (including TCs Debby in 2012 and Joaquin in 2015), and found that the steering flow near the TC (within 500 km) was important for determining the track outcome. Alaka et al. (2019) used an ensemble of the Hurricane Weather Research and Forecasting (HWRF) Model to understand the uncertainty associated with the track of Hurricane Joaquin in 2015. That analysis found that Joaquin's track was highly sensitive to the evolution of several key synoptic-scale features, including troughs over North America and to the northeast of the TC. These features altered the steering flow near the TC center, with small-scale track variations at shorter lead times having large implications for the TC position at longer lead times. Nystrom et al. (2018) also used an ensemble of high-resolution forecasts to examine Hurricane Joaquin, and found that while the intensity was sensitive to the initial conditions near the TC, the track was most sensitive to the steering flow in the region 600–900 km away from the TC center.

For Hurricane Dorian, the period of largest track uncertainty occurred as the storm slowly approached the populated Bahamas and East Coast of Florida. In particular, it was unclear whether Dorian would make landfall in Florida or would remain offshore, underscoring the importance of understanding the environmental factors that influenced the eventual track offshore of Florida. In addition, as ensemble modeling becomes more important for forecasting and analysis due to its ability to represent a range of possible outcomes and the forecast uncertainty, it is important to examine techniques that can be applied to high-resolution, large ensemble datasets to understand both large-scale and small-scale factors that influence TC track in uncertain patterns like those associated with Dorian as well as other cases with large spread and/or high potential impact. This analysis makes use of a new model system in an ensemble framework to examine these topics.

2. Data and model setup

This study uses an ensemble of the Hurricane Analysis and Forecast System (HAFS), specifically the global-nested version of HAFS (HAFS-globalnest; Hazelton et al. 2021a), which features a global-nested configuration of the FV3 dynamical core with two-way feedback between the global and nested domains (e.g., Harris and Lin 2013). The ensemble set used in this study contains 80 members, initialized with the 80 members of the Global Ensemble Forecast System (GEFS) EnKF data assimilation system (Zhou et al. 2017). The HAFS forecasts were run at 3-km grid spacing over the large Atlantic Ocean nested domain, with 13-km grid spacing over the global domain. The forecasts were initialized at 0000 UTC 27 August 2019, while Dorian was just east of the Lesser Antilles.

Hazelton et al. (2021b) used the same set of ensemble forecasts to analyze the processes leading to the early intensification of Dorian in the Caribbean Sea and described the model configuration, including the grid layout and physics schemes utilized. That study demonstrated that some of the initial moisture fields, including humidity in the northeast quadrant downshear of the TC, were important for the observed intensification and apparent "reformation" of the center northeast of some of the initial forecasts. Many of the ensemble members did not fully capture the observed intensification, although there were a few that were very close to observations during the first 48 h, and one of these more accurate members is documented in detail in that study. The reformation and intensification processes in Dorian were also extensively documented observationally in Alvey et al. (2022).

3. Results

The results are laid out as follows. First, we present the overall ensemble results, particularly focused on the final ~84 h of the forecast period when Dorian was approaching the Bahamas and Florida. Next, we use correlation analyses to help determine what the connection was, if any, between the early evolution of the ensemble forecasts (Hazelton et al. 2021b; Alvey et al. 2022) and the final track, and determine which periods in the forecast were most critical in determining the final track outcome. Then, we separate the groups into four different clusters based on the final position at 168 h, and dive deeper into the divergence in position and speed between the groups, which further illustrates the time periods that were key for ensemble divergence and gives clues as to the important large-scale processes determining Dorian's track near the Bahamas. A sensitivity analysis provides further evidence of the synoptic processes that allowed Dorian to stall and turn before reaching Florida, and these synoptic processes are analyzed in detail through composites of the different groups. Finally, four representative members of the ensemble (one from each cluster) are analyzed to further illustrate the speed

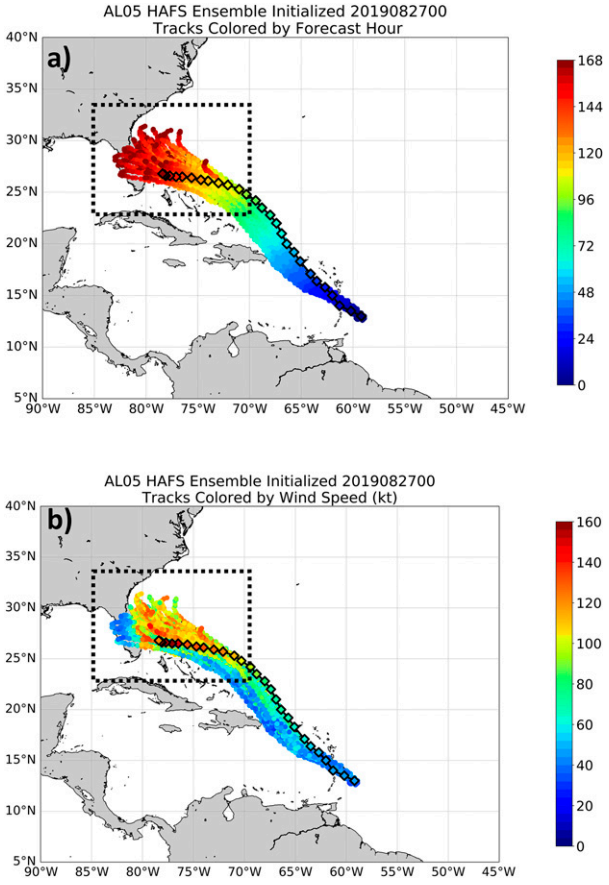

FIG. 1. (a) Tracks of all 80 ensemble members initialized at 0000 UTC 27 Aug 2019 colored by forecast hour (0–168). The observed track is shown in black with forecast hour in colored diamonds. (b) Forecast tracks for the 80 Dorian members shaded by the maximum forecast 10-m wind speed (kt; 1 kt $\approx$ 0.51 m s^{-1}). The observed track is shown in black, with intensity shown in colored diamonds. The dash-outlined black box in (a) and (b) approximately shows the 84–168-h period of the forecast.

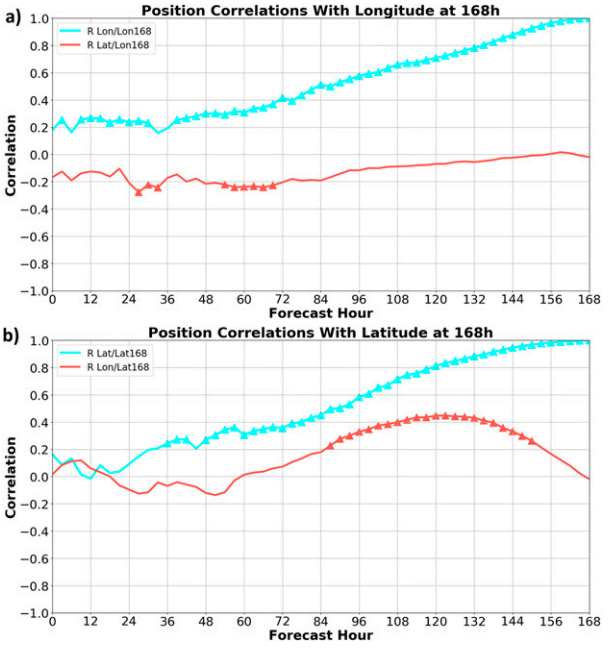

FIG. 2. (a) Correlation coefficient between Dorian's ensemble longitude (light blue) and latitude (coral) at each forecast hour and the ensemble longitude at 168 h. (b) Correlation coefficient between Dorian's ensemble latitude (light blue) and longitude (coral) at each forecast hour and the ensemble latitude at 168 h. The triangles indicate times when the correlations are statistically significant ($p < 0.05$).

differences in the ensemble and the synoptic patterns that led to these differences, and eventually determined the final track of Dorian.

a. Ensemble results

Figure 1 shows the track forecasts of all 80 ensemble members, shaded by both forecast hour (Fig. 1a), to illustrate the change in track spread over time, and by intensity (Fig. 1b), to highlight the track and intensity connection. The observed track was generally toward the northern edge of the ensemble suite for the first 48–72 h, but in the 84–168-h period in the western Atlantic and Bahamas (the focus of this study, shown in the black-outlined box in Fig. 1), the spread increased greatly.

b. Ensemble track correlations

One question that arises due to some of the early track biases is whether a connection exists between the position of Dorian during the first 1–2 days of the forecasts and the ultimate TC position near Florida and the Bahamas, such as the connection seen in the early forecasts of cases like Joaquin in 2015 and Debby in 2012 (Torn et al. 2018). To help address this question, the serial correlations of latitude and longitude at each forecast hour with the longitude (Fig. 2a) and latitude (Fig. 2b) at 168 h are examined. The early longitude and latitude showed little correlation with the final longitude (Fig. 2a). Obviously, these relationships will get stronger as the difference between the lead times decreases, but the time periods where the relationship increases most quickly will point to important periods for further examination, since we are interested in addressing the factors that affected the final track position at 168 h, to see whether they were mainly driven by early pattern differences and biases or differences that appeared later on. These relationships will help to illuminate the sensitivity between the early period of Dorian (e.g., Hazelton et al. 2021b) and the later track. The longitude became more correlated to the final longitude as time went on, as expected, but latitude throughout the forecast showed little correlation with longitude at 168 h [there are some weakly correlated periods (less than ~10% of variance explained) at 27–33 and 54–69 h.] The early longitude and latitude are correlated with the latitude at 168 h (Fig. 2b), but the relationship stayed below $r = 0.4$ until 72–84 h as the TC moved north of the Antilles. The longitude at hour 87–150 h was correlated (statistically significantly, based on a t test) with

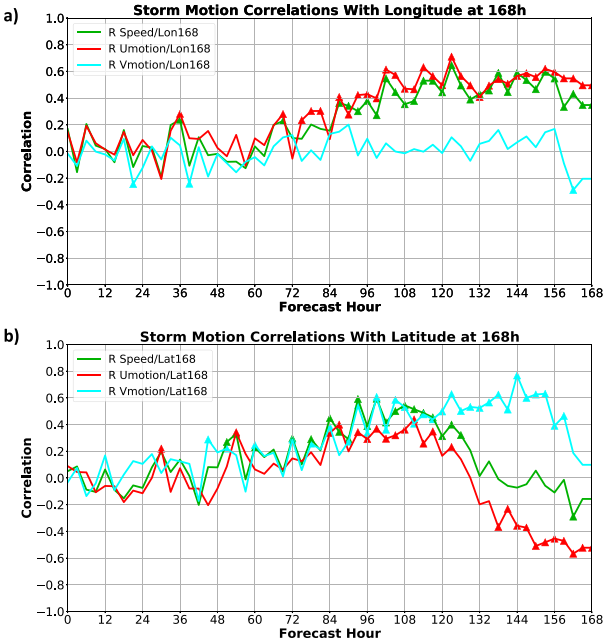

FIG. 3. As in Fig. 2, but for zonal motion (red), meridional motion (cyan), and speed (green).

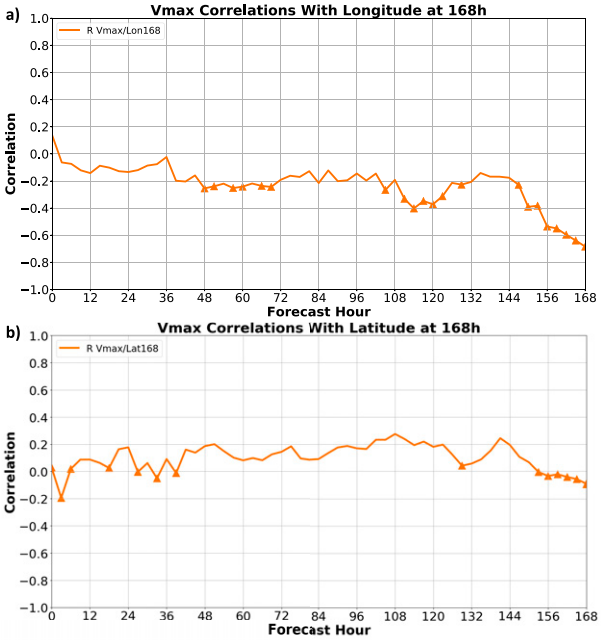

FIG. 4. As in Fig. 2, but for TC intensity.

the latitude at 168 h (Fig. 2b, pink line), indicating that some of the speed differences in the middle of the forecast led to some differences in the speed with which the TC turned north ahead of a midlatitude trough (discussed later). The generally weak relationship between the 168-h position and the 0–48-h position indicates that the early track errors were likely not a major factor affecting the long-term track of Dorian. Rather, Fig. 2 suggests that synoptic features driving the motion of the TC starting 2–3 days into the forecasts were likely more critical for determining the final position. These will be examined in more detail.

To further examine some of the important variables that may have influenced the final position of Dorian at the end of the ensemble runs, Fig. 3 shows the correlation between TC motion (zonal motion, meridional motion, and storm speed) and the longitude (Fig. 3a) and latitude (Fig. 3b) at 168 h. For these calculations (and others for TC motion later), westward movement has positive correlation, while eastward motion has negative correlation. The results paint a picture similar to that shown by the position correlations. First, there is generally little relationship between the early zonal motion (meridional motion at all lead times) and the final longitude (Fig. 3a, red and blue lines). This indicates that it was not necessarily the early northward jump that caused Dorian to end up east of Florida. The correlation between zonal motion and longitude (Fig. 3a) increased to 0.4 around 84 h and stayed strong for the next few days, indicating that the biggest influences on the final longitude of Dorian occurred 2–3 days into the forecast.

One final relationship that was critical to examine was the connection between the early intensity of Dorian and the long-term track, to understand if a stronger storm earlier in the forecast was more likely to stall and turn east of Florida in this ensemble set. Figure 4 shows the correlation between ensemble intensity and the ending longitude (Fig. 4a) and latitude (Fig. 4b) at 168 h. For these plots, positive correlation indicates that the storm was farther west/north.

The correlations are generally less than ±0.2. There was a slight relationship between the intensity after Dorian emerged from the Caribbean (around 48–60 h) and the final longitude, with the stronger storms ending up farther east (Fig. 4a). However, despite statistical significance at a few forecast hours, the variance explained was generally less than 5%–10%. There was a stronger relationship with intensity near the end of the period, since the storms that moved farther west and over Florida ended up weakening earlier. The relationship with the final latitude was even weaker, with very little correlation until the final ~12 h of the forecast period (Fig. 4b). It is possible that if the ensemble had captured the early intensification better (since not all members did), then the correlation would be somewhat stronger. Still, there was enough spread in this ensemble set to indicate that the early intensification (Hazelton et al. 2021b; Alvey et al. 2022) may have been a small piece of the final track outcome but the synoptic evolution around Dorian throughout the forecast period was more important. This is the focus of much of the rest of the paper.

c. *Ensemble clustering*

To parse the ensemble dataset for further analysis of the long-term track, it is helpful to separate the members into groups or clusters of similar positions. Examining how these different clusters evolve should help us see if the time periods analyzed in the correlation analyses above were indeed key divergence points between the members. For the initial analysis, since the forecast spread had increased significantly by the

end of the forecast period (Fig. 1), a *k*-means clustering algorithm (Wilks 2006) was used to group members based on the forecast position at the end of the forecast period (168 h). This technique has been applied previously to analyze clusters of global model forecasts of TCs (Evans et al. 2006). Kowaleski and Evans (2016) used a different clustering method (regression mixture modeling) to group ensemble forecasts of Hurricane Sandy (2012) based on track and structure. Kowaleski and Evans (2020) analyzed ensemble clusters from several different global ensembles, and found that smaller clusters tended to produce the largest TC track errors. Because *k*-means clustering requires the number of clusters to be specified before the clustering is performed, sensitivity tests using $k = 3$–10 clusters were examined. Figure 5 shows the results from applying this algorithm with a cluster size of $k = 3$ and a cluster size of $k = 5$ (these two cluster sizes produced the most cohesive clusters based on the metric discussed below). To determine the cohesiveness of each clustering result, the silhouette score was computed, following Rousseeuw (1987):

$$S = \frac{b - a}{\max(a, b)}, \tag{1}$$

where a is the average intracluster distance and b is the average distance to the nearest cluster for each sample. In essence, as the silhouette score approaches a maximum possible value of 1, the clustering becomes increasingly cohesive. A silhouette score of 1 means that the clusters are very dense and well separated. The score of 0 means that clusters are overlapping.

The $k = 3$ clustering yields clusters that can essentially be classified as "Florida landfall" (cluster 1, green), "fast offshore" (cluster 3, red; staying offshore of Florida and moving north of more quickly), and "slow Bahamas" (cluster 2; orange; staying offshore of Florida and moving more slowly through the northern Bahamas). The "slow Bahamas" group is closest to the observed track of Dorian. The $k = 5$ clustering has the highest "silhouette score" of all the groups. It has similar "slow Bahamas" (cluster 2, orange; hereinafter "southeast") and "fast offshore" (cluster 1, green; hereinafter "northeast") groups, but has two different Florida landfall clusters with a variety of distance across the state by 168 h: 1) "fast Florida landfall" (cluster 4, purple; hereinafter "northwest") and 2) "slow Florida landfall" (cluster 3, red; hereinafter "southwest"). There is also an "outlier" group in this set, with 3–4 members east of the main clusters (cluster 5, brown). Because of the slightly higher silhouette score and clear differences in speed between the two "Florida landfall" groups, the $k = 5$ clustering is chosen for further analysis.

d. Cluster evolution

To use the clusters to help determine the factors that were important for the eventual track of Hurricane Dorian, the point of divergence between the ensemble groups was explored. Based on the correlations shown in Figs. 2–4, the day-3–4 period appears to be critical. To illustrate the ensemble divergence over time, the different groups are shown at several different lead times (Fig. 6). There is little divergence for the first two days of the forecasts (Figs. 6a,b). Around

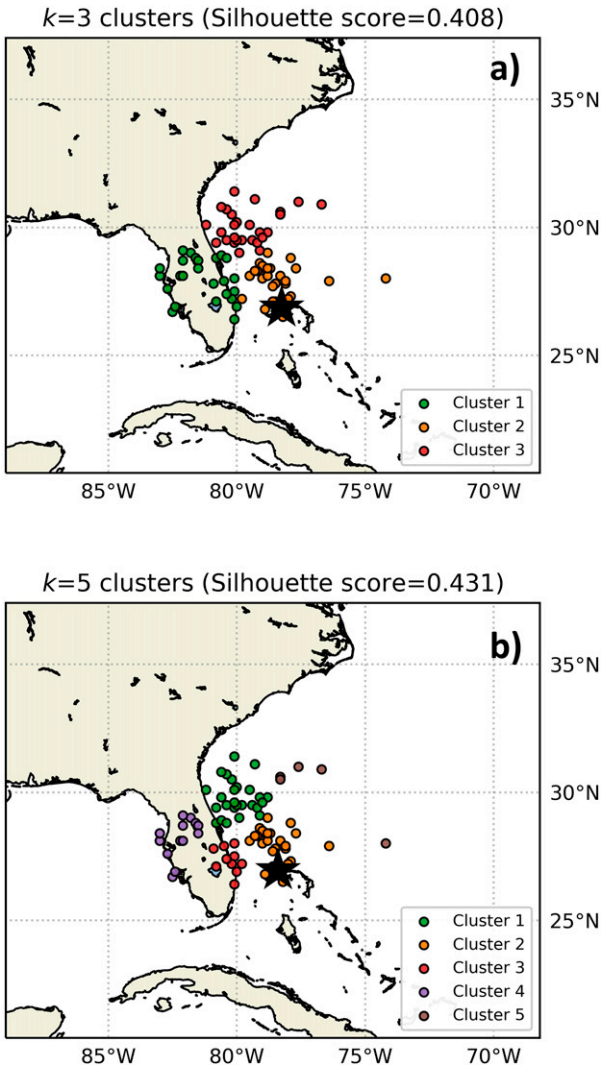

FIG. 5. HAFS ensemble forecast positions of Hurricane Dorian initialized at 0000 UTC 27 Aug 2019, valid at 168 h, grouped by a *k*-means algorithm with (a) $n = 3$ clusters and (b) $n = 5$ clusters. The resulting silhouette score is shown above each panel. The "best track" position of Dorian at 0000 UTC 3 Sep 2019 is shown with the black star.

days 3–4 (Figs. 6c,d), the speed and clustering differences begin to become more apparent, and the 4 key groups diverged farther apart by day 5 (Fig. 6e). One interesting detail to note is some overlap between the southwest and southeast groups at 144 h (Fig. 6g) disappears at the end of the forecast (168 h; Fig. 6h), indicating that there was uncertainty in the westward extent of the forecast tracks late in the period, which is worth exploring further.

To provide further evidence of the divergence periods and help objectively determine when the track forecasts began to diverge, a statistical approach was used to define track spread based on a confidence ellipsoid computed directly from ensemble member locations (Fig. 7). The confidence ellipsoid is calculated using an empirical orthogonal function (EOF) analysis

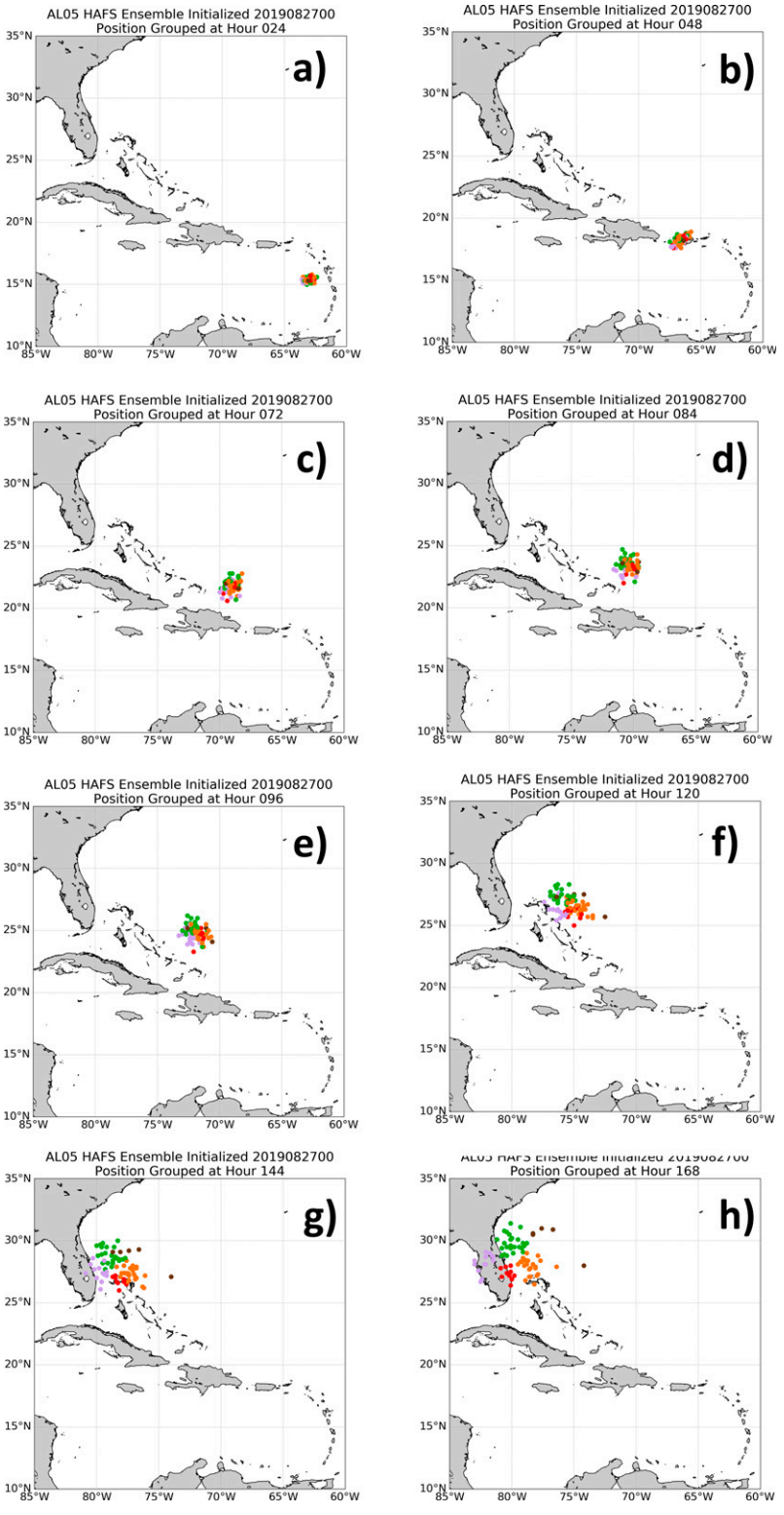

FIG. 6. Dorian HAFS ensemble groups at (a) 24, (b) 48, (c) 72, (d) 84, (e) 96, (f) 120, (g) 144, and (h) 168 h. The group colors are as follows: northwest (purple), southwest (red), northeast (green), southeast (orange), and outlier (brown).

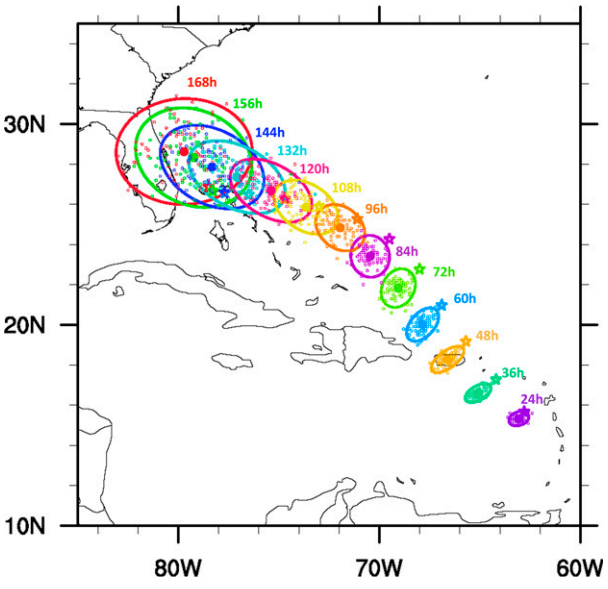

FIG. 7. Confidence ellipsoids illustrate track spread in HAFS-globalnest ensemble forecasts for Hurricane Dorian every 12 h from 24 to 168 h. Also shown are ensemble mean locations (large filled circles), individual ensemble member locations (small open circles), and observed track locations (large stars). Ellipsoids and markers are color coded by forecast lead time.

that uses the TC center latitude and longitude locations of every ensemble member at a given lead time (Alaka et al. 2018, 2019). This method produces qualitatively similar results to the bivariate normal fit approach described in previous studies (Hamill et al. 2011, 2012; Torn et al. 2015). The eigenvalues associated with the first two EOFs are used as a proxy for the variance of the ensemble member locations, and, hence, the semi-major and semi-minor axes are approximated by the square root of the corresponding eigenvalues. Each ellipsoid axis is scaled by a chi-squared probability value to reflect 90% confidence. The rotation of each ellipsoid is calculated as the arctangent of the eigenvectors associated with the first two EOFs.

For the first 2–3 days of the forecast, most of the spread was across-track, with the observed TC track on the northeast edge of the ensemble suite. However, the across-track spread did not increase much from 48 to 108 h, and, around 84 h (purple), the along-track spread began to increase due to speed divergence in different members. The overall spread of the ensemble increased around this time, as well, and the observed TC track was better captured within the ensemble suite starting around 96 h. This, along with the increase in serial correlation of latitude and longitude around this time (Fig. 2), makes 84 h an important divergence point that will be explored in further detail. Toward the end of the forecast period, the across-track spread began to increase again as there were large differences in both the forward speed and also the timing of Dorian's turn to the north.

Another illustration of how the different members and groups diverged over time can be seen in an examination of the TC motion in each group (Fig. 8). In the first 36 h of the forecasts, the forecast and observed TC motions were somewhat divergent as the TC was broad and reformed (Hazelton et al. 2021b; Alvey et al. 2022), and the observed TC moved more northeastward in the first 24 h than almost all of the ensemble members showed. However, as the correlations showed (Figs. 2 and 3), there did not appear to be a strong connection between these early track errors and the eventual track of the TC. Also, after the redevelopment in the first 24–36 h, the ensemble motion was more in line with the observed, although the observed speed was generally closer to the slower ensemble groups (especially the meridional motion).

There were some key differences between the groups that explain the divergence in tracks seen in Fig. 7. For example, around 84 h the northwest (NW; purple) and northeast (NE; green) groups begin to move faster than the southwest (SW; red) and southeast (SE; orange) groups, as the confidence intervals minimally overlap (Fig. 8c). In the case of the NE group, the faster motion is mostly meridional (Fig. 8b), while for the NW group, it is faster zonal motion (Fig. 8a). This result is consistent with where those groups ended up at the end of the forecast (Fig. 5b). The slower motion in the SW/SE groups was more in line with the observed motion at 84 h. Around 144 h, there is another key divergence point, this time between the SE and SW groups. The SW group remains slightly faster (almost all from zonal motion) than the SE group, which is why it moves into or near the coast of Florida rather than staying over the Bahamas like the SE group. Note that the NE group also has very small zonal motion by 168 h, as this cluster has mostly reached the edge of the subtropical ridge due to faster earlier motion, and begun to recurve. This analysis helps to confirm the importance of the 84-h (1200 UTC 30 August 2019) and the 144-h (0000 UTC 2 September 2019) lead times as key time points for further examination of the TC steering and synoptic influences.

e. Composite steering differences

In the section above, we saw that there were key speed differences between the different groups at various points that led to the divergence in forecast position. To see what was responsible for these speed divergences, we next look for differences between the ensemble groups in the synoptic features and steering flow near Dorian when Dorian was east of the Bahamas and as the storm started to turn near Florida. One of the analyses performed involved calculating the layer average flow around the TC at each level from 850 to 200 hPa, with the base of the layer at 850 hPa. The vortex was removed from this data using a method similar to that in Galarneau and Davis (2013), and averaged in a box of size 6° by 6° centered on the TC center location in each member (close to the optimal radius found in Galarneau and Davis 2013). Figure 9 shows the layer-mean zonal and meridional flow for each cluster at 84 and 144 h, as well as the observed flow (from the GFS analysis) at the same valid times.

At 84 h, the NW and NE groups had the strongest easterly flow throughout the troposphere, particularly the NW group (Fig. 9a). The stronger easterlies are consistent with the faster

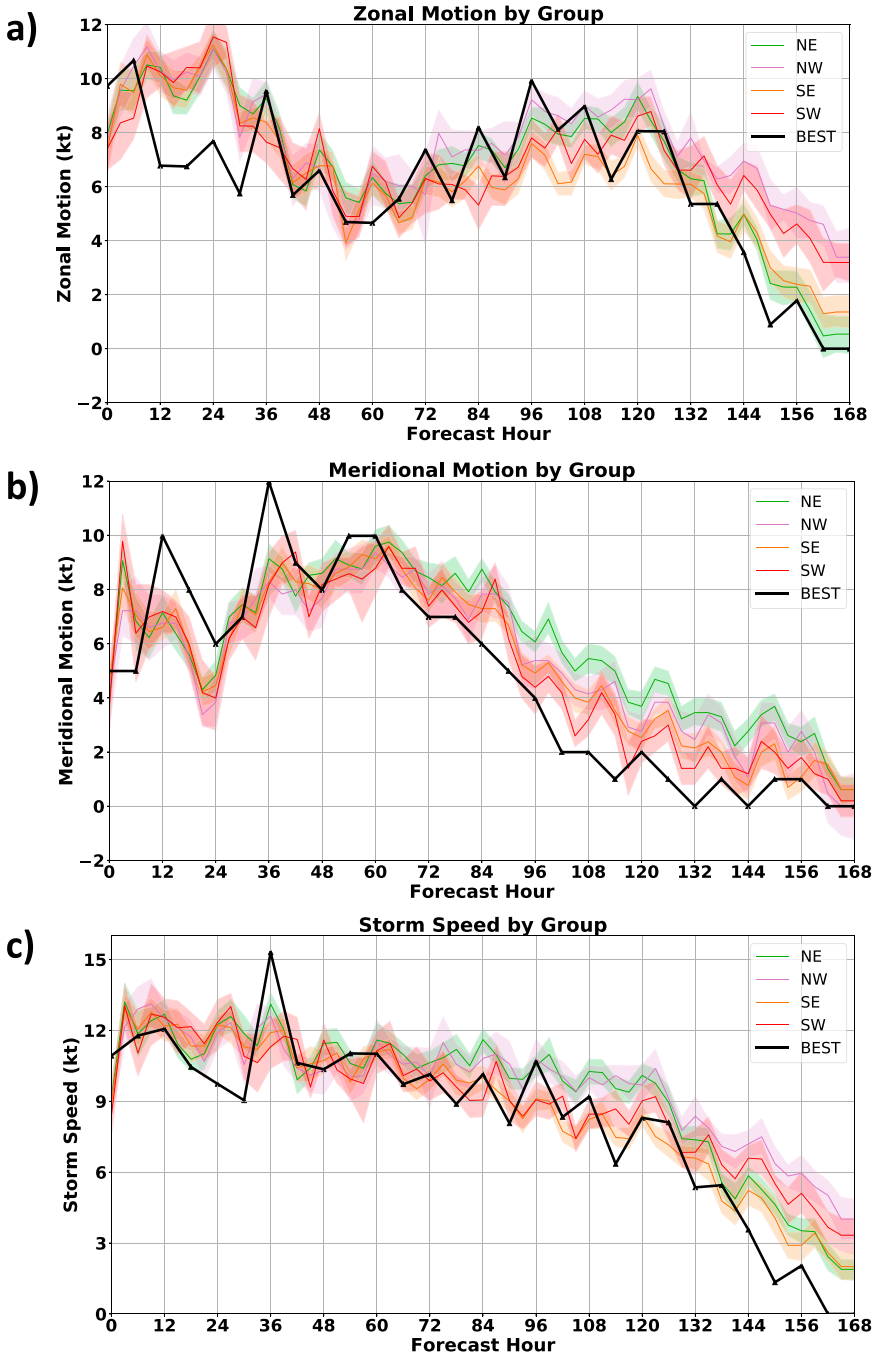

FIG. 8. (a) Zonal TC motion, (b) meridional TC motion, and (c) total TC speed for each forecast hour in HAFS for each group: NW (purple), SW (red), NE (green), and SE (orange). The mean and confidence interval of each group are shown. For (a) and (b), westward and northward motion are respectively shown as positive. The observed TC data (motion or speed) are shown in black.

westward motion in these groups. Interestingly, the zonal winds were similar in the SW and SE groups at this point (SW actually had weaker easterlies), consistent with the fact that the zonal motion had not diverged between these two groups at this point. The meridional winds throughout most of the troposphere (Fig. 9b) were strongest in the NE group, consistent with the faster-than-observed northward motion in this group around 84 h. The meridional winds were weakest in the SW and SE groups through most of the troposphere. Interestingly, there was some divergence between all groups and the

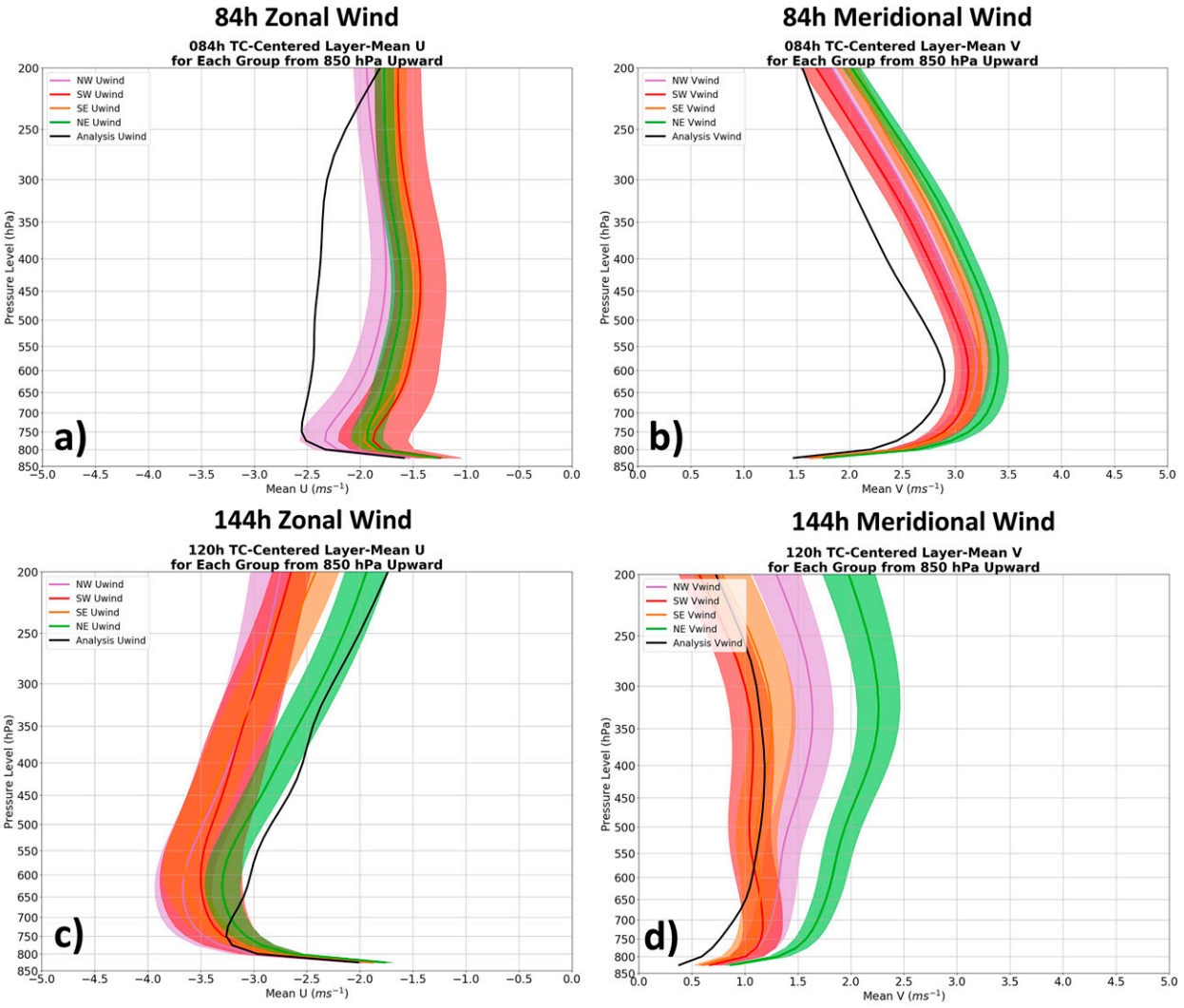

FIG. 9. Mean storm-centered (left) zonal and (right) meridional wind (with the TC vortex removed) from 850 to 200 hPa from the (a),(b) 84- and (c),(d) 144-h forecasts from the NW group (magenta), SW group (red), SE group (orange), and NE group (green). The GFS analysis steering is shown in black. Positive U and V represent flow toward the east and north, respectively.

GFS analysis at 84 h. Obviously there is some observational uncertainty, but this could partially explain why most of the tracks were outside the observed track during the first couple days of the forecast. For example, the weaker meridional flow in the GFS analysis than all groups at 84 h is consistent with the slower poleward motion in all groups at this time (Fig. 8b).

The steering at 144 h is also consistent with the divergence in motion between the different groups. At this point, the NW and SW groups had the most easterly zonal flow through the depth of the troposphere, with the NW having the strongest low-level flow from the east. At 144 h the divergence between the SW and SE groups became more apparent, with notably weaker zonal easterly flow in the SE group (Fig. 9c). This confirms the importance of this second "divergence point" as the different members moved near the Bahamas. The NE group had the smallest easterly flow aloft component, as some of these members had already begun to recurve ahead of a midlatitude trough by 155 h. The meridional wind at 144 h (Fig. 9d) is also consistent with the north–south divergence of the four groups toward the end of the forecast period. The NE group has the strongest southerly flow (especially above 400 hPa) as some of the TCs began to recurve, followed by the NW group, while the SW and SE groups had weaker southerly flow, closer to the GFS analysis. The divergence between the SW group (which got right up to or hit the Florida coast) and the SE group (which stayed offshore) was almost solely due to the zonal flow difference (Fig. 9c). Note that examination of the flow at 120 h (not shown) showed less difference between the SE and SW groups, indicating that 144 h was indeed a key divergence point for these two groups as the members moved slowly near or through the Bahamas.

f. Ensemble sensitivity and synoptic analysis

The previous sections demonstrated some key time periods where the ensemble members diverged, and speed differences showed up during these time periods. Next, we examine the sensitivity of Dorian's track within the ensemble to different synoptic features that evolved near and around the TC throughout its life cycle, which should help identify the aspects of the flow that yielded the speed and position divergence. To help determine which features were most important in the evolution of Dorian's track, we employ ensemble sensitivity analysis (e.g., Torn and Hakim 2008). A similar method has been used to determine key features associated with the track evolution of other TCs within complex steering environments, such as Hurricane Sandy (2012; Torn et al. 2015) and Hurricane Joaquin (2015; Torn et al. 2018). Here, sensitivity is evaluated [which was originally derived in Ancell and Hakim (2007)] as

$$\frac{\partial J}{\partial x} = \frac{\mathrm{cov}(J,x)}{\mathrm{var}(x)}, \tag{2}$$

where J represents the ensemble estimate of the forecast metric (with the mean removed), x is the forecast field of interest (also calculated with the mean removed and normalized by standard deviation), cov defines the covariance, and var defines the variance. For this study, x is defined as the ensemble height field (at 500 and 300 hPa) and J is defined as the anomalous westward distance (from the ensemble mean) at 168 h in the forecast. Thus, the method is being used to identify those regions at each forecast hour where the 168-h longitude/westward position was sensitive to the mid- to upper-level height field. Regions of large sensitivity magnitude indicate where changing the 500-hPa height at that earlier time will yield the largest change in Dorian's track. Statistical significance is assessed using the method of Torn and Hakim (2008), which calculates a confidence interval on the regression coefficient.

Figure 10 shows the sensitivity metric for forecast hours 0, 24, 48, 72, 84, 96, 120, and 144 at 500 hPa, and Fig. 11 shows the same metric evaluated with the 300-hPa height. Red shading indicates that increasing that height yields a more western position (i.e., positive sensitivity) while blue shading indicates that decreasing the height at that location yields a more western position (i.e., negative sensitivity), and interpretation of the graphic relies on the spatial location of the shading. If shading is over a trough, the color dictates whether a weaker trough (red; higher heights) or a stronger trough (blue; lower heights) led to Dorian's anomalous westward distance. If the shading is over a ridge, the color dictates whether a weaker ridge (blue; lower heights) or a stronger ridge (red; higher heights) led to Dorian's anomalous westward distance.

At 0 h, there is little sensitivity of the anomalous westward distance of Dorian at 168 h to the 500-hPa height field (Fig. 10a). Yet there was a slight negative sensitivity (i.e., blue shading; lower heights) to the 300-hPa trough north of Dorian at initialization (Fig. 11a), implying that a stronger trough at 0 h is correlated with Dorian's anomalous westward distance at 168 h. Starting at 72 h, the subtropical ridge north of Dorian becomes a clear source of forecast sensitivity at both 300 and 500 hPa. That ridge has positive forecast sensitivity (i.e., red shading; higher heights) and implies that a stronger ridge is correlated with Dorian's anomalous westward distance at 168 h. The northeastern flank (around 35°–45°N, 45°–55°W) of the subtropical ridge also showed a signal, with a stronger ridge (i.e., red shading, higher heights) in the north-central Atlantic correlated with Dorian's anomalous westward distance at 168 h. Starting around 84–96 h (Figs. 10e,f and 11e,f), heights over the continental United States, particularly the Eastern Seaboard, also exhibit a statistically significant positive sensitivity signal that became increasingly strong later in the period. This was associated with a trough over the United States during this time period.

More specifically, a weaker trough (i.e., red shading; higher heights) over the eastern United States was an indicator that Dorian would end up farther west. Interestingly, there was a negative sensitivity (i.e., blue shading; lower heights) to the trough NE of Dorian, likely as a result of a more progressive trough, and also potentially associated with the strength of the upstream ridge in the more amplified members. Finally, there was a weak but noticeable negative sensitivity (i.e., blue shading; lower heights) in the region of the ridge over the U.S. plains. This indicates that if a stronger ridge were present in this location, it would more likely block Dorian's westward progress toward Florida and/or induce a stronger downstream trough that could lift the storm north. This analysis confirms some of the key evolution of the forecast after 84 h, and also highlights some of the key synoptic features that are examined next.

Figure 12 shows the mean 500- and 300-hPa heights for each group as well as the GFS analysis for forecast hours 84 and 144. To illustrate some of the key details of the large-scale pattern over the western Atlantic and central/eastern United States as Dorian approached from the east, the 590- and 588-dam heights are contoured at 500 hPa and the 972- and 963-dam heights are contoured at 300 hPa.

At 84 h, the subtropical ridge to the NW of Dorian was stronger in the analysis than in all of the group means, both at 500 (Fig. 12a) and at 300 hPa (Fig. 12b). This bias is consistent with a general weak bias in the western flank of the subtropical ridge that has been noted in other studies of HAFS (Hazelton et al. 2022) and suggests that Dorian should have progressed farther west, which was not the case in reality. However, one key difference is the subtropical ridge to the northeast of the TC. The ridge was weaker in the analysis (especially at 300 hPa) in this location than any of the groups, although the SE and SW groups were closer to reality (especially at 500 hPa). This difference in the position of the ridge as the northeast side eroded seemed to be key to the groups that stayed farther south. The much stronger ridge NE of the TC in the NW group was consistent with the stronger 300-hPa easterlies (Fig. 9) relative to the other groups, which lead to Dorian moving west more quickly and to a position to impact Florida. Another subtle, but key difference, was the position and strength of the trough over the central contiguous United States. This feature was more pronounced in the analysis at 500 hPa than in any of the groups at this early lead time. Based on the sensitivity analysis presented

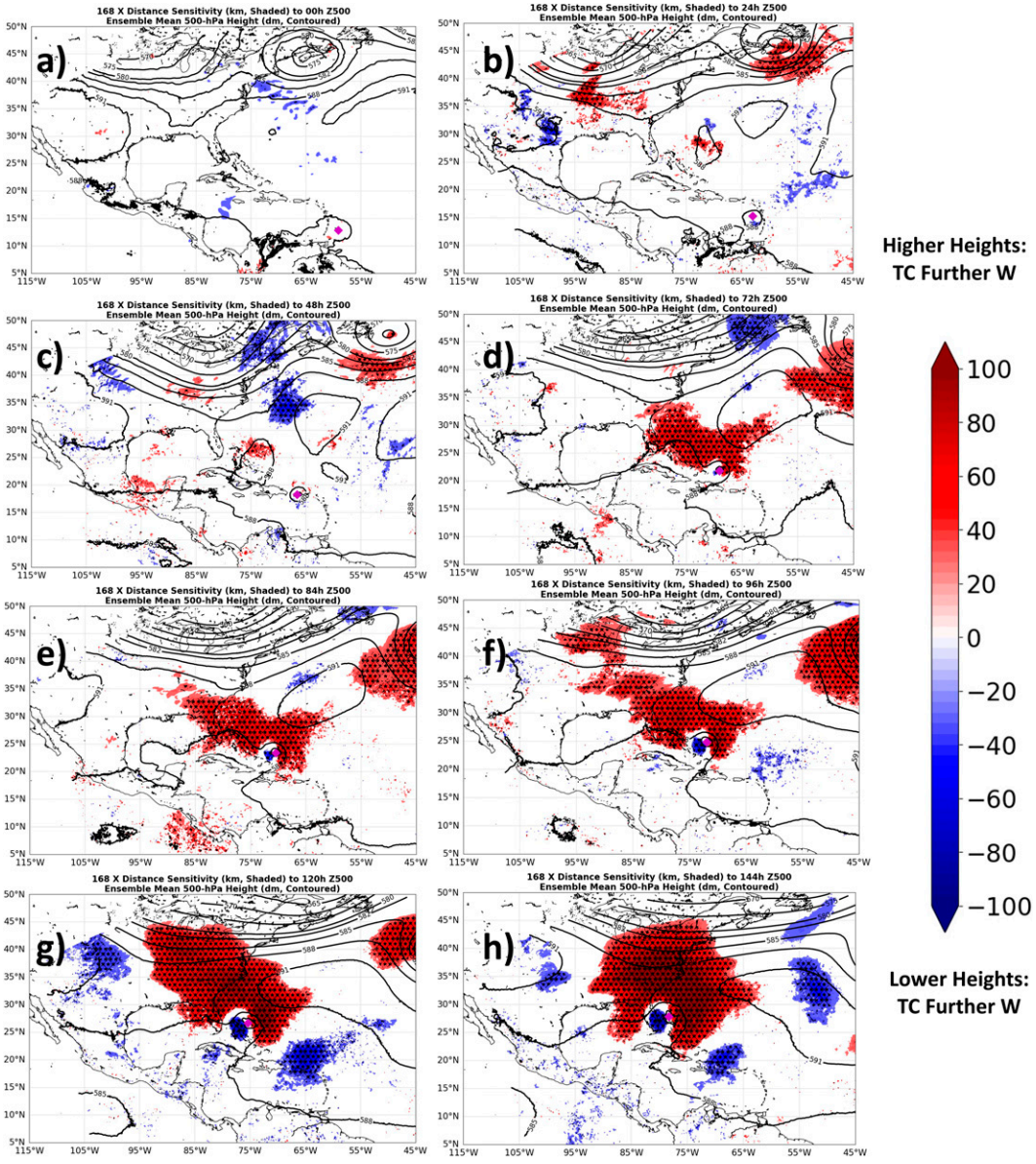

FIG. 10. (a) Ensemble sensitivity (km; shaded) of 168-h forecast zonal center position (relative to the ensemble mean) to the ensemble 500-hPa height field at (a) 0, (b) 24, (c) 48, (d) 72, (e) 84 h, (f) 96, (g) 120 h, and (h) 144 h of the forecast. The ensemble mean height field is contoured. Only regions of 90% significance are shaded, and regions of 95% significance are additionally starred/crosshatched. The ensemble mean TC position is shown as the magenta diamond.

earlier, a stronger trough in this location was associated with the TC staying farther east.

At 144 h, the 500-hPa ridge to the north of Dorian is again stronger in the analysis than in any of the groups, consistent with the aforementioned HAFS bias (Fig. 12c). However, the 500-hPa trough over Kansas and Missouri is most pronounced in the SE group (in comparison with the other groups) and closest to the observed trough, although it is slightly out of phase with the analysis. This shortwave was a key feature that likely allowed the ridge to erode enough for Dorian to stall and turn north. At 300 hPa, the ridging to the north of Dorian was weaker in the analysis than all the groups, although again the SE group was closest to the analysis (Fig. 12d). Also of note was the stronger ridging developing over the Gulf of Mexico in the analysis, which likely aided the slowing of the TC and amplification of the downstream trough.

In short, it appears that although all of the ensemble members suffered from some bias in the synoptic fields (particularly the strength of the subtropical ridge), the SE group (which was closest to the correct track) was correct in two key aspects: 1) the erosion of the upper-level ridge to the NE of Dorian (by an upper level trough to the east)

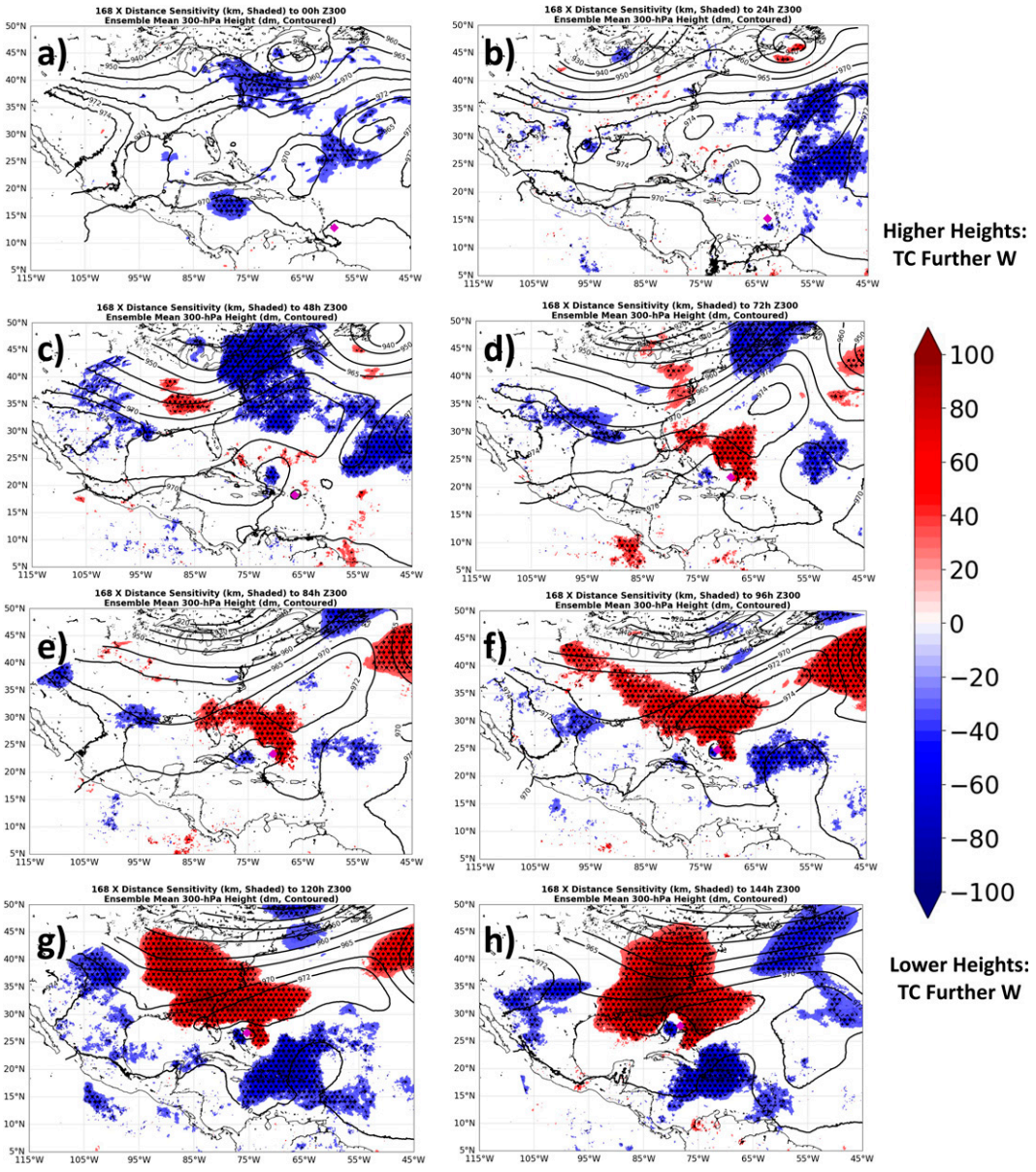

FIG. 11. As in Fig. 10, but for 300-hPa height.

as it was east of the Bahamas, allowing the TC to slow down and 2) the development of a stronger shortwave in the SE group that also allowed the TC to turn north as it later approached the Bahamas and Florida, despite the presence of some ridging to the north of the TC. These features will be explored in more detail in the next section by examining four representative ensemble members.

g. Individual member analysis

The analysis so far considered the ensemble as a whole, or separated into composite groups. To examine some of the speed and steering factors discussed above in more detail and to see whether the key factors are similar or different in individual members in comparison with the composites, several individual members are examined–one representative member from each group. Figure 13 shows the track and intensity evolution of these four individual members:

1) memNW from the NW group, which made landfall near Cape Canaveral and moved across Florida,
2) memNE from the NE group, which turned north of and began to recurve off of the coast of NE Florida,
3) memSE from the SE group, which stalled over Grand Bahama just like the observed TC, and
4) memSW from the SW group, which slowed down significantly but kept crawling west into the coast of Florida.

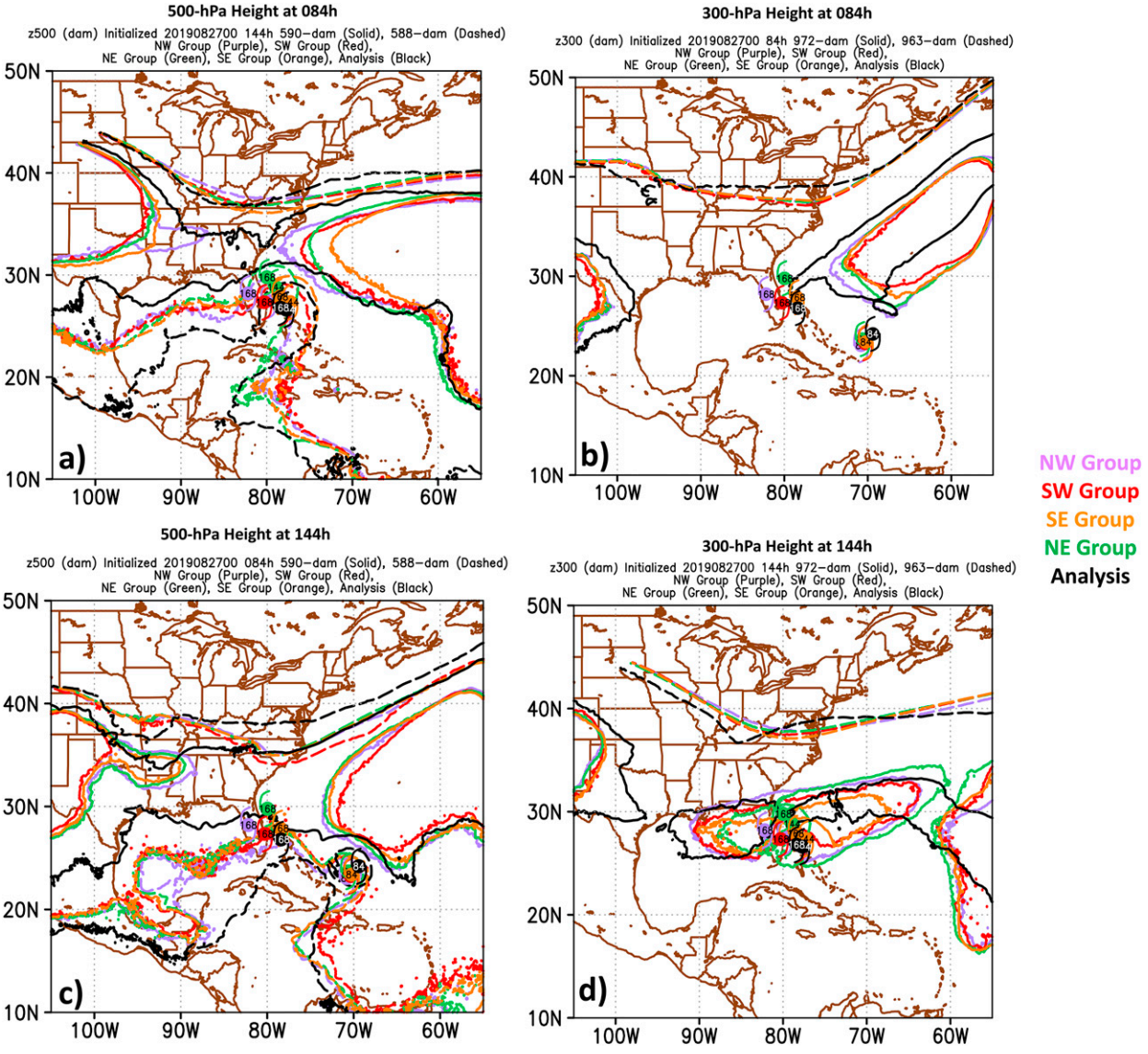

FIG. 12. (a) The 500-hPa height (590-dam contour solid and 588-dam contour dashed) for the NW group (purple), SW group (red), SE group (orange), and NE group (green) at 84 h, as well as the GFS analysis (black) valid at 1200 UTC 30 Aug 2019. The forecast (ensemble mean) and observed positions of Dorian are shown at 84 and 168 h with the forecast hour labels and the same colors. (b) As in (a), but for the 300-hPa height (972-dam contour solid and 963-dam contour dashed). (c),(d) As in (a) and (b), respectively, but for 144 h, valid at 0000 UTC 2 Sep 2019.

These four members were also of note because there were no major differences in the early tracks between them; that is, all four moved near or over NE Puerto Rico. There were slight intensity differences, with memNE being the strongest at 48 h, but the differences were not large. In fact, memSE was the weakest at this early point, suggesting that the early intensification of Dorian was not the major factor in its track in the SW Atlantic. MemSE was also the farthest west during the early period near Puerto Rico, but ended up slower and offshore of FL (close to the observed position of Dorian). The track farther north out of the Caribbean with a sharper bend back to the west in memSE was on the east side of the ensemble envelope, and was closest to the observed track. This is an illustration of the limited connection between the track in the Caribbean and the long-term track of Dorian (Figs. 2–4), and instead points to the importance of the steering flow in the western Atlantic.

The four members were compared with the means of each group to ensure that they were representative examples of each group. To begin examining the key differences between these four members, the speed of all members are compared with each other (and with the observed speed) throughout the forecast period (Fig. 14). The motion from the individual members is smoothed with a 12-h running mean due to the

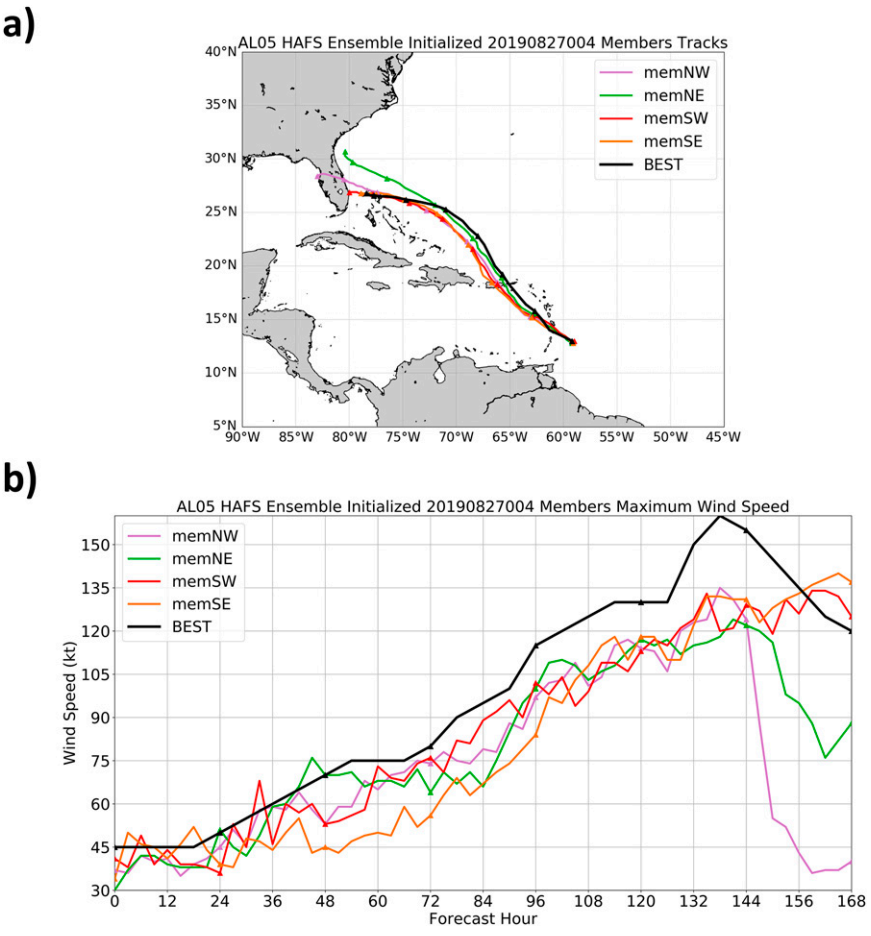

FIG. 13. (a) Track forecasts of the four individual members chosen for further analysis: memNW (purple), memSW (red), memSE (orange), memNE (green), and the observed track (black). Markers are placed every 24 h. (b) Similar to (a), but for intensity forecasts.

noisy nature of the short-term modeled motion. Around 84 h, memSW and memSE move slower to the west than memNE and memNW (Fig. 14a). A bit later, memNE begins to separate from the other members with faster northward motion (Fig. 14b), and there are clear speed differences between the southern and northern members from hours ~84 to 120 (Fig. 14c). Around hours 132–144, the divergence between memSE and memSW begins, specifically the zonal motion (and overall motion) slowing down more in memSE, putting it closer to the observed TC (and thereby remaining over the Bahamas rather than moving into Florida). The overall similarity of these motion patterns to the means further indicates that these members are representative of the four individual clusters and will provide excellent insight into the causes of Dorian's motion.

Next, we examine the synoptic patterns to see whether the four members shown here had a similar pattern as in Fig. 12 (i.e., the cluster means), and to examine the forcing in more detail. Figure 15 shows the 300- and 500-hPa height at 84 and 144 h for memNW, memSW, memSE, and memNE, and also for the GFS analysis. At 84 h (Fig. 15a), the four members were all close in position, indicating that the steering from this point onward was critical for the final location of the TC (memSW was actually slightly farther east than memSE by a small distance). As in the cluster means, the ridge to the NW of the TC was actually stronger in the analysis than all of the members (and closest to memNW). The key difference between the members was that in memSE and memNE at 84 h, the eastern flank of the ridge was more eroded. This can be seen primarily at 300 hPa (Fig. 15b), and contributed to the slower westward propagation during this part of the TC's life cycle. The difference is subtle, but memSE also has a stronger shortwave over Kansas at 500 hPa than the other four members (although not as strong as in the analysis).

At 144 h, the divergence of the members continued. The memNW was making landfall along the coast of FL, with a 500-hPa ridge that was too strong relative to observations leading to the faster westward motion into the coast (Fig. 15c). On the other hand, memNE had too strong of a ridge east of the TC, leading to the unrealistic northward acceleration. An interesting difference from 144 to 168 h was the divergence between memSE and memSW. The two were close in position at

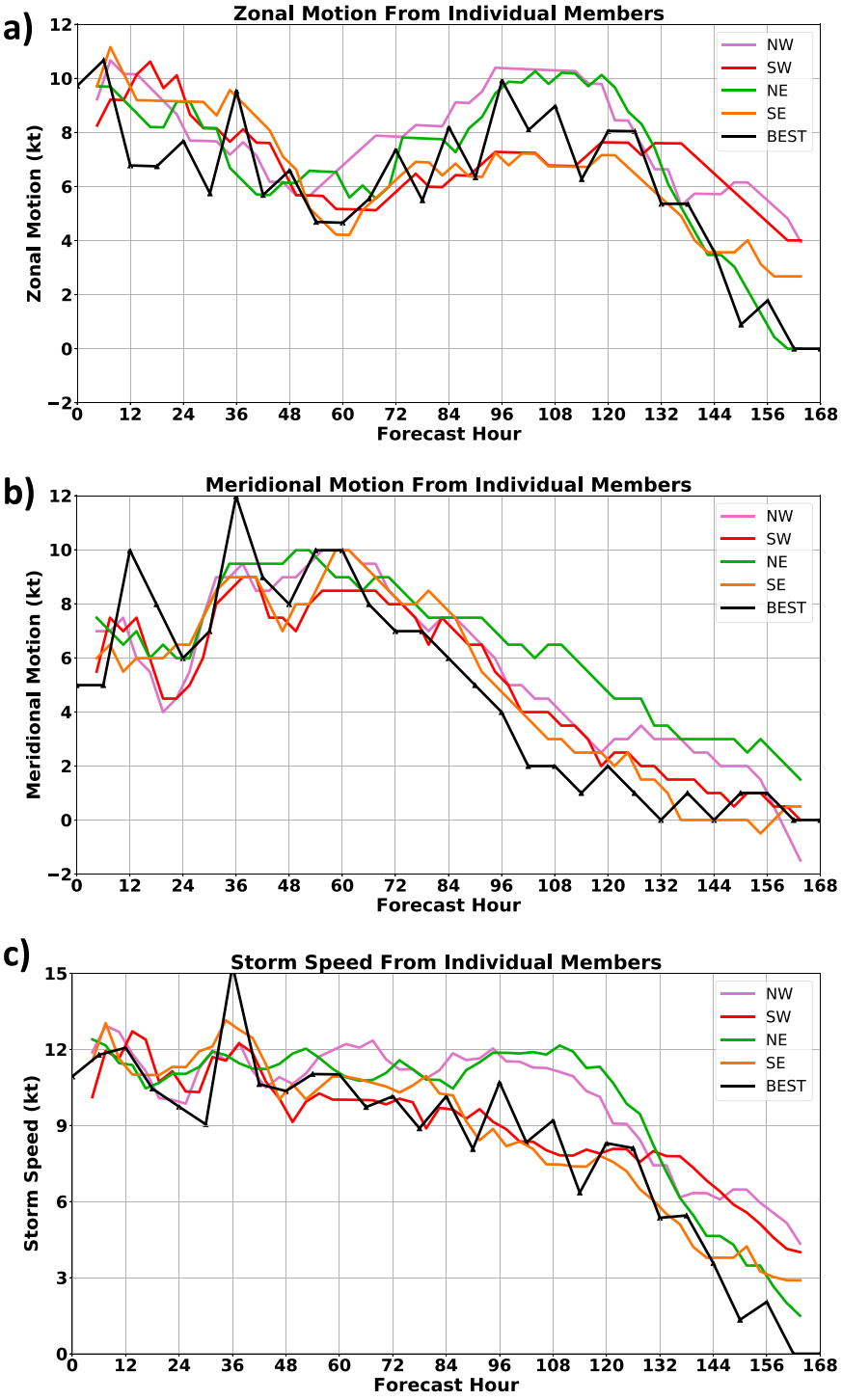

FIG. 14. (a) Zonal storm motion, (b) meridional storm motion, and (c) total storm speed from memNW (purple), memSW (red), memSE (orange), memNE (green), and the observed track (black). For (a) and (b), westward and northward motion are respectively shown as positive.

144 h, but memSW moved slightly faster west and approached the coast of Florida, while memSE moved a little bit slower and stalled out over Grand Bahama (as the observed TC did). The 500-hPa ridge NE of the TC was stronger in memSW, which likely explains the model difference. Interestingly, the analysis showed a stronger ridge than either of these members, but also a stronger 300-hPa ridge to the west of the TC that likely blocked it from moving into Florida (Fig. 15d). In addition, the

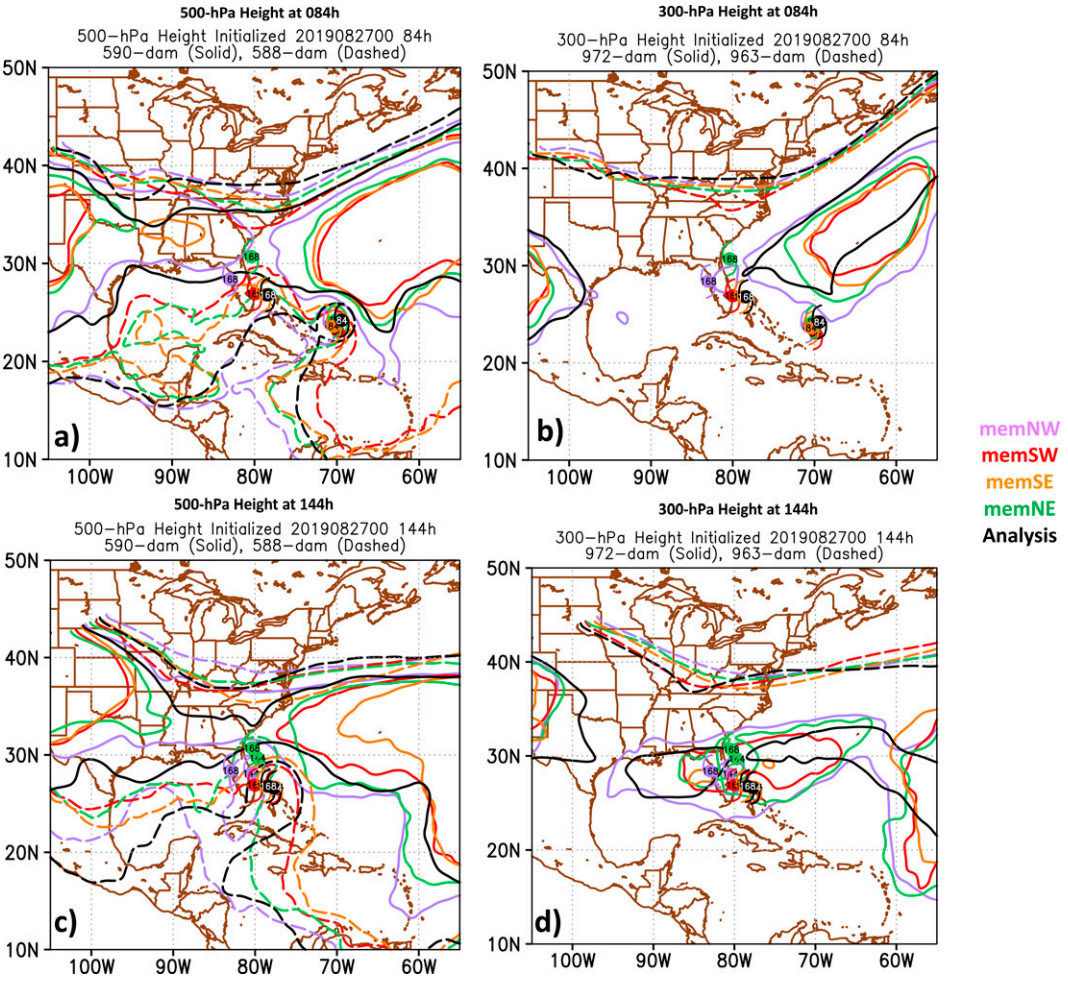

FIG. 15. (a) The 500-hPa height (590-dam contour solid and 588-dam contour dashed) for memNW (purple), memSW (red), memSE (orange), and memNE (green) at 84 h, as well as the GFS analysis (black) valid at 1200 UTC 30 Aug 2019. The forecast and observed positions of Dorian are shown at 84 and 168 h with the forecast hour labels and the same colors. (b) As in (a), but for the 300-hPa height (972-dam contour solid and 963-dam contour dashed). (c),(d) As in (a) and (b), respectively, but for 144 h, valid at 0000 UTC 2 Sep 2019.

300-hPa ridge to the NE of the TC was too strong in memSW. Also, the shortwave off the East Coast continued to be more pronounced in memSE than in memSW, and this weakness is what allowed the TC to stall and eventually move to the north.

4. Discussion and conclusions

An 80-member ensemble of HAFS forecasts of Hurricane Dorian, initialized while the TC was near the eastern Caribbean (early in its life cycle) is used to provide insight into the key large-scale features that were responsible for steering the TC, and also illuminated details of the track sensitivity in terms of key time periods and synoptic features. During the first day or two of the forecasts, the observed track was on the extreme right edge of the ensemble envelope, although the spread increased later and the track near the Bahamas was well within the ensemble suite. The early departure of most of the ensemble suite from the observed track indicates that even a relatively large ensemble of 80 members was not sufficient to fully capture the potential range of outcomes for Dorian, and an even larger ensemble (along with better DA) could potentially be necessary in cases like this to truly show the potential variability. Furthermore, improved initialization of smaller-scale features near the inner core (as discussed in Hazelton et al. 2021b) as well as large-scale steering factors may also help alleviate this issue. Somewhat counterintuitively based on real-time observations, there was little relationship between the early track of Dorian and its position near and Florida. In other words, it was not just the early and somewhat unexpected north-eastward reformation (Hazelton et al. 2021b; Alvey et al. 2022) in the Caribbean that led to the TC eventually remaining offshore of Florida. This was further illustrated by analysis of four individual representative ensemble members, where all four showed a similar initial track but then subsequently diverged over time.

A *k*-means clustering based on the ensemble positions at the end of the run (168 h) showed four distinct clusters, classified as

the NW group (which tended to move faster and make landfall in Florida, with several crossing the state), SW group (which were a bit slower but still approached and impacted the Florida coast), NE group (which stayed offshore but were too fast), and a SE group (which was slower, stayed offshore from Florida, and was the closest to the observed track). Analysis of these four clusters and the TC motion illustrates a couple of key points when the TC track diverged.

The first divergence point occurred around 84 h, where the NE and NW groups were moving faster zonally (to the west) than the SE and SW groups, and the NE group was also moving faster to the north. The key difference at hour 84 was the speed of the TC, and analysis of the mid and upper level ridges showed that the observed subtropical ridge had eroded on its eastern flank. The two slower groups correctly showed this erosion of the eastern part of the ridge, although they also broke down the western flank of the ridge too much. Another key divergence point occurred around 144 h when the SE and SW members began to diverge from each other. This later divergence point was mostly driven by the influence of a shortwave trough over the continental United States, as well as the subtropical ridge to the northeast of the TC starting to weaken (this was reflected better in the SE group than the SW group).

Ultimately, there were key synoptic differences that led to widely divergent track outcomes (including some that were very close to reality) despite a common bias of the western flank of the subtropical ridge being too weak (also noted in many other HAFS forecasts by Hazelton et al. 2021a), indicative of the sensitivity of the track in this case. This finding is similar to "saddle point" steering cases described by Torn et al. (2018) and illustrates how ensemble forecasts can be useful for forecasting and understanding physical processes despite imperfections of the model itself.

Overall, the three key factors that led to Dorian's path that stalled over the Bahamas, rather than Florida, thus appear to be

1) slower westward speed to the north of the Caribbean due to a weakening subtropical ridge,
2) slower westward speed due to the breakdown of the Atlantic ridge and development of a ridge over the Gulf of Mexico, and
3) a northward turn due to the shortwave trough that developed over the United States.

The results indicate that, in some cases, initial structure/intensity differences are not the sole or even key determinant of the long-term track outcome, and remote, synoptic features, such as the strength of the subtropical ridge or upstream troughs can dominate at longer lead times. This contrasts with other previous studies that suggest that near-storm steering flow variability is more critical to creating track variability (e.g., Torn et al. 2018). This makes it critical to take observations in the vicinity of these steering features and to use models that accurately depict the details of their evolution, especially for storms that could threaten land areas.

Note that this study only examines one case at one forecast time using one particular model, so there is a lot of variability that may not be captured even with 80 members perturbed with initial conditions. For other periods in Dorian's life cycle, and for other TC cases, there may have been other sensitivities that were more important. In addition, as the track outside the ensemble envelope early in the period showed, even this large ensemble variability was not totally sufficient to capture the observed track, perhaps due to some biases in the physics and how these affected the placement and strength of the subtropical ridge. Exploring such cases using a different ensemble technique could prove insightful in untangling such biases and providing an even more representative spread. Other future work will make use of new ensemble techniques (including new data assimilation capabilities as well as stochastic physics perturbations, such as those described in Ollinaho et al. 2017) in HAFS in order to improve understanding and prediction of both large-scale and storm-scale processes affecting tropical cyclones. We also plan to continue developing innovative techniques for analyzing and processing ensemble data, building off the artificial intelligence clustering analysis used in this study. As mentioned above, one important caveat in this study is that it is only a single case, and sensitivities for other TCs in similar locations might be different, especially within a less complex steering flow. However, the findings from this study indicate that ensemble techniques can be very valuable for understanding synoptic sensitivity in medium-to-long range forecasts. While some cases may be more sensitive to initial conditions of the TC, there are likely other cases like Dorian where it is the evolution of large-scale steering features ahead of the TC that play the biggest role in modulating the track. The method used in this study can be applied to both types of cases and can help to distinguish between different types of ensemble sensitivity. The tools used in this study, as well as newly developed tools for understanding ensemble forecasts, will be applied to other cases with track forecast uncertainty to help understand and predict future cases. As HAFS continues to develop, work is ongoing to better understand how model physics influence long-range forecasts. For example, the bias in the Atlantic subtropical ridge is something that appears in many HAFS forecasts for cases beyond Dorian (not shown), and we are in the process of examining the impact of PBL and cumulus physics on the strength of this key steering feature.

Acknowledgments. The authors thank Sarah Ditchek, Trey Alvey, and three anonymous reviewers for their helpful comments that led to improvements on an earlier version of this paper. Andrew Hazelton was supported by NOAA Grant NA19OAR0220187.

Data availability statement. Track and .grb2 data files from these ensemble runs are available on the NOAA RDHPCS computer system, or by request to the lead author (andrew.hazelton@noaa.gov).

REFERENCES

Alaka, G. J., Jr., X. Zhang, S. G. Gopalakrishnan, and Z. Zhang, 2018: Investigation of Irma & Maria track forecasts using an ensemble approach within basin-scale HWRF. *33rd Conf. on*

Hurricanes and Tropical Meteorology, Ponte Cedra, FL, Amer. Meteor. Soc., 1C.5, https://ams.confex.com/ams/33HURRICANE/webprogram/Paper339487.html.

——., ——, ——, ——, F. D. Marks, and R. Atlas, 2019: Track uncertainty in high-resolution HWRF ensemble forecasts of Hurricane Joaquin. *Wea. Forecasting*, **34**, 1889–1908, https://doi.org/10.1175/WAF-D-19-0028.1.

Alvey, G. R., III, M. Fischer, P. Reasor, J. Zawislak, and R. Rogers, 2022: Observed processes underlying the favorable vortex repositioning early in the development of Hurricane Dorian (2019). *Mon. Wea. Rev.*, **150**, 193–213, https://doi.org/10.1175/MWR-D-21-0069.1.

Ancell, B., and G. J. Hakim, 2007: Comparing adjoint- and ensemble-sensitivity analysis with applications to observation targeting. *Mon. Wea. Rev.*, **135**, 4117–4134, https://doi.org/10.1175/2007MWR1904.1.

Chan, J. C. L., 2005: The physics of tropical cyclone motion. *Annu. Rev. Fluid Mech.*, **37**, 99–128, https://doi.org/10.1146/annurev.fluid.37.061903.175702.

Evans, J. L., J. M. Arnott, and F. Chiaromonte, 2006: Evaluation of operational model cyclone structure forecasts during extratropical transition. *Mon. Wea. Rev.*, **134**, 3054–3072, https://doi.org/10.1175/MWR3236.1.

Galarneau, T. J., Jr., and C. A. Davis, 2013: Diagnosing forecast errors in tropical cyclone motion. *Mon. Wea. Rev.*, **141**, 405–430, https://doi.org/10.1175/MWR-D-12-00071.1.

George, J. E., and W. M. Gray, 1976: Tropical cyclone motion and surrounding parameter relationships. *J. Appl. Meteor. Climatol.*, **15**, 1252–1264, https://doi.org/10.1175/1520-0450(1976)015<1252:TCMASP>2.0.CO;2.

Hamill, T. M., J. S. Whitaker, M. Fiorino, and S. G. Benjamin, 2011: Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter. *Mon. Wea. Rev.*, **139**, 668–688, https://doi.org/10.1175/2010MWR3456.1.

——, M. J. Brennan, B. Brown, M. DeMaria, E. N. Rappaport, and Z. Toth, 2012: NOAA's future ensemble-based hurricane forecast products. *Bull. Amer. Meteor. Soc.*, **93**, 209–220, https://doi.org/10.1175/2011BAMS3106.1.

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.

Hazelton, A. T., and Coauthors, 2021a: 2019 Atlantic hurricane forecasts from the global-nested hurricane analysis and forecast system: Composite statistics and key events. *Wea. Forecasting*, **36**, 519–538, https://doi.org/10.1175/WAF-D-20-0044.1.

——, G. J. Alaka, L. Cowan, M. Fischer, and S. Gopalakrishnan, 2021b: Understanding the processes causing the early intensification of Hurricane Dorian through an ensemble of the Hurricane Analysis and Forecast System (HAFS). *Atmosphere*, **12**, 93, https://doi.org/10.3390/atmos12010093.

——, Coauthors, 2022: Performance of 2020 real-time Atlantic hurricane forecasts from high-resolution global-nested hurricane models: HAFS-globalnest and GFDL T-SHiELD. *Wea. Forecasting*, **37**, 143–161, https://doi.org/10.1175/WAF-D-21-0102.1.

Kowaleski, A. M., and J. L. Evans, 2016: Regression mixture model clustering of multimodel ensemble forecasts of Hurricane Sandy: Partition characteristics. *Mon. Wea. Rev.*, **144**, 3825–3846, https://doi.org/10.1175/MWR-D-16-0099.1.

——, and ——, 2020: Use of multiensemble track clustering to inform medium-range tropical cyclone forecasts. *Wea. Forecasting*, **35**, 1407–1426, https://doi.org/10.1175/WAF-D-20-0003.1.

Landsea, C. W., and J. P. Cangialosi, 2018: Have we reached the limits of predictability for tropical cyclone track forecasting? *Bull. Amer. Meteor. Soc.*, **99**, 2237–2243, https://doi.org/10.1175/BAMS-D-17-0136.1.

Munsell, E. B., and F. Zhang, 2014: Prediction and uncertainty of Hurricane Sandy (2012) explored through a real-time cloud-permitting ensemble analysis and forecast system assimilating airborne Doppler radar observations. *J. Adv. Model. Earth Syst.*, **6**, 38–58, https://doi.org/10.1002/2013MS000297.

Nystrom, R. G., F. Zhang, E. B. Munsell, S. A. Braun, J. A. Sippel, Y. Weng, and K. Emanuel, 2018: Predictability and dynamics of Hurricane Joaquin (2015) explored through convection-permitting ensemble sensitivity experiments. *J. Atmos. Sci.*, **75**, 401–424, https://doi.org/10.1175/JAS-D-17-0137.1.

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.

Rousseeuw, P. J., 1987: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.*, **20**, 53–65, https://doi.org/10.1016/0377-0427(87)90125-7.

Torn, R. D., and G. J. Hakim, 2008: Ensemble-based sensitivity analysis. *Mon. Wea. Rev.*, **136**, 663–677, https://doi.org/10.1175/2007MWR2132.1.

——, J. S. Whitaker, P. Pegion, T. M. Hamill, and G. J. Hakim, 2015: Diagnosis of the source of GFS medium-range track errors in Hurricane Sandy (2012). *Mon. Wea. Rev.*, **143**, 132–152, https://doi.org/10.1175/MWR-D-14-00086.1.

——, T. J. Elless, P. P. Papin, and C. A. Davis, 2018: Tropical cyclone track sensitivity in deformation steering flow. *Mon. Wea. Rev.*, **146**, 3183–3201, https://doi.org/10.1175/MWR-D-18-0153.1.

Velden, C. S., and L. M. Leslie, 1991: The basic relationship between tropical cyclone intensity and the depth of the environmental steering layer in the Australian region. *Wea. Forecasting*, **6**, 244–253, https://doi.org/10.1175/1520-0434(1991)006<0244:TBRBTC>2.0.CO;2.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 100, Academic Press, 648 pp.

Wu, H.-C., M. K. Lindell, and C. S. Prater, 2012: Logistics of hurricane evacuation in Hurricanes Katrina and Rita. *Transp. Res.*, **15**, 445–461, https://doi.org/10.1016/j.trf.2012.03.005.

Zhang, Z., and T. N. Krishnamurti, 1997: Ensemble forecasting of hurricane tracks. *Bull. Amer. Meteor. Soc.*, **78**, 2785–2796, https://doi.org/10.1175/1520-0477(1997)078<2785:EFOHT>2.0.CO;2.

Zhou, X., Y. Zhu, D. Hou, Y. Luo, J. Peng, and R. Wobus, 2017: Performance of the new NCEP global ensemble forecast system in a parallel experiment. *Wea. Forecasting*, **32**, 1989–2004, https://doi.org/10.1175/WAF-D-17-0023.1.