An Investigation of Large Along-Track Errors in Extratropical Transitioning North Atlantic Tropical Cyclones in the ECMWF Ensemble

NICHOLAS M. LEONARDO AND BRIAN A. COLLE

School of Marine and Atmospheric Sciences, Stony Brook University, State University of New York, Stony Brook, New York

ABSTRACT

The synoptic evolution and mechanisms for the largest medium-range (72–120 h) along-track errors of tropical cyclones (TC) are investigated. The mean along-track errors (ATEs) of the 51-member European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble are evaluated for 393 forecasts (85 TCs) during the 2008 to 2016 North Atlantic seasons. The 27 unique forecasts within the upper quintile of most negative ATEs (i.e., slow bias greater than 500 km by 72 h) are inherently fast-moving TCs that undergo extratropical transition as they recurve and interact with a 300-hPa upstream trough and a downstream ridge. Both the trough and ridge are underamplified by only 5–10 m ~60 h before the time of largest ATE. The height errors then grow rapidly due to underpredicted 300–200-hPa potential vorticity advection by both the nondivergent wind and the irrotational wind from the TC's outflow. Both wind components are underpredicted and result in weak biases in the trough's developing potential vorticity gradient and associated jet streak. The underamplification of the upstream trough is exacerbated by underpredicted 700-hPa cold advection extending from beneath the trough into the TC at 48–36 h before the largest ATE. Standardized differences are consistent with the mean errors and reveal that weaker divergent outflow is driven by underpredicted near-TC precipitation, which corresponds to underpredicted 700-hPa moisture fluxes near the TC at ~108 h before the largest ATE. The ensemble member ATEs at 72–120 h generally show little correlation with their ATEs before 36 h, suggesting that initial position uncertainty is not the primary source of ATE variability later in the forecast.

1. Introduction

a. Background

Tropical cyclones (TCs) are among the costliest natural disasters worldwide (Wirtz et al. 2014) due to their damaging winds, storm surge, and inland flooding. Emergency preparations to an approaching TC, such as the placement of evacuation zones and relief supplies, are often commenced before the National Hurricane Center (NHC) issues a watch or warning (48 and 36 h before the expected arrival of tropical storm force winds, respectively). Thus, skillful forecasts of TCs at lead times of at least 48–72 h are essential, such that NHC issues official forecasts out to 120 h (Cangialosi 2018).

Leonardo and Colle (2017) verified the 72–120 h track forecasts for North Atlantic TCs during the 2008–15 period and found that many numerical models, including the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble mean, had a slow bias. While the 72-h track errors improved by up to 36% during the sample period, the models continued to struggle forecasting certain TCs, such as Cristobal in 2014 and Joaquin in 2015, resulting in individual forecasts with track errors much larger than climatology. Understanding the common causes of anomalous track errors can aid forecasters in recognizing problematic patterns or features that may produce biases in the models. Such biases may be ameliorated by more extensive data assimilation (e.g., Brennan et al. 2015; Nystrom et al. 2018) and improved representation of physical processes (Torn and Davis 2012; Bassill 2014).

There have been few attempts to quantify and compare the mechanisms that cause anomalous track errors among multiple TCs. Carr and Elsberry (2000) qualitatively examined the 72-h forecasts by the U.S. Navy version of the Geophysical Fluid Dynamics Laboratory model (GFDN) and the Navy Operational Global Atmospheric Prediction System (NOGAPS) models for

Corresponding author: Dr. Brian A. Colle, brian.colle@stonybrook.edu

the western North Pacific in 1997. They focused on forecasts with track errors larger than 555 km and divided the cases (i.e., forecasts) based on whether the TC was in the tropics or interacting with midlatitude systems. For midlatitude TCs, 19 of the 38 large error cases forecasted by GFDN overdeepened baroclinic systems north of the TC, while 10 of the 37 NOGAPS TCs were too shallow to be correctly steered by the flow aloft. Kehoe et al. (2007) analyzed the same two models during the 2004 western North Pacific season and found that midlatitude influences caused at least 83% of the largest track errors, with insufficient deepening or excessive weakening of troughs being the most common mechanisms identified. A more recent study by Peng et al. (2017) evaluated official forecasts from five operational centers for the 2004 through 2015 western North Pacific seasons. They found that TCs entering midlatitudes near Japan after 48 h into the forecast had a slow along-track bias greater than 200 km on average. They concluded that 26 out of the 56 largest track error cases involved the presence of an upstream midlatitude trough, an adjacent upstream cyclonic circulation, and a downstream anticyclone.

Other studies have used ensembles composed of 20 or more members to quantify meteorological factors associated with the track errors of particular cases. For example, Munsell and Zhang (2014) ran a Weather Research and Forecasting (WRF) Model (Skamarock et al. 2008) ensemble to simulate Hurricane Sandy (2012) and showed that differences in the initial midlatitude environment had a much smaller impact on the 96-h track compared to differences in the initial near-TC steering flow. Torn et al. (2015) analyzed an experimental Global Forecast System (GFS) ensemble for Sandy (2012) and found that the initial near-TC 450-hPa specific humidity for the more incorrect eastward-moving members was smaller than for the westward members. The incorrect members had less latent heat released and weaker negative potential vorticity (PV) advection aloft, resulting in less amplification of the synoptic ridge north of the TC. Meanwhile, Munsell et al. (2015) ran a WRF ensemble to simulate Hurricane Nadine (2012). The 300–200-hPa steering trough was ~200 km farther west in the poor members compared to the better members by 30 h, with the difference likely associated with the initial upper-level westerlies advecting the trough.

Both Sandy (2012) and Nadine (2012) interacted with a baroclinic environment at higher latitudes and underwent extratropical transition (ET; Jones et al. 2003; Evans et al. 2017; Keller et al. 2019). During this process, the TC's warm-core structure usually becomes shallow and is then replaced by a cold-core asymmetric

structure (e.g., Evans and Hart 2003; Hart et al. 2006), which often includes surface fronts (Klein et al. 2000). In many ET cases, the TC's upper-tropospheric divergent outflow impinges on a large PV gradient associated with a midlatitude jet (Riemer and Jones 2010; Grams et al. 2013; Archambault et al. 2013, 2015), which can affect the downstream transfer of Rossby wave packet energy (Riemer et al. 2008; Archambault et al. 2015; Keller 2017). Underestimating this interaction can then result in an underamplified flow that does not properly accelerate the TC (Carr and Elsberry 2000).

b. Motivation

There have been attempts to understand the common mechanisms associated with abnormally large track errors. While Carr and Elsberry (2000), Kehoe et al. (2007), and Peng et al. (2017) analyzed large samples of midlatitude TCs, they focused only on the western North Pacific basin. Their results may not be representative of the North Atlantic basin, which has a different climatology of ET events (Bieli et al. 2019). All but the Peng et al. (2017) study focused on models that are at least 15 years old, such that the extent to which their results apply to the present is questionable. By comparison, studies such as Munsell and Zhang (2014) and Torn et al. 2015 have used ensembles to quantitatively diagnose sources of track errors in the North Atlantic, but only analyzed one or two TCs. It is unclear how frequently the ensemble tracks of large error cases are sensitive to differences in the near-TC environment and synoptic steering features.

Our paper focuses on the causes of largest negative along-track errors (i.e., slow biases) during the medium-range (72 to 120 h) for the 2008 through 2016 North Atlantic seasons. Along-track errors are important in that they can affect the amount of time that people expect to have for emergency preparations. Slow biases can thus correspond to the actual TC arriving sooner than people prepared for. Along-track errors can also determine the tide at which the TC makes landfall, thereby affecting storm surge prediction. A separate paper will focus on the cross-track TC errors at lower latitudes. Different ensemble verification metrics, such as those used by Torn et al. (2015), are used to help answer the following questions:

- What is the geographic distribution of slow-biased cases over the North Atlantic and does it differ from non-slow-biased cases?
- How do the along-track errors of these cases typically grow with time?
- What are the common synoptic features or mechanisms associated with large slow-biased cases in comparison

to the other cases? How do the model errors attached to these features develop over time?

The data and methodology used in this study are described in section 2. Section 3 shows the climatology of the largest track error cases identified and their relationship with track errors at shorter lead times. The most common synoptic-scale patterns related to the track errors in these cases are examined in section 4. The feedback between the TC and the synoptic flow through convection is assessed in section 5. Section 6 contains a summary and future work.

2. Data and methods

a. Track data and definition of large error events

This study diagnoses 72–120-h TC track forecasts with anomalously large slow biases during the 2008 to 2016 North Atlantic seasons. The focus is on the ECMWF ensemble (Buizza et al. 2007), which is composed of 51 members (50 perturbations and one control) and was shown by Leonardo and Colle (2017) to have a 150–250 km mean slow bias at 120 h during the 2008–15 period. Between 2008 and 2016, the perturbations were generated through a combination of singular vectors, differences between the members of an ensemble of data assimilations, and (since 2009) stochastic physics and backscatter methods. The horizontal resolution was ~50, ~32, and ~18 km in 2008, 2010, and 2016, respectively, and the number of vertical levels increased from 62 to 91 in 2013. The ECMWF cyclone tracks are archived by the THORPEX Interactive Grand Global Ensemble (TIGGE; Bougeault et al. 2010) database and are available online through the National Center for Atmospheric Research (NCAR; http://rda.ucar.edu/datasets/ds330.3/). The tracks are verified against the NHC best track data, which is archived by NHC (ftp://ftp.nhc.noaa.gov/atcf/archive/).

Track error is defined as the great circle distance between the model TC and best track TC positions. Ensemble mean track error is thus the distance between the ensemble mean of all the member TC positions and the best track TC position. To perform the diagnostics described in section 2b, a forecast is only included if at least 20 of the ensemble members have tracker data available after 72 h. In cases where at least 20 members are available at 72 h, all later times with fewer than 20 members are excluded. The total track error (TTE) is decomposed into along-track (ATE) and cross-track (CTE) errors relative to the motion of the best track (See Fig. 1 in Leonardo and Colle 2017). By this convention, a positive (negative) ATE corresponds to a forecast TC that is too fast (slow) relative to the observed TC.
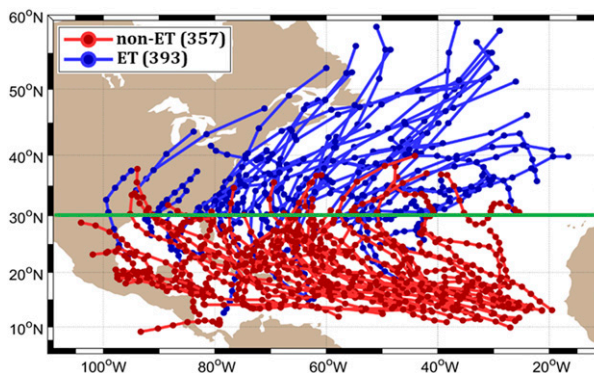

FIG. 1. Best tracks of ET (blue) vs non-ET cases (red). The number of cases are given in parentheses.

Similarly, a positive (negative) CTE corresponds to a forecast TC that is to the right (left) of the observed TC.

To isolate TCs undergoing ET due to interactions with midlatitude baroclinic systems, the forecasts are screened based on the northernmost latitude of the verifying best track and the three cyclone phase space (CPS; Hart 2003) parameters, which are estimated using the $0.5° × 0.5°$ Climate Forecast System Reanalysis (CFSR; Saha et al. 2010) available online through NCAR (https://rda.ucar.edu/datasets/ds093.0/). A TC is considered "ET" if both of the following conditions are met at any time in the forecast: 1) at least one of the three CPS parameters are consistent with the observed TC being extratropical, and 2) the best track crosses 30°N. All other cases are considered "non-ET."

The first CPS parameter B represents the thermal asymmetry of the TC within the 900–600-hPa layer and typically signifies the onset of ET upon reaching or exceeding a value of 10 (Hart 2001). The other parameters, $-V_T^L$ and $-V_T^U$, are proportional to the thermal wind in the 900–600-hPa and the 600–300-hPa layers, respectively. Negative values in $-V_T^L$ and $-V_T^U$ imply that winds are increasing with height within the lower and upper layers, respectively, thereby suggesting the transition to cold-core extratropical systems. Thus, the first condition for an ET case is met if $B \geq 10$, $-V_T^L < 0$, or $-V_T^U < 0$ at any point in the forecast.

The second condition based on latitude is chosen to better isolate TCs transitioning in response to midlatitude systems. The threshold of 30°N is up to 5° south of the lower quartile of latitudes at which North Atlantic TCs typically complete ET (Hart and Evans 2001; Bieli et al. 2019) and is chosen to obtain a slightly larger dataset. Figure 1 shows the verifying best tracks of the resulting 393 ET and 357 non-ET forecasts that will be compared in this study.

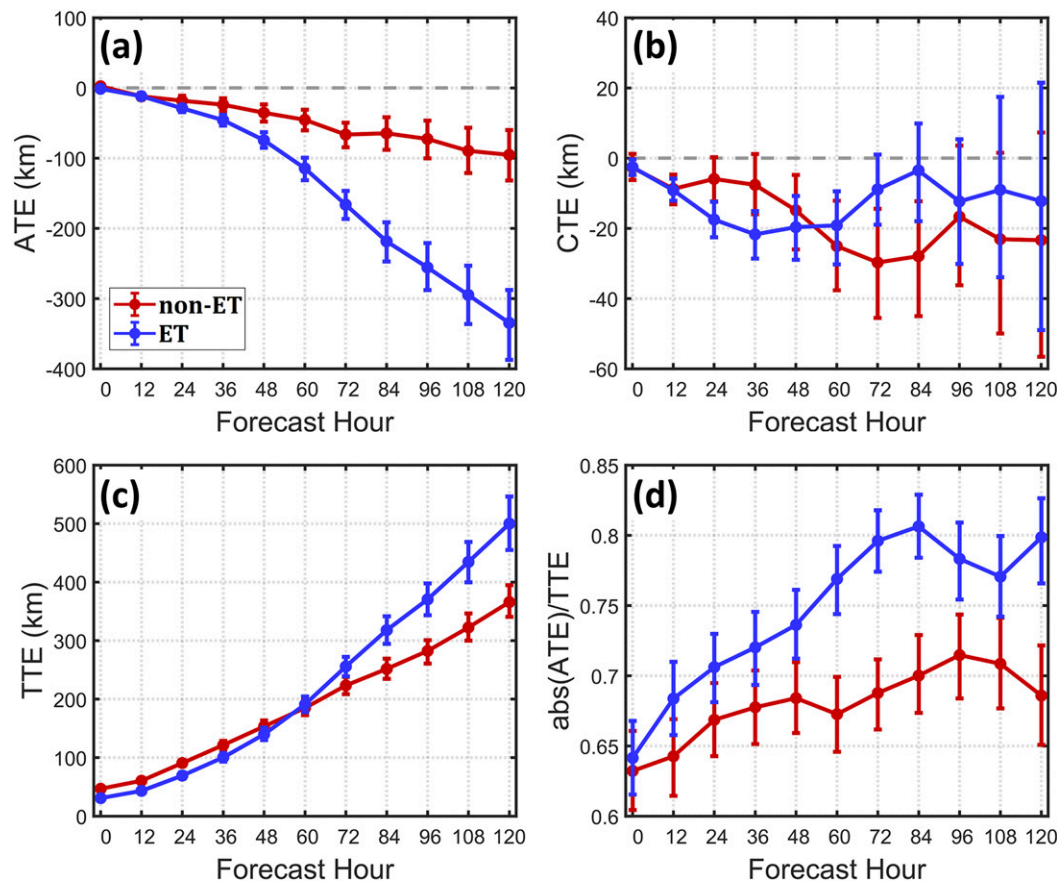Figure 2 shows the ECMWF ensemble mean ATE, CTE, and TTE, averaged as a function of forecast hour.

FIG. 2. ECMWF mean (a) ATEs, (b) CTEs, (c) TTEs, and (d) the ratios of absolute ATE to TTE for ET and non-ET cases averaged as a function of lead time. Error bars indicate the 95% confidence interval using bootstrap resampling without replacement.

The error bars in these and other time series represent the 95% confidence intervals of the sample averages. They are calculated using a bootstrap method (Zwiers 1990), in which values are resampled from the original dataset 1000 times. The ATEs illustrate that the ECMWF underpredicts the forward speed of ET cases significantly more than of the non-ET cases after 36 h (Fig. 2a). By comparison, there are only marginally negative CTEs (left-of-track biases) of less than 50 km magnitude for both ET and non-ET cases (Fig. 2b). The TTEs for the ET cases are indeed larger than non-ET cases by 72 h (Fig. 2c).

The contribution of the along-track component is estimated by the ratio of the magnitude of ATE to TTE (Fig. 2d). By 72 h, 77%–82% of the TTE in the ET cases is from the magnitude of the ATE, versus 65%–70% in non-ET cases. Hence, most of the track error in ET cases is in the along-track component, which on average is significantly negative (slow-biased). The remainder of this paper will therefore focus on the uniquely large slow biases of ET cases, determining how they may be related to TC interactions with baroclinic systems.

The largest ensemble-mean ATE between 72 and 120 h is found for each forecast in which the TC undergoes ET. The distribution of these ATEs is skewed toward negative values (Fig. 3), with only ~50 out of the 393 forecasts having ATEs that are positive and larger than 100 km. The lower (slowest) and upper (fastest) quintiles of the distribution are compiled and called "lower 20" (L20) and "upper 20%" (U20) cases, respectively. The quintiles are chosen as a compromise between maintaining adequate sample sizes in each subset and ensuring that the ATEs in the L20 cases are at least 500 km in magnitude. The L20 cases comprise the tail of significantly negative ATEs in Fig. 3 and are of greater interest in this study compared to the small ATEs of the U20.

The largest ATEs from each ET forecast were averaged as a function of year to assess any trends throughout the sample period (not shown). The slow bias of ET cases overall improves by up to ~200 km between 2008 and 2016. The percentage of ET cases per year that are within the L20 decreases from 50% to ~10% between 2009 and 2016. However, the number of
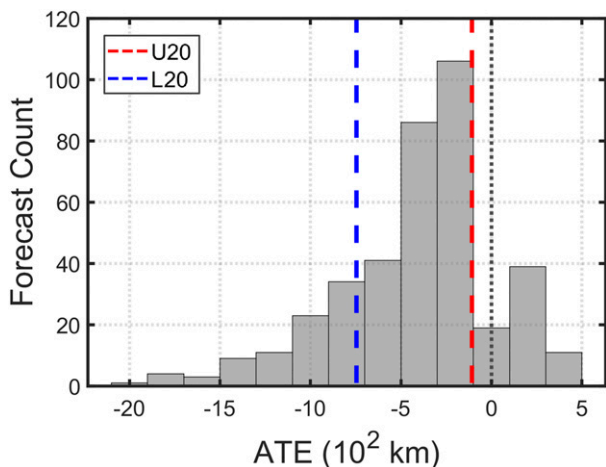
FIG. 3. Histogram of the largest day 3–5 ECMWF mean ATEs, with blue and red lines showing the thresholds for the L20 and U20 cases.

ET cases per year varies between 12 in 2009 and 96 in 2012, making it difficult to assess statistical significance. Nevertheless, it is hypothesized that the overall improvement in both the average slow bias and percentage of forecasts with track errors larger than 500 km corresponds to the major upgrades in the ECMWF.

It is important to note that many of the forecasts in either the L20 or U20 subsets are for the same TCs and are from successive initializations of the ECMWF. The contribution of these potentially autocorrelated forecasts in a subset is minimized in the following manner. For each of the L20 TCs, the most negative ATE forecast within the 72–120 h period is included. The second most negative ATE is also included if it was initialized at least 48 h before or after the first. The third most negative ATE is also included if initialized at least 48 h before or after both the first and second, and so on for the next. The procedure is repeated for the U20 cases but considering the most positive ATEs. The result is 27 L20 cases and 34 U20 cases.

Analogous to Archambault et al. (2015), the cases are compared in a "$T - X$" hour framework. In our paper, the forecast hour of largest ensemble-mean ATE serves as the reference time ("$T - 0$ h") to compare the prior evolutions of different cases. $T - 0$ h can occur at any forecast hour between 72 and 120 h. While the number of cases is therefore limited from $T - 120$ h to $T - 72$ h, at least 20 of the 27 L20 cases are available by $T - 108$ h. The various time series and composites later shown are averaged in this temporal framework.

b. Meteorological fields and compositing techniques

The ECMWF forecasts are archived at 0.5° resolution and are available online through TIGGE (https://apps.ecmwf.int/datasets/data/tigge/levtype=sfc/type=cf/).

The forecast fields for the identified cases are analyzed using various ensemble verification metrics. These metrics are composited using two different approaches. The first ("best track–relative") approach simply takes a grid centered on the best track position in both the ECMWF and CFSR fields of a case. Thus, for each individual case and time, the ensemble member and CFSR fields are compared at matching geographical (latitude/longitude) points. The second ("TC-relative") approach more closely follows Zhang and Colle (2017) and takes the grids of the ECMWF members each centered on their respective TC positions, compared to the reanalysis centered on the best track position. Hence, the grid points in the individual ensemble member and reanalysis fields are no longer necessarily at matching latitudes and longitudes but are at matching positions (in degrees longitude and latitude) relative to their respective TC centers.

Figure 4 shows an example of the errors (model–reanalysis) given by the best track– and TC-relative frameworks for a single case at $T - 0$ h. For the best track–relative approach (Fig. 4a), there are positive biases in the 300–200-hPa divergent wind speeds collocated with the model TC positions south of the best track position. The TC-relative comparison is achieved in Fig. 4b by superimposing the orange boxes around each of the ECMWF TC positions (only two of the 51 members are shown for visual clarity) over the black box around the best track position in Fig. 4a. Comparing Fig. 4a with Fig. 4b, the positive divergent wind speed biases near the model TC in the best track–relative framework are not present in the TC-relative framework. These positive biases in the best track–relative framework are therefore more likely a consequence of the southward position bias of the TC, as opposed to an overprediction of the TC outflow. However, the apparent northward bias of the model's 300-hPa trough (e.g., the 930-dam height contour) in the TC-relative framework is partially caused by the model TC being too far south, as opposed to a large geographical location error in the trough itself. Therefore, the best track–relative framework captures the differences in the large-scale environment (e.g., synoptic features moving independently of the TC), while the TC-relative framework better captures differences in the structure of the TC. Both approaches will thus be used in this study.

Two different metrics are calculated and composited using the two aforementioned frameworks. For the first metric, the mean error in a field for any one case is calculated by subtracting the CFSR from the ECMWF ensemble mean. Given that each reanalysis has biases, other datasets were tested to verify the ECMWF fields. However, the results were similar (not shown), such that the errors that will be shown are larger than the differences
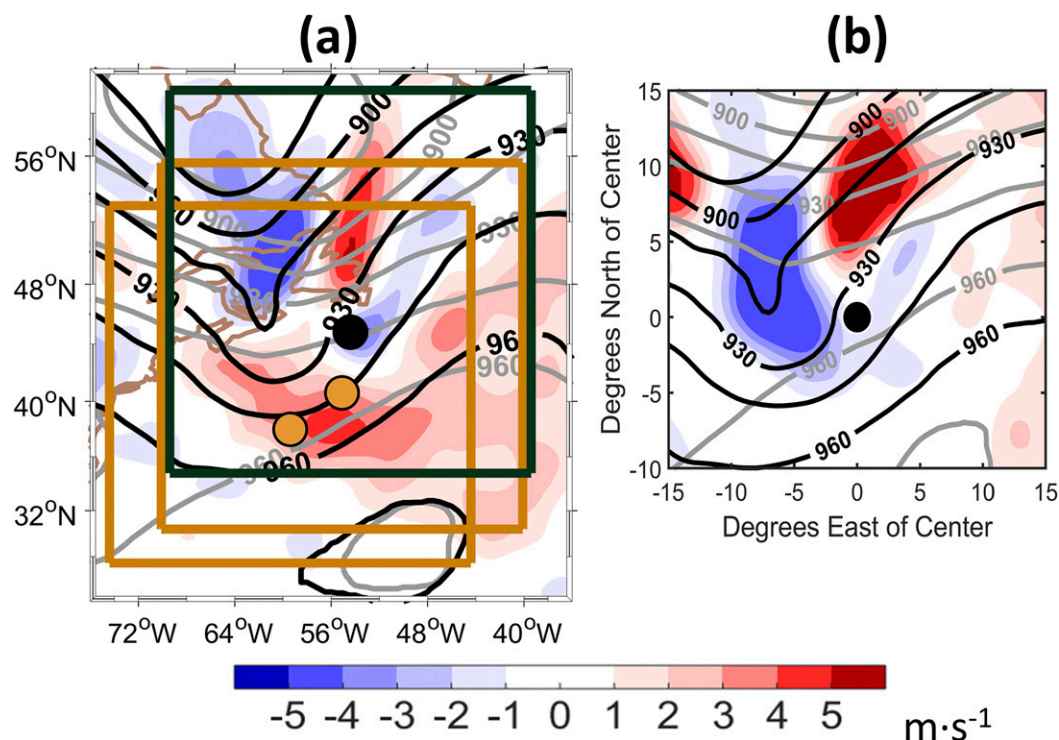
FIG. 4. ECMWF mean 200–300-hPa divergent wind speed errors (shaded) and 300-hPa geopotential heights of the CFSR (black contours in dam) and ECMWF (gray contours) at forecast hour 120 ($T - 0$ h before the time of largest ATE) from the 0000 UTC 18 Sep 2010 forecast of TC Igor. The black and orange dots in (a) indicate the best track and two of the ECMWF member TC positions, respectively. The boxes surrounding the TC positions in (a) are the grids used for the TC-relative differences shown in (b).

between reanalyses. Bootstrap resampling is used to determine whether the mean error composited from multiple cases is statistically different from zero. Specifically, the 95% confidence bounds are created by randomly resampling the cases 1000 times, each time extracting 20 ensemble members from each case at random.

The second metric subsets the 10 "slowest" (most negative) and 10 "fastest" (most positive) ensemble members of each case based on their individual ATEs at T–0 h. Note that almost all ensemble members in the L20 cases have negative ATEs (not shown), such that the 10 fastest members are usually still slow biased, but most closely match the best track. Comparing the differences with mean errors confirms which features are influencing the TC track and that these features behave more like the reanalysis in the fastest members. Following Torn et al. (2015), the standardized differences between these two subgroups are calculated as follows:

$$\Delta x_i = \frac{\overline{x}_i^{\text{Slowest}} - \overline{x}_i^{\text{Fastest}}}{\sigma_{x_i}}, \quad (1)$$

where $\overline{x}_i^{\text{Slowest}}$ ($\overline{x}_i^{\text{Fastest}}$) denotes the mean of the ith state variable for the slowest (fastest) ensemble members and

σ_{x_i} is the ensemble standard deviation of x_i computed from all members. The normalization by ensemble spread allows for comparisons between different fields, vertical levels, and times. To assess the statistical significance of Δx_i for a single case, two subsets of 10 ensemble members are randomly drawn from the full ensemble 1000 times. Each time, the difference of the two means is calculated, thereby giving the 95% confidence bounds on Δx_i for a single case.

3. Climatology of TCs with large along-track errors

The inherent differences in the geographical locations and trajectories of the L20 and U20 best tracks are illustrated in Fig. 5. Most of the L20 are associated with TCs that recurve and accelerate to the northeast at midlatitudes (Fig. 5a). In contrast, most of the U20 TCs on average travel total distances of less than 3000 km throughout the forecasts and stay south of 45°N (Fig. 5b)

Figure 6 shows the ensemble mean of various metrics averaged as a function of lead time before largest mean ATE. The ATEs (Fig. 6a) for the L20 cases first become significantly different from the U20 cases by $\sim T - 72$ h. Afterward, the L20 ATEs grow exponentially more
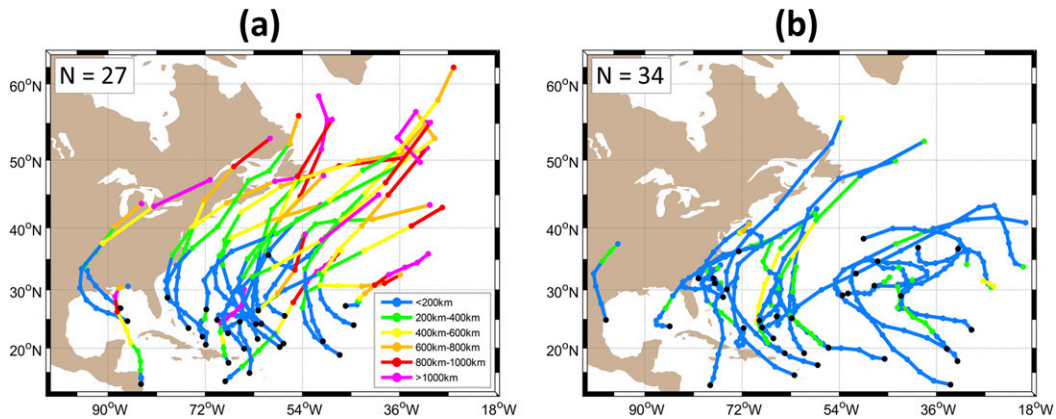
FIG. 5. Best tracks of (a) L20 and (b) U20 cases color-coded by ATE. The positions at the forecast initializations are given by the black dots.

negative (slow-biased), while the U20 ATEs only become significantly positive after $\sim T - 24$ h. The forward speeds of the model TCs are also calculated and verified against those of the best tracks. Comparing the best track forward speeds of the L20 and U20 cases (Fig. 6b),

the L20 cases are up to $4\,\mathrm{m\,s^{-1}}$ faster than the U20 at $\sim T - 48$ h, which is significant at the 95% level. The L20 cases continue to accelerate afterward, their averaged observed forward speeds reaching $15\,\mathrm{m\,s^{-1}}$ by $T - 12$ h, compared to $\sim 6\,\mathrm{m\,s^{-1}}$ for the U20 cases. Comparing the
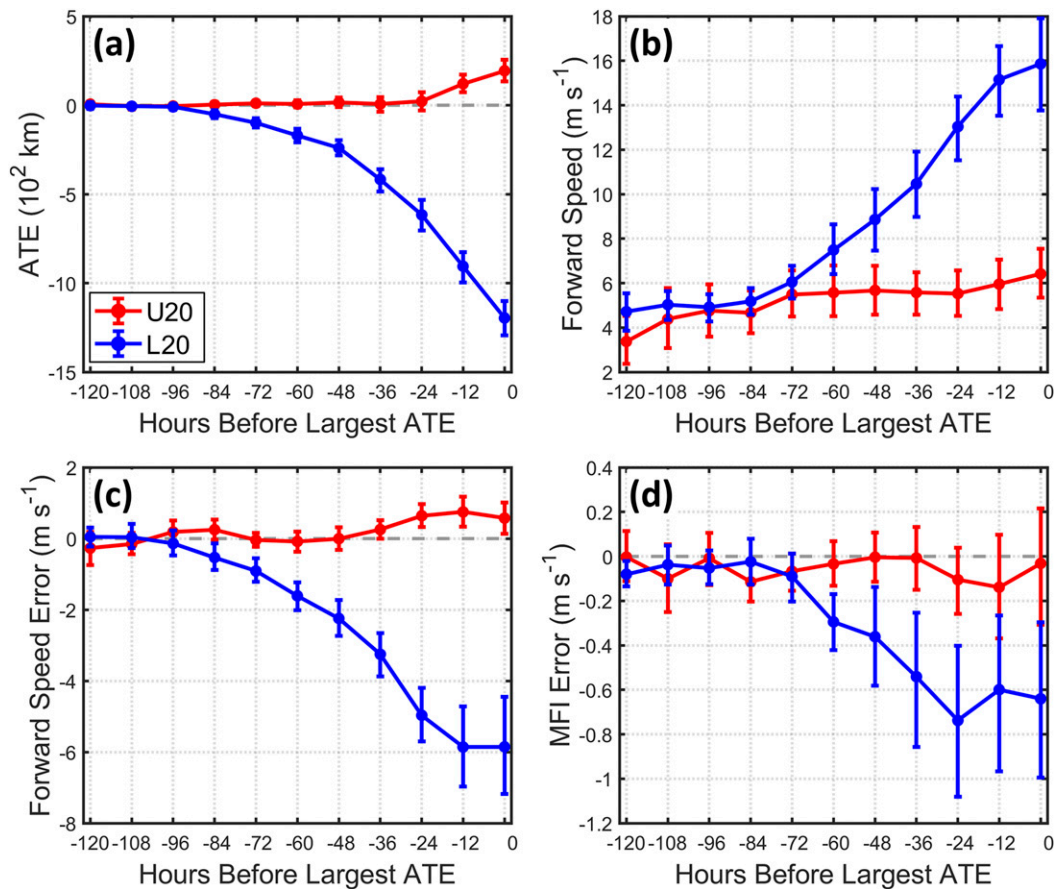

FIG. 6. (a) ATEs, (b) observed forward speed and (c) forward speed errors, and (d) meridional flow index errors composited for L20 and U20 cases.

forward speed errors (Fig. 6c), the L20 cases become increasingly negative (i.e., too slow), reaching $-6\,\mathrm{m\,s^{-1}}$ by $T - 12\,\mathrm{h}$.

The slow bias of the L20 cases may correspond to an underamplification of the extratropical flow. Following Archambault et al. (2013) and Fowler and Galarneau (2017), the meridional flow index (MFI) is computed from the area-average magnitude of the meridional component of the wind at the dynamic tropopause (the 2.0 potential vorticity unit (PVU) surface, where 1.0 PVU $= 1.0 \times 10^{-6}\,\mathrm{K\,kg^{-1}\,m^2\,s^{-1}}$ (e.g., Hoerling et al. 1991; Holton et al. 1995). In this paper, the MFI is calculated over the midlatitude North Atlantic and North America encompassing 35°–65°N and 110°–10°W. While larger than the domain used by Fowler and Galarneau (2017), this region encompasses the many different tracks of the U20 and L20 cases. Subtracting the MFI of the CFSR from those of the individual ECMWF members and averaging the results gives the mean MFI error. The MFI errors of the L20 and U20 cases are similar before $T - 60\,\mathrm{h}$ (Fig. 6d). Afterward, the L20 cases show an increasingly negative (underamplified) bias, which reaches $-0.8\,\mathrm{m\,s^{-1}}$ by $T - 24\,\mathrm{h}$. By comparison, the errors of the U20 cases are not significantly different from zero at any time.

The accelerated growth of the L20 ATEs motivates investigation into a potential correlation between the ensemble member ATEs at $T - 0\,\mathrm{h}$ and the ATEs earlier in the forecast. For example, are the members with the most negative ATEs at $T - 0\,\mathrm{h}$ also the most negative earlier at $T - 96\,\mathrm{h}$? Fig. 7a shows the average Pearson R correlations between member ATEs at $T - 0\,\mathrm{h}$ and at different lead times for L20 and U20. About 50% of the L20 do not have positive correlations significantly different from 0 to $T - 84\,\mathrm{h}$. However, these correlations quickly grow afterward, exceeding 0.60 on average and reaching statistical significance in ~90% of L20 cases by $T - 60\,\mathrm{h}$. In contrast, the U20 only have average correlations of 0.10 by $T - 60\,\mathrm{h}$, with statistical significance in only ~30% of the cases.

4. Large-scale error evolution

a. Steering flow errors

To determine how much of the ATEs originate from errors associated with synoptic steering features, such as troughs and ridges, the L20 and U20 are compared using best track–relative composites of the 300-hPa geopotential height errors and 850–200-hPa layer-averaged wind errors (Fig. 8). In this section, the wind errors are calculated by first removing the CFSR (ECMWF member) TC winds within 8° from the best track (ECMWF member) TC position (e.g., Galarneau and Davis 2013).
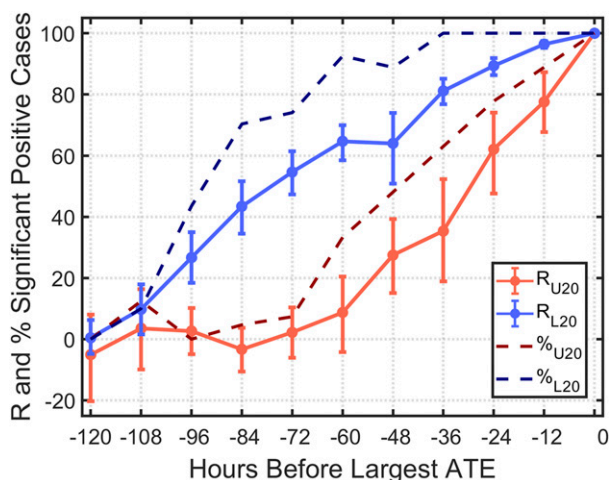

FIG. 7. Pearson correlations ($R \times 100\%$) between member ATEs at $T - 0\,\mathrm{h}$ and member ATEs at different leads before $T - 0\,\mathrm{h}$, averaged for L20 and U20 cases. The dashed lines are the percentages of cases that have statistically significant positive correlations.

At $T - 84\,\mathrm{h}$, the average heights for the L20 and U20 are similar (Figs. 8a,b). Both sets of cases have a trough more than 2000 km northwest of the TC, though the trough in the L20 appears to be ~200–500 km farther east than in the U20. The L20 have a more defined subtropical high centered east of the TC given by a broad closed 970-dam contour, as well as an incipient midlatitude ridge more than 1500 km to the northeast. The composite height errors for both sets show no coherent significance patterns and are only ~5 m in magnitude over a few small regions.

By $T - 60\,\mathrm{h}$, the L20 cases have a deepening trough about 1500 km northwest of the TC and a ridge 1000–1500 km northeast of the TC (Fig. 8c). Statistically significant positive height biases of ~10 m are associated with the eastern flank of trough, suggesting that the model is underpredicting the trough and tilting its axis too far to the west. Meanwhile, a broad region of negative height errors of over 5 m is situated over the downstream midlatitude ridge. As a result, 850–200-hPa steering flow errors are developing east of the trough, the vectors implying an underprediction of the southerly flow. By comparison, the $T - 60\,\mathrm{h}$ height errors of the U20 are still largely below ~5 m in magnitude (Fig. 8d).

By $T - 36\,\mathrm{h}$, the adjacent trough–ridge couplet in the L20 amplifies rapidly (Fig. 8e). Both the trough and ridge are underpredicted by more than 25 m, consistent with an anticyclonic (cyclonic) 850–200-hPa wind error northwest (northeast) of the TC. As a result, the southerly flow steering the TC is underpredicted by more than $5\,\mathrm{m\,s^{-1}}$. Meanwhile, the U20 only begin to show height biases of up to 10 m attached to an approaching trough to the north, though the biases are not
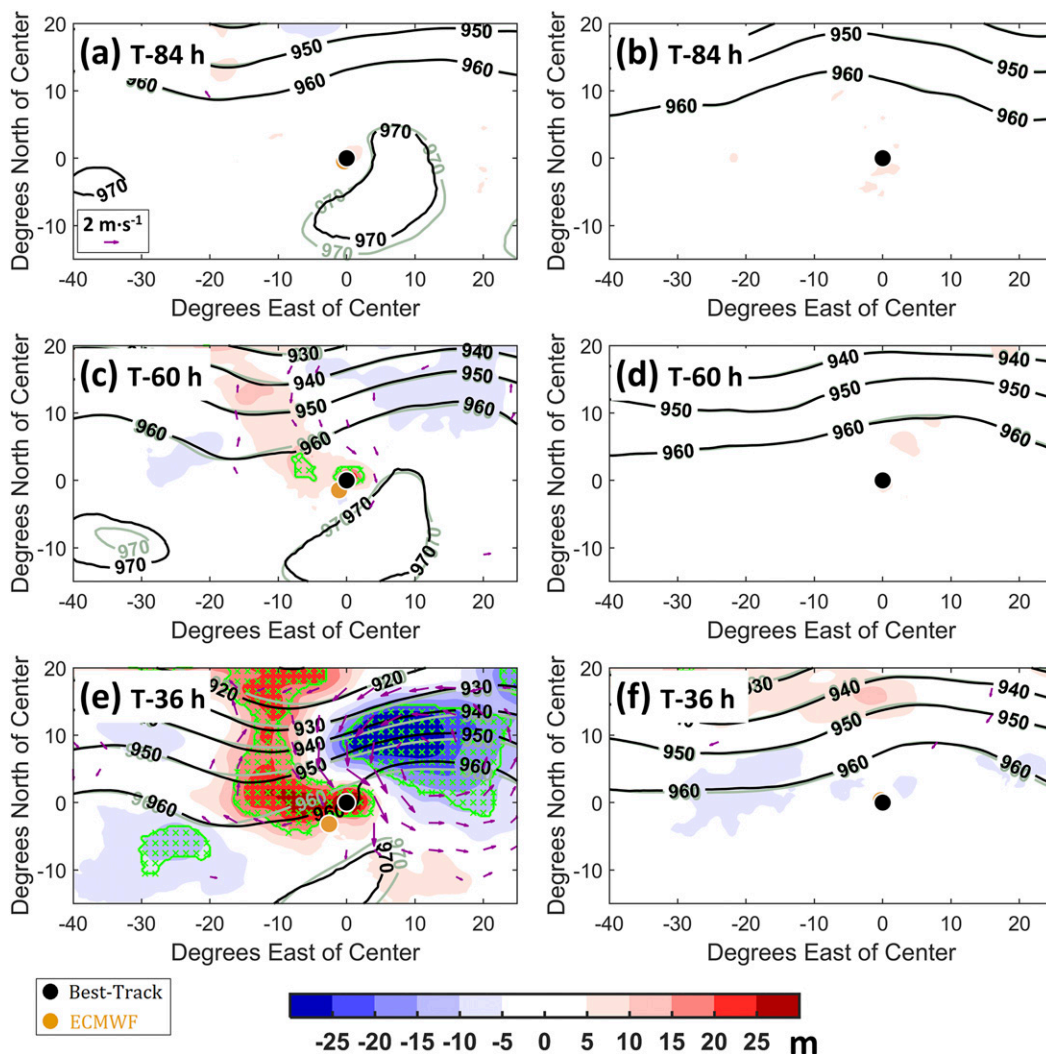
FIG. 8. Best track–relative composites of mean 300-hPa geopotential height errors (shaded) and 300-hPa geopotential heights of the CFSR (black contours in dam) and ECMWF mean (gray contours) for L20 cases at (a) $T - 84$ h, (c) $T - 60$ h, and (e) $T - 36$ h before the time of largest ATE. (b),(d),(f) As in (a),(c),(e), but for U20. Green dotted regions indicate where the composited mean height error is statistically significant. Vectors indicate errors in the 850–200-hPa mean steering flow (after removing the TC circulation) greater than $1\,\mathrm{m\,s^{-1}}$.

statistically significant (Fig. 8f). The U20 cases thereby have significantly smaller systematic biases in the large-scale steering and will not be shown in the remainder of this paper.

The role of the synoptic environment in the L20 cases is further analyzed with best track–centered composites of standardized differences (10 slowest − 10 fastest members) in 300-hPa geopotential height and 850–200-hPa steering flow (Fig. 9). At $T - 108$ h, there is some suggestion that the slowest members have lower heights along a developing ridge north of the TC in 50% of the cases (Fig. 9a). By $T - 84$ h, 50% of the cases also suggest that the 10 slowest members have higher heights near the trough located up to 2000 km west of the TC

(Fig. 9b). The height differences associated with the deepening trough and ridge grow in magnitude between $T - 84$ h and $T - 60$ h (Fig. 9c). Consistent with Fig. 8c, the higher heights along the trough and lower heights over the ridge result in the slowest members erroneously having a more northerly (less southerly) 850–200-hPa wind than the fastest members. By $T - 36$ h (Fig. 9d), the height differences grow to over one standard deviation in magnitude and the wind differences resemble the mean error composites for the L20 cases. The height differences attached to the trough and ridge are statistically significant in more than 70% of the cases. Hence, the height and wind fields of the fastest members are more amplified and more closely resemble the CFSR.
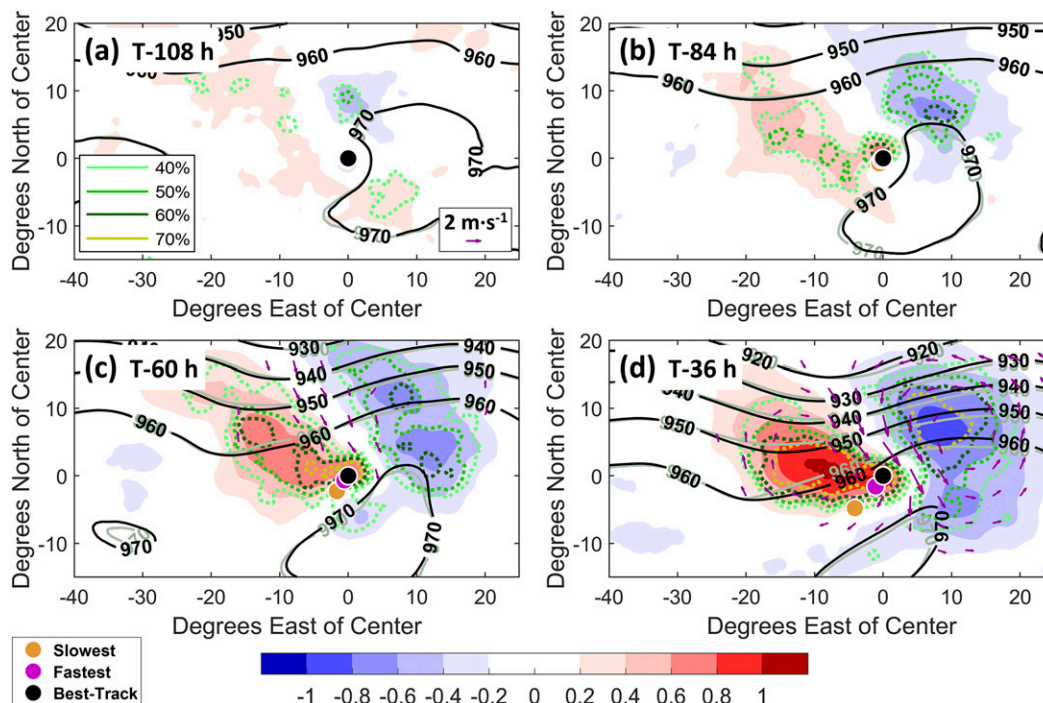
FIG. 9. Best track–relative composites of standardized differences (10 slowest − 10 fastest ensemble members) in 300-hPa geopotential height for L20 cases at (a) $T − 108$ h, (b) $T − 84$ h, (c) $T − 60$ h, and (d) $T − 36$ h. Percentages of L20 cases in which the difference at each grid point is statistically significant are contoured in different shades of green to gold. Black and gray contours are the composited 300-hPa geopotential heights (in dam) of the 10 fastest and 10 slowest ensemble members of each case, respectively, and vectors indicate the slowest − fastest member differences in 850–200-hPa mean steering flow (after removing the TC circulation) greater than $1 \, \mathrm{m \, s^{-1}}$.

A closer qualitative inspection of the individual 27 L20 cases reveals that at least 20 of them have a similar pattern to the one described by Peng et al. (2017): an adjacent upstream trough or cutoff low and a downstream area of ridging. These and other L20 cases are listed in Table 1. There also tends to be another ridge west of the trough in these cases, though this upstream ridge typically does not have significant biases attached to it. The resulting ridge–trough–ridge segment possibly suggests an incipient Rossby wave packet like those shown occurring 72 h prior to "strong interaction" cases by Archambault et al. (2015).

In four of the seven cases not involving an upstream trough and downstream ridge, a steering trough or cutoff low is within 500 km east of the TC at the forecast initialization. The observed TC phases with the trough and progresses northeastward, while the model TC meanders behind. In these cases, the TC appears to be close to a bifurcation point similar to the ones described by Riemer and Jones (2014), such that a slight initial drift in the TC and/or the trough can determine whether the two systems phase with or completely miss each other. Hence, some ensemble members have the TC correctly interact with the downstream trough and closely match

the best track. Hurricane Nadine (2012) was already shown by Munsell et al. (2015) to have this behavior.

b. Upper-level potential vorticity interaction

The development of the trough–ridge couplet in the L20 cases is further examined by analyzing 300–200-hPa layer-averaged potential vorticity advection (PVA), which is composited in the best track–relative framework in Fig. 10. PVA by the nondivergent and irrotational winds are compared side by side. Starting when the trough heights first become underamplified at $T − 60$ h, PVA by the nondivergent wind shows a region of negative biases 700–1000 km northwest of the TC (Fig. 10a). This negative PVA bias grows from −3 to −5 PVU·day^{-1} between $T − 48$ h (Fig. 10c) and $T − 36$ h (Fig. 10e), connecting the trough with the northern portion of the TC. During this time, the PV associated with the trough changes consistently with the PVA by the nondivergent wind. The ECMWF's weaker PVA results in PV lines that are less meridionally oriented north of the TC than in the CFSR. The negative PVA bias corresponds to an underprediction of positive PVA east of the trough's tip. Hence, the underprediction of the trough's amplitude can be partially explained by PVA errors.

TABLE 1. L20 forecasts sorted by synoptic features associated with their track errors. The initialization dates of the forecasts are in parentheses.

L20 error interaction types			
Underamplified upstream trough and downstream ridge	Downstream trough bifurcation point	Erroneous interaction with downstream cut-off low	Not explained by synoptic flow
Gustav (0000 UTC 1 Sep 2008)	Omar (0000 UTC 15 Oct 2008)	Bertha (0000 UTC 17 Sep 2008)	Ida (0000 UTC 6 Nov 2009)
Hanna (1200 UTC 3 Sep 2008)	Nadine (1200 UTC 12 Sep 2012)		Gabrielle (1200 UTC 5 Sep 2013)
Ike (0000 UTC 11 Sep 2008)	Jerry (1200 UTC 30 Sep 2013)		
Bill (0000 UTC 20 Aug 2009)	Cristobal (0000 UTC 24 Aug 2014)		
Danielle (0000 UTC 27 Aug 2010)			
Igor (0000 UTC 18 Sep 2010)			
Otto (1200 UTC 6 Oct 2010)			
Tomas (0000 UTC 7 Nov 2010)			
Irene (0000 UTC 25 Aug 2011)			
Katia (1200 UTC 5 Sep 2011)			
Ophelia (0000 UTC 30 Sep 2011)			
Kirk (0000 UTC 29 Aug 2012)			
Leslie (0000 UTC 7 Sep 2012)			
Rafael (1200 UTC 13 Oct 2012)			
Rafael (0000 UTC 16 Oct 2012)			
Arthur (1200 UTC 2 Jul 2014)			
Cristobal (1200 UTC 26 Aug 2014)			
Gonzalo (1200 UTC 14 Oct 2014)			
Ian (1200 UTC 12 Sep 2016)			
Karl (1200 UTC 21 Sep 2016)			

Within the region of negative PVA errors, the non-divergent wind error vectors point southeastward from the trough to the TC. These wind errors are caused by the ECMWF winds being both weaker and more east-ward than the CFSR. PV increases westward going from the TC to the trough, such that the model wind is going down the PV gradient. Thus, the negative bias in PVA mainly comes from the model underpredicting the PV gradient.

Meanwhile, at $T - 60$ h (Fig. 10b), the TC's divergent outflow interacts with the trough–ridge system by advecting relatively low PV north and northwestward. The irrotational wind errors converge north of the TC center, implying that the ECMWF's outflow is too weak. The irrotational wind speeds are underpredicted by 1 to $2 \, \mathrm{m \, s^{-1}}$ within an 800-km radius from the TC center. The negative PVA is thus underestimated, corresponding to a region of positive biases east of the trough and partially into the downstream ridge. This positive bias in PVA by the irrotational wind persists from $T - 48$ h (Fig. 10d) to $T - 36$ h (Fig. 10f), consistent with the 2-PVU contour of the ECMWF failing to fold north-westward 500–700 km northwest of the TC.

After $T - 60$ h, the negative biases in PVA by the nondivergent wind begin to cancel-out the positive biases in PVA by the irrotational wind from the TC. However, the irrotational wind plays a significant role in amplifying the PV gradient, as demonstrated by the schematic in Fig. 11. East of the trough, the non-divergent wind is east-northeastward and the irrota-tional wind is northwestward (Fig. 11a). Focusing on the PV gradient east of the trough, the nondivergent wind crosses the PV contours, advecting the contours eastward (Fig. 11b). Meanwhile, the irrotational winds are perpendicular to the contours, advecting lower values of PV northwestward. The resulting deforma-tion of the total wind (i.e., the sum of the irrotational and nondivergent wind) increases the PV gradient. Note that the nondivergent wind alone would only advect the PV contours eastward without increasing the gradient.

From Fig. 10, the irrotational winds are under-predicted, while the nondivergent winds are both un-derpredicted and too eastward. The model PV lines are thereby not concentrated enough, causing the PV gra-dient and subsequent steering flow to become further underpredicted, such that the model TC is too slow. The combination of these PVA errors may correspond to the model trough becoming less negatively tilted than ob-served, similar to what was shown by Atallah and Bosart (2003) for Hurricane Floyd (1999).

The 300–200-hPa PVA by the divergent wind is fur-ther analyzed through best track–relative composites of standardized differences between the slowest and fastest
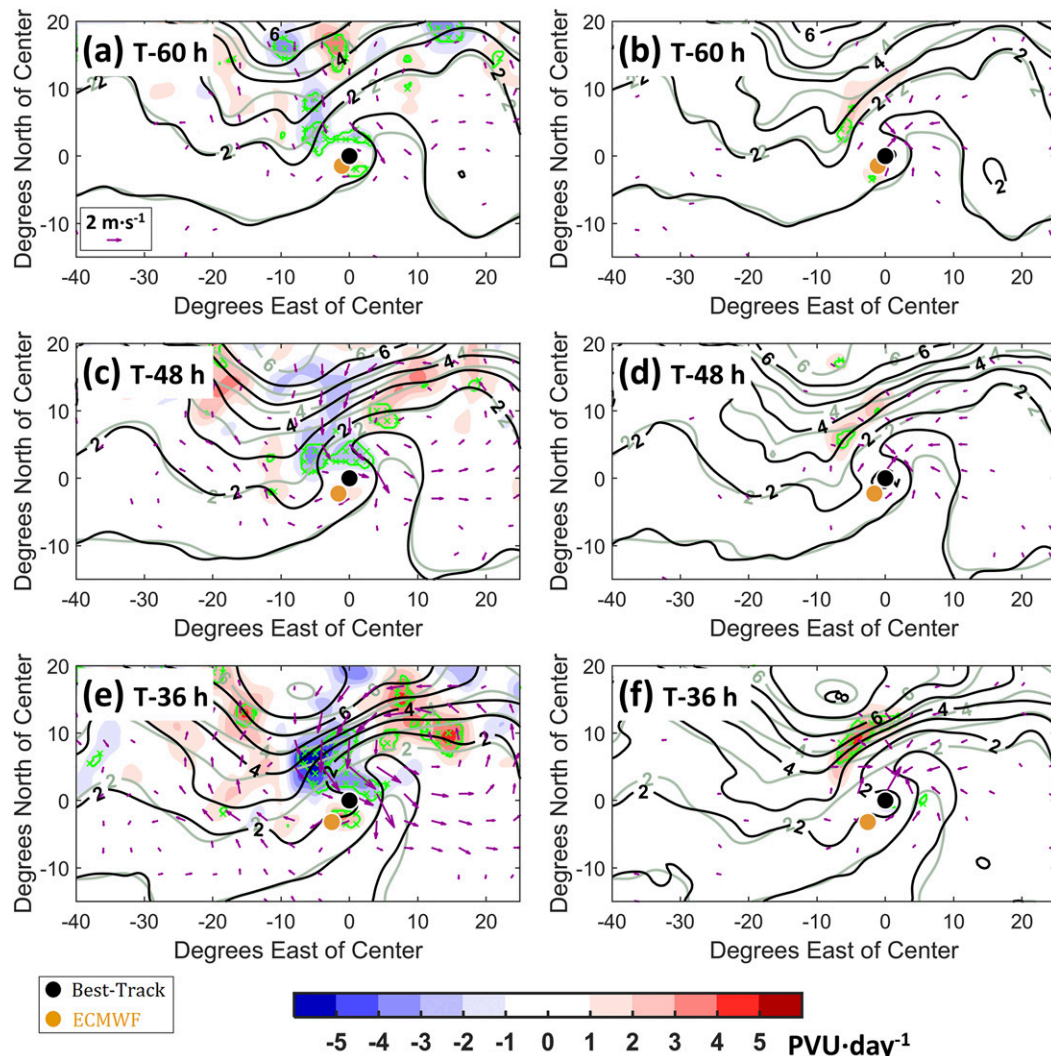
FIG. 10. L20 best track–relative composites of mean errors in 300–200-hPa potential vorticity advection by the nondivergent wind (shaded), nondivergent wind errors (vectors greater than $1\,\mathrm{m\,s}^{-1}$), and potential vorticity of the CFSR (black contours in PVU) and ECMWF mean (gray contours) for L20 cases at (a) $T - 60$ h, (c) $T - 48$ h, and (e) $T - 36$ h before the time of largest ATE. (b),(d),(f) As in (a),(c),(e), but using the divergent wind. Green dotted regions indicate where the composited mean advection errors are statistically significant.

ensemble members (Fig. 12). About 40% of the cases have statistically significant positive PVA differences in a small region 500–1000 km north-northeast of the TC center at $T - 84$ h (Fig. 12a). The PVA of the slowest members over this region is on average ~0.2 standard deviations less negative than the fastest members, consistent with the slowest members having divergent winds that are ~0.5 m s^{-1} weaker. The positive PVA difference shifts to the northwest of the TC and along the incoming trough by $T - 60$ h (Fig. 12b). While less pronounced than the biases in Fig. 10b, this region of positive PVA differences suggests that in 40% of the cases the outflow of the fastest members more correctly interacts with the trough than that of the slowest

members. The region where the slowest member divergent winds are 0.5 m s^{-1} weaker than the fastest extends more than 1000 km northwest of the TC by $T - 48$ h (Fig. 12c). As a result, the PV lines along the trough of the fastest members become more concentrated and meridionally oriented like the CFSR. By $T - 36$ h (Fig. 12d), the PVA along the trough of the slowest members is 0.6 standard deviations less negative than the fastest members, the difference reaching significance in ~60% of the cases. Thus, there is a dependence on the TC's outflow starting around $T - 60$ h, in which weaker divergent winds correspond to less negative PVA along the eastern flank of the trough. The trough does not become as negatively tilted as observed and the downstream
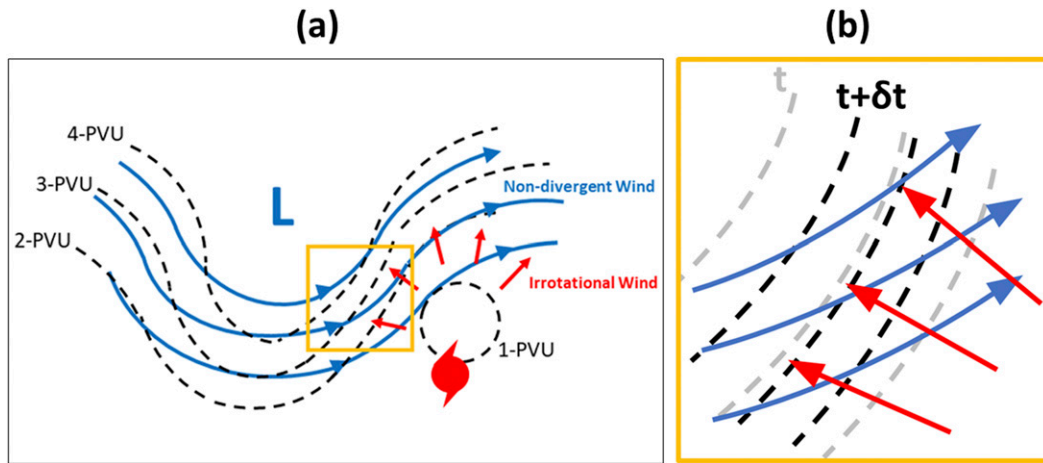
FIG. 11. Schematic of (a) PV (black dashed lines), nondivergent wind (blue streamlines), and irrotational wind (red vectors). (b) A close-up within the orange box in (a), showing PV at times t (gray dashed lines) and $t + \delta t$ (black dashed lines).

ridge underamplifies. The northward flow east of the trough is too weak and does not accelerate the TC.

During ET, upper-level jet streaks are often enhanced as the TC's outflow impinges on the PV gradient (Riemer and Jones 2010; Grams et al. 2013; Archambault et al. 2013, 2015). The associated ageostrophic circulations can enhance the temperature advection along the trough (Cammas and Ramond 1989; Clark et al. 2009) and therefore affect the steering trough's amplitude. The development of
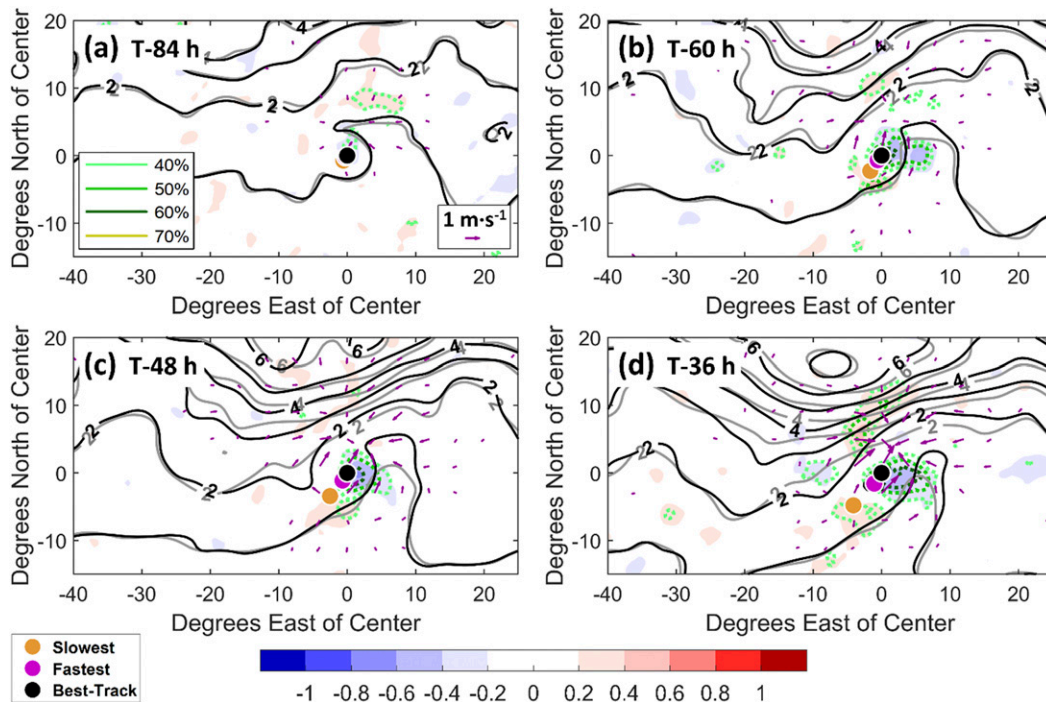

FIG. 12. L20 best track–relative composites of standardized differences in 300–200-hPa potential vorticity advection by the divergent wind at (a) $T - 84$ h, (b) $T - 60$ h, (c) $T - 48$ h, and (d) $T - 36$ h. Percentages of L20 cases in which the difference at each grid point is statistically significant are contoured in different shades of green to gold. Black and gray contours are the composited 300–200-hPa potential vorticity (in PVU) of the fastest and slowest ensemble members, respectively, and vectors indicate the member differences in 300–200-hPa divergent wind greater than $0.2 \, \mathrm{m \, s^{-1}}$.
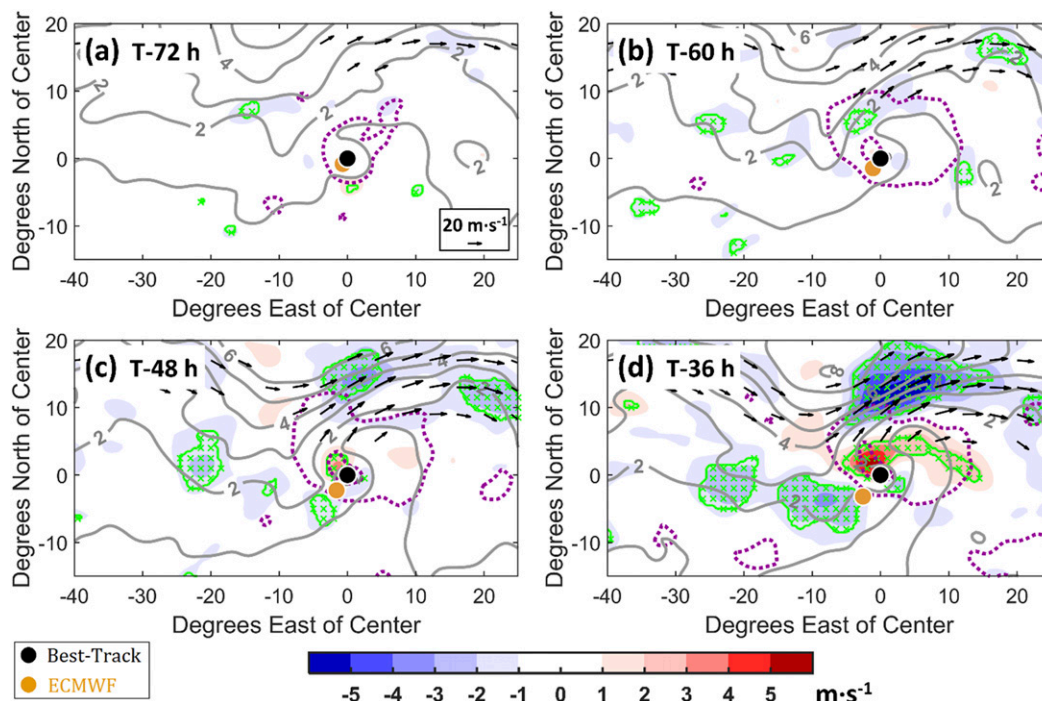
FIG. 13. L20 best track–relative composites of mean 300–200-hPa wind speed errors (shaded), and CFSR 300–200-hPa winds (vectors greater than $20 \, \mathrm{m \, s^{-1}}$), CFSR divergent wind speeds (dashed purple contours every $2 \, \mathrm{m \, s^{-1}}$), and CFSR potential vorticity (gray contours) at (a) $T - 72 \, \mathrm{h}$, (b) $T - 60 \, \mathrm{h}$, (c) $T - 48 \, \mathrm{h}$, and (e) $T - 36 \, \mathrm{h}$ before the time of largest ATE. Green dotted regions indicate where the composited wind speed error is statistically significant.

this jet is examined through best track–relative composites of 300–200-hPa winds and wind speed errors (Fig. 13). At $T - 72 \, \mathrm{h}$ (Fig. 13a), a $25–30 \, \mathrm{m \, s^{-1}}$ jet streak is ~1800–2000 km north of the TC center. As the TC and jet streak draw closer at $T - 60 \, \mathrm{h}$ (Fig. 13b), the divergent outflow, as given by the $2 \, \mathrm{m \, s^{-1}}$ contour, expands by more than 500 km in all directions and begins to interact with the trough's PV gradient. By $T - 48 \, \mathrm{h}$ (Fig. 13c), the entrance region of the jet near this region of interaction intensifies to $35 \, \mathrm{m \, s^{-1}}$, but the ECMWF underestimates the intensification by ~$3–4 \, \mathrm{m \, s^{-1}}$. The jet further intensifies to more than $40 \, \mathrm{m \, s^{-1}}$ by $T - 36 \, \mathrm{h}$ (Fig. 13d), with the ECMWF underpredicting the speed by $5–10 \, \mathrm{m \, s^{-1}}$. There is also a region of positive wind speed biases to the south of the jet, corresponding to the model jet also having a slight southward bias, which may be associated with differences in the TC position affecting the location of the TC's outflow. The L20 cases are thus associated with strong ET interactions that amplify upper-level jet streaks. This interaction is significantly underpredicted, again consistent with the trough becoming further underamplified.

c. Upstream trough thermodynamic interactions

The rapid growth of the underamplification of the trough after $T - 60 \, \mathrm{h}$ (Fig. 8c) may correspond to errors in temperature advection. From a dry perspective, insufficient 300-hPa height falls in the trough may result from an underprediction of cold advection decreasing with height below this level per the quasigeostrophic height tendency equation (Holton and Hakim 2013). Errors in the TC's location and intensity can impact when and how strongly the TC's surface circulation becomes embedded in the approaching trough's temperature gradient (Veren et al. 2009), thereby causing temperature advection errors at low to midlevels. Meanwhile, the underprediction of the jet streak corresponds to a weak bias in its associated ageostrophic circulation, whose lower branch can also impact cold air advection below the jet entrance region.

Figure 14 shows best track–relative and TC-relative errors in 700-hPa temperature advection of the L20 composited side by side. Compared to 700-hPa, the temperature advection errors at 500 and 300 hPa are small (not shown). Thus, the errors in differential temperature advection, and the subsequent errors in height falls along the upper-level trough, are largely from errors at 700 hPa. In the best track–relative composites, 700-hPa temperature advection errors of ~1 to $2 \, \mathrm{K \, day^{-1}}$ are found within 500 km of the TC center at $T - 60 \, \mathrm{h}$ (Fig. 14a), with insufficient warm and cold advection northeast and southwest of the TC,
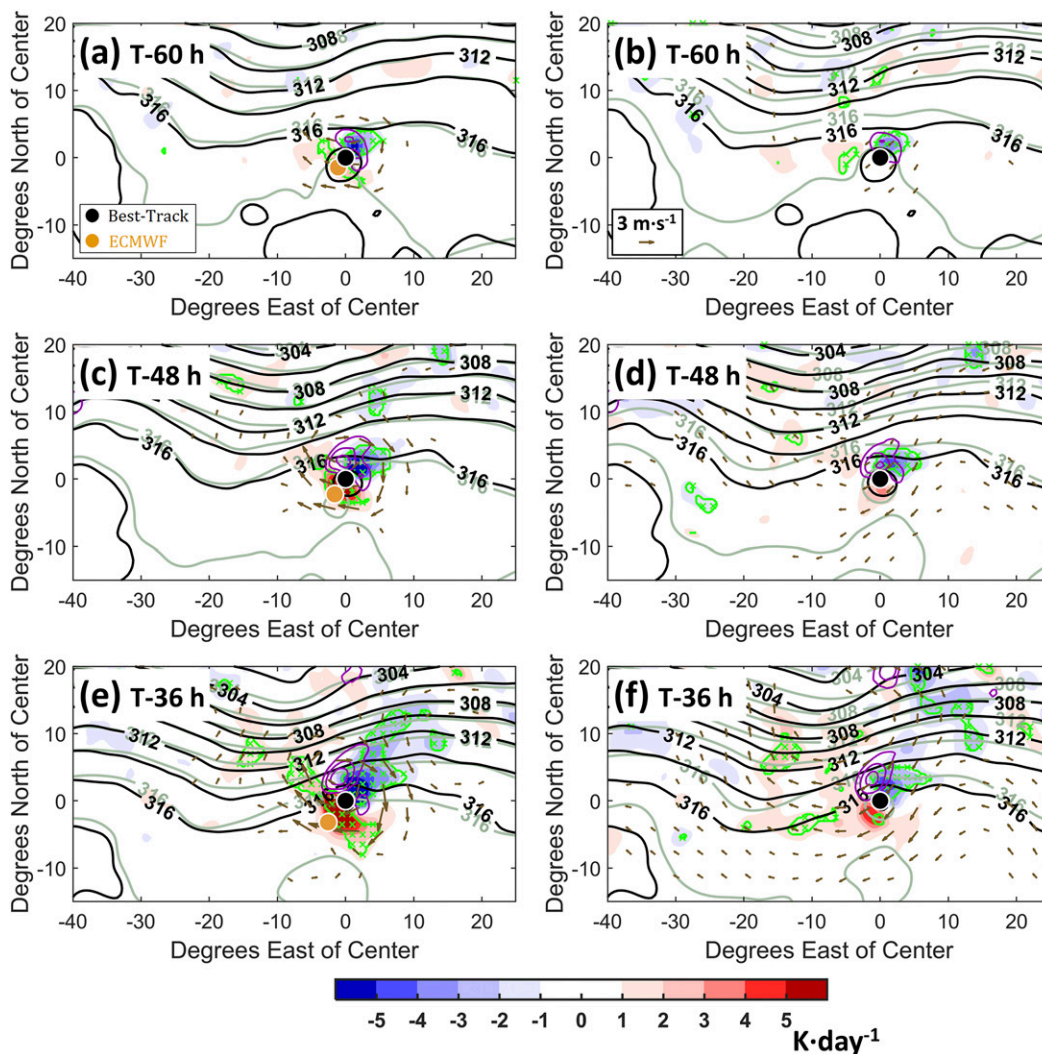
FIG. 14. L20 best track–relative composites of mean 700-hPa temperature advection errors (shaded), frontogenesis errors (purple contours every 10^{-2} K km^{-1} day^{-1}, only showing negative values), and temperatures of the CFSR (black contours in K) and ECMWF mean (gray contours) at (a) $T - 60$ h, (c) $T - 48$ h, and (e) $T - 36$ h before the time of largest ATE. Green dotted regions indicate where the composited temperature advection error is statistically significant. (b),(d),(f) As in (a),(c),(e), but using the TC-relative framework.

respectively. The development of these advection errors is consistent with ~2 to 3 m s^{-1} errors in the 700-hPa winds circulating around the TC center. There is also a region of underpredicted 700-hPa frontogenesis inside of 500 km north-northeast of the TC. By $T - 48$ h (Fig. 14c), the region of underpredicted cold advection expands to more than 700 km northwest of the TC center, extending from the trough to the TC. The wind errors northwest of the TC are up to ~5 m s^{-1} and point northwestward into the trough. This region of southeasterly wind errors is more apparent by $T - 36$ h (Fig. 14e), at which point the cold advection is underpredicted by ~5 K day^{-1} over an area extending 1000 km northwest of the TC. During this time, the weak biases in

frontogenesis expand over 700 km north of the TC, while a 2–3 K day^{-1} underprediction of warm advection extends more than 1000 km northeast of the TC.

A comparison with TC-relative composites helps to determine whether the wind errors, and hence advection errors, are a consequence of the TC position errors. At $T - 60$ h (Fig. 14b), the TC-relative composites have ~1–2 m s^{-1} northeasterly wind errors inside of 500 km from the TC center, compared to the southeasterly wind errors northwest of the TC in Fig. 14a. There is still an underprediction of both warm advection and frontogenesis north of northeast of the TC. West of the TC, there are ~1 m s^{-1} easterly wind errors and a small region of underpredicted cold advection. The wind
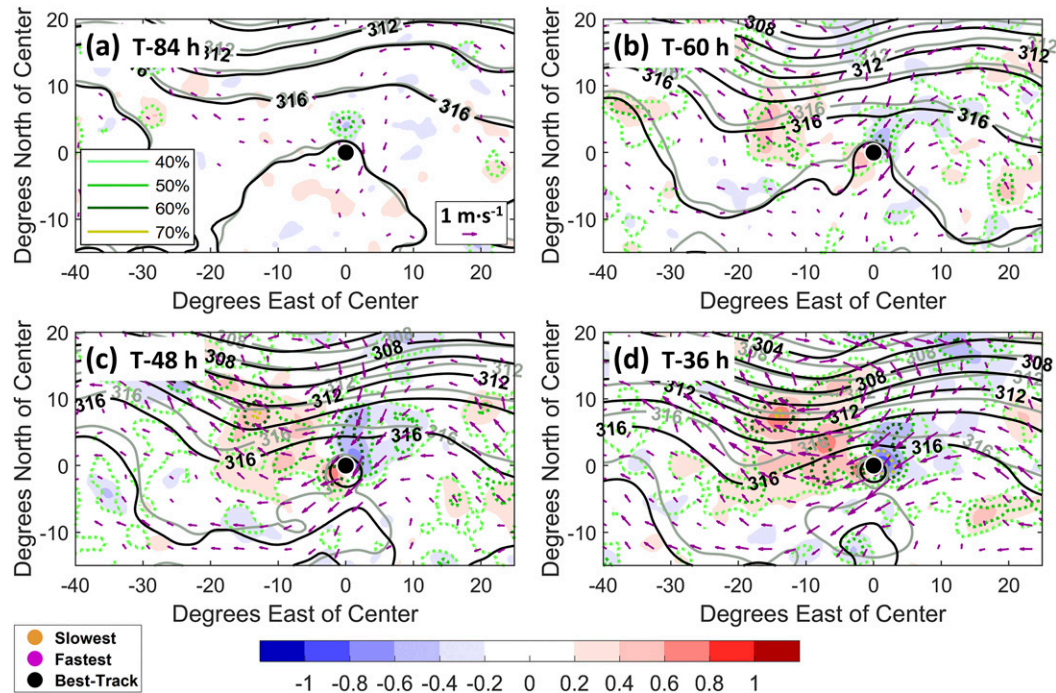
FIG. 15. L20 TC-relative composites of standardized differences in 700-hPa temperature advection at (a) $T - 84$ h, (b) $T - 60$ h, (c) $T - 48$ h, and (e) $T - 36$ h before the time of largest ATE. Black and gray contours are the composited 700-hPa temperatures (in K) of the fastest and slowest ensemble members, respectively. Vectors indicate slowest − fastest member differences in 700-hPa winds greater than 0.5 m s^{-1}. Percentages of L20 cases in which the difference at each grid point is statistically significant are contoured in different shades of green to gold.

errors are southeasterly between ~1000 and 2000 km northwest of the TC, crossing into the trough. From $T - 48$ h (Fig. 14d) to $T - 36$ h (Fig. 14f), these southeasterly wind errors grow and draw closer to the TC, corresponding to ~2 K day^{-1} warm biases in advection over regions similar to Fig. 14e. However, the advection errors within 500 km south of the TC are significant over a smaller region than in Fig. 14e. During this time, the weak biases in warm advection and frontogenesis again expand farther north of northeast of the TC. Thus, the temperature advection errors inside of 500 km from the TC are caused by TC displacement errors and weak biases in the TC wind. However, southeasterly wind errors farther northwest of the TC are associated with the trough and may correspond to the underprediction of the ageostrophic circulation induced by the upper-level jet and low-level frontogenesis.

The thermodynamics of the trough–ridge system are further examined through TC-relative standardized differences in 700-hPa temperature advection (Fig. 15). At $T - 84$ h (Fig. 15a), 50% of the cases have a significantly negative difference ~500 km north of the TC. By $T - 60$ h (Fig. 15b), the differences along the trough to the west become significantly positive in at least 50%

of the cases, the fastest members having 0.4 standard deviations more cold advection than the slowest members. Between $T - 48$ h (Fig. 14c) and $T - 36$ h (Fig. 14d), this positive difference increases to ~0.6 standard deviations and is significant in 60% of the cases. The slowest–fastest member wind differences west of the TC and across the trough are southeasterly and 1–2 m s^{-1} in magnitude. Thus, the underamplification of the trough in the slowest members is accelerated by the underprediction of midlevel cold air advection beneath the trough. The growing slow bias of the TC further reduces the southeastward extent of the cold advection.

5. Impact of moisture fluxes and precipitation

Section 4b demonstrated that the errors in the steering trough are caused by errors in PVA from the nondivergent and irrotational winds, both of which are underpredicted. The irrotational winds are largely driven by from convection near the TC. Thus, the standardized differences in precipitation rate are composited for all L20 cases in Fig. 16. To isolate the differences in TC structure, the composites in this section are within the TC-relative framework. At $T - 108$ h (Fig. 16a), at least
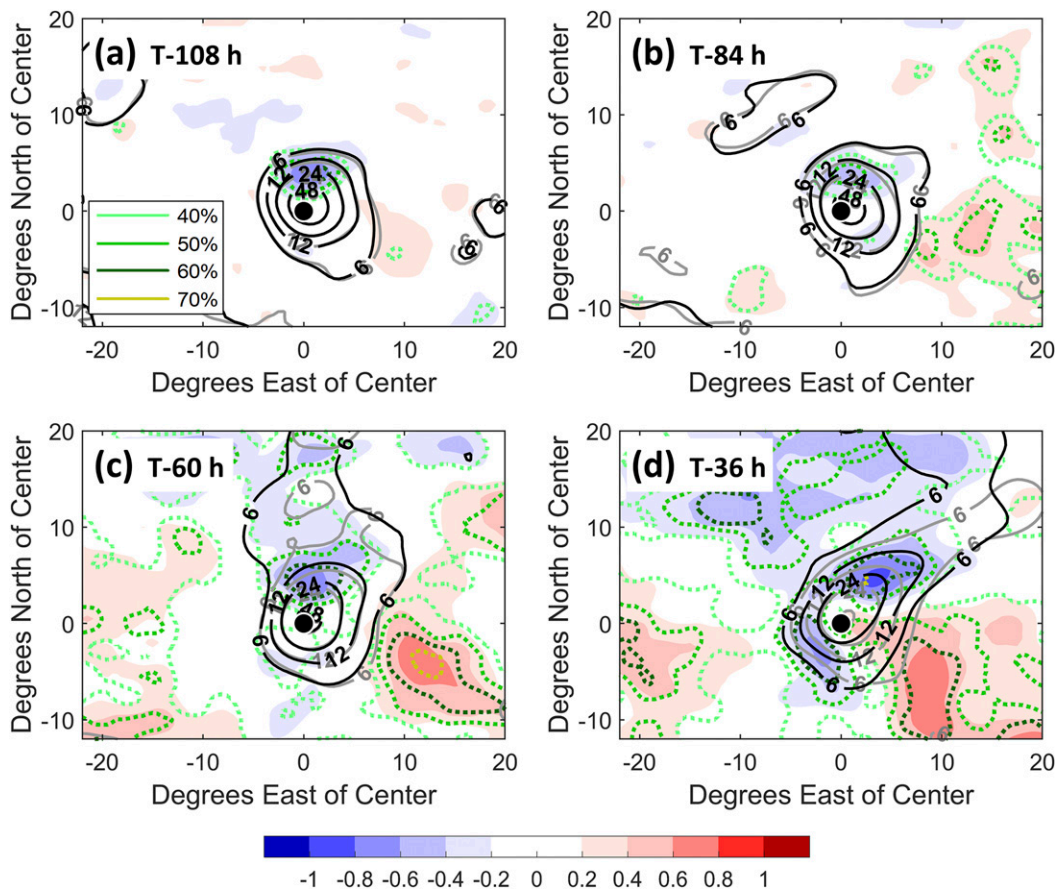
FIG. 16. L20 TC-relative composites of standardized differences in precipitation rate at (a) $T - 108$ h, (b) $T - 84$ h, (c) $T - 60$ h, and (e) $T - 36$ h. Black and gray contours are the composited precipitation rates (in mm·h^{-1}) of the fastest and slowest ensemble members, respectively. Percentages of L20 cases in which the difference at each grid point is statistically significant are contoured in different shades of green to gold.

60% of cases have statistically significant negative differences inside of ~500 km north of the TC center, the slowest members having rainfall rates that are ~0.6 standard deviations lower than the fastest. Rainfall associated with the upstream trough is visible by $T - 84$ h (Fig. 16b). Between $T - 60$ h (Fig. 16c) and $T - 36$ h (Fig. 16d), the precipitation along the trough merges with the TC and the precipitation region elongates northeastward. This merger happens sooner in the fastest members, corresponding at least in part to growing differences in the forward speeds of the TCs. The region of negative differences north of the TC also spreads northeastward by $T - 36$ h, the difference averaging ~0.8 standard deviations in magnitude and reaching significance in up to 70% of the cases. Therefore, the fastest members have higher rainfall rates and stronger divergent outflow, which amplifies the steering flow.

Torn (2010) showed that the ridging downstream of two western Pacific ET cases was significantly correlated with the lower-tropospheric horizontal moisture fluxes on the east side of the TCs. These moisture fluxes increased the precipitation along the baroclinic zone and hence enhanced the divergent outflow interacting with the ambient PV, which strengthened the steering flow. To test his hypothesis on the L20 cases, TC-relative standardized differences in 700-hPa moisture fluxes are plotted in Fig. 17. Consistent with Fig. 16a, ~60% of the cases at $T - 108$ h have statistically significant negative difference inside of 500 km northeast of the TC, the fastest members having moisture fluxes up to 0.6 standard deviations greater than the slowest members (Fig. 17a). Between $T - 96$ h (Fig. 17b) and $T - 84$ h (Fig. 17c), these differences continue to grow in magnitude and extend more than 500 km along the eastern flank of the TC, reaching statistical significance in up to 70% of the cases. By $T - 36$ h, the fastest members have fluxes that are more than one standard deviation greater than the slowest members (Fig. 17d). Thus, the differences in moisture fluxes affect the ATEs by modulating
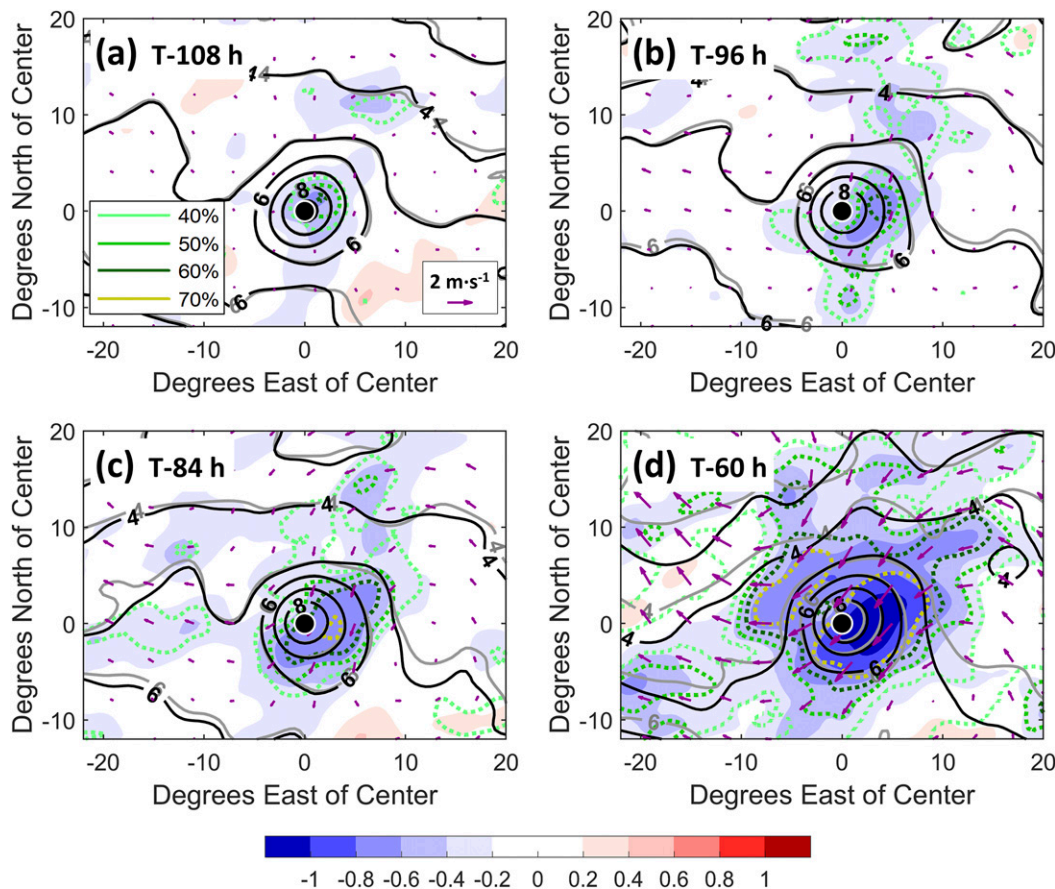
FIG. 17. TC-relative composites of standardized differences in 700-hPa moisture flux at (a) $T - 108$ h, (b) $T - 96$ h, (c) $T - 84$ h, and (e) $T - 60$ h. Black and gray contours are the composited 700-hPa specific humidity (in g·kg^{-1}) of the fastest and slowest ensemble members, respectively. Vectors indicate slowest–fastest member differences in 700-hPa wind. Percentages of L20 cases in which the difference at each grid point is statistically significant are contoured in different shades of green to gold.

the near-TC precipitation rates and hence the diabatic outflow interacting with the ambient extratropical flow.

6. Conclusions

The goal of this study is to determine the most common causes of large ATE forecasts in North Atlantic TCs from the ECMWF ensemble during the 2008–16 period. This study is the first to apply quantitative ensemble-based diagnostics to multiple North Atlantic TCs, correlating relevant meteorological fields with the track errors of each case and compositing the statistics in geographic and storm-centered frameworks. We currently focus on the ECMWF's slow bias for TCs undergoing ET. These ET forecasts are sorted by their 72–120 h ATEs, with the lower and upper quintiles considered L20 and U20 cases, respectively. L20 cases involve observed TCs that accelerate northeastward, reaching speeds ~5–10 m s^{-1} faster than those associated

with U20. The ECMWF significantly underpredicts this acceleration of L20 TCs, such that the L20 ATEs grow rapidly with time after the first 36 h. For each case, the correlation is calculated between ensemble member ATEs at 72–120 h and their ATEs at earlier lead times. The correlations are small for lead times earlier than 48 h, but become significantly positive for a large percentage of L20 cases, implying that the slowest ensemble members at ~48 h remain the slowest at 72–120 h.

Composites reveal that the L20 cases tend to be associated with more amplified flow patterns that are more significantly underpredicted than the U20 by $T - 60$ h before the largest mean ATE. More specifically, 20 of the 27 L20 cases underpredict the amplitudes of an upstream 300-hPa trough and a downstream 300-hPa ridge. In the L20 cases, the ECMWF ensemble members underpredict the TC's 300–200-hPa divergent outflow throughout the forecast and its subsequent role in enhancing the trough–ridge couplet through the following

sequence: the TC's underpredicted outflow does not enhance the potential vorticity (PV) gradient along the eastern flank of the trough between $T - 72$ h and $T - 60$ h. The model trough remains too positively tilted, with an underprediction of both positive PV advection around the trough's tip and negative PV advection along the ridge. The positive PV advection is driven by the nondivergent wind, which acts with the irrotational wind to enhance the PV gradient. The 700-hPa cold air advection is underpredicted beneath the trough, corresponding to the insufficient 300-hPa height falls and hence the trough's underamplification. The extent of this cold air advection is influenced by the TC's underpredicted circulation and position errors. The underamplification of both the trough and the ridge result in a weaker meridional flow that does not adequately accelerate the TC northeastward.

The insufficient divergent outflow of the L20 cases can be attributed to lower precipitation rates near and north of the TC center. In agreement with past published case studies, the lower precipitation rates are a consequence of weaker 700-hPa moisture transport east of the TC at $T - 108$ h. Future work will further examine the error mechanisms in greater detail, determining the contributions of observation errors and model parameterizations in individual cases. Numerical simulations will also be run for select cases, in which small perturbations will be applied to lower-tropospheric moisture and wind fields east of the TC to assess their impacts on the atmospheric evolution and TC track.

The standardized differences between the 10 slowest and 10 fastest ensemble members of each forecast show consistency with the mean errors in depicting the sequence of events, implying that the more correct ensemble members behave more like the reanalyzed atmosphere. Thus, similar ensemble diagnostics can be adapted to aid forecasters in focusing on the more likely ensemble members for these cases, as demonstrated by Dong and Zhang (2016) and Ancell (2016). For example, when the 72-h track forecast of a potential ET case shows sensitivity to the 24-h amplitude of an upstream trough–ridge couplet, forecasters may give more weight to the faster members if observations 24 h later indicate that the couplet is underamplified. However, developing a statistical framework would require a larger sample size.

REFERENCES

Ancell, B. C., 2016: Improving high-impact forecasts through sensitivity-based ensemble subsets: Demonstration and initial tests. *Wea. Forecasting*, **31**, 1019–1036, https://doi.org/10.1175/WAF-D-15-0121.1.

Archambault, H. M., L. F. Bosart, D. Keyser, and J. M. Cordeira, 2013: A climatological analysis of the extratropical flow response to recurving western North Pacific tropical cyclones. *Mon. Wea. Rev.*, **141**, 2325–2346, https://doi.org/10.1175/MWR-D-12-00257.1.

——, D. Keyser, L. F. Bosart, C. A. Davis, and J. M. Cordeira, 2015: A composite perspective of the extratropical flow response to recurving western North Pacific tropical cyclones. *Mon. Wea. Rev.*, **143**, 1122–1141, https://doi.org/10.1175/MWR-D-14-00270.1.

Atallah, E. H., and L. F. Bosart, 2003: The extratropical transition and precipitation distribution of Hurricane Floyd (1999). *Mon. Wea. Rev.*, **131**, 1063–1081, https://doi.org/10.1175/1520-0493(2003)131<1063:TETAPD>2.0.CO;2.

Bassill, N. P., 2014: Accuracy of early GFS and ECMWF Sandy (2012) track forecasts: Evidence for a dependence on cumulus parameterization. *Geophys. Res. Lett.*, **41**, 3274–3281, https://doi.org/10.1002/2014GL059839.

Bieli, M., S. J. Camargo, A. H. Sobel, J. L. Evans, and T. Hall, 2019: A global climatology of extratropical transition. Part I: Characteristics across basins. *J. Climate*, **32**, 3557–3582, https://doi.org/10.1175/JCLI-D-17-0518.1.

Bougeault, P., and Coauthors, 2010: The THORPEX Interactive Grand Global Ensemble. *Bull. Amer. Meteor. Soc.*, **91**, 1059–1072, https://doi.org/10.1175/2010BAMS2853.1.

Brennan, M. J., D. T. Kleist, K. Howard, and S. J. Majumdar, 2015: The impact of supplemental dropwindsonde data on the structure and intensity of Tropical Storm Karen (2013) in the NCEP Global Forecast System. *Wea. Forecasting*, **30**, 683–691, https://doi.org/10.1175/WAF-D-15-0002.1.

Buizza, R., J.-R. Bidlot, N. Wedi, M. Fuentes, M. Hamrud, G. Holt, and F. Vitart, 2007: The new ECMWF VAREPS (Variable-Resolution Ensemble Prediction System). *Quart. J. Roy. Meteor. Soc.*, **133**, 681–695, https://doi.org/10.1002/qj.75.

Cammas, J.-P., and D. Ramond, 1989: Analysis and diagnosis of the composition of ageostrophic circulations in jet-front systems. *Mon. Wea. Rev.*, **117**, 2447–2462, https://doi.org/10.1175/1520-0493(1989)117<2447:AADOTC>2.0.CO;2.

Cangialosi, J., 2018: National Hurricane Center forecast verification report: 2017 hurricane season. NOAA/National Hurricane Center, 73 pp., https://www.nhc.noaa.gov/verification/pdfs/Verification_2017.pdf.

Carr, L. E., III, and R. L. Elsberry, 2000: Dynamical tropical cyclone track forecast errors. Part II: Midlatitude circulation influences. *Wea. Forecasting*, **15**, 662–681, https://doi.org/10.1175/1520-0434(2000)015<0662:DTCTFE>2.0.CO;2.

Clark, A. J., C. J. Schaffer, W. A. Gallus Jr., and K. Johnson-O'Mara, 2009: Climatology of storm reports relative to upper-level jet streaks. *Wea. Forecasting*, **24**, 1032–1051, https://doi.org/10.1175/2009WAF2222216.1.

Dong, L., and F. Zhang, 2016: OBEST: An observation-based ensemble subsetting technique for tropical cyclone track prediction. *Wea. Forecasting*, **31**, 57–70, https://doi.org/10.1175/WAF-D-15-0056.1.

Evans, C., and Coauthors, 2017: The extratropical transition of tropical cyclones. Part I: Cyclone evolution and direct impacts. *Mon. Wea. Rev.*, **145**, 4317–4344, https://doi.org/10.1175/MWR-D-17-0027.1.

Evans, J. L., and R. E. Hart, 2003: Objective indicators of the life cycle evolution of extratropical transition for Atlantic tropical cyclones. *Mon. Wea. Rev.*, **131**, 909–925, https://doi.org/10.1175/1520-0493(2003)131<0909:OIOTLC>2.0.CO;2.

Fowler, J. P., and T. J. Galarneau Jr., 2017: Influence of storm–storm and storm–environment interactions on tropical cyclone formation and evolution. *Mon. Wea. Rev.*, **145**, 4855–4875, https://doi.org/10.1175/MWR-D-17-0131.1.

Galarneau, T. J., 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.

Grams, C. M., S. C. Jones, C. A. Davis, P. A. Harr, and M. Weissmann, 2013: The impact of Typhoon Jangmi (2008) on the midlatitude flow. Part I: Upper-level ridgebuilding and modification of the jet. *Quart. J. Roy. Meteor. Soc.*, **139**, 2148–2164, https://doi.org/10.1002/qj.2091.

Hart, R. E., 2001: The extratropical transition of Atlantic tropical cyclones: Climatology, lifecycle definition, and a case study. Ph.D. thesis, The Pennsylvania State University, 176 pp. [Available from The Pennsylvania State University, 503 Walker Bldg., University Park, PA 16802.]

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

——, and J. L. Evans, 2001: A climatology of the extratropical transition of Atlantic tropical cyclones. *J. Climate*, **14**, 546–564, https://doi.org/10.1175/1520-0442(2001)014<0546:ACOTET>2.0.CO;2.

——, ——, and C. Evans, 2006: Synoptic composites of the extratropical transition life cycle of North Atlantic tropical cyclones: Factors determining posttransition evolution. *Mon. Wea. Rev.*, **134**, 553–578, https://doi.org/10.1175/MWR3082.1.

Hoerling, M. P., T. D. Schaack, and A. J. Lenzen, 1991: Global objective tropopause analysis. *Mon. Wea. Rev.*, **119**, 1816–1831, https://doi.org/10.1175/1520-0493(1991)119<1816:GOTA>2.0.CO;2.

Holton, J. R., and G. J. Hakim, 2013: *An Introduction to Dynamic Meteorology.* 5th ed. Elsevier, 532 pp.

——, P. H. Haynes, M. E. McIntyre, A. R. Douglass, R. B. Hood, and L. Pfister, 1995: Stratosphere–troposphere exchange. *Rev. Geophys.*, **33**, 403–439, https://doi.org/10.1029/95RG02097.

Jones, S. C., and Coauthors, 2003: The extratropical transition of tropical cyclones: Forecast challenges, current understanding, and future directions. *Wea. Forecasting*, **18**, 1052–1092, https://doi.org/10.1175/1520-0434(2003)018<1052:TETOTC>2.0.CO;2.

Kehoe, R. M., M. A. Boothe, and R. L. Elsberry, 2007: Dynamical tropical cyclone 96- and 120-h track forecast errors in the western North Pacific. *Wea. Forecasting*, **22**, 520–538, https://doi.org/10.1175/WAF1002.1.

Keller, J. H., 2017: Amplification of the downstream wave train during extratropical transition: Sensitivity studies. *Mon. Wea. Rev.*, **145**, 1529–1548, https://doi.org/10.1175/MWR-D-16-0193.1.

——, and Coauthors, 2019: The extratropical transition of tropical cyclones. Part II: Interaction with the midlatitude flow, downstream impacts, and implications for predictability. *Mon. Wea. Rev.*, **147**, 1077–1106, https://doi.org/10.1175/MWR-D-17-0329.1.

Klein, P. M., P. A. Harr, and R. L. Elsberry, 2000: Extratropical transition of western North Pacific tropical cyclones: An overview and conceptual model of the transformation stage. *Wea. Forecasting*, **15**, 373–395, https://doi.org/10.1175/1520-0434(2000)015<0373:ETOWNP>2.0.CO;2.

Leonardo, N. M., and B. A. Colle, 2017: Verification of multimodel ensemble forecasts of North Atlantic tropical cyclones. *Wea. Forecasting*, **32**, 2083–2101, https://doi.org/10.1175/WAF-D-17-0058.1.

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.

——, J. A. Sippel, S. A. Braun, Y. Weng, and F. Zhang, 2015: Dynamics and predictability of Hurricane Nadine (2012) evaluated through convection-permitting ensemble analysis and forecasts. *Mon. Wea. Rev.*, **143**, 4514–4532, https://doi.org/10.1175/MWR-D-14-00358.1.

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.

Peng, X., J. Fei, X. Huang, and X. Cheng, 2017: Evaluation and error analysis of official forecasts of tropical cyclones during 2005–14 over the Western North Pacific. Part I: Storm tracks. *Wea. Forecasting*, **32**, 689–712, https://doi.org/10.1175/WAF-D-16-0043.1.

Riemer, M., and S. C. Jones, 2010: The downstream impact of tropical cyclones on a developing baroclinic wave in idealized scenarios of extratropical transition. *Quart. J. Roy. Meteor. Soc.*, **136**, 617–637, https://doi.org/10.1002/qj.605.

——, and ——, 2014: Interaction of a tropical cyclone with a high-amplitude, midlatitude wave pattern: Waviness analysis, trough deformation and track bifurcation. *Quart. J. Roy. Meteor. Soc.*, **140**, 1362–1376, https://doi.org/10.1002/qj.2221.

——, ——, and C. A. Davis, 2008: The impact of extratropical transition on the downstream flow: An idealized modelling study with a straight jet. *Quart. J. Roy. Meteor. Soc.*, **134**, 69–91, https://doi.org/10.1002/qj.189.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, https://doi.org/10.1175/2010BAMS3001.1.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., https://doi.org/10.5065/D68S4MVH.

Torn, R. D., 2010: Diagnosing the factors that influence downstream ridging during extratropical transition. *J. Atmos. Sci.*, **67**, 817–833, https://doi.org/10.1175/2009JAS3093.1.

——, and C. A. Davis, 2012: The influence of shallow convection on tropical cyclone track forecasts. *Mon. Wea. Rev.*, **140**, 2188–2197, https://doi.org/10.1175/MWR-D-11-00246.1.

——, 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.

Veren, D., J. L. Evans, S. Jones, and F. Chiaromonte, 2009: Novel metrics for evaluation of ensemble forecasts of tropical cyclone structure. *Mon. Wea. Rev.*, **137**, 2830–2850, https://doi.org/10.1175/2009MWR2655.1.

Wirtz, A., W. Kron, P. Löw, and M. Steuer, 2014: The need for data: Natural disasters and the challenges of database management. *Nat. Hazards*, **70**, 135–157, https://doi.org/10.1007/s11069-012-0312-4.

Zhang, Z., and B. A. Colle, 2017: Changes in extratropical cyclone precipitation and associated processes during the twenty-first century over eastern North America and the western Atlantic using a cyclone-relative approach. *J. Climate*, **30**, 8633–8656, https://doi.org/10.1175/JCLI-D-16-0906.1.

Zwiers, F. W., 1990: The effect of serial correlation on statistical inferences made with resampling procedures. *J. Climate*, **3**, 1452–1461, https://doi.org/10.1175/1520-0442(1990)003<1452:TEOSCO>2.0.CO;2.