Verification of Rapid Refresh and High-Resolution Rapid Refresh Model Variables in Tornadic Tropical Cyclones

LELAND M. MACDONALD[a] AND CHRISTOPHER J. NOWOTARSKI[a]

[a] *Texas A &M University, College Station, Texas*

(Manuscript received 24 June 2022, in final form 25 February 2023, accepted 4 March 2023)

ABSTRACT: Tropical cyclone tornadoes (TCTORs) are a hazard to life and property during landfalling tropical cyclones (TCs). The threat is often spread over a wide area within the TC envelope and must be continually evaluated as the TC moves inland and dissipates. To anticipate the risk of TCTORs, forecasters may use high-resolution, rapidly updating model analyses and short-range forecasts such as the Rapid Refresh (RAP) and High-Resolution Rapid Refresh (HRRR), and an ingredients-based approach similar to that used for forecasting continental midlatitude tornadoes. Though RAP and HRRR errors have been identified in typical midlatitude convective environments, this study evaluates the performance of the RAP and the HRRR within the TC envelope, with particular attention given to sounding-derived parameters previously identified as useful for TCTOR forecasting. A sample of 1730 observed upper-air soundings is sourced from 13 TCs that made landfall along the U.S. coastline between 2017 and 2019. The observed soundings are paired with their corresponding model gridpoint soundings from the RAP analysis, RAP 12-h forecast, and HRRR 12-h forecast. Model errors are calculated for both the raw sounding variables of temperature, dewpoint, and wind speed, as well as for the selected sounding-derived parameters. Results show a moist bias that worsens with height across all model runs. There are also statistically significant underpredictions in stability-related parameters such as convective available potential energy (CAPE) and kinematic parameters such as vertical wind shear.

KEYWORDS: Atmosphere; North America; Tornadoes; Tropical cyclones; Model errors; Model evaluation/performance

1. Introduction

Tropical cyclone tornadoes (TCTORs) present a critical public hazard and forecasting challenge within landfalling and remnant tropical cyclones (TCs). TCTORs are distributed over wide spatial scales and temporal windows within the TC envelope (e.g., Edwards 2012), and the evolving tornado threat must be monitored as the TC approaches the coast, moves inland, and dissipates. To track current conditions or anticipated short-term changes in mesoscale features and variables critical to midlatitude, supercellular tornado production, forecasters often rely on models such as the Rapid Refresh (RAP) and High-Resolution Rapid Refresh (HRRR), which have high spatial and temporal resolution and frequently assimilate new environmental data (Smith et al. 2008; Benjamin et al. 2016).

As described below, many studies have examined observations and distributions of TCTORs, the characteristics of their near-cell environments, and the characteristics of the broad TC environment. A few have examined RAP and HRRR model performance in continental environments, but not within tropical cyclone envelopes as was recommended as future work by Edwards (2012). Thus, we use observed and model atmospheric soundings to evaluate the representation of TCTOR-related variables in the widely used RAP and HRRR forecast models specifically within TC envelopes. The existing literature (described in section 2) on TCTORs and RAP and HRRR model errors in continental environments motivate the following research questions:

1) What are the height-dependent model biases in temperature, dewpoint, and winds in the RAP analysis, and 12-h RAP and HRRR forecasts relative to radiosonde observations within a TC envelope?
2) How does model performance vary as a function of distance/azimuth from the TC center, TC intensity, and time relative to TC landfall?
3) How do model biases translate to errors in sounding-derived parameters specifically related to TCTOR forecasting?

To address these questions, this study evaluates model biases in the RAP analysis, and RAP and HRRR forecasts, within the envelopes of TCs that made landfall during the 2017–19 Atlantic hurricane seasons. Errors in vertical profiles of temperature, dewpoint, and wind, as well as errors in sounding-derived parameters used in tornado forecasting, are examined spatially and temporally with respect to the TC. This model verification is the foundation of an ongoing larger effort to improve the TCTOR forecasting and warning process by better understanding differences in the near-cell environments of tornadic and nontornadic cells within TC rainbands.

2. Background

While TCTORs comprise less than 10% of overall U.S. tornado activity, they account for ~10%–25% of the overall

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/WAF-D-22-0117.s1.

Corresponding author: Christopher J. Nowotarski, cjnowotarski@tamu.edu

DOI: 10.1175/WAF-D-22-0117.1

tornado activity in each state bordering the Gulf and Atlantic coasts from Louisiana to Maryland (Edwards 2010; Schultz and Cecil 2009). From 1995 to 2021 1746 TCTORs occurred within the contiguous United States, resulting in at least 29 fatalities and 403 injuries (Edwards and Mosier 2022). These tornadoes are typically concentrated within 500 km of the coastline, but the threat of TCTORs can persist much farther inland and many days after the TC landfall (Edwards and Mosier 2022). Hurricane Ivan (2004) provided an extreme example of this, spawning 118 tornadoes in a multiday outbreak that occurred in three distinct geographical clusters along a path from the Gulf Coast of Florida to Maryland (Edwards 2012). Even in less extreme cases, however, the risk of TCTORs remains elevated for a substantial period, with the majority of TCTORs from 1950 to 2007 occurring between 12 h prior to landfall and 48 h after landfall (Schultz and Cecil 2009).

TCTORs most commonly occur in the sector from north to southeast of the TC center with the greatest frequency in the northeast quadrant (Edwards 2010; Edwards and Mosier 2022). TCTOR reports tend to shift toward the southeast as the TC weakens because this quadrant usually moves over land during the remnant phase and typically contains higher values of CAPE (Edwards 2012; McCaul 1991). Motion-relative quadrants are also often used to describe TCTOR occurrence; however, the distinctions between the two distribution patterns are minimal, outside unusual southward TC translation components (Schultz and Cecil 2009; Edwards and Mosier 2022). Schenkel et al. (2020) investigated TCTOR occurrence relative to the background deep-layer shear vector, finding the majority of TCTORs occur in the downshear-left quadrant of TCs, which often overlaps with the northeast or right-front quadrant for typical synoptic patterns and TC motions (though exceptions exist). Most reports fall within 100–500 km of the TC center (Edwards 2012). Thus, even in highly destructive TCs with large hurricane-force wind radii, TCTORs can increase local risk because they occur outside of the area that prepared for hurricane-force winds. Hence, forecasters must closely monitor the favorability of the environment for TCTOR production at the regional scale.

As in nontropical environments, the most prevalent tornadic storm modes within TCs are supercellular. The three most common tornadic storm modes in TCs are: supercells in clusters, discrete right-moving supercells, and supercells embedded in QLCSs. These account for around 80% of TCTOR events (Edwards et al. 2012); however, supercells within TCs are generally shallower and narrower, with weaker mesocyclones than those of their nontropical counterparts (Spratt et al. 1997; Edwards 2012). TCTORs also tend to be weaker than nontropical tornadoes. Edwards and Mosier (2022) found that 95.4% of TCTORs from 1995 to 2021 were classified as weak (F/EF0–1), while just 87.4% of overall U.S. tornadoes fit this category. Stronger tornadoes (F/EF2+) occurred only 6.3% of the time in TCs, compared to 11.2% of overall U.S. tornadoes. The diurnal distribution of TCTORs varies from non-TC tornadoes. Most TCTORs occur between 0900 and 1800 local time, with a distinct peak between 1500 and 1800 local time; this is earlier than the early evening maximum in non-TC tornado frequency (McCaul 1991). Regardless, nocturnal tornadoes are more common in TCs than in the overall U.S. tornado record (Schultz and Cecil 2009).

Systematic differences between TC and non-TC tornado environments emerge in moisture, instability, and shear characteristics of the near-cell environment of their parent cells. Moisture is more abundant throughout the column in the TC environment, as demonstrated by higher values of precipitable water compared to those in non-TC atmospheric profiles (Edwards et al. 2012). CAPE values, on the other hand, tend to be lower throughout the TC envelope than in typical midlatitude convective environments (McCaul 1991). In a study of non-TC versus TC supercell tornadoes from 2003 to 2011, non-TC tornado environments had a median MLCAPE of 1240 J kg^{-1} as opposed to 547 J kg^{-1} in TCTOR environments (Edwards et al. 2012). CAPE is largest in the outer regions of the TC on the right side of the track, especially toward the right-rear quadrant (generally the southeast quadrant), and CAPE decreases inward toward the TC center to values less than 400 J kg^{-1} (McCaul 1991). Surface-based CAPE (SBCAPE) can be enhanced beneath cloud-free slots within the TC envelope that develop through midtropospheric drying (Curtis 2004). However, the warm-core nature of TCs produces weak thermal lapse rates aloft that tend to cap the vertical extent of the buoyancy at just a few kilometers above ground level, around the 600-mb (1 mb = 1 hPa) level (Edwards 2012; McCaul 1991).

Low-level vertical wind shear is generally stronger in TC environments than in non-TC environments (McCaul 1991). Moreover, low-level shear is stronger and tornado production is more likely in more intense TCs (Verbout et al. 2007) and larger TCs (Paredes et al. 2021). Vertical shear decreases with radial distance from the TC center, and the strongest shear and largest storm-relative helicity (SRH) values occur in the northeast or right-front quadrant (McCaul 1991) or downshear side of the TC (Molinari and Vollaro 2010). In these locations, and particularly the downshear left quadrant, the background synoptic-scale flow in the middle and upper troposphere most often overlaps favorably with the TC wind structure to enhance deep-layer shear and support wind profiles that veer continuously with height (Schenkel et al. 2020). As TCs move farther into the midlatitudes after landfall, they often encounter stronger westerly flow aloft. This can maintain and even enhance vertical wind shear with time when coupled with frictional slowing of the wind near the surface. Helicity parameters also tend to increase with time after landfall (Gentry 1983; McCaul 1991).

Despite differences in the distributions of meteorological variables between the two environment types, the traditional ingredients-based approach to non-TC tornado forecasting is also used in TCTOR forecasting. Since low-level moisture is abundant within the TC envelope, TCTOR forecasting relies heavily on identifying regions of favorable instability and shear in proximity to mesoscale boundaries or other lifting mechanisms. Identifying short-term trends, recent or forecasted, in such variables is crucial when diagnosing TCTOR potential (Edwards 2012). Forecasters also examine mid and upper-level features, such as the superpositioning of background flow to produce wind profiles conducive to TCTOR

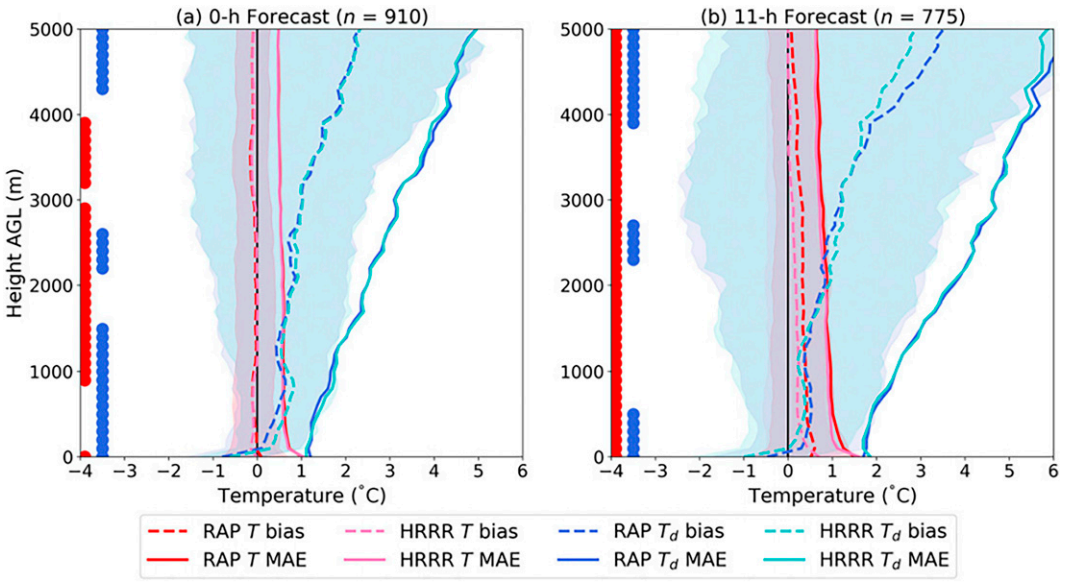

FIG. 1. (a) Vertical profiles of sample-mean bias (°C; dashed lines; defined as model minus observations) and mean absolute error (MAE; °C; solid lines) between 0 and 5 km AGL for RAP and HRRR 0-h temperature (red and pink lines, respectively) and dewpoint temperature (dark blue and light blue lines, respectively) analyses. Shading depicts the interquartile ranges of the error distributions for each variable and model. Solid red and blue dots indicate vertical levels at which the temperature and dewpoint temperature error distributions, respectively, between the RAP and HRRR are significantly different to at least 95% confidence, as assessed using the two-tailed, nonparametric Wilcoxon signed-rank test. The number of observations contributing to each sample is depicted above each panel. (b) As in (a), but for 11-h forecasts. [Figure from Fig. 4 and caption from Fig. 3 in Evans et al. (2018).]

production, or the progression of midtropospheric drying visible on 700- or 500-mb relative humidity maps that could be an early sign of a TCTOR outbreak (Schenkel et al. 2020; Curtis 2004). Nowotarski et al. (2021) showed that, among kinematic parameters, 0–6-km shear and 0–1-km SRH best discriminated between tornadic and nontornadic cells within Hurricane Harvey. Mean values of 100-mb MLCAPE, 0–3-km lapse rate, and the significant tornado parameter (STP) were greater for tornadic versus nontornadic environments in their study. These findings align with those of Edwards et al. (2012), who showed larger values of MLCAPE, 0–6-km shear, and STP in the near-cell environments of strong TCTORs (EF2 and EF3) than in the near-cell environments of weak TCTORs (EF0 and EF1). Additionally, Davies (2006) found that the 0–1-km energy helicity index (EHI), which combines parcel CAPE and 0–1-km SRH into a single parameter, was notably greater in cases with TCTORs rated EF2 or higher than in nontornadic environments.

To evaluate these rapidly evolving conditions on small spatial scales within the TC envelope, forecasters employ high-resolution models such as the Rapid Refresh model, or RAP (Benjamin et al. 2016). The RAP model domain covers the entire North American region at 13-km horizontal resolution and at 51 vertical levels. It assimilates radiosonde data as well as observations from surface weather stations, ships, aircraft, radars, and satellites. The RAP model analysis is issued every hour, and the RAP forecast is issued every 3 h. Each forecast is for at least 18 h with hourly time resolution (Benjamin et al. 2016). The RAP analysis forms the basis of the Real-Time Mesoscale Analysis (RTMA), produced hourly by the National Centers for Environmental Prediction (NCEP), and the hourly Storm Prediction Center (SPC) mesoanalysis fields, both of which are important tools for now-casting (De Pondeca et al. 2011; Storm Prediction Center 2016). The RAP analysis also provides the initial and lateral boundary conditions for the 3-km High-Resolution Rapid Refresh (HRRR) model, which then assimilates radar data before generating hourly convection-allowing forecast grids that are used in short-term severe weather forecasting (Evans et al. 2018). The HRRR forecasts are generally preferred over RAP forecasts in complex convective environments due to their hourly updates and superior spatial resolution.

Evans et al. (2018) calculated model biases for both RAP and HRRR analyses and 11-h forecasts in environments conducive to convection, as defined by SPC day 1 convective outlooks issued during May 2017. In the lowest 5 km of the atmosphere, they found small biases in temperature ($<\sim$0.25°C) in both of the analyses (Fig. 1a). Both of the 11-h forecast temperature profiles were consistently warm-biased from 0 to 5 km, with the largest warm bias in both models occurring near the surface and the RAP warm bias slightly greater than that of the HRRR above the surface layer (Fig. 1b). Surface dewpoints were slightly dry biased across both models and lead times. However, there was a pronounced moist bias above the surface level in both models that worsened with increasing altitude and lead time, exceeding

TABLE 1. The names, landfall dates, maximum intensities, and Saffir–Simpson categories at maximum intensity of the TCs used in this study, and the number of TCTORs associated with each. The minimum number of sounding pairs (observed sounding with its corresponding model sounding) analyzed for each storm is also presented. For all TCs except Dorian, this is the number of HRRR 12-h forecast sounding pairs; for Dorian it is the number of RAP 12-h forecast sounding pairs.

TC name	Mainland U.S. landfall date	No. of TCTORs	Peak intensity (kt)	Category at peak intensity	Min No. of sounding pairs analyzed
Cindy	22 Jun 2017	18	50	TS	97
Emily	31 Jul 2017	1	50	TS	52
Harvey	25 Aug 2017	52	115	4	261
Irma	10 Sep 2017	28	155	5	104
Nate	8 Oct 2017	21	80	1	102
Alberto	29 May 2018	4	55	TS	163
Florence	14 Sep 2018	44	130	4	161
Gordon	4 Sep 2018	7	60	TS	117
Michael	10 Oct 2018	16	140	5	118
Barry	13 Jul 2019	1	65	1	195
Dorian	6 Sep 2019	25	160	5	162
Imelda	17 Sep 2019	2	40	TS	45
Nestor	19 Oct 2019	6	50	TS	27

3°C in the RAP 11-h dewpoint forecast around an altitude of 5 km (Fig. 1b).

Evans et al. (2018) also examined sounding-derived parameter biases. MLCAPE was overestimated in both the analyses and 11-h forecasts due to the moist bias present above the surface, but surface-based (SB) and most-unstable (MU) CAPE were underpredicted due to the surface dry bias. Laflin (2013) performed similar verification of the RAP model in springtime preconvective environments, although restricted to the central and northern Great Plains. They found underestimations in SBCAPE and MUCAPE for both 12- and 6-h model forecasts, which likely resulted from large dry model biases in the near-surface layer of the model. However, these dry biases extended farther above the surface than in the findings of Evans et al. (2018), so MLCAPE was also found to be underpredicted in the Laflin (2013) study. Fovell and Gallagher (2020) analyzed the representation in the contiguous United States of the lowest 1 km of the atmosphere in the HRRR model. They found small near-surface temperature biases that depended on station elevation. They also found small wind speed biases at the analysis time that became increasingly positive throughout the 24-h forecast period. All of these verification studies focused on continental conditions, such that it is unclear if these errors also exist within the envelope of landfalling TCs where moisture is generally much more prevalent, and the TC circulation itself results in atypical wind profiles. Thus, this study aims to quantify RAP analysis and RAP and HRRR forecast errors specifically within the context of the TC envelope and TCTOR forecasting using a similar methodology to Evans et al. (2018).

3. Data and methods

a. Tropical cyclone data

We analyze the environments of TCs from the 2017 to 2019 Atlantic hurricane seasons. Only the TCs that made landfall along the coastline of the contiguous United States are used o ensure adequate observed radiosonde data to compare to the model profiles. This yields 5 TCs from the 2017 season, 4 from 2018, and 4 from 2019, for a total of 13 TCs. A summary of the names, landfall dates, maximum intensities, and categories at peak intensity of these TCs is presented in Table 1, along with the number of TCTORs associated with each TC or its remnants. These data come from the publicly available official Tropical Cyclone Reports released by the National Hurricane Center, except for the TCTOR frequency, which is taken from the Edwards TCTOR database (Edwards 2010). Four of the TCs in this study have fewer than five TCTORs attributed to them. These TCs are still relevant because the purpose of this study is not to capture specific TCTOR cases or analyze near-cell environments. Rather, we broadly examine the distributions of environmental variables and model errors critical to TCTOR production, both spatially within the TC envelope and temporally as the TC moves inland and weakens.

TC track data are acquired from the Atlantic HURDAT2 dataset, which comprises poststorm analyzed TC best track data (Landsea and Franklin 2013). These are documented observations of the TC every 6 h at 0000, 0600, 1200, and 1800 UTC. An extra data point is often recorded near the time of landfall or at the TC's maximum intensity if these points do not coincide with a standard 6-hourly time point. The information collected at each point includes the date and time, the latitude and longitude of the TC center at that time, the mean sea level pressure at the TC center, and the maximum sustained wind (intensity) of the TC in knots. For this research, any points in the TC track data not occurring at 0000, 0600, 1200, or 1800 UTC are discarded, and the latitudes, longitudes, pressures, and intensities from the remaining 6-hourly points are interpolated linearly to yield hourly track data (Fig. 2).

b. Observed soundings

The TC data are used alongside a National Weather Service AWIPS2 dataset containing the latitudes, longitudes, and elevations of upper-air observing sites, hereafter "stations."

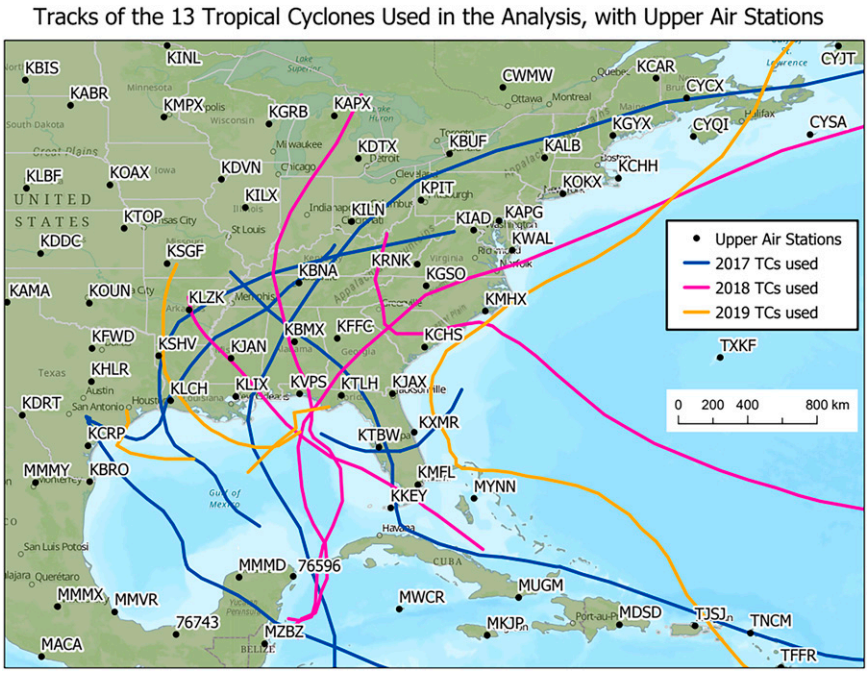

FIG. 2. The tracks of the 13 TCs used in this study, along with the locations of all available upper-air observing stations.

At each hourly time step along each TC track, the distance from the TC center to all the stations in the dataset is computed with the Haversine distance formula (Sinnott 1984), which accounts for the curvature of Earth over long spatial scales in its calculation of the distance between two locations. All stations falling within a radius of 800 km from the TC center at that given time are considered in this study and eligible soundings from these stations are collected (University of Wyoming 2020). This choice is consistent with the 800-km radius used by McCaul (1991) in their assessment of the spatial distributions of TCTOR-related environmental variables, and few TCTORs occur outside this radius (Edwards and Mosier 2022); however, the radius of influence of individual TCs may vary considerably.

Any available special soundings are collected in addition to the standard soundings at 0000 and 1200 UTC. Most of the soundings are obtained from National Weather Service forecast offices located in the United States, but data from upper-air observing stations in Mexico and the Caribbean are included as permitted by the 800-km radius criterion and the domains of the RAP and HRRR models. Each observed sounding contains pressure, height, temperature, dewpoint, and wind speed and direction, which are interpolated to 100-m intervals from 0 to 16 km above ground level (AGL).

c. *Model soundings*

The NCEP RAP model with 13-km horizontal grid spacing and the HRRR model with 3-km horizontal grid spacing are compared in this study. For each observed sounding, the RAP analysis vertical profile from the same time and the nearest grid point to the station is selected from the analysis grid downloaded from the National Centers for Environmental Information (NCEI) model data archive (NOAA/National Centers for Environmental Information 2020). Then, the corresponding 12-h RAP and HRRR forecasts valid at the sounding observation time are downloaded, and the vertical profile from the nearest model grid point to the station is selected. These analysis and forecast soundings contain pressure, height, temperature, relative humidity [converted to dewpoint via NCAR Command Language (NCL) software], and wind data and are interpolated in the same way as the observed soundings. RAP version (v) 4 and HRRRv3 were implemented at NCEP on 12 July 2018, so the model soundings from the beginning of this study come from RAPv3 and HRRRv2 while the remainder come from RAPv4 and HRRRv3 (NOAA/Global Systems Laboratory 2020). While we combine different model versions in our presentation of results, raw model variable errors are also computed separately for each version (supplemental Fig. 1 in the online supplemental material). The RAP analysis and forecasts are missing from 0300 UTC 5 September to 1200 UTC 7 September 2019, which limits the model data and the number of sounding pairs analyzed during Hurricane Dorian.

d. *Sounding-derived parameters*

The Sounding and Hodograph Analysis and Research Program in Python (SHARPpy; Blumberg et al. 2017) software is used to compute sounding-derived parameters for the observed, model analysis, and model forecast soundings. Though we considered a variety of common sounding-derived parameters available in SHARPpy, we focus our analysis below on the following variables. Bulk wind differences (representing vertical wind shear, and hereafter used interchangeably with

"shear") and lapse rates are calculated over various layers, along with SRH values. Without observed convective cell motions for each sounding pair, we calculate SRH using a sounding-derived right-moving supercell motion estimate (Bunkers et al. 2000) modified with the parcel-based approach consistent with Bunkers et al. (2014). Though convective cell motion is likely to deviate from this estimate to varying degrees in the TC envelope, this approach recognizes the often shallow top of TCTOR parent cells. CAPE, convective inhibition (CINH), lifted condensation level (LCL), level of free convection (LFC), and lifted index (LI) are determined for the surface-based, mixed-layer, and most-unstable parcels. The mixed-layer parcel has the averaged properties of the lowest 100 mb of the sounding, and the most-unstable parcel is the parcel with the highest equivalent potential temperature in the lowest 300 mb of the sounding. We also examine both the original fixed-layer STP (Thompson et al. 2003) and the CINH-adjusted effective layer STP (Thompson et al. 2012). All sounding-derived parameters are computed from the vertically interpolated soundings to remove the potential influence of the different native vertical levels of the datasets.

e. *Data analysis methods*

Quality control of the eligible model/observation sounding pairs includes removal of soundings with deep contiguous layers of missing data or clearly erroneous wind data, observations occurring outside of a particular model domain, or missing analysis/forecast data. This results in 1730 corresponding RAP analysis profiles, 1681 RAP 12-h forecast profiles, and 1647 HRRR 12-h forecast profiles. For each sounding pair, model errors are calculated for all raw sounding variables and sounding-derived parameters as the model value minus the observed value, such that a positive model error represents an overprediction of that quantity by the model and a negative model error represents an underprediction.

4. Results

a. *Errors in raw sounding variables*

Mean error (i.e., bias) and mean absolute error profiles of temperature, dewpoint, and wind speed are generated for the RAP analysis, RAP 12-h forecast, and HRRR 12-h forecast. The mean error profiles are calculated by subtracting each observed value from its corresponding model value at each interpolated height, then averaging across all available sounding pairs at that height. The two-tailed, nonparametric Wilcoxon signed-rank test (Wilcoxon 1945; Wilks 2011) is performed at each height of each mean error profile to test whether the median errors at each height are significantly different from zero at the 95% confidence level.

Mean temperature errors do not exceed a magnitude of 0.4°C at any height in any of the three error profiles (Fig. 3a). The RAP analysis (−0.36°C), RAP 12-h forecast (−0.16°C), and HRRR 12-h forecast (−0.12°C) all have cool biases at the surface. This surface cool bias is largest in magnitude for the RAP analysis, remains negative throughout much of the column up to about 9 km before changing to a warm bias in the upper troposphere. The HRRR 12-h forecast has a low-level warm bias starting just above the surface and extending up to 4 km; this changes to a cool bias between 4 and 11 km and then back to a warm bias above 11 km. The temperature errors in the RAP 12-h forecast are the largest of any of the three profiles within the troposphere. The surface cool bias in the RAP 12-h forecast persists up to 2 km and is followed by a warm bias from 2 to 5.5 km and a strong cool bias above that through the remainder of the troposphere. These results differ from those presented by Evans et al. (2018; Fig. 1b); both the RAP and HRRR forecasts in continental convective environments in their study were continuously warm biased in the 0–5-km layer.

Profiles of the mean absolute temperature errors (Fig. 3b) reveal the smallest mean absolute temperature errors in the RAP analysis. The RAP 12-h forecast exhibits the greatest mean absolute temperature error throughout the profile, and the magnitudes of HRRR 12-h forecast absolute temperature errors fall between those of the RAP 12-h forecast and RAP analysis.

The mean dewpoint errors (Fig. 3c) have larger magnitudes than the temperature errors and increase with height. The RAP analysis and HRRR 12-h forecasts have surface dry biases, but the HRRR bias is larger with a magnitude of nearly 0.5°C. Above the surface, the RAP analysis and both forecasts have moist biases that worsen with height, maximizing at nearly 4°C between 12 and 13 km. The difference between the RAP and HRRR 12-h dewpoint forecasts is significant at fewer height levels than for the mean temperature errors. The mean absolute dewpoint errors (Fig. 3d) are up to 1°C greater in the forecasts than in the RAP analysis, although no clear distinction exists between the two forecasts, similar to the mean absolute temperature error profiles (Fig. 3b).

The error profiles for the continental convective environments in Evans et al. (2018) were truncated at an altitude of 5 km, and our analysis shows a similar bias pattern below that height within TCs. In their study, both the RAP and HRRR forecasts as well as the RAP analysis had surface dry biases that quickly transitioned to moist biases increasing with height. The moist bias in the RAP analysis reached just over 2°C at 5 km (Fig. 1a), while the moist biases in the forecasts increased more rapidly to around 3°C at the same height (Fig. 1b). In this study of TC environments, the dewpoint biases are smaller within the lowest 5 km and the behavior of the forecasts and the RAP analysis is more similar; all three biases have a magnitude of around 1°C at 5 km (Fig. 3c). When contrasted with the continental results in Fig. 1, the results presented in Fig. 3c suggest the models predict moisture better within TC envelopes than in continental environments, at least below 5 km AGL.

The RAP analysis and both forecasts overpredict shallow, near-surface wind speeds (Fig. 3e). The largest near-surface overpredictions, exceeding 2.5 kt (1 kt ≈ 0.51 m s^{-1}), occur in the HRRR 12-h forecast. Above 1 km, all three error profiles transition to an underprediction of wind speed, with the largest underpredictions in the RAP 12-h forecast. The smallest underpredictions, never exceeding a magnitude of 0.75 kt, occur in the HRRR 12-h forecast with some layers having errors that are insignificant at the 95% confidence level. Despite these small biases in the HRRR 12-h forecast, the mean absolute

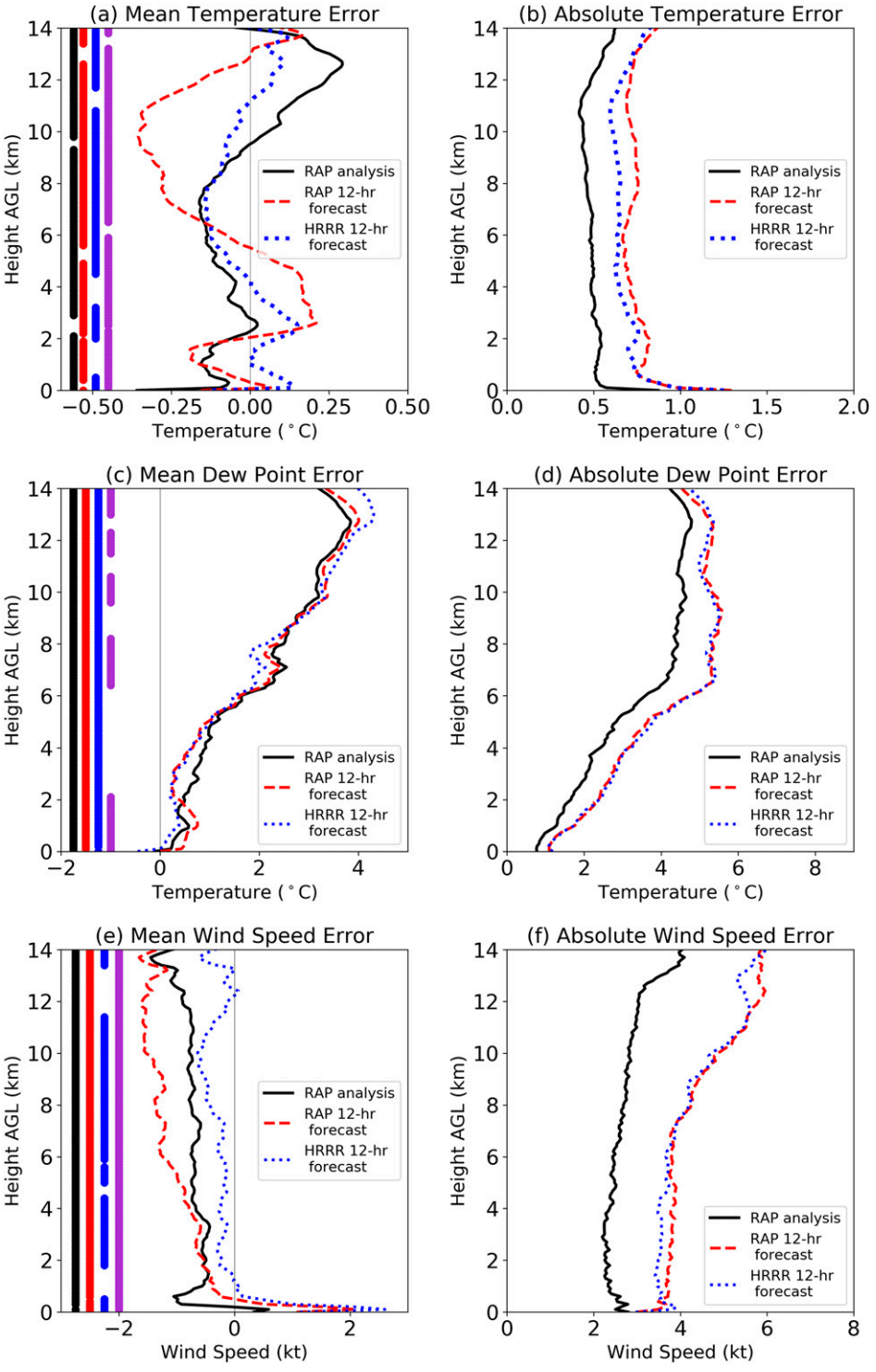

FIG. 3. Vertical profiles of mean error and mean absolute error for (a),(b) temperature; (c),(d) dewpoint; and (e),(f) wind speed from 0 to 14 km AGL for the RAP analysis, RAP 12-h forecast, and HRRR 12-h forecast. Solid black, red, and blue dots at the left edge of (a), (c), and (e) denote the levels at which the mean errors are significant to at least 95% confidence. Purple dots at the left edge of those panels denote the levels at which the RAP and HRRR 12-h forecast errors are significantly different from each other. The number of observed–model sounding pairs contributing to each error profile are as follows: RAP analysis—1730, RAP 12-h forecast—1681, HRRR 12-h forecast—1647.

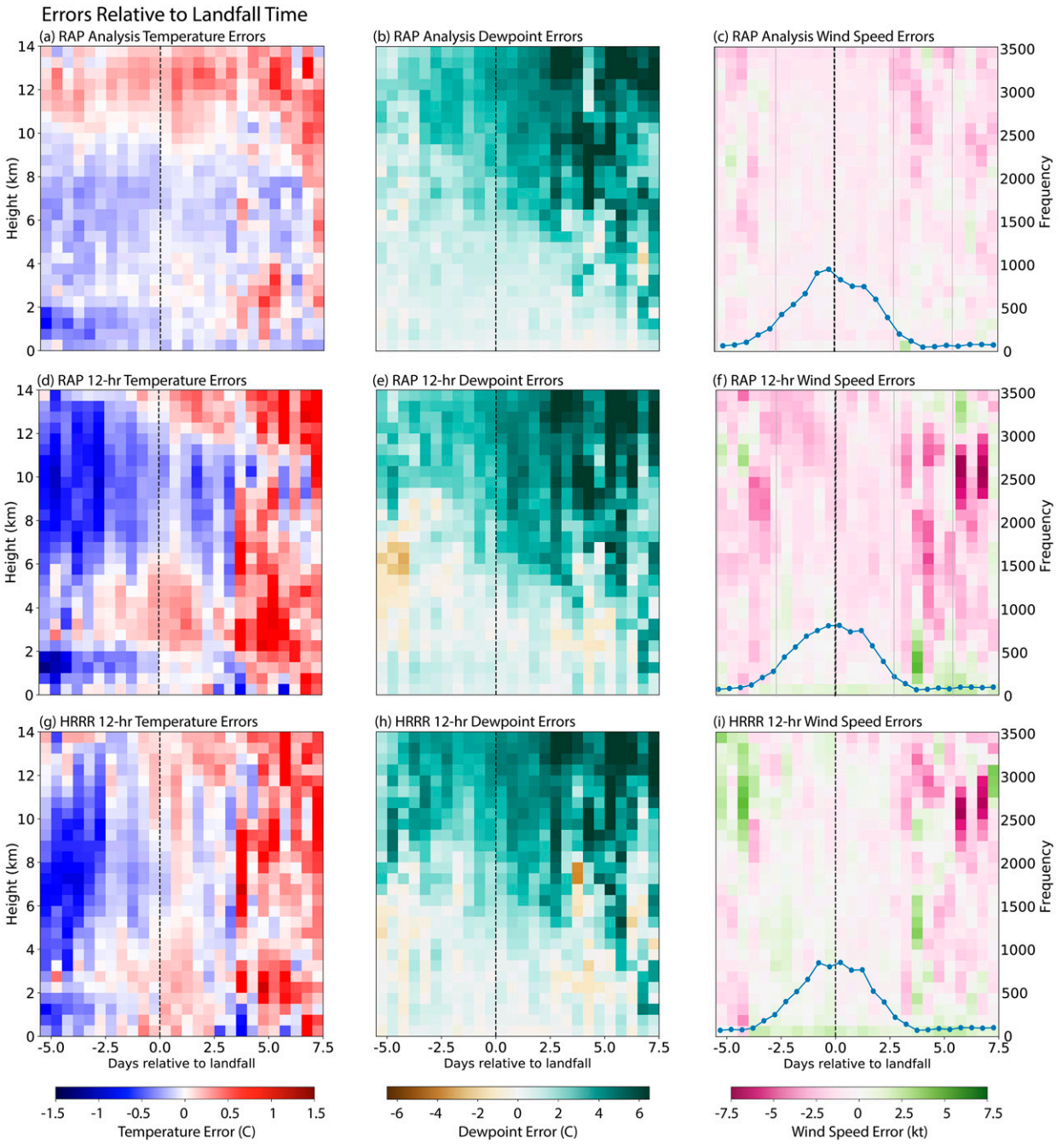

FIG. 4. (a)–(c) RAP analysis, (d)–(f) RAP 12-h forecast, and (g)–(i) HRRR 12-h forecast mean errors in temperature, dewpoint, and wind speed, respectively, with respect to days relative to landfall and height. The vertical dashed line at $x = 0$ represents the TC landfall time. The frequency of observation/model pairs (blue lines) is plotted for each bin and model in the right column figures.

error profiles (Fig. 3f) reveal that RAP analysis absolute wind speed errors tend to be smaller than those of both forecasts by 1–2.5 kt. Positive wind speed biases at the surface and negative wind speed biases aloft are expected to result in underpredictions of vertical wind shear quantities for layers based at the surface; shear will be explored along with other sounding-derived parameters in section 4b below.

Comparison of the RAP analysis, RAP 12-h forecast, and HRRR 12-h forecast errors as a function of height and time relative to TC landfall in Fig. 4 reveals that the temperature and dewpoint biases have larger magnitudes overall in the two forecasts than in the RAP analysis. Before landfall, the RAP analysis is cold biased up to about 10 km (Fig. 4a), consistent with the overall model temperature errors seen in Fig. 3a.

However, this cold bias switches to an unusual low-level warm bias, from the surface up to about 4 km, starting about 3 days after landfall. A strong warm bias is also seen in both the RAP and HRRR 12-h forecasts beginning about three days after landfall, but in both forecasts this occurs after a weaker warm bias has already appeared closer to, and even slightly before, the landfall time (Figs. 4d,g). This bias may be an artifact of the reduced sample size starting three days after landfall, where four TCs (Harvey, Irma, Florence, and Barry) had several model/observation pairs with considerable (>1°C) warm biases. Similar to the RAP analysis, both the RAP and the HRRR 12-h forecasts exhibit a cold bias in the days leading up to landfall, which is retained in the RAP 12-h forecast between 6 and 12 km for 3–4 days after landfall (Fig. 4d).

The moist bias evident in the dewpoint errors of the analysis and both forecasts descend with time, suggesting that the RAP and HRRR are too slow to dry out the environment as the TC moves farther inland and conditions become more typical of the midlatitudes (Figs. 4b,e,h). Wind speed (Fig. 4c) is weakly underpredicted by the RAP analysis for all times relative to the TC landfall. The RAP and HRRR 12-h forecasts slightly overpredict wind speed in the bins from 0 to 0.5 km, with slightly stronger underpredictions than those of the RAP analysis above that and across the rest of the time relative to landfall (Figs. 4f,i).

We also analyze model errors as a function of distance from the center of the TC, which ranges from 0 to 800 km in this dataset. With the exception of the innermost ring within 200 km, RAP analysis temperature and dewpoint errors (Figs. 5a,b) are nearly uniform as distance from the TC center increases. They mimic the RAP analysis mean error profiles for temperature and dewpoint (Figs. 3a,b), with negative temperature errors from the surface up to about 10 km and dewpoint errors gradually increasing with height throughout the depth of the troposphere. The temperature errors are more variable with height within the innermost 200 km, where cold biases from roughly 0–4 to 6–8 km create an alternating pattern with the warm biases, which extend from 4–6 to 8–14 km. We suspect this pattern of temperature errors is due to both the relative lack of observations in this region, as well as the complex nature of the interior region of a TC, with quickly evolving intense rainbands, isolated convective cells, and patches of clearing. The RAP and HRRR 12-h forecast dewpoint errors behave similarly to those of the RAP analysis, showing little to no variation with distance (Figs. 5e,h). Both forecasts show larger overpredictions of temperature within the innermost 200 km. Between 200 and 800 km from the TC center, temperature is generally overpredicted at low levels and underpredicted aloft in the HRRR 12-h forecast (Fig. 5g). In the RAP 12-h forecast, the temperature errors have a slightly larger magnitude and take on a more layered structure: temperature is underpredicted at low levels, overpredicted at midlevels, and underpredicted again above 6 km (Fig. 5d).

For both forecasts and the analysis, the magnitudes of the temperature errors are smaller at weaker tropical storm intensities and increase as the TC intensity increases (Figs. 6a,d,g). However, the dewpoint errors are larger at weaker intensities for both models, with mid- to upper-tropospheric moisture (above 6 km) more severely overpredicted at intensities weaker than hurricane strength. At stronger hurricane intensities, midlevel moisture is occasionally underpredicted (Figs. 6b,e,h). At intensities greater than 110 kt, however, the frequency of observations drops off considerably such that results may be skewed toward individual events (Figs. 6c,f,i). For instance, the atypical warm "streak" in the RAP and HRRR forecasts at 110–120-kt intensities (Figs. 6a,d) are driven by a handful of sounding pairs with large warm forecast errors in Texas and Louisiana during Hurricane Harvey at 0000 UTC 26 August 2017.

With respect to azimuth from the TC center, the RAP analysis has a relatively weak low-level cold bias in all quadrants (Fig. 7a). This bias is worse in the southern quadrants (from 90° to 270°) in the forecasts (Figs. 7d,g), but switches to a low-level warm bias in the northeast and north-northwest directions. Both forecasts are warm biased from 2 to 5 km on the east side of the TC (from 0° to 180°). In the analysis and both forecasts, the upper-level moist bias is stronger on the western half of the TC (Figs. 7b,e,h). Azimuthal trends in wind speed errors are less clear, though the HRRR forecast has a greater tendency to overpredict mid and upper-level wind speeds on the western half of the TC relative to the RAP analysis and forecasts (Figs. 7c,f,i).

b. *Errors in sounding-derived parameters*

We analyze errors in selected sounding-derived parameters that have been shown in previous studies to display predictive utility for TCTORs: MLCAPE, 0–6-km shear, and the significant tornado parameter (Edwards et al. 2012; Nowotarski et al. 2021), as well as 0–1-km SRH and 0–3-km lapse rate (Nowotarski et al. 2021). Sounding-derived parameter errors are computed by first calculating the values of the parameters for each observed and model sounding. Then, the observed value is subtracted from the model value to generate the dataset of error values, discarding pairs where either value is undefined. The statistical significance of the errors is assessed similarly for the raw variable profiles. In general, errors in sounding-derived parameters are consistent with those in their parent variables discussed above.

Surface-based, most-unstable, and mixed-layer CAPE errors are examined first as a function of time relative to landfall and TC intensity (Fig. 8). MLCAPE errors (Figs. 8c,f) tend to be smaller in magnitude, but positive as opposed to the generally negative biases in SBCAPE and MUCAPE (Figs. 8a,b,d,e). The HRRR 12-h forecast has larger negative biases than the RAP analysis and forecast in SBCAPE and MUCAPE within the region of statistically significant results (Figs. 8a,b,d,e). This is likely because the HRRR 12-h forecast has the largest dry bias at the surface (Fig. 3c). Conversely, the HRRR 12-h forecast outperforms the RAP 12-h forecast in its prediction of MLCAPE; this could be attributed to the slightly smaller dewpoint errors in the HRRR 12-h forecast throughout the typical depth of the mixed layer (Fig. 3c) and generally smaller temperature errors aloft (Figs. 3a,b).

Absolute CAPE errors tend to decrease with time following landfall (Figs. 8a–c) and increase for stronger TCs

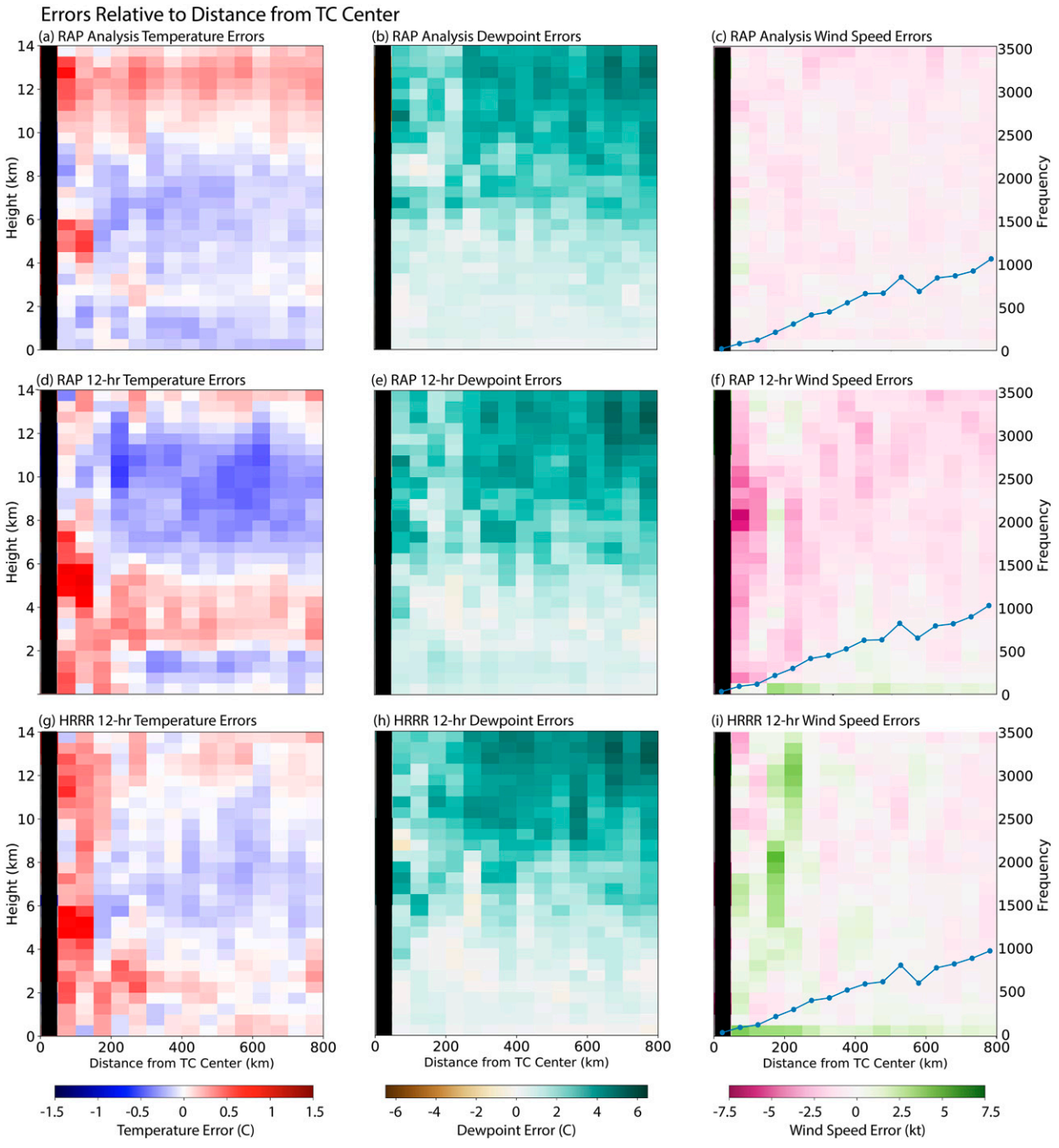

FIG. 5. As in Fig. 4, but plotted relative to distance from the TC center. Columns shown in black have a sample size of fewer than four sounding pairs.

(Figs. 8d–f). Similar to raw temperature and moisture errors (Figs. 3b,d), the mean absolute error in CAPE is typically smaller in the RAP analysis than either of the forecasts, however the RAP analysis is often more biased than the RAP forecast in SBCAPE and MUCAPE. While the magnitudes of the MLCAPE mean errors decrease during the four days following landfall, the magnitudes of the SBCAPE and MUCAPE mean errors remain steadier as time after landfall increases. The mean absolute error decreases for all CAPE varieties and all model runs, as the TC approaches its landfall time and moves inland and more land-based observations can be assimilated into the model (Figs. 8a–c). Though biases (when statistically significant) become larger (and more erratic) at stronger intensities, this may be in part due to smaller sample sizes at intensities over 110 kt.

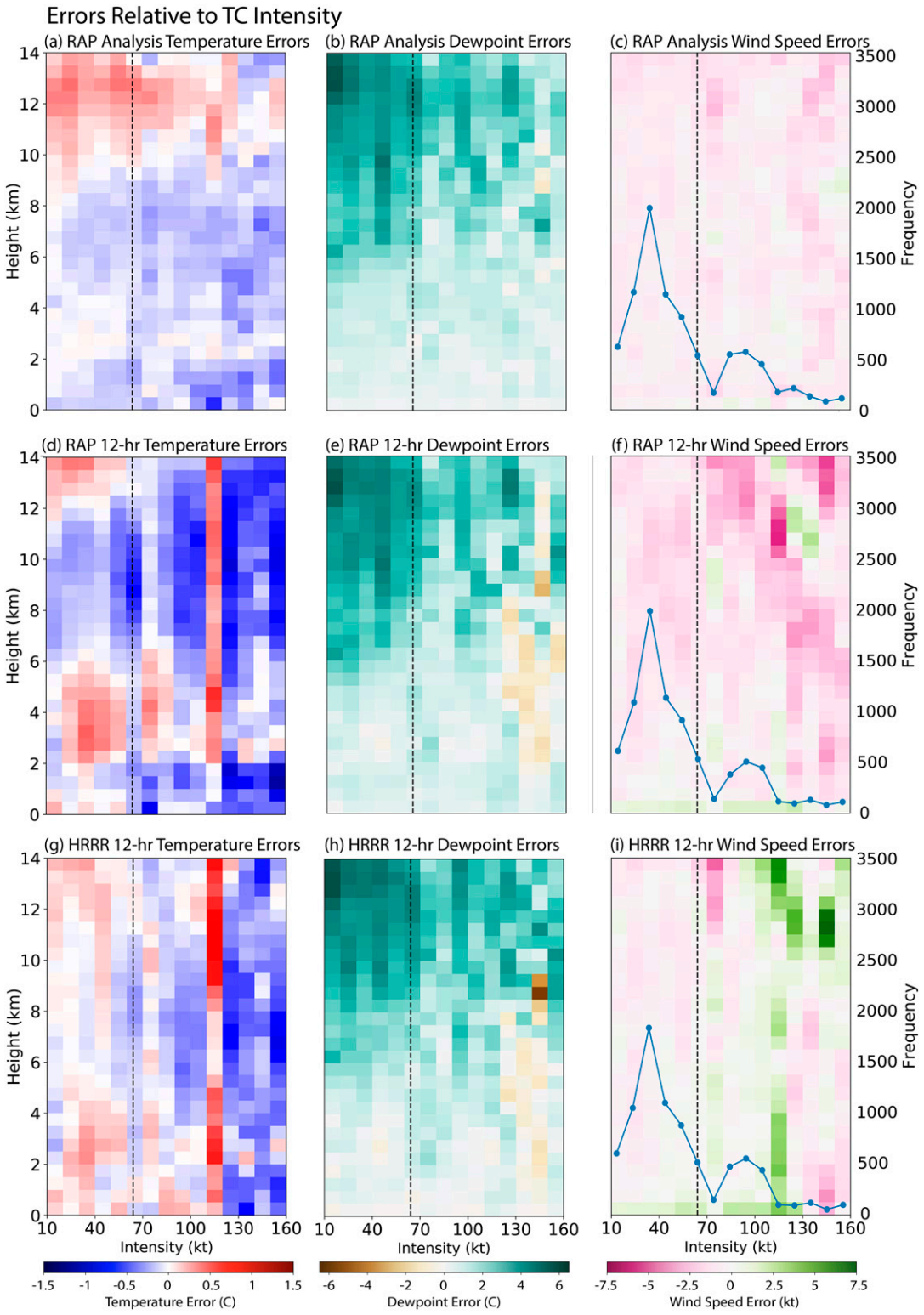

FIG. 6. As in Fig. 4, but plotted relative to TC intensity. The vertical dashed line indicates the threshold for hurricane intensity.

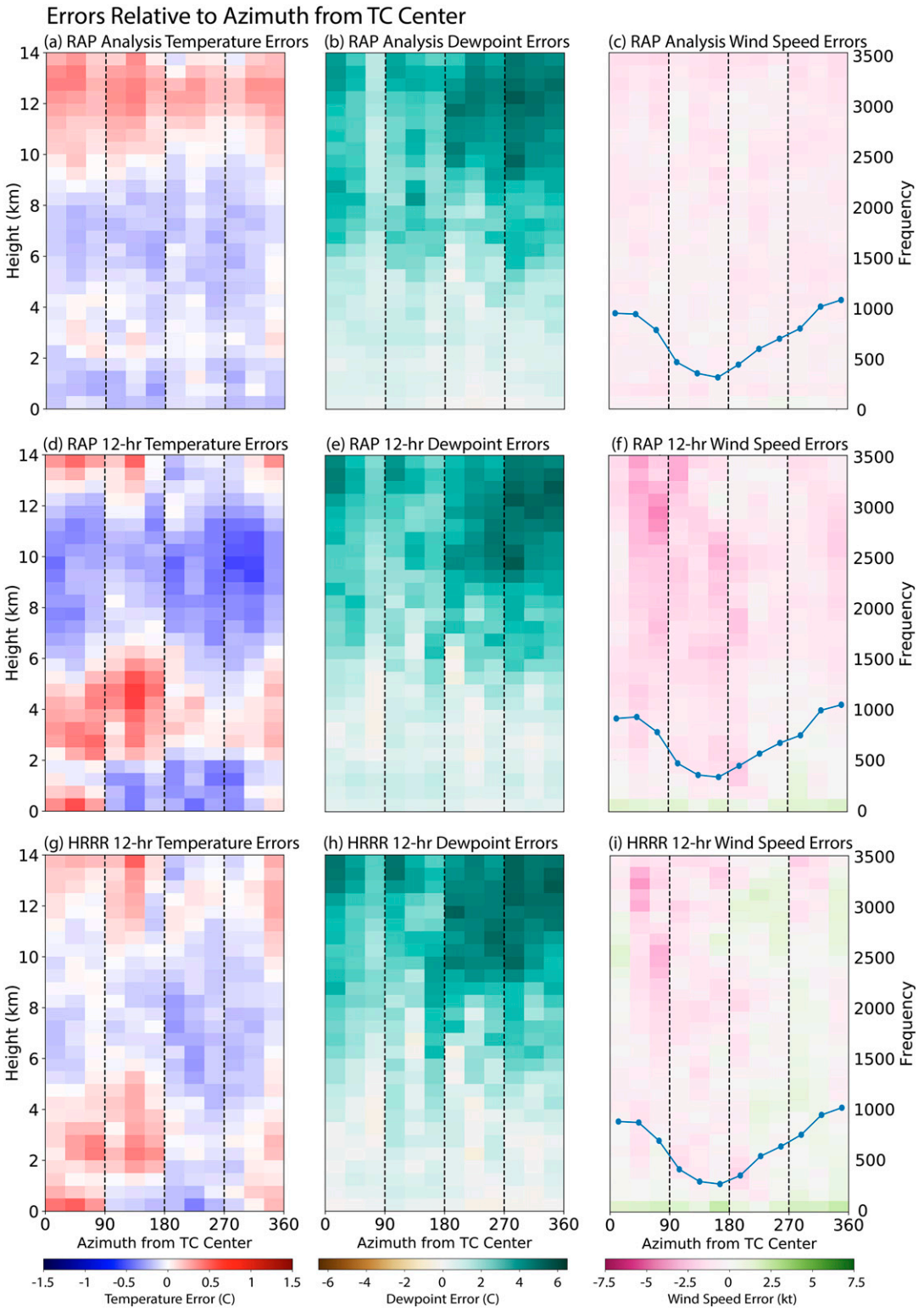

FIG. 7. As in Fig. 4, but plotted relative to azimuth from TC center. The vertical dashed lines separate each north-relative quadrant.

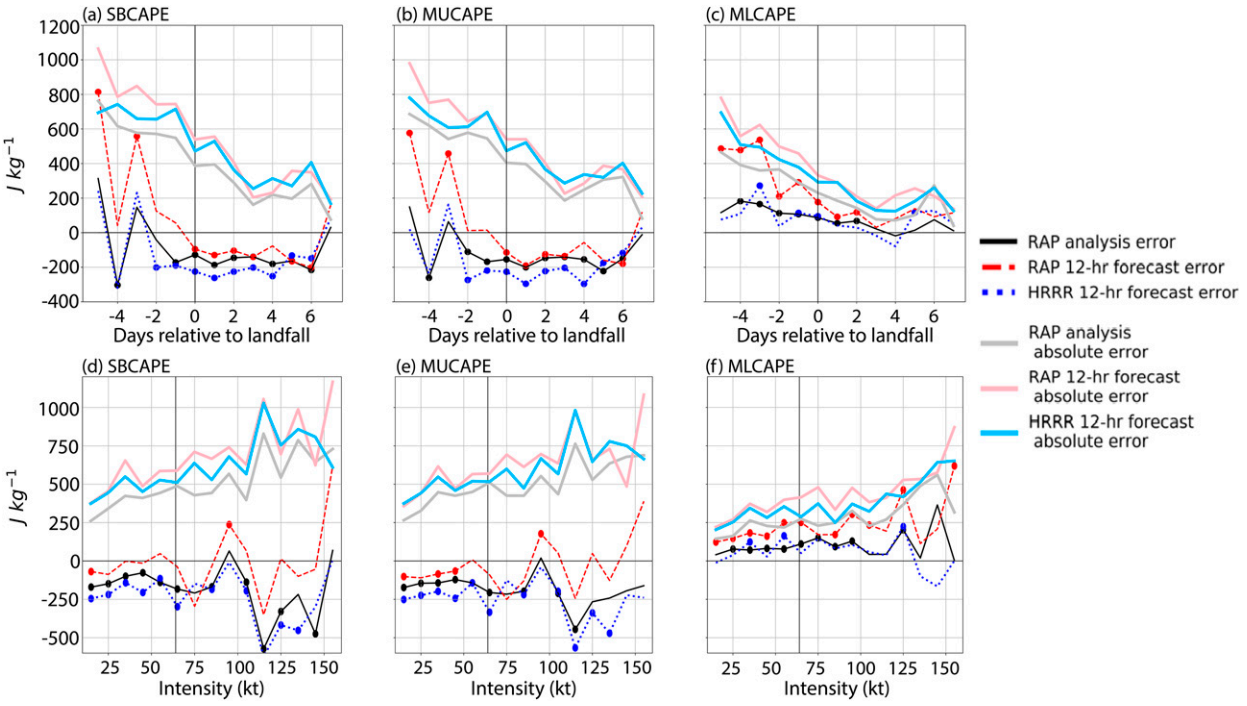

FIG. 8. (a)–(c) SBCAPE, MUCAPE, and MLCAPE mean errors and mean absolute errors (J kg^{-1}) plotted as functions of time relative to landfall for the RAP analysis, RAP 12-h forecast, and HRRR 12-h forecast. The vertical line at $x = 0$ represents the time of the TC landfall. Solid dots are placed along the mean error lines at the midpoint of each bin in which the errors are significant at the 95% confidence level. (d)–(f) As in (a)–(c), but plotted with respect to TC intensity. The vertical line represents the hurricane intensity threshold.

MLCAPE errors are plotted as a function of distance and azimuth relative to the TC center on north-relative polar plots to examine the TC-relative spatial distribution of the errors. Overpredictions of MLCAPE by the RAP 12-h forecast (Fig. 9b) are significant in the entire northern sector of the TC, and these errors range from 100 to over 300 J kg^{-1}. The RAP analysis overpredicts MLCAPE nearly everywhere within the 800-km radius (Fig. 9a). These overpredictions are significant east of the center in a sector from 45° to 135°, which tends to be where most TCTORs occur (Edwards 2012). The HRRR MLCAPE errors are generally weaker and less significant in most regions relative to the TC center (Fig. 9c).

Overall, the signs of the SBCAPE, MUCAPE, and MLCAPE biases examined parallel those of Evans et al. (2018) in the continental convective environments they analyzed. In their study, the RAP and HRRR forecasts as well as the RAP analysis had surface dry biases, which likely contributed to the slight low biases they observed in the SBCAPE and MUCAPE. MLCAPE was biased high in their study as well, likely due to the moist biases present just above the surface.

Lapse rates (LRs) are often used in ingredients-based tornado forecasting as a measure of static stability supplementing the integrated measure of lifted parcel buoyancy provided by CAPE. Because errors are the model value minus the observed value, a positive lapse rate error means that the layer is more unstable in the model than in the observations, while a negative lapse rate error corresponds to a stable bias in the model. Investigation of the lapse rate errors as a function of time relative to TC landfall shows a decreasing trend in 0–1-km lapse rate errors for both forecasts and the analysis, with underpredictions reaching magnitudes of 0.5°–1°C km^{-1} after two days postlandfall (Fig. 10a). The 0–3-km lapse rate errors (Fig. 10b) tend to be negative, with slight decreases after two days postlandfall and smaller magnitudes than the 0–1-km lapse rates. Most of the significant errors for both the 0–1 and 0–3-km layers occur in the period beginning two days before landfall. This significant stable bias in the models in layers starting at ground level is consistent with the CAPE errors, because SBCAPE and MUCAPE were both underpredicted by the models during this period relative to TC landfall (Figs. 8a,b). The overarching stable biases in the 0–1 and 0–3-km lapse rates are consistent with the mean temperature error profiles in Fig. 3a and driven largely by cool biases at the surface for 0–1-km lapse rates, but also a warm bias at 3 km for the 0–3-km lapse rates. The magnitudes of both the 0–1 and 0–3-km lapse rate biases worsen as time after landfall increases and the TC moves farther inland. Mean absolute errors in both lapse rates are generally smaller in the analysis than the forecasts, and they are consistent across times and intensities.

When examined as a function of TC intensity (Fig. 10d), the 0–3-km lapse rate errors are generally underpredictions of similar magnitude (<0.5°C km^{-1}) to the errors as a function of time relative to landfall regardless of intensity. The 0–1-km lapse rate errors are also negative as a function of intensity (at least when statistically significant, Fig. 10c), though with larger magnitudes (up to 1°C km^{-1}) and more variability

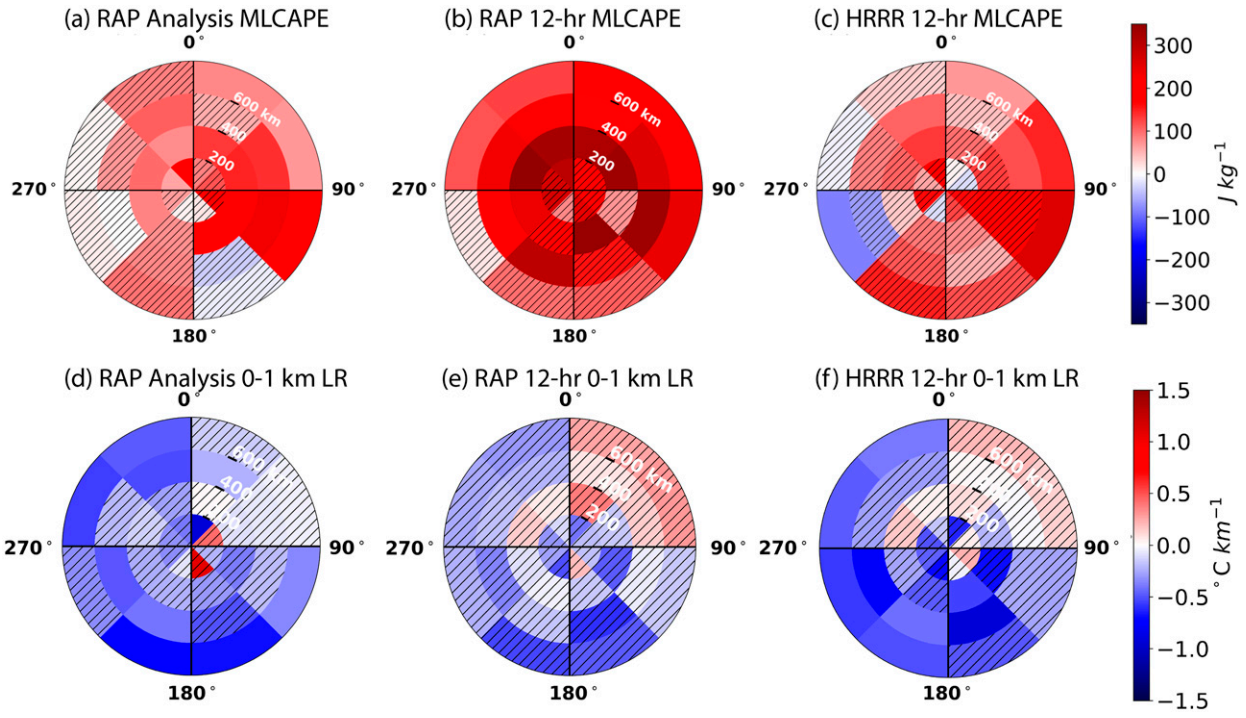

FIG. 9. MLCAPE mean errors (J kg^{-1}) plotted as a function of distance and azimuth relative to the TC center for (a) the RAP analysis, (b) the RAP 12-h forecast, and (c) the HRRR 12-h forecast. (d)–(f) As in (a)–(c), but 0–1-km lapse rate (LR) errors are plotted. Bins in which the errors are not statistically significant at the 95% confidence level are hatched.

across the range of intensities than the 0–3-km lapse rate errors. Few lapse rate errors in either of the forecasts or the analysis are statistically significant at intensities above 100 kt, in either the 0–1 or 0–3-km layer. Moreover, the 0–1-km lapse rate errors are generally negative at most distances and azimuths from 90° to 360° (Figs. 9d–f), which corresponds to a stable bias. However, the RAP and HRRR 12-h forecasts exhibit unstable, but not statistically significant 0–1-km lapse rate biases in the northeast quadrant, consistent with the atypical surface warm bias in that quadrant (Figs. 7d,g).

Shifting to kinematic sounding-derived parameters, bulk wind difference (BWD, representing vertical wind shear) in the layers 0–1, 0–3, and 0–6 km is generally underpredicted as a function of intensity, local time of day, time relative to landfall, and location relative to the TC center for all model runs examined. These underpredictions align with the shear errors expected to result from the wind error profiles (Fig. 3c) and wind error heat maps (Figs. 4–8c,f,i) for each of the forecasts and the analysis, which show generally overpredicted winds at the surface and underpredicted winds aloft. BWD in the layer from 0 to 1 km is underpredicted by 1.5–3 kt for both the forecasts and the RAP analysis before TC landfall (Fig. 11a), though the magnitude of this error decreases with time during the first three days after landfall. This is physically consistent with the behavior of the 0–1-km BWD errors as a function of intensity (Fig. 11b); the largest underpredictions occur for stronger TCs and the magnitude of the underpredictions decreases as the TC weakens. The 0–6-km BWD errors exhibit similar magnitudes to the 0–1-km BWD errors, but they do not notably decrease as time after landfall decreases (Fig. 11c), nor as TC intensity weakens (Fig. 11d). Absolute BWD error magnitudes are 3–6 kt for both the 0–1- and 0–6-km layers, which is about 15%–33% of the average BWD values of around 18–19 kt observed in those layers (unlike typical midlatitude severe storms environments, the mean 0–1- and 0–6-km BWD are of a similar magnitude in TC envelopes).

The RAP analysis significantly underpredicts 0–1-km BWD in all but two range–azimuth spatial bins between 200 and 800 km from the TC center (Fig. 12a). Within the same range ring in the forecasts, significant underprediction of 0–1-km BWD is confined to smaller continuous swaths: across the entire northeast half of the TC from 315° to 135° in the RAP 12-h forecast (Fig. 12b) and within a smaller sector from 315° to 90° in the HRRR 12-h forecast (Fig. 12c). The RAP analysis and forecast continue to significantly underpredict 0–6-km BWD in most of the eastern half of the TC (Figs. 13a,b). The HRRR forecast 0–6-km BWD errors are also negative across most of the domain, but less statistically significant (Fig. 13c).

McCaul (1991) showed that the strongest shear and largest SRH values tend to occur in the right-front quadrant of the TC, whose characteristics tend to align with those of the northeast quadrant in the north-relative reference frame (Schultz and Cecil 2009). In the north-relative polar plots created for BWD errors (Figs. 12a–c), the largest continuous swaths of significant errors tend to include or center around the northeast quadrant, where mean observed values of shear are strongest and TCTORs are the most common relative to the TC center (Edwards 2012). Overall, 0–1-km SRH tends to be underpredicted by models in

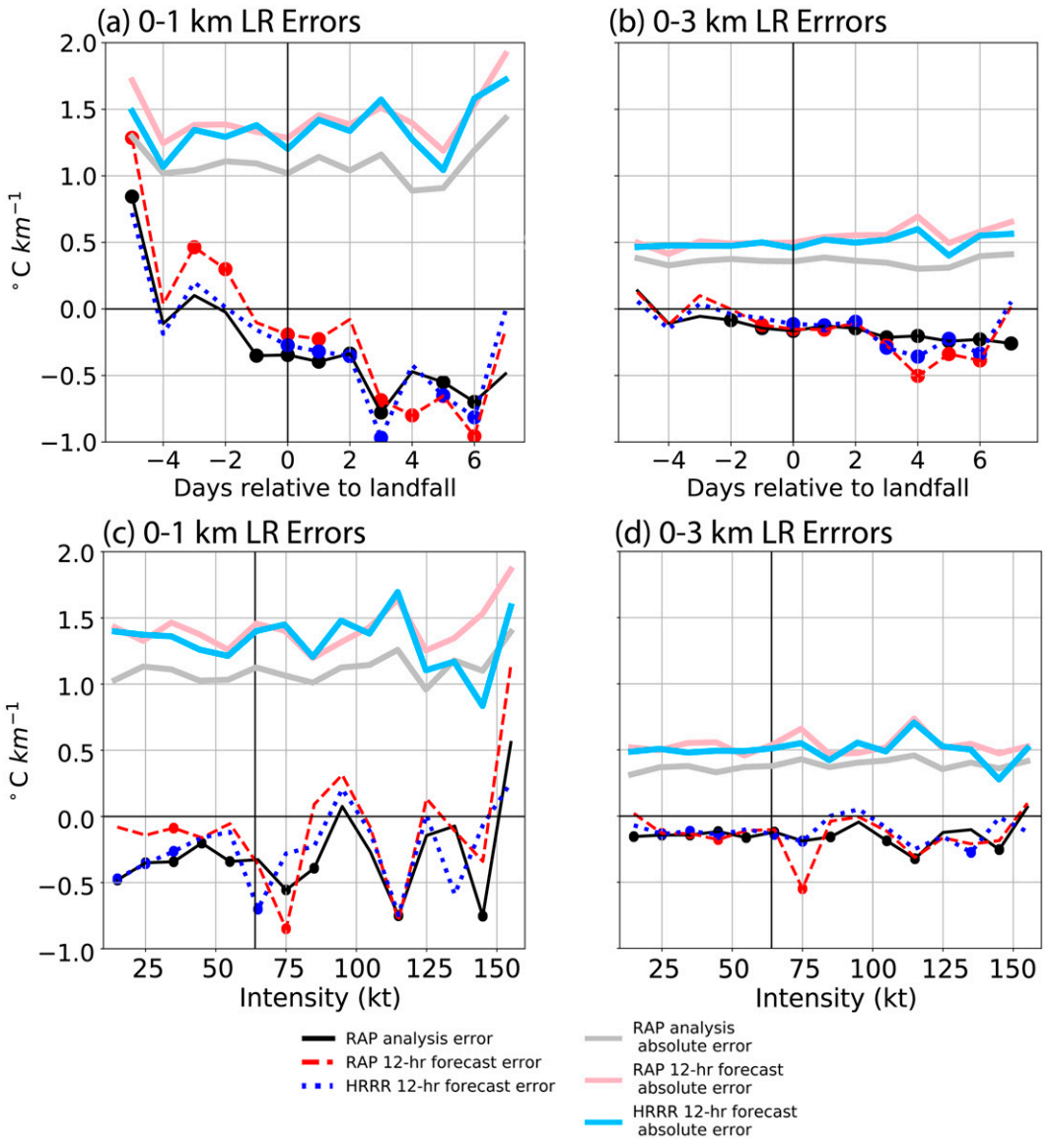

FIG. 10. (a) 0–1-km and (b) 0–3-km lapse rate mean errors and mean absolute errors for the RAP analysis, RAP 12-h forecast, and HRRR 12-h forecast (°C km^{-1}), plotted as functions of time relative to landfall. The vertical line at $x = 0$ represents the time of the TC landfall. (c),(d) As in (a) and (b), but errors are plotted relative to TC intensity and the vertical line represents the threshold for hurricane intensity. Solid dots are placed along the mean error lines at the midpoint of each bin in which the errors are significant at the 95% confidence level.

this quadrant (outside of 200-km range), but this difference is statistically significant in fewer spatial bins than low-level shear (Figs. 12d–f).

Values of selected thermodynamic and kinematic variables are used to calculate the STP, which was shown by Nowotarski et al. (2021) to discriminate between tornadic and nontornadic cells within Hurricane Harvey. Their study used the fixed-layer STP, which is computed using the surface-based lifting condensation level, or SBLCL, along with SBCAPE, 0–1-km SRH, and 0–6-km BWD. The CINH-adjusted STP depends on the mixed-layer thermodynamic quantities MLLCL, MLCAPE, and MLCINH, in addition to values of effective-layer SRH and effective-layer shear.

Fixed-layer STP is significantly underpredicted in a small sector to the east-northeast of the TC center by the RAP 12-h forecast (Fig. 13b). These underpredictions have the largest magnitude within 200–400 km of the TC center and decrease outside of that range. Fixed-layer STP is also significantly underpredicted on the east side of the TC center by the HRRR 12-h forecast through a narrow ring from 45° to 180° and 200–400 km (Fig. 13c), and by the RAP analysis in a small sector just south of east from 200 to 600 km (Fig. 13a). These

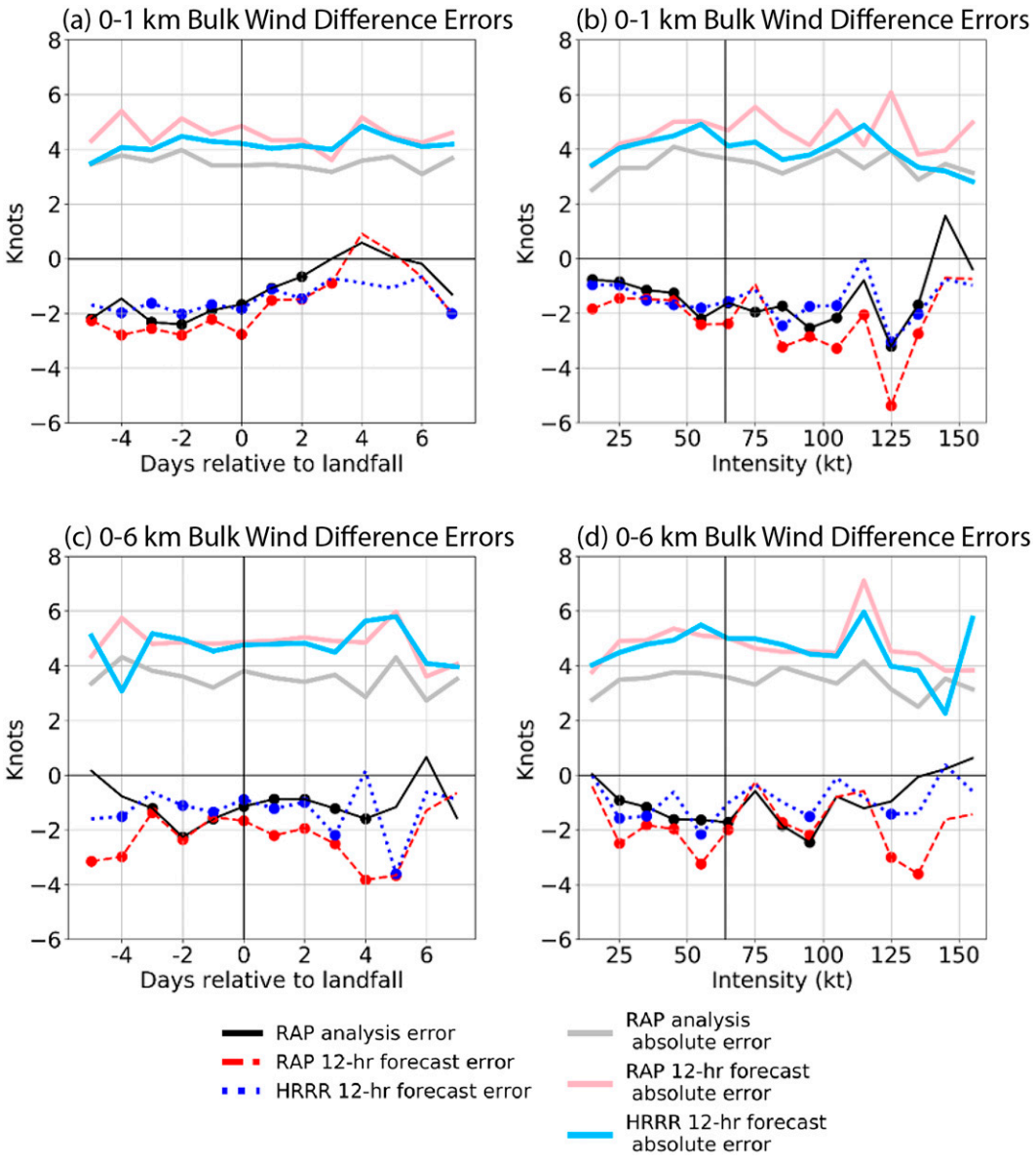

FIG. 11. 0–1- and 0–6-km bulk wind difference mean errors and mean absolute errors (kt) and plotted as a function of (a),(c) time relative to landfall (days) and (b),(d) TC intensity (kt) for the RAP analysis, RAP 12-h forecast, and HRRR 12-h forecast. Solid dots are placed along the mean error lines at the midpoint of each bin in which the errors are significant at the 95% confidence level.

results are consistent with the overlapping negative errors already seen in SBCAPE, 0–6-km BWD, and 0–1-km SRH on the east side of the TC center, while SBLCL errors are generally small given the competing effects of the surface cool and dry bias on the LCL (Fig. 3). Notably, this region of STP underprediction also overlaps with the region where TCTORs most tend to occur. Fixed-layer STP errors remain generally (but often insignificantly) negative at most times relative to TC landfall (Fig. 14b) and are only significant and slightly negative for weaker TCs (Fig. 14a). Largely because of the use of ML thermodynamic variables, which are less sensitive to the localized surface temperature and moisture biases of the models (e.g., MLCAPE biases are positive; Figs. 8 and 9) while effective-layer SRH and shear values are still biased negative, CINH-adjusted STP errors tend to be smaller and not statistically significant.

5. Summary and conclusions

This study explores model errors in the RAP analysis and RAP and HRRR 12-h forecasts within the TC envelope up to 800 km from the center. Up to 1730 sounding pairs are used in the analysis, from 13 TCs that made landfall along the coast of the contiguous United States during the 2017, 2018, and

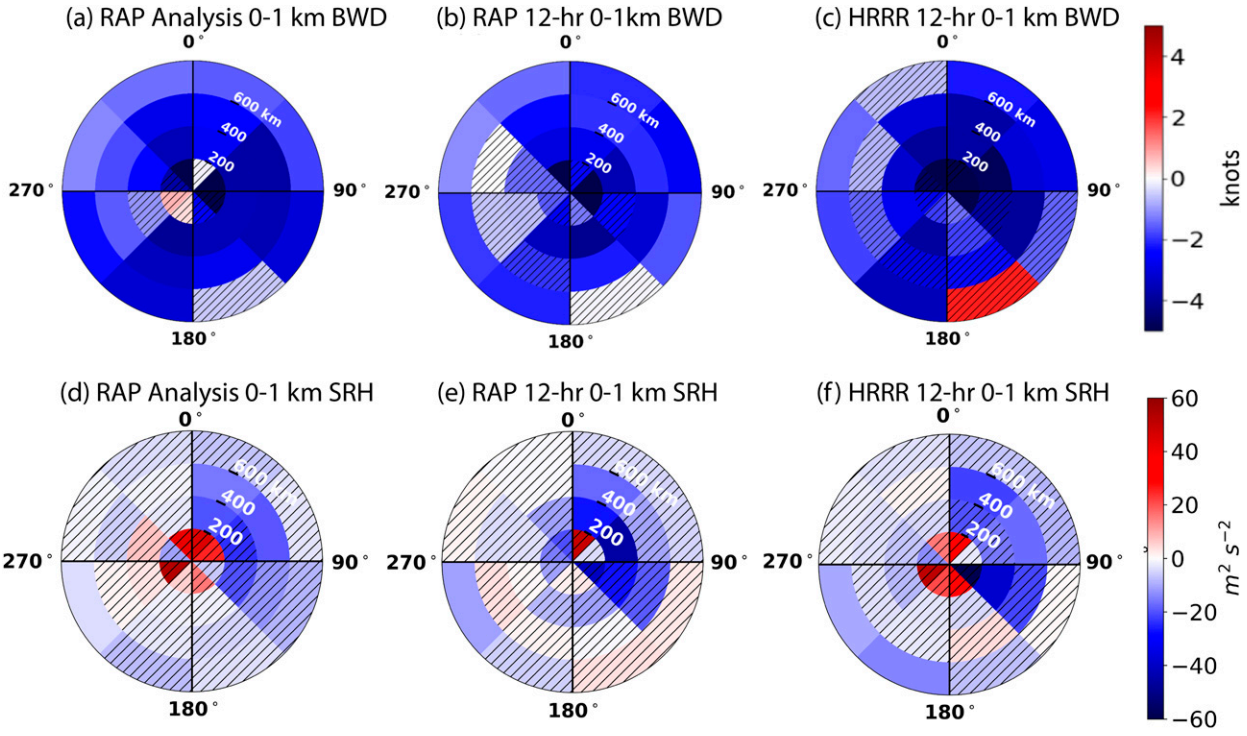

FIG. 12. As in Fig. 9, but for (a)–(c) 0–1-km BWD errors and (d)–(f) 0–1-km SRH.

2019 Atlantic hurricane seasons. First, mean error profiles are created to investigate the model representation of temperature, dewpoint, and wind throughout the depth of the troposphere. Then, we explore how these model errors vary with height as functions of variables related to TCs such as time relative to landfall, intensity, and distance/azimuth from the center. Finally, model errors in sounding-derived parameters are examined, paying particular attention to variables previously

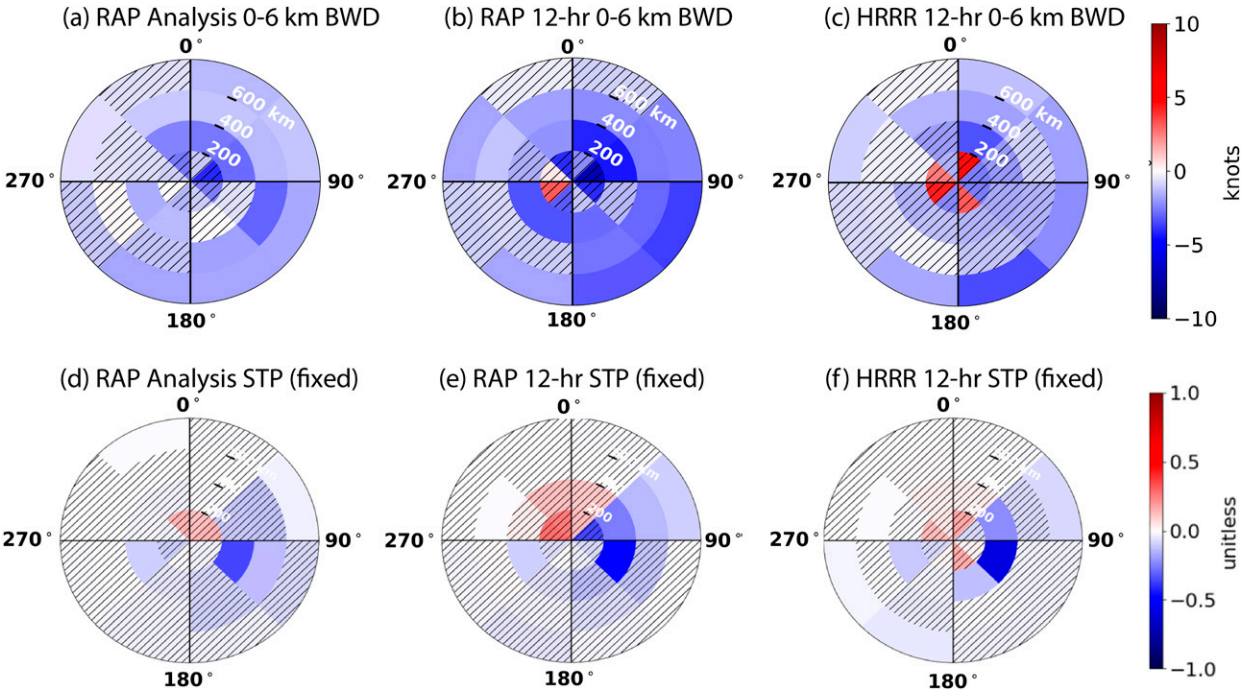

FIG. 13. As in Fig. 9, but for (a)–(c) 0–6-km BWD and (d)–(f) fixed-layer STP.

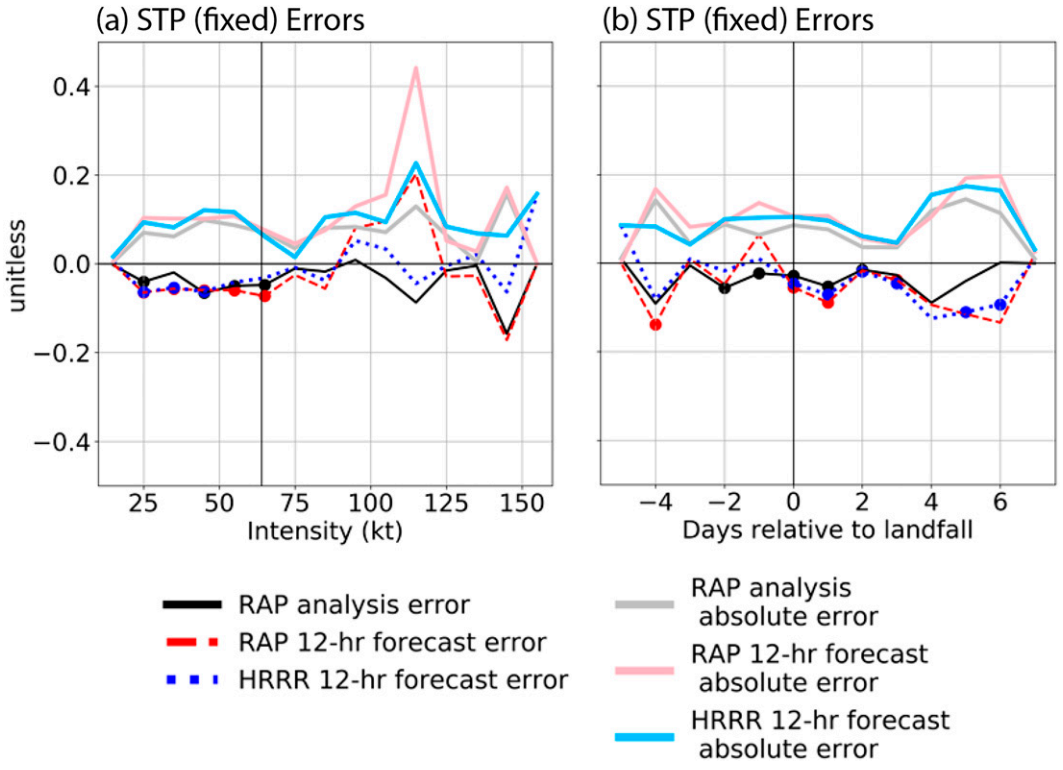

FIG. 14. Fixed-layer STP mean errors and mean absolute errors plotted as a function of (a) TC intensity and (b) time relative to landfall. Solid dots are placed along the line at the midpoint of each bin in which the errors are significant at the 95% confidence level.

identified as useful for TCTOR forecasting. Results are also compared and contrasted with the RAP and HRRR model errors found by Evans et al. (2018) in their examination of continental convective environments.

Our conclusions in terms of the main research questions motivating this study are as follows:

1) What are the height-dependent model biases in temperature, dewpoint, and winds in the RAP analysis, and 12-h RAP and HRRR forecasts relative to radiosonde observations within a TC envelope?
   Surface cool biases are evident in the mean error profiles, which often leads the models to be stable-biased in surface-based parameters used to assess static stability and parcel buoyancy. The temperature errors from 0 to 5 km in the RAP and HRRR 12-h forecasts are not consistent with the results of Evans et al. (2018), who found persistent warm biases within that layer in continental environments. Dewpoint errors have larger magnitudes and standard deviations than the temperature errors, and a pronounced moist bias worsens with height for both the forecasts and the analysis. This moist bias was also evident in the continental environments analyzed by Evans et al. (2018). However, the magnitude of the moist bias from 0 to 5 km tends to be smaller in the TC environments analyzed in this study, suggesting that models predict moisture more accurately within TC envelopes. Wind speeds are overpredicted in a shallow layer near the surface and then underpredicted throughout the rest of the depth of the profile, which leads to underpredictions of vertical wind shear in surface-based layers. Overall, mean absolute errors of temperature, dewpoint, and wind speed in both the 12-h forecasts exceed those in the RAP analysis throughout the profile.
2) How does model performance vary as a function of distance/azimuth from the TC center, TC intensity, and time relative to TC landfall?
   Low-level temperature errors tend to transition from cool to warm biases as time from landfall increases while a moist bias descends with time from landfall, suggesting both the RAP and HRRR may be slow to dry out the TC envelope as it moves inland and weakens. No notable trends in temperature or dewpoint errors exist as distance from the TC center increases, but low-level warm biases occur in the north-northwest and northeast sectors of TCs, while low-level cool biases tend to occur on the southern half of the TC envelope. Upper-level moist biases are amplified at weaker TC intensities and in the western half of TCs. Wind speed errors tend to be consistent with their average vertical profiles at most TC intensities and times after landfall, but the HRRR forecasts tend to overpredict mid and upper-level wind speeds on the western half of the TC compared to the RAP.

3) How do model biases translate to errors in sounding-derived parameters specifically related to TCTOR forecasting?

   SBCAPE and MUCAPE are generally underpredicted across all model runs and as a function of time relative to landfall, intensity, and distance and azimuth relative to the TC center, while MLCAPE is generally overpredicted. These broad trends in model predictions of CAPE align with what could be expected from a cool and dry bias at the surface that transitions to a weaker cool bias and moist bias in the mixed layer just above the surface. The signs of the biases in SBCAPE, MUCAPE, and MLCAPE are also consistent with the findings of Evans et al. (2018). Stable biases also exist in the lapse rate errors, with more variability and larger error magnitudes for the lapse rates calculated from 0 to 1 km than those calculated for the 0–3-km layer.

   The 0–1-, 0–3-, and 0–6-km bulk wind difference (BWD) are generally underpredicted for all model runs examined, consistent with overpredictions of surface wind speed in the forecast and to some extent underpredictions of winds aloft. The underpredictions of 0–1-km BWD decrease after landfall and at weaker intensities, while the underpredictions of 0–3- and 0–6-km BWD remain steadier as a function of time relative to landfall and intensity. Likewise, SRH tends to be underpredicted, although with fewer significant results across the model runs and TC-related variables. Fixed-layer STP is slightly underpredicted, especially in areas where TCTORs are most common, while trends are less clear in the CINH-adjusted STP.

The frequent underprediction of surface-based measures of instability and shear by the RAP and HRRR models is important for forecasters to note. The underpredictions of these variables in the model forecasts and analysis have the potential to lead to an underestimation of the tornado threat during landfalling TCs compared to mixed-layer values. Several of the variables that are significantly over- or underpredicted within sectors of the TC have been identified as favorable for TCTOR production. MLCAPE is significantly overpredicted by 100–200 J kg^{-1} within a sector on the east side of the TC in the RAP analysis, and by 200–300 J kg^{-1} throughout the northern half of the TC in the RAP 12-h forecast. The 0–3-km BWD values are significantly underpredicted by 2–5 kt in the RAP 12-h forecast throughout the eastern half of the TC, outside of 200 km from the center. The 0–6-km BWD is also significantly underpredicted in the RAP 12-h forecast, by the same magnitude within the majority of the eastern half of the TC. Finally, the mid and upper-level moisture biases shown here may lead to significant biases in entrainment CAPE, which has been shown to be relevant within TCs (Molinari et al. 2012) and a skillful predictor of TCTORs (Sueki and Niino 2016).

Our methodology is limited in a few ways that may affect the results. First, we do not filter for the potential for sounding pairs that may be heavily influenced by local effects of convection, since all of these observations occur within the TC envelope. Based on analysis of saturated layers within the soundings, we estimate less than 10% of all soundings occur within convectively active regions. We also assume that the radiosonde observations match vertical model gridpoint profiles, ignoring potential effects of horizontal advection of radiosondes. This choice was made to be consistent with Evans et al. (2018), and because RAP and HRRR data are only available hourly, an accurate spatiotemporal interpolation would be difficult to achieve.

Sample size is another potential limitation of these results. While we have analyzed ~1500 sounding pairs, these occur over three years and 13 TCs, such that results may vary over a longer period. Within these results, we also caution the reader that some portions of the parameter space with fewer observations may lead to results that are overly influenced by one model analysis/forecast period or TC (e.g., results for TC intensities over 110 kt, within 200 km of the TC center, or several days after landfall). Moreover, when sample size is limited there may be notable covariance between aspects of the parameter space. For example, observations in southern TC-relative azimuths tend to be more likely when TCs are well inland and generally weaker.

Finally, model errors may not be consistent for different versions of the RAP/HRRR model. Model physics parameterizations, numerics, and data assimilation are updated in ways that may affect low-level errors, in particular (e.g., PBL and land surface schemes). Both models were updated roughly halfway through our analysis period. Supplemental Fig. 1 shows changes in mean error profiles for the RAP3 versus RAP4 and HRRR2 versus HRRR3 periods. In general, newer model versions have a cooler temperature bias (though the magnitude of this bias is similar for both model versions), a slight reduction in the upper-level moisture bias, and little systematic changes in the wind speed bias. However, it is unclear how many of these differences are due to model changes compared to differences in the nature of the TCs within each subset of events. In December 2020, the RAP and HRRR were updated to versions 4 and 5, respectively, such that the errors presented here may not fully agree with current or future operational analyses or forecasts.

While this study presents a broad overview of RAP and HRRR errors within TC envelopes that may be particularly relevant for TCTOR forecasting, considerable future work in this area is needed. Additional forecasting insights could be gleaned by investigating the errors in the raw variables and sounding-derived parameters at longer model lead times. Examining the errors in different models such as the Hurricane Weather Research and Forecasting model (HWRF), the North American Mesoscale Forecast System (NAM), and the Global Forecast System (GFS) would be beneficial since forecasters often compare forecasts from a variety of models within the complex TC envelope and regional and global models provide estimates of potential tornado environments within TCs at longer lead times. Also, the dataset could be expanded to include TCs from more Atlantic hurricane seasons and more recent versions of the RAP/HRRR to address limitations of this study discussed above. Finally, work is needed to determine the physical mechanisms and flaws in model initial conditions or parameterizations leading to the errors

presented here in order to reduce these errors in future versions of operational models.

Acknowledgments. This work is funded by NWS CSTAR Grant NA19NWS4680007. We thank Roger Edwards and two anonymous reviewers for their constructive criticism that greatly improved the presentation of this work. The authors wish to thank Drs. Matthias Katzfuss, Robert Korty, and Don Conlee for insightful comments at various stages of this work and our larger efforts in improving TCTOR prediction. We are also indebted to Justin Spotts, Dr. Matthew Brown, Dr. Troy Arcomano, Dr. Kyle Wodzicki, and Nicholas Bogen for their advice and assistance with coding and GIS. Many sounding-derived parameters were computed using the Sounding and Hodograph Analysis and Research Program in Python (SHARPpy) software (Blumberg et al. 2017). An earlier version of this work appeared as L. MacDonald's M.S. thesis.

Data availability statement. All RAP, HRRR, radiosonde, and tropical cyclone track and intensity data are freely available via NCEI data servers. Spreadsheets of analyzed data and analysis code are archived locally and available upon request to the corresponding author.

REFERENCES

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

Blumberg, W. G., K. T. Halbert, T. A. Supinie, P. T. Marsh, R. L. Thompson, and J. A. Hart, 2017: SHARPpy: An open source sounding analysis toolkit for the atmospheric sciences. *Bull. Amer. Meteor. Soc.*, **98**, 1625–1636, https://doi.org/10.1175/BAMS-D-15-00309.1.

Bunkers, M. J., B. A. Klimowski, J. W. Zeitler, R. L. Thompson, and M. L. Weisman, 2000: Predicting supercell motion using a new hodograph technique. *Wea. Forecasting*, **15**, 61–79, https://doi.org/10.1175/1520-0434(2000)015<0061:PSMUAN>2.0.CO;2.

——, D. A. Barber, R. L. Thompson, R. Edwards, and J. Garner, 2014: Choosing a universal mean wind for supercell motion prediction. *J. Oper. Meteor.*, **2**, 115–129, https://doi.org/10.15191/nwajom.2014.0211.

Curtis, L., 2004: Midlevel dry intrusions as a factor in tornado outbreaks associated with landfalling tropical cyclones from the Atlantic and Gulf of Mexico. *Wea. Forecasting*, **19**, 411–427, https://doi.org/10.1175/1520-0434(2004)019<0411:MDIAAF>2.0.CO;2.

Davies, J. M., 2006: Hurricane and tropical cyclone tornado environments from RUC proximity soundings. *23rd Conf. on Severe Local Storms,* St. Louis, MO, Amer. Meteor. Soc., P8.1, https://ams.confex.com/ams/pdfpapers/115483.pdf.

De Pondeca, M. S. F. V., and Coauthors, 2011: The real-time mesoscale analysis at NOAA's National Centers for Environmental Prediction: Current status and development. *Wea. Forecasting*, **26**, 593–612, https://doi.org/10.1175/WAF-D-10-05037.1.

Edwards, R., 2010: Tropical cyclone tornado records for the modernized NWS era. Preprints, *25th Conf. on Severe Local Storms*, Denver, CO, Amer. Meteor. Soc., P3.1, https://ams.confex.com/ams/pdfpapers/175269.pdf.

——, 2012: Tropical cyclone tornadoes: A review of knowledge in research and prediction. *Electron. J. Severe Storms Meteor.*, **7** (6), https://ejssm.com/ojs/index.php/site/article/view/42/41.

——, and R. M. Mosier, 2022: Over a quarter century of TCTOR: Tropical cyclone tornadoes in the WSR-88D ERA. *30th Conf. Severe Local Storms*, Santa Fe, NM, Amer. Meteor. Soc., P171, https://ams.confex.com/ams/30SLS/meetingapp.cgi/Paper/407018.

——, A. R. Dean, R. L. Thompson, and B. T. Smith, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part III: Tropical cyclone tornadoes. *Wea. Forecasting*, **27**, 1507–1519, https://doi.org/10.1175/WAF-D-11-00117.1.

Evans, C., S. J. Weiss, I. L. Jirak, A. R. Dean, and D. S. Nevius, 2018: An evaluation of paired regional/convection-allowing forecast vertical thermodynamic profiles in warm-season, thunderstorm-supporting environments. *Wea. Forecasting*, **33**, 1547–1566, https://doi.org/10.1175/WAF-D-18-0124.1.

Fovell, R. G., and A. Gallagher, 2020: Boundary layer and surface verification of the High-Resolution Rapid Refresh, version 3. *Wea. Forecasting*, **35**, 2255–2278, https://doi.org/10.1175/WAF-D-20-0101.1.

Gentry, R. C., 1983: Genesis of tornadoes associated with hurricanes. *Mon. Wea. Rev.*, **111**, 1793–1805, https://doi.org/10.1175/1520-0493(1983)111<1793:GOTAWH>2.0.CO;2.

Laflin, J. M., 2013: Verification of RAP model soundings in preconvective environments. *J. Oper. Meteor.*, **1**, 66–70, https://doi.org/10.15191/nwajom.2013.0106.

Landsea, C. W., and J. L. Franklin, 2013: Atlantic hurricane database uncertainty and presentation of a new database format. *Mon. Wea. Rev.*, **141**, 3576–3592, https://doi.org/10.1175/MWR-D-12-00254.1.

McCaul, E. W., Jr., 1991: Buoyancy and shear characteristics of hurricane–tornado environments. *Mon. Wea. Rev.*, **119**, 1954–1978, https://doi.org/10.1175/1520-0493(1991)119<1954:BASCOH>2.0.CO;2.

Molinari, J., and D. Vollaro, 2010: Distribution of helicity, CAPE, and shear in tropical cyclones. *J. Atmos. Sci.*, **67**, 274–284, https://doi.org/10.1175/2009JAS3090.1.

——, D. M. Romps, D. Vollaro, and L. Nguyen, 2012: CAPE in tropical cyclones. *J. Atmos. Sci.*, **69**, 2452–2463, https://doi.org/10.1175/JAS-D-11-0254.1.

NOAA/Global Systems Laboratory, 2020: Rapid Refresh (RAP). NOAA, accessed 22 April 2021, https://rapidrefresh.noaa.gov/.

NOAA/National Centers for Environmental Information, 2020: Rapid Refresh (RAP). NOAA, accessed 17 April 2023, https://www.ncei.noaa.gov/products/weather-climate-models/rapid-refresh-update.

Nowotarski, C. J., J. Spotts, R. Edwards, S. Overpeck, and G. R. Woodall, 2021: Tornadoes in Hurricane Harvey. *Wea. Forecasting*, **36**, 1589–1609, https://doi.org/10.1175/WAF-D-20-0196.1.

Paredes, M., B. A. Schenkel, R. Edwards, and M. Coniglio, 2021: Tropical cyclone outer size impacts the number and location of tornadoes. *Geophys. Res. Lett.*, **48**, e2021GL095922, https://doi.org/10.1029/2021GL095922.

Schenkel, B. A., R. Edwards, and M. Coniglio, 2020: A climatological analysis of ambient deep-tropospheric vertical wind shear impacts upon tornadoes in tropical cyclones. *Wea. Forecasting*, **35**, 2033–2059, https://doi.org/10.1175/WAF-D-19-0220.1.

Schultz, L. A., and D. J. Cecil, 2009: Tropical cyclone tornadoes, 1950–2007. *Mon. Wea. Rev.*, **137**, 3471–3484, https://doi.org/10.1175/2009MWR2896.1.

Sinnott, R. W., 1984: Virtues of the haversine. *Sky Telescope*, **68**, 158–159.

Smith, T. L., S. G. Benjamin, J. M. Brown, S. S. Weygandt, T. Smirnova, and B. Schwartz, 2008: Convection forecasts from the hourly updated, 3-km High Resolution Rapid Refresh (HRRR) model. *24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 11.1, https://ams.confex.com/ams/24SLS/techprogram/paper_142055.htm.

Spratt, S. M., D. W. Sharp, P. Welsh, A. C. Sandrik, F. Alsheimer, and C. Paxton, 1997: A WSR-88D assessment of tropical cyclone outer rainband tornadoes. *Wea. Forecasting*, **12**, 479–501, https://doi.org/10.1175/1520-0434(1997)012<0479:AWAOTC>2.0.CO;2.

Storm Prediction Center, 2016: SPC mesoscale analysis pages. NOAA, accessed 22 April 2021, https://www.spc.noaa.gov/exper/mesoanalysis/.

Sueki, K., and H. Niino, 2016: Toward better assessment of tornado potential in typhoons: Significance of considering entrainment effects for CAPE. *Geophys. Res. Lett.*, **43**, 12 597–12 604, https://doi.org/10.1002/2016GL070349.

Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18**, 1243–1261, https://doi.org/10.1175/1520-0434(2003)018<1243:CPSWSE>2.0.CO;2.

——, B. T. Smith, J. S. Grams, A. R. Dean, and C. Broyles, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part II: Supercell and QLCS tornado environments. *Wea. Forecasting*, **27**, 1136–1154, https://doi.org/10.1175/WAF-D-11-00116.1.

University of Wyoming, 2020: Upper air soundings. University of Wyoming, accessed 22 January 2020, http://weather.uwyo.edu/upperair/sounding.html.

Verbout, S. M., D. M. Schultz, L. M. Leslie, H. E. Brooks, D. J. Karoly, and K. L. Elmore, 2007: Tornado outbreaks associated with landfalling hurricanes in the North Atlantic Basin: 1954–2004. *Meteor. Atmos. Phys.*, **97**, 255–271, https://doi.org/10.1007/s00703-006-0256-x.

Wilcoxon, F., 1945: Individual comparisons by ranking methods. *Biom. Bull.*, **1**, 80–83, https://doi.org/10.2307/3001968.

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