VERIFICATION OF THE RAPID REFRESH (RAP) AND HIGH-RESOLUTION RAPID REFRESH (HRRR) MODELS WITHIN LANDFALLING TROPICAL CYCLONES TOWARD THE IMPROVEMENT OF RAINBAND TORNADO FORECASTING

A Thesis

by

LELAND MARIE MACDONALD

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Christopher J. Nowotarski
Committee Members,	Don T. Conlee
	Matthias Katzfuss
Head of Department,	Ramalingam Saravanan

May 2022

Major Subject: Atmospheric Sciences

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 diagnose the risk of TCTORs, forecasters use high-resolution, rapidly-updating models such as the Rapid Refresh (RAP) and High-resolution Rapid Refresh (HRRR), and an ingredients-based approach similar to that used for forecasting continental mid-latitude tornadoes. Though RAP and HRRR model errors have been identified in typical midlatitude convective environments, this study seeks to evaluate the performance of the RAP and the HRRR within the TC envelope up to 800 km from the TC center, with particular attention paid to variables identified in previous studies as useful for TCTOR forecasting.

A sample of 1,730 observed upper-air soundings is sourced from 13 TCs that made landfall along the US coastline between 2017-2019. The observed soundings are paired with their corresponding model grid point soundings from the RAP analysis, RAP 12-hour forecast, and HRRR 12-hour forecast. Model errors are calculated for both the raw sounding variables of temperature, dew point, and wind speed, as well as for the quantities of selected sounding-derived parameters. Results show a moist bias that worsens with height across all model runs. There are also significant underpredictions in stability-related parameters such as convective available potential energy (CAPE) and kinematic parameters such as vertical wind shear.

ACKNOWLEDGEMENTS

As with any long-term project, there are many people to thank. First and foremost, I would like to thank Dr. Nowotarski for this opportunity and for his guidance, patience, and belief in me even during the times when I didn't believe in myself. I would like to thank Dr. D for telling me "You can still do it!" on the afternoon of a major setback in the data analysis, and I would like him to know how much I reflected on those words as I worked doubly hard to bring the project to completion. I would like to thank Dr. Katzfuss for taking the time to listen to project updates even while I was in the early stages of the analysis, and for his willingness to answer all of my statistics questions.

To the many, many wonderful friends I have made during my time here: thank you from the bottom of my heart for all of the support you have given me, all of the laughter and experiences (and/or dances!) we have shared, and the way you have made me feel accepted and loved. To Savannah Jorgensen and Justin Spotts: I literally could not have made it through this last month without you. To Alex Smith and Sydney Butler: may the vibe never die! To Troy Arcomano and Dr. Kyle Wodzicki: your enthusiasm for good code is unmatched and so is your willingness to give good advice. To Dylan Leathe: hey look, we've almost made it. And last but not least, to Dr. Matt Brown and Marc Bremenkamp: I couldn't have asked for better big brothers than you guys.

I would also like to thank my fiancé Luis for his love and support throughout this entire process. Our relationship began during my second week of grad school, and he has stuck with me through the good and the bad, the celebrations and the tears, and the

victory dances and the late nights. I will always be grateful for having had you by my side during these past two and a half years, and I can't wait to see where life takes us together!

Finally, I would like to thank my parents for nurturing me and raising me to appreciate learning and hard work. I would like to thank my dad for always reminding me about school: "Just try your best." I would like to thank my mom for being an amazing listener always. I look up to her both as a statistics nerd and as my beautiful mother. Working on our respective projects together from home when I visited for a month in June 2020 is one of my favorite memories of grad school! I love you both. Thank you for everything.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Professors Christopher J. Nowotarski and Don T. Conlee of the Department of Atmospheric Sciences and Professor Matthias Katzfuss of the Department of Statistics.

Additional supervision was provided by Dr. Robert Korty of the Department of Atmospheric Sciences. Computer programs written by Professor Nowotarski and Matthew Brown, and modified by Justin Spotts, were adapted to acquire the model sounding data and to perform the initial calculations of sounding-derived parameters. Nicholas Bogen provided a smoothed coastline shape file created in ArcGIS Pro for the purposes of analysis of variables as a function of station distance from the coast.

All other work conducted for the thesis was completed by the student independently.

Funding Sources

This work was made possible by the NOAA/NWS Collaborative Science, Technology, and Applied Research (CSTAR) Award under Grant Number NA19NWS4680007. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NOAA/NWS.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

Tropical cyclone tornadoes (TCTORs) present a critical public hazard and forecasting challenge within landfalling and remnant tropical cyclones (TCs). Observations of 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 short-term changes in mesoscale features and variables critical to tornado production, forecasters 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 (Benjamin et al. 2016). Numerous studies have verified these models in continental environments, but their performance specifically within landfalling TCs has not been investigated. This study evaluates model biases in the RAP analysis, and RAP and HRRR forecasts, within the envelopes of TCs that made landfall during the 2017-2019 Atlantic hurricane seasons. Errors in vertical profiles of temperature, dew point, 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 comprises 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 non-tornadic cells within TC rainbands.

2. BACKGROUND AND HYPOTHESES

While TCTORs comprise less than 10% of overall U.S. tornado activity, they account for ~10-25% of the overall tornado activity in each state bordering the Gulf and Atlantic coasts from Louisiana to Maryland (Edwards 2010; Schultz and Cecil 2009). It is estimated that nearly 1,800 TCTORs occurred during the years 1950-2007, with every coastal state from Texas to Virginia experiencing more than 100 of these events during that period. Just over 60% of these tornadoes occurred within 100 km of the coastline, but the threat of TCTORs can persist much farther inland and many days after the TC landfall (Schultz and Cecil 2009). Hurricane Ivan (2004) provided an extreme example of this, spawning 118 tornadoes in a multi-day 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 of time, with the majority of TCTORs occurring between 12 hours prior to landfall and 48 hours after landfall (Schultz and Cecil 2009).

TCTORs most commonly occur in the right-front quadrant relative to the motion of the TC while the hurricane or tropical storm is in its mature phase (Figs. 2.1b, c). TCTOR reports tend to shift toward the right-rear quadrant as the TC weakens (Fig. 2.1d) due to the fact that this quadrant of the TC usually moves over land during the remnant phase and usually contains higher values of convective available potential energy (CAPE; Edwards 2012; McCaul 1991). Combining the reports from all phases over the TC life cycle (Fig. 2.1a), 80% of all TCTORs occur in a sector from 350 to 120

meteorological degrees, where 0 degrees represents the TC's direction of motion (Schultz and Cecil 2009). Although Figure 2.1 shows the north-relative distributions of TCTORs as opposed to motion-relative, Schultz and Cecil (2009) showed that the distinctions between the two distribution patterns are minimal.

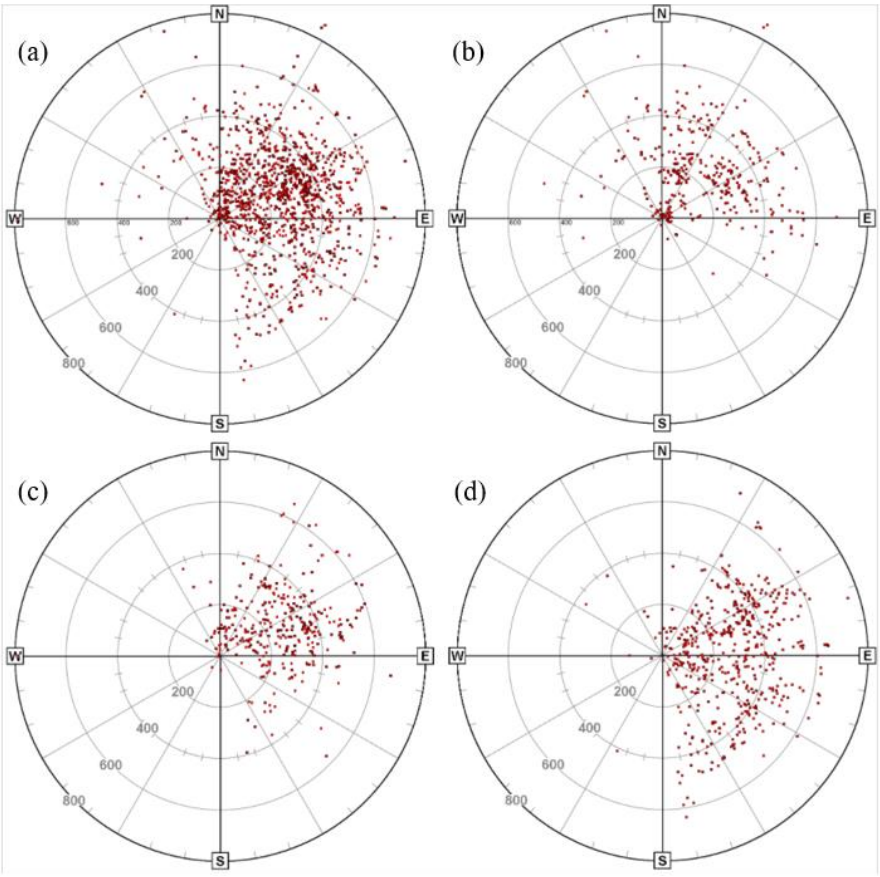

Figure 2.1: TC center-relative plots of 1995–2010 TCTOR records: a) all TCs, b) hurricanes, c) tropical storms, and d) tropical depressions, remnant lows and TC remnants. Events plotted with respect to north-relative azimuth and range (km as labeled) from center position, at the time of each tornado. [Figure and caption reprinted from Figure 6, Edwards (2012).] Reprinted with permission from R. Edwards.

Further examination of the spatial distribution of TCTORs in Figure 2.1 reveals that the overwhelming majority of reports fall within a range of 100-500 km from the center (Edwards 2012). This means that even in highly destructive TCs with large hurricane-force wind radii, many of the TCTORs that develop can cause a significant deterioration of local conditions 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 non-tropical 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; altogether these account for around 80% of TCTOR events (Edwards et al. 2012). However, the supercells within TCs are generally shallower and narrower, with more weakly rotating mesocyclones than those of their non-tropical counterparts (Spratt et al. 1997; Edwards 2012). TCTORs also tend to be weaker than non-tropical tornadoes. The TCTOR climatology compiled by Schultz and Cecil (2009) examined the distribution of their damage ratings and found that 81% of TCTORs were classified as weak (EF0-EF1), while just 74% of overall U.S. tornadoes were classified in this category. Conversely, strong tornadoes (EF2-EF3) occurred just 14% of the time in TCs, while 21% of overall U.S. tornadoes were rated as such. The diurnal distribution of TCTOR frequency is also shifted from that of non-tropical tornadoes. The bulk of TCTORs occur between 9 a.m. and 6 p.m. local time, with a distinct peak in reports between 3 and 6 p.m.; this is earlier than the early-evening maximum in non-tropical

tornado frequency (McCaul 1991). Despite this fact, nocturnal tornadoes are still more common in TCs than in the overall U.S. tornado record (Schultz and Cecil 2009).

Systematic differences between TC and non-tropical tornado environments emerge in the moisture, instability, and shear characteristics of the near-cell environment of their parent cells. Moisture is much more abundant throughout the column in the TC environment, as demonstrated by much higher values of precipitable water compared to those observed in non-tropical atmospheric profiles (Edwards et al. 2012). CAPE values, on the other hand, tend to be lower throughout the TC envelope than in typical mid-latitude convective environments (McCaul 1991). A comparison of mixed-layer CAPE (MLCAPE) in non-tropical versus TC supercell tornado cases from 2003-2011 found a median MLCAPE of 1240 J kg^{-1} present in the non-tropical environments as opposed to just 547 J kg^{-1} in the TCTOR environments (Edwards et al. 2012). The highest values of CAPE are found in the outer regions of the TC on the right side of the track, especially toward the right-rear quadrant, and CAPE values then decrease 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 are created and maintained by mid-tropospheric 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 AGL, around the 600-mb level (Edwards 2012; McCaul 1991).

Much stronger low-level vertical wind shear values are observed in TC environments than in non-tropical environments (McCaul 1991). In general, low-level

shear is stronger and tornado production is more likely in more intense TCs (Verbout et al. 2007). Vertical shear decreases with radial distance from the TC center, and the strongest shear and largest storm-relative helicity (SRH) values occur in the right-front quadrant of the TC (McCaul 1991). This is due to the fact that this quadrant is typically co-located with the downshear-left quadrant of the TC, where 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 mid-latitudes after landfall, they encounter stronger westerly flow aloft. This can maintain and even enhance vertical wind shear with time when coupled with frictional slowing of the wind speeds near the surface. Helicity parameters also tend to increase with time after the landfall (Gentry 1983; McCaul 1991).

Despite all of the aforementioned differences in the values and distributions of meteorological variables between the two environment types, the traditional ingredients-based approach to non-tropical tornado forecasting is also used in TCTOR forecasting. Since low-level moisture is in plentiful supply 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 temporal trends, recent or forecasted, in such variables is crucial when diagnosing TCTOR potential (Edwards 2012). Mid- and upper-level data are also examined for supporting features, such as the super-positioning of background flow to produce wind profiles conducive to TCTOR production, or the progression of mid-tropospheric drying

visible on 700- or 500-mb relative humidity maps that could be an early warning sign of a TCTOR outbreak (Schenkel et al. 2020; Curtis 2004). Nowotarski et al. (2021) showed that 0-6 km shear and 0-1 km SRH were the kinematic parameters that best discriminated between tornadic and non-tornadic cells within Hurricane Harvey. Also, mean values of lowest 100-mb MLCAPE, 0-3 km lapse rate, and the significant tornado parameter (STP) were significantly greater for tornadic versus non-tornadic environments in their study. These variables that display forecasting utility 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 non-tornadic environments.

In order 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, and 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 three hours. Each forecast is for 18 hours 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 relied upon heavily 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. An understanding of RAP analysis errors is critical to understanding forecast errors that may arise, since the RAP model analysis provides the foundation for the HRRR and the other valuable forecasting tools mentioned above.

Evans et al. (2018) calculated model biases for both RAP and HRRR analyses and 11-hr forecasts in environments conducive to convection, as defined by SPC Day 1 convective outlooks issued during May 2017. They plotted their results for the lowest 5 km of the atmosphere and found small biases in temperature (< ~0.25°C) for all altitudes in both of the analyses, as seen in Figure 2.2a. Both of the 11-hr forecast temperature profiles were consistently warm-biased from 0-5 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. 2.2b). The dew point data showed a slight surface dry bias 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 3°C in the RAP 11-hr forecast around an altitude of 5 km (Fig. 2.2b).

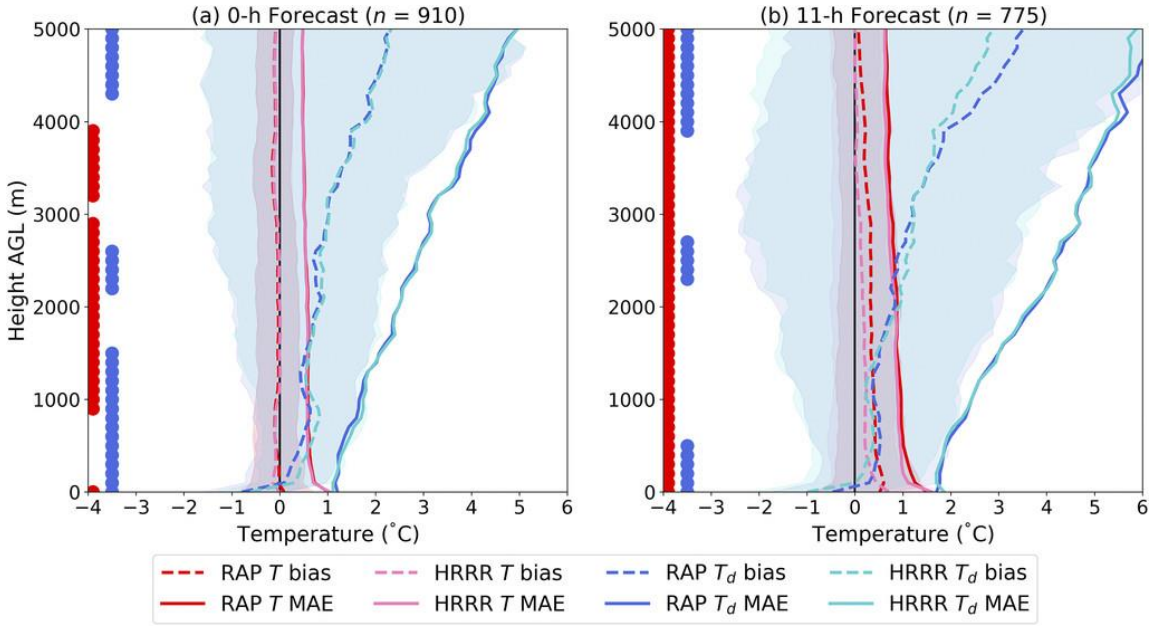

Figure 2.2: (a) Vertical profiles of sample-mean bias (°C; dashed lines; defined as model minus observations) and MAE (°C; solid lines) between 0 and 5 km AGL for RAP and HRRR 0-h temperature (red and pink lines, respectively) and dewpoint temperature (dark blue and light blue lines, respectively) analyses. 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 reprinted from Figure 4 and caption reprinted from Figure 3 in Evans et al. (2018).] © American Meteorological Society. Used with permission.

The Evans et al. (2018) study also examined sounding-derived parameter biases. CAPE was overestimated for the mixed-layer parcels in both the analyses and 11-hr forecasts due to the moist bias present above the surface, but slightly underestimated for

surface-based and most-unstable parcels due to the surface dry bias. Previously, Laflin (2013) had performed similar verification of the RAP model in springtime pre-convective environments, although restricted to the central and northern Great Plains. They found underestimations in surface-based and most-unstable CAPE for both 12- and 6-hour model forecasts, which likely resulted from large dry model biases in the near-surface layer 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 underestimated in the Laflin (2013) study. Fovell and Gallagher (2020) analyzed the representation in the contiguous U.S. of the lowest 1 km of the atmosphere in the HRRR model. They found small near-surface temperature biases, but small wind biases at the analysis time became increasingly positive throughout the 24-hour forecast period and the temperature biases displayed a dependence on station elevation. 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, the proposed study aims to understand RAP analysis and RAP and HRRR forecast errors exclusively within landfalling TCs.

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. This study seeks to bridge the gap between these two types of studies by using observed and model atmospheric

sounding data within the TC envelope to investigate the representation of TCTOR-related variables in the widely-used RAP and HRRR forecast models. The existing literature on TCTORs and RAP and HRRR model errors in continental environments motivates testing of the following hypotheses:

1. Model performance (both in the analysis and forecasts) will improve as the TC environment becomes more characteristic of a typical mid-latitude continental environment (i.e., as distance from the TC center increases and as TC intensity weakens).
2. Model performance will improve as the TC moves inland and more observations from the TC envelope are assimilated into the model analysis.
3. Forecast errors in both the RAP and the HRRR model will exceed RAP analysis errors.
4. Model errors in sounding-derived parameters specifically related to TCTOR forecasting will reflect the model errors seen in the raw sounding variables (temperature, dew point, wind speed and direction).

This work is part of a larger study, which is currently being conducted in partnership with the National Weather Service, aimed at better understanding differences in the near-cell environments of tornadic and non-tornadic cells within the rainbands of TCs to ultimately improve the TCTOR forecasting and warning process. However, before such analysis can be conducted, this proposed study is necessary to characterize

model analysis and forecast parameters in the representation of these near-cell environments, particularly in sounding-derived parameters that are often used in tornado forecasting.

3. DATA AND METHODS

3.1. TC Data

The environments of TCs that occurred during the 2017-2019 Atlantic hurricane seasons are analyzed in this research. From these seasons, only the TCs that made landfall along the coastline of the contiguous United States are used to ensure adequate observational radiosonde data to compare to the model profiles. Applying this landfall constraint yields five TCs from the 2017 season, four from 2018, and four 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 3.1, 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 sourced from the Edwards TCTOR database (Edwards 2010). It should be noted that four of the TCs used in this study have fewer than five TCTORs attributed to them. This reinforces that the purpose of this study is not to capture specific TCTOR cases or analyze near-cell environments, but rather to 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 Name	Mainland U.S. Landfall Date	Number of TCTORs	Peak Intensity (knots)	Category at Peak Intensity
Cindy	6/22/2017	18	50	TS
Emily	7/31/2017	1	50	TS
Harvey	8/25/2017	52	115	4
Irma	9/10/2017	28	155	5
Nate	10/8/2017	21	80	1
Alberto	5/29/2018	4	55	TS
Florence	9/14/2018	44	130	4
Gordon	9/4/2018	7	60	TS
Michael	10/10/2018	16	140	5
Barry	7/13/2019	1	65	1
Dorian	9/6/2019	25	160	5
Imelda	9/17/2019	2	40	TS
Nestor	10/19/2019	6	50	TS

Table 3.1: 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.

Track data for each of the 13 TCs is acquired from the Atlantic HURDAT2 dataset, which is comprised of post-storm analyzed TC best track data (Landsea and Franklin 2013). These are documented observations of the TC every six hours at 00Z, 06Z, 12Z, and 18Z. 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 six-hourly time point. The information collected at each point along the track 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 intensity of the TC in knots. For this research, any points in the TC track data not occurring at 00Z, 06Z, 12Z, or 18Z are discarded and the latitudes, longitudes, pressures, and intensities from the remaining 6-hourly points are interpolated linearly to yield hourly track data (Fig. 3.1).

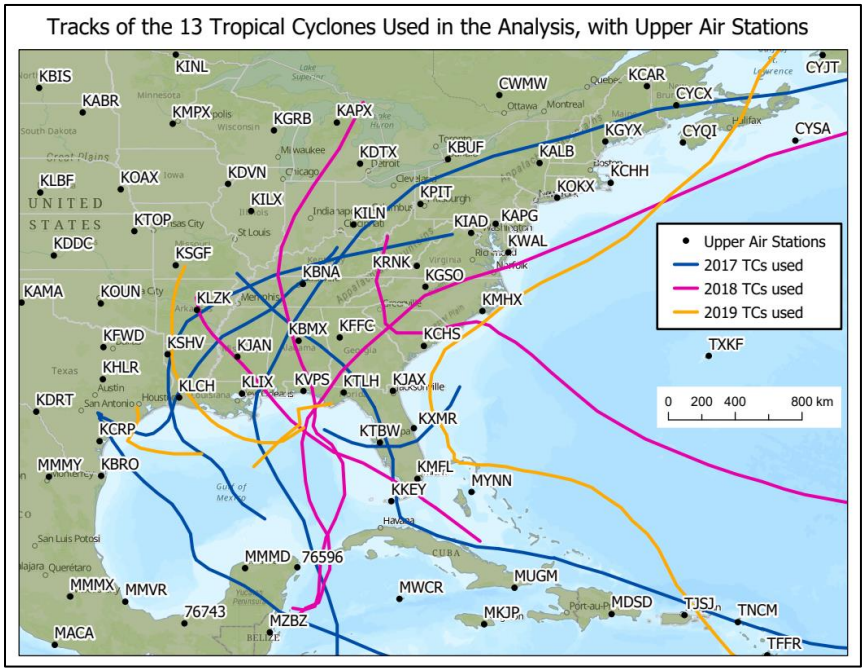

Figure 3.1: The tracks of the 13 TCs used in this study, along with the locations of all available upper air observing stations.

3.2. 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," from which radiosondes are launched. 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, which accounts for the curvature of the earth over long spatial scales in its calculation of the distance between two locations. Then, all stations falling within a radius of 800 km from the TC center at that given time are marked as eligible to be included in this study (Fig. 3.2). Finally, the University of Wyoming archive of observed soundings is checked for all of the eligible soundings and those that exist are downloaded (University of Wyoming 2020). The radius of 800 km is chosen in order to fully encompass the broad TC environment, and is consistent with the 800-km radius used by McCaul (1991) in their assessment of the spatial distributions of TCTOR-related environmental variables.

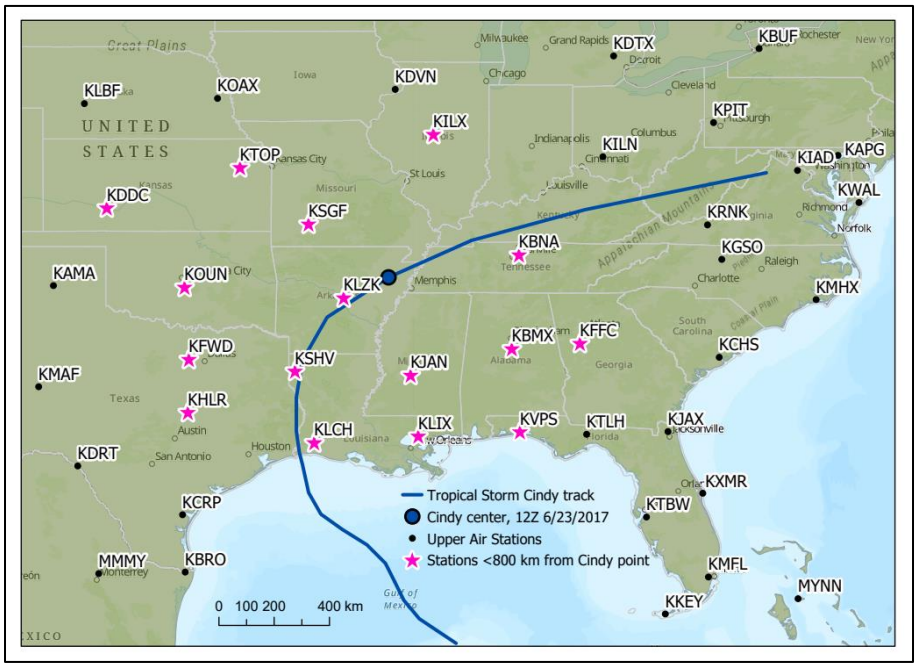

Figure 3.2: Demonstrating the process of selecting the eligible observed soundings for analysis using one of the points along the track of Tropical Storm Cindy (shown in blue). The stations indicated by pink stars are within 800 km of Cindy's center at this time step and will be included in the analysis, while the stations indicated by black dots fall outside of the 800-km radius and are therefore excluded from the analysis.

Due to the hourly nature of the interpolated TC track data, any available special soundings are collected with this method in addition to the standard soundings at 00Z and 12Z. The vast majority 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, dew point, and wind speed and direction, which are interpolated to 100-meter intervals from 0 to 16 km above ground level (AGL).

3.3. Model Soundings

As discussed in Chapter 2, the models used in this study are the RAP model with 13-kilometer grid spacing and the HRRR model with 3-kilometer grid spacing, both run by NCEP. For each observed sounding, the RAP analysis vertical profile from the same time, and the nearest grid point to the station, is downloaded from the National Centers for Environmental Information (NCEI) model data archive (NOAA NCEI 2020). Then, the corresponding 12-hour RAP and HRRR forecast soundings valid at the analysis time and the nearest model grid point to the station are downloaded. These analysis and forecast soundings contain pressure, height, temperature, dew point, and wind data and are interpolated in the same way as the observed soundings, to 100-meter intervals from 0 to 16 kilometers AGL. It should be noted that RAPv4 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). Also, the RAP analysis and forecast initializations are missing for (inclusive) 9/5/2019 03Z - 9/7/2019 12Z, which limits the model data and therefore the number of sounding pairs available for Hurricane Dorian.

3.4. Sounding-derived Parameters

The Sounding and Hodograph Analysis and Research Program in Python (SHARPpy) software is used to compute sounding-derived parameters for the observed, model analysis, and model forecast soundings (Blumberg et al. 2017). Shear magnitudes and lapse rates are calculated over various layers, along with SRH values. 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. Finally, precipitable water, downdraft CAPE (DCAPE), and 0-3 km CAPE are computed, in addition to composite parameters such as the supercell composite parameter (SCP), significant tornado parameter (STP), and energy helicity index (EHI). CINH-scaled SCP values are calculated according to the method detailed by Gropp and Davenport (2018).

3.5. Data Analysis Methods

After quality control, 1,730 observed soundings are suitable for comparison to the model profiles. Model domain restrictions and the aforementioned missing forecast

data result in 1,730 corresponding RAP analysis profiles, 1,681 RAP 12-hour forecast profiles, and 1,647 HRRR 12-hour 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.

The investigation of the model errors takes the form of a two-part analysis, the first of which is the examination of the RAP analysis, RAP 12-hour forecast, and HRRR 12-hour forecast errors in the raw sounding variables: temperature, dew point, and wind. These errors are initially explored by making two-dimensional plots of the error as a function of height only, and later with heat maps that plot the error as a function of height and another variable of interest. Five variables of interest are chosen to test the hypotheses about the model errors in relation to the TC spatially and temporally; they are: distance from the TC center, bearing relative to the TC center, intensity of the TC at that time, proximity of the TC to its landfall point, and the amount of time before or after landfall. A sixth and final variable of interest is the proximity of the station to the coast.

The second part of the analysis examines the model analysis and forecast errors in the sounding-derived parameters. The same variables of interest from the first part are used to gain a better understanding of the errors in relation to the TC, with the only difference being the one-dimensional nature of the sounding-derived parameter errors as opposed to the two-dimensionality of the height-dependent errors in the raw sounding variables.

4. RESULTS

4.1. Errors in Raw Sounding Variables

Mean error profiles of temperature, dew point, and wind speed are generated for the RAP analysis, RAP 12-hour forecast, and HRRR 12-hour forecast. The mean error profiles are calculated by taking each model value at a certain height minus its corresponding observed value, and then averaging across all available sounding pairs at that height. The two-tailed, non-parametric Wilcoxon signed-rank test is performed at each height of each mean error profile to test whether the errors are significant at the 95% confidence level, that is, if the median error at that height is significantly different from zero. Profiles of mean absolute error for each raw sounding variable are computed similarly to the mean error profiles, but the absolute value of each difference is taken before averaging at each height. These errors in the raw sounding variables are examined in the context of Hypothesis 3 from Chapter 2, which states that the forecast errors are expected to exceed the errors in the RAP analysis.

Mean temperature errors do not exceed a magnitude of 0.4°C at any height in any of the three error profiles (Fig. 4.1a). The RAP analysis (-0.36°C), RAP 12-hour forecast (-0.16°C), and HRRR 12-hour forecast (-0.12°C) all have cool biases at the surface. This surface cool bias is largest in magnitude for the RAP analysis, and the RAP analysis temperature errors remain negative throughout much of the column up to about 9 km before changing to a warm bias in the upper troposphere. The HRRR 12-hour 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-hour forecast have the largest magnitude of any of the three profiles within the troposphere. The surface cool bias in the RAP 12-hour forecast profile is generally maintained up to 2 km, and is followed by a warm bias from 2-5.5 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) in their analysis of RAP and HRRR thermodynamic errors in continental convective environments (Fig. 2.2b); both the RAP and HRRR 11-hour forecasts in their study were found to be continuously warm-biased in the layer from 0-5 km.

Profiles of the mean absolute temperature errors (Fig. 4.1b) show a distinct separation between the RAP analysis and the forecasts, with the smallest mean absolute temperature errors in the RAP analysis and the largest in the RAP 12-hour forecast. The RAP 12-hour forecast exhibits the greatest mean absolute temperature error of the three model runs throughout the profile, and the magnitudes of HRRR-12-hour forecast absolute temperature errors fall in between those of the RAP 12-hour forecast and RAP analysis. These results support Hypothesis 3.

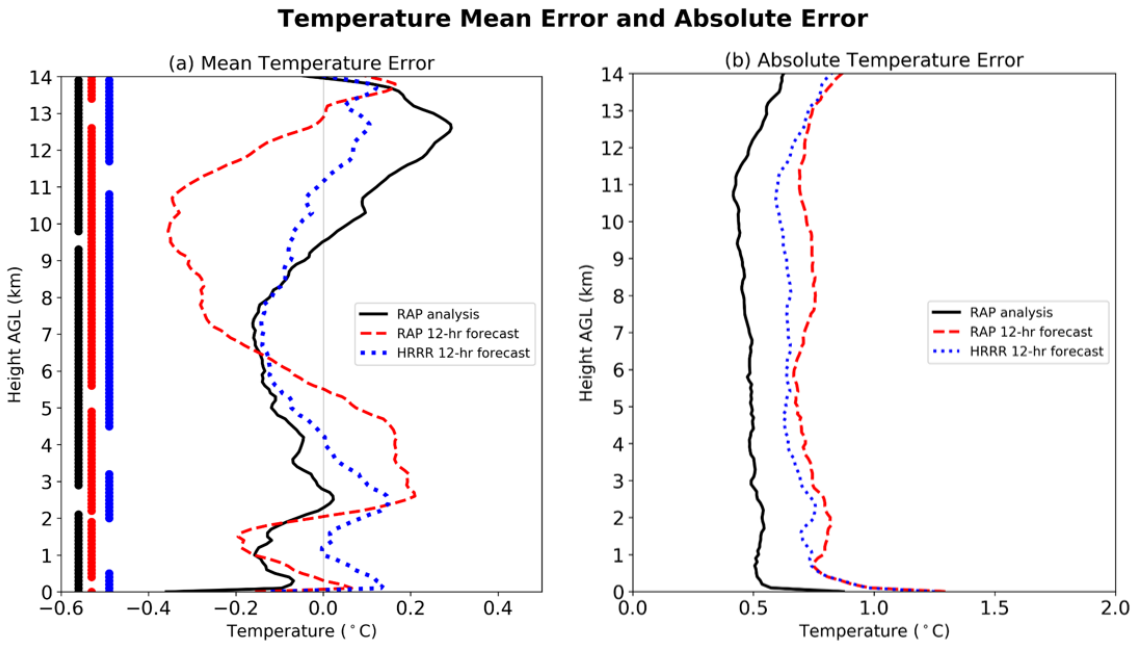

Figure 4.1: Vertical profiles of temperature (a) mean error and (b) mean absolute error from 0-14 km AGL for the RAP analysis, RAP 12-hour forecast, and HRRR 12-hour forecast. Solid black, red, and blue dots at the left edge of panel (a) denote the levels at which the temperature errors are significant to at least 95% confidence, assessed using the two-tailed, non-parametric Wilcoxon signed-rank test. The number of observed-model sounding pairs contributing to each error profile are as follows: RAP analysis – 1730, RAP 12-hour forecast – 1681, HRRR 12-hour forecast – 1647.

The mean dew point errors (Fig. 4.2a) have larger magnitudes than the temperature errors, and increase with height. At the surface, the RAP analysis and HRRR 12-hour forecast have dry biases, with that of the HRRR 12-hour forecast being larger with a magnitude of nearly 0.5°C. The RAP 12-hour forecast, in contrast, has a surface dew point bias that is near zero. Above the surface, the RAP analysis and both forecasts have moist biases that worsen with height for the remainder of the vertical profile and maximize at nearly 4°C between 12 and 13 km. The mean absolute dew point

errors (Fig. 4.2b) are up to 1°C greater in the forecasts than in the RAP analysis, supporting Hypothesis 3, although there is no obvious distinction between the two forecasts as was seen in the profiles of the mean absolute temperature errors (Fig. 4.1b).

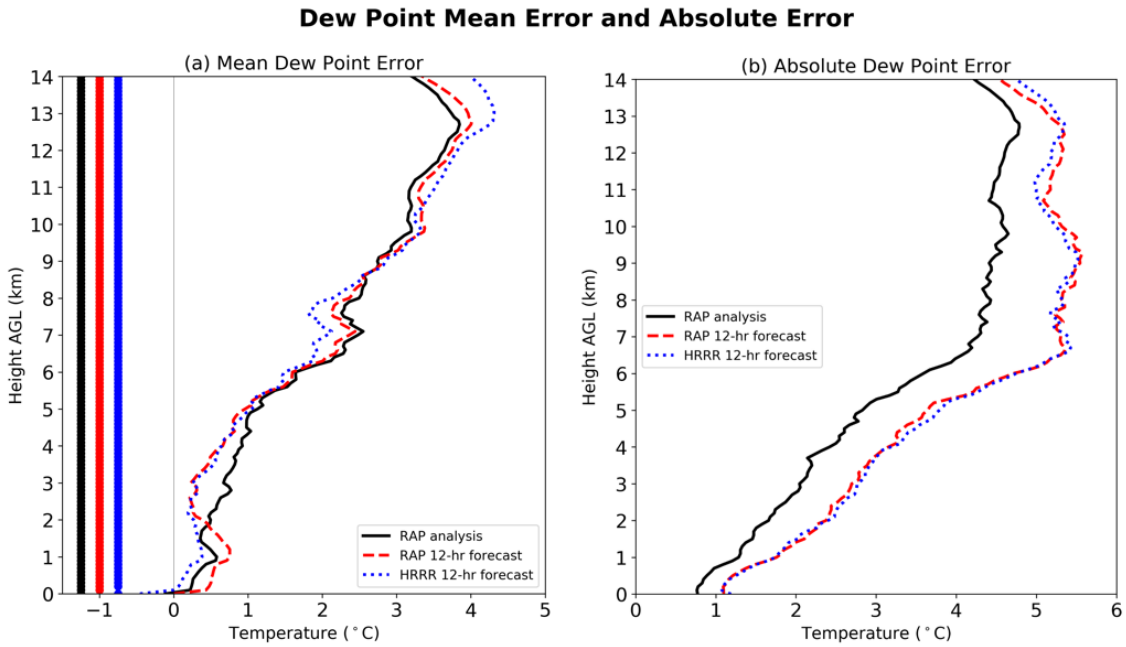

Figure 4.2: Vertical profiles of dew point (a) mean error and (b) mean absolute error from 0-14 km AGL for the RAP analysis, RAP 12-hour forecast, and HRRR 12-hour forecast. Significance testing results shown as in Figure 4.1, but for dew point errors. Sample sizes as in Figure 4.1.

While the error profiles for the continental convective environments in Evans et al. (2018) were truncated at an altitude of 5 km, similar patterns to those of the dew point errors in this study can be observed within this height range. In their study, both the RAP and HRRR 11-hr 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. 2.2a), while the moist biases in the 11-hour forecasts increased more rapidly to around 3°C by the same height (Fig. 2.2b). In this study of TC environments, the dew point 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. 4.2a). These results contradict Hypothesis 1, which posited that the models would perform better under typical mid-latitude continental conditions. When contrasted with the continental results in Figure 2.2, the results presented in Figure 4.2a suggest that model predictions of moisture are better within TC envelopes than in continental environments, at least within the limited vertical extent of 5 km AGL.

Examination of the mean wind speed errors with height (Fig. 4.3a) reveals that wind speeds are overpredicted by the RAP analysis and both forecasts in a shallow layer near the surface. The largest near-surface overpredictions, exceeding 2.5 kt, occur in the HRRR 12-hour forecast. Above 1 km, all three error profiles transition to an underprediction of wind speed, with the largest underpredictions in the RAP 12-hour forecast. The smallest underpredictions, never exceeding a magnitude of 0.75 kt, occur in the HRRR 12-hour forecast such that there are layers where the error is found to be insignificant at the 95% confidence level. Despite these small biases in the HRRR 12-hour forecast, the mean absolute error profiles (Fig. 4.3b) reveal that RAP analysis absolute wind speed errors tend to be smaller than those of both forecasts throughout the depth of the profile by 1-2.5 kt, which supports Hypothesis 3. 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 originating at the surface; shear will be explored along with other sounding-derived parameters in a later subsection of these results.

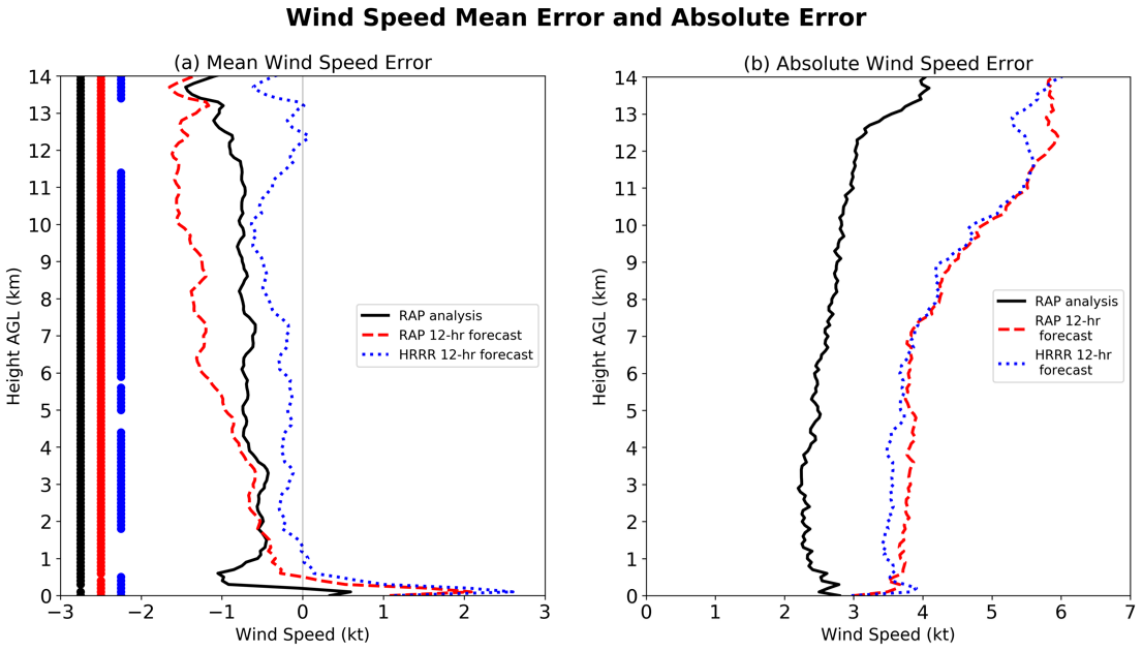

Figure 4.3: Vertical profiles of wind speed (a) mean error and (b) mean absolute error from 0-14 km AGL for the RAP analysis, RAP 12-hour forecast, and HRRR 12-hour forecast. Significance testing results shown as in Figure 4.1, but for wind speed errors. Sample sizes as in Figure 4.1.

Proper investigation of Hypotheses 1 and 2 requires analysis of how model errors vary as a function of both height and other TC-related variables. Figures 4.4-4.6 show the RAP analysis, RAP 12-hour forecast, and HRRR 12-hour forecast temperature

and dew point errors as a function of height on the y-axis and time relative to TC landfall on the x-axis. Time before or after the TC landfall is reported as a negative or positive number of days, respectively. The errors are binned according to selected intervals along the axes and the mean of each bin is shown by its color.

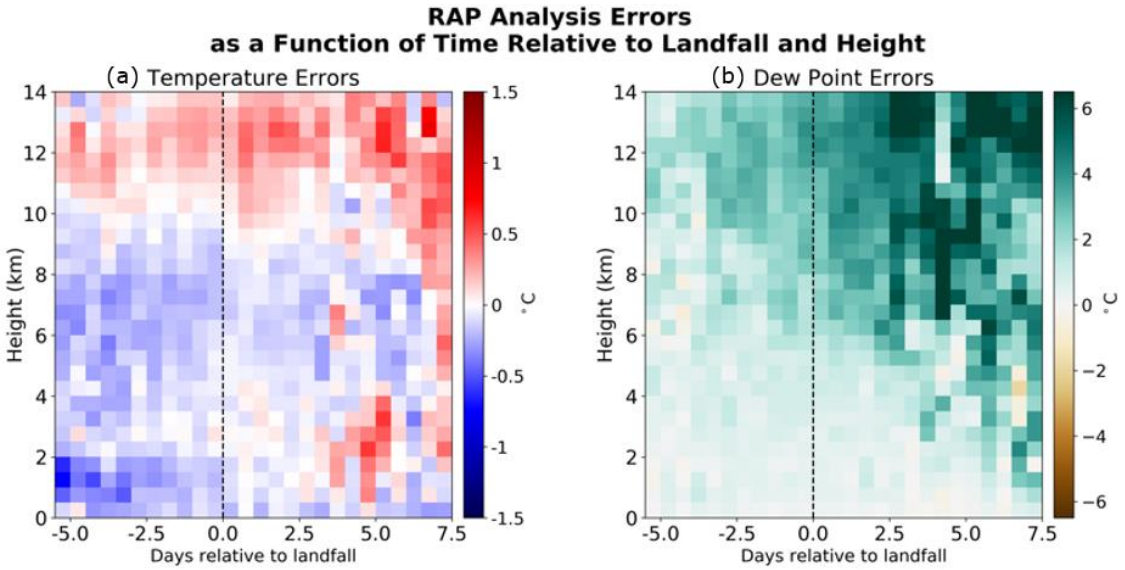

Figure 4.4: RAP model analysis errors in (a) temperature in °C and (b) dew point in °C. The errors are plotted as a function of height in kilometers and time relative to landfall in days, such that an x-value of -2.5 corresponds to 2.5 days before landfall and an x-value of +2.5 corresponds to 2.5 days after landfall. The vertical dashed line represents the TC landfall time.

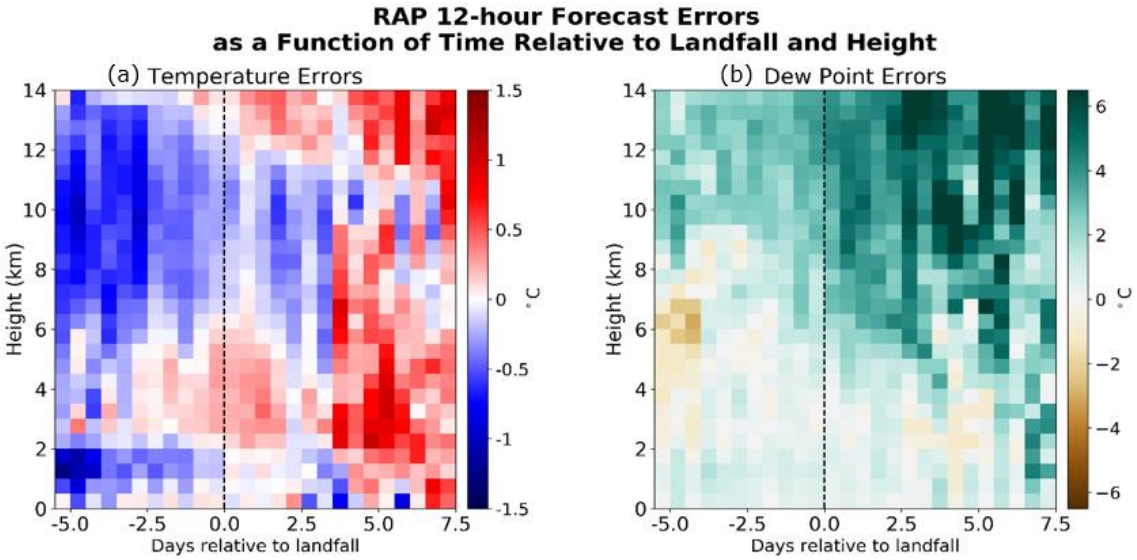

Figure 4.5: As in Figure 4.4, but for the RAP 12-hour forecast.

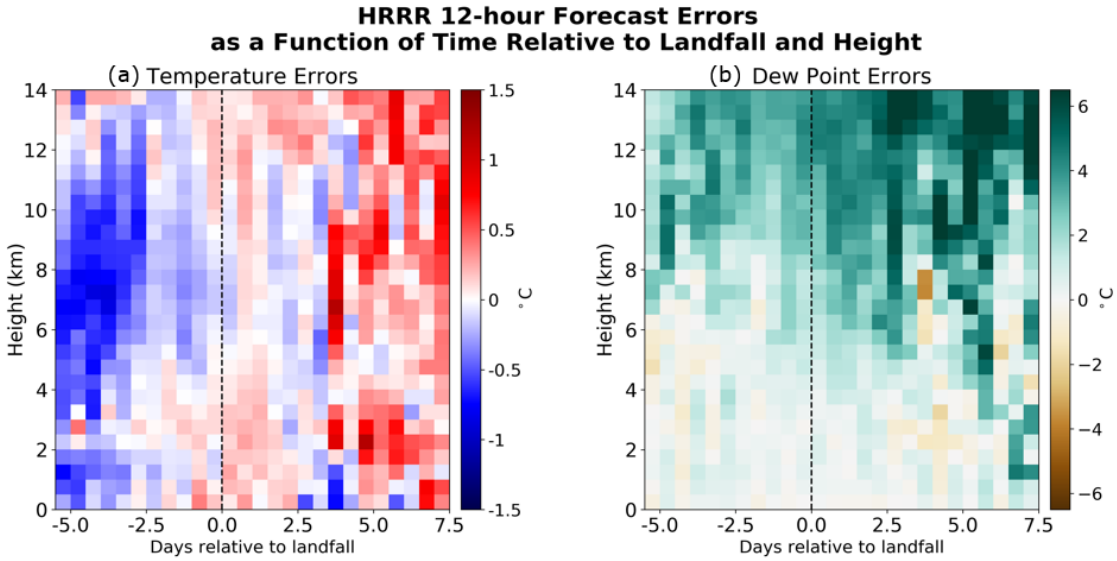

Figure 4.6: As in Figure 4.4, but for the HRRR 12-hour forecast.

Comparison of the RAP analysis, RAP 12-hour forecast, and HRRR 12-hour forecast errors in Figs. 4.4-4.6 reveals that the temperature and dew point biases have larger magnitudes overall in the two forecasts than in the RAP analysis, supporting Hypothesis 3. Wind speed is weakly underpredicted by the RAP analysis for all times relative to the TC landfall (not shown). There are slight overpredictions of wind speed by the RAP and HRRR 12-hour forecasts in the bins from 0-0.5 km, with slightly stronger underpredictions than those of the RAP analysis above that and across the rest of the time relative to landfall (not shown).

Before landfall, there is a cold bias in the RAP analysis throughout the column up to about 10 km (Fig. 4.4a) that is consistent with the overall model temperature errors seen in Figure 4.1a. However, this cold bias switches to an unusual low-level warm bias, from the surface up to about 4 km, starting about three days after landfall. A strong warm bias is also seen in both the RAP and HRRR 12-hour 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. 4.5a, 4.6a). Similar to the RAP analysis, both the RAP and the HRRR 12-hour forecast exhibit a cold bias in the days leading up to landfall, which is retained in the RAP 12-hour forecast between 6 and 12 km for a period of days past the landfall (Fig. 4.5a).

The moist bias evident in the dew point errors of the analysis and both forecasts descends closer to the surface 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 mid-latitudes (Figs. 4.4b, 4.5b, 4.6b). According to Hypotheses 1 and

2, the errors would be expected to diminish in the days following the TC landfall, but neither the temperature nor the dew point errors for any of the model runs follow this anticipated pattern.

To further test Hypothesis 1, errors are also examined as a function of the intensity of the TC at the time of the observed sounding (Figs. 4.7-4.9). The vertical dashed line on each plot marks the 64-knot wind speed threshold between tropical storm and category 1 hurricane. The slight cold temperature bias between the surface and 10 km that was evident in the RAP analysis mean error profile is generally present across the entire range of intensity values (Fig. 4.7a), but the cold temperature bias between 0-2 km worsens with increasing intensity. The RAP and HRRR 12-hour forecasts both have a warm bias between 2-6 km when the intensities are sub-hurricane strength, which is overlaid by an upper-level cold bias from 6-12 km in the RAP 12-hour forecast (Figs. 4.8a, 4.9a). At higher intensities of about 90 kt and above, temperatures are generally underpredicted throughout the column by both the RAP and the HRRR 12-hour forecasts, slightly more so by the RAP.

For both forecasts and the analysis, the magnitudes of the temperature errors are smaller at weaker intensities and increase as the intensity increases, which supports Hypothesis 1. However, the dew point errors do not support Hypothesis 1 because they are larger at weaker intensities for all model runs, with mid- to upper-tropospheric moisture (above 6 km) more severely overpredicted at intensities less than 64 knots (Figs. 4.7b, 4.8b, 4.9b). Hypothesis 3 follows the same pattern of being supported by the temperature errors and not by the dew point errors: the temperature errors in the two 12-

hour forecasts are generally larger in magnitude than the RAP analysis temperature errors, but a difference in the magnitude of the dew point errors between the analysis and the forecasts is not discernible.

The columns showing mean temperature errors in the RAP and HRRR 12-hour forecasts for sounding pairs valid while the TC intensity was within the interval [110, 120 kt) are representative of a small subset of 19 sounding pairs (Figs. 4.8a, 4.9a). Eight of these were valid at a single analysis time during Hurricane Harvey, and nine were valid within a 24-hour period during Hurricane Irma. Mean temperature errors in this column for both forecasts appear inconsistent with the errors in neighboring columns, and could be sensitive to the lack of diversity in the sample.

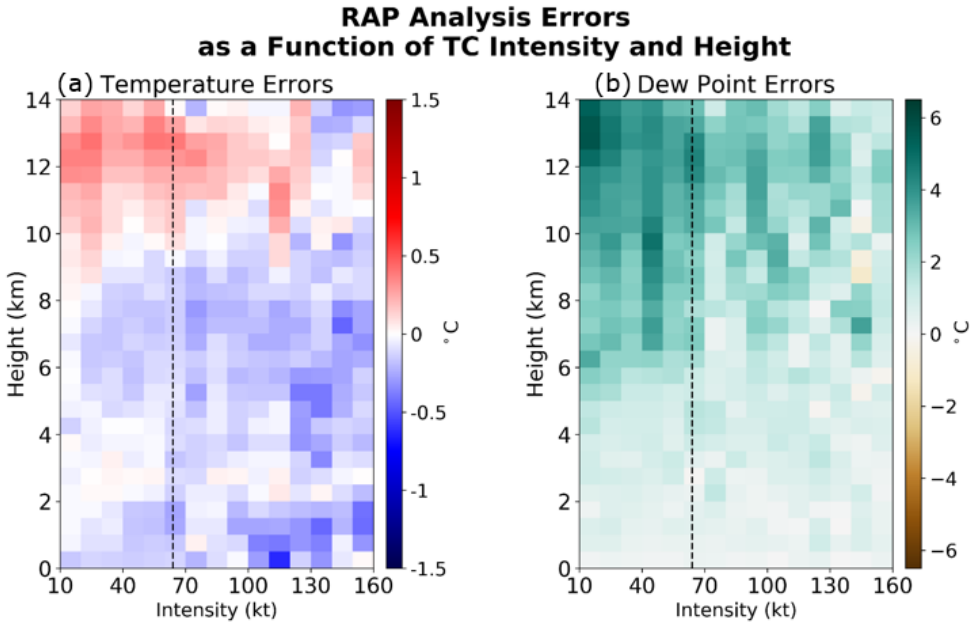

Figure 4.7: RAP model analysis errors in (a) temperature in °C and (b) dew point in °C. The errors are plotted as a function of height in kilometers and TC intensity in knots at the time the sounding pair is valid. The vertical dashed line represents the threshold between tropical storm and category 1 hurricane on the Saffir-Simpson scale.

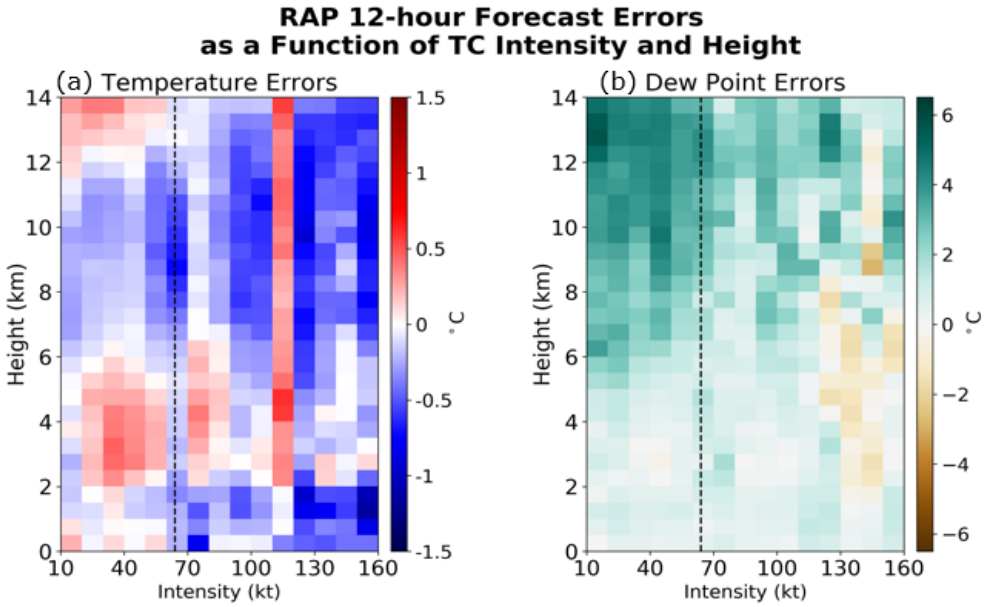

Figure 4.8: As in Figure 4.7, but for the RAP 12-hour forecast.

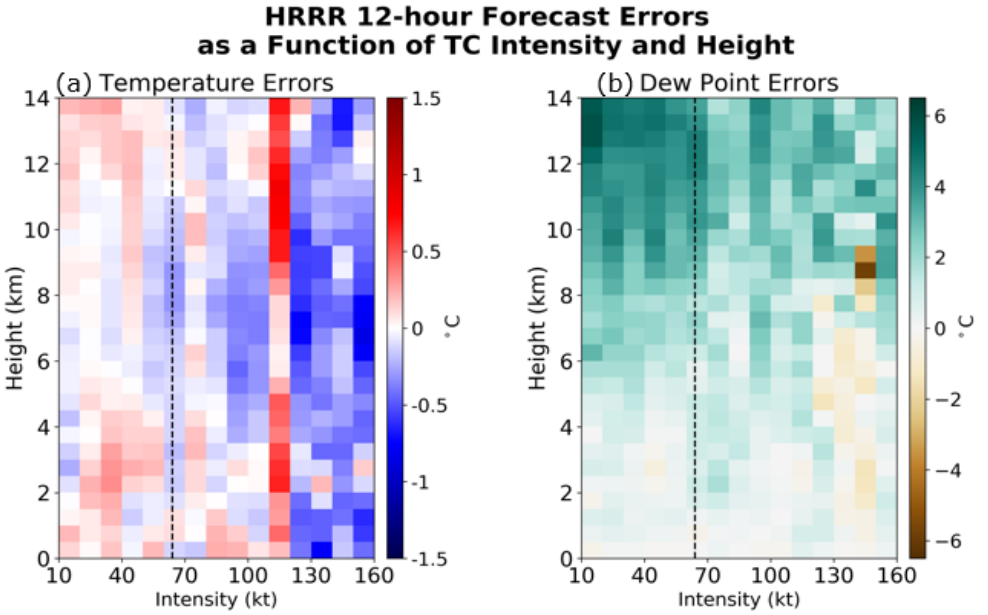

Figure 4.9: As in Figure 4.7, but for the HRRR 12-hour forecast.

As a final test of Hypothesis 1, errors are plotted as functions of distance from the center of the TC, which ranges from 0 to 800 km in this dataset. The magnitudes of the temperature and dew point errors in these heat maps are smaller than their magnitudes in the heat map figures presented so far. With the exception of the innermost ring within 200 km of the TC center, RAP analysis temperature and dew point errors (Figs. 4.10a, b) are nearly uniform as distance from the TC center increases. They mimic the RAP analysis mean error profiles for temperature and dew point (Figs. 4.1a, 4.2a), with negative temperature errors from the surface up to about 10 km and dew point errors gradually increasing with height throughout the depth of the troposphere. The temperature errors are more variable with height within the innermost 200 km of the TC center, where cold biases from roughly 0-4 and 6-8 km create an alternating pattern with the warm biases, which extend from 4-6 km and 8-14 km. This breakdown of temperature errors is most likely due to the complex nature of that region of a TC and the coexistence of quickly evolving intense rain bands, isolated convective cells, and patches of clearing. Similar chaos in the dew point errors within the innermost 200 km is likely not seen because of the pervasive presence of moisture that permeates the TC envelope.

The RAP and HRRR 12-hour forecast dew point errors behave similarly to those of the RAP analysis, showing little to no variation with distance (Figs. 4.11b, 4.12b). Both forecasts show larger overpredictions of temperature within the innermost 200 km. In the range of 200 to 800 km from the TC center, temperature is generally overpredicted at low levels and underpredicted aloft in the HRRR 12-hour forecast (Fig.

4.12a). This could translate to overpredictions of CAPE. In the RAP 12-hour forecast, the temperature errors have a slightly larger magnitude and take on a more stratified structure: temperature is underpredicted at low levels, overpredicted at mid-levels, and underpredicted again above 6 km (Fig. 4.11a). Conversely, this would likely lead to underpredictions of CAPE by the RAP 12-hour forecast.

Hypothesis 1 is supported only by the slight decrease in magnitude of the temperature errors as distance from the TC center increases in the HRRR 12-hour forecast. In all other variables and in the RAP analysis and RAP 12-hour forecast, errors are too uniform with increasing distance from the center to conclude that model performance improves as the environment becomes more characteristic of typical mid-latitude conditions.

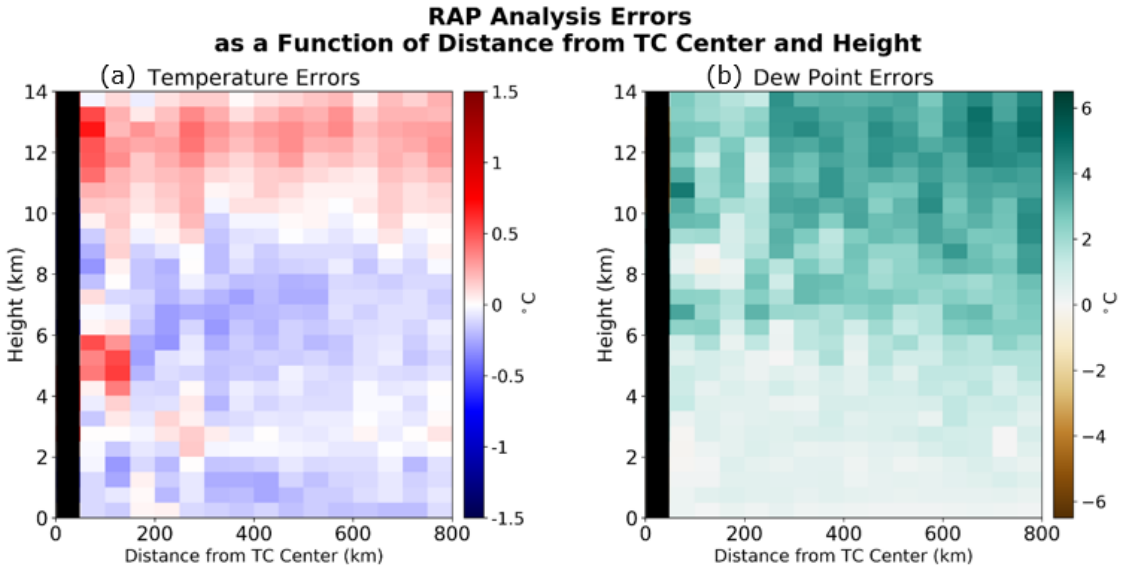

Figure 4.10: RAP model analysis errors in (a) temperature in °C and (b) dew point in °C. The errors are plotted as a function of height and distance from the TC center, both in kilometers. Columns shown in black have a sample size of less than 4 sounding pairs.

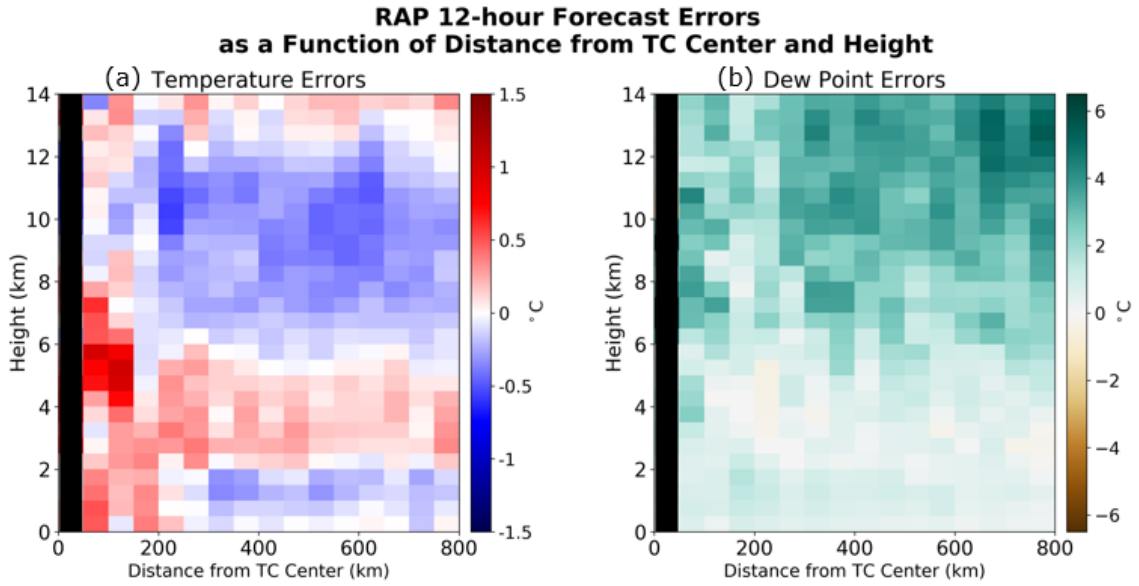

Figure 4.11: As in Figure 4.10, but for the RAP 12-hour forecast errors.

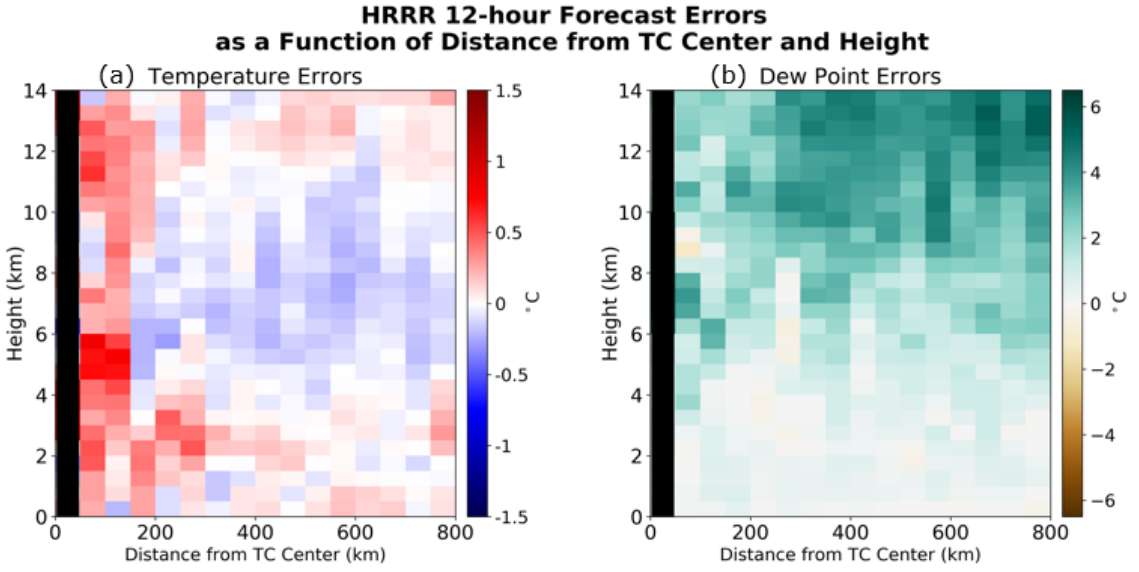

Figure 4.12: As in Figure 4.10, but for the HRRR 12-hour forecast errors.

Errors are also plotted as a function of north-relative azimuth based on their bearing relative to the TC center (not shown). These plots reveal a low-level cold bias on the south side of the TC (from 90 to 270 degrees) that is worse in the forecasts than the analysis, thereby supporting Hypothesis 3. It is overlaid by a warm bias from 2-5 km on the east side of the TC (from 0 to 180 degrees) that only appears in the two forecasts. When sounding-derived parameter errors are explored, they will be plotted on polar plots with a maximum radius of 800 km so that they can be examined as functions of azimuth and distance together and spatial trends relative to the TC center can be more intuitively analyzed.

4.2. Errors in Sounding-derived Parameters

The ingredients-based approach used by forecasters to discern when and where tornadoes are probable employs numerous thermodynamic and kinematic sounding-derived parameters such as CAPE, lapse rates, shear, and storm-relative helicity. Favorable conditions need to align in time and space for tornado production, and tornado environments can be much more subtle within the TC envelope than in a classic mid-latitude convective environment. Therefore, understanding common error patterns in the widely used RAP and HRRR models specifically in TC environments is beneficial. Errors in selected sounding-derived parameters are presented here, being sure to include those 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 quantities from corresponding model and observed soundings are subtracted from each other (model value minus observed value) to generate the dataset of error values. "Nan" values are produced and later discarded whenever one of the parameters is missing or undefined in either the model sounding or the observed sounding. The values are then binned along the x axis according to another variable of interest and the mean and standard deviation are calculated for each bin. Finally, the two-tailed, non-parametric Wilcoxon signed-rank test is performed for each bin along the x-axis to test whether the errors are significant at the 95% confidence level. This is the same statistical test performed at each height of the mean error profiles, and it tests for significance by testing if the median error in that bin is significantly different from zero. Significant results in a bin are depicted with a solid circle plotted along the line at the midpoint of that bin. All four of the aforementioned hypotheses from Chapter 2 are applicable to this section of the results, but most of the focus will be on whether the model errors in sounding-derived parameters reflect the errors expected to result from the errors in the raw sounding variables (Hypothesis 4).

Surface-based, most-unstable, and mixed-layer CAPE errors are examined first as a function of time relative to landfall (Fig. 4.13). MLCAPE errors (Fig. 4.13e) tend to be smaller in magnitude, but positive as opposed to the negative biases more consistently seen in SBCAPE and MUCAPE. The HRRR 12-hour forecast has larger negative biases

than the RAP analysis and forecast in SBCAPE and MUCAPE within the region of significant results. This could be partially due to the fact that the HRRR 12-hour forecast has the largest dry bias at the surface (Fig. 4.2a). Conversely, the HRRR 12-hour forecast outperforms the RAP 12-hour forecast in its prediction of MLCAPE; this could be attributed to the slightly smaller dew point errors in the HRRR 12-hour forecast throughout the typical depth of a surface-based mixed layer (Fig. 4.2a). These subtle distinctions in the results support Hypothesis 4.

While the magnitudes of the MLCAPE errors decrease during the four days following landfall, the magnitudes of the SBCAPE and MUCAPE errors remain more steady as time after landfall increases. So, Hypothesis 2 is only supported by the MLCAPE errors here. The standard deviation decreases for all CAPEs and all model runs as the TC approaches its landfall time and moves inland and more land-based observations are able to be assimilated into the model (Figs. 4.13b, d, f). Despite the large magnitudes of the standard deviations compared to the mean error values, the median errors are significantly different from zero in most of the bins for most of the model runs from one day before landfall to four days after landfall.

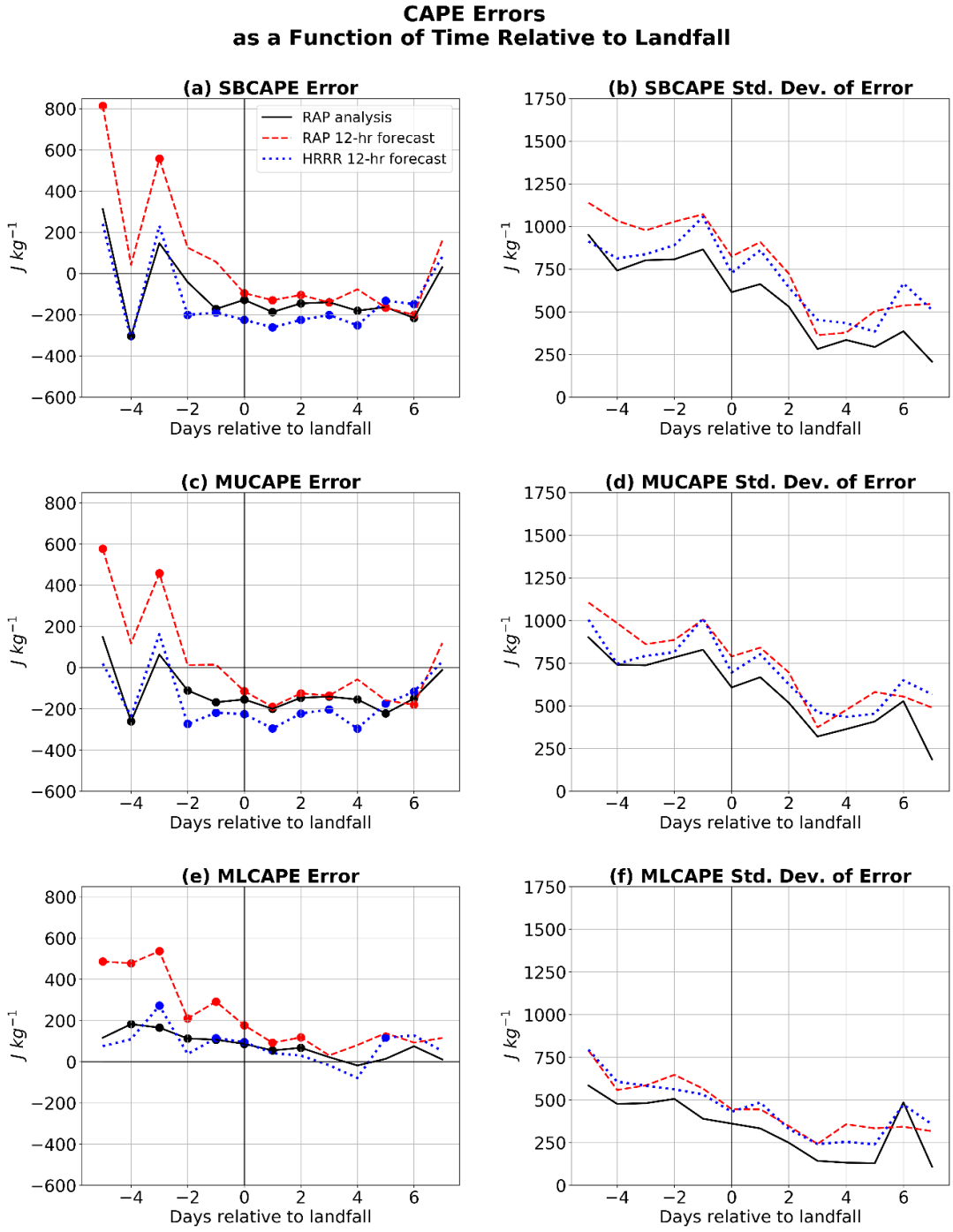

Figure 4.13: SB-, MU-, and MLCAPE mean errors (a, c, e) and standard deviations of the mean error (b, d, f), both in J kg^{-1}, plotted as functions of time relative to landfall for the RAP analysis, RAP 12-hour forecast, and HRRR 12-hour forecast. The vertical dashed line represents the time of the TC 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.

SBCAPE and MUCAPE are also generally underpredicted, and MLCAPE overpredicted, as a function of TC intensity at the intensity values for which the errors are significant (not shown). The magnitudes of the CAPE errors at lower intensities tend to be smaller than at higher intensities, so Hypothesis 1 is supported by these results.

CAPE errors are plotted as a function of distance and azimuth relative to the TC center on north-relative polar plots to allow for easy examination of the spatial distribution of the errors. Overpredictions of MLCAPE by the RAP 12-hour forecast (Fig. 4.14a) 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. 4.14b). These overpredictions are significant east of the center in a sector from 45 to 135 degrees, which tends to be where most TCTORs occur (Fig. 2.1; Edwards 2012). Hypothesis 3 is supported by the fact that the RAP analysis errors tend to be smaller in magnitude than the RAP 12-hour forecast errors when plotted relative to the TC center.

Overall, the signs of the SB-, MU-, 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 11-hour forecasts as well as the RAP analysis had surface dry biases, which likely contributed to the slight stable biases they observed in the SBCAPE and MUCAPE. MLCAPE displayed an unstable bias in their study as well, likely due to the moist biases present just above the surface.

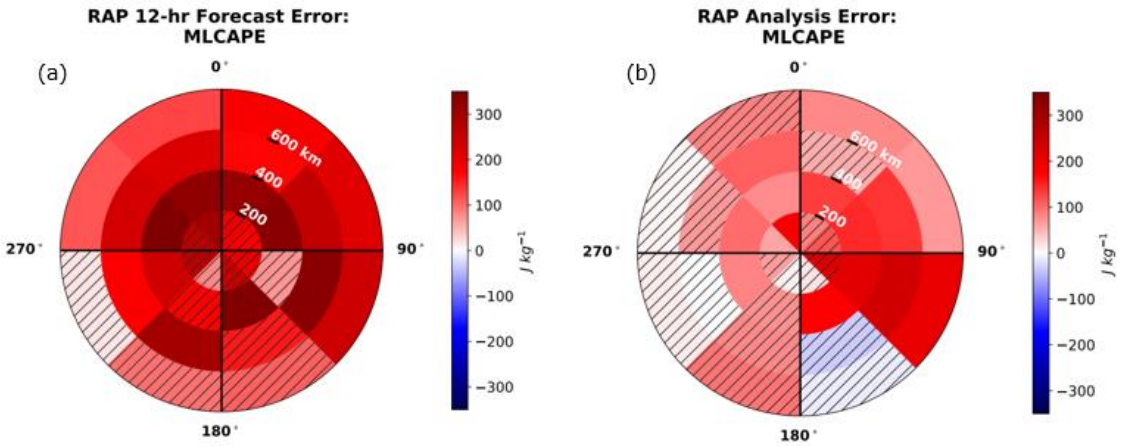

Figure 4.14: MLCAPE errors plotted in units of J kg^{-1} as a function of distance and azimuth relative to the TC center for (a) the RAP 12-hour forecast and (b) the RAP analysis. Bins in which the errors are not statistically significant at the 95% confidence level are hatched.

Lapse rates are often used alongside CAPE in ingredients-based tornado forecasting as an additional measure to diagnose stability. Lapse rates for the 0-1 km and 0-3 km layers are calculated as the temperature difference between the top and bottom of each layer, normalized by the thickness of the layer in kilometers. They are positive when the environment is cooling with height, so a layer with a more positive lapse rate is more unstable. Since errors are calculated in this study as model value minus 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.

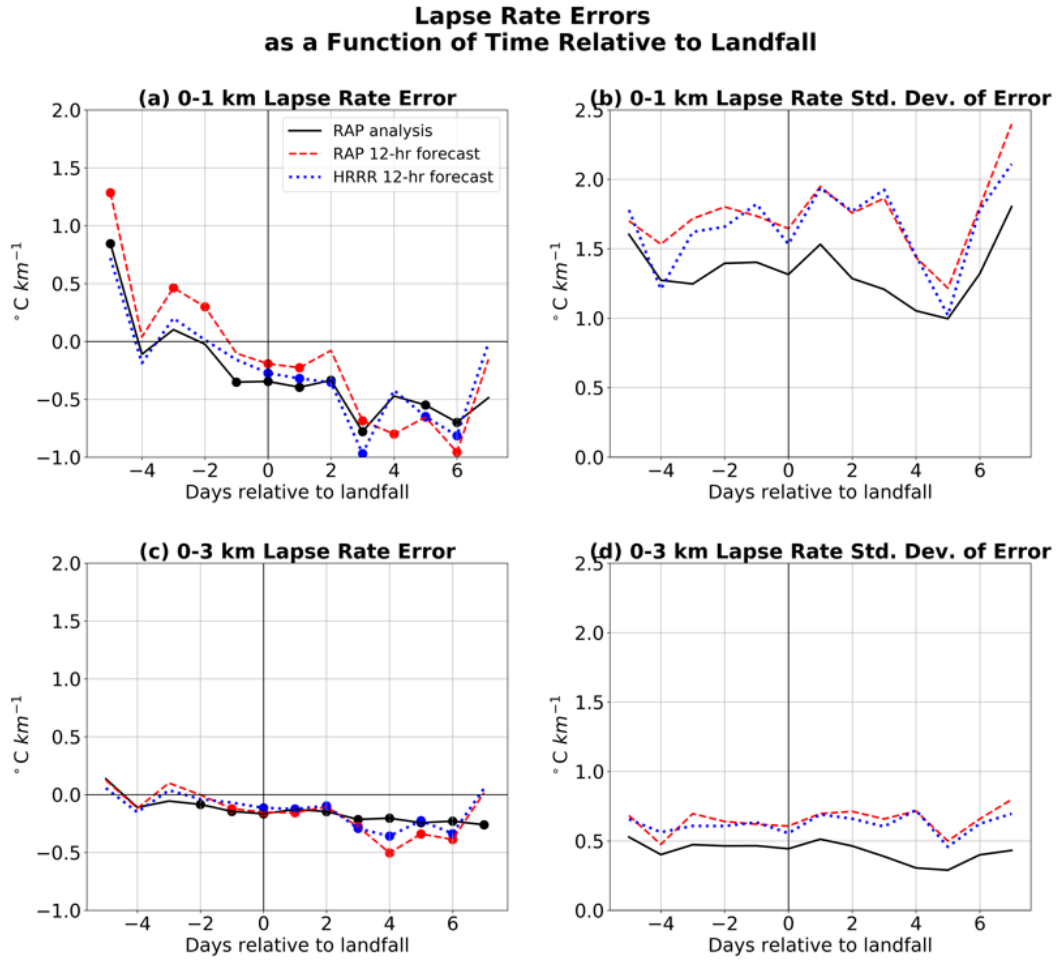

Figure 4.15: 0-1 km and 0-3 km lapse rate mean errors (a, c) and standard deviations of the mean error (b, d), both in °C km^{-1}, plotted as functions of time relative to landfall for the RAP analysis, RAP 12-hour forecast, and HRRR 12-hour forecast. The vertical dashed line represents the time of the TC 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.

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 to 1°C km^{-1} after two days post-landfall (Fig. 4.15a). The 0-3 km lapse rate errors (Fig. 4.15c) tend to be negative, with slight decreases after two days post-landfall 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 the landfall. This significant stable bias in the models in layers starting at ground level is consistent with the CAPE errors (Fig. 4.13) because SBCAPE and MUCAPE were both underpredicted by the models during this time period relative to TC landfall. The tendency of the lapse rates to be stable-biased also aligns with the surface cool bias seen in the mean temperature error profiles of both forecasts and the analysis, supporting Hypothesis 4 (Fig. 4.1a). Hypothesis 2, however, is not supported because the magnitudes of both the 0-1 and 0-3 km lapse rate errors worsen as time after landfall increases and the TC moves farther inland.

When examined as a function of TC intensity, the 0-3 km lapse rate errors (Fig. 4.16c) are generally underpredictions of similar magnitude (<0.5°C km^{-1}) to the errors as a function of time relative to landfall (Fig. 4.15c). The 0-1 km lapse rate errors (Fig. 4.16a) are also negative as a function of intensity, though with larger magnitudes (up to 1°C km^{-1}) and more variability 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 knots, in either the 0-1 or 0-3 km layer. There is not a clear enough trend as intensity decreases to conclude that Hypothesis 1 is supported, i.e., that model performance improves as intensity decreases.

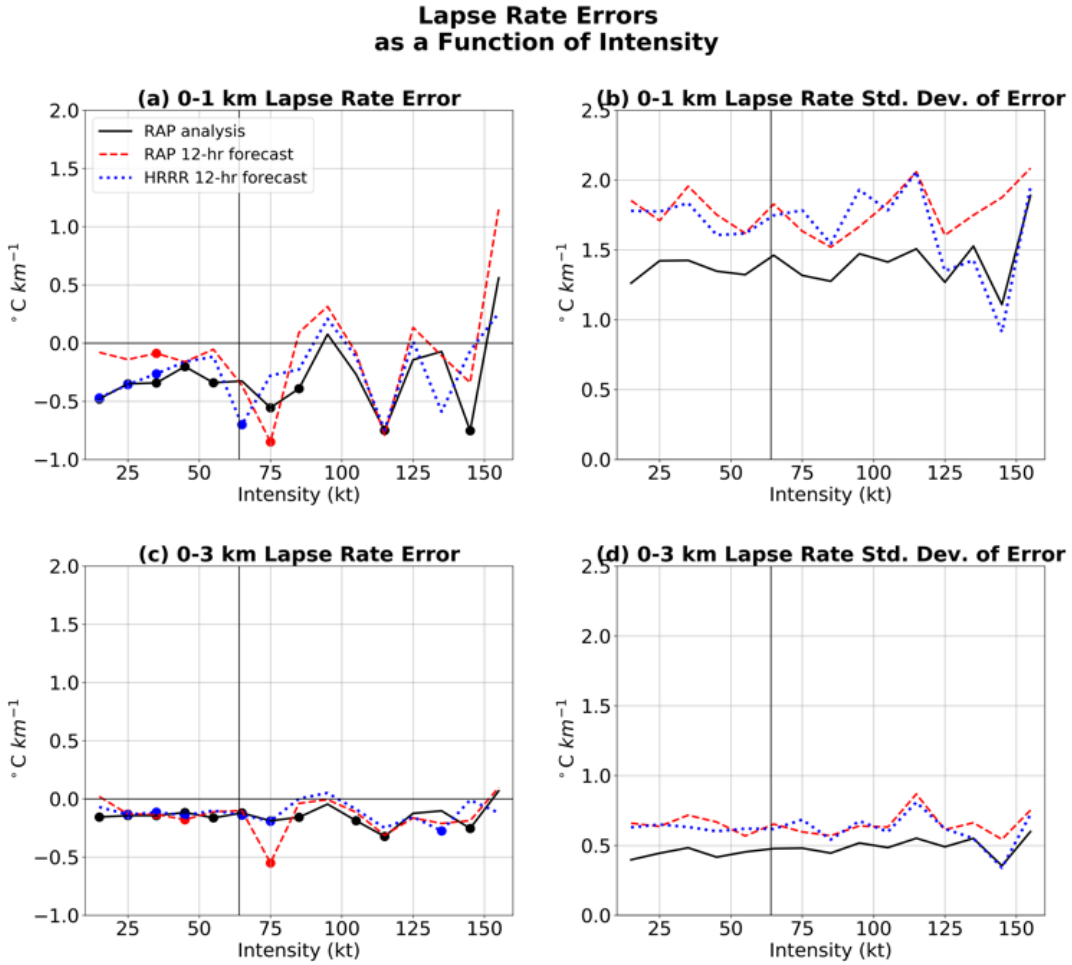

Figure 4.16: 0-1 km and 0-3 km lapse rate mean errors (a, c) and standard deviations of the mean error (b, d), both in °C km^{-1}, plotted as functions of TC intensity for the RAP analysis, RAP 12-hour forecast, and HRRR 12-hour forecast. The vertical dashed line represents the threshold between a tropical storm and category 1 hurricane on the Saffir-Simpson scale. Solid dots are placed along the line at the midpoint of each bin in which the errors are significant at the 95% confidence level.

The 0-1 km lapse rate errors are generally negative with respect to distance and azimuth, which corresponds to a stable bias. However, the RAP and HRRR 12-hour forecasts exhibit unstable 0-1 km lapse rate biases in the northeast quadrant (Fig. 4.17a, b). These unstable biases are insignificant like most of the errors within 800 km of the

TC center, but notable due to the rarity of unstable biases in the lapse rate errors examined across the model runs. Figure 4.15a shows unstable 0-1 km lapse rate biases in the period between two and five days before landfall, and the unstable biases in the northeast quadrant could have contributed to the magnitude of this unstable bias due to the fact that the northeast quadrant is more likely to be moving over land areas with observing stations during that time than quadrants on the southern side of the TC.

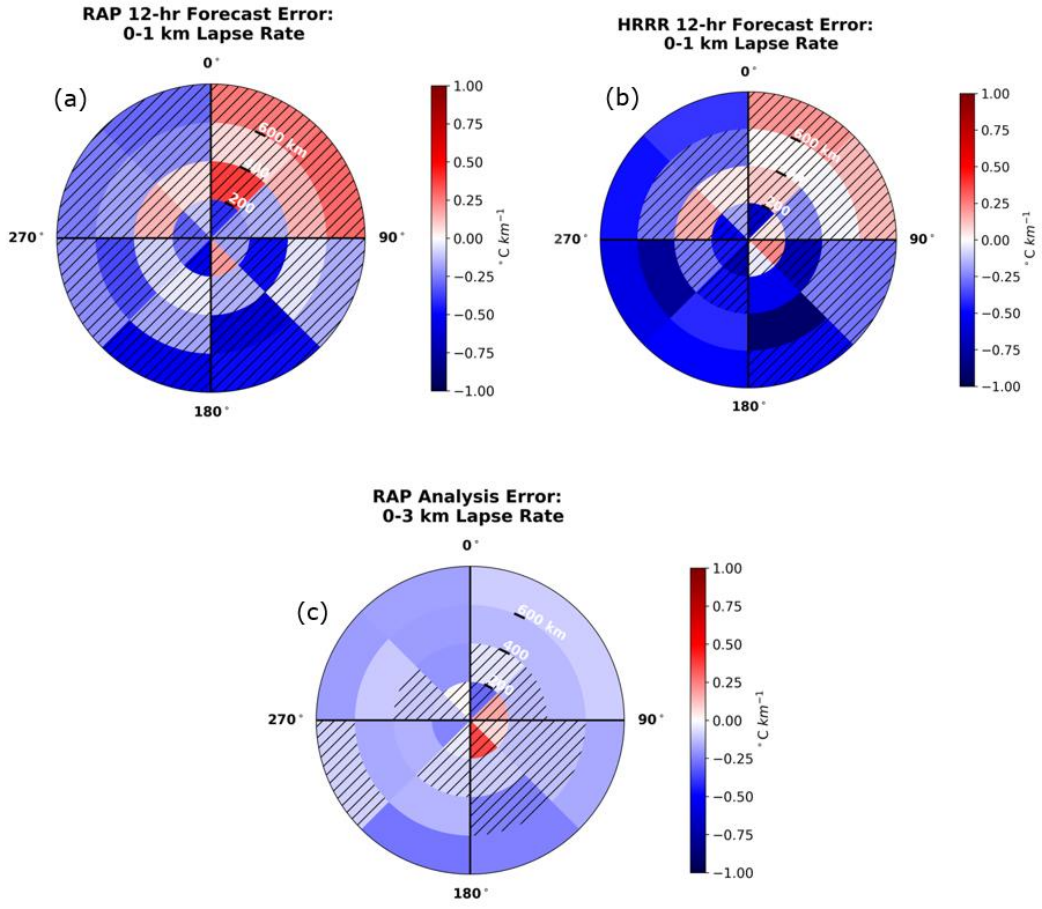

Figure 4.17: 0-1 km lapse rate errors for (a) the RAP 12-hour forecast and (b) the HRRR 12-hour forecast, and 0-3 km lapse rate errors for (c) the RAP analysis, plotted in units of °C km^{-1} as a function of distance and azimuth relative to the TC center. Bins in which the errors are not statistically significant at the 95% confidence level are hatched.

The 0-3 km lapse rate errors tend to be negative, i.e. stable-biased, when plotted relative to the TC center, but are smaller in magnitude than the 0-1 km lapse rate errors. The plot of RAP analysis 0-3 km lapse rate errors (Fig. 4.17c) has the largest swath of significant lapse rate errors of any of the models for either layer; errors are significantly negative in a semicircular arc spanning 270 to 90 meteorological degrees and extending 400-800 km away from the center. The overarching stable biases seen throughout this section in the 0-1 and 0-3 km lapse rate errors agree with the mean temperature error profiles in Figure 4.1a, thereby supporting Hypothesis 4. Surface temperatures in both forecasts and the analysis have a cool bias, which acts to stabilize lapse rates originating at the surface.

Kinematic parameters are used by forecasters along with information about instability to diagnose the likelihood of any thunderstorms that form to rotate and potentially become tornadic. Vertical wind shear and storm-relative helicity (SRH) are two prominently used kinematic parameters, and the RAP and HRRR errors in these quantities within the TC envelope are discussed here.

Vertical wind shear (in this case, "shear" refers to the bulk wind difference) in the layers 0-1 km, 0-3 km, 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. Hypothesis 4 is supported by the shear results because these underpredictions align with the shear errors expected to result from the wind error profiles (Fig. 4.3a) and wind error heat maps (not shown) for each of the forecasts and the analysis, which show overpredicted winds at the surface and

underpredicted winds aloft. Shear in the layer from 0-1 km is underpredicted by 1.5-3 knots for both the forecasts and the RAP analysis before TC landfall (Fig. 4.18a). The magnitude of the 0-1 km shear error decreases with time during the first three days after landfall. This is physically consistent with the behavior of the 0-1 km shear errors as a function of intensity (Fig. 4.18b); the largest underpredictions occur at higher intensities and the magnitude of the underpredictions decreases as the TC weakens. Both Hypothesis 2 and Hypothesis 1 are supported by these results.

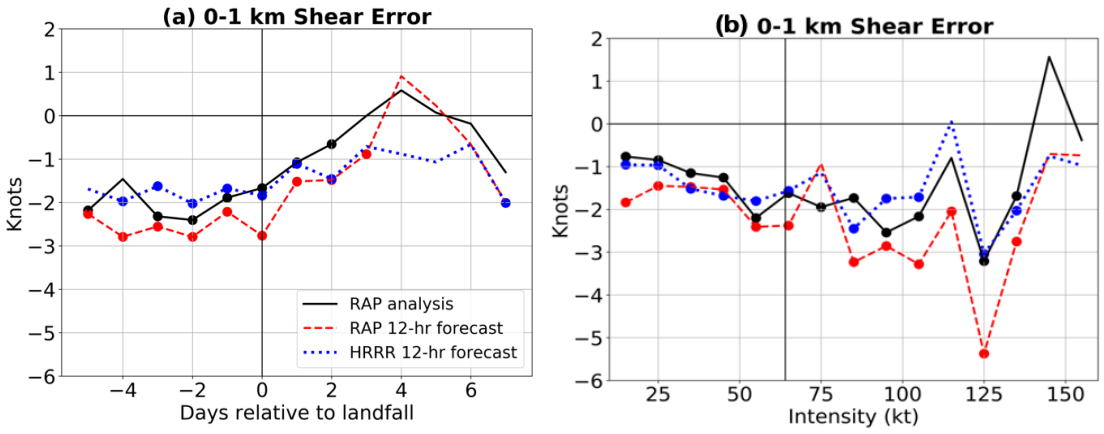

Figure 4.18: 0-1 km shear errors, measured in knots and plotted as a function of (a) time relative to landfall in days and (b) TC intensity in knots for the RAP analysis, RAP 12-hour forecast, and HRRR 12-hour forecast. Solid dots are placed along the line at the midpoint of each bin in which the errors are significant at the 95% confidence level.

The RAP analysis significantly underpredicts 0-1 km shear in all but two range-azimuth spatial bins between 200 and 800 km from the TC center (Fig. 4.19a). Within the same range ring in the forecasts, significant underprediction of 0-1 km shear is confined to smaller continuous swaths: across the entire northeast half of the TC from 315 to 135 degrees in the RAP 12-hour forecast (Fig. 4.19b) and within a smaller sector

from 315 to 90 degrees in the HRRR 12-hour forecast (Fig. 4.19c). Underpredictions of similar magnitude are seen in the 0-3 km shear errors when examined relative to the TC center. Most notably, the RAP 12-hour forecast significantly underpredicts 0-3 km shear throughout the eastern half of the TC outside of 200 km from the TC center (Fig. 4.19d). In the layer from 0-6 km, the RAP 12-hour forecast continues to significantly underpredict shear in the majority of the eastern half of the TC. The HRRR 12-hour forecast 0-6 km shear errors across most of the spatial domain are also negative, but insignificant.

McCaul (1991) showed that the strongest shear and largest SRH values tend to occur in the motion-relative 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 shear errors (Fig. 4.19), 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). The regions of significant SRH errors on similar north-relative polar plots appear small and discontinuous when compared to the regions of significant shear errors. The RAP analysis and HRRR 12-hour forecast have significant 0-1 km SRH errors in the northeast quadrant at a range of 400 to 600 km, while the same significant errors in the RAP 12-hour forecast are generally confined to a small sector from 0 to 45 degrees between 200 and 600 km from the TC center (not shown). Within these regions, 0-1 km SRH is underpredicted by about 20 $m^2 s^{-1}$. The 0-3 km SRH errors are widely insignificant as a function of location

relative to the TC center for the RAP 12-hour forecast (not shown), and across both forecasts and the analysis when plotted as a function of time relative to landfall, TC intensity, and local time of day.

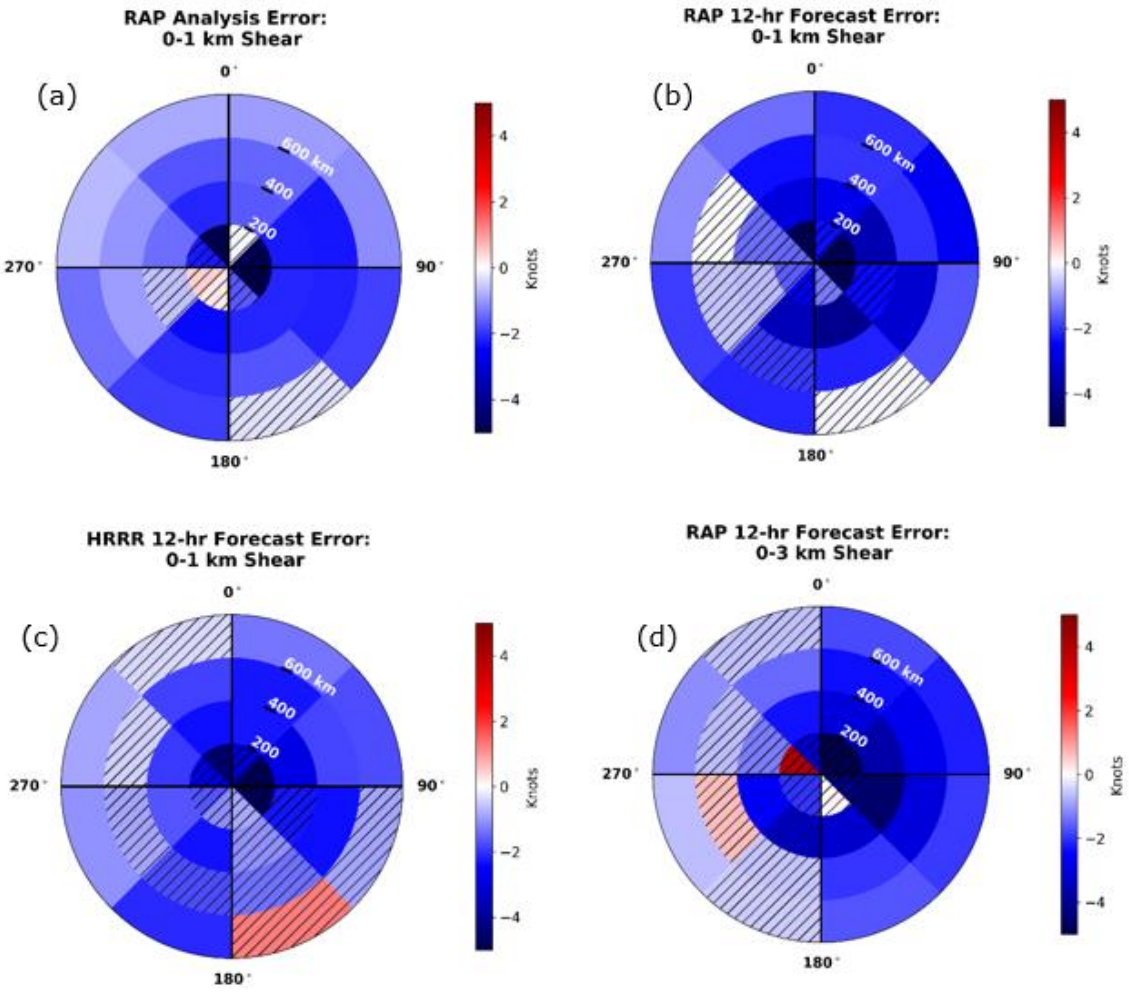

Figure 4.19: 0-1 km shear errors in knots for (a) the RAP analysis, (b) the RAP 12-hour forecast, and (c) the HRRR 12-hour forecast, and 0-3 km shear errors for (d) the RAP 12-hour forecast, plotted as a function of distance and azimuth relative to the TC center. Bins in which the errors are not statistically significant at the 95% confidence level are hatched.

When 0-1 km SRH errors are plotted as a function of time relative to landfall, statistically significant underpredictions emerge in the two days on either side of the landfall time (Fig. 4.20a). However, the magnitude of these errors does not lessen as time after landfall increases, so Hypothesis 2 is not supported by these results. The 0-1 km SRH underpredictions also tend to be significant in the RAP analysis from the early morning through the mid-afternoon hours when examined as a function of local time of day (Fig. 4.20b). Overall, 0-1 km SRH errors are more frequently significant across the model runs and variables of interest than the 0-3 km SRH errors.

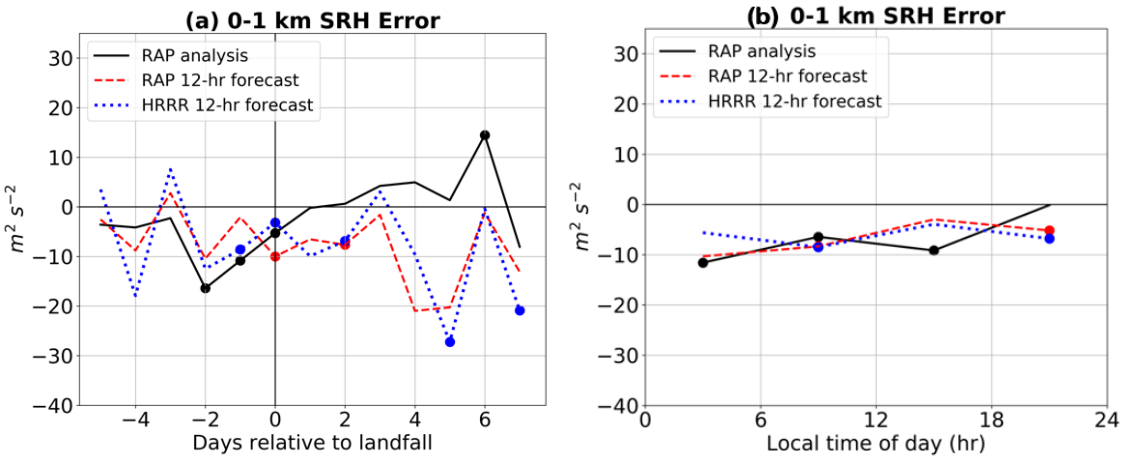

Figure 4.20: 0-1 km SRH errors in $m^2\ s^{-2}$, plotted as functions of (a) time relative to landfall in days and (b) local time of day. Solid dots are placed along the line at the midpoint of each bin in which the errors are significant at the 95% confidence level.

Values of selected thermodynamic and kinematic variables are used to calculate the significant tornado parameter (STP), which was shown by Nowotarski et al. (2021) to discriminate between tornadic and non-tornadic cells within Hurricane Harvey. Their study utilized 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 shear. The CINH-scaled STP depends on the mixed-layer thermodynamic quantities MLLCL, MLCAPE, and MLCINH, in addition to values of effective-layer SRH and effective layer shear. These are the SRH and shear as calculated for a unique effective inflow layer, or the layer from which a convective storm sources its updraft. Errors in both the fixed-layer STP and the CINH-scaled STP are examined here.

When fixed-layer STP errors are plotted relative to the TC center, few bins display significant errors. Fixed-layer STP is significantly underpredicted in a small sector to the east-northeast of the TC center by the RAP 12-hour forecast (Fig. 4.21a). 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-hour forecast through a narrow ring from 45-180 degrees and 200-400 km (Fig. 4.21b), and by the RAP analysis in a small sector just south of east from 200-600 km (Fig. 4.21c). These results act in accordance with the overlapping negative errors already seen in SBCAPE, 0-6 km shear, and 0-1 km SRH on the east side of the TC center. In contrast, CINH-scaled STP errors are largely insignificant when plotted as a function of position relative to the TC center (not shown). Errors in CINH-scaled STP are also widely insignificant across the full range of intensities and times relative to landfall. A lack of consistent positive or negative errors in CINH-scaled STP could be due to counteracting effects of some of the variables used to calculate it: MLCAPE which is typically overpredicted, and effective layer SRH and shear whose errors behave erratically as a function of different TC-related variables.

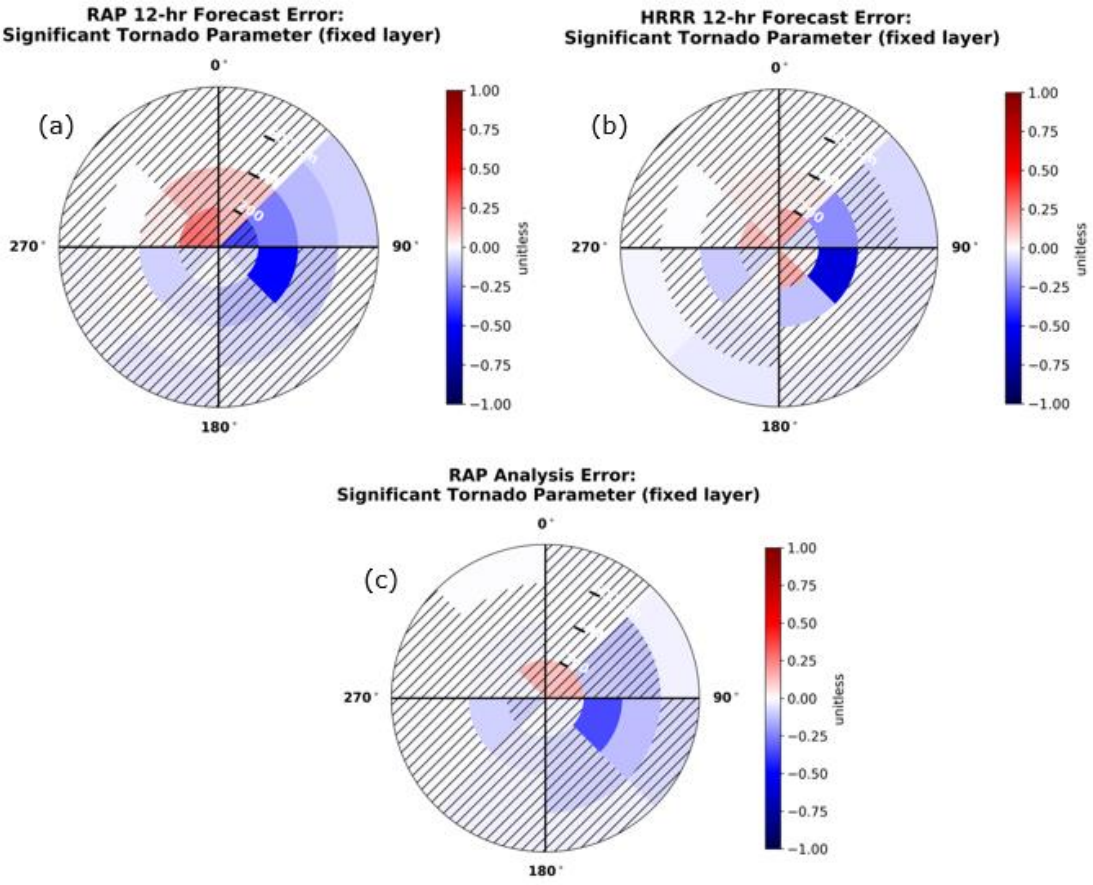

Figure 4.21: Fixed-layer STP errors for (a) the RAP 12-hour forecast, (b) the HRRR 12-hour forecast, and (c) the RAP analysis, plotted as a function of distance and azimuth relative to the TC center. Bins in which the errors are not statistically significant at the 95% confidence level are hatched.

Fixed-layer STP is significantly underpredicted both at low TC intensities up to hurricane strength and from the landfall time up to two days after landfall (Fig. 4.22a, b). The results in Figure 4.22a support Hypothesis 1 in that errors are smaller in magnitude at the lower intensities. With respect to local time of day, CINH-scaled STP is overpredicted in the local midafternoon, and slightly less so in the evening, by both forecasts and the RAP analysis (Fig. 4.22c). These positive errors are only significant in

the RAP 12-hour forecast and RAP analysis for the midafternoon, and the RAP 12-hour forecast for the evening period, but they should be noted as it is not common for STP to be overpredicted in any model run or as a function of any variable.

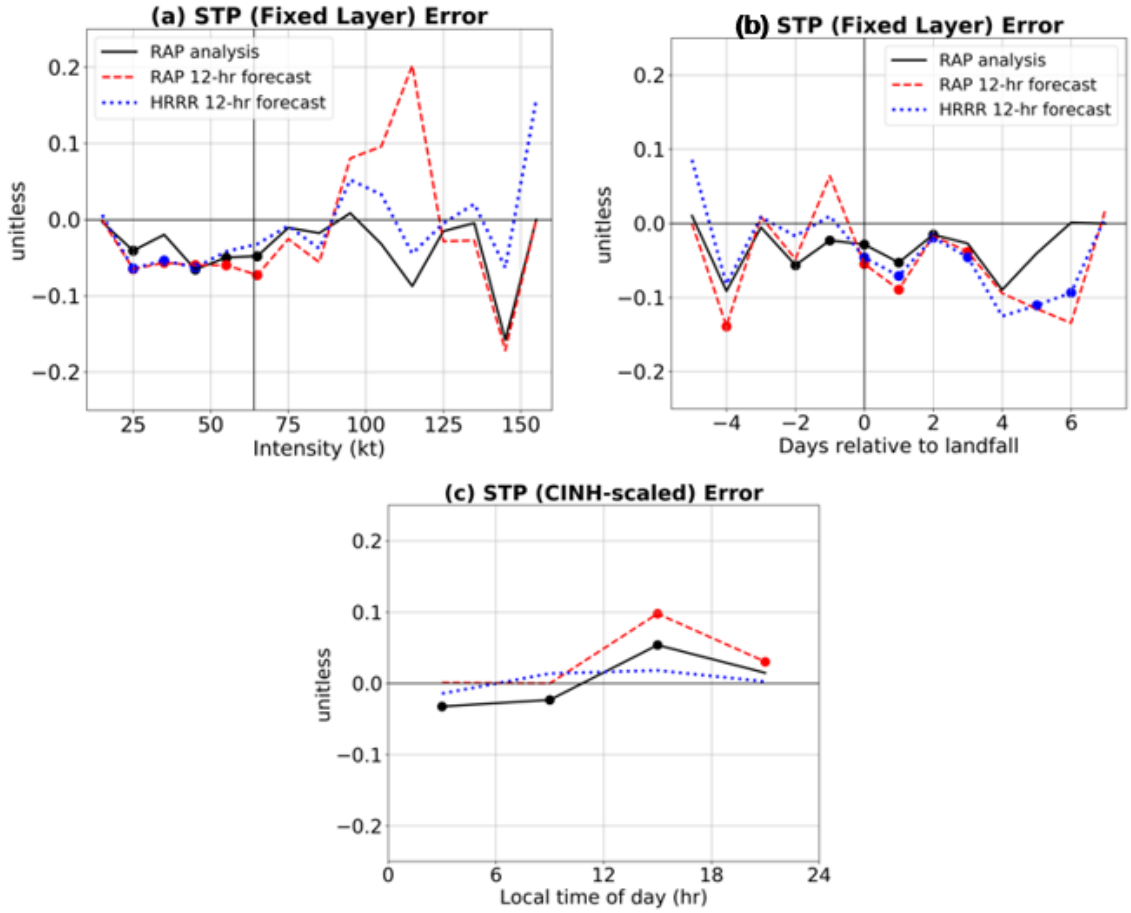

Figure 4.22: Fixed-layer STP plotted as a function of (a) TC intensity and (b) time relative to landfall, and (c) CINH-scaled STP plotted as a function of local time of day. Solid dots are placed along the line at the midpoint of each bin in which the errors are significant at the 95% confidence level.

5. CONCLUSIONS

This study explores model errors in the RAP analysis and RAP and HRRR 12-hour forecasts within the TC envelope up to 800 km from the center. Up to 1,730 sounding pairs are used in the analysis, sourced from 13 TCs that made landfall along the coast of the contiguous U.S. during the 2017, 2018, and 2019 Atlantic hurricane seasons. First, mean error profiles are created to investigate the representation of the raw sounding variables temperature, dew point, and wind speed in the models throughout the depth of the troposphere. Then, an understanding is developed of how these model errors vary with height as functions of variables related to TCs such as time relative to landfall, intensity, and distance from the center. Finally, model errors in sounding-derived parameters are explored, paying particular attention to variables previously 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.

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. The temperature errors from 0-5 km in the RAP and HRRR 12-hour forecasts are not consistent with the results of Evans et al. (2018), who found persistent warm biases within that layer in continental environments. Dew point errors have larger magnitudes and standard deviations than the temperature errors, and there is a pronounced moist bias that worsens with height for both of 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 0-5 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.

The four hypotheses tested in this study are listed in Chapter 2. In paraphrased form, they are (1) Model performance will improve as the environment becomes more characteristic of a typical continental environment (i.e. as intensity decreases and distance from the TC center increases), (2) Model performance will improve as the TC moves further inland, (3) Forecast errors will exceed analysis errors, and (4) Sounding-derived parameter errors will reflect errors in the raw sounding variables.

Hypothesis 3 is supported by the mean error profiles described above; absolute errors of temperature, dew point, and wind speed in the 12-hour forecasts exceed those in the RAP analysis throughout the profile.

Hypothesis 1 is tested by examining plots of the errors as functions of intensity and distance from the TC center. As expected, temperature errors tend to be smaller at lower TC intensities, and at their largest within 200 km of the TC center. However, only the HRRR 12-hour forecast shows any marked decrease in temperature error magnitudes as distance from the TC center increases outside of 200 km, and dew point errors are relatively uniform at all radial distances from the center. Thus, Hypothesis 1 is only partially supported by these results.

Hypothesis 2 is contradicted by temperature and dew point errors that do not diminish as time after landfall increases. SBCAPE and MUCAPE errors show a similar failure to decrease in magnitude after landfall. However, MLCAPE errors support Hypothesis 2.

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 support Hypothesis 4 because they 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 0-1 km than those calculated for the 0-3 km layer.

The 0-1, 0-3, and 0-6 km shear values are generally underpredicted for all model runs examined. The underpredictions of 0-1 km shear decrease after landfall and at weaker intensities, supporting Hypotheses 1 and 2, while the underpredictions of 0-3 and 0-6 km shear remain more steady as a function of time relative to landfall and intensity. Likewise, 0-1 and 0-3 SRH tend to be underpredicted, although with fewer significant results across the model runs and TC-related variables. Fixed-layer STP is largely underpredicted, while trends are less clear in the CINH-scaled STP.

The frequent underprediction of measures of instability and shear by the RAP and HRRR 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. It is worth pointing out several variables that are significantly over- or underpredicted within sectors of the TC that 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-hour forecast. The 0-3 km shear values are significantly underpredicted by 2-5 kt in the RAP 12-hour forecast throughout the eastern half of the TC, outside of 200 km from the center. The 0-6 km shear is also significantly underpredicted in the RAP 12-hour forecast, by the same magnitude within the majority of the eastern half of the TC.

Additional forecasting insights could be gleaned by investigating the errors in the raw variables and sounding-derived parameters at longer model lead times, to get a sense of how the errors tend to evolve leading up to the analysis time. 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. Also, the dataset could be expanded to include TCs from the most recent Atlantic hurricane seasons in 2020 and 2021. Any future work will be completed with the same strong focus on operational utility and supporting the effort to improve the TCTOR forecasting and warning process.

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://github.com/sharppy/SHARPpy

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%3C0411:MDIAAF%3E2.0.CO;2

Davies, J. M., 2006: Hurricane and tropical cyclone tornado environments from RUC proximity soundings. *Preprints, 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. Met. Soc., P3.1, https://ams.confex.com/ams/pdfpapers/175269.pdf

Edwards, R., 2012: Tropical cyclone tornadoes: A review of knowledge in research and prediction. *Electronic J. Severe Storms Meteor.*, **7**, 1-61, [Available online at: https://www.spc.noaa.gov/publications/edwards/ejssmtct.pdf]

Edwards, R., 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. [Available online at: https://journals.ametsoc.org/downloadpdf/journals/mwre/111/9/1520-0493_1983_111_1793_gotawh_2_0_co_2.pdf]

Gropp, M. E., and C. E. Davenport, 2018: The impact of the nocturnal transition on the lifetime and evolution of supercell thunderstorms in the Great Plains. *Wea. Forecasting*, **33**, 1045-1061, https://doi.org/10.1175/WAF-D-17-0150.1

Laflin, J. M., 2013: Verification of RAP model soundings in preconvective environments. *J. Operational 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%3C1954:BASCOH%3E2.0.CO;2

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

NOAA National Centers for Environmental Information, 2020: Rapid Refresh (RAP). Accessed 27 May 2020, https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/rapid-refresh-rap

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

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

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%3C0479:AWAOTC%3E2.0.CO;2

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

University of Wyoming, 2020: Upper air soundings. 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