Carroll-Smith, D. L., L. C. Dawson, and R. J. Trapp, 2019: High-resolution real-data WRF modeling and verification of tropical cyclone tornadoes associated with Hurricane Ivan (2004). *Electronic J. Severe Storms Meteor.*, **14** (2), 1–36.

Electronic Journal of
SEVERE STORMS METEOROLOGY

High-Resolution Real-Data WRF Modeling and Verification of Tropical Cyclone Tornadoes Associated with Hurricane Ivan (2004)

DEREKA L. CARROLL-SMITH
Jackson State University, Jackson, Mississippi and
National Center for Atmospheric Research, Boulder, Colorado

LOGAN C. DAWSON
I.M. Systems Group, Inc. and
NOAA/NWS/NCEP/Environmental Modeling Center, College Park, Maryland

ROBERT J. TRAPP
University of Illinois at Urbana-Champaign, Urbana, Illinois

(Submitted 3 April 2018; in final form 17 April 2019)

ABSTRACT

This study used high-resolution real-data WRF simulations of Hurricane Ivan (2004) to document the structure of potentially tornadic supercells embedded in tropical cyclone (TC) rainbands. The simulated TC track and intensity matched well with observations post landfall, and the simulated TC structure closely replicated the observed shape and rainbands, as evident by an assessment of observed composite and simulated radar reflectivity. TC tornado surrogates (TCT-Ss) were identified and calibrated using thresholds based on percentile values of maximum updraft helicity (UH_{max}) and simulated radar reflectivity. Although the magnitude of UH_{max} generally decreased as the simulated TC moved inland, sensitivity testing revealed that a threshold based on the 99.95% percentile value of UH_{max} achieved optimal TCT-S coverage and agreement with observed TC tornado (TCT) tracks on domains with 3-km and 1-km grid spacings. Three cells were identified at three different stages of the TC inland evolution and were found to have hook appendages resembling supercells found in midlatitudes. The cells produced storm tops of 13 km, with rotating cores with depths of 4 km above ground level. Although results in this study are unique to Hurricane Ivan, the authors believe the use of convection-allowing models and UH-based surrogate methodologies can be applied to other TC cases.

1. Introduction

In addition to coastal storm surge, tropical cyclones (TCs) produce strong winds, tornadoes, and flooding rains as they move inland (Czajkowski and Kennedy 2010). Hazards such as these can lead to loss of life and property within inland communities (Rappaport 2000; 2014). They also can complicate coastal evacuations, particularly if the coastal evacuees flee to inland locations affected by one or more of the aforementioned hazards.

Indeed, successful predictions of TC impact must exhibit skill in the TC track and intensity after landfall, and contemporaneously provide accurate and timely guidance on smaller-scale convective storms as the TC evolves inland. Implied here is the need for convection-allowing numerical prediction models (CAMs) (e.g., Kain et al. 2008, 2010), which have been proven to be valuable for predictions of convective hazards associated with midlatitude weather systems. Although research exists using CAMs to predict tornadoes developing in environments similar to

Corresponding author address: Dereka Carroll-Smith, 1400 J. R. Lynch, Jackson, MS 39217, E-mail: dcarroll@ucar.edu

those spawned by tropical cyclones (Cohen et al. 2015; Cohen et al. 2017), such models have not been proven useful for convective-hazard predictions within landfalling TCs, to our knowledge. Such is the basic interest of this research, with a particular focus placed herein on tropical-cyclone tornadoes (TCTs).

A comprehensive review of TCT knowledge can be found in Edwards (2012), so only particularly relevant background information is included here. For example, although it is uncommon to have more than a few tornadoes per TC on average, there are cases in which over one hundred tornadoes were spawned. Such cases include Hurricane Beulah in 1967 (115 TCTs), Hurricane Frances in 2004 (103 TCTs), and Hurricane Ivan in 2004 (118 TCTs) (Edwards 2012). The majority of these TCTs were rated at the lower end of the Fujita and Enhanced Fujita scales (Edwards et al. 2013). Stronger TCTs have been documented, however, such as in association with Hurricane Ivan 2004, which produced 18 F2 tornadoes and one F3 tornado (Edwards 2010).

TCTs typically develop from "miniature supercells," which are smaller and shallower than those found within the U.S. Great Plains, (McCaul 1991; McCaul and Weisman 1996; Baker et al. 2009; McCaul et al. 2004). For example, Eastin and Link (2009) showed that the dimensions of these miniature supercells have an average mesocyclone diameter of about 5–7 km and an average depth of 4 km. These dimensions can make TCT detection difficult with operational Doppler radar and associated automated algorithms, compared to the Great Plains supercells that have a mesocyclone diameter of ≈10 km and depth of 9 km (Eastin and Link 2009). Edwards et al. (2012) also noted that while 88% of TCTs form out of miniature supercells (in the modes of discrete cells, cell in lines, and in clusters), the remaining 12% form out of nonsupercellular convective modes, such as quasi-linear convective systems and clusters.

Tornadic miniature supercells are commonly found in the outer TC rainbands, 200–400 km from the TC center (Schultz and Cecil 2009), although some tornadoes are known to have formed within the eyewall and inner core (McCaul 1991). The northeast quadrant of the TC is favored for tornado development due to ample CAPE and vertical wind shear (McCaul 1991; Verbout et al. 2007). McCaul (1991) found that helicity and bulk shear are maximized in the northeast quadrant. That sort of parameter space, in addition to moderate CAPE values, contributed to Ivan's three-day, center-relative tornado distribution plotted in Fig. 1 (Baker et al. 2009). TCs that make landfall along the U.S. Gulf Coast are more likely to have a higher number of reported tornadoes compared to TCs that make landfall along the U.S. Atlantic Coast. This is because the Gulf Coast is exposed to the northeast quadrant of the TC longer than is the Atlantic Coast when TCs make landfall along the respective coastlines (Verbout et al. 2007).

The curvature of the TC track is also known to affect tornado development. Using synoptic composites of 83 TCs, Verbout et al. (2007) showed that midlatitude troughs provide additional deep-layer and low-layer vertical wind shear, which favor mesocyclogenesis and tornadogenesis, respectively, when a TC recurves. A diurnal signal associated with the number and strength of TCTs also exists. More inland TCTs have been found to occur during daytime hours since the TC supercell environment experiences a larger cycle of sensible heat flux and thus higher values of surface-based CAPE during the daytime (Curtis 2004; Shultz and Cecil 2009).

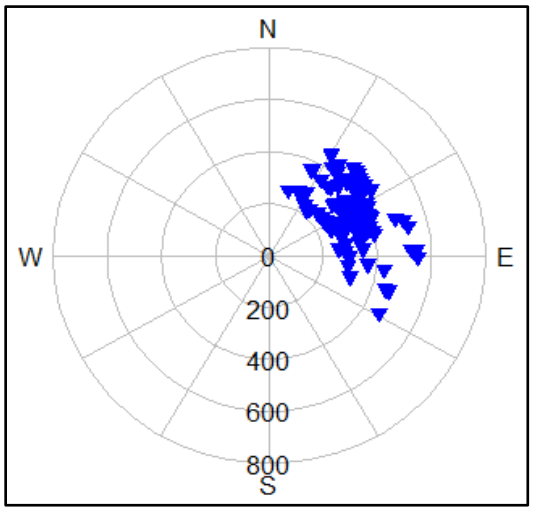


Figure 1: North-relative polar plot of tornadoes (blue inverted triangles) spawned by Hurricane Ivan over a three-day period, with respect to azimuth (plotted every 30°) and range (radials every 200 km from the storm center). Adapted from Fig. 1b in Edwards (2012).

This diurnal effect on TC-supercell formation has been explored further using quasi-idealized numerical modeling, as have the effects of dry air intrusions and the sea-to-land transition of the TC (Morin and Parker 2011). They found that TC landfall and dry air intrusions are not necessary conditions for supercell development, and that the most optimal ingredients are found offshore, which is consistent with previous literature (Baker et al. 2009; Eastin and Link 2009; Molinari and Vollaro 2008). Dry air intrusions have been shown to provide thermodynamic environments favorable for tornadic activity (Curtis 2004; Edwards 2012), however, and the sea-to-land transition contributes to supercell development and intensification, due to the magnitude of the diurnal temperature changes over land versus over water. Finally, Morin and Parker (2011) found that the time of day of TC landfall impacts the strength, number and lifetime of the idealized supercells. For example, the experiments with nighttime landfall produced fewer and shorter-duration mesocyclones compared to experiments with daytime landfall (consistent with Curtis 2004 and Shultz and Cecil 2009).

The quasi-idealized approach of Morin and Parker (2011) was used to isolate the impacts of the environment on simulated supercells; this method was considered "quasi" because it allowed the simulated supercells to form more naturally than in typical idealized studies. Our work presented here differs in that it seeks to model an actual event using initial and boundary conditions supplied by observations and reanalysis data. Indeed, Morin and Parker as well as Green et al. (2011) provide motivation to explore the use of high-resolution numerical weather prediction models to generate forecast guidance on TCT formation.

With horizontal grid spacings on the order of 1–4 km, CAMs are capable of nominally resolving convective-scale phenomena such as the mesocyclones that characterize supercell thunderstorms. Grid-scale updraft helicity (UH) is now widely used to identify potential (midlevel) mesocyclones in CAM forecasts (e.g., Kain et al. 2010; Carley et al. 2011; Clark et al. 2012, 2013; Sobash et al. 2011, 2016; Schwartz et al. 2015a,b; Dawson et al. 2017). Additionally, low-level vertical vorticity can diagnose near-surface rotation that may indicate tornado-bearing low-level mesocyclones (e.g., Skinner et al. 2016). CAMs are now heavily used to aid in the forecasting process when tornadoes and other severe weather phenomena threaten.

This work uses the Weather Research and Forecasting model (WRF; Skamarock et al. 2008). We describe the WRF configuration and a TCT "surrogate" methodology in section 2. In section 3, we demonstrate the ability of such surrogates to reveal the tornado occurrences in association with Hurricane Ivan (2004), and then further demonstrate the capability of WRF with 1-km and even 3-km grid spacings to represent the bulk structure of TCT-generating supercells. The success of WRF in this specific case of Hurricane Ivan leads us to conclude in section 4 that CAMs configured for the TC environment indeed do appear to hold promise for TCT forecasting.

2. Methodology

Hurricane Ivan, which was responsible for 25 deaths in the U.S. and $18.8 billion (2004 USD) in damages, served as the case study for this research. Ivan (2004) made landfall at 07 UTC on 16 September 2004 in Gulf Shores, AL, as a Category 3 (Simpson 1974) hurricane. Ivan produced large rainfall amounts, which contributed to vast inland flooding, and also spawned 118 tornadoes (see Table 1), per the TCTOR dataset (Edwards 2010). The tornadoes were concentrated during three periods: 36 TCTs during the landfall (LF) period (18 UTC 15 September 2004 to 18 UTC 16 September 2004); 23 TCTs during the mid-landfall (ML) period (18 UTC 16 September 2004 to 18 UTC 17 September 2004); and 59 TCTs during the extratropical transition (ET) (18 UTC 17 September 2004 to 18 UTC 18 September 2004). These periods guided some aspects of the model configuration, as described next.

Table 1: Observed TCT counts by date.

Day/Time (UTC)	TCT counts
15/18–16/18	36
16/18–17/18	23
17/18–18/18	59
Total	118

a. Model configuration and data

Ivan's landfall and subsequent evolution were simulated over the interval 00 UTC 14

September 2004 to 00 UTC 19 September 2004 using version 3.7 of the Advanced Research (ARW) core of the WRF model (WRF-ARW; Skamarock et al. 2008). The computational area consisted of an outer domain with 9-km horizontal grid spacing, and two, two-way nested domains with 3-km and 1-km horizontal grid spacing (hereinafter referred to as the 9-km, 3-km, 1-km domains, respectively) (Fig. 2). The center of the 1-km innermost domain depended on the three different periods (LF, ML, ET) in Ivan's inland progression, and thus allowed for an assessment of differences in supercell structure at each stage. For the 9-km and 3-km domains, WRF was integrated over the entire five-day interval; for the 1-km domain, WRF was integrated over one-day intervals only, beginning at 18 UTC 15 September 2004, 18 UTC 16 September 2004 and 18 UTC 17 September 2004, respectively, to correspond to the LF, ML and ET periods (Fig. 2).

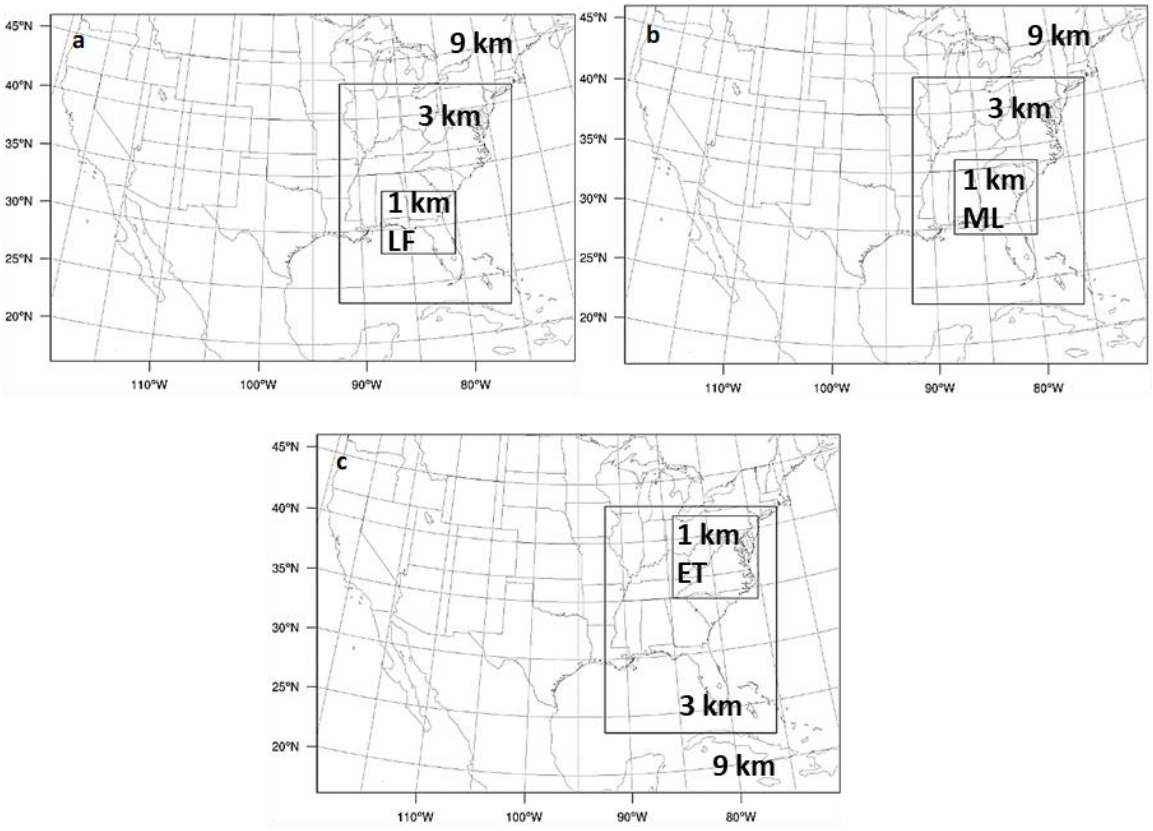


Figure 2: Configuration of WRF computation domain: 9 km parent domain, nested 3 km domain, and three separate nested 1-km domains for the a) LF, b) ML, and c) ET periods (see text). *Click image to enlarge.*

Data for initial and boundary conditions are from the 6-hourly National Center for Environmental Prediction Final (NCEP FNL) Operational Global Analysis, which is on a 1°×1° grid, at the surface and at 26 mandatory levels from 1000 to 10 hPa (NCAR 2000). The NCEP FNL data are the same as that of the Global Forecast System (GFS); however, the FNL data are delayed in order to incorporate more observations.

Based on previous tropical cyclone studies (e.g. Gentry and Lackmann 2010; Sun and Barros 2012; Sun and Barros 2014; Lackmann 2015) and sensitivity tests by Carroll-Smith (2018), the WRF model configuration was set as follows. The Kain-Fritsch (Kain 2004) scheme was used to parameterize cumulus convection on the outer domain only; convective processes were represented explicitly on the 3-km and 1-km nests. On all three grids: the updated Rapid Radiative Transfer Model (RRTMG; Iacono et al. 2008) scheme parameterized longwave and

shortwave radiation; the Eta surface layer scheme and Noah land surface model (Chen and Dudhia 2001) were used for surface layer interactions and surface physics, respectively; the Mellor-Yamada-Janjić (MYJ; Janjić 1994) scheme parameterized planetary boundary layer processes; and the Thompson scheme (Thompson et al. 2008) was used to parameterize the cloud and precipitation physics. To improve upon the intensity forecasts, sea surface temperatures were updated every 6 h, and the "isftcflx" is activated using the Garratt formulation, option 2 (Lackmann 2015). This physics parameter supplies different sensible and latent heat values, correcting the surface bulk drag and enthalpy coefficients to provide more realistic TC intensities and TC inner core structure (Green and Zhang 2013; Parker et al. 2017 and references therein offer details on "isftcflx").

Note that the sensitivity experiments of Carroll-Smith (2018) tested the Yonsei University (YSU; Hong et al. 2006) planetary boundary layer (PBL) scheme and the WRF single moment 6-class microphysics scheme (WSM6; Hong and Lim 2006). Results showed that TCT surrogate (section 2b) generation was most sensitive to the choice of PBL scheme, even though the WSM6 experiments had stronger rotating updrafts, owing in part to stronger updraft speeds. Nevertheless, the combination of Thompson microphysics and MYJ PBL led to the best agreement between TCT surrogates and observed TCTs.

a. TCT surrogates

TCT surrogates (TCT-S) were diagnosed on land points only, using a simple exceedance algorithm based on simulated radar reflectivity factor (SRF) and updraft helicity (UH). Updraft helicity is a measure of the intensity of a rotating updraft and is defined as the vertical integral of the product of vertical velocity and vertical vorticity (Kain et al. 2008). This work examined run-time maximum updraft helicity (UH_{max}) computed between 2 and 5 km AGL, at hourly (5-min) frequencies on the 3-km (1-km) domain(s). UH_{max} has been used widely in severe weather applications to identify supercells and associated hazards (e.g., Kain et al. 2010; Carley et al. 2011; Clark et al. 2012, 2013; Sobash et al. 2011, 2016; Dawson et al. 2017).

Green et al. (2011) found UH to be of little value in identifying TC supercells spawned just offshore by Hurricane Katrina (2004), due to the cells being shallow and tilted. Our preliminary findings with simulations of Hurricane Ivan suggested the opposite, thus motivating the work herein. In Dawson et al. (2017), percentiles of the UH_{max} distribution were used as thresholds to identify surrogates. We applied this approach herein in an effort to account for the resolution dependence of UH_{max} and to account for the relative infrequency of tornado-generating convective storms within TCs. Accordingly, the 99.5%, 99.9%, 99.95%, and 99.99% percentile values of the distribution of all non-zero UH_{max} values in the Ivan simulation were used to define and then test four different threshold values (see Table 2). As noted, our TCT-S identification algorithm required exceedance of a UH_{max} threshold as well as a SRF threshold, set here to 30 dB*Z*. Sensitivity tests were conducted to determine the ideal layer of UH to document potentially tornadic storms. UH calculations over a 1–4-km layer yielded only minor differences in TCT-S occurrence; UH calculations over a 0–3-km layer resulted in a significant over-representation in TCT-S occurrence as indicated quantitatively by the percentage of domain covered (detailed in section 3b). Thus, we have chosen UH calculations over a 2–5-km layer.

Percentile values of UH_{max} were assessed on the 3-km domain over the entirety of the simulation as well as over each of the three periods (LF, ML, and ET; not shown) (Table 2). Because of the possibility that TCT-generating storms would be inadequately represented with 3-km grid spacings, we also assessed percentile values of UH_{max} on the 1-km domains over each period (LF, ML, and ET), from which a composite value was determined (1-km_COMP; Table 2). Not surprisingly, higher percentiles consistently yielded higher threshold values, and the use of higher threshold values in the TCT-S algorithm decreased the number of TCT-Ss on both the 3-km and 1-km domains, for all periods (Table 3). A discussion of the rationale for choosing one threshold over another is provided in section 3.

Table 2: Percentile values of UH_{max} (m^2 s^{-2}) for the 1-km and 3-km domains. 1-km_COMP and 1-km_IND represent the composite and individual thresholds, respectively.

	Period	99.50%	99.90%	99.95%	99.99%	Max
3-km	LF,ML,ET	32.01	65.71	82.66	146.75	356.05
1-km_COMP	LF,ML,ET	158.34	377.66	512.42	933.20	2129.07
1-km_IND	LF	265.39	603.35	785.34	1248.19	2129.07
	ML	133.67	311.70	416.77	828.26	2126.08
	ET	114.92	250.07	320.41	483.44	1110.96

Table 3: Numbers of TCT-Ss for the 3-km and 1-km domains for each threshold. The 1-km_COMP threshold was applied to the LF, ML and ET periods.

	PERIOD	99.50%	99.90%	99.95%	99.99%
3-KM	LF,ML,ET	8414	1875	928	190
1-KM_COMP	LF	22079	6632	3793	841
	ML	11740	1992	939	202
	ET	9891	1053	352	16
1-KM_IND	LF	11334	2635	1381	297
	ML	15380	3116	1547	262
	ET	16775	3640	1849	437

3. Results

a. Simulation overview

The WRF-simulated Ivan made landfall near Dauphin Island (Mobile County), AL, about 48 km west of the observed landfall location of Orange Beach (Baldwin County), AL (Fig 3a). Thereafter, the simulated inland track agrees well with the inland track from the IBTrAC data (Knapp et al. 2010). Osuri et al. (2013) noted TC track biases in WRF-ARW occur despite higher resolutions and model initialization closest to when TCs are at the "severe cyclone" stage. In the case of WRF-simulated Ivan, the westward track bias could be due to the initialization time, in addition to the tracking algorithm used (tracked the lowest 6 hourly mean sea level pressure within the model boundary), and lack of data assimilation methods that have been found to improve TC track forecasts (e.g. Rappaport 2009; Cavallo 2013). The IBTrAC data also show that the simulated storm, which had a Category 2 intensity at landfall, was 16 hPa weaker than that observed, but soon converged upon the observed intensity. This low intensity bias over the ocean has been seen in other WRF simulations of landfalling hurricanes (Gentry and Lackmann 2010; Lackmann 2015). Coupled ocean-atmosphere models have been used to account for this discrepancy; however, we are more focused on the inland tornado hazards, so the present methods are sufficient. Also noted are the weaker simulated winds at landfall compared to observations, although the simulated TC sustained stronger winds than observed towards the end of the simulation; we attribute this to the much higher resolution of wind information in the model than in the observations over land.

SRF within the LF, ML, and ET periods was compared to composite reflectivity observations during these periods (Fig. 4). As would be expected in a single deterministic model solution, there are differences in the details of the reflectivity structures, such as in feeder-band placements. However, there is good qualitative agreement between SRF and observed composite reflectivity, especially in terms of Ivan's evolving shape, size, and rainband structure. The extensive outer rainband during the ML and ET periods is of particular importance because it hosted numerous rotating updraft cores (section 3b). The outer rainband also generated

heavy accumulated rainfall from 00 UTC 14 September to 00 UTC 19 September (Fig. 5). Comparison of simulated and observed rainfall also shows agreement in the overall rain patterns, especially in the general locations with the heaviest rain totals: the Florida Panhandle and southern Alabama, along the Tennessee–North Carolina border, and southwestern Ohio and Pennsylvania (Fig. 5).

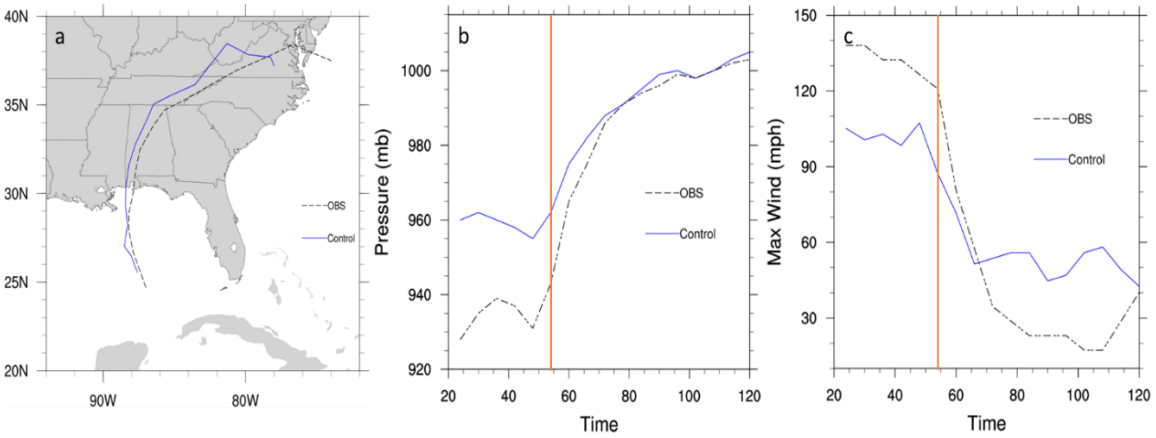


Figure 3: a) Observed (black dashed), simulated (blue) TC track; b) observed (black dashed) and simulated (blue) TC intensity, defined by minimum MSL pressure at 6-h intervals beginning at the 24th forecast hour; c) observed (black dashed) and simulated maximum wind. Orange line represents landfall time. *Click image to enlarge.*

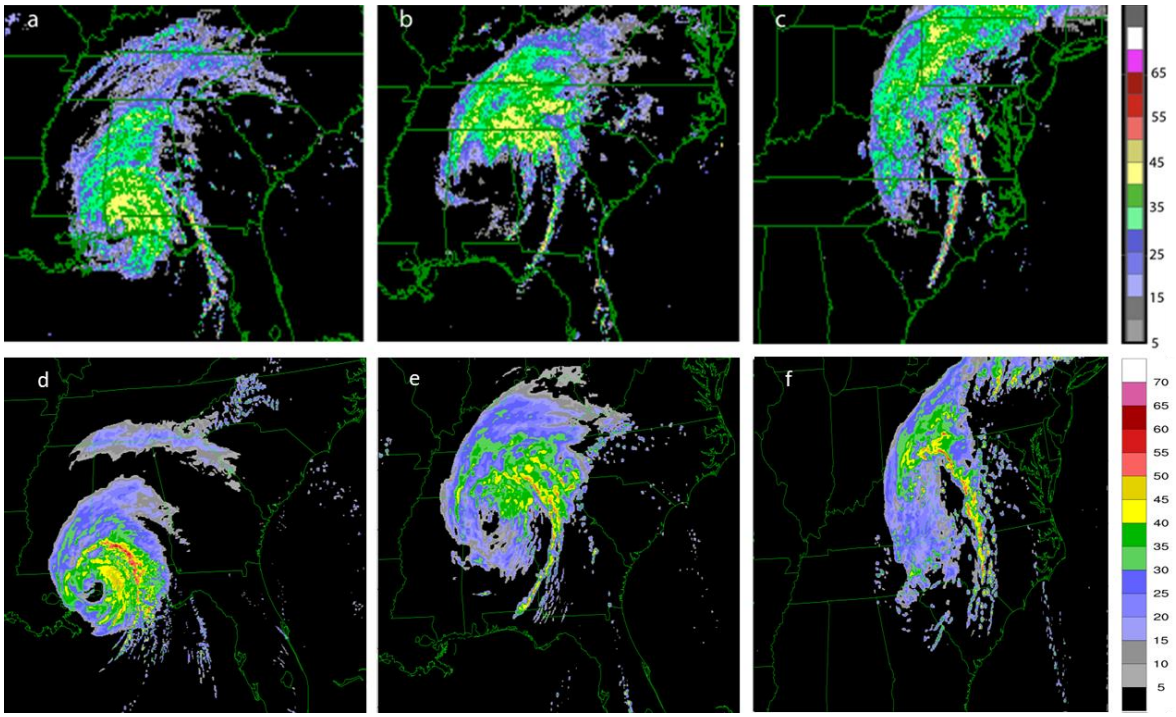


Figure 4: a–c) Observed composite radar reflectivity factor (dBZ) during LF, ML and ET at 06 UTC 16 September, 01 UTC 17 September, and 20 UTC 17 September 2004 respectively, compared to d–f) simulated radar reflectivity factor (dBZ) at 0830 UTC 16 September, 0030 UTC 17 September, and 20 UTC 17 September 2004. *Click image to enlarge.*

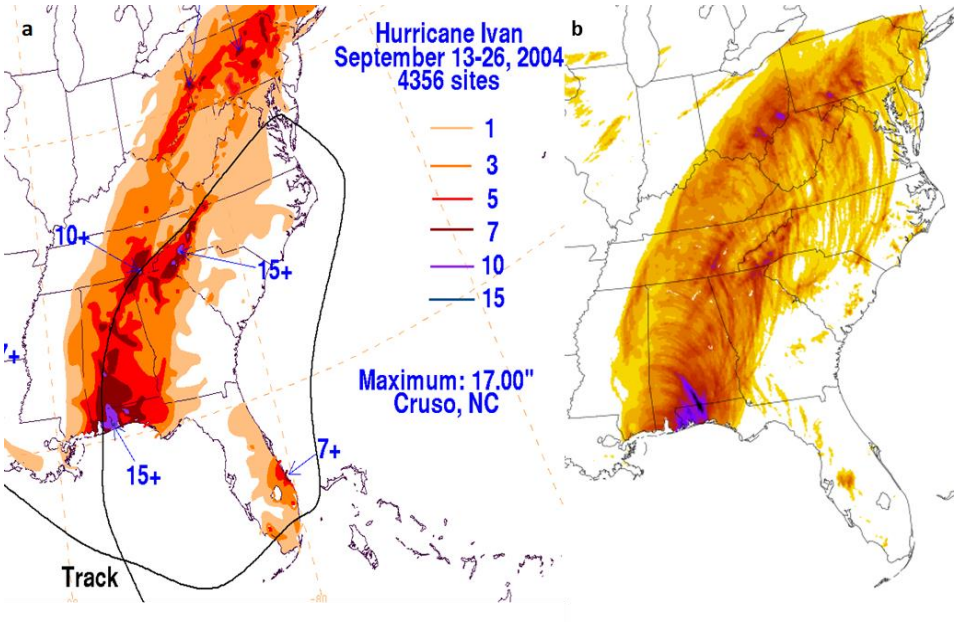


Figure 5: a) Observed and b) simulated accumulated rainfall (in) for Ivan. NOAA Weather Prediction Center (http://www.wpc.ncep.noaa.gov/tropical/rain/ivan2004.html) provided (a). *Click image to enlarge.*

b. TCT surrogates and their sensitivity to UH_{max}

While a demonstrated agreement between simulated and observed TC-scale structure is important, of most relevance is the agreement between TCT-S and observed TCTs. The TCT tracks are from the TCTOR dataset (Edwards 2010). This dataset consists of beginning and ending geo-locations, beginning times, damage and injury information, and other pertinent statistics such as path width and length. Figure 5 shows a comparison of TCT tracks over land and TCT-Ss identified using the 99.5%–99.99% thresholds on the 3-km domain (Table 3). Although there are errors in Ivan's simulated track during the ET period, the spatial distribution of the TCT-Ss is similar to that of the observed TCTs. The majority of the TCTs are located east of the TC track, relative to storm motion, until the ET period when both TCT-Ss and observed TCTs are found to the east and west of the TC track. There are also TCT-Ss in locations where no tornadoes were reported, which we address below. Visually, the 99.95% threshold produced the desired result of sufficient TCT-S coverage without over-prediction (as seen when using the 99.5% threshold; Fig. 6b).

To quantitatively compare the TCT-Ss to the observed TCTs, grid points were assigned to observed TCT tracks using a distance-based selection method in ArcGIS version 10.6.1. Specifically, all grid points in the 3-km domain that were within 3 km of an observed TCT track were counted as "TCT track points" (Table 5; Fig. 7 has a 3-km domain example). A total of 546 points were assigned to the observed TCT tracks on the 3-km domain, compared to the 8414, 1875, 928, and 190 points from the TCT-Ss using the 99.5%–99.99% percentile thresholds (Table 3).

This quantification alone would appear to support the use of the 99.95% or 99.99% thresholds, but confirmation of this is sought by a slightly different quantification, namely, the percentage of domain covered (PDC). PDC is defined herein as the ratio of the number of points exceeding the UH_{max} threshold to the total number of points in the 3-km (and 1-km) domain(s). The goal is to have the simulated PDC (Table 4) as close to or slightly greater than the observed PDC (Table 5), to ensure broad coverage of TCT-Ss without including extraneous or physically unimportant features. Consistent with the qualitative assessment, the 99.95% threshold (UH_{max} of 83 $m^2 s^{-2}$) for the 3-km domain yields a composite PDC of 0.223% for all three periods, which compares favorably to the observed PDC of 0.144%. The 99.9% (99.99%) threshold yields a PDC of 0.451% (0.046%), suggesting a TCT-S overprediction (underprediction) (bias scores in Table 4).

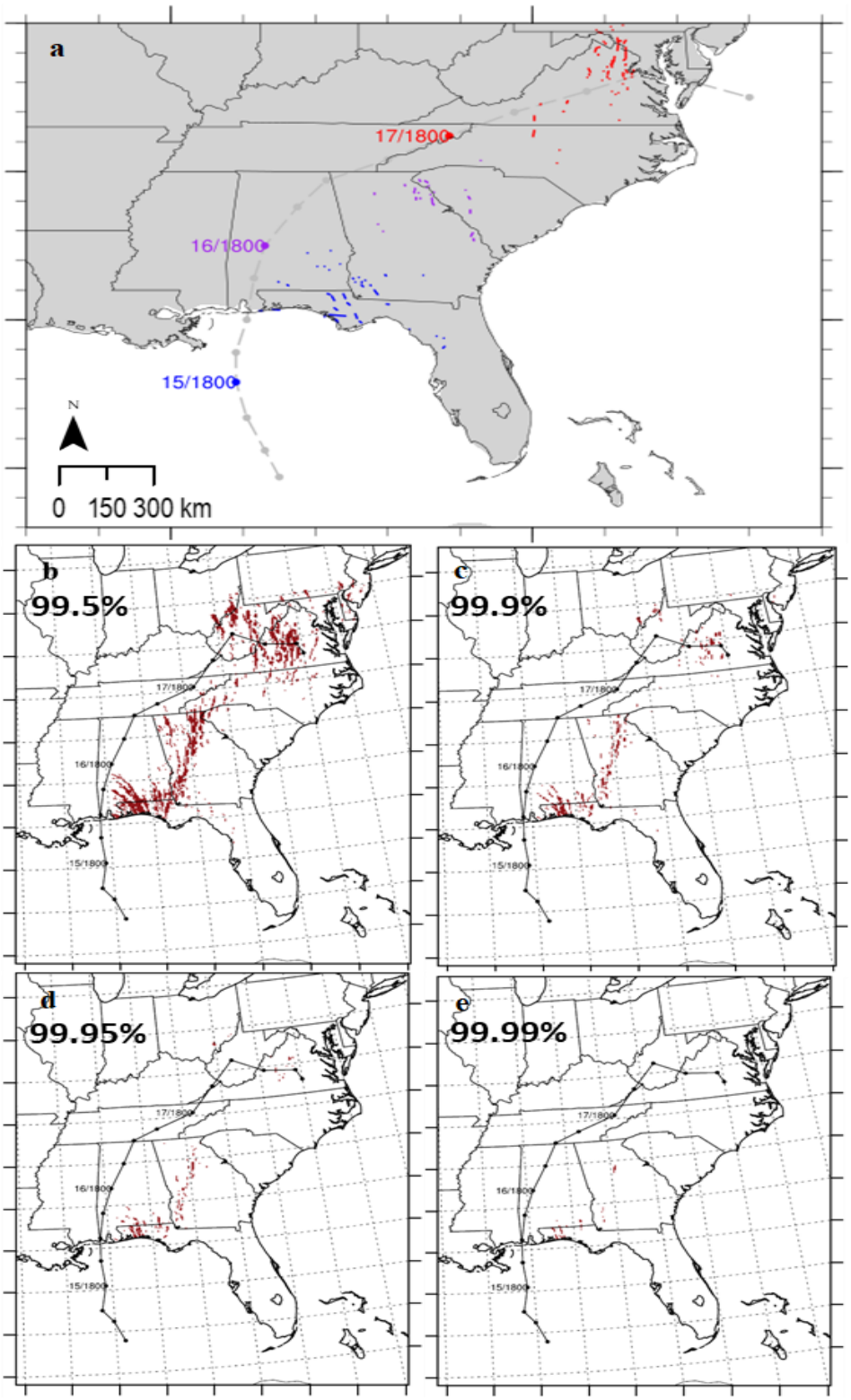


Figure 6: a) Observed TCT tracks (adapted from Edwards 2012) and b–e) simulated TCT-S based on UH_{max} thresholds using the 99.5%–99.99% values. The observed and simulated TC track is overlaid on the respective plots, annotated by the period. *Click image to enlarge*.

Table 4: Same as Table 3, except bias of PDC values (%), where bias is the difference between the simulated and observed PDC values.

	Period	99.50%	99.90%	99.95%	99.99%
3-km	LF,ML,ET	1.881	0.307	0.079	–0.098
1-km_COMP	LF	4.711	1.343	0.724	0.080
	ML	1.947	0.306	0.129	0.005
	ET	1.868	0.110	–0.029	–0.096
1-km_IND	LF	2.368	0.472	0.198	–0.038
	ML	2.560	0.496	0.231	0.015
	ET	3.238	0.625	0.269	–0.012

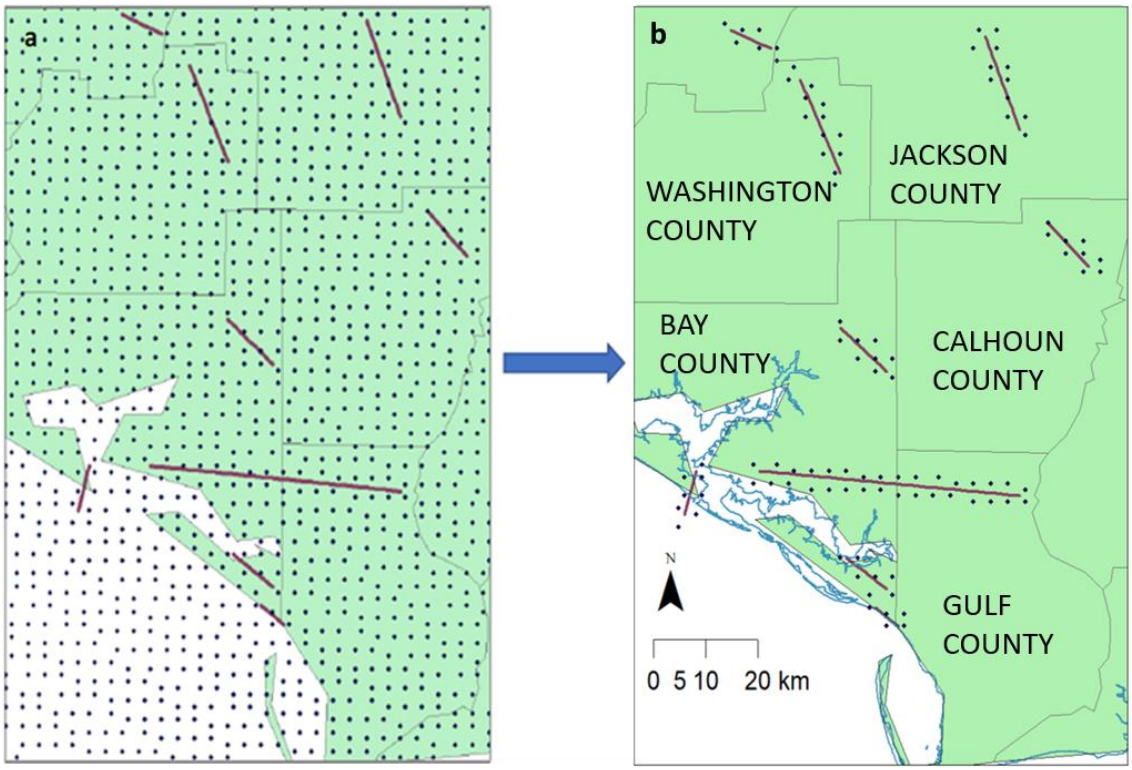


Figure 7: Example of gridpoint assignment for observed TCT tracks in the Florida Panhandle: a) 3-km grid superimposed on tracks, b) points within 3 km of (and assigned to) the observed tracks.

We next consider the TCT-Ss identified using the UH_{max} thresholds on 1-km domains (Table 3). When using composite values (i.e., thresholds based on the distribution of UH_{max} pooled from the three 1-km domains used for each period), the TCT-Ss diagnosed using the 99.9% threshold value of UH_{max} (378 m^2 s^{-2}) have the best overall spatial agreement with the observed TCTs (Fig. 8). Separated by period, the TCT-Ss diagnosed during LF and ML using the 99.99% and 99.95% composite threshold values, respectively (933 m^2 s^{-2} and 512 m^2 s^{-2}; Fig. 8), compare best to the observations, whereas the TCT-Ss diagnosed during ET using the 99.9% value (378 m^2 s^{-2}; Fig. 8) have the best qualitative comparison.

For a quantitative assessment, we again begin by assigning track points to the observed TCTs, except now relative to grid points in each of the

1-km domains, using a search radius of 1 km. This yields 471 points for LF, 173 points for ML, and 498 points for ET periods (Table 5). We found it somewhat surprising to have nearly the same number of track points during the LF period as during the ET period, with fewer tornadoes reported during LF than ET. This suggests that the TCTs near Ivan's landfall, though mostly weak (F/EF0–1), were longer-tracked than the stronger (F/EF2–3) TCTs that occurred during Ivan's extratropical transition. The observed TCT track lengths (Table 5) support this statement, where the maximum track length during LF (ET) was 50.1 km (39.0 km), and the mean track length during LF (ET) was 7.1 km (5.3 km).

The PDC for the observed TCTs on the 1-km domains are 0.103%, 0.029%, and 0.099% for the LF, ML, and ET periods, respectively (Table 5). The PDC for TCT-Ss on the 1-km grid generally agree with results from the spatial assessment, with the 99.99% threshold resulting in 0.183% and 0.034% coverage for the LF and ML periods, and the 99.9% threshold resulting in 0.209% coverage for the ET period. Unlike the observed TCTs, which generally increase inland, the TCT-Ss decrease as Ivan moves inland for all percentile thresholds (1km_COMP; Table 3). We attributed this in part to the use of a single UH_{max} threshold for all periods, versus one unique to each period, as discussed next.

Given the result in Table 2 showing a general decrease in UH_{max} from LF to ET, here we use thresholds based on the distributions of UH_{max} within each period, and thus over each 1-km domain. The most favorable spatial representation of TCT-Ss for each period is provided by the use of the 99.95% thresholds (785 $m^2 s^{-2}$, 417 $m^2 s^{-2}$, and 320 $m^2 s^{-2}$ for LF, ML, and ET respectively). The PDC values support the use of the period-specific 99.95% thresholds for LF and ET, and the 99.99% thresholds for the ML period. The period-specific thresholds (1-km_IND; Table 2) for the 1-km domains also yield an increase in TCT-Ss (1-km_IND; Table 3) as Ivan traverses inland, which was loosely shown in the TCT reports (Table 1). There is a decrease in reports from LF to ML and an increase from ML to ET.

Ultimately, the best threshold to use overall is based on the 99.95% value of UH_{max}, which provides favorable results for two of the three periods using the composite threshold, and all three periods using the individual UH_{max} threshold. This recommendation comes despite the 99.9% percentile threshold showing favorable results for the ET period for the 1-km composite UH_{max} threshold. These findings highlight the complexity of using the same threshold from a distribution of all the UH_{max} values verses calibrated thresholds based on the stage of Ivan's inland life propagation. This is evidenced by the ability to use higher individual thresholds for all periods, and still produce favorable results (Fig. 9). Implications of these results could suggest that using a "one-size-fits-all" UH_{max} threshold could lead to over-detection near landfall and under detection as a TC moves inland. The authors caution however, that these results are unique to Ivan, and additional testing is needed to determine the suitability of this methodology to other tornadic TCs.

Table 5: Statistical values of observed TCT track lengths for 1-km domain by period, and track points and percentage of domain covered for 1-km domain by period, and 3-km domain.

Period (domain)	Max Track Length (km)	Mean Track Length (km)	Sum Track Length (km)	Track Points	PDC
LF (1-km)	50.1	7.1	257.2	471	0.103
ML(1-km)	22.3	5.3	150.3	173	0.029
ET(1-km)	39.0	5.3	286.1	498	0.099
LF,ML,ET(3km)				546	0.144

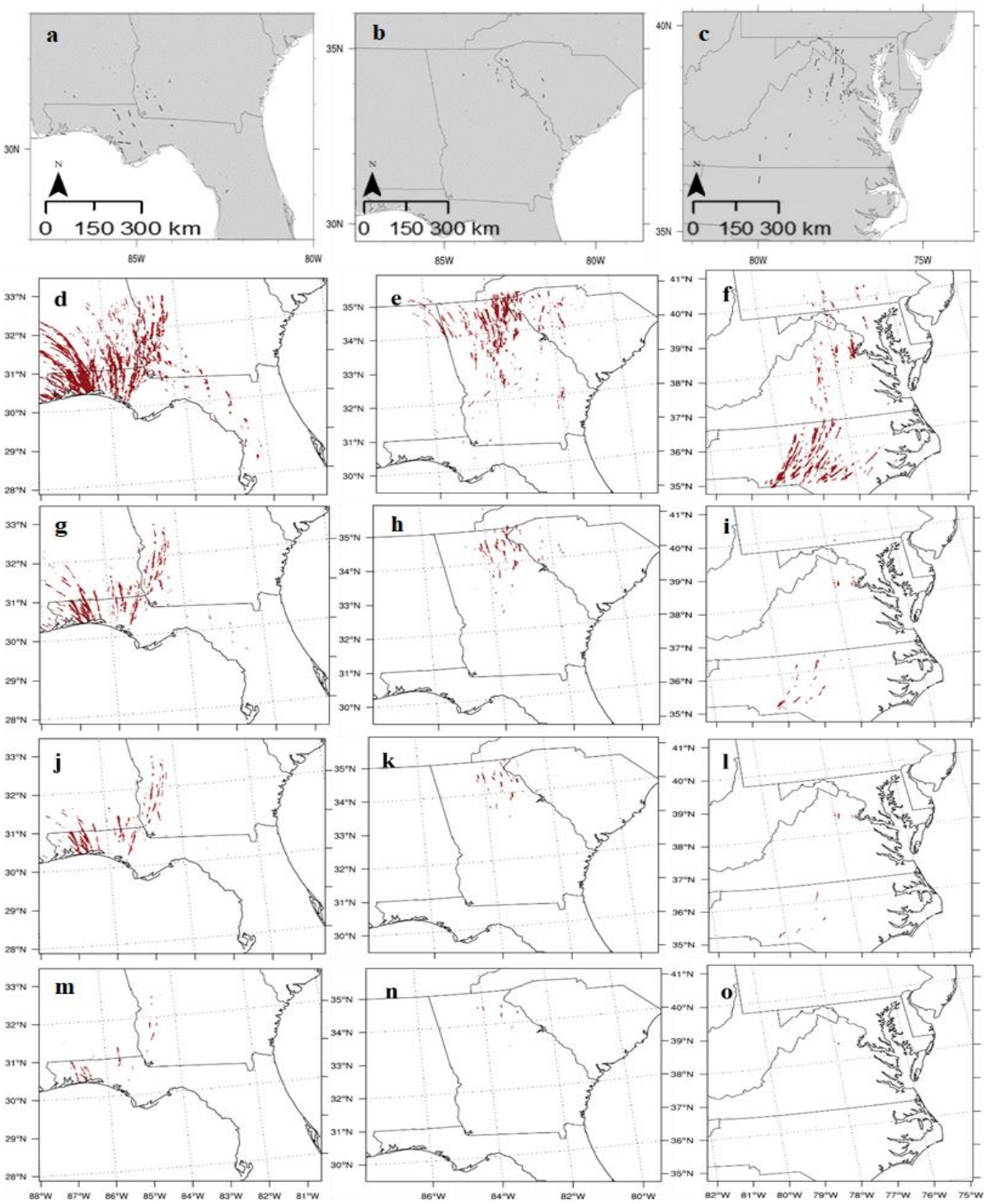


Figure 8: Comparison of a–c) the raw observed TCT tracks (black) to TCT-S (red) for the d–f) 99.5%, g–i) 99.9%, j–l) 99.95%, and m–o) 99.99% percentile thresholds for 1-km_COMP. *Click image to enlarge.*

c. Evolution of supercell intensity, and associated environmental parameters, with inland progression of TC

Based on Table 2 we hypothesize that the potentially tornadic storms have rotating-updraft intensities that vary with TC evolution over land. This is supported by a time series of maximum values of UH, vertical velocity, and vertical vorticity, over the 3-km domain, which indeed shows a decrease in simulated storm intensity over time (Fig. 10).

An analysis of the observed convective environment over the three periods perhaps helps explains this decrease. Note here that the definition of "environment" for TCTs is more complicated than for other tornadoes. This is because the TC itself (especially its wind field) contributes to the environmental conditions supporting TCTs, yet the TC is also comprised of deep convective clouds that locally reduce CAPE and otherwise modify wind shear.

With this caveat in mind, we collected the 1200 UTC sounding data from two stations that fell within the domain of each period (LF: Tampa Bay, FL, TBW, and Tallahassee, FL, TLH; ML: Peachtree City, GA, FFC, and Charleston, SC, CHS; ET: Washington/Dulles International Airport, VA, IAD, and Wallops Island, VA, WAL). The following environmental parameters were considered AGL: 0–1 and 0–3-km storm- relative environmental helicity (SRH01 and SRH03), 0–6-km bulk wind difference (BWD), mean-layer CAPE (MLCAPE), surface-based CAPE (SBCAPE), and the supercell composite parameter (SCP; Thompson et al. 2003) (Table 6). We then normalized the parameter values using the maximum and minimum values of each parameter across all stations, and then averaged the normalized values to determine a characteristic value for each period (Fig. 11).

The 1200 UTC analysis (≈6 h prior to each period's temporal boundary of 18 UTC) show that the largest values of BWD are during the ET period, thus reflecting a transition to a more extratropical environment. The variants of CAPE decrease with period and thus with TC evolution over land. SCP, which depends partly on CAPE, also decreases with TC evolution. Although the SRH values do not show the same sort of gradual decrease, they are the lowest during the ET period. This analysis suggests that environmental changes could explain the decrease in convective-cell intensity. Future work will further assess these environmental complexities, as we seek to understand the transitions of (and mechanisms for) different convective modes and intensities as the TC moves inland.

d. Structure of supercells during LF, ML, and ET

To address the aforementioned hypothesis of dependence of rotating-storm structure on inland TC evolution, we next analyzed such structure of various convective storms during each of the three periods. Specific cells meeting both the composite UH_{max} and reflectivity criteria in the hourly output were identified and further scrutinized using 5-min output on the 1-km domains. Horizontal cross sections were then used to document the size, intensity, and depth of a representative cell exhibiting a peak in UH_{max} for each period.

The following three cells were considered: Cell one (C1; Fig. 12) occurred during LF at 0850 UTC 16 September 2004, Cell two (C2; Fig. 13) occurred during ML at 2250 UTC 16 September 2004, and Cell three (C3; Fig. 14) occurred during ET at 0025 UTC 18 September. C1 and C3 had tops reaching up to 13 km as defined by the height at which vertical velocity $<10 ms^{-1}$ and SRF >20 dB*Z*; C2 had a storm top of 12 km. All three cells had relatively deep rotating updraft cores, 6 km in depth for C1 and 4 km in depth for C2–3. The rotating updraft cores were identified based on overlapping vertical velocity and vertical vorticity maxima >10 m s^{-1} and >0.01 s^{-1} respectively. Examples of this overlap are shown at 1- and 3-km AGL (Figs. 12–14). All three cells also exhibited vertical vorticity >0.01 s^{-1} at 500 m AGL, providing confidence that these storms could result in tornadic activity (Figs. 12–14). Cell C2 had the strongest low-level rotation, with 500-m vertical vorticity exceeding 0.02 s^{-1}. Several other storms in each period (not shown) also exhibited maximized vertical vorticity and velocity over 3–4 km AGL in depth, with some storm tops reaching up to 10 km AGL.

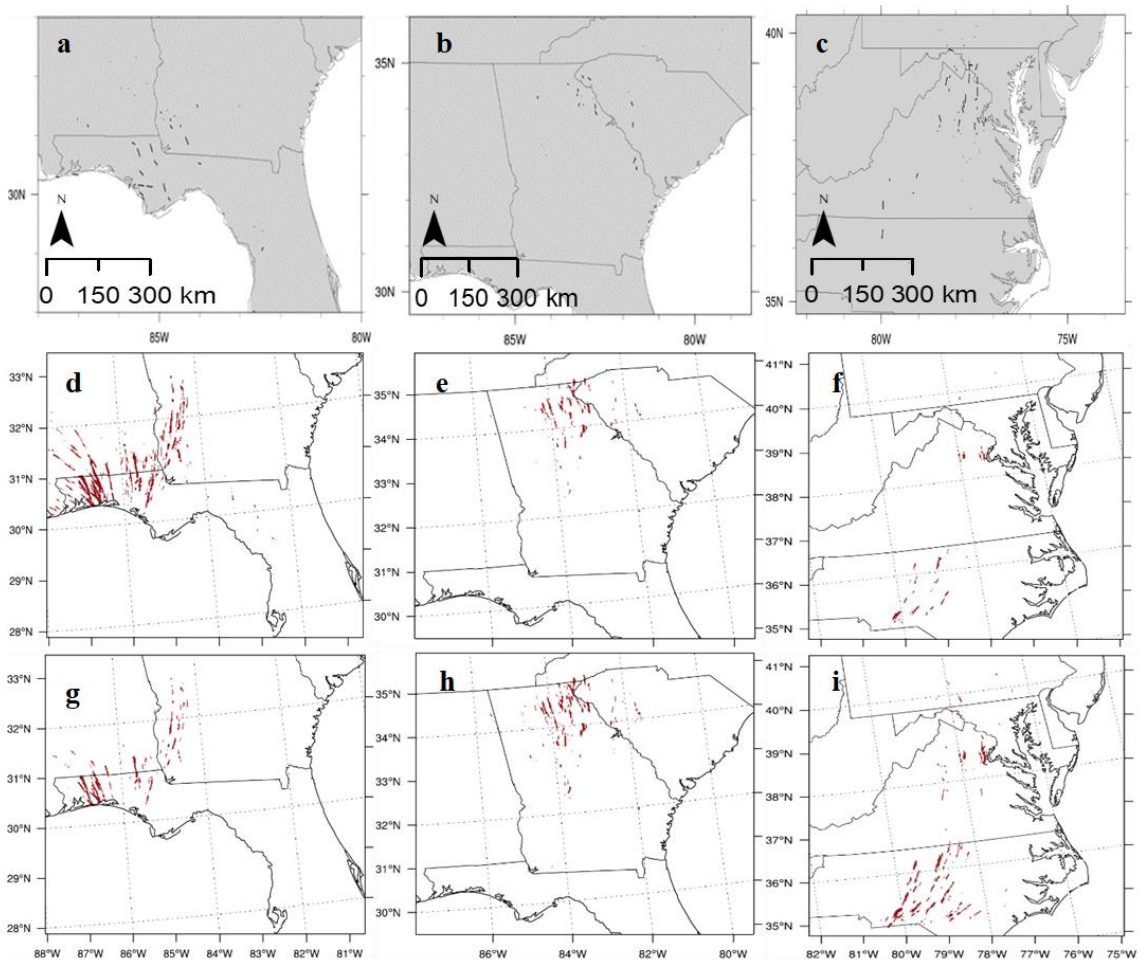


Figure 9: Comparison of a–c) observed TCT tracks to d–f) TCT-Ss based on 1 km_COMP and g–i) 1 km_IND for the 99.9% percentile thresholds for each period. *Click image to enlarge.*

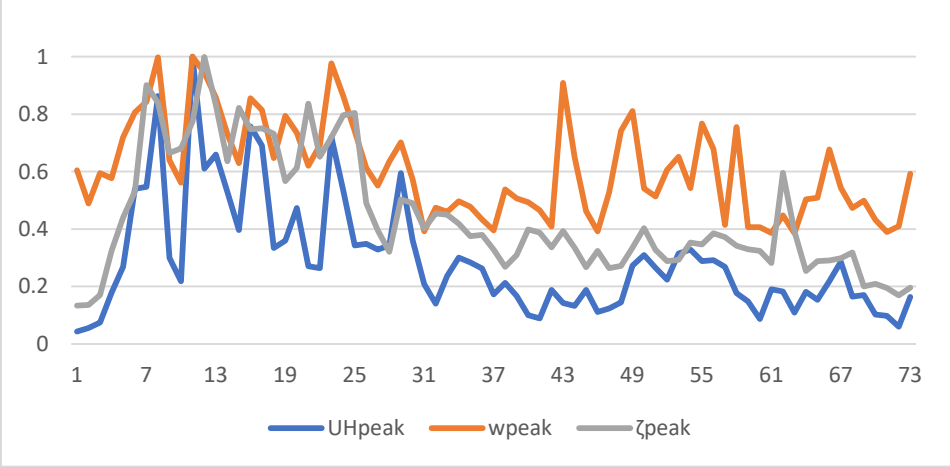


Figure 10: Hourly time series of normalized max values of UH (dark blue line), vertical velocity (w; orange line), and vertical vorticity (ζ; gray line). Values are normalized using the maximum value across all temporal and spatial domains. Peak values of UH, w, and ζ are shown to decrease over time, indicating weakening of simulated convective cells as the simulated TC moves inland.

Table 6: Raw environmental data values extracted from six stations defined in the text (two stations falling within the LF period): TBW, TLH; ML period: FFC, CHS; and ET period: IAD, WAL) at 1200 UTC, and averaged values assigned to each period (LF, ML, and ET) at 1200 UTC.

Date/Time (UTC)	Station/ Period	BWD	SRH01	SRH03	SBCAPE	MLCAPE	SCP
15 Sept. 1200	TBW	37	134	219	1809	3578	1
	TLH	15	222	296	1154	1854	1
16 Sept. 1200	FFC	43	281	433	339	1505	1
	CHS	25	92	75	1623	3149	0.19
17 Sept. 1200	IAD	49	63	123	0	542	0.21
	WAL	37	62	21	576	1446	0.03
1200	LF	26	178	257.5	1481.5	2716	1
	ML	34	186.5	254	981	2327	0.595
	ET	43	62.5	72	288	994	0.12

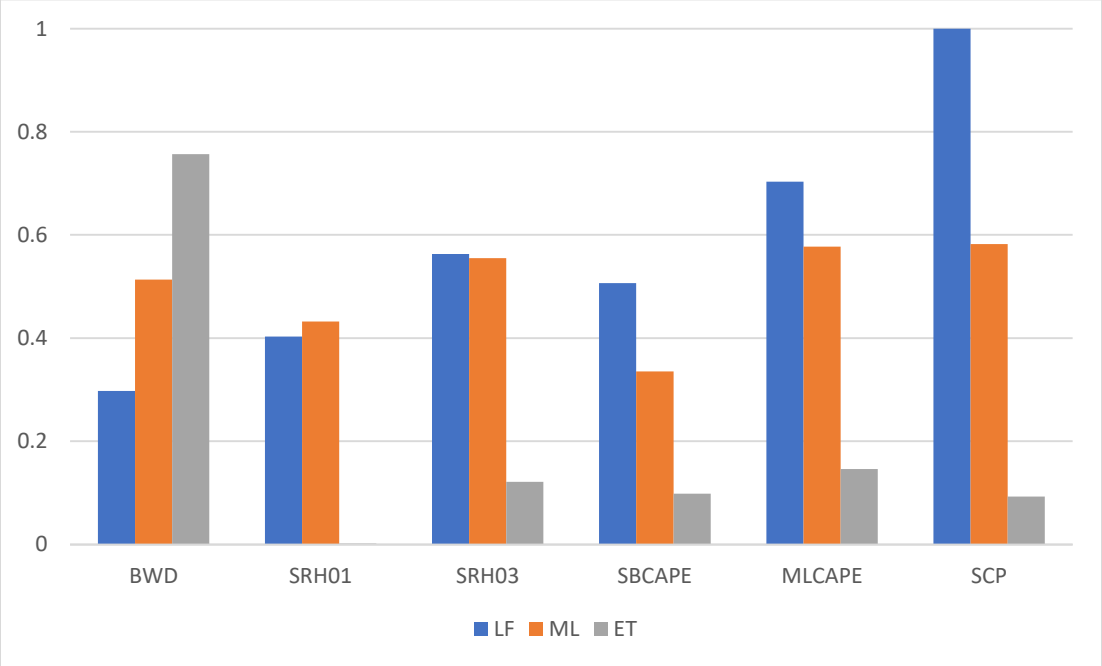


Figure 11: Environmental data values normalized using feature scaling with the overall maximum and minimum values for each variable; values for LF (blue), ML (orange), and ET (gray) periods are defined as in Table 6 at 1200 UTC.

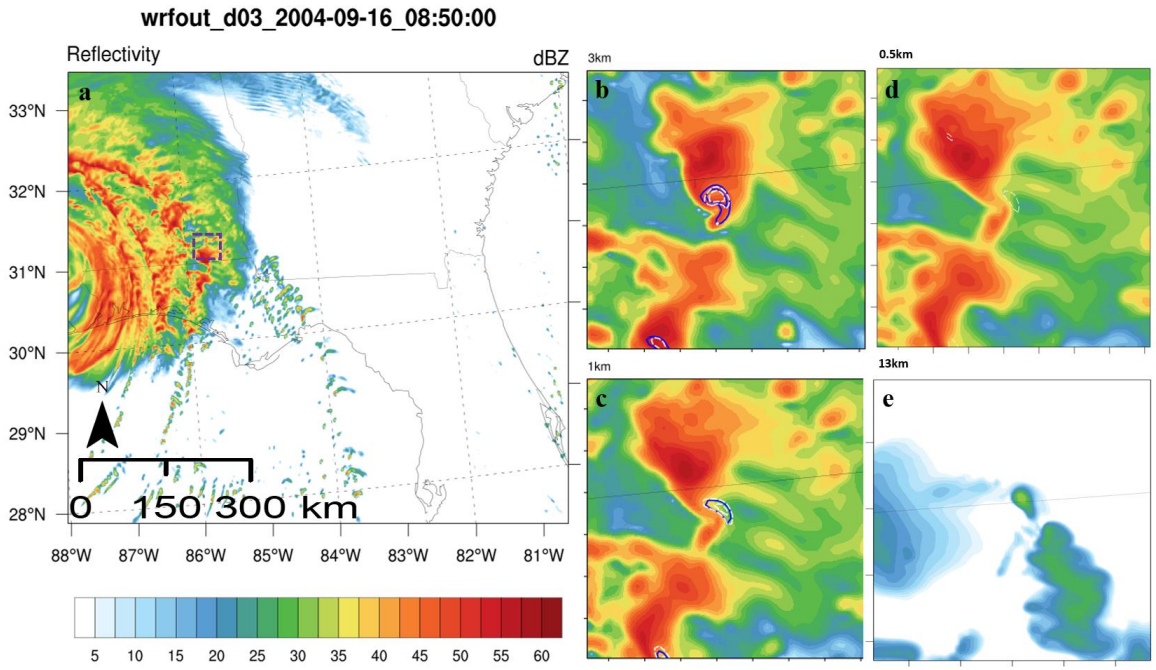


Figure 12: Panned view of a) a cell from 1-km domain at 0850 UTC 16 September 2004, with a dashed box signifying location where horizontal cross sections were taken at: b) 3 km, c) 1 km, and d) 500 m AGL (top). SRF (dBZ) is represented by shaded contours, vertical vorticity >0.01 s^{-1} by white dashed contour, and vertical velocity >10 m s^{-1} by solid blue line. Evidence of low-level rotation is shown in (d) and e) storm tops ≈13 km, indicated by height at which vertical velocity is <10 ms^{-1} and SRF >20 dBZ. *Click image to enlarge.*

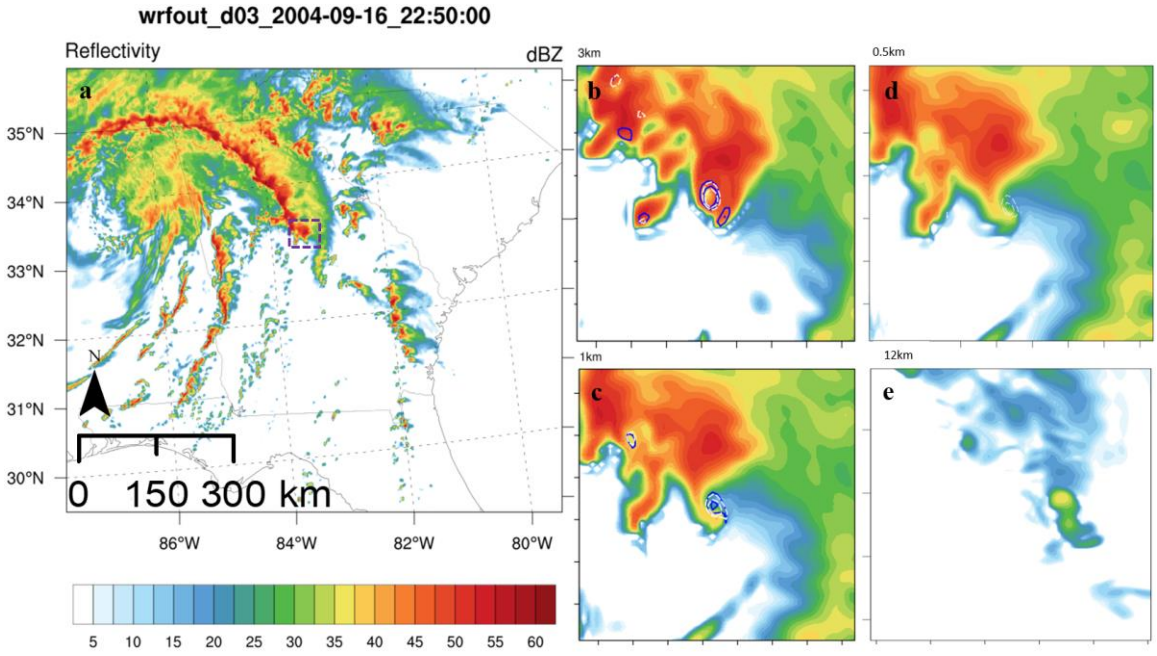


Figure 13: Same as Fig. 12 except 2250 UTC 16 September 2004, and e) storm top at 12 km. *Click image to enlarge.*

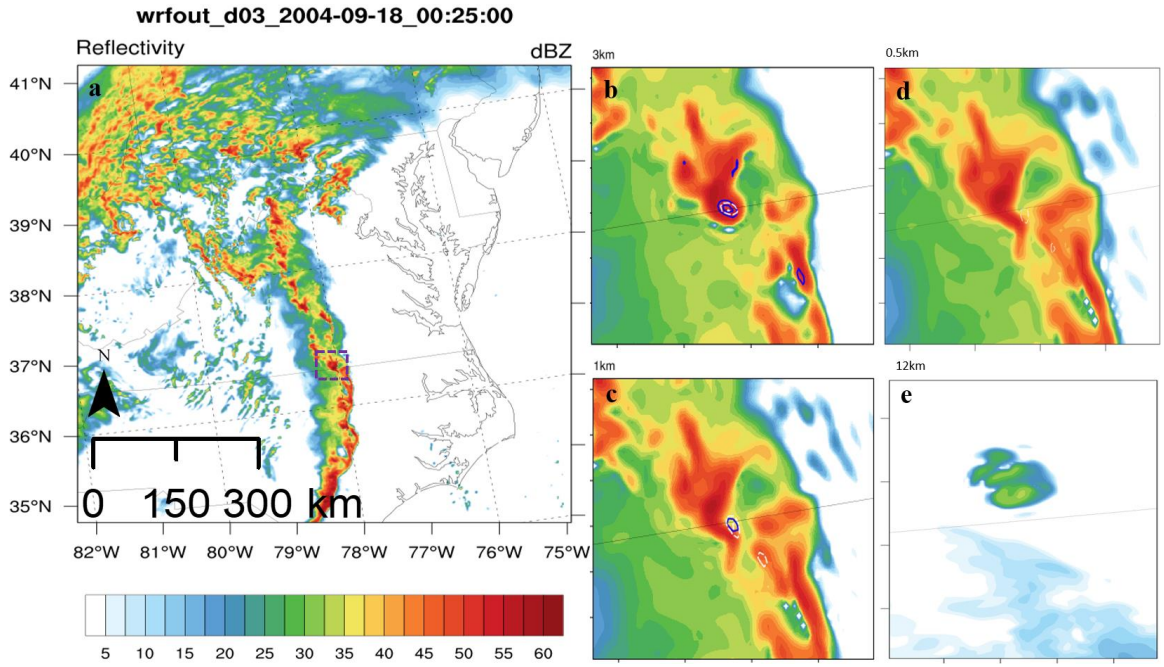


Figure 14: Same as Fig. 13 except ET 0025 UTC 18 September 2004. *Click image to enlarge.*

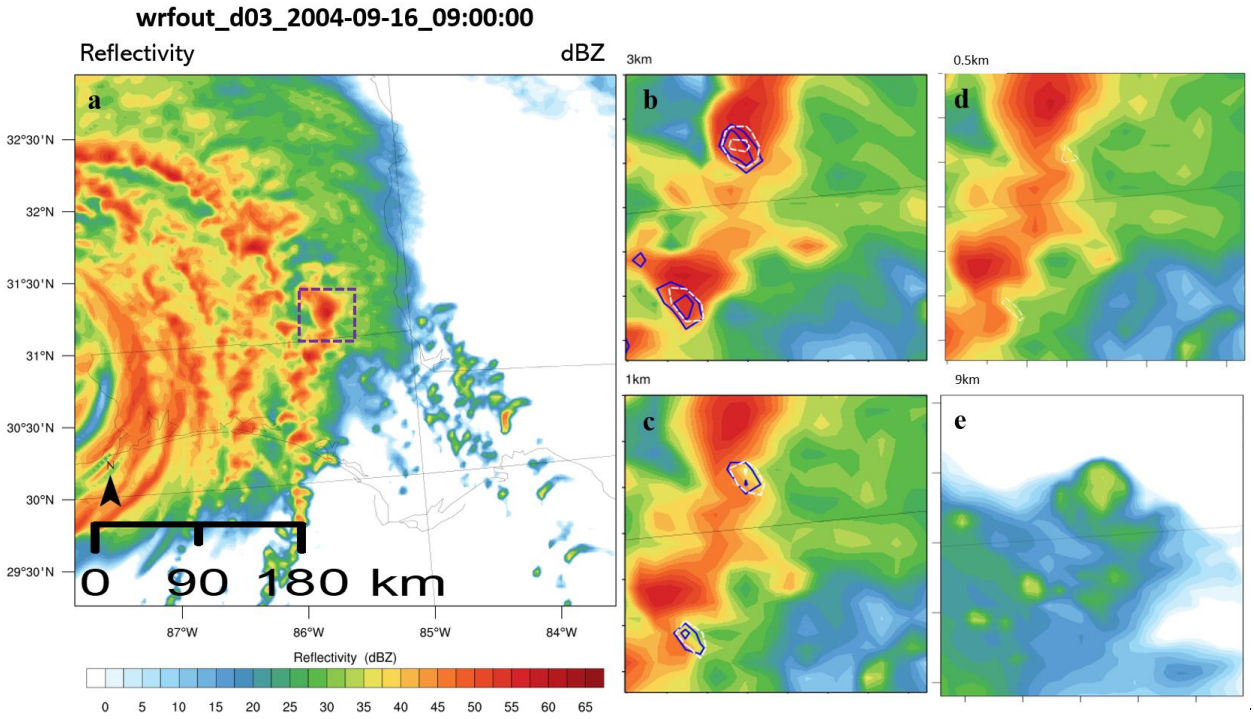


Figure 15: Same as Fig. 12 except for the 3-km grid, and for vertical vorticity >0.005 s^{-1}, vertical velocity >5 m s^{-1}, and e) storm top 9 km. *Click image to enlarge.*

Overall, the types of convective modes associated with TCTs, as identified by Edwards et al. (2012) (who adapted metrics from Smith et al. 2012), are represented in this model simulation. C1 has discrete supercell characteristics, defined as a deep rotating cell distinct from surrounding cells with radar echoes ≥35 dBZ and rotational velocities ≥10 m s^{-1} (Edwards et al. 2012). However, C2 and C3 represent cells with rotating cores, embedded within a quasilinear convective system and defined by continuous reflectivity >35 dBZ for a length ≥100 km at ≥3:1 aspect ratio (Edwards et al. 2012). Additionally, each of these cells have reflectivity appendages with a likeness to hook echoes, and thus at least have some resemblance to the more isolated cells and cells within lines that occur in the Great Plains (e.g., Smith et al. 2012).

These cell structures are based on output on the 1-km grids, but a relevant question for operational applicability is whether these structures would also be identifiable on the relatively coarser grids used in operational CAMs. Therefore, in Fig. 15, we show C1 as realized on the 3-km grid from hourly output. Cell C1 on the 3-km grid has the same basic features as on the 1-km grid (Fig. 12). The suggestion of this analysis is that potentially tornadic cells in TCs are better resolved with 1-km grid spacings, but such cells can still be represented on 3-km grids.

4. Conclusions

This study used high-resolution real-data WRF simulations of Hurricane Ivan to document the existence and structure of potentially tornadic supercells within tropical cyclone (TC) rainbands. The simulated TC track and intensity matched well with observations post-landfall, and the simulated TC structure closely replicated the observed rainbands and overall shape. TC tornado surrogates (TCT-Ss) were identified and calibrated using thresholds based on percentile values of maximum updraft helicity (UH_{max}) and simulated radar reflectivity.

Although the magnitude of UH_{max} generally decreased as the simulated TC moved inland, sensitivity testing revealed that a threshold based on the 99.95% percentile value of UH_{max} achieved optimal TCT-S coverage, and agreement with observed TCT tracks, on domains with 3-km and 1-km grid spacings. These values include: 83 m^2 s^{-2} for the 3-km domain, 512 m^2 s^{-2} for the 1-km composite domain, 785, 417, and 320 m^2 s^{-2} for the individual period specific, LF, ML, and ET 1-km domains, respectively. *These values are specific to Ivan*, and further testing is needed to determine the applicability of the TCT-S methodology to other TCs. Nonetheless, the complexity in choosing a proper threshold was revealed through the sensitivity tests. Three cells were identified at three different stages of the TC inland evolution and had hook-like appendages and other characteristics of supercells found in the Great Plains and elsewhere around the world. The cells produced storm tops of 12–13 km AGL, with rotating cores with depths of 4 km AGL.

This study highlights UH_{max} in a CAM as a potentially useful tool in forecasting TCT occurrence. Although improvement in storm-scale forecasts, using a higher resolution domain, agrees with Schwartz et al. (2017), the results from the 3-km grid show promise for CAM applications in operational settings (Fig. 13). Our future work will include simulations of recent TC events. We will also explore sensitivities to model physics, including exploration of the Asymmetric Convective Model version 2 (Cohen et al. 2017) for parameterizing boundary layer processes. Finally, we will explore the applicability of Doppler-radar-derived rotation tracks for TCT identification (as suggested by Carroll-Smith 2018) and CAM evaluation (as in, e.g., Dawson et al. 2017).

ACKNOWLEDGMENTS

The authors would like to thank the National Science Foundation Graduate Research Fellowship dev- 00053298, and the Blue Waters Supercomputing facility for providing the resources to complete this research. The authors would also like to thank the reviewers of this article for their thought-provoking comments and suggestions to improve upon this work.

REFERENCES

Baker, A. K., M. D. Parker, and M. D. Eastin, 2009: Environmental ingredients for supercells and tornadoes within Hurricane Ivan. *Wea. Forecasting*, **24**, 223–244.

Carley, J. R., B. R. J. Schwedler, M. E. Baldwin, R. J. Trapp, J. Kwiatkowski, J. Logsdon, and S. J. Weiss, 2011: A model-based methodology for feature-specific prediction for high-impact weather. *Wea. Forecasting*, **26**, 243–249.

Carroll-Smith, D., 2018: "If it happened in…": A pseudo-global warming assessment of tropical cyclone tornadoes. Ph.D. dissertation, University of Illinois at Urbana-Champaign, 100 pp.

Cavallo, S. M., R. D. Torn, C. Snyder, C. Davis, W. Wang, and J. Done, 2013: Evaluation of the Advanced Hurricane WRF data assimilation system for the 2009 Atlantic hurricane season. *Mon. Wea. Rev.*, **141**, 523–541.

Cohen, A. E., S. M. Cavallo, M. C. Coniglio, and H. E. Brooks, 2015: A review of planetary boundary layer parameterization schemes and their sensitivity in simulating southeastern U.S. cold season severe weather environments. *Wea. Forecasting*, **30**, 591–612.

——, ——, ——, ——, and I. L. Jirak, 2017: Evaluation of multiple planetary boundary layer parameterization schemes in southeast U.S. cold season severe thunderstorm environments. *Wea. Forecasting*, **32**, 1857–1884.

Chen, F., and J. Dudhia, 2001: coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev*., **129**, 569–585.

Clark, A. J., J. S. Kain, P. T. Marsh, J. Correia Jr., M. Xue, and F. Kong, 2012: Forecasting tornado pathlengths using a three-dimensional object algorithm applied to convection-allowing forecasts. *Wea. Forecasting*, **27**, 1090–1113.

——, J. Gao, P. T. Marsh, T. Smith, J. S. Kain, J. Correia, M. Xue, and F. Kong, 2013: Tornado pathlength forecasts from 2010 to 2011 using ensemble updraft helicity. *Wea. Forecasting*, **28**, 387–407.

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.

Czajkowski, J., and E. Kennedy, 2010: Fatal tradeoff? Toward a better understanding of the costs of not evacuating from a hurricane in landfall counties. *Pop. Env.*, **31**, 121–149.

Dawson, L. C., G. S. Romine, R. J. Trapp, and M. E. Baldwin, 2017: Verifying supercellular rotation in a convection-permitting ensemble forecasting system with radar-derived rotation track data. *Wea. Forecasting*, **32**, 781–795.

Eastin, M. D., and M. C. Link, 2009: Miniature supercells in an offshore outer rainband of Hurricane Ivan (2004). *Mon. Wea. Rev*., **137**, 2081–2104.

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.

——, 2012: Tropical cyclone tornadoes : A review of knowledge in research and prediction. *Electronic J. Severe Storms Meteor*., **7** (6), 1–61.

——, 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.

——, J. G. LaDue, J. T. Ferree, K. Scharfenberg, C. Maier, and W. L. Coulbourne, 2013: Tornado intensity estimation: Past, present and future. Bull. Amer. Meteor. Soc., **94**, 641–653.

Gentry, M. S., and G. M. Lackmann, 2010: Sensitivity of simulated tropical cyclone structure and intensity to horizontal resolution. *Mon. Wea. Rev*., **138**, 688–704.

Green, B. W., and F. Zhang, 2013: Impacts of air–sea flux parameterizations on the intensity and structure of tropical cyclones. *Mon. Wea. Rev*., **141**, 2308–2324.

——, ——, and P. M. Markowski, 2011: Multiscale processes leading to supercells in the landfalling outer rainbands of Hurricane Katrina (2005). *Wea. Forecasting*, **26**, 828–847.

Hong, S. Y., and J. O. J. Lim, 2006: The WRF single–moment 6–class microphysics scheme (WSM6). *J. Kor. Meteor. Soc.*, **42**, 129–151.

——, Y. Noh, J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long–lived greenhouse gases: Calculations with the AER radiative transfer models. *Geophys. Res. Lett.*, **113**, D13103.

Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945.

Kain, J. S., 2004: The Kain–Fritsch convective parameterization: An update. *J. Appl. Meteor.*, **43**, 170–181.

——, and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931–952.

——, S. R. Dembek, S. J. Weiss, J. L. Case, J. J. Levit, and R. A. Sobash, 2010: Extracting unique information from high-resolution forecast models: Monitoring selected fields and phenomena every time step. *Wea. Forecasting*, **25**, 1536–1542.

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

Lackmann, G. M., 2015: Hurricane Sandy before 1900 and after 2100. *Bull. Amer. Meteor. Soc.*, **96**, 547–560.

McCaul, E. W. Jr., 1991: Buoyancy and shear characteristics of hurricane-tornado environments. *Mon. Wea. Rev.*, **119**, 1954–1978.

——, and M. L. Weisman, 1996: Simulations of shallow supercell storms in landfalling hurricane environments. *Mon. Wea. Rev.*, **124**, 408–429.

——, D. E. Buechler, S. J. Goodman, and M. Cammarta, 2004: Doppler radar and lightning network observations of a severe outbreak of tropical cyclone tornadoes. *Mon. Wea. Rev.*, **132**, 1747–1763.

Molinari, J., and D. Vollaro, 2008: Extreme helicity and intense convective towers in Hurricane Bonnie. *Mon. Wea. Rev.*, **136**, 4355–4372.

Morin, M. J., and M. D. Parker, 2011: A numerical investigation of supercells in landfalling tropical cyclones. *Geophys. Res. Lett.*, **38**, L10801.

NCAR, 2000, cited 2015: NCEP FNL operational model global tropospheric analyses, continuing from July 1999. [Available online at https://rda.ucar.edu/datasets/ds083.2/.]

Osuri, K. K., U. C. Mohanty, A. Routray, M. Mohapatra, and D. Niyogi, 2013: Real-time track prediction of tropical cyclones over the North Indian Ocean using the ARW model. *J. Appl. Meteor. Climatol.*, **52**, 2476–2492.

Parker, C. L., A. H. Lynch, and P. A. Mooney, 2017: factors affecting the simulated trajectory and intensification of Tropical Cyclone Yasi (2011). *Atmos. Res.*, **194**, 27–42.

Rappaport, E. N., 2000: Loss of life in the United States associated with recent Atlantic tropical cyclones. *Bull. Amer. Meteor. Soc.*, **81**, 2065–2073.

——, 2014: Fatalities in the United States from Atlantic tropical cyclones: New data and interpretation. *Bull. Amer. Meteor. Soc.*, **95**, 341–346.

——, and Coauthors, 2009: Advances and challenges at the National Hurricane Center. *Wea. Forecasting*, **24**, 395–419.

Schultz, L. A., and D. J. Cecil, 2009: Tropical cyclone tornadoes, 1950–2007. *Mon. Wea. Rev.*, **137**, 3471–3484.

Schwartz, C. S., G. S. Romine, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2015: NCAR's experimental real-time convection-allowing ensemble prediction system. *Wea. Forecasting*, **30**, 1645–1654.

——, G. S. Romine, M. L. Weisman, R. A. Sobash, K. R. Fossell, K. W. Manning, and S. B. Trier, 2015: A real-time convection-allowing ensemble prediction system initialized by mesoscale ensemble Kalman filter analyses. *Wea. Forecasting*, **30**, 1158–1181.

——, ——, K. R. Fossell, R. A. Sobash, and M. L. Weisman, 2017: Toward 1-km ensemble forecasts over large domains. *Mon. Wea. Rev*., **145**, 2943–2969.

Simpson, R. H., 1974: The hurricane disaster-potential scale. *Weatherwise*, **27**, 169–186.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF Version 3. NCAR tech. note NCAR/TN-4751STR, 113 pp., doi:10.5065/D68S4MVH.

Skinner, P. S., L. J. Wicker, D. M. Wheatley, and K. H. Knopfmeier, 2016: Application of two spatial verification methods to ensemble forecasts of low-level rotation. *Wea. Forecasting*, **31**, 713–735.

Smith, B. T., R. L. Thompson, J. S. Grams, C. Broyles, and H. E. Brooks, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part I: Storm classification and climatology. *Wea. Forecasting*, **27**, 1114–1135.

Sobash, R. A., J. S. Kain, D. R. Bright, A. R. Dean, M. C. Coniglio, and S. J. Weiss, 2011: Probabilistic forecast guidance for severe thunderstorms based on the identification of extreme phenomena in convection-allowing model forecasts. *Wea. Forecasting*, **26**, 714–728.

——, C. S. Schwartz, G. S. Romine, K. R. Fossell, and M. L. Weisman, 2016: Severe weather prediction using storm surrogates from an ensemble forecasting system. *Wea. Forecasting*, **31**, 255–271.

Sun, X., 2013: High resolution simulation of Tropical Storm Ivan (2004) in the southern Appalachians: Role of planetary boundary layer schemes and cumulus parameterization. *Quart. J. Roy. Meteor. Soc*., **140**, 1847–1865.

——, and A. P. Barros, 2012: The impact of forcing datasets on the high-resolution simulation of Tropical Storm Ivan (2004) in the southern Appalachians. *Mon. Wea. Rev*., **140**, 3300–3326.

Thompson, G., P. R. Field, R. M. Rasmussen, W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev*., **136**, 5095–5115.

Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. M. Markowski, 2003: Close proximity soundings within supercell environments obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18**, 1243–1261.

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.

REVIEWER COMMENTS

[Authors' responses in *blue italics.*]

REVIEWER A (Daniel T. Dawson):

Initial Review:

Recommendation: Accept with major revisions.

Synopsis: This study describes analyses of CAM simulations of rotating convective storms in Hurricane Ivan (2004), focusing on verification with observed tornado occurrences as a first step toward assessing the potential for CAMs for TC tornado prediction. Overall, the paper is well-motivated and –organized, and addresses an important aspect of tropical cyclone hazards that has received relatively little attention from a numerical prediction standpoint, as compared with studies that focus on track, overall intensity, and rainfall hazards from TCs. My overall assessment of the study is that it has the potential to be an important contribution to this area, but requires major revisions before being accepted for publication.

My biggest concern has to do with the experiment design and analyses not adequately supporting the idea that changes in the structure of Ivan as it moved inland and made its extratropical transition are responsible for the temporal variation in UH_{max} thresholds that are required to get the best "match" with observations. This concern is tied up with other related concerns regarding a lack of clarity as to what sorts of structural changes (both on the larger TC scale and on the scale of the convective storms themselves) are important and how these should be diagnosed. Also, some of the description of methodology for both the experiment design and analyses could be improved. Overall, I think that these concerns could be adequately addressed (or contended, as the case may be) by the authors through revisions over a relatively short time period, and thus my recommendation is for acceptance pending major revisions.

Substantive/Major Comments: An important conclusion of your study seems to be that the proper choice of percentile threshold for identifying TCTS that match well with observed radar-derived rotation tracks and/or actual observed tornado tracks depends on the period over which the data is collected, corresponding to different intensities of rotating updrafts that themselves depend on different stages in the large-scale structural evolution of Ivan. I have two basic thoughts about this. First, I don't feel as if you have convincingly shown that this dependence on period is directly mappable to these structural changes, either on the storm-scale or the TC scale. The results of your study seem at least somewhat consistent with this, but 1) some physical justification or explanation for the computed trend in UH_{max} thresholds in terms of either the TC-scale or storm-scale changes should be provided, and 2) other possible reasons should be ruled out, including the possibility that the dependence is spurious (i.e., owing to limited statistics). One suggestion I would have to test this would be to choose random subperiods that don't correspond neatly to the three stages chosen in your study, to see if comparable changes in the "best" threshold occur for these as well. Another possibility is to perform an ensemble of simulations and assess the variability across the ensemble (though care would have to be taken to ensure that the periods in question had similar ensemble spread, which would require careful experiment design and different initialization periods). Still another is to repeat the experiment on other landfalling tornado-producing TCs, (though this may be too much to ask for the current study!).

Second, it seems to me that there are many other potential dependencies that aren't accounted for or discussed in this paper. One would be the choice of physics packages, particularly the microphysics scheme, which, among other things, affects the appropriateness of the chosen threshold of reflectivity for the TCTS algorithm (see more on this below). While clearly an exhaustive study of the possible sensitivities is too much to ask for, I think this study would benefit from at least a more thorough discussion of these possibilities and preferably one or more sensitivity experiments along the above or similar lines.

We appreciate your comments and suggestions. Physical justification to explain the trends in the UH_{max} values are beyond the scope of this research and would be best explored in a separate study. We feel this

study highlights the complexity of using a "one-size fits all" threshold, while shedding light on the need for more research into the physical mechanisms governing tornadic rotating storms as the TC moves inland. We agree that repeating this experiment on other landfalling tornado-producing TCs would be desirable, however as you mentioned, that is beyond the capabilities of the current study. Incorporating more cases will be the focus of future work.

We conducted sensitivity experiments in an adjacent study using the YSU PBL and the WSM6 MP schemes on one two-way nested domain at 9 and 3 km. Results showed that TCT surrogate generation was most sensitive to the choice of PBL scheme, even though the WSM6 experiments had stronger rotating updrafts, owing in part to stronger updraft speeds. A more thorough examination of these sensitivities is needed and therefore we did not include the results in this study at this time.

It's not clear to me what conclusion should be drawn from the analysis of the convective mode during the three periods chosen (i.e., section 3c). You initially state that you did so to "address the aforementioned hypothesis of a dependence of rotating-storm structure on inland TC evolution", yet the evidence shown for this dependence strikes me as relatively weak, at least from the results shown here. Indeed, you don't seem to make a claim one way or the other about how your hypothesis fares, other than to suggest that the mode in the first period is more discrete (but see my comment below). Choosing just three cells from each period has limited usefulness (unless care has been taken to ensure representativeness), as one doesn't know if the differences are robust or merely due to chance. Moreover, from just the horizontal cross-section plots alone, one can at best derive rough qualitative differences.

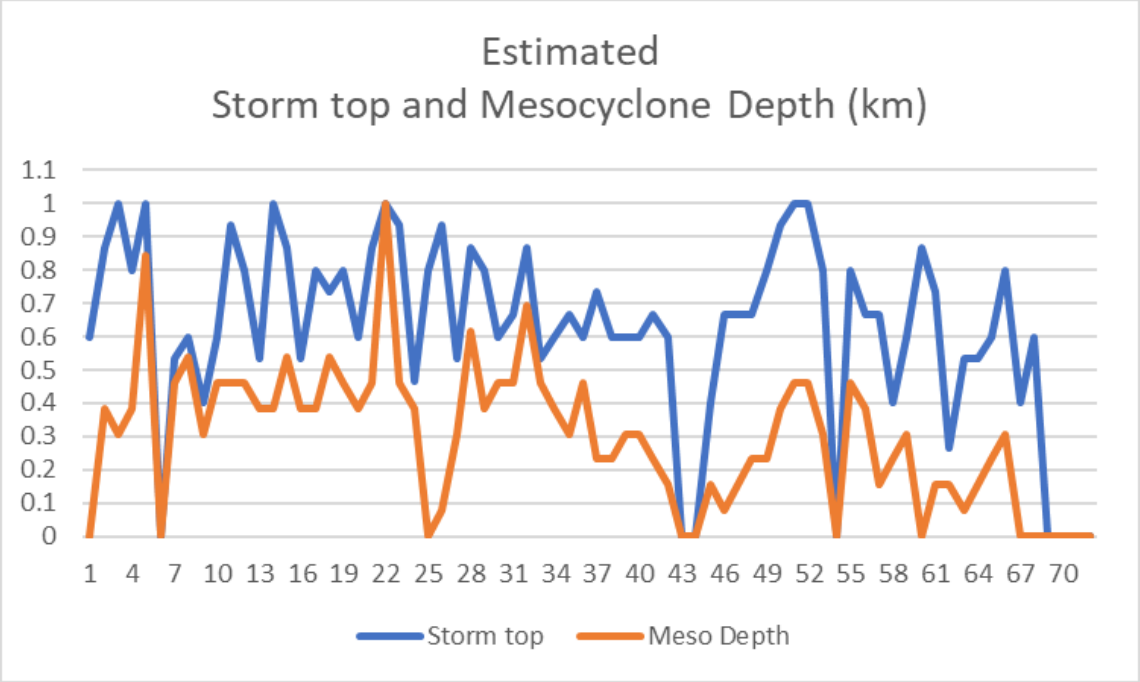
To really get at the question of whether there is a dependence on inland TC evolution requires additional quantitative analyses (i.e., such as trends in updraft and mesocyclone strength, depth, areal coverage, etc., not to mention analyses of environmental characteristics such as CAPE and bulk shear). It seems to me that there is potential here, with a modest amount of additional effort, to greatly improve the analyses in this section and thus to better vet this hypothesis. The results could also feed back to your UH_{max} threshold analysis in the previous section. Along these lines, it occurs to me upon looking back at Table 2, that there is a clear decrease in the UH_{max} values for each of the percentiles on the 1-km domain across the three periods, which could easily be unpacked in more detail in this context.

We agree, a quantitative analysis of storm scale and TC scale characteristics would improve this analysis. However, what we've found in testing this so far is that doing this analysis presents a couple of issues. 1) Capturing pre-convective values of environmental variables such as CAPE as often done for mesoanalysis of severe weather from non-tropical origins, is complicated. We struggle with how to resolve wanting to capture the pre-convective/pre-TC environment while also acknowledging that the TC environment provides the ingredients that lead to tornadic activity. 2) The second issue that we were working to resolve is having to average environmental variables over such a large area. It seems that carrying out the analysis you mentioned, would require a total re-work of the methodology, shifting the focus instead on one cell from each period.

We included a trend analysis below of peak values of UH from the 3-km grid analysis. Just as expected, peak values of UH, vertical vorticity and velocity, decrease as the storm moves inland.

[Editor's Note: What became Fig. 10 in the final manuscript was shown here. See Fig. 10 for illustration.]

An objective analysis of mesocyclone depth and storm top (defined here as the height above which vertical velocity <10 ms^{-1}), show a similar trend, though not as apparent.



Methodology, section a: Some more detail about the initialization of the model simulations is needed. Did you perform a single initialization of the 9- and 3-km grids, and only updated the boundary conditions thereafter for the entire 5-day period? Or were multiple initializations made for each stage of the event (i.e., those corresponding to the three 1-km domains)? (I know you say that the 9- and 3-km grids were integrated for the entire 5-day period, but what's not clear is if there was any stopping and restarting with fresh FNL analysis during this period). This has implications for how to interpret the second and third (ML and ET) periods, since if only a single initialization is used, the accuracy of the entire simulation would presumably suffer as one gets further from the initialization time. On a related note, how much time was given for the "spin-up" of the simulation prior to the analysis period for the landfall stage?

The authors made a single initialization for the 9- and 3-km grids beginning at 0000 UTC 14 September, corresponding to roughly 48 h prior to Ivan making landfall, 0600 UTC 16 September. Integrations of the different 1-km were done using the boundary conditions from the 9- and 3-km domains.

Methodology, section b: First, regarding the use of UH between 2 and 5 km AGL: since TCTs typically form from relatively low-topped convection (with mesocyclone depths ~4 km as described in the introduction), it seems odd to choose a range of heights that was developed to diagnose mid-level mesocyclones in typical continental supercell environments. What was the rationale for this decision? Second, while you state that a TCTS is defined based on the joint exceedance of a UH and Z threshold, it's not clear how you use this exceedance information to diagnose a given TCTS. Is there some feature-tracking methodology that is also applied, or is it done grid point-by-grid point? Presumably your methodology is based on that of Dawson et al. (2017) (which, by the way, is missing from the reference list), but it would be very helpful if you briefly described it here, as well, given its central importance to your study.

The TCTS are defined by grid point exceedances of a defined threshold of UH and simulated radar reflectivity. Feature tracking is not a part of the methodology at this time.

Methodology, section c: The use of the same dBZ threshold (30 dBZ) for the rotation tracks derived from the radar data as that for the model TCTS algorithm strikes me as potentially problematic, because you are then (apparently) not accounting for potentially large systematic biases between the simulated and observed reflectivity. Evidence for this large bias comes from Fig. 3, where at least as far as composite reflectivity

goes, the simulated reflectivity from the Thompson scheme is systematically quite a bit higher than the observed over large portions of the storm. I strongly recommend considering calculating and correcting in some way for this bias when attempting to compare the rotation tracks and TCTS. Also, this brings up another question: is the reflectivity threshold used based on the composite reflectivity, or the lowest elevation, or some other elevation?

Thank you, you make a very good point, however this analysis would not be feasible with the present study. Bias correction of the simulated reflectivity would require a larger sample of simulated and observed values to understand which bias characteristics of the model needs to be adjusted. You also make a good point regarding the Thompson microphysics scheme; however, the authors believe the resolution and other physics parameterizations would factor into determining the bias characteristics of these specific forecasts. In addition, bias correction would not be feasible for this study given the limited radar observations available for this case. If you recall, the authors were limited to using rotation tracks data for the ET period, or last 24-h period. Given these reasons, an appropriate bias correction would not be possible for this case.

For the rotation tracks, the reflectivity threshold is applied along radar elevation angles throughout the 0−3-km layer. The raw radar data is saved on different elevation angles, and the azimuthal shear is computed along those elevation angles. The vertical max (i.e., composite) is then taken to get the maximum value in the 0−3-km layer.

The reflectivity threshold for the TCTS analysis is based on the lowest model level.

Results: regarding the assigning of grid points to TCT tracks: this is more of a suggestion, but what about making a figure showing the results of this distance-based selection method for each of the domains as compared with the simulated TCTS track coverage?

Please see Fig. 6.

Results, regarding the PDC calculation. Another suggestion: instead of the raw values of the simulated PDCs, why not calculate the bias scores between the various simulated PDCs and the observed PDCs and show those directly in Table 4 (and Table 5 for the 1-km results)?

The authors calculated the PDC value and updated the table accordingly. Table 5 consists of the observed PDC and was not updated with bias calculations.

Results, section c: how representative were the storms chosen in each of the three periods to the other rotating storms in their respective periods? Also, how did the overall structure of these storms compare to their observed counterparts? I'm thinking particularly of the mesocyclone depth and overall storm depth. This has implications for the choice of the 2–5-km layer for the UH calculations mentioned above.

The authors chose three cells based on the criteria that they had a maximum value of UH for the 5min output on the across the 1-km domain, or hourly output on the 3-km domain. In addition, the cells had to meet the criteria of overlapping vertical velocity and vorticity maxima greater than 10 m s^{-1} and 0.01 s^{-1} respectively, and low-level vertical vorticity >0.01 s^{-1}. Many of the storms met the maxima in vertical vorticity and vertical velocity but did not meet the low-level vertical vorticity criteria. These cells were also much smaller and less structured than the representative cell. They are also situated in clusters within the distant rainbands.

Results, section c: description of convective modes. It's not obvious to me that there is a substantial difference in the convective mode between cell C1 on the one hand, and cells C2 and C3 on the other. They all appear to be in relatively close proximity to their neighbors that are (more-or-less) arranged in a line. Could you elaborate on the reasons for your claim? It's also worth pointing out that while it's true that supercells in the Great Plains *tend* to be more isolated, they do still come in a variety of modes, and there are many examples of at least as close spacing and similar linear relationship as seen here (see, e.g., Smith et al. 2012).

Smith, B. T., R. L. Thompson, J. S. Grams, C. Broyles, and H. E. Brooks, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part I: Storm classification and climatology. *Wea. Forecasting*, **27**, 1114–1135.

Thank you for this reference, the text has been updated accordingly. Regarding the cell's convective modes, the authors can see how there is ambiguity in determining which cell fell under a certain category. We objectively determined the convective mode, based on the cell's position relative to other cells around it, and using Edwards et al. 2012, Part III of the referenced series as guidance.

Edwards, R., A. R. Dean, R. L. Thompson, 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.*

[Minor comments omitted...]

Second Review:

Reviewer recommendation: Accept with major revisions.

Synopsis: The authors have improved the manuscript in response to my comments, and overall I think it is closer to publication. However, based on a careful reading of the revised paper as well as the responses to my (and the other reviewers) original comments, I think there are still a lot of unanswered questions and potentially problematic issues that still need to be answered and addressed, and I have described these below.

Substantive/Major Comments: I see that the authors removed the discussion of how the observed TCTs were derived from an analysis of radar-based rotation tracks as a response to one of the other reviewers' criticisms, but I admit now to being rather confused. In my previous review I was laboring under the (evidently mistaken) assumption that the "observed TCTs" were in fact those derived from the authors rotation track analyses, but now that those have been removed, it's not at all obvious to me now where the observed TCT data are coming from. Is it from *Storm Data* or other storm reports? The only clue I could find is in the caption to Fig. 6 where it is stated that the tracks are adapted from Edwards (2012). Digging into this paper, it looks like the tracks may ultimately come from SPCs "TCTOR" dataset. Is this the case? Either way, the origin and basic characteristics (i.e., does it consist of start and end points, etc.) of the TCT track data the authors use for verification needs to be clearly explained in the paper.

The tracks are indeed from the SPCs TCTOR dataset and the text has been updated to reflect this information, including basic characteristics of the TCT track data.

In response to my previous major comment #1, the authors argue that "physical justification to explain the trends in the UH_{max} values are beyond the scope of this research and would be best explored in a separate study." I think this is fine as far as it goes, but the current manuscript still claims that "Implications of these results could suggest that using a 'one-size-fits-all' UHmax threshold could lead to over detection near landfall and under detection as a TC moves inland." My ongoing concern is that this statement cries out for at least some expansion by: 1) giving a plausible physical explanation based on the analyses the authors have already performed (see my second comment below) or could be performed with modest additional effort, and 2) emphasizing that this may only apply to Ivan. I realize that the latter point is made later on in the conclusions, but I strongly suggest that it also be emphasized in the text where the statement is first made. Also, the authors discuss some sensitivity experiments that they performed but elected to not include the results of in the current paper. I understand and agree that a more thorough analysis could be left for future work, but why not at least mention that some of these experiments have already been performed and briefly discuss the preliminary results in the Summary section?

The authors added a statement in section 2a addressing the sensitivity experiments for physical parameters conducted in a previous study, as well as experiments testing the optimal level of UH max in section 2b paragraph 1. We also emphasized the results are unique to Ivan and can be applicable to storms like Ivan.

In regard to the authors' response to my previous major comment #2 and associated revisions: First, thanks for your explanation regarding the evolution of environment variables; I understand that this is more complicated than it seems on the surface, and that there really isn't a good way to separate the "mesoscale pre-convective" (if by this we mean before the TC arrives on the scene) from the TC-induced mesoscale environment (with likely already ongoing convection). In fact, it was the latter sense I had in mind when I made my original comment, but admittedly, I wasn't clear on this point. That is, by "environment" I meant the mesoscale circulation associated with TC, which necessarily includes the effects of the convection itself to some degree (at least its upscale influences). I think we can agree that from this perspective, the "preconvective" environment in the above sense may have limited relevance to the subsequent tornadic activity. Nevertheless, it seems that it should still be possible to calculate relevant environmental parameters on the mesoscale directly from the model fields (with suitable filtering) and/or leverage the SPC mesoanalysis (or some other comparable analysis dataset) fields to help shed light on some of these issues.

Unfortunately, the SPC mesoanalysis is unavailable for 2004, when Ivan occurred, however, the authors assessed the TC-induced mesoscale environment using SPC observed soundings. Please see results in section 3c.

But, leaving aside the question of analyzing environmental characteristics, the authors have provided some interesting plots of trends of peak UH/w/low-level ζ in their previous response that I think would go a long way toward improving the paper by giving at least some context to (e.g.) the finding that using the "'one-size-fits-all' UH_{max} threshold could lead to over detection near landfall and under detection as a TC moves inland". That is, the finding that the period specific thresholds for a given percentile decrease for each subsequent period seems to be consistent with the finding that the overall intensity of the simulated storms decreases for each subsequent period (as shown in the figures given in the authors' previous response). In short, I strongly recommend including these figures (or something like them) in the manuscript itself, along with appropriate description of their implications (whether or not the authors agree with my assessment).

These results have been added to section 3c.

Is it safe to say now that the three simulated supercells chosen for further analysis are in fact the cells with the strongest rotation during their respective periods? If so, I think this should be described clearly in the manuscript. In section 3d, the authors state that "Horizontal cross sections are then used to document the size, intensity, and depth of a *representative* cell exhibiting a peak in UH_{max} for each period" (emphasis mine). In my mind, the strongest cell (or at least the one with the strongest rotation) is (almost by definition) not representative of its peers. Can the authors clarify this?

The cells chosen were not the most intensely rotating cells within each period, although they were among the hourly maximum and 5-min maximum UH values within that period and met the threshold requirements of overlapping contours of vertical vorticity >0.01 s^{-1}, vertical velocity >10 m s^{-1}, and/or low low-level (500-m) vertical vorticity >0.01 s^{-1}. The cells with the max value of UH, however, did not meet the low-level vertical vorticity requirement.

What is meant by an "over-representation in TCT-S occurrence"? How was this determined?

Over-representation in TCT-S occurrence is indicated quantitatively by the percentage of domain covered for a particularly temporal and spatial domain. The text has been updated accordingly and directs viewers to the section detailing the calculations of the percentage of domain covered.

[Minor comments omitted...]

REVIEWER B (Ariel E. Cohen):

Initial Review

Reviewer recommendation: Accept with major revisions.

Overall summary: This submission provides an excellent first start to the application of simulated convective attributes from convection-allowing model guidance to assessing the potential for tornadoes to accompany tropical cyclones. I commend the authors for their very thorough investigation and use of integrated datasets to formulate thresholds from numerical guidance to assess tornado potential in a repeatable manner. My overall concern lies with the potential implication of more widespread applicability of this work to a larger sample of tropical cyclones. The inland thermodynamic profile with Ivan was characterized by a rather substantial amount of buoyancy, potentially unrepresentative of a broader distribution of tropical cyclones that also produce tornadoes—perhaps with less buoyancy. This calls to question the global application of the UH thresholds nominally stated in this work—and even the layer of computed UH—especially given the characteristically low-RCAPE (reversible CAPE; see Edwards and Thompson's paper on RCAPE) part of the parameter space that many tropical-cyclone tornadoes occupy. While the authors clearly suggest that their work effectively has motivated additional research and future work to investigate some of their findings in more detail, I believe it is very important that findings from this work—in its present form—be very specifically referred to as emanating from a single case study and may not be applicable to many other systems. Unique characteristics of the buoyancy distribution for this system may render findings only applicable to this system, and this system alone, without additional investigation. Nevertheless, the authors have laid out a fantastic first approach to CAM storm-attribute evaluation for tropical cyclone tornado threat assessment—undeniably a novel effort well deserving of incorporation into meteorological literature.

Major Comments: Because the *typical* TCT is characterized by a convective-inflow environment marked by limited RCAPE, and often muted traditional CAPE, in association with the deep moist-neutral thermal profile, it is often the shallowest layers of buoyancy and related vorticity stretching that are critical to the amplification of surface-layer vorticity supporting tornadoes. As a result, I am greatly concerned with widespread use of the 2–5-km layer for assessing updraft helicity for TCTS identification.

Sobash et al. (2016) highlight two subsets of simulated convective elements: one with substantial 1-km relative vorticity, and another with ratios of 0–3-km UH to 2–5-km UH above unity. In the case of the first subset, Sobash et al. (2016) suggest a linear relationship between 2–5-km UH and 0–3-km UH, and this established relationship would support the use of either layer of UH. However, this assumes substantial 1-km relative vorticity, which has not been established by the authors of the present work as being characteristic of simulated environments producing TCTSs—for Ivan or other TCs.

Now, it is entirely possible that the usage of 2–5-km UH can be justified for the buoyant sector of—specifically—Ivan. Observed soundings at Dulles and Wallops Island at 00Z peripheral to Ivan (link) indicated MLCAPE upwards of 2000–3000 J kg^{-1} in association with the buoyant convective inflow. This profile will have the potential to support a deeper rotating updraft, making 2–5-km UH more relevant as a proxy for tornado potential, than a Tropical Cyclone Harvey-like case (SPC Mesoanalysis MLCAPE: link). As a result, it may be coincidental that the 2–5-km UH is found to be relevant for adequately identifying TCTSs for Ivan, especially if these TCTSs are associated with larger 1-km relative vorticity supporting the correlation of 0–3-km UH to 2–5-km UH—consistent with the first subset identified by Sobash et al. (2016). Indeed, Sobash et al. (2016) indicate surface-based CAPE associated with this first subset as being around 1600 J kg^{-1} larger than that associated with the second subset.

So, if the more global distribution of TCTs—less representative of the greater-buoyancy and prolific-tornado production for Ivan—were characterized by Sobash et al. (2016)'s second subset identified, then the linear relationship between 0–3-km and 2–5-km UH no longer holds the way it does for the first subset. This would be substantiated by the lower buoyancy characteristic of the second subset. In this second subset, higher ratios of 0–3-km UH to 2–5-km UH are asymmetrically (little change in 1-km relative vorticity) explained by lower magnitudes of 2–5-km UH. As a result, these cases—perhaps more typical TCT-associated cases where comparatively shallower circulations and associated dynamic amplification of

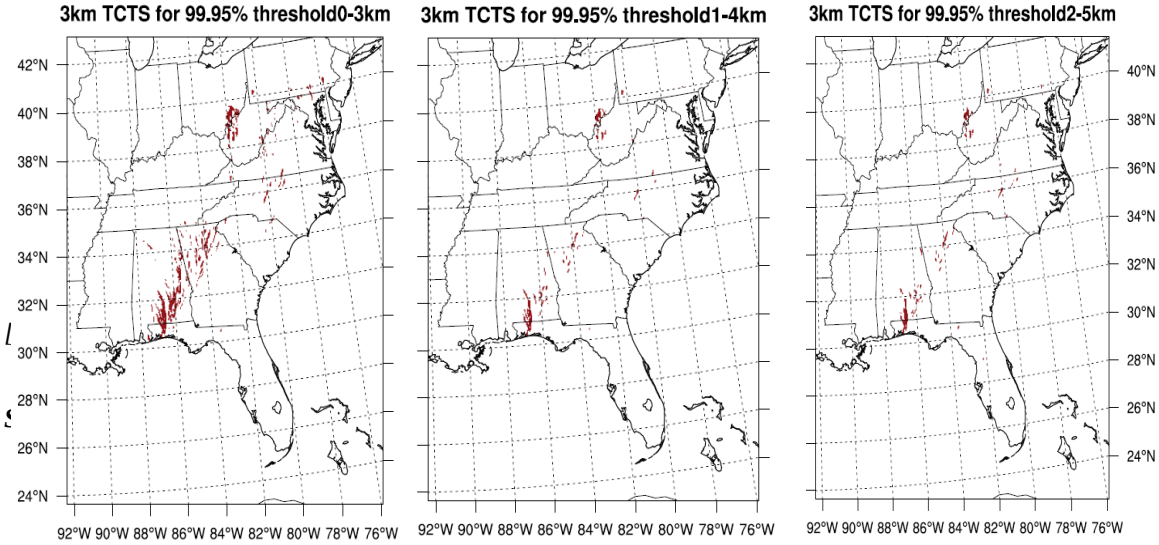
near-surface vorticity support tornado potential—may require separate investigation of UH in the lowest 3 km. In other words, variability of 2–5-km UH in the typical TC regime of weaker buoyancy may not necessarily explain the potentially critical variability of 0–3-km UH governing TCT potential. Using 2–5-km UH may work fine for Ivan, but 0–3-km UH may be much more appropriate for systems bearing less midlevel buoyancy.

Because of the aforementioned notions, I'd like to how this analysis changes for the use of 0–3-km UH in place of 2–5-km UH. I realize that this may potentially be less of an issue for Ivan—in which case that needs to be clearly stated. It needs to be clearly stated that this layer of UH analysis may be tied to the unique buoyancy distribution of Ivan, that a shallower layer of UH may be more relevant in other cases, and because of the deeper/stronger buoyancy characteristic of Ivan the resulting magnitudes of 2–5-km UH may potentially be too large if they were sought to be accurately applied to other systems. However, ultimately, I'd prefer to see how the analysis of UH in the 0–3-km layer would appear.

The authors are appreciative of your thorough explanation of the importance of assessing low level 0–3 km UH as a proxy for tornadic supercells, and the correlation of 2–5-km UH pending substantial 1-km relative vorticity. The authors put in a substantial amount of time and effort to add 0–3 km and 1–4 km to the WRF NWP diagnostics, to address this comment, and conducted an analysis on the 3-km domain to save computing resources.

As you suggested, coincidentally due to the buoyancy characteristics of Ivan being deeper and stronger than that of the "classic" tornado-producing tropical cyclones, the authors found that the magnitudes of UH in the 2–5-km layer better represented TCT surrogates in this study. The 1–4 km layer of UH showed negligible differences in TCT surrogate generation compared to the 2–5-km layer. The 0–3- km [layer] however, produced a strong positive bias, suggesting the thresholds generated from the statistics for this layer of UH were too low, resulting in an overproduction of TCT surrogates. The authors acknowledge that the results from this study are anomalous to what is typically found in other tornado producing TCs and will temper the text accordingly.

Please see figures and table below for results from this analysis.



UH_{max} level	Peak UH ($m^2 s^{-2}$)	99.95% threshold value	TCTS	PDC bias (%) (PDC_{sim}–PDC_{obs})
2–5 km	298.24	94.37	1106	0.122
1–4 km	241.74	97.23	1076	0.115
0–3 km	188.06	48.76	3399	0.674

****Keep in mind these values are based on simulations using model runs from a two-way nested grid with 9-km parent grid and 3-km nest; they will not be identical to the values listed in the manuscript since there is no 1-km grid feeding back into it.*

The choice of the MYJ PBL parameterization scheme is somewhat concerning owing to its local-type configuration. The effects of turbulent kinetic energy generated by the vertical shear of the horizontal wind can be under-represented in this type of parameterization yet can be very important for their influence in low-CAPE, high-shear severe-thunderstorm environments. Please reference Cohen et al. (2015) for additional details. Cohen et al. (2017) especially highlight the importance of using a hybrid PBL parameterization scheme in low-CAPE, high-shear southeast U.S. tornado environments—one that simultaneously incorporates the non-local effects related to shear-produced turbulence and the local effects from limited-instability-related muting of this turbulence. Granted, these works were developed for the Southeast U.S. low-CAPE, high-shear regimes and not specifically for TCTs. Regardless, there are ultimately elements of similarity between these environments. Ideally, it would be appropriate to determine whether another PBL parameterization scheme [i.e., one of the hybrid variety such as the Asymmetric Convective Model, version 2 or any of its variants as Cohen et al. (2017) describe] may result in a more realistic environment. After all, the simulated convective activity and related foundation upon which proxy thresholds are built can be undermined by a poorly simulated environment. However, at the very least, acknowledgement of these concerns needs to be addressed in the context of model configuration.

The authors fully acknowledge your recommendation; however, priority was given to the reviewer's comments addressing the ideal level of UH to analyze, to achieve optimal results for Ivan. The text will be updated, reflecting the need to address such concerns in future studies using CAMs to model tropical cyclone tornadoes.

There should be mention added, within the introduction, that there indeed exists previous research on convection-allowing-model guidance regarding thermodynamic/kinematic environments that are similar to those favoring severe thunderstorms/tornadoes with TCs. The tornado potential is ultimately governed by the convective environment in which their parent storms evolve, which can be brought about by multiple larger-scale pattern regimes. That is, while the authors' work is unique to TCT environments, similar environments have been investigated from a convection-allowing-model perspective—specifically low-CAPE, high-shear southeast U.S. tornado environments. Both Cohen et al. (2015) and Cohen et al. (2017) are very relevant in terms of simulating foundational environmental conditions and resulting convective morphologies in somewhat-similar-type regimes.

Thank you for providing this reference, the text has been updated accordingly.

There needs to be much clearer distinction made between radar-driven TCTs and actual-reported TCTs. Whether this is accomplished through a "proxy/P" prefix on the TCTs for the non-reported TCTs, or a "reported/R" prefix on the TCTs for reported TCTs, there needs to be much more deliberate stating of what corresponds to reported tornadoes and what corresponds to proxy tornadoes. I completely agree with the discussion regarding problems in the tornado database for consistent evaluation of tornadoes and corresponding verification. However, in Fig. 5, when actual tornadoes are plotted and used for comparison from Edwards (2012), there is an inconsistent application of the "actual" dataset—whether it's reported or proxy. The comparison dataset needs to be consistent throughout the manuscript, or it needs to be much more clearly stated when deviations from that dataset are made.

We apologize for the ambiguity, TCT surrogates are now annotated as TCT-S. The text was also edited to emphasis what is observed and surrogate (proxies).

Across the board from the abstract through the conclusions—and for the reasons cited in comment 1 and in the overall summary—there needs to be much more deliberate clarification that the findings of this study apply specifically to Ivan. Multiple TCs representing a variety of environments need to be investigated in order to solidify UH thresholds for assessing tornado potential from convection-allowing models. Ideally, it would be best to apply the UH thresholds found from Ivan to another system to see how they perform for a different TC—as a part of an attempt to perform an independent evaluation/validation. The use of 2–5-

km UH alone, let alone the exact magnitudes in that layer, may result in a substantially different/reduced probability of detection for tornadoes for a TC bearing much less buoyancy. However, if testing on another system were not feasible, deliberate mention of the single-case-study limitations needs to be more clearly and frequently cited throughout this work.

The 2–5-km UH may not be the best metric to capture tornadic storms in cases unlike that of Ivan. Given that testing on another case is not feasible at this time, the authors have updated the text to stress the uniqueness of this variable to Ivan.

[Minor comments omitted...]

Second Review:

Reviewer recommendation: Accept with minor revision.

General comment: I want to commend the authors for doing a truly fantastic job addressing my previously mentioned concerns. I very much appreciate their detailed evaluation of the sensitivities of layer choice for UH and description in the document. This is an excellent contribution, and I feel that the authors qualified their previous statements in a respectable manner.

I only have a few relatively minor comments, and I believe that the manuscript is very close to being ready for publication.

[Editor's Note: One comment appeared substantive enough for inclusion in the review record, below.]

In vertically sheared environments, convection is typically tilted. So, is the weakness in UH offering value for supercell threat assessment with Katrina mainly associated with the convective-scale upward accelerations being generally shallow?

Green et al. (2011) does not elaborate on this assertion other than to say UH was an insufficient metric to identify supercells, given the cells in Katrina's rainbands were "tilted and shallow". However, given your assessment of convective environments and general knowledge of TC supercells and their "miniature" characterization, the cells were likely too shallow. It is also possible Green et al. (2011) used the standard UH measured between 2 and 5 km verses using low level 0−3 km UH. Again, there is not enough discussion to give a definitive answer.

[Minor comments omitted...]

REVIEWER C (Matthew D. Eastin):

Initial Review:

Recommendation: Accept with major revisions.

Summary: A high-resolution WRF simulation of Hurricane Ivan (2004) is used to evaluate the use of updraft helicity (UH) maxima to identify potential tropical cyclone tornadoes (TCTs) for application in a short-term prognostic setting. Overall, the study is well-motivated and -written, but suffers from poor methods and validation of the overall TC structure (especially the low-level outer rainband wind fields) and representative outer rainband supercells (e.g., storm tops, mesocyclone depth, longevity, etc.). Moreover, the results are confused by counting each grid point that exhibits extreme UH values (at each model output time) as a unique TCT, rather than identifying and tracking spatially coherent TCTs over sequential output times (a method more similar to reality). Recommendations of how to address my concerns are provided below.

1. Your validation and documentation of Ivan's overall intensity and structure, as well as individually simulated outer rainband supercells could be significantly improved:

a. Given your focus on supercells and tornadoes, combined with the well-known role that low-level vertical wind shear plays in supercell/tornado evolution, greater emphasis should be placed on validating the simulated three-dimensional wind field. For example, the simulated offshore surface winds could be validated against observational HWIND analyses, while the simulated onshore winds could be validated against surface station and rawinsonde observations. Such validation would be more constructive than the qualitative comparison of observed and simulated radar reflectivity for overall storm "shape" and the simple "presence" of an outer rainband east of center. Moreover, I do not find any value in comparing the overall storm total precipitation—Fig. 4 could be replaced with a wind field comparison/validation.

If you wish to retain the validation using reflectivity, provide more quantitative results. For example, the total area occupied by dBZ >30 could be computed for both the overall storm and target outer rainband from both the observational composites and simulations. Lastly, any figure depicting the observed and simulated reflectivity fields (e.g., Fig. 3) should have the same color scale.

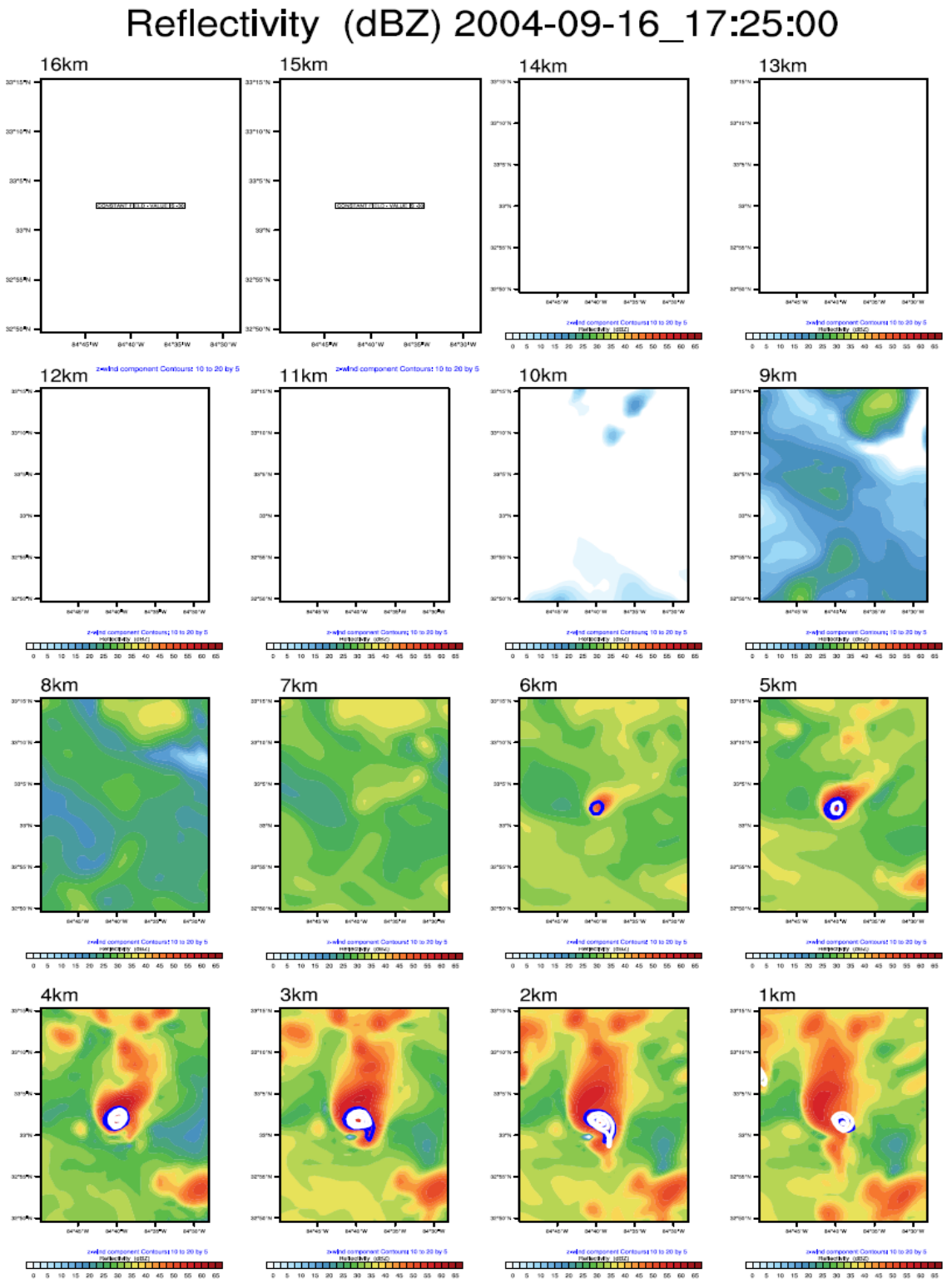
Thank you for your suggestion and the resource, however a quantitative analysis of radar reflectivity is not feasible right now. Much focus was placed on validating the use of 2–5-km UH to validate potentially tornadic storms, as a result we were not able to revisit this issue and opt to retain the current analysis as is. We did however correct the color scale of the simulated radar plots to match that of the observed composite reflectivity.

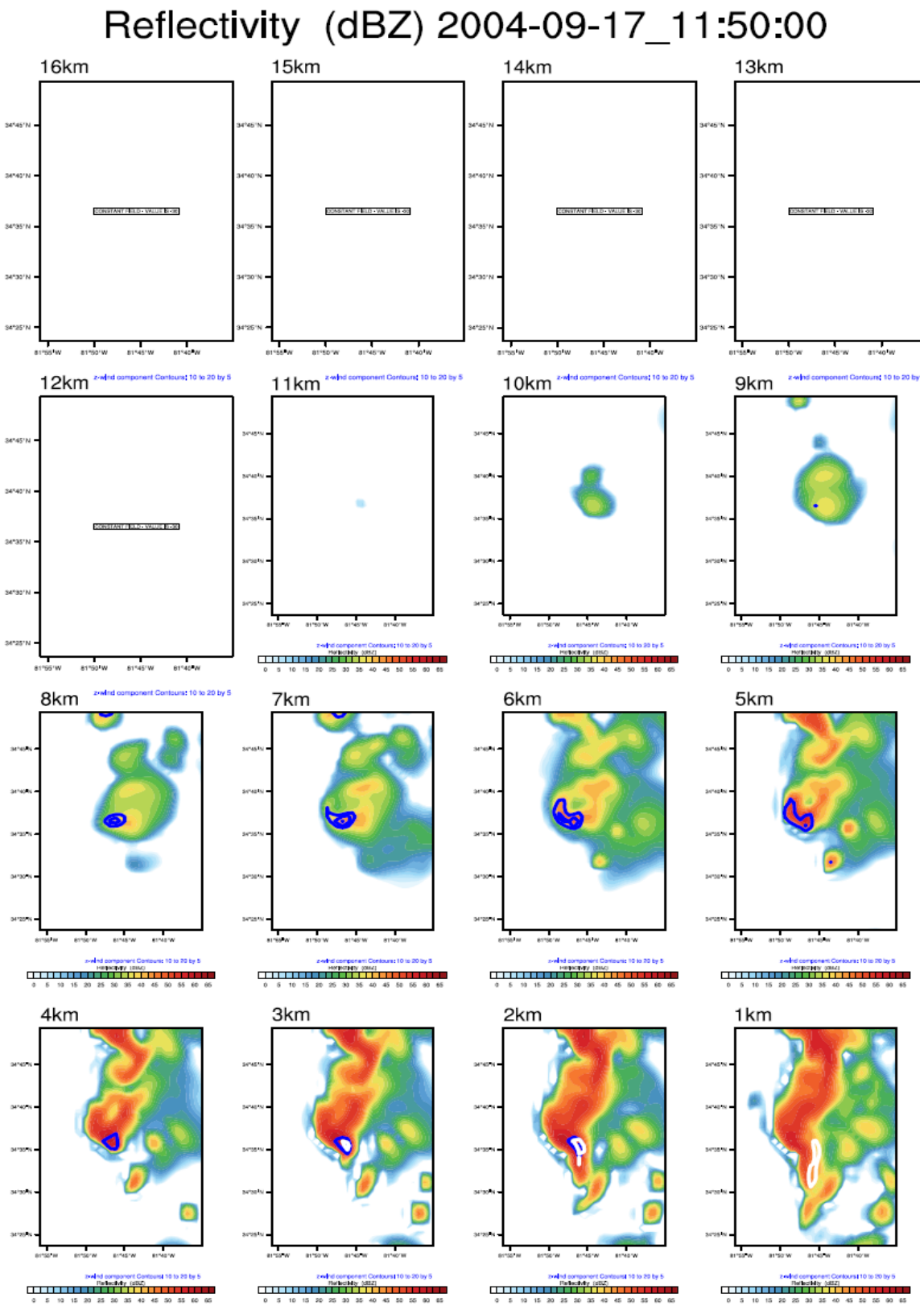
b. Tables 3, 4, and 5 would benefit from adding a column/row summarizing the same respective statistics for Ivan storm reports (i.e., the observations).

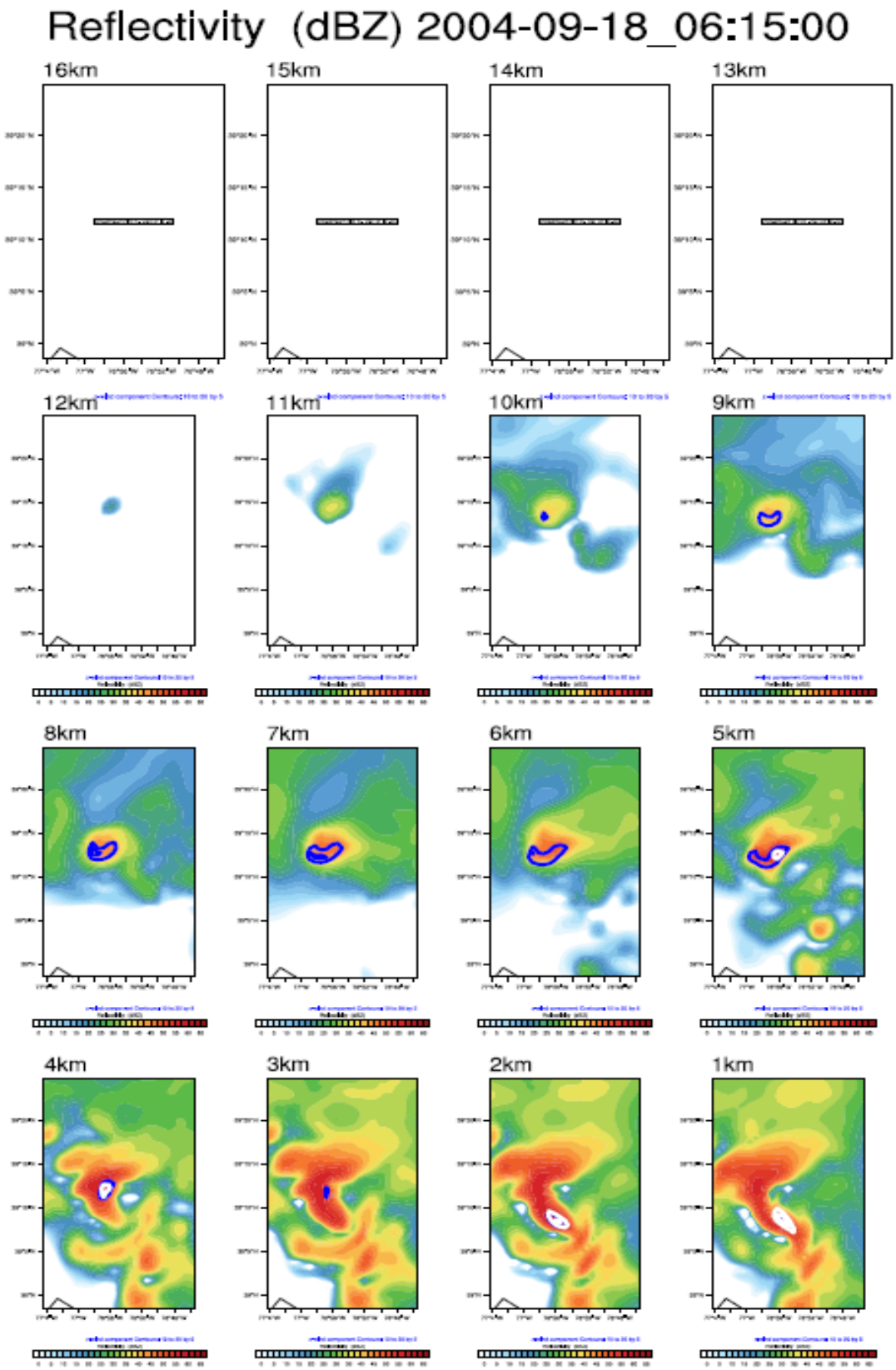
The authors have given your comment much thought, and we decided to refrain from including this information.

c. Your choice of the "representative" supercells during each time period does not provide confidence the simulation was effectively producing the often-observed and previously documented "miniature" supercells. However, you make a brief note that smaller supercells were observed (i.e., "…maximized vertical vorticity and velocity over 3–4 km in depth, with some storm tops reaching up to 10 km"). Therefore, I recommend providing examples of the apparently more common "miniature" supercells produced by the simulation.

We understand your concern and decided for the sake of brevity to focus on those particular cells. Below are a few examples of possible miniature supercells, all of which had low-level rotation at 500 m (not shown):


Reflectivity (dBZ) 2004-09-16_17:25:00
16km
15km
14km
13km
12km
11km
10km
9km
8km
7km
6km
5km
4km
3km
2km
1km


Reflectivity (dBZ) 2004-09-17_11:50:00
16km
15km
14km
13km
12km
11km
10km
9km
8km
7km
6km
5km
4km
3km
2km
1km


Reflectivity (dBZ) 2004-09-18_06:15:00
16km
15km
14km
13km
12km
11km
10km
9km
8km
7km
6km
5km
4km
3km
2km
1km

d. As briefly noted on Page 5 (RHS), your preliminary results suggested that updraft helicity (UH) was an effective metric to identify TC supercells, which contrasted with Green et al. (2011) and their analysis of simulated outer rainband supercells in Katrina. More importantly, the reasons why Green et al. found UH to be a poor identifier were not discussed—both their simulated and observed supercells were shallow and exhibited considerable vertical tilt. In your case, the high storm tops (~13 km) for each representative supercell raises concern as to whether your Ivan simulation was effectively simulating the "miniature" supercells observed in TCs.

Your concern is valid, however we were able to simulate miniature supercells and included examples, as suggested by comment 1c.

2. Your TC tornado surrogate (TCTS) identification algorithm would benefit from more carefully selected criteria that ensure the identified UH maxima are indeed strong rotating updrafts (rather than primarily a strong updraft or primarily a strong vertical vorticity maximum). For example, Morin and Parker (2011) set minimum thresholds for both vertical vorticity (>0.01 s^{-1}) and vertical velocity (>10 m s^{-1}), resulting in a minimum UH of 550 m^2 s^{-2} for identification of an outer rainband supercell. Based on your tabular statistics, only a very small fraction of your "data points" exhibited UH greater than such value. What fraction of your identified TCTS satisfy these two criteria (or even lower combined thresholds)? Also, how does the application of a lower simulated radar reflectivity threshold (i.e., less than 30 dBZ) affect your results?

UH is resolution-dependent. Morin and Parker (2011) chose that threshold based on the distribution of their UH values in their highest-resolution domain (667 m). The highest-resolution domain within the present study is 1 km, and therefore will not have UH values directly comparable to what was found in Morin and Parker. We did employ similar thresholds for vertical vorticity and vertical velocity when identifying potentially tornadic supercells.

Reducing the reflectivity threshold from 30 dBZ to 25 and 20 dBZ increased the number of TCTSs, but not significantly. The largest increase in surrogate generation was found when removing the criterion.

3. The authors appear to assume that each grid point that exhibits extreme UH at each 5-min output interval constitutes a unique TCTS, without accounting for contiguous grid points that exceed the extreme UH threshold at a given time and/or the same simulated TCTS that persists for multiple output intervals. In other words, observations of TC supercells/tornadoes suggest: a) unique supercells have horizontal dimensions of 4–6 km (based on the diameter of the updraft and/or vorticity maximum) and thus a simulated supercell (identified via UH) on a 1-km or 3-km grid should embody 4–36 contiguous grid cells; and b) tornadoes are NOT observed in "families" with each family member located 1–3 km apart. Rather than focusing on an optimal UH threshold based on a grid cell count, the authors need to focus on identifying unique (but realistic) TCTS events, and then comparing the simulated and observed event counts.

This is a very good point, however the authors do not intend to imply that each gridpoint is suggested to constitute a unique tornado, for the very reasons you described in your comment. TCT surrogates are qualitatively compared to the observed reports for their spatial location, not quantitative value. Instead, the PDC for both observed TCT tracks and simulated TCT surrogates are compared. Focusing on specific events is beyond the scope of this research at this point in time, given the lack of resolution required to explicitly resolve simulated tornado events for fair comparison to observations.

4. The rotation tracks discussed in Section 3d are tangential to this overall study and should be removed (including its methodology discussed in section 2c). All previous discussion in this manuscript focused on the numerical simulation, and then you briefly switch to an analysis of operational Doppler radar data. Moreover, the provided discussion is very weak and incomplete. For example, there is a clear range dependence to your results that was documented and discussed decades ago by Spratt et al. (1997). The logical next step would be to develop a range-dependent algorithm that would more

effectively identify rotation signatures (possible TCTs) of similar intensity but at far ranges. Therefore, a quality evaluation of using rotation tracks as an alternative source for TCT identification would be better addressed in a separate manuscript.

The authors appreciated this honest recommendation. Doing a more comprehensive evaluation of the rotation tracks data during landfalling TCs was beyond the scope of this current work. Therefore, the authors agreed to remove the discussion of the rotation tracks.

[Minor comments omitted...]

Second Review:

Reviewer recommendation: Accept.

Summary: The authors have either adequately addressed my previous concerns or explained why addressing such concerns were beyond the scope of the current study. The revised manuscript now provides a concise, case-specific example of how CAMs may provide useful forecast information regarding TC tornadoes once additional detailed simulations and verifications can be conducted. I have no further significant concerns at this time, and I support its publication.