Weather and Forecasting

Predictability of Tropical Cyclone Intensity Evaluated through 5-year Forecasts with a Convection-permitting Regional-scale Model in the Atlantic Basin

--Manuscript Draft--

Manuscript Number:	WAF-D-13-00085
Full Title:	Predictability of Tropical Cyclone Intensity Evaluated through 5-year Forecasts with a Convection-permitting Regional-scale Model in the Atlantic Basin
Article Type:	Article
Corresponding Author:	Zhiyong Meng School of Physics, Peking University Beijing , CHINA
Corresponding Author's Institution:	School of Physics, Peking University
First Author:	Yunji Zhang
Order of Authors:	Yunji Zhang
	Zhiyong Meng
	Fuqing Zhang
	Yonghui Weng
Abstract:	The practical predictability of tropical cyclone (TC) intensity in terms of mean absolute forecast error with respect to different conditions at forecast initialization was explored through convection-permitting hindcasts of all Atlantic storms during the 2008-2012 hurricane seasons using the Weather Research and Forecast (WRF) model. Averaged over a total of 2190 simulations, the day 1-5 performance of these WRF hindcasts was comparable to two operational regional-scale hurricane prediction models used by the National Hurricane Center (NHC) but was slightly inferior to the NHC official forecasts. It was found that the prediction accuracy of TC intensity, both at the initialization time and the targeted forecast hours, was strongly correlated with the TC intensity. On average, for both the WRF hindcasts and the NHC official forecasts, stronger intensities and larger intensity variations led to larger forecast errors. A number of synoptic-scale environmental parameters, such as vertical wind shear, sea surface temperature (SST) and the underlying surface condition (land vs. sea), affected the intensity forecast errors of TCs in part due to their influence on intensity changes, while other thermodynamic environmental parameters such as moisture and instability had relatively minor effect. The accuracy of the intensity prediction was also found to be sensitive to the translation speed of the TCs. A moderate TC translation speed of 11-15 knots (kt) corresponded to the largest intensity errors during forecast lead-times less than 60 h, while the slowest translation speed (< 7 kt) was associated with the largest errors after the 60-h forecast lead-time.

Predictability of Tropical Cyclone Intensity Evaluated through 5-year Forecasts with a Convection-permitting Regional-scale Model in the Atlantic Basin

Yunji Zhang[1, 2], Zhiyong Meng[1], Fuqing Zhang[2], and Yonghui Weng[2]

[1] Laboratory for Climate and Ocean-Atmosphere Studies, Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing, China

[2]Department of Meteorology, Pennsylvania State University, University Park, Pennsylvania

Submitted to Weather and Forecasting

April 23, 2014

Corresponding author address: Dr. Zhiyong Meng, Laboratory for Climate and Ocean-Atmosphere Studies, Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing, China.

E-mail: zymeng@pku.edu.cn

Abstract

The practical predictability of tropical cyclone (TC) intensity in terms of mean absolute forecast error with respect to different conditions at forecast initialization was explored through convection-permitting hindcasts of all Atlantic storms during the 2008–2012 hurricane seasons using the Weather Research and Forecasting (WRF) model. Averaged over a total of 2190 simulations, the day 1–5 performance of these WRF hindcasts was comparable to two operational regional-scale hurricane prediction models used by the National Hurricane Center (NHC) but was slightly inferior to the NHC official forecasts.

It was found that the prediction accuracy of TC intensity, both at the initialization time and the targeted forecast hours, was strongly correlated with the TC intensity. On average, for both the WRF hindcasts and the NHC official forecasts, stronger intensities and larger intensity variations led to larger forecast errors. A number of synoptic-scale environmental parameters, such as vertical wind shear, sea surface temperature (SST) and the underlying surface condition (land vs. sea), affected the intensity forecast errors of TCs in part due to their influence on intensity changes, while other thermodynamic environmental parameters such as moisture and instability had relatively minor effect. The accuracy of the intensity prediction was also found to be sensitive to the translation speed of the TCs. A moderate TC translation speed of 11–15 knots (kt) corresponded to the largest intensity errors during forecast lead-times less than 60 h, while the slowest translation speed (< 7 kt) was associated with the largest errors after the 60-h forecast lead time.

1. Introduction

It is well known in the tropical cyclone (TC) research and operational forecast communities that track forecasts have experienced large improvements during the past few decades, while there has been virtually no improvement in the intensity forecasts in both the North Atlantic (Cangialosi and Franklin 2013; Houze et al. 2007) and western North Pacific (Yu et al. 2013) basins, especially during stages of rapid intensity change (Elsberry et al. 2007). TC track is primarily controlled by the large-scale environment, including for example the steering flow (Chan and Gray 1982), β-effect (Holland 1983) and Fujiwhara effect (Fujiwhara 1921), and better track predictions have been achieved with the advances in numerical weather prediction (NWP) models, observing systems and data assimilation methods. The intensity, on the other hand, is mainly determined by the internal dynamics and moist processes. These dynamics and processes are smaller in scale and more chaotic, making intensity less predictable.

The predictability of weather systems consists of intrinsic and practical predictability (Lorenz 1963), and the lack of improvement in the accuracy of intensity forecasts has resulted from both the limited intrinsic predictability of underlying dynamics and the limited practical predictability due to deficiencies in the current generation of intensity-forecast guidance tools. The practical limitations arise from insufficient model resolutions, inaccurate initial conditions (ICs) and uncertainties in the representations of various physical processes. Using finer convection-resolving resolutions in NWP models, applying advanced ensemble-based or hybrid ensemble-variational data assimilation methods and assimilating inner-core observations such as ground-based or airborne Doppler radar radial velocity observations will greatly mitigate the practical predictability limitation and improve intensity forecast accuracy (Zhang et al. 2009, 2011; Weng and Zhang 2012; Li et al 2012; Aksoy et al. 2013; Cavallo et al. 2013). Davis et al.

(2010) found a statistically significant improvement in TC intensity forecasts using a nested domain of 1.33-km resolution rather than 12-km in an NWP model. Based on real-time analyses and forecasts using an EnKF and the advanced Hurricane WRF from the 2009 North Atlantic hurricane season, Cavallo et al. (2013) found that the EnKF data assimilation can systematically reduce the TC position and intensity errors except for strong TCs. Because of the chaotic nature of moist convection and internal dynamics that dominate TC intensity, it may be intrinsically less predictable than track. Zhang and Sippel (2009) found that small unobservable initial condition perturbations could lead to large divergence in TC forecasts, as similarly observed in studies concerning the mesoscale predictability of continental mesoscale convective systems (Melhauser and Zhang 2012; Wu et al. 2013) and midlatitude extratropical cyclones (Zhang et al. 2002, 2003). Van Sang et al. (2008) found that small random moisture perturbations in the boundary layer might greatly change the structure and intensity of TCs, implying large intrinsic uncertainties associated with TC intensity prediction. Using similar initial perturbations but with the inclusion of environmental vertical wind shear, Zhang and Tao (2013) found that larger magnitudes of shear decrease the intrinsic predictability of the TC, especially during periods of genesis or rapid intensification (RI). Furthermore, Hakim (2013) and Brown and Hakim (2013) explored the possible time scale of intrinsic predictability of several TC characters under an idealized equilibrium framework and found that most features near eyewall region maintained their predictability for no longer than 48 hours without considering the ambient environment. It is noteworthy that most studies on TC intensity predictability focused on intrinsic predictability of single or idealized cases. The factors controlling intensity forecast error growth in realistic or operational NWP models remain largely unknown.

Although as stated in Brown and Hakim (2013) that intrinsic predictability of many TC characteristics is lost after about two days, with the help of realistic dynamical models as well as statistical models, another major branch in TC operational forecast community, forecasts up to 120-h are routinely issued at various major operational centers, including the National Hurricane Center (NHC). One of these statistical models, the Statistical Hurricane Intensity Prediction Scheme (SHIPS; DeMaria and Kaplan 1994), uses a combination of climatological, persistence and synoptic parameters to provide intensity forecasts of Atlantic and eastern North Pacific TCs (DeMaria and Kaplan 1999). This model has also been updated to include effects of land surface after landfall [decay SHIPS (DSHIPS); DeMaria et al. 2005]. Another widely used statistical model, the 5-day Statistical Hurricane Intensity Forecast Model (SHF5; Knaff et al. 2003), uses a combination of only climatological and persistence predictors and their products. The performance of SHF5 is generally inferior to SHIPS (Elsberry et al. 2007), possibly resulting from the exclusion of synoptic parameters in SHF5, which implies the importance of considering synoptic conditions in predicting TC intensity. Although statistical models have provided valuable guidance in operational forecasts, these models still lack skill in predicting tropical cyclogenesis and RI (Elsberry et al. 2007). According to surveys, about 31% of all TCs that formed in the Atlantic basin between 1989 and 2000 underwent RI [defined as an intensity increase of more than 30 knots (kt) within a 24-h period] at least once during their lifetimes (Kaplan and DeMaria 2003), and 6% of Atlantic TC 24-h intensity changes were greater than 30 kt (Kaplan et al. 2010). Thus, the capability of accurately predicting RI is essential for intensity forecasting. Another statistical model designed specifically for intensity predictions of RI applied different combinations of predictors with respect to TC intensity (hurricane category) and stage of RI (Law and Hobgood 2007), indicating that the controlling factors of TC intensity may vary

during different stages. With all these statistical models demonstrating various factors that may affect TC intensity, it is worth exploring whether these factors, such as wind shear, moisture and translation speed, might also impact intensity forecast errors. Most recently, Bhatia and Nolan (2013) examined statistically the uncertainties of TC intensity forecasts with respect to the synoptic environment conditions. They focused on the forecast uncertainty in the NHC's operational forecast guidance models (both statistical and dynamical) as well as in the NHC official forecasts. Given the configurations of operational guidance models often change from year to year, one potential limitation of their study is that some of the forecast sensitivities may have come from yearly variations in the model rather than the true forecast uncertainties due to environmental factors.

This article focuses on using one specific NWP model with the same configuration throughout a 5-year period. We seek to explore the mean absolute forecast error characteristics from day one to five with respect to different factors at simulation initialization time. Understanding practical intensity predictability of TCs in an NWP model can help us not only examine possible causal links between dynamical and thermodynamical processes within TCs, but may also provide guidance to forecasters. Further knowledge on the less predictable TCs will help us to find solutions to improve their forecasts through the development of better models, better data assimilation and vortex initialization methods and/or the deployment of better observations. Section 2 introduces the NWP model and methodology used in this study. After comparing the performance of the models in section 3, a statistical analysis of mean absolute intensity forecast errors is provided in section 4. Finally, section 5 provides the conclusions of this study.

2. Model, data and methodology

The Advanced Research (ARW) core of the Weather and Research Forecasting (WRF) model, version 3.4.1 (Skamarock et al. 2008), was used in this study. The three two-way nested domains had 379×244 (D01), 304×304 (D02) and 304×304 (D03) horizontal grid points with resolutions of 27 km, 9 km and 3 km, respectively. D01 was fixed to cover the central to eastern three-quarters of the CONUS, tropical and subtropical North Atlantic as shown in Fig. 1. All Atlantic TCs named by the NHC during 2008–2012, especially the formation and intensification stages of these storms, occurred inside D01, which was large enough to minimize the influence of lateral boundary conditions (BCs) in most cases while remaining computationally affordable. The inner domains (D02 and D03) moved automatically following the location of the TC vortex. There were 44 terrain-following hydrostatic-pressure levels vertically, decreasing to 50 hPa at the top. The time step used for the model integration was 90 s in the outermost domain. Physical parameterization schemes included the Grell–Devenyi ensemble scheme for cumulus parameterization (Grell and Devenyi 2002; only in D01), the WRF single-moment (WSM) 6-class scheme for microphysical processes (Hong and Lim 2006), the Yonsei University (YSU) planetary boundary layer scheme (Hong et al. 2006) with the fifth-generation Penn State/NCAR mesoscale model (MM5) similarity surface layer scheme (Zhang and Anthes 1982) and the MM5-based 5-layer thermal diffusion land-surface model (Duhdia 1996), the Rapid Radiative Transfer Model (RRTM) scheme (Mlawer et al. 1997) for longwave and the Dudhia scheme (Duhdia 1989) for shortwave atmospheric radiation. An empirical scheme implemented in Green and Zhang (2013; denoted therein as the "PSU" scheme) was used to estimate the bulk drag and enthalpy coefficients (C_D/C_K). This ad-hoc scheme has been found to be effective in improving the TC wind-pressure relationship forecasts.

During the years of 2008–2012, there were 17, 11, 21, 20 and 19 TCs respectively, which totaled to 88 TCs in the North Atlantic basin. These numbers were based on the Tropical Cyclone Vital Database (TCVitals), which contains TC location, intensity and structure information and were generated in real time every 6 h by forecasters (Trahan and Sparling, 2012). A 126-h deterministic forecast was conducted when a TCVitals report was issued, with ICs and BCs provided by the analysis and forecasts of the operational Global Forecast System (GFS) at the forecast initialization time, which had a horizontal resolution of 0.5°×0.5°. Sea surface temperature (SST) condition is included in the GFS analysis at model initialization and was held constant over the 5-day forecast period. A total of 2190 hindcast cases were collected and used for the statistical analysis in this work, and will be referred to as the ANPS (*A*rw *N*o data assimilation forecast by the *P*enn *S*tate university) forecasts hereafter, which is also the identification acronym in the Automatic Tropical Cyclone Forecast system (ATCF; Sampson and Schrader 2000) designated by NHC for the Penn State University experimental real-time forecasts during 2011–2012. Note that because of the coarse resolution of the ICs that directly come from the GFS analysis, the initial TCs often differed from the real storms with weaker intensity and smoother structure, and experienced rapid adjustment for the first several hours of model integration. Thus the nested model domains (D02 and D03) were fixed in location during 0–6 h and began moving with the TC center afterwards.

The maximum sustained 10-m wind speed (V_{max}) of each TC was chosen to represent TC intensity. Comparisons between results using minimum central sea level pressure (P_{min}) and V_{max} showed general consistencies. Intensity and positions of TCs were determined by the Vortex Tracker program of the Geophysical Fluid Dynamics Laboratory (GFDL) model (Gopalakrishnan et al. 2012) every 6 h. The Vortex Tracker calculates the position based on the

average of extrema of several parameters of the model forecasts in the vicinity of an input first guess position. These parameters include relative vorticity and geopotential height at 850 hPa and 700 hPa, wind speed at 850 hPa, 700 hPa and 10 m, and mean sea level pressure (MSLP). The first guess position was determined by TCVitals at the model initialization time, and by a weighted average of advecting the previous TC center using the 500-hPa, 700-hPa, 850-hPa wind speed from previous position during the model forecast.

In order to calculate model forecast errors, the NHC post-storm "best-track" analysis (referred to as BEST) from ATCF was used as the observations. Two other major multi-layer regional dynamical models that are operationally running at NHC, the National Weather Service (NWS)/GFDL model (Bender et al. 2007) and the NWS/Hurricane WRF (HWRF) model (Gopalakrishnan et al. 2012) were used for verification of the performance of the ANPS that will be presented in the next section. The NHC official forecasts were also examined (referred to as OFCL). GFDL, HWRF and OFCL forecasts were all acquired through ATCF. Note that ANPS, GFDL and HWRF generated intensity and position information every 6 h during the 126-h forecast, while OFCL forecasts were available every 6 h during 0–12 h, every 12 h during 12–48 h, and every 24 h during 48–120 h.

For all results, the statistical significance was verified using the bootstrap method (Wilks 2006). In this method, to verify the difference between the mean values of two groups that have a and b members respectively, a new group of a members and a new group of b members are randomly selected with replacement from the $(a+b)$-member union of the original two groups, and the difference of the mean values of the two new groups is recorded. This procedure was repeated for 10000 times, and the original difference that was to be verified was regarded as significant if it achieved 0.05 level (95% confidence; when its value falls outside the 2.5th–

97.5th percentiles of the 10000-difference distribution) in our study. Significances were verified for all pairs of two individual groups within one categorization. For the remaining part of this article, "significant" and "significance" are all referred in a statistical sense.

3. Model performance

Before further analyzing the ANPS forecasts, it is desirable to examine whether the NWP model used can provide dynamically reliable TC forecasts with an acceptable error magnitude. Fig. 2a shows that the ANPS track forecasts were comparable in accuracy to the operational forecasts by the HWRF and GFDL models as well as OFCL for 2008–2012. The bootstrap test confirmed that the track accuracies of these forecasts were statistically indistinguishable. However, since the ANPS forecasts with the WRF model were directly initialized with the coarse GFS analysis without additional TC vortex initialization[1], the intensity forecast errors of ANPS were larger than HWRF and GFDL at first (Fig. 2b), and the differences between ANPS and HWRF/GFDL were significant untill 24 h. The OFCL intensity forecast error saturated around 13–15 m s^{-1} after approximately 48 h. On the contrary, the errors of the dynamical model forecasts continuously increased over time, mostly due to the continuously increasing forecast biases (Fig. 2c). It is also worth noting that the ANPS forecast achieved a decent wind-pressure relationship (often used as a measure of dynamical reliability), as the least-square linear fit between the forecasted V_{max} and P_{min} closely followed those of HWRF, GFDL and the best-track estimates (Fig. 2d).

[1] However, a TC vortex relocation scheme is used in the operational GFS analyses that are used to initialize the ANPS hindcasts (Liu et al. 2000).

Despite the configuration uniformity, there was a strong variability in the intensity forecast errors from year to year by ANPS (Fig. 3a). The accuracy of the OFCL forecasts also experienced strongly year-to-year variability though not always in the same direction as ANPS (not shown). Some of the ANPS forecast performance variability might come from the strongly year-to-year variability in the forecast biases (Fig. 3b). All yearly averaged intensity biases increased over forecast lead-times, and over the five years the biases generally shifted from being negative to positive (intensity forecasts from being weaker to stronger than best-track estimates). Some of these yearly shifts in forecast biases might be a reflection of the changes in the GFS analyses and forecasts (both in terms of the forecast model and the data assimilation system; NCEP 2013) that were used as ICs and BCs for the ANPS forecasts. The yearly variability in the ANPS forecast biases clearly revealed the influence of global analysis and forecasts that were used to initialize regional scale models: although we strived to maintain uniformity of the ANPS hindcasts, the simulations were nonetheless sensitive to changes and quality of its ICs and BCs.

4. Dependence of intensity forecast error on various storm-scale and environmental factors

a. TC intensity

First, we examined how different initial TC intensities might result in different characteristics of intensity forecast error growth. Each ANPS forecast was categorized according to the Saffir-Simpson hurricane wind scale based on the best-track observations at the initialization time. Fig. 4a shows the mean absolute forecast error growth of intensity categorized into tropical depressions and tropical storms (TD and TS; V_{max} < 64 kt), non-major hurricanes

(category 1 and 2 hurricanes, 64 kt ≤ V_{max} < 96 kt), and major hurricanes (category 3 to 5, V_{max} ≥ 96 kt) as well as the 24-hourly first and third quartiles of each group. The spin-up periods were clearly revealed, with much larger initial error for major hurricanes and decreasing to a similar magnitude of TD and TS after about 72 h. On the contrary, the intensity forecast errors for TD/TS increased with time. The intensity forecast errors for non-major hurricanes also increased with time but at a lesser growth rate. Major hurricanes had the potential to have larger forecast errors at least for 0–60 h, and this difference was significant for 0–48 h. In OFCL, major hurricanes contained larger errors during 12–96 h (Fig. 4b), and were significant untill 72 h. When the mean absolute errors were normalized into mean relative error percentages (mean of absolute intensity error divided by best-track intensity), the differences among groups in OFCL became almost statistically indistinguishable throughout the 5-day period (Fig. 4d). However, the percentage-based intensity forecast error of ANPS was different among forecasts with different initial intensity (Fig. 4c). For example, the intensity forecast error for initially major hurricanes was significantly larger than the other two groups at the beginning of the forecast period, partly due to model initialization and subsequent spin-ups, and quickly decreased to near a quasi-steady value of ~ 20% at around 24 h. This error subsequently became significantly smaller than the forecasts initialized with weaker initial TCs after 90 h.

Fig. 5 further stratifies the forecast errors according to the best-track TC intensity at each verification time. It is quite apparent that stronger TCs (at the verification time) had larger intensity forecast errors both for ANPS (Fig. 5a) and OFCL (Fig. 5b) at almost all lead times. This characteristic was significant during 0–102 h for ANPS and 12–120 h for OFCL. Given initially weaker TCs on average tend to strengthen to stronger storms, while initially stronger TCs tend to weaken (Fig. 6a), the strong dependence of forecast error on the intensity of the

observed storms at the verification times shown in Fig. 5a was thus consistent with the dependence of forecast error on the initial storm intensity shown in Fig. 4a. On the other hand, the percentage-based forecast errors (categorized by the concurrent intensity) for both ANPS and OFCL (Figs. 5c and 5d) had much larger variabilities among groups than those categorized by the initial intensity shown in Figs. 4c and 4d. For ANPS, major hurricanes had significantly larger errors for 0–54 h, while the weakest TD and TS became the significantly largest after 78 h. OFCL tended to have persistently larger error magnitudes for TS and TD than hurricanes (Fig. 5d; significant after 12 h), which might be in part due to the difficulty in forecasting rapid intensity changes. Given the absolute error remains the primary verification metrics in various studies (e.g., Elsberry et al. 2007, Bhatia and Nolan 2013) and operational centers including the NHC, the mean absolute errors were used in the remainder of this study to make it more comparable to other studies.

Next, we examined the dependence of intensity forecast error on the initial bias. Since there was a sharp decrease of overall intensity forecast error for ANPS (Fig. 2b) during the first 6 h of simulations, the initial bias for the ANPS forecasts was therefore considered to be the bias at t = 6 h. Not surprisingly, larger initial intensity biases resulted in larger forecast errors at most times. The noteworthy result in the ANPS forecasts in Fig. 7a is that forecast errors with initially positive bias (stronger than the observations) larger than 10 kt decreased quickly after initialization and became indistinguishable from errors in TCs with much smaller initial biases. Meanwhile, the forecast errors with initially large negative bias (blue lines in Fig. 7a) remained significantly larger than the other errors until 54 h. However, with a much smaller initial error (by human forecasters), the initial biases in OFCL generally did not affect the subsequent OFCL forecasts (Fig. 7b). The much longer-sustained larger errors of TCs with initially large negative

biases could result from a strong TC with longer time for model spin-up or a TC experiencing RI, which will be examined next.

RI has long been recognized as one of the most challenging elements of TC prediction. Both dynamical and statistical models lack sufficient skill in accurately forecasting both the timing and magnitude of RI (Elsberry et al. 2007). RI herein was defined as the process in which the V_{max} of a TC intensifies by more than 30 kt within a 24-h period as in Kaplan and DeMaria (2003) and used operationally by NHC. Based on this definition, we divided the forecast dataset into five groups based on the observed ΔV_{max} during the first 24 h of forecasts. These groups were RI ($\Delta V_{max} > 30$ kt), normal intensification (30 kt > $\Delta V_{max} > 5$ kt), maintenance (5 kt > $\Delta V_{max} > -5$ kt), normal decay (-5 kt > $\Delta V_{max} > -30$ kt) and rapid decay ($\Delta V_{max} < -30$ kt). The results showed limited practical predictability of RI for ANPS model, as the intensity forecast error experienced an increase of over 20 kt within the first 24 h (Fig. 7c). The forecast errors of RI TCs were significantly larger than all other TCs during 12–66 h, while the forecast errors of rapidly decaying TCs decreased dramatically after initialization and remained comparable to (or slightly smaller than) others. A similar practical predictability limitation of TCs undergoing RI also appeared in the OFCL forecasts (Fig. 7d); those TCs had significantly larger errors during 24–48 h. These characteristics might result from the fact that rapid decay occurs more frequently in strong TCs that may already have large errors at initialization, while RI happens mostly in weak TCs. This was confirmed by Fig. 6b during 0–24 h, when the RI TCs on average intensified by nearly 40 kt from TS strength, while the rapidly decaying TCs weakened by a similar magnitude from a category-2 hurricane.

The forecasts that underwent RI were further investigated. All cases that went through RI during the 126-h simulation had been categorized by the forecast lead-time when their RI began

(e.g. forecasts in category “Day 1” began their RI in forecast lead-times of 0–24 h). Only the earliest RI start time was counted for each forecast. It is clear that the pattern of intensity error of ANPS (Fig. 7e) resembled the pattern of best-track intensity (Fig. 6c). Furthermore, the peak errors of all groups during their respective RI were also significantly larger than those before RI at the same time. OFCL was similar to ANPS (Fig. 7f) for RI before 48 hours, but the error of RI after 48 hours ("Day 3" to "Day 5" in the figure) for OFCL behaved differently from those of ANPS even considering the much coarser time resolution of 24 h of OFCL, and RI after 96 h ("Day 5" in the figure) had much larger errors in OFCL than in ANPS. Strong correlation between intensity change and error growth of ANPS throughout the entire forecast period indicates that large errors during RI periods are not due to the spin-up effect of ANPS (bias resulting from initializing the TC directly from the GFS analysis), otherwise intensity forecast errors would be smaller in subsequent RIs beyond the spin-up periods. Also, the much larger intensity errors after RI than before might be a consequence of the increased intensity as well as the generally increasing trend of intensity forecast errors during the simulations (Fig. 2b).

b. Environmental Factors

1) Vertical wind shear

Vertical wind shear (referred to as shear hereafter) is one of the most influential environmental parameters throughout the entire lifetime of a TC. It affects tropical cyclogenesis and RI (Molinari et al. 2004; Molinari and Vollaro 2010b; Nguyen and Molinari 2012; Zhang and Tao 2013), TC structure (Cavallo et al. 2013) and more specifically the distribution or asymmetries of convection (Corbosiero and Molinari 2002, 2003; Molinari and Vollaro 2010a; Reasor et al. 2013) and precipitation (Chen et al. 2006; Gao et al. 2009; Wingo and Cecil 2010),

as well as changes in TC intensity (DeMaria 1996; Molinari et al. 2006; Zeng et al. 2010). However, Zehr (1993) argued that the influence of shear might not be reliably quantified and consistently related. Here we examined the impact of shear on TC intensity prediction and practical predictability within both ANPS and OFCL.

The 1°×1° GFS final analysis (FNL) was used to calculate environmental shear. The FNL analysis is based on the GFS analyses, but assimilates more observations after the synoptic time using a +3-h cut off window (rather than a +1-h cut off window for the operational GFS analysis). Many different ways to average wind speed at a specific isobaric level had been utilized in the aforementioned literature. In this study, the deep-layer shear was calculated as the difference between the mean wind vectors within an annular area between 200 and 800 km from the TC center at 850 and 200 hPa (Table 1), which had been widely used in previous studies (e.g., Gao et al. 2009).

The mean TC intensity evolution categorized by the amplitude of shear at forecast initialization times for the subsequent 5 days of the forecast is given in Fig. 6d. The average shear at the initial time for all storms was 8.84 m s^{-1} (Table 2). Somewhat surprisingly, the mean initial intensity of the TCs (Fig. 6d) for the weak to moderate shear values (0–10 m s^{-1}) was the weakest (around 50 kt) while the storms with strong shears (> 10 m s^{-1}) were more intense on average (55–60 kt). This shows that the instantaneous shear values were not strongly correlated with the instantaneous TC intensity. Nevertheless, the initial shear amplitude was directly related to the subsequent mean intensity changes. The storms with averaged initial shear < 5 m s^{-1} intensified the fastest, followed by storms with initial shear of 5–10 m s^{-1}, both of which approached a near steady intensity between 48–72 h. The TCs with the largest initial shear (> 15 m s^{-1}) on average had the largest decrease in intensity, followed by the storms with shears of 10–

15 m s^{-1}, both of whose average intensities changed from decaying to strengthening at around 72–96 h. The strengthening of these strong initially sheared storms might result from the vortex precession/alignment processes that counter against the environmental shear (e.g., Rappin and Nolan 2012; Zhang and Tao 2013); it is also possible that shear changed with time.

The effect of the initial shear on the intensity forecast error was rather complicated (Figs. 8a and 8b). The TCs with initial shear of 0–5 m s^{-1} had the largest intensity forecast error by ANPS (until about 84 h), and its differences from the other groups were significant during 18–72 h. The storms that had the largest initial shear (> 15 m s^{-1}) on average had the smallest intensity forecast errors at nearly all times, even though the error differences might not always be significant. However, a 10–15 m s^{-1} strong shear produced larger errors than the moderate 5–10 m s^{-1} shear in ANPS, although their difference was statistically indistinguishable for 15 out of the 22 verified forecast lead-times (Fig. 8a). OFCL also had the least forecast errors for initial shear > 15 m s^{-1} that was significant during 24–96 h, while its errors for weak (0–5 m s^{-1}) and moderate (5–10 m s^{-1}) shears were statistically indistinguishable (Fig. 8b).

The above findings on practical predictability with respect to initial shear were complementary to the recent studies of Zhang and Tao (2013). They found that stronger deep-layer environmental shear leads to smaller intrinsic predictability of tropical cyclone intensity during RI. An extension of their work (Zhang and Tao, personal communication) revealed similar conclusions for shear magnitude up to 12.5 m s^{-1}. Because Zhang and Tao (2013) applied an idealized framework that explored only the impacts of shear and focused primarily on the intrinsic predictability of a TC's formation and RI (the NHC best track database does not include any genesis forecasts, and thus neither ANPS or OFCL forecasts demonstrate the difficulty of producing genesis forecasts under different shear), the discrepancy of the impact from

environmental shear on TC intensity predictability between these two studies is expected. In addition, the error characteristics under different shear conditions (Figs. 8a and 8b) were different from corresponding intensity evolutions (Fig. 6d) as decay of TCs under shear > 10 m s^{-1} did not lead to a decrease of intensity forecast errors, which indicates further links between shear and predictability of TC intensity. However, this is beyond the scope of this study.

2) Sea surface temperature (SST) and center latitude

Sufficient SST has long been recognized as a prerequisite of TC formation and maintenance since the sea surface is its major energy source (Emanuel 1986). It has been proven to be an important factor for both TC track and intensity forecasts as well (Kunii and Miyoshi 2013). The National Oceanic and Atmospheric Administration (NOAA) Optimum Interpolation (OI) 1/4 Degree Daily Sea Surface Temperature Analysis (OISST; Reynolds et al. 2007) produced from the Advanced Very High Resolution Radiometer (AVHRR) aboard the NOAA-series polar orbiting satellites was used in this study. If more than half the grid points within a 200-km radius from the TC center were located over the sea, SST values of these grid points were averaged to give the SST value (Table 1).

TCs under different initial SSTs generally had similar initial intensities of 50–55 kt, or medium TS intensity (Fig. 9a). However, the storms subsequently underwent different intensity changes. Not surprisingly, environments with SSTs lower than 27 °C were not favorable for the intensification of TCs. On the contrary, warmer SSTs fueled the intensification, except for SSTs higher than 29 °C where TCs became weaker after 48 h. In agreement with previous results that stronger TCs have larger errors, the strongest TCs with SSTs of 28–29 °C generally had the largest errors, although its differences in error amplitude from storms with SSTs of 27–28 °C

were not significant (Fig. 8c). A higher or lower SST than these values both resulted in smaller errors, but only SST lower than 27 °C had significantly lower errors between 18–66 h. The distribution of SST had a mean value of about 27 °C (Table 2) with a long tail in the distribution towards the lower temperatures (with the coldest SSTs of 14 °C), and an even lower SST would further increase the practical predictability (figure not shown). This might be related to extratropical transitioning (ET) systems moving northward over cold sea surfaces at higher latitudes. Forecast errors of OFCL also showed a substantial increase of error with respect to warmer SSTs (Fig. 8d), although only the coldest conditions continuously passed the significance test after 12 h.

Categorization by initial latitude of forecasts (Fig. 8e) showed similar characteristics to categorization by initial SST. TC forecasts from 15–25 °N had the largest forecast errors, and moving both poleward and equatorward in initial position could reduce the error, although only errors of the northernmost cases during 18–60 h were significantly different. OFCL categorized by initial latitude (Fig. 8e) was also similar to its categorization by initial SST (Fig. 8d) with decreasing errors resulting from increasing latitude. However, the partial correlation coefficient[2] between latitude and forecast errors while holding SST constant for ANPS was no larger than 0.05 at all times after initialization, and the time-average is -0.0005, indicating that latitude effects on intensity forecast errors are nearly solely due to the changing SSTs with respect to latitudes.

[2] When two variables A and B are both affected by other variable(s), the partial correlation coefficient between A and B measures the degree of their association with removing the effect of other variable(s).

3) Convective parameters

Doswell et al. (1996) proposed three necessary ingredients for deep convection: lifting, moisture and instability. Since TCs are convective phenomena and moist convection has been proven to play a crucial role in limiting the intrinsic predictability of TC intensity (e.g. Zhang and Sippel 2009; Zhang and Tao 2013), effects of moisture and instability on intensity forecast errors were explored. The FNL analysis was again used to calculate moisture and instability parameters. Moisture was averaged over an annular area of 200–800 km from the TC center (same as calculation of shear) to provide an estimation of the environment. For instability parameters, an observational analysis in Molinari et al. (2012) found that both values of convective available potential energy (CAPE) and convective inhibition (CIN) remained nearly unchanged within 400–1000 km from the TC center (their Fig. 6) and can be roughly regarded as the environment. The instability parameters in this study were averaged over an annular area between these two radii. We used relative humidity (RH) at 925-hPa and CAPE (calculated using surface air parcels) as representations of moisture and instability respectively. Calculations using different variables including CIN, lifted index (LI; both calculated using surface air parcels), mixing ratios and precipitable water, within different radii or different levels were also examined and the results were generally consistent. Wu et al. (2012) also showed that low-level (1000–925 hPa) RH has the most horizontal homogeneity compared with other levels. The calculations of all environmental factors are included in Table 1.

It was found that the initial environmental RH of 81–88% had nearly no impact on the subsequent average TC intensity (Fig. 9b). The mean intensity remained nearly unchanged during the 126-h forecast, although a higher environmental RH tended to have stronger TCs at

initialization, and only the wettest or driest TCs (of RH greater than 88% or less than 81%, respectively) had distinct intensity changes. Interestingly, the driest environment (less than 81%) had the smallest forecast errors during 36–84 h among all categories for ANPS (Fig. 10a), and this difference was significant during 42–66 h. This more or less coincided with its accompanying decrease in intensity during 0–72 h. All other RH ranges had statistically indistinguishable errors throughout the forecast. This outlier with a dry environment also appeared in the OFCL forecast error (Fig. 10b), although the error differences were much smaller in magnitude and significant only during the 24–48 h period.

Initial environmental CAPE values ranged from 30 to nearly 2200 J kg^{-1}, leading to a large standard deviation of 405 J kg^{-1} (Table 2). Although TCs under different initial CAPE values all started from an average initial intensity of 50–55 kt (Fig. 9c), a CAPE value of less than 900 J kg^{-1} was not conducive to intensification, while when the initial value of CAPE exceeded 900 J kg^{-1}, the TCs had a chance to intensify on average. The impacts of CAPE on forecast errors were more or less similar to moisture. On average, the larger values of CAPE did have larger intensity forecast errors by ANPS likely because they were associated with stronger TCs (Fig. 10c), but sometimes the most stable environments lead to forecast errors comparable to those in the most unstable environment. The significances of differences were acquired mostly for longer forecast lead times when the differences became sufficiently large with their continuous increasing during the whole forecast period, which indicates that environmental instability may have a continuous and long-time impact on intensity practical predictability. On the other hand, forecast error of OFCL seemed to be more stratified in that larger CAPE was more likely to be practically less predictable (Fig. 10d), and forecast errors of the smallest CAPE were significantly different from others after 12 h.

The comparably smaller changes of TC intensity forecast errors in response to environmental instability and moisture in ANPS was expected, and this complex relationship might result from numerous reasons. Intrinsically, convective conditions, and consequently moist convection processes, are meso- to microscale phenomena and less predictable than synoptic-scale conditions. Moisture and instability also might be greatly modified by the interaction of TC circulation and neighboring synoptic systems during the forecasts, like the mid-level dry air intrusion, which might result in a shorter effective predictability time scale for these parameters.

c. Translation characteristics

Landfall typically alters a TC's intensity, especially when a TC crosses over the coastline where the underlying surfaces have different frictional and thermal characteristics (Rappaport et al. 2010). When focusing only on the TCs that made landfall during the 126-h forecast, it was clear from the best-track observations that TCs would experience a rapid weakening by approximately 20 kt during landfall (Fig. 9d) on average. Indicative of previous results where a stronger intensity led to a larger intensity forecast error, it is not surprising that a similarly steep drop in error during landfall was observed (Fig. 10e). The forecast errors of landfallen TCs were significantly different from those still over the sea. However, for OFCL, the error decrease during landfall was not observed and there was no significant difference between groups. The similar magnitude of error between OFCL and TCs after landfall in ANPS indicates that the performance of ANPS, as well as other NWP models, was worse before TCs made landfall when over the sea surface than other state-of-the-art forecasts. This might result from the complexity and uncertainty of sea surface parameterization schemes under high wind speed conditions (Green and Zhang 2013).

The relationship between translation speed and intensity forecast error was complicated and limited over shorter periods of time. Figure 11 shows the intensity forecast errors categorized by the observed translation speed of the TCs at different forecast lead-times (calculated using the two latest 6-hourly best-track positions at each verified time). The characteristics that appear in all panels are smaller errors associated with fast-moving TCs (faster than 15 kt) and a more continuous increase of forecast errors of the slowest TCs (slower than 7 kt). Focusing on a shorter time period (24 h) after each categorization time (indicated by the horizontal range line at the top in each panel of Fig. 11), we can divide the 126-h forecast time into two periods: before about 60 h (Figs. 11a–d) and after about 60 h (Figs. 11e–f). Before 60 h, TCs with moderate speeds of 11–15 kt led to the largest intensity errors for the following 24 h (but the error differences are insignificant), while all other TCs except for the fastest-movers had similar smaller errors. After 60 h, the much larger forecast errors of the slowest TCs continued to increase at a faster rate than the rest, while the intensity forecast errors of all TCs except the fastest-movers maintained similar errors as in the first 60 h. The forecast superiority of the fastest TCs attained significance at nearly all forecast lead-times within the time windows of Figs. 11c–f.

These error characteristics were quite different from the intensity changes under different translation speeds (Fig. 12a). The slowest-moving TCs were the weakest and this was likely due to strong upwelling that cooled the underlying SSTs, subsequently preventing the storms from intensifying. Numerous simulation and statistical studies proved the negative feedback that when a TC becomes stronger, it will induce stronger upwelling in the mixing layer below sea surface, subsequently resulting in a cooler SST that is less favorable for further intensification of the TC (Chang 1985, Schade and Emanuel 1999, Chan et al. 2001, Dare and McBride 2011, Liu et al. 2011). TCs with a moderate speed of 11–15 kt on average had the strongest intensity and the

fastest intensification, while the fastest TCs generally maintained their intensity. The relationship between forecast error and translation speed at each forecast lead-time again indicates the impact of translational speed on practical predictability (Fig. 12b), although most of the differences were not significant except for the fastest TCs. However, model bias may also play a role. The slowest TCs had the largest intensity biases, especially in longer forecast lead times (Fig. 12c). TCs of 11–15 kt translation speed also experienced constantly negative bias (forecasts weaker than observations). Biases of these two groups were significantly different from the others after 36 h and 30 h respectively. Since ANPS is not coupled with any ocean model, the cooling effect of TCs on SST through upwelling cannot be represented in the ANPS forecasts. Sandery et al. (2010) showed that a coupling atmosphere-ocean model produces smaller TC intensity forecast errors than a stand-alone atmosphere model. For the slowest-moving TCs, the NWP model may artificially extract too much energy from the sea surface, which results in a positive intensity forecast bias compared to best-track observations. If all biases were removed, the forecast errors of the slowest TCs were reduced (Fig. 12d); storms with moderate speeds of 11–15 kt had the largest errors and both faster- and slower-moving TCs had smaller errors. Again, only the errors of the fastest TCs remained significantly different (after 24 h), and the largest errors for TCs with the 11–15 kt translation speed acquired significance for a few longer lead-times. These results indicate that the practical predictability of TC intensity might be improved through the coupling with an ocean model, especially for slowing moving storms.

5. Conclusions

This study explored the practical predictability of TC intensity through convection-permitting hindcasts with WRF/ARW model and real-time 5-day official forecasts issued by

NHC of all Atlantic TCs from the 2008–2012 seasons. The focus was the impact of initial conditions on the intensity forecast errors during the 126-h TC hindcasts. The day 1–5 performance of these WRF hindcasts of TC intensity was comparable to two operational regional-scale hurricane prediction models used by NHC, and slightly inferior to the NHC official forecasts.

The TC intensity itself was found to be the most important factor related with the accuracy of the intensity forecasts among all metrics explored in this study. Since the initially stronger TCs tend to weaken in the forecasts, their intensity forecast errors generally decreased through time, while the initially weaker TCs on average had increasing values of V_{max} and thus increasing errors. Forecast errors were also positively correlated with initial errors. TCs with a large negative initial bias (the initialized TC in the NWP model was much weaker than the observations) maintained significantly larger forecast errors. The inability to accurately capture the timing and magnitude of RI was also revealed, as there were steep increases of forecast errors during RI.

Many other internal or environmental variables, including environmental vertical wind shear, SST, latitude, underlying surface condition etc., affected the forecast errors through their impacts on intensity. Deep-layer (850–200 hPa) vertical wind shear was an important factor that might affect TC intensity forecast accuracy. The practical predictability of TC intensity generally increased when shear became stronger for both ANPS and OFCL. Since the sea surface is the major energy source for TC development and maintenance, higher SSTs can fuel stronger TCs and therefore lead to larger forecast errors. However, in a statistical sense, TCs do not intensify when the SST is higher than 29 °C on average, possibly because higher SSTs are usually located in areas near the equator where the Coriolis force is insufficient for development. Because of this,

smaller intensity forecast errors were observed in this range of SSTs than in the 27–29 °C group in the ANPS model. The latitude at forecast initialization was also strongly related with intensity predictability. However, a calculation of partial correlation revealed that this was almost exclusively because of the strong relationship between latitude and SST. As a consequence of the chaotic nature of mesoscale convection, the impacts of environmental moisture and instability on intensity forecast errors were comparably smaller than some of the other parameters examined in this study, and most of their impacts were insignificant. The intensity forecast errors decreased sharply during landfall, partly because of the rapid weakening of TCs during landfall.

The impact of translation speed on intensity predictability was apparent over much shorter periods of time than other parameters. Besides the largest predictability of the fastest-moving TCs, a moderate speed of 11–15 kt had the largest forecast errors during the following 24 h for forecast lead times shorter than 60 h, while TCs with the slowest speeds (smaller than 7 kt) were the least predictable after 60 h. Analyses on the forecast biases indicate that the slowest TCs have the largest biases and that the simulated TCs in this group were stronger than the observed ones. When the model biases of different translation speeds were removed, the largest errors of the slowest TCs disappeared and the TCs with speeds between 11–15 kt consistently had the significantly largest forecast errors. Analysis of the de-biased intensity forecast errors as well as model bias characteristics suggest that systematic model bias of TC intensity under different translation speeds might also contribute to the forecast errors.

In summary, larger TC intensity forecast errors are more likely in the ANPS model configuration described here, when the environment has a weak to moderate vertical shear of 0–10 m s^{-1}, a warm SST of 27–29 °C, when the TC is moving at a translation speed of 11–15 kt or when the TC is experiencing RI, resulting from a higher chance of producing stronger TCs under

these conditions. Changes of other environmental thermodynamic parameters may contribute less to the practical predictability. It should be pointed out that improvements in ICs and BCs may also alter practical predictability of TC intensity, and the characteristics here were analyzed under the framework of the WRF model without using a regional scale cycling data assimilation system. Therefore, the results presented in this study should be explained and applied with caution, especially in circumstances where a different physical parameterization scheme is used or there are different model conditions[3]. Real world tropical cyclones are influenced by plenty of atmospheric and oceanic variables, and parameters examined in this work are somewhat limited. However, our study did reveal some features affecting TC intensity forecast errors in an NWP model, and will be helpful in further improving TC intensity predictions by increasing the in-depth understanding of practical predictability.

Acknowledgement

We thank Erin Munsell for her insightful comments and proofreading of an earlier version of the manuscript. The first author is supported by the Chinese Scholarship Council (CSC). This work is partially supported by NOAA under the Hurricane Forecast Improvement Project (HFIP), China National Basic Research Program 2013CB430104, the R&D Special Fund for Public Welfare Industry (meteorology) GYHY201306004 and NSFC41375048. The simulations were performed on the High Performance Computing System of NOAA and the Texas Advanced Computing Center (TACC).

[3] A comparison between intensity forecast errors of ANPS, GFDL and HWRF shows generally consistent characteristics among them.

References:

Aksoy, A., S. D. Aberson, T. Vukicevic, K. J. Sellwood, S. Lorsolo, and X. Zhang, 2013: Assimilation of high-resolution tropical cyclone observations with an ensemble Kalman filter using NOAA/AOML/HRD's HEDAS: Evaluation of the 2008–11 vortex-scale analysis. *Mon. Wea. Rev.*, **141**, 1842–1865.

Bender, M.A., I. Ginis, R. Tuleya, B. Thomas, and T. Marchok, 2007: The operational GFDL coupled hurricane-ocean prediction system and summary of its performance. *Mon. Wea. Rev.*, **135**, 3965–3989.

Bhatia, K. T., and D. S. Nolan, 2013: Relating the skill of tropical cyclone intensity forecasts to the synoptic environment. *Wea. Forecasting*, **28**, 961–980.

Brown, B. R., and G. J. Hakim, 2013: Variability and predictability of a three-dimensional hurricane in statistical equilibrium. *J. Atmos. Sci.*, **70**, 1806–1820.

Cangialosi, J. P., and J. L. Franklin, 2013: 2012 National Hurricane Center verification report. Tropical Prediction Center, National Hurricane Center, National Center for Environmental Prediction, National Weather Center, NOAA, 79pp. [Available online at http://www.nhc.noaa.gov/verification/pdfs/Verification_2012.pdf.]

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.

Chan, J. C. L., William M. Gray, 1982: Tropical cyclone movement and surrounding flow relationships. *Mon. Wea. Rev.*, **110**, 1354–1374.

Chan, J. C. L., Y. Duan, and L. Shay, 2001: Tropical cyclone intensity change from a simple ocean-atmosphere coupled model. *J. Atmos. Sci.*, **58**, 154–172.

Chang, S. W., 1985: Deep ocean response to hurricanes as revealed by an ocean model with free surface. Part I: Axisymmetric case. *J. Phys. Oceanogr.*, **15**, 1847–1858

Chen, S. S., J. A. Knaff, and F. D. Marks, Jr., 2006: Effects of vertical wind shear and storm motion on tropical cyclone rainfall asymmetries deduced from TRMM. *Mon. Wea. Rev.*, **134**, 3190–3208.

Corbosiero, K. L., and J. Molinari, 2002: The effect of vertical wind shear on the distribution of convection in tropical cyclones. *Mon. Wea. Rev.*, **130**, 2110–2123.

Corbosiero, K. L., and J. Molinari, 2003: The relationship between storm motion, vertical wind shear, and convective asymmetries in tropical cyclones. *J. Atmos. Sci.*, **60**, 366–376.

Dare, R. A., and J. McBride, 2011: Sea surface temperature response to tropical cyclones. *Mon. Wea. Rev.*, **139**, 3798–3808.

Davis, C., W. Wang, J. Dudhia, and R. Torn, 2010: Does increased horizontal resolution improve hurricane wind forecasts? *Wea. Forecasting*, **25**, 1826–1841.

DeMaria, 1996: The effect of vertical shear on tropical cyclone intensity change. *J. Atmos. Sci.*, **53**, 2076–2087.

DeMaria, M., and J. Kaplan, 1994: A statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic basin. *Wea. Forecasting*, **9**, 209–220.

DeMaria, M., and J. Kaplan, 1999: An updated statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic and eastern North Pacific basins. *Wea. Forecasting*, **14**, 326–337.

DeMaria, M., M. Mainelli, L. K. Shay, and J. Kaplan, 2005: Further improvements to the statistical hurricane intensity prediction scheme (SHIPS). *Wea. Forecasting*, **20**, 531–543.

Doswell, C. A., H. E. Brooks, and R. A. Maddox, 1996: Flash flood forecasting: an ingredients-based methodology. *Wea. Forecasting*, **11**, 560–581.

Duhdia, J., 1989: Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci*., **46**, 3077–3107.

Duhdia, J., 1996: A multi-layer soil temperature model for MM5. *Preprints*, *6th Annual MM5 Users Workshop*, Boulder, CO. [Available online at http://www.mmm.ucar.edu/mm5/lsm/soil.pdf]

Elsberry, R. L., T. D. B. Lambert, and M. A. Boothe, 2007: Accuracy of Atlantic and eastern North Pacific tropical cyclone intensity forecast guidance. *Wea. Forecasting*, **22**, 747–762.

Emanuel, K. A., 1986: An Air-Sea Interaction Theory for Tropical Cyclones. Part I: Steady-State Maintenance. *J. Atmos. Sci.*, 43, 585–604.

Fujiwhara, S., 1921: The natural tendency towards symmetry of motion and its application as a principle in meteorology. *Quart. J Roy. Meteor. Soc.*, **47**, 287–292.

Gao, S., Z. Meng, F. Zhang, and L. F. Bosart, 2009: Observational analysis of heavy rainfall mechanisms associated with severe tropical storm Bilis (2006) after its landfall. *Mon. Wea. Rev.*, **137**, 1881–1897.

Gopalakrishnan, S., Q. Liu, T. Marchok, D. Sheinin, V. Tallapragada, M. Tong, R. Tuleya, R. Yablonsky, and X. Zhang, 2012: Hurricane Weather Research and Forecasting (HWRF) model: 2012 scientific documentation. L. Bernardet, Ed., 96pp. [Available online at http://www.dtcenter.org/HurrWRF/users/docs/scientific_documents/HWRFScientificDocumentation_v3.4a.pdf]

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.

Grell, G. A., and D. Devenyi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Lett.*, **29**, 1693.

Hakim, G. J., 2013: The variability and predictability of axisymmetric hurricanes in statistical equilibrium. *J. Atmos. Sci.*, **70**, 993–1005.

Holland, Greg J., 1983: Tropical Cyclone Motion: Environmental Interaction Plus a Beta Effect. *J. Atmos. Sci.*, **40**, 328–342.

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

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

Houze, R. A., S. S. Chen, B. F. Smull, W.-C. Lee, and M. M. Bell, 2007: Hurricane intensity and eyewall replacement. *Science*, **315**, 1235–1238.

Kaplan, J., and M. DeMaria, 2003: Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin. *Wea. Forecasting*, **18**, 1093–1108.

Kaplan, J., M. DeMaria, and J. A. Knaff, 2010: A revised tropical cyclone rapid intensification index for the Atlantic and eastern North Pacific basins. *Wea. Forecasting*, **25**, 220–241.

Knaff, J., M. DeMaria, C. R. Sampson, and J. M. Gross, 2003: Statistical, 5-day tropical cyclone intensity forecasts derived from climatology and persistence. *Wea. Forecasting*, **18**, 80–92.

Kunii, M., and T. Miyoshi, 2013: Including uncertainties of sea surface temperature in an ensemble Kalman filter: a case study of typhoon Sinlaku (2008). *Wea. Forecasting*, **27**, 1586–1597.

Law, K. T., and J. S. Hobgood, 2007: A statistical model to forecast short-term Atlantic hurricane intensity. *Wea. Forecasting*, **22**, 967–980.

Li, Y., X. Wang, and M. Xue, 2012: Assimilation of radar radial velocity data with the WRF hybrid ensemble-3DVar system for the prediction of hurricane Ike (2008). *Mon. Wea. Rev.*, **140**, 3507–3524.

Liu, Q., T. Marchok, H. Pan, M. Bender, and S. Lord, 2000: Improvements in hurricane initialization and forecasting at NCEP with global and regional (GFDL) models. NCEP/EMC Tech. Procedures Bull. 472, 7 pp. [Available online at http://205.156.54.206/om/tpb/472.htm.]

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.

Melhauser, C., and F. Zhang, 2012: Practical and intrinsic predictability of severe and convective weather at the mesoscales. *J. Atmos. Sci.*, **69**, 3350–3371.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16663–16682.

Molinari, J., and D. Vollaro, 2010a: Distribution of helicity, CAPE, and shear in tropical cyclones. *J. Atmos. Sci.*, **67**, 274–284.

Molinari, J., and D. Vollaro, 2010b: Rapid intensification of a sheared tropical storm. *Mon. Wea. Rev.*, **138**, 3869–3885.

Molinari, J., D. Vollaro, and K. L. Corbosiero, 2004: Tropical cyclone formation in a sheared environment: a case study. *J. Atmos. Sci.*, **61**, 2493–2509.

Molinari, J., P. Dodge, D. Vollaro, K. L. Corbosiero, and F. Marks, Jr., 2006: Mesoscale aspects of the downshear reformation of a tropical cyclone. *J. Atmos. Sci.*, **63**, 341–354.

Molinari, J, D. M. Romps, D. Vollaro, and L. Nguyen, 2012: CAPE in tropical cyclones. *J. Atmos. Sci.*, **69**, 2452–2463.

NCEP, cited 2013: GFS/GDAS changes since 1991. [Available online at http://www.emc.ncep.noaa.gov/gmb/STATS/html/model_changes.html]

Nguyen, L. T., and J. Molinari, 2012: Rapid intensification of a sheared, fast-moving hurricane over the Gulf Stream. *Mon. Wea. Rev.*, **140**, 3361–3378.

Rappaport, E. N., J. L. Franklin, A. B. Schumacher, M. DeMaria, L. K. Shay, and E. J. Gibney, 2010: Tropical cyclone intensity change before U.S. Gulf coast landfall. *Wea. Forecasting*, **25**, 1380–1396.

Rapping, E. D., and D. S. Nolan, 2012: The effect of vertical shear orientation on tropical cyclogenesis. *Q. J. R. Meteorol. Soc.*, **138**, 1035–1054.

Reasor, P. D., R. Rogers, and S. Lorsolo, 2013: Environmental flow impacts on tropical cyclone structure diagnosed from airborne Doppler radar composites. *Mon. Wea. Rev.*, **141**, 2949–2969.

Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, 2007: Daily high-resolution-blended analysis for sea surface temperature. *J. Climate*, **20**, 5473–5496.

Sampson, C. R., and A. J. Schrader, 2000: The Automated Tropical Cyclone Forecasting system (Version 3.2). *Bull. Amer. Meteor. Soc.,* **81**, 1231–1240.

Schade, L. R., and K. A. Emanuel, 1999: The ocean's effect of the intensity of tropical cyclones: Results from a simple coupled atmosphere–ocean model. *J. Atmos. Sci.*, **56**, 642–651.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang, W. Wang, and J. G. Powers, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp.

Trahan, Sam, Lynn Sparling, 2012: An Analysis of NCEP Tropical Cyclone Vitals and Potential Effects on Forecasting Models. *Wea. Forecasting*, **27**, 744–756.

Van Sang, N., R. K. Smith, and M. T. Montgomery, 2008: Tropical-cyclone intensification and predictability in three dimensions. *Quart. J Roy. Meteor. Soc.*, **134**, 563–582.

Weng, Y., and F. Zhang, 2012: Assimilating airborne Doppler radar observations with an ensemble Kalman filter for convection-permitting hurricane initialization and prediction: Katrina (2005). *Mon. Wea. Rev.*, **140**, 841–859.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences.* 2nd ed. Academic Press, 627 pp.

Wingo, M. T., and D. J. Cecil, 2010: Effects of vertical wind shear on tropical cyclone precipitation. *Mon. Wea. Rev.*, **138**, 645–662.

Wu, D., Z. Meng, and D. Yan, 2013: The predictability of a squall line in South China on 23 April 2007. *Advances in Atmos. Sci.*, **30**, 485–502.

Wu, L., and coauthors, 2012: Relationship of environmental relative humidity with North Atlantic tropical cyclone intensity and intensification rate. *Geophys. Res. Lett.*, **39**, L20809.

Yu, H., P. Chen, Q. Li, and B. Tang, 2013: Current capability of operational numerical models in predicting tropical cyclone intensity in the western North Pacific. *Wea. Forecasting*, **28**, 353–367.

Zehr, R. M., 2003: Environmental vertical wind shear with hurricane Bertha (1996). *Wea. Forecasting*, **18**, 345–356.

Zeng, Z., Y. Wang, and L. Chen, 2010: A statistical analysis of vertical shear effect on tropical cyclone intensity change in the North Atlantic. *Geophys. Res. Lett.*, **37**, L02802.

Zhang, D., and R. A. Anthes, 1982: A high-resolution model of the planetary boundary layer–sensitivity tests and comparisons with SESAME-79 data. *J. Appl. Meteor.*, **21**, 1594–1609.

Zhang, F., and J. A. Sippel, 2009: Effects of moist convection on hurricane predictability. *J. Atmos. Sci.*, **66**, 1944–1961.

Zhang, F., and D. Tao, 2013: Effects of vertical wind shear on the predictability of tropical cyclones. *J. Atmos. Sci.*, **70**, 975–983.

Zhang, F., C. Snyder, and R. Rotunno, 2002: Mesoscale predictability of the "surprise" snowstorm of 24–25 January 2000. *Mon. Wea. Rev.*, **130**, 1617–1632.

Zhang, F., C. Snyder, and R. Rotunno, 2003: Effect of moist convection on mesoscale predictability. *J. Atmos. Sci.*, **60**, 1173–1185,

Zhang, F., Y. Weng, J. F. Gamache, and F. D. Marks, 2011: Performance of cloud-resolving hurricane initialization and prediction during 2008–2010 with ensemble data assimilation of inner-core airborne Doppler radar observations. *Geophys. Res. Lett.*, **38**, L15810.

Zhang, F., Y. Weng, J. A. Sippel, Z. Meng, and C. H. Bishop, 2009: Cloud-resolving hurricane initialization and prediction through assimilation of Doppler radar observations with an ensemble Kalman filter: Humberto (2007). *Mon. Wea. Rev.*, **137**, 2105–2125.

Tables

Table 1. Calculations of environmental parameters

Variables	Data	Algorithm
Shear ($m\ s^{-1}$)	NCEP FNL	Difference between 200 and 850 hPa, each layer averaged within an annulus area of 200–800 km from the TC center
SST (°C)	NOAA OISST	Averaged within 200 km from the TC center, if more than half grids of this area are over sea
925-hPa RH	NCEP FNL	Averaged within an annulus area of 200–800 km from the TC center
CAPE ($J\ kg^{-1}$)	NCEP FNL	Averaged within an annulus area of 400–1000 km from the TC center

1 Table 2. Mean values and standard deviations of environmental parameters.

Variables	Shear ($m\ s^{-1}$)	SST (°C)	925-hPa RH	CAPE ($J\ kg^{-1}$)
Mean	8.84	27.46	83.7%	958
St. D.	4.99	2.03	5.7%	405

2

Figure Captions

Figure 1. Domain setting as of 0000 UTC 26 Oct 2012 used for a forecast of hurricane Sandy (2012), with tracks of all simulated TCs during 2008–2012. Different colors of tracks stand for different hurricane intensity categories as labeled at the bottom of this figure.

Figure 2. Mean absolute errors during 2008–2012 of (a) track and (b) maximum wind speed (V_{max}) for selected regional NWP models and NHC official forecasts, as well as (c) their mean biases of V_{max} and (d) relationships between V_{max} and P_{min} for NWP model forecasts and best-track observations. In (a) and (b), the numbers at the top of each panel indicate sample size of different models at that time, the two-color squares at the bottom indicate significantly different pairs at that time. In (a), (b) and (c), the vertical lines accompanying error or bias trends are ranges between the first and the third quartiles of each group every 24 h. In (d), the scatters are each single V_{max}-P_{min} pair of best track.

Figure 3. (a) Yearly mean absolute intensity forecast errors and (b) yearly mean intensity biases of ANPS. In each panel, the vertical lines accompanying error or bias trends are ranges between the first and the third quartiles of each group every 24 h.

Figure 4. (a) ANPS and (b) OFCL mean absolute intensity forecast errors, and (c) ANPS and (d) OFCL mean relative intensity forecast errors, all categorized by best-track TC intensity categories at forecast initialization time (t = 0 h). In each panel, the vertical lines accompanying

error trends are ranges between the first and the third quartiles of each group every 24 h; the two-color squares at the bottom indicate significantly different pairs at that time.

Figure 5. Similar to Figure 4, but for (a) ANPS and (b) OFCL mean absolute intensity forecast errors, and (c) ANPS and (d) OFCL mean relative intensity forecast errors, all categorized by best-track TC intensity categories at each verification forecast lead-times.

Figure 6. Best track mean intensity categorized by (a) best-track TC intensity categories at forecast initialization time, (b) best-track intensity changes during the first 24 h of forecasts, (c) days that rapid intensification began, and (d) 850–200-hPa vertical wind shear magnitudes. In each panel, the vertical lines accompanying intensity trends are ranges between the first and the third quartiles of each group every 24 h.

Figure 7. Similar to Figure 4, but for mean absolute intensity forecast errors of (a), (c) and (e) ANPS and (b), (d) and (f) OFCL, categorized by intensity bias at (a) 6 h and (b) 0 h, (c) and (d) best-track intensity changes during the first 24 h of forecasts, and (e) and (f) days that rapid intensification began, except for that the quartile ranges are plotted during 6–126 h every 24 h in (a).

Figure 8. Similar to Figure 4, but categorized by (a) and (b) 850–200-hPa vertical wind shear magnitudes, (c) and (d) SST at forecast initialization time, and (e) and (f) center latitude at forecast initialization time, for (a), (c) and (e) ANPS and (b), (d) and (f) OFCL forecasts.

Figure 9. Similar to Figure 6, but categorized by (a) SST at forecast initialization time, (b) 925-hPa RH at forecast initialization time, (c) CAPE at forecast initialization time, and (d) days that the TC made landfall in best track during the forecast.

Figure 10. Similar to Figure 4, but categorized by (a) and (b) 925-hPa RH at forecast initialization time, (c) and (d) CAPE and forecast initialization time, and (e) and (f) days that the TC made landfall in best track during the forecast, for (a), (c) and (e) ANPS and (b), (d) and (f) OFCL forecasts.

Figure 11. Similar to Figure 4, but categorized by the best-track translation speed of the TC at forecast lead-times of (a) 0 h, (b) 12 h, (c) 24 h, (d) 48 h, (e) 72 h and (f) 96 h for ANPS forecasts. In each panel, the horizontal range lines at the top indicate 0–24-h time period after categorization time.

Figure 12. (a) Best track mean intensity, (b) ANPS mean absolute intensity forecast errors, (c) ANPS mean intensity forecast bias and (d) ANPS de-biased mean absolute intensity forecast errors, all categorized by translation speed of the TC at each verification forecast lead-times. In each panel, the vertical lines accompanying intensity, error or bias trends are ranges between the first and the third quartiles of each group every 24 h. In (b) and (d), the two-color squares at the bottom indicate significantly different pairs at that time.

Figures

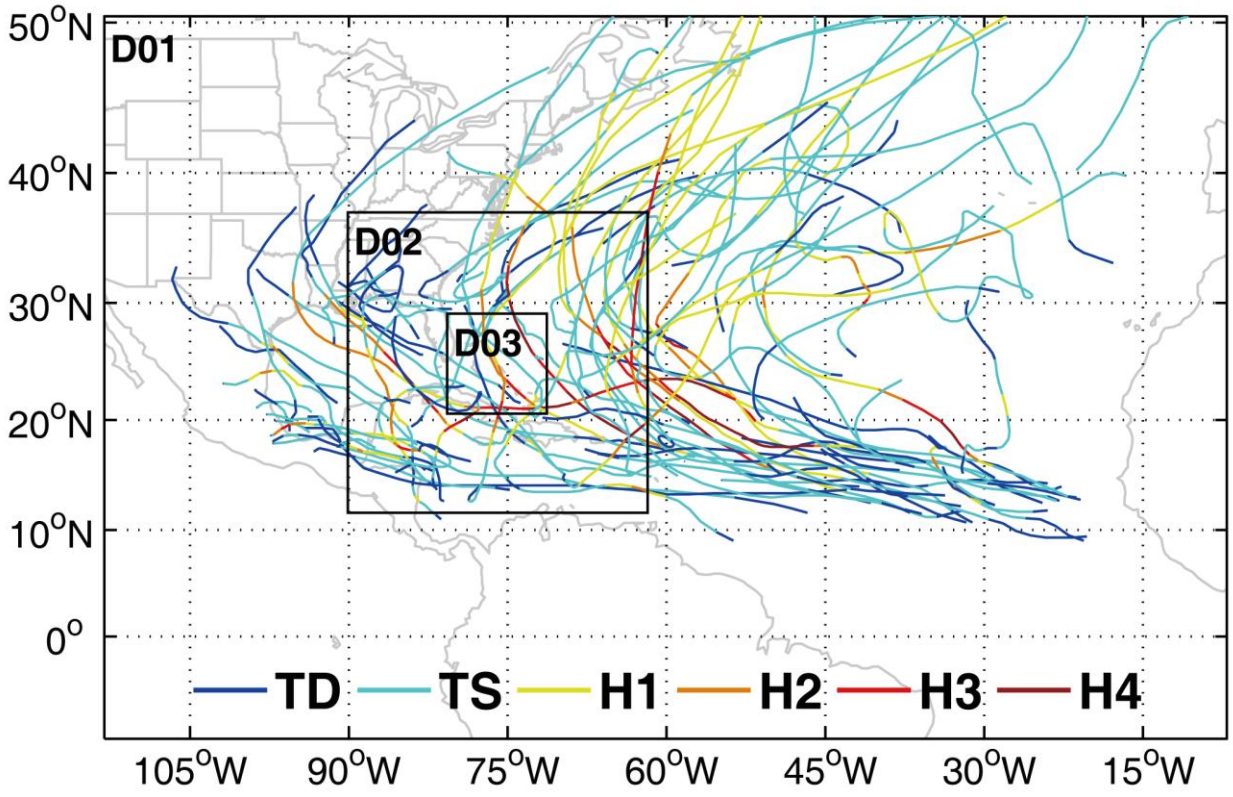

Figure 1. Domain setting as of 0000 UTC 26 Oct 2012 used for a forecast of hurricane Sandy (2012), with tracks of all simulated TCs during 2008–2012. Different colors of tracks stand for different hurricane intensity categories as labeled at the bottom of this figure.

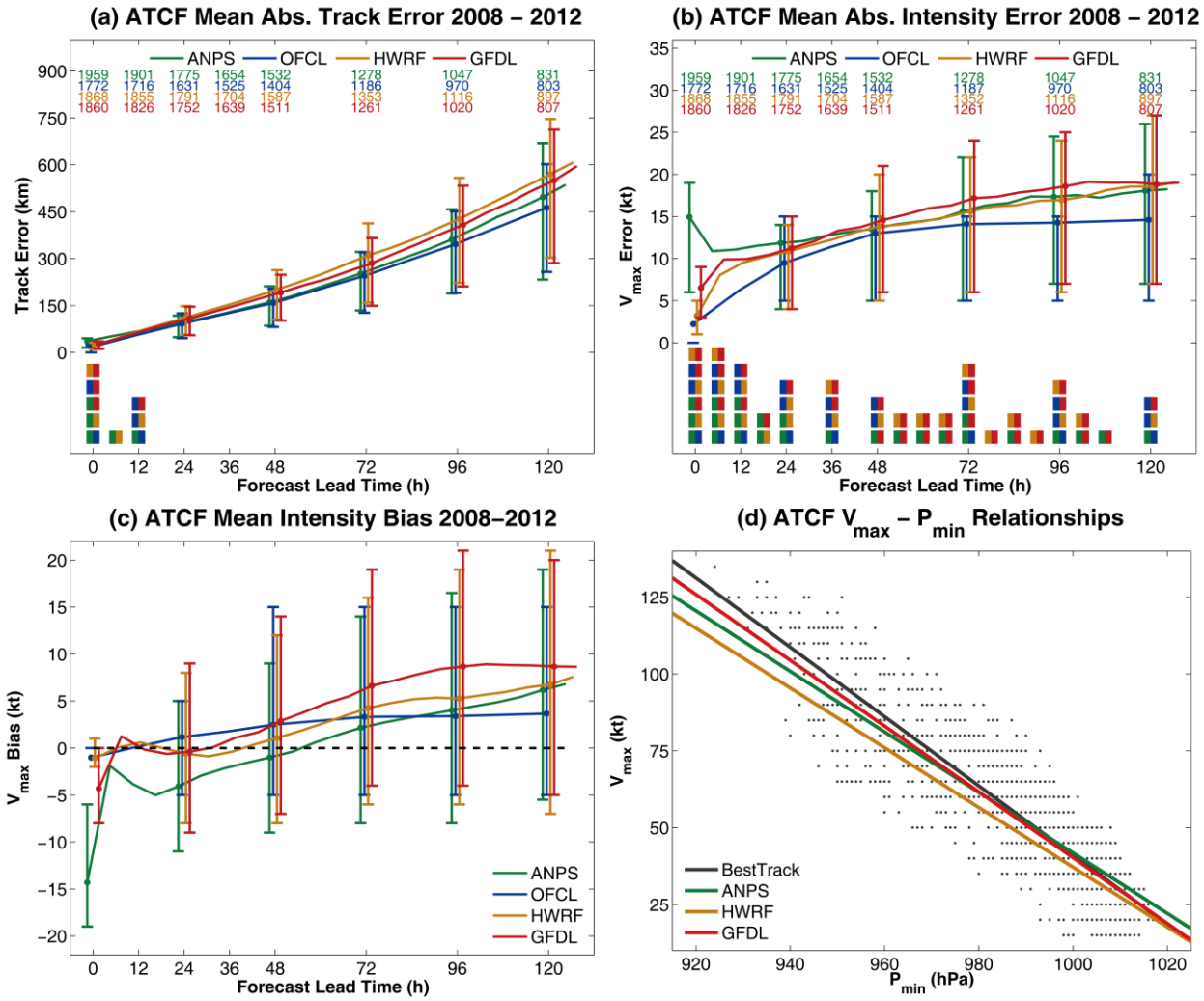

Figure 2. Mean absolute errors during 2008–2012 of (a) track and (b) maximum wind speed (V_{max}) for selected regional NWP models and NHC official forecasts, as well as (c) their mean biases of V_{max} and (d) relationships between V_{max} and P_{min} for NWP model forecasts and best-track observations. In (a) and (b), the numbers at the top of each panel indicate sample size of different models at that time, the two-color squares at the bottom indicate significantly different pairs at that time. In (a), (b) and (c), the vertical lines accompanying error or bias trends are ranges between the first and the third quartiles of each group every 24 h. In (d), the scatters are each single V_{max}-P_{min} pair of best track.

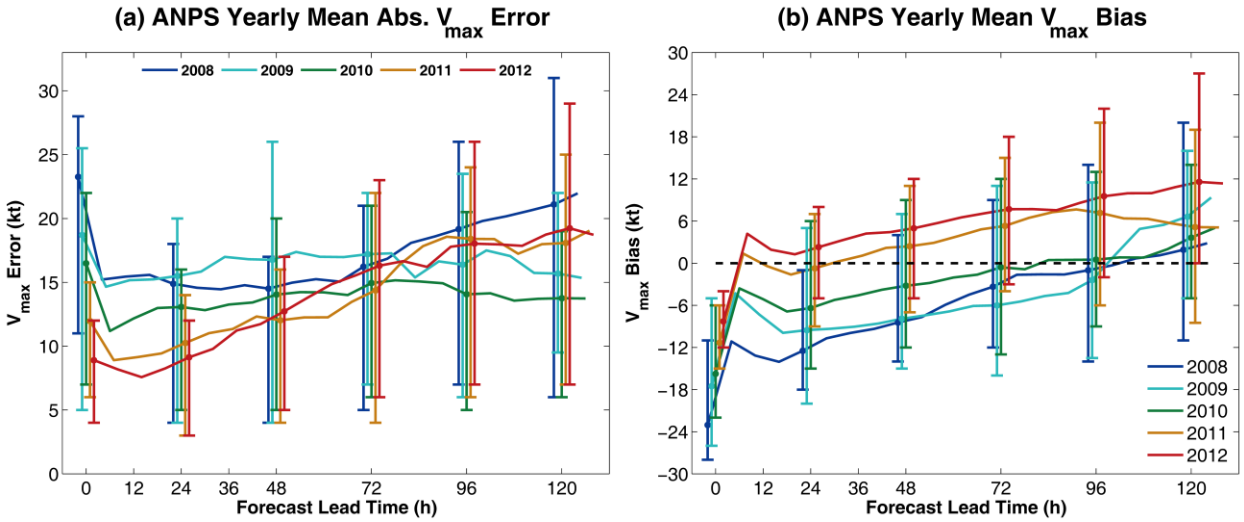

Figure 3. (a) Yearly mean absolute intensity forecast errors and (b) yearly mean intensity biases of ANPS. In each panel, the vertical lines accompanying error or bias trends are ranges between the first and the third quartiles of each group every 24 h.

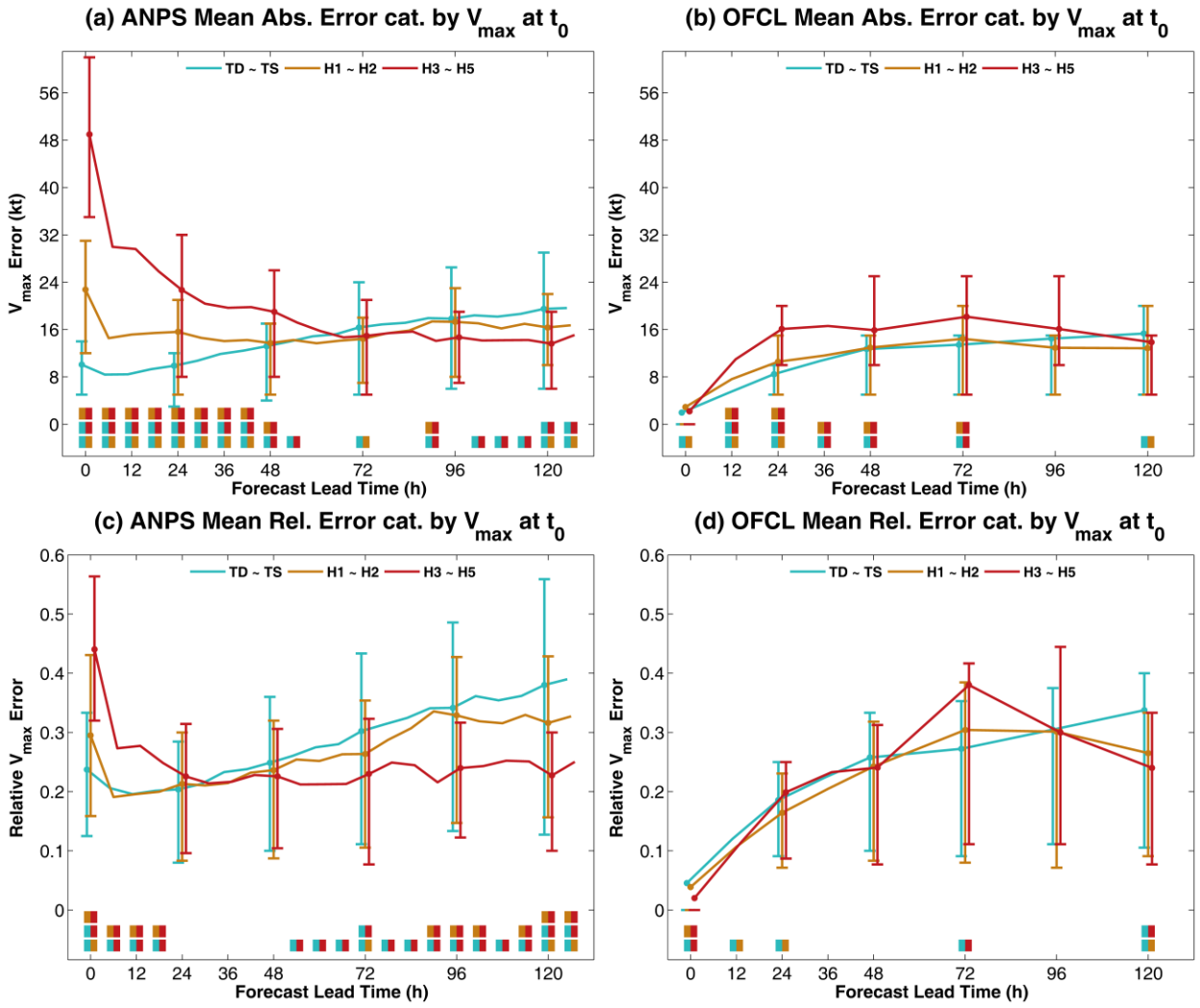

Figure 4. (a) ANPS and (b) OFCL mean absolute intensity forecast errors, and (c) ANPS and (d) OFCL mean relative intensity forecast errors, all categorized by best-track TC intensity categories at forecast initialization time (t = 0 h). In each panel, the vertical lines accompanying error trends are ranges between the first and the third quartiles of each group every 24 h; the two-color squares at the bottom indicate significantly different pairs at that time.

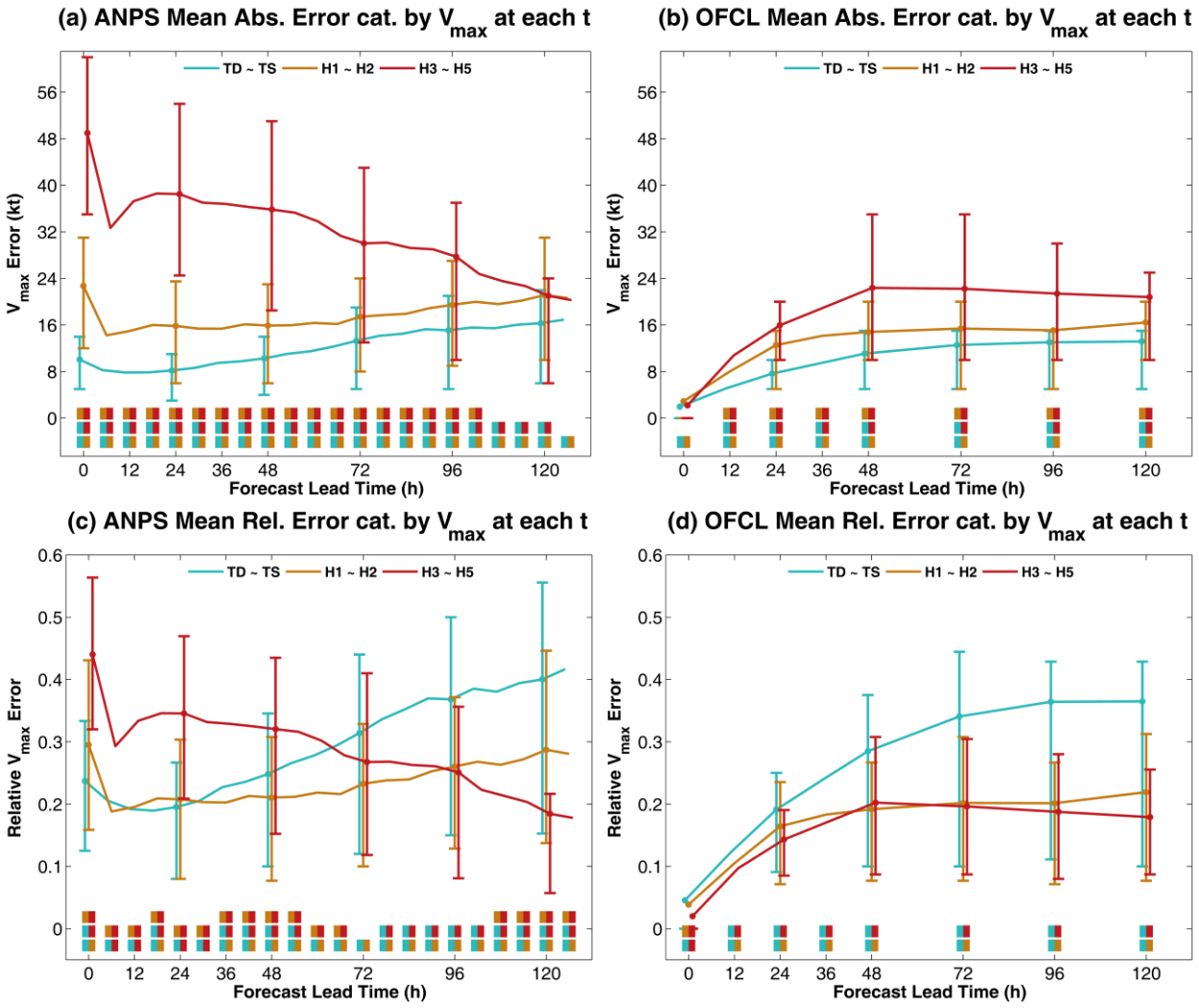

Figure 5. Similar to Figure 4, but for (a) ANPS and (b) OFCL mean absolute intensity forecast errors, and (c) ANPS and (d) OFCL mean relative intensity forecast errors, all categorized by best-track TC intensity categories at each verification forecast lead-times.

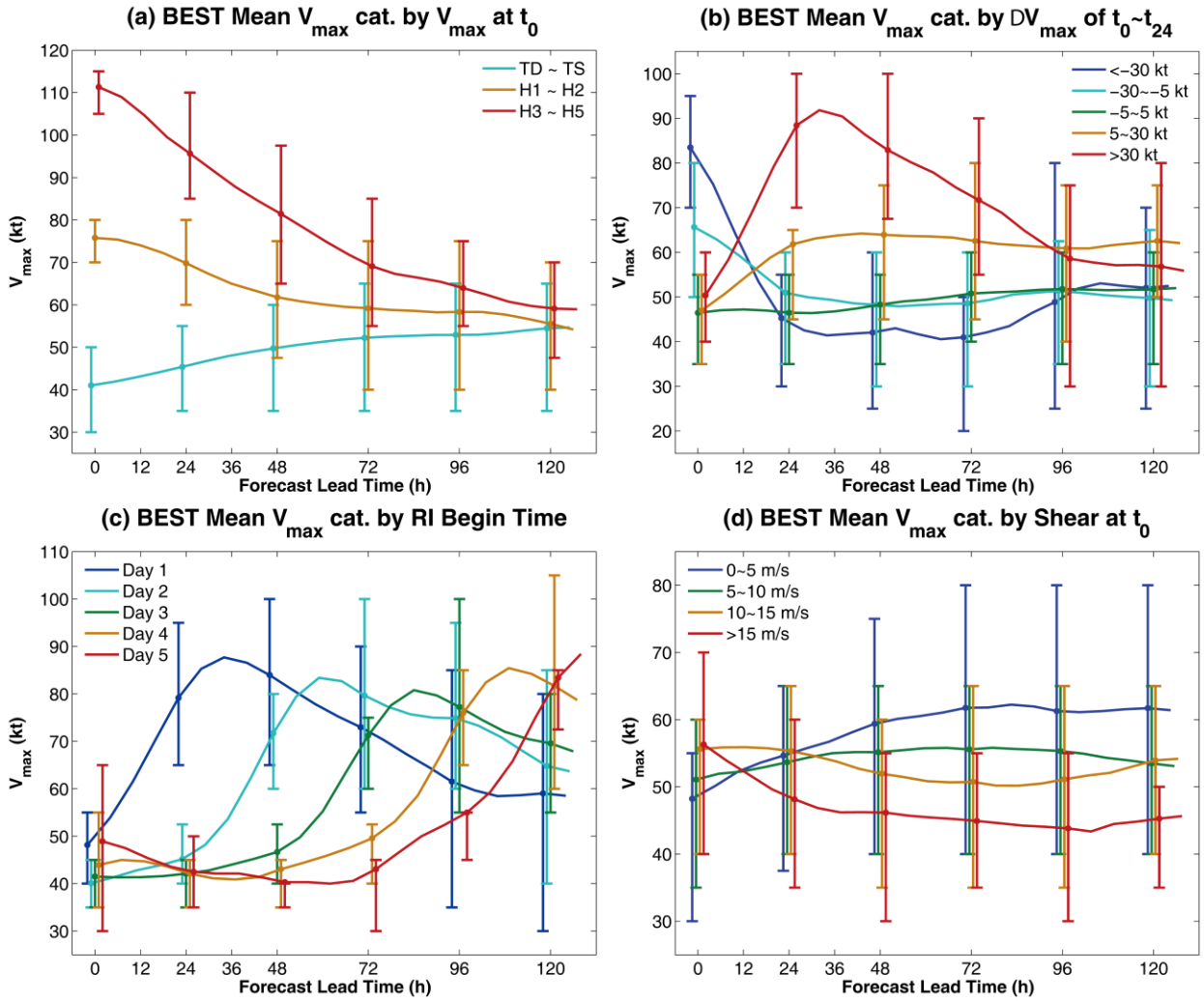

Figure 6. Best track mean intensity categorized by (a) best-track TC intensity categories at forecast initialization time, (b) best-track intensity changes during the first 24 h of forecasts, (c) days that rapid intensification began, and (d) 850–200-hPa vertical wind shear magnitudes. In each panel, the vertical lines accompanying intensity trends are ranges between the first and the third quartiles of each group every 24 h.

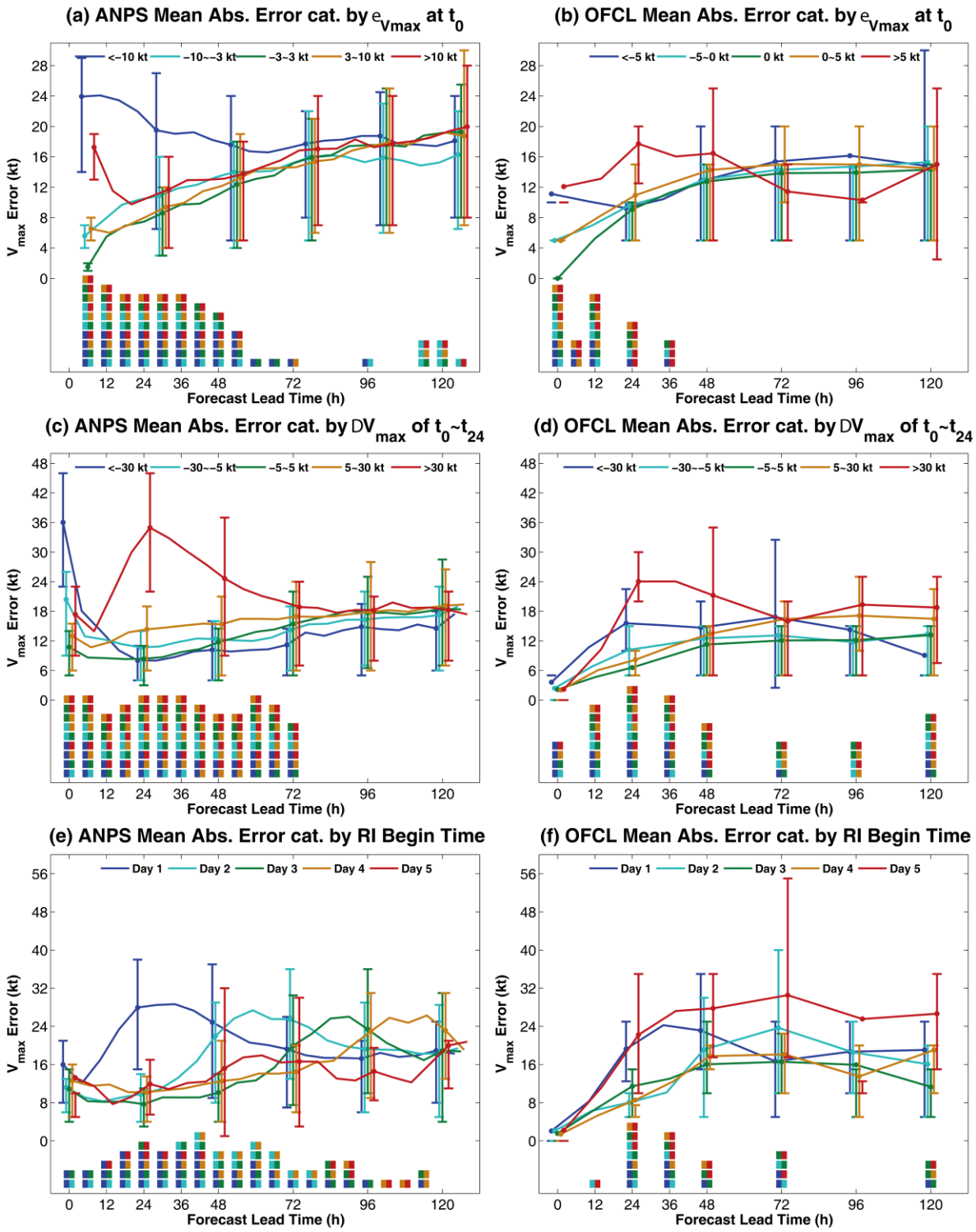

Figure 7. Similar to Figure 4, but for mean absolute intensity forecast errors of (a), (c) and (e) ANPS and (b), (d) and (f) OFCL, categorized by intensity bias at (a) 6 h and (b) 0 h, (c) and (d) best-track intensity changes during the first 24 h of forecasts, and (e) and (f) days that rapid intensification began, except for that the quartile ranges are plotted during 6–126 h every 24 h in (a).

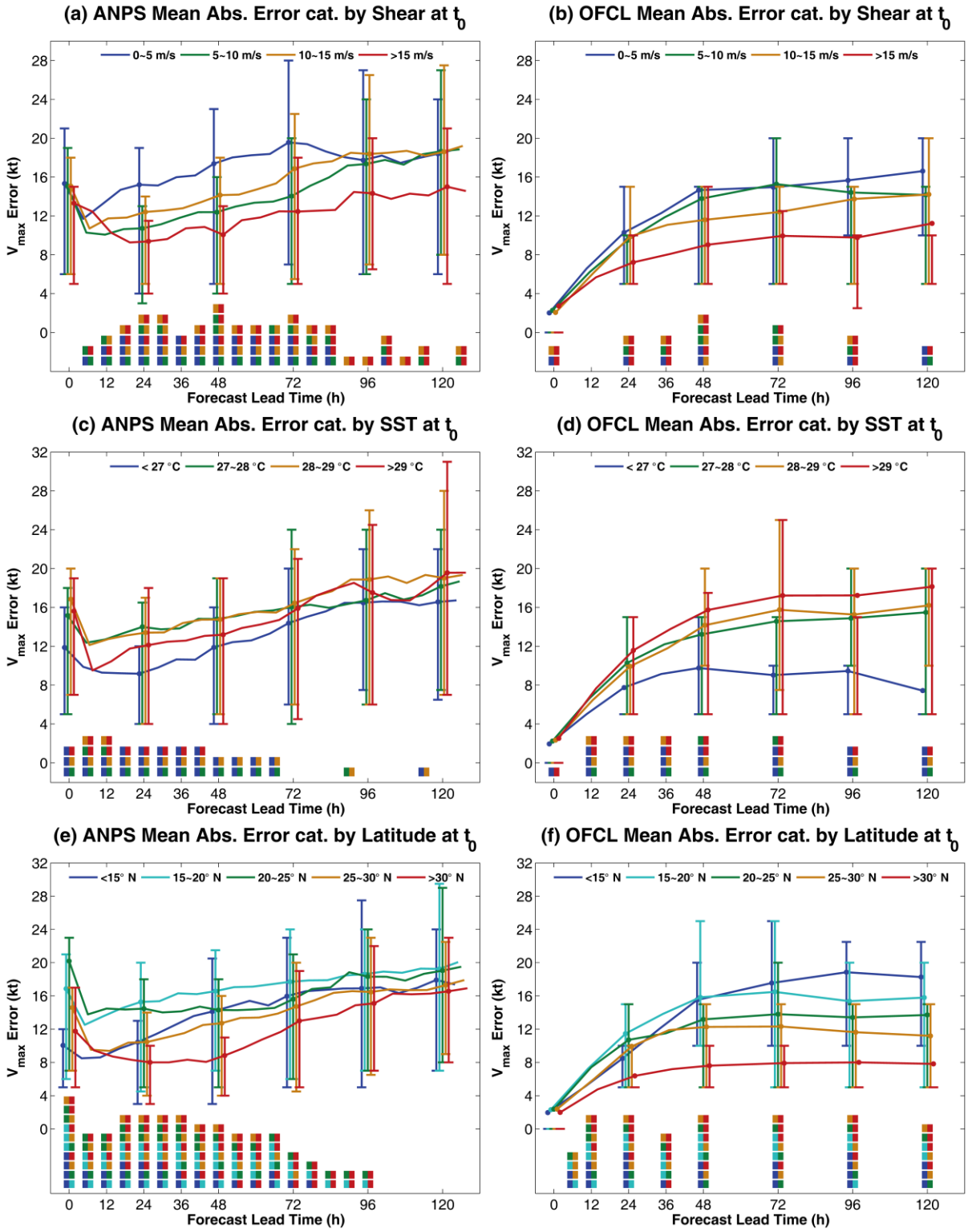

Figure 8. Similar to Figure 4, but categorized by (a) and (b) 850–200-hPa vertical wind shear magnitudes, (c) and (d) SST at forecast initialization time, and (e) and (f) center latitude at forecast initialization time, for (a), (c) and (e) ANPS and (b), (d) and (f) OFCL forecasts.

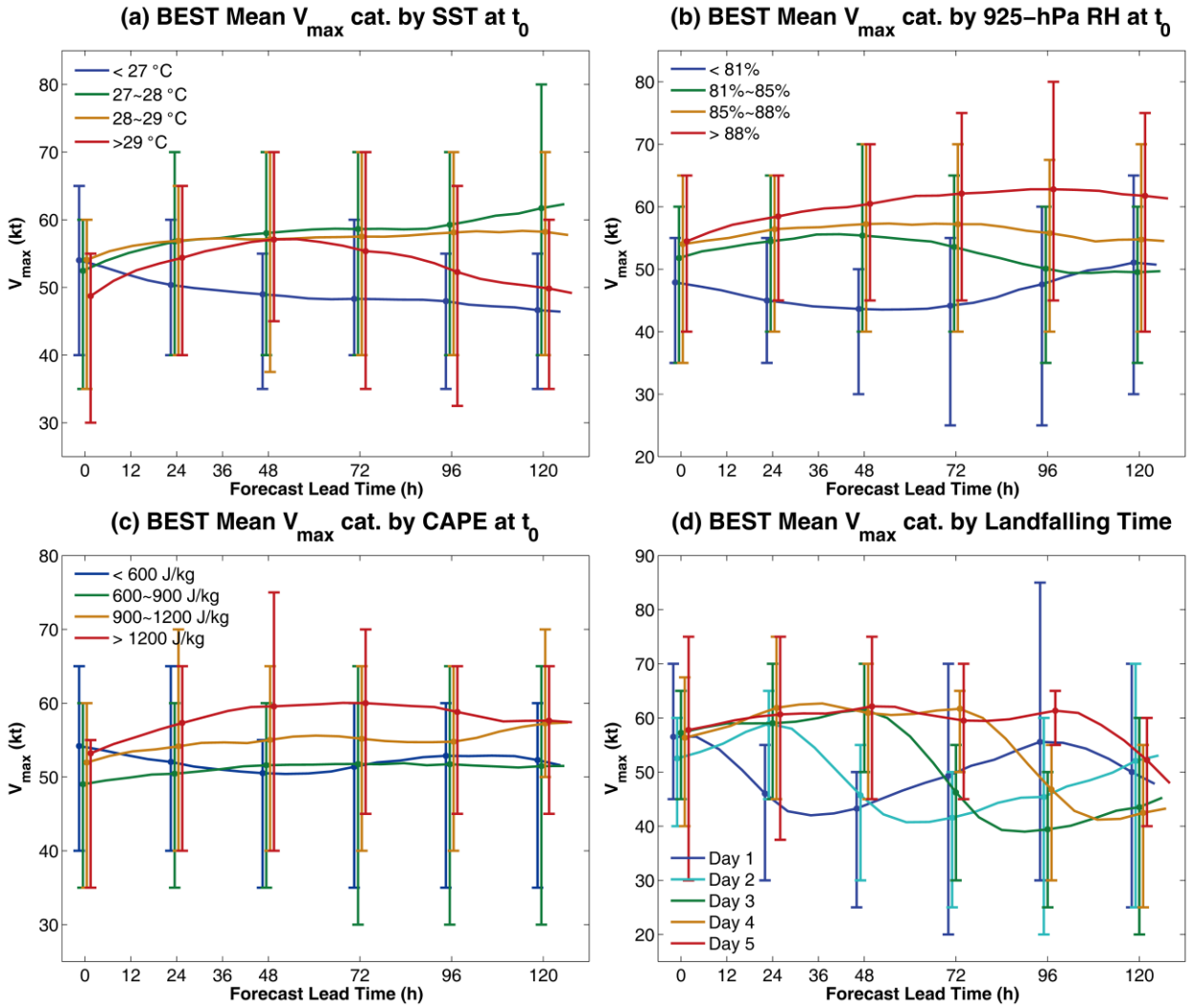

Figure 9. Similar to Figure 6, but categorized by (a) SST at forecast initialization time, (b) 925-hPa RH at forecast initialization time, (c) CAPE at forecast initialization time, and (d) days that the TC made landfall in best track during the forecast.

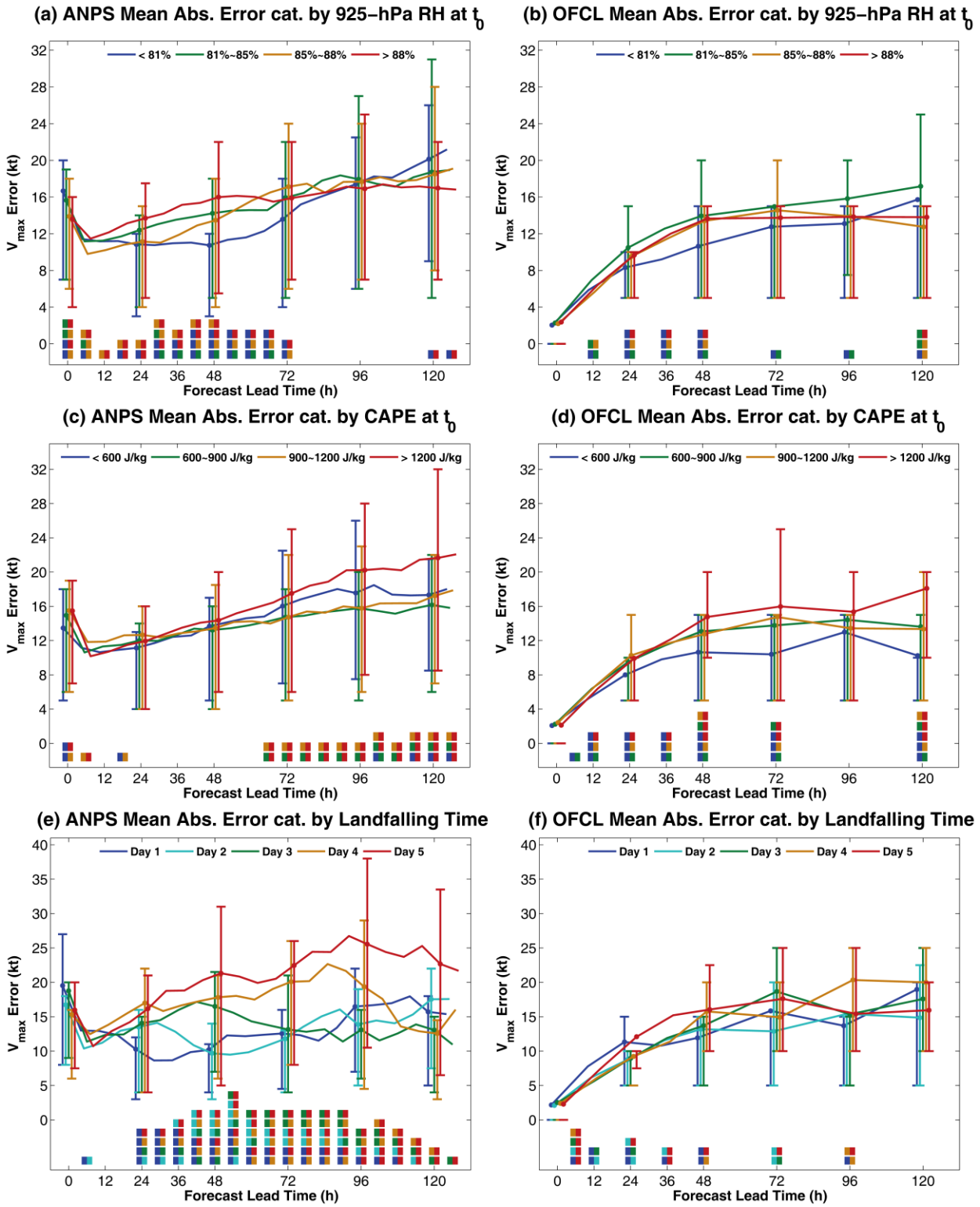

Figure 10. Similar to Figure 4, but categorized by (a) and (b) 925-hPa RH at forecast initialization time, (c) and (d) CAPE and forecast initialization time, and (e) and (f) days that the TC made landfall in best track during the forecast, for (a), (c) and (e) ANPS and (b), (d) and (f) OFCL forecasts.

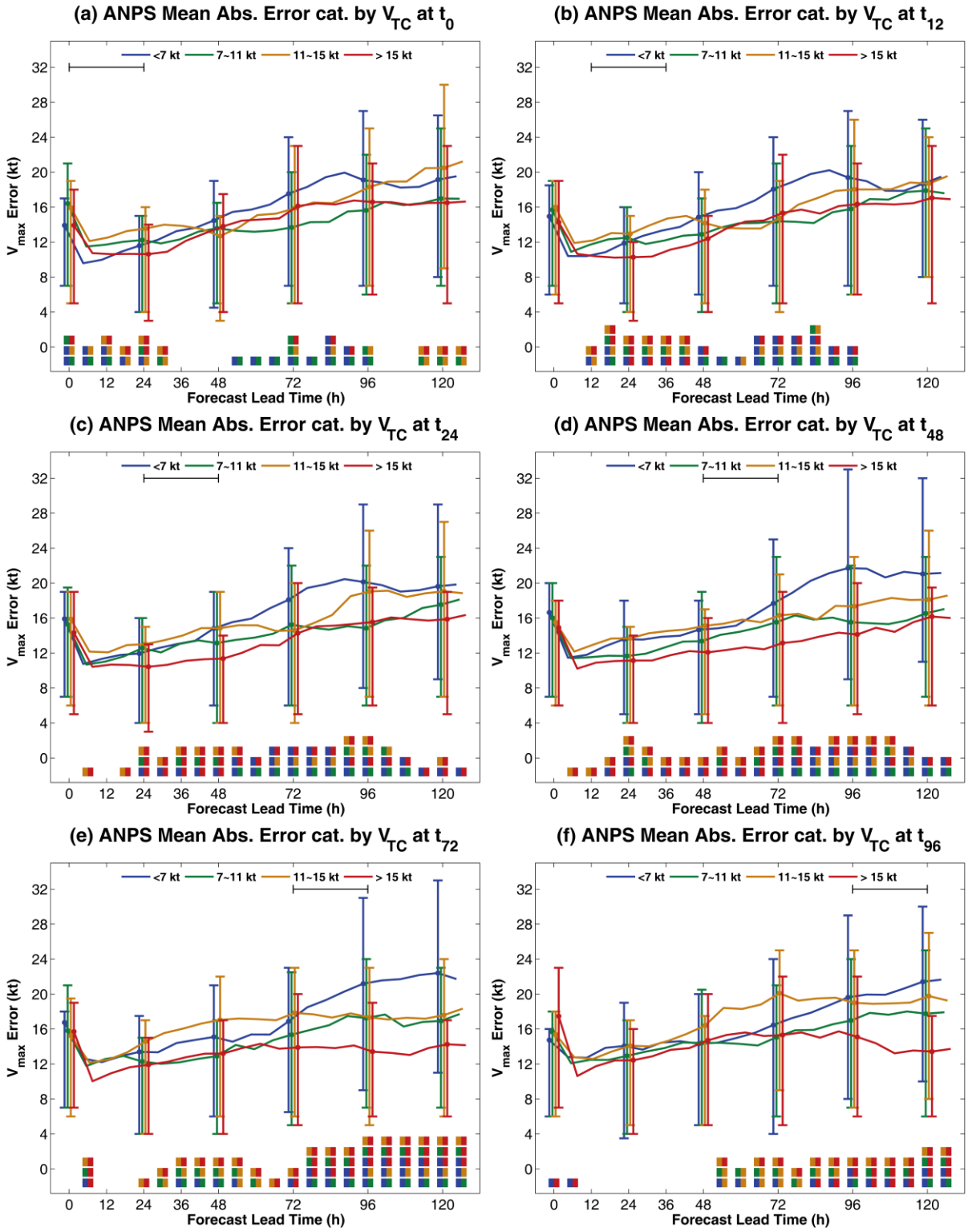

Figure 11. Similar to Figure 4, but categorized by the best-track translation speed of the TC at forecast lead-times of (a) 0 h, (b) 12 h, (c) 24 h, (d) 48 h, (e) 72 h and (f) 96 h for ANPS forecasts. In each panel, the horizontal range lines at the top indicate 0–24-h time period after categorization time.

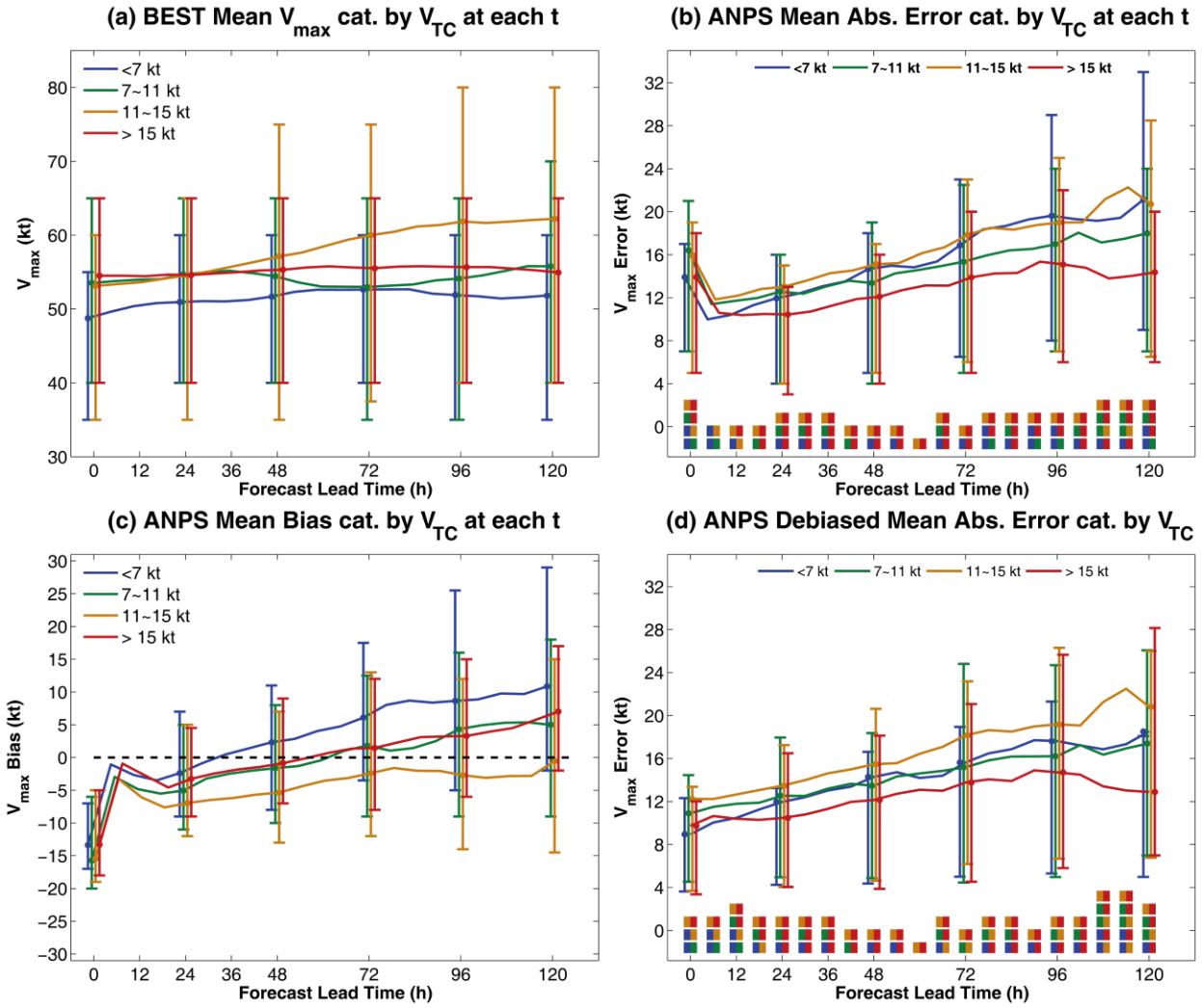

Figure 12. (a) Best track mean intensity, (b) ANPS mean absolute intensity forecast errors, (c) ANPS mean intensity forecast bias and (d) ANPS de-biased mean absolute intensity forecast errors, all categorized by translation speed of the TC at each verification forecast lead-times. In each panel, the vertical lines accompanying intensity, error or bias trends are ranges between the first and the third quartiles of each group every 24 h. In (b) and (d), the two-color squares at the bottom indicate significantly different pairs at that time.