

**Track and Intensity Forecasting of Hurricanes: Impact of Convection-Permitting
Resolution and Global Ensemble Kalman Filter Analysis on 2010 Atlantic Season Forecasts**

Ming Xue^{1,2*}, Jordan Schleif^{1,2}, Fanyou Kong¹, Kevin W. Thomas¹, Yunheng
Wang¹, and Kefeng Zhu¹

¹Center for Analysis and Prediction of Storms and ²School of Meteorology
University of Oklahoma, Norman, Oklahoma

June 6, 2012

Submitted to Weather and Forecasting

Revised November 3, 2012

*Corresponding author address:
Dr. Ming Xue

Center for Analysis and Prediction of Storms
University of Oklahoma
120 David Boren Blvd
Norman, Oklahoma, OK 73072.
mxue@ou.edu

Abstract

Twice daily 48-hour tropical cyclone (TC) forecasts were produced in real time for the fall 2010 Atlantic hurricane season using the WRF model on a single, large 4 km grid covering much of the northern Atlantic basin. Parallel forecasts were initialized from operational GFS analyses based on the GSI method, and from experimental global ensemble Kalman filter (EnKF) analyses and are compared to corresponding GFS forecasts.

For the track, WRF forecasts are significantly better than GFS forecasts using either initial condition (IC). The EnKF-initialized GFS and WRF are also better than the corresponding GSI-initialized forecasts. At all lead times, the 4-km WRF track errors are comparable to or smaller than the NHC official track forecast error, with those of EnKF WRF being even smaller. For weaker TCs, more improvement comes from the model (resolution) than from the IC. For hurricane intensity TCs, EnKF ICs produce better track forecasts than GSI ICs, with the WRF being even better at most lead times.

For intensity forecast, the EnKF ICs consistently outperform GSI ICs in both models for weaker TCs. For hurricane-strength TCs, EnKF ICs produce intensity forecasts statistically indistinguishable from GSI ICs in either model. For all TCs combined, WRF produces about half the error of corresponding GFS beyond 24 hours, and at 36 and 48 hours, the errors are smaller than NHC official forecasts. The improvement is even greater for hurricane strength TCs. Overall, the high-resolution WRF forecasts with EnKF ICs have the smallest intensity error, showing statistically significant positive impacts of convection-permitting resolution and advanced data assimilation.

1. Introduction

The National Hurricane Center (NHC) forecasts of tropical cyclone (TC) position have improved significantly since 1990, thanks to improvements in observing systems, data assimilation (DA) techniques, and numerical weather prediction models (Rappaport et al. 2009). However, despite the improvement in track forecasts, forecasts of TC intensity have not improved much. Intensity forecasts are difficult because small-scale, inner-core processes are very important when predicting changes in TC intensity, and global models typically lack the resolution necessary to resolve the intense vortex circulation in the TC inner core region. It has been hypothesized that a high-resolution grid would improve TC forecast, in particular the forecast of TC intensity.

During the fall 2010 hurricane season, the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma produced twice daily (00 and 12 UTC) 48-hour experimental hurricane forecasts in real time for the Atlantic Ocean on a single high-resolution 4 km grid using the Advanced Research WRF (WRF-ARW) (Skamarock et al. 2008). Two sets of forecasts were produced, one initialized from the operational NCEP Global Forecast System (GFS) analyses and one from the experimental global ensemble Kalman filter (EnKF) ensemble mean analyses produced by the NOAA Earth System Research Laboratory (ESRL) (Hamill et al. 2011b). These forecasts allow us to examine the impact of the high resolution and initial conditions (ICs) on hurricane forecasts, including the forecasts of track and intensity. Verification statistics for all hurricane forecasts from 11 September through 9 October 2010 were calculated against the NHC best track data. The two WRF forecasts were also compared with the two global GFS forecasts, one using the operational GSI (Grid-point Statistical Interpolation

3DVAR system, Kleist et al. 2009) analyses and one using the ESRL EnKF analyses. As a reference, we also include the official track forecast from NHC along with our track verification.

Our forecast results will be discussed in light of a limited number of earlier studies involving high-resolution realtime forecasting of TCs; they are Davis et al. (2010, D10 hereafter) and those documented in the NOAA High-Resolution Hurricane Forecast Test Report (HRHFT, DTC 2009). One important difference is that their forecasts all used relatively small, TC-following nested high-resolution grids. Their results are briefly summarized here.

The real time forecasts in D10 used a version of the WRF ARW model called the Advanced Research Hurricane WRF (AHW). It included 10 Atlantic TCs: six from 2005, and four from 2007. Their 120 h forecasts used ICs produced by a cycling regional ensemble Kalman filter run on a 36 km grid. Two sets of forecasts were made: one with a single 12 km grid, and one with triple nested 12, 4 and 1.33 km grids. The 12 km grid is fixed while the 4 and 1.33 km grids follow the TCs. D10 found that there was no meaningful difference between storm position errors in their 12 km and nested higher-resolution forecasts. However, TC intensity (in terms of the maximum 10 m wind) was somewhat better forecasted on the nested grids than on the single 12 km grid, and the difference was statistically significant at 72 h and beyond. In particular, the intensity forecast for hurricanes category 3 and stronger benefitted the most from the nested grids. The 12 km forecasts tended to exhibit a negative intensity bias for stronger TCs, and the nested grid forecasts showed a positive intensity bias for weaker TCs.

The NOAA HRHFT was a study aimed at improving hurricane intensity forecasts. Six modeling groups (from which, only five produced usable results in realtime) produced forecasts for ten tropical storms of interest from 2005 and 2007. The study found that the use of higher resolution did not necessarily lead to an improvement in TC forecasting, and further research into the use of physics schemes, ocean-atmosphere coupling, initialization, etc. is needed. The following is a summary of forecasts produced by each modeling group, according to the HRHFT report (DTC 2009).

The NOAA Atlantic Oceanographic and Meteorological Laboratory (AOML) ran the Experimental Hurricane WRF (HWRFX, Gopalakrishnan et al. 2010) on two configurations: one with 27 and 9 km nested grids; the other with 9 and 3 km nested grids. The track error of configuration 1 increased more quickly than configuration 2, and the difference was statistically significant between 30 h and 48 h. Both configurations tended to under-predict TC intensity at lead times up to 48 h, with configuration 2 exhibiting a smaller negative bias. Configuration 2 showed a statistically significant improvement in intensity forecasts at 6, 24, and 30 h lead times.

The Mesoscale and Microscale Meteorology (MMM) division of the National Center for Atmospheric Research (NCAR) ran the Advanced Hurricane WRF (AHW, Davis et al. 2008) in two configurations: one with a single 12 km grid, and another which adds storm-following 4 km and 1.33 km nested grids. Like AOML, the track error for configuration 1 increased more quickly than configuration 2, with a statistically significant improvement between 84 and 114 h lead times. TC intensity was under-predicted using configuration 1, and over-predicted with configuration 2 at early lead times, with statistically significant improvement only at the 18 h lead time.

The Naval Research Laboratory (NRL) ran a TC-optimized version of the Coupled Ocean/Atmosphere Mesoscale Prediction System (Hendricks et al. 2011) model in two configurations: one having 81, 27, and 9 km nested grids, and the other adds a 3 km grid. The track errors for each configuration increased at different rates, leading to the 3 km configuration being significantly better at 24 h, while the 9 km configuration was significantly better at 24, 42, 54, and 96 h. Both configurations under-predicted TC intensity, however, the absolute intensity error in the 3 km configuration was significantly better at 6, 24, and 48 h lead times, due to a reduction in negative wind speed bias.

The University of Rhode Island (URI) used the Geophysical Fluid Dynamics Laboratory (GFDL) ocean-atmosphere model (Bender et al. 2007) with two configurations. The first used a static grid with 1/2 degree spacing, and two TC-following grids with 1/6 and 1/12 degree spacing, and the second replaced the 1/12 degree grid with a 1/18 degree grid. The track errors of both configurations were similar, except for cross track error improvement in the higher-resolution configuration at 48 and 72 h lead times. Both configurations tended to under-predict TC intensity at early lead times, and over-predict intensity at lead times beyond 4 days. The difference in intensity error was not statistically significant.

The University of Wisconsin-Madison (UWM) group ran the University of Wisconsin Nonhydrostatic Modeling System (UW-NMS, Tripoli 1992) with two configurations. The first used a single 12-km grid, and the second added a nested, TC-following 3 km grid. The track errors favored the low-resolution forecasts at one lead time (78 h) but were statistically indistinguishable elsewhere. The low-resolution

forecasts under-predicted TC intensity at most lead times, whereas the high-resolution forecast over-predicted intensity at short lead times and under-predicted intensity at long lead times. The high resolution forecast was statistically favorable in the 48 to 84 h time frame.

As can be seen, results from the above modeling groups, using grid spacings down to 1.33 km, are inconclusive with respect to the prediction errors of both TC track and intensity. The high-resolution grids improve the forecasts sometimes and degrade the forecasts at other times. The use of relatively small nested high-resolution grids in these studies might have hampered the expected positive impact of the convection-permitting/resolving resolutions. In the next section, we present results of our forecast experiment, and will discuss the results in reference to the findings of these earlier studies when appropriate.

The rest of this paper is organized as follows. Section 2 will discuss the forecast models and their configuration, as well as the initial and boundary conditions used. Section 3 will present the intensity and track verification results, and section 4 will provide a summary and conclusions.

2. Forecast models and configurations

The WRF-ARW version 3.1 was used in the CAPS forecasts. The forecast grid has 1801×901 grid points and is centered at (24°N , 63°W), spanning 70×30 degrees in longitude-latitude over the North Atlantic Ocean. The grid has a 4 km spacing in the horizontal, and 51 vertical levels (see Figs. 1 and 2). This large 4 km domain was chosen to cover most of the tropical Atlantic, to track TCs from their genesis through possible

recurvature or landfall, and the use of a single grid avoids complication and uncertainty related to the use of multiple nested grids and the associated, often large, lateral boundary condition influence. The use of a single grid permits a cleaner comparison between WRF and the global model running at two different resolutions, while at the same time noting the fact that we are using the WRF model, not the GFS model so forecast difference can be due to both model and resolution. A 4 km resolution has been successfully used for continental storm-scale realtime forecasting for several years (Xue et al. 2010) and has been shown to produce *second-day* forecasting guidance similar to corresponding 2 km forecasts (Schwartz et al. 2009) but is much better than the same model run at 20 km resolution (Clark et al. 2009). A 4 km resolution is often referred to as convection-permitting or marginally cloud-resolving.

The WRF model was configured in the following way. It employs the Thompson microphysics scheme (Thompson et al. 2006; Thompson et al. 2008), Goddard shortwave radiation parameterizations (Chou et al. 1998; Chou and Suarez 1999), the rapid radiative transfer model (RRTM) for longwave radiation (Mlawer et al. 1997), Noah land-surface model (Ek et al. 2003), Mellor-Yamada-Janjić boundary layer physics (Janjić 1990), no cumulus scheme, and monotonic moisture advection (Skamarock and Weisman 2008).

Two WRF-ARW forecasts were initialized each day at 00 UTC and 12 UTC: one using the operational GFS analysis and forecasts to provide the initial and lateral boundary conditions, respectively (referred to as WRF-GSI, where GSI is the Gridpoint Statistical Interpolation 3DVAR DA system used by GFS), and another using the experimental global EnKF ensemble mean analysis produced by the NOAA ESRL (Whitaker et al. 2008; Hamill et al. 2011b) as the IC and the corresponding deterministic

GFS model forecasts for lateral boundary conditions (referred to as WRF-EnKF). The global EnKF analysis assimilates the same observations as the GSI analysis, plus NHC advisory minimum sea level pressure (TCVitals, Hamill et al. 2011a; Hamill et al. 2011b). The GSI analysis employs vortex relocation to move the TC from its analyzed position to its actual location (Liu et al. 2000).

Within the global EnKF DA cycles, the GFS ensemble was run at T254L64 resolution (~47 km at 25°N). The ensemble mean EnKF analyses (Hamill et al. 2011b) was used to initialize the GFS deterministic forecasts (GFS-EnKF), run at T574L64 resolution (~21 km at 25°N). The operational GFS model forecasts based on GSI 3DVAR ICs (Kleist et al. 2009) also had T574L64 resolution (referred to as GFS-GSI). The WRF-GSI and WRF-EnKF forecasts will be evaluated together with the GFS-GSI and GFS-EnKF forecasts.

The primary difference between the 3DVAR and EnKF methods is with their determination of background error covariance (e.g., Li et al. 2012). The GSI 3DVAR uses static background error covariance derived from historical forecasts using the so-called NWC method (Parrish and Derber 1992). Hence, the covariance is basically unaware of the presence of tropical cyclones in the forecast background, and is unable to produce dynamically consistent TC analyses unless comprehensive TC vortex-scale observations are available. The EnKF, however, derives the background error covariance from a forecast ensemble that is specific for the analysis time, and this flow-dependent, TC-aware error covariance gives EnKF the ability to produce dynamically consistent TC analyses from a limited number of observations. For example, EnKF is able to correct errors in the wind field by using the TC minimum sea-level pressure advisory data

(Hamill et al. 2011b).

Figure 1 shows an example of 48-hour WRF-EnKF (Fig. 1a) and WRF-GSI (Fig. 1b) forecasts, with the forecast tracks overlaid, as compared to the best track and GFS analysis (Fig. 1c) valid at the same time. This example shows that the EnKF-initialized WRF predicted the path of Hurricanes Karl and Julia more accurately than the GSI-initialized WRF, while the projected tracks of Hurricane Igor were similar. Another feature to note is the capability of the high-resolution forecasts to resolve greater wind speeds than coarse GFS fields, and this is particularly evident in Hurricane Igor's wind field. Figure 2 shows sample model output of sea level pressure, reflectivity, and streamlines from the convection-permitting WRF-EnKF forecast at 48 hours, that captures the eyewall and rainband structures within three TCs, in particular the intense eyewall within Igor (Fig. 2b), which was of category 4 at this time. The CAPS WRF forecasts were produced for seven Atlantic TCs, Igor through Otto, from 11 September through 9 October 2010.

The WRF and GFS forecasts were verified against the best track data from NHC. The two WRF forecasts will be compared to each other and to the GFS-GSI and GFS-EnKF forecasts to see how the grid resolution and ICs impact TC track and intensity forecasts. In addition, the forecasts were split into two groups: those for weaker cyclones initially at tropical depression or tropical storm strength, and those for stronger cyclones initially at hurricane strength. This is to determine how the ICs and the high-resolution grid affected forecasts for TCs of different initial intensity.

Throughout this paper, statistical significance is determined by using block bootstrap resampling, as in Hamill et al. (2011a). Because different forecasts for a

particular TC are correlated, each set of TC forecasts (for Igor, Julia, etc.) are put into a block, and blocks are randomly selected 1000 times with replacement to construct a new set of bootstrap observations. From this sample, the mean is calculated, along with a 90% confidence interval. To determine whether the difference between two means is statistically significant, one set of bootstrap samples is subtracted from the other, and a mean difference and 90% confidence interval is computed. If the mean difference confidence interval does not include zero, the difference is statistically significant at the 10% level. These mean differences and confidence intervals are shown as an example in Fig. 4 for absolute track error. Mean differences for other verification statistics were calculated and their statistical significance is noted but not shown in figures for clarity.

3. Results

a. Absolute track error

Absolute track error (ATE) is defined as the great circle distance between the forecast position and the best track position of a TC center. In Fig. 3, the ATEs averaged over all TCs are shown for each of the four models from the initial analysis time (0 h) every six hours until 48 h. In the initial time, the operational GFS analyses are better at determining the initial position of the TC center positions compared to EnKF ICs, apparently due to the use of the vortex relocation technique (Liu et al. 2000). While the EnKF analyses also benefited from the assimilation of advisory minimum sea level pressure (MSLP) observations (TCVitals, Hamill et al. 2011a; Hamill et al. 2011b), the EnKF does not force the analyzed location to exactly match that of best track. In the 6 h forecast, the track error of both WRF forecasts is slightly worse than either of the global forecasts, which may be due to an adjustment caused by interpolating the global analysis

to the high-resolution grid. However, in the 12 h forecast and beyond, the error growth for the two global forecasts exceeds that of the WRF forecasts. These results are similar to the AOML and MMM experiments in the HRHFT (DTC 2009), which report that track error in low-resolution forecasts increase faster than in high-resolution forecasts, while the other HRHFT experiments and D10 showed little change or degradation of track forecasts using a high-resolution grid.

At 48 h, WRF-EnKF forecasts have the smallest ATE while GFS-GSI forecasts have the largest ATE, about 50% larger than that of WRF-EnKF. This difference is statistically significant. The ATE uncertainty (as indicated by the error bars) for the GFS-GSI forecasts is similar in size to that of WRF-GSI forecasts, while the GFS-EnKF forecasts have a larger uncertainty than WRF-EnKF forecasts, based on block bootstrap resampling (Hamill et al. 2011a). Between 24 and 42 h, both WRF forecasts outperform the GFS-GSI forecasts, and this difference is statistically significant at 24 and 30 h. Compared to the NHC official track forecasts, the WRF forecast ATEs are comparable or smaller (except for WRF-GSI at 48 h), while the global forecast ATEs are comparable to or higher than the official forecasts at all four times.

To better delineate the impacts of forecast models and the DA methods, the differences in mean ATE between forecast pairs are plotted in Fig. 4. At 6 h, WRF performs worse than the GFS model. At 12 h and 18 h, WRF performs better than GFS, but the difference between models is smaller than the difference between EnKF and GSI ICs when using the same model. In the 24 through 42 h forecasts, WRF does much better at predicting the TC tracks than GFS, and the benefit of convection-permitting resolution clearly outweighs the benefit of advanced EnKF DA. The much larger absolute error

differences between forecasts using GSI and EnKF ICs in both WRF and GFS at 48 h (see also Fig. 3) are actually due to the poor predictions of Hurricane Lisa when initialized using GSI (not shown). Overall, the improved track forecasts in WRF-EnKF, especially when compared with GFS-GSI, show the benefit of using the high-resolution WRF model with EnKF ICs.

To see further the track forecast performance for TCs of different intensity, the ATE was calculated separately for TCs above and below the hurricane strength threshold at the initial time (Fig. 5). For weaker TCs, WRF-EnKF performs worst before 18 h. By 18 h, WRF-EnKF ATE decreases and becomes the second-best performer next to GFS-EnKF, and the difference is statistically significant compared to GFS-GSI. From 24-36 h and at 48 h, the high-resolution WRF produces smaller track errors than GFS. Forecasts from the same model are comparable using either IC: EnKF is slightly better some of the times and worse at other times (Fig. 5a).

For TCs of hurricane strength, the advantage of the EnKF IC is much clearer, with the EnKF-initialized forecast errors being always smaller than the GSI-initialized forecast errors with the same model. The improvement from using EnKF ICs in the GFS (WRF) model is statistically significant at 48 h (12 h). There is also a clear advantage with WRF-EnKF over GFS-EnKF, especially from 12 to 30 h when the difference is statistically significant. For hurricane-strength TCs, there is little uncertainty about the initial center locations – the initial track errors in both GSI and EnKF are about 20 km (Fig. 5b). The initial track errors are larger for weaker TCs, ranging between 50 and 70 km, consistent with the larger difficulty locating the center of circulation in weak TCs. The mean track forecast errors for weaker TCs are also larger, exceeding 250 km in GFS-GSI forecasts at

48 hours, while those of stronger hurricanes stayed below 200 km. Structural asymmetry and larger influence of environmental steering flows are thought to be the cause of the larger ATE with weaker TCs.

b. Along and cross track error

As illustrated in Fig. 6, along track error (ATE) is defined as the component of absolute error in the direction of the actual track of the TC. Similarly, cross track error (CrTE) is defined as the component of absolute error in the direction perpendicular to the actual track.

The average absolute along and cross track errors for TC tracks are plotted in Fig. 7. Between 12 and 36 hours, the GFS forecasts have larger ATEs than the WRF forecasts, with the difference being largest at 24 h (about 50% larger, Fig. 7a). At the later times, the differences become smaller and the GFS actually produced slightly smaller ATEs than WRF. Neither IC exhibits a clear advantage in either model. The CrTEs of the 4 model forecasts are similar before 30 h (Fig. 7b), with no clear advantage with any one forecast. After 30 h, the GFS-EnKF forecasts are somewhat sporadic: its CrTE becomes the smallest at 36 h, rises to a maximum at 42 h, then drops below the two GSI-based forecasts at 48 h. In general, cross track errors are larger than along track errors, especially for longer lead times. This indicates that forecasts have more difficulty determining the direction of a TC than its speed of movement.

At the 48 h lead time, both WRF-GSI and WRF-EnKF forecasts are slightly faster than, and to the left of, the best track (Fig. 8). This is the same as found by D10 for both of their 12 km and nested grids, for all lead times except 120 h. Figure 8 further shows that the track error is predominantly in the cross track direction, i.e., forecasts have more

difficulty determining the direction of a TC than its speed of movement. At 48 h, the standard deviation of CrTE in GSI (EnKF) initialized WRF forecasts is more than 70% (20%) larger than that of AITE. The increased resolution of WRF forecasts tends to decrease the occurrence times of large cross track errors, which is very important when forecasting the track of a potentially landfalling TC.

c. Absolute wind speed error

The use of high-resolution WRF has the potential to significantly reduce TC intensity forecast error. Fig. 9 shows the absolute errors in the TC maximum 10 m wind speed from the four model forecasts. It is clear that EnKF does a much better job in analyzing the initial TC intensity in terms of the maximum surface wind than the GSI analyses. The global EnKF analyses include TC Vitals estimates of cyclone locations and minimum central pressure (Hamill et al. 2011b), which, through flow-dependent cross covariance, can directly update the wind fields and reduce surface wind speed error. However, the absolute wind speed error (AWSE) grows rapidly in GFS-EnKF from about 6 m s^{-1} at the initial time to above 10 m s^{-1} at 6 h. This is due to the inability of the coarse resolution GFS model to support intense vortices analyzed by the EnKF using TC Vitals data (Hamill et al. 2011a). This rapid error growth is also observed in WRF-EnKF, but at a somewhat slower rate. Direct assimilation of TC Vitals data on the high-resolution WRF grid may help (but is not done here). After the initial adjustment period, the AWSE remains above 10 m s^{-1} in GFS forecasts, with the errors of GFS-EnKF remaining lower than GFS-GSI except for the final time. In contrast, the AWSE in WRF forecasts decreases slightly with time until 42 h, with the errors of WRF-EnKF being smaller than those of WRF-GSI except at 6 h and 12 h. The reduction in wind speed forecast error

with time indicates the ability of the high-resolution WRF model to spin up strong TCs that may not have been properly analyzed at the initial time, producing a dynamically more balanced cyclone. By contrast, the GFS forecasts lack the resolution necessary to produce a dynamically balanced cyclone of realistic intensity, resulting in increasing error differences between the GFS and WRF forecasts. By 48 h, the GFS-EnKF error is almost twice that of WRF-EnKF and the GFS-GSI error is about 50% more than the WRF-GSI error.

Overall, the AWSEs of WRF forecasts are much smaller than those of GFS forecasts, with statistical significance, while the EnKF-initialized forecasts are slightly better than corresponding GSI-initialized forecasts, but the differences are not statistically significant. For comparison, three experiments in the HRHFT (DTC 2009) exhibited statistically significant improvement in wind speed error at some lead times when using a high-resolution grid. The AOML 3 km experiment exhibited improvement for lead times between 30 and 48 h. The NCAR MMM 1.33 km experiment only saw significant improvement at the 18 h lead time. In the UWM experiment, the 3 km experiment showed improvement at 48 to 84 h lead times. Neither D10 nor the remainder of the HRHFT experiments found a statistically significant improvement in intensity forecasts.

The wind speed forecast errors for strong and weak TCs evolve quite differently. After 6 h, the error generally increases with time for tropical depressions and tropical storms (Fig. 11a), but generally decreases with time for hurricane intensity TCs (Fig. 11b), especially before 36 h. For the weaker TCs, the wind speed error remains at similar levels in the first 6 hours, except for WRF-GSI, in which the error decreases noticeably due to spin-up. There are also small decreases in wind speed error in WRF-EnKF out to

18 h. These are indications of vortex spin-up in the high resolution model, starting from initial vortices that are somewhat too weak. Similar spin-up does not happen in the GFS forecasts due to the lack of resolution, and the error increases monotonically with time (Fig. 11a) except at 42 h. The WRF errors also increase with time after the initial spin-up period. For the weak TCs, the wind speed errors are consistently ranked in a descending order for GFS-GSI, GFS-EnKF, WRF-GSI and WRF-EnKF after 12 h, again showing the benefits of both high resolution and EnKF DA. At 36 through 48 h, WRF-EnKF forecasts for weak TCs perform the best, especially compared to GFS-GSI where the difference is statistically significant.

For hurricane intensity TCs (Fig. 11b), there is a large jump in the wind speed error from an initially low level to more than twice as much at 6 h in both EnKF-initialized forecasts. As discussed earlier, this is mainly due to the low resolution at which the EnKF analyses were produced (at T256L64, ~48 km at 25 °N), and the lack of full dynamic consistency and balance among state variables in the ICs. In stronger TCs, assimilation of TC Vitals observations in the EnKF analysis creates a central pressure that is in a good agreement with the minimum sea level pressure but not necessarily in a very good balance with the vortex circulation and temperature field (even although it can do much better than 3DVAR can). Therefore, the hurricanes undergo a spin-down process before they are spun up again. The spin-up is especially clear in the WRF forecasts: the errors decrease from 12-13 m s^{-1} at 6 h to 6-7 m s^{-1} at 36 h before they increase again (Fig. 11b). A similar pattern happens in the GFS forecasts, except that their error levels are higher, and there is no initial spin-down in GFS-GSI. Overall, the stronger TCs benefit much more from higher resolution than the weaker TCs for intensity forecast, and

at 48 h, the WRF-EnKF forecast error is slightly smaller for hurricanes than for weaker TCs. This is different from the results of D10, which showed that the wind errors for weak TCs were larger than those for strong TCs on a 1.3 km nested grid.

TC wind speed as forecasted by global models is known to have a large negative bias; that is, the maximum wind speed is forecasted to be much lower than the best track wind speed. This is due to the inability of global models to resolve small scale TC structures and properly capture their intensity changes. The wind speed biases of the four sets of forecasts are plotted in Fig. 10. Both GFS forecasts have consistently negative biases of about 10 m s^{-1} or larger at all forecast times, and the GFS-EnKF forecast biases are slightly smaller than those of GFS-GSI except at 48 h. Starting from 6 h, the WRF forecast biases are generally 2/3 to 1/3 of the GFS forecasts, and the EnKF-initialized biases are smaller than the GSI-initialized forecasts for the WRF model between 6 and 42 h. At 42 h, the negative bias of WRF-EnKF is about 2 m s^{-1} , well within the best track wind speed estimation uncertainty. The results of D10 for wind speed bias on a 12 km grid are similar: there is a negative bias through 48 h lead time. However, their 1.33 km nested grid had a positive bias. The AOML experiment in HRHFT yielded under-forecasting of intensity for both low- and high-resolution forecasts at all lead times up to 48 h, however, the negative bias was reduced in high-resolution forecasts. Both MMM and UWM systems resulted in under-forecasting (over-forecasting) of TC intensity in low- (high-) resolution forecasts.

d. Absolute minimum sea level pressure error

The minimum sea level pressure (MSLP) gives another measure of TC intensity; we look at both MSLP and maximum surface wind speed because conclusions based on

the two are not always the same, at least quantitatively. The mean absolute MSLP errors are shown in Fig. 12. In our case, the MSLP errors tell a similar story to the wind speed error in general, except that the GFS-EnKF forecast errors are higher than those of GFS-GSI from 18 h onward, but they are statistically indistinguishable. Both EnKF-based forecasts went through error increase before 6 h due to spin-down. Beyond 6 h, the GFS forecast errors show an overall increasing trend while the WRF forecast errors show an overall decreasing trend. The error differences between GFS and WRF runs are clearly statistically significant while those between the GSI and EnKF forecasts of the same model are not. We note that most of the WRF MSLP forecast errors are below 8 hPa in day 2 forecasts. At 48 h, the WRF MSLP forecast error uncertainty is smaller than those of global forecasts, and the WRF forecasts cut the MSLP error in half, compared to the global forecasts.

The MSLP forecast bias is related to the wind speed forecast bias. The large negative wind speed bias for global models in Fig. 10 is accompanied by a large positive MSLP bias in Fig. 13. The high-resolution WRF model reduces the MSLP forecast bias to smaller than 3 hPa from 30 h onward, compared to over 10 hPa in the global forecasts over the same period. Figure 13 also shows that the MSLP biases are larger in the EnKF-initialized forecasts in both GFS and WRF models at the later times, despite generally larger (though statistically insignificant) absolute MSLP errors with WRF-GSI than WRF-EnKF (Fig. 12). This suggests that there are more MSLP error cancelations in WRF-GSI forecasts than in WRF-EnKF when bias is calculated.

The MSLP forecast bias was also calculated separately for weak and strong TCs (Fig. 14). For tropical storms and tropical depressions, the MSLP bias for both GSI and

EnKF ICs is initially positive and under 5 hPa. After staying relatively constant for the first 24 h, GFS-GSI and GFS-EnKF MSLP biases grow to over 10 hPa at 48h. Meanwhile, WRF-GSI and WRF-EnKF biases decrease to near zero at 42 h, before increasing back to near 5 hPa at 48h (Fig. 14a). For hurricanes, the initial positive bias is larger (near 10 hPa for EnKF ICs, and 13 hPa for GFS ICs) because the resolution of the analyses is too coarse to resolve the actual intensity of the hurricane. For all lead times, the MSLP bias for GFS-GSI stayed between 12 and 15 hPa, and the bias for GFS-EnKF increased to about 15-17 hPa from 12 h onwards. For the high-resolution forecasts, WRF-GSI biases generally decrease with time, becoming negative at 30 h, and decreasing to -5 hPa at 48 h. WRF-EnKF bias increases at first as the model variables become more dynamically consistent, and then decreases to near zero at 48 h (Fig. 14b). The high-resolution model predicts hurricanes of more accurate intensity, with small biases at longer lead times.

4. Summary and conclusions

This study examines the impacts of high, convection-permitting model resolution and EnKF data assimilation on the track and intensity forecasting of 2010 tropical cyclones in the Atlantic basin. The twice daily, 48 h forecasts used the WRF-ARW model and a single large 4 km grid covering much of the North Atlantic. In addition, two sets of ICs were used for parallel WRF forecasts: one set is the NCEP operational GFS analyses (produced by the operational GSI 3DVAR data assimilation scheme), and another is the experimental ESRL global EnKF ensemble mean analyses. The use of the 4 km high-resolution grid was hypothesized to improve hurricane forecasting, particularly intensity forecasting, due to the ability of the high-resolution grid to resolve the inner core

structures and processes important to the forecasting of TC intensity. The single large grid used also avoids the complications (e.g., discontinuities and interactions across domain boundaries; the initialization of fine grids using coarse resolution solutions when the grid moves) and uncertainties (e.g., domain movement and effects on track forecasting) involved with the use of multiple movable nested grids. The use of a single grid also allows the representation of TCs' environment entirely on the high-resolution grid. Such a practice was also hypothesized to help improve track forecasting. The use of two sets of ICs in the 4 km WRF forecasts and the availability of global GFS forecasts from the same sets of ICs, provide us with an opportunity to evaluate the impacts of the 3DVAR-based operational global analyses and experimental EnKF global analyses on forecasts at both low convection-parameterizing and high convection-permitting resolutions. Using best track data from the National Hurricane Center, verification statistics were calculated for both sets of WRF forecasts, and compared with those of GFS model forecasts. The main conclusions are listed below. Overall, the findings are encouraging, and demonstrate that the use of a high-resolution model in hurricane forecasting is an avenue worthy of further exploration while theoretically more advanced EnKF data assimilation method does perform better at both coarse global-model and high convection-permitting resolutions.

- Significant improvement to track forecasts was observed in the high-resolution WRF forecasts initialized with EnKF, compared to global GFS model forecasts initialized with GSI. This difference is significant from 12-30 h (every six hours) and at 48 h. There is a slight improvement in the track forecast in EnKF-initialized GFS over GSI-initialized GFS at all lead times for all TCs combined, but the improvement is not

statistically significant. At all lead times 12 hours apart, WRF-EnKF forecast track errors are smaller than the NHC official forecasts. Most of the error in track forecasts is found to be due to cross track error, indicating a greater difficulty pinpointing the direction of the TC movement than the forward speed.

- For weaker TCs below hurricane strength, more improvement in the track forecasts comes from the high resolution from 24 hours on, while the differences due to the ICs are smaller in both WRF and GFS models. While EnKF ICs improve the track forecasts for weak TCs in both models at 12 and 18 h, GSI ICs produce smaller errors initially and at 36 and 42 h. For hurricane intensity TCs, the EnKF ICs improve both GFS and WRF forecasts, and high-resolution WRF track forecasts are better than GFS forecasts except at 48 h.
- The high-resolution WRF forecasts produce significantly improved intensity forecasts at 24 hours and beyond, both in terms of maximum 10 m wind speed and minimum sea level pressure. For all TCs combined together, the maximum surface wind speed errors the high-resolution WRF are about one-third smaller than those of GFS forecasts beyond 24 hours. The error differences are even larger for TCs of hurricane strength, with the wind speed errors in the high-resolution forecasts being less than half of the GFS forecast error. An improvement in intensity for weak TCs is also observed in the high-resolution forecasts, but the improvement is not as dramatic as that for strong TCs. EnKF ICs improve TC intensity forecasts in both GFS and WRF models, for hurricane-strength TCs and all TCs together.
- Global forecasts have negative wind speed biases, underestimating hurricane intensity, while the high-resolution WRF forecasts reduce this bias by up to two-

thirds compared to GFS forecasts. The GSI-initialized WRF over-intensifies hurricane-strength TCs in terms of MSLP error at the 30 h lead time and beyond.

- For all TCs combined together, the EnKF ICs improve the wind speed forecasts in both GFS and WRF models for most of the lead times compared to GSI ICs, but the improvements are much smaller than those due to increased model resolution. Intensity improvements associated with ICs, both in terms of wind speed and MSLP, are found to be mostly associated with weaker TCs.

Further studies are possible to explore the impacts of resolution and data assimilation on hurricane forecasting. The spin-down followed by spin-up observed in this study from dynamically imbalanced TCs represented in the EnKF ICs leads to the question of whether performing EnKF data assimilation directly on the high-resolution grid would lead to a better intensity forecast, especially at shorter lead times. Forecasts of longer ranges than the 48 h forecasts examined here should also be explored, which would require more computational resources. Related questions of interest to the TC forecasting community include if a high-resolution grid improves forecasts of TC genesis and rapid intensification. Other avenues worthy of study include which physics packages are best for high-resolution TC forecasting, and examining the potential benefit of ocean-atmosphere coupling.

Similar forecasts are being carried out for the Pacific basin. Large forecast samples will also increase the statistical robustness of the conclusions. Even higher convection-resolving resolutions as well as high-resolution assimilation of any available TC inner core observations may be beneficial to further improving the intensity or even track forecasting. These are topics for future research.

Acknowledgments This research was primarily supported by grants from the Office of Naval Research (ONR) Defense EPSCOR program N00014-10-1-0133 and ONR funding N00014-10-1-0775. The global EnKF runs were supported by the NOAA THORPEX and Hurricane Forecast Improvement Program (HFIP) to ESRL. The WRF forecasts were carried out at the University of Oklahoma Supercomputing Center for Education and Research (OSCER). A mass storage system at the Pittsburgh Supercomputer Center was used for data archival. The first author was also supported by NSF grants AGS-0802888, OCI-0905040, AGS-0941491, AGS-1046171, AGS-1046081. Wei Li performed the data archiving, Xuguang Wang and Keith Brewster provided input on the forecast grid configuration.

References

- Bender, M. A., I. Ginis, R. Tuleya, B. Thomas, and T. Marchok, 2007: The Operational GFDL Coupled Hurricane–Ocean Prediction System and a Summary of Its Performance. *Mon. Wea. Rev.*, **135**, 45.
- Chou, M.-D. and M. J. Suarez, 1999: A shortwave radiation parameterization for atmospheric studies, 15, NASA/TM-104606, 40 pp.
- Chou, M.-D., M. J. Suarez, C.-H. Ho, M. M. H. Yan, and K.-T. Lee, 1998: Parameterizations for cloud overlapping and shortwave single-scattering properties for use in general circulation and cloud ensemble models. *J. Clim.*, **11**, 202-214.
- Clark, A. J., W. A. Gallus, Jr., M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small convection-permitting and large convection-parameterizing ensembles. *Wea. Forecasting*, **24**, 1121-1140.
- Davis, C., W. Wang, J. Dudhia, and R. Torn, 2010: Does Increased Horizontal Resolution Improve Hurricane Wind Forecasts? *Wea. Forecasting*, **25**, 1826-1841.
- Davis, C., W. Wang, S. S. Chen, Y. S. Chen, K. Corbosiero, M. DeMaria, J. Dudhia, G. Holland, J. Klemp, J. Michalakes, H. Reeves, R. Rotunno, C. Snyder, and Q. N. Xiao, 2008: Prediction of landfalling hurricanes with the Advanced Hurricane WRF model. *Mon. Wea. Rev.*, **136**, 1990-2005.
- DTC, 2009: Developmental Testbed Center High Resolution Hurricane Test Final Report, http://www.dtcenter.org/plots/hrh_test/HRH_Report_30Sept.pdf.
- Ek, M. B., K. E. Mitchell, Y. Lin, P. Grunmann, E. Rogers, G. Gayno, V. Koren, and J. D. Tarpley, 2003: Implementation of the upgraded NOAA land-surface model in the

- NCEP operational mesoscale Eta model. *J. Geophys. Res.*, **108**, 8851,
doi:10.1029/2002 JD003296.
- Gopalakrishnan, S. G., F. Marks Jr., X. Zhang, J.-W. Bao, K.-S. Yeh, and R. Atlas, 2010:
The Experimental HWRF System: A Study on the Influence of Horizontal
Resolution on the Structure and Intensity Changes in Tropical Cyclones Using an
Idealized Framework. *Mon. Wea. Rev.*, **139**, 23.
- Hamill, T. M., J. S. Whitaker, M. Fiorino, and S. G. Benjamin, 2011a: Global ensemble
predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter.
Mon. Wea. Rev., **139**, 668-688.
- Hamill, T. M., J. S. Whitaker, D. T. Kleist, M. Fiorino, and S. G. Benjamin, 2011b:
Predictions of 2010's tropical cyclones using the GFS and ensemble-based data
assimilation methods. *Mon. Wea. Rev.*, **139**, 3243-3247.
- Hendricks, E. A., M. S. Peng, X. Ge, and T. Li, 2011: Performance of a Dynamic
Initialization Scheme in the Coupled Ocean–Atmosphere Mesoscale Prediction
System for Tropical Cyclones (COAMPS-TC). *Wea. Forecasting*, **26**, 650-663.
- Janjić, Z. I., 1990: The step-mountain coordinate: Physical package. *Mon. Wea. Rev.*, **118**,
1429-1443.
- Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009:
Introduction of the GSI into the NCEP global data assimilation system. *Wea.*
Forecasting, **24**, 1691-1705.
- Li, Y., X. Wang, and M. Xue, 2012: Assimilation of radar radial velocity data with the
WRF ensemble-3DVAR hybrid system for the prediction of hurricane Ike (2008).
Mon. Wea. Rev., **140**, 3507-3524.

- Liu, Q., T. Marchok, H.-L. Pan, M. Bender, and S. Lord, 2000: Improvements in Hurricane Initialization and Forecasting at NCEP with Global and Regional (GFDL) models. *NOAA Tech. Procedures Bull.*, **472**, 7pp.
- 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.
- Parrish, D. F. and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747-1763.
- Rappaport, E. N., J. L. Franklin, L. A. Avila, S. R. Baig, J. L. Beven, E. S. Blake, C. A. Burr, J.-G. Jiing, C. A. Juckins, R. D. Knabb, C. W. Landsea, M. Mainelli, M. Mayfield, C. J. McAdie, R. J. Pasch, C. Sisko, S. R. Stewart, and A. N. Tribble, 2009: Advances and challenges at the National Hurricane Center. *Wea. Forecasting*, **24**, 395-419.
- Schwartz, C., J. Kain, S. Weiss, M. Xue, D. Bright, F. Kong, K. Thomas, J. Levit, and M. Coniglio, 2009: Next-day convection-allowing WRF model guidance: A second look at 2 vs. 4 km grid spacing. *Mon. Wea. Rev.*, **137**, 3351-3372.
- Skamarock, W. C. and M. L. Weisman, 2008: The impact of positive-definite moisture transport on NWP precipitation forecasts. *Mon. Wea. Rev.*, **137**, 488-494.
- 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, 113 pp.
- Thompson, G., P. R. Field, W. D. Hall, and R. M. Rasmussen, 2006: A New Bulk Microphysical Parameterization for WRF (& MM5). *WRF User's Workshop*,

Paper 5.3.

Thompson, G., P. R. Field, R. M. Rasmussen, and 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.

Tripoli, G. J., 1992: A nonhydrostatic mesoscale model designed to simulate scale interaction. *Mon. Wea. Rev.*, **120**, 1342-1359.

Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136**, 463-482.

Xue, M., F. Kong, K. W. Thomas, Y. Wang, K. Brewster, J. Gao, X. Wang, S. Weiss, A. Clark, J. Kain, M. Coniglio, J. Du, T. Jensen, and Y.-H. Kuo, 2010: CAPS Realtime Storm Scale Ensemble and High Resolution Forecasts for the NOAA Hazardous Weather Testbed 2010 Spring Experiment. *25th Conf. Severe Local Storms*, Amer. Meteor. Soc., Paper 7B.3.

List of figures

Fig. 1. Surface wind speed (shaded), wind vectors, and forecast track from 48 hour 4-km WRF forecasts (a) using global EnKF analysis produced by ESRL and (b) the NCEP GFS analysis, and (c) surface wind speed (shaded), wind vectors from the GFS analysis valid at the same time with the best track overlaid. Hurricanes Karl, Igor, and Julia are lined up from west to east.

Fig. 2. (a) Surface wind vectors, mean sea-level pressure (contours) and composite (column maximum) reflectivity, and (b) surface streamlines and simulated radar reflectivity from the 48 hour forecast of WRF-EnKF, valid at 1200 UTC, September 17, 2010.

Fig. 3. Absolute errors of TC track forecasts. Error bars represent the two-tailed 90% confidence interval using the block bootstrap distribution method. The number n at the bottom represents the number of forecasts, which decreases in time due to dissipation or movement of TCs outside of the forecast domain. The errors of NHC official forecasts are provided for comparison purpose.

Fig. 4. Difference in absolute track error between models, specified in the legend. Error bars represent the two-tailed 90% confidence interval of the bootstrapped difference between two means.

Fig. 5. As in Figure 3, for tropical depressions and tropical storms (a), and for hurricanes (b) at time of forecast.

Fig. 6. Schematic of cross track and along track error in relation to actual and forecast hurricane tracks.

Fig. 7. Absolute along track error (a) and cross track error (b) for hurricane track forecasts.

Fig. 8. Scatter plot of cross track error vs. along track error for TC forecasts. Positive (negative) along track error represents a forecast track faster (slower) than the best track. Positive (negative) cross track error represents a forecast track to the right (left) of the actual track. The bold crosshairs represent one standard deviation above and below the mean in both the along-track and cross-track directions, while the center of the crosses represents the error values.

Fig. 9. Absolute wind speed error of TC forecasts, for maximum 10 m winds. Error bars represent the two-tailed 90% confidence interval using the block bootstrap distribution method (Hamill et al., 2010). The number n represents the number of forecasts, which decreases in time due to dissipation or movement outside of the forecast domain.

Fig. 10. Biases of TC maximum surface wind speed forecasts.

Fig. 11. As in Figure 9, for tropical depressions and tropical storms (a), and for hurricanes (b) at time of forecast.

Fig. 12. Absolute MSLP error of TC forecasts. Error bars represent the two-tailed 90% confidence interval using the block bootstrap distribution method (Hamill et al. 2010). The number n represents the number of forecasts, which decreases in time due to dissipation or movement outside of the WRF forecast domain.

Fig. 13. Biases of TC MSLP forecasts. The n numbers at the bottom are the forecast sample sizes.

Fig. 14. As in Figure 12, for tropical depressions and tropical storms (a), and for hurricanes (b) at time of forecast.

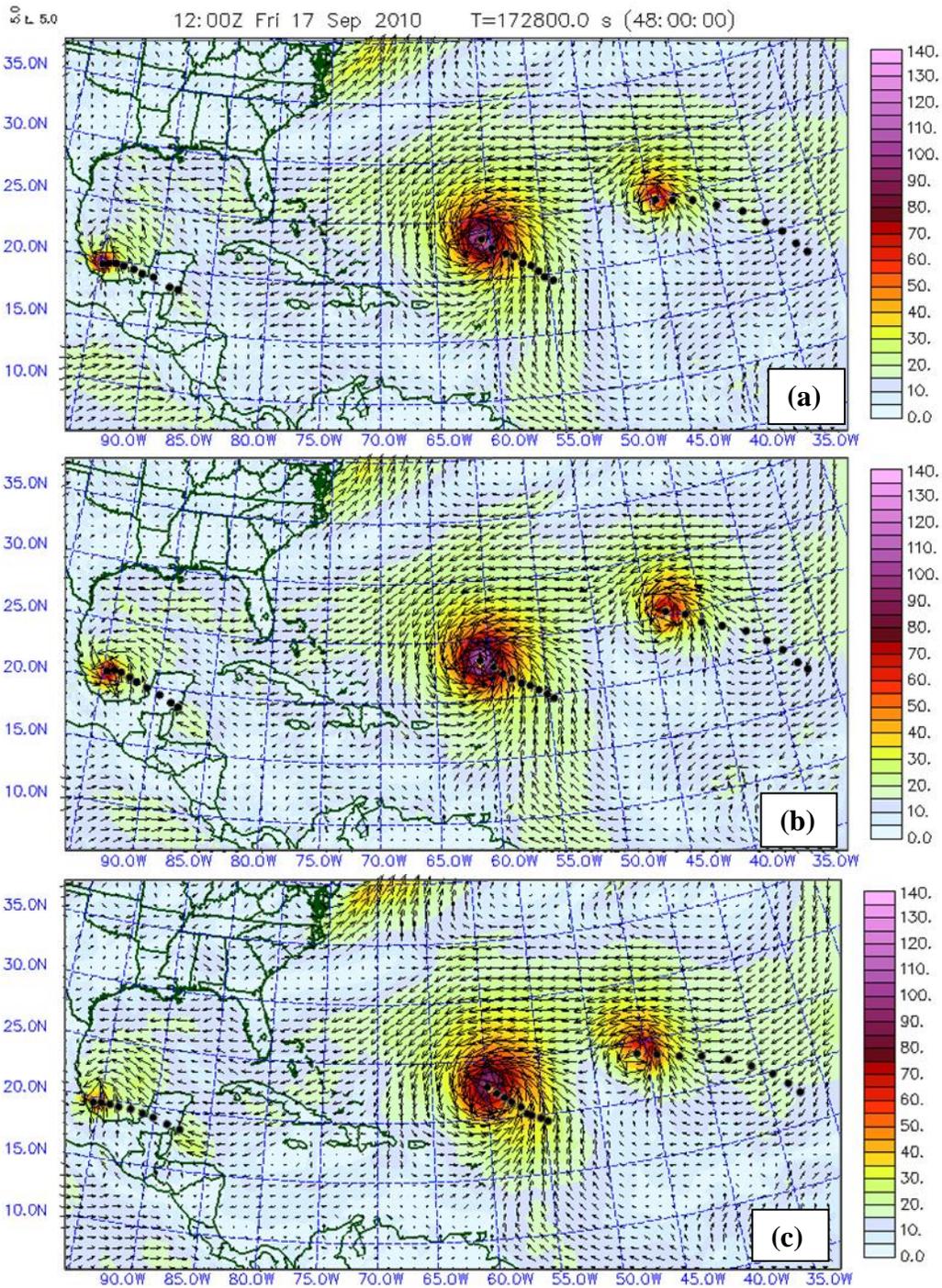


Fig. 1. Surface wind speed (shaded), wind vectors, and forecast track from 48 hour 4-km WRF forecasts (a) using global EnKF analysis produced by ESRL and (b) the NCEP GFS analysis, and (c) surface wind speed (shaded), wind vectors from the GFS analysis valid at the same time with the best track overlaid. Hurricanes Karl, Igor, and Julia are lined up from west to east.

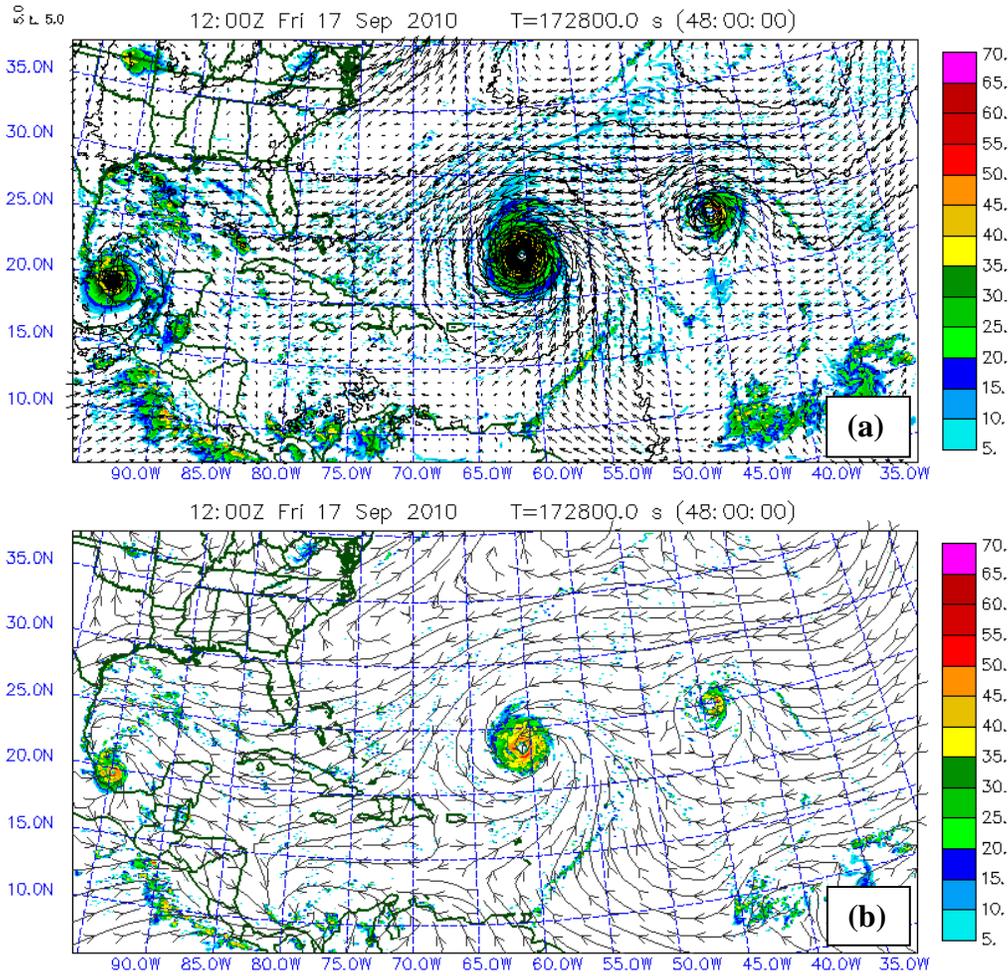


Fig. 2. (a) Surface wind vectors, mean sea-level pressure (contours) and composite (column maximum) reflectivity, and (b) surface streamlines and simulated radar reflectivity from the 48 hour forecast of WRF-EnKF, valid at 1200 UTC, September 17, 2010.

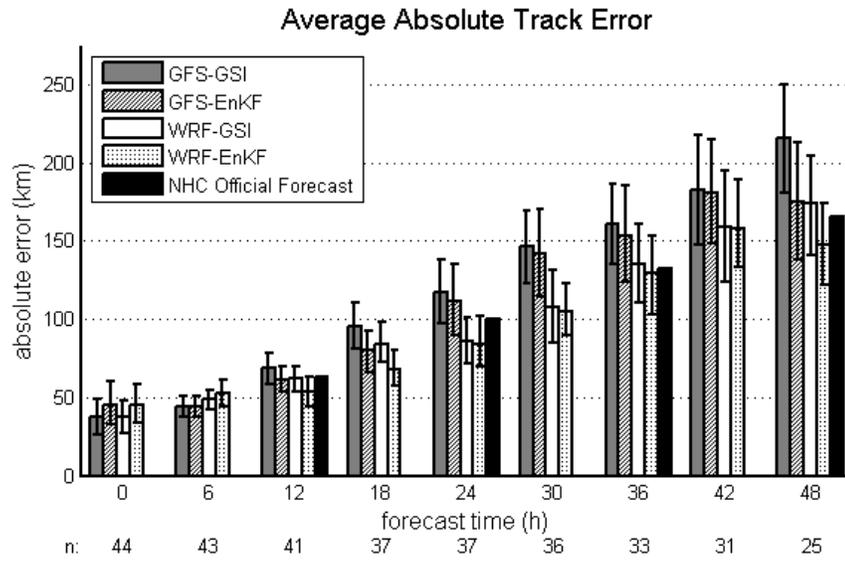


Fig. 3. Absolute errors of TC track forecasts. Error bars represent the two-tailed 90% confidence interval using the block bootstrap distribution method. The number n at the bottom represents the number of forecasts, which decreases in time due to dissipation or movement of TCs outside of the forecast domain. The errors of NHC official forecasts are provided for comparison purpose.

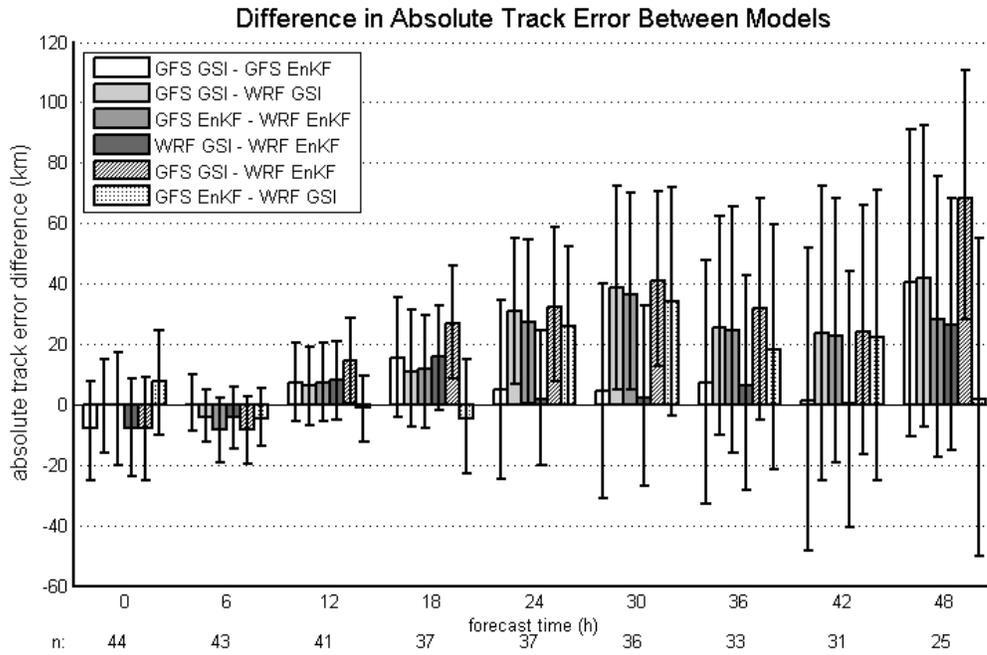


Fig. 4. Difference in absolute track error between models, specified in the legend. Error bars represent the two-tailed 90% confidence interval of the bootstrapped difference between two means.

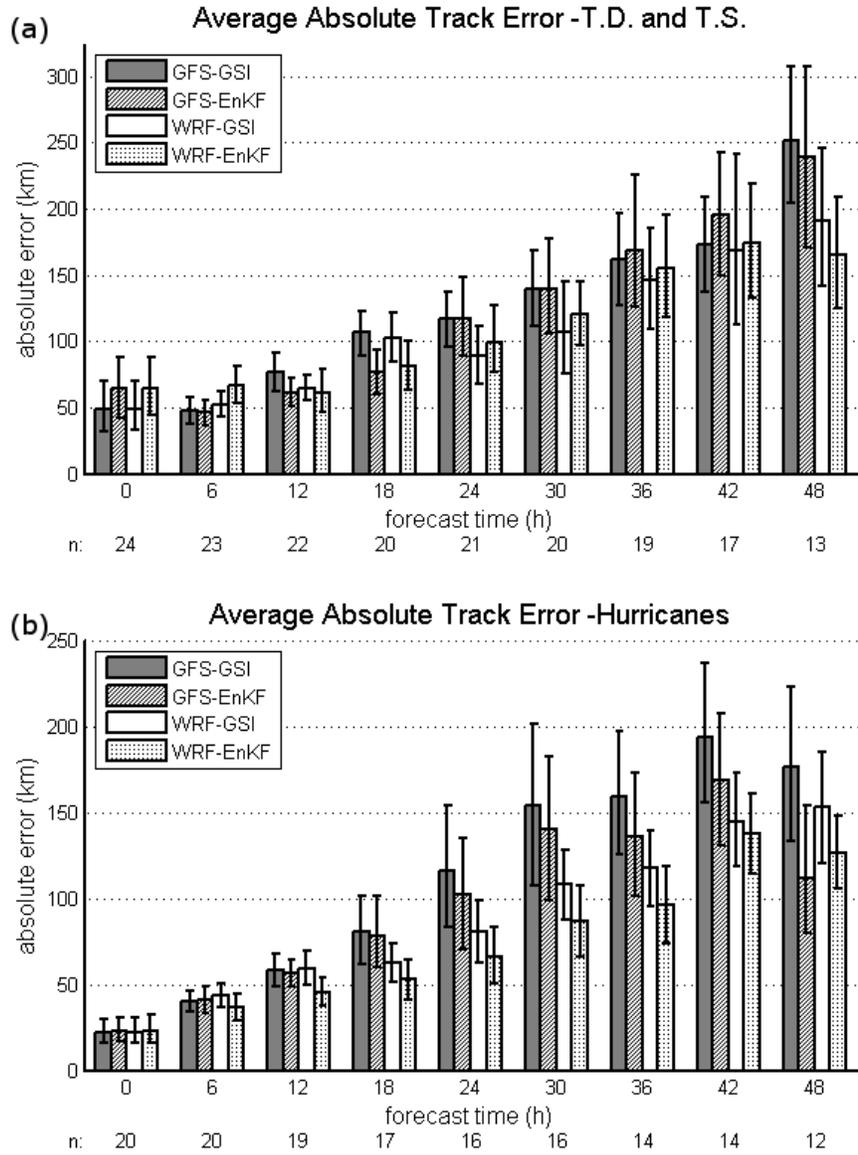


Fig. 5. As in Figure 3, for tropical depressions and tropical storms (a), and for hurricanes (b) at time of forecast.

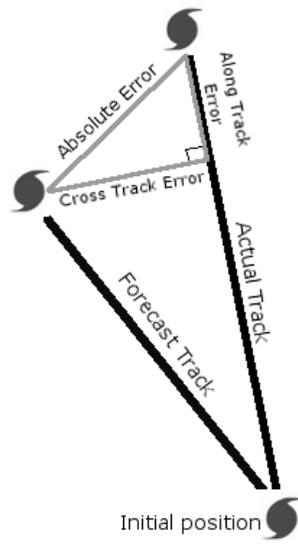


Fig. 6. Schematic of cross track and along track error in relation to actual and forecast hurricane tracks.

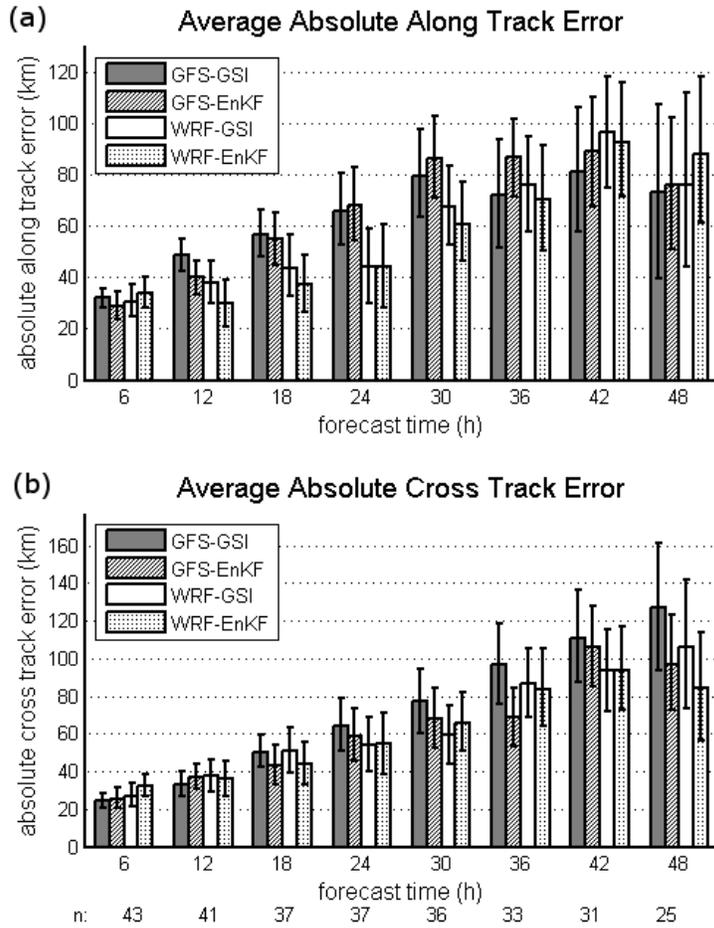


Fig. 7. Absolute along track error (a) and cross track error (b) for hurricane track forecasts.

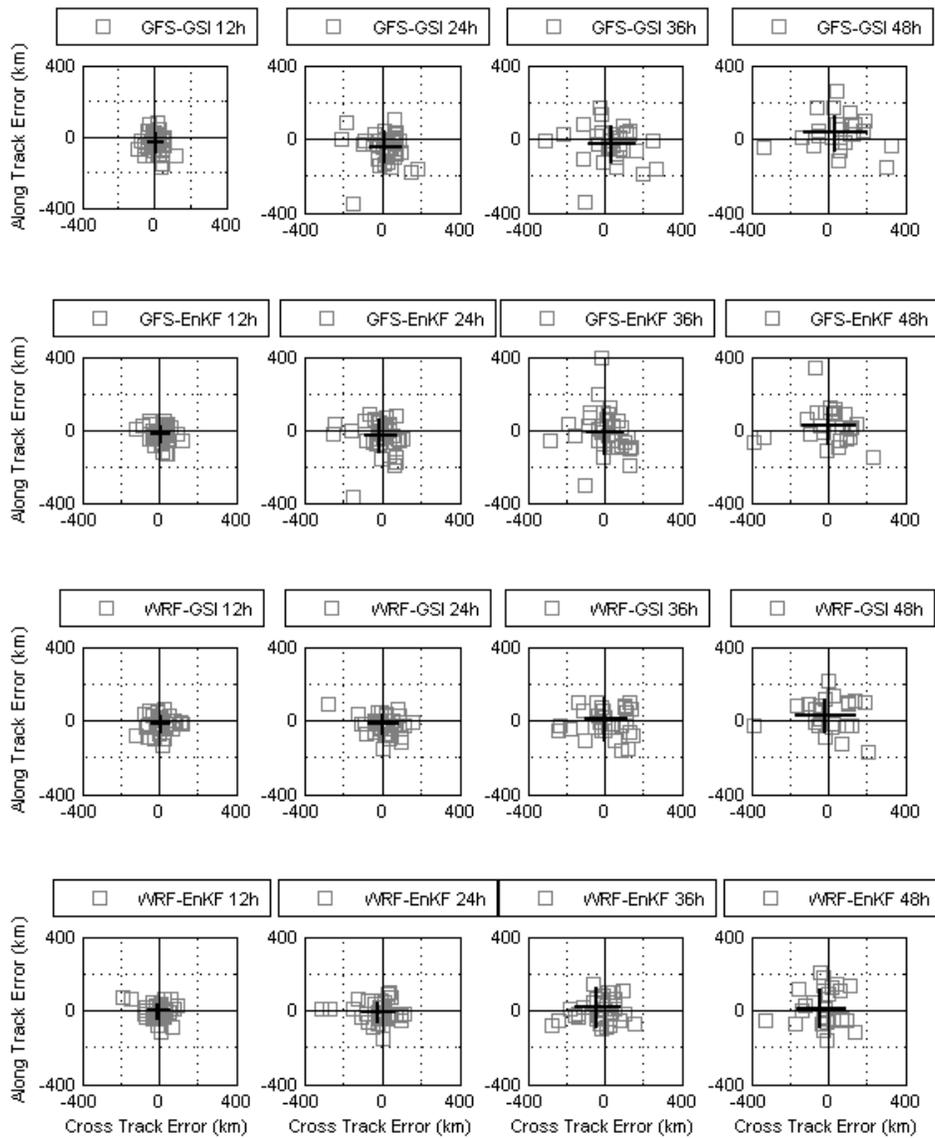


Fig. 8. Scatter plot of cross track error vs. along track error for TC forecasts. Positive (negative) along track error represents a forecast track faster (slower) than the best track. Positive (negative) cross track error represents a forecast track to the right (left) of the actual track. The bold crosshairs represent one standard deviation above and below the mean in both the along-track and cross-track directions, while the center of the crosses represents the error values.

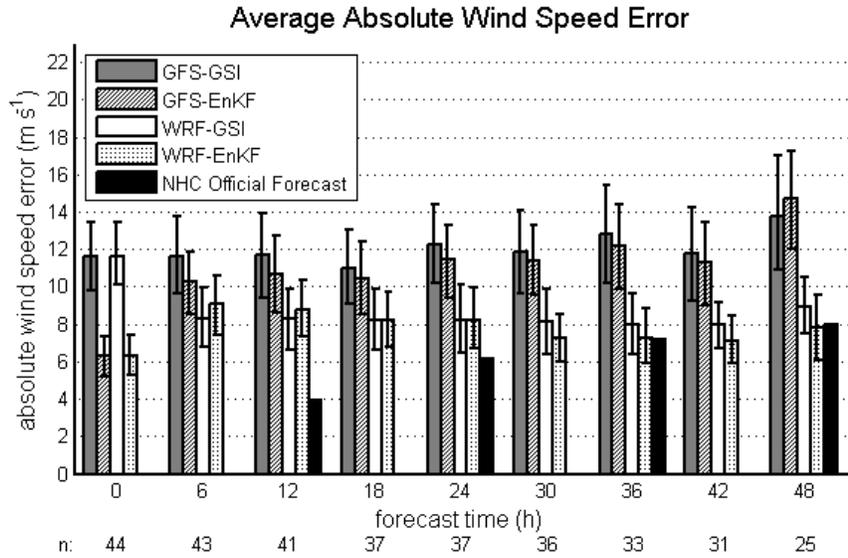


Fig. 9. Absolute wind speed error of TC forecasts, for maximum 10 m winds. Error bars represent the two-tailed 90% confidence interval using the block bootstrap distribution method (Hamill et al., 2010). The number n represents the number of forecasts, which decreases in time due to dissipation or movement outside of the forecast domain.

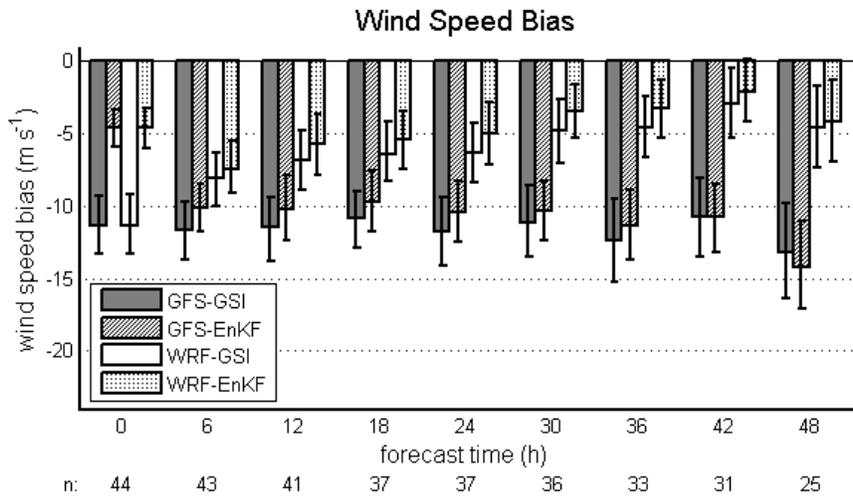


Fig. 10. Biases of TC maximum surface wind speed forecasts.

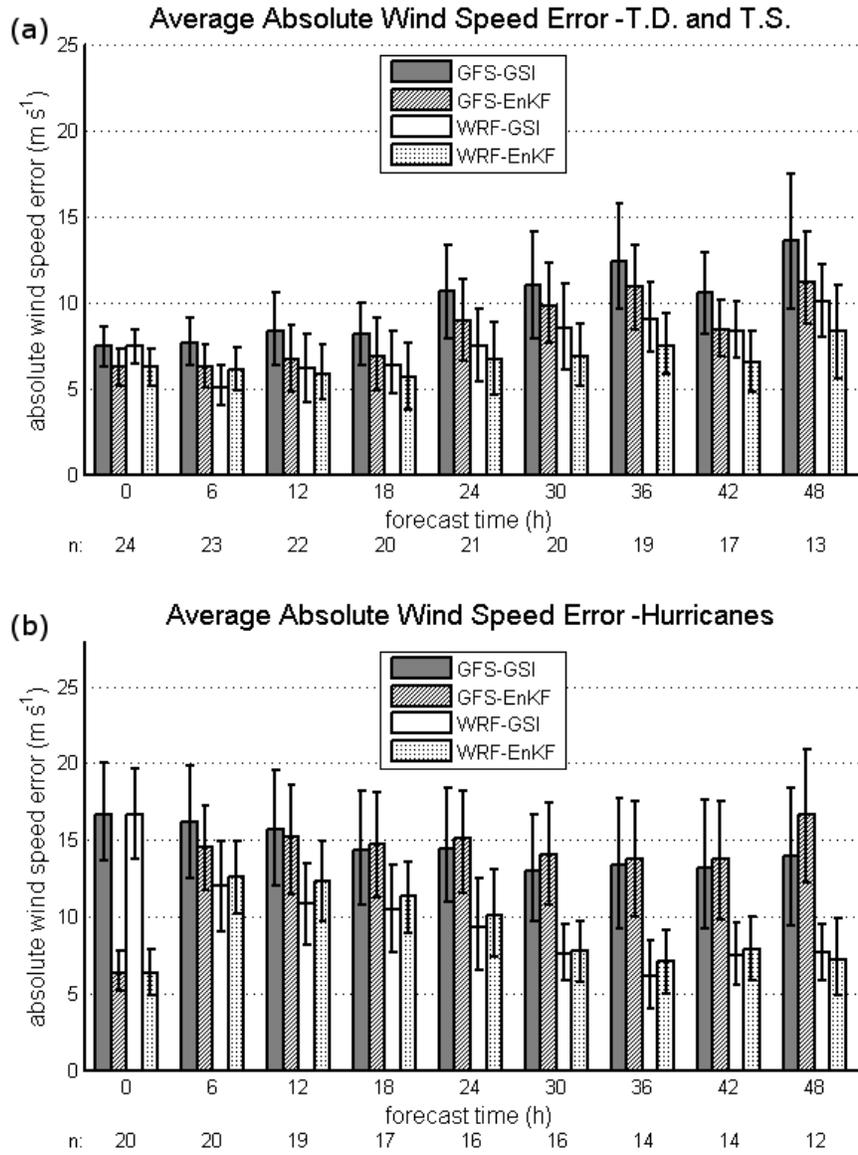


Fig. 11. As in Figure 9, for tropical depressions and tropical storms (a), and for hurricanes (b) at time of forecast.

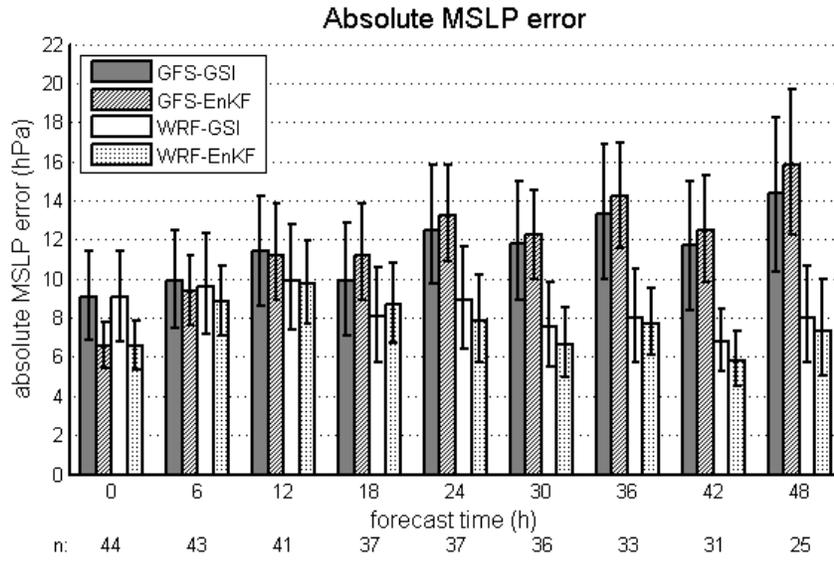


Fig. 12. Absolute MSLP error of TC forecasts. Error bars represent the two-tailed 90% confidence interval using the block bootstrap distribution method (Hamill et al. 2010). The number n represents the number of forecasts, which decreases in time due to dissipation or movement outside of the WRF forecast domain.

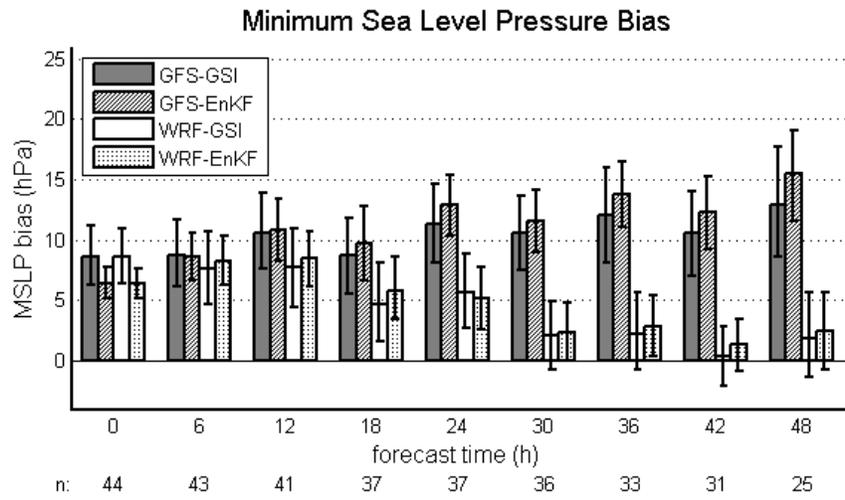


Fig. 13. Biases of TC MSLP forecasts. The n numbers at the bottom are the forecast sample sizes.

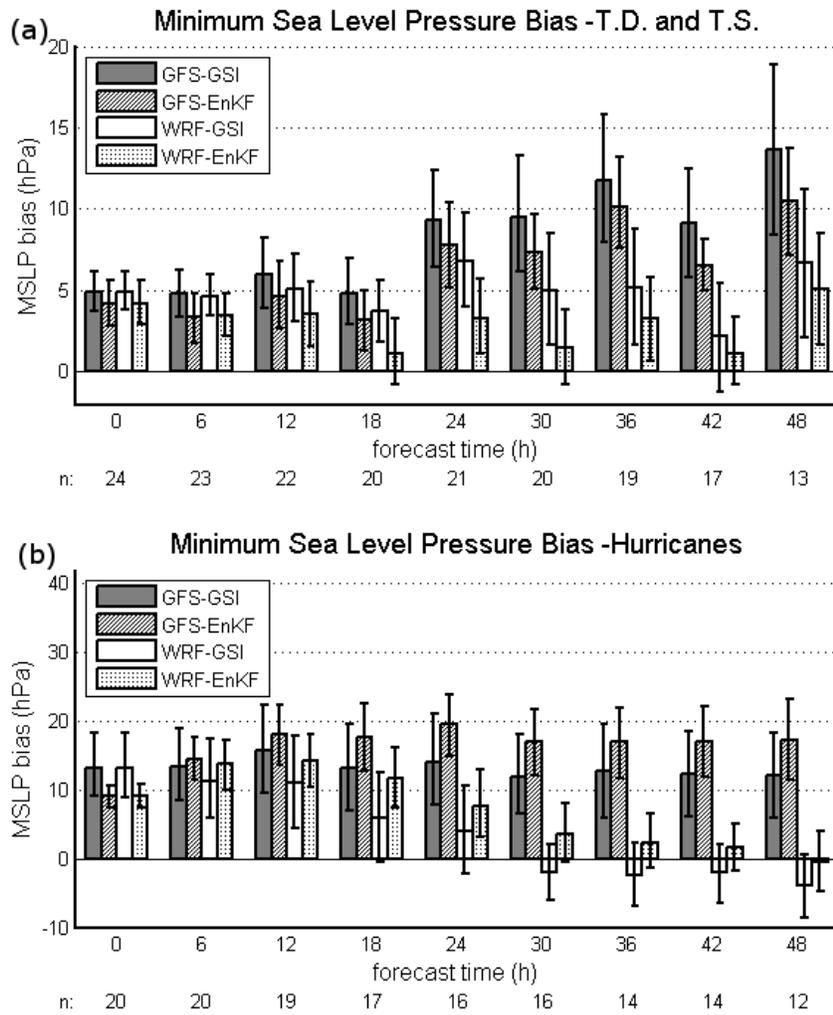


Fig. 14. As in Figure 12, for tropical depressions and tropical storms (a), and for hurricanes (b) at time of forecast.