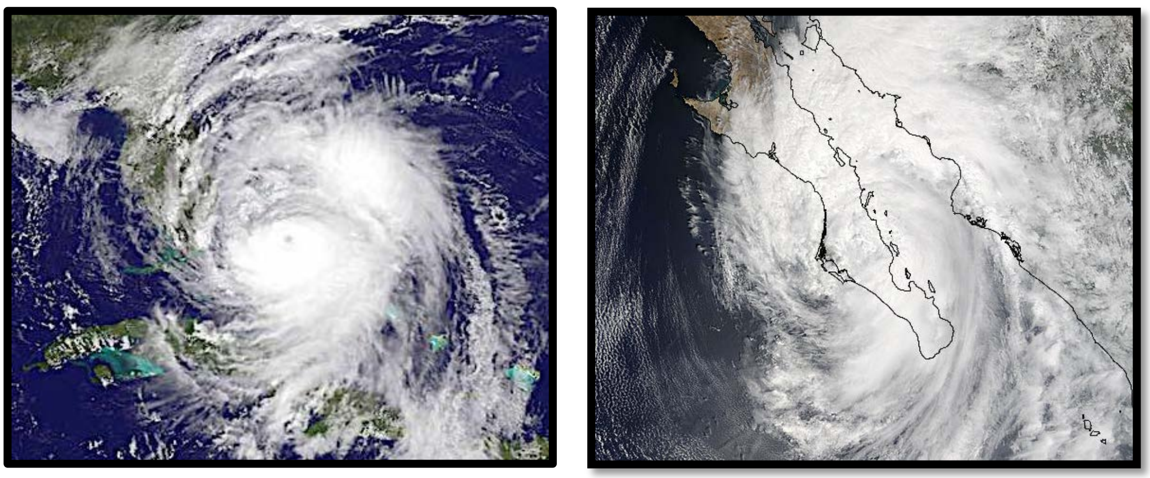
NATIONAL HURRICANE CENTER FORECAST VERIFICATION REPORT

2016 HURRICANE SEASON

John P. Cangialosi and James L. Franklin

4 April 2017



SATELLITE IMAGES OF HURRICANE MATTHEW (LEFT) AND HURRICANE NEWTON (RIGHT).

ABSTRACT

The 2016 Atlantic hurricane season had above-normal activity, with 386 official forecasts issued. The mean NHC official track forecast errors in the Atlantic basin were smaller than the previous 5-yr means at all forecast times. Records for track accuracy were set from 12 to 96 h, and the record was just missed at 120 h. The official track forecasts were highly skillful, but they were slightly beaten by the consensus models. GFSI and EGRI were the best-performing individual models in the short term, but EMXI had the highest skill of any individual model from 72 to 120 h. UEMI was a strong performer as well, and it had similar skill to GFSI, EGRI, and EMXI. AEMI, EGRI, and HWFI were fair performers while GHMI, CMCI, NVGI, and GFNI performed less well. The Government Performance and Results Act of 1993 (GPRA) track goal was met.

Mean official intensity errors for the Atlantic basin in 2016 were below the 5-yr means from 12 to 72 h, but above the means at the longer lead times. Decay-SHIFOR errors in 2016 were higher than their 5-yr means at all forecast times, and were well above the means at 96 and 120 h. The official forecasts were quite skillful and near or better than the top models. No records for intensity accuracy were set in 2016. Among the guidance, IVCN and HCCA were the best

performers. HWFI and CTCI showed increased skill with forecast time and were the best models at 120 h. DSHP and LGEM were fair performers while GHMI and GFNI lagged behind. The GPRA intensity goal was met.

There were 391 official forecasts issued in the eastern North Pacific basin in 2016, although only 119 of these verified at 120 h. This level of forecast activity was above average. Records for track accuracy were set at all forecast times, except 120 h. The official forecasts were very skillful, but they were outperformed by HCCA and by TVCE and FSSE at some forecast times. GFSI and AEMI were the next best models, followed by EMXI, HWFI, and CTCI.

For intensity, the official forecast errors in the eastern North Pacific basin were lower than the 5-yr means through 48 h, but slightly higher than the 5-yr means at the longer lead times. Decay-SHIFOR errors in 2016 were slightly lower or similar to their 5-yr means, indicating the season's storms were a little easier than average to predict. A record for intensity accuracy was set at 36 h. The official forecasts were more skillful than the models through 36 h, but were beaten by HCCA, HWFI, and IVCN at 48 h and beyond. FSSE was competitive with the best aids in the short term, and CTCI was skillful at the longer lead times. GHMI, GFNI, GFSI, and EMXI were not competitive. Nearly all of the guidance and the official forecasts had a low bias in 2016.

An evaluation of track performance during the 2014-16 period in the Atlantic basin indicates that FSSE and TVCA were the best-performing track models in the short term, however, EMXI had equal or greater skill from 72 to 120 h. The official track forecasts for the 3-yr sample had skill that was quite close to the best aids throughout the forecast period. For intensity in the Atlantic basin, the official forecasts have been consistently performing better than all of the guidance for the three years 2014-16. Among the guidance, the top models were FSSE, IVCN, and HWFI. An evaluation of the three years 2014-16 in the eastern North Pacific indicates that the official track forecasts were very skillful, but were slightly outperformed by TVCE. Regarding intensity, the official forecasts have been performing better than all of the guidance during the 2014-16 period from 12 to 48 h, but were slightly beaten by IVCN from 72 to 120 h.

Quantitative probabilistic forecasts of tropical cyclogenesis were publicly expanded from 48 to 120 h in August of 2013. Forecasts were expressed in 10% increments and in terms of categories ("low", "medium", or "high"). In the Atlantic basin, results from 2016 indicate that the 48-h probabilistic forecasts were quite reliable at the low probabilities, but were not well calibrated at the medium probabilities. The 120-h probabilities also had a low bias at the high and low probabilities, but were reliable at the medium probabilities. In the eastern North Pacific basin, the 48-h probabilistic forecasts were generally well calibrated, but the 120-h probabilities had a low bias at the low probabilities.

TABLE OF CONTENTS

1. Introduction

For all operationally designated tropical or subtropical cyclones in the Atlantic and eastern North Pacific basins, the National Hurricane Center (NHC) issues an official forecast of the cyclone's center location and maximum 1-min surface wind speed. Forecasts are issued every 6 h, and contain projections valid 12, 24, 36, 48, 72, 96, and 120 h after the forecast's nominal initial time (0000, 0600, 1200, or 1800 UTC)[1]. At the conclusion of the season, forecasts are evaluated by comparing the projected positions and intensities to the corresponding post-storm derived "best track" positions and intensities for each cyclone. A forecast is included in the verification only if the system is classified in the final best track as a tropical (or subtropical[2]) cyclone at both the forecast's initial time and at the projection's valid time. All other stages of development (e.g., tropical wave, [remnant] low, extratropical) are excluded[3]. For verification purposes, forecasts associated with special advisories do not supersede the original forecast issued for that synoptic time; rather, the original forecast is retained[4]. All verifications in this report include the depression stage.

It is important to distinguish between *forecast error* and *forecast skill*. Track forecast error, for example, is defined as the great-circle distance between a cyclone's forecast position and the best track position at the forecast verification time. Skill, on the other hand, represents a normalization of this forecast error against some standard or baseline. Expressed as a percentage improvement over the baseline, the skill of a forecast s_f is given by

$$s_f\ (\%) = 100 * (e_b - e_f) / e_b$$

where e_b is the error of the baseline model and e_f is the error of the forecast being evaluated. It is seen that skill is positive when the forecast error is smaller than the error from the baseline.

To assess the degree of skill in a set of track forecasts, the track forecast error can be compared with the error from CLIPER5, a climatology and persistence model that contains no information about the current state of the atmosphere (Neumann 1972, Aberson 1998)[5]. Errors from the CLIPER5 model are taken to represent a "no-skill" level of accuracy that is used as the baseline (e_b) for evaluating other forecasts[6]. If CLIPER5 errors are unusually low during a given season, for example, it indicates that the year's storms were inherently "easier" to forecast than normal or otherwise unusually well behaved. The current version of CLIPER5 is based on developmental data from 1931-2004 for the Atlantic and from 1949-2004 for the eastern Pacific.

Particularly useful skill standards are those that do not require operational products or inputs, and can therefore be easily applied retrospectively to historical data. CLIPER5 satisfies

[1] The nominal initial time represents the beginning of the forecast process. The actual advisory package is not released until 3 h after the nominal initial time, i.e., at 0300, 0900, 1500, and 2100 UTC.
[2] For the remainder of this report, the term "tropical cyclone" shall be understood to also include subtropical cyclones.
[3] Possible classifications in the best track are: Tropical Depression, Tropical Storm, Hurricane, Subtropical Depression, Subtropical Storm, Extratropical, Disturbance, Wave, and Low.
[4] Special advisories are issued whenever an unexpected significant change has occurred or when watches or warnings are to be issued between regularly scheduled advisories. The treatment of special advisories in forecast databases changed in 2005 to the current practice of retaining and verifying the original advisory forecast.
[5] CLIPER5 and SHIFOR5 are 5-day versions of the original 3-day CLIPER and SHIFOR models.
[6] To be sure, some "skill", or expertise, is required to properly initialize the CLIPER model.

this condition, since it can be run using persistence predictors (e.g., the storm's current motion) that are based on either operational or best track inputs. The best-track version of CLIPER5, which yields substantially lower errors than its operational counterpart, is generally used to analyze lengthy historical records for which operational inputs are unavailable. It is more instructive (and fairer) to evaluate operational forecasts against operational skill benchmarks, and therefore the operational versions are used for the verifications discussed below.[7]

Forecast intensity error is defined as the absolute value of the difference between the forecast and best track intensity at the forecast verifying time. Skill in a set of intensity forecasts is assessed using Decay-SHIFOR5 (DSHIFOR5) as the baseline. The DSHIFOR5 forecast is obtained by initially running SHIFOR5, the climatology and persistence model for intensity that is analogous to the CLIPER5 model for track (Jarvinen and Neumann 1979, Knaff et al. 2003). The output from SHIFOR5 is then adjusted for land interaction by applying the decay rate of DeMaria et al. (2006). The application of the decay component requires a forecast track, which here is given by CLIPER5. The use of DSHIFOR5 as the intensity skill benchmark was introduced in 2006. On average, DSHIFOR5 errors are about 5-15% lower than SHIFOR5 in the Atlantic basin from 12-72 h, and about the same as SHIFOR5 at 96 and 120 h.

It has been argued that CLIPER5 and DSHIFOR5 should not be used for skill benchmarks, primarily on the grounds that they were not good measures of forecast difficulty. Particularly in the context of evaluating forecaster performance, it was recommended that a model consensus (see discussion below) be used as the baseline. However, an unpublished study by NHC has shown that on the seasonal time scales at least, CLIPER5 and DSHIFOR5 are indeed good predictors of official forecast error. For the period 1990-2009 CLIPER5 errors explained 67% of the variance in annual-average NHC official track forecast errors at 24 h. At 72 h the explained variance was 40% and at 120 h the explained variance was 23%. For intensity the relationship was even stronger: DSHIFOR5 explained between 50 and 69% of the variance in annual-average NHC official errors at all time periods. Given this, CLIPER5 and DSHIFOR5 appear to remain suitable, if imperfect, baselines for skill, in the context of examining forecast performance over the course of a season (or longer). However, they're probably less useful for interpreting forecast performance with smaller samples (e.g., for a single storm).

The trajectory-CLIPER (TCLP) model is an alternative to the CLIPER and SHIFOR models for providing baseline track and intensity forecasts (DeMaria, personal communication). The input to TCLP [Julian Day, initial latitude, longitude, maximum wind, and the time tendencies of position and intensity] is the same as for CLIPER/SHIFOR, but rather than using linear regression to predict the future latitude, longitude and maximum wind, a trajectory approach is used. For track, a monthly climatology of observed storm motion vectors was developed from a 1982-2011 sample. The TCLP storm track is determined from a trajectory of the climatological motion vectors starting at the initial date and position of the storm. The climatological motion vector is modified by the current storm motion vector, where the influence of the current motion vector decreases with time during the forecast. A similar approach is taken for intensity, except that the intensity

[7] On very rare occasions, operational CLIPER or SHIFOR runs are missing from forecast databases. To ensure a completely homogeneous verification, post-season retrospective runs of the skill benchmarks are made using operational inputs. Furthermore, if a forecaster makes multiple estimates of the storm's initial motion, location, etc., over the course of a forecast cycle, then these retrospective skill benchmarks may differ slightly from the operational CLIPER/SHIFOR runs that appear in the forecast database.

tendency is estimated from the logistic growth equation model with climatological input. Similar to track, the climatological intensity tendency is modified by the observed tendency, where the influence decreases with forecast time. The track used for the TCLP intensity forecast is the TCLP track forecast. When the storm track crosses land, the intensity is decreased at a climatological decay rate. A comparison of a 10-yr sample of TCLP errors with those from CLIPER5 and DSHIFOR5 shows that the average track and intensity errors of the two baselines are within 10% of each other at all forecast times out to five days for the Atlantic and eastern North Pacific. One advantage of TCLP over CLIPER5/DSHIFOR5 is that TCLP can be run to any desired forecast time.

NHC also issues forecasts of the size of tropical cyclones; these “wind radii” forecasts are estimates of the maximum extent of winds of various thresholds (34, 50, and 64 kt) expected in each of four quadrants surrounding the cyclone. Unfortunately, there is insufficient surface wind information to allow the forecaster to accurately analyze the size of a tropical cyclone’s wind field. As a result, post-storm best track wind radii are likely to have errors so large as to render a verification of official radii forecasts unreliable and potentially misleading; consequently, no verifications of NHC wind radii are included in this report. In time, as our ability to measure the surface wind field in tropical cyclones improves, it may be possible to perform a meaningful verification of NHC wind radii forecasts.

Numerous objective forecast aids (guidance models) are available to help the NHC in the preparation of official track and intensity forecasts. Guidance models are characterized as either *early* or *late*, depending on whether or not they are available to the forecaster during the forecast cycle. For example, consider the 1200 UTC (12Z) forecast cycle, which begins with the 12Z synoptic time and ends with the release of an official forecast at 15Z. The 12Z run of the National Weather Service/Global Forecast System (GFS) model is not complete and available to the forecaster until about 16Z, or about an hour after the NHC forecast is released. Consequently, the 12Z GFS would be considered a late model since it could not be used to prepare the 12Z official forecast. This report focuses on the verification of early models.

Multi-layer dynamical models are generally, if not always, late models. Fortunately, a technique exists to take the most recent available run of a late model and adjust its forecast to apply to the current synoptic time and initial conditions. In the example above, forecast data for hours 6-126 from the previous (06Z) run of the GFS would be smoothed and then adjusted, or shifted, such that the 6-h forecast (valid at 12Z) would match the observed 12Z position and intensity of the tropical cyclone. The adjustment process creates an “early” version of the GFS model for the 12Z forecast cycle that is based on the most current available guidance. The adjusted versions of the late models are known, mostly for historical reasons, as *interpolated* models[8]. The adjustment algorithm is invoked as long as the most recent available late model is not more than 12 h old, e.g., a 00Z late model could be used to form an interpolated model for

[8] When the technique to create an early model from a late model was first developed, forecast output from the late models was available only at 12 h (or longer) intervals. In order to shift the late model’s forecasts forward by 6 hours, it was necessary to first interpolate between the 12 h forecast values of the late model – hence the designation “interpolated”.

the subsequent 06Z or 12Z forecast cycles, but not for the subsequent 18Z cycle. Verification procedures here make no distinction between 6 and 12 h interpolated models.[9]

A list of models is given in Table 1. In addition to their timeliness, models are characterized by their complexity or structure; this information is contained in the table for reference. Briefly, *dynamical* models forecast by solving the physical equations governing motions in the atmosphere. Dynamical models may treat the atmosphere either as a single layer (two-dimensional) or as having multiple layers (three-dimensional), and their domains may cover the entire globe or be limited to specific regions. The interpolated versions of dynamical model track and intensity forecasts are also sometimes referred to as dynamical models. *Statistical* models, in contrast, do not consider the characteristics of the current atmosphere explicitly but instead are based on historical relationships between storm behavior and various other parameters. *Statistical-dynamical* models are statistical in structure but use forecast parameters from dynamical models as predictors. *Consensus* models are not true forecast models *per se*, but are merely combinations of results from other models. One way to form a consensus is to simply average the results from a collection (or "ensemble") of models, but other, more complex techniques can also be used. The FSU "super-ensemble", for example, combines its individual components on the basis of past performance and attempts to correct for biases in those components (Williford et al. 2003). A consensus model that considers past error characteristics can be described as a "weighted" or "corrected" consensus. Additional information about the guidance models used at the NHC can be found at http://www.nhc.noaa.gov/modelsummary.shtml.

The verifications described in this report are based on forecast and best track data sets taken from the Automated Tropical Cyclone Forecast (ATCF) System[10] on 2 March 2017 for the Atlantic basin, and on 6 March 2017 for the eastern North Pacific basin. Verifications for the Atlantic and eastern North Pacific basins are given in Sections 2 and 3 below, respectively. Section 4 discusses NHC's probabilistic genesis forecasts. Section 5 summarizes the key findings of the 2016 verification and previews anticipated changes for 2017.

2. Atlantic Basin

a. 2016 season overview – Track

Figure 1 and Table 2 present the results of the NHC official track forecast verification for the 2016 season, along with results averaged for the previous 5-yr period, 2011-2015. In 2016, the NHC issued 386 Atlantic basin tropical cyclone forecasts[11], a number above the long-term average of 316 (Fig. 2) Mean track errors ranged from 24 n mi at 12 h to 168 n mi at 120 h. The mean official track forecast errors in 2016 were smaller than the previous 5-yr means by up to 24%, but the CLIPER errors were also lower than average, indicating that the season's storms

[9] The UKM and EMX models are only available through 120 h twice a day (at 0000 and 1200 UTC). Consequently, roughly half the interpolated forecasts from these models are 12 h old.

[10] In ATCF lingo, these are known as the "a decks" and "b decks", respectively.

[11] This count does not include forecasts issued for systems later classified to have been something other than a tropical cyclone at the forecast time.

were a little easier to forecast than normal. Records for accuracy were set in 2016 at all time periods, except at 120 h, although the 120-h error was only 3 n mi larger than the record set in 2013. The official track forecast vector biases were northwesterly (i.e., the official forecast tended to fall to the northwest of the verifying position) at all forecast periods, and it increased with time. Track forecast skill ranged from 45% at 12 h to 72% at 72 h (Table 2). The track errors in 2016 decreased from the 2015 values, and over the past 25 yr, the 24−72-h track forecast errors have been reduced by about 70% (Fig. 3). Track forecast error reductions of about 55% have occurred over the past 15 yr for the 96- and 120-h forecast periods. On average, the NHC track errors decrease as the initial intensity of a cyclone increases (Fig. 4).

Note that the mean official error in Fig. 1 is not precisely zero at 0 h (the analysis time). This non-zero difference between the operational analysis of storm location and best track location, however, is not properly interpreted as "analysis error". The best track is a subjectively smoothed representation of the storm history over its lifetime, in which the short-term variations in position or intensity that cannot be resolved in a 6-hourly time series are deliberately removed. Thus the location of a strong hurricane with a well-defined eye might be known with great accuracy at 1200 UTC, but the best track may indicate a location elsewhere by 5-10 miles or more if the precise location of the cyclone at 1200 UTC was unrepresentative. Operational analyses tend to follow the observed position of the storm more closely than the best track analyses, since it is more difficult to determine unrepresentative behavior in real time. Consequently, the t=0 "errors" shown in Fig. 1 contain both true analysis error and representativeness error.

Table 3a presents a homogeneous[12] verification for the official forecast along with a selection of early models for 2016. In order to maximize the sample size, a guidance model had to be available at least two-thirds of the time at both 48 and 120 h to be included in this comparison. The performance of the official forecast and the early track models in terms of skill are presented in Fig. 5. The figure shows that the official forecasts were highly skillful, but they were slightly bested by the consensus models HCCA and TVCA from 12 to 96 h. GFSI and EGRI were the best individual models in the short range, while EMXI was the most skillful individual model from 72 to 120 h. The UKMET ensemble mean (UEMI) was also very skillful and had similar or slightly lower errors than GFSI, EMXI, and EGRI. The strong performance of EGRI and UEMI was boosted by their correct predictions of Hurricane Matthew moving parallel to and just offshore of the east coast of Florida. AEMI, CTCI, and HWFI were not quite as skillful and made up the next tier of models. GHMI, CMCI, NVGI, and GFNI were the poorest-performing dynamical models again in 2016. An evaluation over the three years 2014-16 (Fig. 6) indicates that FSSE and TVCA were the best-performing models through 48 h, however, EMXI had equal or greater skill from 72 to 120 h. The official forecast had skill that was quite close to the best aids throughout the forecast period. GFSI, AEMI, EGRI, and HWFI were strong performers and made up the next best models. GHMI and CMCI performed less well.

Vector biases of the guidance models for 2016 are given in Table 3b. The table shows that the official forecast had similar biases to the consensus aids. Most of the guidance, like the official forecast, had a northwest bias at most forecast time periods.

[12] Verifications comparing different forecast models are referred to as *homogeneous* if each model is verified over an identical set of forecast cycles. Only homogeneous model comparisons are presented in this report.

A separate homogeneous verification of the primary consensus models for 2016 is shown in Fig. 7. The figure shows that the skill of the consensus aids were similar to one another for the short lead times, but TVCX was the most skillful, albeit by a small amount, at 48 h and beyond. AEMI and UEMI had less skill than the consensus aids, but UEMI was competitive with the other aids at 96 and 120 h. An examination of AEMI over the 2014-16 period (not shown) indicates that the ensemble mean has become increasingly skillful in the Atlantic basin, and it is even more skillful than the GFS deterministic run at 96 and 120 h.

Atlantic basin 48-h official track error, evaluated for all tropical cyclones, is a forecast measure tracked under the Government Performance and Results Act of 1993 (GPRA). In 2016, the GPRA goal was 71 n mi and the verification for this measure was 61.8 n mi.

b. 2016 season overview – Intensity

Figure 8 and Table 4 present the results of the NHC official intensity forecast verification for the 2016 season, along with results averaged for the preceding 5-yr period. Mean forecast errors in 2016 ranged from 5 kt at 12 h to 16 kt at 120 h. These errors were below the 5-yr means from 12 to 72 h, but above the 5-yr means at 96 and 120 h. No records for accuracy were set in 2016. The official forecasts had a slight low bias at all forecast time periods. Decay-SHIFOR5 errors were above their 5-yr means at all times and were much larger than average at 96 and 120 h, implying that the season's storms were challenging to predict. Figure 9 shows that the errors decreased at most forecast times in 2016, and there has been a notable decrease in error that began at the start of the decade. Over the long term, the intensity predictions seem to be improving as the forecasts are mostly more skillful in the current decade than they were in the previous one.

Table 5a presents a homogeneous verification for the official forecasts and the primary early intensity models for 2016. Intensity biases are given in Table 5b, and forecast skill is presented in Fig. 10. The official forecasts were quite skillful at all times, and beat all of the models, including the consensus aids, through 48 h. Among the guidance, the most accurate intensity models in 2016 were the consensus aids IVCN and HCCA. FSSE was competitive with IVCN and HCCA early, but its skill lagged at the longer leads. Conversely, HWFI and CTCI showed increased skill with forecast time and were the best models at 96 and 120 h. DSHP and LGEM were fair performers, but were not as skillful as the regional models CTCI and HWFI. GFSI was competitive at 48 h and beyond, while GFNI, GHMI, and EMXI trailed. An inspection of the intensity biases (Table 5b) indicates that majority of the models had a low bias at most forecast times. HWFI and GFSI had the smallest biases overall (less than 3 kt). An evaluation over the three years 2014-16 (Fig. 11) indicates that the official forecasts have been consistently performing better than all of the guidance. Similar to 2016, the top models for the 3-yr sample were FSSE, IVCN, and HWFI.

The 48-h official intensity error, evaluated for all tropical cyclones, is another GPRA measure for the NHC. In 2016, the GPRA goal was 12 kt and the verification for this measure was 10.6 kt. This was the sixth time in nine years that the intensity goal was met.

c. Verifications for individual storms

Forecast verifications for individual storms are given in Table 6. Of note are the large track errors for Tropical Storm Fiona, which at 96 and 120 h were about 50% larger than the long-term mean. The primary reason for the higher-than-average errors for this storm was that the official forecasts incorrectly predicted Fiona to turn poleward, which did not occur when the cyclone stayed shallow and moved more westward. Conversely, the official track forecast errors were quite low for Hurricane Earl and Hurricane Otto, where the track errors for those storms where significantly lower than the long-term mean at most forecast times. Figure 12 shows an illustration of the official track errors stratified by storm.

With regards to intensity, Tropical Storm Karl was one of the more challenging cyclones to predict in 2016. Although the official intensity forecasts were lower than the 5-yr averages for the short lead times, the NHC errors were considerably larger-than-average at the 96- and 120-h forecast times. The primary reason for the large errors at the longer leads was because the models incorrectly predicted strengthening while Karl remained relatively steady state. In general, global models incorrectly forecast the upper-level winds to become more conducive for strengthening while Karl was over the tropical Atlantic. This resulted in significant high biases in the official forecast. Another challenging case was Hurricane Matthew. The early official forecasts for Matthew missed the 24-h period of significant strengthening that occurred from 30 September to 1 October, resulting in a marked low bias. Figure 13 shows an illustration of the official intensity errors stratified by storm. Additional discussion on forecast performance for individual storms can be found in NHC Tropical Cyclone Reports available at **http://www.nhc.noaa.gov/data/tcr/index.php?season=2016&basin=atl**

3. Eastern North Pacific Basin

a. 2016 season overview – Track

The NHC track forecast verification for the 2016 season in the eastern North Pacific, along with results averaged for the previous 5-yr period is presented in Figure 14 and Table 7. There were 391 forecasts issued for the eastern Pacific basin in 2016, although only 119 of these verified at 120 h. This level of forecast activity is above the long-term average of 331 (Fig. 15). Mean track errors ranged from 19 n mi at 12 h to 104 n mi at 120 h. These errors were notably lower (up to 35%) than the 5-yr means at all forecast times. The CLIPER5 errors were also lower than its 5-yr means, indicating that the season's storms were easier to forecast than average, likely due to the relatively few recurving storms in 2016. Records for accuracy were set at all forecast periods, except at 120 h. The official track forecast vector biases were small for all lead times.

Figure 16 shows recent trends in track forecast accuracy and skill for the eastern North Pacific. Errors have been reduced by 60-70% for the 24 to 72 h forecasts since 1990. Similar improvements have been made for the 96 and 120 h forecasts since 2001. Forecast skill in 2016 increased from 2015, and the overall trend shows a significant increase in skill over the past couple of decades.

Table 8a presents a homogeneous verification for the official forecast and the early track models for 2016, with vector biases of the guidance models given in Table 8b. Skill comparisons of selected models are shown in Fig. 17. The official forecasts were very skillful and were near the best models, the consensus aids. Among the consensus models, HCCA had the lowest errors, and it was the only aid that beat the official forecast at all time periods. GFSI and AEMI were the next best models with EMXI, HWFI, and CTCI not far behind. EGRI and UEMI were fair performers, while GHMI, GFNI, NVGI, and CMCI lagged behind. An evaluation of the three years 2014-16 (Fig. 18) indicates that the official forecast was very skillful, but it was slightly outperformed by the consensus aid TVCE. FSSE was the second best aid and it had similar skill values to the official forecast. Among the individual models, EMXI was the best performer, closely followed by AEMI, GFSI, and HWFI.

A separate verification of the primary consensus aids is given in Figure 19. The skill of the consensus models was tightly clustered with HCCA edging out the other models through 48 h, and TVCE/TVCX being the most skillful models at the longer leads. AEMI was competitive with the consensus models, but UEMI had considerably less skill.

b. 2016 season overview – Intensity

Figure 20 and Table 9 present the results of the NHC eastern North Pacific intensity forecast verification for the 2016 season, along with results averaged for the preceding 5-yr period. Mean forecast errors were 6 kt at 12 h and increased to 17 kt at 96 h and 120 h. The errors were lower than the 5-yr means from 12 to 48 h, but slightly higher than the means at the longer lead times. The Decay-SHIFOR5 forecast errors were slightly lower or similar to their 5-yr means, indicative that the season's storms were a little easier than normal to predict. A record for accuracy was set at 36 h in 2016. A review of error and skill trends (Fig. 21) indicates that although there is considerable year-to-year variability in intensity errors, there has been a slight decrease in error over the past couple of decades at 48 h and beyond. No trend, however, is apparent for 24 h. Forecast skill has changed little during the past 5 or 6 years at most forecast time periods. Intensity forecast biases were low (under-forecast) and increased with forecast time.

Figure 22 and Table 10a present a homogeneous verification for the primary early intensity models for 2016. Forecast biases are given in Table 10b. The official forecasts outperformed the models from 12 to 36 h, but were beaten by the best-performing models HWFI, HCCA, and IVCN at 48 h and beyond. FSSE was competitive with the other consensus aids at the shorter leads, but it trailed at longer range. Conversely, CTCI was not very skillful through 36 h, but was among the more skillful guidance at 96 and 120 h. LGEM was a fair performer, while DSHP lagged beyond 48 h. GHMI, GFNI, and GFSI had little skill in 2016, and EMXI was not skillful at any forecast time. Nearly all of the models suffered from a low bias, with GHMI and GFNI having the most significant bias of about 10 kt at 48 and 72 h. It should be noted that the official forecast had less bias than most of the guidance. An evaluation over the three years 2014-16 (Fig. 23) indicates that the official forecast have been performing better than all of the guidance from 12 to 48 h, but were slightly beaten by IVCN from 72 to 120 h. FSSE had similar skill to the official forecast in the short term, but the model lagged at the longer leads. HWFI was a strong performer and the best individual model. DSHP and LGEM were fair performers, while GHMI and GFSI had little skill. EMXI was not skillful for this sample.

c. Verifications for individual storms

Forecast verifications for individual storms are given for reference in Table 11. Additional discussion on forecast performance for individual storms can be found in NHC Tropical Cyclone Reports available at http://www.nhc.noaa.gov/data/tcr/index.php?season=2016&basin=epac.

4. Genesis Forecasts

The NHC routinely issues Tropical Weather Outlooks (TWOs) for both the Atlantic and eastern North Pacific basins. The TWOs are text products that discuss areas of disturbed weather and their potential for tropical cyclone development. Since 2007, forecasters subjectively assigned a probability of genesis (0 to 100%, in 10% increments) to each area of disturbed weather described in the TWO, where the assigned probabilities represented the forecaster's determination of the chance of tropical cyclone formation during the 48 h period following the nominal TWO issuance time. In 2009, the NHC began producing in-house (non-public) experimental probabilistic tropical cyclone forecasts for 120 h, which became public in August of 2013. Verification is based on NHC best-track data, with the time of genesis defined to be the first tropical cyclone point appearing in the best track.

Verifications of the 48-h outlook for the Atlantic and eastern North Pacific basins for 2016 are given in Table 12 and illustrated in Fig. 24. In the Atlantic basin, a total of 586 genesis forecasts were made. These 48-h forecasts were quite reliable for the low probabilities, but were less reliable at the medium probabilities. An under-forecast (low) bias for a small sample existed at the high probabilities. In the eastern Pacific, a total of 881 genesis forecasts were made. The forecasts were well calibrated at the low and medium probabilities, but a low bias also existed for a small sample at the high probabilities.

Verification of the 120-h outlook for the Atlantic and eastern North Pacific basins for 2016 are given in Table 13 and illustrated in Fig. 25. In the Atlantic basin, the 120-h forecasts were well calibrated at the medium probabilities, but a slight low bias existed at both the high and low categories. In the eastern North Pacific, a slight low bias was noted at the low probabilities, but the forecasts were generally well calibrated at the remaining probabilities. The diagrams also show the refinement distribution, which indicates how often the forecasts deviated from (a perceived) climatology. Sharp peaks at climatology indicate low forecaster confidence, while maxima at the extremes indicate high confidence; the refinement distributions shown here suggest an intermediate level of forecaster confidence.

5. Looking Ahead to 2017

a. Track Forecast Cone Sizes

The National Hurricane Center track forecast cone depicts the probable track of the center of a tropical cyclone, and is formed by enclosing the area swept out by a set of circles along the forecast track (at 12, 24, 36 h, etc.). The size of each circle is set so that two-thirds of historical official forecast errors over the most-recent 5-yr sample fall within the circle. The circle radii defining the cones in 2017 for the Atlantic and eastern North Pacific basins (based on error distributions for 2012-16) are given in Table 14. In the Atlantic basin, the cone circles will be smaller (by up to 11 %) at all forecast time periods. In the eastern Pacific basin, the cone circles will also be smaller (by up to 15 %) at all forecast time periods.

b. Consensus Models

In 2008, NHC changed the nomenclature for many of its consensus models. The new system defines a set of consensus model identifiers that remain fixed from year to year. The specific members of these consensus models, however, will be determined at the beginning of each season and may vary from year to year.

Some consensus models require all of their member models to be available in order to compute the consensus (e.g., TCOA), while others are less restrictive, requiring only two or more members to be present (e.g., TVCA). The terms “fixed” and “variable” can be used to describe these two approaches, respectively. In a variable consensus model, it is often the case that the 120 h forecast is based on a different set of members than the 12 h forecast. While this approach greatly increases availability, it does pose consistency issues for the forecaster.

The consensus model composition for 2017 is given in Table 15. The 2017 model composition for all consensus aids (except GFEX) have been modified to remove GHMI, which was retired at the conclusion of the 2016 hurricane season. An evalution of GHMI’s replacement, Hurricanes in a Multi-scale Ocean coupled Non-hydrostatic model (HMON), is underway and its potential inclusion in the consensus models will be assessed later this year. No other modifications were made to the previous configurations of the consensus models.

Acknowledgments

The authors gratefully acknowledge Michael Brennan, Monica Bozeman, and Craig Mattocks of NHC, managers of the NHC forecast databases.

6. References

Aberson, S. D., 1998: Five-day tropical cyclone track forecasts in the North Atlantic basin. Wea. Forecasting, 13, 1005-1015.

DeMaria, M., J. A. Knaff, and J. Kaplan, 2006: On the decay of tropical cyclone winds crossing narrow landmasses, J. Appl. Meteor., 45, 491-499.

Jarvinen, B. R., and C. J. Neumann, 1979: Statistical forecasts of tropical cyclone intensity for the North Atlantic basin. NOAA Tech. Memo. NWS NHC-10, 22 pp.

Knaff, J.A., M. DeMaria, B. Sampson, and J.M. Gross, 2003: Statistical, five-day tropical cyclone intensity forecasts derived from climatology and persistence. Wea. Forecasting, 18, 80-92.

Neumann, C. B., 1972: An alternate to the HURRAN (hurricane analog) tropical cyclone forecast system. NOAA Tech. Memo. NWS SR-62, 24 pp.

Williford, C.E., T. N. Krishnamurti, R. C. Torres, S. Cocke, Z. Christidis, and T. S. V. Kumar, 2003: Real-Time Multimodel Superensemble Forecasts of Atlantic Tropical Systems of 1999. Mon. Wea. Rev., 131, 1878-1894.

List of Tables

Table 1. National Hurricane Center forecasts and models.

ID	Name/Description	Type	Timeliness (E/L)	Parameters forecast
OFCL	Official NHC forecast			Trk, Int
GFDL	NWS/Geophysical Fluid Dynamics Laboratory model	Multi-layer regional dynamical	L	Trk, Int
HWRF	Hurricane Weather and Research Forecasting Model	Multi-layer regional dynamical	L	Trk, Int
GFSO	NWS/Global Forecast System (formerly Aviation)	Multi-layer global dynamical	L	Trk, Int
AEMN	GFS ensemble mean	Consensus	L	Trk, Int
UKM	United Kingdom Met Office model, automated tracker	Multi-layer global dynamical	L	Trk, Int
EGRR	United Kingdom Met Office model with subjective quality control applied to the tracker	Multi-layer global dynamical	L	Trk, Int
UEMN	UKMET ensemble mean	Consensus	L	Trk, Int
NVGM	Navy Global Environmental Model	Multi-layer global dynamical	L	Trk, Int
GFDN	Navy version of GFDL	Multi-layer regional dynamical	L	Trk, Int
CMC	Environment Canada global model	Multi-level global dynamical	L	Trk, Int
NAM	NWS/NAM	Multi-level regional dynamical	L	Trk, Int
CTX	COAMPS-TC using GFS initial and boundary conditions	Multi-layer regional dynamical	L	Trk, Int
EMX	ECMWF global model	Multi-layer global dynamical	L	Trk, Int

ID	Name/Description	Type	Timeliness (E/L)	Parameters forecast
EEMN	ECMWF ensemble mean	Consensus	L	Trk
BAMS	Beta and advection model (shallow layer)	Single-layer trajectory	E	Trk
BAMM	Beta and advection model (medium layer)	Single-layer trajectory	E	Trk
BAMD	Beta and advection model (deep layer)	Single-layer trajectory	E	Trk
LBAR	Limited area barotropic model	Single-layer regional dynamical	E	Trk
CLP5	CLIPER5 (Climatology and Persistence model)	Statistical (baseline)	E	Trk
SHF5	SHIFOR5 (Climatology and Persistence model)	Statistical (baseline)	E	Int
DSF5	DSHIFOR5 (Climatology and Persistence model)	Statistical (baseline)	E	Int
OCD5	CLP5 (track) and DSF5 (intensity) models merged	Statistical (baseline)	E	Trk, Int
TCLP	Trajectory-CLIPER model	Statistical (baseline)	E	Trk, Int
SHIP	Statistical Hurricane Intensity Prediction Scheme (SHIPS)	Statistical-dynamical	E	Int
DSHP	SHIPS with inland decay	Statistical-dynamical	E	Int
OFCI	Previous cycle OFCL, adjusted	Interpolated	E	Trk, Int
GFDI	Previous cycle GFDL, adjusted	Interpolated-dynamical	E	Trk, Int

ID	Name/Description	Type	Timeliness (E/L)	Parameters forecast
GHMI	Previous cycle GFDL, adjusted using a variable intensity offset correction that is a function of forecast time. Note that for track, GHMI and GFDI are identical.	Interpolated-dynamical	E	Trk, Int
HWFI	Previous cycle HWRF, adjusted	Interpolated-dynamical	E	Trk, Int
CTCI	Previous cycle CTCX, adjusted	Interpolated-dynamical	E	Trk, Int
GFSI	Previous cycle GFS, adjusted	Interpolated-dynamical	E	Trk, Int
UKMI	Previous cycle UKM, adjusted	Interpolated-dynamical	E	Trk, Int
EGRI	Previous cycle EGRR, adjusted	Interpolated-dynamical	E	Trk, Int
NVGI	Previous cycle NVGM, adjusted	Interpolated-dynamical	E	Trk, Int
GFNI	Previous cycle GFDN, adjusted	Interpolated-dynamical	E	Trk, Int
EMXI	Previous cycle EMX, adjusted	Interpolated-dynamical	E	Trk, Int
CMCI	Previous cycle CMC, adjusted	Interpolated-dynamical	E	Trk, Int
GUNA	Average of GFDI, EGRI, and GFSI	Consensus	E	Trk
CGUN	Version of GUNA corrected for model biases	Corrected consensus	E	Trk
AEMI	Previous cycle AEMN, adjusted	Consensus	E	Trk, Int
UEMI	Previous cycle UEMN, adjusted	Consensus	E	Trk, Int
FSSE	FSU Super-ensemble	Corrected consensus	E	Trk, Int
GFEX	Average of GFSI and EMXI	Consensus	E	Trk
TCON	Average of GHMI, EGRI, GFSI, and HWFI	Consensus	E	Trk

ID	Name/Description	Type	Timeliness (E/L)	Parameters forecast
TCCN	Version of TCON corrected for model biases	Corrected consensus	E	Trk
TVCN	Average of at least two of GFSI EGRI GHMI HWFI EMXI CTCI	Consensus	E	Trk
TVCA	Average of at least two of GFSI EGRI GHMI HWFI EMXI CTCI	Consensus	E	Trk
TVCE	Average of at least two of GFSI EGRI GHMI HWFI EMXI CTCI	Consensus	E	Trk
TVCX	EMXI and average of at least two of GFSI EGRI GHMI HWFI EMXI CTCI	Consensus	E	Trk
TVCC	Version of TVCN corrected for model biases	Corrected consensus	E	Trk
HCCA	Weighted average of AEMI, GFSI, CTCI, DSHP, EGRI, EMNI, EMXI,HWFI, LGEM	Corrected consensus	E	Trk, Int
ICON	Average of DSHP, LGEM, GHMI, and HWFI	Consensus	E	Int
IVCN	Average of at least two of DSHP LGEM GHMI HWFI GFNI	Consensus	E	Int

Table 2. Homogenous comparison of official and CLIPER5 track forecast errors in the Atlantic basin for the 2016 season for all tropical cyclones. Averages for the previous 5-yr period are shown for comparison.

	Forecast Period (h)						
	12	24	36	48	72	96	120
2016 mean OFCL error (n mi)	24.3	36.5	47.7	61.8	88.8	133.1	168.2
2016 mean CLIPER5 error (n mi)	44.4	90.6	145.3	204.0	317.6	416.0	478.3
2016 mean OFCL skill relative to CLIPER5 (%)	45.3	59.7	67.2	69.7	72.0	68.0	64.8
2016 mean OFCL bias vector (°/n mi)	314/003	307/008	306/013	311/017	318/022	336/034	329/051
2016 number of cases	354	322	290	258	204	161	131
2011-2015 mean OFCL error (n mi)	26.3	41.8	56.7	73.6	109.7	159.5	221.7
2011-2015 mean CLIPER5 error (n mi)	46.4	99.0	159.6	218.9	324.4	419.2	498.1
2011-2015 mean OFCL skill relative to CLIPER5 (%)	43.3	57.8	64.5	66.4	66.2	62.0	55.5
2011-2015 mean OFCL bias vector (°/n mi)	339/003	325/005	316/007	316/011	316/018	326/029	331/045
2011-2015 number of cases	1233	1085	945	816	607	463	361
2016 OFCL error relative to 2011-2015 mean (%)	-7.6	-12.7	-15.9	-16.0	-19.1	-16.6	-24.1
2016 CLIPER5 error relative to 2011-2015 mean (%)	-4.3	-8.5	-9.0	-6.8	-2.1	-0.8	-4.0

Table 3a. Homogenous comparison of Atlantic basin early track guidance model errors (n mi) for 2016. Errors smaller than the NHC official forecast are shown in bold-face.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	21.7	33.7	44.5	56.3	84.6	127.7	161.6
OCD5	42.2	89.9	146.3	208.7	320.2	403.4	429.2
GFSI	23.1	36.6	48.9	59.9	101.7	157.7	221.2
GHMI	25.5	43.7	64.8	86.4	140.2	202.3	276.5
HWFI	25.2	41.0	59.1	78.5	122.2	174.4	263.3
EMXI	24.4	39.5	53.3	65.3	94.4	135.8	172.7
CMCI	29.6	50.7	68.9	93.4	140.5	206.7	288.0
EGRI	23.6	37.5	49.2	62.1	96.2	156.3	208.8
NVGI	33.1	56.8	80.9	109.2	155.3	211.7	284.4
GFNI	30.9	55.1	82.0	108.0	161.0	235.5	317.0
CTCI	23.4	39.6	56.4	73.7	117.3	148.1	209.4
AEMI	24.0	38.8	53.3	66.9	112.3	165.1	218.5
UEMI	24.4	39.3	51.0	64.4	100.2	143.8	163.2
FSSE	**20.9**	**32.5**	**42.7**	**54.3**	86.1	129.6	178.6
TVCA	**21.1**	**32.6**	**43.0**	**53.5**	**81.6**	**125.5**	173.3
HCCA	**20.9**	**32.5**	**42.6**	**53.0**	**81.8**	**123.1**	184.1
BAMD	43.9	83.9	124.8	158.2	224.9	290.3	341.1
BAMM	37.1	67.0	93.5	114.8	145.7	189.1	243.5
BAMS	52.6	97.3	137.0	175.3	230.9	286.8	385.1
TCLP	43.5	95.0	154.9	216.1	326.7	425.0	506.4
# Cases	199	187	178	157	125	104	82

Table 3b. Homogenous comparison of Atlantic basin early track guidance model bias vectors (º/n mi) for 2016.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	312/004	290/008	295/011	304/019	318/032	325/049	310/065
OCD5	330/007	318/011	331/018	332/021	356/052	355/074	297/073
GFSI	303/005	286/009	288/012	290/019	297/030	309/055	305/098
GHMI	204/001	295/004	327/011	324/024	334/042	356/075	357/132
HWFI	323/007	309/011	312/016	317/022	348/042	002/068	002/102
EMXI	308/005	294/009	298/010	311/014	327/029	321/042	279/050
CMCI	290/010	286/020	291/023	302/032	324/060	339/102	356/153
EGRI	298/003	277/006	281/010	277/018	284/028	305/029	305/038
NVGI	297/012	285/020	278/022	278/029	281/036	299/034	275/031
GFNI	262/007	270/015	282/025	287/039	305/063	333/099	335/133
CTCI	358/005	357/006	357/007	345/010	344/026	337/039	312/054
AEMI	283/007	270/012	277/014	290/021	309/034	320/055	307/081
UEMI	265/005	256/012	259/018	264/028	279/040	283/051	267/072
FSSE	303/002	267/003	288/003	320/007	320/017	338/026	330/052
TVCA	313/004	301/007	308/010	309/016	326/030	337/047	331/068
HCCA	248/003	228/007	236/009	263/011	293/027	311/033	308/054
BAMD	041/010	047/020	048/030	031/033	342/040	293/046	265/100
BAMM	266/011	249/019	241/025	238/031	242/024	290/026	300/076
BAMS	275/030	264/057	254/080	246/102	234/127	239/150	235/233
TCLP	267/011	258/030	258/050	254/065	258/090	263/140	251/247
# Cases	199	187	178	157	125	104	82

Table 4. Homogenous comparison of official and Decay-SHIFOR5 intensity forecast errors in the Atlantic basin for the 2016 season for all tropical cyclones. Averages for the previous 5-yr period are shown for comparison.

	Forecast Period (h)						
	12	24	36	48	72	96	120
2016 mean OFCL error (kt)	5.4	7.6	9.4	10.5	11.7	14.7	16.1
2016 mean Decay-SHIFOR5 error (kt)	7.2	10.9	13.4	15.4	18.2	20.4	25.7
2016 mean OFCL skill relative to Decay-SHIFOR5 (%)	25.0	30.3	29.9	31.8	35.7	27.9	37.4
2016 OFCL bias (kt)	-1.5	-2.1	-2.7	-2.4	-2.0	-3.0	-3.8
2016 number of cases	354	322	290	258	204	161	131
2011-15 mean OFCL error (kt)	5.6	8.3	10.6	12.4	14.1	14.1	14.4
2011-15 mean Decay-SHIFOR5 error (kt)	6.7	10.0	12.3	14.3	16.8	16.7	16.4
2011-15 mean OFCL skill relative to Decay-SHIFOR5 (%)	16.4	17.0	13.8	13.3	16.1	15.6	12.2
2011-15 OFCL bias (kt)	-0.9	-0.6	-0.2	0.2	0.0	0.7	1.6
2011-15 number of cases	1233	1085	945	816	607	463	361
2016 OFCL error relative to 2011-15 mean (%)	-3.6	-8.4	-11.3	-15.3	-17.0	4.3	11.8
2016 Decay-SHIFOR5 error relative to 2011-15 mean (%)	7.5	9.0	8.9	7.7	8.3	22.2	56.7

Table 5a. Homogenous comparison of selected Atlantic basin early intensity guidance model errors (kt) for 2016. Errors smaller than the NHC official forecast are shown in boldface.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	6.1	8.5	10.1	10.3	11.4	16.1	17.4
OCD5	7.8	11.4	13.5	15.5	18.5	21.1	29.0
HWFI	7.5	9.7	11.5	12.6	13.2	**15.3**	**16.5**
GHMI	8.8	11.7	14.1	13.6	14.3	17.9	20.2
GFNI	8.6	11.9	13.7	15.1	16.8	19.4	23.8
DSHP	7.4	9.9	11.5	12.2	13.9	18.6	19.6
LGEM	7.6	10.1	12.0	12.7	13.6	19.3	24.3
IVCN	7.1	8.9	10.1	10.6	11.7	**14.6**	**17.2**
FSSE	6.8	9.0	10.7	11.3	13.1	17.0	21.5
HCCA	7.1	9.1	10.4	10.7	12.2	**15.0**	18.0
GFSI	8.2	10.9	12.2	12.5	13.4	16.9	18.2
EMXI	7.9	11.2	12.7	14.0	16.5	18.9	17.9
CTCI	7.8	10.4	11.8	12.6	14.5	**15.4**	**16.0**
TCLP	7.7	11.5	14.7	16.6	19.6	21.3	20.8
# Cases	217	204	193	175	137	109	89

Table 5b. Homogenous comparison of selected Atlantic basin early intensity guidance model biases (kt) for 2016. Biases smaller than the NHC official forecast are shown in boldface.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	-1.8	-2.2	-2.4	-2.1	-1.7	-2.9	-3.6
OCD5	-2.1	**-2.1**	**-1.3**	**-1.4**	-3.2	-6.5	-17.1
HWFI	-2.9	-2.8	**-1.5**	**0.3**	**-0.5**	**-1.4**	**-1.5**
GHMI	-4.6	-5.2	-3.6	-2.4	**-0.3**	**1.8**	**1.6**
GFNI	-4.4	-6.4	-5.8	-5.2	-4.5	-5.8	-7.6
DSHP	-2.7	-3.0	-2.9	-2.9	-2.1	**-2.7**	-3.9
LGEM	-3.3	-4.7	-5.5	-5.8	-6.7	-7.8	-9.2
IVCN	-2.9	-3.5	-2.9	-2.2	-2.3	-3.1	**-3.3**
FSSE	-1.9	**-1.4**	**-0.8**	**-1.1**	-4.5	-9.6	-13.9
HCCA	-3.3	-4.6	-4.8	-2.6	-4.4	-8.2	-11.0
GFSI	-2.8	**-1.9**	**-0.7**	**0.3**	2.0	**1.8**	**0.0**
EMXI	-1.8	**-0.5**	**0.7**	**1.9**	2.7	**1.4**	**-3.4**
CTCI	-2.6	-3.4	-2.6	**-1.6**	-3.5	-7.1	-4.1
TCLP	-3.1	-4.9	-5.1	-4.6	-4.4	-5.7	-8.7
# Cases	217	204	193	175	137	109	89

Table 6. Official Atlantic track and intensity forecast verifications (OFCL) for 2016 by storm. CLIPER5 (CLP5) and SHIFOR5 (SHF5) forecast errors are given for comparison and indicated collectively as OCD5. The number of track and intensity forecasts are given by NT and NI, respectively. Units for track and intensity errors are n mi and kt, respectively.

Verification statistics for: AL012016 ALEX

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	8	5.3	8.3	8	3.1	3.1
012	6	30.8	53.2	6	15.0	22.5
024	4	44.2	121.6	4	18.8	47.5
036	2	70.3	259.1	2	10.0	45.0
048	0	-999.0	-999.0	0	-999.0	-999.0
072	0	-999.0	-999.0	0	-999.0	-999.0
096	0	-999.0	-999.0	0	-999.0	-999.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: AL022016 BONNIE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	21	7.5	7.8	21	1.2	1.4
012	17	26.9	47.1	17	3.2	4.3
024	13	31.7	81.5	13	3.1	5.1
036	9	35.7	123.3	9	4.4	3.6
048	5	51.8	139.2	5	6.0	7.0
072	2	69.4	177.2	2	2.5	8.0
096	6	78.4	345.0	6	3.3	13.2
120	10	79.0	395.3	10	6.0	6.5

Verification statistics for: AL032016 COLIN

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	8	8.0	8.0	8	1.3	1.3
012	6	39.5	87.5	6	1.7	6.8
024	4	28.2	184.9	4	1.3	8.8
036	2	15.9	270.7	2	0.0	12.0
048	0	-999.0	-999.0	0	-999.0	-999.0
072	0	-999.0	-999.0	0	-999.0	-999.0
096	0	-999.0	-999.0	0	-999.0	-999.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: AL042016 DANIELLE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	6	10.5	11.2	6	3.3	3.3
012	4	36.5	42.6	4	3.8	6.3
024	2	55.4	51.9	2	2.5	2.5
036	0	-999.0	-999.0	0	-999.0	-999.0
048	0	-999.0	-999.0	0	-999.0	-999.0
072	0	-999.0	-999.0	0	-999.0	-999.0
096	0	-999.0	-999.0	0	-999.0	-999.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: AL052016 EARL

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	17	3.1	3.1	17	2.1	2.1
012	15	28.1	38.5	15	7.3	6.3
024	13	32.6	65.2	13	7.3	8.7
036	11	21.2	76.9	11	11.8	13.8
048	9	25.0	84.5	9	11.7	9.0
072	5	26.2	111.8	5	6.0	8.8
096	1	66.3	270.8	1	0.0	3.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: AL062016 FIONA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	26	6.7	6.3	26	2.3	2.1
012	24	24.7	34.8	24	3.5	4.2
024	22	46.1	74.8	22	4.8	6.2
036	20	60.3	128.6	20	5.5	8.6
048	18	70.7	186.8	18	6.4	12.7
072	14	116.2	245.5	14	2.9	17.6
096	10	220.6	295.3	10	4.0	25.1
120	6	311.3	365.2	6	8.3	29.2

Verification statistics for: AL072016 GASTON

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	44	6.2	6.5	44	1.9	2.2
012	42	16.1	43.8	42	4.9	7.1
024	40	22.4	90.8	40	7.0	10.3
036	38	27.8	150.8	38	7.8	13.2
048	36	34.8	220.9	36	7.5	16.6
072	32	53.6	362.6	32	9.7	20.4
096	28	84.3	435.2	28	11.1	19.8
120	24	117.8	416.4	24	8.3	17.7

Verification statistics for: AL082016 EIGHT

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	15	10.4	10.5	15	0.7	0.7
012	13	14.8	31.6	13	1.5	4.5
024	11	18.3	54.7	11	5.0	9.7
036	9	27.0	72.2	9	5.6	19.1
048	7	34.8	72.5	7	8.6	25.4
072	3	64.4	80.1	3	8.3	35.3
096	0	-999.0	-999.0	0	-999.0	-999.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: AL092016 HERMINE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	23	9.1	9.4	23	2.2	2.6
012	21	21.4	34.0	21	5.0	7.5
024	19	37.6	69.6	19	6.6	10.8
036	17	59.6	121.4	17	7.6	13.5
048	15	86.2	198.3	15	7.0	12.3
072	11	124.8	327.6	11	8.2	8.0
096	7	160.6	393.1	7	8.6	6.6
120	3	150.3	495.8	3	1.7	13.0

Verification statistics for: AL102016 IAN

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	16	13.5	13.5	16	0.6	0.6
012	14	30.2	62.3	14	4.6	2.7
024	12	48.6	131.7	12	5.8	4.3
036	10	65.3	191.4	10	4.5	5.6
048	8	98.0	317.9	8	4.4	8.0
072	4	213.8	555.7	4	3.8	7.5
096	0	-999.0	-999.0	0	-999.0	-999.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: AL112016 JULIA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	20	8.6	9.3	20	4.8	4.8
012	18	37.4	53.1	18	4.4	5.1
024	16	69.2	116.2	16	4.4	9.1
036	14	104.5	193.8	14	5.0	14.1
048	12	137.0	293.1	12	6.7	20.4
072	5	116.5	482.0	5	1.0	30.8
096	1	137.8	740.2	1	5.0	37.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: AL122016 KARL

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	44	11.0	11.3	44	0.6	0.9
012	42	32.6	53.9	42	2.9	3.9
024	40	48.8	103.3	40	5.1	6.9
036	38	57.4	149.6	38	8.8	9.7
048	36	66.2	190.0	36	11.5	12.4
072	32	75.1	278.2	32	15.9	16.3
096	28	94.2	435.3	28	22.0	21.4
120	24	132.6	538.0	24	25.8	24.3

Verification statistics for: AL132016 LISA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	22	5.9	6.2	22	1.1	1.6
012	20	31.1	37.3	20	4.8	5.4
024	18	58.7	60.5	18	5.6	8.3
036	16	84.8	90.1	16	4.4	7.6
048	14	113.5	122.6	14	5.7	11.1
072	10	166.6	193.5	10	5.5	13.8
096	6	214.0	284.9	6	4.2	19.7
120	2	268.2	386.5	2	2.5	22.0

Verification statistics for: AL142016 MATTHEW

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	44	3.9	3.9	44	3.6	4.0
012	42	17.6	34.6	42	7.7	11.6
024	40	27.6	82.1	40	12.5	15.8
036	38	42.4	141.3	38	16.8	16.6
048	36	58.8	202.8	36	19.6	18.4
072	32	87.2	343.7	32	21.1	22.0
096	28	135.1	451.5	28	22.5	27.9
120	24	170.8	486.4	24	22.1	55.5

Verification statistics for: AL152016 NICOLE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	56	6.0	6.2	56	2.9	3.2
012	54	20.7	49.3	54	7.2	9.6
024	52	29.5	106.5	52	10.3	14.7
036	50	37.5	178.4	50	11.9	18.5
048	48	51.1	244.5	48	11.6	18.6
072	44	93.4	329.4	44	11.6	18.3
096	40	160.4	393.0	40	14.0	18.4
120	36	214.9	496.7	36	17.1	19.0

Verification statistics for: AL162016 OTTO

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	16	6.8	7.3	16	4.1	4.1
012	16	18.4	35.1	16	9.7	10.3
024	16	21.7	77.7	16	11.9	13.1
036	16	27.7	129.2	16	11.6	13.8
048	14	30.9	195.1	14	10.4	14.2
072	10	53.6	402.0	10	12.5	21.6
096	6	158.0	624.1	6	16.7	14.2
120	2	272.1	892.1	2	10.0	12.0

Table 7. Homogenous comparison of official and CLIPER5 track forecast errors in the eastern North Pacific basin in 2016 for all tropical cyclones. Averages for the previous 5-yr period are shown for comparison.

	Forecast Period (h)						
	12	24	36	48	72	96	120
2016 mean OFCL error (n mi)	18.5	28.7	37.4	46.1	63.7	82.5	104.2
2016 mean CLIPER5 error (n mi)	31.0	63.9	98.5	130.5	175.3	213.2	266.7
2016 mean OFCL skill relative to CLIPER5 (%)	40.3	55.1	62.0	64.7	63.7	61.3	60.9
2016 mean OFCL bias vector (°/n mi)	315/003	301/006	275/008	260/009	241/009	204/013	209/012
2016 number of cases	359	323	288	256	196	154	119
2011-2015 mean OFCL error (n mi)	23.6	36.2	46.9	59.2	87.8	122.0	159.7
2011-2015 mean CLIPER5 error (n mi)	36.8	73.9	115.8	157.9	240.6	315.3	402.5
2011-2015 mean OFCL skill relative to CLIPER5 (%)	35.9	51.0	59.5	62.5	63.5	61.3	60.3
2011-2015 mean OFCL bias vector (°/n mi)	341/002	351/002	034/003	060/006	073/016	067/032	050/044
2011-2015 number of cases	1488	1335	1184	1043	797	592	416
2016 OFCL error relative to 2011-2015 mean (%)	-21.6	-20.7	-20.3	-22.1	-27.4	-32.4	-34.8
2016 CLIPER5 error relative to 2011-2015 mean (%)	-15.8	-13.5	-14.9	-17.4	-27.1	-32.4	-33.7

Table 8a. Homogenous comparison of eastern North Pacific basin early track guidance model errors (n mi) for 2016. Errors smaller than the NHC official forecast are shown in boldface.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	17.3	26.7	35.1	42.9	57.5	77.3	105.2
OCD5	28.9	61.4	96.1	128.0	174.6	223.7	303.6
GFSI	19.2	29.5	38.8	47.6	66.2	88.2	124.6
GHMI	22.1	37.2	52.0	65.9	94.0	139.2	200.6
HWFI	20.0	30.7	41.1	50.4	71.4	92.4	115.3
EMXI	18.7	30.8	41.8	51.1	68.3	94.1	131.4
CMCI	22.3	37.0	49.7	64.6	94.7	127.4	201.2
EGRI	22.3	36.7	49.2	59.3	77.1	103.7	155.3
NVGI	24.2	39.7	54.9	67.8	91.6	125.4	198.9
GFNI	23.6	39.5	55.6	71.6	114.4	166.5	233.9
CTCI	19.9	32.6	44.3	53.1	70.4	98.4	136.0
AEMI	18.5	29.0	38.5	48.9	65.9	86.2	127.2
UEMI	22.3	37.6	52.0	63.7	82.4	107.5	156.3
FSSE	**17.2**	**26.4**	35.3	**42.5**	57.9	78.6	116.3
TVCE	**17.2**	26.7	35.6	**42.1**	**53.4**	**68.8**	**91.2**
HCCA	**16.6**	**25.3**	**33.5**	**40.7**	**56.2**	**74.1**	**104.9**
BAMD	31.8	57.6	82.1	106.1	150.1	217.6	320.3
BAMM	27.5	48.2	67.3	85.7	120.0	154.9	191.6
BAMS	31.9	56.8	79.2	99.6	131.4	156.2	180.7
TCLP	29.1	61.8	96.5	129.5	177.9	221.0	274.9
# Cases	255	234	213	188	142	104	69

Table 8b. Homogenous comparison of eastern North Pacific basin early track guidance model bias vectors (º/n mi) for 2016.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	328/005	315/008	296/009	284/010	268/008	221/009	137/007
OCD5	357/003	042/004	073/009	067/014	044/046	054/103	068/218
GFSI	340/004	326/006	304/006	283/003	073/001	124/011	110/022
GHMI	075/007	061/017	054/026	054/034	063/056	073/097	078/148
HWFI	327/007	317/011	302/016	299/019	290/020	227/020	185/036
EMXI	313/007	304/011	289/014	275/019	265/028	248/037	252/027
CMCI	052/003	071/007	080/011	084/018	102/031	127/057	124/096
EGRI	298/008	286/016	279/022	277/028	284/035	281/032	307/035
NVGI	355/008	344/011	328/010	318/011	333/022	019/044	026/079
GFNI	106/005	090/016	077/028	069/038	059/066	066/115	066/156
CTCI	359/008	358/012	356/013	356/015	347/022	346/029	329/054
AEMI	334/003	320/004	315/002	151/001	107/008	128/017	146/028
UEMI	295/009	285/015	275/021	271/026	263/029	237/027	173/026
FSSE	333/002	332/004	300/004	277/005	256/006	153/011	107/032
TVCE	341/005	335/008	324/010	323/011	330/012	023/005	059/016
HCCA	332/002	307/004	287/006	273/008	259/013	206/018	146/028
BAMD	018/015	018/031	015/048	012/066	007/095	010/146	021/243
BAMM	356/013	353/025	351/038	346/050	334/068	340/078	003/098
BAMS	339/015	332/028	325/040	315/055	293/082	272/087	250/079
TCLP	320/003	310/004	305/002	327/004	358/026	025/054	062/120
# Cases	255	234	213	188	142	104	69

Table 9. Homogenous comparison of official and Decay-SHIFOR5 intensity forecast errors in the eastern North Pacific basin for the 2016 season for all tropical cyclones. Averages for the previous 5-yr period are shown for comparison.

	Forecast Period (h)						
	12	24	36	48	72	96	120
2016 mean OFCL error (kt)	5.8	8.7	10.4	11.8	15.6	16.8	16.5
2016 mean Decay-SHIFOR5 error (kt)	7.6	11.4	14.4	16.7	21.0	21.6	19.3
2016 mean OFCL skill relative to Decay-SHIFOR5 (%)	23.7	23.7	27.8	29.3	26.2	22.2	14.5
2016 OFCL bias (kt)	-0.6	-1.2	-2.2	-3.8	-4.8	-4.0	-4.9
2016 number of cases	359	323	288	256	196	154	119
2011-15 mean OFCL error (kt)	6.0	10.0	12.6	14.0	15.4	15.9	14.7
2011-15 mean Decay-SHIFOR5 error (kt)	7.8	12.9	16.6	19.1	21.2	22.0	20.6
2011-15 mean OFCL skill relative to Decay-SHIFOR5 (%)	23.1	22.5	24.1	26.7	27.4	27.7	28.6
2011-15 OFCL bias (kt)	-0.8	-1.2	-2.1	-2.9	-2.8	-2.7	-1.9
2011-15 number of cases	1488	1335	1184	1043	797	592	416
2016 OFCL error relative to 2011-15 mean (%)	-3.3	-13.0	-17.5	-15.7	1.3	5.4	12.2
2016 Decay-SHIFOR5 error relative to 2011-15 mean (%)	-2.6	-11.6	-13.3	-12.6	-0.9	-1.8	-6.3

Table 10a. Homogenous comparison of eastern North Pacific basin early intensity guidance model errors (kt) for 2016. Errors smaller than the NHC official forecast are shown in boldface.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	6.0	8.9	10.5	12.2	15.1	18.0	16.4
OCD5	8.0	11.4	14.3	16.7	19.8	21.0	18.8
HWFI	7.3	9.5	**10.3**	**10.4**	**12.9**	**16.1**	**15.6**
GHMI	8.7	12.4	15.4	16.5	17.7	**17.7**	18.1
GFNI	8.0	11.5	13.8	15.4	16.8	19.8	22.2
DSHP	7.0	10.0	12.0	13.4	17.9	21.9	20.8
LGEM	7.1	9.9	12.0	13.4	16.3	18.5	16.8
IVCN	6.9	9.2	10.7	**11.8**	**13.7**	**16.1**	**15.2**
FSSE	6.6	9.3	10.7	**12.0**	16.1	19.1	18.1
HCCA	6.8	9.7	11.2	**10.8**	**14.1**	**16.8**	**15.4**
GFSI	7.6	10.8	12.7	14.6	18.1	19.8	21.4
EMXI	8.2	12.1	15.0	17.2	20.5	23.7	25.1
CTCI	7.5	11.2	13.2	14.6	15.6	**16.5**	17.1
TCLP	8.1	11.9	15.1	17.7	20.6	21.5	20.9
# Cases	269	244	215	186	142	105	72

Table 10b. Homogenous comparison of eastern North Pacific basin early intensity guidance model biases (kt) for 2016. Biases smaller than the NHC official forecast are shown in boldface.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	-0.3	-0.7	-1.7	-3.7	-4.2	-3.6	-5.1
OCD5	-1.6	-3.6	-6.2	-9.7	-12.8	-14.5	-14.9
HWFI	-3.4	-3.5	-2.6	**-2.4**	**-1.6**	**0.2**	**1.7**
GHMI	-5.0	-7.9	-9.2	-10.7	-10.2	-7.5	-5.4
GFNI	-4.1	-7.0	-8.3	-9.5	-10.3	-8.5	-6.5
DSHP	-0.7	-1.5	-2.2	**-3.3**	-4.7	-6.0	-8.0
LGEM	-1.3	-3.6	-6.0	-8.4	-11.3	-12.2	-13.4
IVCN	-2.2	-4.2	-5.4	-6.4	-7.0	-6.1	-5.6
FSSE	-0.8	-0.7	**-0.5**	**-1.3**	**-3.6**	-5.6	-9.0
HCCA	-0.7	-0.8	**-0.8**	**-1.1**	**-3.9**	-4.1	-6.3
GFSI	-0.8	**0.4**	**1.2**	**1.3**	**2.3**	5.4	8.0
EMXI	**0.1**	1.4	2.2	**2.3**	5.0	9.4	13.4
CTCI	-2.4	-5.9	-8.2	-9.0	-8.5	-6.8	**-3.4**
TCLP	-1.8	-5.2	-8.7	-12.2	-16.3	-17.9	-19.2
# Cases	269	244	215	186	142	105	72

Table 11. Official eastern North Pacific track and intensity forecast verifications (OFCL) for 2016 by storm. CLIPER5 (CLP5) and SHIFOR5 (SHF5) forecast errors are given for comparison and indicated collectively as OCD5. The number of track and intensity forecasts are given by NT and NI, respectively. Units for track and intensity errors are n mi and kt, respectively.

Verification statistics for: EP012016 ONE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	5	6.9	10.2	5	0.0	0.0
012	3	22.4	38.8	3	0.0	4.3
024	1	25.0	98.7	1	0.0	1.0
036	0	-999.0	-999.0	0	-999.0	-999.0
048	0	-999.0	-999.0	0	-999.0	-999.0
072	0	-999.0	-999.0	0	-999.0	-999.0
096	0	-999.0	-999.0	0	-999.0	-999.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: EP022016 AGATHA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	13	7.8	8.5	13	1.5	1.5
012	11	22.1	25.8	11	1.8	3.5
024	9	34.5	46.9	9	2.8	5.2
036	7	42.7	76.3	7	3.6	5.6
048	5	55.6	124.4	5	5.0	4.2
072	1	132.5	247.9	1	5.0	8.0
096	0	-999.0	-999.0	0	-999.0	-999.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: EP032016 BLAS

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	29	6.3	6.7	29	1.4	1.4
012	27	16.4	27.6	27	4.8	8.1
024	25	24.6	52.7	25	6.4	12.0
036	23	28.4	80.8	23	6.1	16.0
048	21	33.9	104.3	21	7.6	19.2
072	17	50.7	142.0	17	9.1	24.6
096	13	82.3	144.3	13	9.6	25.5
120	9	108.1	100.2	9	6.7	16.9

Verification statistics for: EP042016 CELIA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	35	9.1	9.0	35	1.1	1.3
012	35	20.7	35.3	35	4.1	5.6
024	33	32.6	68.0	33	6.4	8.1
036	31	43.9	101.1	31	7.4	10.8
048	29	49.3	125.5	29	9.0	12.2
072	25	61.5	165.3	25	12.4	13.7
096	21	68.5	189.7	21	13.8	15.6
120	17	67.6	204.9	17	15.9	11.6

Verification statistics for: EP052016 DARBY

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	36	5.7	5.8	36	1.8	2.1
012	36	13.7	25.0	36	5.8	7.1
024	36	25.1	53.1	36	7.8	10.4
036	36	36.3	78.2	36	9.0	14.8
048	36	44.1	102.1	36	8.9	17.6
072	36	68.0	153.1	36	11.9	21.6
096	36	95.8	201.3	36	15.1	20.3
120	36	127.1	241.3	36	14.2	15.5

Verification statistics for: EP062016 ESTELLE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	26	7.6	7.6	26	1.3	1.3
012	24	15.5	20.7	24	4.0	4.0
024	22	24.5	41.0	22	6.6	6.0
036	20	29.4	62.1	20	9.5	5.9
048	18	40.4	85.5	18	13.3	6.3
072	14	54.3	122.1	14	17.1	3.5
096	10	51.2	155.9	10	17.5	7.6
120	6	33.4	198.5	6	13.3	6.8

Verification statistics for: EP072016 FRANK

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	27	7.2	7.9	27	1.3	1.1
012	25	19.9	28.4	25	6.6	6.5
024	23	33.2	58.9	23	9.8	11.0
036	21	41.9	92.5	21	10.7	10.9
048	19	48.0	124.6	19	13.9	12.4
072	15	55.3	159.2	15	18.3	11.6
096	11	70.5	161.2	11	16.8	15.9
120	7	120.4	160.7	7	14.3	21.3

Verification statistics for: EP082016 GEORGETTE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	22	4.5	4.5	22	0.5	0.5
012	20	17.5	29.4	20	6.3	10.4
024	18	28.4	64.8	18	13.1	18.7
036	16	27.8	92.8	16	15.0	22.5
048	14	24.4	115.3	14	13.2	22.8
072	10	41.0	169.0	10	17.5	25.5
096	6	54.2	209.2	6	18.3	17.7
120	2	65.8	221.1	2	7.5	8.0

Verification statistics for: EP092016 HOWARD

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	12	9.4	8.4	12	0.8	1.3
012	10	17.4	25.6	10	2.5	3.9
024	8	26.5	68.5	8	3.1	6.6
036	6	27.9	131.3	6	2.5	10.7
048	4	31.9	185.2	4	3.8	8.3
072	0	-999.0	-999.0	0	-999.0	-999.0
096	0	-999.0	-999.0	0	-999.0	-999.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: EP102016 IVETTE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	21	8.9	8.9	21	0.7	0.7
012	21	16.5	23.6	21	3.8	4.5
024	19	26.2	42.4	19	6.3	4.9
036	17	33.6	56.2	17	11.2	7.8
048	15	43.3	74.3	15	16.3	9.0
072	11	64.7	115.4	11	25.9	14.8
096	7	89.3	141.4	7	32.9	16.4
120	3	137.3	195.0	3	35.0	17.3

Verification statistics for: EP112016 JAVIER

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	10	25.3	27.1	10	2.0	2.0
012	8	33.5	60.1	8	8.8	9.8
024	6	43.9	95.6	6	14.2	8.2
036	4	60.8	150.0	4	17.5	12.8
048	2	68.7	215.6	2	17.5	10.5
072	0	-999.0	-999.0	0	-999.0	-999.0
096	0	-999.0	-999.0	0	-999.0	-999.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: EP122016 KAY

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	19	11.3	11.1	19	1.6	1.8
012	17	21.5	28.6	17	2.9	4.3
024	15	28.4	52.6	15	4.7	6.1
036	13	28.0	68.7	13	5.0	5.1
048	11	41.8	83.2	11	5.5	3.5
072	7	114.5	59.3	7	11.4	7.9
096	3	273.3	81.6	3	10.0	10.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: EP132016 LESTER

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	29	5.4	5.4	29	1.7	1.9
012	29	14.4	25.8	29	9.5	11.2
024	29	21.3	50.5	29	15.5	17.8
036	29	28.1	84.5	29	16.2	20.6
048	29	31.8	119.1	29	16.6	27.1
072	29	44.1	205.5	29	20.5	38.2
096	29	56.6	311.6	29	22.1	38.0
120	29	82.8	439.9	29	24.0	37.2

Verification statistics for: EP142016 MADELINE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	6	15.0	15.0	6	0.8	0.8
012	6	25.0	43.2	6	6.7	7.3
024	6	24.6	58.5	6	10.0	8.7
036	6	26.3	71.1	6	11.7	9.8
048	6	34.9	82.2	6	16.7	25.8
072	6	51.5	113.5	6	36.7	49.3
096	6	87.2	157.4	6	15.8	29.3
120	6	164.4	219.7	6	8.3	6.3

Verification statistics for: EP152016 NEWTON

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	12	7.0	7.0	12	4.2	3.8
012	10	20.6	58.9	10	9.5	13.3
024	8	40.0	155.4	8	10.6	15.0
036	6	71.0	280.6	6	14.2	19.5
048	4	118.8	419.5	4	10.0	14.0
072	0	-999.0	-999.0	0	-999.0	-999.0
096	0	-999.0	-999.0	0	-999.0	-999.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: EP162016 ORLENE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	24	6.4	6.4	24	0.2	0.2
012	22	18.1	38.2	22	6.8	7.9
024	20	27.3	89.3	20	9.8	13.6
036	18	41.3	158.3	18	10.3	15.4
048	16	62.8	232.3	16	10.0	15.9
072	12	100.2	324.7	12	4.6	8.9
096	8	138.4	308.1	8	11.9	12.9
120	4	181.1	310.6	4	20.0	3.5

Verification statistics for: EP172016 PAINE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	11	11.9	11.9	11	1.4	1.8
012	9	25.2	35.0	9	10.0	14.4
024	7	27.4	73.7	7	14.3	20.0
036	5	30.5	140.9	5	17.0	23.6
048	3	49.4	183.8	3	15.0	19.0
072	0	-999.0	-999.0	0	-999.0	-999.0
096	0	-999.0	-999.0	0	-999.0	-999.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: EP182016 ROSLYN

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	16	8.2	8.6	16	1.9	2.2
012	14	27.5	41.2	14	4.3	4.8
024	12	52.5	96.3	12	5.0	5.3
036	10	80.4	164.4	10	5.5	6.1
048	8	102.8	256.5	8	6.3	7.4
072	4	136.0	420.0	4	5.0	8.8
096	0	-999.0	-999.0	0	-999.0	-999.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: EP192016 ULIKA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	10	4.2	4.2	10	3.5	3.5
012	10	17.3	26.9	10	6.0	6.4
024	8	34.9	71.9	8	4.4	10.4
036	6	47.6	117.8	6	6.7	18.7
048	4	59.7	157.6	4	10.0	27.5
072	1	89.7	396.4	1	5.0	11.0
096	0	-999.0	-999.0	0	-999.0	-999.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: EP202016 SEYMOUR

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	20	4.3	4.3	20	1.8	1.8
012	18	15.3	28.4	18	6.4	12.8
024	16	22.3	71.2	16	11.9	23.9
036	14	36.2	118.8	14	19.3	33.9
048	12	51.0	164.8	12	24.6	39.3
072	8	70.5	245.2	8	23.8	39.9
096	4	105.9	362.7	4	18.8	19.5
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: EP212016 TINA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	2	2.8	2.8	2	2.5	2.5
012	0	-999.0	-999.0	0	-999.0	-999.0
024	0	-999.0	-999.0	0	-999.0	-999.0
036	0	-999.0	-999.0	0	-999.0	-999.0
048	0	-999.0	-999.0	0	-999.0	-999.0
072	0	-999.0	-999.0	0	-999.0	-999.0
096	0	-999.0	-999.0	0	-999.0	-999.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: EP222016 OTTO

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	6	8.1	8.1	6	5.0	5.0
012	4	18.6	60.7	4	8.8	14.0
024	2	19.1	144.5	2	17.5	19.0
036	0	-999.0	-999.0	0	-999.0	-999.0
048	0	-999.0	-999.0	0	-999.0	-999.0
072	0	-999.0	-999.0	0	-999.0	-999.0
096	0	-999.0	-999.0	0	-999.0	-999.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Table 12a. Verification of 48-h probabilistic genesis forecasts for the Atlantic basin in 2016.

Atlantic Basin Genesis Forecast Reliability Table		
Forecast Likelihood (%)	**Verifying Genesis Occurrence Rate (%)**	**Number of Forecasts**
0	5	251
10	12	131
20	17	64
30	33	40
40	59	27
50	48	29
60	42	12
70	92	13
80	100	12
90	100	7
100	-	0

Table 12b. Verification of 48-h probabilistic genesis forecasts for the eastern North Pacific basin in 2016.

Eastern North Pacific Basin Genesis Forecast Reliability Table		
Forecast Likelihood (%)	**Verifying Genesis Occurrence Rate (%)**	**Number of Forecasts**
0	4	445
10	17	139
20	17	96
30	33	52
40	39	36
50	57	35
60	48	23
70	76	34
80	100	14
90	100	7
100	-	0

Table 13a. Verification of 120-h probabilistic genesis forecasts for the Atlantic basin in 2016.

Atlantic Basin Genesis Forecast Reliability Table		
Forecast Likelihood (%)	**Verifying Genesis Occurrence Rate (%)**	**Number of Forecasts**
0	8	25
10	20	195
20	34	83
30	60	52
40	35	37
50	47	51
60	59	44
70	85	39
80	83	35
90	100	25
100	-	0

Table 13b. Verification of 120-h probabilistic genesis forecasts for the eastern North Pacific basin in 2016.

Eastern North Pacific Basin Genesis Forecast Reliability Table		
Forecast Likelihood (%)	**Verifying Genesis Occurrence Rate (%)**	**Number of Forecasts**
0	0	7
10	26	252
20	28	174
30	43	75
40	47	59
50	48	65
60	49	61
70	60	47
80	88	75
90	91	66
100	-	0

Table 14. NHC forecast cone circle radii (n mi) for 2017. Change from 2016 values expressed in n mi and percent are given in parentheses.

Track Forecast Cone Two-Thirds Probability Circles (n mi)		
Forecast Period (h)	**Atlantic Basin**	**Eastern North Pacific Basin**
3	16 (0: 0%)	16 (0: 0%)
12	29 (-1: -3%)	25 (-2: -7%)
24	45 (-4: -8%)	40 (-2: -5%)
36	63 (-3: -5%)	51 (-4: -7%)
48	78 (-6: -7%)	66 (-4: -6%)
72	107 (-8: -7%)	93 (-7: -7%)
96	159 (-6: -4%)	116 (-21: -15%)
120	211 (-26: -11%)	151 (-21: -12%)

Table 15. Composition of NHC consensus models for 2017. It is intended that TCOA/TVCA would be the primary consensus aids for the Atlantic basin and TCOE/TVCE would be primary for the eastern Pacific.

NHC Consensus Model Definitions For 2017			
Model ID	**Parameter**	**Type**	**Members**
GFEX	Track	Fixed	GFSI EMXI
TCOA	Track	Fixed	GFSI EGRI HWFI
TCOE*	Track	Fixed	GFSI EGRI HWFI
ICON	Intensity	Fixed	DSHP LGEM HWFI
TVCA**	Track	Variable	GFSI EGRI HWFI EMXI CTCI
TVCE**	Track	Variable	GFSI EGRI HWFI EMXI CTCI
TVCX	Track	Variable	EMXI (double weight) GFSI EGRI HWFI
IVCN	Intensity	Variable	DSHP LGEM HWFI CTCI

* TCON will continue to be computed and will have the same composition as TCOE.
** TVCN will continue to be computed and will have the same composition as TVCA and TVCE. GPCE circles will continue to be based on TVCN.

LIST OF FIGURES

20. NHC official and Decay-SHIFOR5 (OCD5) eastern North Pacific basin average intensity errors for 2016 (solid lines) and 2011-2015 (dashed lines).

21. Recent trends in NHC official intensity forecast error (top) and skill (bottom) for the eastern North Pacific basin.

22. Homogenous comparison for selected eastern North Pacific basin early intensity guidance models for 2016. This verification includes only those models that were available at least 2/3 of the time (see text).

23. Homogenous comparison for selected eastern North Pacific basin early intensity guidance models for 2014-16.

24. Reliability diagram for Atlantic (top) and eastern North Pacific (bottom) probabilistic tropical cyclogenesis 48-h forecasts for 2016. The solid lines indicate the relationship between the forecasts and verifying genesis percentages, with perfect reliability indicated by the thin diagonal black line. The dashed lines indicate how the forecasts were distributed among the possible forecast values.

25. As described for Fig. 24, but for 120-h forecasts.

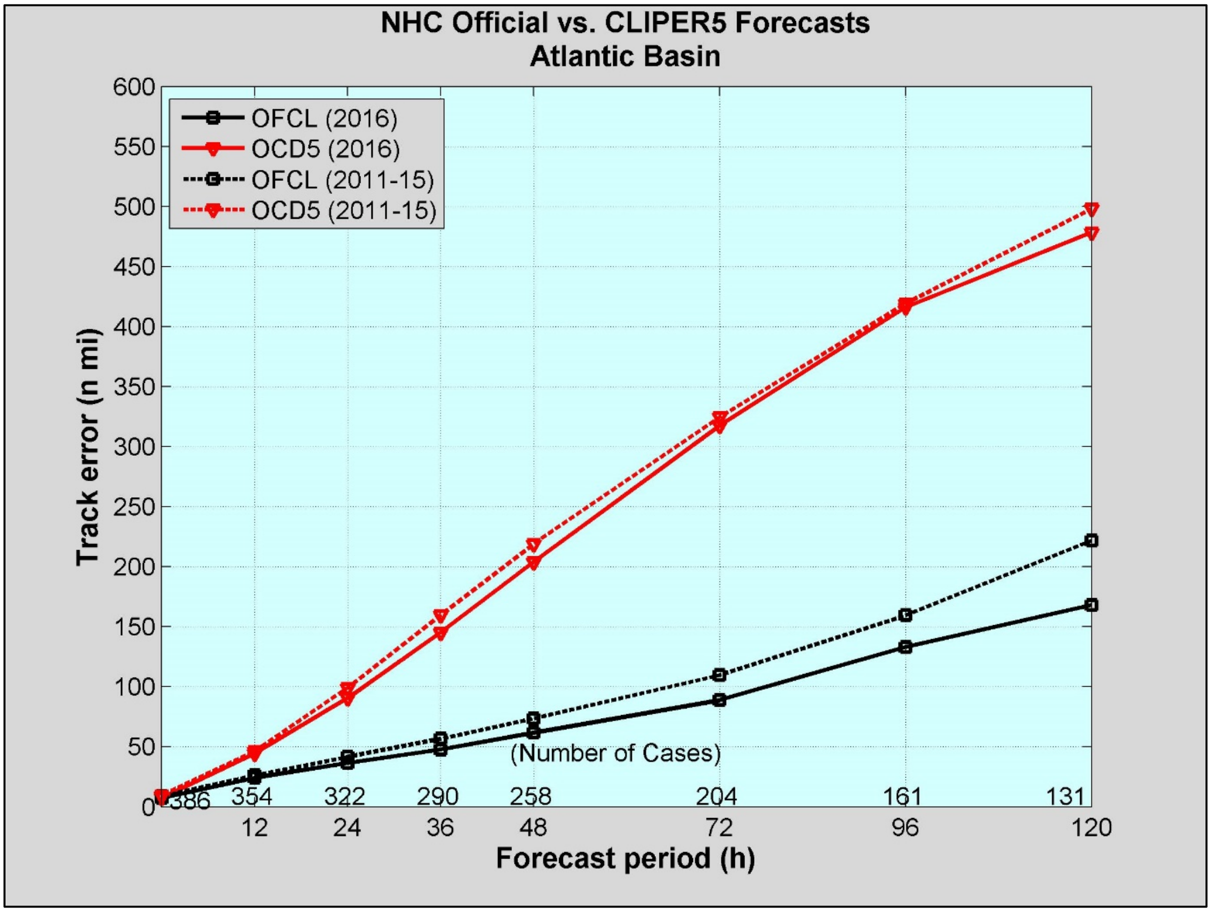


Figure 1. NHC official and CLIPER5 (OCD5) Atlantic basin average track errors for 2016 (solid lines) and 2011-2015 (dashed lines).

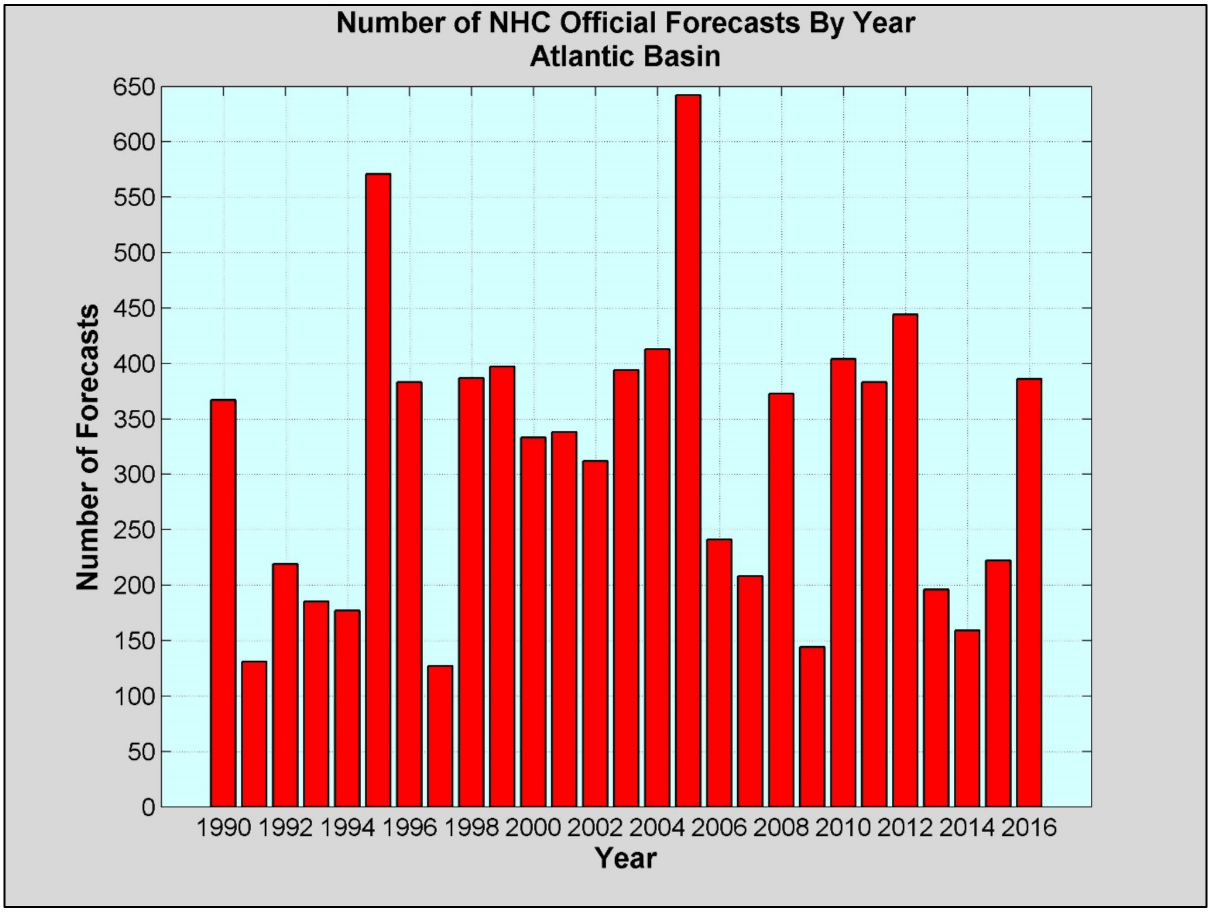


Figure 2. Number of NHC official forecasts for the Atlantic basin stratified by year.

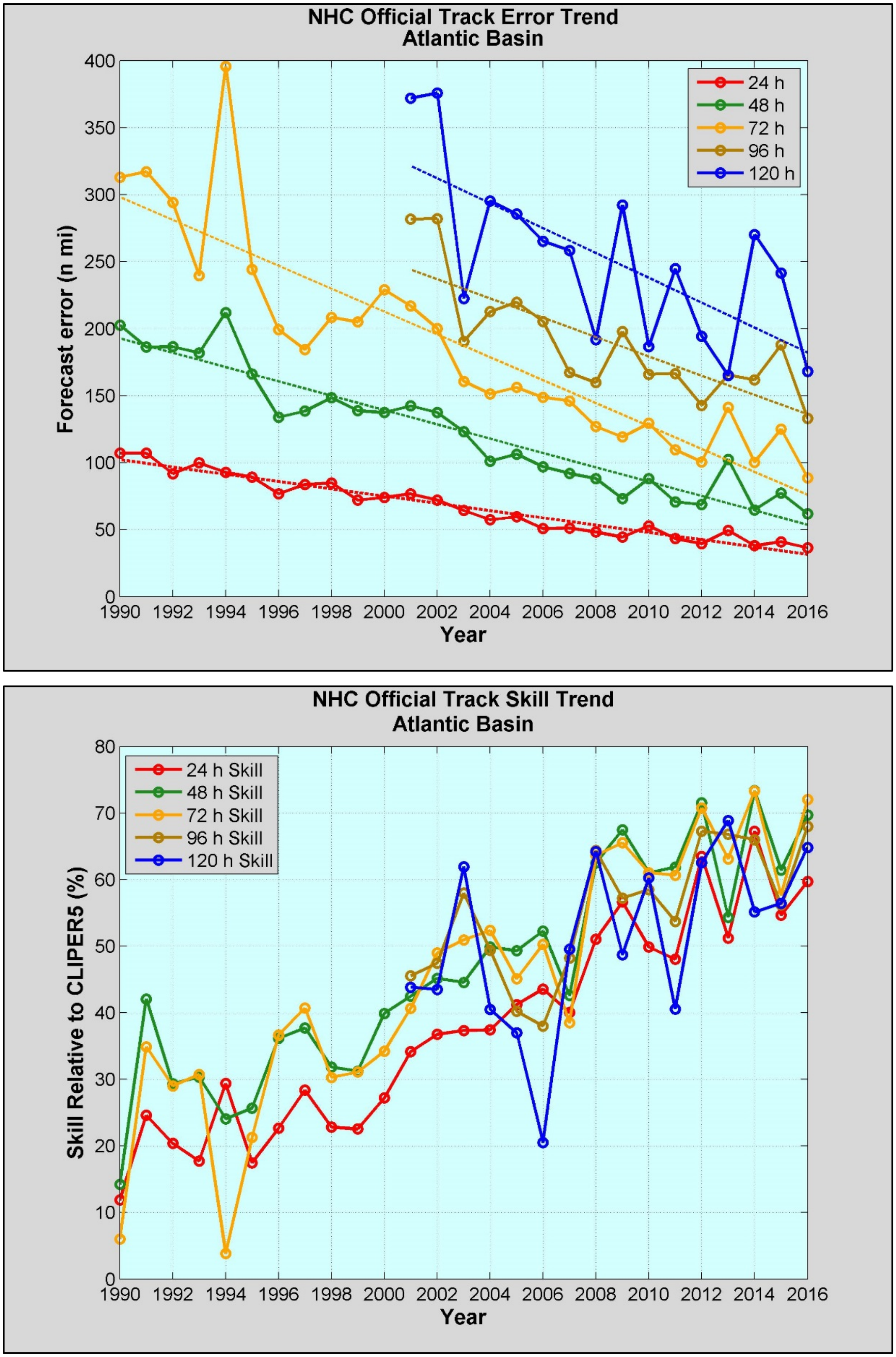


Figure 3. Recent trends in NHC official track forecast error (top) and skill (bottom) for the Atlantic basin.

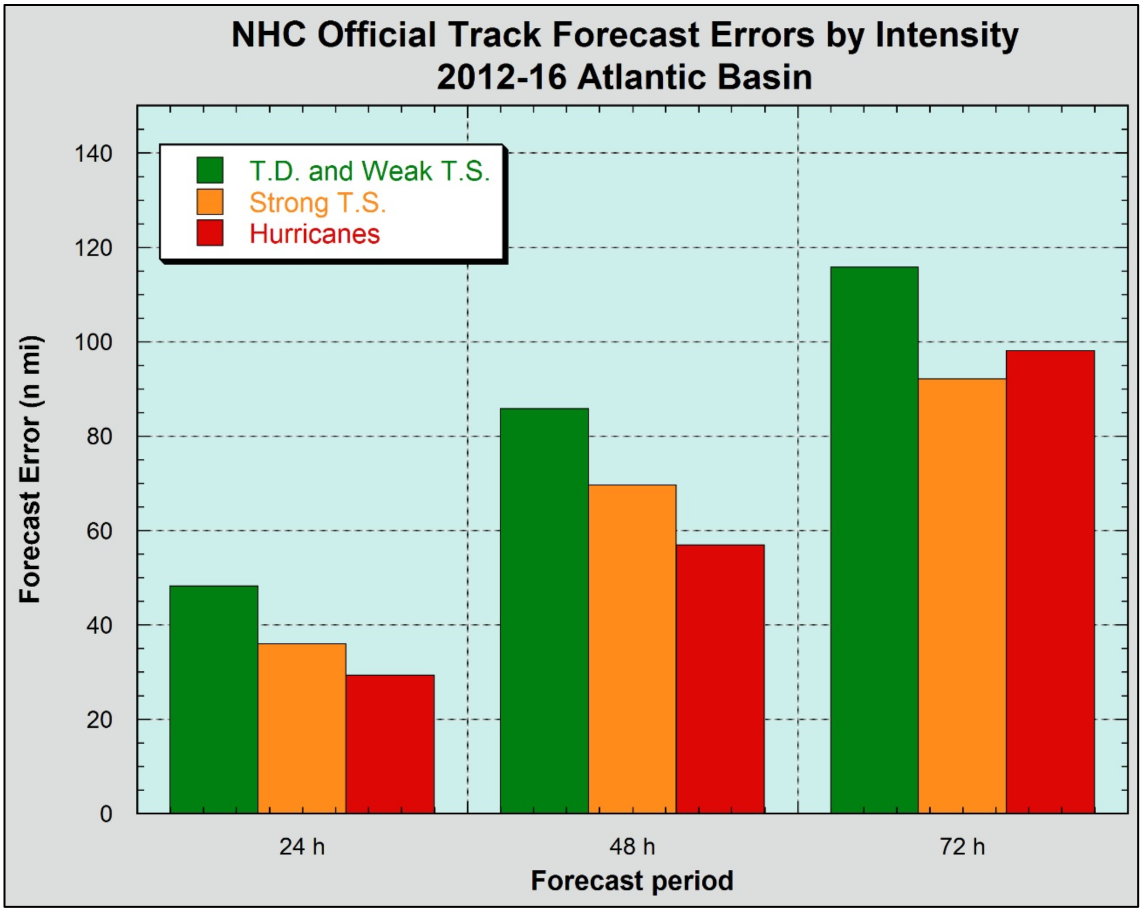


Figure 4. 2012-16 NHC official track forecast error binned by initial intensity for the Atlantic basin. Weak tropical storms are in the 35-45 kt range and strong tropical storms are in the 50-60 kt range.

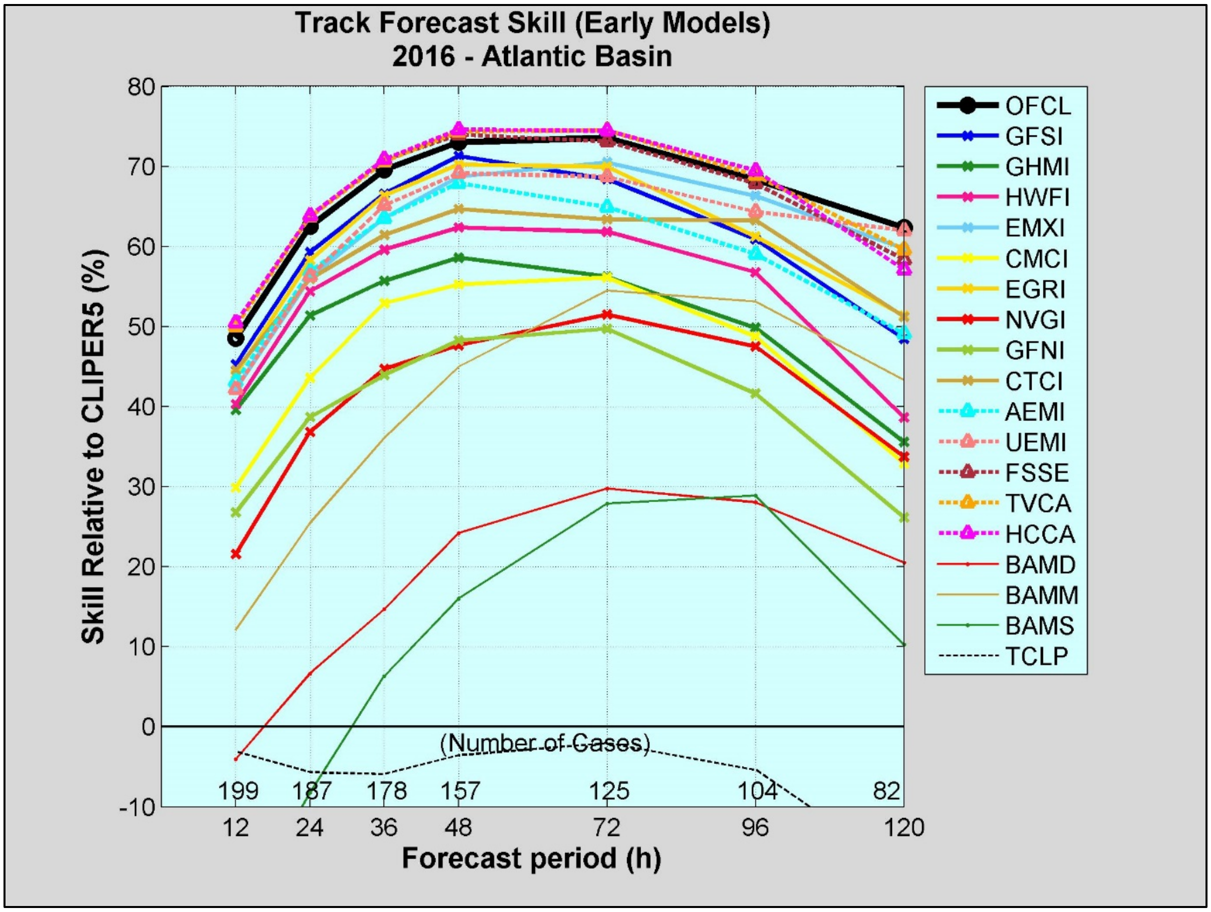


Figure 5. Homogenous comparison for selected Atlantic basin early track models for 2016. This verification includes only those models that were available at least 2/3 of the time (see text).

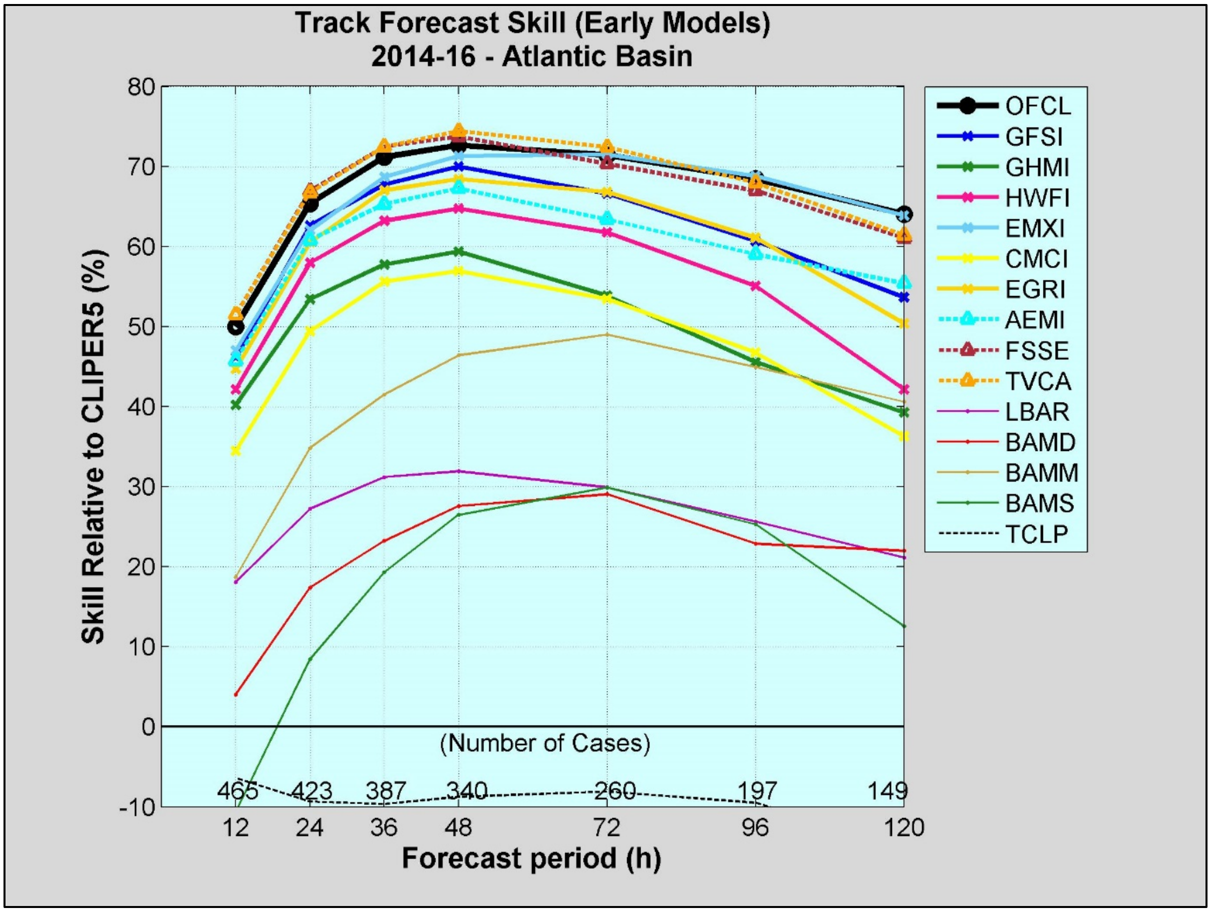


Figure 6. Homogenous comparison for selected Atlantic basin early track models for 2014-2016.

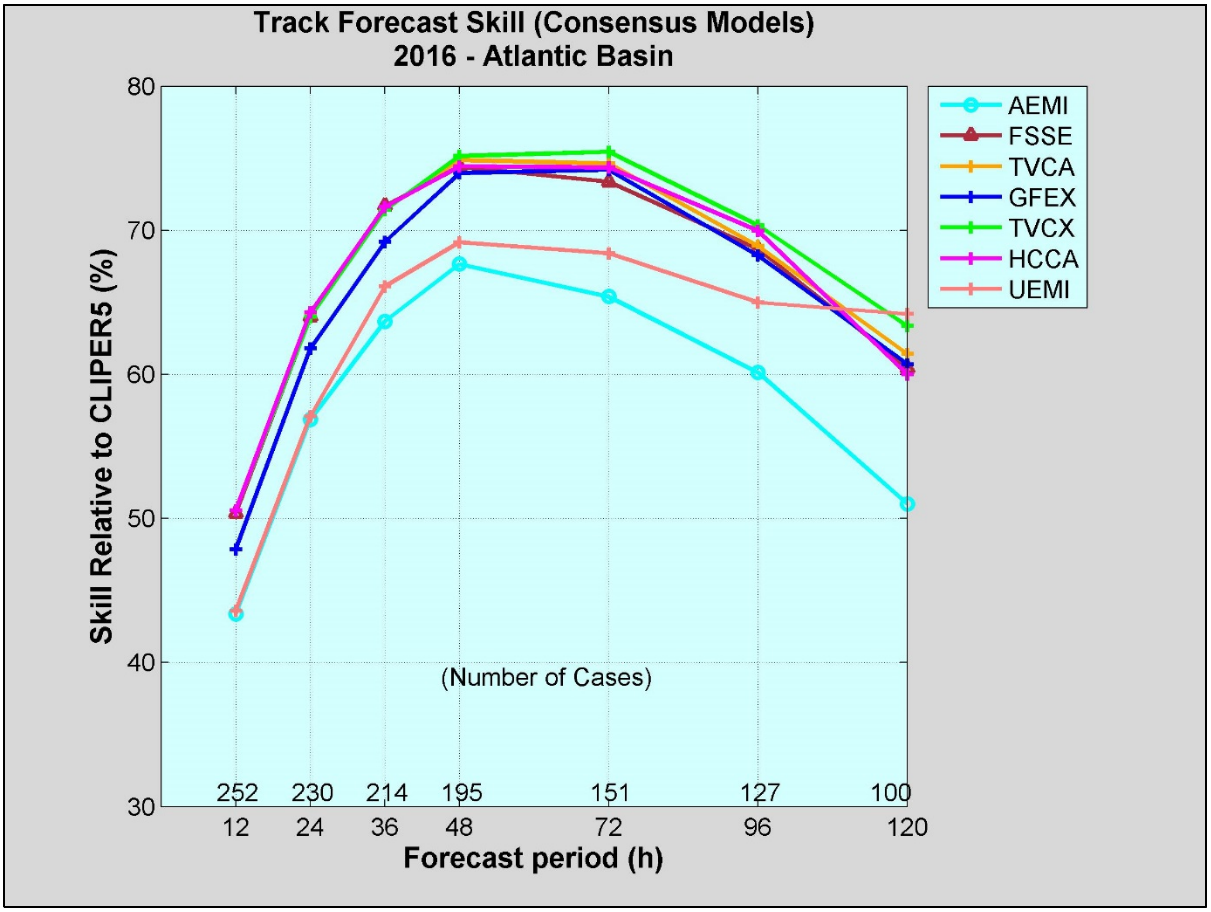


Figure 7. Homogenous comparison of the primary Atlantic basin track consensus models for 2016.

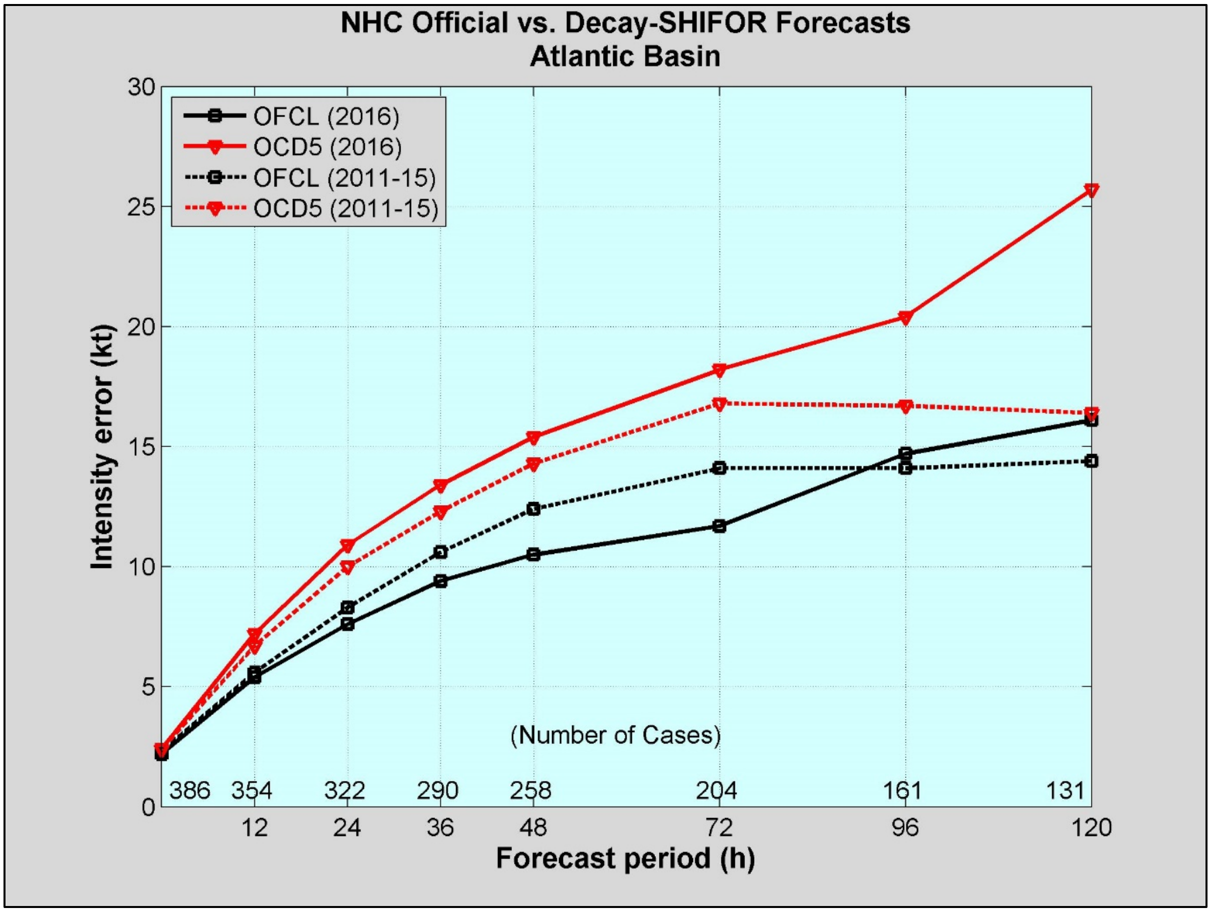


Figure 8. NHC official and Decay-SHIFOR5 (OCD5) Atlantic basin average intensity errors for 2016 (solid lines) and 2011-2015 (dashed lines).

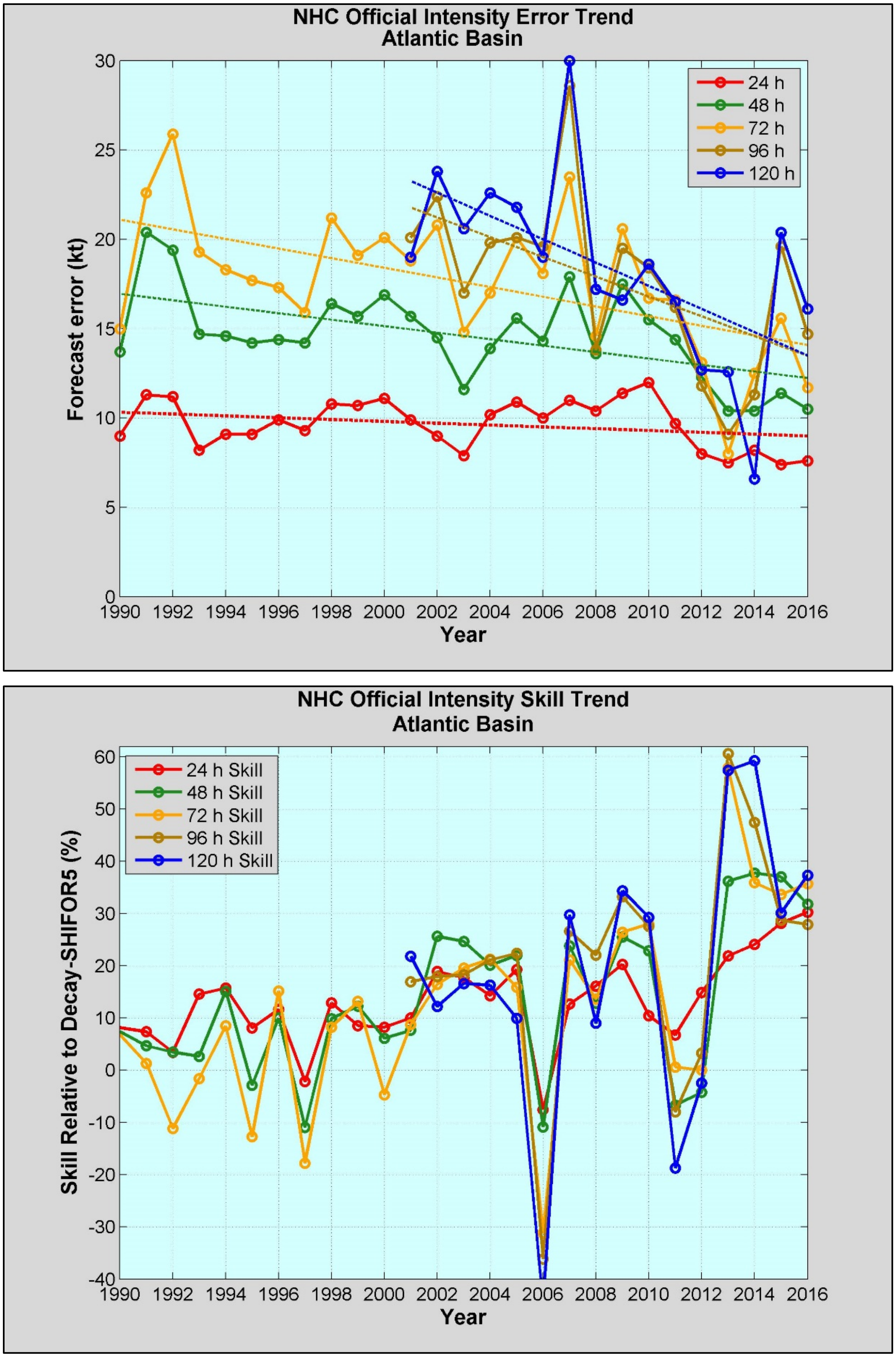


Figure 9. Recent trends in NHC official intensity forecast error (top) and skill (bottom) for the Atlantic basin.

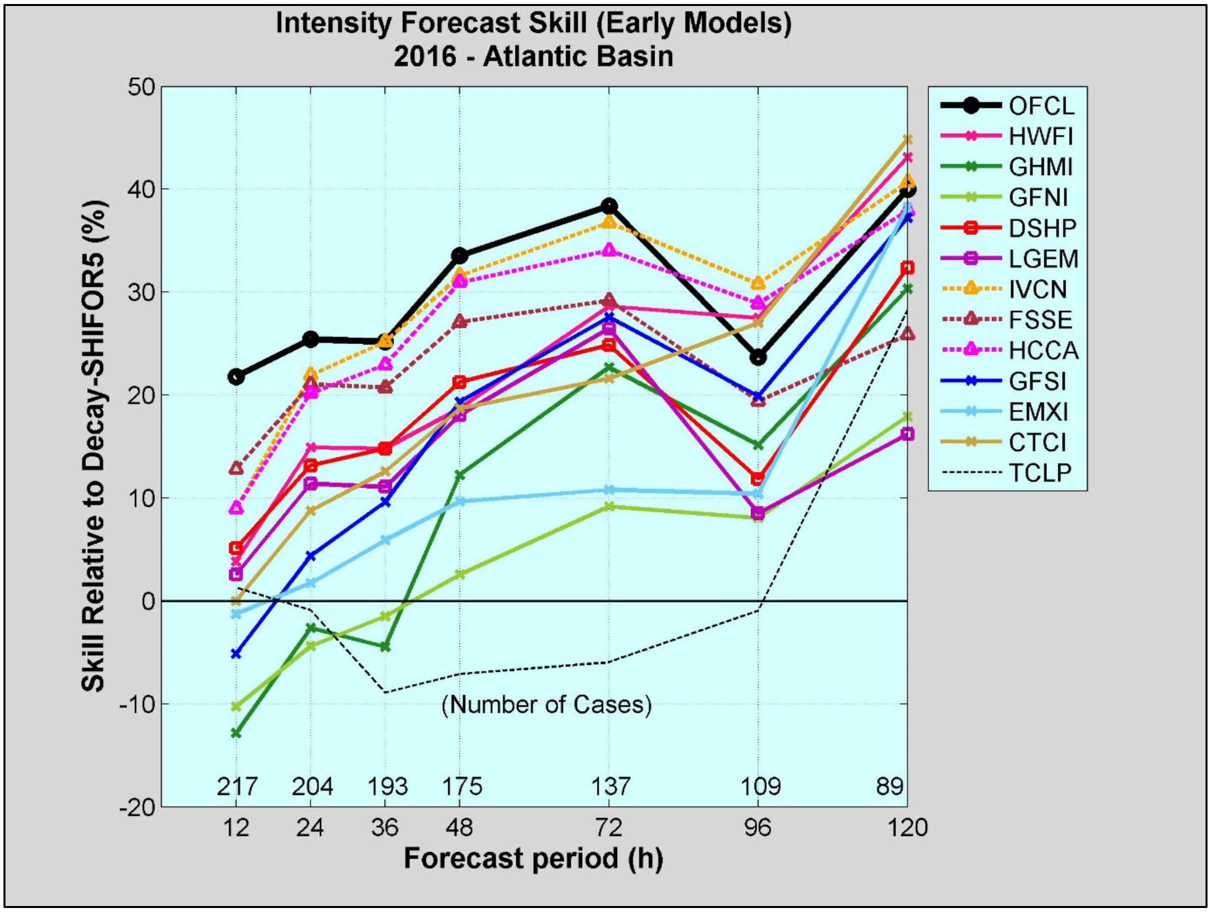


Figure 10. Homogenous comparison for selected Atlantic basin early intensity guidance models for 2016.

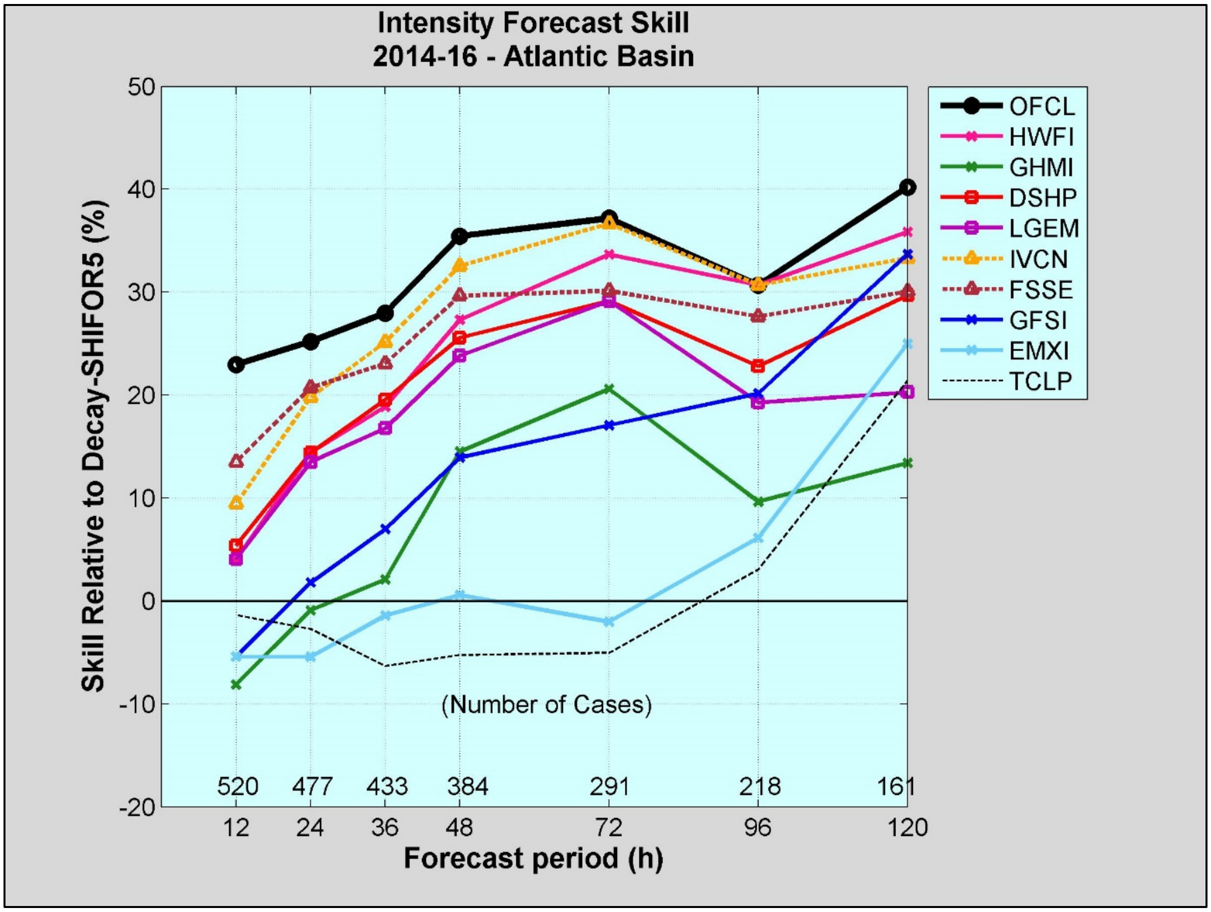


Figure 11. Homogenous comparison for selected Atlantic basin early intensity guidance models for 2014-2016.

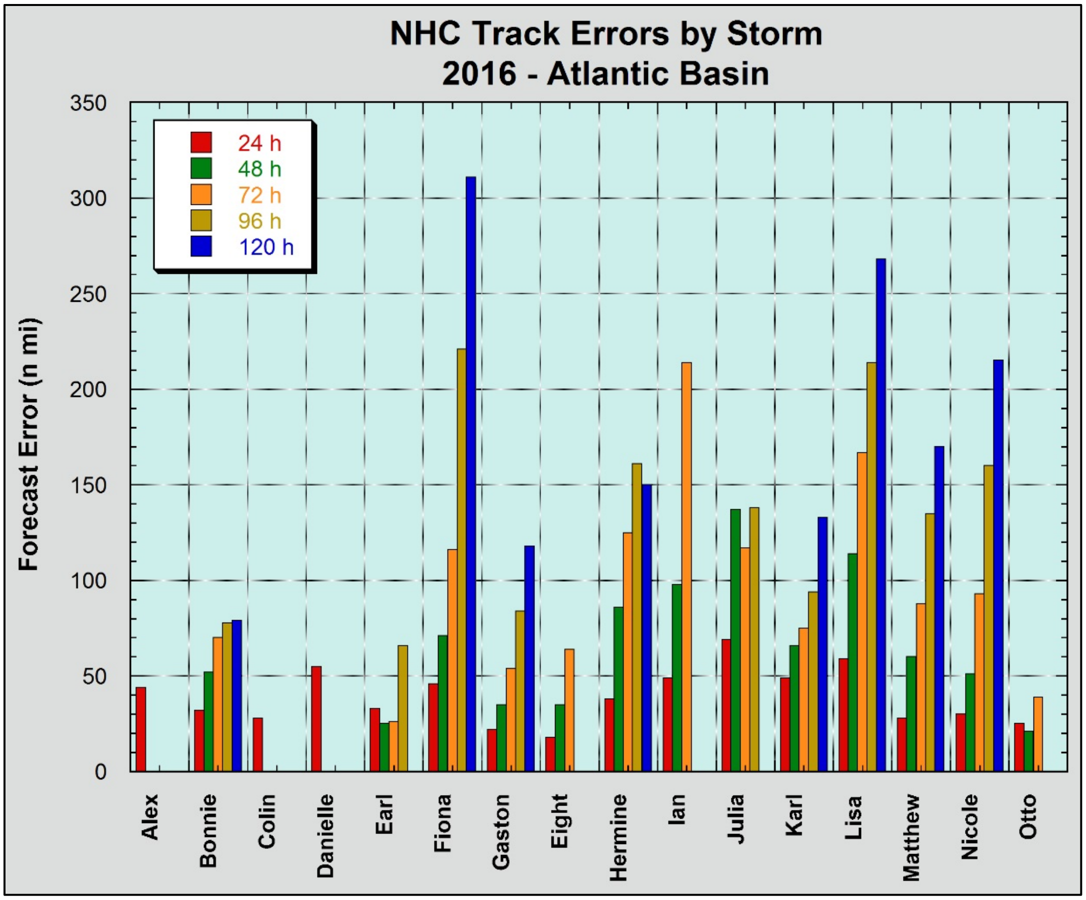


Figure 12. 2016 NHC official track errors by tropical cyclone.

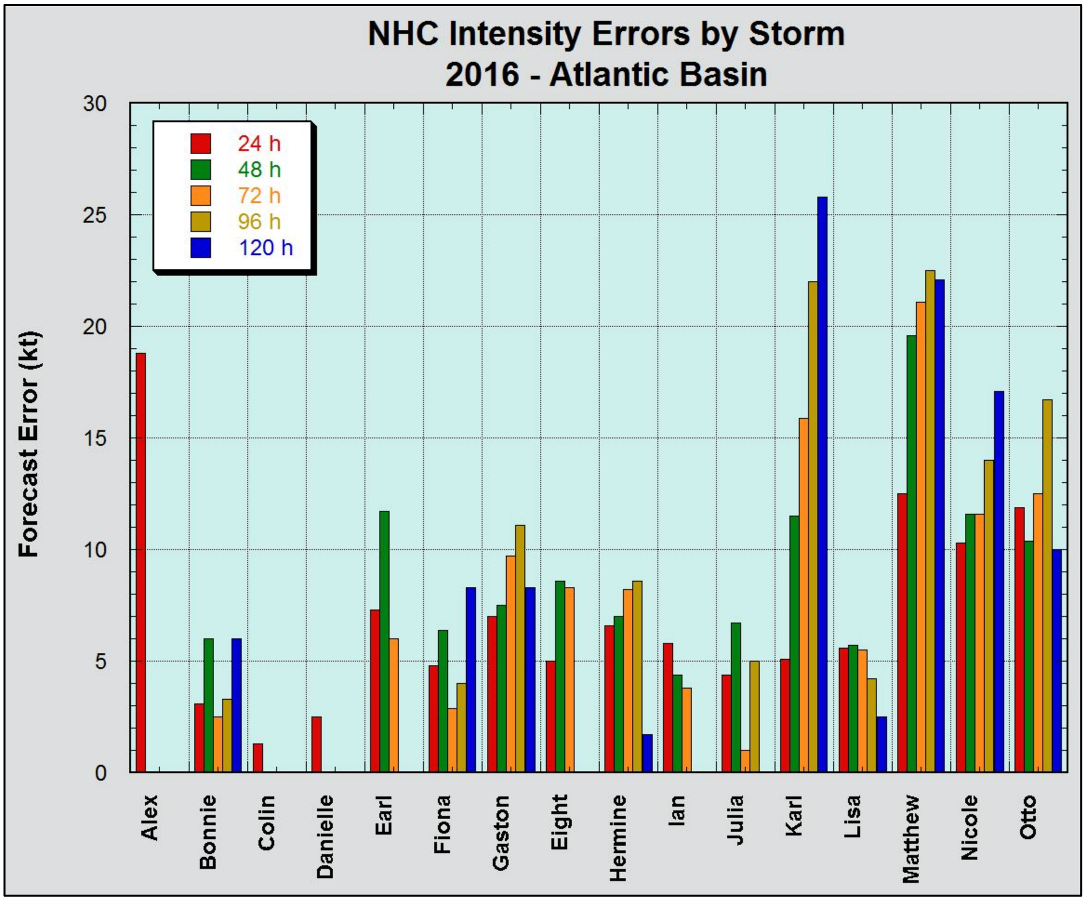


Figure 13. 2016 NHC official intensity errors by tropical cyclone.

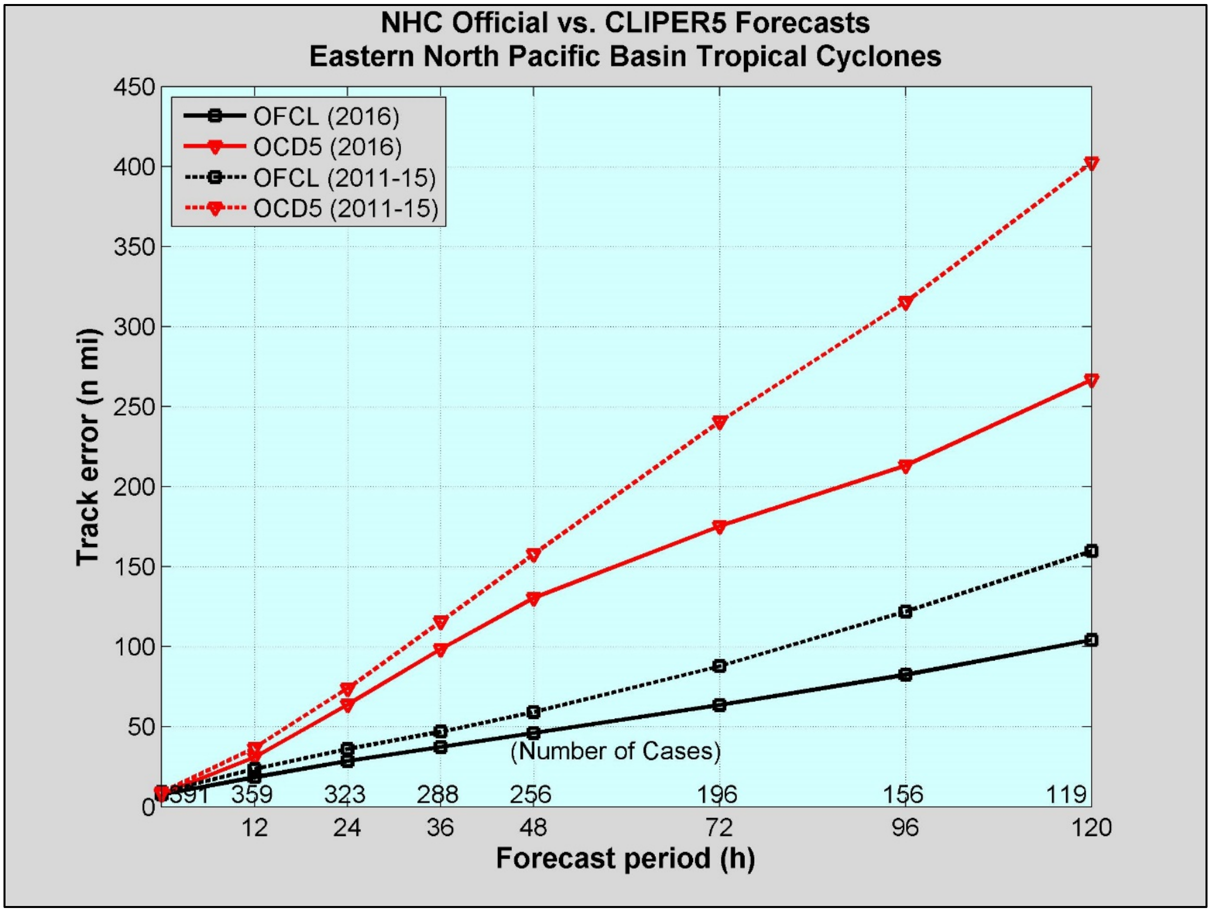


Figure 14. NHC official and CLIPER5 (OCD5) eastern North Pacific basin average track errors for 2016 (solid lines) and 2011-2015 (dashed lines).

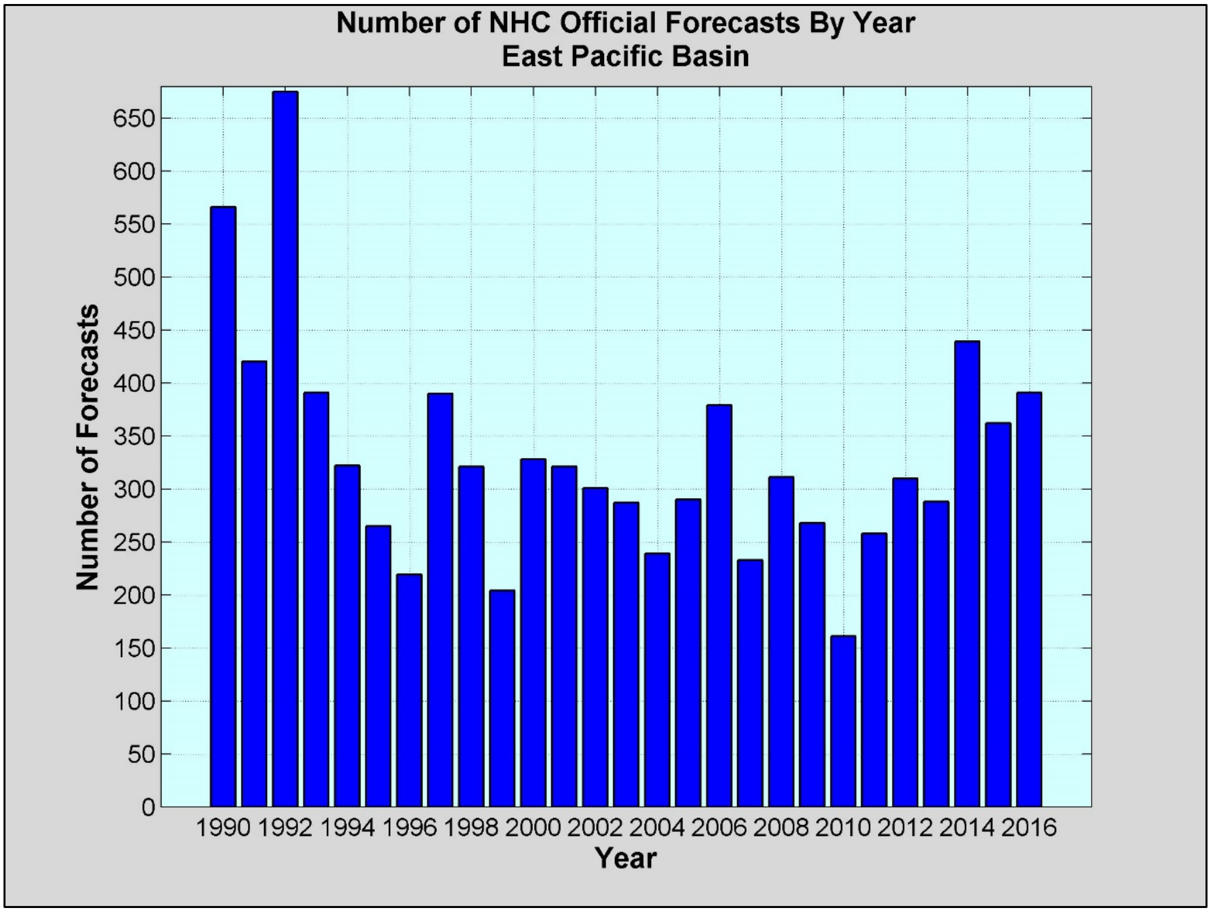


Figure 15. Number of NHC official forecasts for the Atlantic basin stratified by year.

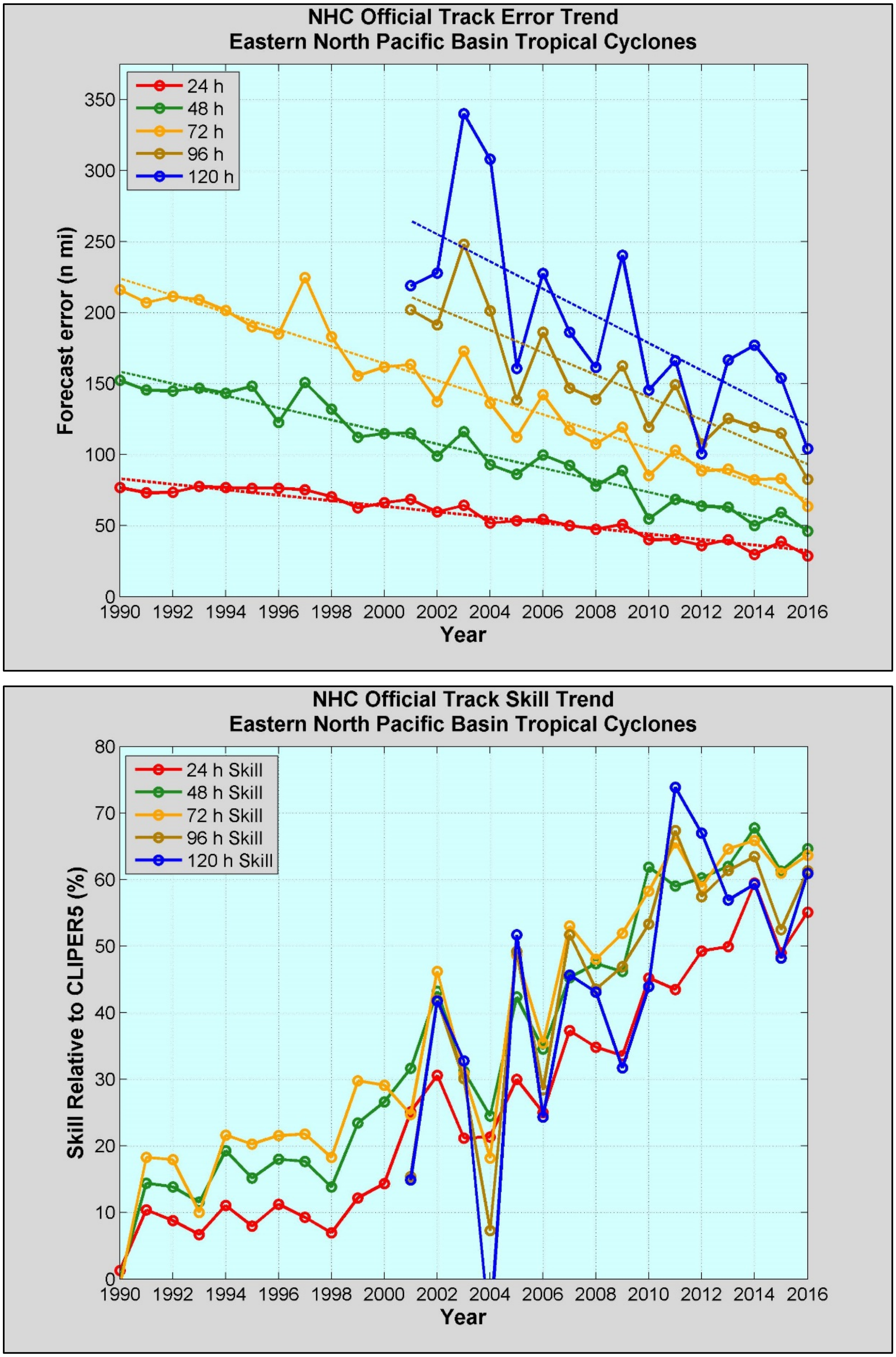


Figure 16. Recent trends in NHC official track forecast error (top) and skill (bottom) for the eastern North Pacific basin.

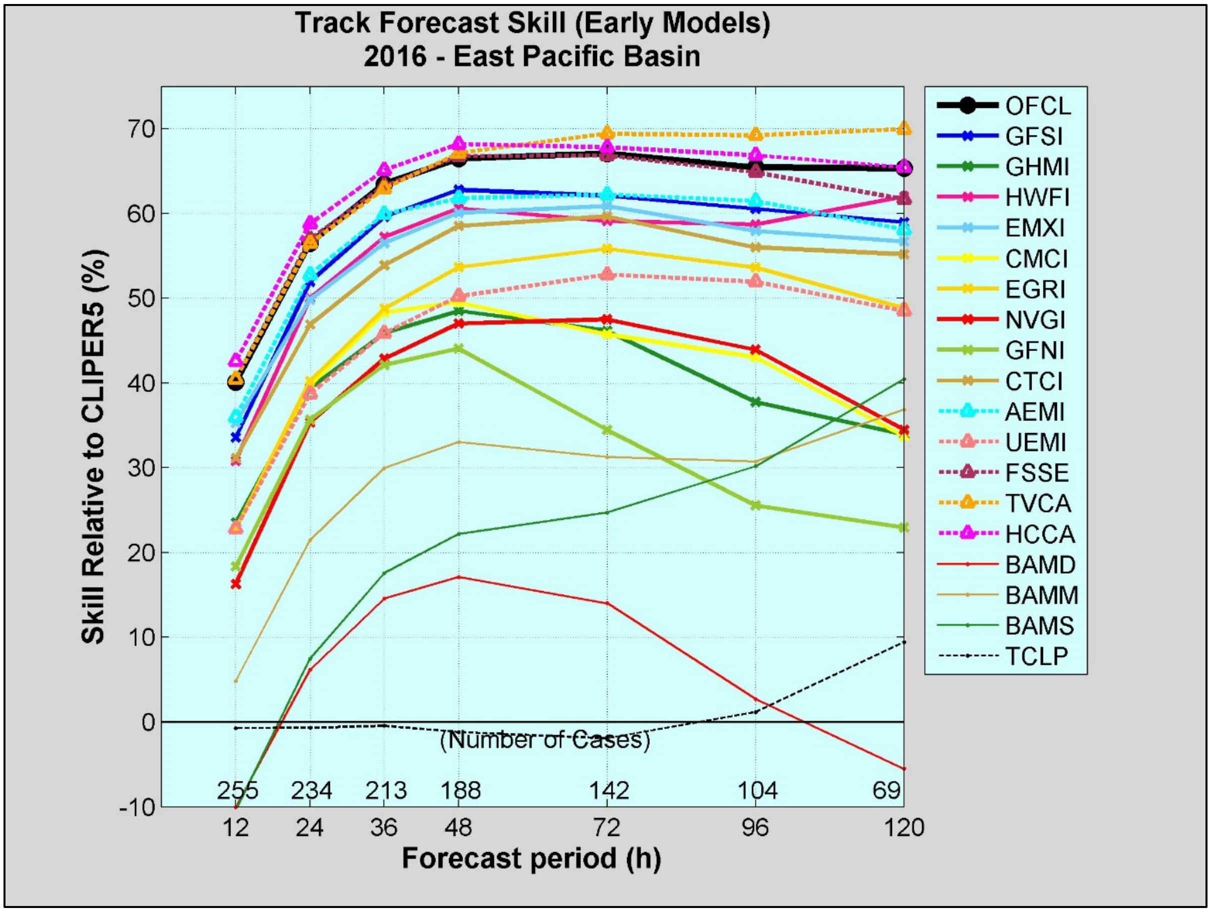


Figure 17. Homogenous comparison for selected eastern North Pacific early track models for 2016. This verification includes only those models that were available at least 2/3 of the time (see text).

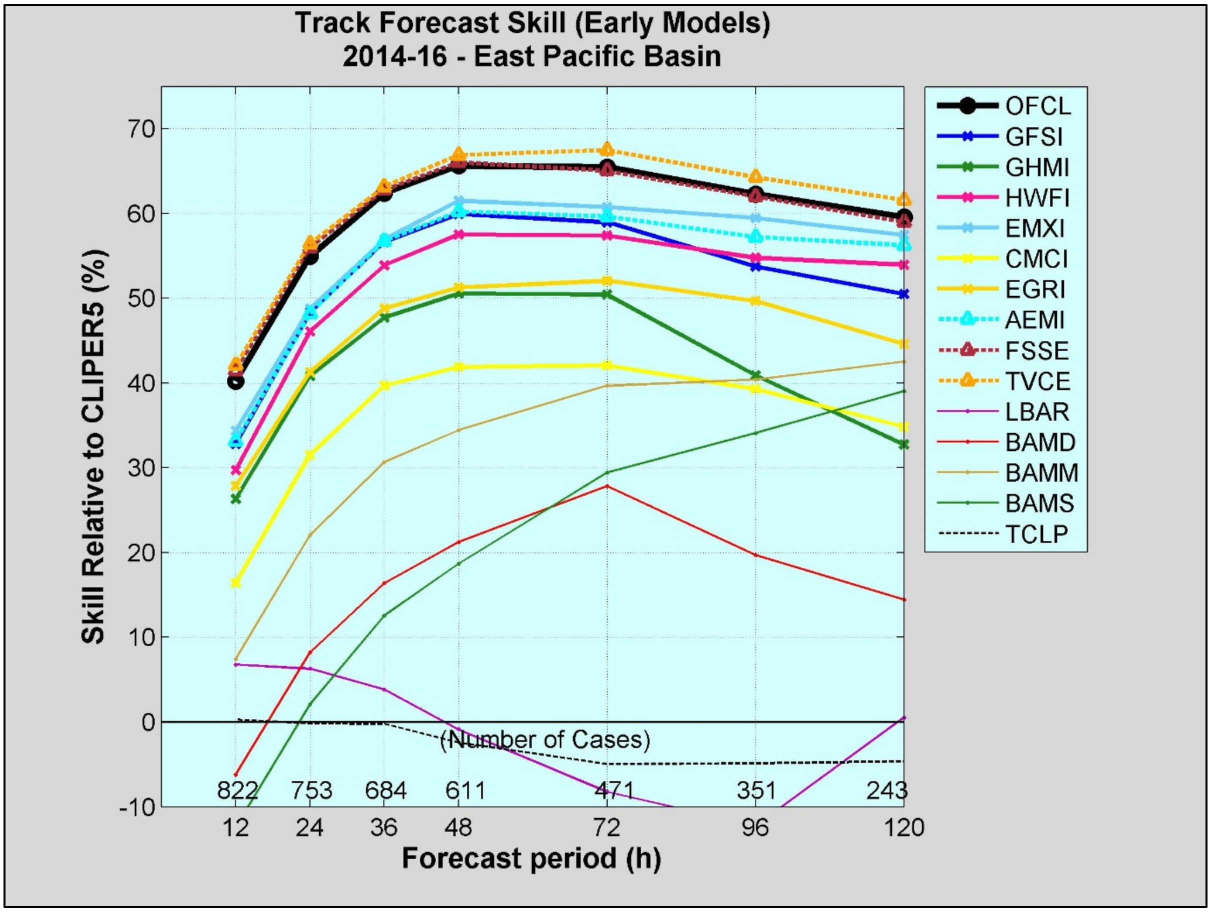


Figure 18. Homogenous comparison for selected eastern North Pacific basin early track models for 2014-2016.

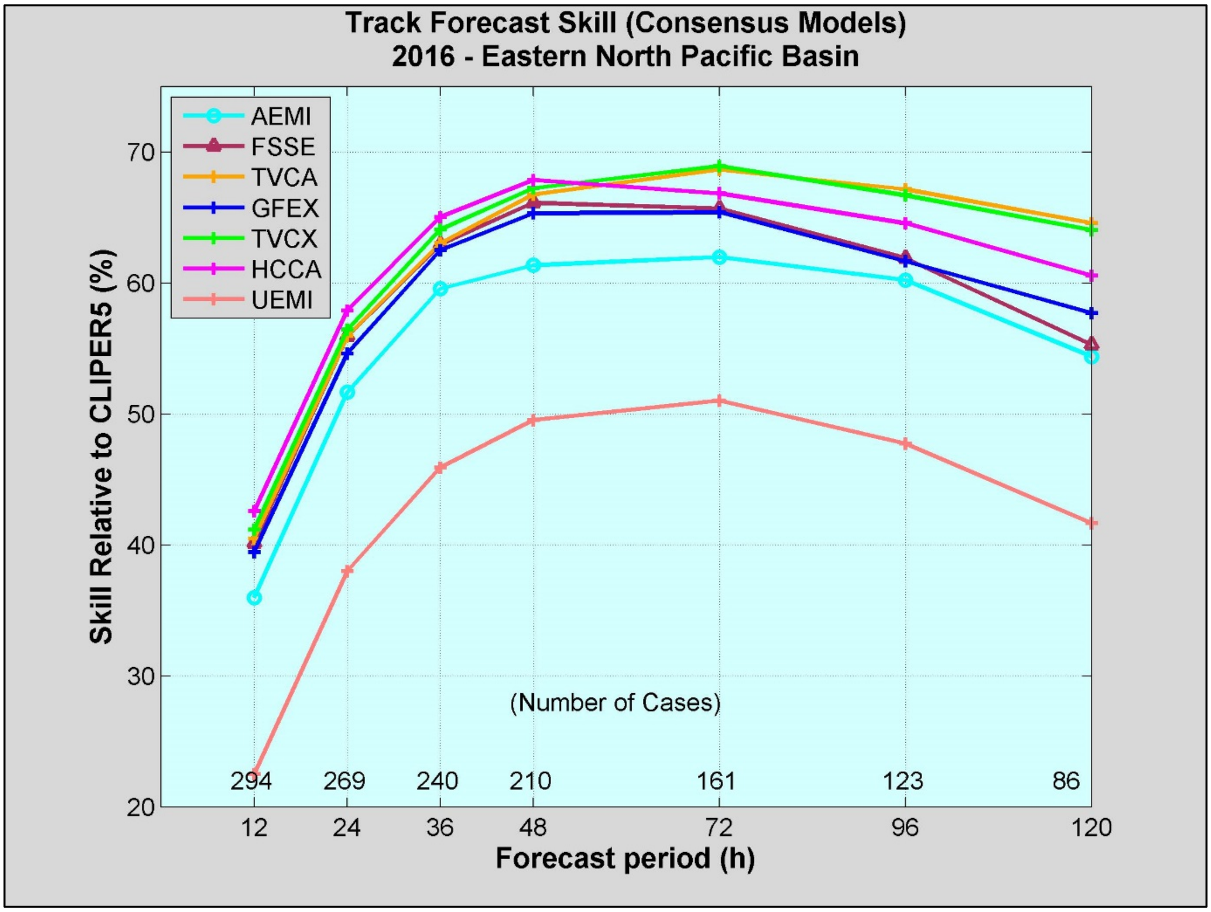


Figure 19. Homogenous comparison of the primary eastern North Pacific basin track consensus models for 2016.

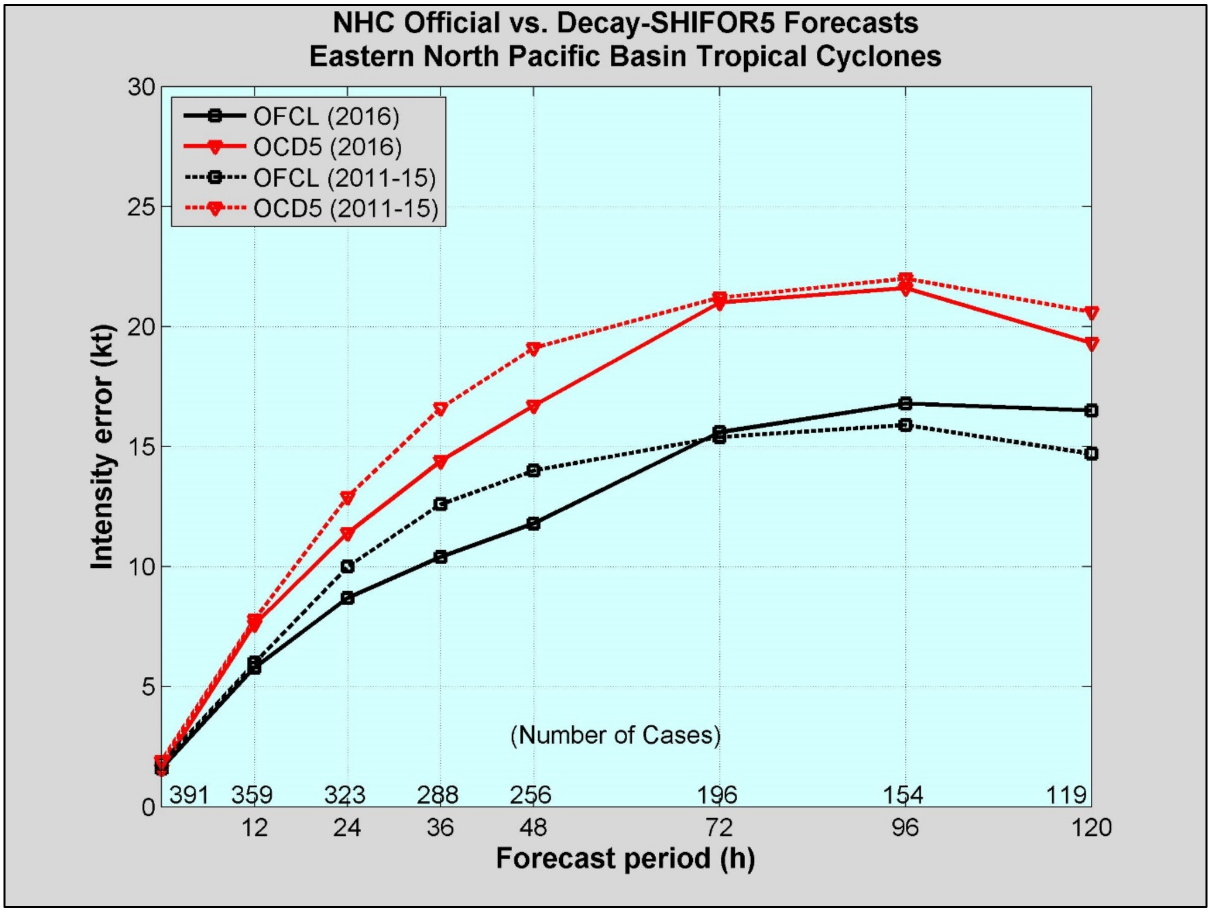


Figure 20. NHC official and Decay-SHIFOR5 (OCD5) eastern North Pacific basin average intensity errors for 2016 (solid lines) and 2011-2015 (dashed lines).

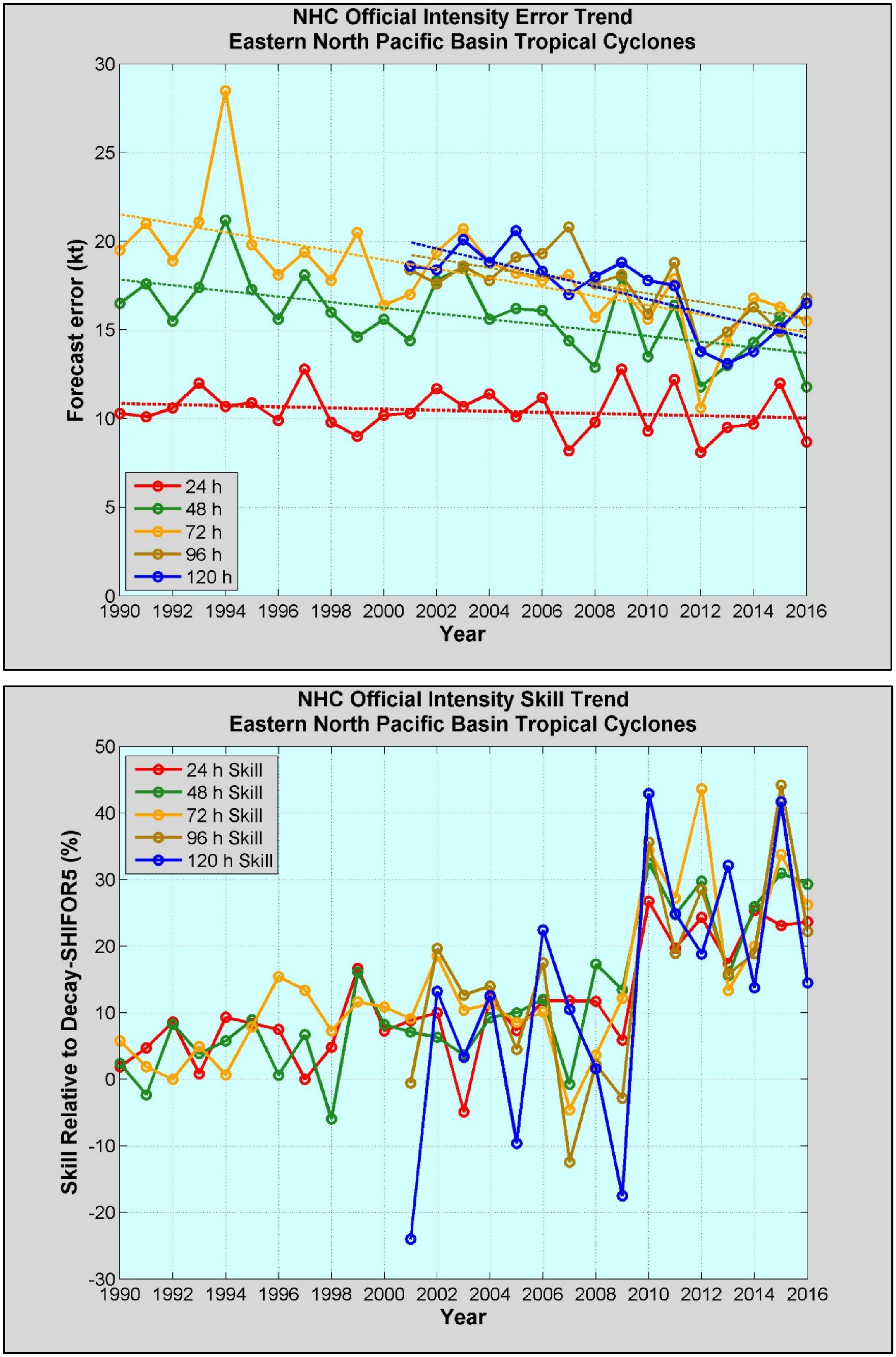


Figure 21. Recent trends in NHC official intensity forecast error (top) and skill (bottom) for the eastern North Pacific basin.

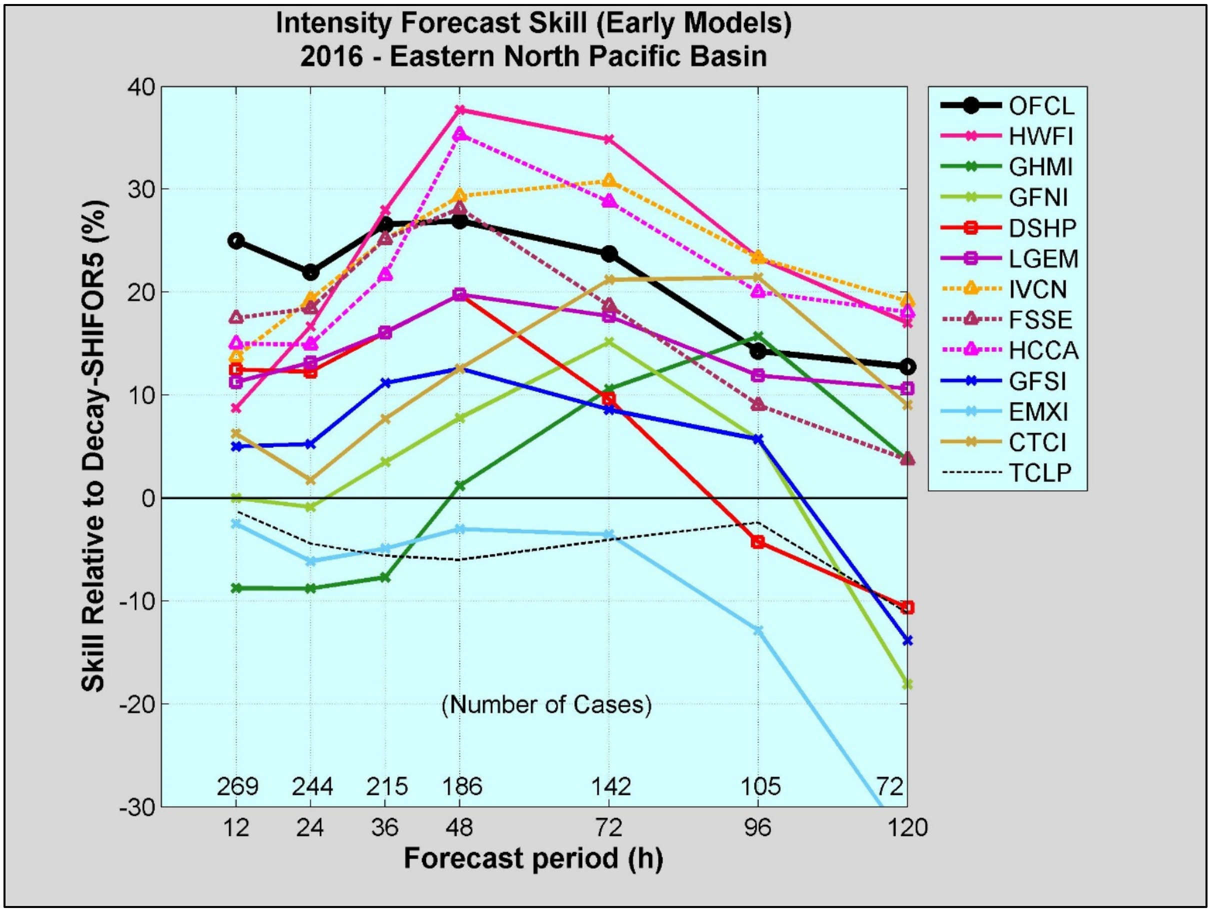


Figure 22. Homogenous comparison for selected eastern North Pacific basin early intensity guidance models for 2016.

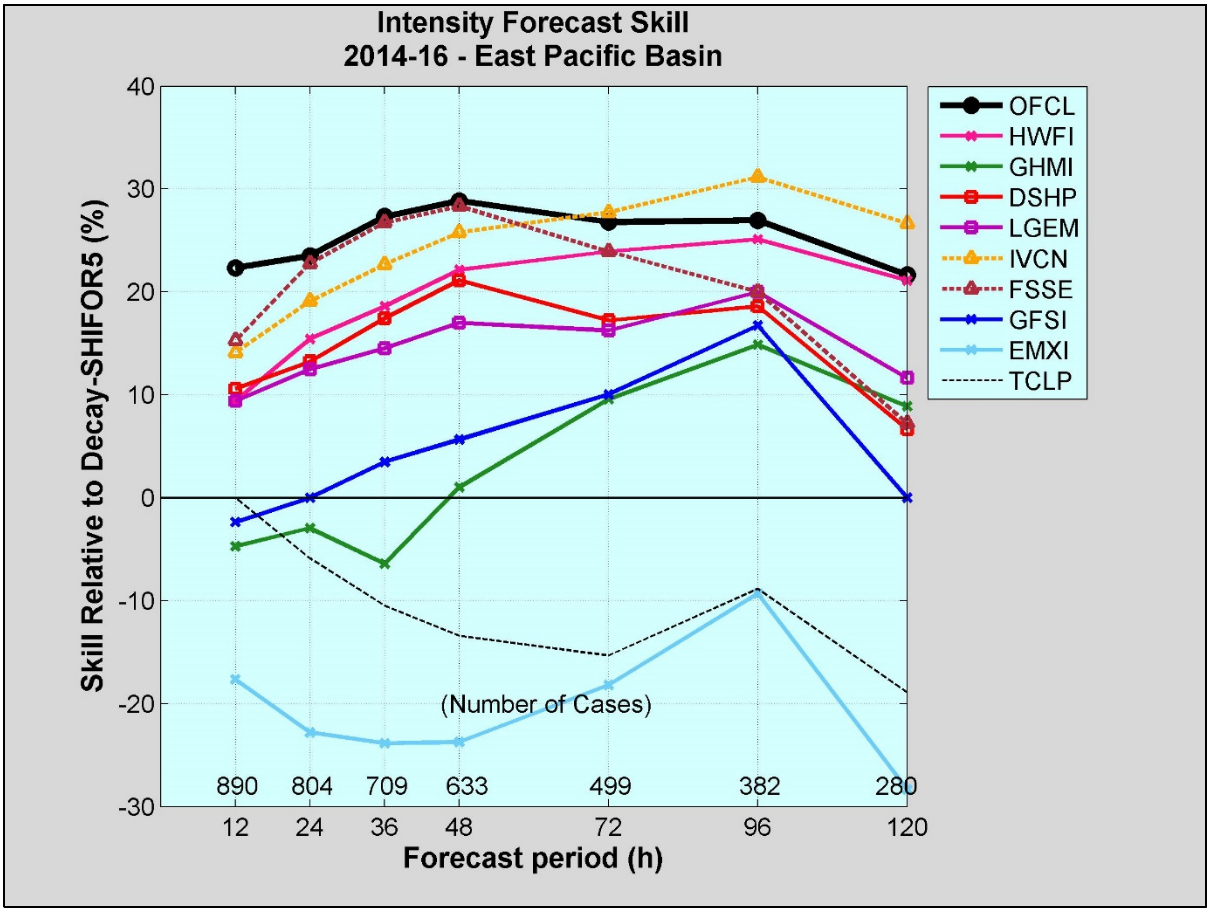
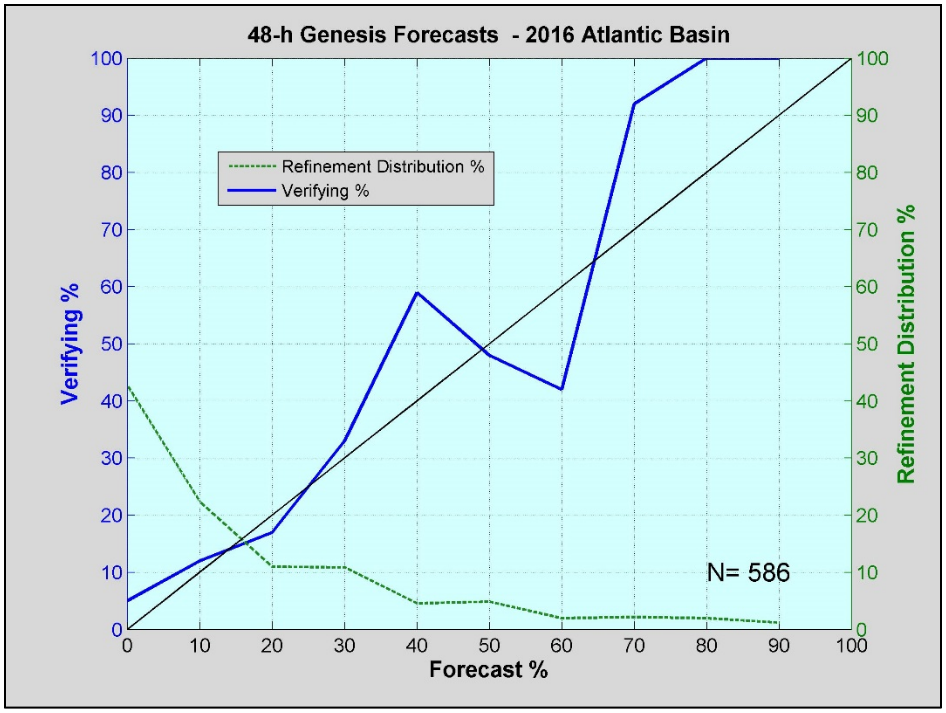


Figure 23. Homogenous comparison for selected eastern North Pacific basin early intensity guidance models for 2014-16.



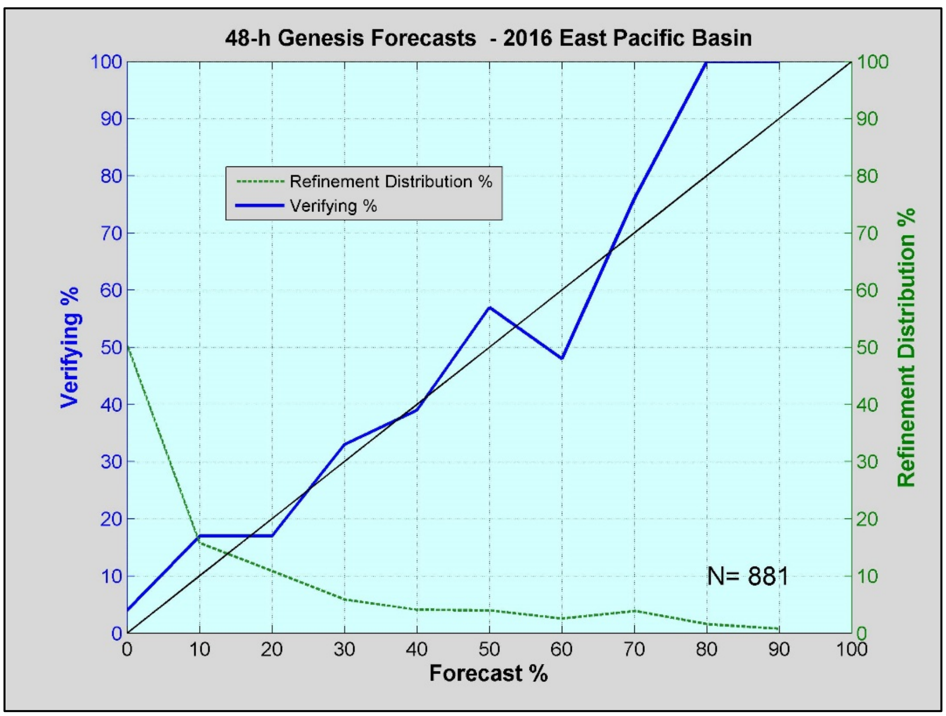
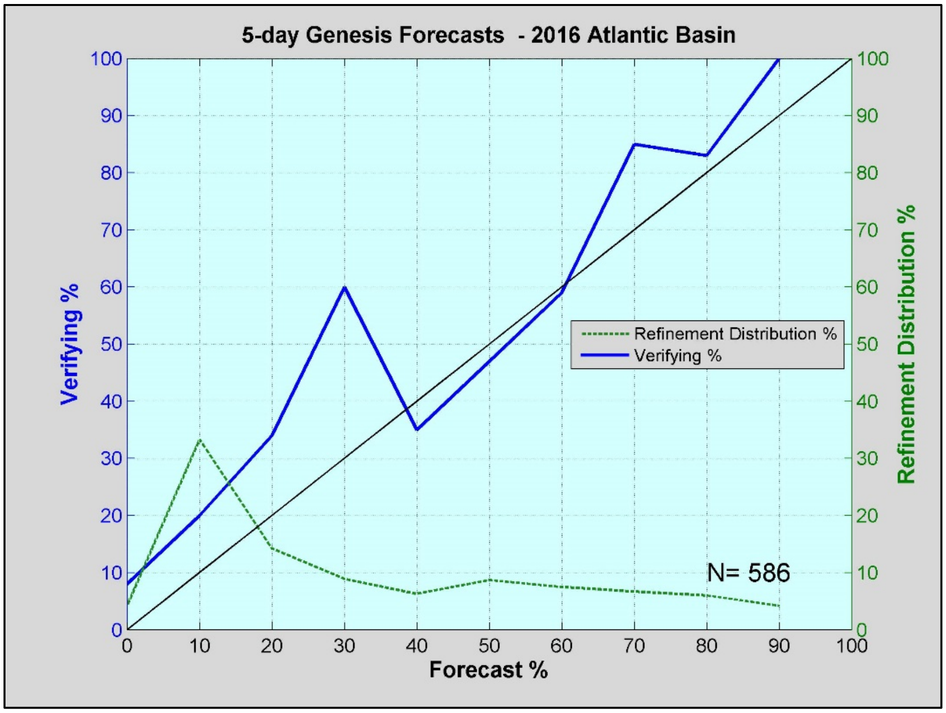


Figure 24. Reliability diagram for Atlantic (top) and eastern North Pacific (bottom) probabilistic tropical cyclogenesis 48-h forecasts for 2016. The solid lines indicate the relationship between the forecasts and verifying genesis percentages, with perfect reliability indicated by the thin diagonal black line. The dashed lines indicate how the forecasts were distributed among the possible forecast values.



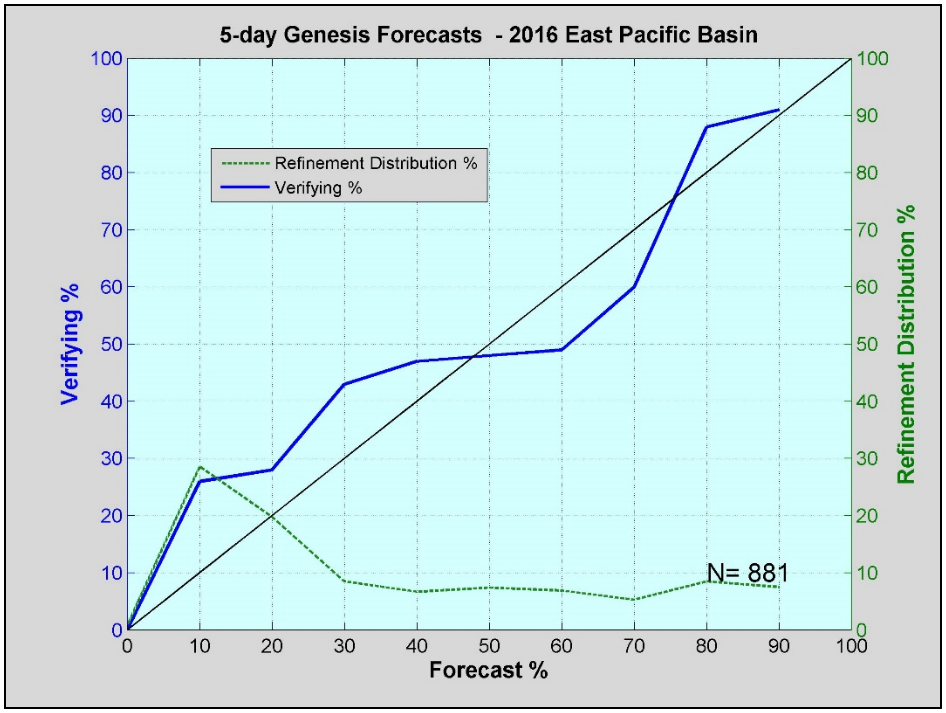


Figure 25. As described for Fig. 24, except for 120-h forecasts.