2012 National Hurricane Center Forecast Verification Report

John P. Cangialosi and James L. Franklin
NOAA/NWS/NCEP/National Hurricane Center

22 March 2013

ABSTRACT

The 2012 Atlantic hurricane season had above-normal activity, with 444 official forecasts issued. The mean NHC official track forecast errors in the Atlantic basin were lower than the previous 5-yr means at all times, and set records for accuracy at all forecast times except 120 h. The official track forecasts were very skillful and performed close to or better than the TVCA consensus model and the best-performing dynamical models. The FSSE had the highest skill and was the only guidance that consistently beat the official forecast. GFSI, AEMI, and EXMI were very good performers, with the HWFI and EGRI making up the second tier. The NGXI was the poorest-performing major dynamical model, and the CMCI and GHMI had similar skill to NGXI at 96 and 120 h. The Government Performance and Results Act of 1993 (GPRA) track goal was met.

Mean official intensity errors for the Atlantic basin in 2012 were below the 5-yr means at all lead times. Decay-SHIFOR errors in 2012 were also lower than their 5-yr means at all forecast times, indicating the season's storms were easier to forecast than normal. The consensus models ICON and FSSE were the best performers, and were the only models that had skill throughout most or all of the forecast period. The HWFI was a poor performer and had no skill throughout the entire period. The GPRA intensity goal was met.

There were 310 official forecasts issued in the eastern North Pacific basin in 2012, although only 39 of these verified at 120 h. This level of forecast activity was near normal. NHC official track forecast errors set a new record for accuracy at the 12-, 24 -, 48-, 96-, and 120-h forecast times, and track forecast skill was at or near all-time highs. The official forecast outperformed all of the guidance except for TVCE, which beat the official forecast at the 12-, 72-, and 96-h periods. Among the guidance models with sufficient availability, EMXI was the best individual model, and GFSI and HWFI performed fairly well. The skill of FSSE was close to that of TVCE, but it trailed TVCE by 5-10 % at 96 and 120 h.

For intensity, the official forecast errors in the eastern North Pacific basin were lower than the 5-yr means at all times. Decay-SHIFOR errors in 2012 were slightly lower than their 5-yr means at all forecast times, indicating the season's storms were a little easier to forecast than normal. The official forecasts, in general, performed as well as or better than all of the eastern Pacific guidance throughout the forecast period. The ICON and DSHP were the best performers from 12 to 72 h. The LGEM was the best individual model and beat the official forecast at 96 and 120 h. HWFI struggled late in the forecast period and was the worst performer at the longer forecast times.

Quantitative probabilistic forecasts of tropical cyclogenesis (i.e., the likelihood of tropical cyclone formation from a particular disturbance within 48 h) were made public for the first time in 2010. Forecasts were expressed in 10% increments and in terms of categories ("low", "medium", or "high"). Results from 2012 indicate that these probabilistic forecasts had a low (under-forecast) bias in the Atlantic basin. An under-forecast bias was also present in the eastern North Pacific basin at the middle probabilities with an over-forecast (high) bias at the high probabilities.

The Hurricane Forecast Improvement Project (HFIP) and the National Hurricane Center agreed in 2009 to establish a pathway to operations known as "Stream 1.5". The performance of the Stream 1.5 models in 2012 was generally poor. However, the FM9I was competitive with the top-tier dynamical models for track, and SPC3 performed better than much of the intensity guidance.

Table of Contents

1. Introduction **4**

2. Atlantic Basin **11**

a. 2012 season overview – Track *11*
b. 2012 season overview – Intensity *14*
c. Verifications for individual storms *16*

3. Eastern North Pacific Basin **17**

a. 2012 season overview – Track *17*
b. 2012 season overview – Intensity *18*
c. Verifications for individual storms *19*

4. Genesis Forecasts **19**

5. HFIP Stream 1.5 Activities **20**

6. Looking Ahead to 2013 **23**

a. Track Forecast Cone Sizes *23*
b. Consensus Models *23*

7. References **25**

List of Tables **26**

List of Figures **55**

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

---

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.

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 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

[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.

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

[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.

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 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 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 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]

---

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".

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.

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 29 January 2013 for the eastern North Pacific basin, and on 5 February 2013 for the Atlantic

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

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, which began experimentally in 2007 and became operational in 2010. Section 5 discusses the Hurricane Forecast Improvement Project (HFIP) Stream 1.5 activities in 2012. Section 6 summarizes the key findings of the 2012 verification and previews anticipated changes for 2013.

2. Atlantic Basin

a. 2012 season overview – Track

Figure 1 and Table 2 present the results of the NHC official track forecast verification for the 2012 season, along with results averaged for the previous 5-yr period, 2007-2011. In 2012, the NHC issued 444 Atlantic basin tropical cyclone forecasts[11], a number well above the average over the previous 5 yr (302). Mean track errors ranged from 25 n mi at 12 h to 194 n mi at 120 h. It is seen that mean official track forecast errors in 2012 were smaller than the previous 5-yr mean at all forecast times, even though the season's storms were harder than average to forecast. The official track forecast errors also set records for accuracy at all forecast times except 120 h. Over the past 15-20 yr, 24−72-h track forecast errors have been reduced by about 60% (Fig. 2). Track forecast error reductions of about 50% have occurred over the past 10 yr for the 96- and 120-h forecast periods. The official track forecast vector biases were small and generally westward through 72 h (i.e., the official forecast tended to fall to the west of the verifying position), and northeastward at 96 and 120 h. An examination of the track errors shows

[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.

that the biases were primarily along-track and slow, but there was a slight cross-track bias as well. Track forecast skill in 2012 ranged from 50% at 12 h to 72% at 48 h (Table 2).

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 2012. 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. Vector biases of the guidance models are given in Table 3b. The table shows that the official forecast had similar biases to TVCA, but the biases were generally smaller than most of the model guidance. Among the typically high-performing models, the EMXI had a slight southwestward bias, except at 120 h when it was northwestward, and GFSI had a pronounced northeastward bias at 96 and 120 h.

[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.

The performance of the official forecast and the early track models in terms of skill are presented in Fig. 3. The figure shows that official forecast was highly skillful, and even slightly better than TVCA, the primary Atlantic basin consensus aid. The only model that consistently beat the official forecast was FSSE, which had the highest skill of any model at all forecast times. The best-performing individual dynamical model in 2012 was GFSI, followed by EMXI. The HWFI and EGRI made up the second tier of the three-dimensional dynamical models; while NGXI[13], GHMI, and CMCI performed less well. The more simplistic BAMM was a relatively good performer in the 72 to 120 h forecast period, and beat the second tier of the three-dimensional models at those times. An evaluation over the three years 2010-12 (Fig. 4) indicates that FSSE, TVCN, EMXI, and GFSI are the best-performing models and have about equivalent skill from 12 to 72 h. At the longer leads, EMXI is the most skillful. The official forecasts are as good as or better than the best-performing models.

A separate homogeneous verification of the primary consensus models for 2012 is shown in Fig. 5. The figure shows that the skill of FSSE was superior to TVCA and the GFS ensemble mean (AEMI) by about 5 % at all forecast times. The skill of AEMI was only slightly worse than that of its respective deterministic model GFSI (Fig. 3), and represents an improvement in performance compared to the previous years. An examination of the verification of AEMI over the past few years (not shown) indicates that the ensemble mean has become increasingly skillful in the Atlantic basin, and it is quite competitive with the deterministic run and even slightly more skillful than GFSI at the longer forecast times.

[13] Communication problems prevented transmission of NGPS to NHC in 2012. NGXI is computed at NHC using NGPS fields. Historically the performance of NGXI is very similar to NGPI.

Atlantic basin 48-h official track error, evaluated for all tropical cyclones[14], is a forecast measure tracked under the Government Performance and Results Act of 1993 (GPRA). In 2012, the GPRA goal was 84 n mi and the verification for this measure was 68.8 n mi.

b. 2012 season overview – Intensity

Figure 6 and Table 4 present the results of the NHC official intensity forecast verification for the 2012 season, along with results averaged for the preceding 5-yr period. Mean forecast errors in 2012 ranged from about 5 kt at 12 h to about 13 kt at 72 and 120 h. These errors were below the 5-yr means at all forecast times, and the official forecasts had little bias in 2012. Decay-SHIFOR5 errors were well below their 5-yr means at all forecast times, however, indicating the season's storms were significantly easier than normal to forecast. Figure 7 shows that there has been a decrease in the intensity errors over the past few years; however, these recent improvements are likely due to a lack of rapidly intensifying hurricanes, which are typically the source of the large forecast errors. Over the long term there has been virtually no net change in error at the shorter leads, although forecasts during the current decade, on average, have been more skillful than those from the previous one. Comparison of Figs 7a and 7b suggests that the downward trend in the 96- and 120-h error does not represent an increase in forecast skill.

Table 5a presents a homogeneous verification for the official forecast and the primary early intensity models for 2012. Intensity biases are given in Table 5b, and forecast skill is presented in Fig. 8. The intensity models were not very skillful in 2012. The best performers were the consensus aids ICON/IVCN and FSSE, but even these

[14] Prior to 2010, the GPRA measure was evaluated for tropical storms and hurricanes only.

models only had marginal skill through the forecast period. The LGEM, typically one of the better individual models, lacked skill in 2012 and was one of the poorer performing models. HWFI was the worst model at the longer leads, and had skill near -60% at 120 h. The top-performing global models, GFSI and EMXI, were included in the intensity verification for completeness, although they are typically not considered by forecasters. EMXI was not skillful at any time, but still performed better than HWFI at 96 and 120 h. GFSI had some skill early and was better than much of the standard guidance from 12 to 36 h. Beyond that, however, the skill of GFSI decreased and was similar to GHMI, DSHP, and LGEM. An inspection of the intensity biases (Table 5b) indicated that the HWFI suffered from a high bias, but not to the degree that it had in 2011. The official forecast biases, in contrast, were generally small. An evaluation over the three years 2010-12 (Fig. 9) indicates that the consensus models have been superior to all of the individual models throughout the entire forecast period. However, a separate verification including only the pre-landfall cases reveals that DSHP and LGEM are slightly more skillful than the consensus models at the longer forecast times when land interactions are not involved.

The 48-h official intensity error, evaluated for all tropical cyclones, is another GPRA measure for the NHC. In 2012, the GPRA goal was 15 kt and the verification for this measure was 12.3 kt, with this year's success attributed mostly to low forecast difficulty. This was only the second time in five years that the intensity goal was met. The GPRA goal itself was established based on the assumption that the HWRF model would immediately lead to forecast improvements, which has not occurred. It is reasonable to assume that until there is some modeling or conceptual breakthrough, annual official intensity errors are mostly going to rise and fall with forecast difficulty, and therefore often fail to meet GPRA goals.

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 Debby, which were more than triple the long-term mean at 72 h. In the case of Debby, early track guidance indicated a dichotomy in the model forecast tracks, with almost as many model solutions taking Debby toward the Texas coast as solutions showing a northeastward track toward north Florida. Early official forecasts placed more weight on EMXI, which incorrectly predicted Debby to track toward Texas. Large track errors were also made for Hurricane Kirk during the 72- to 120-h forecast periods. An examination of the individual forecasts indicates that the first two forecasts called for a more westerly motion before recurvature than actually occurred, which caused large errors at the longer forecast times. On the other hand, track errors were very low for Beryl, Michael, and Sandy. Regarding the intensity forecasts, Kirk and Michael were the sources of the largest error. For both of these storms, the official forecast did not correctly anticipate the rapid intensification periods. Additional discussion on forecast performance for individual storms can be found in NHC Tropical Cyclone Reports available at http://www.nhc.noaa.gov/2012atlan.shtml.

3. Eastern North Pacific Basin

a. 2012 season overview – Track

The NHC track forecast verification for the 2012 season in the eastern North Pacific, along with results averaged for the previous 5-yr period is presented in Figure 10 and Table 7. There were 310 forecasts issued for the eastern Pacific basin in 2012, although only 39 of these verified at 120 h. This level of forecast activity was about average. Mean track errors ranged from 23 n mi at 12 h to 108 n mi at 96 h, and were unanimously lower than the 5-yr means. New records were set for forecast accuracy at the 12-, 24-, 48-, 96-, and 120-h forecast times. CLIPER5 errors were similar to their long-term means from 12 to 48 h, but below those values beyond 48 h. A small westward or west-northwestward track bias in the official forecasts was noted from 12 to 96 h, with a moderate northeastward bias present at 120 h.

Figure 11 shows recent trends in track forecast accuracy and skill for the eastern North Pacific. Errors have been reduced by roughly 45-60% for the 24 to 72 h forecasts since 1990, a somewhat smaller but still substantial improvement relative to what has occurred in the Atlantic. Forecast skill in 2012 set a new record high at 24 h and was near all-time highs at the remaining forecast times.

Table 8a presents a homogeneous verification for the official forecast and the early track models for 2012, with vector biases of the guidance models given in Table 8b. Skill comparisons of selected models are shown in Fig. 12. Note that the sample becomes rather small by 120 h (only 15 cases). FSSE was eliminated from this evaluation because that model did not meet the two-thirds availability threshold. The official forecast outperformed all of the guidance except for TVCE, which beat the official forecast by a small margin at the 12-, 72-, and 96-h forecast times. EMXI was the best individual model at all times, but it had about 5% less skill than TVCE and the official forecast. GSFI, HWFI, and AEMI made up the second tier of models, with GHMI and EGRI not far behind. NGXI was a poor performer and had similar skill to the simple BAMM and BAMD models.

A separate verification of the primary consensus aids is given in Figure 13. TVCE and FSSE had comparable skill from 12 to 72 h, but FSSE trailed TVCE at the longer forecast times. The skill of AEMI was noticeably smaller than that of FSSE and TVCE, and it was near zero at 120 h.

b. 2012 season overview – Intensity

Figure 14 and Table 9 present the results of the NHC eastern North Pacific intensity forecast verification for the 2012 season, along with results averaged for the preceding 5-yr period. Mean forecast errors were 5 kt at 12 h and increased to 14 kt at 96 and 120 h. The errors were lower than the 5-yr means, by up to 38%, at all times. The Decay-SHIFOR5 forecast errors were also lower than their 5-yr means (by up to 18%); this implies that forecast difficulty in 2012 was lower than normal. A review of error and skill trends (Fig. 15) indicates that the intensity errors have decreased slightly over the past 15-20 yr at 48 h and beyond. Forecast skill had generally increased in 2012, and reached an all-time high at 72 h. Intensity forecast biases in 2012 were slightly negative throughout the forecast period.

Figure 16 and Table 10a present a homogeneous verification for the primary early intensity models for 2012. Forecast biases are given in Table 10b. The official forecasts were more skillful than all of the models except for LGEM, which performed slightly better than the official forecasts at 96 and 120 h, and met or exceeded the skill of ICON at all times. DSHP and ICON were the best models through 72 h, and LGEM was the best aid at 96 and 120 h. HWFI had some skill early, but its performance was worse than Decay-SHIFOR5 at 96 and 120 h. GHMI performed slightly better than HWFI at the longer forecast times, but it lacked skill from 12 to 36 h. The performance of the global models for intensity prediction was poor. EMXI had skill between -20 and -35% throughout the forecast period. The GFSI errors were very close to the errors of Decay-SHIFOR5 from 12 to 120 h.

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/2012epac.shtml.

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 during the following 48 hours. In 2007, the NHC began producing in-house (non-public) experimental probabilistic tropical cyclone genesis forecasts. 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 TC formation during the 48 h period following the nominal TWO issuance time. These probabilities became available to the public in 2010. 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 for the Atlantic and eastern North Pacific basins for 2012 are given in Table 12 and illustrated in Fig. 17. In the Atlantic basin, a total of 397 genesis forecasts were made. These forecasts exhibited a slight under-forecast (low) bias in 2012, and were not as reliable as 2011, when little bias was present. In the eastern Pacific, the forecasts were reliable at the lower probabilities, but an under-forecast bias existed in the middle probabilities and an over-forecast (high) bias was present at the high probabilities. Another way to interpret this result is that once the forecast likelihood exceeded 40%, there was minimal correlation between the forecast and actual verifying rates. 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. HFIP Stream 1.5 Activities

The Hurricane Forecast Improvement Project (HFIP) and the National Hurricane Center agreed in 2009 to establish a pathway to operations known as "Stream 1.5". Stream 1.5 covers improved models and/or techniques that the NHC, based on prior assessments, wants to access in real-time during a particular hurricane season, but which cannot be made available to NHC by the operational modeling centers in conventional production mode. HFIP's Stream 1.5 supports activities that intend to bypass operational limitations by using non-operational resources to move forward the delivery of guidance to NHC by one or more hurricane seasons. Stream 1.5 projects are run as part of HFIP's annual summertime "Demo Project".

Eight models/modeling systems were provided to NHC in 2012 under Stream 1.5; these are listed in Table 13. Note that most models were admitted into Stream 1.5 based on the models' performance forecasting either track or intensity, but generally not both. For example, forecasters were instructed to consult the COTI intensity forecasts but not the COTI track forecasts. Two HFIP Stream 1.5 consensus aids were constructed: the track consensus TV15 comprised the operational models GFSI, EGRI, GHMI, HWFI, GFNI[15], EMXI and the Stream 1.5 models AHWI, APSI, and FM9I, while the intensity consensus IV15 comprised the operational models DSHP, LGEM, GHMI, HWFI and the Stream 1.5 models AHWI, COTI, APSI, and UWNI.

Figure 18 presents a homogeneous verification of the primary operational models against the AHWI Stream 1.5 track model (top) and a homogenous verification that includes the FM9I (bottom), which had limited availability. The figure shows that in 2012 the AHWI was not competitive with the top-tier dynamical models, and in fact, had skill that was comparable to the rather poor-performing NGXI and CMCI. Conversely, for a smaller sample FM9I was competitive with the top-tier operational models, with skill similar to or higher than EMXI. Figure 19 shows that there was very little impact from adding the Stream 1.5 models to the track consensus through 48 h, and then a slight negative effect from 72 to 120 h.

[15] GFNI is formally part of the Stream 1.5 TV15 consensus and TVCA, but it was unavailable in 2012.

Figure 20 presents the track and intensity forecast skill of GHMI and the Stream 1.5 GFDL ensemble mean (GPMI) and an unbogused GFDL ensemble member (G01I). G01I performed better than GHMI and GPMI for track at all times except 120 h, on the order of about 5% more skill, than the operation GFDL and its ensemble mean for track in 2012. Regarding intensity prediction, the GFDL ensemble mean was not consistently better than its deterministic run, and G01I performed worse than GHMI and GPMI at most times. It should be noted, than none of these models had any skill for intensity prediction throughout the forecast period.

Intensity results are shown in Fig. 21, for a sample that excludes the PSU Doppler runs due to limited availability. The Stream 1.5 models COTI and AHWI performed very poorly. These models had no skill throughout the forecast period and performed worse than all of the operational models. UWNI was a better intensity model in 2012, but its skill was still near the poor-performing HWFI at 96 and 120 h. The SPC3 was the best performing Stream 1.5 intensity model, and that result is not surprising, given that it represents an intelligent consensus of the already top-tier dynmamical-statistical models LGEM and DSHP. The impact of the Stream 1.5 models was slightly positive to the intensity consensus from 12 to 36 h, but noticeably negative at the longer forecast times (Fig. 22).

The performance of the Stream 1.5 models in 2012 was generally disappointing. The FM9I, however, for a smaller subset of cases did show about equivalent skill to the high performing operational models for track, and the dynamical-statistical consensus SPC3 did have more intensity skill than its individual components.

6. Looking Ahead to 2013

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 2013 for the Atlantic and eastern North Pacific basins (based on error distributions for 2008-12) are in Table 14. In the Atlantic basin, the cone circles will be slightly smaller than they were last year, with the biggest decrease at 72 h. In the eastern Pacific basin, the cone circles will be about 10 % smaller than they were last year at most forecast times.

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 2013 is given in Table 15. Several changes have been made to the consensus models because of the retirement of NOGAPS. Therefore, NGPI and GFNI were removed for all consensus compositions. The Navy Global Environmental Model (NAVGEM) will be replacing NOGAPS in 2013, but this model will not be included in the consensus models until its performance for tropical cyclones is better understood. Of note, the GUNA consensus model has been retired.

Acknowledgments:

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

7. 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

1. National Hurricane Center forecasts and models.
2. Homogenous comparison of official and CLIPER5 track forecast errors in the Atlantic basin for the 2012 season for all tropical cyclones.
3. (a) Homogenous comparison of Atlantic basin early track guidance model errors (n mi) for 2012. (b) Homogenous comparison of Atlantic basin early track guidance model bias vectors (º/n mi) for 2012.
4. Homogenous comparison of official and Decay-SHIFOR5 intensity forecast errors in the Atlantic basin for the 2012 season for all tropical cyclones.
5. (a) Homogenous comparison of Atlantic basin early intensity guidance model errors (kt) for 2012. (b) Homogenous comparison of a selected subset of Atlantic basin early intensity guidance model errors (kt) for 2012. (c) Homogenous comparison of a selected subset of Atlantic basin early intensity guidance model biases (kt) for 2012.
6. Official Atlantic track and intensity forecast verifications (OFCL) for 2012 by storm.
7. Homogenous comparison of official and CLIPER5 track forecast errors in the eastern North Pacific basin for the 2012 season for all tropical cyclones.
8. (a) Homogenous comparison of eastern North Pacific basin early track guidance model errors (n mi) for 2012. (b) Homogenous comparison of eastern North Pacific basin early track guidance model bias vectors (º/n mi) for 2012.
9. Homogenous comparison of official and Decay-SHIFOR5 intensity forecast errors in the eastern North Pacific basin for the 2012 season for all tropical cyclones.
10. (a) Homogenous comparison of eastern North Pacific basin early intensity guidance model errors (kt) for 2012. (b) Homogenous comparison of eastern North Pacific basin early intensity guidance model biases (kt) for 2012.
11. Official eastern North Pacific track and intensity forecast verifications (OFCL) for 2012 by storm.
12. Verification of experimental in-house probabilistic genesis forecasts for (a) the Atlantic and (b) eastern North Pacific basins for 2012.
13. HFIP Stream 1.5 models for 2012.
14. NHC forecast cone circle radii (n mi) for 2013. Change from 2012 values in n mi and percent are given in parentheses.
15. Composition of NHC consensus models for 2013. 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.

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
NGPS	Navy Operational Global Prediction System	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
AFW1	Air Force MM5	Multi-layer regional dynamical	L	Trk, Int
EMX	ECMWF global model	Multi-layer global dynamical	L	Trk, Int
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

ID	Name/Description	Type	Timeliness (E/L)	Parameters forecast
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
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
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
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
NGXI	Previous cycle NGPS, 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, NGPI, and GFSI	Consensus	E	Trk
CGUN	Version of GUNA corrected for model biases	Corrected consensus	E	Trk

ID	Name/Description	Type	Timeliness (E/L)	Parameters forecast
AEMI	Previous cycle AEMN, adjusted	Consensus	E	Trk, Int
FSSE	FSU Super-ensemble	Corrected consensus	E	Trk, Int
TCON*	Average of GHMI, EGRI, NGPI, GFSI, and HWFI	Consensus	E	Trk
TCCN*	Version of TCON corrected for model biases	Corrected consensus	E	Trk
TVCN*	Average of at least two of GFSI EGRI NGPI GHMI HWFI GFNI EMXI	Consensus	E	Trk
TVCA*	Average of at least two of GFSI EGRI GHMI HWFI GFNI EMXI	Consensus	E	Trk
TVCE*	Average of at least two of GFSI EGRI NGPI GHMI HWFI GFNI EMXI	Consensus	E	Trk
TVCC*	Version of TVCN corrected for model biases	Corrected consensus	E	Trk
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

* The composition of the consensus aids can change from year to year; the table lists the composition used during the 2012 season.

Table 2. Homogenous comparison of official and CLIPER5 track forecast errors in the Atlantic basin for the 2012 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
2012 mean OFCL error (n mi)	24.6	39.7	53.6	68.8	100.6	142.8	194.4
2012 mean CLIPER5 error (n mi)	48.8	108.7	177.9	241.7	344.3	436.2	518.6
2012 mean OFCL skill relative to CLIPER5 (%)	49.6	63.5	69.9	71.5	70.8	67.3	62.5
2012 mean OFCL bias vector (°/n mi)	306/004	281/009	273/013	274/017	274/014	024/018	046/060
2012 number of cases	404	364	324	289	232	188	148
2007-2011 mean OFCL error (n mi)	30.4	48.4	65.9	83.1	124.4	166.5	213.4
2007-2011 mean CLIPER5 error (n mi)	46.9	95.2	151.7	211.6	316.8	404.3	485.2
2007-2011 mean OFCL skill relative to CLIPER5 (%)	35.2	49.2	56.6	60.7	60.7	58.8	56.0
2007-2011 mean OFCL bias vector (°/n mi)	328/003	326/006	321/008	325/010	301/008	020/007	030/019
2007-2011 number of cases	1347	1181	1027	896	706	543	422
2012 OFCL error relative to 2007-2011 mean (%)	-19.1	-18.0	-18.7	-17.2	-19.1	-14.2	-8.9
2012 CLIPER5 error relative to 2007-2011 mean (%)	4.1	14.2	17.3	14.2	8.7	7.9	6.9

Table 3a. Homogenous comparison of Atlantic basin early track guidance model errors (n mi) for 2012. 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	23.4	37.8	50.0	64.0	94.2	139.2	182.4
OCD5	46.7	104.2	172.2	234.1	342.7	419.8	504.7
GFSI	**22.9**	**37.3**	50.9	**63.1**	95.4	157.6	209.4
GHMI	30.1	50.0	69.6	90.3	141.7	221.2	329.4
HWFI	27.7	49.6	67.9	85.5	130.2	184.7	260.0
NGXI	34.2	64.6	90.7	112.0	170.4	220.5	315.1
EGRI	30.3	49.7	71.2	94.6	138.2	192.7	251.3
EMXI	**22.5**	38.3	53.4	68.5	115.0	155.0	199.1
CMCI	31.5	52.5	69.0	87.8	144.8	224.3	334.8
AEMI	23.8	40.3	53.7	65.3	96.7	155.0	227.5
FSSE	**20.3**	**32.2**	**44.7**	**55.4**	**82.5**	**128.4**	**172.9**
TVCA	**22.8**	**37.7**	50.9	**63.8**	94.7	147.7	212.0
LBAR	39.5	74.8	114.4	164.7	292.3	424.2	446.9
BAMD	46.7	82.6	118.3	150.2	217.1	351.8	484.8
BAMM	36.2	61.4	87.1	102.1	122.5	180.1	212.7
BAMS	49.3	90.4	127.2	150.5	177.7	214.2	255.5
TCLP	37.6	88.6	150.6	211.5	332.0	435.1	542.9
# Cases	277	251	230	208	167	122	88

Table 3b. Homogenous comparison of Atlantic basin early track guidance model bias vectors (º/n mi) for 2012.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	304/004	297/009	290/013	285/015	323/012	023/034	015/066
OCD5	299/005	334/015	355/031	010/053	011/104	016/168	005/191
GFSI	324/005	315/007	319/007	316/002	068/018	064/058	056/103
GHMI	268/005	297/012	310/021	314/029	350/052	001/083	353/165
HWFI	304/007	299/014	291/020	286/025	305/022	025/034	028/089
NGXI	296/012	293/023	291/032	286/043	301/058	335/094	345/177
EGRI	280/006	266/012	265/018	258/027	281/027	359/045	352/117
EMXI	239/002	236/006	222/012	216/020	214/034	227/002	314/036
CMCI	291/010	291/018	299/020	313/023	352/041	015/087	011/142
AEMI	298/005	286/009	284/008	273/004	060/017	055/053	042/103
FSSE	246/001	207/004	196/010	191/016	179/016	051/018	005/053
TVCA	288/005	288/009	251/014	275/017	310/015	019/039	008/089
LBAR	070/014	055/021	065/026	079/048	088/136	083/272	087/313
BAMD	054/022	049/045	047/060	048/069	060/112	063/233	064/342
BAMM	303/003	329/007	321/009	303/013	279/015	043/025	060/057
BAMS	283/022	275/040	264/053	260/059	240/069	187/050	110/096
TCLP	250/009	274/019	285/032	303/044	321/068	341/100	324/150
# Cases	277	251	230	208	167	122	88

Table 4. Homogenous comparison of official and Decay-SHIFOR5 intensity forecast errors in the Atlantic basin for the 2012 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
2012 mean OFCL error (kt)	5.4	8.0	10.2	12.3	13.1	11.8	12.7
2012 mean Decay-SHIFOR5 error (kt)	6.6	9.4	10.9	11.8	13.1	12.2	12.4
2012 mean OFCL skill relative to Decay-SHIFOR5 (%)	18.2	14.9	6.4	-4.2	0.0	3.3	-2.4
2012 OFCL bias (kt)	-1.2	-1.1	-0.4	0.7	1.6	1.7	2.4
2012 number of cases	404	364	324	289	232	188	148
2007-11 mean OFCL error (kt)	7.1	10.8	13.0	15.0	16.9	17.1	18.1
2007-11 mean Decay-SHIFOR5 error (kt)	8.4	12.4	15.4	17.7	20.5	21.5	21.2
2007-11 mean OFCL skill relative to Decay-SHIFOR5 (%)	15.5	12.9	15.6	15.3	17.6	20.5	14.6
2007-11 OFCL bias (kt)	0.0	0.7	1.0	1.5	1.7	0.9	0.6
2007-11 number of cases	1347	1181	1027	896	706	543	422
2012 OFCL error relative to 2007-11 mean (%)	-23.9	-25.9	-21.5	-18.0	-22.5	-31.0	-29.8
2012 Decay-SHIFOR5 error relative to 2007-11 mean (%)	-21.4	-24.2	-29.2	-33.3	-36.1	-43.3	-41.5

Table 5a. Homogenous comparison of selected Atlantic basin early intensity guidance model errors (kt) for 2012. Errors smaller than the NHC official forecast are shown in boldface.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	5.5	8.2	10.5	12.3	12.0	11.2	12.6
OCD5	6.8	9.7	11.2	**11.7**	**11.9**	11.6	**12.5**
HWFI	7.0	9.7	11.5	12.9	14.4	16.3	19.8
GHMI	6.9	9.5	11.9	12.5	**11.8**	12.8	15.1
DSHP	6.4	9.4	11.5	12.7	13.3	13.4	14.2
LGEM	6.5	9.4	11.6	13.5	14.4	13.7	13.8
ICON	6.1	8.2	**9.6**	**10.6**	**10.9**	11.5	13.6
IVCN	6.1	8.2	**9.6**	**10.6**	**10.9**	11.5	13.6
FSSE	6.1	8.7	10.5	**11.6**	**11.4**	11.4	12.9
GFSI	6.5	9.2	10.8	**12.2**	12.4	13.4	14.8
EMXI	7.2	10.4	12.2	13.8	14.8	15.8	15.6
TCLP	6.7	9.6	11.4	**12.2**	12.9	13.9	13.7
# Cases	339	304	270	243	196	149	112

Table 5b. Homogenous comparison of selected Atlantic basin early intensity guidance model biases (kt) for 2012. Biases smaller than the NHC official forecast are shown in boldface.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	-1.2	-0.8	1.5	1.7	3.8	3.7	3.8
OCD5	-1.7	-1.9	-2.5	-2.7	**-1.9**	**-0.3**	**1.7**
HWFI	**-0.9**	**-0.2**	**1.2**	3.4	5.6	5.7	7.9
GHMI	**-0.8**	-3.7	-5.9	-3.8	**0.2**	**1.1**	**0.2**
DSHP	**-1.0**	**0.3**	1.4	2.8	5.4	**3.3**	**0.6**
LGEM	-1.4	-1.3	**-1.0**	**0.0**	**2.6**	**2.7**	**3.0**
ICON	**-0.8**	-1.0	**-0.8**	**0.9**	**3.7**	**3.5**	**3.0**
IVCN	**-0.8**	-1.0	**-0.8**	**0.9**	**3.7**	**3.5**	**3.0**
FSSE	-1.4	-1.9	-2.7	-2.3	**-2.6**	-3.7	-5.0
GFSI	-1.4	**-0.3**	**0.7**	2.0	**3.6**	3.7	**2.1**
EMXI	-2.3	-2.3	-2.5	-1.8	**-1.5**	**-0.7**	**-0.1**
TCLP	-1.6	-1.8	-2.6	-2.3	**-1.6**	**-0.3**	**1.8**
# Cases	339	304	270	243	196	149	112

Table 6. Official Atlantic track and intensity forecast verifications (OFCL) for 2012 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: AL012012 ALBERTO

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	11	3.7	3.0	11	1.4	1.4
012	9	26.8	55.4	9	1.7	3.0
024	7	55.4	151.2	7	4.3	8.1
036	5	84.4	217.2	5	7.0	9.0
048	3	119.3	213.3	3	10.0	7.7
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: AL022012 BERYL

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	19	5.0	5.0	19	1.1	1.3
012	17	18.6	53.1	17	4.7	5.4
024	15	19.9	117.3	15	5.3	6.8
036	13	28.9	197.0	13	6.9	8.1
048	11	31.3	274.6	11	6.4	8.1
072	7	33.6	425.2	7	5.0	14.3
096	3	53.0	691.4	3	5.0	8.3
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: AL032012 CHRIS

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	11	2.9	3.8	11	5.0	5.5
012	9	26.8	110.1	9	10.0	12.3
024	7	42.9	233.2	7	17.1	20.3
036	5	52.8	340.0	5	21.0	23.4
048	3	66.0	353.0	3	15.0	18.0
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: AL042012 DEBBY

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	16	12.5	10.1	16	1.6	1.3
012	14	39.8	47.8	14	3.6	5.0
024	12	77.3	91.1	12	6.3	5.7
036	10	127.7	140.2	10	11.5	8.6
048	8	196.0	194.5	8	19.4	7.4
072	4	456.0	283.4	4	31.3	7.0
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: AL052012 ERNESTO

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	35	8.2	8.7	35	2.9	3.0
012	33	30.8	44.9	33	7.4	8.9
024	31	50.3	93.2	31	6.9	10.4
036	29	62.2	135.3	29	8.1	10.6
048	27	71.9	166.3	27	13.7	15.0
072	23	88.1	219.8	23	11.7	16.7
096	19	141.7	276.1	19	9.7	11.1
120	15	215.1	380.6	15	11.7	11.7

Verification statistics for: AL062012 FLORENCE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	10	9.6	9.6	10	1.0	1.0
012	8	27.5	31.0	8	8.1	9.6
024	6	46.4	65.4	6	13.3	17.3
036	4	47.1	78.4	4	8.8	17.0
048	2	67.7	93.7	2	2.5	7.5
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: AL072012 HELENE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	12	14.3	14.3	12	0.8	1.7
012	8	33.4	50.6	8	6.3	5.6
024	4	66.4	127.8	4	7.5	8.3
036	1	99.4	293.4	1	10.0	16.0
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: AL082012 GORDON

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	20	3.4	3.9	20	1.8	1.8
012	18	19.4	46.6	18	7.8	9.2
024	16	33.8	119.1	16	12.8	16.3
036	14	48.6	229.0	14	13.9	18.1
048	12	62.8	355.6	12	17.9	19.2
072	8	82.9	613.2	8	24.4	20.1
096	4	95.0	852.4	4	13.8	8.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: AL092012 ISAAC

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	39	12.8	13.0	39	2.8	2.7
012	39	31.2	44.2	39	3.1	5.7
024	39	44.1	84.0	39	5.3	7.3
036	39	52.5	134.4	39	8.8	7.3
048	37	60.3	189.5	37	11.2	8.1
072	33	77.2	269.0	33	10.2	10.7
096	29	132.7	342.0	29	10.0	11.4
120	25	219.6	459.3	25	11.8	9.5

Verification statistics for: AL102012 JOYCE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	8	10.0	10.0	8	0.0	0.0
012	6	38.6	30.4	6	5.8	6.0
024	4	45.6	38.9	4	7.5	5.8
036	2	56.6	37.6	2	17.5	16.0
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: AL112012 KIRK

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	21	6.0	6.0	21	1.4	1.9
012	19	22.7	48.8	19	6.6	7.1
024	17	41.3	122.0	17	14.1	12.8
036	15	68.8	229.8	15	21.0	17.7
048	13	96.3	341.2	13	24.2	21.9
072	9	199.9	531.7	9	19.4	14.4
096	5	420.0	710.9	5	13.0	8.8
120	1	755.1	688.3	1	10.0	11.0

Verification statistics for: AL122012 LESLIE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	48	9.7	9.8	48	3.4	3.6
012	46	23.6	37.7	46	4.2	5.4
024	44	33.8	72.1	44	6.9	7.4
036	42	49.9	115.6	42	10.5	8.6
048	40	69.4	156.8	40	14.9	9.5
072	36	96.6	240.6	36	18.6	10.1
096	32	127.4	294.0	32	17.8	7.4
120	28	165.8	346.3	28	18.6	8.9

Verification statistics for: AL132012 MICHAEL

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	32	4.1	4.1	32	2.2	1.9
012	30	16.7	40.0	30	8.0	8.3
024	28	31.4	88.8	28	11.3	11.1
036	26	42.7	147.8	26	16.0	15.7
048	24	55.2	210.7	24	19.8	19.0
072	20	89.7	329.4	20	19.8	21.1
096	16	111.5	431.5	16	21.9	19.6
120	11	93.5	521.4	11	21.8	16.6

Verification statistics for: AL142012 NADINE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	83	7.2	7.4	83	1.5	1.6
012	79	18.6	48.0	79	4.1	4.9
024	75	31.9	123.0	75	6.9	7.2
036	72	45.0	212.9	72	7.2	7.8
048	70	62.0	299.7	70	7.3	7.6
072	66	105.0	418.0	66	8.3	8.7
096	62	161.1	514.0	62	8.2	11.1
120	58	209.2	637.2	58	8.5	11.9

Verification statistics for: AL152012 OSCAR

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	9	6.0	5.3	9	3.9	3.9
012	7	21.8	78.3	7	5.0	4.4
024	5	36.7	198.2	5	7.0	4.2
036	3	58.7	331.5	3	6.7	5.0
048	1	61.5	401.5	1	5.0	3.0
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: AL162012 PATTY

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	8	12.5	13.2	8	2.5	2.5
012	6	34.3	35.4	6	4.2	4.0
024	4	72.7	69.1	4	5.0	12.3
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: AL172012 RAFAEL

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	20	2.9	2.9	20	2.8	2.5
012	18	26.2	55.1	18	4.2	5.9
024	16	41.3	131.3	16	5.3	8.8
036	14	48.5	203.8	14	6.8	11.9
048	12	67.7	248.9	12	7.5	14.6
072	8	116.9	351.7	8	11.9	23.1
096	4	145.7	731.8	4	15.0	21.5
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: AL182012 SANDY

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	30	9.6	9.5	30	3.2	3.7
012	28	23.9	56.7	28	8.0	10.6
024	26	33.2	118.2	26	10.6	14.0
036	24	39.6	189.7	24	11.0	16.9
048	22	41.6	252.1	22	10.9	17.2
072	18	61.3	360.8	18	10.3	18.0
096	14	88.3	477.9	14	8.9	22.9
120	10	148.9	647.3	10	14.5	27.9

Verification statistics for: AL192012 TONY

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	12	22.6	24.3	12	0.0	0.4
012	10	28.2	77.6	10	3.5	4.2
024	8	68.7	186.7	8	5.0	7.4
036	6	133.4	324.3	6	4.2	7.3
048	4	219.1	482.7	4	3.8	4.0
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 7. Homogenous comparison of official and CLIPER5 track forecast errors in the eastern North Pacific basin in 2012 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
2012 mean OFCL error (n mi)	23.2	36.1	49.2	63.8	88.6	107.6	100.7
2012 mean CLIPER5 error (n mi)	35.1	71.2	117.8	160.5	215.7	252.9	304.9
2012 mean OFCL skill relative to CLIPER5 (%)	33.9	49.3	58.2	60.2	58.9	57.5	67.0
2012 mean OFCL bias vector (°/n mi)	317/003	293/007	284/012	285/014	276/019	268/013	043/031
2012 number of cases	278	246	214	183	125	76	39
2007-2011 mean OFCL error (n mi)	28.6	46.3	62.7	78.1	108.0	145.3	181.1
2007-2011 mean CLIPER5 error (n mi)	38.5	74.8	116.0	159.8	246.1	324.2	392.8
2007-2011 mean OFCL skill relative to CLIPER5 (%)	25.7	38.1	45.9	51.1	56.1	55.2	53.9
2007-2011 mean OFCL bias vector (°/n mi)	243/001	174/001	166/003	151/004	127/011	107/024	103/41
2007-2011 number of cases	1091	953	824	712	523	367	237
2012 OFCL error relative to 2007-2011 mean (%)	-18.9	-22.0	-21.5	-18.3	-18.0	-25.9	-44.4
2012 CLIPER5 error relative to 2007-2011 mean (%)	-8.8	-4.8	1.6	0.4	-12.4	-22.0	-22.4

Table 8a. Homogenous comparison of eastern North Pacific basin early track guidance model errors (n mi) for 2012. Errors smaller than the NHC official forecast are shown in boldface.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	22.1	34.3	47.4	61.4	71.0	87.4	95.8
OCD5	33.9	69.5	116.5	163.9	208.2	278.0	272.5
GFSI	23.7	38.4	55.6	82.9	107.2	139.2	164.8
GHMI	28.1	47.6	64.8	82.9	107.2	139.2	164.8
HWFI	25.4	41.6	56.5	75.0	90.6	130.8	152.6
NGXI	31.9	55.4	82.2	106.2	128.1	173.4	208.3
EGRI	31.6	53.8	73.0	89.3	112.4	143.5	159.2
EMXI	24.0	36.4	50.0	62.8	84.6	105.6	109.6
CMCI	31.3	54.4	77.3	96.8	144.9	174.6	125.2
AEMI	25.0	41.4	61.1	79.6	94.5	137.1	192.6
TVCE	**21.7**	34.4	47.9	61.6	**68.5**	**86.6**	96.6
LBAR	32.8	65.2	106.5	147.8	208.3	261.1	259.7
BAMD	36.9	65.5	93.9	117.9	149.3	195.6	283.3
BAMM	33.3	57.5	81.2	103.0	127.1	178.0	278.9
BAMS	40.8	75.8	112.3	144.2	173.5	213.8	325.2
TCLP	29.4	59.7	101.6	148.3	199.8	263.4	252.9
# Cases	212	184	158	133	81	39	15

Table 8b. Homogenous comparison of eastern North Pacific basin early track guidance model bias vectors (º/n mi) for 2012.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	330/003	304/005	294/010	301/011	353/014	008/025	030/075
OCD5	278/005	258/018	259/039	263/059	297/055	006/066	036/198
GFSI	342/005	332/009	326/014	326/018	341/036	331/068	341/182
GHMI	087/006	081/010	049/011	045/017	055/052	063/070	079/123
HWFI	356/007	340/007	301/009	291/011	296/016	350/013	092/090
NGXI	297/008	275/012	269/023	274/034	308/050	324/117	355/204
EGRI	245/008	237/023	241/036	237/046	232/054	229/051	133/107
EMXI	001/006	352/006	320/005	338/003	055/021	032/034	057/066
CMCI	336/009	318/017	311/030	316/037	352/051	326/095	304/090
AEMI	305/006	300/013	296/021	294/029	314/049	321/076	334/166
TVCE	346/003	292/004	277/010	277/011	354/009	004/021	061/065
LBAR	352/009	323/036	315/072	314/104	319/163	320/232	341/244
BAMD	332/012	319/022	310/036	311/045	329/068	317/095	340/207
BAMM	318/015	303/027	293/041	291/052	308/070	323/107	336/242
BAMS	305/018	294/033	286/055	285/071	302/103	326/139	349/277
TCLP	231/007	233/019	242/036	248/054	279/050	006/042	033/177
# Cases	212	184	158	133	81	39	15

Table 9. Homogenous comparison of official and Decay-SHIFOR5 intensity forecast errors in the eastern North Pacific basin for the 2012 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
2012 mean OFCL error (kt)	5.1	8.1	10.7	11.8	10.6	13.8	13.8
2012 mean Decay-SHIFOR5 error (kt)	6.5	10.7	14.6	16.8	18.8	19.3	17.0
2012 mean OFCL skill relative to Decay-SHIFOR5 (%)	21.5	24.3	26.7	29.8	43.6	28.5	18.8
2012 OFCL bias (kt)	-0.7	-1.5	-3.4	-5.0	-4.1	-2.9	0.0
2012 number of cases	278	246	214	183	125	76	39
2007-11 mean OFCL error (kt)	6.4	10.6	13.7	15.1	17.0	18.5	17.8
2007-11 mean Decay-SHIFOR5 error (kt)	7.5	12.4	16.1	18.4	20.1	20.1	20.8
2007-11 mean OFCL skill relative to Decay-SHIFOR5 (%)	14.7	14.5	14.9	17.9	15.4	8.0	14.4
2007-11 OFCL bias (kt)	0.4	0.7	0.7	0.1	0.4	-0.1	-0.4
2007-11 number of cases	1091	953	824	712	523	367	237
2012 OFCL error relative to 2007-11 mean (%)	-20.3	-23.6	-21.9	-21.9	-37.6	-25.4	-22.5
2012 Decay-SHIFOR5 error relative to 2007-11 mean (%)	-13.3	-13.7	-9.3	-8.7	-6.5	-4.0	-18.3

Table 10a. Homogenous comparison of eastern North Pacific basin early intensity guidance model errors (kt) for 2012. Errors smaller than the NHC official forecast are shown in boldface.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	5.0	8.2	10.8	11.8	10.5	14.9	15.0
OCD5	6.5	10.8	14.7	16.8	18.3	19.2	17.7
HWFI	6.3	9.5	12.1	14.6	15.8	20.4	20.7
GHMI	6.7	10.8	14.8	16.2	15.2	16.5	16.5
DSHP	5.8	8.9	11.4	13.1	13.2	15.9	18.6
LGEM	6.3	9.8	12.6	14.3	13.2	**13.7**	**14.3**
ICON	5.9	8.6	11.4	13.1	12.4	**14.8**	16.2
IVCN	5.8	8.6	11.4	13.1	12.3	14.9	16.2
GFSI	7.3	11.3	14.9	17.0	18.4	18.3	18.7
EMXI	8.4	14.6	19.4	22.7	23.3	23.8	23.4
TCLP	6.8	11.5	15.9	18.9	20.0	19.5	16.3
# Cases	261	233	202	171	115	63	30

Table 10b. Homogenous comparison of eastern North Pacific basin early intensity guidance model biases (kt) for 2012. Biases smaller than the NHC official forecast are shown in boldface.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	-0.7	-1.6	-3.5	-5.1	-3.8	-2.1	-0.3
OCD5	-0.8	**-0.4**	**-0.1**	**0.5**	**3.7**	4.0	6.4
HWFI	-2.2	-3.1	-3.8	**-4.7**	-6.0	-11.6	-10.1
GHMI	-2.4	-5.7	-8.2	-7.5	-3.8	-4.4	-1.9
DSHP	-1.3	-2.0	**-3.3**	**-4.1**	-5.2	-4.9	-2.8
LGEM	-2.0	-4.3	-7.2	-9.2	-9.5	-7.1	-5.3
ICON	-1.7	-3.5	-5.3	-6.0	-5.9	-6.8	-4.7
IVCN	-1.7	-3.6	-5.5	-6.2	-5.8	-6.8	-4.7
GFSI	-2.3	-3.1	-4.1	**-4.6**	**-3.5**	**-2.0**	-1.7
EMXI	-2.2	-3.0	-3.8	**-3.6**	**-1.7**	2.8	9.9
TCLP	-1.2	-2.4	-3.9	**-4.8**	**-3.1**	**-0.7**	2.8
# Cases	261	233	202	171	115	63	30

Table 11. Official eastern North Pacific track and intensity forecast verifications (OFCL) for 2012 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: EP012012 ALETTA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	19	11.2	10.6	19	2.1	2.1
012	17	29.8	41.8	17	2.6	3.1
024	15	57.8	95.6	15	2.3	5.2
036	13	73.5	154.2	13	4.2	10.6
048	11	99.1	193.7	11	4.5	18.9
072	7	238.2	283.3	7	3.6	28.1
096	3	419.7	333.7	3	5.0	28.7
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: EP022012 BUD

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	21	7.5	7.5	21	3.3	3.3
012	19	28.4	42.1	19	8.7	11.2
024	17	50.0	93.5	17	14.1	18.6
036	15	70.2	183.0	15	19.0	23.1
048	13	91.0	267.7	13	18.5	22.8
072	9	116.4	367.4	9	11.1	25.0
096	5	147.3	426.7	5	11.0	19.8
120	1	262.3	461.4	1	25.0	6.0

Verification statistics for: EP032012 CARLOTTA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	12	5.9	5.9	12	0.8	0.8
012	10	29.0	29.3	10	11.5	8.7
024	8	48.7	63.1	8	20.0	12.6
036	6	68.2	107.0	6	26.7	25.2
048	4	90.8	162.0	4	32.5	33.5
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: EP042012 DANIEL

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	29	5.1	5.6	29	1.6	2.1
012	29	16.5	24.1	29	4.8	6.2
024	29	22.4	45.3	29	7.1	9.6
036	27	29.7	72.7	27	10.7	12.8
048	25	36.4	110.5	25	13.0	16.6
072	21	43.4	192.6	21	19.0	22.2
096	17	53.3	286.3	17	21.8	23.2
120	13	85.5	372.7	13	15.4	16.2

Verification statistics for: EP052012 EMILIA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	32	13.1	13.1	32	2.3	2.3
012	30	17.8	25.5	30	6.5	8.5
024	28	24.7	46.8	28	10.9	14.9
036	26	33.2	72.2	26	13.5	18.0
048	24	45.3	99.5	24	11.7	17.3
072	20	63.4	149.3	20	4.8	17.5
096	16	84.8	193.0	16	14.1	22.7
120	12	95.0	217.8	12	11.7	16.6

Verification statistics for: EP062012 FABIO

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	25	5.9	5.6	25	2.6	2.6
012	23	16.3	26.4	23	5.2	4.7
024	21	29.0	61.1	21	7.9	10.0
036	19	40.2	98.5	19	11.8	15.5
048	17	59.4	128.8	17	13.8	18.2
072	13	91.7	151.2	13	14.2	22.1
096	9	124.6	187.2	9	8.3	17.1
120	5	121.3	328.6	5	7.0	17.6

Verification statistics for: EP072012 GILMA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	17	8.9	8.9	17	0.9	0.9
012	15	22.3	35.5	15	4.0	5.9
024	13	39.9	76.9	13	6.2	8.8
036	11	59.9	138.6	11	7.3	13.5
048	9	77.5	201.4	9	5.6	15.8
072	5	112.0	269.3	5	6.0	14.2
096	1	183.6	333.9	1	15.0	13.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: EP082012 HECTOR

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	22	10.7	11.2	22	1.6	1.8
012	20	31.1	41.3	20	4.5	6.7
024	18	40.5	71.8	18	5.0	9.5
036	16	50.5	110.0	16	6.3	11.6
048	14	69.0	147.4	14	7.9	12.4
072	10	98.1	239.0	10	14.0	20.3
096	6	90.0	376.0	6	23.3	23.7
120	2	37.0	420.0	2	20.0	23.5

Verification statistics for: EP092012 ILEANA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	23	7.9	9.0	23	1.3	1.7
012	21	20.7	28.7	21	2.1	4.6
024	19	29.4	58.5	19	3.9	6.9
036	17	44.4	95.8	17	4.1	8.9
048	15	60.3	132.6	15	5.7	9.5
072	11	78.6	153.8	11	2.7	11.1
096	7	91.8	196.7	7	5.0	13.4
120	3	77.3	196.6	3	16.7	20.7

Verification statistics for: EP102012 JOHN

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	7	9.7	10.5	7	3.6	3.6
012	5	22.9	36.2	5	1.0	3.8
024	3	20.4	79.1	3	5.0	3.0
036	1	21.2	138.3	1	10.0	8.0
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: EP112012 KRISTY

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	19	8.6	10.7	19	2.1	2.1
012	17	19.3	30.0	17	3.2	4.0
024	15	30.2	54.6	15	4.7	3.1
036	13	39.2	77.0	13	5.0	4.5
048	11	51.5	102.9	11	5.5	5.5
072	7	73.6	136.8	7	5.0	12.4
096	3	140.4	199.0	3	5.0	13.3
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: EP122012 LANE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	15	13.2	13.2	15	1.0	1.0
012	13	25.4	44.0	13	5.0	7.2
024	11	32.4	78.8	11	6.8	11.2
036	9	49.4	144.2	9	10.0	13.1
048	7	64.6	218.9	7	12.9	17.1
072	3	99.5	349.4	3	15.0	12.3
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: EP132012 MIRIAM

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	23	11.0	11.4	23	1.5	1.7
012	21	24.6	36.8	21	5.2	6.9
024	19	34.2	71.7	19	11.1	14.5
036	17	42.8	103.4	17	14.7	21.5
048	15	51.3	124.6	15	18.7	25.9
072	11	60.3	133.6	11	16.8	20.5
096	7	88.4	137.5	7	12.9	9.6
120	3	167.1	299.4	3	16.7	17.0

Verification statistics for: EP142012 NORMAN

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	4	12.6	12.6	4	3.8	3.8
012	2	21.3	40.5	2	5.0	4.0
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: EP152012 OLIVIA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	10	7.0	6.4	10	0.5	0.5
012	8	33.8	54.3	8	8.1	6.5
024	6	50.8	112.3	6	11.7	10.7
036	4	67.4	199.3	4	15.0	14.5
048	2	68.3	217.6	2	12.5	6.0
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: EP162012 PAUL

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	14	11.7	12.5	14	2.9	2.9
012	12	30.9	66.8	12	8.8	12.8
024	10	51.5	160.3	10	16.0	22.1
036	8	73.6	323.7	8	18.8	24.4
048	6	91.7	502.5	6	24.2	30.2
072	2	115.4	887.1	2	7.5	12.5
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: EP172012 ROSA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	18	8.5	10.1	18	1.9	1.7
012	16	23.4	34.5	16	1.3	3.5
024	14	49.3	79.4	14	2.1	5.1
036	12	74.9	134.0	12	5.0	7.8
048	10	99.1	192.0	10	5.0	8.4
072	6	146.1	333.9	6	5.8	9.7
096	2	196.7	461.7	2	7.5	8.5
120	0	-999.0	-999.0	0	-999.0	-999.0

Table 12a Verification of experimental in-house probabilistic genesis forecasts for the Atlantic basin in 2012.

Atlantic Basin Genesis Forecast Reliability Table		
Forecast Likelihood (%)	**Verifying Genesis Occurrence Rate (%)**	**Number of Forecasts**
0	8	49
10	15	129
20	24	74
30	50	46
40	58	31
50	56	25
60	75	16
70	100	16
80	100	10
90	100	1
100	-	0

Table 12b. Verification of experimental in-house probabilistic genesis forecasts for the eastern North Pacific basin in 2012.

Eastern North Pacific Basin Genesis Forecast Reliability Table		
Forecast Likelihood (%)	**Verifying Genesis Occurrence Rate (%)**	**Number of Forecasts**
0	6	18
10	12	86
20	17	59
30	46	46
40	50	26
50	92	36
60	64	22
70	50	8
80	60	15
90	38	8
100	-	0

Table 13. HFIP Stream 1.5 models for 2012.

ID	Description	Parameter	NHC Application
APSI	PSU ARW with radar data assimilated. Early version of APSU.	Trk, Int	Direct use. Include in TV15 and IV15 consensus.
FM9I	ESRL FIM 15-km global model. Early version of FIM9.	Trk	Include in TV15 consensus.
UWNI	University of Wisconsin non-hydrostatic. Early version of UWN8.	Int	Include in IV15 consensus.
SPC3	CIRA statistical intensity consensus.	Int	Direct use.
AHWI	SUNY Advanced Hurricane WRF. Early version of AHW4.	Trk, Int	Include in TV15 and IV15 consensus.
COTI	NRL COAMPS-TC regional model. Early version of COTC.	Int	Include in IV15 consensus.
GPMI	GFDL ensemble mean. Early version of GPMN.	Trk, Int	Direct use.
G01I	Unbogussed GFDL ensemble member. Early version of GP01.	Trk, Int	Direct use.

Table 14. NHC forecast cone circle radii (n mi) for 2013. Change from 2012 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**
12	33 (-3: -8%)	30 (-3: -9%)
24	52 (-4: -7%)	49 (-3: -6%)
36	72 (-3: -4%)	66 (-6: -8%)
48	92 (-3: -3%)	82 (-7: -8%)
72	128 (-13: -8%)	111 (-10: -8%)
96	177(-3: -3%)	157 (-13: -8%)
120	229 (-7: -3%)	197 (-19: -9%)

Table 15. Composition of NHC consensus models for 2013. 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 2013			
Model ID	**Parameter**	**Type**	**Members**
TCOA	Track	Fixed	GFSI EGRI GHMI HWFI
TCOE*	Track	Fixed	GFSI EGRI GHMI HWFI
ICON	Intensity	Fixed	DSHP LGEM GHMI HWFI
TVCA	Track	Variable	GFSI EGRI GHMI HWFI EMXI
TVCE**	Track	Variable	GFSI EGRI GHMI HWFI EMXI
IVCN	Intensity	Variable	DSHP LGEM GHMI HWFI

* 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 TVCE.
GPCE circles will continue to be based on TVCN.

List of Figures

1. NHC official and CLIPER5 (OCD5) Atlantic basin average track errors for 2012 (solid lines) and 2007-2011 (dashed lines).
2. Recent trends in NHC official track forecast error (top) and skill (bottom) for the Atlantic basin.
3. Homogenous comparison for selected Atlantic basin early track guidance models for 2012. This verification includes only those models that were available at least 2/3 of the time (see text).
4. Homogenous comparison of the primary Atlantic basin track consensus models for 2010-2012.
5. Homogenous comparison of the primary Atlantic basin track consensus models for 2012.
6. NHC official and Decay-SHIFOR5 (OCD5) Atlantic basin average intensity errors for 2012 (solid lines) and 2007-2011 (dashed lines).
7. Recent trends in NHC official intensity forecast error (top) and skill (bottom) for the Atlantic basin.
8. Homogenous comparison for selected Atlantic basin early intensity guidance models for 2011. This verification includes only those models that were available at least 2/3 of the time (see text).
9. Homogenous comparison for selected Atlantic basin early intensity guidance models for 2010-2012 (top) and for pre-landfall verifications only from 2010-12 (bottom). The pre-landfall verification sample is defined by excluding any portion of a model forecast that occurs after either the model forecast track or the verifying best track encounters land.
10. NHC official and CLIPER5 (OCD5) eastern North Pacific basin average track errors for 2012 (solid lines) and 2007-2011 (dashed lines).
11. Recent trends in NHC official track forecast error (top) and skill (bottom) for the eastern North Pacific basin.
12. Homogenous comparison for selected eastern North Pacific early track models for 2012. This verification includes only those models that were available at least 2/3 of the time (see text).
13. Homogenous comparison of the primary eastern North Pacific basin track consensus models for 2012.
14. NHC official and Decay-SHIFOR5 (OCD5) eastern North Pacific basin average intensity errors for 2012 (solid lines) and 2007-2011 (dashed lines).
15. Recent trends in NHC official intensity forecast error (top) and skill (bottom) for the eastern North Pacific basin.

16. Homogenous comparison for selected eastern North Pacific basin early intensity guidance models for 2012. This verification includes only those models that were available at least 2/3 of the time (see text).
17. Reliability diagram for Atlantic (a) and eastern North Pacific (b) probabilistic tropical cyclogenesis forecasts for 2012. The solid blue line indicates the relationship between the forecast and verifying genesis percentages, with perfect reliability indicated by the thin diagonal black line. The dashed green line indicates how the forecasts were distributed among the possible forecast values.
18. Homogeneous comparison of HFIP Stream 1.5 track models and selected operational models for 2012 (top), including FM9I (bottom).
19. Impact of adding Stream 1.5 models to the variable track consensus TVCA.
20. Homogeneous comparison of HFIP Stream 1.5 GFDL ensemble mean and GFDL unbogused ensemble member for track (top) and intensity (bottom).
21. Homogeneous comparison of HFIP Stream 1.5 intensity models and selected operational models for 2012.
22. Impact of adding Stream 1.5 models to the fixed intensity consensus ICON.

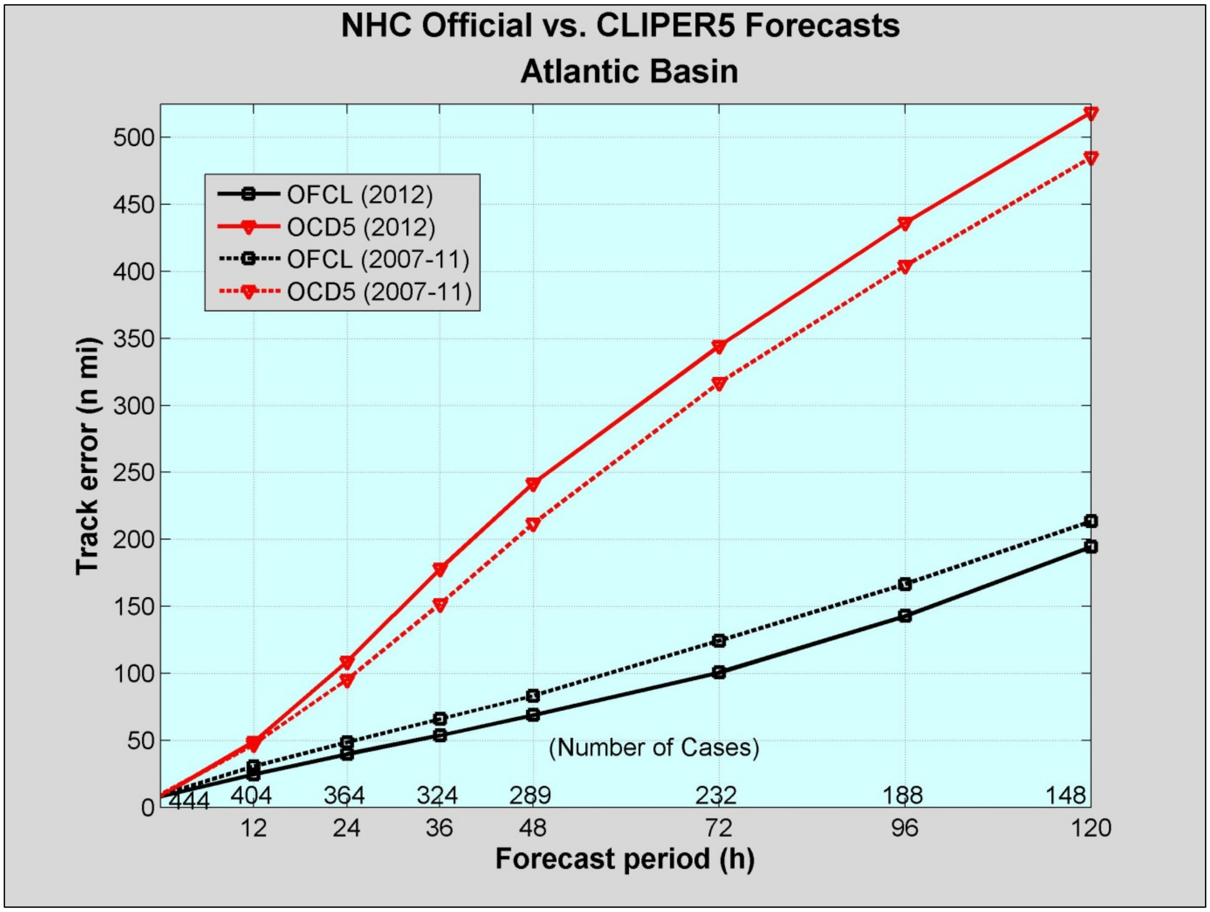


Figure 1. NHC official and CLIPER5 (OCD5) Atlantic basin average track errors for 2012 (solid lines) and 2007-2011 (dashed lines).

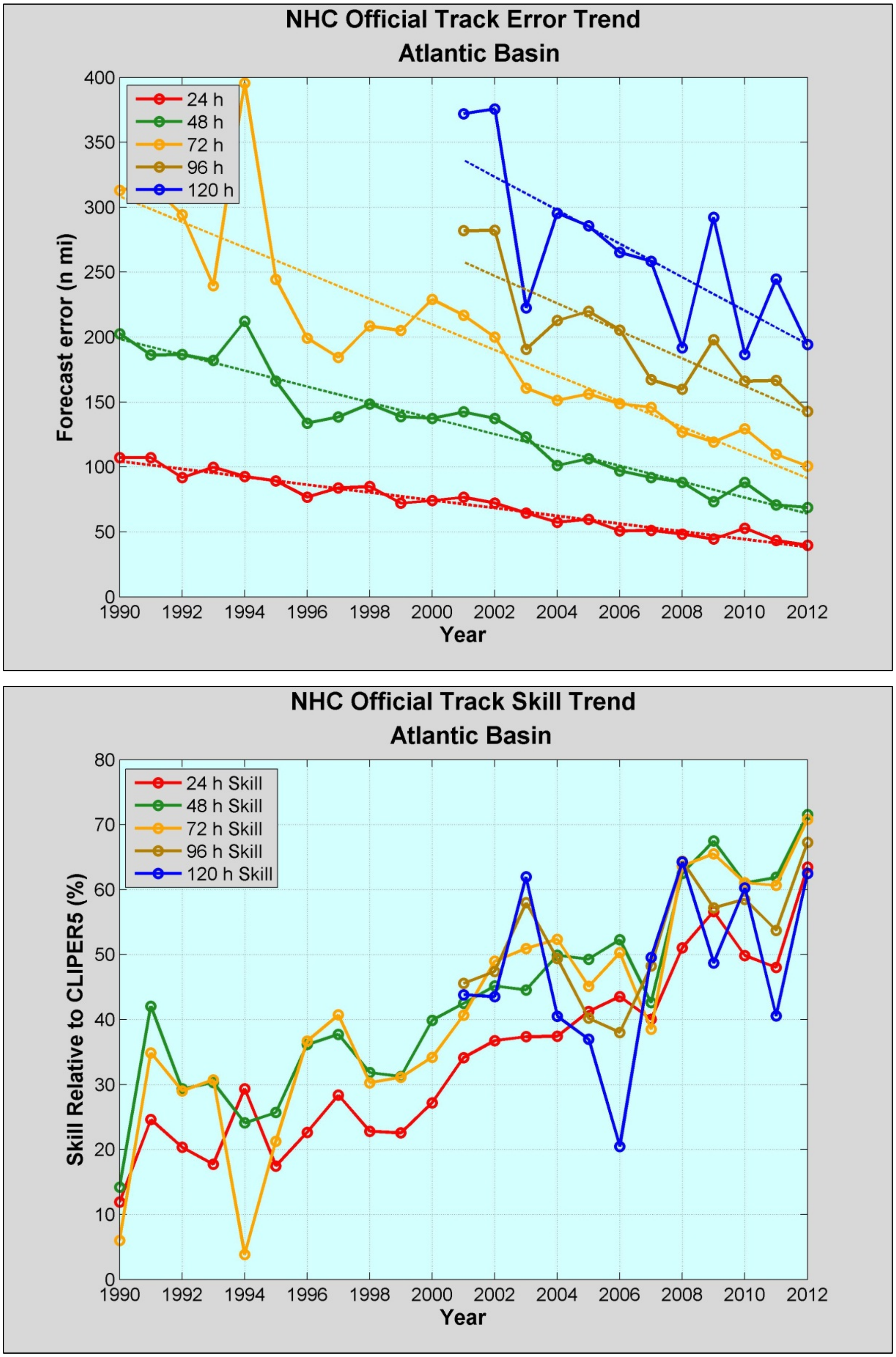


Figure 2. Recent trends in NHC official track forecast error (top) and skill (bottom) for the Atlantic basin.

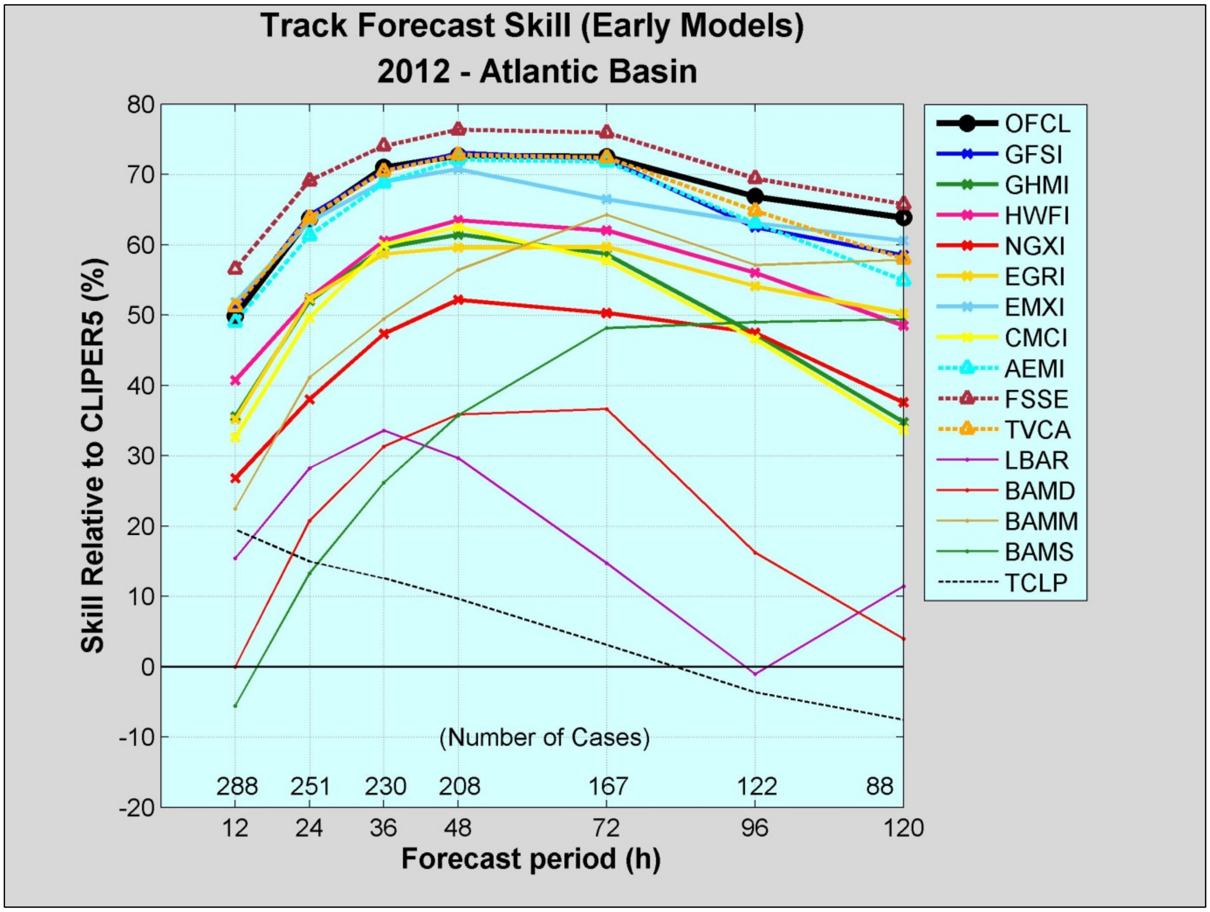


Figure 3. Homogenous comparison for selected Atlantic basin early track models for 2012. This verification includes only those models that were available at least 2/3 of the time (see text).

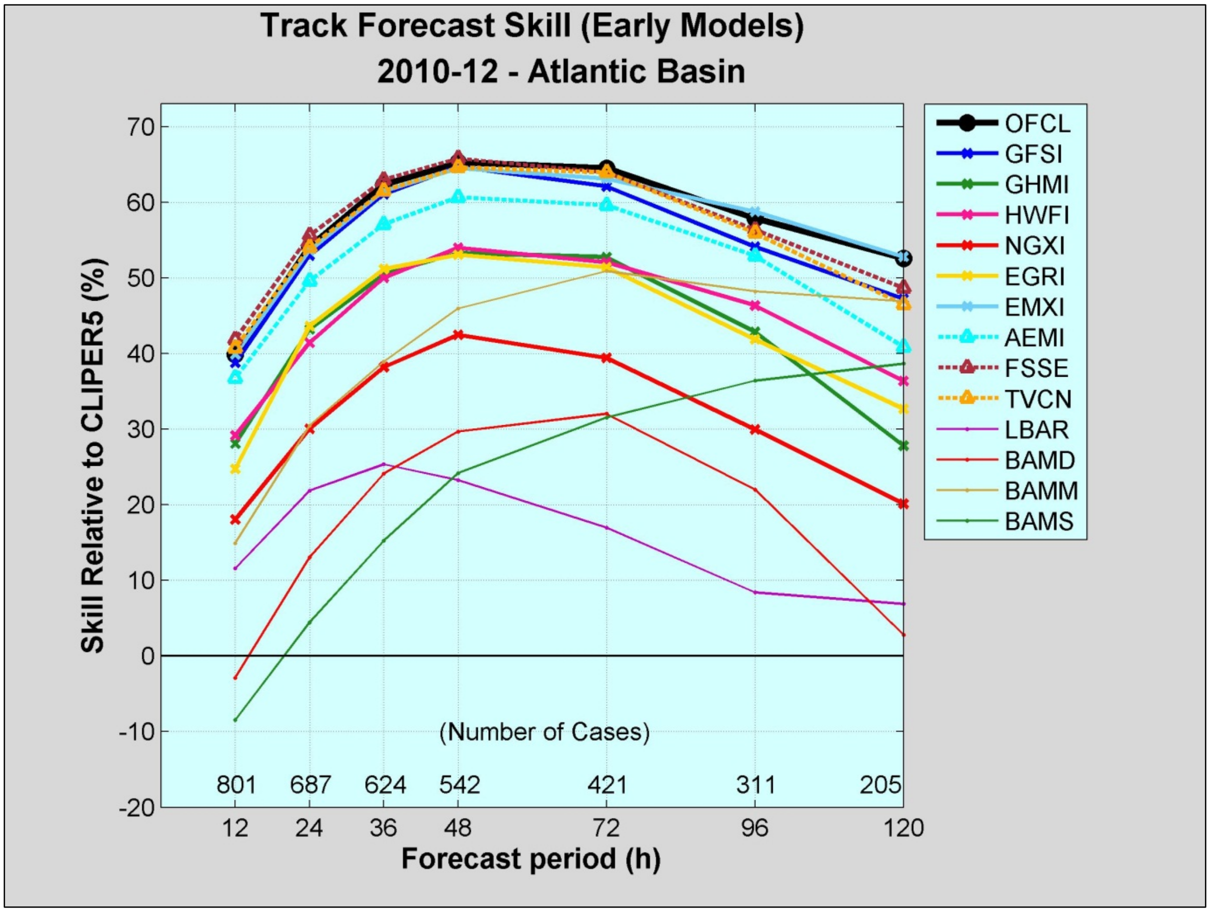


Figure 4. Homogenous comparison for selected Atlantic basin early track models for 2010-2012.

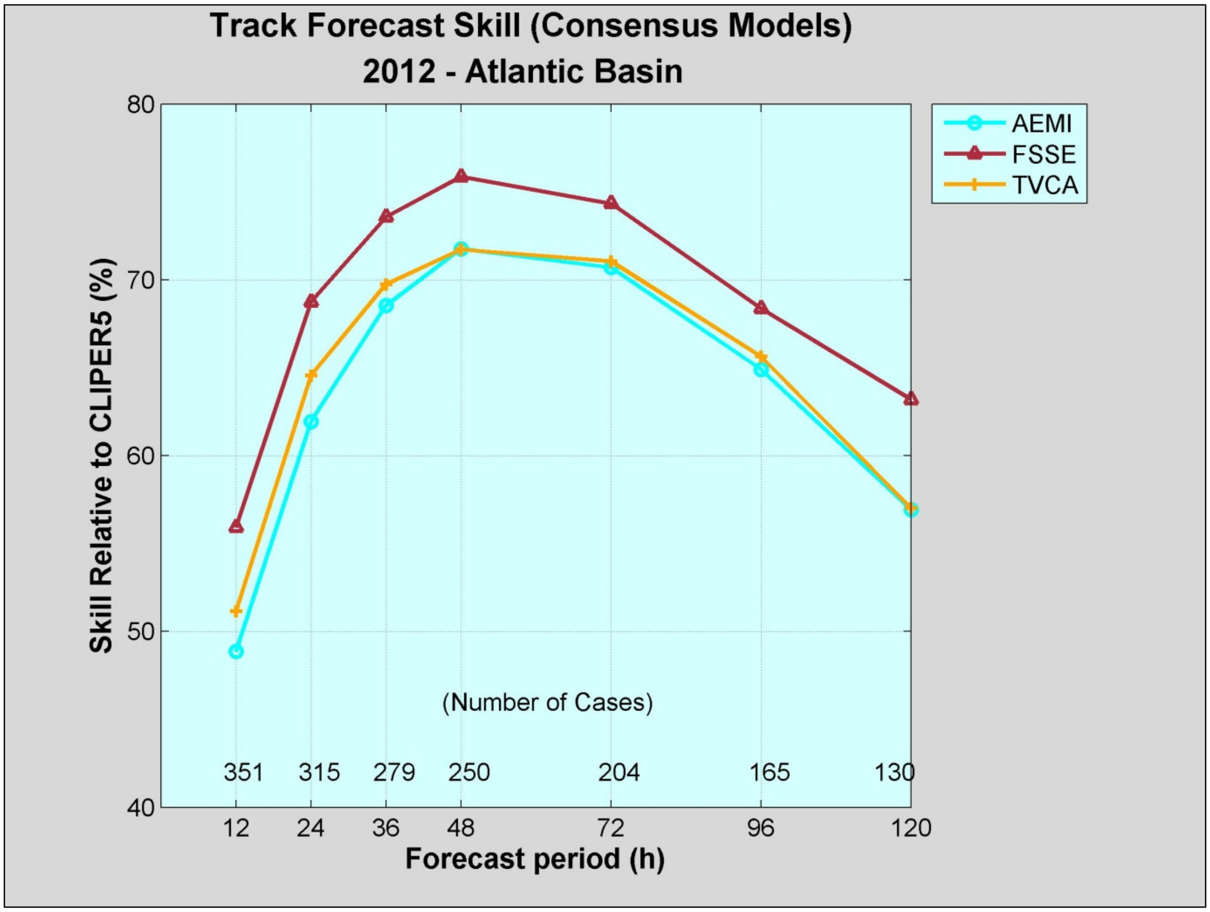


Figure 5. Homogenous comparison of the primary Atlantic basin track consensus models for 2012.

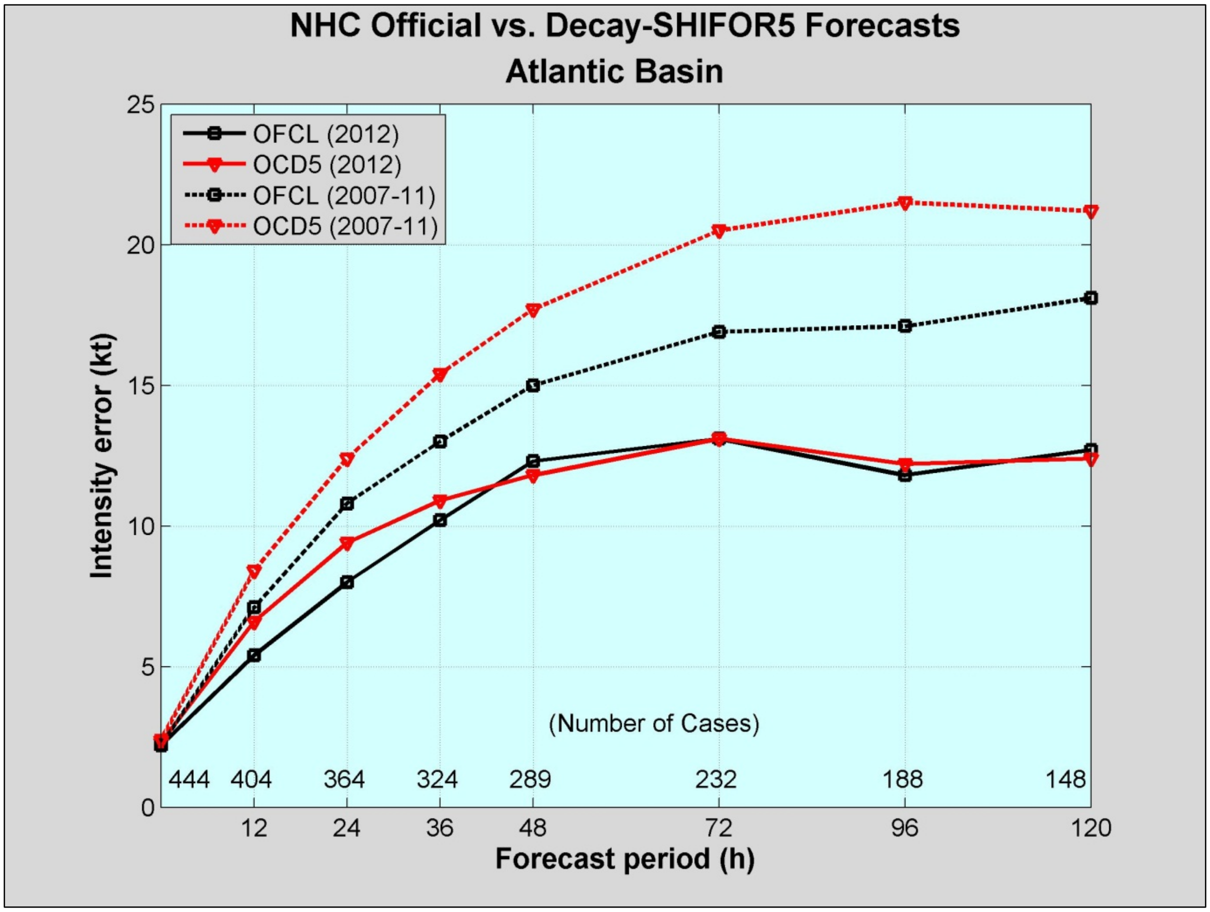
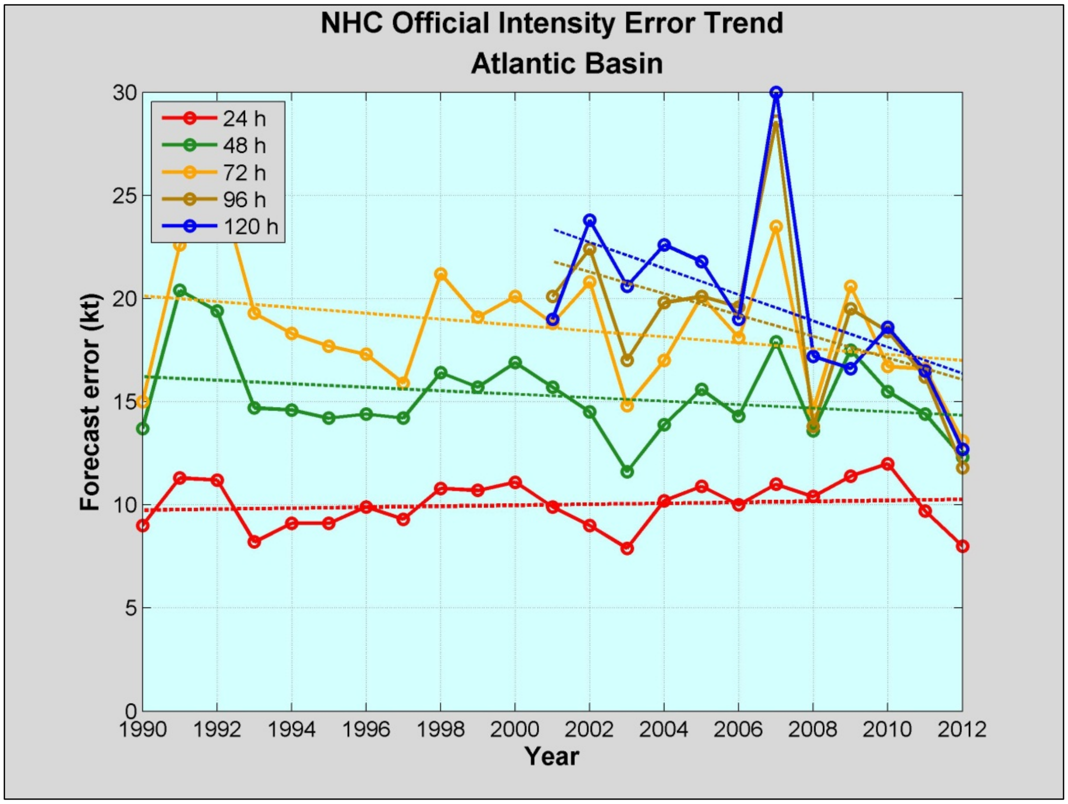


Figure 6. NHC official and Decay-SHIFOR5 (OCD5) Atlantic basin average intensity errors for 2012 (solid lines) and 2007-2011 (dashed lines).



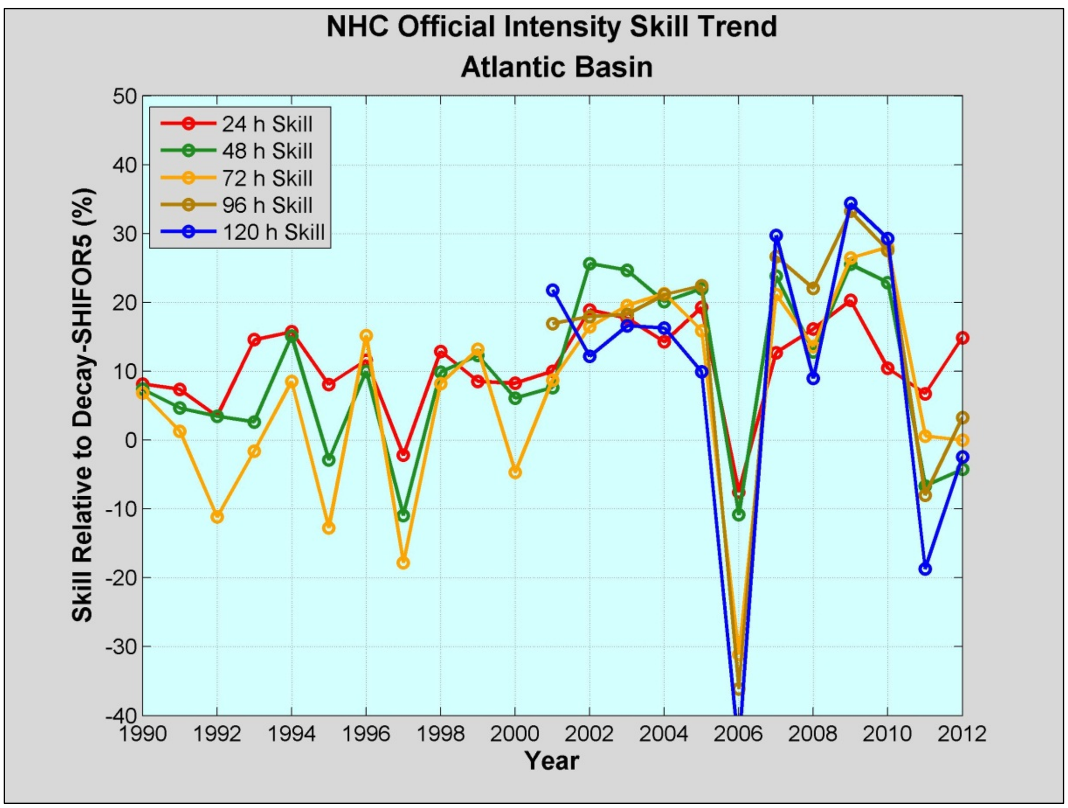


Figure 7. Recent trends in NHC official intensity forecast error (top) and skill (bottom) for the Atlantic basin.

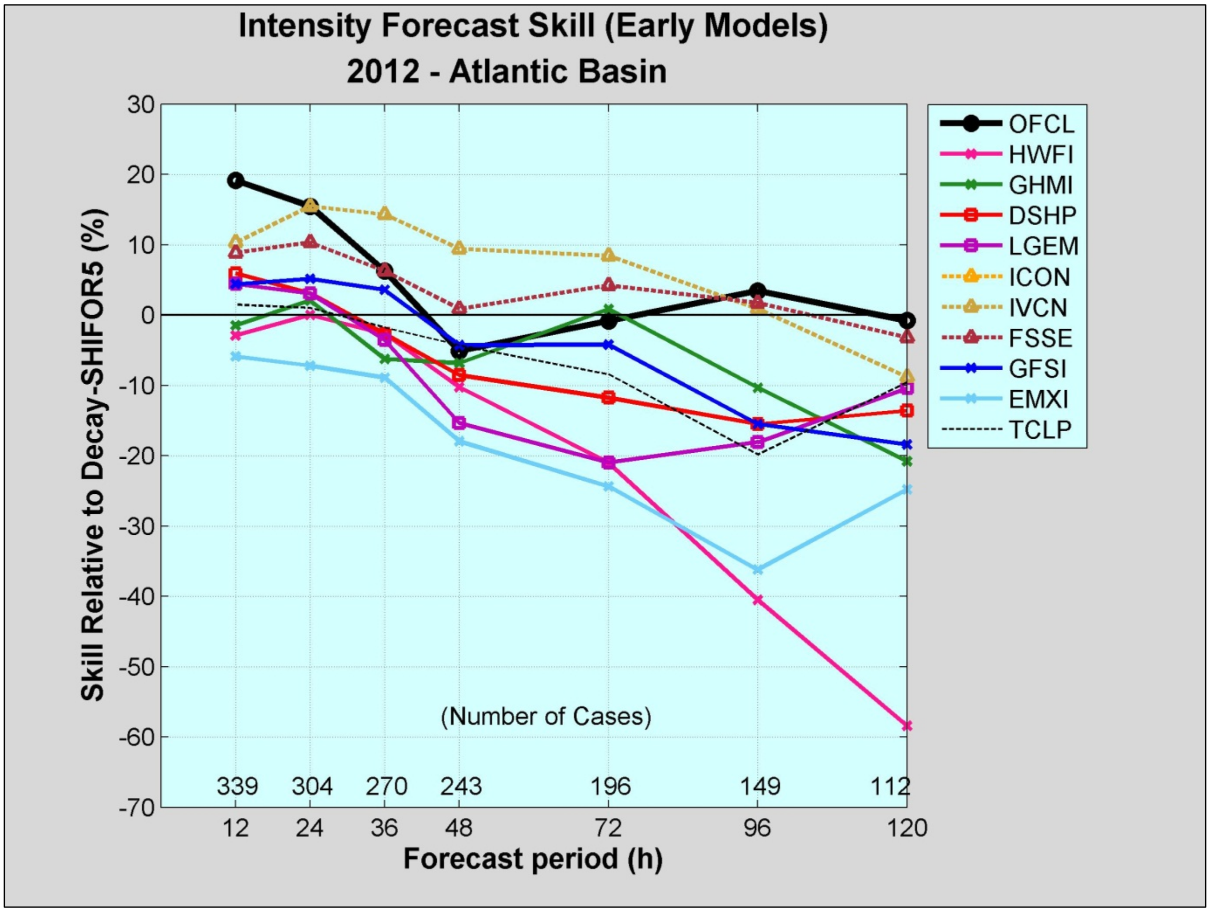
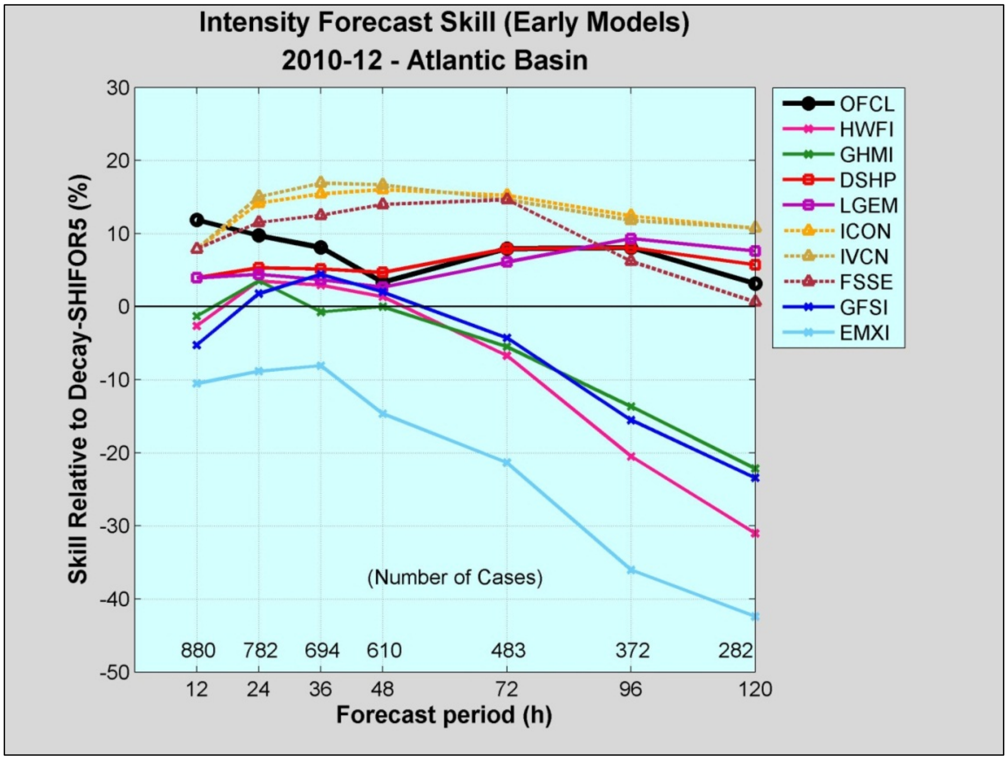


Figure 8. Homogenous comparison for selected Atlantic basin early intensity guidance models for 2012.



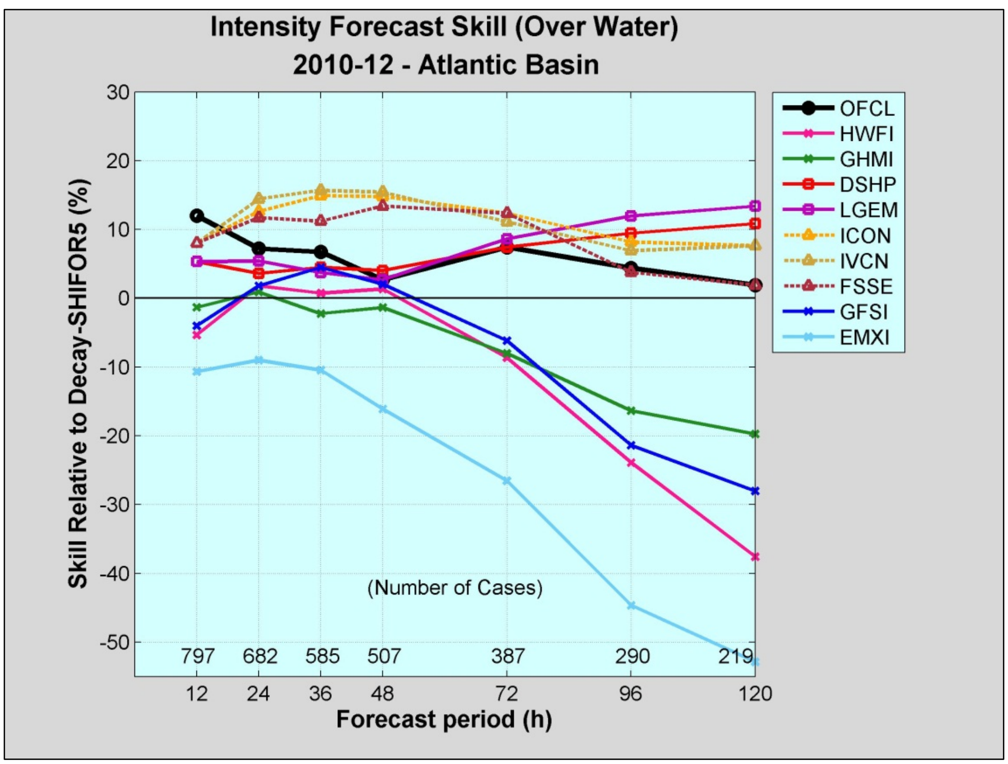


Figure 9. Homogenous comparison for selected Atlantic basin early intensity guidance models for 2010-2012 (top) and for pre-landfall verifications only from 2010-12 (bottom). The pre-landfall verification sample is defined by excluding any portion of a model forecast that occurs after either the model forecast track or the verifying best track encounters land.

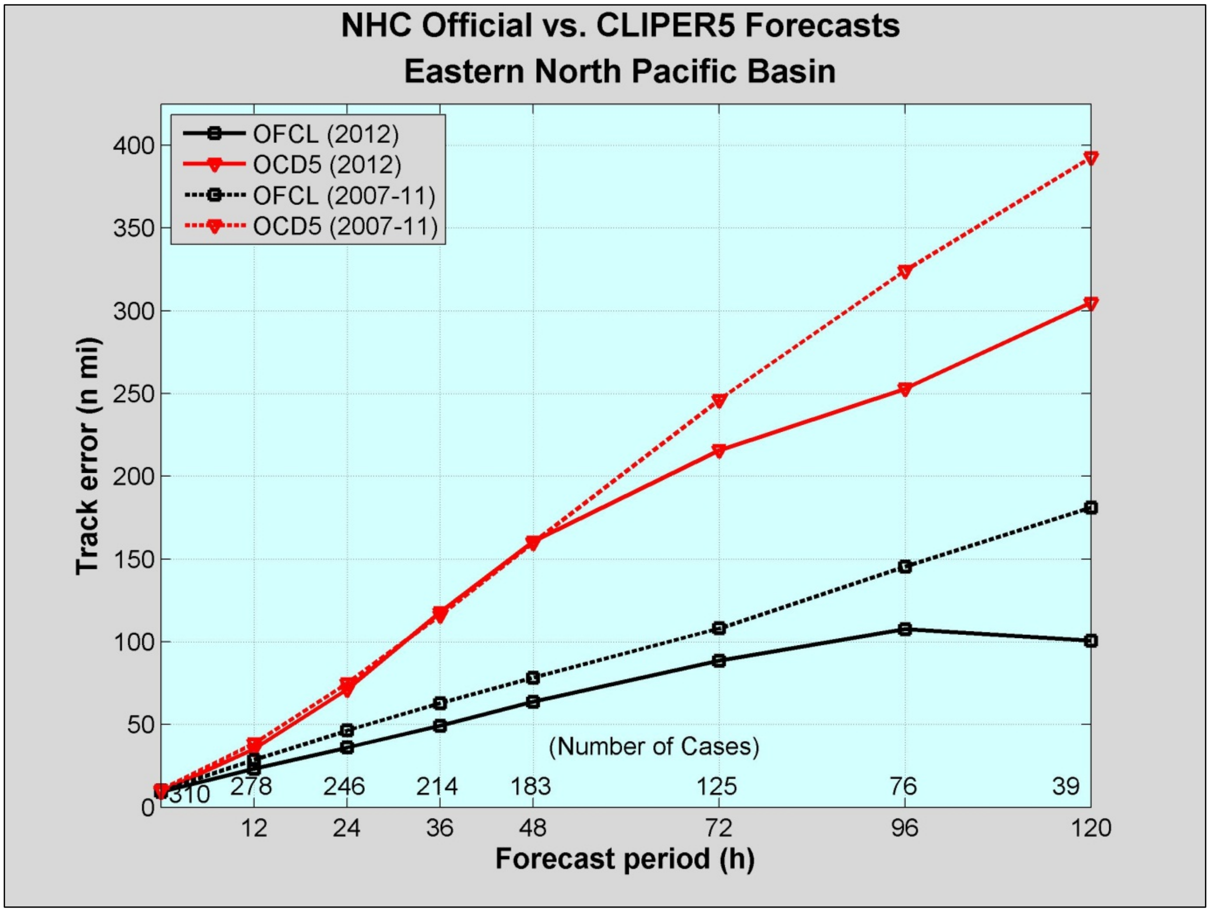
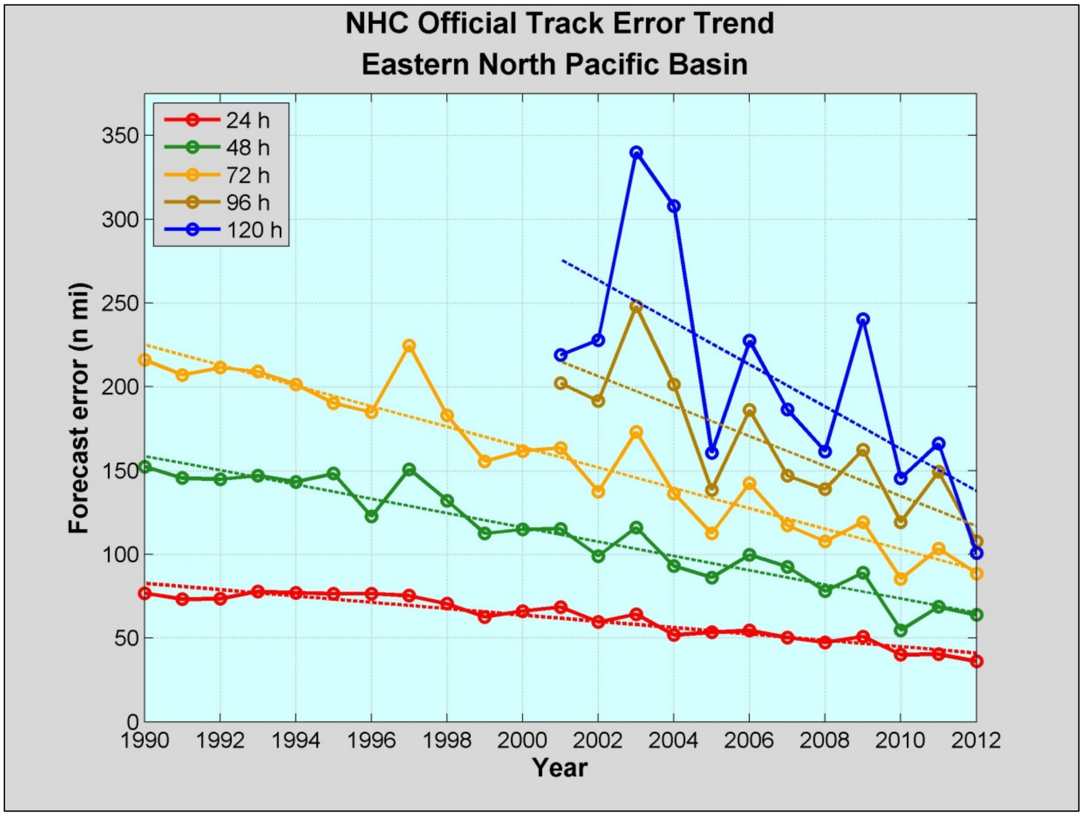


Figure 10. NHC official and CLIPER5 (OCD5) eastern North Pacific basin average track errors for 2012 (solid lines) and 2007-2011 (dashed lines).



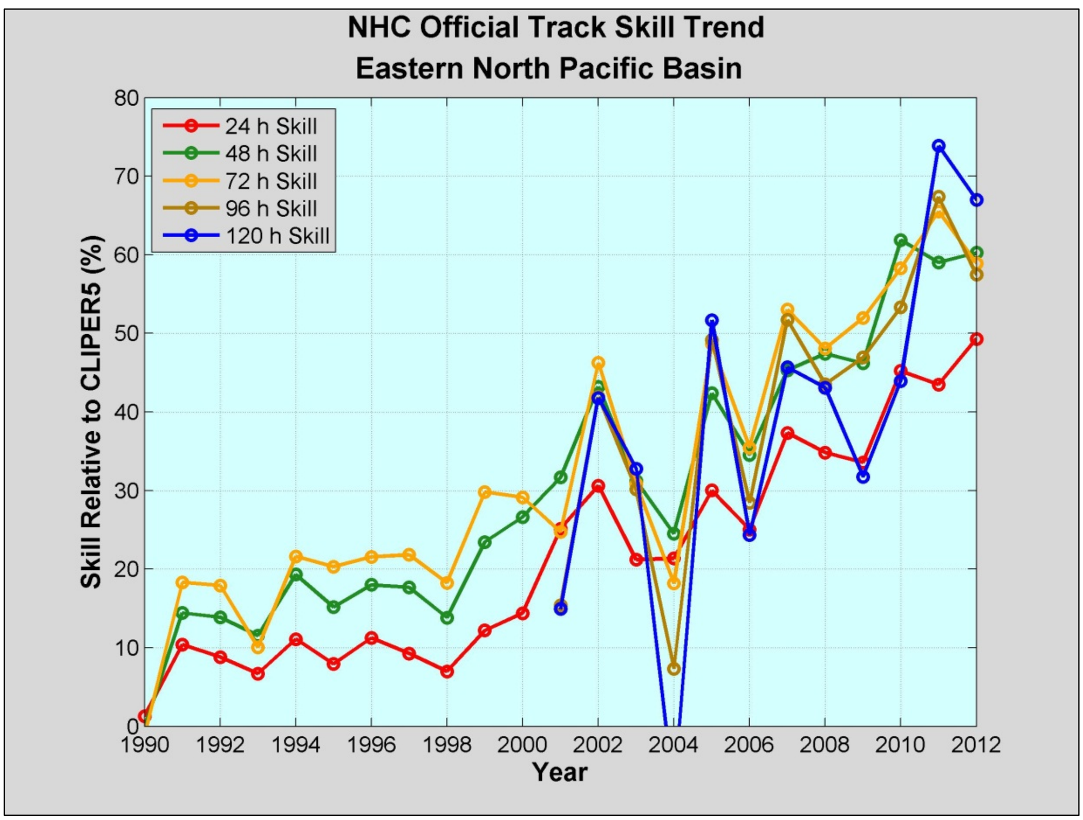


Figure 11. Recent trends in NHC official track forecast error (top) and skill (bottom) for the eastern North Pacific basin.

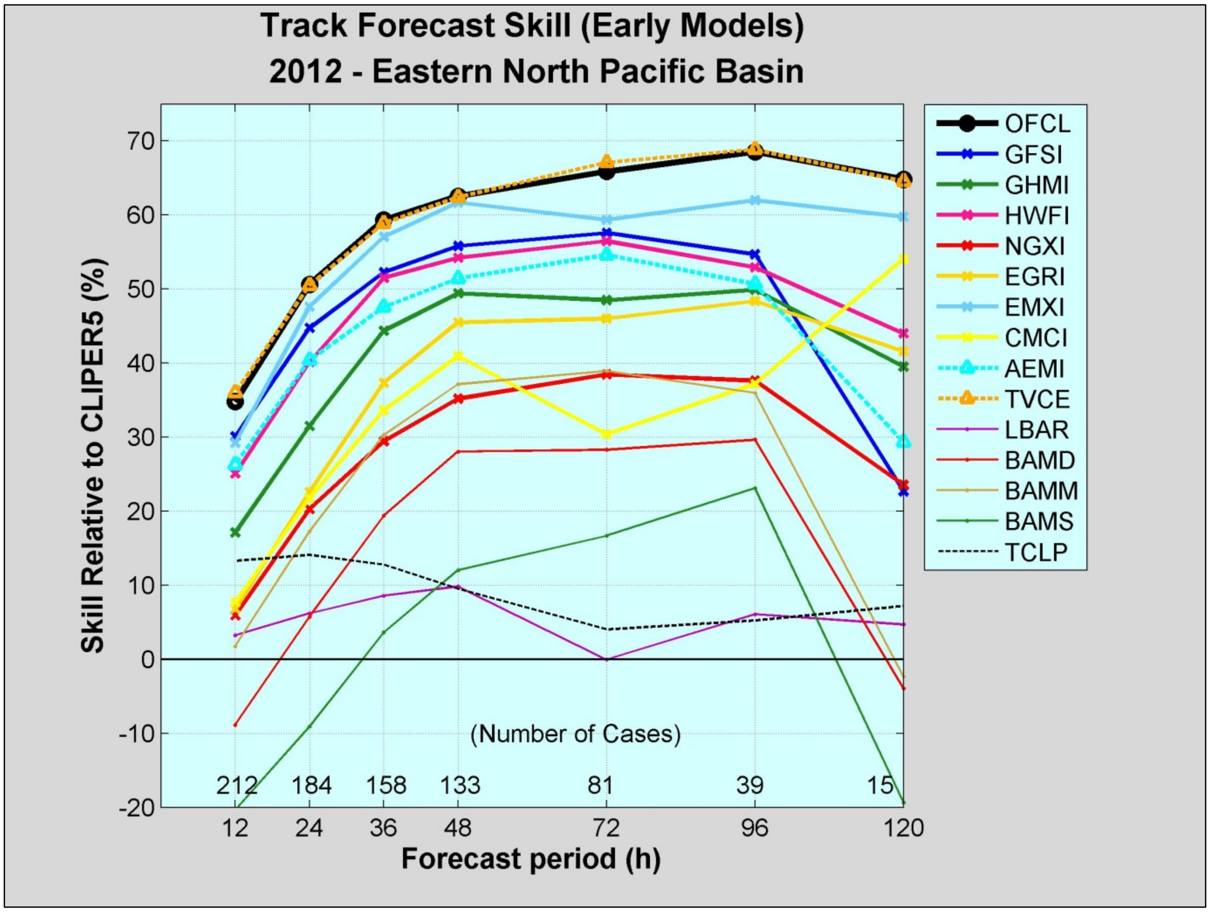


Figure 12. Homogenous comparison for selected eastern North Pacific early track models for 2012. This verification includes only those models that were available at least 2/3 of the time (see text).

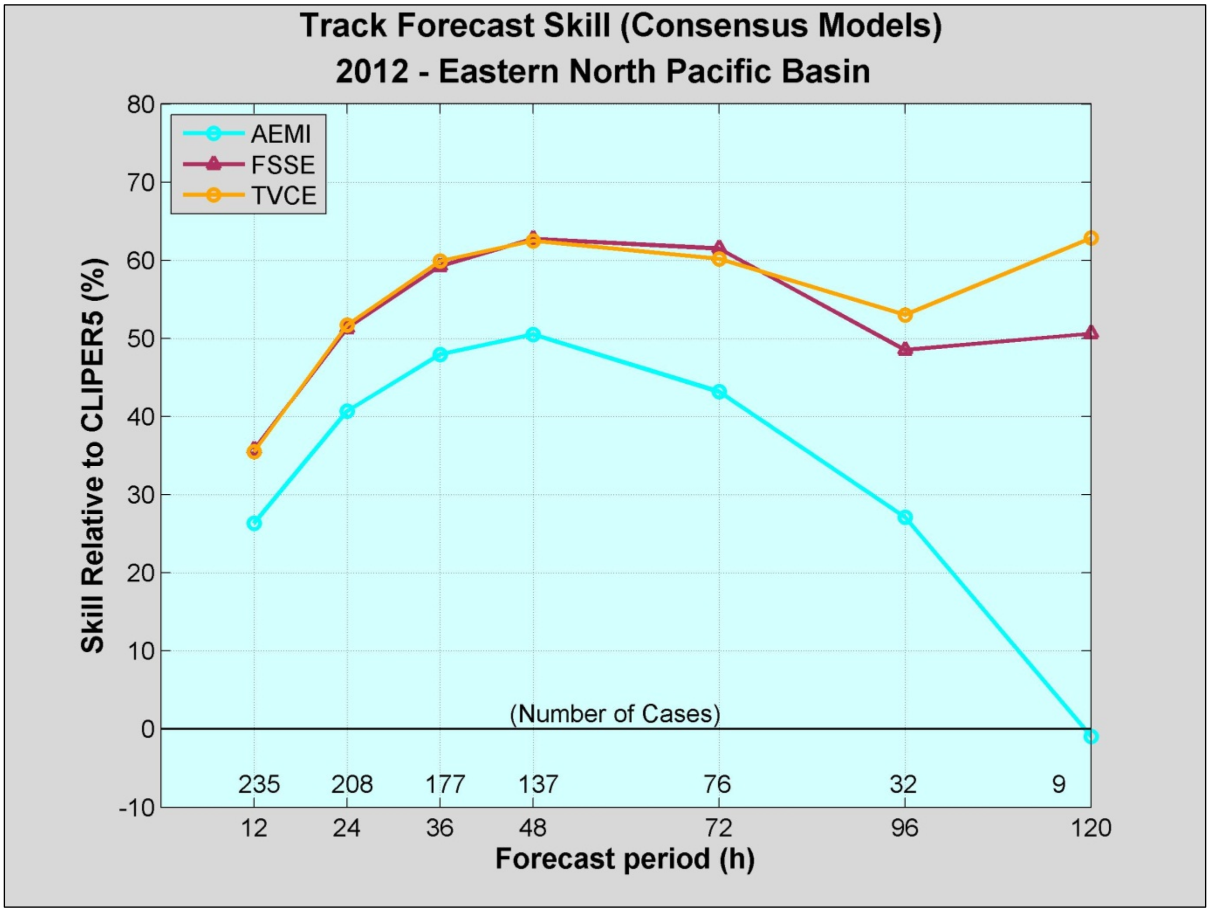


Figure 13. Homogenous comparison of the primary eastern North Pacific basin track consensus models for 2012.

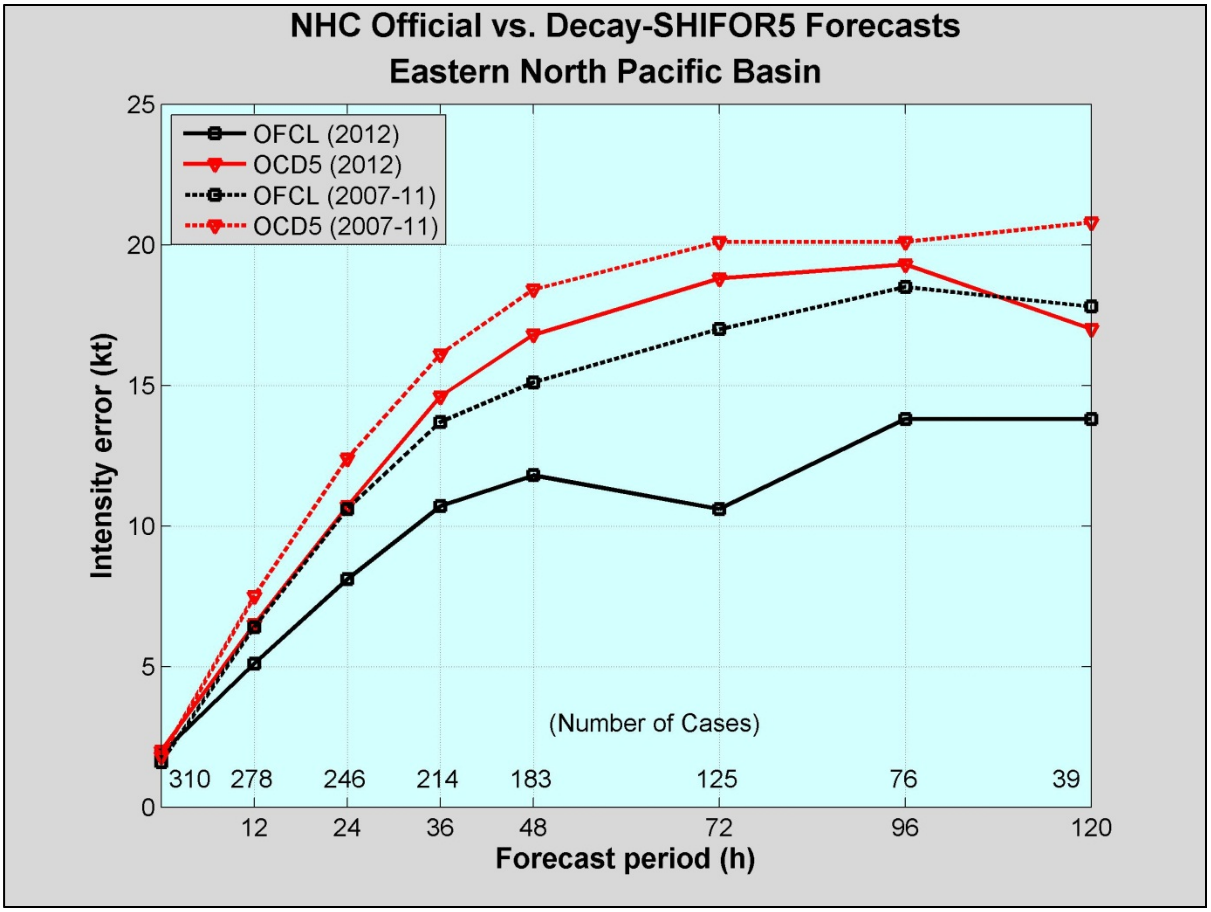
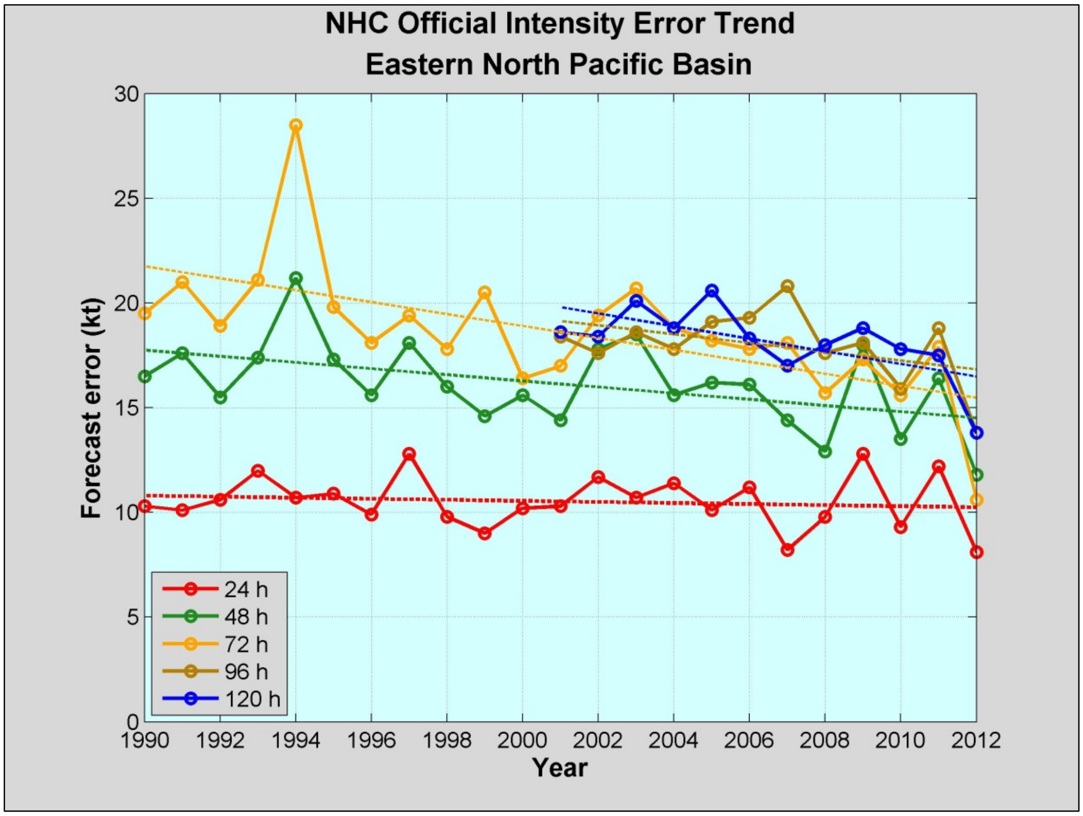


Figure 14. NHC official and Decay-SHIFOR5 (OCD5) eastern North Pacific basin average intensity errors for 2012 (solid lines) and 2007-2011 (dashed lines).



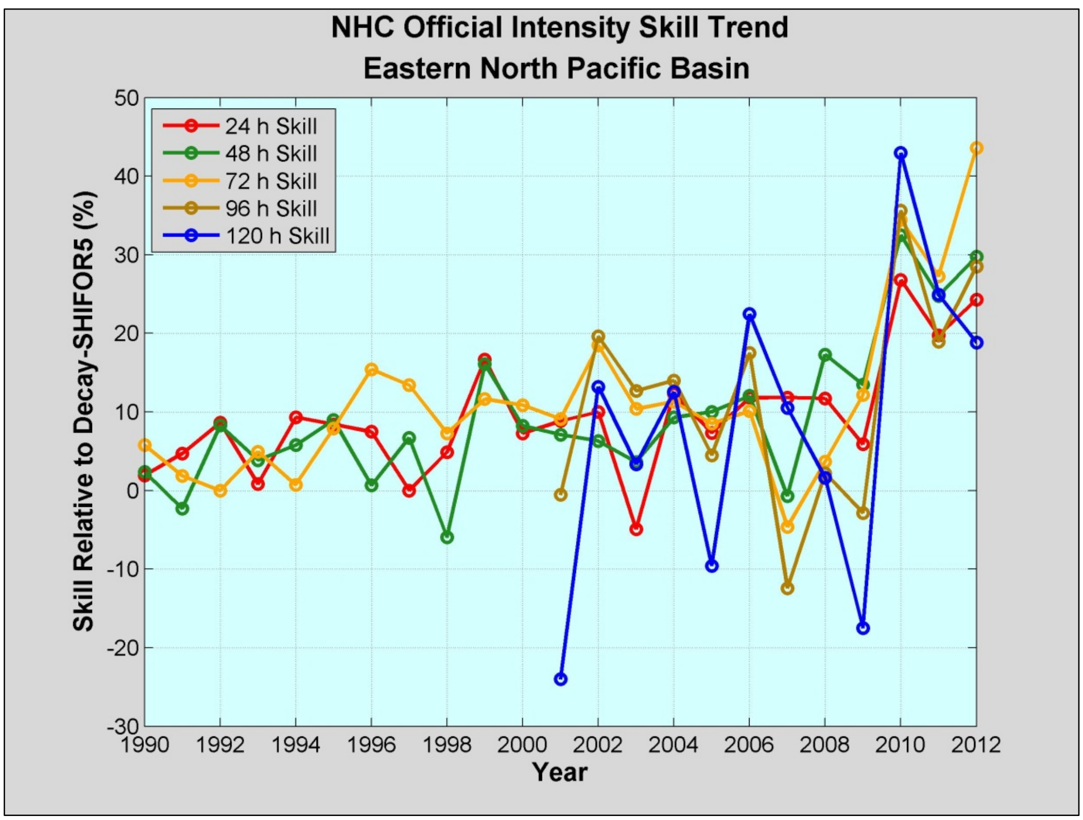


Figure 15. Recent trends in NHC official intensity forecast error (top) and skill (bottom) for the eastern North Pacific basin.

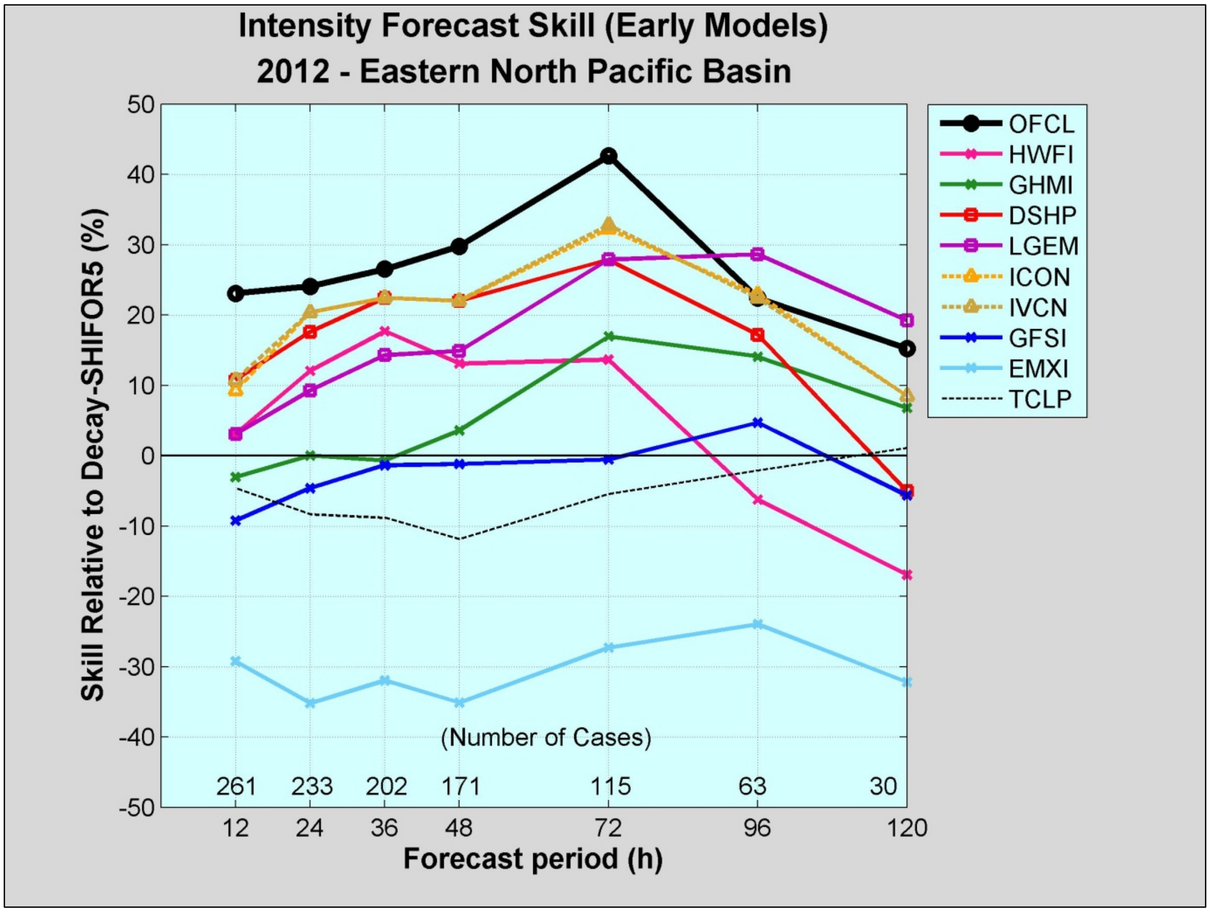


Figure 16. Homogenous comparison for selected eastern North Pacific basin early intensity guidance models for 2012.

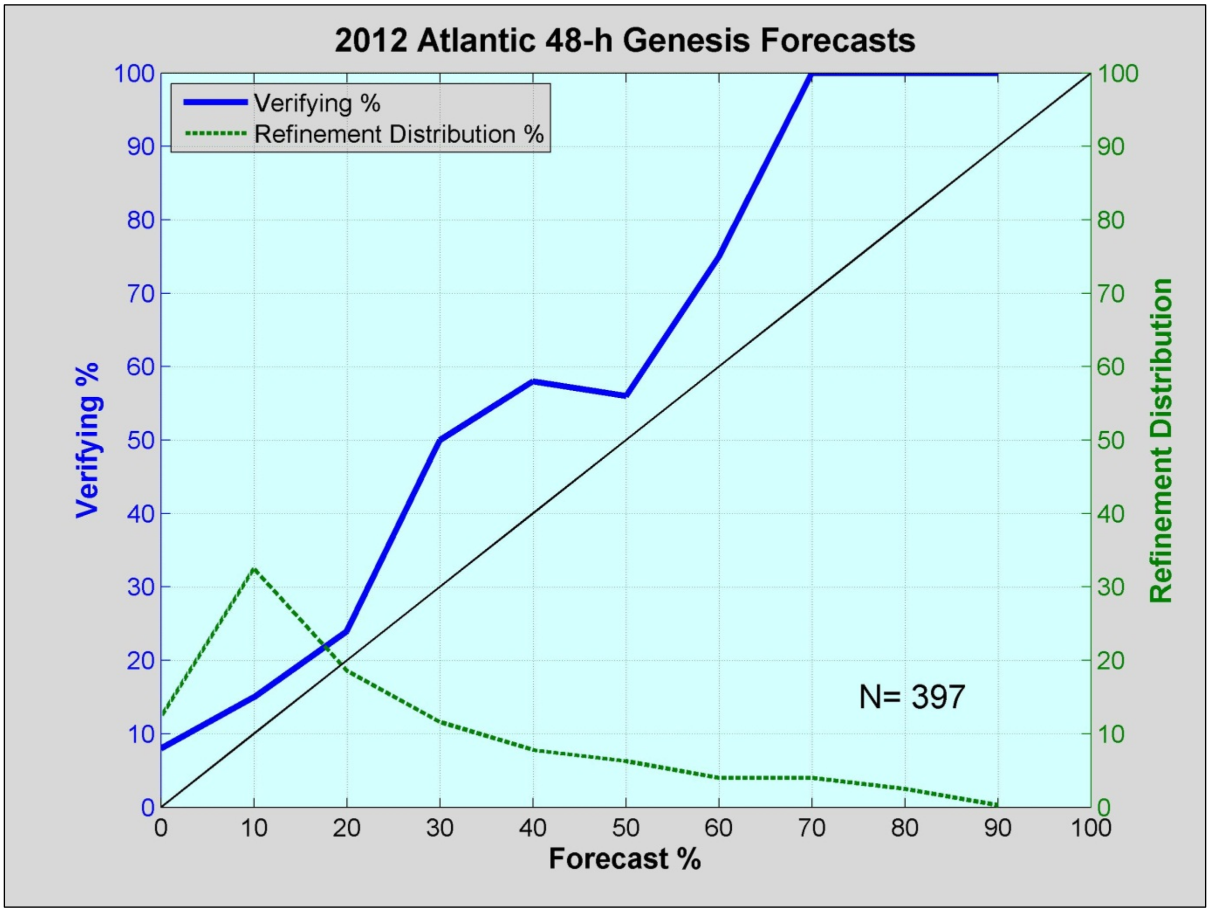


Figure 17a. Reliability diagram for Atlantic probabilistic tropical cyclogenesis forecasts for 2012. The solid blue line indicates the relationship between the forecast and verifying genesis percentages, with perfect reliability indicated by the thin diagonal black line. The dashed green line indicates how the forecasts were distributed among the possible forecast values.

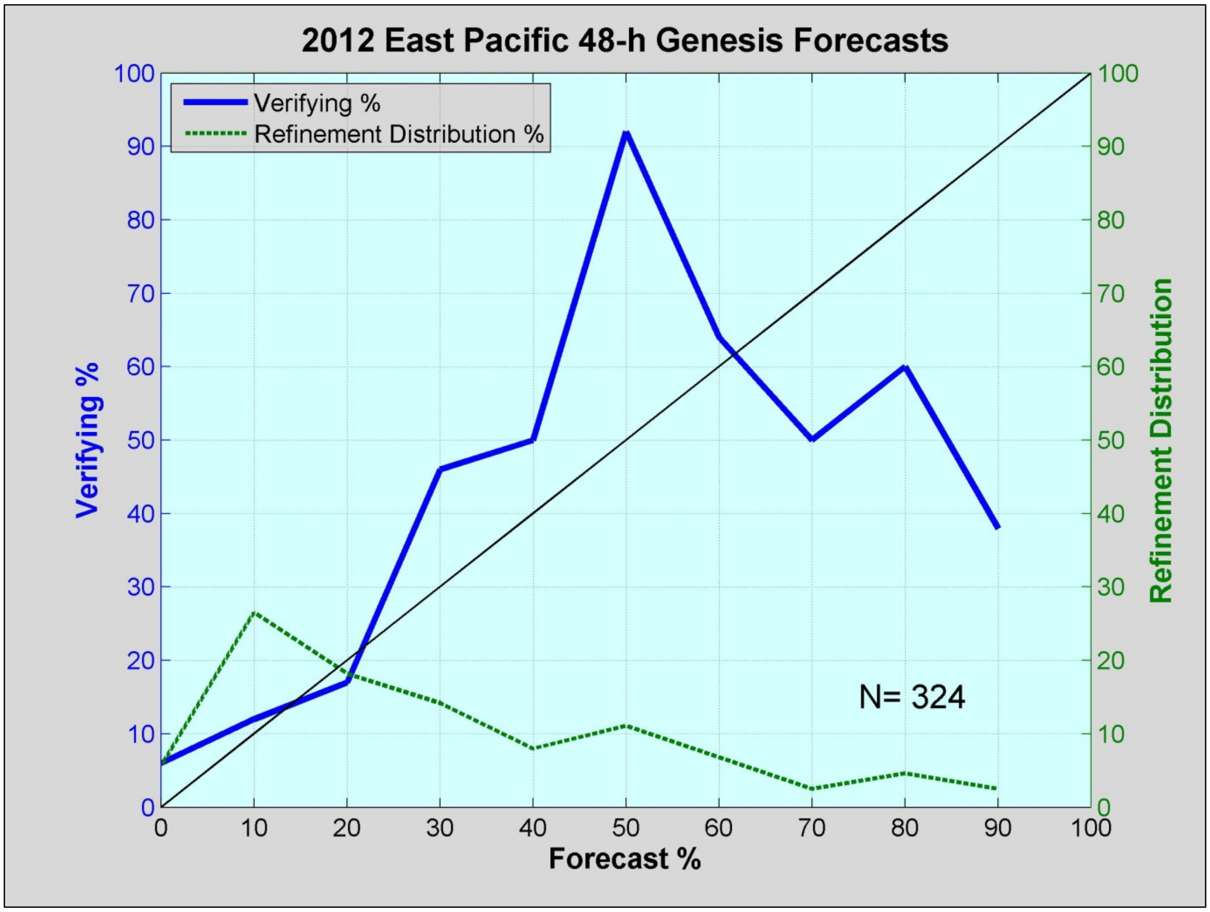
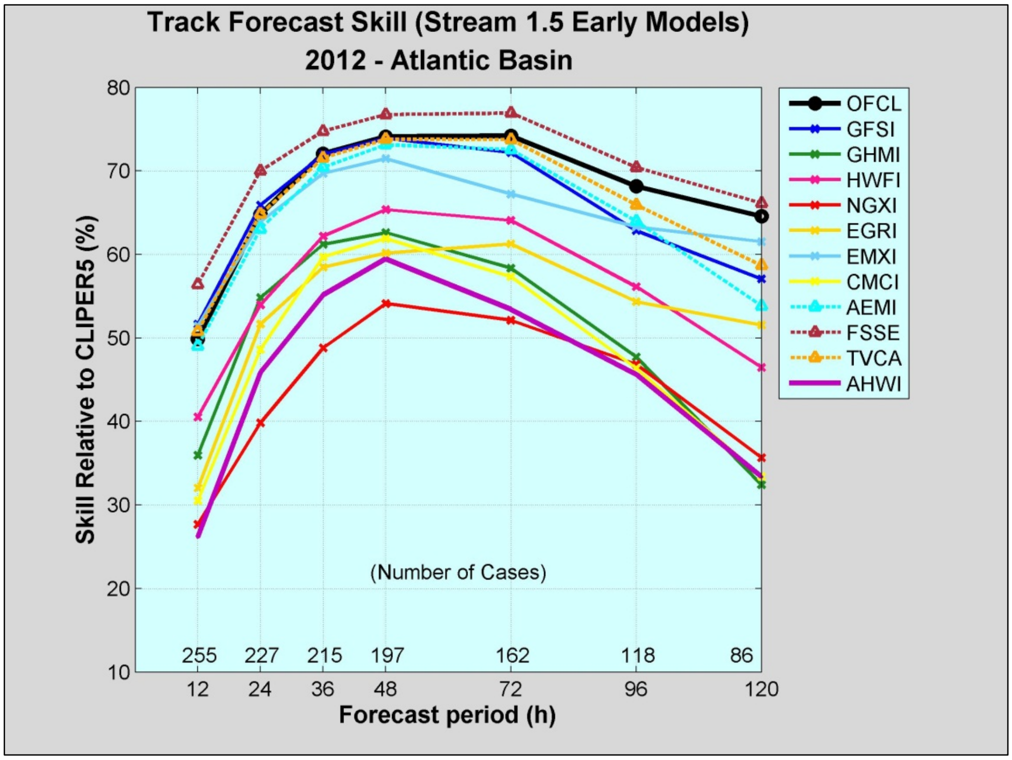


Figure 17b. As described for Fig. 17a, except for the eastern North Pacific basin.



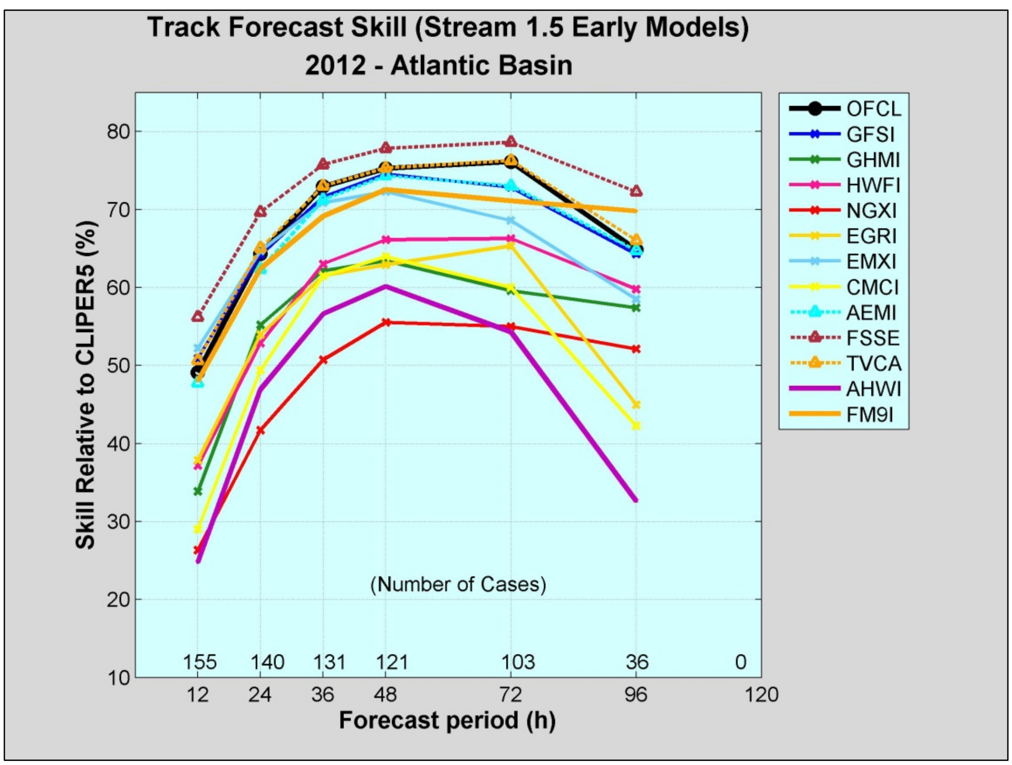


Figure 18. Homogeneous comparison of HFIP Stream 1.5 track models and selected operational models for 2012 (top), including FM9I (bottom).

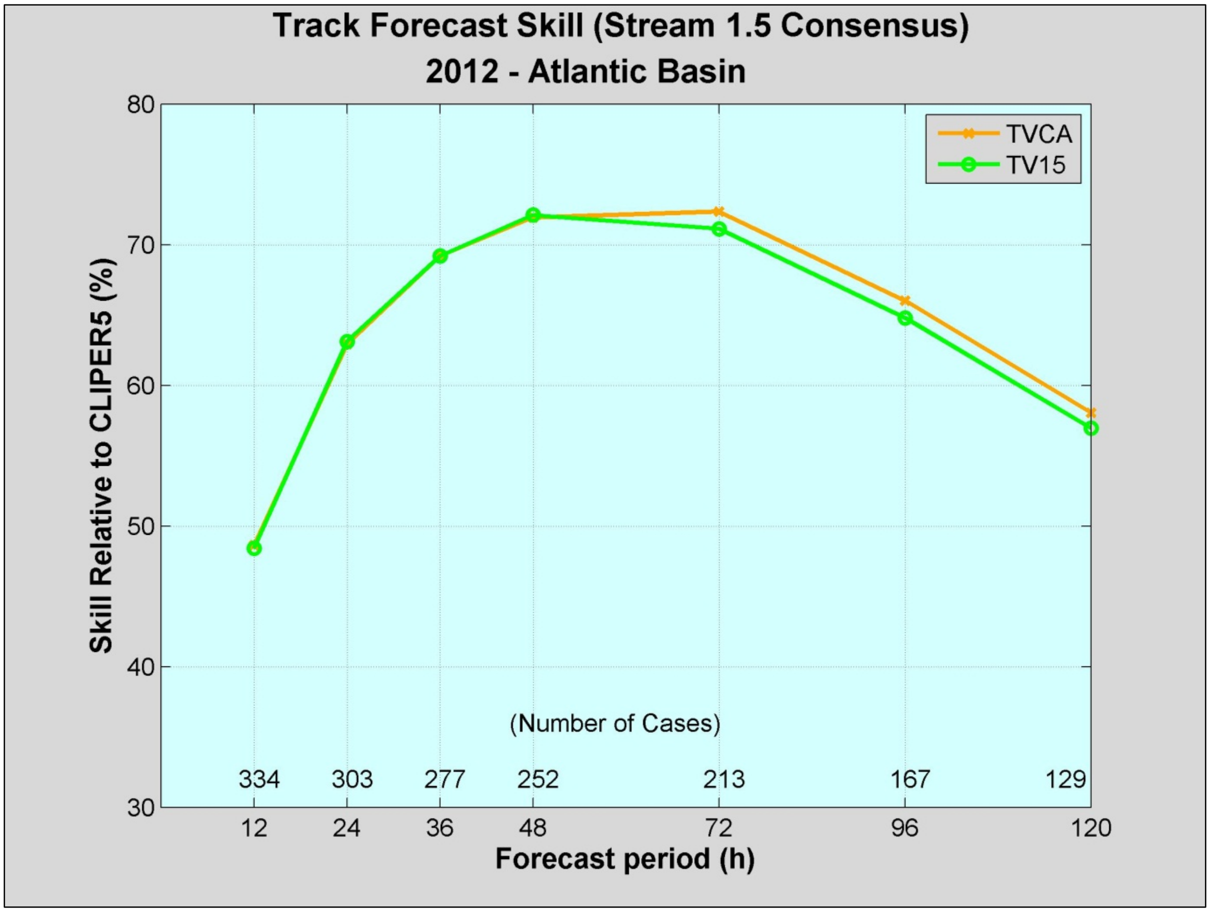
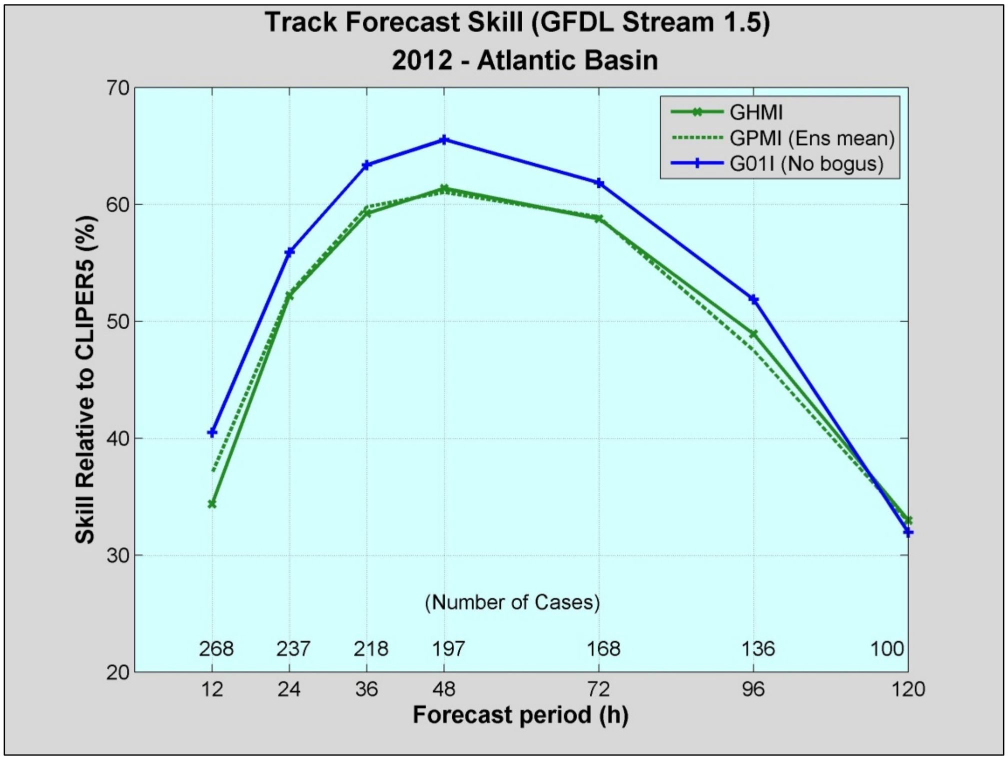


Figure 19. Impact of adding Stream 1.5 models to the variable track consensus TVCA.



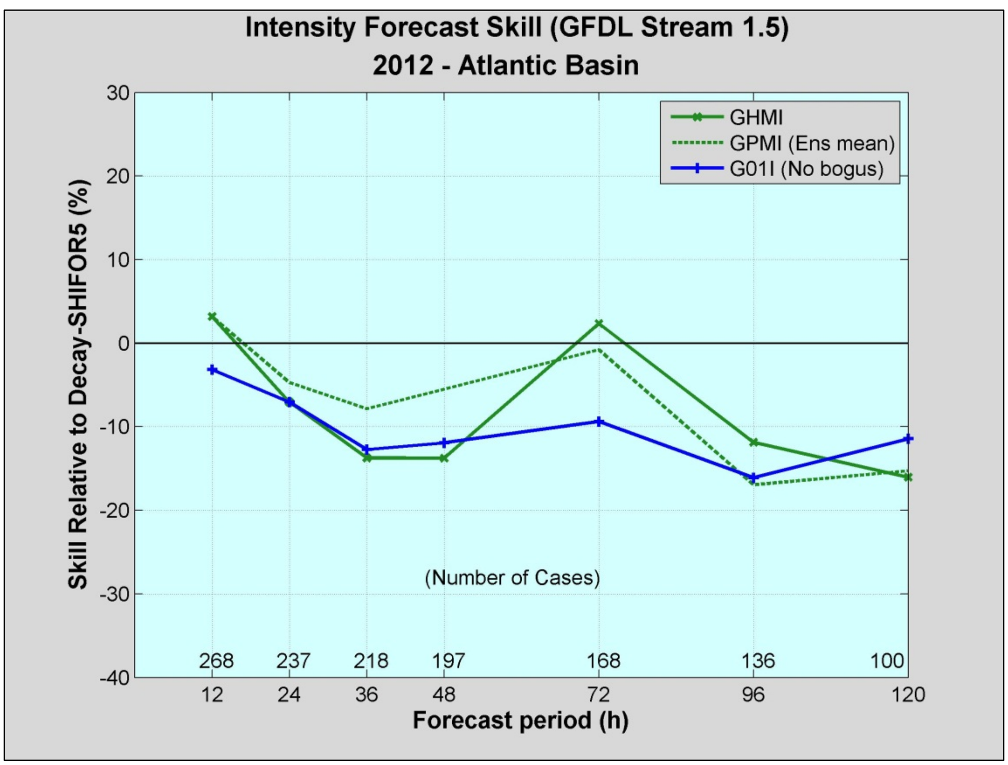


Figure 20. Homogeneous comparison of HFIP Stream 1.5 GFDL ensemble mean and GFDL unbogused ensemble member for track (top) and intensity (bottom).

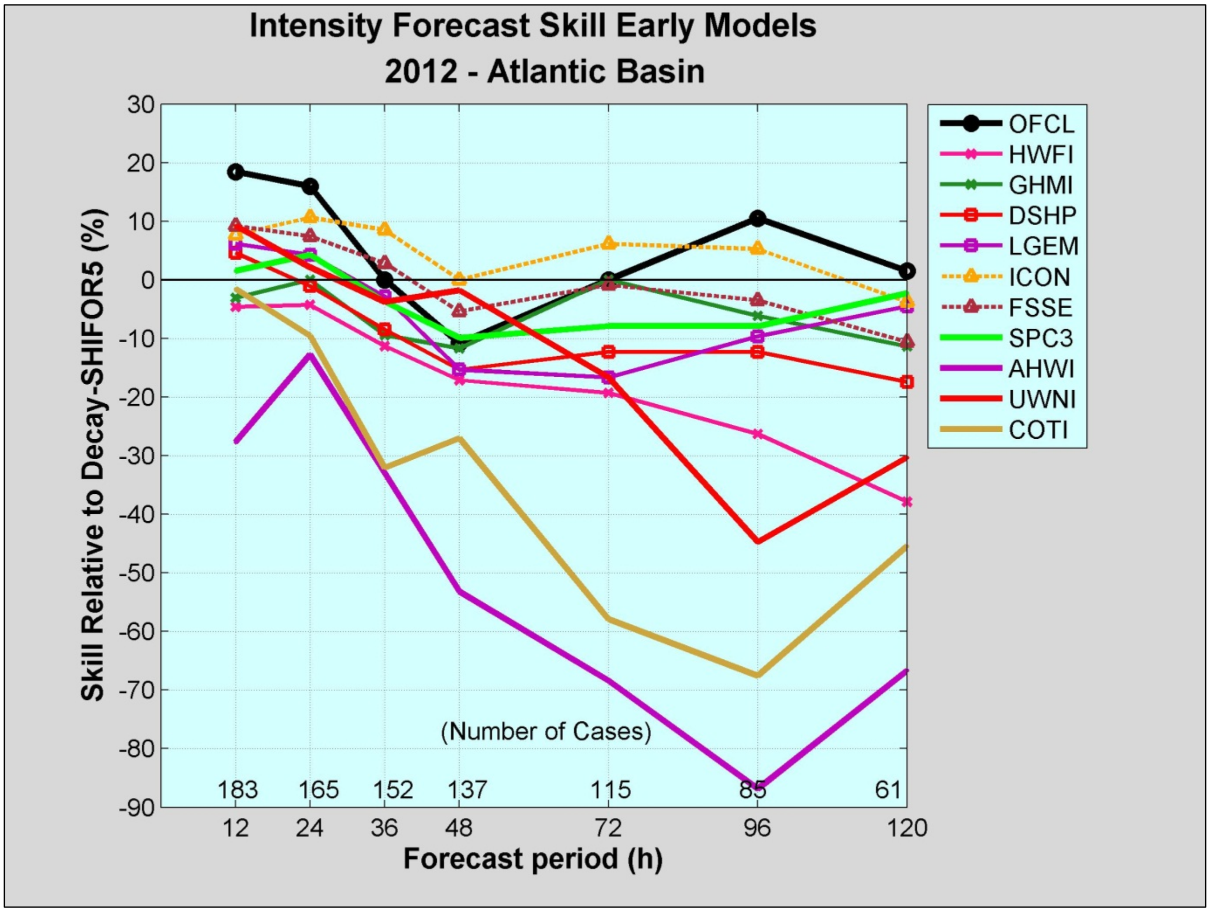


Figure 21. Homogeneous comparison of HFIP Stream 1.5 intensity models and selected operational models for 2012.

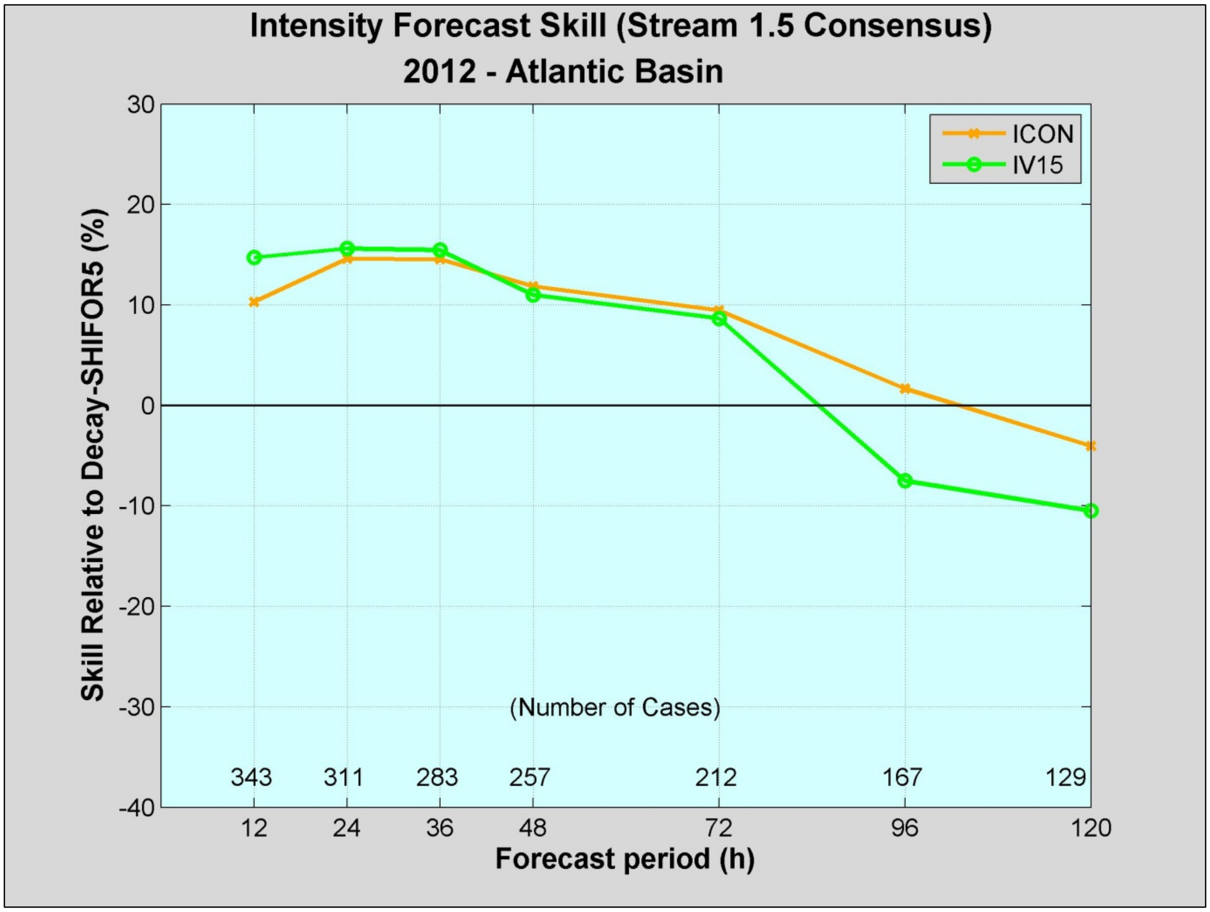


Figure 22. Impact of adding Stream 1.5 models to the fixed intensity consensus ICON.