2011 National Hurricane Center Forecast Verification Report

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

1 March 2012

ABSTRACT

The 2011 Atlantic hurricane season had above-normal activity, with 383 official forecasts issued. The NHC official track forecast errors in the Atlantic basin were lower than the previous 5-yr means at all times, except for 120 h, and set a record for accuracy at the 24-, 36-, 48-, and 72-h forecast times. 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 EMXI and GFSI exhibited the highest skill, with the GHMI and HWFI making up the second tier. The NGPI and GFNI were among the poorer-performing major dynamical models, and the EGRI was the worst model at 120 h. Among the consensus models, TVCA performed the best overall. The corrected versions of TCON, TVCA, and GUNA, however, did not perform as well as their parent models. The Government Performance and Results Act of 1993 (GPRA) track goal was met.

Official intensity errors for the Atlantic basin in 2011 were below the 5-yr means at all lead times. Decay-SHIFOR errors in 2011 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/IVCN were among the best performers at 12-48 h, with the LGEM showing similar or superior skill at 72-120 h. The dynamical models were the worst performers and had little or no skill beyond 48 h. The GPRA intensity goal was not met.

There were 258 official forecasts issued in the eastern North Pacific basin in 2011, although only 58 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 12 h. Track forecast skill was at an all-time high at 72-120 h. The official forecast outperformed all of the guidance through 36 h and was near the skill of the best aids after that. Among the guidance models with sufficient availability, EMXI and EGRI were the best individual models overall, and FSSE and AEMI performed very well at 96-120 h. There was a significant eastward bias in the official forecasts and in some of the more reliable models.

For intensity, the official forecast errors were lower than the 5-yr means at all times except 120 h. This result is particularly impressive since the 2011 Decay-SHIFOR errors were up to 30% larger than their long-term mean. The official forecasts, in general, performed as well as or better than all of the eastern Pacific guidance throughout the forecast period. The GFNI was the most skillful individual model overall, while the HWFI and GHMI struggled.

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 the 5-yr period 2007-11 indicate that these probabilistic forecasts are quite reliable in the Atlantic basin, with forecasts being particularly well calibrated in 2011. A low (under-forecast) bias, however, was present in the eastern North Pacific basin.

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

---

[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).

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.

[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".

Verification procedures here make no distinction between 6 h 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.

[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.

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 27 January 2012 for the Atlantic basin, and on 7 February 2012 for the eastern 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, which began experimentally in 2007 and became operational in 2010. Section 5 discusses the Hurricane Forecast Improvement Project (HFIP) Stream 1.5 activities in 2011. Section 6 summarizes the key findings of the 2011 verification and previews anticipated changes for 2012.

2. Atlantic Basin

a. 2011 season overview – Track

Figure 1 and Table 2 present the results of the NHC official track forecast verification for the 2011 season, along with results averaged for the previous 5-yr period, 2006-2010. In 2011, the NHC issued 383 Atlantic basin tropical cyclone forecasts[11], a number well above the average over the previous five years (274). Mean track errors ranged from 28 n mi at 12 h to 245 n mi at 120 h. It is seen that mean official track forecast errors in 2011 were smaller than the previous 5-yr mean at all forecast times except 120 h. In addition, the official track forecast errors set a record for accuracy at the 24-, 36-, 48-, and 72-h forecast times. Over the past 15 years or so, 24–72-h track forecast errors have been reduced by about 50% (Fig. 2), although it appears that track forecast skill has leveled off during the past few years. Track forecast error reductions of about 40% have occurred over the past 10 years for the 96-120 h forecast periods. Vector biases were consistently

[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.

north-northwestward in 2011 (i.e., the official forecast tended to fall to the north-northwest of the verifying position). An examination of the track errors shows that the biases were primarily along-track and fast, but there was a cross-track bias as well. Track forecast skill in 2011 ranged from 33% at 12 h to 62% 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 2011. In order to maximize the sample size for comparison with the official forecast, a guidance model had to be available at least two-thirds of the time at both 48 h and 120 h. Vector biases of the guidance models are given in Table 3b. This table shows that the official forecast had similar biases to the EMXI and the consensus models from 12-72 h, but smaller biases than most of the model

[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.

guidance beyond 72 h. Results in terms of skill are presented in Fig. 3. The figure shows that official forecast skill was slightly higher than that of the consensus models TVCA, TVCC, and FSSE. In the Atlantic basin it is not uncommon for the best of the dynamical models to beat TVCA, and such was the case in 2011 beyond 72 h. The best-performing dynamical model in 2011 was EMXI, followed by GFSI. The GHMI and HWFI made up the second tier of three-dimensional dynamical models; while NGPI, GFNI, and EGRI performed less well, with skill comparable to or even lower than the two-dimensional BAM collection. The EGRI was the worst performer at 120 h, at which time the skill was strongly negative. The official forecast beat almost all of the guidance in 2011, with only EXMI having lower errors at 96 and 120 h, and BAMM at 120 h.

A separate homogeneous verification of the primary consensus models is shown in Fig. 4. The figure shows that skill was about equal among the models through 36 h, with the exception of the GFS ensemble mean (AEMI), whose skill was about 5-10% lower at those forecast times. TVCA was the best consensus aid at 36 h and beyond, and it beat TVCE (same model but with the removal of NGPI) at all forecast times. The AEMI, which was the least skillful in the short-term, had comparable skill to the TVCA at 96-120 h. The corrected-consensus models (TVCC and CGUN) showed less skill than their respective parent models again in 2011, and because of their poor performance over the past several years these models have been discontinued. In general, it has proven difficult to use the past performance of models to derive operational corrections; the sample of forecast cases is too small, the range of meteorological conditions is too varied, and model characteristics are insufficiently stable to produce a robust developmental data sample on which to base the corrections.

The AEMI trailed its respective deterministic model (GFSI) at all time periods during 2011 (Fig. 3). While multi-model ensembles continue to provide consistently useful tropical cyclone guidance, the same cannot yet be said for single-model ensembles (although a five-year comparison of AEMI and GFSI shows roughly equivalent skill at 120 h).

Atlantic basin 48-h official track error, evaluated for all tropical cyclones[13], is a forecast measure tracked under the Government Performance and Results Act of 1993 (GPRA). In 2011, the GPRA goal was 87 n mi and the verification for this measure was 70.8 n mi.

b. 2011 season overview – Intensity

Figure 5 and Table 4 present the results of the NHC official intensity forecast verification for the 2011 season, along with results averaged for the preceding 5-yr period. Mean forecast errors in 2011 ranged from about 6 kt at 12 h to about 17 kt at 72 and 120 h. These errors were below the 5-yr means at all forecast times. Official forecasts had little bias in 2011. Decay-SHIFOR5 errors were also below their 5-yr means at all forecast times, indicating the season's storms were easier than normal to forecast. Figure 6 shows that there has been virtually no net change in error over the past 15-20 years, although forecasts during the current decade, on average, have been more skillful than those from the previous one.

Table 5a presents a homogeneous verification for the official forecast and the primary early intensity models for 2011. Intensity biases are given in Table 5b, and forecast skill is presented in Fig. 7. The intensity models were not very skillful in 2011. The best performers were the statistical-dynamical and consensus aids, but even these

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

models had negative skill by 72 h. The best individual model overall was LGEM, which hovered around the zero skill line throughout the forecast period. The dynamical models were the worst performers, all having negative skill at 48 h and beyond, with the GHMI and GFNI having skill lower than -100 % at 120 h. An inspection of the intensity biases (Table 5b) indicated that the dynamical models suffered from an extraordinary high bias − up to 80 % of the mean error. The official forecast biases, in contrast, were generally small. An evaluation over the three years 2009-11 indicates that the consensus models have been superior to all of the individual models at 12-48 h, with LGEM surpassing the consensus aids at 72 h and beyond (Fig. 8).

The 48-h official intensity error, evaluated for all tropical cyclones, is another GPRA measure for the NHC. In 2011, the GPRA goal was 13 kt and the verification for this measure was 14.4 kt. Failure to reach the GPRA goal can be attributed in part to the very poor performance of the dynamical models. The GPRA goal itself was established based on the assumption that the HWRF model would immediately lead to forecast improvements. This has not occurred, of course, and only in 2003 were seasonal mean errors as low as the current GPRA goal of 13 kt. (And as it happens, the forecast *skill* in 2003 was not particularly high.) 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 routinely 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 at 96-120 h for Ophelia, which were nearly double the long-term mean. These large errors were associated with difficulty in predicting the dissipation and reformation of this tropical cyclone. On the other hand, track errors were very low for Rina and Sean. Regarding the intensity forecasts, there was a high bias in the operational analysis of Irene's intensity during much of 25-28 August, a period when the typical surface to flight-level wind ratio did not apply. Intensity forecasts for Rina had large errors, primarily because the early forecasts were too conservative in forecasting intensification, and later forecasts held onto the high wind speeds for too long after the peak intensity. Additional discussion on forecast performance for individual storms can be found in NHC Tropical Cyclone Reports available at http://www.nhc.noaa.gov/2011atlan.shtml.

3. Eastern North Pacific Basin

a. 2011 season overview – Track

The NHC track forecast verification for the 2011 season in the eastern North Pacific, along with results averaged for the previous 5-yr period is presented in Figure 9 and Table 7. There were 258 forecasts issued for the eastern Pacific basin in 2011, although only 58 of these verified at 120 h. This level of forecast activity was about average. Mean track errors ranged from 25 n mi at 12 h to 166 n mi at 120 h, and were unanimously lower than the 5-yr means. A new record was set for forecast accuracy at 12 h. CLIPER5 errors were below their long-term means at 12-36 h, but above those values beyond 36 h. In fact, the 120-h CLIPER5 error was more than double its long-term mean. Hurricanes Irwin and Jova were the biggest contributors to the large CLIPER5 errors at the long-range forecast times. An eastward track bias in the official forecasts was noted at every forecast time. This bias was quite considerable, explaining more than 60 % of the mean error at 36 h and beyond. Greg and Jova were major contributors to these biases.

Figure 10 shows recent trends in track forecast accuracy and skill for the eastern North Pacific. Errors have been reduced by roughly 35-60% for the 24-72 h forecasts since 1990, a somewhat smaller but still substantial improvement relative to what has occurred in the Atlantic. Forecast skill in 2011 set new records at 72-120 h. The forecast skill at 24 and 48 h edged lower when compared to 2010, but these values are still the second highest on record.

Table 8a presents a homogeneous verification for the official forecast and the early track models for 2011, with vector biases of the guidance models given in Table 8b. Skill comparisons of selected models are shown in Fig. 11. Note that the sample becomes rather small (only 27 cases) by 120 h. A couple of models (GUNA and TCON) were eliminated from this evaluation because they did not meet the two-thirds availability threshold. The official forecast outperformed virtually all of the guidance for the first 36 h, at which time the consensus aid TVCE was the best model. The EMXI had the lowest errors at 48-96 h. The EGRI, CMCI, AEMI, and FSSE showed increased skill at the longer ranges and fared the best among the guidance at 120 h. The GFSI had considerably less skill than its ensemble mean and is in the middle of the pack with the NGPI and GHMI. The GFNI and HWFI were poorer performers and even lagged the relatively simple BAMS and BAMM at the longer-range forecast times.

A separate verification of the primary consensus aids is given in Figure 12. TVCE performed best at 12-72 h, but AEMI and FSSE had the highest skill at 96-120 h. An evaluation over the three years 2009-11 (not shown) indicates that the superior performance of AEMI over the GFSI in 2011 was not an anomaly, which is in contrast to the Atlantic where GFSI beats AEMI at most forecast times. The corrected consensus model TVCC was the worst consensus aid and had 15-20 % less skill than its parent model.

b. 2011 season overview – Intensity

Figure 13 and Table 9 present the results of the NHC eastern North Pacific intensity forecast verification for the 2011 season, along with results averaged for the preceding 5-yr period. Mean forecast errors were 7 kt at 12 h and increased to 19 kt by 96 h. The errors were lower than the 5-yr means, by up to 16%, at all times except 120 h. The Decay-SHIFOR5 forecast errors were substantially higher than their 5-yr means; this implies that forecast difficulty in 2011 was higher than normal. A review of error and skill trends (Fig. 14) indicates that the intensity errors have decreased slightly over the past 15-20 years at the 48- and 72-h forecast times. Forecast skill has decreased in 2011, but it was still quite high when compared to historical values. Intensity forecast biases in 2011 were small through 48 h and modestly positive thereafter.

Figure 15 and Table 10a present a homogeneous verification for the primary early intensity models for 2011. Forecast biases are given in Table 10b. The official forecasts, in general, were about as skillful as the best models throughout the forecast period. The GFNI was the best individual model, while the other dynamical models (GHMI and HWFI) performed the worst and had negative skill at 96 and 120 h. The statistical-dynamical guidance (DSHP and LGEM) and the intensity consensus models (ICON, IVCN, and FSSE) were competitive with one another, all having positive skill between 10 and 35 % throughout the forecast period.

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/2011epac.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 was 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 2011 are given in Table 12 and illustrated in Fig. 16. In the Atlantic basin, the forecasts were well calibrated throughout the entire probability range. This result was much improved from the past few years, especially at the mid-range probabilities. In the eastern Pacific, the forecasts were reliable at the low and high probabilities, but an under-forecast bias existed in the middle 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.

Combined results for the five-year period 2007-11 are given in Table 13. For the 5-yr sample, the Atlantic basin forecasts were well calibrated overall. Results for the eastern North Pacific were not quite as good, with a general under-forecast bias in the middle probabilities. Even so, the forecasters were clearly able to distinguish gradations in genesis likelihood (evidenced by the nearly monotonic increase of the verifying percentage with forecast percentage).

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 2011 under Stream 1.5; these are listed in Table 14. 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. Some models (e.g., COTI) were admitted into Stream 1.5 because in retrospective testing they contributed positively to the consensus, even though individually the model performance might not have been sufficient to qualify for direct use by forecasters. Two HFIP Stream 1.5 consensus aids were constructed: the track consensus TV15 comprised the operational models GFSI, EGRI, GHMI, GFNI, HWFI, EMXI and the Stream 1.5 models AHWI and FIMI, while the intensity consensus IV15 comprised the operational models DSHP, LGEM, GHMI, HWFI and the Stream 1.5 models AHQI, COTI, A4QI, and UWQI. It should be noted that the standard interpolator, rather than the GFDL version, was inadvertently applied to some of these models operationally in 2011; the results shown here were based on aids regenerated post-storm using the interpolators as indicated in Table 14.

Figure 17 presents a homogeneous verification of the primary operational models against the Stream 1.5 track models (excluding the GFDL ensemble because of its limited availability). The figure shows that in 2011 the FIMI was competitive with the top-tier operational models, while the AHWI and H3GI performed less well. Figure 18 shows that there was a small positive impact from adding the Stream 1.5 models to the track consensus.

Intensity results are shown in Fig. 19, for a sample that excludes the GFDL ensemble and the PSU Doppler runs due to limited availability. The Stream 1.5 models COTI and SPC3 generally outperformed the operational models. The strong performance of SPC3 is not surprising, given that it represents an intelligent consensus of the already top-tier statistical models LGEM and DSHP. The strong performance of COTI largely derives from less aggressive forecasts of the intensity of Irene, and it is not clear whether these results will prove to be representative in a season with more rapidly intensifying storms. The Stream 1.5 models also contributed positively to the intensity consensus (Fig. 20), although the differences in terms of error were all less than 1 kt.

The Stream 1.5 activity in 2011 was highly successful. The number of participating models was greatly enhanced over 2010, when only two models were presented to the forecasters, and as noted above some of the Stream 1.5 models performed very well. Forecasters were able to gain experience with these new aids, which should greatly enhance their impact on operations in 2012.

6. Looking Ahead to 2012

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 2012 for the Atlantic and eastern North Pacific basins (based on error distributions for 2007-11) are in Table 15. In the Atlantic basin, the cone circles will be slightly smaller than they were last year, with the biggest decrease at 96 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., GUNA), 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 2012 is unchanged from 2011 and is given in Table 16.

Acknowledgments:

The authors gratefully acknowledge Chris Sisko of NHC, keeper 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

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
NGPI	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
AEMI	Previous cycle AEMN, adjusted	Consensus	E	Trk, Int

ID	Name/Description	Type	Timeliness (E/L)	Parameters forecast
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

Table 2. Homogenous comparison of official and CLIPER5 track forecast errors in the Atlantic basin for the 2011 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
2011 mean OFCL error (n mi)	28.2	43.4	57.1	70.8	109.7	166.6	244.7
2011 mean CLIPER5 error (n mi)	42.3	82.5	133.6	185.8	278.9	360.0	411.7
2011 mean OFCL skill relative to CLIPER5 (%)	33.3	47.4	57.2	61.9	60.7	53.7	40.5
2011 mean OFCL bias vector (°/n mi)	332/3	347/8	347/11	350/17	336/28	335/38	332/57
2011 number of cases	339	297	260	226	176	140	113
2006-2010 mean OFCL error (n mi)	30.8	50.2	69.4	89.2	133.2	174.2	214.8
2006-2010 mean CLIPER5 error (n mi)	47.5	97.7	155.3	216.9	323.3	402.2	476.1
2006-2010 mean OFCL skill relative to CLIPER5 (%)	35.1	48.6	55.3	58.9	58.8	56.7	54.9
2006-2010 mean OFCL bias vector (°/n mi)	322/3	315/6	312/9	319/11	300/6	098/4	104/2
2006-2010 number of cases	1231	1089	954	839	662	503	387
2011 OFCL error relative to 2006-2010 mean (%)	-8.4	-13.5	-17.7	-20.6	-17.6	-4.4	13.9
2011 CLIPER5 error relative to 2005-2009 mean (%)	-10.9	-15.6	-14.0.	-14.3	-14.0	-10.5	-13.5

Table 3a. Homogenous comparison of Atlantic basin early track guidance model errors (n mi) for 2011. 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	25.8	40.0	53.8	68.5	110.9	167.6	239.3
OCD5	37.8	73.2	125.9	184.3	292.6	347.5	334.4
GFSI	27.2	44.3	59.1	75.6	128.2	175.9	245.4
GHMI	29.6	50.7	74.8	102.1	152.4	217.5	304.6
HWFI	31.6	53.1	72.2	92.1	155.8	234.3	310.0
GFNI	33.9	57.8	85.1	113.7	169.1	254.0	341.8
NGPI	34.7	60.1	89.2	117.1	194.4	303.1	399.3
EGRI	33.1	49.7	67.8	92.8	172.0	309.3	453.0
EMXI	27.1	43.4	57.2	71.5	112.0	**156.9**	**227.8**
CMCI	34.8	57.9	87.7	121.5	186.2	285.3	353.0
AEMI	27.5	47.3	68.9	92.9	144.8	179.6	275.8
FSSE	28.0	43.8	58.4	77.0	135.8	210.4	310.6
TCON	26.1	41.5	57.3	74.4	122.4	192.1	279.4
TVCA	**25.7**	40.5	54.5	70.0	111.6	171.5	250.3
TVCC	25.8	40.3	55.3	72.6	118.5	184.6	270.5
LBAR	34.7	63.1	97.4	140.2	236.6	307.5	403.6
BAMD	43.7	76.4	107.2	130.0	202.7	254.5	367.6
BAMM	34.5	55.4	79.3	104.3	174.3	219.7	**234.6**
BAMS	43.4	78.4	118.5	161.6	257.2	285.6	254.4
# Cases	212	182	169	145	114	83	52

Table 3b. Homogenous comparison of Atlantic basin early track guidance model bias vectors (°/n mi) for 2011.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	343/006	347/014	345/019	347/025	348/046	346/056	265/017
OCD5	314/001	030/005	244/003	226/020	010/023	038/088	048/094
GFSI	013/007	019/013	021/018	027/026	010/046	006/045	069/076
GHMI	004/002	315/003	303/029	304/048	309/081	330/101	344/090
HWFI	021/007	034/015	023/018	017/024	012/060	020/093	075/081
GFNI	289/008	305/018	309/030	310/039	344/066	019/112	056/146
NGPI	326/016	330/030	326/042	321/050	349/079	020/132	056/137
EGRI	308/010	293/015	278/022	261/030	282/034	293/032	206/091
EMXI	357/007	353/016	344/019	331/022	335/039	311/041	239/061
CMCI	327/012	321/020	319/030	311/042	318/072	334/112	007/080
AEMI	341/007	335/014	326/019	320/023	335/039	011/034	152/042
FSSE	336/007	338/011	341/016	347/022	351/063	350/089	047/059
TCON	341/007	339/014	327/021	322/027	340/051	001/071	058/049
TVCA	340/006	337/012	326/018	320/024	337/047	355/061	058/032
TVCC	341/003	344/007	323/012	310/020	332/043	002/042	098/041
LBAR	039/008	012/023	349/034	339/046	001/086	038/143	063/252
BAMD	041/004	044/010	008/005	003/005	057/009	089/056	091/167
BAMM	261/010	250/017	237/030	222/042	216/041	203/010	088/061
BAMS	278/029	269/052	255/074	242/092	237/099	268/050	002/013
# Cases	212	182	169	145	114	83	52

Table 4. Homogenous comparison of official and Decay-SHIFOR5 intensity forecast errors in the Atlantic basin for the 2011 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
2011 mean OFCL error (kt)	6.3	9.7	12.0	14.4	16.6	16.2	16.5
2011 mean Decay-SHIFOR5 error (kt)	6.9	10.4	12.2	13.5	16.7	17.3	13.9
2011 mean OFCL skill relative to Decay-SHIFOR5 (%)	8.7	6.7	1.6	-6.7	0.6	-8.0	-18.7
2011 OFCL bias (kt)	-0.5	0.3	0.4	0.7	0.2	0.2	-1.0
2011 number of cases	339	297	260	226	176	140	113
2006-10 mean OFCL error (kt)	7.3	11.0	13.2	15.1	17.2	17.9	18.7
2006-10 mean Decay-SHIFOR5 error (kt)	8.5	12.3	15.5	17.9	20.2	21.9	21.7
2006-10 mean OFCL skill relative to Decay-SHIFOR5 (%)	14.1	10.6	14.8	15.6	14.9	18.3	13.8
2006-10 OFCL bias (kt)	0.2	1.1	1.5	2.1	2.7	2.4	2.2
2006-10 number of cases	1231	1089	954	839	662	503	387
2011 OFCL error relative to 2006-10 mean (%)	13.7	11.8	9.1	4.6	3.5	9.5	11.8
2011 Decay-SHIFOR5 error relative to 2006-10 mean (%)	18.8	15.4	20.6	24.6	17.3	21.0	35.9

Table 5a. Homogenous comparison of selected Atlantic basin early intensity guidance model errors (kt) for 2011. Errors smaller than the NHC official forecast are shown in boldface.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	6.6	10.2	12.7	15.4	17.0	15.8	16.9
OCD5	7.2	10.9	12.8	**13.5**	**16.1**	**13.1**	**12.9**
HWFI	6.8	**9.6**	**11.9**	**14.6**	19.7	20.3	23.0
GHMI	7.0	**10.1**	12.9	15.8	23.2	25.1	26.0
GFNI	7.4	10.3	13.0	16.0	21.7	24.1	27.0
DSHP	7.1	10.3	**12.4**	**14.3**	**16.6**	**15.5**	**15.6**
LGEM	7.2	10.6	12.8	**15.0**	**16.1**	**14.1**	**14.9**
ICON	**6.5**	**9.1**	**10.9**	**12.8**	**16.2**	**15.1**	**14.5**
IVCN	**6.5**	**8.9**	**10.7**	**12.5**	**16.5**	15.8	**14.9**
FSSE	6.8	**9.7**	**11.5**	**12.7**	**15.6**	**15.7**	**15.7**
# Cases	278	249	214	186	148	118	97

Table 5b. Homogenous comparison of selected Atlantic basin early intensity guidance model biases (kt) for 2011. Biases smaller than the NHC official forecast are shown in boldface.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	-0.1	1.4	1.6	1.2	0.7	-0.3	-2.5
OCD5	-0.2	1.6	2.0	1.7	2.9	2.2	3.6
HWFI	-1.0	**-0.5**	**1.4**	3.4	7.7	6.7	6.5
GHMI	0.3	1.5	4.4	7.8	16.0	18.6	20.8
GFNI	1.1	1.8	3.9	6.3	12.6	16.8	21.9
DSHP	-0.6	**0.5**	**0.9**	**0.2**	**0.1**	-1.7	-7.6
LGEM	-0.8	**-0.7**	**-1.3**	-2.6	-3.6	-4.4	-8.3
ICON	-0.3	**0.5**	1.6	2.4	5.3	5.1	3.0
IVCN	0.1	**0.9**	2.2	3.3	6.9	7.5	6.6
FSSE	-1.8	**-1.0**	**-0.5**	**-0.6**	1.5	2.3	**-0.6**
# Cases	278	249	214	186	148	118	97

Table 6. Official Atlantic track and intensity forecast verifications (OFCL) for 2011 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: AL012011 ARLENE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	10	11.9	11.3	10	1.5	1.5
012	8	28.8	32.3	8	4.4	3.9
024	6	31.7	58.1	6	6.7	10.2
036	4	20.6	107.3	4	5.0	2.5
048	2	34.2	154.5	2	5.0	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: AL022011 BRET

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	19	5.9	6.1	19	2.1	2.1
012	17	16.1	27.3	17	3.5	3.9
024	15	23.1	49.6	15	6.0	8.5
036	13	32.8	74.2	13	5.4	12.8
048	11	42.7	109.4	11	3.6	16.1
072	7	66.1	189.9	7	5.7	21.3
096	3	133.4	408.7	3	6.7	19.3
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: AL032011 CINDY

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	9	11.0	11.0	9	4.4	5.6
012	7	33.5	48.3	7	6.4	5.7
024	5	67.0	95.8	5	6.0	11.0
036	3	71.7	141.5	3	5.0	10.3
048	1	97.4	268.1	1	5.0	7.0
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: AL042011 DON

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	10	7.0	8.9	10	0.0	0.5
012	8	22.1	31.8	8	4.4	4.6
024	6	29.4	77.3	6	5.0	7.8
036	4	46.1	164.6	4	7.5	10.5
048	2	56.5	257.3	2	15.0	15.5
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	-999.9

Verification statistics for: AL052011 EMILY

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	14	16.6	18.5	14	2.1	2.1
012	10	51.7	59.0	10	2.0	3.8
024	7	86.5	88.3	7	4.3	6.7
036	5	91.6	87.8	5	7.0	7.6
048	3	127.8	154.7	3	6.7	6.7
072	3	49.1	370.6	3	21.7	19.7
096	3	116.1	278.6	3	21.7	4.0
120	2	134.7	414.1	2	27.5	38.5

Verification statistics for: AL062011 FRANKLIN

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	4	18.6	17.1	4	0.0	0.0
012	2	47.2	52.4	2	2.5	2.5
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: AL072011 GERT

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	10	15.7	17.2	10	4.5	4.5
012	8	27.8	53.9	8	8.1	8.1
024	6	47.8	105.1	6	9.2	10.8
036	4	61.0	213.0	4	5.0	6.5
048	2	95.8	284.9	2	2.5	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: AL082011 HARVEY

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	15	17.0	17.9	15	2.3	2.7
012	13	37.0	48.5	13	7.3	6.8
024	11	49.5	63.7	11	9.1	9.2
036	8	76.8	107.4	8	10.6	9.5
048	7	100.1	165.0	7	1.4	3.6
072	3	146.2	214.4	3	6.7	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: AL092011 IRENE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	32	3.5	6.0	32	4.2	4.1
012	30	23.8	38.1	30	10.2	9.1
024	28	38.0	65.5	28	13.9	11.9
036	26	53.0	92.1	26	18.3	13.4
048	24	68.4	112.7	24	21.5	15.4
072	20	101.2	174.1	20	28.5	17.8
096	16	132.2	222.1	16	25.9	17.6
120	12	237.1	245.5	12	22.9	12.7

Verification statistics for: AL102011 TEN

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	7	16.7	17.1	7	0.0	0.0
012	5	24.6	56.6	5	1.0	2.6
024	3	43.5	127.9	3	6.7	6.7
036	1	59.3	167.2	1	15.0	14.0
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: AL112011 JOSE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	3	5.3	5.3	3	3.3	6.7
012	1	10.1	93.6	1	5.0	6.0
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: AL122011 KATIA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	49	8.3	8.6	49	3.0	2.9
012	47	20.7	37.0	47	6.3	7.5
024	45	31.6	72.3	45	9.8	10.5
036	43	43.4	116.7	43	12.3	12.2
048	41	48.7	161.2	41	15.0	13.4
072	37	65.3	208.5	37	15.0	14.8
096	33	100.2	215.0	33	14.5	10.6
120	29	164.8	226.4	29	13.6	9.9

Verification statistics for: AL132011 LEE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	13	7.6	8.9	13	2.7	1.9
012	11	32.9	38.9	11	3.6	3.1
024	9	41.8	73.7	9	8.9	2.9
036	7	43.8	102.7	7	10.0	7.7
048	5	37.3	98.7	5	14.0	11.8
072	1	49.4	69.5	1	15.0	2.0
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: AL142011 MARIA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	37	14.3	14.1	37	2.2	2.3
012	33	35.7	51.1	33	3.5	5.1
024	29	61.8	125.4	29	4.5	7.6
036	27	91.9	221.9	27	5.9	8.1
048	25	115.8	300.5	25	8.4	7.8
072	22	156.4	356.1	22	8.6	14.5
096	21	210.1	397.1	21	6.9	15.7
120	17	268.8	442.1	17	10.9	21.2

Verification statistics for: AL152011 NATE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	17	6.9	6.6	17	2.6	2.9
012	15	23.8	30.4	15	10.7	11.7
024	13	33.7	53.0	13	14.2	14.2
036	11	57.7	91.0	11	17.3	13.7
048	9	88.4	139.5	9	26.7	18.3
072	5	188.0	252.2	5	25.0	17.8
096	1	193.7	409.3	1	40.0	18.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: AL162011 OPHELIA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	41	17.6	18.4	41	2.2	2.4
012	37	30.1	45.2	37	6.6	8.9
024	33	40.4	85.9	33	10.8	13.6
036	29	46.6	141.5	29	12.9	17.6
048	25	53.0	196.6	25	14.4	23.4
072	21	135.4	363.9	21	18.1	33.5
096	17	271.4	560.0	17	14.1	31.7
120	15	428.8	628.0	15	12.3	23.1

Verification statistics for: AL172011 PHILIPPE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	58	13.0	13.6	58	4.1	4.1
012	56	35.1	48.4	56	6.5	6.5
024	54	56.2	99.5	54	9.4	9.3
036	52	71.0	163.3	52	11.1	10.5
048	50	88.2	227.7	50	13.4	9.5
072	46	132.4	348.5	46	16.3	10.2
096	42	181.0	434.3	42	19.2	11.1
120	38	230.6	506.4	38	20.4	9.3

Verification statistics for: AL182011 RINA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	20	3.7	4.2	20	4.3	4.3
012	18	12.0	22.6	18	10.6	10.6
024	16	21.7	52.7	16	20.3	18.8
036	14	31.9	80.3	14	28.6	24.1
048	12	40.1	111.5	12	33.8	25.0
072	8	29.2	172.5	8	25.6	24.8
096	4	81.4	295.3	4	15.0	12.8
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: AL192011 SEAN

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	15	9.6	10.0	15	1.3	1.3
012	13	24.8	50.7	13	4.6	4.1
024	11	42.7	90.5	11	8.2	6.7
036	9	46.7	122.7	9	6.7	8.0
048	7	36.5	183.1	7	7.1	7.6
072	3	70.5	193.9	3	5.0	6.0
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 2011 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
2011 mean OFCL error (n mi)	25.1	40.4	53.2	68.6	103.3	149.2	166.1
2011 mean CLIPER5 error (n mi)	35.7	71.5	113.7	167.4	299.7	457.2	635.6
2011 mean OFCL skill relative to CLIPER5 (%)	29.7	43.5	53.2	59.0	65.5	67.4	73.9
2011 mean OFCL bias vector (°/n mi)	101/7	098/20	098/33	095/47	088/73	076/106	067/131
2011 number of cases	234	213	193	171	129	93	58
2006-2010 mean OFCL error (n mi)	29.7	49.9	69.0	86.6	119.0	155.8	197.7
2006-2010 mean CLIPER5 error (n mi)	38.4	74.8	115.3	155.9	226.3	273.7	310.4
2006-2010 mean OFCL skill relative to CLIPER5 (%)	22.7	33.3	40.2	44.5	47.4	43.1	36.3
2006-2010 mean OFCL bias vector (°/n mi)	292/3	291/5	279/7	273/10	235/11	183/14	148/22
2006-2010 number of cases	1198	1042	895	769	553	381	250
2011 OFCL error relative to 2006-2010 mean (%)	-15.5	-19.0	-22.9	-20.8	-13.2	-4.2	-16.0
2011 CLIPER5 error relative to 2006-2010 mean (%)	-7.0	4.4	-1.4	7.4	32.4	67.0	104.8

Table 8a. Homogenous comparison of eastern North Pacific basin early track guidance model errors (n mi) for 2011. Errors smaller than the NHC official forecast are shown in boldface.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	21.2	32.5	42.9	55.8	91.9	144.9	168.9
OCD5	32.2	68.6	115.1	177.4	328.2	496.3	632.9
GFSI	24.4	41.6	61.1	84.0	142.8	199.5	193.4
GHMI	26.9	48.0	68.6	90.6	144.5	252.0	382.9
HWFI	32.3	57.4	81.8	112.0	184.1	290.2	383.0
GFNI	31.1	53.2	75.5	99.7	170.9	254.0	352.6
NGPI	32.0	54.0	71.2	82.7	128.5	196.3	275.2
EGRI	26.8	43.6	59.6	70.8	92.5	**120.9**	180.9
EMXI	22.9	35.2	44.8	**52.6**	**75.1**	**112.9**	**158.9**
CMCI	36.1	61.8	91.2	117.7	151.6	180.3	**167.7**
AEMI	25.7	42.1	56.6	71.2	106.1	**138.9**	**125.1**
FSSE	21.6	34.4	45.5	58.7	92.5	**129.3**	**149.2**
TVCE	**21.1**	33.1	43.6	**53.6**	94.8	153.7	217.5
TVCC	23.0	34.1	46.9	61.5	118.2	245.5	353.5
LBAR	30.2	57.9	90.2	121.9	190.5	273.9	374.2
BAMD	33.9	62.1	94.0	120.1	187.8	288.3	424.1
BAMM	30.1	54.8	83.0	112.1	172.2	219.6	278.3
BAMS	36.3	63.2	97.5	132.1	201.7	229.1	290.1
# Cases	138	123	113	95	68	51	27

Table 8b. Homogenous comparison of eastern North Pacific basin early track guidance model bias vectors (º/n mi) for 2011.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	079/007	083/014	079/024	075/037	075/057	067/098	062/141
OCD5	308/005	297/019	302/042	298/080	285/195	285/309	293/358
GFSI	067/009	076/020	078/036	076/055	078/095	073/130	084/102
GHMI	085/014	086/025	083/041	084/059	080/106	060/190	057/332
HWFI	030/018	036/038	042/059	046/086	053/147	061/248	060/354
GFNI	137/006	143/020	139/035	138/051	121/102	101/165	074/285
NGPI	063/006	087/014	097/022	107/028	112/049	104/085	086/156
EGRI	223/004	245/009	269/010	275/014	243/020	214/019	252/030
EMXI	101/006	115/012	122/016	120/021	111/023	069/046	076/129
CMCI	076/011	086/016	095/015	115/014	193/020	211/032	253/042
AEMI	092/004	078/009	079/015	078/023	081/018	014/022	268/030
FSSE	071/006	070/009	060/014	055/020	051/029	030/056	027/092
TVCE	071/006	084/014	084/024	084/034	086/065	075/115	068/187
TVCC	022/006	036/009	063/028	078/032	100/067	146/205	150/260
LBAR	035/010	353/031	345/057	345/087	347/137	348/187	344/232
BAMD	058/006	050/014	045/023	041/032	035/059	025/103	020/110
BAMM	036/010	042/020	044/033	050/051	057/095	052/130	045/090
BAMS	032/016	029/024	029/033	038/046	069/076	064/081	100/048
# Cases	138	123	113	95	68	51	27

Table 9. Homogenous comparison of official and Decay-SHIFOR5 intensity forecast errors in the eastern North Pacific basin for the 2011 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
2011 mean OFCL error (kt)	7.2	12.2	15.0	16.4	17.9	18.8	17.5
2011 mean Decay-SHIFOR5 error (kt)	9.0	15.2	19.5	21.8	24.6	23.2	23.3
2011 mean OFCL skill relative to Decay-SHIFOR5 (%)	20.0	19.7	23.1	24.8	27.2	19.0	24.9
2011 OFCL bias (kt)	-0.3	-0.4	-0.4	-0.4	2.9	2.8	2.8
2011 number of cases	234	213	193	171	129	93	58
2006-10 mean OFCL error (kt)	6.3	10.5	13.7	15.1	17.1	18.6	18.0
2006-10 mean Decay-SHIFOR5 error (kt)	7.3	11.9	15.3	17.6	19.0	20.3	21.1
2006-10 mean OFCL skill relative to Decay-SHIFOR5 (%)	13.7	11.8	10.5	14.2	10.5	8.4	14.7
2006-10 OFCL bias (kt)	0.7	1.2	1.2	0.5	0.9	0.1	-1.1
2006-10 number of cases	1198	1042	895	769	553	381	250
2011 OFCL error relative to 2006-10 mean (%)	14.3	16.2	9.5	8.6	4.7	1.1	-2.8
2011 Decay-SHIFOR5 error relative to 2006-10 mean (%)	23.3	27.7	27.5	23.9	29.5	14.3	10.4

Table 10a. Homogenous comparison of eastern North Pacific basin early intensity guidance model errors (kt) for 2011. Errors smaller than the NHC official forecast are shown in boldface.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	7.2	11.8	13.7	14.8	16.1	17.7	16.6
OCD5	9.2	15.3	19.5	22.4	24.4	24.7	23.3
HWFI	9.6	14.0	17.8	19.1	24.2	29.0	32.2
GHMI	8.5	12.2	13.8	15.9	22.3	30.1	31.2
GFNI	9.2	13.2	14.0	**14.3**	17.1	**17.5**	**15.8**
DSHP	8.0	12.8	16.2	17.9	20.3	19.9	19.0
LGEM	7.9	12.2	15.2	17.8	20.2	18.2	**14.6**
ICON	7.9	**11.4**	**13.6**	15.2	19.5	22.1	19.5
IVCN	7.9	**11.1**	**12.6**	**14.2**	18.6	20.7	18.0
FSSE	7.2	**10.1**	**12.6**	14.8	18.3	24.8	21.6
# Cases	179	160	142	121	84	57	40

Table 10b. Homogenous comparison of eastern North Pacific basin early intensity guidance model biases (kt) for 2011. Biases smaller than the NHC official forecast are shown in boldface.

	Forecast Period (h)						
Model ID	12	24	36	48	72	96	120
OFCL	0.4	2.6	5.3	7.4	8.2	3.9	4.6
OCD5	0.5	**2.5**	**4.8**	**5.4**	**0.2**	**-0.8**	**-0.1**
HWFI	-1.5	**-1.0**	**1.5**	**3.9**	**6.5**	8.5	13.9
GHMI	-2.6	-3.0	**-1.6**	**2.9**	9.2	8.3	**2.9**
GFNI	-3.4	-7.1	-7.9	**-5.8**	**-5.3**	**-3.0**	**-0.5**
DSHP	**0.0**	**1.9**	**4.6**	**6.3**	**5.4**	**2.2**	6.1
LGEM	-0.4	**-0.4**	**1.8**	**2.2**	**0.4**	**-2.1**	**0.1**
ICON	-0.9	**-0.2**	**1.9**	**4.0**	**5.5**	4.4	5.8
IVCN	-1.3	**-1.5**	**0.0**	**2.2**	**3.4**	**3.0**	4.6
FSSE	-0.5	**0.8**	**3.0**	**3.9**	**0.5**	-8.0	-11.3
# Cases	179	160	142	121	84	57	40

Table 11. Official eastern North Pacific track and intensity forecast verifications (OFCL) for 2011 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: EP012011 ADRIAN

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	20	12.0	12.1	20	1.0	1.0
012	18	25.1	29.8	18	6.7	10.6
024	16	52.3	56.4	16	15.9	21.7
036	14	83.2	75.5	14	22.1	31.9
048	12	122.8	102.3	12	27.1	36.8
072	8	240.7	176.3	8	31.9	30.0
096	4	357.0	237.2	4	30.0	14.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: EP022011 BEATRIZ

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	11	15.4	16.1	11	5.0	5.0
012	9	37.6	55.1	9	15.6	13.3
024	7	43.6	82.2	7	20.7	18.4
036	5	45.8	115.4	5	26.0	22.6
048	3	46.2	167.9	3	25.0	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: EP032011 CALVIN

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	11	17.8	15.8	11	0.0	0.0
012	9	26.0	31.9	9	7.8	11.8
024	7	51.3	68.0	7	10.7	23.1
036	5	72.8	93.8	5	11.0	21.6
048	3	115.3	76.6	3	6.7	8.7
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: EP042011 DORA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	26	6.5	6.5	26	3.7	4.4
012	24	18.0	25.2	24	9.2	12.4
024	22	27.3	54.6	22	14.3	18.2
036	20	31.5	88.2	20	16.5	23.8
048	18	31.7	125.0	18	17.2	24.9
072	14	37.2	200.7	14	22.1	29.8
096	10	63.7	272.1	10	22.5	25.9
120	6	71.3	306.7	6	15.0	27.5

Verification statistics for: EP052011 EUGENE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	25	7.5	7.2	25	2.2	2.0
012	23	17.8	24.8	23	3.9	7.0
024	21	30.3	43.8	21	8.6	11.0
036	19	49.6	65.6	19	13.4	15.3
048	17	75.3	91.1	17	16.8	17.5
072	13	151.4	168.5	13	21.9	23.1
096	9	278.9	279.3	9	13.9	18.9
120	5	387.6	373.8	5	6.0	12.8

Verification statistics for: EP062011 FERNANDA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	11	8.3	8.8	11	0.9	0.9
012	11	20.5	35.0	11	5.5	5.8
024	11	31.9	65.4	11	8.2	8.7
036	11	34.6	87.8	11	5.9	8.7
048	9	37.4	102.9	9	9.4	11.1
072	5	66.2	84.9	5	15.0	18.2
096	1	123.1	108.5	1	10.0	31.0
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: EP072011 GREG

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	19	3.5	4.5	19	0.3	0.3
012	17	25.5	35.8	17	4.1	5.9
024	15	45.6	75.3	15	10.0	12.8
036	13	64.6	119.2	13	11.9	20.0
048	11	84.2	163.8	11	12.7	22.1
072	7	126.4	279.3	7	25.0	25.6
096	3	196.3	413.5	3	31.7	28.3
120	0	-999.0	-999.0	0	-999.0	-999.0

Verification statistics for: EP082011 EIGHT

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	2	8.3	8.3	2	0.0	0.0
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: EP092011 HILARY

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	37	6.2	6.2	37	2.0	2.3
012	35	21.8	29.8	35	6.3	6.0
024	33	34.3	50.8	33	9.1	12.0
036	31	49.0	85.5	31	12.3	17.7
048	29	66.3	127.3	29	13.3	21.8
072	25	110.6	208.1	25	12.6	28.0
096	21	166.9	264.2	21	12.1	30.4
120	17	219.5	315.4	17	15.6	32.5

Verification statistics for: EP102011 JOVA

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	28	11.4	11.7	28	2.3	2.5
012	26	30.0	40.0	26	6.9	8.5
024	24	39.1	77.8	24	9.8	13.6
036	22	43.5	125.7	22	9.3	15.5
048	20	54.2	199.5	20	10.3	21.5
072	16	79.9	397.2	16	9.7	27.8
096	12	131.8	629.4	12	33.8	34.8
120	5	128.4	787.0	5	45.0	35.8

Verification statistics for: EP112011 IRWIN

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	42	11.4	11.9	42	3.0	3.2
012	40	29.1	48.3	40	6.8	7.4
024	38	50.6	115.7	38	12.4	12.4
036	36	63.0	198.4	36	15.1	13.9
048	34	77.2	296.6	34	14.3	13.5
072	30	94.9	506.6	30	14.8	14.8
096	26	117.1	711.7	26	15.0	16.0
120	22	128.4	1003.3	22	15.7	14.4

Verification statistics for: EP122011 TWELVE

VT (h)	NT	OFCL	OCD5	NI	OFCL	OCD5
000	3	35.9	35.9	3	0.0	0.0
012	1	77.6	54.1	1	5.0	0.0
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:    EP132011              KENNETH

     VT (h)       NT    OFCL     OCD5       NI    OFCL     OCD5
     000          23    10.1     10.5       23     2.4      2.6
     012          21    26.6     37.8       21    11.0     15.9
     024          19    44.2     72.1       19    20.3     25.2
     036          17    57.5    103.8       17    27.4     33.6
     048          15    68.2    158.1       15    32.7     40.3
     072          11    72.7    282.9       11    27.3     32.5
     096           7    65.8    483.6        7    17.1     12.3
     120           3    23.2    594.5        3    20.0     24.0
```

Table 12a Verification of experimental in-house probabilistic genesis forecasts for the Atlantic basin in 2011.

Atlantic Basin Genesis Forecast Reliability Table		
Forecast Likelihood (%)	**Verifying Genesis Occurrence Rate (%)**	**Number of Forecasts**
0	6	67
10	11	169
20	24	100
30	31	59
40	25	36
50	37	19
60	62	29
70	65	17
80	83	6
90	83	6
100	100	4

Table 12b. Verification of experimental in-house probabilistic genesis forecasts for the eastern North Pacific basin in 2011.

Eastern North Pacific Basin Genesis Forecast Reliability Table		
Forecast Likelihood (%)	**Verifying Genesis Occurrence Rate (%)**	**Number of Forecasts**
0	3	32
10	8	105
20	29	55
30	21	14
40	57	14
50	77	22
60	80	10
70	63	19
80	71	7
90	80	5
100	100	3

Table 13a. Verification of experimental in-house probabilistic genesis forecasts for the Atlantic basin for the period 2007- 2011.

Atlantic Basin Genesis Forecast Reliability Table		
Forecast Likelihood (%)	**Verifying Genesis Occurrence Rate (%)**	**Number of Forecasts**
0	3	547
10	7	966
20	17	489
30	32	282
40	43	164
50	43	110
60	57	129
70	56	75
80	70	50
90	79	29
100	100	5

Table 13b. Verification of experimental in-house probabilistic genesis forecasts for the eastern North Pacific basin for the period 2007-2011.

Eastern North Pacific Basin Genesis Forecast Reliability Table		
Forecast Likelihood (%)	**Verifying Genesis Occurrence Rate (%)**	**Number of Forecasts**
0	3	203
10	15	568
20	29	350
30	40	140
40	58	100
50	73	105
60	74	72
70	71	66
80	71	41
90	82	11
100	100	4

Table 14. HFIP Stream 1.5 models for 2011.

ID	Description	Parameter	NHC Application
AHQI	NCAR/MMM – SUNY 4-km WRF AHW. Early version of AHW4 using GFDL interpolator. Note that AHWI (standard interpolator) is identical to AHQI for track and was used in TV15.	Trk, Int	Direct use, include in TV15 and IV15 consensus.
COTI	NRL COAMPS-TC. Early version of COTC using GFDL interpolator.	Int	Include in IV15 consensus.
A4QI	PSU 4.5 km WRF-EnKF systm with Doppler data assimilated. Early version of A4PS using GFDL interpolator.	Int	Include in IV15 consensus.
FIMI	ESRL 30-km FIM. Early version of FIMY.	Trk	Direct use, include in TV15 consensus.
UWQI	University of Wisconsin 8 km. Early version of UWN8 using GFDL interpolator.	Int	Include in IV15 consensus.
H3GI	EMC-HRD 3-km HWRF. Early version of H3GP.	Trk	Direct use.
GPMI	GFDL ensemble mean; early version of GPMN using GFDL interpolator	Trk, Int	Direct use.
G01I	GFDL ensemble member with no bogus vortex.	Trk	Direct use.
SPC3	CIRA SPICE 6-member statistical consensus of DSHP and LGEM with different initial conditions.	Int	Direct use.

Table 15. NHC forecast cone circle radii (n mi) for 2012. Change from 2011 values (n mi) given in parentheses.

Track Forecast Cone Two-Thirds Probability Circles (n mi)		
Forecast Period (h)	**Atlantic Basin**	**Eastern North Pacific Basin**
12	36 (0)	33 (0)
24	56 (-3)	52 (-7)
36	75 (-4)	72 (-7)
48	95 (-3)	89 (-9)
72	141 (-3)	121 (-13)
96	180(-10)	170 (-17)
120	236 (-3)	216 (-14)

Table 16. Composition of NHC consensus models for 2012. 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 2012			
Model ID	**Parameter**	**Type**	**Members**
GUNA	Track	Fixed	GFSI EGRI NGPI GHMI
TCOA	Track	Fixed	GFSI EGRI GHMI HWFI
TCOE*	Track	Fixed	GFSI EGRI NGPI GHMI HWFI
ICON	Intensity	Fixed	DSHP LGEM GHMI HWFI
TVCA	Track	Variable	GFSI EGRI GHMI HWFI GFNI EMXI
TVCE**	Track	Variable	GFSI EGRI NGPI GHMI HWFI GFNI EMXI
IVCN	Intensity	Variable	DSHP LGEM GHMI HWFI GFNI

* 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

reliability indicated by the thin diagonal black line. The dashed green line indicates how the forecasts were distributed among the possible forecast values.

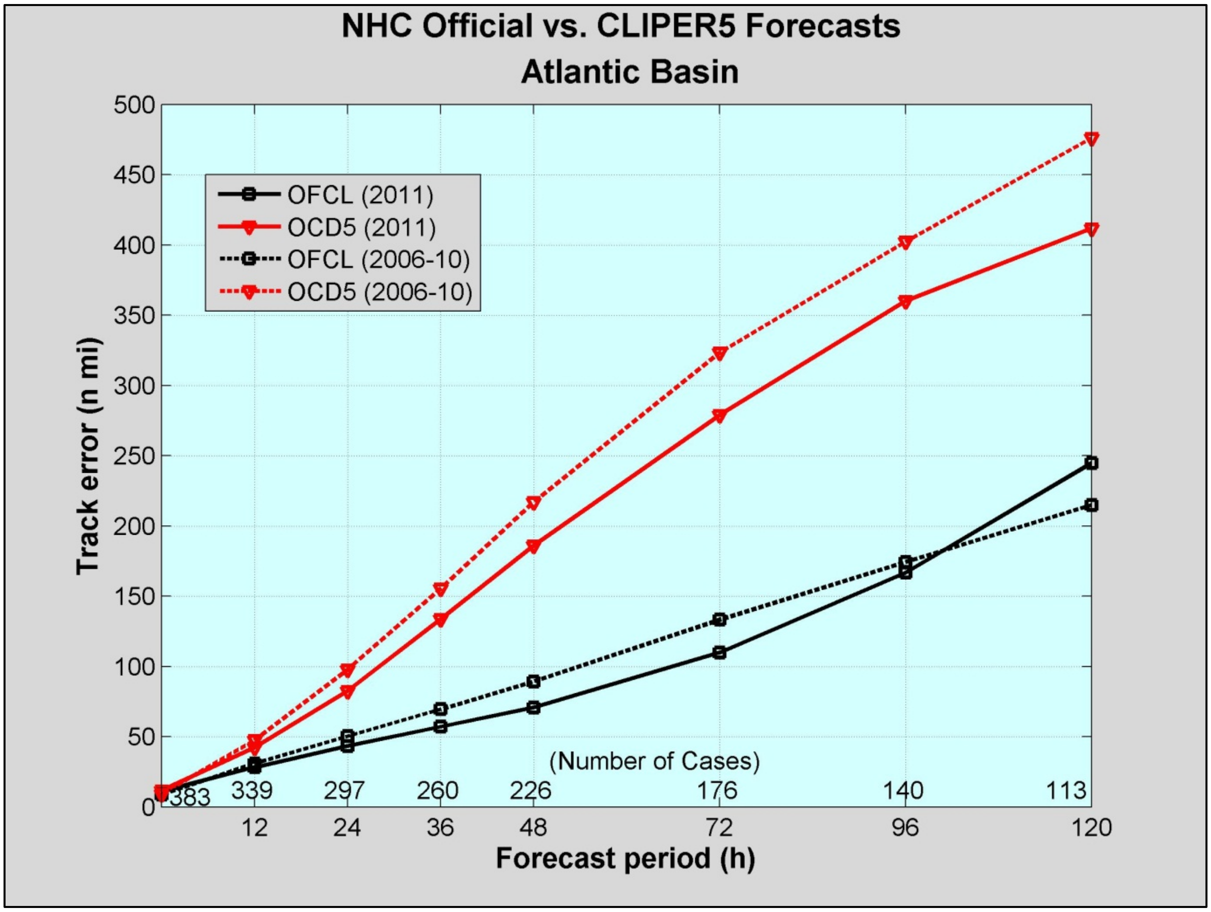


Figure 1. NHC official and CLIPER5 (OCD5) Atlantic basin average track errors for 2011 (solid lines) and 2006-2010 (dashed lines).

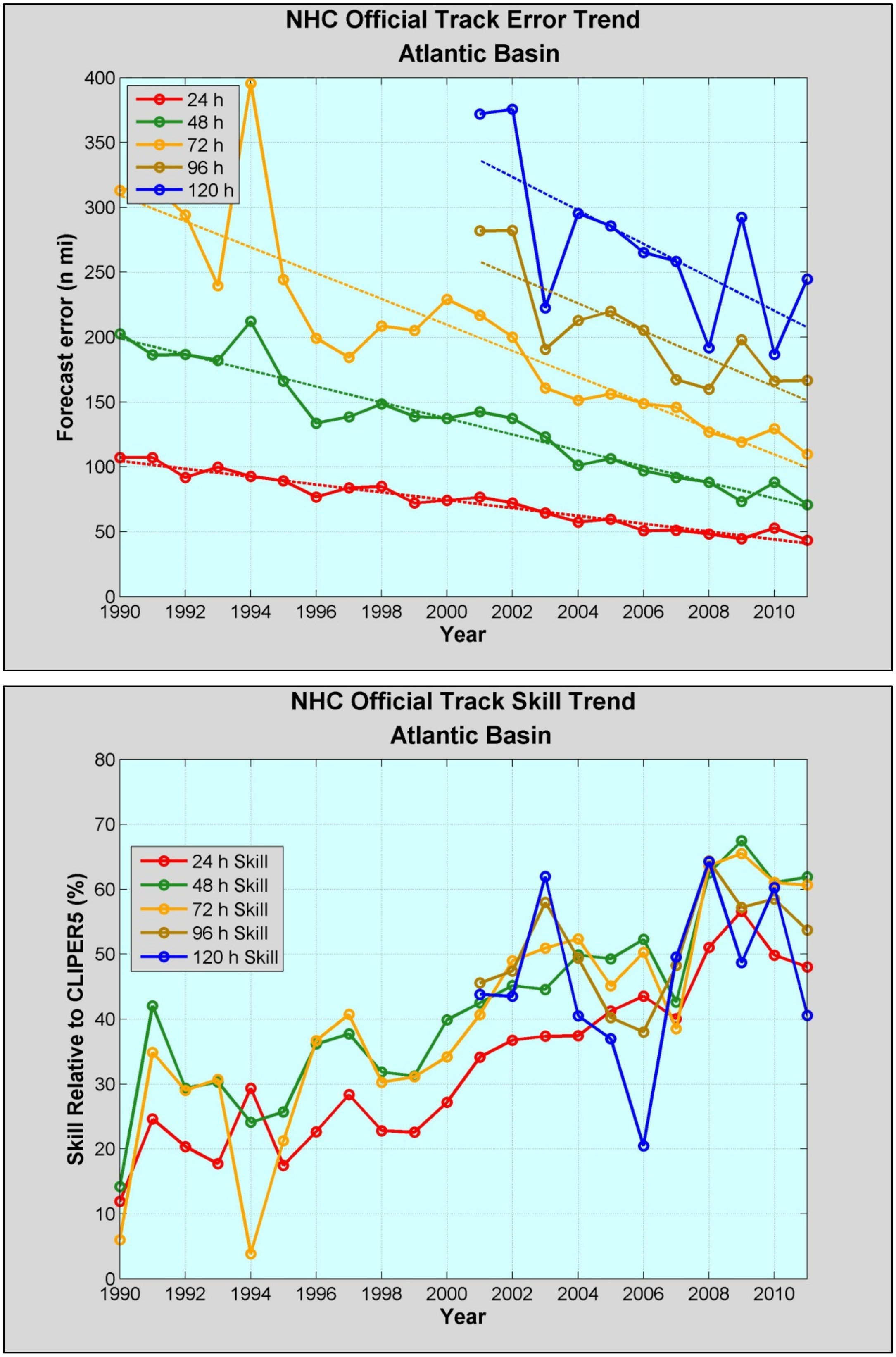


Figure 2. Recent trends in NHC official track forecast error (top) and skill (bottom) for the Atlantic basin.

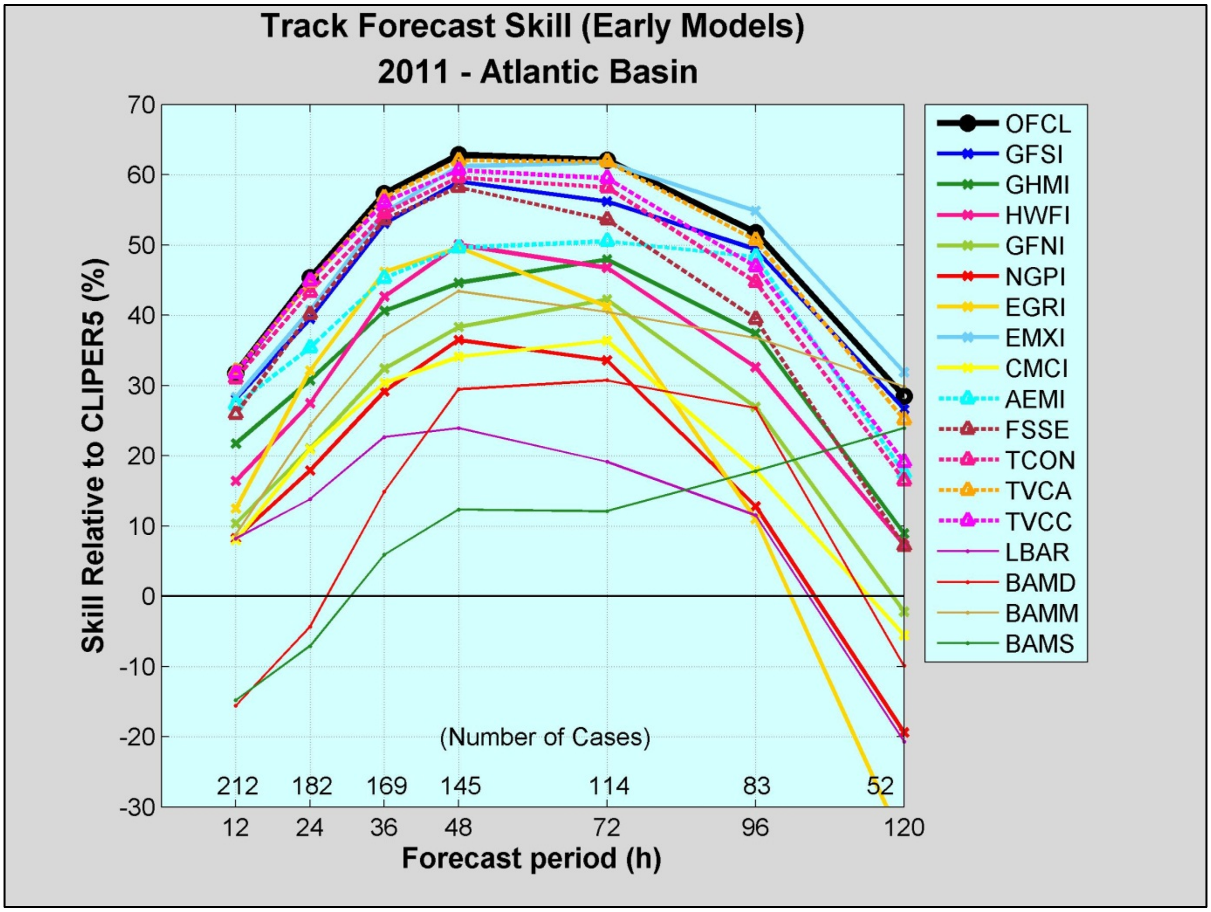


Figure 3. Homogenous comparison for selected Atlantic basin early track models for 2011. This verification includes only those models that were available at least 2/3 of the time (see text).

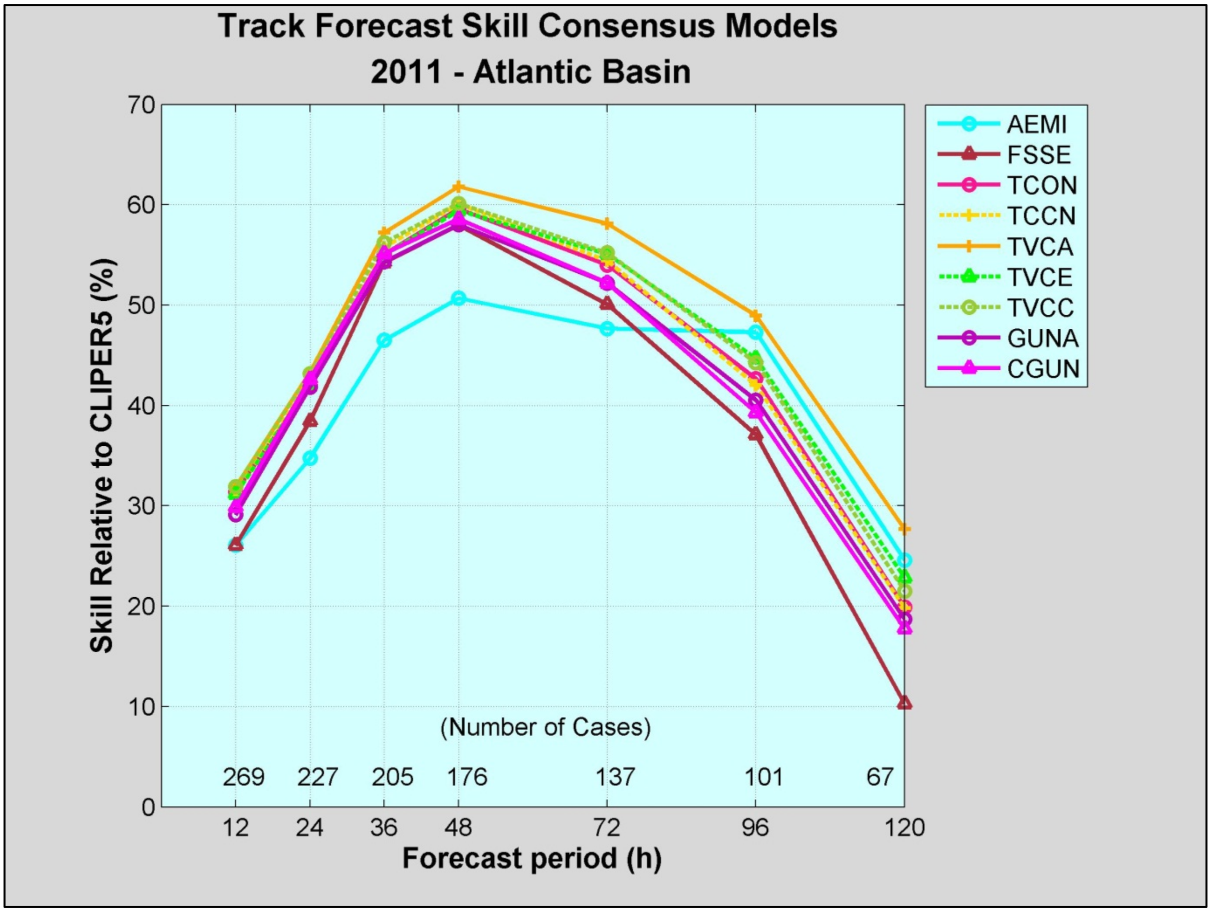


Figure 4. Homogenous comparison of the primary Atlantic basin track consensus models for 2011.

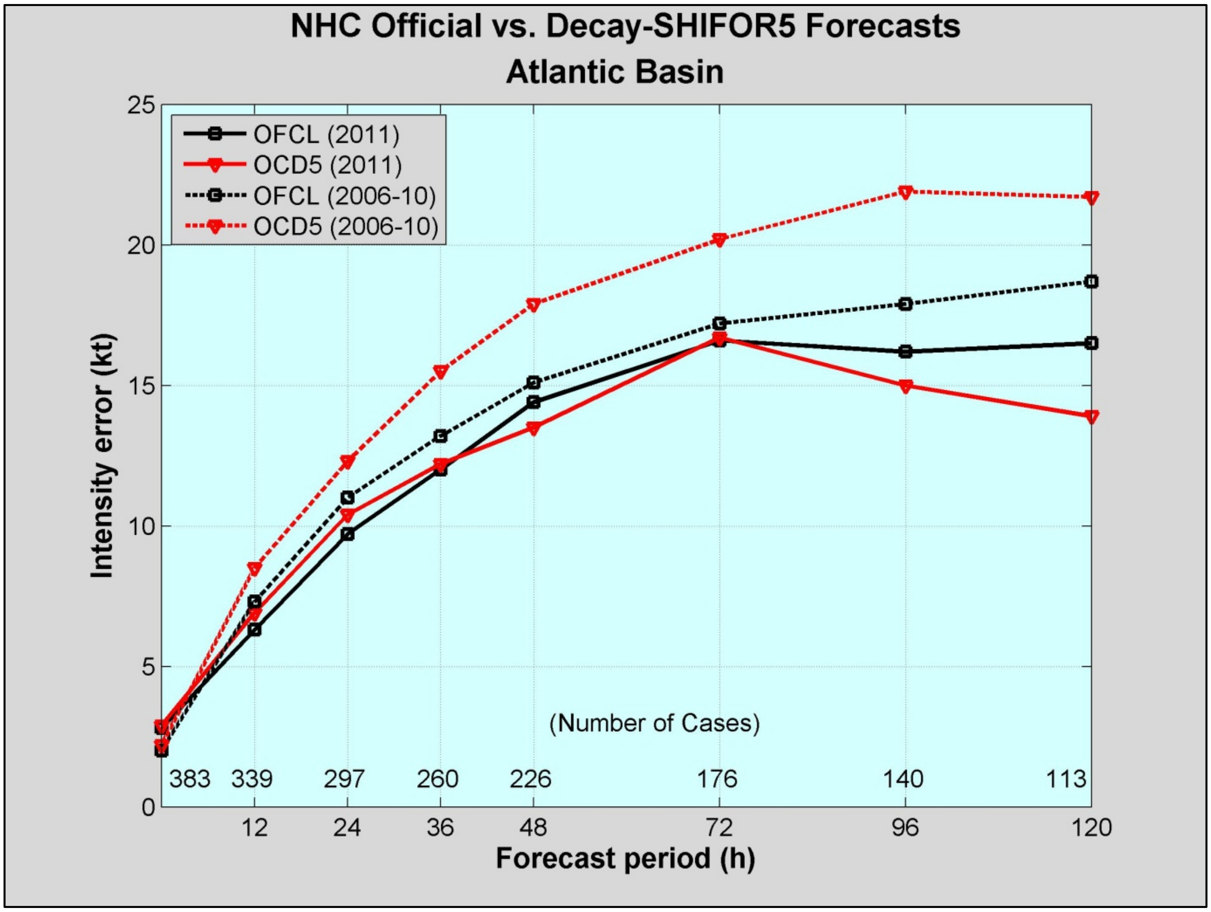


Figure 5. NHC official and Decay-SHIFOR5 (OCD5) Atlantic basin average intensity errors for 2011 (solid lines) and 2006-2010 (dashed lines).

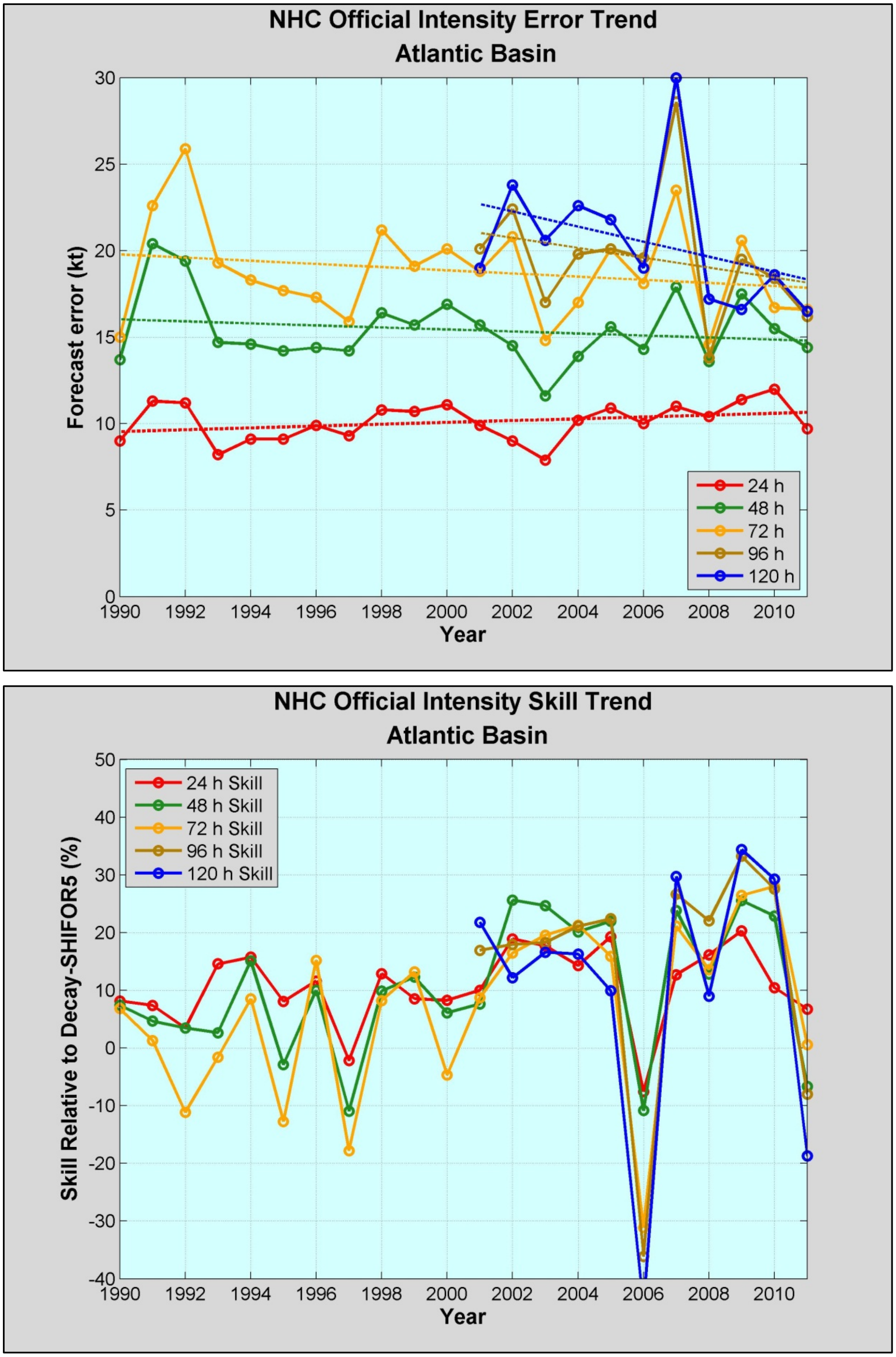


Figure 6. Recent trends in NHC official intensity forecast error (top) and skill (bottom) for the Atlantic basin.

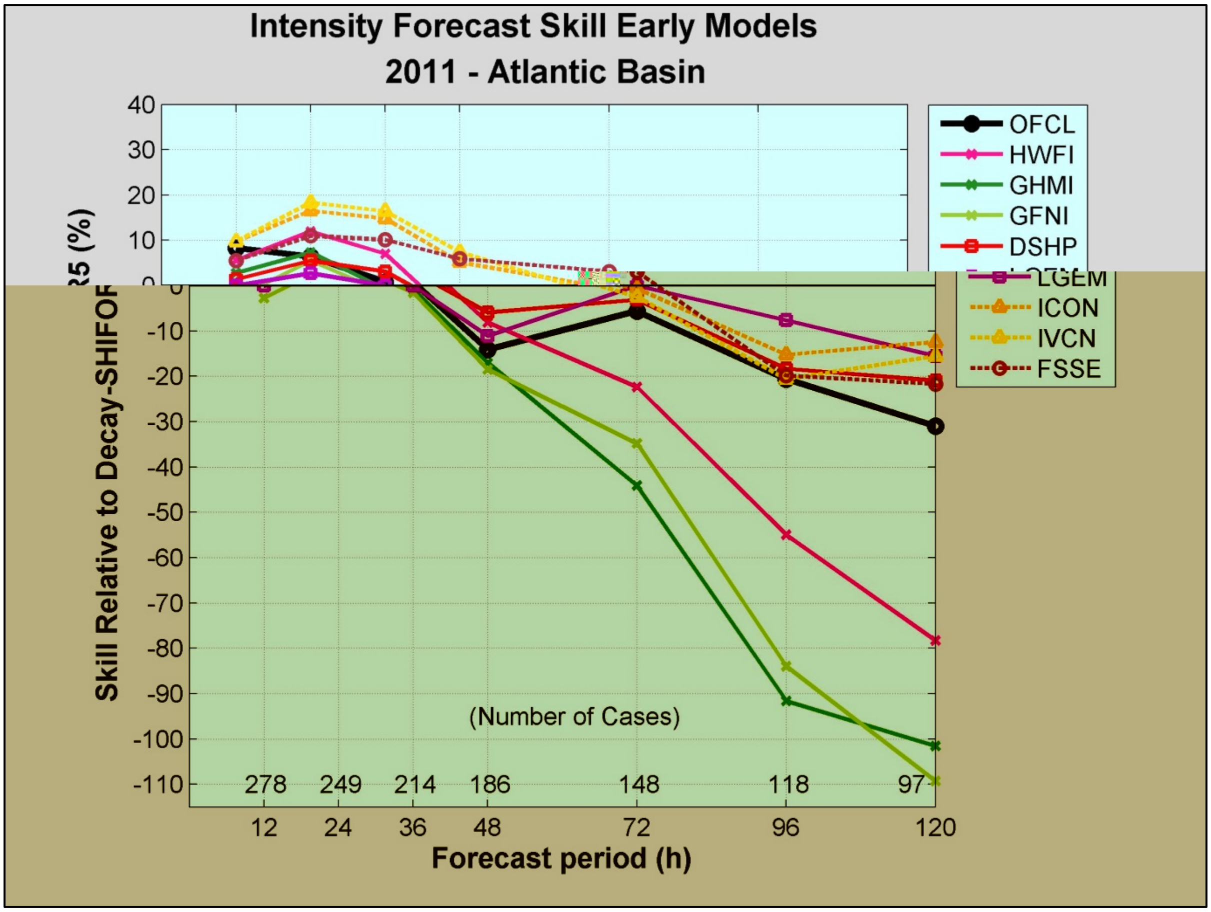


Figure 7. Homogenous comparison for selected Atlantic basin early intensity guidance models for 2011.

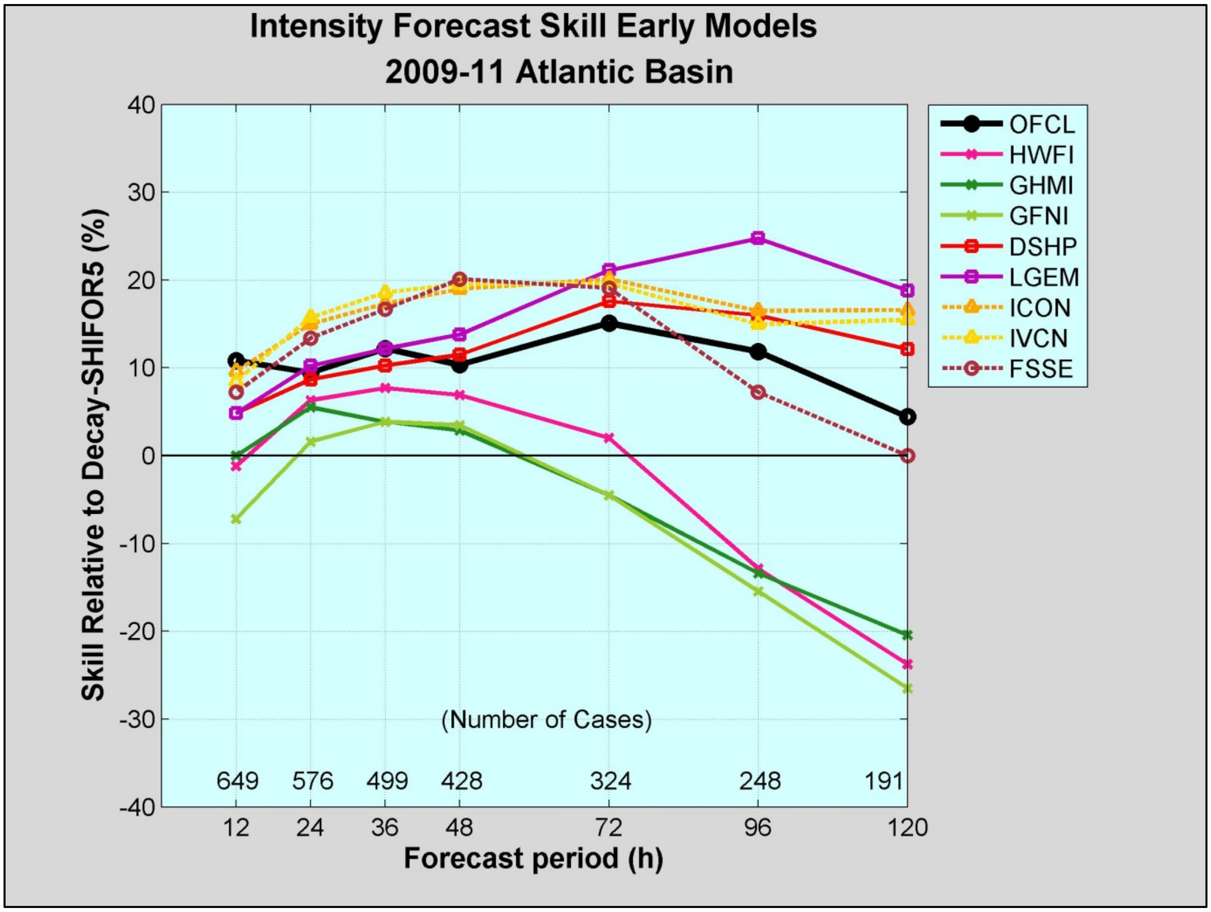


Figure 8. Homogenous comparison for selected for Atlantic basin early intensity guidance models for 2009-2011.

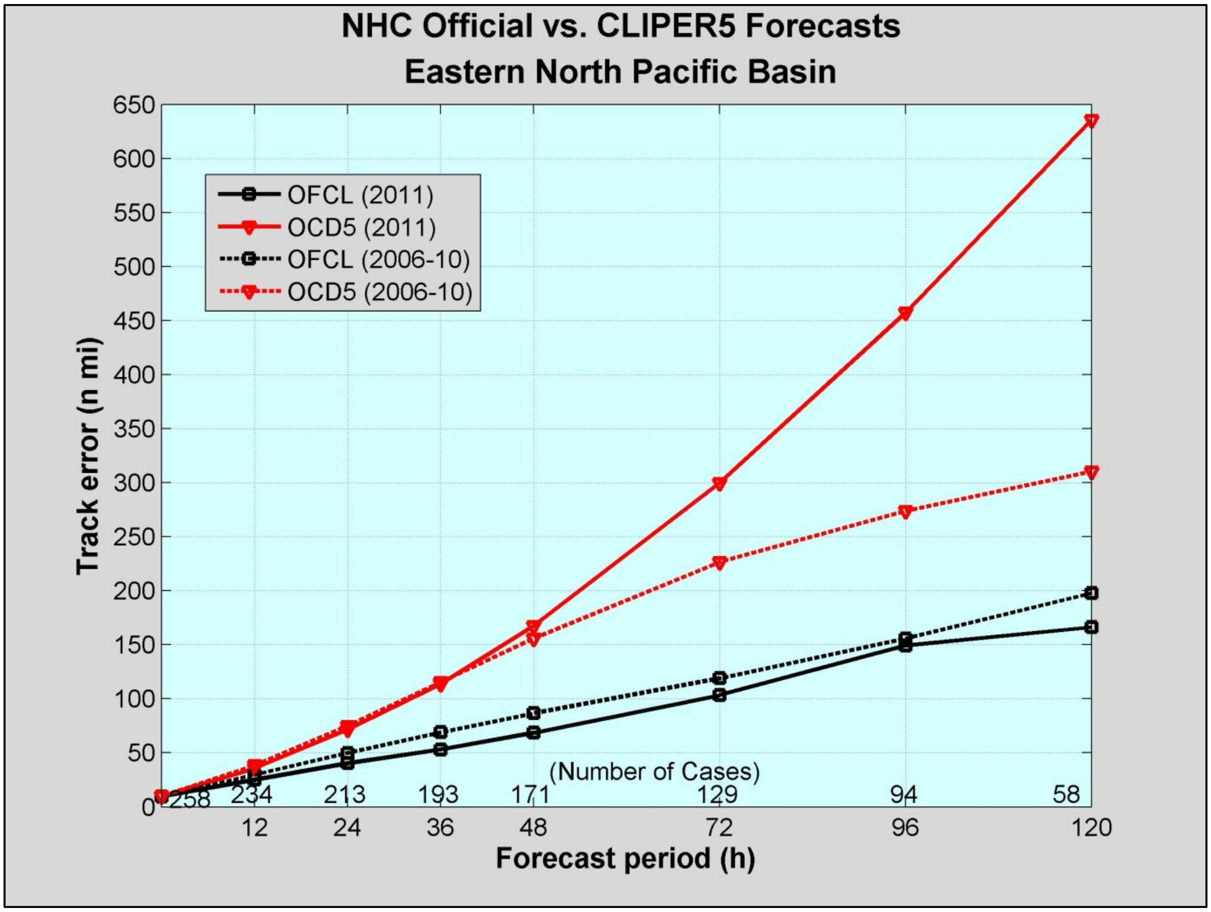
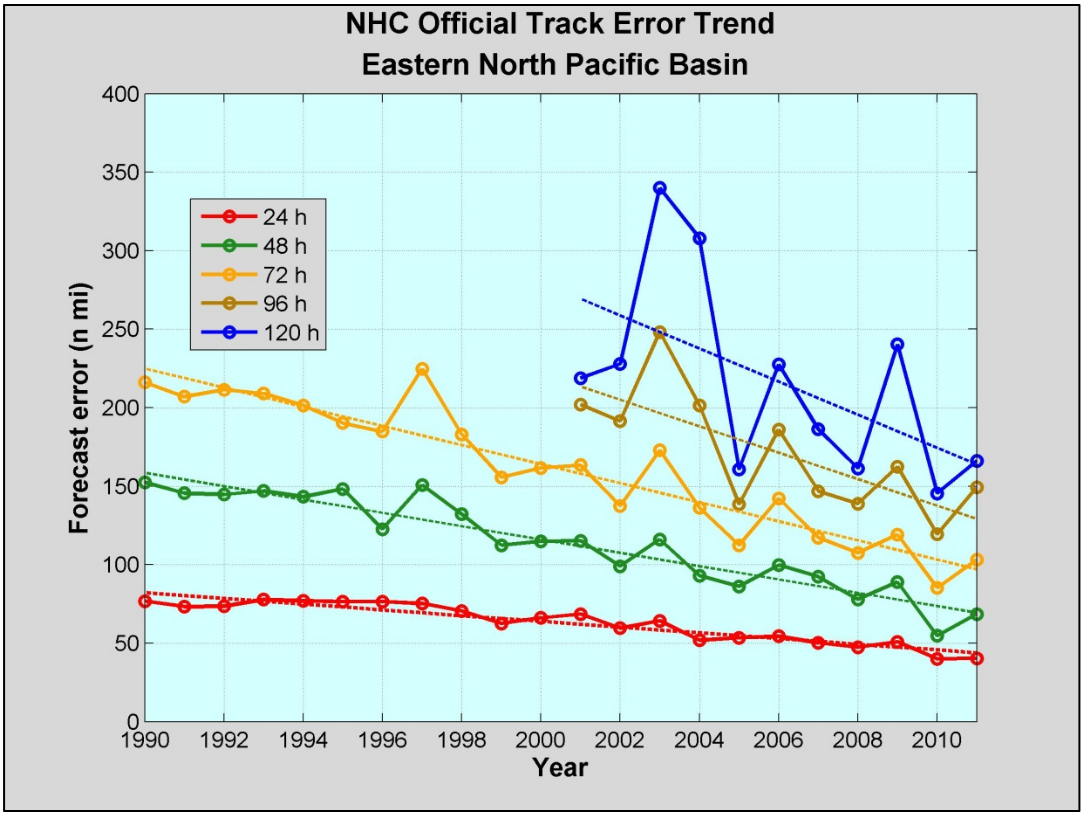


Figure 9. NHC official and CLIPER5 (OCD5) eastern North Pacific basin average track errors for 2011 (solid lines) and 2006-2010 (dashed lines).



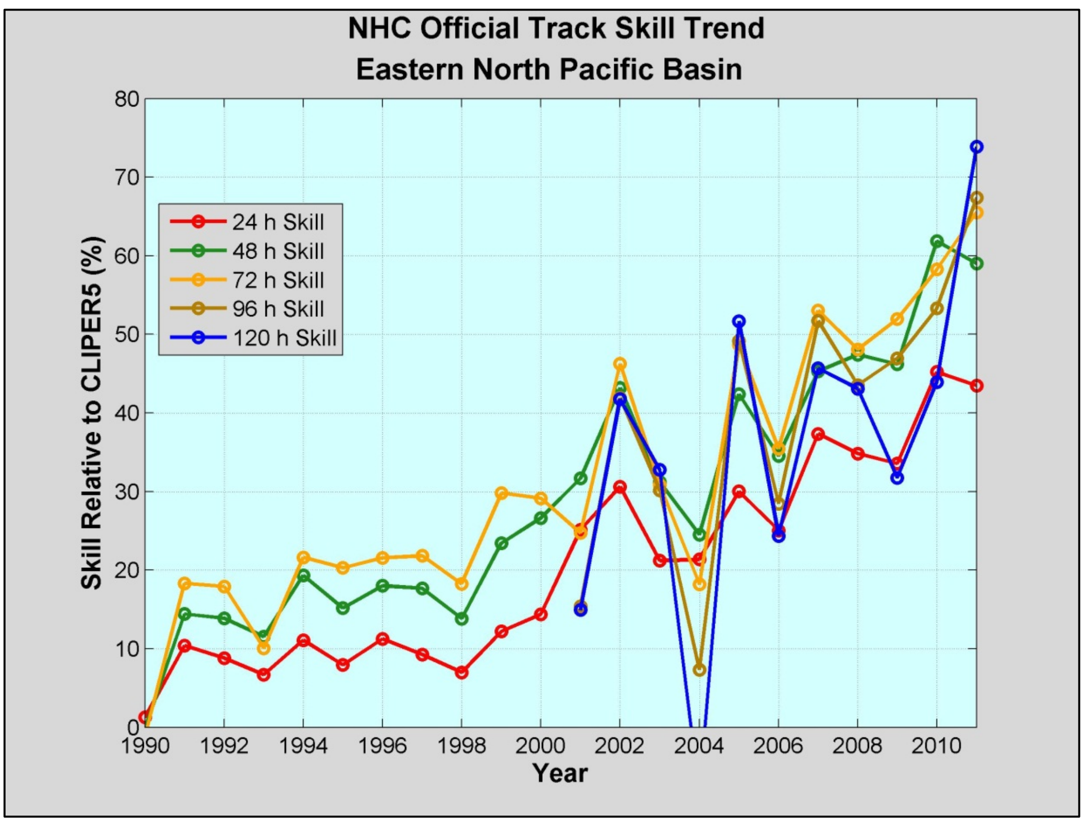


Figure 10. Recent trends in NHC official track forecast error (top) and skill (bottom) for the eastern North Pacific basin.

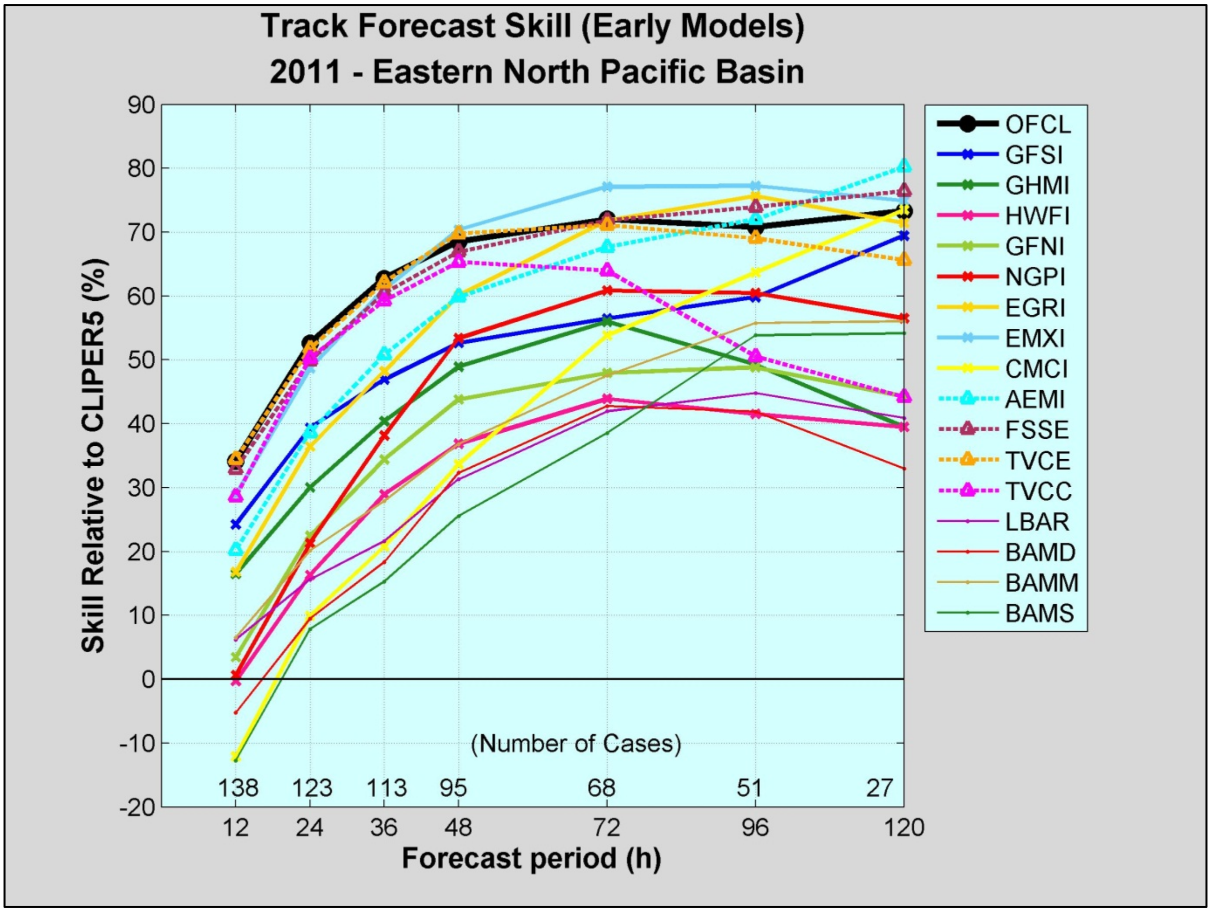


Figure 11. Homogenous comparison for selected eastern North Pacific early track models for 2011. This verification includes only those models that were available at least 2/3 of the time (see text).

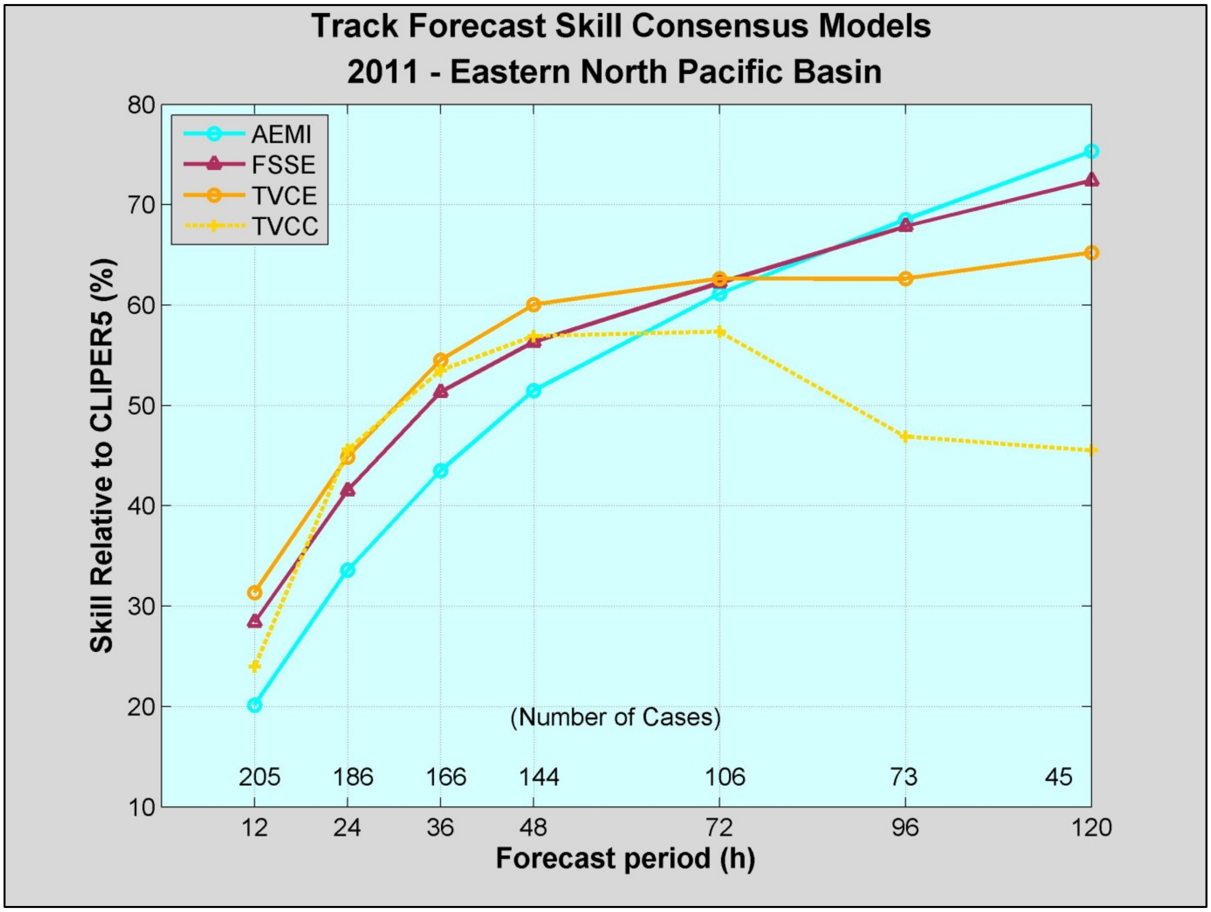


Figure 12. Homogenous comparison of the primary eastern North Pacific basin track consensus models for 2011.

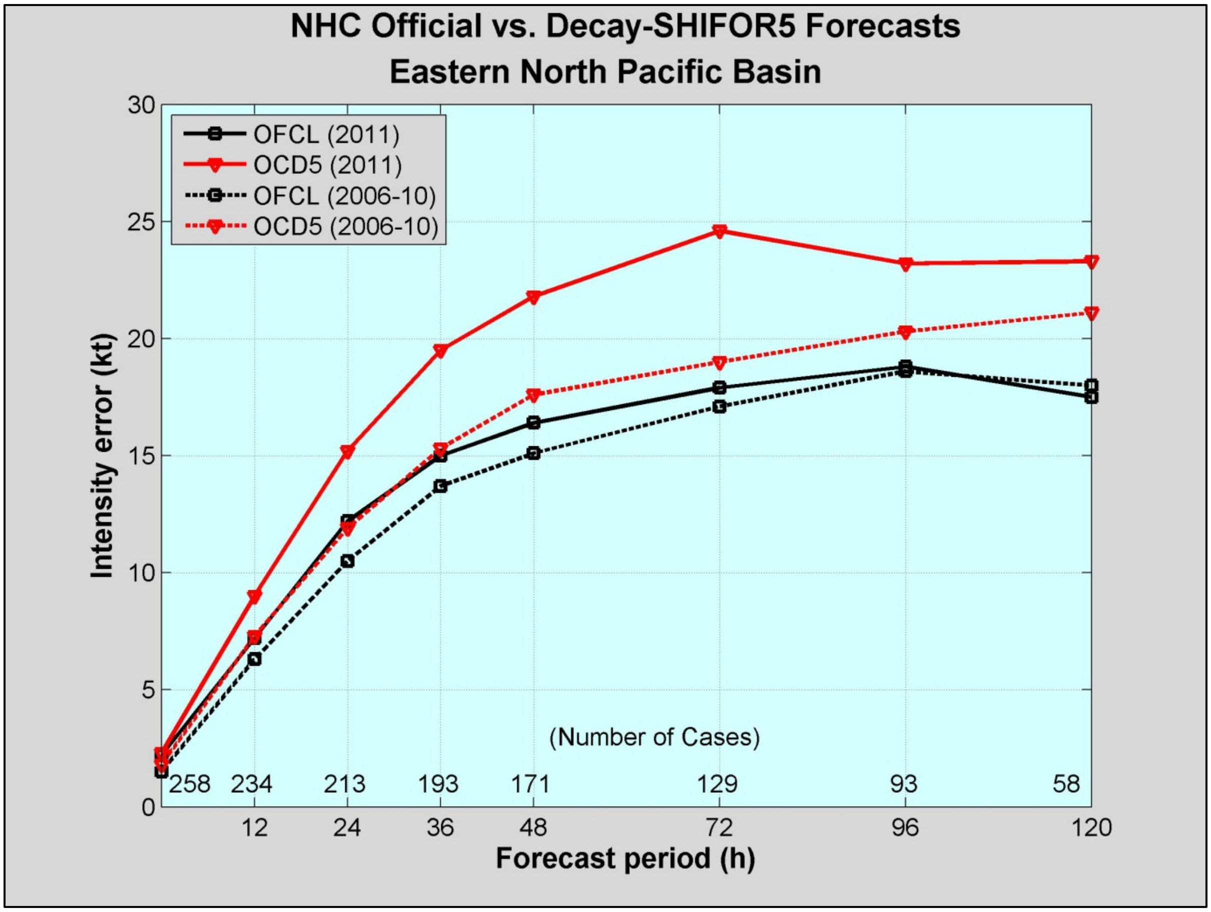
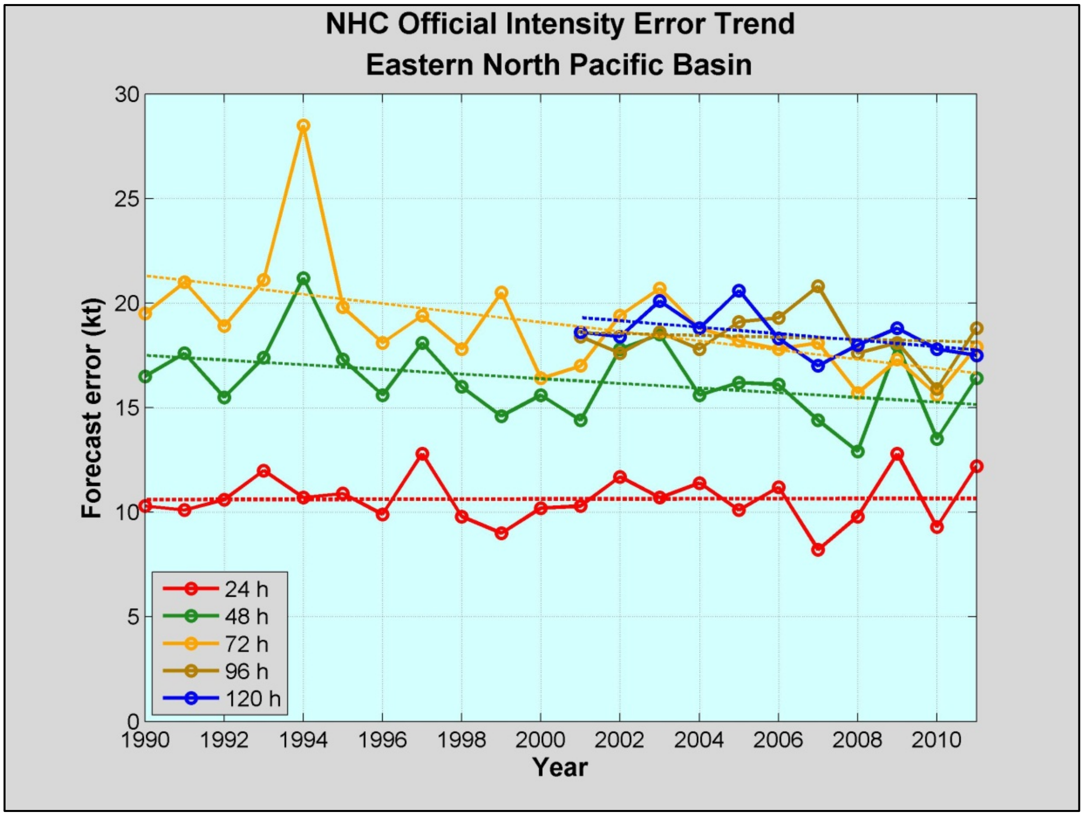


Figure 13. NHC official and Decay-SHIFOR5 (OCD5) eastern North Pacific basin average intensity errors for 2011 (solid lines) and 2006-2010 (dashed lines).



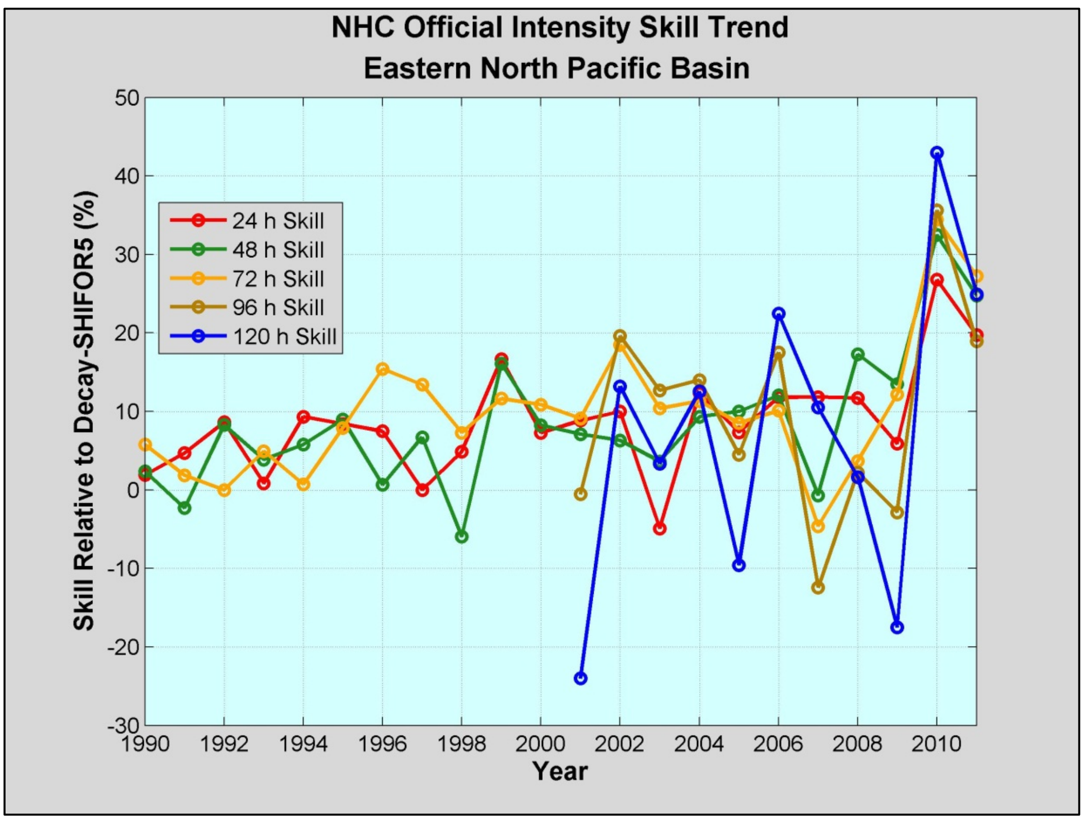


Figure 14. Recent trends in NHC official intensity forecast error (top) and skill (bottom) for the eastern North Pacific basin.

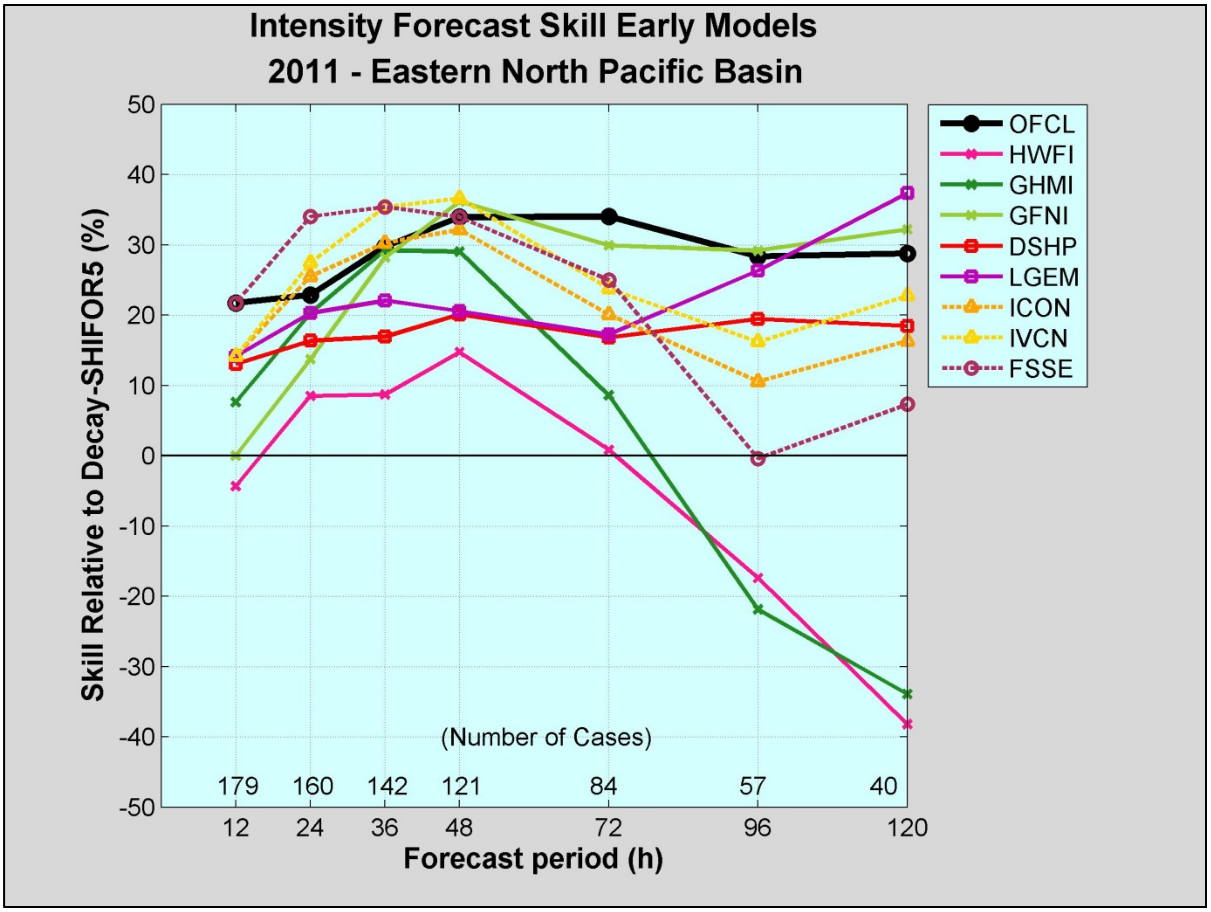


Figure 15. Homogenous comparison for selected eastern North Pacific basin early intensity guidance models for 2011.

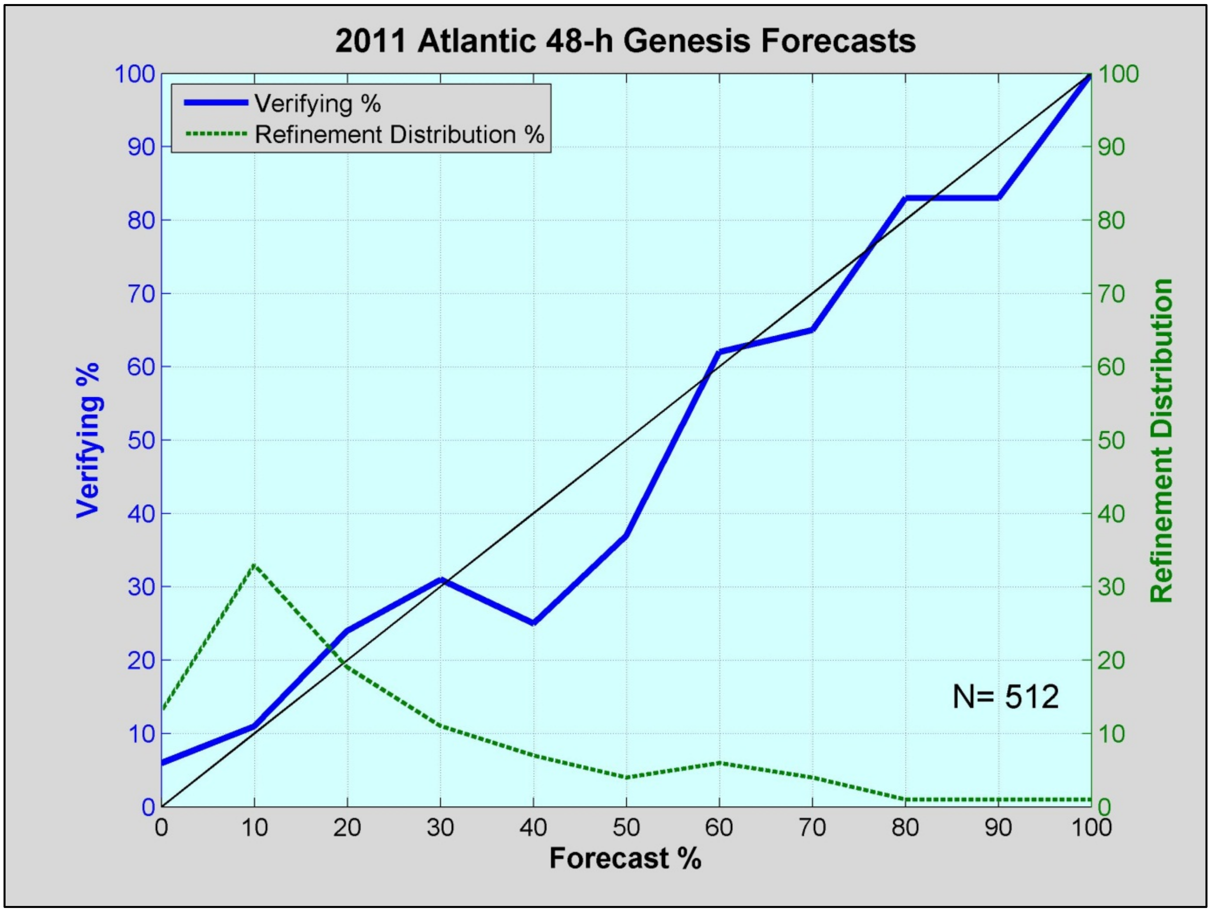


Figure 16a. Reliability diagram for Atlantic probabilistic tropical cyclogenesis forecasts for 2011. The solid blue line indicates the relationship between the forecast and verifying genesis percentages, with perfect reliability indicated by the thin diagonal black line. The dashed green line indicates how the forecasts were distributed among the possible forecast values.

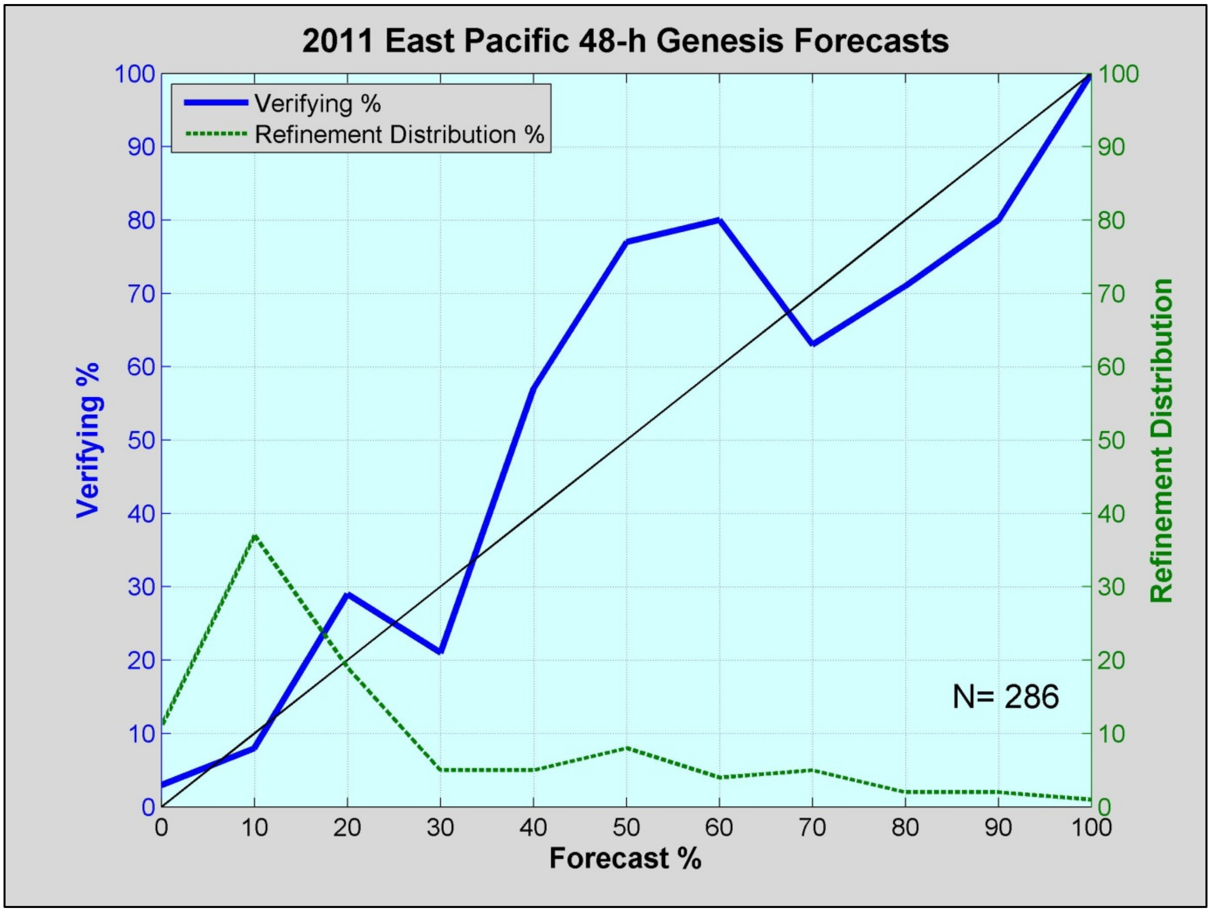


Figure 16b. As described for Fig. 16a, except for the eastern North Pacific basin.

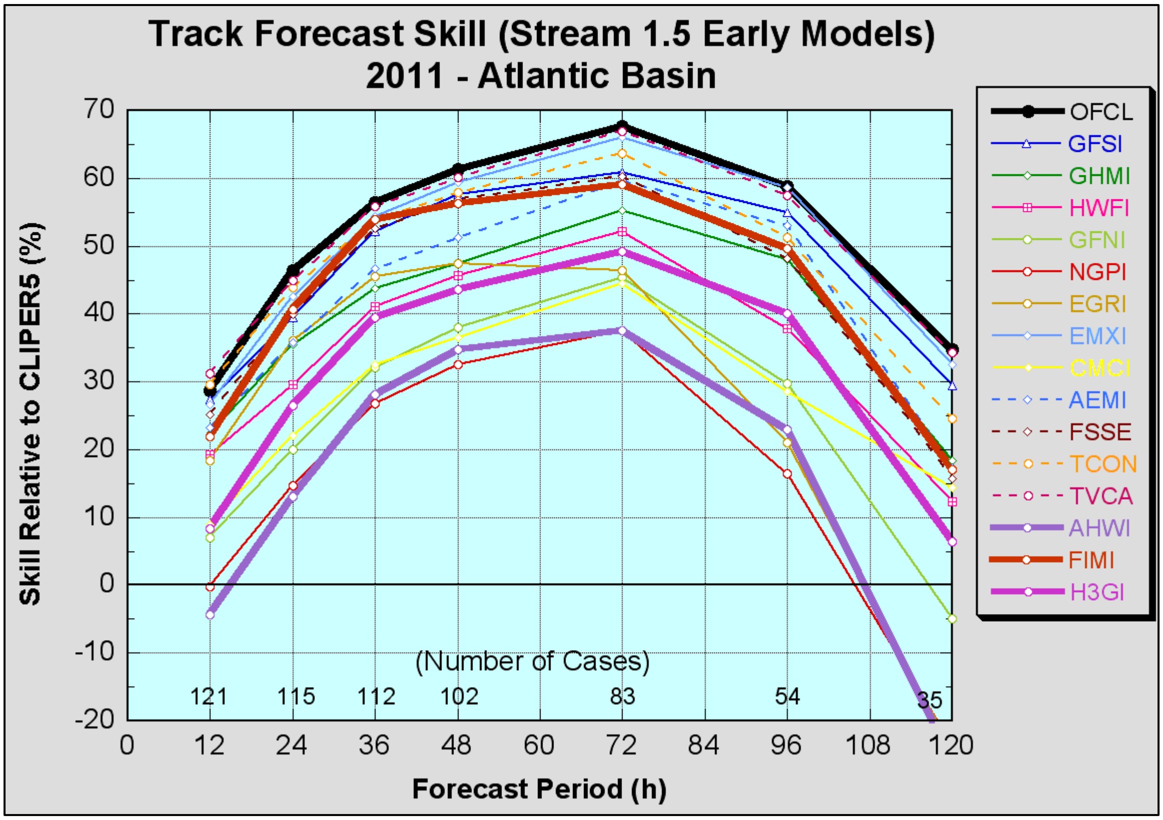


Figure 17. Homogeneous comparison of HFIP Stream 1.5 track models and selected operational models for 2011.

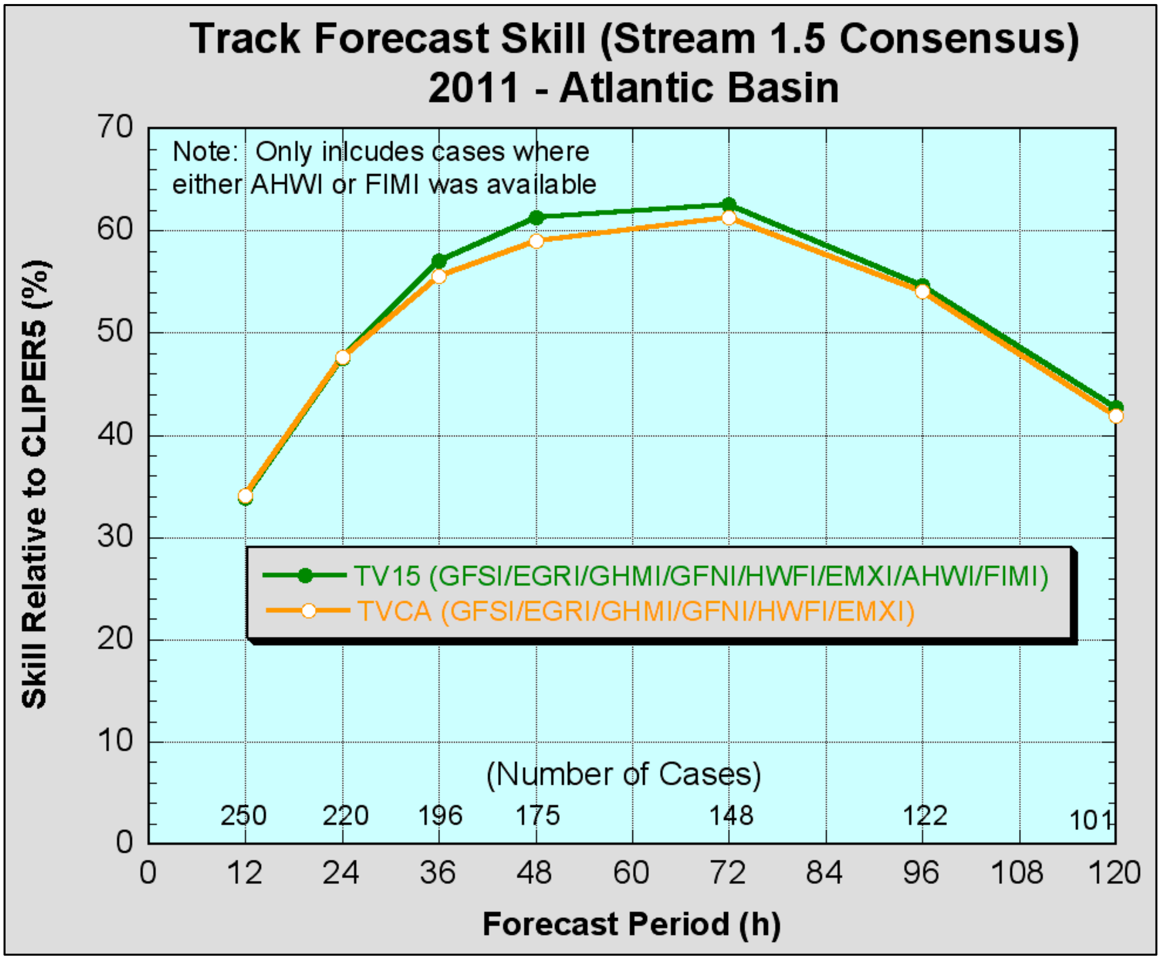


Figure 18. Impact of adding Stream 1.5 models to the variable track consensus TVCA.

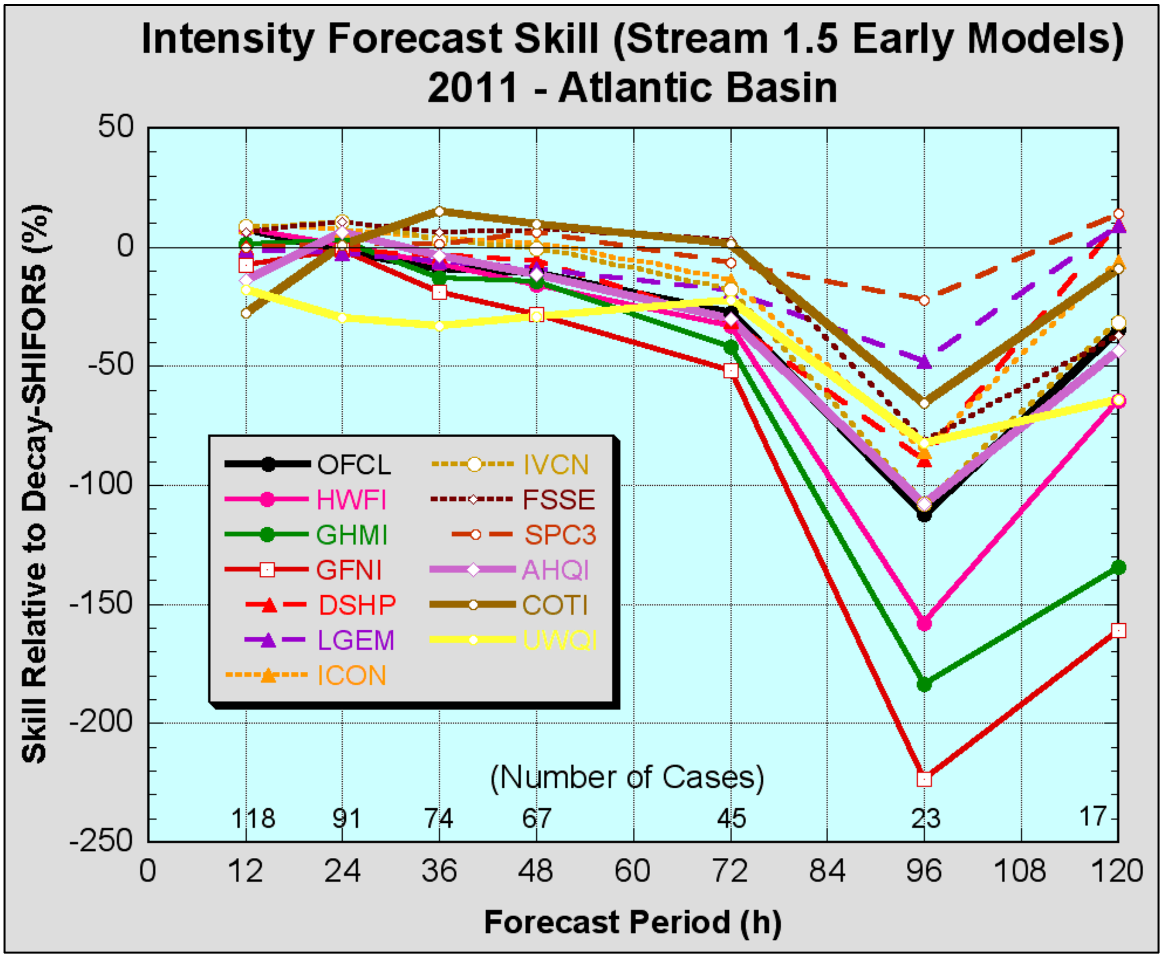


Figure 19. Homogeneous comparison of HFIP Stream 1.5 intensity models and selected operational models for 2011.

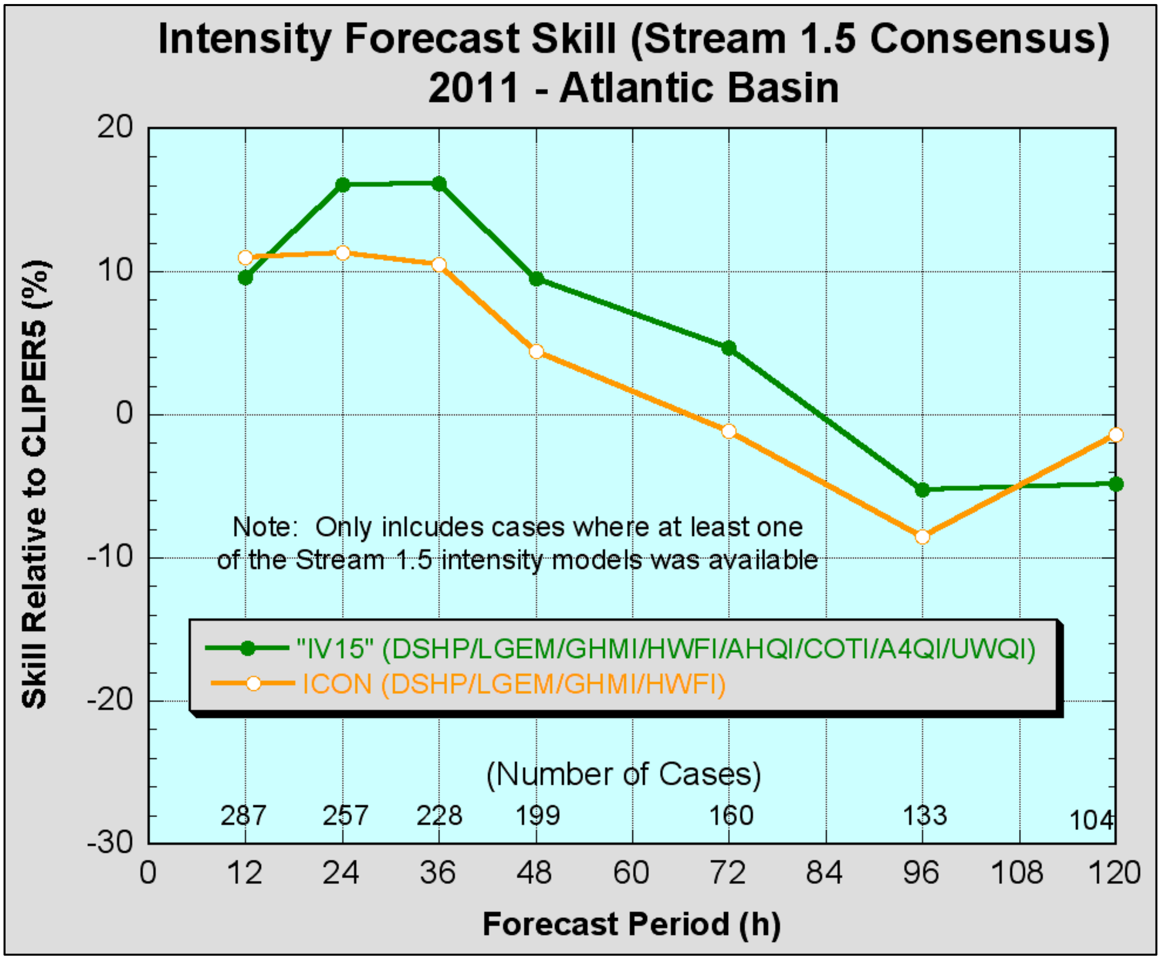


Figure 20. Impact of adding Stream 1.5 models to the fixed intensity consensus ICON.