

## Assessing the Skill of Operational Atlantic Seasonal Tropical Cyclone Forecasts

BRIAN F. OWENS\*

*University of Miami/RSMAS, Miami, Florida*

CHRISTOPHER W. LANDSEA

*NOAA/AOML/Hurricane Research Division, Miami, Florida*

(Manuscript received 21 May 2001, in final form 9 August 2002)

### ABSTRACT

Since 1984, W. Gray of Colorado State University and a team of researchers have been issuing seasonal tropical cyclone forecasts for the North Atlantic Ocean. Prior to this, little work had been done in the area of long-term tropical cyclone forecasting because researchers saw minimal potential skill in any prediction models and no obvious benefits to be gained. However, seasonal forecasts have been attracting more attention as economic and insured losses from hurricane-related catastrophes rose sharply during recent decades. Initially, the forecasts issued by Gray consisted of output from simple statistical prediction models. Over time, the models became increasingly more complex and sophisticated, with new versions being introduced in 1992, 1993, 1994, 1996, and 1997. In addition, based on a combination of experience with the statistical models and other qualitative considerations such as examinations of analog years, the statistical forecasts were modified to create adjusted seasonal forecasts. This analysis assessed the skill demonstrated, if any, of both the statistical and adjusted forecasts over the benchmarks of climatology and persistence and examined whether the adjusted forecasts were more accurate than the statistical forecasts. The analysis indicates that, over the past 18 yr, both the statistical and adjusted forecasts demonstrated some skill over climatology and persistence. There is also evidence to suggest that the adjusted forecast was more skillful than the statistical model forecast.

### 1. Introduction

Hurricanes rank among the most destructive and costliest of natural phenomena (Pielke and Landsea 1998). In 1992, a new watermark was reached when Hurricane Andrew caused \$26.5 billion in losses in south Florida (Jarrell et al. 2001). Even more alarming, total loss estimates for a category-4 hurricane striking the Miami region were revised upward to over \$60 billion (Pielke and Landsea 1998). There has been much speculation concerning the reasons for the sharp increase in damage estimates, even adjusting for inflation. In 1995, the U.S. Senate issued a report expressing concern at the increasing severity of Atlantic storms (U.S. Senate 1995), a belief not supported by the meteorological record (Landsea et al. 1996). According to Pielke and Landsea (1998), the principal reasons for the increase in damages were demographics of population and wealth increases.

---

\* Current affiliation: Risk Management Solutions Limited, London, United Kingdom.

---

*Corresponding author address:* Brian F. Owens, Risk Management Solutions Limited, 10 Eastcheap, London EC3M 1AJ, United Kingdom.  
E-mail: brian.owens@rms.com

The trend in damages highlights the importance of better hurricane forecasting as a tool to be used by emergency managers, insurance commissioners, state and federal legislators, insurance and construction industry executives, and the general public.

An important product to emerge over the past two decades has been the seasonal tropical cyclone forecast for the North Atlantic Ocean. Seasonal forecasts of tropical cyclone activity are used by corporations and governments to better prepare for hurricane activity and generate considerable interest among the general public, thereby prompting greater awareness. Development of seasonal forecast models has also given researchers greater insight into the linkages that may exist in our climate system, and provides the possibility that future models may be improved upon over time. Several parties have either issued or issue seasonal forecasts, including Elsner and colleagues (Elsner and Schmertmann 1993), Gray and colleagues (Gray et al. 1984–2001), and the National Oceanic and Atmospheric Administration (NOAA; [www.nhc.noaa.gov](http://www.nhc.noaa.gov)). Of these parties, only Gray's group has produced at least 10 yr of real-time seasonal forecasts, and this analysis focused on these forecasts only.

Gray initially began issuing seasonal Atlantic hurricane forecasts in 1984 (Gray 1984a,b) and has since

TABLE 1. Current statistical model forecast predictive variable groups. The number of available predictive variables for each group is in parentheses.

Dec
Quasi-biennial oscillation (three)
West African rainfall (two)
Forecast ENSO conditions (four)
Azores SLP anomalies (one)
Apr
Quasi-biennial oscillation (four)
West African rainfall (two)
Current ENSO conditions (three)
Azores SLP anomalies (two)
Atlantic SST anomalies (two)
Jun
Quasi-biennial oscillation (three)
West African rainfall (two)
Current ENSO conditions (two)
Caribbean SLP anomalies and 200-hPa winds (two)
Azores SLP anomalies (two)
Atlantic SST anomalies (three)
African Sahel temperature gradient (one)
Aug
Quasi-biennial oscillation (three)
West African rainfall (two)
Current ENSO conditions (two)
Caribbean SLP anomalies and 200-hPa winds (two)
Azores SLP anomalies (two)
Atlantic SST anomalies (three)
African Sahel temperature gradient (one)
Named storm days in the Tropics (one)

assembled a team of hurricane researchers to assist in the forecasting process (Gray et al. 1992, 1993, 1994). Early models were simple: forecasts were issued in early June and August of each year and were based on statistical adjustments to climatological data. Over time, the statistical scheme was revised and expanded (Gray et al. 1984–2001; Landsea et al. 1994). From 1992, a longer-term forecast was issued in December of the year preceding the forecast year (Gray et al. 1992) and, from 1995, an updated version of this forecast was issued in the following April. Many new measures of tropical activity have been included in the latest versions of the models; however, this analysis focuses on the frequency forecasts for the following measures:

- named storms, a tropical cyclone with maximum sustained surface winds of  $\geq 18 \text{ m s}^{-1}$ ;
- hurricanes, a tropical cyclone with maximum sustained surface winds of  $\geq 33 \text{ m s}^{-1}$ ;
- major hurricanes, a hurricane with maximum sustained surface winds of  $\geq 50 \text{ m s}^{-1}$ .

Up to 16 predictor variables were available to the statistical model, including several El Niño–Southern Oscillation (ENSO)-related variables, and various sea surface temperature (SST) and sea level pressure (SLP) variables. Not all variables, however, were used in each forecast. In Table 1, the predictor variables for each of

TABLE 2. Number of years for which tropical cyclone activity forecasts are available.

Month of forecast issue	Named storms	Hurricanes	Major hurricanes
Dec	10 (1992–2001)	10 (1992–2001)	10 (1992–2001)
Apr	7 (1995–2001)	7 (1995–2001)	7 (1995–2001)
Jun	18 (1984–2001)	18 (1984–2001)	12 (1990–2001)
Aug	18 (1984–2001)	18 (1984–2001)	12 (1990–2001)

the four annually issued forecasts are categorized into a number of predictive groups.

The statistical forecasts were adjusted each year to generate adjusted forecasts. The adjustments were made to account for factors that were not explicitly addressed in the statistical models, for example, regime shifts in tropical activity (Landsea et al. 1999), and were based on a better understanding of both the contributing variables and the known biases within the models. From 1984 to 1991, the adjustments were smaller than in subsequent years, and from 1994, adjustments were also based on an examination of tropical cyclone activity in years similar to the year being forecast, that is, an analog-years technique.

Many of the parties referenced earlier, that is, emergency managers, the insurance and construction industries, legislators, and the public, factor the seasonal forecasts into their decision-making processes and, as such, would generally find them useful [M. Ernst, Office of the U.S. Foreign Disaster Assistance (USAID), 2002, personal communication]. A higher standard test was required, however, to genuinely assess the skill, that is, statistically significant improvements, of Gray et al.'s statistical and adjusted forecasts over the benchmarks of climatology and persistence (defined in section 2) and to examine whether the adjusted forecasts improved upon the statistical forecasts. The analysis continued work initially performed by Landsea (2000) in which some skill in the seasonal forecasting models through 1996 was determined to exist.

## 2. Data sources

Forecasts, issued in June and August, for the number of named storms, hurricanes, and major hurricanes in each year available were tabulated. In Table 2 (Fig. 5a), the number of years of available data for each of these activity measures is outlined. To assess the skill of the model forecasts produced by Gray, it was first necessary to establish appropriate benchmarks against which to compare model forecasts. Two benchmarks were selected: climatology and persistence.

The first benchmark, climatology, was defined by the mean seasonal activity for the years from 1944 [the first year for which complete and reliable tropical cyclone records are considered to exist for the North Atlantic (Neumann et al. 1999)] to the year prior to that for which

the first forecast for a particular tropical cyclone activity measure was issued. For named storms and hurricanes, this period was defined for the 40 yr from 1944 to 1983. For major hurricanes, climatology was defined for the 46 yr from 1944 to 1989. The defined climatological values for named storms, hurricanes, and major hurricanes were 9.2, 5.7, and 2.2 systems per season, respectively. A rolling-year version of the climatology was also tested, in which the climatological values were updated each year to include the prior year's observed activity. This did not substantially alter the results of the analysis. The climatology data were based on the format outlined in Jarvinen et al. (1984) and were supplemented by data from Neumann et al. (1999). The second benchmark, persistence, is a simple forecasting technique in which the forecast activity for each tropical cyclone activity measure in a given year was equal to the prior year's observed activity for that measure.

### 3. Methodology

Two sets of forecast data were analyzed. In the case of named storms and hurricanes, these were the statistical and adjusted forecast data from 1984 to 2001, and in the case of major hurricanes, for which forecasts have been made only since 1990, these were the statistical and adjusted forecast data from 1990 to 2001. For each tropical cyclone activity measure, comparisons with observed activity were made over corresponding time periods. The mean and standard deviation of observed activity were calculated on a seasonal basis.

A number of different analyses were performed using the data. As explained below, each analysis provided its own insight, from which conclusions on the relative strengths and weaknesses of the statistical and adjusted forecasts were drawn. An underlying assumption of these analyses is that the data were distributed normally. Therefore, the data were first tested for normality, and the following analyses were considered appropriate given the results of these tests:

- natural categories analysis—tropical cyclone activity was analyzed from a frequency perspective, that is, the differences between observed and predicted activity were aggregated into unit frequency intervals;
- root-mean-square error (rmse) analysis—the rmse of the annual differences between observed and predicted activity was compared; and
- regression analysis—observed tropical cyclone activity was regressed against that forecast using each of persistence, the statistical forecast, and the adjusted forecast to determine which forecasting method explained the greatest amount of variance in observed activity. The amount of variance explained was expressed by the resulting  $r^2$  value—higher  $r^2$  values indicated a stronger predictive relationship between forecast and observed activity.

In the case of the rmse analyses, the results were

examined for statistical significance at the 90th and 95th percentiles, using an  $F$ -statistic test for each of the following comparisons: statistical forecast versus climatology, adjusted forecast versus climatology, and adjusted forecast versus statistical model. The more rigorous benchmark, that is, climatology, was selected over persistence for the statistical comparisons.

For the regression analyses, the results were examined for statistical significance at the 90th and 95th percentiles, using a one-tailed  $t$ -value test for each of the following: observed activity regressed against persistence, observed activity regressed against the statistical forecast, and observed activity regressed against the adjusted forecast. The relative performance of each of the forecasts and persistence, as measured by an increase in  $r^2$  values, was also tested for statistical significance using a one-tailed  $t$ -value test, as follows: increase in  $r^2$  value using the statistical forecast over persistence, increase in  $r^2$  value using the adjusted forecast over persistence, and increase in  $r^2$  value using the adjusted forecast over the statistical forecast.

As a general illustration of hurricane activity levels and forecasts since 1984, Fig. 1 illustrates a histogram of statistical and adjusted forecasts, persistence, and observed activity for the number of hurricanes predicted in the June and August forecasts. The horizontal lines in this figure indicate climatology. The period from 1984 to 2001 saw a wide variety in hurricane seasons, from very busy in 1995 (11 hurricanes) and 1998 (10 hurricanes) to very quiet in 1987, 1994, and 1997 (3 hurricanes each).

### 4. Results and discussion

#### a. Natural categories analysis

This analysis aggregated the differences between the observed tropical cyclone activity and that predicted using climatology, persistence, and the statistical and adjusted forecasts into frequency intervals, each interval representing a unit error between forecast and observed activity. There were three sets of intervals, one each for named storms, hurricanes, and major hurricanes, as outlined in Tables 3a, 3b, and 3c, respectively, in which each interval frequency was expressed as a percentage of the total number of forecast years. The intervals were cumulative, for example, the interval into which all errors of  $\pm 2$  named storms were grouped also included all errors of  $\pm 1$  named storms. As an illustration of the errors being considered, Fig. 2 outlines the errors between the observed number of hurricanes and the number of hurricanes predicted using persistence, and both the statistical and adjusted forecasts. The purpose of this analysis is to give the reader a sense of how often errors of various magnitudes occurred with each forecasting method.

For named storm forecasts, persistence had a higher frequency of smaller errors than climatology, that is,

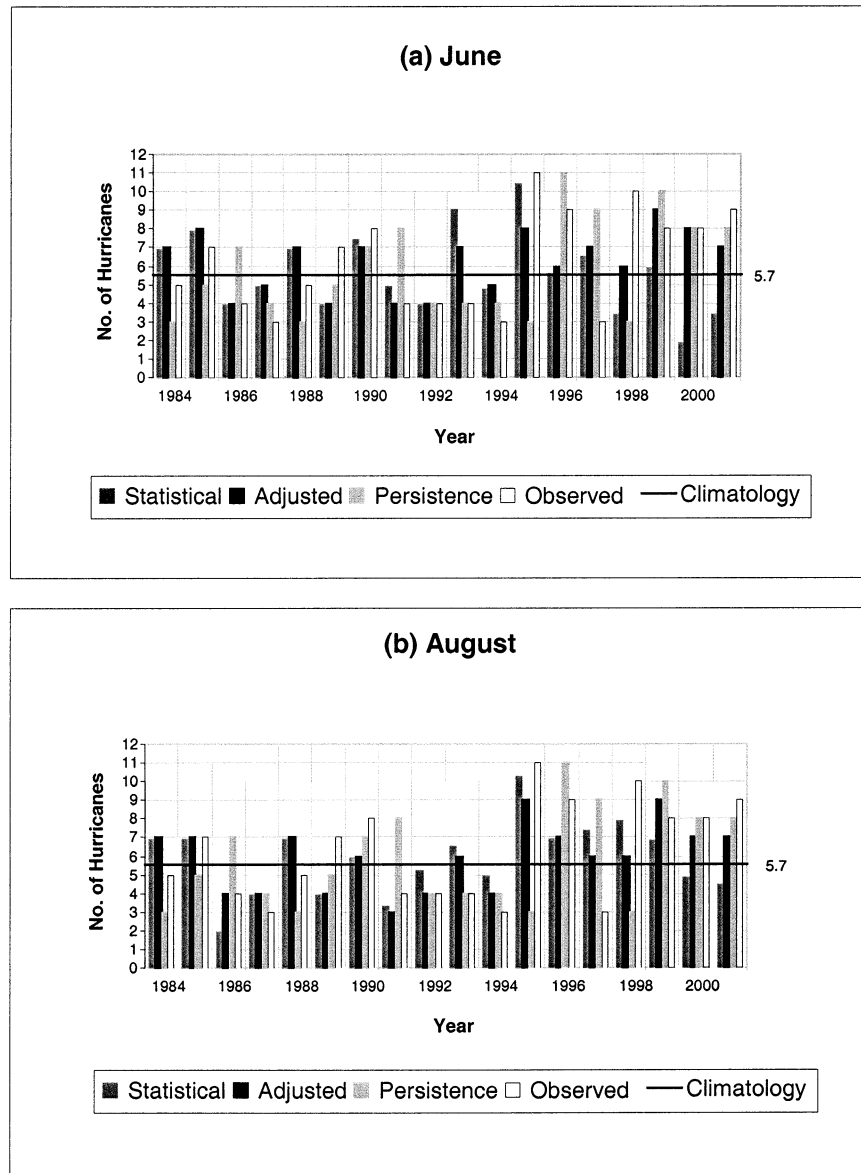


FIG. 1. Comparison of the number of observed hurricanes with the number of hurricanes predicted by each of the statistical forecast, the adjusted forecast, and persistence, for the forecasts issued in (a) Jun and (b) Aug.

$\leq 2$ , reflecting a number of forecasts in which persistence benefited from successive years of low and high activity. However, in years in which there were sudden shifts in the levels of named storm activity, for example, from 1994 to 1995 and 1996 to 1997, persistence generated very large errors, that is,  $\geq 4$ , which climatology did not, and these were also reflected in their respective frequency intervals. The statistical and adjusted forecasts in June and August almost always had a higher frequency of lower magnitude errors than for the corresponding forecasts using either climatology or persistence, with the exception of errors of  $\pm 1$  named storms using the June statistical and adjusted forecasts.

In general, the maximum magnitude errors between the statistical and adjusted forecasts were also less than those of climatology and persistence, again with the exception of the June statistical forecast, when the maximum errors were equal. The adjusted forecasts performed particularly well, having a maximum error of  $\pm 4$  named storms, and had a much lower incidence of higher-magnitude errors than the statistical forecasts.

In the case of hurricane forecasts, both the statistical and adjusted forecasts had more low-error forecasts than climatology. The statistical forecast, however, particularly for June forecasts, tended to have a lower frequency of smaller errors than persistence, reflecting a

TABLE 3. Results of the natural categories analysis, outlining the percentage of times that climatology, persistence, the Jun forecasts, and the Aug forecasts each were within a number of units of the observed activity of (a) named storms, (b) hurricanes, and (c) major hurricanes.

(a) Named storms	$\pm 0$	$\pm 1$	$\pm 2$	$\pm 3$	$\pm 4$	$\pm 5$	$\pm 6$	$\pm 7$	$\pm 8$
Climatology (%)	0	0	22	56	72	89	94	94	100
Persistence	0	28	50	56	56	67	83	94	100
Jun									
Statistical (%)	0	11	50	56	78	78	83	94	100
Adjusted	11	22	56	78	94	94	94	100	100
Aug									
Statistical (%)	6	33	61	67	72	89	94	100	100
Adjusted	11	63	67	89	100	100	100	100	100
(b) Hurricanes	$\pm 0$	$\pm 1$	$\pm 2$	$\pm 3$	$\pm 4$	$\pm 5$	$\pm 6$	$\pm 7$	$\pm 8$
Climatology (%)	0	11	44	78	89	94	100	10	100
Persistence	17	39	72	78	83	83	89	94	100
Jun									
Statistical (%)	0	33	56	61	78	83	89	100	100
Adjusted	22	39	67	89	100	100	100	100	100
Aug									
Statistical (%)	0	22	50	78	89	100	100	100	100
Adjusted	17	44	83	94	100	100	100	100	100
(c) Major hurricanes	$\pm 0$	$\pm 1$	$\pm 2$	$\pm 3$	$\pm 4$	$\pm 5$			
Climatology (%)	0	25	67	92	100	100			
Persistence	8	58	83	83	83	100			
Jun									
Statistical (%)	0	50	58	83	92	100			
Adjusted	8	67	92	92	100	100			
Aug									
Statistical (%)	0	25	67	92	100	100			
Adjusted	17	75	92	100	100	100			

series of low activity years, for example, the early 1990s, or a series of high activity years, for example, the late 1990s, from which persistence as a forecasting method benefited. Persistence had a higher frequency of high-magnitude errors also, reflecting regime shifts in activity, for example, from 1994 to 1995. The statistical forecast in August had a maximum error of  $\pm 5$  hurricanes, which compared favorably to both climatology and persistence. The adjusted forecast clearly had the highest frequency of low-magnitude errors, relative to the statistical forecast, climatology, and persistence. Furthermore, all forecast years were accounted for by the fifth interval, that is, the maximum error in any year was only  $\pm 4$  hurricanes.

Last, for major hurricane forecasts, climatology performed well with a maximum error of  $\pm 4$  major hurricanes, which occurred only once. Persistence, in a pattern that is repeated for all of the tropical cyclone activity measures, had a high frequency of lower errors as well as some very high errors. In two of the years, the error between persistence and observed activity was 5 or more major hurricanes, a very large number given that the average number in a typical year was about 2.2. The adjusted forecast always had the highest frequency of smaller errors for forecasts issued in August. This was true also for June forecasts, except for the number

of zero-error forecasts. Finally, when compared to the statistical forecast, the adjusted forecast always had a higher number of forecasts in each interval.

*b. Root-mean-square error analysis*

In the next analysis, the rmse of climatology, persistence, the statistical forecast, and the adjusted forecast, relative to observed activity, were compared. The results are outlined in Table 4. In this analysis, a forecast in which the rmse was less than that of climatology, being the most appropriate “best guess” of tropical cyclone activity, was subjectively considered to be useful.

The rmse of climatology was always lower than that of persistence, reflecting the large errors generated from time to time when using persistence, particularly when transitions from quiet to active years, for example, 1994 to 1995, or from active to quiet years, for example, 1996 to 1997, occurred. It had a higher rmse than the adjusted forecast, particularly for the number of named storms predicted by the June and August forecasts, but also with respect to hurricane and major hurricane activity predicted in the August forecasts. For the number of hurricanes and major hurricanes predicted by the June seasonal forecast models, climatology performed con-



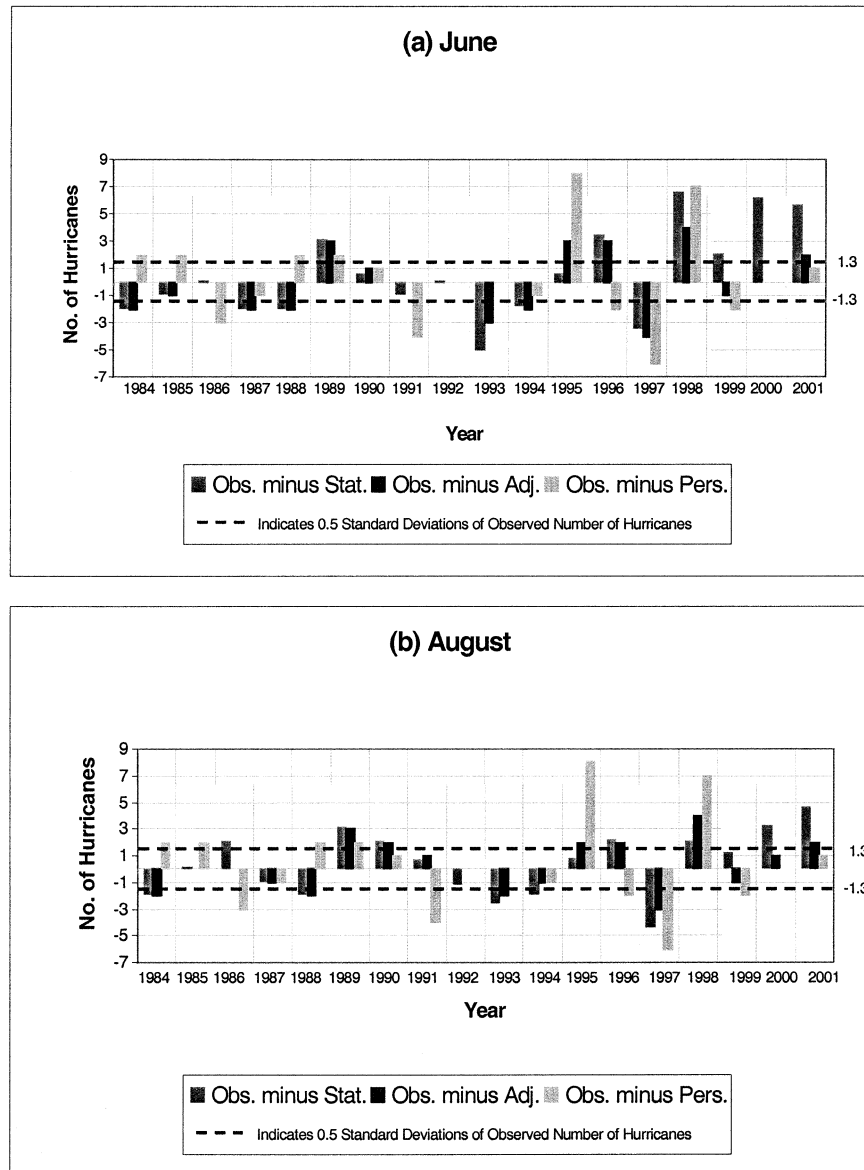


FIG. 2. Comparison of the differences between the mean number of observed hurricanes and the number of hurricanes predicted by each of the statistical forecast, the adjusted forecast, and persistence, for the forecasts issued in (a) Jun and (b) Aug.

TABLE 4. Rmse relative to observed activity for climatology, persistence, the Jun forecasts, and the Aug forecasts.

	Named storms	Hurricanes	Major hurricanes
Climatology	4.0	2.6	1.9
Persistence	5.0	3.4	2.4
Jun			
Statistical	4.0	3.3	2.3
Adjusted	3.0	2.2	1.7
Aug			
Statistical	3.0	2.4	1.8
Adjusted	2.3	1.9	1.4

siderably better than the statistical forecast, but not as well as the adjusted forecast.

For each forecast category of tropical cyclone activity, persistence had the highest rmse, particularly with respect to the number of named storms. In this case, the rmse was over 25% higher than that of climatology, at least 25% higher than that of the model forecasts issued in June, and almost 67% higher than the rmse of the model forecasts issued in August.

In all cases the rmse of the adjusted forecast was less than that of the statistical model. However, for the number of major hurricanes predicted by the June and August forecasts, and the number of named storms and

TABLE 5. Statistical results of the rmse analysis, outlining  $F$  statistics and their level of statistical significance, if any, for the forecasts issued in Jun and Aug. Numbers  $<1$  indicate that the rmse of the second forecast is less than the rmse of the first forecast.

	Named storms	Hurricanes	Major hurricanes
Jun			
Statistical vs climatology	0.99	0.63	0.69
Adjusted vs climatology	1.71	1.34	1.29
Adjusted vs statistical	1.73	2.11*	1.88
Aug			
Statistical vs climatology	1.69	1.22	1.21
Adjusted vs climatology	2.93**	1.82	1.89
Adjusted vs statistical	1.74	1.49	1.55

\* Significant at the 90% level.  
 \*\* Significant at the 95% level.

hurricanes predicted by the August forecasts, the differences were all 0.7 or less. In the remaining cases, that is, for the forecasts of named storm and hurricane activity issued in June, the adjusted forecast was substantially superior to the statistical forecast, that is, it had an rmse of at least one unit less than that of the statistical forecast.

The rmse analysis results were tested for statistical significance, using an  $F$ -statistic test. The results of these statistical tests are outlined in Table 5. In this situation, climatology was the more rigorous standard and was selected over persistence as the appropriate benchmark against which to compare the statistical and adjusted forecasts.

For the number of named storms forecast, the improvement in rmse of the statistical and adjusted forecasts over climatology was not significant for forecasts made in June, but was significant at the 95% level in the case of the adjusted forecast made in August. In all other cases, while some of the rmse differences between the statistical and adjusted forecasts and climatology were substantial (many were significant at the 85% level), none was significant at the 90% level.

For predicted hurricane activity, neither the statistical forecast nor the adjusted forecast provided any statistically significant improvement over climatology. As noted above, the adjusted forecast consistently outperformed the statistical model forecast from an rmse perspective. Again, while many of these improvements were significant at the 85% level, only the improvement for hurricane forecasts issued in June was significant at the 90% level.

With respect to predicted major hurricane activity, although the statistical and adjusted forecasts were both better than climatology, neither forecast provided an improvement that was statistically significant at the 90% level.

*c. Regression analysis*

The final analysis performed, the results of which are outlined in Table 6, regressed persistence, the statistical

TABLE 6. Results of the statistical regression analyses against observed activity, outlining  $r^2$  values and their level of statistical significance, if any, using a one-tailed  $t$  test, for persistence, the Jun forecasts, and the Aug forecasts.

	Named storms	Hurricanes	Major hurricanes
Persistence	0.01	0.02	0.04
Jun			
Statistical	0.03	0.00	0.01
Adjusted	0.36**	0.24**	0.21*
Aug			
Statistical	0.42**	0.24**	0.25**
Adjusted	0.65**	0.44**	0.54**

\* Significant at the 90% level.  
 \*\* Significant at the 95% level.

forecast, and the adjusted forecast against observed activity to determine which forecast's measures of tropical cyclone activity, if any, correlated most strongly with observed activity. A forecast that correlated with observed activity with an  $r^2$  of 0.10 or greater in the case of named storms and hurricanes, and 0.17 or greater in the case of major hurricanes, was considered skillful, being the minimum  $r^2$  required to be significant at the 90% level. The skill threshold was higher in cases of major hurricanes to account for the shorter dataset and fewer degrees of freedom. Given the relatively short lengths of the datasets being analyzed, forecasts were subjectively considered to be useful if they satisfied the criterion for being skillful.

The correlations between persistence and observed activity, as represented by the  $r^2$  values, were very low, accounting for, at most, less than 4% of the variance in annual observed activity. Figure 3 shows a scatterplot of persistence versus observed hurricane activity with the resulting regression line illustrated.

Figures 4 and 5 illustrate the results of the regression

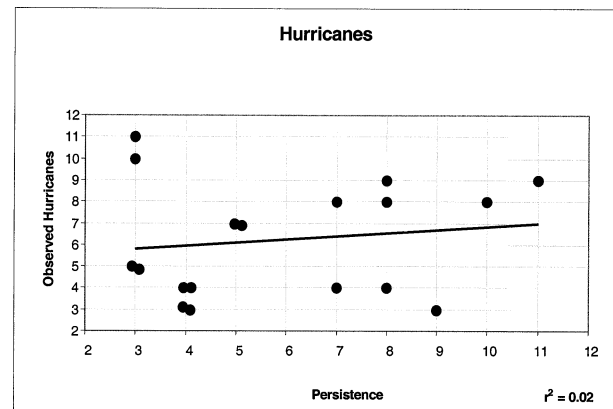


FIG. 3. Regression of the number of observed hurricanes against the number of hurricanes predicted using persistence. (Note: In Figs. 3–5, some points have been offset slightly from their actual position so that all data may be viewed more clearly. This does not alter the  $r^2$  values presented.)

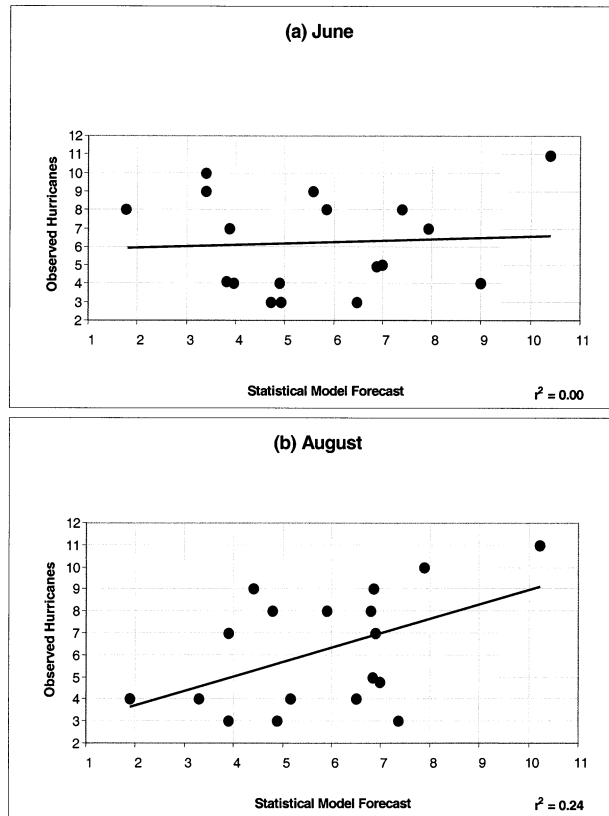


FIG. 4. Regression of the number of observed hurricanes against the number of hurricanes predicted by the statistical forecast, for the forecasts issued in (a) Jun and (b) Aug.

analyses for the statistical and adjusted forecasts of hurricane activity. With the exception of Fig. 4a, the slopes of the trend lines in these plots are close to 1, indicating an approximate one-to-one relationship between predicted hurricanes and observed hurricanes.

In the case of June forecasting schemes, the  $r^2$  values were generally smaller than those for August. The statistical forecasts issued in June had very low  $r^2$  values, none of which was significant (e.g., see Fig. 4a). In contrast, the adjusted forecasts performed well, particularly for the number of named storms and hurricanes (Fig. 5a) forecast, accounting for approximately 36% and 24% of the variance in observed activity, respectively. These explanations of variances were each significant at the 95% level. For major hurricane forecasts, the June adjusted model explains 21% of the variance in observed activity, significant at the 90% level.

For seasonal forecasts issued in August, the forecast models, in particular the adjusted forecast, were significantly more skillful than those issued in June. The adjusted forecast accounted for 65% of the variance in observed activity for the number of named storms. The adjusted forecast also explained 44% of the variance in observed activity in the case of the number of hurricanes forecast (Fig. 5b), and 54% of the variance in observed

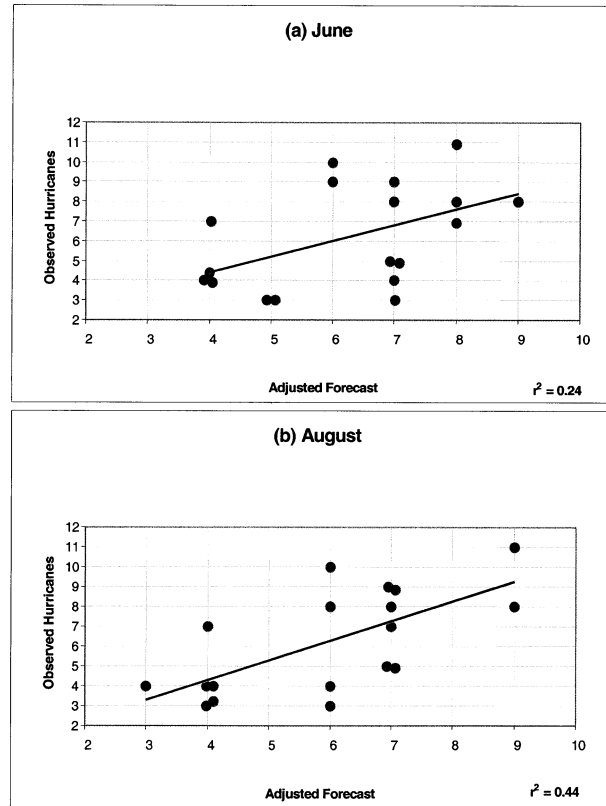


FIG. 5. Regression of the number of observed hurricanes against the number of hurricanes predicted by the adjusted forecast, for the forecasts issued in (a) Jun and (b) Aug.

major hurricane activity. These  $r^2$  values were all significant at the 95% level. The statistical forecasts, while not as highly correlated with observed activity as the adjusted forecasts, were nonetheless skillful predictors of observed activity. With respect to the number of named storms forecast, the statistical forecast accounted for almost 42% of the variance of observed activity. It also accounted for 24% of the variance with respect to hurricane activity (Fig. 4b), and 25% of the variance in major hurricane activity. These values were all statistically significant at the 95% level.

The improvement in  $r^2$  values was also interesting when comparing persistence to both the statistical and adjusted forecasts. This was a useful consideration when trying to determine which of the forecasts, that is, which of persistence, the statistical forecast, or the adjusted forecast, if any, provided the greatest improvement over each of the remaining forecasts. In this case, an analysis of the improvement of the statistical forecast over persistence, and the improvements of the adjusted forecast over both the statistical forecast and persistence, was performed. In each case, the improvement was measured by the corresponding increase or decrease in  $r^2$  values. The results of this analysis are outlined in Table 7.

From these results, it is clear that, with respect to the number of named storms predicted by the June and Au-



TABLE 7. Improvement in  $r^2$  values and their level of statistical significance, if any, using a one-tailed  $t$  test for the forecasts issued in Jun and Aug. Negative numbers indicate that the second forecast improves upon the first forecast.

	Named storms	Hurricanes	Major hurricanes
Jun			
Statistical vs persistence	0.02	-0.02	-0.03
Adjusted vs persistence	0.35**	0.22**	0.17*
Adjusted vs statistical	0.33**	0.24**	0.20*
Aug			
Statistical vs persistence	0.41**	0.22**	0.21*
Adjusted vs persistence	0.64**	0.42**	0.50**
Adjusted vs statistical	0.24**	0.20**	0.29**

\* Significant at the 90% level.  
 \*\* Significant at the 95% level.

gust seasonal forecasts, the statistical and adjusted forecasts generally represented an improvement over persistence. With the exception of the statistical forecast issued in June, the improvement was always significant at the 95% level.

In the case of forecast hurricane activity, the improvements in  $r^2$  of the statistical forecast were significant only with respect to forecasts of seasonal activity issued in August, at the 95% level. For forecasts made using the adjusted model, the improvements over persistence were again significant at the 95% level for forecasts issued both in June and August.

For the number of major hurricanes forecast, persistence was slightly better than the statistical model for forecasts issued in June, but this difference was not statistically significant. In all other cases, however, the improvements of the statistical and adjusted forecasts over persistence were statistically significant. In the case of the adjusted forecast issued in August, the improvement over persistence was significant at the 95% level.

It is clear from the results that the adjusted forecast was always more accurate than the statistical forecast. These improvements were all statistically significant at the 95% level except for the major hurricane forecasts issued in June, when the significance level was 90%.

**5. Conclusions**

The purpose of this study was to determine if Gray et al.'s statistical and adjusted seasonal forecasts were more skillful than some simple benchmarks, and to investigate if the adjusted forecast improved upon the purely statistical forecast. Three separate analyses were performed, comparing the number of named storms, hurricanes, and major hurricanes predicted by the statistical and adjusted forecasts with the benchmarks of climatology and persistence. In addition, the statistical and adjusted forecasts were also compared with each other.

The first analysis performed was the natural category

TABLE 8. Summary results of the three analyses performed, showing that the adjusted model (AM) had the highest skill in each of the three tropical cyclone activity measures when compared with the statistical model and the benchmarks.

	Named storms	Hurricanes	Major hurricanes
Jun			
Natural categories analysis	AM	AM	AM
Rmse analysis	AM	AM	AM
Regression analysis	AM**	AM**	AM**
Aug			
Natural categories analysis	AM	AM	AM
Rmse analysis	AM*	AM	AM
Regression analysis	AM**	AM**	AM**

\* Significant over the benchmark of climatology/persistence at the 90% level.  
 \*\* Significant over the benchmark of climatology/persistence at the 95% level.

analysis. There was no testing of the results for statistical significance but there are some clear observations that can be made. First, the adjusted forecast in almost all cases had the highest frequency of smaller errors and was certainly the best forecasting technique using this analysis. Second, the statistical model in general had quite a large number of medium-size errors (in the three to five range for named storms and hurricanes, two to four range for major hurricanes), which limited its usefulness as a forecasting technique. Third, persistence had high frequencies of small and very large errors, reflecting how persistence was useful during times when activity did not change much from year to year, but experienced very large errors when activity shifted sharply between years.

In the rmse analysis, the statistical and adjusted forecasts appeared to be most skillful in the forecasting of named storms when held to the more rigorous standard benchmark of climatology; this was particularly true for the adjusted forecasts. Additionally, the adjusted forecast nearly always had smaller errors or variances (relative to observed activity) than those of the statistical forecast. This suggests that the adjusted forecast was in general better than the statistical forecast, but the differences between them were seldom significant at the 90% level.

In the regression analysis, however, the adjusted forecast always explained a statistically significant larger portion of the variance in observed tropical cyclone activity than the statistical forecast, for forecasts issued in both June and August. The same was true of the adjusted forecast with respect to persistence. For the statistical forecasts, the regression analysis was statistically significantly better than persistence only for forecasts issued in August, and was no better than persistence in June (both were very poor predictors of variance in observed activity).

The three analyses highlighted the strengths of each forecasting method and are summarized in Table 8. In

many cases, the statistical and adjusted forecasts provided statistically significant improvements over the benchmarks of persistence and climatology, particularly in the regression analysis, and therefore demonstrated skill. However, it is clear that, of all the forecasts, the adjusted forecast was consistently the best. In summary, we observed the following:

- depending on the analysis performed, either climatology or persistence provided the best benchmark for determining skill in seasonal forecasting;
- relative to the benchmarks, the statistical forecasts issued in June were in general not very skillful; however, the forecasts issued in August showed substantial skill, particularly in the regression analysis;
- in contrast, both the June and August adjusted forecasts were better than the benchmark forecasts; many of these improvements were significant at the 90% level or higher;
- when comparing the statistical and adjusted forecasts, the adjusted forecasted was almost always superior to the statistical forecast; and
- seasonal forecast skill was highest for the predicted number of named storms and was somewhat lower for the predicted number of hurricanes and major hurricanes.

*Acknowledgments.* The authors thank Dr. Bruce Albrecht and Dr. Chidong Zhang of the University of Miami/RSMAS and three anonymous reviewers for their valuable comments during the preparation of this paper. Brian F. Owens is grateful for the funding provided through the Rosenstiel Fellowship from the University of Miami/RSMAS.

#### REFERENCES

- Elsner, J. B., and C. P. Schmertmann, 1993: Improving extended-range seasonal predictions of intense Atlantic hurricane activity. *Wea. Forecasting*, **8**, 345–351.
- Gray, W. M., 1984a: Atlantic seasonal hurricane frequency: Part I: El Niño and 30 mb quasi-biennial oscillation influences. *Mon. Wea. Rev.*, **112**, 1649–1668.
- , 1984b: Atlantic seasonal hurricane frequency: Part II: Forecasting its variability. *Mon. Wea. Rev.*, **112**, 1669–1683.
- , C. W. Landsea, P. W. Mielke Jr., and K. J. Berry, 1984–2001: Seasonal Atlantic hurricane forecasts and verifications. Department of Atmospheric Science, Colorado State University, Fort Collins, CO. [Available online at <http://tropical.atmos.colostate.edu/forecasts/>.]
- , ———, ———, and ———, 1992: Predicting Atlantic seasonal hurricane activity 6–11 months in advance. *Wea. Forecasting*, **7**, 440–455.
- , ———, ———, and ———, 1993: Predicting Atlantic basin seasonal tropical cyclone activity by 1 August. *Wea. Forecasting*, **8**, 73–86.
- , ———, ———, and ———, 1994: Predicting Atlantic basin seasonal tropical cyclone activity by 1 June. *Wea. Forecasting*, **9**, 103–115.
- Jarrell, J. D., M. Mayfield, E. N. Rappaport, and C. W. Landsea, 2001: The deadliest, costliest, and most intense United States hurricanes from 1900 to 2000 (and other frequently requested hurricane facts). NOAA Tech. Memo. NWS TPC-3, Miami, FL, 29 pp.
- Jarvinen, B. R., C. J. Neumann, and M. A. S. Davis, 1984: A tropical cyclone data tape for the North Atlantic Basin, 1886–1983: Contents, limitations, and uses. NOAA Tech. Memo. NWS NHC 22, Miami, FL, 21 pp.
- Landsea, C. W., 2000: El Niño–Southern Oscillation and the seasonal variability of tropical cyclones. *El Niño: Impact of Multiscale Variability on Natural Ecosystems and Society*, H. F. Diaz and V. Markgraf, Eds., Cambridge University Press, 149–181.
- , W. M. Gray, P. W. Mielke Jr., and K. J. Berry, 1994: Seasonal forecasting of Atlantic hurricane activity. *Weather*, **49**, 273–284.
- , N. Nicholls, W. M. Gray, and L. A. Avila, 1996: Downward trends in the frequency of intense Atlantic hurricanes during the past five decades. *Geophys. Res. Lett.*, **23**, 1697–1700.
- , R. A. Pielke Jr., A. M. Mestas-Núñez, and J. A. Knaff, 1999: Atlantic basin hurricanes: Indices of climatic changes. *Climatic Change*, **42**, 89–129.
- Neumann, C. J., B. R. Jarvinen, C. J. McAdie, and G. R. Hammer, 1999: Tropical cyclones of the North Atlantic Ocean, 1871–1998. Historical Climatology Series, No. 6-2, National Climatic Data Center, Asheville, NC (in cooperation with the National Hurricane Center, Miami, FL), 206 pp.
- Pielke, R. A., Jr., and C. W. Landsea, 1998: Normalized hurricane damages in the United States: 1925–95. *Wea. Forecasting*, **13**, 621–631.
- U.S. Senate, 1995: Federal disaster assistance. Report of the Bipartisan Task Force on Funding Disaster Relief, 104th Cong., 250 pp.