

The Prediction of the Performance of a Nested Barotropic Hurricane Track Forecast Model

SIM D. ABERSON

Hurricane Research Division, NOAA/AOML, Miami, Florida

(Manuscript received 24 August 1995, in final form 13 August 1996)

ABSTRACT

Linear multiple regression and discriminant analyses provide estimates of the errors of track forecasts from a nested barotropic hurricane track forecast model (VICBAR), which was run in the North Atlantic Basin during the 1989–94 hurricane seasons. Predictors are determined from the synoptic situation, the magnitude of atmospheric changes in the environment of the tropical cyclone, the consistency between current and past predictions, and the past performance of the model for each particular storm. This technique distinguishes cases in which VICBAR performs well from those for which it performs poorly and can provide skillful operational predictions of model performance to forecasters.

1. Introduction

Tropical cyclone track prediction continues to be one of the most difficult problems facing forecasters. The various operational hurricane track forecast models commonly provide skillful predictions, but none performs well consistently, and they often give solutions that diverge widely from each other (Sheets 1990). A significant task, then, is to supply accurate estimates of the track forecast errors for each model in addition to the track forecasts themselves.

Attempts to predict the accuracy or skill of forecast models began with the stochastic dynamic modeling technique (Epstein 1969). Leith (1974) approximated this with a less-expensive Monte Carlo method that examined the sensitivity of each particular forecast to small perturbations in the initial conditions; ensembles of forecasts with similar solutions were shown to be relatively skillful. Both of these techniques, however, require large amounts of computer resources.

Several simpler and less-expensive methods for the prediction of forecast skill have been developed. Hoffman and Kalnay (1983) introduced lagged average forecasting, in which model runs initialized at different synoptic times and verifying at the same time are compared; as in the Monte Carlo method, sets of forecasts that are similar are likely to be skillful. Palmer and Tibaldi (1986) discovered that variables derived from the current synoptic situation could predict the skill of model runs, and Kistler et al. (1988) found that forecasts of

cases in which the atmosphere is slowly changing tend to be more skillful than others in which the atmosphere changes more rapidly. Palmer and Tibaldi (1988) listed the methods for predicting forecast skill, noting the four important types of predictors: those derived from 1) the synoptic situation, 2) the amount of persistence in the atmosphere, 3) the consistency of forecasts initialized at different times and verifying at the same time, and 4) the recent past performance of the model.

These four types of variables are used for the prediction of model performance (POMP) of a barotropic hurricane track forecast model (VICBAR). Aberson and DeMaria (1994, hereafter AD) showed that during the 1989–93 Atlantic hurricane seasons VICBAR had skill that was competitive with the best prediction models available but that the forecasts were not consistently good. The version of VICBAR described in AD supplies the forecast sample for this study. For a detailed description of VICBAR model results, the reader is referred to that paper. A discussion of the predictors used in this study is given in section 2. Descriptions of the discriminant analyses that provide POMP are given in section 3, with results in section 4.

2. Predictors used for POMP

Linear multiple regression and discriminant analyses are used to examine the possibility of providing an operational estimate of the performance of each VICBAR track forecast. All potential predictors tested for POMP are summarized in Table 1. Aberson and DeMaria (1994) showed that a number of synoptic and climatological variables can be important in operationally assessing the performance of individual VICBAR fore-

Corresponding author address: Sim D. Aberson, 8, NOAA/ERL/HRD/AOML, 4301 Rickenbacker Causeway, Miami, FL 33149.
E-mail: aberson@aoml.noaa.gov

TABLE 1. Predictors tested by the multiple linear regression and linear discriminant analyses. Predictors positively correlated with forecast error are marked with a plus sign, those negatively correlated with a minus sign, and those not chosen by the regression with an "N." Predictors marked with an asterisk are calculated 0–500 km from the current storm center. Those marked with a double asterisk are calculated 500–1000 km from the current storm center. Mean values (in km) of the consistency and past performance predictors are given. The left- and right-hand columns signify use in the 0000 and 1200 UTC and the 0600 and 1800 UTC samples, respectively.

(A) Synoptic predictors				
1) Latitude	+	+		
2) Longitude	–	–		
3) Initial intensity	N	–		
4) Julian date	N	N		
5) Difference between initial direction of motion and westward motion	+	N		
6) Magnitude difference between CSMV and DLM steering**	N	+		
7) Magnitude of the 850–200-mb wind shear*	+	+		
8) Thermal advection measurement*	+	+		
9) Magnitude of the horizontal deformation*	N	+		
10) Cumulative distance between BAM forecasts at each forecast time	+	+		
11) Forecast latitude at each forecast time	+	+		
12) Forecast longitude at each forecast time	–	–		
13) Forecast difference between current motion and westward motion	+	N		
(B) Persistence predictors				
1) Initial speed of motion	+	+		
2) 12-h intensity change	–	N		
3) 12-h change in storm motion	+	+		
4) Distance between NHC90 early and late forecasts at each forecast time	+	–		
5) Forecast speed of motion at each forecast time	+	+		
(C) Consistency predictors				
1) Distance between 12-h-old, 24-h, and current 12-h forecast positions	N	+	134	137
2) Distance between 12-h-old, 36-h, and current 24-h forecast positions	N	N	179	176
3) Distance between 12-h-old, 48-h, and current 36-h forecast positions	N	+	227	218
4) Distance between 24-h-old, 36-h, and current 12-h forecast positions	N	+	224	231
5) Distance between 24-h-old, 48-h, and current 24-h forecast positions	N	+	287	288
6) Distance between 24-h-old, 72-h, and current 48-h forecast positions	N	+	427	459
7) Distance between 36-h-old, 48-h, and current 12-h forecast positions	N	N	325	332
8) Distance between 36-h-old, 72-h, and current 36-h forecast positions	+	+	489	526
9) Distance between 48-h-old, 72-h, and current 24-h forecast positions	N	N	519	557
10) Distance between 60-h-old, 72-h, and current 12-h forecast positions	N	N	537	582
(D) Past performance predictors				
1) Most recent 12-h forecast error	+	N	90	96
2) Most recent 24-h forecast error	+	+	163	172
3) Most recent 36-h forecast error	+	+	243	257
4) Most recent 48-h forecast error	N	N	343	351
5) Most recent 72-h forecast error	+	+	537	587

casts. For example, VICBAR has smaller forecast errors for tropical cyclone cases that were initially of hurricane strength than for those of weaker storms, probably because weak storms tend to be steered by the flow averaged over a layer shallower than the 850–200-mb deep layer mean (DLM) used by VICBAR (Velden 1993). Forecasts made in early summer tend to be better than those of late fall, since baroclinic processes reach farther into tropical regions during the latter period. Storms originating in the southern areas of the basin tend to have smaller errors than those farther north, because the atmosphere is more likely to be barotropic in lower latitudes than farther north. A weak relationship between the size of VICBAR errors and the initial longitude of tropical cyclones is also noted in AD, since data coverage is more extensive over the American continent than over the Atlantic Ocean. Also, storms with large northward or eastward current storm motions (those storms that are recurving or have recurved) tend

to have larger forecast errors than those moving southward or westward. Therefore, initial intensity, Julian date, latitude, longitude, and the difference between the initial direction of storm motion and due west are included as potential predictors; the VICBAR-forecasted latitude, longitude, and direction of storm motion are also included. However, the regression does not choose the Julian date because it is highly correlated with some of the other potential predictors.

Aberson and DeMaria (1994) also showed that the majority of the very worst forecasts were due to high vertical shear of the horizontal wind in the vicinity of the tropical cyclone. Therefore, the difference between the flow at 850 and 200 mb within 500 km of the storm center is used to predict VICBAR forecast accuracy. In addition, a measure of the baroclinicity of the atmosphere (the magnitude of the temperature advection) within 500 km of the current storm center is tested as a potential predictor. Baroclinicity is associated with the

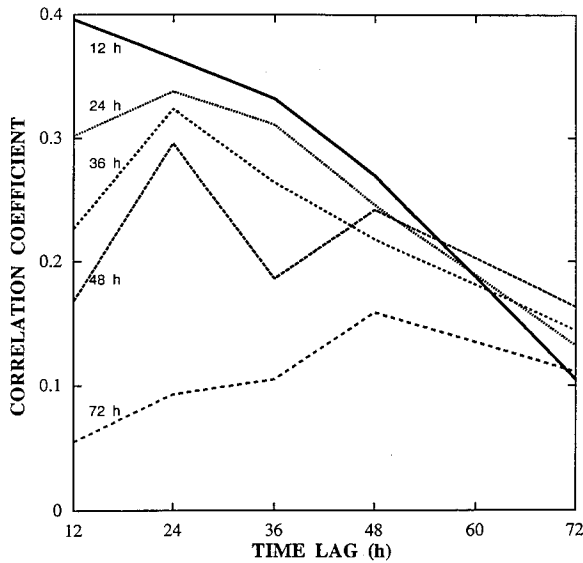


FIG. 1. Correlation coefficients between lagged forecasts for 0000 and 1200 UTC VICBAR forecasts.

vertical shear of the geostrophic wind. The most important sources of wind shear in the Atlantic tropical cyclone basin are the amplification of middle-latitude upper-tropospheric troughs into the Tropics and upper-tropospheric cyclonic circulations. To a first approximation, both of these can be considered geostrophic so, assuming geostrophic and hydrostatic balance in the vicinity of the tropical cyclone, this predictor is given by the equation

$$\int_{p_T}^{p_B} |V \cdot \nabla T| dp \cong \frac{f}{R} \int_{p_T}^{p_B} p \left| u \frac{\partial v}{\partial p} - v \frac{\partial u}{\partial p} \right| dp, \quad (1)$$

where p_T and p_B are the top and bottom of the DLM, respectively; f is the Coriolis parameter; R the ideal gas constant; and u and v the zonal and meridional wind components, respectively. Also, the beta and advection model (BAM, Marks 1992) is run for three different levels in the troposphere (shallow-layer, middle-layer, and deep-layer mean). The distance between the three BAM model forecasts at each forecast time may provide information on the future vertical shear in the vicinity of the tropical cyclone.

Other types of potential predictors not discussed by AD are also included. Cases in which the current storm motion vector (CSMV) closely matches the DLM flow between 500 and 1000 km from the storm center are likely to have smaller forecast errors than those in which the difference between these two vectors is large. (The large annulus is chosen because the VICBAR initialization removes all data within 500 km of the storm center and replaces them with a synthetic vortex, an axisymmetric vortex with the CSMV added.) Also, tropical cyclones embedded in highly deformed DLM flow may be more difficult to forecast than those in less-deformed flow. A measure of the total horizontal de-

formation (D) of the flow is given by the combination of the shearing deformation (H) and stretching deformation (T):

$$\begin{aligned} H &= \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \\ T &= \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \\ D &= (T^2 + H^2)^{1/2}, \end{aligned} \quad (2)$$

where x and y are the zonal and meridional distances, respectively.

The predictors that describe the amount of persistence in the atmosphere are the initial and VICBAR-forecast forward speed of the storm, the 12-h intensity change of the storm, and the 12-h change in the CSMV. Additionally, the distance between track forecasts from the two versions of the NHC90 model (McAdie 1991) initialized at the same time, one run from the 12-h-old global spectral model run and the other from the current run, can quantify the amount of persistence in the atmosphere. Cases in which any of these values, except the intensity change, are large may have large track forecast errors. Curiously, the correlation between the difference between the two NHC90 runs and forecast error is positive for the 0000 and 1200 UTC cases, and negative for the 0600 and 1800 UTC cases, and both correlations are found to be statistically significant at the 90% level.

All of the above predictors are available for every synoptic time except the vertical shear, the temperature advection, and the NHC90 predictor, which are available only for 0000 and 1200 UTC cases. The previous values of these predictors are used for the 0600 and 1800 UTC forecasts.

The consistency between current and past VICBAR forecasts may provide an indication of future model performance. Figure 1 shows the correlation coefficients of time-lagged VICBAR forecasts. The correlations never surpass 0.4 and are generally highest around 24 h, possibly due to many rawinsonde sites in the Caribbean Basin reporting only every 24 h. The calculation of the statistical separation time between forecasts (AD) also confirms that correlations between successive forecasts are not very large. Despite these relatively small correlations, all possible combinations of the lagged-average predictors, and also those related to the recent performance of VICBAR, are tested by the linear regression analysis, though only some are included in the final analyses. If the necessary previous VICBAR run is not available, the value of the mean of the particular predictor is substituted. Table 1 shows the mean values of the lagged-average and recent past performance predictors.

TABLE 2. Classification tables for VICBAR forecast errors for both the 0000 and 1200 UTC samples (left) and the 0600 and 1800 UTC samples (right) for the linear multiple regression analysis. Columns signify the predicted group; rows the actual group. Here, N is the number of cases, and S is the skill score for each table.

	0000 and 1200 UTC			0600 and 1800 UTC		
	Good	Average	Poor	Good	Average	Poor
12 h	(N = 567, S = 0.07)			(N = 520, S = 0.12)		
Good	15.5	14.3	3.4	18.5	7.9	6.5
Average	11.6	9.0	12.7	10.0	16.8	7.9
Poor	5.8	10.1	17.6	6.5	7.2	18.7
24 h	(N = 527, S = 0.10)			(N = 498, S = 0.08)		
Good	15.6	13.7	4.0	22.4	5.5	5.0
Average	12.0	10.6	10.8	12.1	13.6	8.0
Poor	5.3	8.9	19.2	7.5	7.3	18.6
36 h	(N = 422, S = 0.03)			(N = 453, S = 0.05)		
Good	15.9	11.6	5.7	13.2	12.0	6.6
Average	12.6	12.1	8.5	10.2	15.7	8.2
Poor	4.7	9.5	19.4	6.6	8.8	18.7
48 h	(N = 382, S = 0.10)			(N = 407, S = 0.08)		
Good	15.2	11.3	7.1	18.2	8.0	5.7
Average	13.4	10.5	8.6	11.7	14.8	6.8
Poor	4.5	11.3	18.1	7.4	7.7	19.7
72 h	(N = 328, S = 0.10)			(N = 318, S = 0.15)		
Good	13.2	13.2	6.8	16.5	10.0	6.0
Average	14.2	9.8	9.1	8.4	17.7	6.8
Poor	5.7	10.1	17.9	7.6	8.4	18.6

3. Methods for POMP

a. Linear multiple regression analyses

Multiple linear regression analyses are performed on the 5-yr (1989–93) sample of VICBAR runs to predict the performance of individual forecasts. Because 0000 and 1200 UTC VICBAR forecasts have different error characteristics than those from 0600 and 1800 UTC, the regression analyses of the two sets are done separately. A stepwise regression technique includes those predictors that are statistically significant at the 90% level at any of the five forecast times (12, 24, 36, 48, and 72 h). The most important predictors in both samples are the initial speed of motion of the storm, the initial thermal advection in the vicinity of the tropical cyclone, the VICBAR-forecasted latitude of the storm, and the forecast horizontal wind shear in the vicinity of the storm (BAM dispersion predictor). Additional important predictors in the 0000 and 1200 UTC sample are the 12- and 24-h forecast errors from the most recent VICBAR forecast available, the initial horizontal wind shear in the vicinity of the storm, and the VICBAR-forecasted direction of motion of the storm. Additional important predictors in the 0600 and 1800 UTC samples are the initial and VICBAR-forecasted longitudes of the storm, the initial deformation of the DLM flow in the vicinity of the storm, and the VICBAR-forecasted speed of motion of the storm.

Neumann et al. (1977) note that the significance of

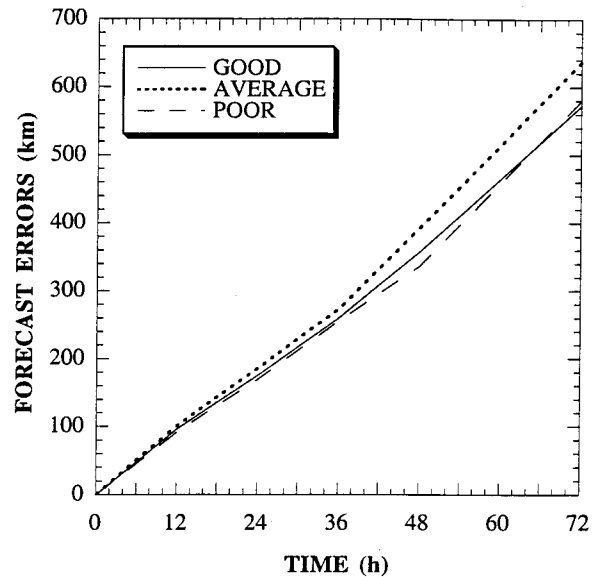


FIG. 2. Average forecast errors for those forecasts that the multiple linear regression analyses predict the forecasts will be good, average, or poor, for the dependent dataset, 0000 and 1200 UTC samples.

the predictors can be spuriously inflated due to the large number of predictors tested. Despite this, the regression analyses generally explain less than one-third of the variance of the errors at the different forecast times, and thus few forecasts of errors deviate significantly from the mean. However, distinguishing cases in which the errors are above or below average still may be possible. Actual forecast errors and those predicted by the regression analyses are each stratified into three categories (good, average, and poor) with equal numbers of cases in each group. Classification tables and a skill score (appendix), shown in Table 2, assess the ability of the regression analyses to group the forecasts correctly. If the classification were random, 33% of the cases would be grouped accurately, and 22% would be predicted good when really poor, and vice versa. The regression analyses are able to classify 44% (0000 and 1200 UTC samples) and 46% (0600 and 1800 UTC samples) of the cases correctly, and only 9% (0000 and 1200 UTC samples) and 12% (0600 and 1800 UTC samples) of the cases are forecast to be good when they are in fact poor, and vice versa. All classification tables are significantly different from chance at the 95% level using the chi-square test (appendix), yet the skill scores are only modest. The regression analysis therefore performs better than a chance forecast.

Figure 2 shows the average actual forecast errors of the cases that are predicted to be good, average, and poor (0000 and 1200 UTC cases only). The regression scheme is unable to clearly distinguish the good, average, and poor cases. A one-sided test of the hypothesis that the means of the errors in the predicted-good group is smaller than in the other two groups, and that the mean of the predicted-average group is lower than that

TABLE 3. Classification tables for VICBAR forecast errors for both the 0000 and 1200 UTC samples (left) and the 0600 and 1800 UTC samples (right) for the linear discriminant analysis. Columns signify the predicted group; rows the actual group. Here, N is the number of cases, and S is the skill score for each table.

	0000 and 1200 UTC			0600 and 1800 UTC		
	Good	Average	Poor	Good	Average	Poor
12 h	(N = 567, S = 0.19)			(N = 520, S = 0.22)		
Good	17.8	9.3	5.8	18.8	7.7	7.3
Average	11.6	13.9	7.8	11.0	14.0	9.2
Poor	7.0	8.8	17.8	5.2	7.3	19.4
24 h	(N = 527, S = 0.19)			(N = 498, S = 0.23)		
Good	17.8	9.1	5.9	20.3	7.2	5.0
Average	14.0	11.0	8.2	11.0	12.4	10.0
Poor	7.2	5.3	21.4	7.2	7.2	19.5
36 h	(N = 484, S = 0.22)			(N = 453, S = 0.23)		
Good	18.4	11.6	5.4	18.5	6.4	7.9
Average	10.5	13.2	7.2	9.5	12.8	11.0
Poor	7.0	8.3	18.4	6.0	8.6	19.2
48 h	(N = 438, S = 0.24)			(N = 407, S = 0.24)		
Good	14.8	12.6	5.7	19.4	7.6	5.4
Average	9.6	15.1	8.9	11.3	16.2	6.1
Poor	5.0	5.9	22.4	4.9	8.8	20.1
72 h	(N = 328, S = 0.28)			(N = 318, S = 0.32)		
Good	21.0	9.4	3.7	18.9	9.4	4.7
Average	11.6	14.6	5.8	8.2	17.9	7.2
Poor	8.2	6.4	19.5	6.9	7.9	18.9

of the predicted-poor group, is performed for these samples. The 0000 UTC forecasts are tested separately from the 1200 UTC forecasts, as are the 0600 and 1800 UTC forecasts, to reduce the effect of serial correlation in the dataset. None of the differences between the means of the groups are statistically significant at any forecast times. Despite the modest skill and the fact that the classification tables are significantly better than chance tables, the regression technique is unsuitable for POMP.

b. Linear discriminant analyses

Unlike the linear regression analyses, linear discriminant analyses (Lachenbruch 1975) are specifically designed to classify cases into different groups (Allen and Marshall 1994). The discriminant analysis classifies each forecast as either good, average, or poor, by maximizing the differences between the three groups based upon values of the predictors. The predictors that the multiple linear regression analysis chose as most effective (Table 1) are used in the linear discriminant analysis.

Table 3 shows the discriminant analysis classification tables, and Fig. 3 shows the average actual forecast errors for the cases predicted to be in the three groups (0000 and 1200 UTC samples only). For both samples, more than half of the cases are correctly classified, and only about one-tenth of the cases are classified as good or poor when they are actually poor or good, respec-

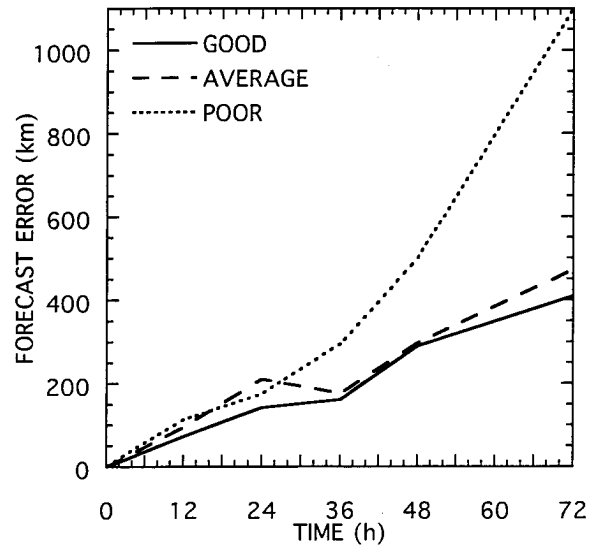


FIG. 3. Average forecast errors for those forecasts that the linear discriminant analyses predict the forecasts will be good, average, or poor, for the dependent dataset, 0000 and 1200 UTC samples.

tively. The skill is much higher than the multiple regression analyses in all cases, and the chi-square test shows that all classification tables are statistically significantly different than chance at the 95% level. The differences between the means of the predicted-poor and predicted-average groups in Fig. 3 are statistically significant at all times. The differences between the means of the predicted-good and predicted-average groups are only significant at 36 and 72 h in the 0000 and 1200 UTC samples. This scheme therefore seems best suited to distinguishing cases that are likely to be poorly forecast from others. Linear discriminant analyses seem better able to provide accurate POMP than regression analyses.

These results would not be attainable in an operational setting. Cases from the 1994 hurricane season are used as an independent dataset to test how well this scheme might perform operationally. Table 4 shows classification tables for all forecast times for this independent dataset. These discriminant analyses correctly classify 40%–42% of both samples. Classification as good when poor, and vice versa, occurs 8% and 18% of the time in the 0000 and 1200 UTC and the 0600 and 1800 UTC samples, respectively. The early forecast times of both samples have no skill, but the skill increases to 15%–17% in the 0000 and 1200 UTC sample by 48 and 72 h but is lower in the 0600 and 1800 UTC samples. These skill values are comparable to those from the dependent sample in the linear multiple regression. The classification tables for these later times are statistically significantly different from chance at the 95% level. Statistical significance tests also show that the differences between the means of the predicted-average and predicted-poor groups at all times except 72 h, and between the predicted-good and predicted-poor types at all times

TABLE 4. Classification tables for VICBAR forecast errors for both the 0000 and 1200 UTC samples (left) and the 0600 and 1800 UTC samples (right) for the independent sample for the linear discriminant analysis. Columns signify the predicted group; rows the actual group. Here, N is the number of cases, and S is the skill score for each table.

	0000 and 1200 UTC			0600 and 1800 UTC		
	Good	Average	Poor	Good	Average	Poor
12 h	(N = 90, S = 0.00)			(N = 90, S = 0.00)		
Good	6.7	8.9	3.3	6.7	4.4	3.3
Average	16.7	13.3	13.3	12.2	10.0	13.3
Poor	6.7	11.1	20.0	14.4	15.6	14.4
24 h	(N = 83, S = 0.01)			(N = 75, S = 0.00)		
Good	8.4	4.8	2.4	12.0	2.7	6.7
Average	15.7	13.3	14.5	12.0	1.3	12.0
Poor	6.5	18.1	19.3	25.3	12.0	16.0
36 h	(N = 73, S = 0.01)			(N = 69, S = 0.16)		
Good	12.3	5.5	6.8	10.1	2.9	1.4
Average	16.4	16.4	12.3	7.2	5.8	1.4
Poor	1.4	15.1	13.7	21.7	18.8	30.4
48 h	(N = 67, S = 0.17)			(N = 62, S = 0.14)		
Good	6.0	6.0	1.5	9.7	1.6	3.2
Average	20.9	13.4	10.4	17.7	14.5	16.1
Poor	4.5	14.9	22.4	6.5	12.9	17.7
72 h	(N = 56, S = 0.15)			(N = 48, S = 0.10)		
Good	8.9	7.1	3.6	4.2	6.3	2.1
Average	14.3	19.6	12.5	16.7	22.9	14.6
Poor	5.4	10.7	17.9	4.2	10.4	18.8

are statistically significant in the 0000 and 1200 UTC samples (Fig. 4). Similarly, the differences between the means of the predicted-average and predicted-poor groups at 36, 48, and 72 h, and between the predicted-good and predicted-poor groups at all forecast times, are statistically significant in the 0600 and 1800 UTC samples. Therefore, the scheme is best able to distinguish the worst cases from the others in this relatively small independent sample. Since forecasters have previously had no method to accurately discern poor model forecasts from others, this tool can be especially helpful in operational forecasting.

Since the value of forecasts that are inconsistent from one forecast time to the next may be uncertain, all forecasts from the independent sample reaching 72 h are examined for consistency. Only about 7% of the predictions varied from good to poor, or vice versa, between forecast times. Therefore, the results of the discriminant analyses are usually consistent through the forecasts and may present an accurate prognosis of the ability of each VICBAR forecast to predict the track of tropical cyclones.

4. Summary

A linear discriminant analysis scheme using predictors that are important in forecasting VICBAR model behavior has been developed to assess the performance of individual VICBAR forecasts operationally. This

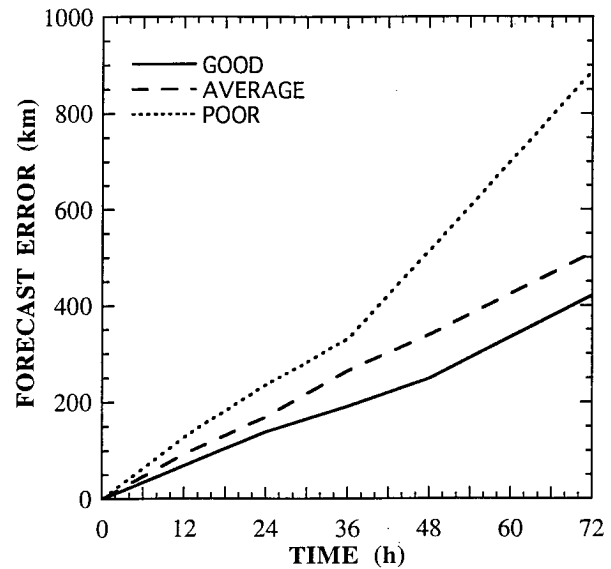


FIG. 4. Average forecast errors for those forecasts that the linear discriminant analyses predict the forecasts will be good, average, or poor, for the independent dataset, 0000 and 1200 UTC samples.

technique uses predictors based upon the synoptic situation, the persistence of the atmosphere, the consistency of model runs initialized at different times but verifying at the same time, and the previous performance of the model. These analyses are effective in predicting the model's performance, and such forecasts can be provided with the actual prediction to enable the hurricane forecaster to better utilize the forecasts. Discriminant analyses are being developed for the model output of the other operational hurricane track forecast models in the Atlantic basin.

Acknowledgments. The author wishes to thank Dr. Lloyd Shapiro for his valuable suggestions during the creations of the POMP scheme and Dr. Robert Burpee for his encouragement during the work. Additional reviews of the manuscript by James Franklin and three anonymous reviewers have helped to make the manuscript clearer and more accessible.

APPENDIX

Skill Scores and Statistical Independence Tests

A skill score

$$S = (C - E)/(T - E), \tag{A1}$$

where C is the number of correct forecasts, T is the total number of forecasts, and E is the number of forecasts expected to be correct based on chance, can be calculated for each classification table (Panofsky and Brier 1958). The skill score will be unity if all cases are correctly predicted, and negative for no skill.

A chi-square test can be used to discern whether the classification table is statistically significantly different

from chance, assuming that each entry in the table is independent of all others. Chi square is defined as

$$\chi^2 = \sum \frac{(O - H)^2}{H}, \quad (\text{A2})$$

where H stands for the numbers in a chance classification table and O represents the observation classification table. The number of degrees of freedom is $(m - 1)$, where m is the number of rows or columns in the classification table.

REFERENCES

- Aberson, S. D., and M. DeMaria, 1994: Verification of a nested barotropic hurricane track forecast model (VICBAR). *Mon. Wea. Rev.*, **122**, 2804–2815.
- Allen, G., and J. F. Le Marshall, 1994: An evaluation of neural networks and discriminant analysis methods for application in operational rain forecasting. *Aust. Meteor. Mag.*, **43**, 17–28.
- Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus*, **21**, 739–759.
- Hoffman, R. N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus*, **35A**, 100–118.
- Kistler, R. E., E. Kalnay, and M. S. Tracton, 1988: Forecast agreement, persistence, and forecast skill. Preprints, *Eighth Conf. on Numerical Weather Prediction*, Baltimore, MD, Amer. Meteor. Soc., 641–646.
- Lachenbruch, P. A., 1975: *Discriminant Analysis*. Hafner Press, 128 pp.
- Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418.
- Marks, D. G., 1992: The beta and advection model for hurricane track forecasting. NOAA Tech. Memo. NWS NMC 70, 89 pp. [Available from National Center for Environmental Prediction, 5200 Auth Rd., Camp Springs, MD 20746-4304.]
- McAdie, C. J., 1991: A comparison of tropical cyclone track forecasts produced by NHC90 and an alternate version (NHC90A) during the 1990 hurricane season. Preprints, *19th Conf. on Hurricanes and Tropical Meteorology*, Miami, FL, Amer. Meteor. Soc., 290–294.
- Neumann, C. J., M. B. Lawrence, and E. L. Caso, 1977: Monte Carlo significance testing as applied to statistical tropical cyclone prediction models. *J. Appl. Meteor.*, **16**, 1165–1174.
- Palmer, T. N., and S. Tibaldi, 1986: Forecast skill and predictability. ECMWF Tech. Memo. 139. [Available from ECMWF, Shinfield Park, Reading, RG2 9AX, United Kingdom.]
- , and —, 1988: On the prediction of forecast skill. *Mon. Wea. Rev.*, **116**, 2453–2480.
- Panofsky, H. A., and G. W. Brier, 1958: *Some Applications of Statistics to Meteorology*. The Pennsylvania State University, 224 pp.
- Sheets, R. C., 1990: The National Hurricane Center—Past, present, and future. *Wea. Forecasting*, **5**, 185–232.
- Velden, C., 1993: The relationship between tropical cyclone motion, intensity, and the vertical extent of the environmental steering layer in the Atlantic basin. Extended Abstracts, *20th Conf. Hurricanes and Tropical Meteorology*, San Antonio TX, Amer. Meteor. Soc., 31–34.