ACCURACY OF ATLANTIC TROPICAL CYCLONE FORECASTS

JACK D. TRACY

National Hurricane Research Laboratory, Environmental Science Services Administration, Miami, Fla.

ABSTRACT

During the past several years a number of techniques have been developed for forecasting the motion of tropical cyclones over areas of the Atlantic, Caribbean, and Gulf of Mexico. Many of these have been tested on an operational basis. These forecasts have been collected and verified and probability ellipses have been constructed. While direct comparisons between the forecast methods are difficult (because of inhomogeneity of sample size, differences in times of availability of the forecasts, etc.), it has been found that two of the objective techniques tested result in slightly better forecasts than the others. The accuracy of the forecasts depends somewhat on the geographical area and the forecast error varies almost linearly with the length of the forecast period.

1. INTRODUCTION

A continuing and developing program for the evaluation of the performance of the various techniques utilized in predicting the future course of hurricanes has been in existence at the National Hurricane Center (NHC), Miami, Fla., for several years. Sufficient data have now been accumulated to permit meaningful conclusions to be drawn as to the relative merits of some of the numerous systems which have been used to forecast hurricane motion. In this paper, attention will be focused on the performance of eight such techniques during the years 1959–64, although not all methods were used throughout the entire period. The first part deals with a number of statistical measures of the forecast errors. The second part presents verification information in the form of probability ellipses for four levels of probability, namely, 20, 40, 60, and 80 percent.

It is not the purpose of this paper to discuss in detail the forecast techniques which have been evaluated, since most of these have been described elsewhere [4, 5, 8, 9, 10]. However, in order to facilitate reading of the paper a brief summary of the main features of each method will be given. The methods range from simple straight-line extrapolation to numerical prediction with a simple barotropic model. Several are statistical in nature and were derived by a screening and multiple linear regression technique. As predictors, these use such parameters as sea level pressures or heights of constant pressure surfaces, geostrophic wind components, the past motion of the cyclone center, thicknesses between constant pressure surfaces, and height changes measured in a moving coordinate system. The official forecasts issued by the Weather Bureau Forecast Centers represent the professional judgment and experience of the hurricane forecasters, and are essentially subjective in nature, although consideration is given to the numerical forecasts, objective aids, climatology, etc. The basic characteristics of each method are listed in table 1.

In verifying the forecasts a "best track" prepared by the Hurricane Forecast Centers was used, the writer having no part in determining this track. With one exception, only those forecasts prepared at the time of the event for operational forecast purposes were verified. In an earlier paper the writer [7] described a set of "simulated" operational forecasts prepared by use of the T–60 method. The verification statistics for this method in-

TABLE 1.—*Forecast methods which were verified. Symbols in column 2 have the following meaning:*

P_x, P_y	Past motion of the cyclone center
P	Sea level pressure
H_7	700-mb. height
Z_5	500-mb. height
u, v	geostrophic wind component
$\Delta Z_{10/7}$	1000–700-mb. thickness
$\Delta Z_{7/5}$	700–500-mb. thickness
$\delta H_7/\delta t$	700-mb. height change
$\delta Z_5/\delta t$	500-mb. height change

Method	Predictors used	Forecast periods
Official Weather Bureau Advisory (WB)	Professional judgment plus numerical products, objective aids, and climatology.	12, 24, 48, 72 hr.
Numerical Weather Prediction (NWP)	Barotropic model (500 mb.)	12 to 72 hr. by 12-hr. steps.
Travelers Synoptic Climatological 1959 (T–59)	P, P_x, P_y	24 hr.
Travelers-National Hurricane Research Laboratory-1960 (T–60)	P, P_x, P_y, Z_5, u, v	24 and 36 hr.
National Hurricane Center-1964 (NHC–64)	P, P_x, P_y, H_7, Z_5, u, v, $\Delta Z_{10/7}$, $\Delta Z_{7/5}$, $\delta H_7/\delta t$, $\delta Z_5/\delta t$	12 to 48 hr. by 12-hr. steps.
Riehl-Haggard-Sanborn (RHS)	u, v (500 mb.)	24 hr.
Miller-Moore (MM)	u, v (700 mb.), P_x, P_y, H_7, $\delta H_7/\delta t$	24 hr.
Persistence (PERS)	P_x, P_y	12 to 72 hr. by 12-hr. steps.

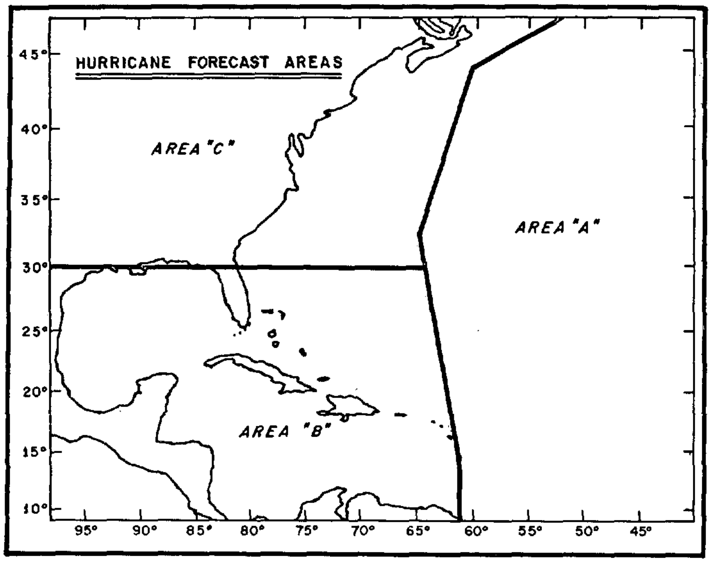

FIGURE 1.—Hurricane forecast areas.

clude the simulated forecasts for the years 1959–60 in addition to those prepared for the other years. Most of the statistical measures were obtained by computer methods, which virtually eliminated the possibility for computational errors.

Previous studies by Gentry [3], Dunn [2], and Tracy [7] have indicated that the accuracy of hurricane forecasts varies from one geographical area to another. For this reason the verification statistics have been stratified according to area (fig. 1). Area B has the largest number of oceanic upper-air stations and area A has the fewest. Differences in the verification statistics have been attributed to the differences in the data density, although other factors (such as climatology) may contribute to these differences.

Before a detailed examination of the statistical quantities, some comments and a word of caution are in order concerning the interpretation of these results. While comparisons as to the relative merits of the techniques are inevitable, these comparisons may not be fully justified. The sample sizes are unequal, and not all methods were used during all the years. For example, NHC–64 was used for the first time during the 1964 season, but routine preparation of the RHS and T–60 forecasts was discontinued at Miami in 1964. This made it impossible to obtain a homogeneous sample for comparison of all techniques, although a small homogeneous sample was available for comparing five forecast methods based on the 1964 season. In addition, the official Weather Bureau forecasts are issued at 6-hour intervals, beginning at 0400 GMT. This means that the 24-hr. forecasts issued at 0400 GMT and 1600 GMT are based on surface and upper-air data observed 4 hours earlier, while those issued at 1000 GMT and 2200 GMT are based on upper-air data observed 10 hours earlier. On the other hand, the forecast period for the objective systems is measured from observation time, even though the forecast may not be available until several hours later. This places the official forecasts at a disadvantage whenever comparisons between methods are made, but this disadvantage is partly offset by the availability of later data, particularly aircraft reconnaissance reports.

TABLE 2.—*Statistical data for 24-hr. forecast period, 1959 through 1964.*

Name of technique	NHC–64*			NWP			T–59			M–M			T–60#			RHS			PERS		
Area	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
Number of forecasts	7	46	15	75	126	52	114	254	127	8	103	38	102	234	51	11	111	29	280	414	199
Mean (n. mi.)	98	88	156	190	133	159	147	116	164	155	96	140	150	112	168	251	130	145	235	134	206
Standard deviation (n. mi.)	44	51	86	155	85	98	106	66	92	69	57	82	126	67	97	120	84	100	192	86	121
Median (n. mi.)	102	83	142	153	114	139	126	106	143	177	87	117	114	97	169	266	120	108	175	115	179
Lower quartile (n. mi.)	48	47	95	96	80	80	78	67	91	91	54	75	75	64	77	147	73	76	101	73	116
Upper quartile (n. mi.)	105	116	173	244	179	213	177	158	230	215	125	167	180	151	229	297	164	165	317	169	277
Range (n. mi.)	48–189	13–245	34–370	13–949	13–614	26–496	13–747	0–430	6–401	26–228	6–271	25–343	21–793	8–365	25–377	79–545	13–544	12–401	16–1270	8–578	10–611

*1964, only.
#1959–1963.

TABLE 3.—*Statistical data for 36-hr. Forecast Period, 1959 through 1964.*

Name of technique	NHC–64*			NWP			T–60#			PERS		
Area	A	B	C	A	B	C	A	B	C	A	B	C
Number of forecasts	7	46	14	65	121	37	88	221	44	117	269	70
Mean (n. mi.)	186	145	252	281	210	272	212	195	231	309	209	317
Standard deviation (n. mi.)	83	74	105	259	158	169	165	124	125	253	134	194
Median (n. mi.)	176	151	231	238	183	237	178	181	245	234	174	268
Lower quartile (n. mi.)	123	75	174	124	114	142	115	98	107	134	112	173
Upper quartile (n. mi.)	209	195	290	344	256	350	251	253	327	408	269	424
Range (n. mi.)	45–331	32–306	113–514	12–1445	30–1332	24–830	19–1177	0–755	21–496	6–1329	19–730	36–941

*1964 only.
#1959–1963.

TABLE 4.—*Statistical data for 48-hr. forecast period*

Name of technique	NHC-64*			NWP†			PERS#		
Area	A	B	C	A	B	C	A	B	C
Number of forecasts	7	45	13	45	97	36	79	131	56
Mean (n. mi.)	264	211	420	314	273	407	313	279	473
Standard deviation (n. mi.)	107	106	154	203	161	212	203	138	284
Median (n. mi.)	288	197	407	283	247	405	254	254	434
Lower quartile (n. mi.)	151	119	310	155	160	234	156	174	217
Upper quartile (n. mi.)	302	288	435	440	345	529	429	351	686
Range (n. mi.)	68–430	35–408	212–747	51–786	43–902	75–1093	33–1100	37–633	78–1212

*1964 only.
†1959–64.
#1962–64.

TABLE 5.—*Statistical data for 72-hr. forecast period, 1964*

Name of technique	NWP			PERS		
Area	A	B	C	A	B	C
Number of forecasts	10	24	4	49	66	12
Mean (n. mi.)	467	423	525	436	449	752
Standard deviation (n. mi.)	218	244	306	293	232	375
Median (n. mi.)	528	346	556	377	430	812
Lower quartile (n. mi.)	248	235	83	181	271	353
Upper quartile (n. mi.)	651	608	684	608	557	1038
Range (n. mi.)	54–698	71–925	83–906	0–1188	48–1086	157–1285

TABLE 6.—*Official Weather Bureau forecasts, 1959 through 1964*

Forecast period	Area	Number of forecasts	Mean (n. mi.)	Standard deviation (n. mi.)	Median (n. mi.)	Lower quartile (n. mi.)	Upper quartile (n. mi.)	Range (n. mi.)
24 hr	A	289	183	138	138	85	237	5–812
24 hr	B	406	113	71	96	60	152	6–395
24 hr	C	194	167	94	150	93	216	7–502
48 hr.*	A	104	271	164	243	140	346	8–787
48 hr	B	153	237	132	212	142	304	21–708
48 hr	C	57	381	218	344	189	519	33–880
72 hr.†	A	53	390	252	332	212	472	102–1236
72 hr	B	65	362	217	335	210	470	55–998
72 hr	C	10	403	164	410	254	448	130–670

*Began in 1961.
†Started in 1964.

TABLE 7.—*Statistical data for 24-hr. forecast period. Homogeneous for 1964*

Name of technique	Number of forecasts	Mean (n. mi)	Standard deviation (n. mi.)	Mean (n. mi.)	Lower quartile (n. mi.)	Upper quartile (n. mi.)	Range (n. mi.)
PERS	23	120	85	105	48	128	24–357
M-M	23	101	56	97	54	142	6–219
NWP	23	127	71	94	80	168	26–278
NHC-64	23	107	60	88	60	124	44–321
T-59	23	127	52	136	73	163	42–202

TABLE 8.—*Forecast system having the least mean error by areas and forecast periods*

Forecast period	24 hr.			36 hr.		
Area	A	B	C	A	B	C
Method	NHC-64	NHC-64	M-M	NHC-64	NHC-64	T-60
Number of forecasts	7	46	38	7	46	44
Mean (n. mi.)	98	88	140	186	145	231
Median (n. mi.)	102	83	117	176	151	245

Forecast period	48 hr.			72 hr.		
Area	A	B	C	A	B	C
Method	NHC-64	NHC-64	NWP	PERS	NWP	NWP
Number of forecasts	7	45	36	49	24	4
Mean (n. mi.)	264	211	407	436	423	525
Median (n. mi.)	288	197	405	377	346	556

OFFICIAL WB FORECASTS

Forecast period	24 hr.			48 hr.			72 hr.		
Forecast period when referred to synoptic map time	28 hr.			52 hr.			76 hr.		
Area	A	B	C	A	B	C	A	B	C
Number of forecasts	289	406	194	104	153	57	53	65	10
Mean (n. mi.)	183	113	167	271	237	381	390	362	403
Median (n. mi.)	138	96	150	243	212	344	332	335	410

2. DISCUSSION OF RESULTS BY GEOGRAPHICAL AREAS

The results of the statistical analyses are shown in tables 2–8, with the verification numbers stratified by geographical areas (fig. 1). Table 2 presents the results of the error-record of the seven methods for a 24-hr. forecast period. One of the most striking facts brought out by this table is that for *all* systems the errors were for the most part substantially less in area B than in either areas A or C. The standard deviations were generally smaller in area B than in the other areas. This indicates that the errors in the distribution were more closely clustered about the mean. Observing the upper quartiles, one notes that there is quite a large reduction in the magnitudes of these errors in area B when compared with the other areas. This shows that there were considerably fewer large-error types of forecasts made in area B than in A or C. This same fact is noticed when most of the other statistical items in table 2 are examined.

The data in table 2 seem to indicate that the NHC–64 and M–M techniques produced 24-hr. forecasts with the smallest vector errors. For example, the mean and median vector errors for NHC–64 and M–M in area B are 88 and 83 n. mi. and 96 and 87 n. mi., respectively. These are substantially less than those of the other systems.

The 36-hr. forecast errors for the four systems which produce 36-hr. forecasts are shown in table 3. These data indicate that NHC–64 produced forecasts with the smallest vector errors in area B, where the mean was 145 n. mi. The T–60 was the next best with a mean error of 195 n. mi. In area C the mean error for T–60 was lower, but the median for NHC–64 was lower. However, the results of the two systems are not strictly comparable since they are for different years.

For a 48-hr. forecast three systems could be compared (table 4). The NHC–64 errors are less in A and B than either the NWP or PERS errors, while in C the NWP has a better record.

For a 72-hr. forecast (table 5), the NWP is better than persistence in areas B and C, while persistence is better in A. This latter fact is supported by Colón's [1] finding that a persistence forecast in the area east of the Lesser Antilles has a high probability of success.

Table 6 presents the statistical items for the official Weather Bureau forecasts. These data are listed separately because of the differences in the times these forecasts are issued. While no direct comparisons between the WB forecasts and the various objective systems should be made, it is of interest to note that the two best objective systems (NHC–64 and M–M) have demonstrated forecast skill comparable to that of the official forecasts. Table 6 also shows that forecasts for area B are better than those in areas A or C.

Table 7 shows the result of a homogeneous comparison of forecasts obtained by five of the objective forecast systems during the hurricane season of 1964. Homogeneous comparison as used here indicates that a forecast was "simultaneously" obtained by each of the five methods employed for the same synoptic times for the hurricanes of 1964. It was only possible to obtain 23 cases meeting this criterion. The data in this table show that for these 23 cases the NHC–64 and M–M techniques demonstrated somewhat superior performance over the NWP, T–59, and PERS systems. This is evident from a comparison of the means and medians of the vector errors.

The results of this section may be summarized briefly. Most of the forecast systems verified produce better forecasts in area B than elsewhere. This is the area with the greatest density of oceanic upper-air stations, and part of the increased accuracy is probably due to the better analyses which can be produced with more data. However, if one examines the persistence forecasts one finds that they are also significantly better in area B for 24-, 36-, and 48-hr. forecasts. This probably indicates that some of the increased forecast accuracy in area B is due to climatological factors. Finally, a 6-year record for the Miller-Moore technique and a 1-year record for the NHC–64 method indicate that these two systems are capable of producing somewhat better forecasts than the other objective methods tested. This is shown by the homogeneous sample (table 7) as well as table 8, which lists the forecast system having the least mean error by areas and forecast periods.

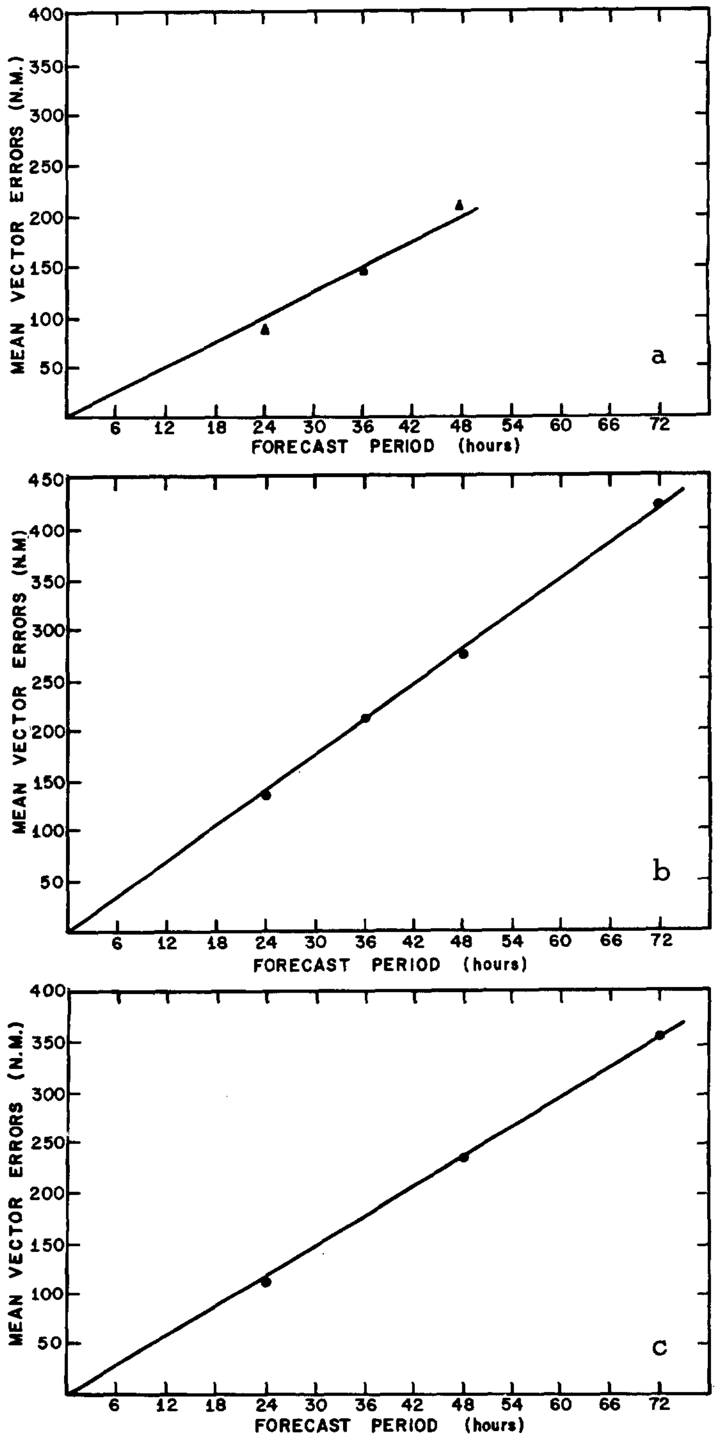

FIGURE 2.—Vector errors as a function of a change in the forecast period (a) NHC–64 method, (b) NWP method, (c) Weather Bureau official forecasts.

3. CHANGE IN VECTOR ERRORS WITH TIME

Figures 2a, 2b, and 2c contain plots of the mean vector error as a function of the forecast period for three of the hurricane track forecasting techniques. The three methods that are plotted are the NHC–64, NWP, and WB. The data are for area B, and include 1959–1964 for WB and NWP, but 1964 only for NHC–64.

The figures show that for all three methods there is very nearly a linear relationship between mean vector errors and length of forecast period.

4. PROBABILITY ELLIPSES

In this section the forecast vector errors will be displayed by use of probability ellipses. Figures 3 through 21 show probability ellipses for a number of techniques for 24-hr. forecast periods.

In the figures the origin (cross) represents the forecast position of the storm and the small dots represent the observed position of the storm at verification time relative to the forecast position. Thus, the error distribution represented by these figures may be interpreted in the following way:

Quadrant [1] 1—forecast position too far south and east.
Quadrant 2—forecast position too far south and west.
Quadrant 3—forecast position too far north and west.
Quadrant 4—forecast position too far north and east.

The numbers appearing on the probability ellipses represent the probabilities that that percentage of the distribution of the vector errors will be contained within these ellipses.

Before proceeding with a summary discussion of the figures, some comments regarding probability ellipses will be offered. Following Veigas, Miller, and Howe [9], the vector errors were resolved into latitude and longitude components and were plotted on suitable figures. Then, if it is assumed that these components, when considered jointly, may be displayed as a bivariate normal distribution (see Appendix), the resulting equations (1)–(3) from the probability density function are as follows:

$$a^2=\frac{2(1-\rho^2)\ln S}{\frac{1}{\sigma^2_{long}}-\frac{\rho}{\sigma_{long}\sigma_{lat}}\tan\phi} \tag{1}$$

$$b^2=\frac{2(1-\rho^2)\ln S}{\frac{1}{\sigma^2_{lat}}+\frac{\rho}{\sigma_{long}\sigma_{lat}}\tan\phi} \tag{2}$$

where a is the semi-major axis and b is the semi-minor axis of an ellipse. The other terms in equations (1) and (2) are defined as follows: σ_{long} and σ_{lat} are the population standard deviations of the longitude and latitude components, respectively; ρ is the linear correlation coefficient between the longitude and latitude components; and S is a measure of the probability and is given by $S=1/(1-p)$ where p is a certain probability, e.g., 10 percent, 28 percent, etc.

$$\phi=\frac{1}{2}\arc\tan\frac{2\rho\sigma_{long}\sigma_{lat}}{\sigma^2_{long}-\sigma^2_{lat}} \tag{3}$$

where ϕ is the angle which the semi-major axis of an ellipse makes with the horizontal coordinate axis of the figure depicting the error distribution. To solve the above equations, estimates of the various population parameters were obtained from the data sample represented in each figure.

Table 9 is a summary of some of the relationships represented by the probability ellipses depicted in figures 3 through 21. This table identifies the name of the forecast method being considered, the number of the figure which gives the detailed data for that forecasting technique, the area in which the data represented was acquired, a comparison of areas for each system depicted, and an observation of the degree to which the probability ellipses depart from ellipticity. Table 10 contains a section showing the percentage of forecasts which should be contained within the areas of the probability ellipses if the assumption is met that the joint distribution of the latitude and longitude components of the vector errors is a bivariate normal one, and the actual percentage of these forecasts which are contained within these probability ellipses. Also appearing in this table are the allowable differences in percentages permitted between a bivariate normal distribution and an actual distribution.

In conclusion, tables 9–10, illustrate several facts concerning figures 3 through 21.

First, for all techniques the forecasts made while storms were in area B yielded the smallest errors.

Second, most of the methods examined exhibited some sort of bias, even though it was quite small for some of the methods. For example, forecasts made while storms were in area A with the WB, T–59, NWP, and PERS methods produced forecasts which, on the average, tended to be too far south and too fast with reference to the observed positions of the storm centers.

Forecasts made when storms were in area B with the WB and RHS systems produced forecasts which, on the average, tended to be too far north and too fast, while those using the M–M and T–60 techniques produced "on the average" forecasts which tended to be too far north and too slow. Forecasts made in area B by the T–59 and NWP methods exhibited primarily a bias toward being too fast.

TABLE 9.—*Summary of some of the relationships represented by the probability ellipses depicted in figures 3 through 21. 24-hr. forecast periods represented.*

			Average bias of the forecasts					Relative areas covered by ellipses for a particular technique			Probability contours approximately	
Name of technique	Fig. No.	Area	Too far N	Too far S	Too fast	Too slow	None	Smallest	In between	Largest	Circular	Elliptical
WB	3	A		X	X					X	X	
WB	4	B	X		X			X				X
WB	5	C					X		X			X
M-M	6	B	X			X		X				X
M-M	7	C				X			X			X
R-H-S	8	B	X		X			X			X	
R-H-S	9	C	X			X				X		X
T-59	10	A		X	X				X			X
T-59	11	B			X			X				X
T-59	12	C		X		X				X		X
NWP	13	A		X	X					X		X
NWP	14	B			X			X				X
NWP	15	C		X	X				X			X
T-60	16	A	X			X				X		X
T-60	17	B	X			X		X				X
T-60	18	C		X		X			X			X
PERS	19	A		X	X					X	X	
PERS	20	B		X	X			X				X
PERS	21	C		X	X				X			X

[1] Quadrants are numbered by considering Quadrant 1 to be the upper lefthand quadrant, Quadrant 2 the upper righthand quadrant, Quadrant 3 the lower righthand quadrant, and Quadrant 4 the lower lefthand quadrant. These quadrants are with reference to the coordinate axes *not* with reference to the ellipses.

TABLE 10.—*Percentage of forecasts contained within probability ellipses. Bivariate normal percentages (5 percent level)*

Name of technique	Figure No.	Area	20% probability ellipse		40% probability ellipse		60% probability ellipse		80%probability ellipse	
			Actual percentage	Maximum allowable absolute difference	Actual percentage	Maximum allowable absolute difference	Actual percentage	Maximum allowable absolute difference	Actual percentage	Maximum allowable absolute difference
			Percent	*Percent*	*Percent*	*Percent*	*Percent*	*Percent*	*Percent*	*Percent*
WB	3	A	34	8.0	55	8.0	69	8.0	83	8.0
WB	4	B	25	6.7	46	6.7	62	6.7	77	6.7
WB	5	C	21	9.8	43	9.8	59	9.8	77	9.8
M-M	6	B	22	13.4	41	13.4	60	13.4	73	13.4
M-M	7	C	18	22.1	45	22.1	61	22.1	76	22.1
R-H-S	8	B	27	13.1	47	13.1	71	13.1	88	13.1
R-H-S	9	C	28	24.6	59	24.6	72	24.6	79	24.6
T-59	10	A	27	12.8	51	12.8	70	12.8	85	12.8
T-59	11	B	21	5.4	40	5.4	56	5.4	76	5.4
T-59	12	C	22	12.1	46	12.1	61	12.1	72	12.1
NWP	13	A	32	15.7	55	15.7	67	15.7	79	15.7
NWP	14	B	22	12.1	48	12.1	67	12.1	79	12.1
NWP	15	C	21	18.9	38	18.9	56	18.9	73	18.9
T-60	16	A	39	13.5	59	13.5	76	13.5	84	13.5
T-60	17	B	22	8.9	44	8.9	63	8.9	78	8.9
T-60	18	C	29	19.0	41	19.0	55	19.0	69	19.0
PERS	19	A	36	9.4	59	9.4	71	9.4	83	9.4
PERS	20	B	25	7.6	46	7.6	67	7.6	79	7.6
PERS	21	C	21	11.1	47	11.1	60	11.1	75	11.1

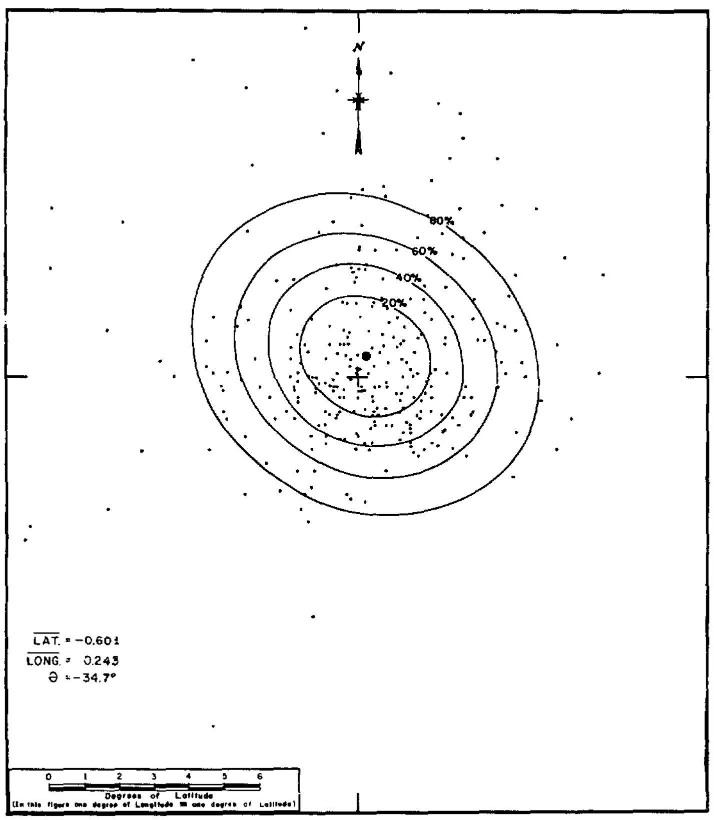

FIGURE 3.—Probability ellipses for Weather Bureau 24-hr. forecasts, area A (1959–64). Numbers in lower left are as follows: $\overline{\text{LAT.}}$—average of latitude component (in °lat.); $\overline{\text{LONG.}}$—average of longitude component (in °lat.); θ—angle formed by semi-major axis with the horizontal axis of the figure. Bold dot indicates center of ellipses; cross indicates origin of coordinate axes.

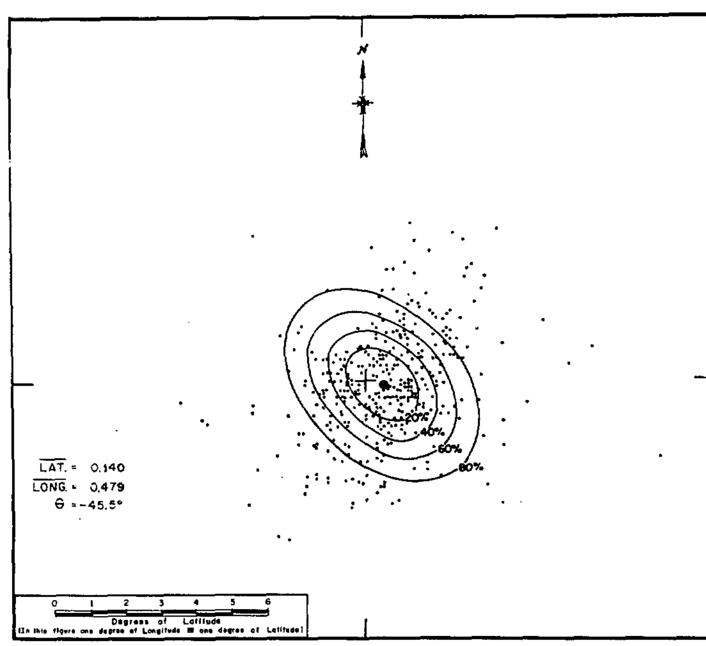

FIGURE 4.—Probability ellipses for Weather Bureau 24-hr. forecasts, area B (1959–64). For explanation see figure 3.

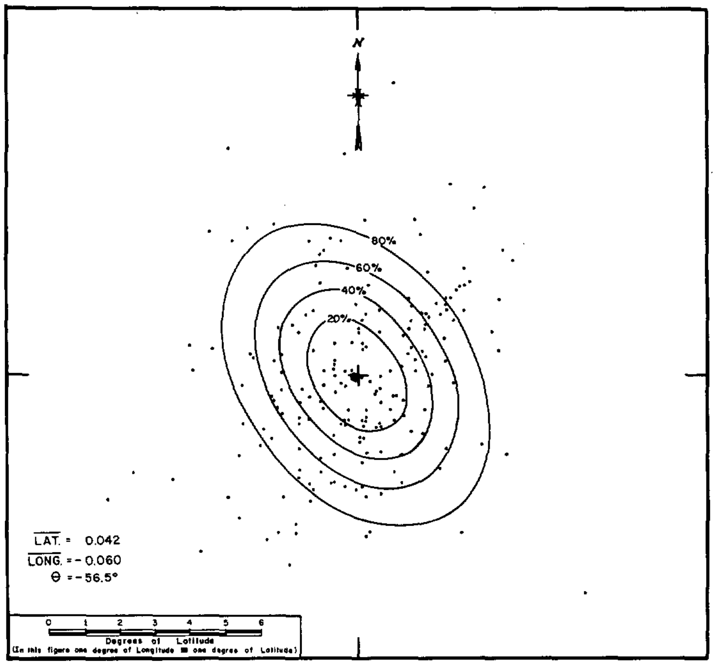

FIGURE 5.—Probability ellipses for Weather Bureau 24-hr. forecasts, area C (1959–64). For explanation see figure 3.

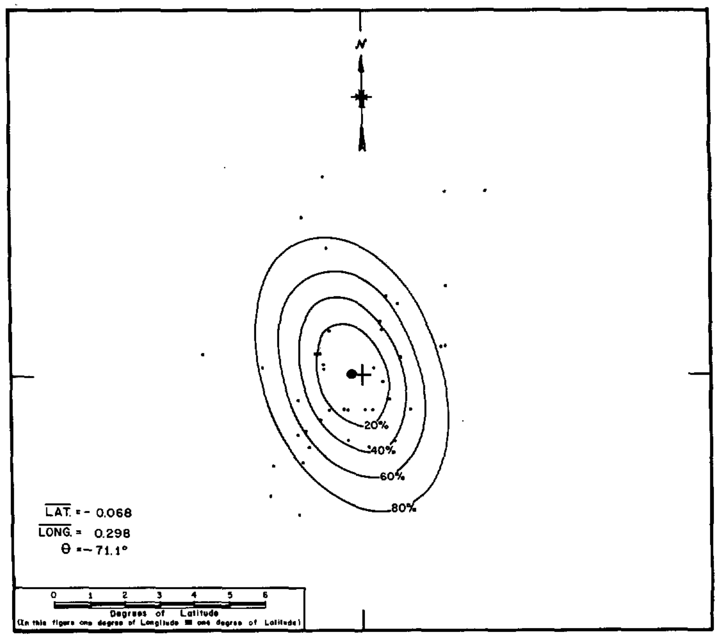

FIGURE 7.—Probability ellipses for modified Miller-Moore method, area C (1959–64). For explanation see figure 3.

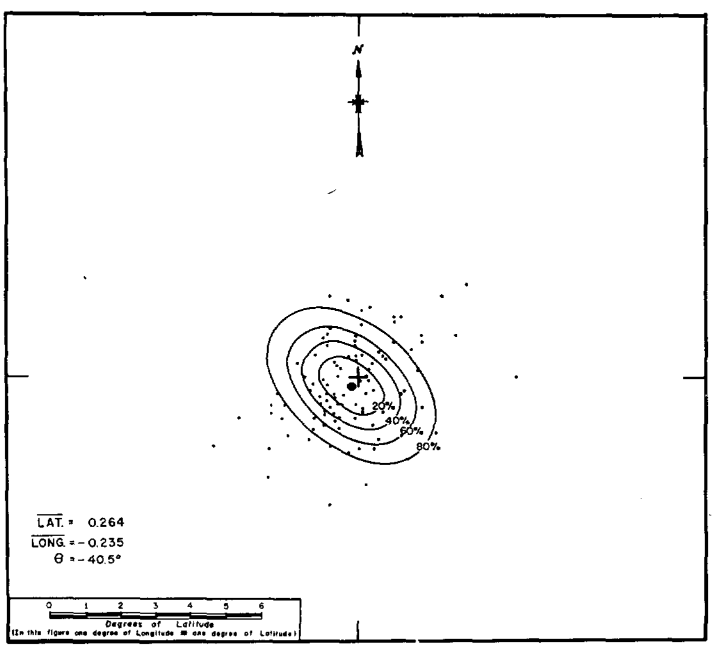

FIGURE 6.—Probability ellipses for modified Miller-Moore method, area B (1959–64). For explanation see figure 3.

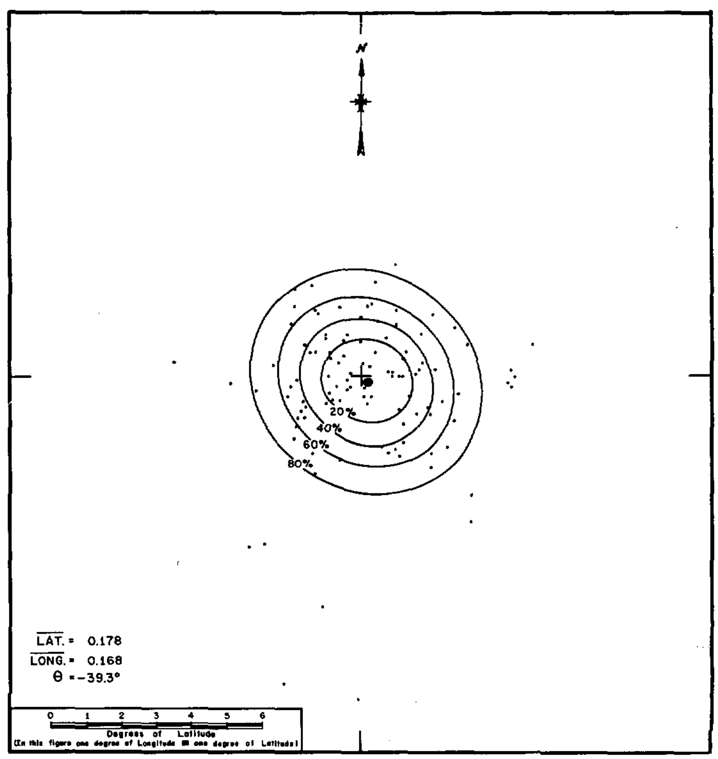

FIGURE 8.—Probability ellipses for Riehl-Haggard-Sanborn 500-mb. grid method, 24-hr. forecasts, area B (1959–63). For explanation see figure 3.

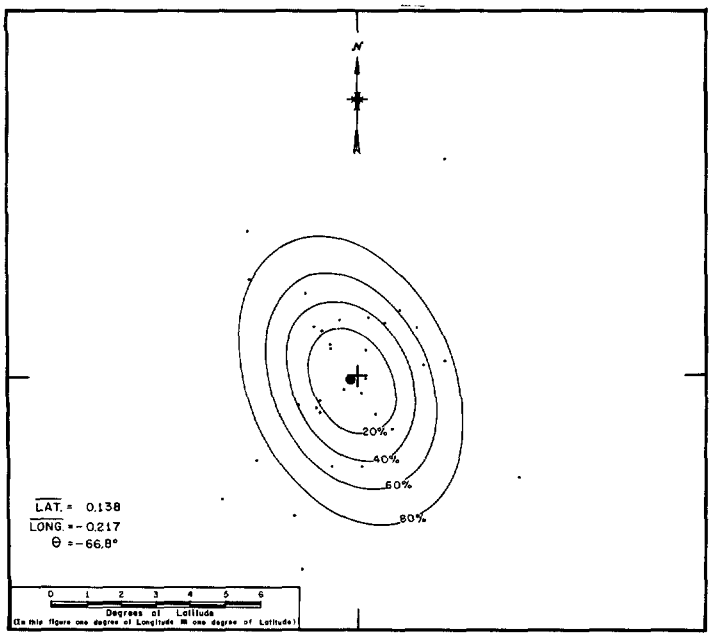

FIGURE 9.—Probability ellipses for Riehl-Haggard-Sanborn 500-mb. grid method, 24-hr. forecasts, area C (1959–63). For explanation see figure 3.

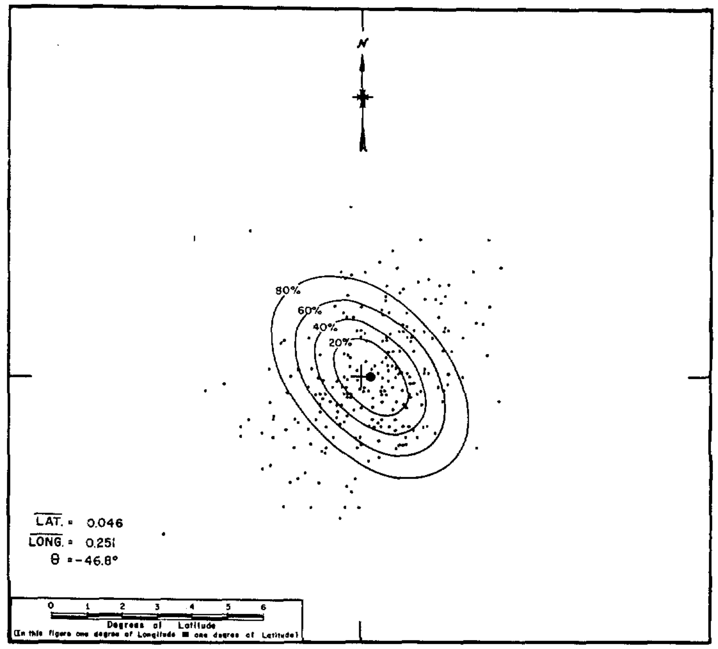

FIGURE 11.—Probability ellipses for Travelers 1959 (T–59) 24-hr. forecasts, area B (1959–64). For explanation see figure 3.

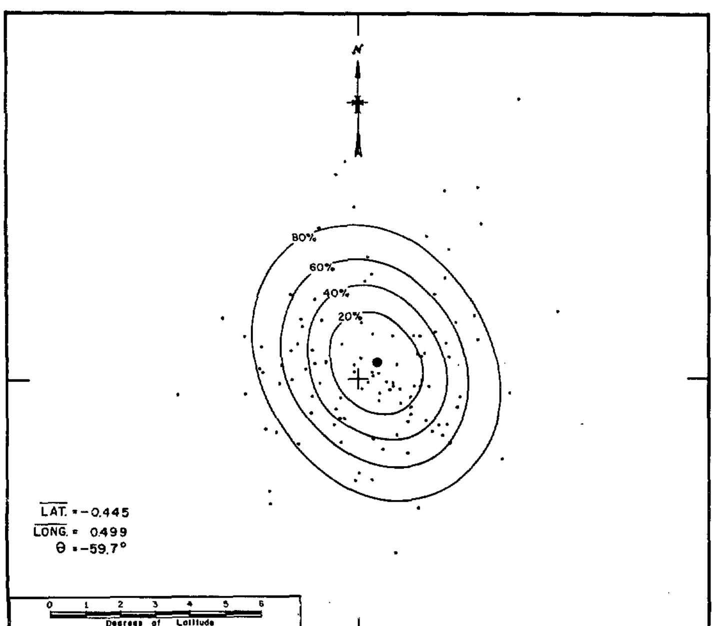

FIGURE 10.—Probability ellipses for Travelers 1959 (T–59) 24-hr. forecasts, area A (1959–64). For explanation see figure 3.

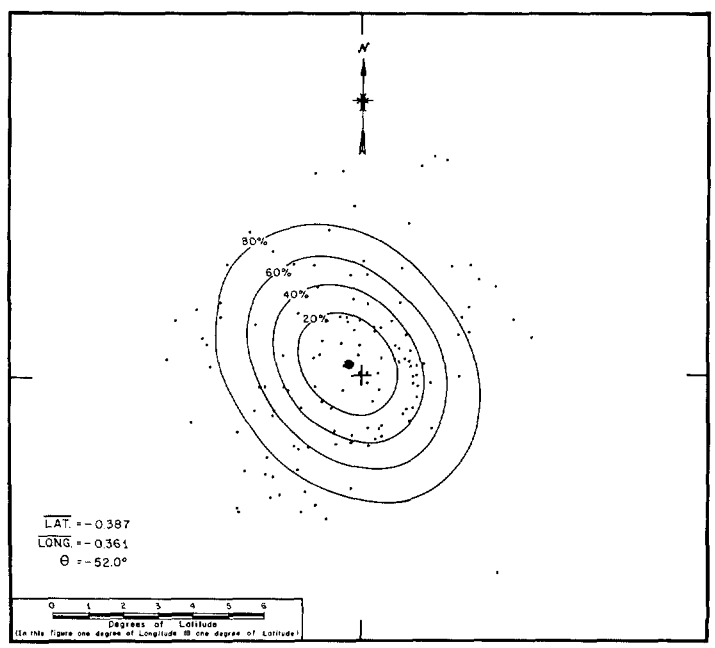

FIGURE 12.—Probability ellipses for Travelers 1959 (T–59) 24-hr. forecasts, area C (1959–64). For explanation see figure 3.

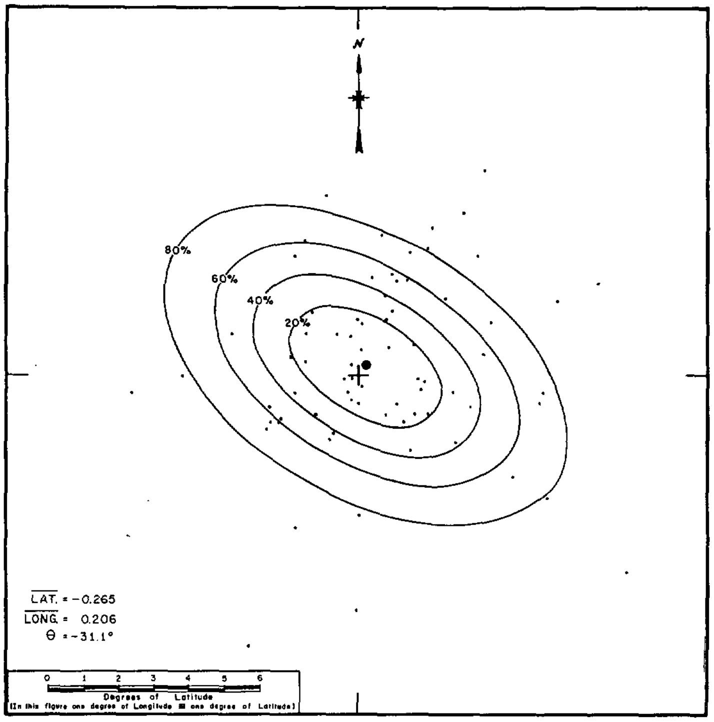

FIGURE 13.—Probability ellipses for numerical (NWP) 24-hr. forecasts, area A (1959–64). For explanation see figure 3.

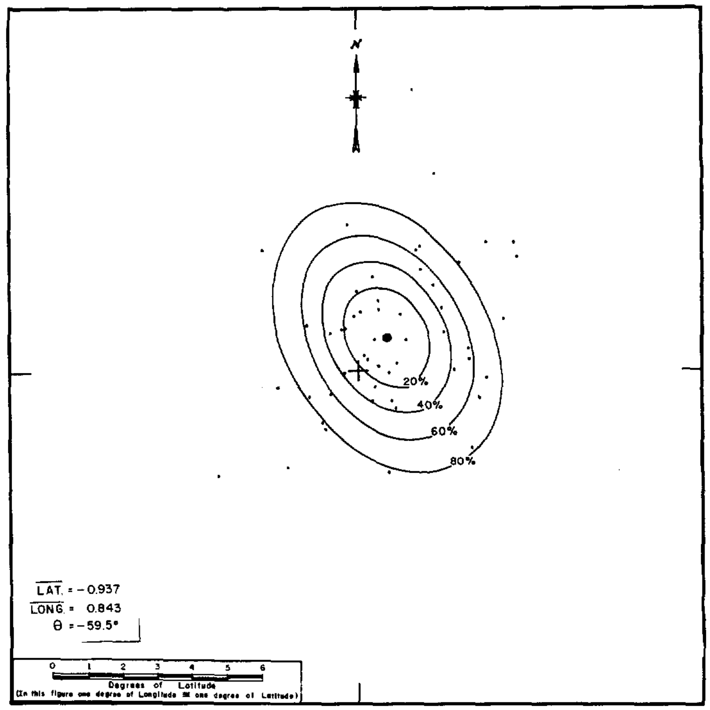

FIGURE 15.—Probability ellipses for numerical (NWP) 24-hr. forecasts, area C (1959–64). For explanation see figure 3.

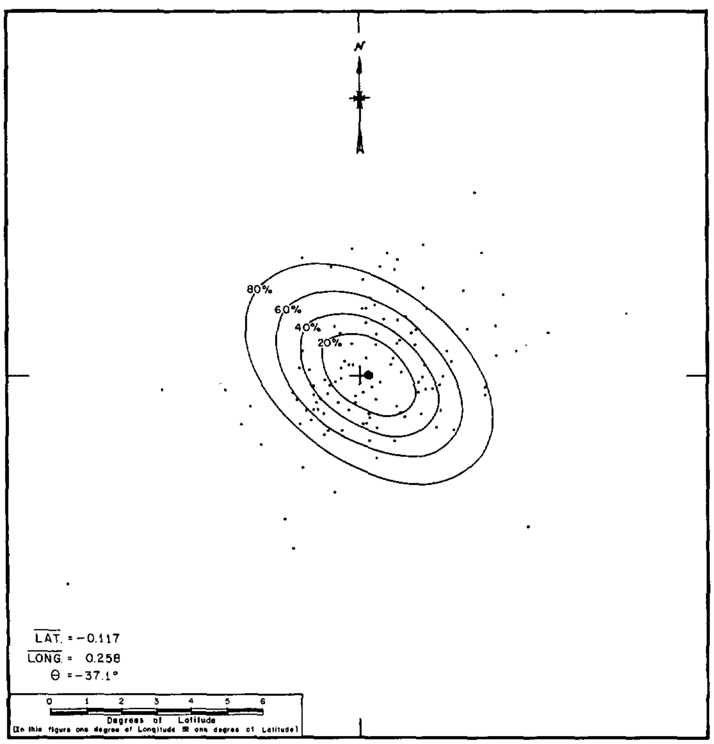

FIGURE 14.—Probability ellipses for numerical (NWP) 24-hr. forecasts, area B (1959–64). For explanation see figure 3.

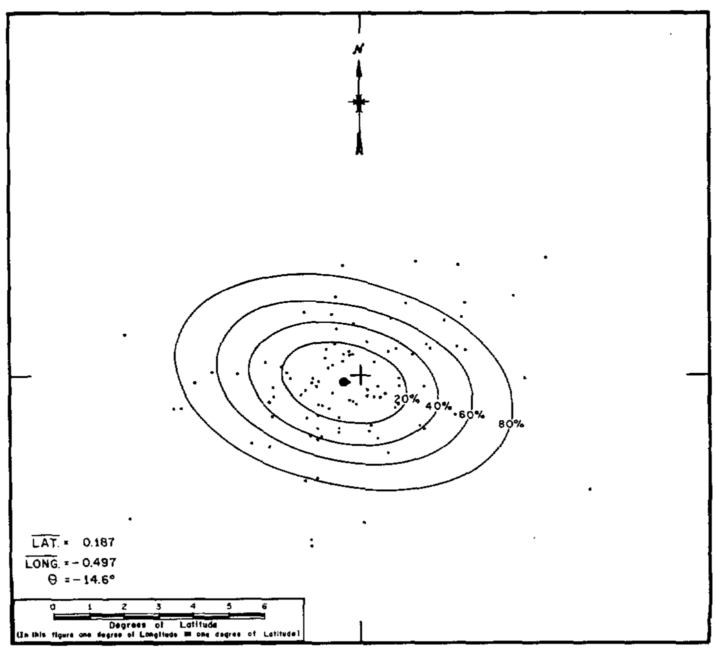

FIGURE 16.—Probability ellipses for Travelers 1960 method (T-60) 24-hr. forecasts, area A (1959–64). For explanation see figure 3.

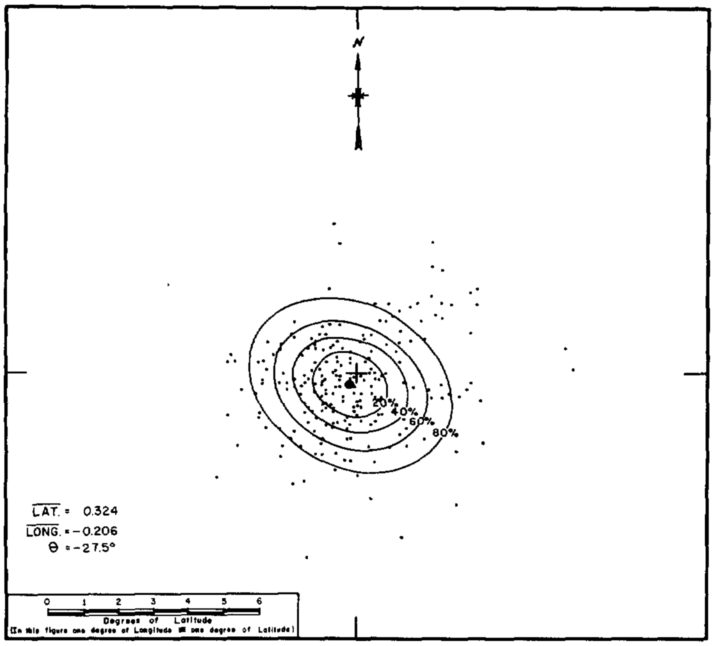

FIGURE 17.—Probability ellipses for Travelers 1960 method (T-60) 24-hr. forecasts, area B (1959–64). For explanation see figure 3.

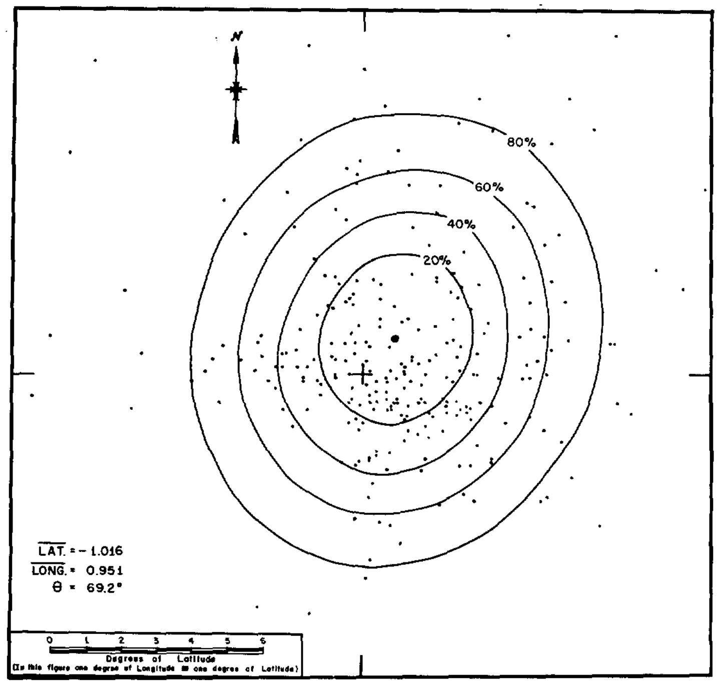

FIGURE 19.—Probability ellipses for persistence forecasts (24-hr.), area A (1959–64). For explanation see figure 3.

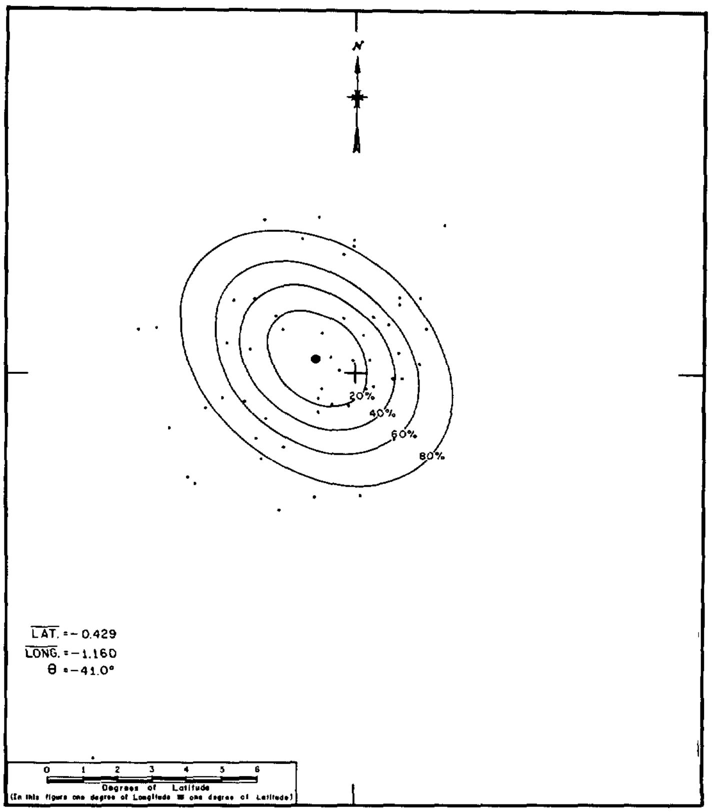

FIGURE 18.—Probability ellipses for Travelers 1960 method (T-60) 24-hr. forecasts, area C (1959–64). For explanation see figure 3.

Forecasts made while storms were in area C by the T-59 and T-60 systems yielded "on the average" forecasts that were too far south and too slow while those using the RHS system were too far north and too slow. Also in the area C, "on the average" forecasts produced by the NWP method were too far south and too fast, those by the M-M method were primarily too slow, those by the PERS method were too fast, and in those by the WB method little bias was shown.

Third, all but three of the techniques (WB-Area A, RHS-Area B, PERS-Area A) for which probability ellipses were determined had shapes that were elliptical. The three which did not, had shapes approximating that of a circle. Contour ellipses having a circular configuration indicate a random distribution of the quantities involved. The more eccentric the ellipses becomc, the closer the relationship between latitude and longitude errors, and one can see that as the distribution of the vector errors comprising the sample distribution more closely approximates that of a straight line, then the linear correlation coefficient between latitude and longitude errors approaches one.

Fourth, it is apparent, from the section in table 10 containing the percentage of forecasts within probability ellipses, that the probability ellipses for all the figures except figures 3, 16, and 19 represent their respective distributions as valid bivariate normal distributions within

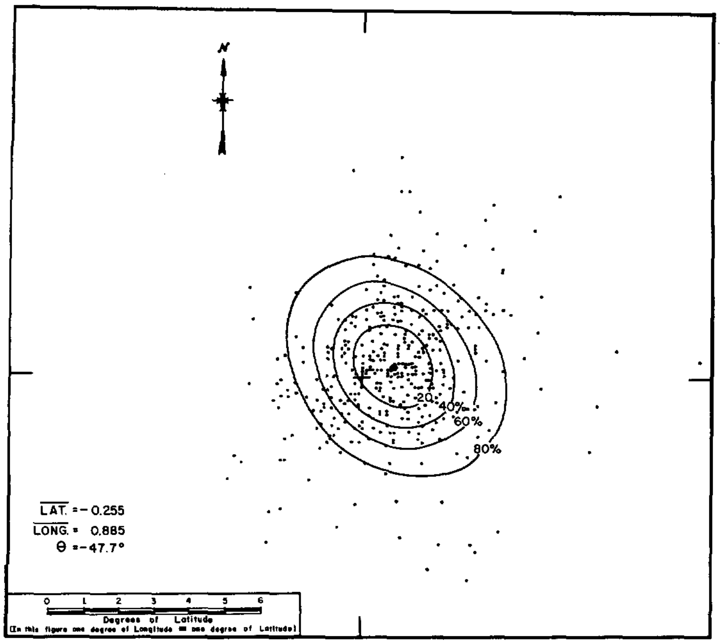

FIGURE 20.—Probability ellipses for persistence forecasts (24-hr.), area B (1959–64). For explanation see figure 3.

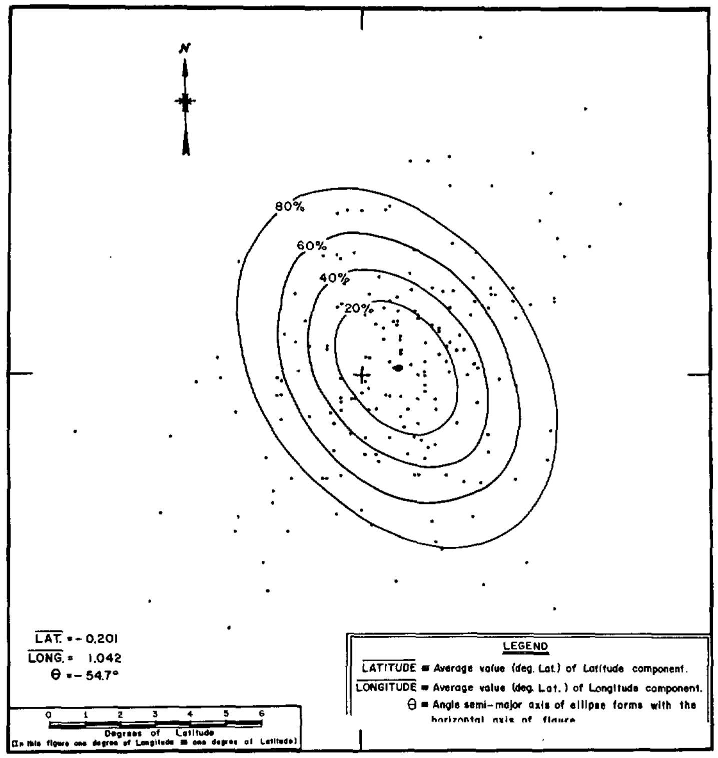

FIGURE 21.—Probability ellipses for persistence forecasts (24-hr.), area C (1959–64). For explanation see figure 3.

the limits of the Kolmogorov-Smirnov test. Strictly speaking, figures 3, 16, and 19 should not appear, as they do not represent close agreement with a bivariate normal distribution. However for the sake of completeness they are presented.

ACKNOWLEDGMENTS

The author thanks Dr. Banner I. Miller, Dr. R. Cecil Gentry, and Mr. Harry F. Hawkins for their helpful criticisms and comments in the preparation of this paper.

Appreciation is also extended to Mr. Keith W. Veigas of the Travelers Research Center, Inc., Hartford, Conn., for his assistance in certain statistical aspects of the paper.

The author also thanks Messrs. Robert C. Carrodus and Charles H. True for drafting and photographing the figures in this paper.

Final thanks go to Miss Debbie Nejman for typing the manuscript and tables.

APPENDIX

To test a sample of data to see whether it has a bivariate normal distribution, one can utilize the Kolmogorov-Smirnov one-sample test [9] and [6]. This test " . . . is concerned with the degree of agreement between the distribution of a set of sample values [observed vector errors] and some specified theoretical distribution. . . ." It determines whether the scores in the sample can reasonably be thought to have come from a population having the theoretical distribution. Briefly, the test involves specifying the cumulative frequency distribution which would occur under the theoretical distribution and comparing that with the observed cumulative frequency distribution. The point at which these two distributions, theoretical and observed, show the greatest divergence is determined. Reference to the sampling distribution indicates whether such a large divergence is likely on the basis of chance. That is, the sampling distribution indicates whether a divergence of the observed magnitude would probably occur if the observations were really a random sample from the theoretical distribution.

The Kolmogorov-Smirnov test focuses on the largest of the deviations. The largest value of $F_0(X)-S_N(X)$ is called the maximum deviation, D:

$$D=maximum\,|F_0(X)-S_N(X)|$$

where F_0 is a specified cumulative frequency distribution—the theoretical distribution—and $S_N(X)$ is the observed cumulative frequency distribution of a random sample of N observations.

Values of D at the 5 percent level of significance were determined from the components of the vector errors for each technique and for each area for which forecasts were made. These values of D were then compared with certain critical values which are contained in table E on page 251 of Siegel's [6] book. The results of these tests are presented in table 9 in a different and, perhaps, more interpretative form.

REFERENCES

1. J. A. Colón, "A Study of Hurricane Tracks for Forecasting Purposes," *Monthly Weather Review*, vol. 81, No. 3, Mar. 1953, pp. 53–66.
2. G. E. Dunn, "Hurricane Prediction," Paper presented at the American Meteorological Society Annual Meeting, New York, Jan. 22, 1962.
3. R. C. Gentry, "A Re-evaluation of the Problem of Predicting Hurricane Movement," *Proceedings* of the Technical Conference on Hurricanes, Miami, Florida, American Meteorological Society, 1958.
4. B. I. Miller and P. L. Moore, "A Comparison of Hurricane Steering Levels," *Bulletin of the American Meteorological Society*, vol. 41, No. 2, Feb. 1960, pp. 59–63.
5. H. Riehl, W. H. Haggard, and R. W. Sanborn, "On the Prediction of 24-hour Hurricane Motion," *Journal of Meteorology*, vol. 13, No. 5, Oct. 1956, pp. 415–420.
6. S. Siegel, *Nonparametric Statistics*, McGraw-Hill Book Co., Inc., New York, N.Y., 1956.
7. J. D. Tracy, "Verification of Forecasts of Hurricane Motion Using Various Techniques," pp. 373–382 in "Proceedings of the Second Technical Conference on Hurricanes," *National Hurricane Research Project Report* No. 50, U.S. Weather Bureau, 1962.
8. L. W. Vanderman, "An Improved NWP Model for Forecasting the Paths of Tropical Cyclones," *Monthly Weather Review*, vol. 90, No. 1, Jan. 1962, pp. 19–22.
9. K. W. Veigas, R. G. Miller, and G. M. Howe, "Probabilistic Prediction of Hurricane Movements by Synoptic Climatology," *Occasional Papers in Meteorology*, No. 2, Travelers Weather Research Center, Hartford, Conn., 1959, 54 pp.
10. K. W. Veigas, "Prediction of Twelve, Twenty-four and Thirty-six Hour Displacement of Hurricanes by Statistical Methods," *Final Report*, Contract No. Cwb–9807, The Travelers Research Center, Inc., Hartford, Conn., 1961, 36 pp.

[*Received March 3, 1966; revised April 22, 1966*]