Typhoon Forecast with the GFDL Hurricane Model: Forecast Skill and Comparison of Predictions using AVN and NOGAPS Global Analyses

By Chun-Chieh Wu

Department of Atmospheric Sciences, National Taiwan University, Taipei, Taiwan

Morris A. Bender and Yoshio Kurihara[1]

Geophysical Fluid Dynamics Laboratory, NOAA, Princeton, NJ, U.S.A.

(Manuscript received 26 October 1999, in revised form 17 August 2000)

Abstract

A hurricane model developed at GFDL, NOAA, was combined with each of AVN and NOGAPS global analyses to construct typhoon prediction systems GFDS and GFDN, respectively. The GFDS system performed 125 (178) forecast experiments for 16 (24) storms in the western North Pacific basin during 1995 (1996). It exhibited considerable skill in the forecast of tropical cyclone tracks. The average forecast position errors at 12, 24, 36, 48 and 72 h in 1995 (1996) were 95 (108), 146 (178), 193 (227), 249 (280), and 465 (480) km. The improvement with GFDS in the typhoon position forecast over CLIPER was roughly 30 %. The reduction of position errors in both average and standard deviations indicates superior forecast accuracy and consistency of GFDS, although there existed systematic northward bias in the forecast motion at low latitudes. On the other hand, intensity forecast was not satisfactory, showing a tendency to overpredict weak storms and underpredict strong storms, similar to the tendency in the Atlantic.

Two sets of forecasts performed in the 1996 season, the one by GFDS and the other by GFDN, were compared with each other. Forecast skills of the storm position with the two systems were comparable. However, the two forecast positions tended to be systematically biased toward different directions. As a result, when the two forecasts were averaged, the mean error was 10 % smaller than that of each forecast. Also, overall improvement in track forecast was obtained in supplemental experiments in which individual forecasts were corrected for systematic biases. Though systematic bias is not steady, there may be ways to utilize it for improvement of tropical cyclone forecasts.

1. Introduction

Improvement in tropical cyclone forecasting represents one of the greatest challenges in numerical weather prediction. A hurricane forecast system was developed at NOAA's (National Oceanic and Atmospheric Administration) Geophysical Fluid Dynamics Laboratory (GFDL) to provide forecast guidance on storms in the Atlantic basin. The system, an official operational hurricane prediction tool of the U.S. National Weather Service (NWS) since 1995, consists of the GFDL's primitive equation hurricane model, a model initialization scheme, the global analysis called AVN (Derber et al. 1998) of the National Center for Environmental Prediction (NCEP), and the storm information from the National Hurricane Center. Its remarkable performance has been demonstrated in Atlantic hurricane prediction. For example, the mean track forecast errors for 255 cases in 1995 in the Atlantic at 36, 48, and 72 hours were reduced by 14, 19, and 25 %, respectively, relative to the next best NWS dynamical prediction model (Kurihara et al. 1998).

Ideally, an improved hurricane prediction system should demonstrate improvement in forecast skill in all cases in all ocean basins. The GFDL system was experimentally run in 1995 for 125 cases in the

Corresponding author: Chun-Chieh Wu, Department of Atmospheric Sciences, National Taiwan University, 61, Ln. 144, Sec. 4, Keelung Rd., Taipei 10772, Taiwan. E-mail: cwu@typhoon.as.ntu.edu.tw

1 Current affiliation: Frontier Research System for Global Change, Sumitomo-Hamamatsucho Bldg. 4F, Minato-ku, Tokyo, 105-0013, Japan.

western North Pacific basin (hereafter, referred to as GFDS), where typhoons evolve in an environment of higher sea surface temperature, warmer and more humid air than the Atlantic and monsoon circulation.

In 1996, the U.S. Navy implemented a new operational system of typhoon prediction in which the GFDL model was combined with the NOGAPS (Navy Operational Global Atmospheric Prediction System) global analysis (Goerss and Phoebus 1992; Goerss and Jeffries 1994). This system (hereafter, referred to as GFDN) had been successfully tested in the preceding season for a limited number of western North Pacific cases with the result indicating that mean track forecast accuracy was similar to that of GFDS through 48 h forecast period. The GFDS was also run in 1996 in a parallel mode with GFDN for 178 cases, which is 75 % of the total number of GFDN forecasts. The parallel forecast experiments for such a large number of cases could not be repeated in subsequent years due to limited computer availability at NCEP.

Rennick (1999) analyzed the GFDN performance in 1996. In this study, typhoon forecasts using GFDS in 1995 and 1996 are analyzed to make more extensive evaluation of the GFDL model performance. Also, forecasts with GFDS in 1996 are compared against those with GFDN. Various factors can contribute to the spread of forecasts, such as difference in the environmental conditions at initial time, difference in specification of initial storm vortex, difference in forecast model, and combination of these. Differences in the initial environmental condition and time-dependent lateral boundary values caused a forecast difference between GFDS and GFDN. In this respect, our comparison study is different from others, e.g., Rennick (1999), Zhang and Krishnamurti (1999) and Goerss (2000).

In the present analysis, systematic biases in forecasts with GFDS and GFDN are emphasized. An ensemble forecast using these two systems is attempted as well. Furthermore, supplemental experiments are carried out to suggest applicability of information on systematic bias to operational forecasts.

In Section 2, the forecast experiment and the data used in the forecast evaluation are described. The results of error analysis for the 1995 and 1996 typhoon seasons are respectively presented in Sections 3 and 4. Comparison of forecasts between GFDN and GFDS are included in Section 4. Summary and remarks are made in Section 5.

2. Forecast experiments and verification data

2.1 Forecast experiment

The GFDL multiply-nested movable mesh model (Kurihara et al. 1995) was used in GFDN and GFDS forecasts. The integration domain of the GFDL model spans 75° latitude by 75° longitude, with a triply-nested grid system with resolutions of 1°, 1/3°, and 1/6° (Table 1 of Bender et al. 1993). The outermost domain extends from 15°S to 60°N in the meridional direction (for GFDN, the northern and southern boundaries may sometimes shift slightly southward), and varies in the zonal direction depending on the storm's condition at initial time.

The time integration for each GFDS (GFDN) forecast was carried out for 72 hours using the lateral boundary values which were updated hourly by linear interpolation of the forecasts from the AVN (NOGAPS) global spectral model (Kurihara et al. 1989). In 1995 each GFDS experiment run with initial times of 00 or 12 UTC, while in 1996 with initial times of 06 or 18 UTC. The same storm information was used for the storm vortices initialization (Kurihara et al. 1995) in GFDN and GFDS systems.

Numbers of GFDS forecast made in the western North Pacific basin in 1995 (1996) were 125 (178) for 16 (24) storms that formed during the months of July through early November (late December). The lists of GFDS forecasts made in 1995 and 1996 are summarized separately in Tables 1 and 2.

2.2 Forecast verification

Data for verifying the track position and intensity forecast were obtained from the real-time tropical cyclone message from the Joint Typhoon Warning Center (JTWC). (The same messages were used for specifying the initial vortex in the model.) Tests were conducted to verify both GFDS and GFDN track forecasts against either the JTWC position fixed at real time or the JTWC post-analysis best-track position. Very little differences were found in the mean position error. Verification of GFDS forecasts was not made when JTWC ceased its forecasts because storms had made landfall, or had become extra-tropical disturbances.

To evaluate the performance of GFDS, various error statistics were computed for all cases in Tables 1 and 2. They include the mean forecast position error and its standard deviation for showing the forecast accuracy and consistency, distribution of systematic forecast bias, examination of error scatter, and evaluation of the intensity forecast. Also, for 72 homogeneous cases in 1995, GFDS forecasts were compared against the AVN and NOGAPS global model forecasts, and the official JTWC forecasts. These model forecasts were also compared against the CLIPER (a simple model based on climatology and persistence) as a means of evaluating forecast skill.

3. Results from forecast experiments in 1995

3.1 Average error and comparison with other models

The average forecast position errors (standard deviations) of GFDS in 1995 at 12, 24, 36, 48 and 72 h forecast times were 95 (63), 146 (87), 193 (131), 249

Table 1. List of all GFDS forecasts in the western North Pacific basin in 1995.

STORM	NAME	NUMBER OF CASES	DATES OF FORECASTS
FAYE	(WP05)	11	JULY 18 – JULY 23
GRAY	(WP07)	3	JULY 29 – JULY 30
HELEN	(WP08)	6	AUGUST 8 – AUGUST 11
IRVING	(WP09)	3	AUGUST 17 – AUGUST 18
JANIS	(WP10)	9	AUGUST 22 – AUGUST 26
KENT	(WP12)	8	AUGUST 27 – AUGUST 30
LOIS	(WP13)	4	AUGUST 27 – AUGUST 29
OSCAR	(WP17)	9	SEPT. 12 – SEPT. 16
POLLY	(WP18)	5	SEPT. 16 – SEPT. 20
RYAN	(WP19)	12	SEPT. 17 – SEPT. 23
SIBLY	(WP20)	7	SEPT. 28 – OCTOBER 2
VAL	(WP25)	1	OCTOBER 10
WARD	(WP26)	9	OCT. 18 – OCT. 22
YVETTE	(WP27)	5	OCT. 23 – OCT. 25
ZACK	(WP28)	13	OCT. 25 – NOV. 1
ANGELA	(WP29)	20	OCT. 26 – NOV. 4
TOTAL		125	JULY 18 – NOVEMBER 4

Table 2. List of all GFDS forecasts in the western North Pacific basin in 1996.

STORM	NAME	NUMBER OF CASES	DATES OF FORECASTS
DAN	(WP06)	3	JULY 06 – JULY 08
EVE	(WP07)	11	JULY 13 – JULY 19
GLORIA	(WP09)	5	JULY 22 – JULY 25
HERB	(WP10)	9	JULY 24 – JULY 31
IAN	(WP11)	2	JULY 29 – JULY 29
JOY	(WP12)	3	AUGUST 02 – AUGUST 03
KIRK	(WP13)	15	AUGUST 05 – AUGUST 15
NIKI	(WP18)	5	AUGUST 18 – AUGUST 20
ORSON	(WP19)	17	AUGUST 22 – SEP. 03
SALLY	(WP23)	6	SEP. 05 – SEP. 08
TOM	(WP25)	8	SEP. 11 – SEP. 17
VIOLET	(WP26)	11	SEP. 12 – SEP. 22
WILLIE	(WP27)	6	SEP. 18 – SEP. 22
YATES	(WP28)	7	SEP. 23 – OCT. 01
ZANE	(WP29)	10	SEP. 24 – OCT. 02
ABEL	(WP30)	2	OCT. 13 – OCT. 16
WP31		2	OCT. 13 – OCT. 15
BETH	(WP32)	9	OCT. 14 – OCT. 21
CARLO	(WP33)	7	OCT. 21 – OCT. 25
WP35		1	NOVEMBER 02
DALE	(WP36)	11	NOV. 04 – NOV. 12
ERNIE	(WP37)	13	NOV. 04 – NOV. 16
FERN	(WP42)	8	DEC. 22 – DEC. 26
GREG	(WP43)	7	DEC. 24 – DEC. 27
TOTAL		178	JULY 06 – DECEMBER 27

(158) and 465 (330) km, respectively, for 125, 123, 115, 103 and 77 cases (Table 3). Large values of the standard deviation mean that the forecast error can vary significantly from case to case. As indicated by Rennick (1999), cases with large errors were generally associated with either the erratic timing of recurvature, poor tracking performance over elevated terrain, or poor performance in an environment of strong vertical shear of the wind. A number of researches addressing these problems have been developed at GFDL (Kurihara et al. 1998).

Forecasts of storm position by CLIPER for the same cases as mentioned above showed average errors (standard deviations) of 126 (88), 223 (147), 326 (210), 417 (266), and 633 (400) km for 12, 24, 36, 48 and 72 h, respectively. The frequencies of

Table 3. Average track forecast errors (standard deviations) in km for GFDS forecasts in the western North Pacific basin in 1995.

Unit: km

	12 h	24 h	36 h	48 h	72 h
GFDS	95 (63)	146 (87)	193 (131)	249 (158)	465 (330)
Case number	125	123	115	103	77

Table 4. Average track forecast errors (standard deviations) in km for a homogeneous comparison among various model forecasts in the western North Pacific basin in 1995.

Unit: km

	12 h	24 h	36 h	48 h	72 h
CLIPER	116 (78)	209 (128)	312 (193)	400 (279)	619 (431)
JTWC	99 (65)	177 (89)	263 (127)	369 (196)	629 (317)
GFDS	94 (61)	150 (87)	197 (131)	275 (174)	527 (372)
NOGAPS	122 (77)	201 (118)	286 (169)	400 (260)	702 (447)
AVN	134 (81)	226 (130)	331 (225)	468 (360)	829 (527)
Case number	72	71	64	54	38

superior performance by GFDS over CLIPER were 67 %, 72 % and 68 % for 24, 48 and 72 h, respectively. The reduction in the average error and standard deviation, as well as the high frequencies of superior performance, indicates the superior forecast of storm position by GFDS compared with CLIPER. Customarily, the model's forecast skill is expressed by the difference of the average errors between the model and CLIPER, divided by the average error of CLIPER. The result of calculation indicates reduction of errors by GFDS relative to CLIPER was 25 %, 35 %, 41 %, 40 %, and 27 % at 12, 24, 36, 48, and 72 hours, respectively.

Forecasts of storm position from GFDS, AVN, NOGAPS, and the official JTWC forecast for 72 homogeneous cases are compared in Table 4. Overall, GFDS was the best performer in 1995 with the smallest average error and standard deviation at all forecast times. The differences in the average position error between GFDS and other forecasts increased roughly linearly as the forecast time increased. Compared with CLIPER, JTWC, NOGAPS and AVN, reduction of the average position errors by GFDS at 24 (48) h was 28 (31), 15 (25), 25 (31) and 34 (41) %, respectively. Only GFDS showed improved skill relative to CLIPER at all forecast times. Student t-test indicated that the improvement with GFDS over NOGAPS and AVN was statistically significant at all time levels at the 95 % confidence level, and the improvement over JTWC was at all forecast time levels except at 72 h. (Note that the GFDS forecasts in 1995 were run in an experimental mode at NCEP, and the forecasts were not available at JTWC during that period. JTWC started referring to GFDN forecasts in 1996 when the GFDL hurricane system was adopted in the Navy.) It is also noted here that the statistics for all of 212 NOGAPS track forecasts made in 1995 had similar errors to those for the 72 case samples used in this study. The NOGAPS performance shown here was probably representative of its performance in the western North Pacific in the 1995 season.

The homogeneous comparison presented above suggests that GFDS can produce useful dynamical model guidance for tropical cyclone forecasters in the western North Pacific region. A key factor in the success of the GFDL hurricane forecast system is probably the initialization methodology used (Kurihara et al. 1995). Dramatic reduction in track forecast errors occurred in the typhoon forecast system of the Central Weather Bureau in Taiwan when the filtering technique (Kurihara et al. 1995) was employed in their forecast model initialization (Chen et al. 1997).

It should be emphasized that the forecast comparison between GFDS and other models presented in this paper is not necessarily a strict model comparison, because each model used different analyses to define its initial condition. Also, one should note that the skill of a model is different in different basins and in different seasons, indicating the sensitivity of model performance to changes in environmental conditions. For example, while the overall performance of NOGAPS in the western North Pacific was not higher than GFDS during 1995, its performance in the Atlantic basin during that same year was comparable with GFDL forecasts.

3.2 Systematic bias

Forecast errors of the tropical cyclone position by a model tend to exhibit position dependent systematic bias. Figure 1 shows the distribution at two-degree resolution of the mean position error, i.e., systematic bias, of GFDS for the 24 and 48 h fore-

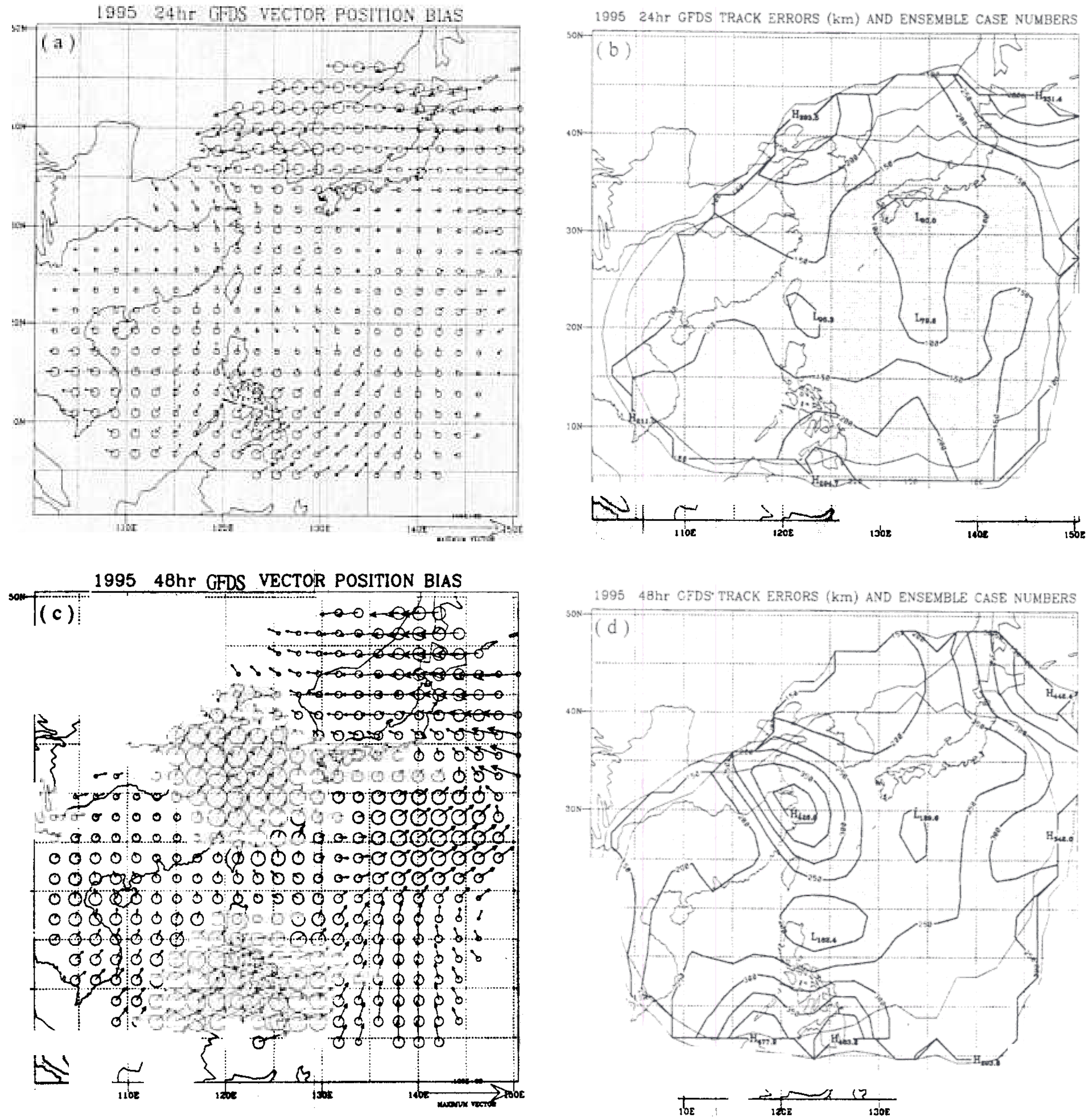

Fig. 1. (a) Systematic bias at 24 h of GFDS forecasts in 1995. The head, tail, and length of each arrow represent GFDS forecast position, the verifying position, and the mean position error in degrees, respectively. The radius of each circle represents the standard deviation of all position errors used in computing average. (b) the magnitude (bold solid line with contour interval of 50 km) of the systematic bias in (a), and the number of cases (thin solid line with contour interval of 5; regions with the number of cases equal to or larger than 10 are shaded); (c) and (d) are as for (a) and (b), but for 48 h.

cast times and the number of forecast cases averaged to obtain a bias. The vector plots in Figs. 1a and 1c were based on an ensemble of error vectors in a 10-degree-radius domain, while the standard deviation of errors for the cases used in bias computation was represented by a circle centered at a vector tail. In general, the radius of each circle was much shorter than the length of the corresponding error vector, indicating that the arrows were representative of the spatial distribution of the forecast position bias.

There exists a general northward bias in the region south of 30°N. Since most storms moved westward or northwestward in the low latitudes, the northward bias might suggest a tendency of GFDS to predict recurvature of storms too early. Figures 1a and 1c also show a general westward bias north of 30°N for both the 24 and 48 h forecast times. This resulted from the predicted motion of northeastward or eastward moving storms being too slow in this region, where eastward acceleration of storms was often observed.

The forecasts at 24 h tended to have a relatively small position error bias in the central part of Fig. 1b. Two regions with minimum systematic bias could be found: one near the strait between Taiwan and the Philippines, the other in the central Pacific between 20°N and 30°N, and 130°E and 136°E. In the latter region, the arrows were well contained in the circles (Fig. 1a), implying random scatter of error vectors in an ensemble. For the 48 h forecasts (Figs. 1c and d), minimum systematic bias was found to the northeast of the Philippines, as well as in the central Pacific between 25°N and 30°N, and 130°E and 133°E. One would expect that, as these minimum biases were calculated in regions with a relatively high number of cases (Figs. 1b and 1d), small systematic bias could mean a small error in individual cases. However, Fig. 1c indicates that the standard deviation was larger than the systematic bias to the northeast of the Philippines. Therefore, such minimum bias was not representative of errors in individual cases. A local maximum bias near eastern China was caused by several bad forecasts of Typhoon Janis. Generally, large bias existed near the boundary of the plotted region for both the 24 and 48 h forecasts. This was due to a few bad forecasts among the limited number of cases occurring there (Figs. 1b and 1d). For the 72 h forecast time (figure not shown), a strong northward bias was also found in the area south of 32°N.

By decomposing forecast errors of the storm motion in the cross-track/along-track directions (in the present analysis, storm's heading direction was estimated from the observed storm positions 6 h before and 6 h later), one may see the tendency of the model storm to move to the right or to the left of the storm's heading direction and, also, to move too fast or too slow [e.g., Fig. 4 in Rennick (1999)]. A scatter plot of forecast position errors relative to the storm's heading direction is presented in Fig. 2. Each quadrant of the diagram corresponds to a particular kind of bias. For example, points in the upper right quadrant represent forecasts that were faster than the actual movement and to the right (taken as positive) of the storm's heading direction. In consistent with the northward bias shown in Fig. 1 of westward moving storms in low latitudes, the plot revealed mean rightward bias relative to the storm's heading direction for all forecast times. The forecast position error vectors were also decomposed into zonal/meridional directions. As shown by the scatter plot in Fig. 3, general northward bias existed for all forecast times, which was consistent again with rightward bias of westward moving storms.

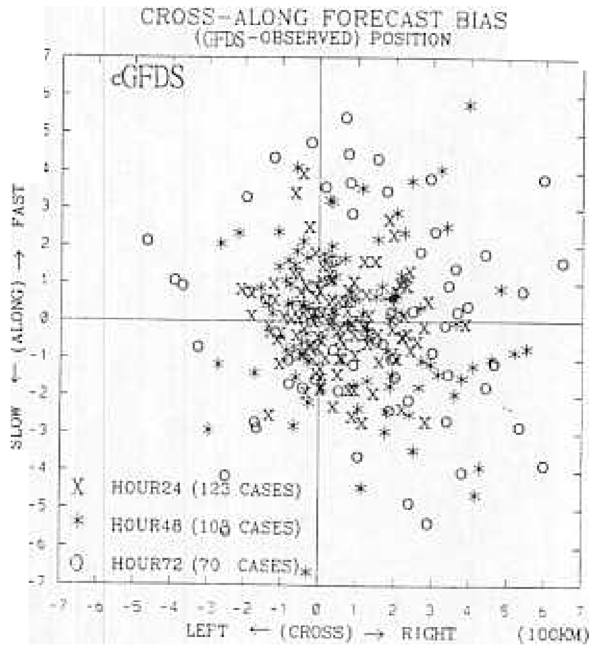

Fig. 2. Scatter diagram of the 24, 48 and 72 h GFDS cross-track (abscissa) and along-track (ordinate) forecast position errors (units of 100 km) for cases shown in Table 1. The ordinate axis points toward the storm's heading direction.

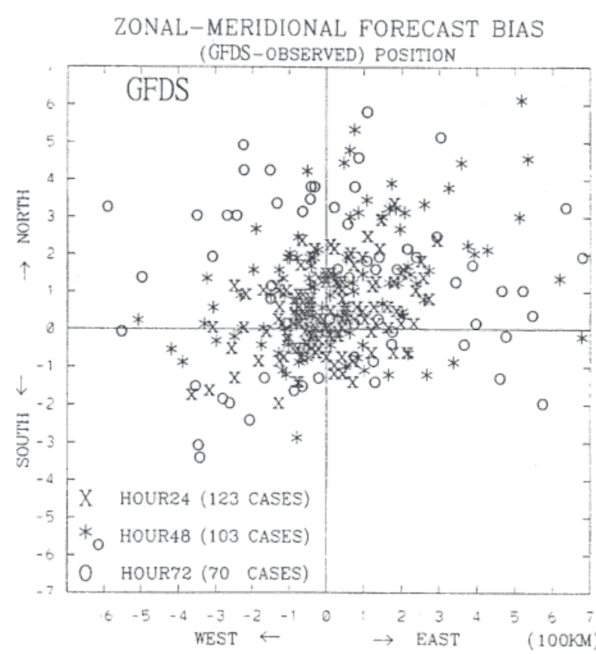

Fig. 3. Scatter diagram of the 24, 48 and 72 h GFDS forecast position errors (units of 100 km) for cases shown in Table 1. The ordinate axis points to the north. The abscissa points to the east.

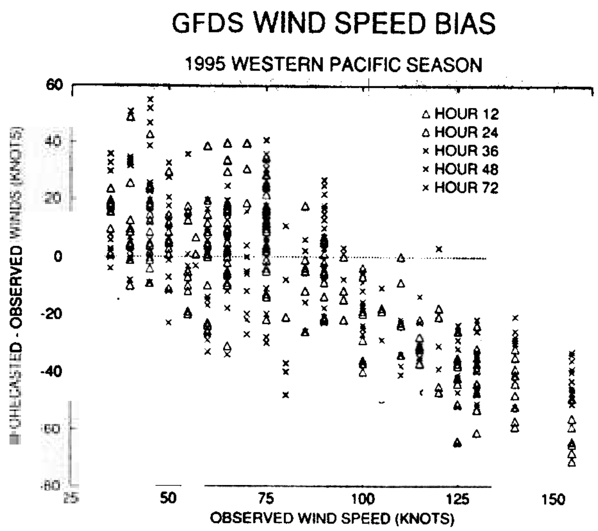

Fig. 4. The GFDS intensity forecast error (knots) as a function of storm intensity for cases shown in Table 1.

3.3 *Intensity forecast*

Intensity prediction remains quite difficult in the tropical cyclone forecasting. The plot of forecast intensity errors (Fig. 4) showed a tendency of GFDS to predict weak (strong) storms too strong (weak). A similar tendency has been observed in the case of the GFDL hurricane prediction system in both the Atlantic and eastern Pacific basins (Kurihara et al. 1998; Bender and Ginis 2000). Probably, the tendency of GFDS to underpredict intense storms is partly due to the insufficient finest resolution of the model (i.e., 1/6 degree) which still cannot adequately resolve the eye and eyewall structure. (It should be noted that in the 1995 version of GFDS, the maximum wind at the 40-m level was taken for the maximum surface wind rather than the conventional 10-m level wind.) The tendency to overpredict weak systems may be partly related to the scheme currently used in the initialization step which produces a vortex without realistic asymmetries.

The average errors of the intensity forecast by GFDS were 20.6, 21.7, and 23.1 knots at 24, 48 and 72 h, compared with 13.1, 19.8 and 25.2 knots for the JTWC official forecast. Although the errors from the GFDS were larger than those from the JTWC forecasts, the error increased by only 2.5 knots from 24 h to 72 h in contrast to a 12.1 knots increase for the same time period in the JTWC forecasts. When the GFDS intensity forecasts were corrected for bias in the early period, the results became more competitive with the JTWC forecasts. Thus, GFDS can still provide useful information for the forecast of storm intensity in some instances, particularly for the later forecast hours.

4. Results from forecast experiments in 1996

In this section, the error statistics of the GFDS forecasts in 1996 are presented and compared with the results for 1995. Detailed comparison of the two sets of forecasts (i.e., GFDN and GFDS) is also shown, emphasizing the differences and similarities in the distribution of the forecast bias.

4.1 *Average error: comparison between GFDN and GFDS*

The average forecast position errors for GFDS in 1996 for forecast times 12, 24, 36, 48 and 72 h were 108, 178, 227, 280 and 480 km, respectively, for 178, 166, 153, 143 and 114 cases (Table 5). Compared with GFDS forecasts in 1995, the position errors in 1996 increased by 10 % through 48 h, but decreased slightly at 72 h. The GFDS forecasts in 1996 were performed at non-synoptic times (06 and 18 UTC). It remains to be investigated whether the increase in position error in 1996 were partly related to the quality of the AVN global analysis at the non-synoptic time. Interestingly, for the same homogeneous cases, the position error of the CLIPER forecasts in 1996 increased by 30 % (20 %) at 24 h (48 h) compared with those in 1995, suggesting greater forecast difficulty in 1996. As a result, in spite of the reduction of position forecast skill of GFDS in 1996, its forecast skill relative to CLIPER in 1996 was about 5–10 % greater than in 1995. Overall, GFDS forecasts both in 1995 and 1996 showed superior forecasts to CLIPER for all forecast periods.

The average position errors of GFDN and GFDS for each forecast time are shown in Table 5, along with the percentage of cases in which GFDN exhibited a performance superior to GFDS. Comparison between these forecasts, which used the same model but based on different global analyses and forecasts, showed about a 3 % smaller average track error for GFDN through 48 h. At 72 h, GFDN forecasts exhibited about 11 % reduction in mean track error over GFDS. This was the only time level at which the difference was found to be statistically significant at the 95 % confidence level.

The spatial distribution of the forecast error (figure not shown) showed that both models had their largest errors over northern Japan and in the South China Sea between the Philippines and Vietnam. The forecast error for GFDS was particularly large in these two regions, with 600 km average at 48 h just east of central Vietnam compared with 450 km for GFDN. Both GFDN and GFDS exhibited smallest errors in the 22° to 34° latitudinal band, which was the same with the results from the GFDS forecasts in 1995.

Experimentally, forecast positions of the GFDN and GFDS were averaged to obtain composite forecasts (GFDA). It was found that the mean forecast error of GFDA at each time level was less than

Table 5. Forecast errors (standard deviations) in km for GFDS (AVN analysis), GFDN (NOGAPS analysis) and % of cases with superior performance of GFDN in 1996.

Unit: Km

	12 h	24 h	36 h	48 h	72 h
GFDS	108 (77)	178 (124)	227 (158)	280 (196)	480 (349)
GFDN	107 (78)	170 (118)	220 (142)	277 (260)	411 (256)
% of Superior performance of GFDN	51 %	51 %	48 %	46 %	58 %
Case number	178	166	153	143	114

Table 6. Forecast errors (standard deviations) in km for GFDA (average of GFDS and GFDN) and % of superior performance of GFDA compared to GFDS and GFDN in 1996.

Unit: Km

	12 h	24 h	36 h	48 h	72 h
GFDA	103(75)	161(117)	205(141)	255(164)	404(261)
% of Superior Performance Against GFDS	61 %	64 %	61 %	55 %	63 %
% of Superior Performance Against GFDN	55 %	58 %	63 %	62 %	51 %

the error of GFDN as well as of GFDS (Table 6) by about 10 %, while the standard deviation also slightly decreased. The improvement of GFDA was found to be statistically significant over GFDS at all forecast times shown, and over GFDN at all forecast times except 72 h.

4.2 Systematic bias: further comparison between GFDN and GFDS

As shown above, the storm position forecast was improved by combination of GFDN and GFDS. Figure 5 shows the distribution of the systematic bias in position forecast at 48 h for both GFDN and GFDS. For GFDN (Fig. 5a), a southward bias was indicated in the region south of 30°N and west of 140°E, except for a slight northward bias just east of Taiwan and over the South China Sea. North of 30°N GFDN had a strong westward bias. For GFDS (Fig. 5b), similar to the 1995 season, a general northward bias was indicated in the region south of 30°N. North of 30°N, the strong westward bias during the 1995 season was somewhat reduced in 1996 with more of a northward bias east of 140°E. Overall, the systematic biases of the two forecasts differed in many places, while similarities were found on the location of maximum and minimum biases. This means that the forecast positions of GFDN and GFDS tended to spread systematically in position-dependent different directions with respect to the best track position. Such a difference in the systematic bias of forecast position between GFDN and GFDS might partly explain why the reduction in position error occurred in GFDA.

Following Aberson et al. (1998), the relation between the error of an ensemble (forecast position error of GFDA) and the error spread within the ensemble (the forecast position difference between GFDN and GFDS), was investigated. In agreement with Aberson et al. (1998) and Goerss (2000), no clear correlation between the spread and the GFDA error at 72 h was found (Fig. 6). Nevertheless, the spread appears to crudely suggest the upper bound of the ensemble forecast error.

4.3 Supplemental experiments

In order to evaluate the impact of systematic bias (e.g., Fig. 1 for GFDS in 1995 or Fig. 5b for GFDS in 1996) on the tropical cyclone track forecast, supplemental experiments were conducted, in which systematic bias associated with each forecast system was subtracted from each forecast. The average position errors and standard deviations after the systematic bias correction are listed in Tables 7 and 8 for GFDS in 1995, and GFDS, GFDN and GFDA in 1996, individually. For each experiment, the reduction in both average position errors and standard deviations as compared to the prediction without the systematic bias correction is mostly more than 20 % for all forecast time. As we combine the GFDS forecasts both in 1995 and 1996 to form a larger forecast sample (Table 9) and to perform a new systematic bias calculation and correction, the improvement of more than 15 % at all forecast time is still evident. Results obtained from the above experiments clearly suggest the utility of the systematic bias in the operational forecasts, and issues related to this point

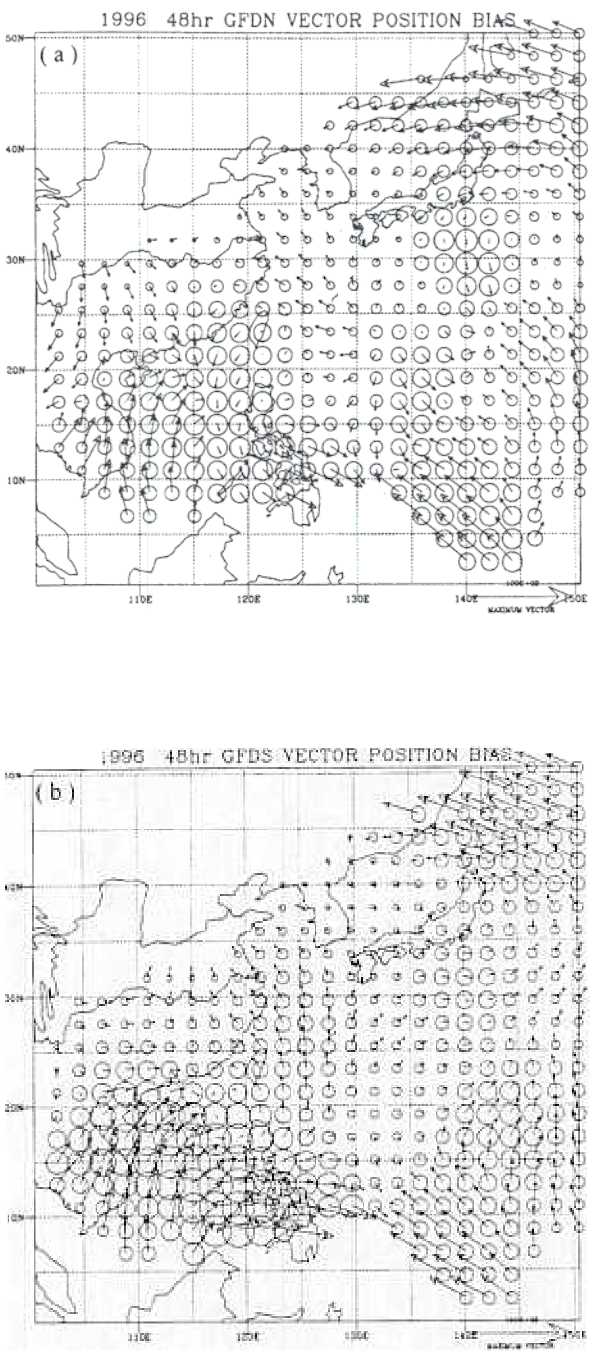

Fig. 5. Systematic bias of (a) GFDN (NOGAPS analysis) and (b) GFDS (AVN analysis) forecasts at 48 h for 143 cases (Table 5) in the western North Pacific in 1996. The head, tail, and length of each arrow represent the forecast position, the verifying position, and the actual mean position error in degrees, respectively. The radius of each circle represents the standard deviation of all position errors used in computing average.

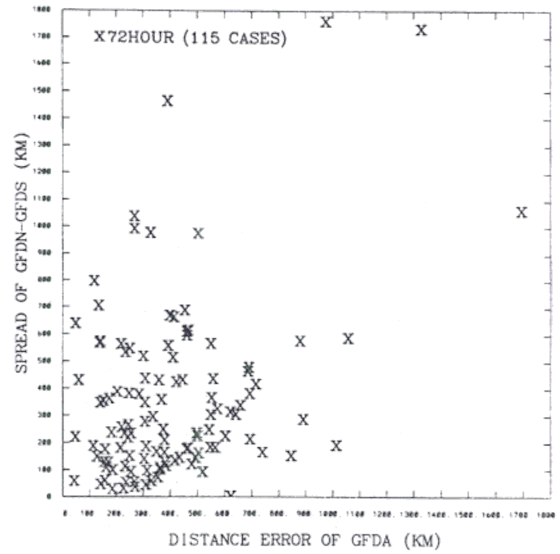

Fig. 6. Relation between forecast position error of an ensemble (abscissa) and the spread among the ensemble members (ordinate) at 72 h for cases shown in Table 5 in western North Pacific in 1996.

are further addressed in the summary.

4.4 Intensity forecast

The storm intensity was represented by the maximum wind at the conventional 10 m level in 1996, rather than by the 40 m level wind in 1995. As a result, each GFDS forecast showed about a 5-knot reduction in intensity from the forecasts in 1995. Still, in agreement with Rennick (1999), a tendency of the GFDL system to predict the intensity of weak storms too strong, and predict strong storms too weak, was indicated (figure not shown).

5. Summary and remarks

The GFDS system, which combines the model developed at the GFDL with the AVN global analysis, was experimentally run in the prediction of tropical cyclones in the western North Pacific basin. The number of cases treated by GFDS in 1995 (1996) was 125 (178) for 16 (24) tropical cyclones. The error analysis indicated superior track prediction skill of GFDS compared with CLIPER. The average position error in 1995 was nearly 30 % less than that of CLIPER.

Intensity forecast with GFDS was not satisfactory. A tendency to overpredict the intensity of weak storms and to underpredict the intensity of strong storms were indicated, similar to the tendency found in the Atlantic and East Pacific basins. Efforts to improve the GFDL hurricane prediction system has been under way with emphasis on the improvement in intensity forecast (e.g., Bender and Ginis 2000).

Table 7. Average position errors (standard deviations) in km for GFDS forecasts with systematic bias correction for cases in 1995, and % of improvement compared to those without bias correction.

Unit: km

	12 h	24 h	36 h	48 h	72 h
GFDS/bias-correction	78 (50)	106 (64)	141 (98)	183 (106)	315 (190)
% of improvement relative to GFDS	18 % (21 %)	27 % (26 %)	27 % (25 %)	27 % (32 %)	32 % (42 %)
Case number	125	123	115	103	77

Table 8. Average position errors (standard deviations) in km for GFDS, GFDN and GFDA forecasts with systematic bias correction for cases in 1996, and % of improvement compared to those without bias correction.

Unit: km

	12 h	24 h	36 h	48 h	72 h
GFDS/bias-correction	87 (60)	142 (104)	173 (123)	211 (146)	347 (218)
% of improvement relative to GFDS	19 % (22 %)	20 % (16 %)	24 % (22 %)	25 % (26 %)	26 % (38 %)
GFDN/bias-correction	85 (65)	133 (100)	176 (115)	212 (130)	311 (211)
% of improvement relative to GFDN	21 % (17 %)	22 % (15 %)	20 % (19 %)	23 % (50 %)	24 % (18 %)
GFDA/bias-correction	81 (59)	126 (97)	156 (113)	189 (127)	291 (182)
% of improvement relative to GFDA	21 % (21 %)	21 % (17 %)	24 % (20 %)	26 % (23 %)	28 % (30 %)
Case number	178	167	155	144	115

Table 9. Average track forecast errors (standard deviations) in km for GFDS forecasts for all cases in 1995 and 1996 without and with systematic bias correction, and % of improvement with bias correction.

Unit: km

	12 h	24 h	36 h	48 h	72 h
GFDS	103 (72)	164 (111)	213 (148)	267 (181)	474 (341)
GFDS/bias-correction	88 (60)	134 (95)	172 (115)	212 (142)	353 (214)
% of improvement relative to GFDS	15 % (17 %)	18 % (14 %)	19 % (22 %)	19 % (22 %)	26 % (37 %)
Case number	303	290	269	247	193

The GFDL model was also combined with the NOGAPS global analysis to establish a prediction system, GFDN. Both GFDS and GFDN were run for 178 homogeneous cases in 1996. It was indicated from the comparison of forecasts by two systems that, although average errors in forecast position were comparable with each other in distance, positions of GFDN and GFDS tended to be systematically biased toward different directions. When GFDN and GFDS forecasts were averaged, the mean error was 10 % smaller than the error for either individual forecast. This suggests that, depending on the spread of systematic biases of ensemble members, the ensemble average of forecasts may make a reliable guidance (Goerss 2000).

As demonstrated by the supplemental experiments, it is likely that the track forecast error would be reduced when the forecast could be corrected by bias of the prediction system used. In practice, however, distribution of systematic bias changes with time and, hence, it cannot be available beforehand. Yet, if the temporal change of the systematic bias is slow, the bias may show up in the forecast error as the season progresses to increase the forecast cases. Information on the forecast bias may be useful to forecasters, particularly to those acquiring only one forecast guidance. Of course, research for reducing bias in the storm track forecast is needed using various approaches such as the potential vorticity diagnostics (Wu and Emanuel 1995; Wu et al. 2000).

Acknowledgments

The authors thank Jerry Mahlman, Robert Tuleya and Jeffrey Anderson for their helpful remarks, and Masashi Nagata (JMA) and the anonymous reviewers for valuable comments. The authors also thank Steven Lord (NCEP) for kindly providing the data used in the present work. Special acknowledgment is given to Mary Alice Rennick of FNMOC for her valuable suggestions, for providing data necessary for this study, and for helping in the implementation of an automated system to provide the GFDN forecasts and storm information in near-real time. The computational support provided by the NCEP computer facility for this study is appreciated. Credit is also given to Yu-Tseng Cho (National Taiwan University), Kang-Ning Huang (Central Weather Bureau) and Jeffrey Varanyak (GFDL) for drafting assistance. The first author is partially supported through the National Science Council of Taiwan by Grants NSC87-2621-P-002-053 and NSC88-2325-Z-002-028.

References

Aberson, S.D., M.A. Bender and R.E. Tuleya, 1998: Ensemble forecasting of tropical cyclone tracks. *Preprints, 12th Conf. on Numerical Weather Prediction*, Phoenix, AZ, Ameri. Meteor. Soc., 290–292.

Bender, M., A., R.J. Ross, R.E. Tuleya and Y. Kurihara, 1993: Improvements in tropical cyclone track and intensity forecasts using GFDL initialization system. *Mon. Wea. Rev.*, **121**, 2046–2061.

——— and I. Ginis, 2000: Real case simulations of hurricane-ocean interaction using a high resolution coupled model:Effects on hurricane intensity. *Mon. Wea. Rev.*, **128**, 917-946.

Chen, D.-S., T.-C. Yeh, K.-N. Huang, M.-S. Peng and S.W. Chang, 1997: Performance of the typhoon-track forecast system in the Central Weather Bureau in Taiwan. *Proc., 22 Conf. on Hurricanes and Tropical Meteorology.* American Meteorological Society, Boston MA., 248–249.

Derber, J., H. Pan, J. Alpert, P. Caplan, G. White, M. Iredell, Y. Hou, K. Campana and S. Moorthi, 1998: Changes to the 1998 NCEP operational MRF model analysis/forecast system. Tech. Procedures Bull. 449. [Available from National Centers for Environmental Prediction, W/NP23, World Weather Building, Washington DC, 20233; or http://www.nws.noaa.gov/om/tpb/index.htm.]

Goerss, J.S. and P. Phoebus, 1992: The navy's operational atmosphric analysis. *Wea. Forecasting*, **7**, 232–249.

——— and R. Jeffries, 1994: Assimilation of synthetic tropical cyclone observations into the Navy Operational Global Atmospheric Prediction System. *Wea. Forecasting*, **9**, 557-576.

———, 2000: Tropical cyclone track forecasts using an ensemble of dynamical models. *Mon. Wea. Rev.*, **128**, 1187–1193.

Kurihara, Y., C.L. Kerr and M.A. Bender, 1989: An improved numerical scheme to treat the open lateral boundary of a regional model. *Mon. Wea. Rev.*, **117**, 2714–2722.

———, ———, R.E. Tuleya and R.J. Ross, 1995: Improvement in the GFDL hurricane prediction system. *Mon. Wea. Rev.*, **123**, 2791-2801.

———, R.E. Tuleya and M.A. Bender, 1998: The GFDL hurricane prediction system and its performance in the 1995 hurricane season. *Mon. Wea. Rev.*, **126**, 1306–1322.

Rennick, M.C., 1999: Performance of the Navy's tropical cyclone prediction model in the western North Pacific basin during 1996. *Weather and Forecasting*, **14**, 297–305.

Wu, C.-C. and K.A. Emanuel, 1995: Potential vorticity diagnostics of hurricane movement. Part I: A case study of Hurricane Bob (1991). *Mon. Wea. Rev.*, **123**, 69–92.

———, T.-S. Huang, K.-N. Huang and T.-C. Yeh, 2000: PV diagnostics of the impact of model initialization on the performance of a typhoon prediction system. *Proc., 24th Conf. on Hurr. And Tropical. Meteor.*, Ameri. Meteor. Soc., Boston, MA. 423–424.

Zhang, Z. and T.N. Krishnamurti, 1999: A perturbation method for hurricane ensemble predictions. *Mon. Wea. Rev.*, **127**, 447–469.

GFDL ハリケーンモデルを用いた台風予報：予報精度と AVN 及び NOGAPS 全球解析を用いた予報の比較

Chun-Chieh Wu
(国立台湾大学)
Morris A. Bender・Yoshio Kurihara[1]
(米国海洋大気庁)

GFDL において開発されたハリケーンモデルを AVN 及び NOGAPS 全球解析のそれぞれと結合して、台風予報システム GFDS と GFDN がつくられた。GFDS システムは、1995(1996) 年に 16(24) 個の台風について、125(178) 例の予報実験を行い、台風経路の予報で非常に良い成績を示した。12、24、36、48 及び 72 時間後の予報位置の平均誤差は、1995(1996) 年の場合、95(108)、146(178)、193(227)、249(280) 及び 465(480) km である。CLIPER 予報と比べると、約 30 %の精度向上となった。平均誤差と同時に誤差の標準偏差も減少したが、これは、低緯度では進路が北に偏る傾向があるものの、GFDS の予報精度と信頼度の高さを示すものである。一方、台風強度の予報は満足出来るものではなく、大西洋における誤差と同様に、弱い (強い) 台風を実際よりも強く (弱く) 予報し過ぎる傾向がみられた。

1996 年には二組の予報システム (GFDS と GFDN) で同時に予報が行われたので、二組の台風位置予報を比較した。台風の予報位置の誤差は、距離については両者は大体同程度であったが、それぞれの予報位置には、異なった方向に系統的に偏る傾向 (系統的偏差、場所に依存する) が認められた。その結果、二つの予報の平均をとると、それぞれの予報にくらべて予報誤差が 10 %減少した。予報の向上は、それぞれの予報を系統的偏差で修正する補足実験でも認められた。系統的偏差は定常ではないが、それを熱帯低気圧の予報精度向上に役立てることができるかもしれない。

[1] 現所属：地球フロンティア研究システム