An error analysis of objective tropical cyclone forecasting schemes used in Australia

T. D. Keenan, Head Office, Bureau of Meteorology, Melbourne
(Manuscript received March 1981; revised June 1981)

The performance of objective tropical cyclone forecasting schemes in Australia is examined on limited samples of 'best track' and 'operational track' independent test data. The mean forecast errors indicate that persistence forecasts were generally more accurate although in some cases the differences were not statistically significant. Biases associated with each technique are documented although, with the exception of NHC-67, they are found to be a relatively minor error component. Simple combinations of the individual forecasts are shown to produce lower errors, although the consistent use of persistence within the techniques limits the overall gain of such approaches.

Introduction

Forecasting the movement of tropical cyclones represents one of the most demanding problems facing a meteorologist. In the Australian region objective techniques are available to assist the forecaster. However, these aids can often place one in the invidious situation of having conflicting results from which some decision must be made. To assist in these situations, the comparative performance of the techniques and their inherent weaknesses must be known. This note attempts to partially satisfy this need by presenting some results pertinent to objective tropical cyclone forecasting schemes available in the Australian region. Specifically, some mean errors and the probability of forecast displacement errors will be presented.

The objective techniques examined were:

(i) **NHC-67** – an imported statistical scheme developed by Miller et al. (1968) for the northwestern Atlantic;

(ii) **CYCLOGUE** – an analogue forecasting model developed by Annette (1978);

(iii) **AUSTCYC-7072** – an **NHC-67** type scheme derived using Australian data by Kuuse (1979);

(iv) the **PERSISTENCE** technique, defined here as a forecast derived by multiplying the previous 12-hour motion of the cyclone by the ratio of the forecast interval to 12 hours.

Since November 1979 **CYCLOGUE** has been the only objective technique available to Bureau of Meteorology forecasters.

The performance of these techniques on 12-hour, 24-hour, 36-hour and 48-hour forecasts is initially examined on some 26 storms occurring in the Australian region between November 1973 and April 1975, providing independent test data for all forecasting schemes. **AUSTCYC-7075** (Kuuse 1979) has not been included as it was developed from storm data between 1970 and 1975. Note that these test data are not strictly homogeneous as some forecast situations were not tested by all techniques. The tracks of the test storms are shown in Fig. 1.

The results presented in Table 1 give the mean vector error, the standard deviation and the number of forecasts used in the derivation of the statistics. The disturbing trend revealed is that the general ranking of the schemes was **PERSISTENCE** first, followed by **CYCLOGUE**, **AUSTCYC-7072** and then **NHC-67.** Changing the test data to a homogeneous set (see Table 3) does not alter the fact that **PERSISTENCE** consistently produced the lowest forecast errors.

Mean errors are subject to sampling variability and thus the differences between the average errors have been evaluated using the standard t test. To allow for serial correlation between successive forecasts, the number of degrees of freedom was adjusted employing the method given by Siegal (1956). The results presented in Table 2 show the level of significance that can be attached to the null hypothesis that any two techniques have an identical mean error. At the 5 per

Fig. 1 Tracks of the 26 tropical cyclones occurring during the period November 1973 to April 1975.

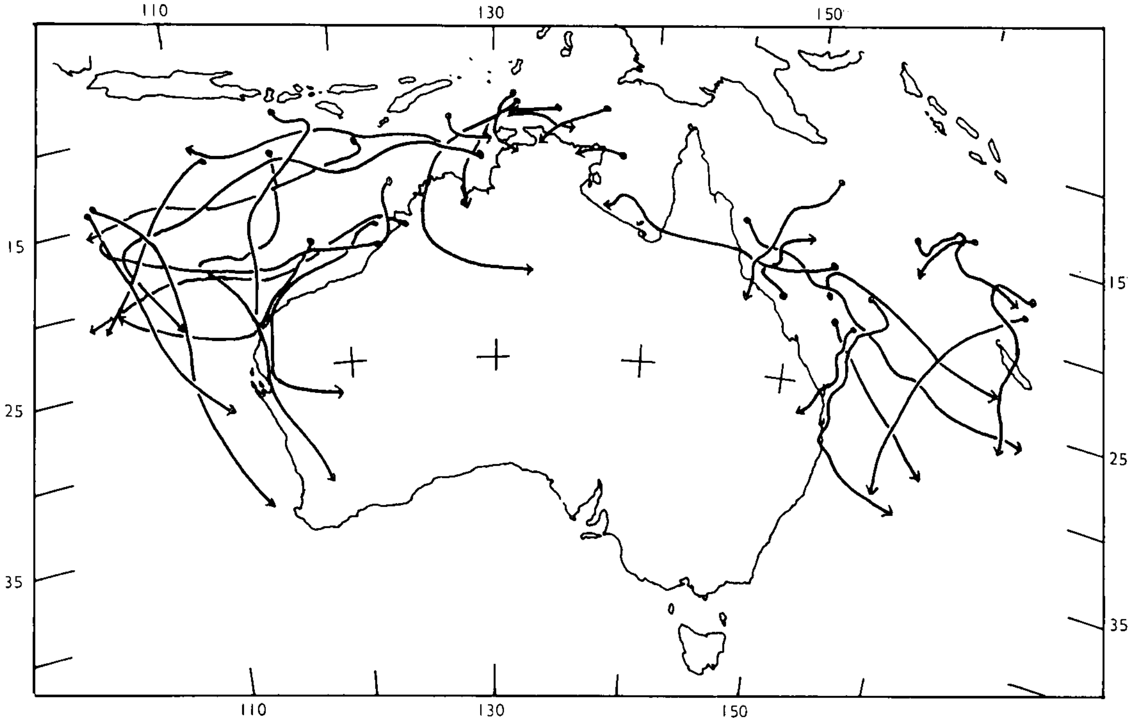


Table 1. Mean errors of objective tropical cyclone forecasts for the two cyclone seasons between 1973 and 1975.

	Forecast interval (hr)	*Mean error (km)*	*Std dev. (km)*	*No. of forecasts*
AUSTCYC-7072	12	77	51	117
NHC-67	12	95	56	117
PERSISTENCE	12	66	56	133
CYCLOGUE	12	74	47	97
AUSTCYC-7072	24	171	111	102
NHC-67	24	207	116	102
PERSISTENCE	24	150	111	117
CYCLOGUE	24	183	99	87
AUSTCYC-7072	36	317	188	90
NHC-67	36	373	189	90
PERSISTENCE	36	252	168	102
CYCLOGUE	36	309	164	80
AUSTCYC-7072	48	497	266	79
NHC-67	48	591	289	79
PERSISTENCE	48	375	252	90
CYCLOGUE	48	481	180	71

cent level of significance **PERSISTENCE** was not better than **CYCLOGUE** at the 12-hour interval, **AUSTCYC-7072** at the 24-hour interval and **CYCLOGUE** at the 48-hour interval. This inability to definitely separate the techniques at the shorter time intervals is expected since both **CYCLOGUE** and **AUSTCYC-7072** then depend significantly on persistence. At the other extreme **NHC-67** was significantly inferior to all the other techniques, at all time intervals.

The relative effectiveness of **PERSISTENCE** is at variance with results in other basins (e.g.

Table 2. Levels of significance for the null hypothesis that any two techniques have an equal mean forecast error.

Forecast interval		*Techniques* AUSTCYC-7072	NHC-67	PERSISTENCE	CYCLOGUE
12-hour	AUSTCYC-7072	0.0	0.007	0.043	0.306
	NHC-67		0.0	<0.001	0.001
	PERSISTENCE			0.0	0.112
	CYCLOGUE				0.0
24-hour	AUSTCYC-7072	0.0	0.019	0.072	0.210
	NHC-67		0.0	<0.001	0.056
	PERSISTENCE			0.0	0.016
	CYCLOGUE				0.0
36-hour	AUSTCYC-7072	0.0	0.027	0.006	0.390
	NHC-67		0.0	<0.001	0.009
	PERSISTENCE			0.0	0.015
	CYCLOGUE				0.0
48-hour	AUSTCYC-7072	0.0	0.038	0.003	0.045
	NHC-67		0.0	<0.001	<0.001
	PERSISTENCE			0.0	0.097
	CYCLOGUE				0.0

Brand et al. 1975) and the smaller study in the Australian basin undertaken by Annette (1978). It is also surprising because all the objective techniques, during the development, incorporated persistence in seeking the best predictor combination. These results indicate that the selected combinations were not optimum for this test sample, i.e. the underlying characteristics of the independent test data are different from the developmental data of the respective techniques.

The characteristics of the 1973–75 test sample are an important point. To assess how typical it was, **PERSISTENCE** was used to derive forecasts for all tropical cyclones occurring in the Australian region during the period November 1968 to March 1980, but excluding the above 1973–75 period. This showed that the 1973–75 period had a mean **PERSISTENCE** error lower at all forecast intervals (and significant at the 5 per cent level) than the longer-term sample. Although this does not show how the other techniques would have performed in comparision, it does imply that the 1973–75 test sample was biased to well-behaved storm tracks. However, even under such circumstances, the failure of the other objective techniques to compete and perform up to their developmental standard represents an important defect.

An alternative analysis of the errors, following Tracy (1966), is presented in Figs 2 to 5. Here for each forecast interval the spatial distribution of the actual forecast position relative to the verifying position (the origin) is shown. Also included is the mean displacement error, 40, 60 and 80 percentile ellipses (fitted on the assumption that the errors have a bivariate normal distribution), the area of the 80 percentile ellipses, and the number of forecasts. This presentation enables one to view any systematic bias associated with the schemes.

The results show that for all techniques, except **NHC-67,** the major errors were introduced in the zonal location of the storms. This is to be expected since the average zonal speed of storms is greater than their meridonal component. The biases introduced by each scheme increased systematically in magnitude with forecast interval.

In general terms the basic behaviour of the four techniques was as follows: **PERSISTENCE** tended to leave storms increasingly northward failing to capture poleward motion; **AUSTCYC-7072** forecasts generally produced an underestimate of the zonal motion; **CYCLOGUE** had a slight and probably insignificant southward bias, although its errors became increasingly concentrated in an east-southeast to west-northwest direction. Errors produced with **NHC-67** were greatly affected by a bias produced by a tendency to recurve the cyclones. The closer proximity of westerly flow to the Australian tropics may cause this tendency with the American **NHC-67** equations.

The relationship of the forecast errors to the cyclone motion is further explored in Figs 6 to 9, where following Neumann and Pelissier (1981) the errors are shown, as in Figs 2 to 5, but now relative to the storm motion at the forecast time. The direction of motion in this case was averaged over a 12-hour period centred at the forecast time. This display attempts to relate to a variable that is continuously varying at each forecast. Errors in the longitudinal sense can be interpreted as a

Fig. 2 Distribution of AUSTCYC-7072 forecast positions relative to the verifying position (the origin) at 12, 24, 36 and 48 hours. The centroid of the error distribution is denoted X and the 40, 60 and 80 percentile ellipses are shown. The numbers in the top right of each frame represent the mean east-west error (km), the mean north-south error (km), the area of the 80 percentile ellipse (km^2) and the number of forecasts.

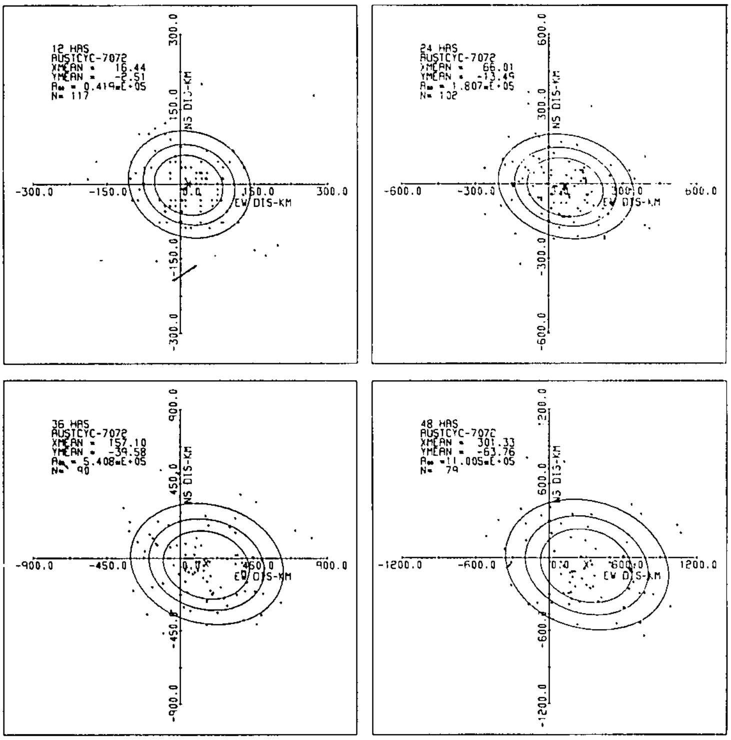


Fig. 3 As in Fig. 2 but for NHC-67.

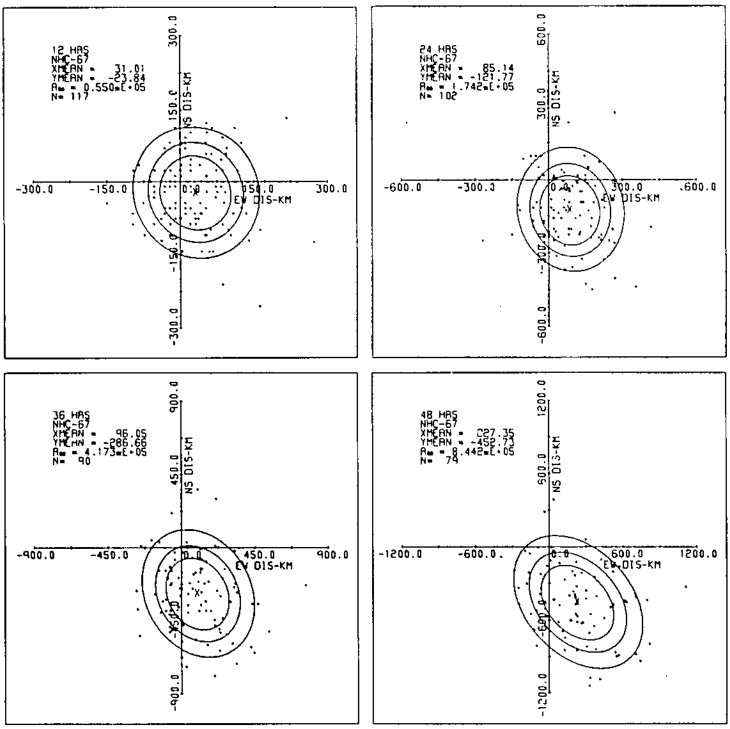


Fig. 4 As in Fig. 2 but for PERSISTENCE.

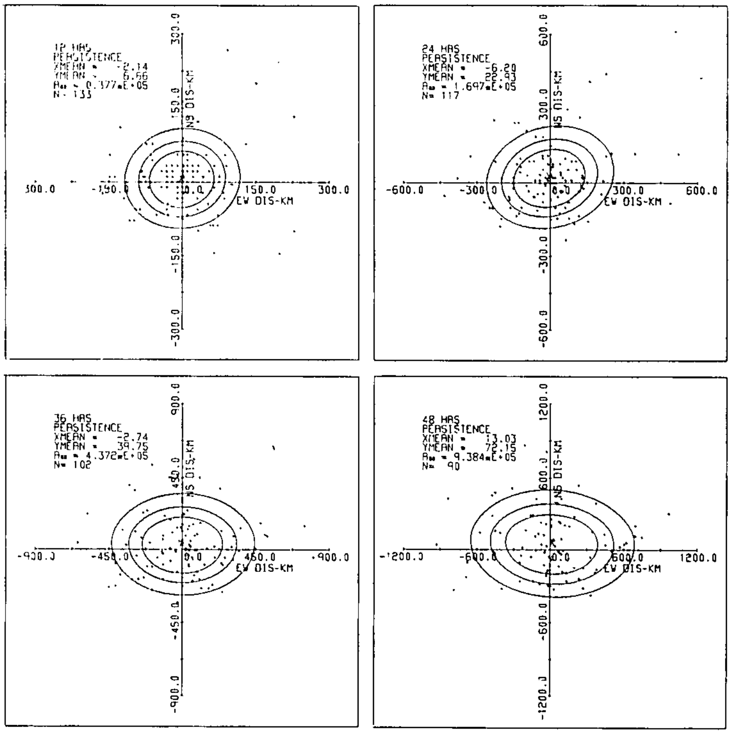


Fig. 5 As in Fig. 2 but for CYCLOGUE.

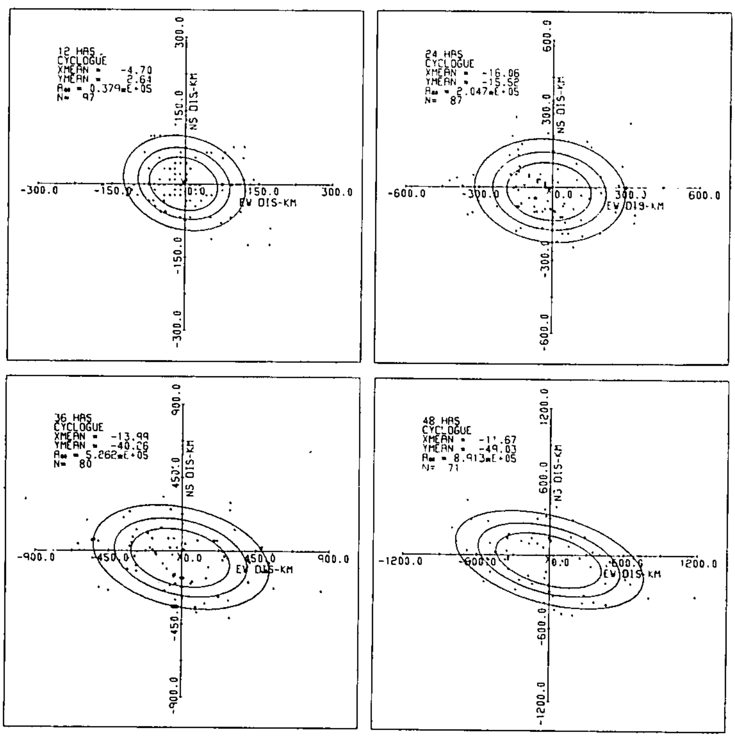


Fig. 6 As for Fig. 2 but with the Y axis now aligned along the direction of storm motion at the forecast interval. LG and TR denote the directions along, and transverse to, the storm motion.

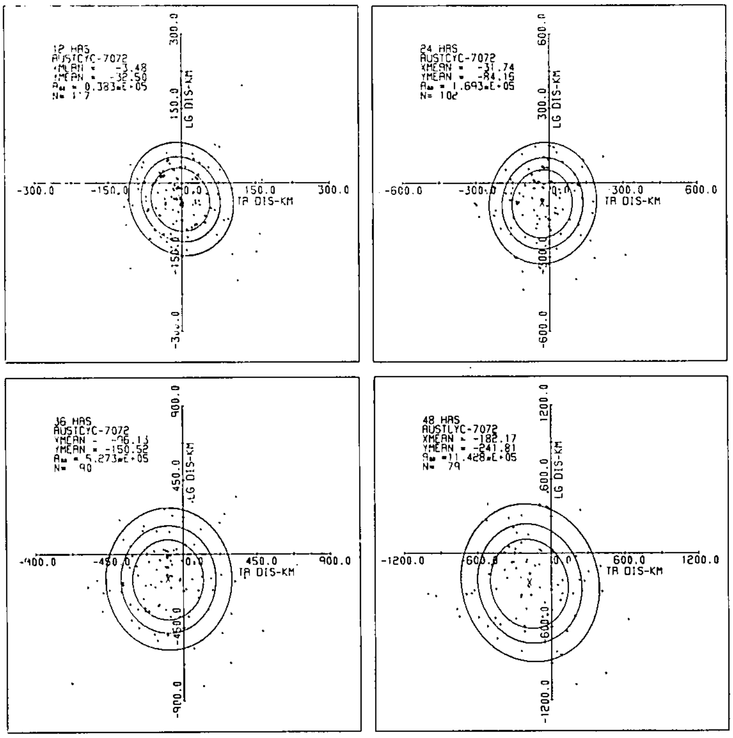


Fig. 7 As for Fig. 6 but for NHC-67.

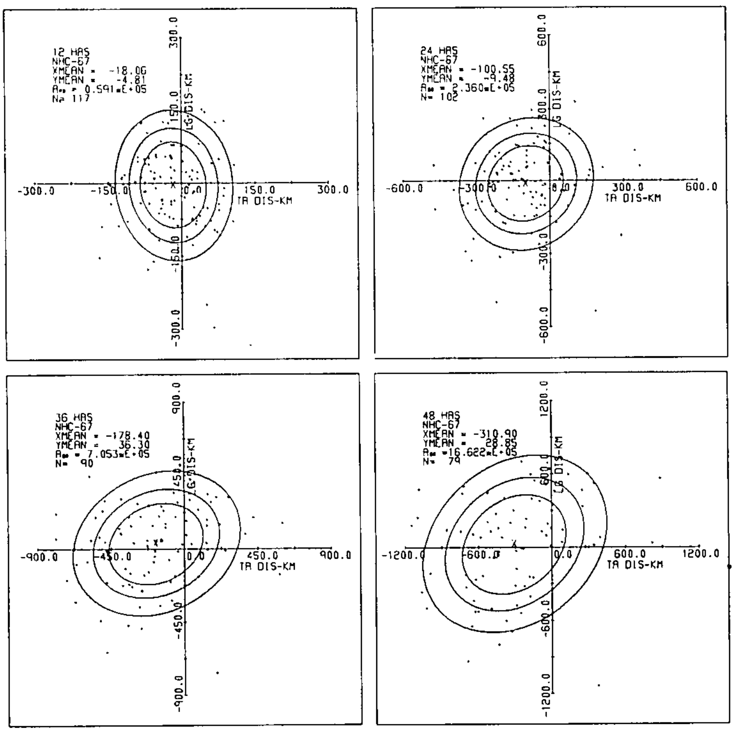


Fig. 8 As for Fig. 6 but for PERSISTENCE.

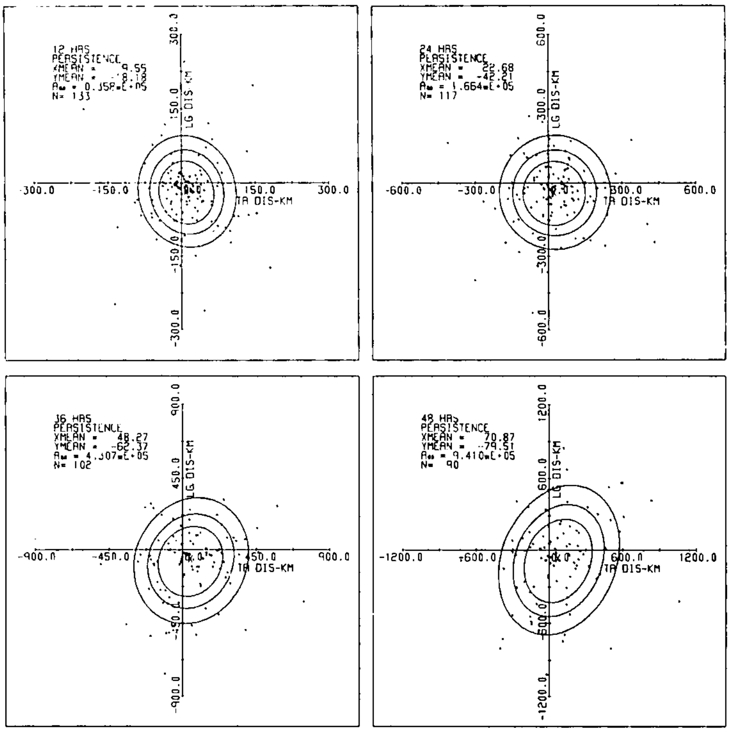
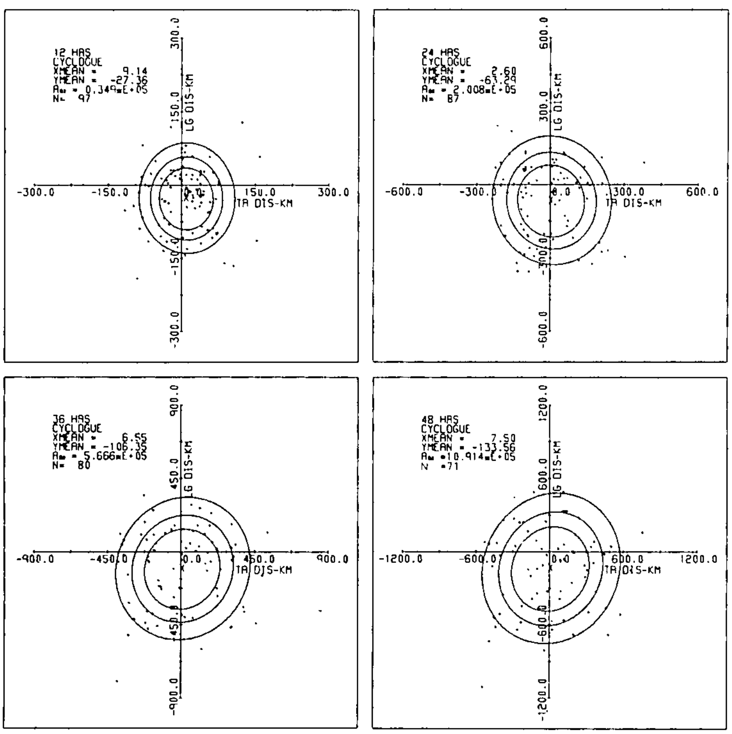


Fig. 9 As for Fig. 6 but for CYCLOGUE.



failure to capture along track motion such as positive or negative acceleration while transverse errors occur by incorrect prediction of turning motion. Note incorrect turning motion may also produce a longitudinal error.

Consistent with the previous distributions, most errors except for the **NHC-67** bias were introduced by the failure to capture along track motion. All the mean biases in that direction, except for **NHC-67,** were negative which a speed analysis confirmed was due to the schemes underestimating the future speed of movement. Both **AUSTCYC-7072** and **NHC-67** exhibited a bias to the left of the direction of motion, this latter bias being the major source of error with **NHC-67. PERSISTENCE** produced a small systematic bias to the right of the direction of motion although the failure to capture longitudinal motion was again indicated to be the major source of forecast errors.

The analysis to date has shown where the forecast errors are most likely to occur with each technique. However, it is of practical value to know whether these known characteristics can be employed to correct the forecasts and or combine them to obtain better results. The relative performance of the techniques corrected for biases — and when combined for a consensus forecast — has therefore been investigated.

The methodology for this study was to derive the basic biases from all the forecast data used previously. These biases were then applied to a homogeneous set of forecasts for the same period and a completely independent homogeneous set, selected randomly from operational forecasts issued during the 1977–79 cyclone seasons. Mean errors were then compared for each technique both with and without the bias corrections. The 'consensus' forecasts in this case were simply derived by averaging the forecast positions of the four techniques, both with and without bias corrections applied and by the use of a multiple linear regression equation approach. The latter technique employed a stepwise screening approach to minimise the mean square error in the zonal and meridional directions when using the bias-corrected component forecasts of the four individual techniques as potential predictors. The regression equations were derived from a homogeneous sub-set of the 1973–75 data and the 1977–79 data were used for the independent testing. This latter regression equation combination was selected because it was a simplistic way to combine already existing forecast data. A single screening of the potential predictors available to all techniques may yield better results.

The results summarised in Table 3 show that the removal of biases and the use of consensus forecasting had an increasingly favourable impact as the forecast interval increased. In fact, on the independent test data, the consensus approach that employed the averaging of the bias corrected forecasts produced the lowest mean errors. As expected from Figs 2 to 5 **NHC-67** and **AUSTCYC-7072** benefited most, and **PERSISTENCE** and **CYCLOGUE** the least, from the bias removal. However, neither the removal of the biases nor the use of the consensus approaches consistently provided significantly more skill than that available through the use of **PERSISTENCE.**

The failure of the bias correction alone to provide a major error reduction is due to the

Table 3. Comparison of errors on homogeneous sets of forecasts before and after the removal of biases and when employing consensus schemes.

	Dependent data Mean error (km)			*Independent data Mean error (km)*		
Forecast interval (hr)	12	24	36	12	24	36
Approach						
AUSTCYC-7072	74	165	310	90	183	310
AUSTCYC-7072 — BR	72	151	274	90	169	261
NHC-67	90	211	378	119	254	391
NHC-67 — BR	87	163	248	122	214	305
PERSISTENCE	61	141	246	93	177	239
PERSISTENCE — BR	61	142	248	93	178	247
CYCLOGUE	76	187	323	112	188	284
CYCLOGUE — BR	76	183	316	112	184	270
CONSENSUS	61	144	254	89	177	258
CONSENSUS — BR	61	135	233	90	167	223
Linear combination	60	128	209	91	189	277
Number of forecasts	86	78	71	39	31	35

BR — Bias removed
CONSENSUS — Average of all forecasts

relatively small contribution these terms make to each forecast. The consensus approach fails to produce substantial gains because the four techniques are not completely independent. This is demonstrated in Table 4 where the correlations between the component forecast errors, produced by each technique at the 24-hour interval, are given. With such positive correlations, the four techniques must on average provide a group forecast similar to that of an individual technique. By combining forecasts with lower correlations (see Thompson 1977) a greater increase in accuracy should be possible. In the present schemes, the consistent use of persistence is a factor contributing to the similarity of the individual forecasts.

Summary

A limited analysis of objective tropical cyclone forecasting schemes available in Australia has been undertaken on two independent test data sets. The average forecast errors have shown that their skill is generally no better than that obtained with persistence, although in some cases the differences were not statistically significant at the 5 per cent level. The analysis has also revealed where systematic errors occur with each technique and that, with the exception of **NHC-67**, these biases are small compared to the total error. Various approaches to combining the forecasts were also evaluated. These were found to result in marginally better performance, although the inherent consistency of the techniques limited the potential gains of such an approach.

The apparent favourable disposition of the 1973–75 test data to persistence forecasts indicates that longer-term comparisons should be conducted. This is presently being done, with the ultimate aim of discriminating the circumstances where objective techniques can be considered reliable.

Table 4. Correlation between the 24-hour forecast errors of the four objective techniques. Zonal error correlations are the upper values and meridional error correlations the lower values.

Technique	AUSTCYC-7072	NHC-67	PERSISTENCE	CYCLOGUE
AUSTCYC-7072	1.000	0.770	0.635	0.787
	1.000	0.713	0.686	0.655
NHC-67		1.000	0.668	0.684
		1.000	0.525	0.484
PERSISTENCE			1.000	0.728
			1.000	0.699
CYCLOGUE				1.000
				1.000

Acknowledgments

I wish to thank John Lorkin, Bernie Shanahan and Barry Southern for providing most of the forecast data used in this analysis and Mary Smedley for typing the manuscript. I also wish to thank the reviewers for their helpful comments.

References

Annette, P. 1978. Cyclogue — Analogic prediction of the course of tropical cyclones by National Meteorological Analysis Centre, Melbourne. *Tech. Rep. 28*, Bur. Met., Australia.

Brand, S., Long, J. M., Belloch, J. W. and Hamilton, G. D. 1975. A tropical cyclone analog program for the north Indian Ocean. *Mon. Weath. Rev.*, *103*, 258–61.

Kuuse, J. 1979. The statistical synoptic prediction of tropical cyclone movement in the Australian region. Paper presented at the International Conference on Tropical Cyclones, Perth, 25–29 November 1979.

Miller, B. I., Hill, E. C. and Chase, P. P. 1968. Revised technique for forecasting hurricane motion by statistical methods. *Mon. Weath. Rev.*, *96*, 540–8.

Neumann, C. J. and Pelissier, J. M. 1981. Models for the prediction of tropical cyclone motion over the North Atlantic: An operational evaluation. *Mon. Weath. Rev.*, *109*, 522–38.

Siegal, S. 1956. *Nonparametric Statistics*. McGraw-Hill Book Company, Inc., New York, 312 pp.

Thompson, P. D. 1977. How to improve accuracy by combining independent forecasts. *Mon. Weath. Rev.*, *105*, 228–9.

Tracy, J. D. 1966. Accuracy of Atlantic tropical cyclone forecasts. *Mon. Weath. Rev.*, *94*, 407–18.