# Performance of Multiple-model Ensemble Techniques in Tropical Cyclone Track Prediction

T.C. Lee & W. M. Leung
The Hong Kong Observatory, Hong Kong, China

## 1. Introduction

In recent years, weather forecasting centres have increasingly adopted single model or multiple model ensemble methods to optimize tropical cyclone (TC) predictions. For the single model approach, ensemble members are generated from a single numerical model using different initial conditions representing the uncertainty of the analysis [1,2,3]. The multiple model ensemble method makes simple averaging of the forecast outputs of several models from different centres [4,5].

The Hong Kong Observatory (HKO) began in 1999 to experiment with a multiple model ensemble method based on the equally weighted (EW) average of the forecast positions of several global numerical models. The forecast positions are extracted from prognostic products of the European Centre for Medium-Range Forecasts (ECMWF), the Japan Meteorological Agency (JMA) and the UK Meteorological Office (UKMO). In view of its superiority over the forecasts of individual models, the EW multiple model ensemble method is put into operationally use in 2002. Official tropical cyclone forecast tracks followed closely the multiple model ensemble forecast unless there were large discrepancies amongst the model forecasts.

In the study presented in this paper, the performance of the EW multiple model ensemble method in forecasting tropical cyclone tracks in the South China Sea (SCS) and the western North Pacific (WNP) for the 3-year period from 1999 to 2001 was evaluated and compared with the forecasts of the three global models. The HKO's subjective TC track forecast in 2002 was also verified to investigate the effect of the ensemble method on operational TC track forecast. Lastly, possible improvements in the ensemble method using weighting schemes based on initial position error and 12-hour forecast errors were briefly discussed.

## 2. Data Sources

Model forecast TC positions within the SCS and WNP (0-45$^o$N, 100-180$^o$E) for the period of 1999-2001 were used in this study. The forecast TC positions of the ECMWF and JMA models were determined from the surface prognoses as the point of minimum mean sea-level pressure (MSLP) identified by overlapping parabolic interpolation [6]. The forecast TC positions of the UKMO global model are extracted from the TC guidance of UK Meteorological Office (Bracknell) received via the WMO Global Telecommunication System. Table 1 lists the forecast start times and forecast intervals of the model outputs available to HKO in the period 1999-2001.

As only the prognoses for 12 UTC (but not 00 UTC) were available from ECMWF, the 36-, 60-, and 84-hour TC forecast positions were obtained by linear interpolation of the forecast positions in the preceding 12 UTC ECMWF model run. They were used to calculate the 24-, 48-, and 72-hour ensemble forecast positions for 00 UTC. The 00 and 12 UTC TC forecast positions of the three models and those of the ensemble forecasts were verified against HKO's 'best track' positions. Verification was conducted only for those cases in which prognostic data from all three global models were available.

The HKO's 24- and 48-hour subjective TC track forecasts in 2002 were also verified against the "best track" within Hong Kong's area of responsibility for issuing TC warning for shipping (10-30$^o$N, 105-125$^o$E).

## 3. Performance of Equally Weighted Ensemble Method against Individual Models

The multiple model ensemble method currently used by HKO assigns equal weights to the forecast TC positions of the three models to determine the average position. Table 2 shows the mean and standard deviations of the 24-, 48-, and 72-hour forecast errors of the EW ensemble method and those of the individual models in 1999-2001.

It can be seen that the mean forecast errors of the JMA and UKMO models were similar for all forecasting periods. The mean forecast error of the ECMWF model was comparatively larger. The poorer performance of the ECMWF model, especially for the 24-hour forecast period, may be partly due to the fact that lower resolution ($2.5^{o}$ * $2.5^{o}$, in contrast to JMA's $1.25^{o}$ * $1.25^{o}$) forecast products were available for this study and that 00 UTC forecast positions were obtained by linear interpolation of the forecasts of the preceding 12 UTC model run.

Comparing with the forecast errors of individual models, the EW ensemble method outperforms the best of the three models in 1999-2001. The overall mean forecast error of the EW ensemble method at 24, 48, and 72 hours is about 7%, 17% and 19% less than the mean forecast errors of the best of the individual models respectively. Moreover, the standard deviation of forecast errors of the ensemble method is smaller than those of the individual models, especially for the 72-hour forecast. These results agree with the findings of recent studies [5,7] in which different combination of models and shorter data periods were used.

## 4. Performance of HKO's Subjective Forecast in 2002

In view of the encouraging results mentioned in Section 3 above, HKO puts the EW multiple model ensemble method into operational use in 2002. Operational guidance has also been established to assist forecasters in formulating the official TC forecast track using the ensemble model as a primary guidance. Forecasters can exercise their judgment to apply meteorological reasoning to adjust the subjective forecast track if the models give widely divergent track forecasts.

To assess the effect of the ensemble method on HKO's TC track forecasting skill, the errors of the HKO's subjective forecasts in 2002 were compared with those obtained by the climatology-persistence (CLIPER) method [8]. In general, the skill of a particular forecast relative to the CLIPER forecast is represented by the skill score, where:

$$\text{skill score} = \frac{(\text{CLIPER error - error of a particular forecast})}{\text{CLIPER error}} \times 100\%$$

The skill score is therefore a measure of normalized improvement over CLIPER. Positive skill indicates that the forecast outperforms the CLIPER forecast and negative skill indicates otherwise.

Table 3 shows the mean forecast errors and skill scores of the HKO subjective forecast and EW ensemble forecast in 2002. There is a significant reduction in the 24- and 48-hour mean forecast errors of the HKO subjective forecast when compared with the corresponding mean forecast errors in the period 1997-2001. Also, the skill score of the 48-hour subjective forecast of HKO reaches an all-time high of 55% in more than two decades (Figure 1). This shows that the skill of ensemble forecast has translated into a noticeable improvement in HKO's official TC forecast.

## 5. Possible Improvement in Multiple Model Ensemble Method

As discussed in Section 3, the UKMO and JMA models had, on average, smaller errors in TC track forecast when compared with the ECMWF model in 1999-2001. However, for 24-hour forecast, there were still over 30% of the cases in which the ECMWF model performed better than the JMA and/or UKMO models (Figures 2 and 3). The percentages for 48- and 72-hour forecasts even increased to some 40% and 50% respectively. This means that none of the three models is superior to the other two on all occasions. It should therefore be possible to reduce the ensemble forecast error by assigning more weight to the models with better performance on a case-by-case basis. Lee et al. [9] recently tested two ensemble weighting schemes based respectively on the direct position error of (a) initial position (DPE00) and (b) the 12-hour forecast position error (DPE12) to see if more accurate ensemble forecasts could be made. The results (Table 4) show that while the EW scheme and DPE00 scheme have similar performance, the DPE12 scheme can improve the ensemble forecasts in the 24- and 48-hour forecast periods.

## 6. Conclusion

This study has shown that the multiple model ensemble method using the EW scheme outperformed all individual models in TC track prediction in 1999-2001. Reductions in 24-, 48-, and 72-hour forecast errors are about 7%, 17% and 19% respectively. The use of the EW multiple model ensemble method has resulted in significant improvement of the performance of HKO's official TC forecast. It is possible to further improve the performance of the ensemble forecast by using a different weighting scheme based on the 12-hour forecast position errors of individual members of the ensemble.

## References


Aberson, S. D., Lord, S. J., DeMaria, M., and Tracton, M. S., (1995). Short-range ensemble forecasting of hurricane tracks. In *Preprints, 21st Conf. Hurr. Trop. Meteorol., Miami*, *Am. Meteorol. Soc.*, 290-292.

Cheung, K. K. W., (2001). A review of ensemble forecasting techniques with a focus on tropical cyclone forecasting. *Meteorol. Appl.* **8**, 315-332.

Zhang, Z., and Krishnamurti, T. N., (1997). Ensemble forecasting of hurricane tracks. *Bull. Am. Meteorol. Soc.*, **78**: 2785-2795.

Aberson, S.D. (2001). The ensemble of tropical cyclone track forecasting models in the North Atlantic Basin (1976-2000), *Bull. Amer. Meteor. Soc.,* **82** , 1895-1904.

Goerss, J. S. (2000). Tropical cyclone track forecasts using an ensemble of dynamical models. *Mon. Wea. Rev.*, **128**, 1187-1193.

Manning, K. W. & Haagenson, P. L. (1992). Data ingest and objective analysis for the PSU/NCAR modeling system: Programs DATAGRID and RAWINS. *NCAR Technical Note, NCAR/TN-376+IA*, 209 pp.

Lam, H., & Ma, M. H. (2002). A review of the use of NWP models for tropical cyclone track prediction at Hong Kong Observatory in 2001 (in Chinese only). 16th Hong Kong-Guangdong-Macau Symposium On Met. Science & Technology (Guangzhou, 30 to 31 Jan 2002).

Bell, G. J. (1962). Predicting the movement of tropical cyclones in the region of the China sea. Proceedings of the Inter-Regional Seminar on Tropical Cyclones, Tokyo, 1962. (Techn. Rep. JMA, No. 21).

Lee, T. C., and Wong, M. S. (2002). The use of multiple model ensemble techniques in tropical cyclone track forecast at the Hong Kong Observatory. WMO Commission for Basic Systems, Technical Conference on Data Processing and Forecasting Systems (Cairns, Australia, 2 to 3 December 2002).

***Table 1*** *Details of model run schedules and forecast intervals of TC forecast products available to HKO in 1999-2001.*

| Models | Forecast start time | Forecast intervals |
|---|---|---|
| ECMWF | 12 UTC | 24-hour |
| JMA | 00 & 12 UTC | 12-hour |
| UKMO | 00 & 12 UTC | 12-hour (24-hour before July 2001) |

***Table 2*** *24-, 48- and 72-hour forecast mean errors (standard deviation) in km*

| Forecast period (hr) | ECMWF Mean error (standard deviation) | UKMO Mean error (standard deviation) | JMA Mean error (standard deviation) | EW Mean error (standard deviation) |
|---|---|---|---|---|
| 24 | 214 (161) | 140 (94) | 147 (95) | 130 (86) |
| 48 | 309 (245) | 256 (168) | 256 (173) | 213 (153) |
| 72 | 417 (302) | 390 (247) | 379 (270) | 308 (204) |

***Table 3*** *Summary of mean forecast errors and Skill Scores of the HKO subjective forecast and EW ensemble forecast in 2002*

| Forecast Hours | 24 hours | | 48 hours | |
|---|---|---|---|---|
| Forecast | HKO | Ensemble | HKO | Ensemble |
| Mean Forecast Error (km) | 119 | 97 | 212 | 165 |
| Skill Score | 50% | 59% | 55% | 65% |
| 5 Year (1997-2001) Mean Error of HKO subjective forecast (km) | 178 | | 341 | |

***Table 4*** *Mean and standard deviations of 24-, 48-, and 72-hour forecasts of the EW, DPE00, and DPE12 schemes*

| Forecast period (hr) | EW Mean error (standard deviation) | DPE00 Mean error (standard deviation) | DPE12 Mean error (standard deviation) |
|---|---|---|---|
| 24 | 130 (86) | 123 (80) | 108 (74) |
| 48 | 213 (153) | 214 (148) | 204 (139) |
| 72 | 308 (204) | 320 (212) | 308 (190) |

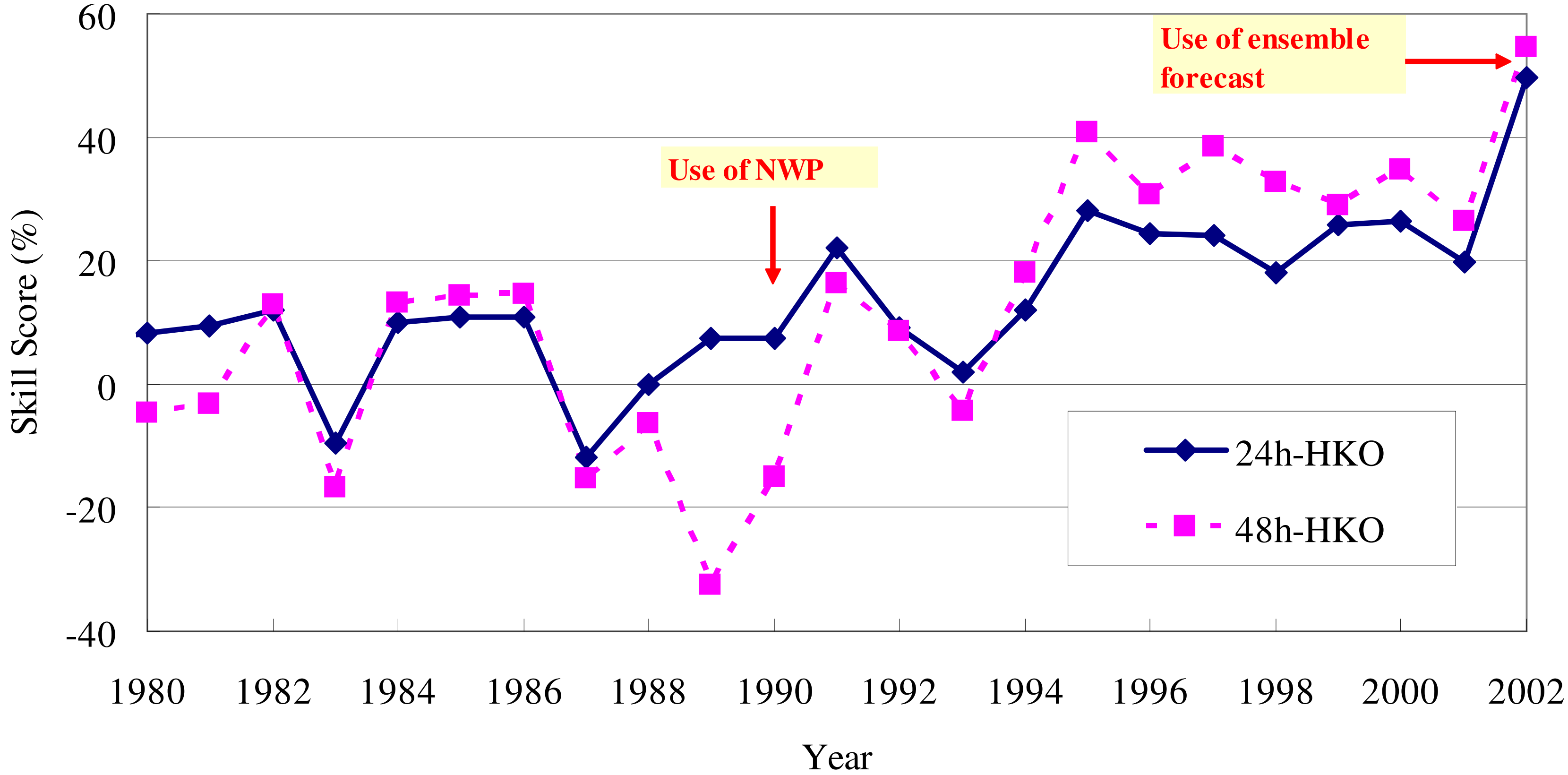


**Figure 1** Skill relative to CLIPER of the 24- and 48-hour HKO subjective TC track forecast, 1980-2002 (Area : 10-30°N, 105-125°E).

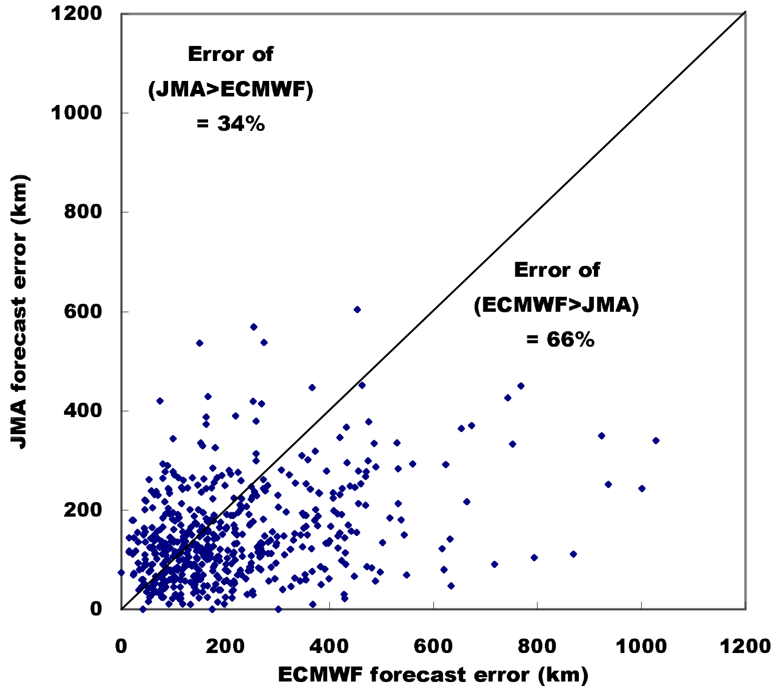


***Figure 2*** *Scatter plot of 24-hour forecast error of JMA against EMCWF.*

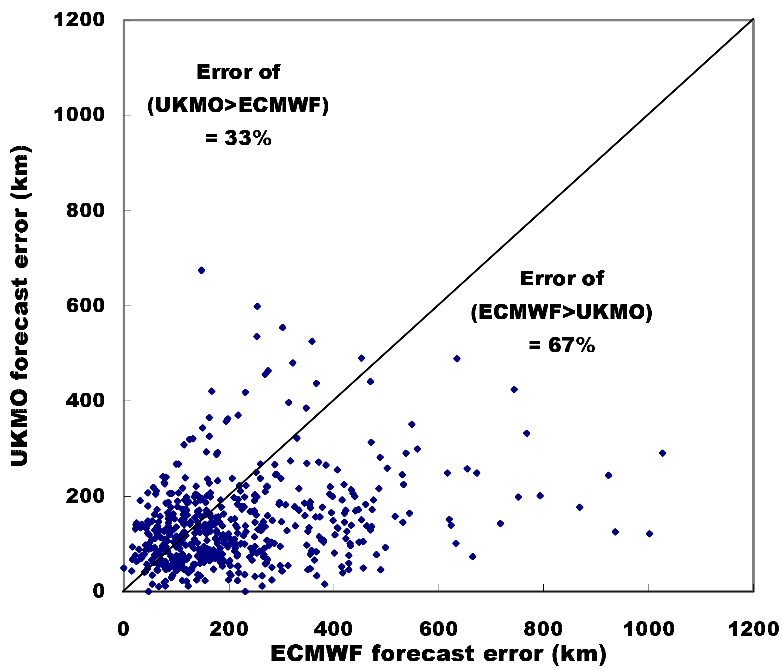


***Figure 3*** *Scatter plot of 24-hour forecast error of UKMO against EMCWF.*