香港天文台
HONG KONG OBSERVATORY

Reprint 1338

Performance of weighted multi-model ensemble and selective ensemble approaches for prediction of TC positions

CHAN Yan-chun

The 32nd Guangdong - Hong Kong - Macao Seminar
on Meteorological Science and Technology
and
The 23rd Guangdong - Hong Kong - Macao Meeting
on Cooperation in Meteorological Operations

(Macau 8-10 January 2018)

多模式權重及篩選性集合預報在熱帶氣旋路徑預測的表現

陳恩進
香港天文台

摘要

熱帶氣旋伴隨強風、強降水和風暴潮，往往對社會造成嚴重影響。準確預測熱帶氣旋位置對於評估並及早應對天氣帶來的影響非常重要。此研究驗證了基於多個確定性模式和集合預報系統（EPS）的熱帶氣旋路徑預測之表現。2012 至 2016 年的結果顯示歐洲中期天氣預報中心（ECMWF）的確定性模式和集合預報系統較其他模式優勝。在採用多模式權重方法後，路徑預測的表現更稍有改善。

此研究亦使用篩選性集合方法來預測近期的熱帶氣旋包括 2017 年的苗柏、天鴿和帕卡的路徑，主要是將 12 小時預測位置誤差較大的成員篩去才計算集合平均。結果顯示，篩選性集合方法能改善個別熱帶氣旋的路徑預測。

Performance of weighted multi-model ensemble and selective ensemble approaches for prediction of TC positions

CHAN Yan-chun
Hong Kong Observatory

Abstract

Tropical cyclone (TC) very often brought severe impact to the society by its accompanying high winds, heavy precipitation and storm surge. Accurate forecast of the tropical cyclone's position is important to the assessment and early response to the weather impact. This study verified the performance of TC position forecast based on multiple deterministic models and ensemble prediction system (EPS) models. Results from 2012 to 2016 showed that ECMWF's deterministic and EPS models outweighed other models. A weighted ensemble approach was also adopted and slight improvement in position forecast was observed.

Study on selective ensemble approach for forecasting TC position for recent TC cases including Merbok, Hato and Pakhar in 2017 was pursued. It involved taking a grand ensemble mean after screening out individual members with larger 12-hour forecast position error. Results showed that this selective ensemble approach improved the position forecast for a few specific cases.

1. Introduction

Over a year, there were on average about 30 tropical cyclones (TC) formed over the western North Pacific and about 10-15 over the South China Sea. Some of the TCs might move towards the coast of Guangdong and pose direct threat to Hong Kong, especially in the peak season from July to September. Accurately forecasting the TC position is undoubtedly important in enabling members of the public, Government departments, public utilities and relevant stakeholders to take necessary precautionary measures in a timely manner. The Hong Kong Observatory (HKO) issues TC forecast track for at least twice daily when a TC develops and centres within the area of responsibility (7-36N, 100-140E). The forecast period of the TC track is up to 120 hours ahead in 24-hourly intervals.

Tropical cyclone is a very complicated weather system that its evolution sometimes cannot be fully simulated by numerical weather prediction (NWP) models despite the advances made in recent decades on physical parameterization and data assimilation, leading to improved performance in general. Not only the structure of a TC was not fully understood, but also the mechanism or physical processes could not be fully represented in the NWP models. With different techniques employed in assimilating different observational data, there are differences in the initial conditions among NWP models. The model physics and parametrizations used by various models are different such that the evolution of weather systems predicted by different models could deviate from each other. As a result, there are errors in forecast TC position that grow with forecast time and that could reach the order of thousands of kilometres in 120-hour forecast.

Past studies showed that the position error could be reduced by using ensemble of TC forecast tracks from different models, including ECMWF, JMA, NCEP and EGRR [1]-[2]. In operation, the ensemble TC track provided a base reference of objective forecast track for constructing a subjective TC forecast track. In order to further improve the TC forecast technique, minimising position error by fine-tuning the method of taking ensemble was the focus of this paper.

2. Discussion

The study first investigated the error behaviour of global models in the past few years. Major global deterministic models including ECMWF, JMA, NCEP and EGRR were considered. For global models, the forecast track was either provided by bulletins through Global Telecommunication System (GTS) or by retrieving the minimum mean sea level pressure through interpolation of model grid point data.

There were occasions where the TCs forecast by models were too weak that the position could hardly be determined by pressure field. The vorticity field at level of 925hPa or 850hPa would be used as replacement for operational purpose. The operational data from HKO between 2009 and 2016 were retrieved to perform the analysis. The forecasts from ensemble prediction system (EPS) were also considered with ensemble mean of ECMWF EPS track members being collected from 2009 to 2016. Meanwhile, data used operationally from 2014 to 2016 for JMA Typhoon EPS, NCEP GFS EPS and EGRR EPS were also included in this study.

The direct position error (DPE) was calculated by comparing the forecast with HKO best track data. To address the unavailability of some model forecasts, common data set with only all models and the corresponding best track positions being available would be considered. TC forecast tracks with initial position located outside the area of responsibility (7-36N, 100E-140E) were excluded in the error analysis. The results were illustrated in Figure 1.

A general decreasing trend in DPE was observed from 2009 to 2016 for global models. The tendency was more obvious at T+24 forecast although there were still fluctuations as the error value was sensitive to different TC cases.

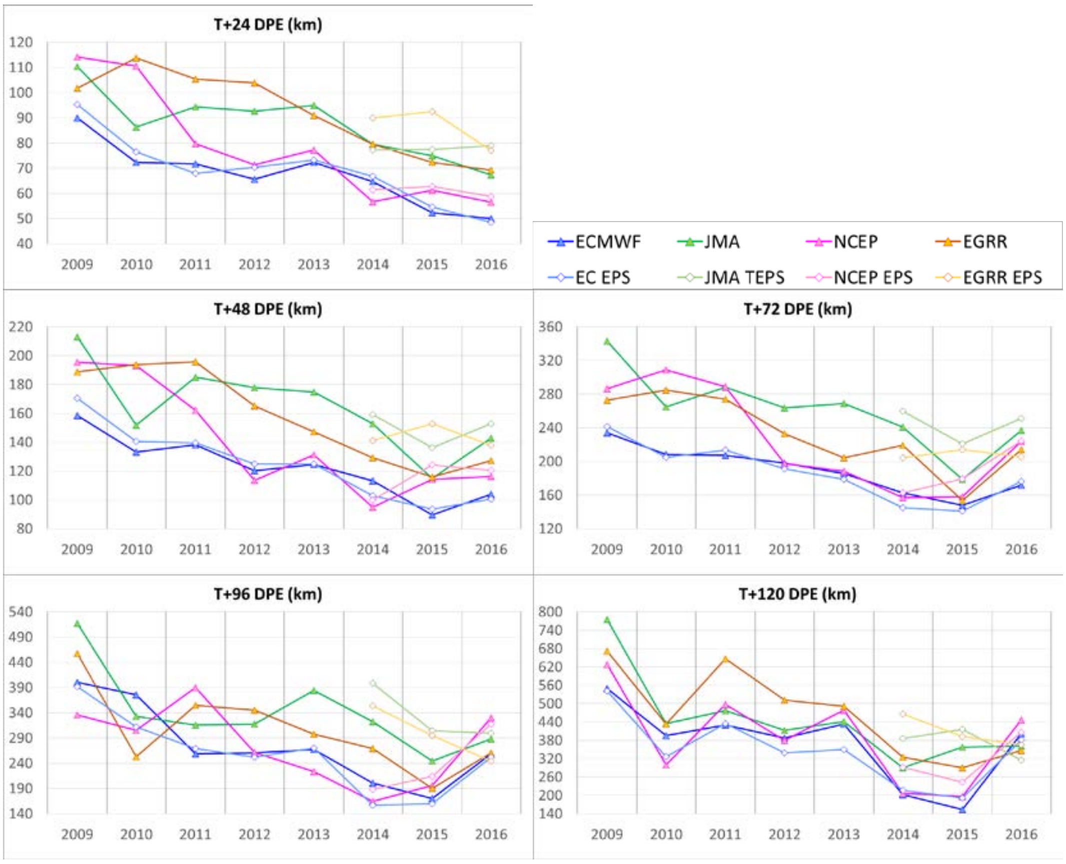


Figure 1 Direct positions error (DPE) at T+24, T+48, …, T+120 hours for various deterministic and EPS forecasts, based on data from 2009 to 2016 within common data set.

The along track error (ATE), which was equivalent to the error in predicting the speed of TC movement was plotted in Figure 2. Positive value means that TC forecast motion was faster than the actual situation. The average value represented the bias in TC forecast speed. Models in general showed a behavior of slow bias, matching with the results in previous study [2]. Among different models, JMA model generally showed the slowest bias, though the bias was once reduced in 2015.

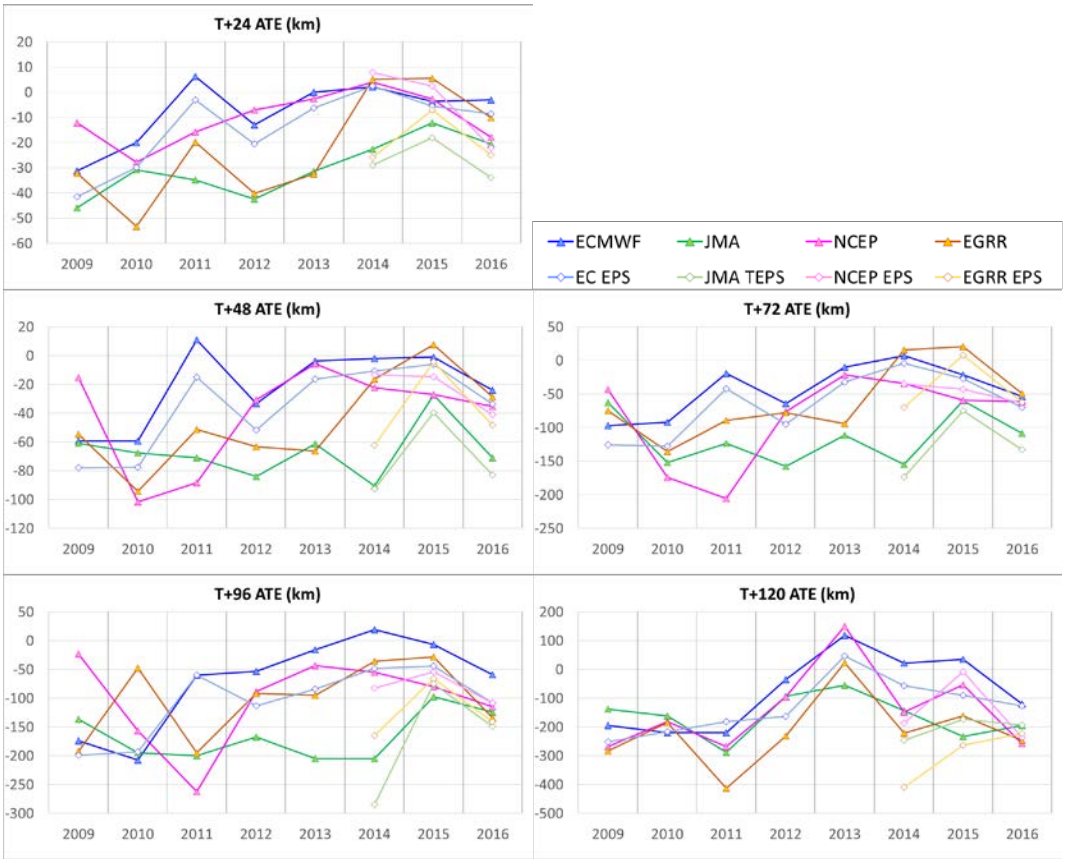


Figure 2 Along track error (ATE) at T+24, T+48, …, T+120 hours for various deterministic and EPS forecasts.

In order to quantify the result for comparison between models, the DPE of T+24, T+48, T+72, T+96 and T+120 were averaged and summarised into Table 1. The averaged value was sensitive to the number of available forecast tracks and the corresponding forecast lengths. As a result, comparison was only made among models within the same year. ECMWF and ECMWF EPS (or denoted as "EC EPS" hereinafter) showed relatively small error in comparison with other models. Moreover, error of JMA deterministic models was originally comparable to the other models in the earlier years. With relatively more improvement of the other models, the error of JMA was comparably large in recent years, while the performance for NCEP and EGRR was between that of ECMWF and JMA models.

Average DPE (km)	ECMWF	JMA	NCEP	EGRR	EC EPS	JMA TEPS	NCEP EPS	EGRR EPS	No. of data
2009	184	251	215	216	190	N/A	N/A	N/A	350
2010	144	167	192	191	142	N/A	N/A	N/A	231
2011	145	191	183	202	146	N/A	N/A	N/A	300
2012	128	174	128	172	128	N/A	N/A	N/A	385
2013	127	174	131	149	125	N/A	N/A	N/A	402
2014	115	161	103	147	107	175	113	162	213
2015	96	132	112	120	97	155	123	162	226
2016	116	150	140	138	115	159	140	142	325

Table 1 Average DPE for various deterministic and EPSs from 2009 to 2016.

Considering the average DPE, EC EPS showed rather good performance out of all available EPSs. In HKO's operational forecast, 4 deterministic models, namely ECMWF, JMA, NCEP and EGRR were used to compose the ensemble track for reference. In addition to the four deterministic models, ensemble mean of EC EPS was also included in the following discussion. Another common data set was used and the resulting average DPE was tabulated in Table 2.

Average DPE (km)	ECMWF	JMA	NCEP	EGRR	EC EPS	No. of data
2009	184	251	215	216	190	350
2010	144	167	192	191	142	231
2011	145	191	183	202	146	300
2012	128	174	128	172	128	385
2013	127	174	131	149	125	402
2014	118	162	116	137	110	383
2015	114	153	131	134	121	421
2016	121	164	149	142	117	376

Table 2 Average DPE of T+24, T+48, T+72, T+96 and T+120 for various deterministic models and EC EPS from 2009 to 2016.

In view of the different performance of various models, a weighted ensemble approach was studied. A simplified formula for the TC position forecast with weighting could be represented by the sum of the position vector *T* with a variable α as the weighting factor.

$$T_{ENS} = \sum_{\mathrm{model}\,i}^{N} \frac{\alpha_i T_i}{N}$$

A first guess for the weighting factor for model *i* to minimize the error of the grand ensemble track, based on *N* number of models including the ensemble mean of EC EPS, would be inversely proportional to the square of the corresponding expected error of model *i*.

$$\alpha_i = \frac{N \times \dfrac{1}{\left\langle E_i^2 \right\rangle}}{\displaystyle\sum_{\text{model } i}^{N} \frac{1}{\left\langle E_i^2 \right\rangle}}$$

In the study, the average DPE for a single model in the previous 3 years was used as the expected DPE for the following year. A summary of the weighting factors by considering the ratio of square of DPE was tabulated in Table 3, with the sum of weighting factors of all models being equal to 1.

Year	ECMWF	JMA	NCEP	EGRR	EC EPS
2012	0.261	0.155	0.171	0.160	0.254
2013	0.259	0.155	0.186	0.141	0.259
2014	0.251	0.137	0.211	0.149	0.253
2015	0.235	0.126	0.233	0.157	0.249
2016	0.238	0.129	0.215	0.175	0.243
2017	0.243	0.133	0.195	0.178	0.251

Table 3 Weighting factors for different models from 2012 to 2016 based on performance in the previous 3 years.

3. Analysis and result

3.1 Weighted ensemble

Depending on the evolution of the weather systems forecast by various models, there were different lengths of forecast period of the resultant model TC tracks. Simply averaging the forecast positions would generate discontinuity between consecutive forecast hours. Instead of positions, the mean of motion vectors was taken to compose the ensemble track [3].

The following three multi-model ensemble tracks using data from 2012 to 2016 were constructed:

(a) A 4-model ensemble track (denoted as “4 model ENS” in the figures or tables below) using equal weighting of ECMWF, JMA, NCEP and EGRR models;
(b) A 5-model ensemble track (5 model ENS) using equal weighting of ECMWF, JMA, NCEP, EGRR and EC EPS; and
(c) A 5-model weighted ensemble track (5 model weighted ENS) using the proposed weighting factors in Table 3.

Common data set was used to verify against the HKO best track and the results were shown in Figure 3.

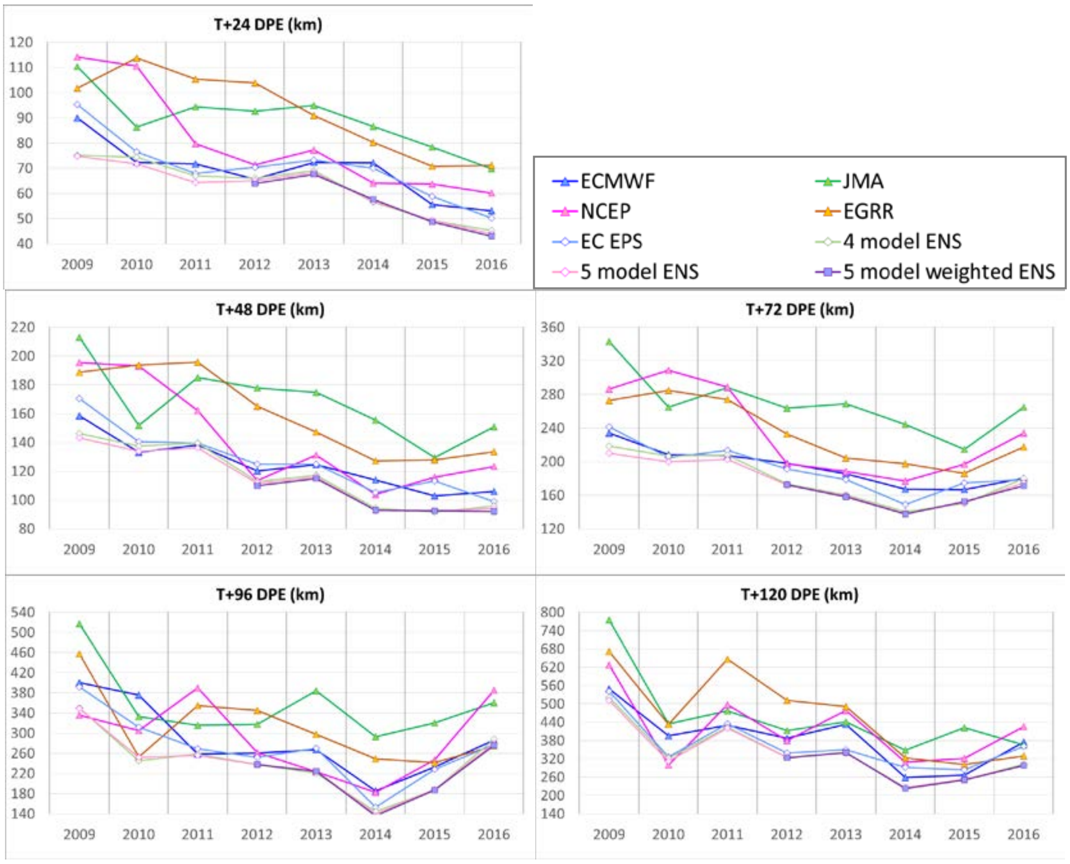


Figure 3 DPE for global models and the three multi-model ensemble tracks.

There was improvement of models in TC track forecast in recent years and the performance of ECMWF was comparably better among different models [4]. However, results from 2009 to 2016 indicated that multi-model ensemble track still performed better than single model track in general. Choosing the average DPE of T+24, T+48, T+72, T+96 and T+120 as the parameter to represent performance, the results were tabulated in Table 4. Similar findings could be obtained using other parameters that the performance of ensemble track was better than single model track.

Even though EC EPS mean was calculated using similar approach of averaging at most 51 ensemble track members, the average DPE was just comparable to that of the deterministic model, and even higher than that of the 4-model ensemble track. The value gained in multi-model ensemble track by taking into account behaviour or model physics of various NWP models still existed up to 2016.

Average DPE (km)	ECMWF	JMA	NCEP	UKMO	EC EPS	4 models ENS	5 models ENS	Weighted ENS	No. of data
2009	184	251	215	216	190	167	164	N/A	350
2010	144	167	192	191	142	137	134	N/A	231
2011	145	191	183	202	146	143	140	N/A	300
2012	128	174	128	172	128	118	117	116	385
2013	127	174	131	149	125	115	114	113	402
2014	118	162	116	137	110	97	96	96	383
2015	114	153	131	134	121	101	102	102	421
2016	121	164	149	142	117	113	110	108	376

Table 4 Average DPE for various models, EPS and three multi-model ensemble tracks from 2009 to 2016.

For the 5-model ensemble track with equal weighting, which included EC EPS mean track as a component, the average DPE was persistently lower than that using 4-model ensemble track. For the 5-model weighted ensemble, as it accounted for the performance in the past 3 years, only results from 2012 to 2016 could be analysed. The DPE was close to but still showed a slightly lower value than that of the 5-model ensemble track (equal weighting).

The availability of model runs of various NWP models in operational forecasting should also be considered. The configuration of the individual model track was summarised in Table 5. For example, during the time of issuing 12Z subjective forecast track, only forecast model tracks of 00Z run were available. The forecast position of ensemble track at T+24 would refer to the forecast position of model at T+36. At the time when issuing the 18Z subjective forecast track, latest 12Z deterministic model tracks would become available while only 00Z tracks were available from EPS models. A resultant combination of positions at T+30 from deterministic models and position at T+42 from EPS model would be used as a reference for the position at T+24 of operational forecast track.

Track	00Z	03Z	06Z	09Z	12Z	15Z	18Z	21Z
ECMWF	(D-1) 12Z	(D-1) 12Z	00Z	00Z	00Z	00Z	12Z	12Z
JMA	(D-1) 12Z	(D-1) 12Z	00Z	00Z	00Z	00Z	12Z	12Z
NCEP	(D-1) 12Z	(D-1) 12Z	00Z	00Z	00Z	00Z	12Z	12Z
UKMO	(D-1) 12Z	(D-1) 12Z	00Z	00Z	00Z	00Z	12Z	12Z
EPS	(D-1) 12Z	(D-1) 12Z	(D-1) 12Z	(D-1) 12Z	00Z	00Z	00Z	00Z

Table 5 The configuration of model runs for various models in constructing the ensemble track. (D-1) represents model run in the previous day.

If the position at a specific forecast hour of an individual model was unavailable, the forecast position would be interpolated using Bezier curve fitting in order to maintain the smoothness of the track as well as the steady change in momentum in short time. If a particular model indicated dissipation of TC at a specific forecast hour, ensemble track at that forecast hour would comprise other available models with weighting factors adjusted on a pro-rata basis.

The forecast tracks including HKO warning track, CMA forecast track, simulated 4-model ensemble track and simulated 5-model weighted ensemble track from 2012 to 2016 were verified against HKO best track using common data set. Results aggregated by forecast hours were tabulated in Table 6. There were fluctuations in performance of the model ensemble tracks and tracks issued by HKO and CMA, but in general the 5-model weighted ensemble track showed a slightly better performance than the 4-model ensemble track. It hinted that there were possibilities of value gain if the proposed 5-model weighted ensemble track was used as reference.

(a)

2012 DPE	HKO	CMA	4 model ENS	weighted ENS	No. of Sample
T+24	96	97	95	91	190
T+48	152	165	150	145	150
T+72	226	245	228	224	117

(b)

2013 DPE	HKO	CMA	4 model ENS	weighted ENS	No. of Sample
T+24	91	93	93	93	250
T+48	146	153	141	140	165
T+72	202	216	197	195	105

(c)

2014 DPE	HKO	CMA	4 model ENS	weighted ENS	No. of Sample
T+24	83	79	76	77	205
T+48	127	137	113	113	150
T+72	167	182	152	151	113
T+96	206	241	188	190	41
T+120	245	330	219	224	22

(d)

2015 DPE	HKO	CMA	4 model ENS	weighted ENS	No. of Sample
T+24	69	71	67	68	250
T+48	122	128	118	120	187
T+72	169	189	166	169	133
T+96	216	225	214	219	85
T+120	367	378	323	327	52

(e)

2016 DPE	HKO	CMA	4 model ENS	weighted ENS	No. of Sample
T+24	62	65	58	56	214
T+48	119	124	115	110	159
T+72	205	208	213	202	104
T+96	286	310	302	294	62
T+120	354	432	346	353	35

(f)

Average DPE	HKO	CMA	4 model ENS	weighted ENS	No. of Sample
2012	148	157	147	143	457
2013	131	137	129	128	520
2014	129	140	117	118	531
2015	141	149	136	138	707
2016	146	156	146	141	574

Table 6 DPE of HKO warning tracks, forecast tracks issued by CMA, 4-model ensemble track and 5-model weighted ensemble track. Results (a) to (e) correspond to 2012 to 2016 respectively and result (f) shows the averaged DPE of each year.

3.2 Selective ensemble approach

Another approach of utilizing ensemble forecast would be analysing each individual member of the EPS, which represents a possible single scenario. Taking the time delay in receiving EPS model track into account, a 00Z EPS model track would be available before issuing a 12Z subjective forecast track. This means that the position fix at 12Z could act as a reference for selectively screening some EPS track members.

In this study, EPS track members of 00Z and 12Z model runs from ECMWF EPS, NCEP EPS and EGRR EPS were used. There were at most 51 track members from ECMWF EPS, 21 track members from NCEP EPS and 36 members from EGRR EPS. Each track was considered as a possible scenario and track members at a total number of at most 108 were combined as a grand EPS cluster.

The ensemble mean of the grand EPS track members could be considered as the "most probable" scenario ("EPS mean"). However, there may be outliers, which could be filtered out by comparing the forecast position at T+12 with the analysed position for each EPS member. Around 50% of the EPS tracks with larger DPE for the T+12 forecast were excluded from the "most probable" scenario to form a selective EPS track member cluster ("selective EPS"). The EPS mean and the selective EPS tracks in 2016 were constructed and discussed in the following.

DPE of forecast tracks issued by HKO and CMA, simulated 4-model ensemble tracks, 5-model weighted ensemble tracks, EPS mean tracks and the selective EPS tracks at different forecast hours in 2016 was plotted in Figure 4. A slightly lower DPE for selective EPS track at T+24 could be seen as compared with EPS mean track. As expected, an increasing trend of DPE could be seen with longer forecast hours for both tracks. However, the performance of the two tracks was found to be reversed, in the sense that DPE of selective EPS tracks was lower than that of EPS mean track for short forecast hours but reversed for longer forecast hours. This means that there was improvement in short forecast hours by neglecting the outliner tracks but the effect diminished or even become not useful for longer forecast hours.

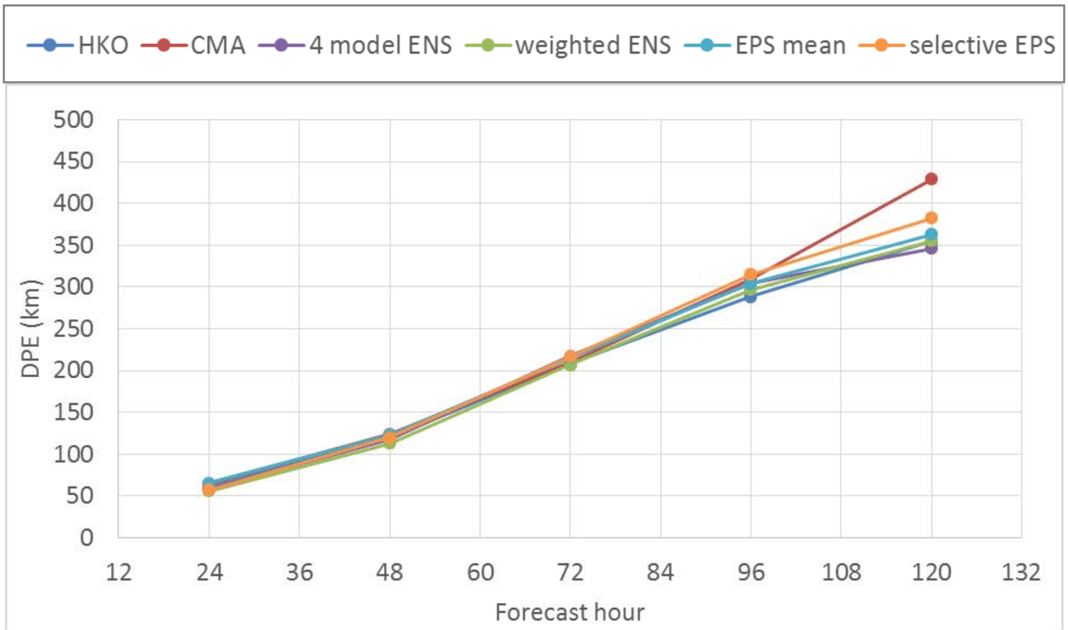


Figure 4 DPE at different forecast hours for various forecast tracks issued by centre and ensemble tracks in 2016.

Further on the performance of various forecast tracks issued by centre and ensemble tracks against best track, the DPE of HKO warning track was taken as the baseline while the DPE of other tracks were subtracted by the baseline for easier comparison. Figure 5 shows the relative DPE with negative value indicating a better performance than the HKO warning track.

Although a slightly better performance could be seen for selective EPS track at T+24 forecast hour, it was just comparable to the 5-model weighted ensemble track. For the forecast hours beyond T+48, 5-model weighted ensemble track showed improvement but not for the selective EPS track. The 5-model weighted ensemble track also showed better performance than selective EPS track in general for different forecast hours.

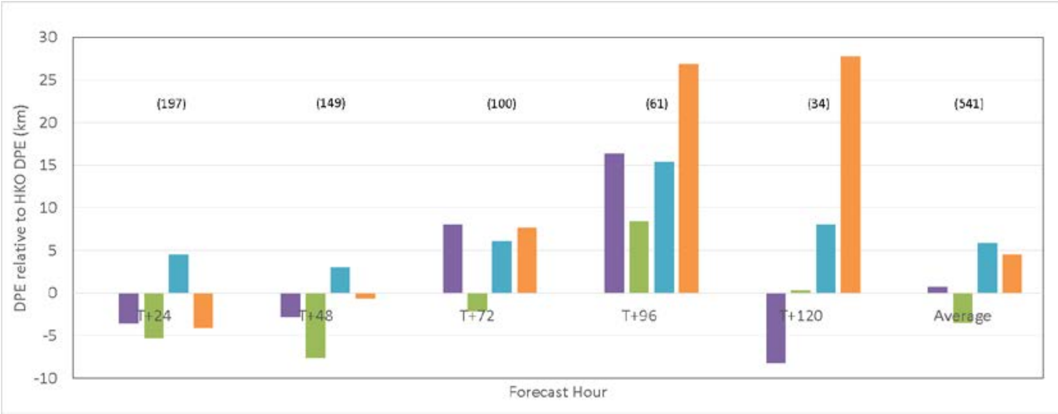


Figure 5 DPE for ensemble tracks in 2016. The values were relative to HKO warning track and grouped by forecast hours (x-axis). The right-most column represented the average value for all forecast hours. Value in blanket represents the number of samples.

The average DPE for different TC cases were shown in Figure 6. There were some cases in which the selective EPS showed better performance than the EPS mean track, indicating the usefulness of neglecting outlier EPS tracks in some cases.

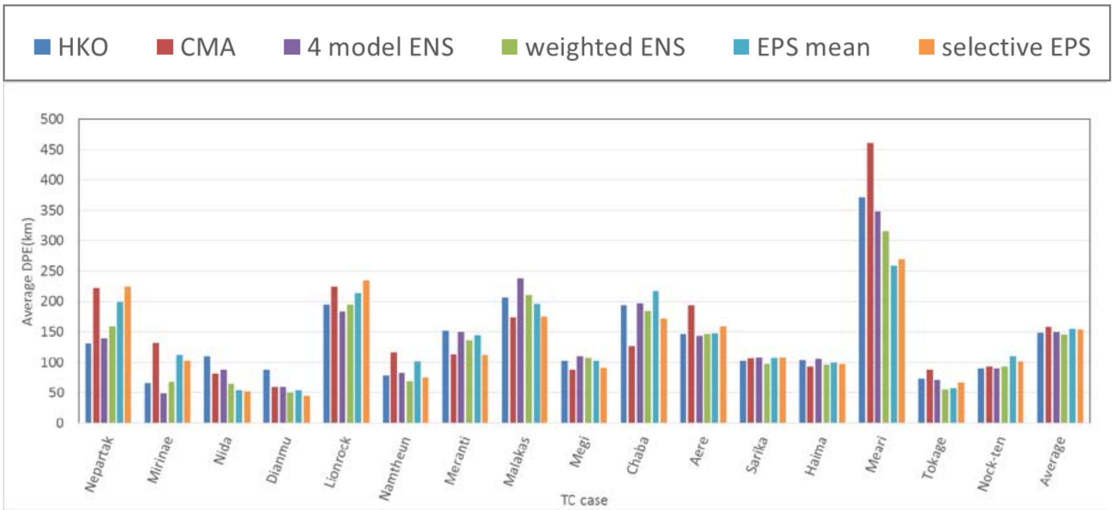


Figure 6 Average DPE for various forecast tracks issued by centre and ensemble tracks in 2016. The average DPE was grouped in different TC cases. The right-most column represented the average value for all TCs.

3.3 Case study

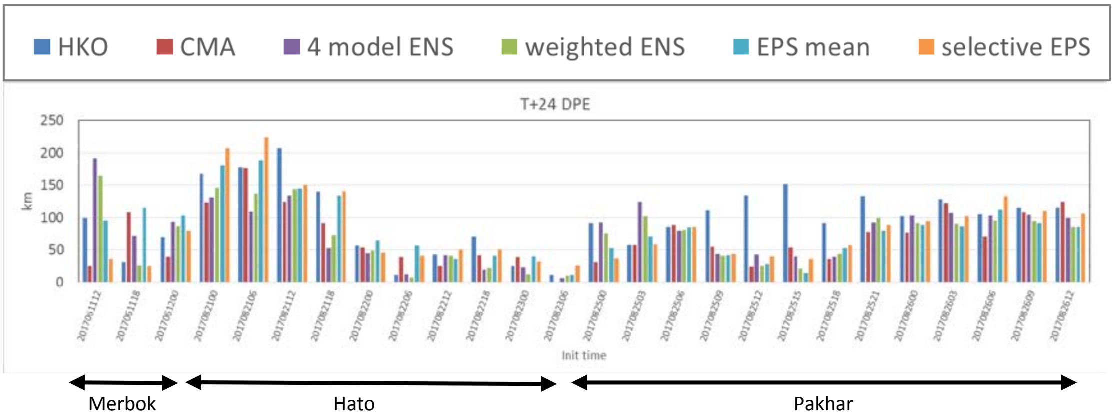
For more detailed comparison of model ensemble tracks and selective EPS approach, cases in 2017 were studied. The DPE at T+24 and T+48 of tracks issued by HKO and CMA as well as ensemble tracks for the cases of Merbok, Hato, and Pakhar in 2017 was plotted in Figure 7. Cases of Merbok showed better result for selective approach with much lower DPE for selective EPS track than EPS mean track. Figure 8 (a) showed the TC tracks at 11 June 12Z based on model runs of 11 June 00Z. Models in general predicted an easterly bias while the position fix at 11 June 12Z was at the west boundary of the EPS member tracks cluster. Selective EPS track accounted for the bias and showed better performance than the other ensemble tracks.

For the case of Hato, selective EPS tracks showed comparatively large error in the earlier initial time. Example could be illustrated for the simulated tracks of 21 August 00Z in Figure 8 (b) based on 20 August 12Z model runs. As Hato showed temporarily northwards motion from 20 August 12Z to 21 August 00Z, the selection of EPS track members biased more for those tracks at higher latitude when compared with the grand EPS mean track. The selective EPS track therefore showed larger departure from the actual track which beared more southerly component on 21 August.

After Hato crossed the Luzon Strait and at the issuance of 22 August 00Z track, as shown in Figure 8 (c), models based on 21 August 12Z were in better consensus. Hato moved in a westerly direction, locating more to the south than most other EPS members. EPS tracks farther north were thus neglected in the selective EPS approach. Selection of 50% EPS track members did rectify the bias of model tracks before landing. However, the accounted improvement was just similar to the 5-model weighted ensemble track in this case.

There was no noticeable improvement with the selective EPS approach for the case of Pakhar as the model tracks and EPS track members were in good consensus. In fact, the usefulness of selective ENS approach was case dependent and the performance was also sensitive to the position fix in operation.

(a)



(b)

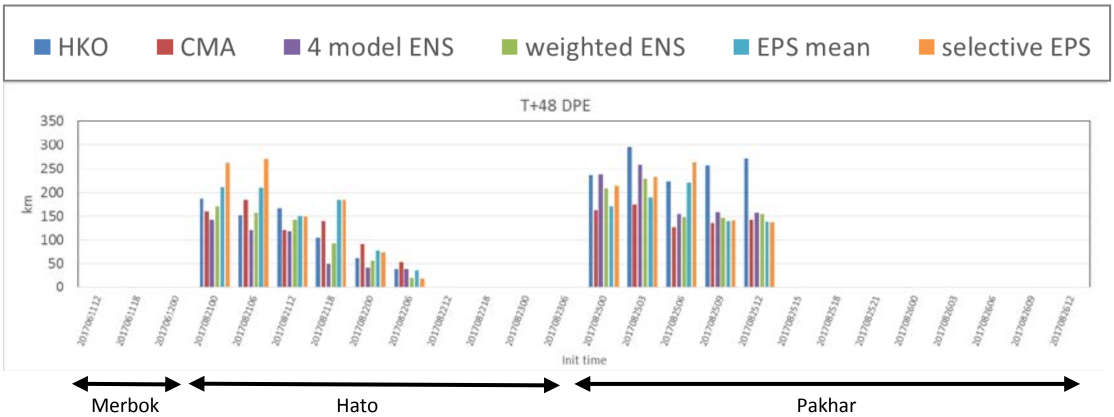
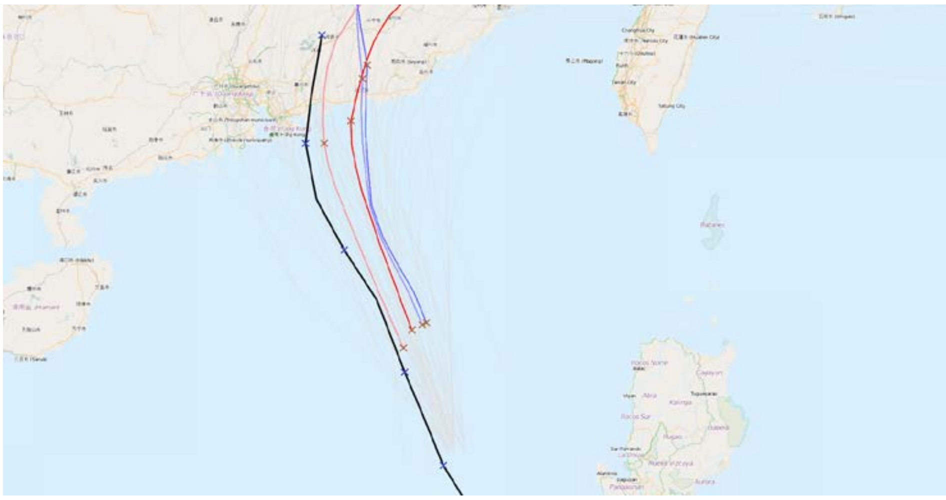
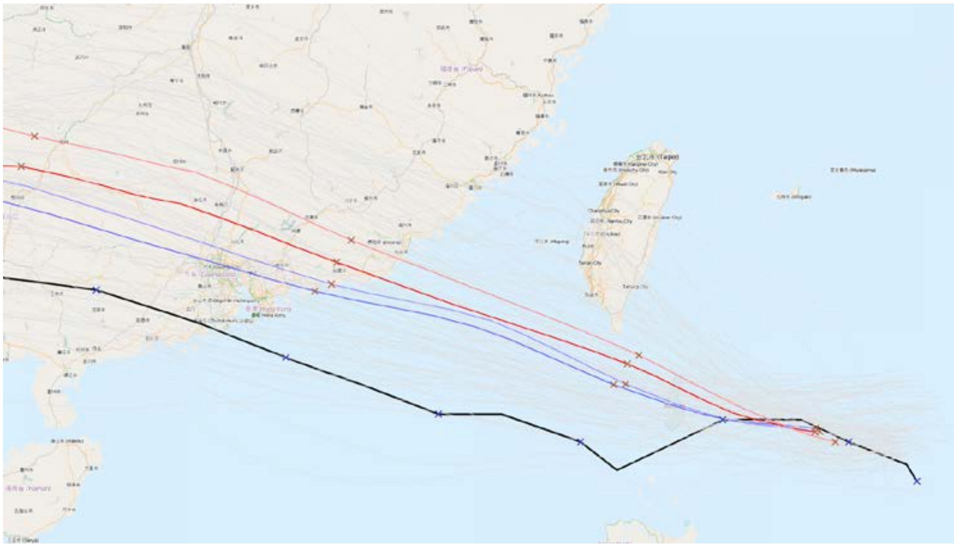


Figure 7 DPE for various forecast tracks issued by centre and ensemble tracks in 2017 at (a) T+24 and (b) T+48. All available forecast hours were listed in each column.

(a) 

(b) 

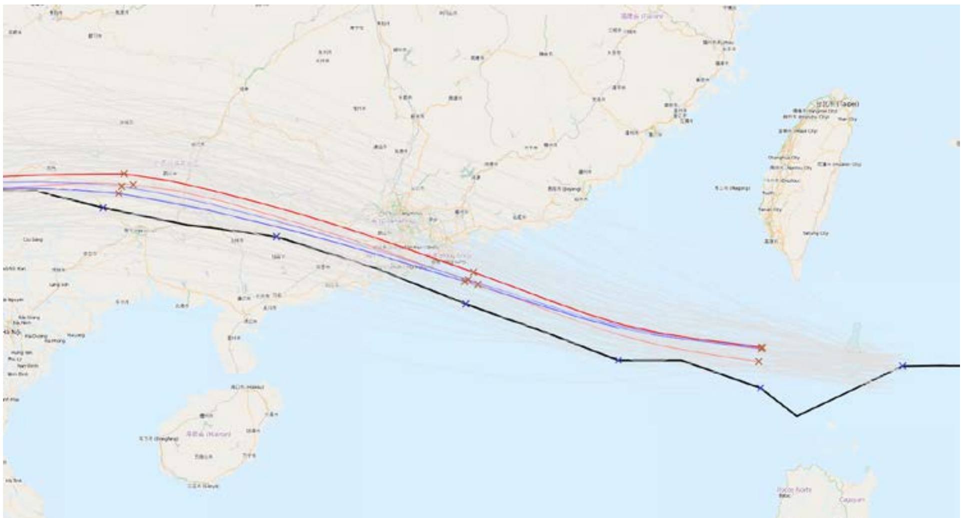
(c) 

Figure 8 Simulated TC track at (a) 11 June 12Z based on 11 June 00Z models for Merbok, (b) 21 August 00Z based on 20 August 12Z models for Hato, and (c) 22 August 00Z based on 21 August 12Z models for Hato. Brown crosses represent the forecast positions in 24-hourly intervals. Grand EPS mean without selection (red), after selection (pink), 4-model ensemble (plain blue) and 5-model weighted ensemble (blue) tracks were shown. All members of the Grand EPS tracks were plotted in light coloured thin lines. Black line indicates HKO best track with blue crosses representing the 00Z and 12Z positions.

4. Conclusion

This study again showed that multi-model ensemble track performed better than single model track. ECMWF model track had a leading performance among various global models. The choice of including EC EPS as a member in constructing multi-model ensemble track gained value when compared with the 4-model (ECMWF, JMA, NCEP and EGRR) ensemble track. Taking past performance into account, 5-model weighted ensemble track has slightly better performance according to the verification from 2012 to 2016 and provided a way to further improve the operational warning tracks.

The selective ensemble approach of EPS track members was found to be useful in some specific cases in 2016 and 2017. Improvement was more apparent in short-term forecast such as T+24 than longer forecast hours. However, the usefulness was situation dependent as demonstrated in the cases of TC Merbok, Hato and Pakhar in 2017.

Acknowledgement

The author would like to thank Messrs L.S. Lee and W.K. Wong for their guidance of this study.

Reference

[1] Lam, H. and H.M. Ma, 2002: Performance of NWP Model Ensemble for Tropical Cyclone Track Prediction in 2001. 16th Guangdong-Hong Kong-Macao Seminar on Meteorological Science and Technology, Guangzhou, 30-31 January 2002.

[2] Tam, Y.T., W.K. Wong and M.Y. Chan, 2015: Error Characteristics of Numerical Weather Prediction Model Ensemble in Tropical Cyclone Track Prediction, 29th Guangdong-Hong Kong-Macao Seminar on Meteorological Science and Technology, Macao, 20-22 January 2015.

[3] Woo W.C. and H. Lam, 2007: Application of the Multi-Model Ensemble technique on Forecast of Tropical Cyclone Tracks, 21st Guangdong-Hong Kong-Macao Seminar on

Meteorological Science and Technology, Hong Kong, 24-26 January 2007.

[4] Chan, S.T. and J.C.W. Lee, 2012: Recent Operational Performance of Multi-model Ensemble on Tropical Cyclone Track Forecasting at the Hong Kong Observatory, The 16th National Workshop on Tropical Cyclones (NWTC-XVI) cum the 7th International Workshop on Tropical Cyclones (IWTC2012) Zhoushan, China.