ESCAP/WMO Typhoon Committee
Forty-ninth Session
21- 24 February 2017
Yokohama, Japan

FOR PARTICIPANTS ONLY
14 February 2017
ENGLISH ONLY

Performance of tropical cyclone forecast in western North Pacific in 2016

Xiaotu LEI, Guomin CHEN, Xiping ZHANG, Peiyan CHEN, Hui YU, Rijin WAN
(Submitted by *Shanghai Typhoon Institute/CMA*)

TABLE OF CONTENTS

1. Introduction

An important key to making better predictions of TC is having an understanding of the forecast errors in current predictions. Subjective and objective verification of TC forecasts give evidence regarding the accuracy and performance characteristics of TC forecasts and warnings. Verification analyses diagnose and quantify the systematic and random errors so that improvements can be made to operational forecasting methodologies and to the underpinning numerical models.

This report is primarily about *performance of typhoon forecast over western North Pacific in 2016*. At the conclusion of the season, forecasts are evaluated by comparing the projected positions and intensities to the corresponding post-storm derived "best track" positions and intensities for each cyclone. A forecast is included in the verification only if the system is classified in the final best track as a tropical cyclone at both the forecast's initial time and at the projection's valid time. In this report, we start with a short discussion of best track datasets, which are the first requirement for verifying TC forecasts. The next section describes deterministic forecast methods, which will be evaluated here including official guidances, global models and regional models. We'll also discuss the deviation of operational real time positioning results for official guidances. Last and most important, we will evaluate the cyclone track, intensity forecast, which will include deterministic and ensemble predictions.

2. Best track datasets

Currently, four agencies provide their own TC best track analyses for the WNP region: 1) Shanghai Typhoon Institute of China Meteorological Administration, 2) the Japan Meteorological Agency (JMA) Regional Specialized Meteorological Center (RSMC) in Tokyo, 3) Joint Typhoon Warning Center, 4) Hong Kong Observatory. Table 1 provide the data period, characteristics and wind averaging time information of these four best track datasets. It should be noted that the TC position, intensity and structural information usually differ among those agencies due to the lack of sufficient surface observations for TCs, as well as the different techniques used to estimate the position and intensity of a TC. Thus, differences in TC forecast performance may be obtained, depending on the best-track dataset used as a reference. As the typhoon center in RSMC-Tokyo is the regional center that carries out specialized activities in analysis and forecasting of WNP TCs within the framework of the World Weather Watch (WWW) Program of WMO, in this verification report, we used RSMC-Tokyo best track-dataset as the reference.

Table 1. Descriptions of western North Pacific best-track datasets.

Agency	Period	Characteristics	Wind
RSMC Tokyo	1951 to present	Includes extratropical cyclone stage, longitude, latitude, MCP and TS markers since 1951; MSW and typical severe wind radii since 1977 (without TD cases).	10 min
CMA	1949 to present	Includes sub-centers, some double eyewall cases/coastal severe wind of landfalling TCs (until 2004); includes TD cases; extratropical cyclone stage; longitude, latitude, MSW and MCP since 1949.	2 min
HKO	1961 to present	Includes TD cases; longitude, latitude, MSW and MCP since 1961 (extratropical cyclone stages are not marked).	10 min
JTWC	1945 to present	Includes TD cases; extratropical cyclone stage since 2000; longitude, latitude, and MSW since 1945; MCP and TC size	1 min

parameters since 2001.

3. TC position and intensity forecast data

In this report, TC position and intensity forecast results from five official guidances, five global models and three regional models are evaluated. These totally 13 methods are deterministic forecast guidance, detail explanations including their abbreviations, short description and source agencies are listed in Table 2. Additional verification on position and intensity of ensemble prediction system will also be show in this report. Parameter details including model resolution, data resolution, ensemble members, perturbation method, forecast time, output interval and forecast hours of each EPS are given in Table 3.

Table 2. Details of deterministic forecast guidance

Category	Abbreviation	Full name or short description	Source agency
Official guidance (5)	**CMA**	*China Meteorological Administration*	CMA
	JMA	*Japan Meteorological Agency*	JMA
	JTWC	*Joint Typhoon Warning Center*	JTWC
	KMA	*Korea Meteorological Administration*	KMA
	HKO	*Hong Kong Observatory*	HKO
Global NWP model (5)	**CMA-T639**	*Global spectral model of CMA at a resolution of T639L60*	CMA
	ECMWF-IFS	*Integrated Forecasting System of ECMWF*	ECMWF
	JMA-GSM	*Global Spectral Model of JMA*	JMA
	NCEP-GFS	*Global Forecast System of NCEP*	NCEP
	UKMO-MetUM	*Unified Model system of UKMO*	UKMO
Regional NWP model (3)	**BoM-ACCESS-TC**	*Tropical cyclone model in the Australian Community Climate and Earth-System Simulator Numerical Weather Prediction systems*	BoM
	STI-GRAPES	*Regional TC-forecasting model based on the Global/Regional Assimilation and PrEdiction System (GRAPES)*	STI/CMA
	CMA-TRAMS	*Tropical Regional Atmosphere Model for the South China Sea based on GRAPES*	ITMM/CMA

Table 3. Details of ensemble forecasts guidances

	ECMWF-EPS	JMA-TEPS	JMA-WEPS	MSC-CENS	NCEP-GEFS	UKMO-EPS
Resolution	TL639 (0-10d) TL319 (10-15d)	TL319L60	TL319L60	0.9°	T126L28	
Data resolution	\	0.5625°	0.5625°	1°	1°	
Members	51	11	51	21	21	24
Perturbation method	Singular Vector	SVD	SVD	Ensemble Kalman	Ensemble Transform	
Forecast time	00:00 12:00	00:00 12:00	12:00	00:00 12:00	00:00 06:00 12:00 18:00	00:00 12:00
Output Interval (h)	12	6	6	6	6	12

Forecast hour	120	132	216	240	240	192

4. Deviation of operational TC initial position

Generally, due to the limitation of different technology, often there could be large variations in TC intensity estimates from difference operational agencies. Compare to the deviation of TC intensity estimates, the variations in TC position estimates are often overlooked. Figure 1 presents deviation of initial positions between two official agencies. Basically, the position deviations for most cases are under 1 degree. However, the fact remains that there is still a significant part of cases which have large deviations. These cases have focused mainly on the early stages of lifetime. An extreme example is the generate location of tropical storm Omais, which is the fifth TC in western North Pacific in 2016. In this case, there exits almost 5 degrees of deviation between JTWC and other three agencies (CMA, JMA and KMA, HKO has no record on that case).

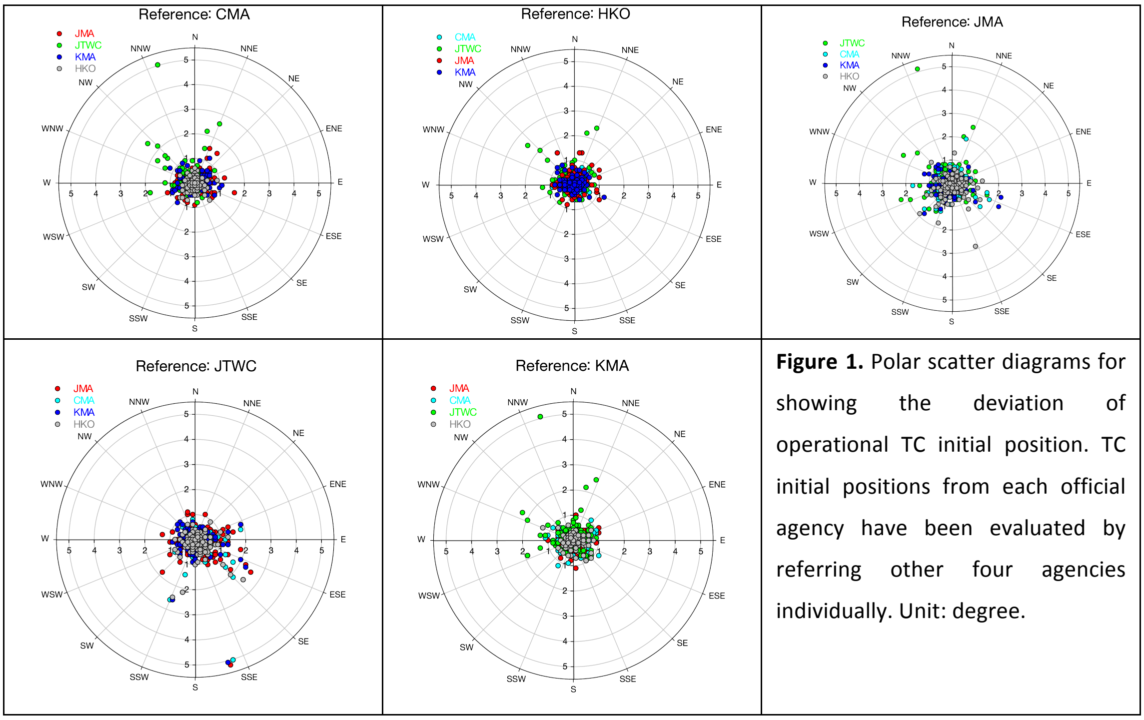


Figure 1. Polar scatter diagrams for showing the deviation of operational TC initial position. TC initial positions from each official agency have been evaluated by referring other four agencies individually. Unit: degree.

5. Performance of TC track forecast

5.1 Deterministic forecast

TC position error is defined as the great-circle difference between a TC's forecast center position and the best track position at the verification time. TC position errors typically are presented as mean errors for a large sample of TCs, as in Figure 2, which shows mean position errors for each official guidances, global models and regional models at the lead time levels of 24, 48, 72, 96 and 120h. The detail numerical values of position error which related to Figure 2 are list in Table 4. Figure 3 presents the radar area diagrams for comparing position error of five official guidances from 2013 to 2016 at lead time levels of 24, 48 and 72h. Encouragingly, for the last four years, official agencies' ability of TC position prediction has been steadily improving. More importantly, the differences in track forecast performance between agencies are falling. In 2016, the position errors for each official agency were under 85km, 150km and 250km at 24, 48 and 72h, respectively.

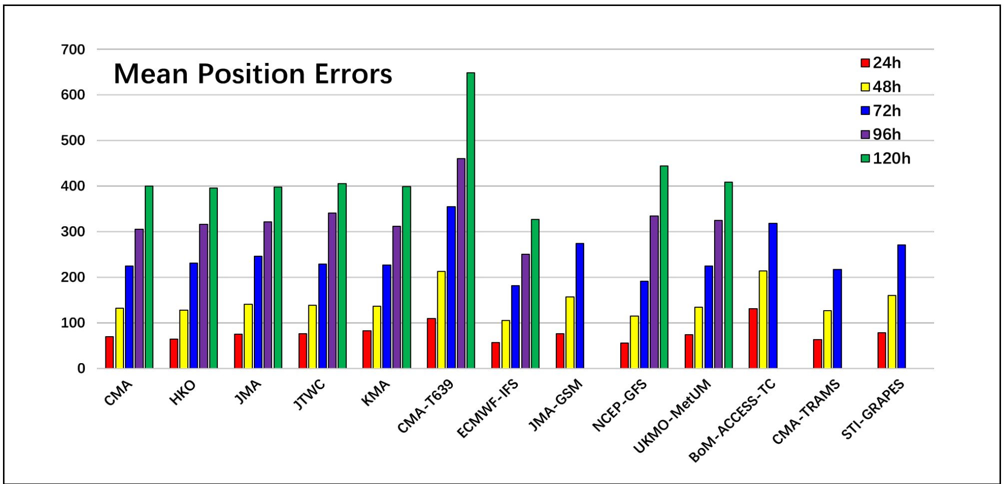


Figure 2. Mean position errors of official guidances, global models and regional models at the lead time levels of 24h(red), 48h(yellow), 72h(blue), 96h(purple) and 120h(olive) in 2016.

Table 4. Average position error for each method at lead time levels of 24, 48, 72, 96 and 120h in 2016. Numbers in bracket are sample sizes. (Unit: km)

Method	Lead times	24h	48h	72h	96h	120h
Official Guidance	CMA	69.2(363)	132.1(272)	224.5(196)	305.3(133)	399.3(85)
	HKO	64.4(214)	127.4(161)	230.8(112)	315.8(72)	395.0(44)
	JMA	74.3(365)	140.2(275)	246.2(200)	320.7(133)	397.1(85)
	JTWC	75.9(306)	137.9(221)	228.6(152)	340.5(84)	405.6(41)
	KMA	81.9(200)	135.9(144)	226.8(96)	311.9(56)	399.2(28)
Global NWP Model	CMA-T639	109.1(360)	212.4(270)	354.9(194)	460.0(128)	648.7(79)
	ECMWF-IFS	56.3(141)	104.7(110)	181.4(83)	249.7(56)	326.7(34)
	JMA-GSM	76.3(336)	156.2(254)	274.1(184)	/	/
	NCEP-GFS	55.5(320)	114.4(205)	191.0(164)	334.6(98)	444.3(65)
	UKMO-MetUM	73.9(169)	134.1(126)	224.3(90)	324.8(58)	407.9(36)
Regional NWP Model	BoM-ACCASS-TC	130.8(52)	213.7(36)	318.5(21)	/	/
	CMA-TRAMS	63.0(158)	126.8(117)	216.8(85)	/	/
	STI-GRAPES	77.6(266)	159.3(212)	271.1(152)	/	/

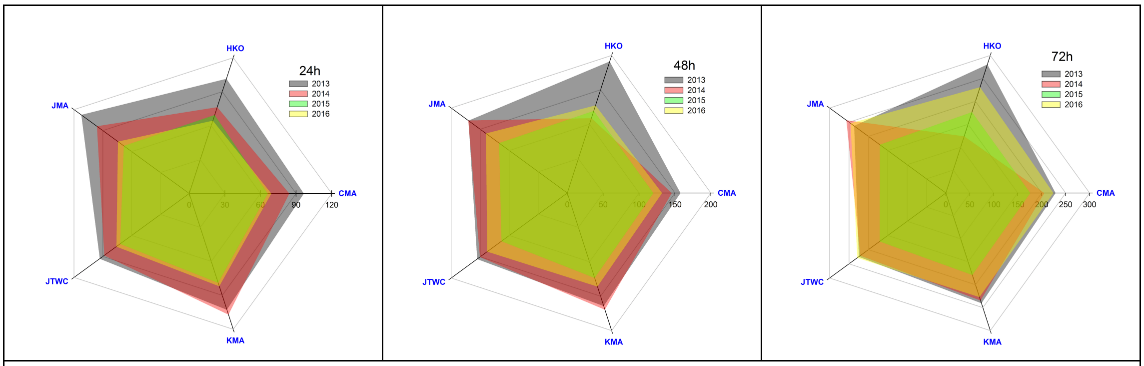


Figure 3. Radar area diagrams for comparing position error of five official guidances from 2013 to 2016.

Along and cross track biases of official guidances from 24 to 120h are given in figure 4. The figure shows that for the lead time which less than 48h, both along and cross track component do not exhibit obvious bias. With forecast lead time growing, forecasted TCs propagate on average too fast and rightward for JTWC and KMA, and propagate on average obviously rightward for HKO and JMA, however, propagate on average a little slow and leftward for CMA.

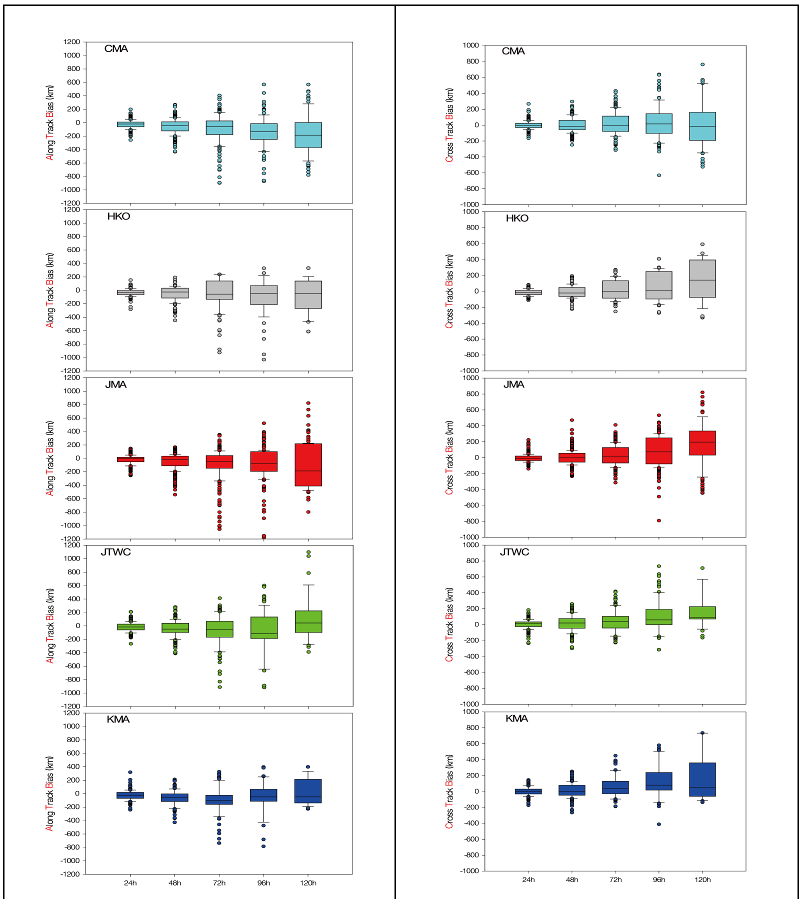
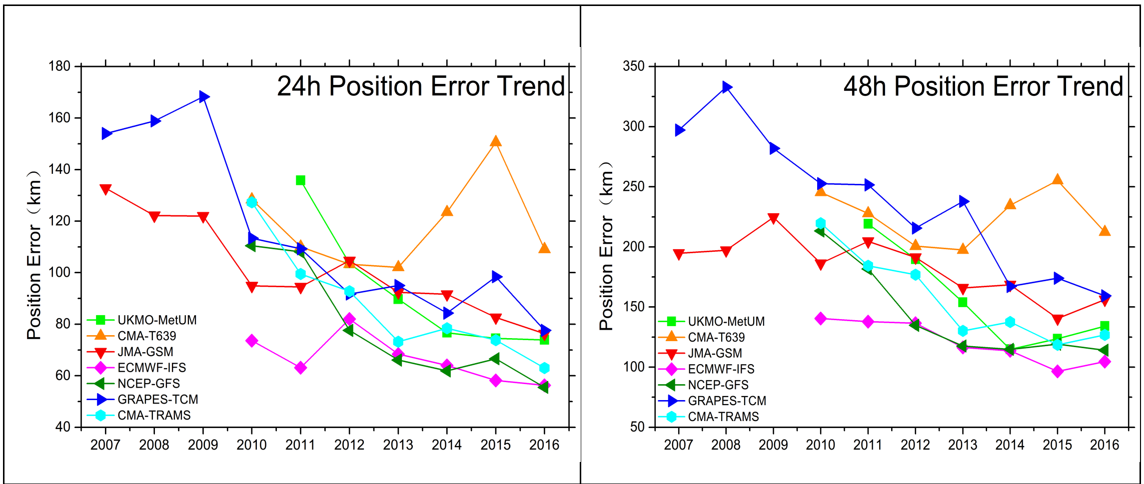


Figure 4. Along (left) and Cross (right) track biases for official guidances.

Compare to official guidances, performance of position prediction in numerical models is rather spotty. The position errors at the lead time of 24h are generally less than 80km for most models, but there still exist two models' mean position errors at the lead time level of 24h are larger than 100km. Impressively, the performance of ECMWF-IFS was stable for the last 6 years and its mean position error at the lead time level of 120h is only 326.7km, which is the smallest error among all the forecast guidances.



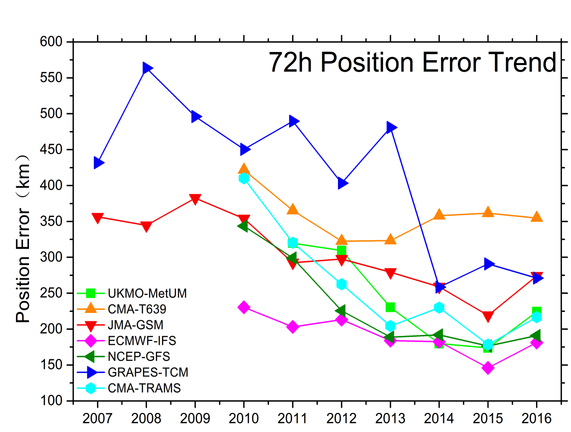
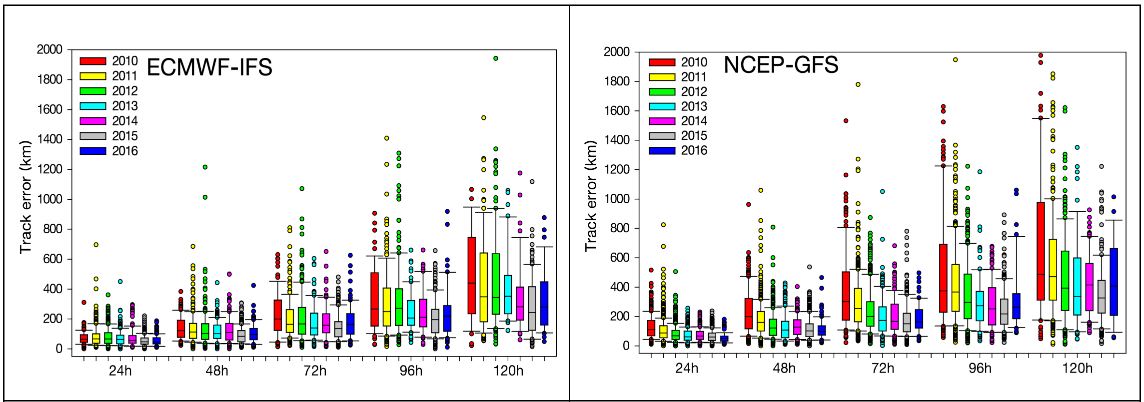
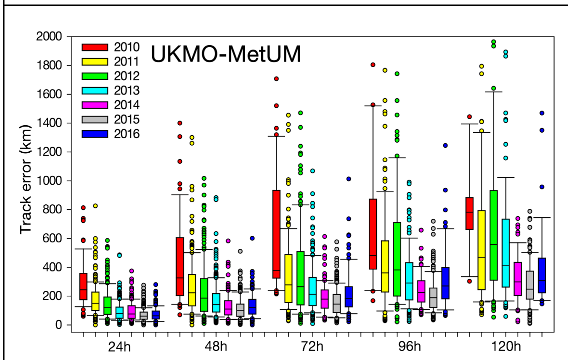


Figure 5. Position error trend for both global and regional models at the lead time levels of 24, 48 and 72h from 2007 to 2016.

An alternative approach to examining the average errors is to consider the distributions of errors, as in Figure 6. In this example, box plots are used to summarize the distributions of errors in track forecasts from 2010 to 2016 for three global models. Such a position error distributional approach not only shows the entire performance of each model's track forecast at each lead time, but also provides a straightforward method of understanding the annual progress of each global model. This methodology is developed to evaluate the uncertainty in verification measures through confidence intervals and paired statistical tests. And it can provide a consistent set of results that allowed the forecasts from the various models to be compared and fairly evaluated. In Figure 6, it clearly shows that stepped decreases in the values of each quantile were made at every lead time level from 2010 to 2015, and the forecast accuracy at 48h (72, 96 and 120h) in 2015 were almost close to or beyond the forecast accuracy at 24h (48, 72 and 96h) in 2010. However, such a progress has been stagnated or even regress in 2016, especially for long lead time levels. Anyway, it should be noted that this is not necessarily a conclusive comparison because the TCs in each year was not the same.





Box plots show specific quantiles and other statistics used to represent a distribution. The bar in the middle of the plot represents the median value, the lower and upper ends of the boxes represent the 25th and 75th quantile values. The bars below and above the box represent the non-outlier extreme values, and the specific circles represent the outliers.

Figure 6. Box plots for representing the distributions of position errors for TC track forecast from 2010 to 2016.

To assess the track forecast skill, the track forecast error can be compared with the error from a climatology and persistence model that contain no information about the current state of the atmosphere. Figure 7 shows the track forecast skill score at the lead time levels of 24 and 48h for subjective method, global and regional models from 2010 to 2016. All the forecast methods obtained positive skill indicating the forecast accuracy of official guidances, global and regional models are better than climatic persistence method in the last 7 years.

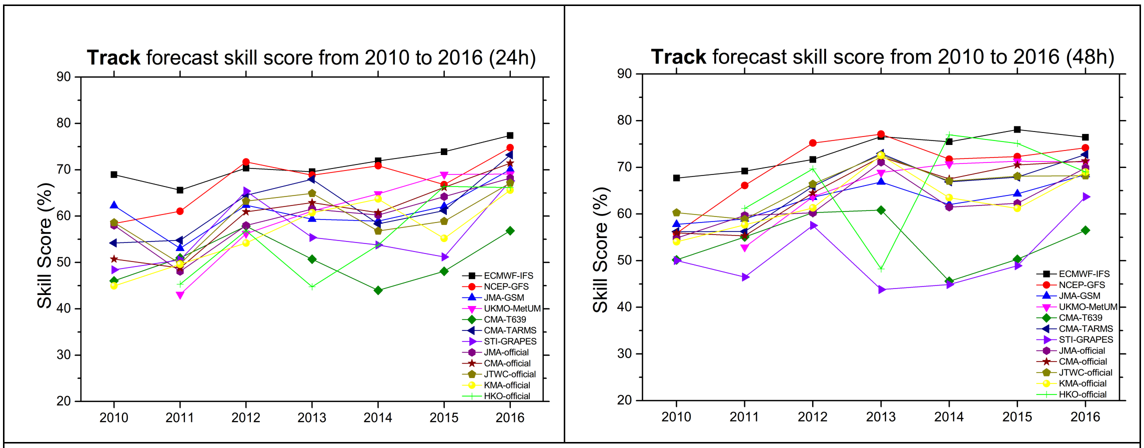


Figure 7. Track forecast skill score trend at the lead time level of 24h (left) and 48h (right) for official guidances, global models and regional models.

Figure 8 presents the polar scatter plots which depicting the mean combined direction and magnitude errors around the actual storm location for global and regional models at different lead time levels in 2016. Each models' systematic biases of track forecast are showed clearly through the Figure 8. The numbers with different colors denote annual mean locations relative to actual typhoon locations which obtain from best track dataset. Most global models track forecast systematic biases are small through 48h, but increased beyond that time and are generally westward at 96 and 120h. However, the systematic bias of

NCEP-GFS is southeastward at 96 and 120h. Three regional models show different systematic characteristics. The bias of CMA-TRAMS is negligible from 24 to 72h. By contrast, biases of BoM-ACCESS-TC and STI-GRAPES are southward and northwestward respectively. Plots like those in Figure 8 provide information that is useful for pre-estimate the bias of a certain method.

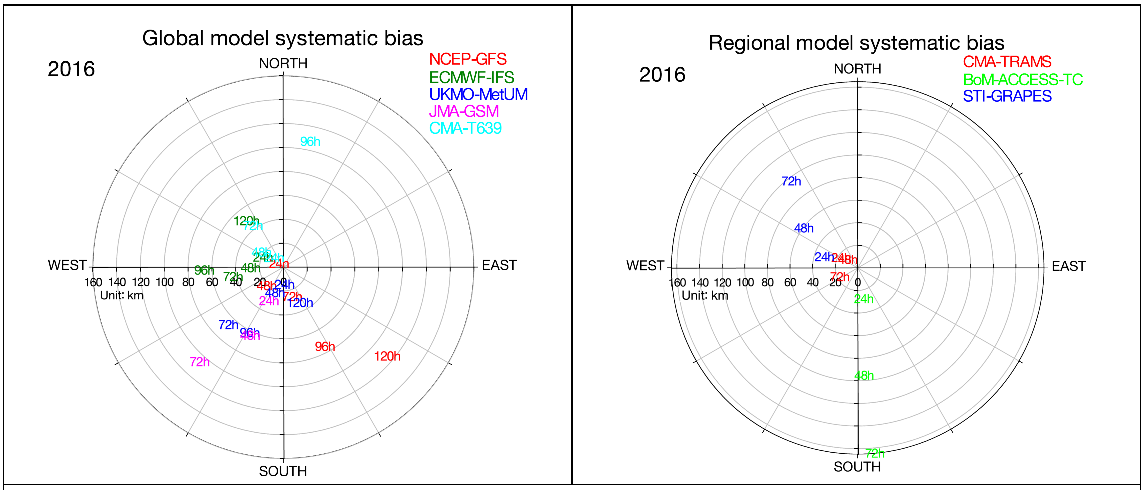
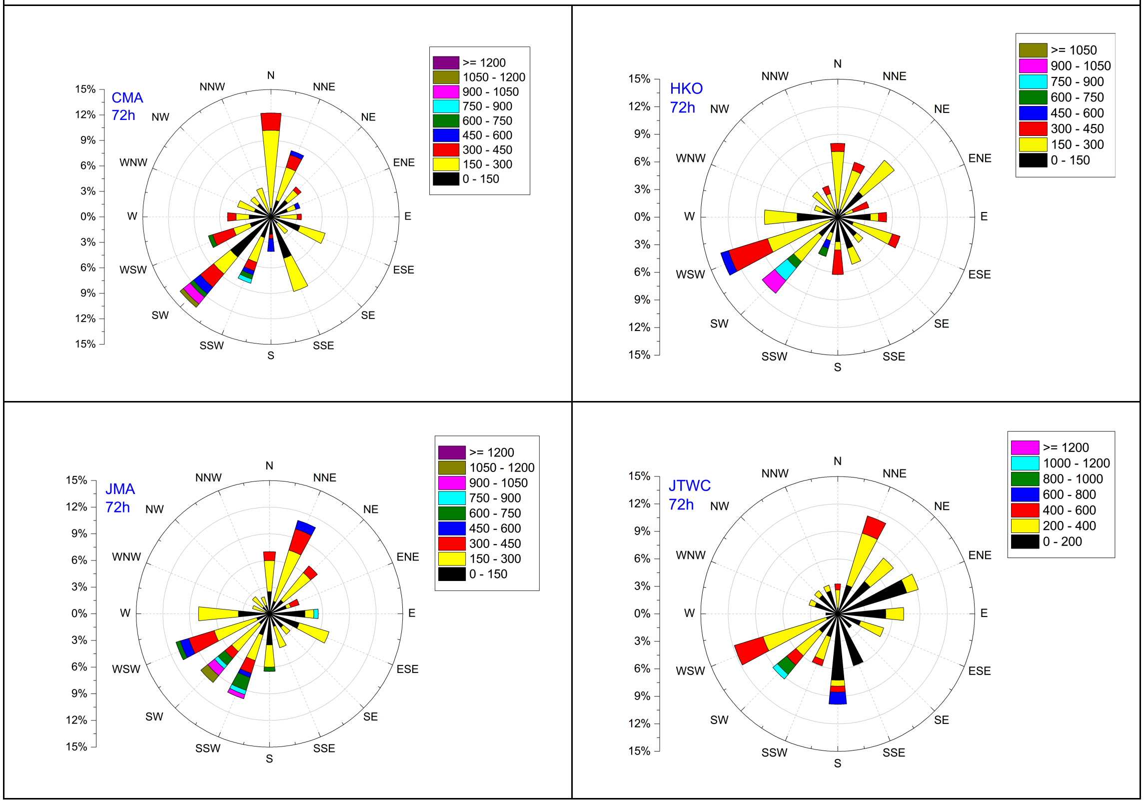


Figure 8. Polar scatter plots depicting the mean combined direction and magnitude errors around the actual storm location for each method at different lead time levels in 2016.



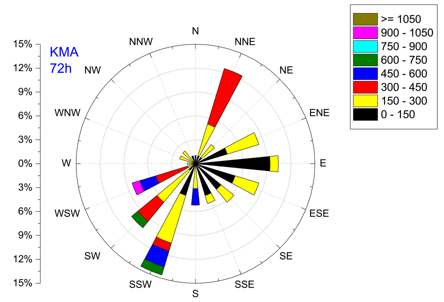


Figure 9. Examples of position error-rose diagram (TER) to represent the direction and magnitude distributions of position errors.

Another useful tool to evaluating the systematic bias of a certain objective track forecast method is name "Position Error Rose". TER uses the same conception of "wind rose" diagram as reference. Figure 9 shows the examples of track TER to represent the direction and magnitude distributions of position errors from five official guidances at 72h in 2016. In this example of TER diagram, each color bar represents different magnitude of position error, and the length of alignment of color bars represent the proportion of each azimuth angles. The TER diagram reveals the position error distribution (both the error magnitude and percentage of sample size) at each azimuth angle.

5.2 EPS forecast

To evaluate the performance of TC track forecast of each EPS (listed in Table 3), we first treat the ensemble forecasts as deterministic by summarizing the ensemble using the mean applied to the members. Figure 10 shows the ensemble mean position errors for six EPSs. It is indicates that ECMWF-EPS, UKMO-EPS and NCEP-GEFS are the top 3 EPSs in 2016. The ensemble mean position error at the lead time level of 120h for both ECMWF-EPS and UKMO-EPS are less than 400km.

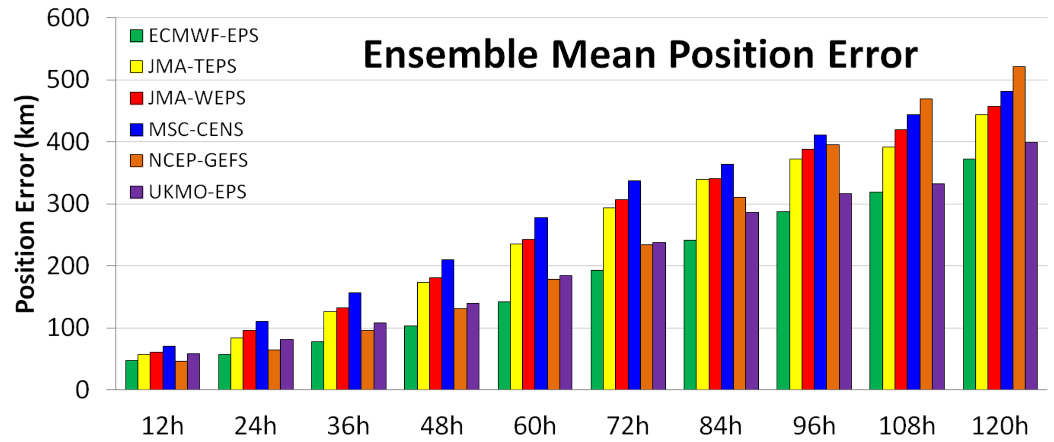


Figure 10. Ensemble mean position error for six EPSs in 2016.

The ensemble spread is an indicator of forecast uncertainties, which is not in linear relationship with mean position error. When the spread is large, the mean position error may be smaller, and vice-versa. Traditionally, researcher applied scatter plot of position error and ensemble spread to analyze the relationship between the forecast uncertainty and the error of a particular EPS. A bi-directional scatter plot is adopted here to re-analyze the traditional scatter plot. In the bi-directional scatter plot (Figure 11), the blocks in the middle of the plot represents the mean value of spread or position error. The lower (left) and upper (right) bars represent the 25th and 75th quantile values. It's found that only JMA-WEPS's median ensemble spreads and position errors are almost the same from 24 to 120h, most ensemble systems' median spreads are larger than position errors.

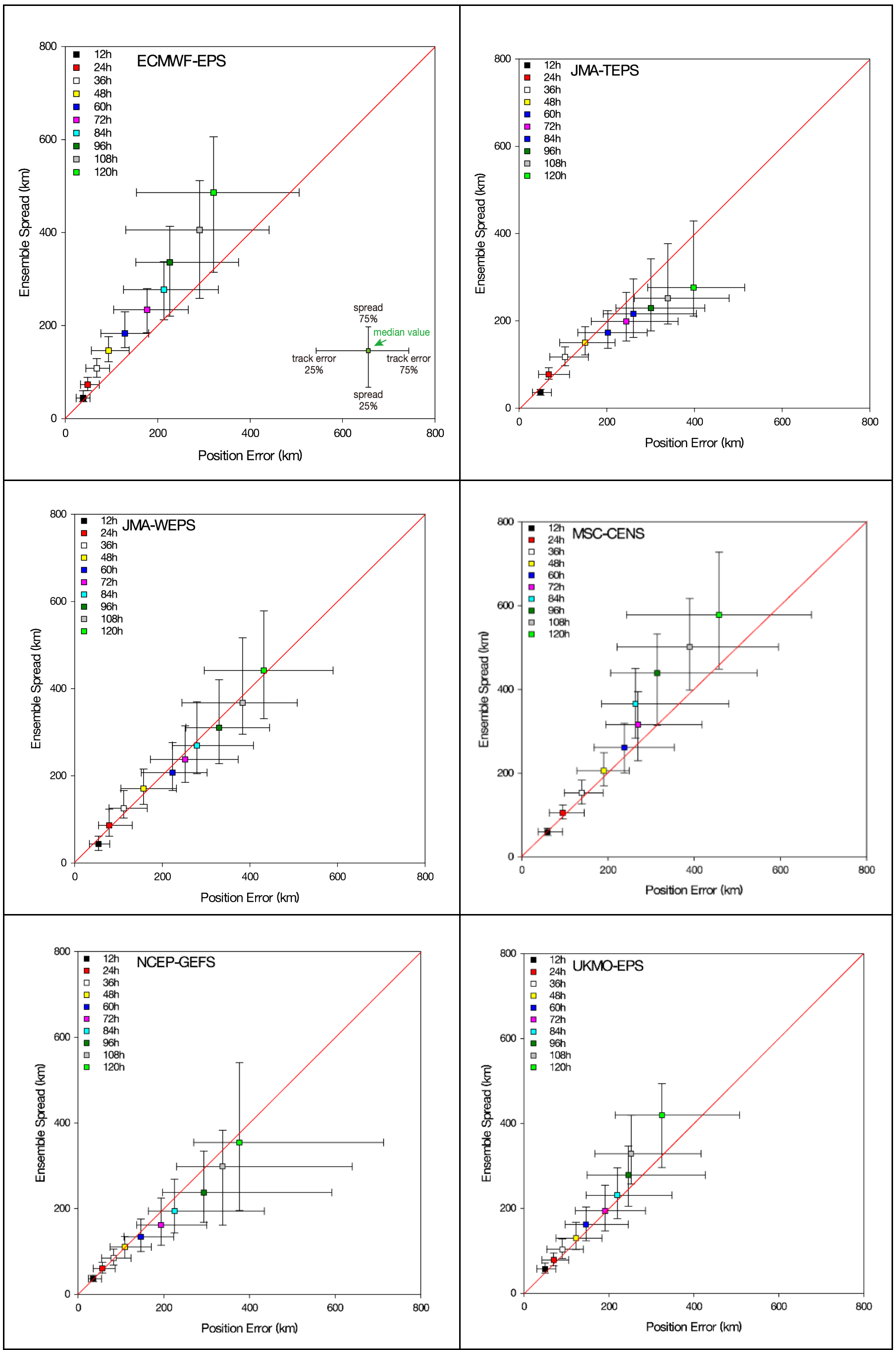


Figure 11. Bi-directional track forecast scatter plot for EPSs. The blocks represent the mean value of spread or position error. The lower (left) and upper (right) bars represent the 25th and 75th quantile values.

6. TC intensity forecast verification

6.1 Deterministic forecast

Forecast intensity error (i.e., maximum wind speed and minimum pressure) is defined as the Mean Absolute Error or Mean Relative Error of the difference between the forecast and best track intensity at the forecast verifying time. MAE provides an indication of the average magnitude of the error, whereas MRE measures the bias in the forecasts. Table 5 show the

MAE of maximum wind speed forecast for each method at each lead time level in 2016. One thing should be remembered that the wind speed of all forecast methods is converted to 10-min average according to the WMO documentation (Harper B A. *et al*, 2010).

Table 5. Mean absolute maximum wind speed error for each method at lead time levels of 24, 48, 72, 96 and 120h in 2016. Numbers in bracket are sample sizes. (Unit: m/s)

Method \ Lead times		24h	48h	72h	96h	120h
Official Guidance	CMA	5.10(363)	7.24(272)	7.13(196)	7.95(133)	9.07(85)
	HKO	5.12(214)	6.98(161)	8.42(112)	7.90(72)	10.16(44)
	JMA	5.30(358)	7.68(275)	8.96(200)	/	/
	JTWC	5.40(306)	6.11(221)	7.48(152)	8.02(84)	12.10(41)
	KMA	5.39(198)	8.33(144)	9.49(96)	8.22(56)	7.96(28)
Global NWP Models	ECMWF-IFS	9.22(141)	10.88(110)	11.28(83)	9.45(56)	7.24(34)
	JMA-GSM	6.92(336)	11.20(254)	13.64(184)	/	/
	NCEP-GFS	8.57(320)	10.81(205)	12.64(164)	15.29(98)	12.48(65)
	UKMO-MetUM	9.48(169)	10.67(126)	10.83(90)	10.45(58)	10.67(36)
Regional NWP Models	BoM-ACCESS-TC	8.19(52)	13.25(36)	14.38(21)	/	/
	CMA-TRAMS	8.60(158)	12.81(117)	13.13(85)	/	/
	STI-GRAPES	5.86(266)	7.35(212)	9.78(152)	/	/

Figure 12 presents two Taylor diagrams (Taylor, 2001) to assess the performance of intensity forecast. Taylor diagram is introduced in the verification of TC intensity forecast to analyze the internal relationship between the standardized deviation and correlation coefficient together with center different root-mean-square. The best prediction always with highest correlation coefficient compared to "OBS", and with standardized deviation and center different root-mean-square closed to "1". According to Figure 12 the RMS error of both minimum surface pressure and maximum wind speed were smallest at 0h for JMA. For most global models, the correlation coefficients of minimum surface pressure between observation and forecast are in the interval of 0.6 to 0.9. The normalized standardized deviations of maximum wind speed forecast for most global models are in the interval 0.75 to 1.25.

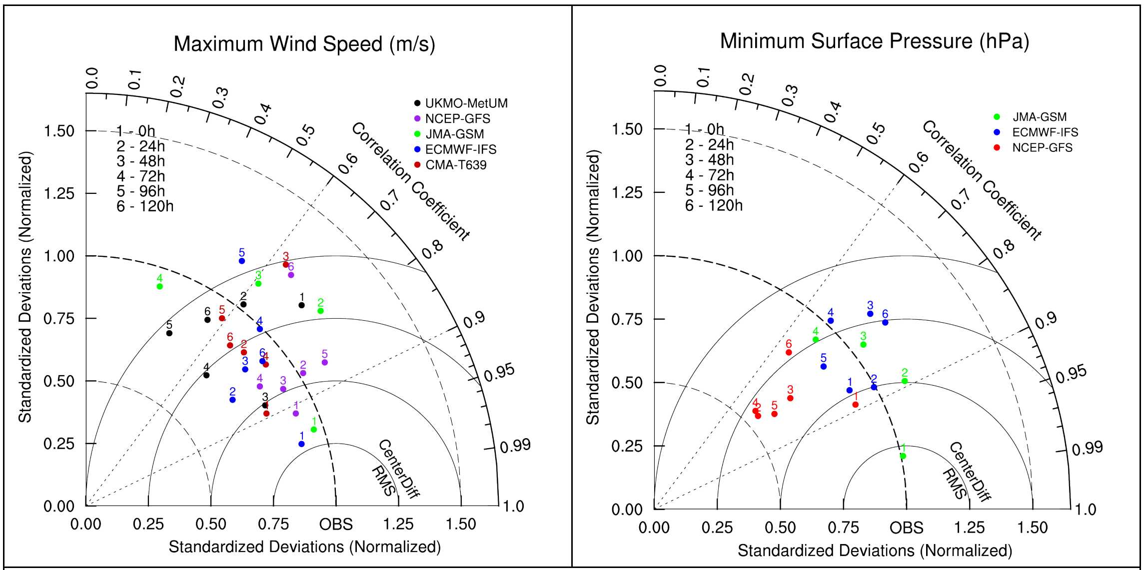


Figure 12. Using Taylor-diagram to evaluate global models' TC intensity forecast. Left: maximum wind speed, right: minimum surface pressure.

Figure 13 shows the intensity forecast skill score at the lead time levels of 24 and 48h for official guidances, global and regional models from 2010 to 2016. All the official guidances obtained positive intensity forecast skill scores for the past 7 years, however, skill scores were still much less than track forecast skill scores. Honestly, compare to official guidance, both global and regional models have been less skillful than official guidances in recent years. But in 2016, some models, such as JMA-GSM, NCEP-GFS, STI-GRAPES and UKMO-MetUM have shown positive intensity forecast skill at 24 and 48h.

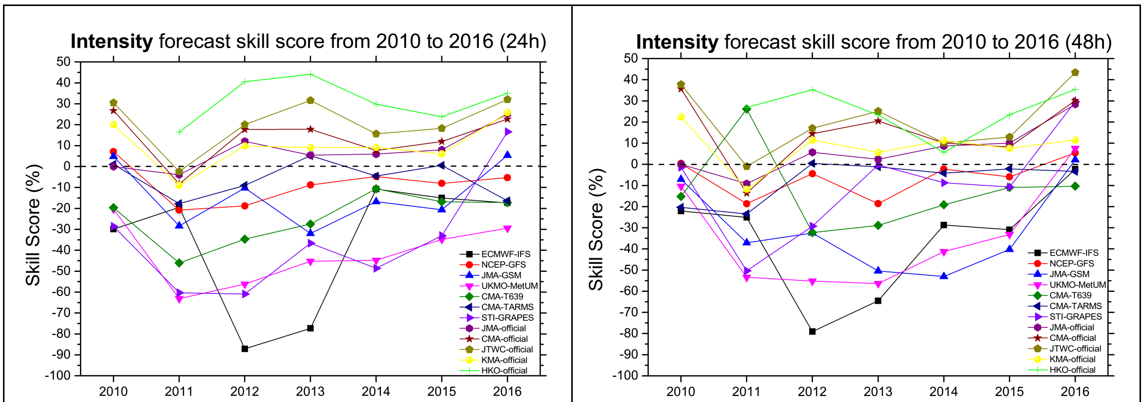


Figure 13. Same as figure 4 but for intensity forecast skill score.

6.2 EPS forecast

The ensemble forecasts of TC intensity from the TIGGE ensemble prediction systems as listed in Table 4 have been evaluated using Brier scores in minimum central pressure (Pmin). Figure 14 presents the non-homogeneous (215-315 cases) and homogeneous (96 cases) brier scores for six EPSs in 2016. It is show that ECMWF-EPS outperforms other systems significantly at short lead times, especially for homogeneous comparison. However, BS difference between six EPSs is narrowed at long lead time levels. The effect of initial correction is in-significant or even negative for some systems after 30 h. Compare to 2015, the ability of EPSs' intensity forecast has made a significant progress in 2016. The range of improvement is almost 5 to 40%. In particular, the improvements of NCEP-EPS are more than 20% at the lead time levels from 48 to 120h.

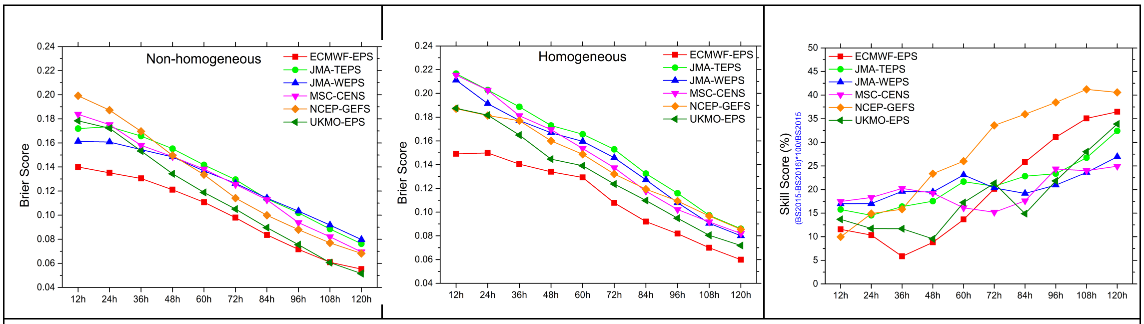


Figure 14. Brier scores for EPS intensity forecast in 2016, left: non-homogeneous comparison, middle: homogeneous comparison, right: BS skill which compare to 2015.

Non-homogeneous mean absolute intensity errors of EPSs' ensemble mean for Pmin are shown in figure 15. The maximum mean absolute errors of each system are located at lead time levels between 48 to 72h. The mean errors of EPSs at all lead times are positive between 5.8 to 28.7. Through estimating the initial intensity bias between EPSs and best track records. It is show that all EPSs are under estimated the TC initial intensity, and the more powerful the TC initial intensity is, the larger deviation between EPSs' starting intensity and observed intensity. The positive contribution of initial correction degrades quickly from 6 to 36h, especially for UKMO-EPS, ECMWF-EPS, and JMA-TEPS. The effect of initial correction is in-significant or even negative for some systems after 30 -54 h.

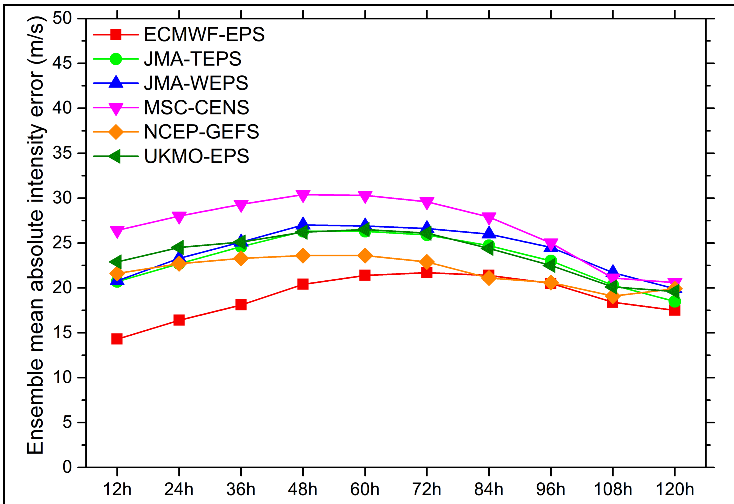


Figure 15. Mean absolute intensity errors of six EPSs for Pmin in 2016.

7. Future plans

Verification of TC forecasts is important for improving the NWP and official guidance that underpins the forecasts, making best use of this guidance in a forecasting context, and assisting the public, emergency managers, and other users of the TC forecasts to develop an appropriate level of confidence in the forecasts.

This report has briefly discussed the performance of typhoon forecast over western North Pacific in 2016. The verification results include TC track, and intensity for both deterministic and ensemble forecast guidance. In the future, for STI, we'll not only focus on evaluation of basic TC attributes such as track, intensity and genesis, but also focus on verifying TC impact variables such as precipitation, wind and storm surge. We'll continue to develop and improve methodologies for verifying forecast aspects of TC formation, structure, evolution, and motion, particularly from high resolution and ensemble NWP which are now the foundation for most operational TC forecasts.

Acknowledgement

This verification report has been effectively supported by ESCAP/WMO Typhoon Committee, CMA, STI/CMA, RSMC Tokyo Typhoon Center/WMO.

Contributors to this report are (in alphabetic order) Dr. BAI Lina (STI/CMA), Ms. Barbara Brown (RAL/NCAR), Ms. CAO Qing (SMC/CMA), Ms. LU Xiaoqin (STI/CMA), Dr. TANG Jie (STI/CMA), Dr. ZENG Zhihua (STI/CMA).

Appendix: acronyms used in this report

BoM	Bureau of Meteorology (Australia)
CMA	China Meteorological Administration
CMC	Canadian Meteorological Center
CSI	Critical Success Index
ECMWF	European Centre for Medium Range Weather Forecasting
EPS	Ensemble Prediction System
FAR	False Alarm Ratio
GEFS	Global Ensemble Forecast System
GFS	Global Forecast System
HKO	Hong Kong Observatory
JMA	Japan Meteorological Agency
JTWC	Joint Typhoon Warning Center
KMA	Korea Meteorological Administration
MAE	Mean Absolute Error
ME	Mean Error
MSE	Mean Squared Error
NWP	Numerical weather prediction
RMSE	Root Mean Squared Error
STI	Shanghai Typhoon Institute
TC	Tropical Cyclone
TIGGE	THORPEX Interactive Grand Global Ensemble
WMO	World Meteorological Organization

Reference

1. Chen G. M., H. Yu, Q. Cao, Z. H. Zeng, 2013: The performance of global models in TC track forecasting over the western North Pacific from 2010 to 2012. *Tropical Cyclone Res. Rev*., 2013, 2(3): 149-158.
2. Elsberry, R.L., T.D.B. Lambert, and M.A. Boothe, 2007: Accuracy of Atlantic and Eastern North Pacific tropical cyclone intensity forecast guidance. *Weather Forecasting*, 22, 747-762.
3. Harper B A, Kepert J D, Ginger J D. 2010, Guidelines for converting between various wind averaging periods in tropical cyclone conditions. Wold Meteorological Organization, *TCP Sub-Project Report*, WMO/TD-No.1555.
4. Lei, X.T. and H. Yu, 2014: Progress of the WMO Typhoon Landfall Forecast Demonstration Project (WMO-TLFDP). *Advances in Met. Sci. & Tech*. (in Chinese with English abstract)
5. Taylor K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *JGR*, vol 106, no. D7, 7183-7192, April 16, 2001
6. Tang, X., X.T. Lei, and H. Yu, 2012a: WMO Typhoon Landfall Forecast Demonstration Project (WMO-TLFDP) – concept and progress. *Tropical Cyclone Res. Rev.*, 1, 89-96.
7. Tang X., X.T. Lei, H. Yu, B.D. Chen, Z.Q. Chen, L.Y. Tao, 2012b: Progress report of the WMO Typhoon Landfall Forecast Demonstration Project (WMO-TLFDP). *WMO Workshop and Training Course on Operational Tropical Cyclone Forecast*, Shanghai, China.
8. Yu, H., G. Chen, and B. Brown, 2013: A new verification measure for tropical cyclone track forecasts and its experimental application. *Tropical Cyclone Res. Rev.*, 2, 185-195.
9. Yu, H., S.T. Chan, B. Brown, and Coauthors, 2012: Operational tropical cyclone forecast verification practice in the western North Pacific region. *Tropical Cyclone Res. Rev.*, 1(3), 361-372.