Front. Earth Sci. 2022, 16(1): 17–33
https://doi.org/10.1007/s11707-021-0900-2

RESEARCH ARTICLE

Evaluation of forecast performance for Super Typhoon Lekima in 2019

Guomin CHEN (✉)[1,2,3], Xiping ZHANG[2,3], Qing CAO[4], Zhihua ZENG[2,3]

1 Key Laboratory of Meteorological Disaster (Ministry of Education)/Joint International Research Laboratory of Climate and Environmental Change/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing 210044, China

2 Shanghai Typhoon Institute of China Meteorological Administration, Shanghai 200030, China

3 Shanghai Typhoon Institute, and Key Laboratory of Numerical Modeling for Tropical Cyclone of China Meteorological Administration, Shanghai 200030, China

4 Shanghai Marine Center Meteorological Observatory, Shanghai 200030, China

Abstract The predictions for Super Typhoon Lekima (2019) have been evaluated from official forecasts, global models, regional models and ensemble prediction systems (EPSs) at lead times of 1–5 days. Track errors from most deterministic forecasts are smaller than their annual mean errors in 2019. Compared to the propagation speed, the propagation direction of Lekima (2019) was much easier to determine for the official agency and numerical weather prediction (NWP) models. The National Centers for Environmental Prediction Global Ensemble Forecast System (NCEP-GEFS), Japan Meteorological Agency Global Ensemble Prediction System (JMA-GEPS) and Meteorological Service of Canada Ensemble System (MSC-CENS) are underdispersed, and the Shanghai Typhoon Institute Typhoon Ensemble Data Assimilation and Prediction System (STI-TEDAPS) is overdispersed, while the ensemble prediction system from European Centre for Medium-Range Weather Forecasts (ECMWF) shows adequate dispersion at all lead times. Most deterministic forecasting methods underestimated the intensity of Lekima (2019), especially for the rapid intensification period after Lekima (2019) entered the East China Sea. All of the deterministic forecasts performed well at predicting the first landfall point at Wenling, Zhejiang Province with a lead time of 24 and 48 h.

Keywords Typhoon Lekima (2019), track, intensity, landfall point, forecast verification

Received July 16, 2020; accepted February 18, 2021

E-mail: chengm@typhoon.org.cn

1 Introduction

Tropical cyclones (TCs) are among the most impressive weather systems in the western North Pacific (WNP) region, not only as serious natural weather phenomenon but also in terms of the intense precipitation, strong winds and storm surges they can cause. Each year, on average, approximately 9 typhoons make landfall in Chinese mainland (Lu and Zhao, 2013), leading to enormous disasters and property losses.

Recent studies have highlighted improved TC track forecasting by global numerical weather prediction (NWP) models (Yamaguchi et al., 2017; Hodges and Klingaman, 2019; Chen et al., 2020). Chen et al. (2020) compared the results from ECMWF-IFS, NCEP-GFS and UKMO-MetUM deterministic forecasts for the 2010–2019 WNP typhoon seasons and found that, on average, the ECMWF had the smallest track errors for all typhoons across the entire WNP region. The three global models all showed obvious improvement for lead times beyond 72 h. However, in terms of central mean sea level pressure and maximum sustained winds, typhoon intensities are still underestimated due to a lack of resolution and the use of parameterized processes for regional and/or global models (DeMaria et al., 2014). Kaplan et al. (2010) demonstrated that it is difficult for models to predict rapid transitions in intensity. Emanuel and Zhang (2016) also found that even though the skill of TC track forecasts has increased in recent years, the TC intensity forecast skill has improved slowly. To improve the performance of TC intensity forecasts, the National Oceanic and Atmospheric Administration (NOAA) initiated the ongoing Hurricane Forecast Improvement Project (HFIP) in 2008 (Gall et al., 2013).

Super Typhoon Lekima in 2019 (Fig. 1), considered as one of the strongest typhoons which made three times of landfall in China, underwent a significant rapid intensification after entering the East China Sea, and resulted in huge damage in Eastern China. Operational forecasting systems, including official forecasts, global and regional models, and ensemble prediction systems (EPSs), issued track and intensity forecasts after the formation of Lekima (2019).

The main purpose of this paper is to verify and compare the track and intensity forecast performance of Super Typhoon Lekima (2019) from several deterministic methods and EPSs. Specifically, this paper focuses on the following 4 questions regarding Lekima (2019):

1) What is the accuracy of the track and intensity forecasts from the official operational, deterministic models and ensemble prediction systems for Super Typhoon Lekima (2019) from day 1 to 5?

2) For each modeling system, how did the 24 to 120 h forecast performance for Lekima (2019) compare to the model's annual mean performance?

3) How does the track ensemble spread compare with the ensemble mean track?

4) What is the landfall point prediction performance of all the deterministic forecast methods?

Section 2 that follows describes data (both best track and forecast data) and the evaluation methodology used in this study. Section 3 presents the verification results for the deterministic track and intensity biases, errors and skill, the ensemble track errors, and the landfall point forecast errors. Section 4 is the summary and conclusions.

2 Data and method

2.1 Best track data

The best track is a subjective assessment of a typhoon's center location and intensity that is identified by trained analysts at the Shanghai Typhoon Institute of China Meteorological Administration (CMA-STI) using all observations available at the time of the analysis (typically some period of time – even several months – after the typhoon occurs). The best track will often differ from the "operational" track that is estimated real-timely during the

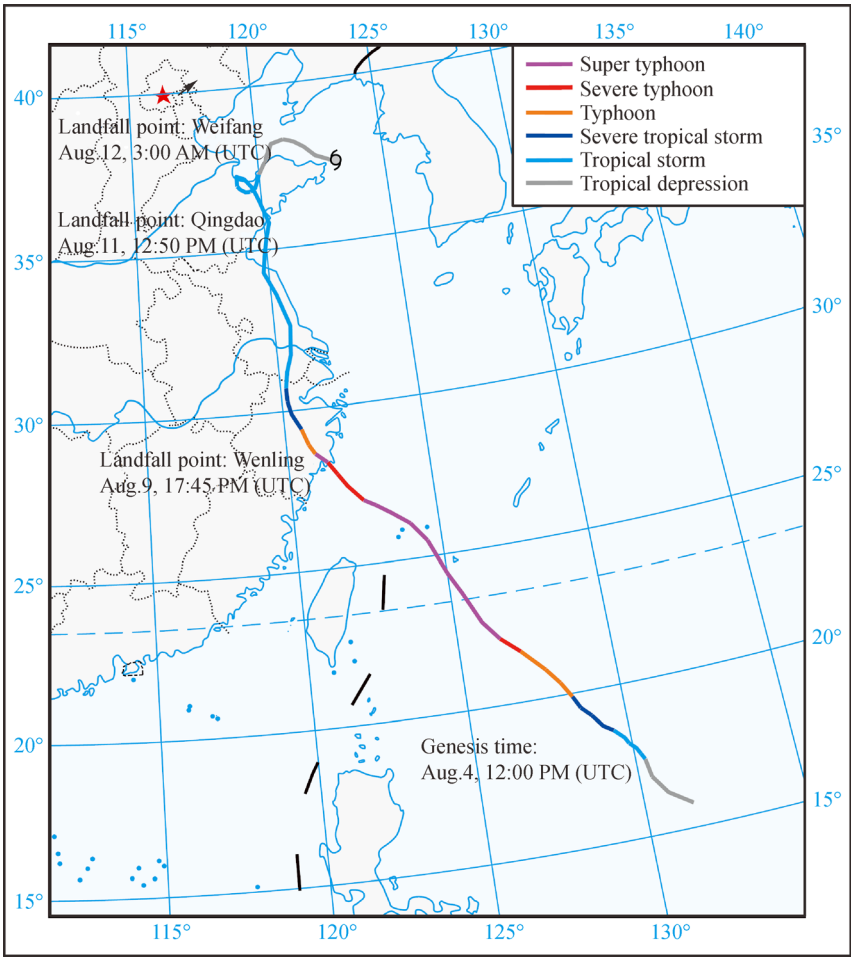

Fig. 1 Genesis time, track path, intensity evolution and three landfall points of Super Typhoon Lekima (2019).

lifetime of the storm. In particular, the best track from the CMA-STI consists of center position, maximum 2-min surface wind speed, and minimum central pressure (MCP) every 6 h during the life cycle of a tropical or subtropical cyclone. The CMA-STI's best tracks also include landfall information: location, time and intensity. The CMA-STI best track data can be accessed online (available at CMA tropical Cyclone data center website) (Ying et al., 2014).

2.2 Forecast data sets

Various types of TC forecast guidance are available in the WNP. This study focuses on forecast performance of operational official forecasts, global models, regional models and EPSs for Super Typhoon Lekima (2019). Table 1 shows detailed information regarding all forecast guidance products used in this study.

Official forecasts provided by the typhoon warning centers and national meteorological forecast agencies consist of human-generated track and intensity information. NWP systems, including both regional and global models, also provide predictions of TCs. The track and intensity forecast data from both the official forecasts and NWP were obtained through the Global Telecommunication System (GTS) of the World Meteorological Organization (WMO) in real time. The track data for Lekima (2019) from European Centre for Medium-Range Weather Forecasts Ensemble Prediction System (ECMWF-EPS), UK Met Office Ensemble Prediction System (UKMO-EPS), National Centers for Environmental Prediction Global Ensemble Forecast System (NCEP-GEFS), Japan Meteorological Agency Global Ensemble Prediction System (JMA-GEPS) and Meteorological Service of Canada Ensemble System (MSC-CENS) were downloaded from The International Grand Global Ensemble (TIGGE) database (Swinbank et al., 2016) in near real-time. These data were available online at Research Data Archive at the National Center for Atmospheric Research (NCAR), Computational and Information Systems Laboratory's website. The track data from Shanghai Typhoon Institute

Table 1 Details of forecast guidance

Category	Abbreviation		Full name or short description
Deterministic guidance	Official forecasts	CMA	China Meteorological Administration
		JMA	Japan Meteorological Agency
		JTWC	Joint Typhoon Warning Center
		KMA	Korea Meteorological Administration
		HKO	Hong Kong Observatory
	Global models	ECMWF-IFS	European Centre for Medium-Range Weather Forecast Integrated Forecasting System
		JMA-GSM	Japan Meteorological Agency Global Spectral Model of JMA
		NCEP-GFS	National Centers for Environmental Prediction Global Forecast System
		UKMO-MetUM	UK Met Office Unified Model system
	Regional models	GRAPES-TCM	Shanghai Typhoon Institute Regional TC-forecasting model based on Global/Regional Assimilation and PrEdiction System (GRAPES)
		GRAPES-TYM	National Meteorological Center of CMA Regional Typhoon forecasting model based on Global/Regional Assimilation and PrEdiction System (GRAPES)
		TRAMS	Tropical Regional Atmosphere Model for the South China Sea based on Global/Regional Assimilation and PrEdiction System (GRAPES)
		Shanghai-TCM	Shanghai Typhoon Institute Regional TC forecasting model based on Weather Research and Forecast modeling system
		HWRF	The atmosphere-ocean coupled Hurricane Weather Research and Forecast modeling system
Ensemble prediction system		ECMWF-EPS	European Centre for Medium-Range Weather Forecasts Ensemble Prediction System
		JMA-GEPS	Japan Meteorological Agency Global Ensemble Prediction System
		MSC-CENS	Meteorological Service of Canada Ensemble System
		NCEP-GEFS	National Centers for Environmental Prediction Global Ensemble Forecast System
		UKMO-EPS	UK Met Office Ensemble Prediction System
		STI-TEDAPS	Shanghai Typhoon Institute Typhoon Ensemble Data Assimilation and Prediction System

Typhoon Ensemble Data Assimilation and Prediction System (STI-TEDAPS) was gathered from the typhoon forecast database of CMA-STI.

2.3 Methods of track, intensity and landfall point forecast error verification

The direct track error (TE) is calculated by the great-circle distance between the forecasted points to the observed point (obtained from best track data set). As described by Froude et al. (2007), TE can be decomposed into two components: the along-track error (ATE) and the cross-track error (CTE). Figure 2 shows the relationship among the TE, ATE, and CTE. Therefore, a negative (positive) ATE indicates the forecast typhoon motion was too slow (fast) compared to the observed typhoon. Similarly, a negative (positive) CTE corresponds to a forecasted typhoon that is left (right) compared with the observed typhoon.

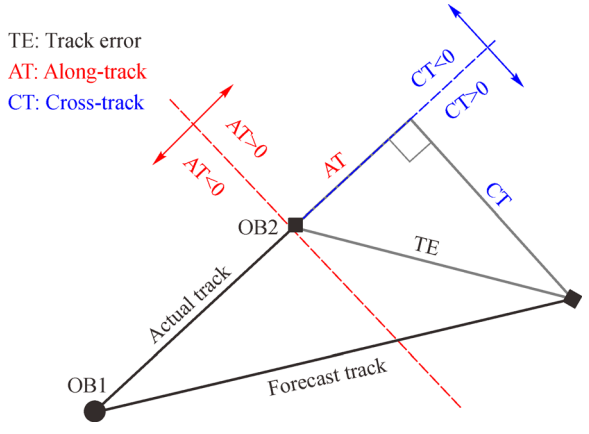

Fig. 2 Schematic of the decomposition of TE into the components CTE and ATE. The circle "OB1" point represents the initial position of the observed typhoon from best track. The diamonds "OB2" and "FC" represent the instantaneous positions of the observed and forecasted typhoons, respectively.

The forecasted landfall point is the intersection between the forecasted track and the coastline. Therefore, the landfall point forecast error is defined as the great-circle distance from the forecasted landfall point to the observed landfall point.

The ensemble spread (*SP*) and ensemble mean track error (EMTE) are compared to assess the statistical reliability of TC ensemble track forecasts. The position of the ensemble mean is the average of all ensemble member positions at a given lead time. Similar to the TE, the EMTE is defined as the great-circle distance between the mean forecasted point of all ensemble members and the position of the best track. The SP can be defined as the root mean square (RMS) average of the distances between the ensemble mean position and the ensemble member positions (Whitaker and Loughe, 1998):

$$SP = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\overline{f} - f(i)\right)^2}, \quad (1)$$

where $\overline{f}$ is the ensemble mean position, N represents the number of ensemble members and $f(i)$ is the track forecasted position of member i.

In this paper, maximum sustained surface wind speed is used to represent typhoon intensity for the evaluation. It should be emphasized that the average time used to define the maximum sustained wind speeds is not the same at different official forecast agencies. To make the verification results consistent, the maximum sustained wind speed in the official forecasts was converted to 2-min average according to WMO documentation (Harper et al., 2010). The recommended wind speed conversion factors for tropical cyclone conditions could be found in Table 1. One of this WMO documentation. However, the wind speed data from global or regional models was not converted, since the wind speed data was always output directly from the models themselves.

The intensity forecast mean relative error (MRE) is then defined as the mean difference between the best track intensity and the forecasted typhoon maximum sustained surface wind speed. As the name suggests, the mean absolute error (MAE) for intensity forecast is an average of the absolute value of the errors.

The skill score (*SS*) allows a comparison of a score for a forecast with the score for a reference forecast (e.g., climatology (Murphy, 1992), persistence (Yu et al., 2007), or random walk (DelSole and Tippett, 2016)). A simple SS for track forecast can be defined as the following equation:

$$SS_{\text{Track}} = \frac{TE_{\text{CLIPER}} - TE_{\text{ForecastMethod}}}{TE_{\text{CLIPER}}} \times 100\%, \quad (2)$$

where TE_{CLIPER} is the mean direct track error from the climatology and persistence track forecasting system (CLIPER) (Murphy, 1992). The CLIPER is considered a no-skill forecast and generally taken as a benchmark for forecast verification of a particular track forecast method. is the mean direct track forecast error for a particular forecast method.

As was the case for the track forecasts, the following equation can be used to compute the score skill for intensity forecast:

$$SS_{\text{Intensity}} = \frac{IE_{\text{TCSP}} - IE_{\text{ForecastMethod}}}{IE_{\text{TCSP}}} \times 100\%, \quad (3)$$

where IE_{TCSP} is the mean intensity error from a tropical cyclone statistical prediction (TCSP) (Yu et al., 2007). The TCSP model can be a useful benchmark for intensity forecast verification. is the mean direct intensity forecast error for a particular forecast method.

The above definition of *SS* allows a simple and direct evaluation of the forecasts, when compared to CLIPER data for the track and to TCSP, respectively (Chen et al.,

2016a) data for the intensity. The CLIPER and TCSP statistical models currently only provide forecasts for lead times out to 3 days; therefore, the SS results discussed below are only focused on 24, 48 and 72 h.

3 Results

3.1 Deterministic track error

First, the general track forecast performance of the deterministic forecast methods (including the official forecasts, global models and regional models) is evaluated for Super Typhoon Lekima (2019) and verified against the CMA-STI's best track.

Table 2 shows the mean track errors (non-homogeneous comparison) of the deterministic forecast methods for Typhoon Lekima (2019) and the annual means of 2019 at the lead times of 24, 48, 72, 96, and 120 h. For official forecasts, the mean track errors of Lekima (2019) from KMA at 24 h and HKO at 120 h are greater than their annual mean errors. The 24, 48, 72, 96, and 120 h track errors of Lekima (2019) from CMA, JMA and JTWC outperform their annual mean errors by 6.5–18.8 km (7.6%–22.1%), 5.8 – 31.9 km (4.1%–21.3%), 34.0–95.2 km (16.4%–40.5%), 47.2–77.4 km (15.7%–26.6%) and 28.4 – 74.5 km (7.7%–19.5%). Evaluation data for non-homogeneous comparisons shown in Table 2 have different sample sizes in each category. Therefore, it is needed for discussing inter-model comparisons to match numbers of each sample. The same forecast dates and lead times for the same category of forecast methods (official forecasts, global models, and regional models) are selected for homogeneous comparisons (Table 3). For the homogeneous comparison, JMA, JTWC, and CMA have the smallest mean track errors at 24 and 96 h, 48 and 120 h, 72 h, respectively.

For global models, the track forecast performance of ECMWF-IFS and NCEP-GFS for forecasting Lekima (2019) is better than their annual mean track forecast performances in 2019 at each lead time level. The track errors of ECMWF-IFS and NCEP-GFS are only 47.7, 86.7, 142.1, 164.9, 230.4 km and 64.9, 114.3, 193.2, 302.1, and 342.8 at 24, 48, 72, 96, 120 h, respectively. However, the mean track errors for Lekima (2019) from UKMO-MetUM are larger than the model's annual mean values from 48 to 120 h. Additionally, the results from homogeneous comparison indicate that ECMWF-IFS outperformed the other three global models, with mean track errors of 35.4, 72.7, 120.8, 166.3, and 188.0 km at 24, 48, 72, 96 and 120 h (Table 4).

For regional models, most models had better performance for forecasting Lekima (2019) compared with their average performance in 2019; only the mean track errors from GRAPES-TYM for Lekima (2019) at 24 and 120 h are greater than the corresponding annual means. Homogeneous comparisons of the mean track errors of regional models indicate that GRAPES-TYM, GRAPES-TCM and HWRF outperformed other regional models at lead time of 24 h, 48 and 72 h, and 96, and 120 h, respectively (Table 3).

Figure 3 shows the ATE and CTE components of the official forecasts, global models and regional models. For lead times less than 48 h, on average, the along-track components only show a small bias of 30 to 60 km for all the regional and global models. However, with lead time increasing, the negative along-track biases increase for all official forecasts (Fig. 3(a)), global models (Fig. 3(c)) and regional models (Fig. 3(e)). The 10% quintile of ATEs from most forecast methods exceed −500 km, indicating that for long forecast lead times, the forecasted Lekima (2019) from most models propagate too slowly, on average, compared to the observed track. The cross-track component does not show a distinct bias for any forecast method (Figs. 3(b), 3(d), and 3(f)).

The propagation direction (resulting in the CTE) of Lekima (2019) is much easier to handle for the official agencies and models. The propagation speed (resulting in the ATE) of Lekima (2019), however, seems to have been more difficult to predict for all the forecasting methods.

3.2 Deterministic intensity error

The non-homogeneous comparison of intensity forecast performance for Lekima (2019) and the whole year of 2019 (Table 4) shows that KMA, UKMO-MetUM and TRAMS have larger MAEs for Lekima (2019) compared to those from the whole year of 2019 based on the maximum sustained wind speeds at each lead time level (Table 4). For four-fifths of the lead times for the JTWC, ECMWF-IFS and HWRF systems, the forecast performance for Lekima (2019) is worse than the annual mean forecast for 2019 for that modeling system.

Table 4 also shows that the intensity forecast relative errors of most methods are negative from 24 to 120 h, indicating these methods underestimate Lekima's intensity. However, the HKO and UKMO-MetUM have positive errors in the maximum sustained wind speed for all 5-day forecasts. CMA, KMA, TRAMS, and GRAPES-TYM have positive errors for lead times of 120, 120, 72, and 120 h, respectively. Furthermore, the detailed intensity relative error results for all forecast methods show that the positive errors are centered upon the first 2 or 3 days of Lekima's life cycle. All the forecast methods evaluated in this paper failed to predict the rapid intensification after Lekima (2019) entered the East China Sea.

The homogeneous comparison of intensity forecast performance (Table 3) shows that, for official forecasts, JTWC, JMA, and CMA have the smallest absolute intensity forecast errors at 24 h, 48 and 96 h, 72 and 120 h, respectively. For global models, NCEP-GFS has the smallest absolute intensity errors within the first 2 days.

Table 2 Summary of non-homogeneous comparison of track location errors from deterministic forecast methods and ensemble prediction systems for Typhoon Lekima (2019) and the annual mean for 2019 at lead times of 24, 48, 72, 96, and 120 h. The highlighted values in bold text indicate that the error of the forecast method for Lekima (2019) is larger than the annual mean

		24 h		48 h		72 h		96 h		120 h	
		Lekima	Annual mean	Lekima	Annual mean	Lekima	Annual mean	Lekima	Annual mean	Lekima	Annual mean
Official forecasts	CMA	66.1(48)	84.9(557)	117.9(40)	149.8(431)	139.6(33)	234.8(324)	214.1(25)	291.5(244)	307.2(21)	381.7(170)
	JMA	78.8(50)	85.3(629)	135.6(31)	141.4(437)	173.2(27)	207.2(330)	220.9(25)	268.1(237)	327.0(22)	365.7(167)
	JTWC	74.0(31)	84.0(533)	140.1(29)	146.2(435)	187.6(27)	224.7(362)	256.7(25)	304.5(277)	341.3(22)	369.7(199)
	KMA	**110.3(31)**	102.2(477)	160.3(28)	172.8(389)	184.6(26)	242.8(312)	237.0(19)	289.2(214)	362.3(16)	366.9(143)
	HKO	47.3(14)	84.9(365)	96.5(14)	151.2(292)	151.3(13)	225.9(208)	134.9(6)	305.6(145)	**364.5(3)**	341.0(89)
Global models	NCEP-GFS	64.9(21)	78.0(364)	114.3(21)	145.9(326)	193.2(21)	222.2(274)	302.1(21)	355.3(211)	342.8(20)	454.4(150)
	ECMWF-IFS	47.7(12)	73.3(216)	86.7(11)	124.1(177)	142.1(10)	182.7(139)	164.9(9)	249.7(100)	230.4(9)	309.3(76)
	UKMO-MetUM	79.3(18)	82.0(281)	**177.4(17)**	141.7(228)	**278.5(16)**	211.0(176)	**280.7(13)**	269.2(136)	**414.9(12)**	378.1(98)
	JMA-GSM	76.9(36)	94.0(563)	156.7(35)	157.5(482)	196.5(32)	230.6(384)	306.8(25)	314.7(288)	**489.1(21)**	441.2(209)
Regional models	Shanghai-TCM	69.3(17)	75.8(196)	133.2(15)	133.5(156)	211.4(14)	213.4(119)	/	/	/	/
	TRAMS	57.7(17)	91.9(247)	135.4(16)	155.8(210)	245.8(15)	252.1(166)	/	/	/	/
	GRAPES-TYM	**85.0(35)**	81.6(513)	144.8(31)	150.2(426)	246.2(27)	253.3(329)	371.8(22)	408.2(232)	**616.1(21)**	548.1(155)
	GRAPES-TCM	84.2(36)	97.0(406)	140.8(34)	172.0(354)	216.5(30)	273.0(285)	/	/	/	/
	HWRF	65.7(19)	92.2(292)	168.8(19)	173.8(258)	204.0(19)	280.4(210)	258.0(19)	456.2(160)	337.4(18)	655.3(112)
Ensemble prediction systems (ensemble mean)	ECMWF-EPS	40.9(10)	85.7(243)	72.5(10)	142.8(211)	150.8(9)	215.7(171)	213.1(8)	325.7(136)	270.3(8)	428.7(101)
	JMA-GEPS	72.8(34)	101.0(523)	144.7(33)	175.6(448)	213.6(30)	247.9(368)	315.4(24)	328.9(286)	**462.1(19)**	430.2(212)
	MSC-CENS	80.0(15)	122.8(224)	170.3(15)	222.1(192)	277.1(14)	352.4(151)	**486.9(12)**	469.4(113)	**886.1(9)**	660.5(81)
	NCEP-GEFS	55.6(29)	80.8(371)	**155.7(29)**	155.0(309)	257.6(28)	246.7(242)	**360.3(24)**	343.3(179)	**474.9(21)**	473.2(131)
	UKMO-EPS	73.9(34)	97.4(496)	144.6(31)	170.5(422)	186.8(28)	244.3(334)	272.1(24)	326.3(256)	**466.8(21)**	413.9(189)
	STI-TEDAPS	72.2(18)	85.5(236)	140.2(17)	153.2(198)	206.4(14)	232.5(153)	/	/	/	/

Note: Numbers in brackets are sample size (unit: km).

Table 3 Homogeneous comparison of mean track location errors and absolute intensity error. The highlighted values in bold text indicate the smallest track and intensity errors in each lead time in each category

		24 h		48 h		72 h		96 h		120 h	
		Track	Intensity	Track	Intensity	Track	Intensity	Track	Intensity	Track	intensity
Official forecasts	CMA	46.8 (7)	3.9 (7)	83.3 (7)	3.9 (7)	**87.6** (6)	**1.5** (6)	82.7 (2)	3.5 (2)	265.1 (1)	**3.0** (1)
	JMA	**41.9** (7)	3.0 (7)	93.2 (7)	**1.4** (7)	98.7 (6)	2.7 (6)	**50.3** (2)	**1.5** (2)	346.3 (1)	6.0 (1)
	JTWC	42.9 (7)	**2.9** (7)	**76.5** (7)	4.0 (7)	94.1 (6)	2.3 (6)	113.2 (2)	4.5 (2)	**145.1** (1)	7.0 (1)
	KMA	70.4 (7)	5.6 (7)	88.9 (7)	3.6 (7)	120.8 (6)	2.5 (6)	61.7 (2)	4.5 (2)	323.6 (1)	4.0 (1)
	HKO	42.0 (7)	6.4 (7)	90.7 (7)	6.1 (7)	133.3 (6)	3.8 (6)	98.0 (2)	3.0 (2)	311.9 (1)	4.0 (1)
Global models	NCEP-GFS	62.5 (6)	**6.7** (6)	93.4 (6)	**8.5** (6)	223.0 (6)	12.0 (6)	342.6 (6)	12.8 (6)	356.0 (6)	5.2 (6)
	ECMWF-IFS	**35.4** (6)	13.7 (6)	**72.7** (6)	18.3 (6)	**120.8** (6)	15.7 (6)	**166.3** (6)	**7.3** (6)	**188.0** (6)	**3.8** (6)
	UKMO-MetUM	96.2 (6)	7.3 (6)	200.3 (6)	9.8 (6)	278.8 (6)	**10.5** (6)	365.9 (6)	9.5 (6)	452.6 (6)	11.3 (6)
	JMA-GSM	61.8 (6)	8.0 (6)	156.0 (6)	12.7 (6)	282.7 (6)	13.8 (6)	437.6 (6)	14.8 (6)	631.5 (6)	6.5 (6)
Regional models	Shanghai-TCM	65.9 (8)	7.9 (8)	147.5 (8)	10.9 (8)	201.4 (8)	12.1 (8)	/	/	/	/
	TRAMS	50.8 (8)	9.8 (8)	116.2 (8)	12.2 (8)	253.7 (8)	15.2 (8)	/	/	/	/
	GRAPES-TYM	**48.7** (8)	7.2 (8)	118.7 (8)	**7.4** (8)	255.5 (8)	10.9 (8)	394.6 (17)	**7.7** (17)	608.2 (16)	9.7 (16)
	GRAPES-TCM	69.0 (8)	**4.5** (8)	**102.3** (8)	8.1 (8)	**168.9** (8)	**10.2** (8)	/	/	/	/
	HWRF	72.4 (8)	5.1 (8)	199.5 (8)	11.8 (8)	216.4 (8)	13.9 (8)	**261.9** (17)	11.1 (17)	**339.7** (16)	**5.4** (16)

Note: Numbers in brackets are sample size.

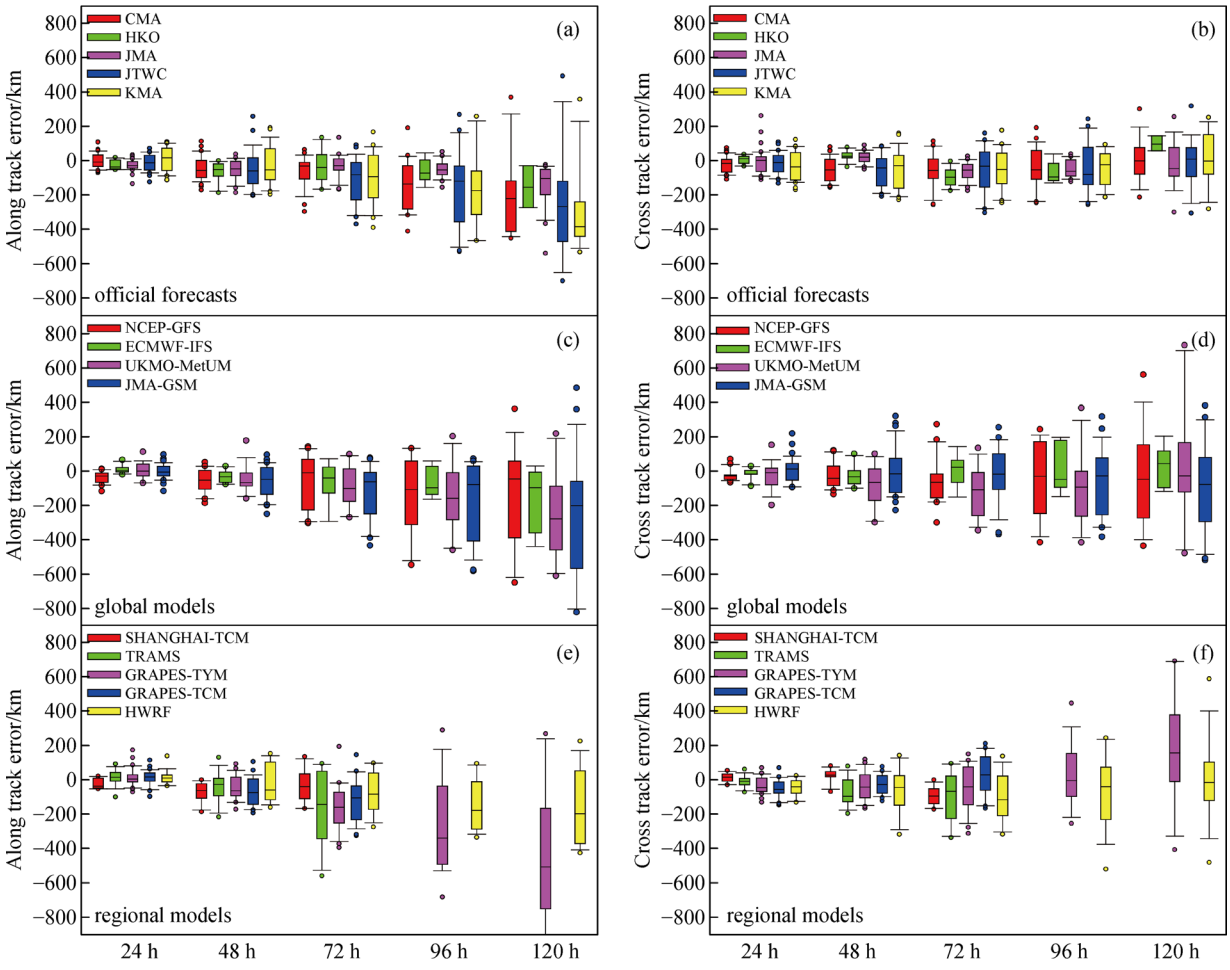

Fig. 3 Along-track and cross-track error components from the official forecast (a, b), global models (c, d) and regional models (e, f) for forecasting Super Typhoon Lekima (2019). The bar in the middle of the plot represents the median values of the errors, and the lower and upper ends of the boxes represent the 25th and 75th quantile values. The bars below and above the box represent the 10% and 90% quantile values, and the circles represent the outliers.

However, ECMWF-IFS has the smallest absolute intensity errors at 96 and 120 h.

The distributions of relative intensity errors from each method are shown in Fig. 4, using maximum sustained wind speed. For official forecasts (Fig. 4(a)), the maximum negative errors appear at 72 and 96 h of lead time, followed by 48 h, but the relative errors are the smallest at 120 h for all the official forecasts. Conversely, the global models have the largest negative errors at 120 h, and the maximum negative error, greater than -40 m·s^{-1}, is from JMA-GSM at 120 h (Fig. 4(b)). TRAM has the smallest forecast deviation among the regional models (Fig. 4(c)), with mean relative intensity error values of -1.1, -1.6, and 0.3 m·s^{-1} at the lead times of 24, 48, and 72 h, respectively (Table 4).

Chen et al. (2016b) used the Taylor-diagram to evaluate the internal relationship between the correlation coefficient (CC) and standard deviation (SD) together with the centered pattern root mean square (CRMS) of the intensity forecast. The Taylor-diagram is first used to display the quality of the model predictions compared directly with observations (Taylor, 2001).

For ease of comparison, the standardized deviation for intensity forecasts from each method is normalized by corresponding observed TC intensity. Therefore, in a typical Taylor-diagram, the best prediction always has the highest CC compared to the "OBS," with SD and CRMS values approaching 1 and 0, respectively. Figure 5 shows the Taylor-diagram results for the evaluation of the maximum sustained wind forecasts for Lekima from official forecasts, global models, and regional models. Most official forecast SD values are concentrated from 0.8 to 0.9, and all six points from HKO exceed the SD value of 1.0 (Fig. 5(a)); additionally, the CC values from CMA at 0, 48 and 96 h and from HKO at 0 h are greater than 0.95. SD values less than 1.0 mean the forecast method underestimate Lekima's intensity, and conversely, high CC values also mean that the forecast method has excellent

Table 4 Summary of non-homogeneous comparison of mean absolute error and relative error for the intensity forecast from deterministic forecasts for typhoon Lekima (2019) and the whole year of 2019 at lead times of 24, 48, 72, 96 and 120 h. The highlighted values in bold text indicate that the error of the forecast method for Lekima (2019) is larger than the annual mean

			24 h		48 h		72 h		96 h		120 h	
			MAE	MRE	MAE	MRE	MAE	MRE	MAE	MRE	MAE	MRE
Official forecasts	CMA	Lekima	4.2	−0.6	5.8	−0.9	6.9	−1.3	**8.8**	−1.0	5.0	1.8
		Annual mean	4.3	−0.5	5.9	−0.6	7.0	−1.2	7.3	−0.2	7.1	1.8
	JMA	Lekima	4.1	−2.8	**7.1**	−4.8	**9.7**	−5.9	**8.9**	−4.9	5.2	−2.1
		Annual mean	4.3	−1.1	6.5	−1.4	7.7	−1.5	7.2	−0.3	7.2	2.6
	JTWC	Lekima	**5.3**	−2.0	**8.5**	−3.1	**12.7**	−4.9	**11.0**	−4.5	5.3	−1.3
		Annual mean	4.8	1.6	6.2	1.2	7.6	0.8	7.8	1.0	7.0	2.0
	KMA	Lekima	**5.7**	−2.8	**7.1**	−2.9	**10.0**	−2.3	**8.0**	−0.2	**6.6**	4.1
		Annual mean	5.3	−1.9	6.6	2.9	7.3	−3.4	6.7	−2.4	6.5	0.5
	HKO	Lekima	**6.8**	6.4	**7.3**	5.7	4.2	2.1	2.7	2.7	3.0	3.0
		Annual mean	4.0	−1.5	5.4	−1.5	6.4	−2.6	6.8	−3.6	6.0	−3.5
Global models	NCEP-GFS	Lekima	**6.1**	−5.0	5.7	−5.0	**9.7**	−5.0	**11.2**	−6.0	**8.7**	−4.8
		Annual mean	5.0	−3.8	6.1	−4.3	6.8	−4.9	7.6	−4.3	7.6	−3.8
	ECMWF-IFS	Lekima	**9.8**	−10.0	**10.8**	−9.3	**10.4**	−8.2	**8.6**	−4.1	5.2	−0.6
		Annual mean	6.0	−4.7	7.1	−4.6	7.4	−3.5	7.2	−2.2	7.6	−1.7
	UKMO-MetUM	Lekima	**9.1**	3.3	**11.1**	4.2	**12.1**	4.8	**12.0**	6.5	**11.4**	9.8
		Annual mean	8.0	−0.2	9.3	−0.2	10.4	0.6	11.0	3.3	10.8	5.6
	JMA-GSM	Lekima	4.9	−2.8	6.3	−4.8	9.8	−5.9	**13.8**	−4.9	**11.6**	−2.1
		Annual mean	5.7	−1.1	8.9	−0.9	10.7	−1.3	10.4	−1.5	9.2	0.0
Regional model	Shanghai-TCM	Lekima	4.3	−6.9	**7.8**	−7.3	7.2	−5.6	/	/	/	/
		Annual mean	6.5	−6.0	7.1	−5.8	7.4	−5.6	/	/	/	/
	TRAMS	Lekima	**7.9**	−1.1	**9.7**	−1.6	**12.8**	0.3	/	/	/	/
		Annual mean	6.4	−4.1	8.8	−5.6	9.3	−5.1	/	/	/	/
	GRAPES-TYM	Lekima	5.6	−3.0	5.4	−3.6	7.1	−1.2	**8.4**	−3.0	**9.0**	5.5
		Annual mean	5.6	−2.7	6.6	−4.2	7.6	−2.3	8.1	−4.6	8.3	2.4
	GRAPES-TCM	Lekima	4.3	−3.0	6.1	−4.2	**7.5**	−3.3	/	/	/	/
		Annual mean	4.9	−1.7	6.5	−2.1	6.9	−1.9	/	/	/	/
	HWRF	Lekima	4.8	−3.0	**9.4**	−9.0	**11.6**	−7.0	**12.5**	−7.0	**7.0**	−5.6
		Annual mean	5.3	0.9	6.2	−1.5	7.3	−3.2	7.4	−4.4	6.1	−9.7

Note: The sample sizes are the same as in Table 2 (unit: $m \cdot s^{-1}$).

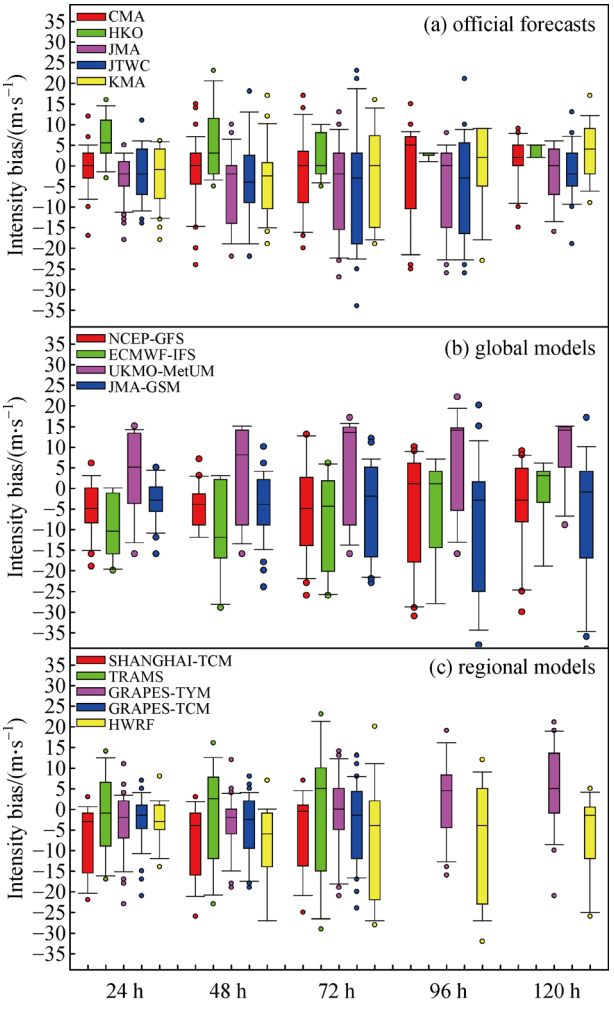

Fig. 4 The distributions of maximum sustained wind relative intensity error from (a) official forecast, (b) global models and (c) regional models for forecasting Super Typhoon Lekima (2019).

intensity forecast capacity. Therefore, Figure 5(a) shows that the intensity forecast results from HKO for Lekima (2019) were overestimates.

For global (Fig. 5(b)) and regional (Fig. 5(c)) models, in general, the CC values are much larger for regional models than global models. However, most models underestimate Lekima's intensity, except for HWRF at 0 h, UKMO-MetUM at 24 to 120 h, Shanghai-TCM at 96 and 120 h, TRAMS at 72 h, and GRAPES-TYM at 120 h.

3.3 Deterministic track and intensity forecasts skill

The SSs of the deterministic track and intensity forecasts for Lekima (2019) are compared in Figs. 6 and 7, respectively. For track forecasts, all the examined official forecasts, global models and regional models have positive skill at lead times of 24, 48, and 72 h. However, the skill of most official forecasts for forecasting Lekima (2019) are less than the skill of their annual mean, except for CMA at

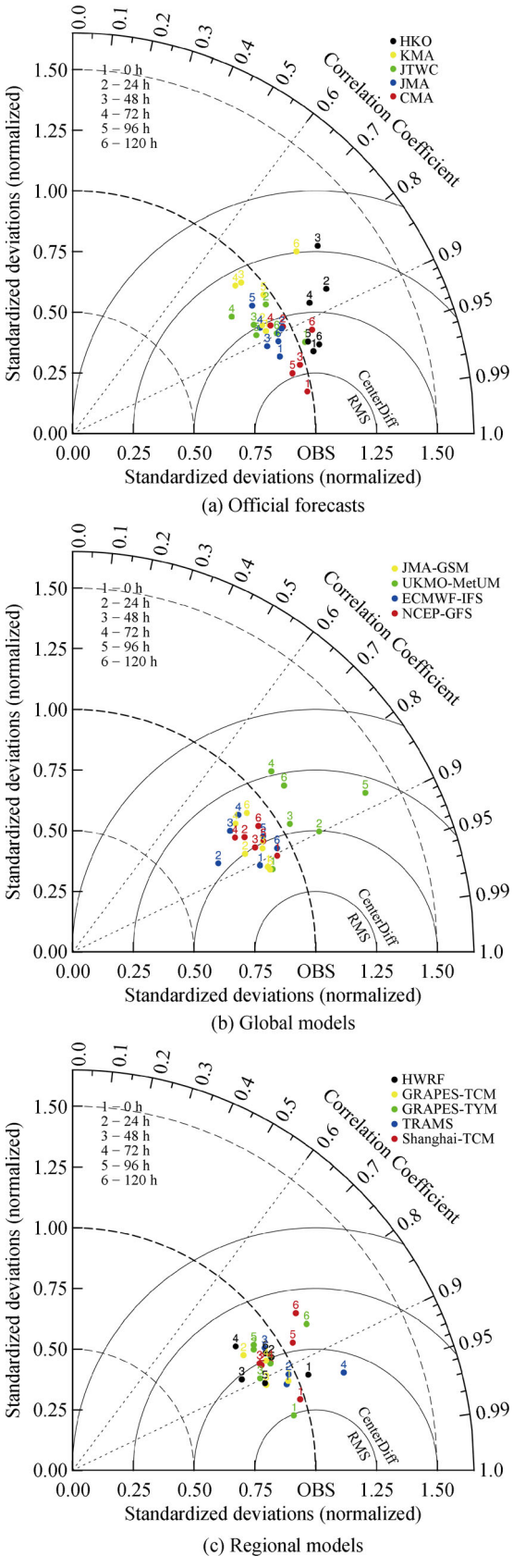

Fig. 5 Taylor diagrams for evaluating the maximum sustained wind forecast performance for Super Typhoon Lekima (2019) from the (a) official forecasts, (b) global models, and (c) regional models.

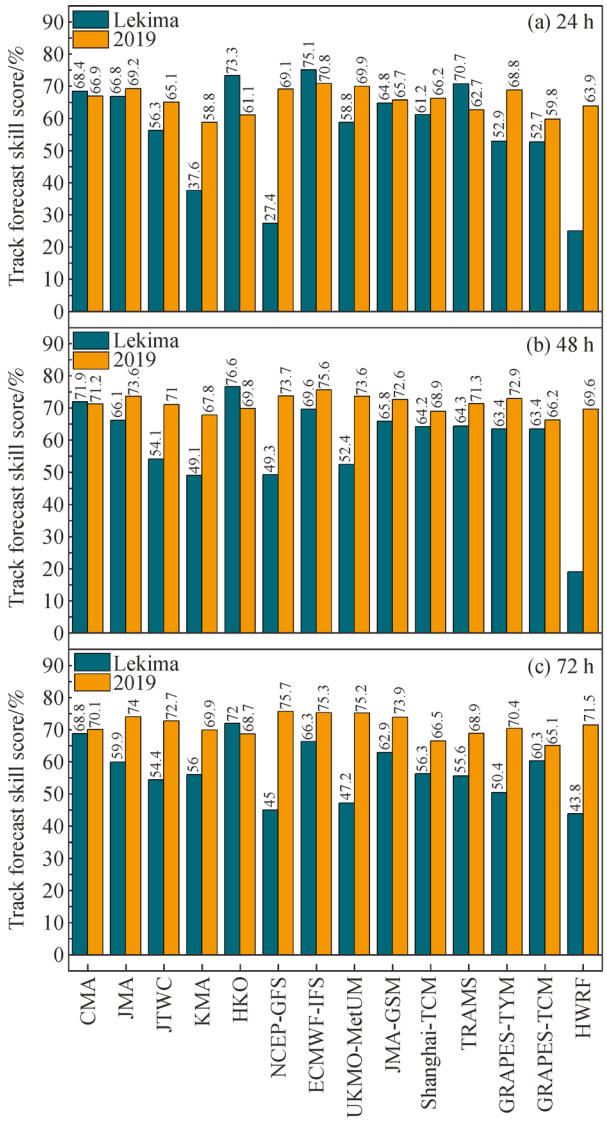

Fig. 6 Track forecast skill score of the official forecast, global models and regional models at lead times of (a) 24 h, (b) 48 h, and (c) 72 h for forecasting Super Typhoon Lekima (2019).

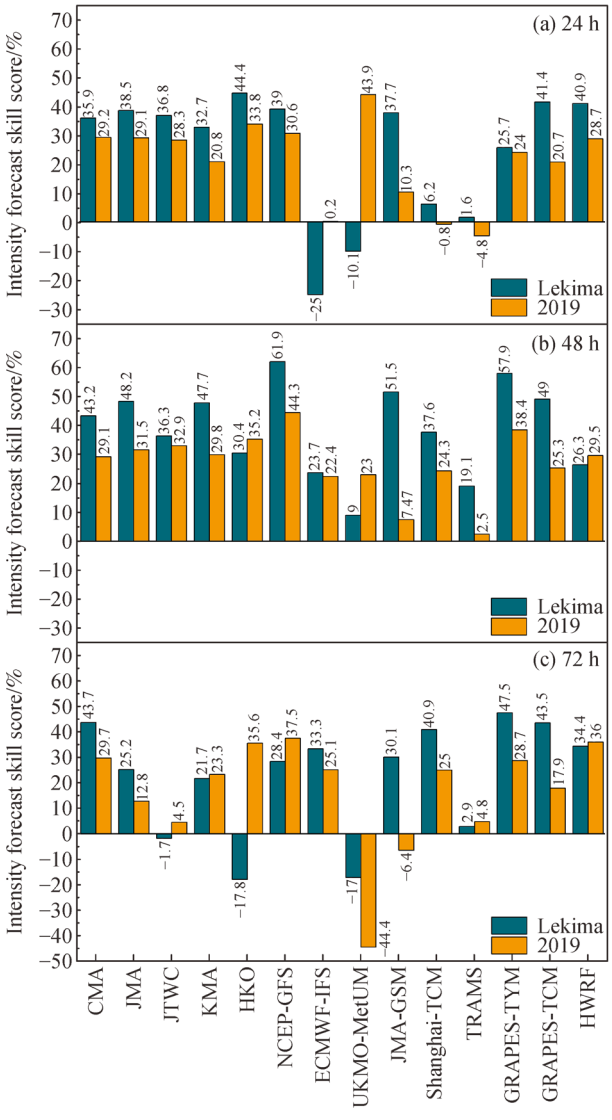

Fig. 7 Intensity forecast skill score from the official forecast, global models and regional models at lead times of (a) 24 h, (b) 48 h, and (c) 72 h for forecasting Super Typhoon Lekima (2019).

24, 48 h and HKO at 24, 48, and 72 h. Additionally, for both global and regional models, only the skill scores for ECMWF-IFS and TRAMS at 24 h exceed their annual mean skill scores.

For intensity forecasts, most official forecasts have positive skills for Lekima (2019) at the lead times of 24, 48, and 72 h, except for JTWC and HKO intensity *SS* at 72 h. At the lead time of 24 h, all 5 official forecasts' *SS*s for forecasting Lekima's intensity were larger than their annual mean intensity *SS*s in 2019. The verification results for the four global models' intensity predictions for Lekima (2019) indicate that ECMWF-IFS, UKMO-MetUM have negative skill at 24, 24, and 72 h, respectively. Unlike the global models, all of the regional models have positive *SS*s for forecasting Lekima's intensity for lead times within 72 h, and for most regional models, the *SS*s for Lekima are larger than the model's annual *SS*.

The increase in intensity forecasting skill for both regional and global models has been substantial in the last few years (Chen et al., 2020), starting from values of approximately −30% in 2010, and reaching values up to 10%–20%, and even 30% in 2018–2019. However, following Chen et al. (2020), since the early years of the 21st Century, the errors of intensity forecast models have evolved only slightly, with maximum decreases of only 2–3 $m \cdot s^{-1}$.

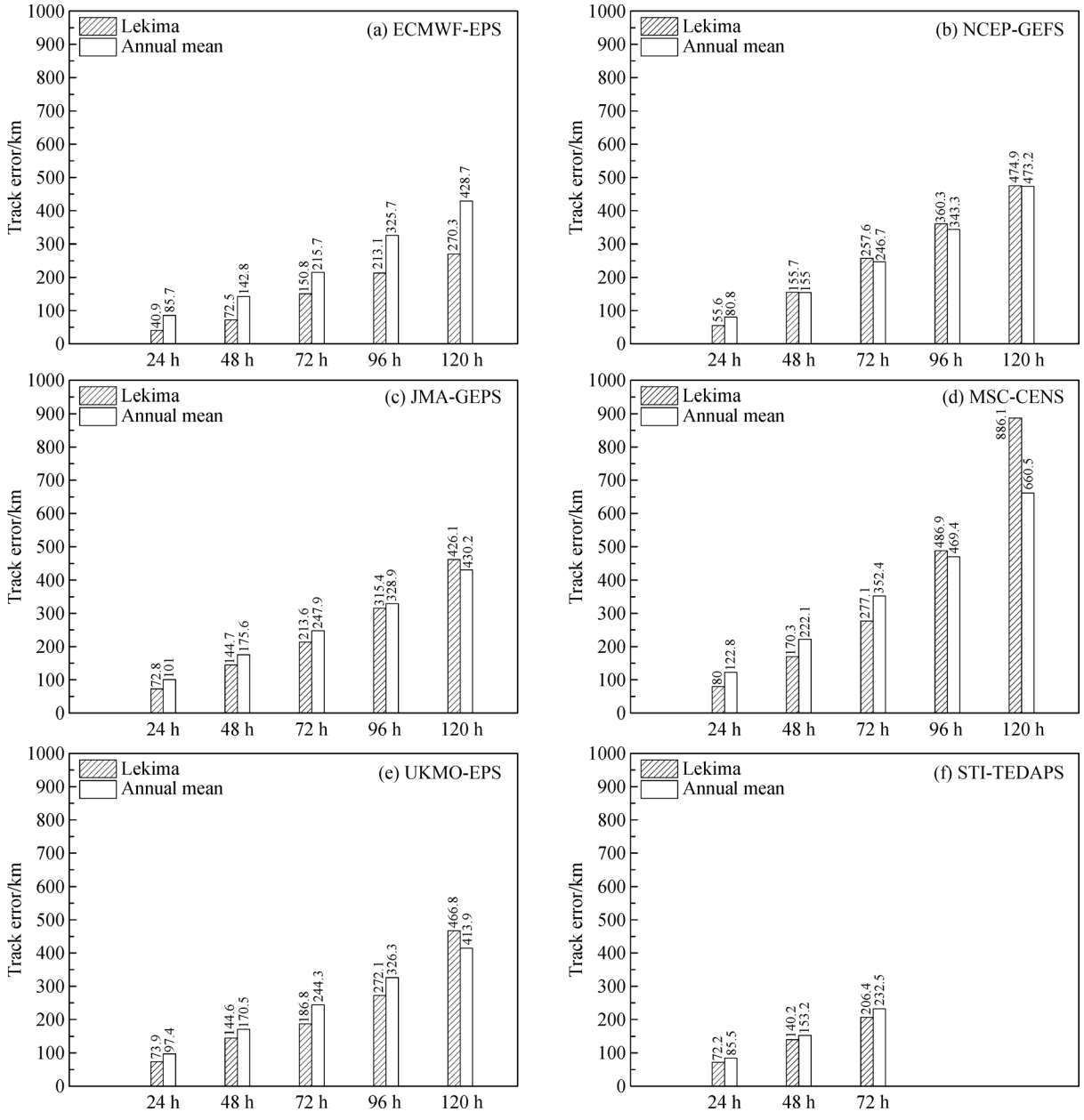

Fig. 8 Annual mean values of ensemble mean track errors in 2019 and ensemble mean track errors of (a) ECMWF-EPS, (b) NCEP-GEFS, (c) JMA-GEPS, (d) MSC-CENS, (e) UKMO-EPS, and (f) STI-TEDAPS for forecasting Super Typhoon Lekima (2019).

3.4 Ensemble track error

Figure 8 shows the EMTEs of ECMWF-EPS, NCEP-GEFS, JMA-GEPS, MSC-CENS, UKMO-EPS and STI-TEDAPS for Typhoon Lekima (2019), corresponding to their annual mean track errors in 2019. Overall, the EMTE of each EPS increases linearly during the entire 120 h forecast period for Typhoon Lekima (2019), and the calculated ranges of the EMTEs are between 270 and 890 km by 120 h.

It is important to note that EPS can provide a TC track probability result that ranges from 0 to 1, while traditional deterministic global and regional models can only provide the result 0 or 1 for a TC position (Leonardo and Colle, 2017). A simple way to measure the reliability of an EPS's probability forecast is to compare EMTE and *SP*. For a perfectly reliable EPS, the ensemble spread should be considered equal to the ensemble mean (Buizza, 1997). For an EPS to be statistically reliable, the EMTE should exhibit the similar value as the *SP*. In other words, the ensemble mean track error value should be reflected in the average difference between the ensemble mean value and the individual ensemble members. If the EMTE is greater than the *SP*, it implies that the EPS is underdispersive, and the

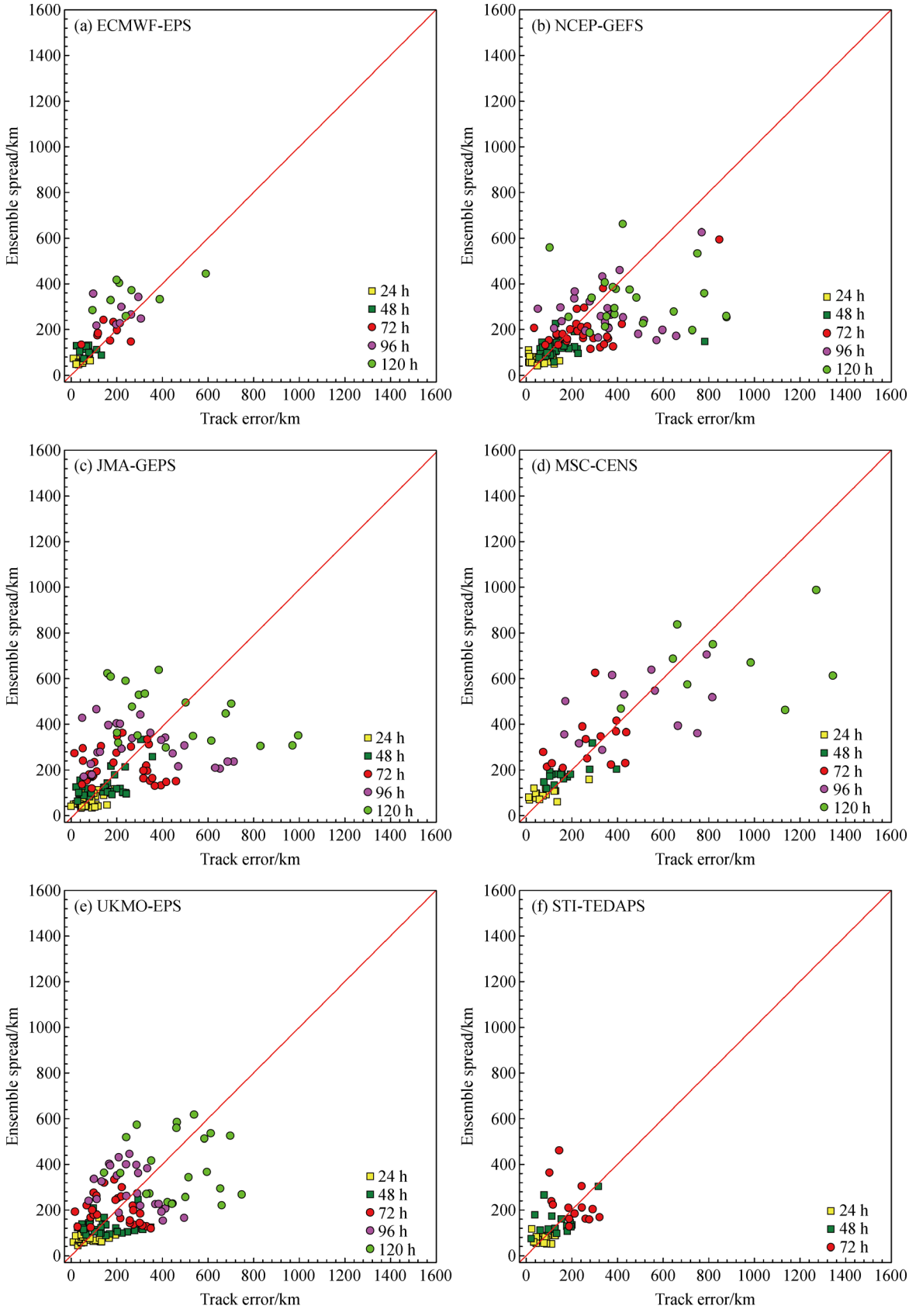

Fig. 9 Ensemble track forecast error and spread scatter plot for (a) ECMWF-EPS, (b) NCEP-GEFS, (c) JMA-GEPS, (d) MSC-CENS, (e) UKMO-EPS, and (f) STI-TEDAPS.

Table 5 The 24, 48, and 72 h forecast errors of the landfall point for Typhoon Lekima (2019) from official forecasts, global models and regional models

Lead time	24 h			48 h			72 h		
Landfall point	Wenling	Qingdao	Weifang	Wenling	Qingdao	Weifang	Wenling	Qingdao	Weifang
CMA	7.9	47.5	200.3	10.3	85.2	435.9	65.8	108.3	422.3
JMA	44.7	68.2	212.1	20.3	75.3	423.1	17.3	134.5	/
JTWC	99.7	90.6	206.1	105.4	99.3	416.9	/	154.3	/
KMA	67.2	91.3	212.7	32.0	110.3	354.6	103.9	199.3	418.9
HKO	14.8	74.8	211.8	48.5	64.4	428.5	47.3	102.3	/
NCEP-GFS	49.5	38.9	/	31.1	81.8	428.7	379.4	187.2	439.7
ECMWF-IFS	21.3	7.0	203.7	71.7	59.4	/	164.1	72.8	/
UKMO-MetUM	0	66.6	201.1	156.9	112.3	422.5	232.6	84.3	537.3
JMA-GSM	33.4	89.3	199.3	16.6	54.4	435.3	/	111.3	330.7
Shanghai-TCM	65.4	106.3	196.7	26.3	87.3	452.2	/	/	419.9
TRAMS	6.6	78.4	199.4	18.9	78.4	419.1	/	/	/
GRAPES-TYM	58.7	115.6	198.8	68.8	45.3	441.2	/	87.9	431.7
GRAPES-TCM	101.3	45.3	15.9	33.4	90.3	457.2	/	/	433.4
HWRF	34.6	89.3	/	11.2	105.4	436.1	894.1	77.4	424.5

Note: The slashes represent the forecasted track did not make landfall at that lead time (unit: km).

model system is overconfident in forecasting TCs. In contrast, if the *SP* is greater than the EMTE, that means the EPS is overdispersive, and the system is underconfident in forecasting TCs.

Figure 9 shows scatterplots of the ensemble mean error and ensemble spread of six EPSs at each lead time for Lekima (2019). In Fig. 9, the abscissa represents the track error, and the ordinate represents the ensemble spread. The maximum EMTE of ECMWF-EPS is only 590.3 km, and the corresponding *SP* is 443.5 km at 120 h of lead time. For the other four EPSs (except for STI-TEDAPS, with a maximum lead time of only 72 h), the EMTEs (*SP*s) all exceed 750 km (600 km) at 120 h. For NCEP-GEPS, JMA-GEPS and MSC-CENS, the EMTEs are larger than the SPs with increasing lead time. For STI-TEDAPS, the *SP*s are larger than the EMTEs with increasing lead time. These results indicate that STI-TEDAPS is overdispersive, and NCEP-GEFS, JMA-GEPS and MSC-CENS are underdispersive for predicting the track of Lekima (2019). However, for ECMWF-EPS and UKMO-EPS, the differences between the EMTEs and *SP*s are smaller.

NCEP-GEFS has unusual forecast results because the observed Lekima (2019) is going to make the third landfall at Weifang Shandong province, however, the NCEP-GEFS still predicted that it would go to north-eastward along with the edge of westerly trough.

3.5 Landfall point forecast error

Typhoon Lekima (2019) made its first landfall in Wenling, Zhejiang Province, as a super typhoon with a maximum 2-min surface sustained wind speed near the typhoon center of 52 $m \cdot s^{-1}$ and the MCP in the typhoon center of 930 hPa at 17:45 on August 9 (UTC). Approximately 43 h later, Lekima (2019) made its second landfall as a tropical storm in Qingdao, Shandong Province, with a maximum 2-min surface sustained wind speed near the typhoon center of 23 $m \cdot s^{-1}$ and the MCP in the typhoon center of 980 hPa at 12:45 on August 11 (UTC). Before merging with a cold front and turning into an extratropical cyclone, Lekima (2019) made the third landfall in Weifang, Shandong Province at 03:00 on August 12 (UTC).

As a super typhoon, Lekima (2019) enters present history as having the third strongest intensity on record among the typhoons that have made landfall in Zhejiang Province. It caused significant wind, rainfall and flooding damage when it made the first landfall in Wenling and caused severe rainfall damage throughout eastern China.

To assess the landfall point forecast errors of each forecast method, three different lead times are chosen for calculating the above three landfall point forecast errors: 24, 48, and 72 h before landfall. Table 5 shows the landfall point forecast errors at the lead times of 24, 48, and 72 h for Lekima (2019) from official forecasts, global models and regional models are shown. For the first landfall point in Wenling, the 24 h landfall point forecast errors for CMA, JMA, HKO, NCEP-GFS, ECMWF-IFS, UKMO-MetUM, JMA-GSM, TRAMS, and HWRF are less than 50 km. Notice that the forecasted tracks issued by JTWC, JMA-GSM, Shanghai-TCM, TRAMS, GRAPES-TYM, and GRAPES-TCM do not made landfall in Wenling at 72 h. For the second landfall point in Qingdao, the landfall point

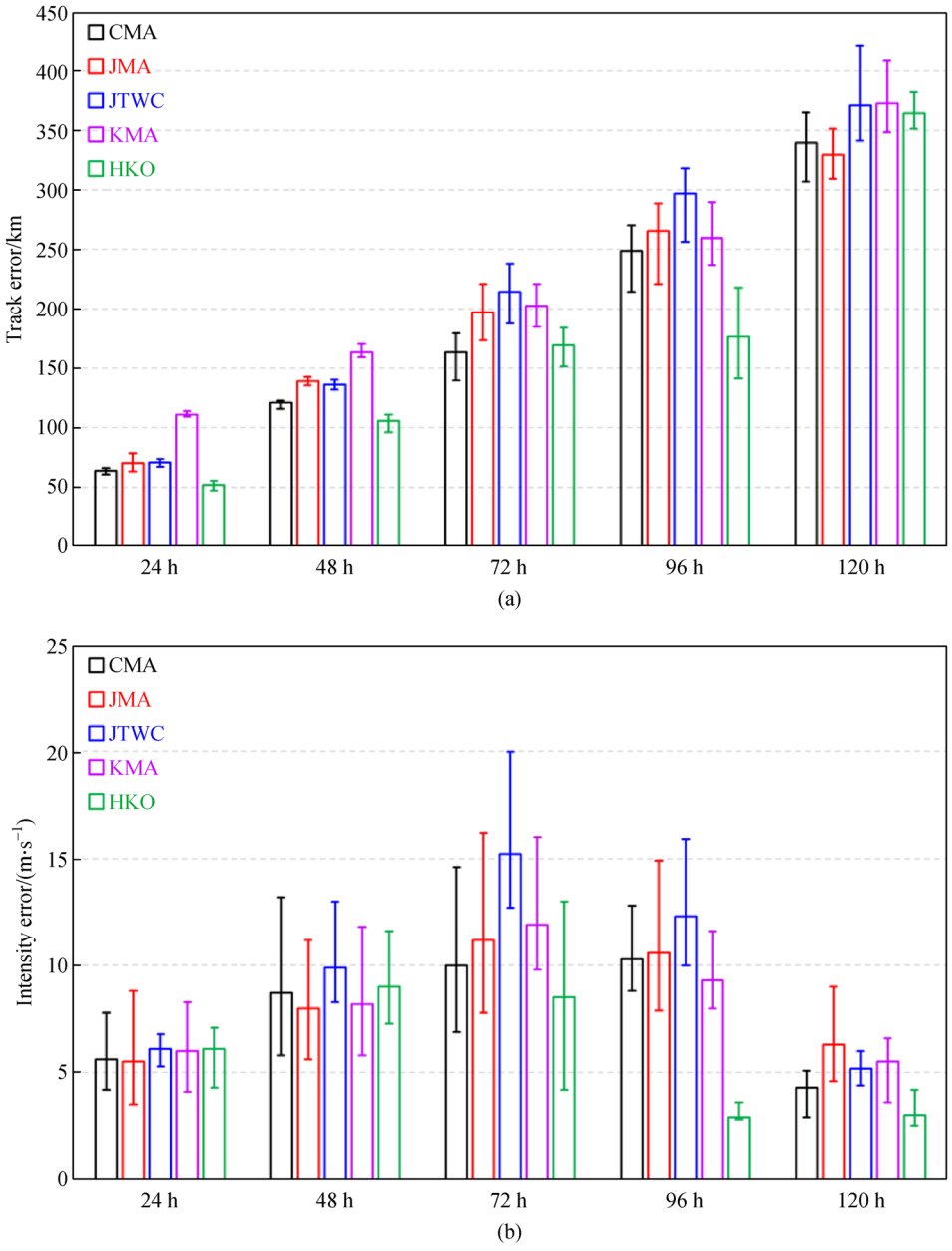

Fig. 10 Error variation intervals of (a) track forecast and (b) intensity forecast for official guidance by referring TC best track data sets form CMA-STI, JMA and JTWC, respectively. The rectangle bars, upper and lower error bars represent the mean, maximum and minimum value of errors by referring different best track data sets, respectively. For instance, the track errors of CMA for Lekima at 24 h are 66.1, 60.7, and 63.1 km after referring CMA-STI, JMA and JTWC best track data sets, the mean, maximum and minimum value of track errors would be 63.3, 60.7, 66.1 km in this figure.

forecast error ranges are from 7.0 to 115.6, 45.3 to 112.3 and 72.8 to 199.3 km at the 24, 48, and 72 h. Furthermore, for the third landfall point in Weifang, the landfall forecast errors from all the deterministic forecast methods approach or exceed 200 and 400 km at 24 and 48 h. The above results indicate that most of the forecast methods successfully predicted the first and second landfall points in Wenling and Qingdao at lead times of 24 and 48 h. However, almost all the methods failed to forecast the third landfall point in Weifang within 72 h.

4 Summary and conclusions

A post-verification of the forecast performance associated with Typhoon Lekima (2019) has been conducted for 5 official forecasts, 4 global models, 5 regional models and 6 ensemble prediction systems. Most official forecasts, such as CMA, JMA, and JTWC, show good performance in terms of mean track errors (TEs), as their forecast accuracies for Lekima (2019) outperform their annual mean errors. However, the mean track errors of Lekima

(2019) from KMA and HKO at 24 and 120 h are larger than their annual mean errors. For the global models, the performance of NCEP-GFS and ECMWF-IFS for forecasting Lekima (2019) is better than their annual mean errors. Homogeneous comparisons among the global models, show that ECMWF-IFS is, moreover, the most accurate global model for forecasting the track of Lekima (2019).

All forecast methods have a remarkable negative bias in the along-track forecasts for Lekima (2019) with increasing lead time. This implies that the Lekima (2019) forecasts by all the deterministic methods propagated too slowly on average. This phenomenon is more pronounced after Lekima (2019) made landfall in Wenling and entered the inland area. As Lekima (2019) underwent extratropical transition (ET), all the forecasts have their maximum slow speed biases in the direction along Lekima's observed track. The systematic slow speed biases in both the regional and global models may be mainly because lower-level flow was decoupled erroneously from the upper-level flow (Carr and Elsberry 2000; Payne et al. 2007).

EPS mean track errors (EMTEs) are evaluated by the scatter plot charts associate with EPS spreads (*SP*s) in this study. The scatter plots of the EMTEs and *SP*s of the EPSs suggest that, for the ECMWF-EPS and UKMO-EPS, the differences between the EMTEs and SPs are very small. However, NCEP-GEFS, JMA-GEPS and MSC-CENS are underdispersed, and the STI-TEDAPS is overdispersed with the lead time increasing for predicting the position of Lekima (2019).

The intensity forecasts are also evaluated in terms of maximum sustained wind speed for Lekima (2019). For most deterministic forecasts, there exist large negative errors compared to the observed intensity form Lekima (2019), which means that most of the deterministic methods underestimated the intensity. Overall, the correlation coefficients of the regional models are larger than those from the global models.

Additionally, this study used *SS* to evaluate the track and intensity forecast skill of the deterministic methods, by comparing an absolute forecast quality score with the score of a climatology, persistence forecast. For track forecasting, all the deterministic methods had positive skill; however, most *SS*s for official forecasts of Lekima (2019) are less than the skill of their annual mean. For intensity forecasting, all regional models have positive *SS*s, while some of the global models (such as ECMWF-IFS, UKMO-MetUM at 24 h and UKMO-MetUM at 72 h) have negative SSs.

Lekima (2019) made landfall at three points in eastern China. Examining the forecast errors for those three landfall points shows that all the deterministic forecasts have good performance in predicting the first landfall point at Wenling, Zhejiang Province at 24 and 48 h lead time. The landfall forecast errors for that point are less than 100 km for most official forecast, global models and regional models at 24 and 48 h. However, for the other two landfall points, at Qingdao and Weifang in Shandong Province, all the landfall forecast errors from deterministic methods exceed 200 km at 24 and 48 h.

It needs to be emphasized that the above conclusions are based on the best track data sets from CMA-STI. However, there are other two best track data sets from JMA and JTWC in the western North Pacific region. The differences regarding TC position and intensity among these three data sets are obvious (Ying et al., 2011). Therefore, the verification results would change if the best track data form JMA and JTWC are used. For instance, as Fig. 10(a) shows there may exist 3% to 10% track error variation intervals for the official agencies forecasting Likema (2019) by referring to TC best track data sets form CMA-STI, JMA and JTWC. In addition to the differences in the best track data sets obtained directly from CMA-STI, JMA and JTWC, it also should be noted that wind speeds of official agencies forecast data are converted for comparison with best track data from CMA-STI, JMA and JTWC. For example, the CMA's forecasted 2-min average maximum sustained surface wind speed data should be converted to 10-min or 1-min average before evaluating its intensity error by referring JMA or JTWC best track data sets. Considering these two effects, the intensity error variation intervals as represented by wind speed are 5% to 35% (Fig. 10(b)).

Future work will aim to evaluate the forecasts of the TC size and precipitation caused by Lekima (2019). In particular, if the track forecast results for Lekima are sufficiently accurate, such that Lekima's positions and landfall points were relatively certain, the maximum wind radius and rainfall information would become more important than the track for operational predictions. Maximum sustained wind and rainfall relate directly to potential damage. The role of the rapid intensification of Lekima (2019) will also be assessed further to better understand the dynamic mechanism of the negative intensity forecast errors associated with almost all the global and regional models and to improve the ability of the TC intensity forecast, especially for the TC rapid intensification process.

Acknowledgements The authors thank Dr. Rijin Wan of CMA-STI for providing us the forecast data for the deterministic method. We also thank Dr. Lina Bai of CMA-STI for providing us the best tracks of Lekima (2019). This paper is supported in part by the National Nature Science Foundation of China (Grant Nos. 41875069 and 41975067), the National Key R&D Program of China (Nos. 2018YFC1506406 and 2020YFE0201900) and the Shanghai S&T Research Program (No. 19dz1200101).

References

Buizza R (1997). Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. Mon Wea Rev, 125: 99–119

Carr L E III, Elsberry R L (2000). Dynamical tropical cyclone track forecast errors. Part II: midlatitude circulation influences. Weather Forecast, 15(6): 662–681

Chen G, Lei X, Zhang X, Chen P, Yu H, Wan R (2016b). Performance of tropical cyclone forecast in western North Pacific in 2015. Trop Cyclone Res Rev, 5(3–4): 47–57

Chen G, Yang M, Zhang X, Wan R (2020). Verification of tropical cyclone operational forecast in 2019. Annual report for ESCAP/WMO Typhoon Committee

Chen P, Yu H, Brown B, Chen G, Wan R (2016a). A probabilistic climatology-based analogue intensity forecast scheme for tropical cyclones. Q Roy Meteor Soc, 142: 2386–2397

DelSole T, Tippett M K (2016). Forecast comparison based on random walks. Mon Weather Rev, 144(2): 615–626

DeMaria M, Sampson C R, Knaff J A, Musgrave K D (2014). Is tropical cyclone intensity guidance improving? Bull Am Meteorol Soc, 95(3): 387–398

Emanuel K, Zhang F (2016). On the predictability and error sources of tropical cyclone intensity forecasts. J Atmos Sci, 73(9): 3739–3747

Froude L S R, Bengtsson L, Hodges K I (2007). The predictability of extratropical storm tracks and the sensitivity of their prediction to the observing system. Mon Weather Rev, 135(2): 315–333

Gall R, Franklin J, Marks F, Rappaport E N, Toepfer F (2013). The hurricane forecast improvement project. Bull Am Meteorol Soc, 94 (3): 329–343

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

Hodges K I, Klingaman N P (2019). Prediction errors of tropical cyclones in the Western North Pacific in the Met Office Global Forecast Model. Weather Forecast, 34(5): 1189–1209

Kaplan J, DeMaria M, Knaff J A (2010). A revised tropical cyclone rapid intensification index for the Atlantic and eastern North Pacific basins. Weather Forecast, 25(1): 220–241

Leonardo N M, Colle B A (2017). Verification of multimodel ensemble forecasts of North Atlantic tropical cyclones. Weather Forecast, 32 (6): 2083–2101

Lu X Q, Zhao B K (2013). Analysis of the climatic characteristics of landing tropical cyclones in 373 east China. J Trop Meteorol, 19(2): 145–153

Murphy A H (1992). Climatology, persistence, and their linear combination as standards of reference in skill scores. Weather Forecast, 7(4): 692–698

Payne K A, Elsberry R L, Boothe M A (2007). Assessment of western North Pacific 96- and 120-h track guidance and present forecastability. Weather Forecast, 22(5): 1003–1015

Swinbank R, Kyouda M, Buchanan P, Froude L, Hamill T M, Hewson T D, Keller J H, Matsueda M, Methven J, Pappenberger F, Scheuerer M, Titley H A, Wilson L, Yamaguchi M (2016). The TIGGE project and its achievements. Bull Am Meteorol Soc, 97(1): 49–67

Taylor K E (2001). Summarizing multiple aspects of model performance in a single diagram. J Geophys Res Atmos, 106(D7): 7183–7192

Whitaker J S, Loughe A F (1998). The relationship between ensemble spread and ensemble mean skill. Mon Weather Rev, 126(12): 3292–3302

Yamaguchi M, Ishida J, Sato H, Nakagawa M (2017). WGNE intercomparison of tropical cyclone forecasts by operational NWP models: a quarter century and beyond. Bull Am Meteorol Soc, 98 (11): 2337–2349

Ying M, Cha E J, Kwon H J (2011). Comparison of three western north pacific tropical cyclone best track datasets in a seasonal context. J Meteorol Soc Jpn, 89(3): 211–224

Ying M, Zhang W, Yu H, Lu X, Feng J, Fan Y, Zhu Y, Chen D (2014). An overview of the China Meteorological Administration tropical cyclone database. J Atmos Ocean Technol, 31(2): 287–301

Yu H, Hu C, Jiang L (2007). Comparison of three tropical cyclone intensity datasets. Acta Meteorol Sin, 28: 121–128