CrossMark

ORIGINAL PAPER

Multi-model ensemble forecasts of tropical cyclones in 2010 and 2011 based on the Kalman Filter method

Chengfei He[1] · Xiefei Zhi[1] · Qinglong You[1] · Bin Song[1] · Klaus Fraedrich[1,2]

Abstract This study conducted 24- to 72-h multi-model ensemble forecasts to explore the tracks and intensities (central mean sea level pressure) of tropical cyclones (TCs). Forecast data for the northwestern Pacific basin in 2010 and 2011 were selected from the China Meteorological Administration, European Centre for Medium-Range Weather Forecasts (ECMWF), Japan Meteorological Agency, and National Centers for Environmental Prediction datasets of the Observing System Research and Predictability Experiment Interactive Grand Global Ensemble project. The Kalman Filter was employed to conduct the TC forecasts, along with the ensemble mean and super-ensemble for comparison. The following results were obtained: (1) The statistical–dynamic Kalman Filter, in which recent observations are given more importance and model weighting coefficients are adjusted over time, produced quite different results from that of the super-ensemble. (2) The Kalman Filter reduced the TC mean absolute track forecast error by approximately 50, 80 and 100 km in the 24-, 48- and 72-h forecasts, respectively, compared with the best individual model (ECMWF). Also, the intensity forecasts were improved by the Kalman Filter to some extent in terms of average intensity deviation (AID) and correlation coefficients with reanalysis intensity data. Overall, the Kalman Filter technique performed better compared to multi-models, the ensemble mean, and the super-ensemble in 3-day forecasts. The implication of this study is that this technique appears to be a very promising statistical–dynamic method for multi-model ensemble forecasts of TCs.

1 Introduction

Tropical storms (TSs) or typhoons in the western Pacific frequently bring strong winds, rainstorms and ocean surges and generally cause serious economic losses and casualties across the entire East Asian region. As a result, the forecasting of typhoons, especially their intensity, rainfall and tracks, has been challenging for operational weather centers in different countries and regions. Some improvements have been achieved. Relying on the development in numerical weather prediction (NWP), the technologies of data assimilation and multi-model ensemble forecasting, as well as satellite observations, the forecasting of tropical cyclone (TC) tracks and intensity have improved considerably, in which multi-model ensemble prediction has played a substantial role in enhancing both the accuracy of deterministic forecasts and providing information on forecast uncertainties (WMO 2008a; Franklin 2012; Yamaguchi 2013).

The multi-model ensemble technique has been widely used in typhoon prediction, successfully reducing forecast errors since the first applications at the beginning of the 1990s (Leslie and Fraedrich 1990). With the ensemble mean and super-ensemble multi-model methods presently being the most frequently utilized, a series of tropical system forecast experiments have been conducted, and both track and intensity forecast errors have been reduced compared with individual ensemble members (Goerss

✉ Xiefei Zhi
zhi@nuist.edu.cn

[1] Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing 210044, China

[2] Max Planck Institute for Meteorology, Hamburg, Germany

2000; Krishnamurti et al. 1999; Kumar et al. 2003; Williford et al. 2003). Advanced and new techniques are constantly emerging in predicting TC tracks. For example, Qi et al. (2014) proposed a selective ensemble mean technique for TC track forecast based on the errors of ensemble prediction system members at short lead times. Similarly, a method of retrieving optimum information of typhoon tracks in a multi-model ensemble of forecasts is explored by Tien et al. (2012). Twenty-four- to seventy-two-hour TC track forecast errors have been reduced by nearly half from 1990 to 2008; however, track prediction errors exceeding 1000 km for 3-day predictions still occur (Franklin 2012; DeMaria et al. 2014). Also, intensity forecasting has shown statistically significant improvement during the past two decades, but little or no progress has been made in improving the methods that produce TC intensity forecasts (DeMaria et al. 2014). Thus, reducing TC track and intensity forecast errors remains an important research topic.

The Kalman Filter was first applied in multi-model forecasting based on satellite rainfall data (Shin and Krishnamurti 2003). Since then, Rixen et al. (2009) and Lenartz et al. (2010) also took advantage of the Kalman Filter approach to forecast 48-h sea surface temperature (SST). To improve surface drift prediction, a hyper-ensemble method was introduced by Vandenbulcke et al. (2009). Nevertheless, the Kalman Filter is still mainly used to revise the temperature, precipitation and other scalar atmospheric fields in the post-processing of single-model forecasts (Libonati et al. 2008; McCollor and Stull 2008). Our goal is to produce improved tropical system forecasts and develop innovative Kalman Filter techniques by using a suite of multi-model forecast outcomes to better predict TC's track and intensity evolution. Success in this area would be of great use to government agencies, especially those involved in decision making for people living near the coast and suffering the regular occurrence of TCs in summer and autumn.

In the present study, Kalman Filter techniques were employed to predict TC tracks and intensities in the western North Pacific basin. The multi-model ensemble mean and super-ensemble method were compared. Section 2 describes the models used and the verification TC data. An analysis of the methods and experimental setup is presented in Sect. 3. The multi-model forecasts, as well as those from the ensemble mean, super-ensemble, and Kalman Filter are compared in Sect. 4. Finally, a summary, discussion, and future outlook are given in Sect. 5.

2 Data

The THORPEX Interactive Grand Global Ensemble (TIGGE) is a key component of The Observing System Research and Predictability Experiment (THORPEX)—a

Table 1 Information on the tropical cyclones (TCs)/tropical storms/typhoons in the training and forecasting periods

Lead time	Training period		Forecasting period		Total	
	TCs	Samples	TCs	Samples	TCs	Samples
24 h	14	115	14	153	28	268
48 h	11	82	12	130	23	212
72 h	6	44	10	104	16	148

worldwide weather research program to enhance the accuracy of 1-day to 2-week high-impact weather forecasts for the benefit of society and humanity. TC ensemble forecasts of the China Meteorological Administration (CMA), European Centre for Medium-Range Weather Forecasts (ECMWF), Japan Meteorological Agency (JMA), and National Centers for Environmental Prediction (NCEP) within the TIGGE dataset were used to investigate the track and intensity (central mean sea level pressure) of TCs based on ensemble mean, super-ensemble, and Kalman Filter methods. The four centers in the TIGGE archives exploit different tracking algorithms to identify and track cyclones in the NWP output. The NCEP system uses a single-pass Bames analysis, in which the TC center is the location of minimum surface pressure after interpolation. In this way, the position for a TC from a model dataset with 1-degree resolution can be refined to within 1/16 of a degree (Marchok 2002). The tracking algorithm currently utilized at the ECMWF is based on extrapolation of past movement and mean sea level pressure or vorticity to determine a cyclonic center (Grijn 2002). The TSs, TCs and typhoons employed in our study are summarized in Table 1. The verification TC tracks and intensities came from the best-track datasets provided by the Joint Typhoon Warning Center (JTWC). The forecast and evaluation data are summarized as follows:

TIGGE datasets: The average TC tracks and central mean sea level pressures within the western Pacific region were exploited during the whole-year periods of 2010 and 2011, whose leading times were 24, 48 and 72 h and twice a day (UTC 00:00 and 12:00).

JTWC best-track datasets: The corresponding TC daily data during the same period were applied to verify the forecasts.

3 Methods

Generally, different forecast models can simulate different aspects of the real field. Therefore, to provide a forecast with higher skill, multi-model ensemble techniques usually "assimilate" more than one model that differ in resolution, physical processes, and data assimilation technique. The

simplest and most widely used technique is the ensemble mean. This is simple to implement without a training period and often achieves a better result than a single model (Goerss 2000). In the ensemble mean, every model weight is equal to $\frac{1}{N}$, where N is the number of participating models. The ensemble mean is one of the methods compared in the present paper.

Other kinds of methods make use of historical data to determine the combination of weights by minimizing the distance between the historical forecast and observation or reanalysis data, and then using the weights to combine different models to predict the future. These methods are mainly of non-equal weight and obtain better forecasts than the ensemble mean in most cases. Figure 1 offers a schematic view of these non-equal weighting methods. Their processes can be split into two parts, with the minimization process referred to as the training period and the second period as the prediction. Obviously, a major problem faced by these multi-model techniques is whether or not the combination achieved during the training period still provides optimal performance in the prediction phase. For instance, whether or not the model combination, and thus the forecast skill, can remain stable over time in a particular forecast period.

Krishnamurti et al. (1999, 2000) presented a multi-model method called super-ensemble forecasting in which the weighting factor for each model is calculated through multiple linear regression analysis between the forecasts and observed values. The method has been successfully applied to forecasting TC tracks and intensity in the Pacific and Atlantic basins (Krishnamurti et al. 1999, 2000; Kumar et al. 2003; Williford et al. 2003) and to climatology (Feng et al. 2011). However, once the models' weighting coefficients have been fixed in the training period by either a linear or nonlinear method, the forecasts will become increasingly inaccurate after a prediction period, because the participating models show seasonal or interseasonal variation, which cannot be completely captured by fixed weights. Based on this view, a running training period weighted super-ensemble technique was introduced by Zhi et al. (2012); also a nonlinear combination model, artificial neural network, was found to lead to a small improvement compared to the super-ensemble with a static training period. Another major challenge faced by the super-ensemble is the continuity of the participating models. Since the

super-ensemble customarily adopts the least-squares method to minimize the distance between observation and forecasts and to calculate the weights of models involved in the ensemble, if each model shows significant changes, e.g., in model configuration and the updating of physical processes between the training and prediction period, the forecast by the super-ensemble would be inferior to the participating model outcomes, because the weighting coefficients may not be able to capture the change even in a running training period. Actually, the models in the TIGGE project do indeed update every 2 or 3 years.

An improved method based on the super-ensemble technique was used in the present study. First, the absolute errors between observation and each model are calculated; then, the weight coefficients are determined according to the relative error size and their time lag relation to forecasts; and lastly, the weighting factors are applied to Eq. (1) for forecasting. With the gradual passage of time, the training period will thereafter run progressively in our algorithm and, accordingly, the weights are dynamically adjusted [see Zhi et al. (2012) for more details]. Approaching the forecast period, the proportion of good-performing models increases, and vice versa. Each forecast step and each variable, such as longitude, latitude and central mean sea level pressure per forecast time, are handled separately:

$$F_{SE} = \overline{O_t} + \sum_{i=1}^{N} \bar{w}_{i,t}(F_{i \cdot t} - \overline{F_i}), \tag{1}$$

where F_{SE} is the forecast of the super-ensemble, $F_{i \cdot t}$ represents the ith model forecast value, $\overline{F_i}$ the mean of the ith model in the running training period, $\bar{w}_{i,t}$ the mean weight of the ith model at time t, $\overline{O_t}$ the mean observed value during the training period, and N is the number of models. During the training period, the weights of each time step satisfy the equation $\sum_{i=1}^{N} w_i = 1$, and according to the lag correlation coefficient, more importance would be given to the several latest samples when calculating the average weight coefficients $\bar{w}_{i,t}$ in the training period. In this way, training samples closer to forecast would gain more importance so that the effect of changing model configuration could be eliminated to some extent (Zhi et al. 2011, 2012; Zhang and Zhi 2015).

However, even with a running training period (sliding window) in the super-ensemble, all observations and forecasts are equally important during this period if no weight is given to the latest information. Actually, the importance of the observed values will rise over time. Therefore, a method that automatically adjusts the weights and adapts model changes can be of great advantage (Shin 2001; Vandenbulcke et al. 2009). This view is somewhat similar to the data assimilation technique: starting from the
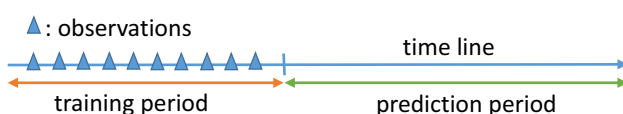

Fig. 1 Schematic representation of the multi-model method of non-equal weights

best guess, weights change during the training period, and thereafter they are instantly updated after the latest observations have been included. And finally, weights remain unchanged and are used for prediction. For another forecast, the cycle begins again. The Kalman Filter has a significant influence on data assimilation and model post-processing, and it has been demonstrated as a powerful tool for eliminating systematic errors (Evensen 1997; Evensen and van Leeuwen 2000). Using data assimilation techniques such as the Kalman Filter and particle filter, Shin and Krishnamurti (2003), Rixen et al. (2009), Vandenbulcke et al. (2009), and Lenartz et al. (2010) introduced dynamically evolving weights in a linear combination of models. This tool can automatically discard old information and dynamically adjust the weights in accordance with the latest information available, and the weights are transformed according to the uncertainty of forecasts and observations.

The Kalman Filter equations can generally be divided into the following two parts:*Forecast:*

$$W_t^f = IW_{t-1}^a, \tag{2}$$

$$P_t^f = IP_{t-1}^a I^T + Q; \tag{3}$$

where superscript f denotes the forecast and a the analysis. W is the state vector, which contains the weights of the models involved in multi-model ensembles. I is the identity matrix, because we have little knowledge about the weight's evolution. In Eq. (2), the latest analysis weights are used as forecast weights, assuming that the combination of member models remains optimal at short term. The weight error covariance matrix is P; and Q is the model error covariance matrix. Initially, both P and Q are diagonal, but non-diagonal elements in P become non-zero with time, whereas Q remains unchanged throughout the process. The sum of the former matrices will approach zero along with the iteration of the algorithm, which shows a sign of convergence. Although initial values in P are random, they had not been given as all-zero to avoid the default W being optimal. A substantial number of trials conducted by us have shown that $Q < 10^{-2}$ provides a relatively optimal result, because if Q is too large, P would never diminish in Eq. (3).*Analysis:*

$$K_t = P_t^f H_t^T (H_t P_t^f H_t^T + R)^{-1}, \tag{4}$$

$$W_t^a = W_t^f + K_t(Y_t - H_t W_t^f), \tag{5}$$

$$P_t^a = P_t^f - K_t H_t P_t^f, \tag{6}$$

Observed values are represented by the vector Y, and H is the observation operator, which contains model forecasts and projects them to Y. Combined with Eq. (8), it can be seen that $Y_t - H_t W_t^f$ yields forecast residual in Eq. (5). Equation (5) is a fundamental formula that connects forecasts and observations, where Kalman gain matrix K is a scale coefficient. Also, K is one of the indicators to measure convergence of the algorithm. R is the observation error covariance matrix, containing instrumental errors and errors caused by the observation operator H. It is quite a difficult task to estimate R, even in a linear system (Kalman 1960), so we empirically take the standard deviation of the observed value as R. Finally, the state vector W is initialized by homogeneous model weights equal to $\frac{1}{N}$.

Typhoon positions can be simply denoted as complex numbers, with the real and imaginary part representing latitude and longitude, respectively. During the experiment, an augmented complex extended Kalman Filter (ACEKF) method was also used (Goh and Mandic 2007; Vandenbulcke et al. 2009). All initial vectors or matrices are expanded in the following form:

$$E^{\text{exp}} = \begin{bmatrix} E^{\text{real}} & 0 \\ 0 & E^{\text{imag}} \end{bmatrix}. \tag{7}$$

Therefore, apart from the length of the observation vector being equal to 2, and the width being double, the rest of the vector lengths are doubled and the widths become 2. As a benefit, the error covariance matrix will evolve as a whole and gradually optimizes.

Finally, the forecast expressions yield

$$F_t = \sum_{i=1}^{N} W_t^f H_t. \tag{8}$$

To summarize, the Kalman Filter was used to allow a dynamic evolution of the model combinations. It was assumed that the error statistics followed a normal distribution; however, owing to a lack of error and model information, we did not test this assumption. Particle Filtering could be another possible way to neglect the Gaussian distribution requirement, but would significantly increase the amount of computation (Rixen et al. 2009).

Non-immediacy of calculated weights, empirical estimates of the observation and model error covariance matrix, and uncertainty of error distribution patterns require a calibration of the final results, especially at the beginning of a typhoon forecast. By studying the error of each model in the training phase, we propose the following relatively simple method of correction: in the forecast period, the range of model forecasts is

$$\Delta d_t = \max(H_t) - \min(H_t) \tag{9}$$

If

$$F_t > \max(H_t) + \Delta d_t \text{ or } F_t < \min(H_t) - \Delta d_t \tag{10}$$

The sum of W will be revised to 1. Although no other mathematical constraint is added to W, we still want the sum of W to be approximately equal to 1, and each component in W to be within the range 0–1. After several trials,

we found that this process effectively reduces the initial TC track forecast uncertainty, decreasing the overall mean absolute track error by about 10 km.

For verification of the single model and evaluation of the multi-model ensemble forecasts, the root-mean-square error was employed:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(F_i - Y_i)^2}, \tag{11}$$

where F_i is the ith sample forecast and Y_i is the corresponding observed value. The RMSE can be decomposed into two terms: (1) the average deviation of the model and the observed (bias); and (2) the overall amplitude and phase disagreement between the variable part of the two variables (unbiased RMSE). The bias and unbiased RMSD are

$$\text{bias} = \frac{1}{n}\sum_{i=1}^{n}|F_i - Y_i|, \tag{12}$$

$$\text{URMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}((F_i - \overline{F_i}) - (Y_i - \overline{Y_i}))^2}, \tag{13}$$

which are linked with RMSE:

$$\text{RMSE}^2 = \text{bias}^2 + \text{URMSE}^2. \tag{14}$$

4 Results

We took TCs, TSs and typhoons from the whole year of 2010 and prior to June 2011 as the training set, and used TCs, TSs and typhoons after June (inclusive) as forecasts. For 2010 and 2011, the western Pacific season had 35 named tropical systems. Due to differences in the length of the data at each center, the 24-, 48- and 72-h lead time forecasts were handled separately; the information is summarized in Table 1.

The ensemble mean, super-ensemble and Kalman Filter approach were all employed to forecast tropical system tracks and intensities in the western North Pacific after June 2011. Figure 2 illustrates the overall forecast skills of the member models, their ensemble mean, super-ensemble, and Kalman Filter. All of these track and intensity forecasts were compared. Note that for all lead times, the meridional deviations (circles) of individual models and multi-model techniques are significantly less than the zonal ones (squares), indicating the ability to forecast tropical systems' west–east trends is relatively weak. Also, for longer lead times, model forecast uncertainties increase. Compared with 24-h predictions, the mean absolute track errors for 72 h are almost doubled (Fig. 2d).

Apparently, the mean absolute errors (MAEs) or RMSEs of the Kalman Filter are less than those of multi-models, the ensemble mean and super-ensemble for a given set of track forecasts during the whole 72 h. In the longer forecast range, the superiority of the Kalman Filter and super-ensemble is more obvious, especially for track forecasts. Therefore, the Kalman Filter can provide more useful and reliable operational guidance than any specific model or other multi-model technique. Comparing the four member models, ECMWF almost always outperforms the others, both in terms of the track and intensity forecast.

The relative improvement in the intensity forecast is less than that of the track forecast. The main reason for this may be that all models adopted in our research are global NWPs, with the highest horizontal grid resolution (in ECMWF) being about 27 km, which is much too coarse to properly resolve the inner core of a mature tropical cyclone. Nevertheless, unlike the ensemble mean and super-ensemble, the Kalman Filter does indeed, to some extent, enhance the intensity forecast both in terms of average intensity deviation (AID) and correlation, with respect to reanalysis data (Fig. 2a–c, shading).

4.1 Track forecasts for typhoons after June 2011

There were 19 named tropical systems after June 2011, including some tropical storms with short lifetimes. Given that the length of the forecast data from each agency is inconsistent, five storms were excluded: Sarika (1103), Tokage (1107), Noru (1113), Banyan (1120), and Washi (1121).

The mean absolute track errors for all of the tropical systems from the participating multi-models and the ensemble mean, super-ensemble and Kalman Filter are summarized in Table 2. To assess the forecast skill, all forecasts from the multi-model technique are homogeneous comparisons under the same conditions and constraints. Similar to Fig. 2, the MAEs of the Kalman Filter are less than those of multi-models, and of the ensemble mean and super-ensemble, during the 3-day forecast. The reduction in MAE compared with the best individual model (ECMWF) is quite substantial—about 100 km for the 72-h forecast. Apart from tropical storm Kulap (1114) and Meari (1105), whose life cycles were relatively short, forecasts from the Kalman Filter regularly outperform those of multi-models, and of the ensemble mean and super-ensemble, significantly improving the forecasting skill and showing stability and consistency. In terms of the weakness in forecasting Kulap and Meari, this is because a convergent process is shown in the Kalman Filter, which takes several steps for K and P to approach zero. Even so, the forecast is still superior to most multi-models. In short, these reductions of MAE provided by the Kalman Filter are encouraging.
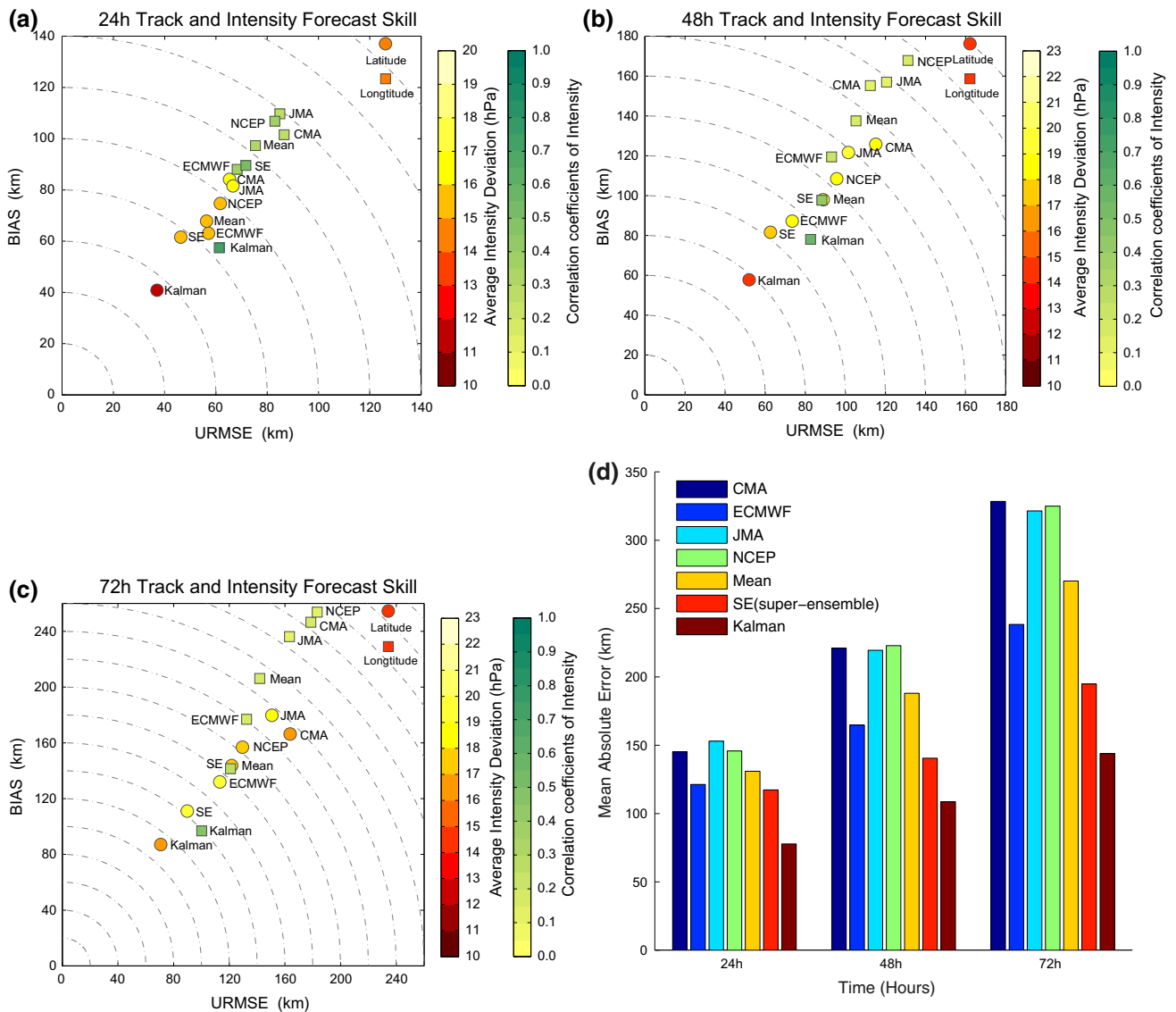
Fig. 2 The track and intensity forecast abilities of the participating forecast models, the ensemble mean, super-ensemble, and Kalman Filter, averaged over the whole forecasting period of 2011. In **a–c**, the average intensity deviation is represented by the *white-to-red* shading, the correlation coefficients between the intensity and verification data by the *yellow-to-green* shading, and the root-mean-square error (RMSE) of the latitude and longitude forecasts by the contours. **d** The mean absolute track errors during the same period

The forecasts obtained from the super-ensemble also significantly outperform the ensemble mean and participating models in most cases. However, sometimes, the prediction from individual models is better than that from the super-ensemble, such as in the forecast for typhoon Nanmado (1111) and Haima (1104). Both cases are of same traits that more than three participating models show similar MAEs. The small variation in MAE between them suggests either limited skill, especially in the longer range, or low correlation with regard to observations, which has a negative effect on the super-ensemble weight coefficients (Williford et al. 2003). Overall, the super-ensemble is more

selective in the bias removal of member models and shows advantage over the equal weight ensemble mean approach.

Although there were 10 typhoons after June 2011, we only illustrate the track forecasts for certain typhoons; namely, Muifa (1109), Talas (1112) and Roke (1115), whose life cycles were longest among the 10 typhoons. The track forecasts for typhoon Muifa are shown in Fig. 3a–c. These figures show the official best track determined by the JTWC in post analysis, along with the forecasts of the best individual model (ECMWF), the ensemble mean, super-ensemble and Kalman Filter forecasts. The Kalman Filter forecast tracks are generally superior to the other models,

Table 2 The mean absolute errors of tropical cyclones (TCs) within the 72-h forecasts during the whole forecasting period of 2011

Name	ID	Leading time	Forecast life	CMA	ECMWF	JMA	NCEP	Ensemble mean	Super-ensemble	Kalman Filter
Average	–	24	–	145.2	121.3	*153.0*	145.8	131.0	117.3	**77.8**
		48		221.0	164.7	219.5	*222.9*	188.0	140.4	**108.6**
		72		*328.4*	238.4	321.5	325.0	270.1	194.8	**143.9**
Haima (TS)	1004	24	4	168.9	129.7	149.6	162.7	146.7	*222.1*	**89.9**
Meari (TS)	1005	24	4	192.8	173.9	199.4	*205.8*	184.8	161.2	**146.3**
		48	3	249.1	**186.3**	*298.2*	269.0	241.8	193.6	216.2
		72	2	240.9	227.6	*469.1*	378.7	279.1	**210.4**	272.4
Ma-on (T)	1106	24	10	139.7	117.1	*151.3*	145.2	126.4	119.9	**82.6**
		48	9	197.3	151.7	*200.8*	190.3	168.0	140.3	**111.3**
		72	8	*330.6*	202.6	268.3	246.6	223.0	187.6	**151.5**
Nock-ten (TS)	1108	24	4	168.0	131.4	*177.8*	174.9	156.1	139.0	**70.1**
		48	4	268.8	213.2	*352.2*	305.5	281.1	137.2	**131.7**
		72	3	371.9	278.7	*502.9*	474.6	401.0	220.0	**154.0**
Muifa (T)	1109	24	11	96.7	106.7	123.3	*127.9*	104.4	93.6	**54.6**
		48	10	184.6	164.6	*192.0*	191.5	150.4	120.9	**79.1**
		72	9	*356.2*	238.8	274.4	275.5	229.7	173.2	**125.1**
Merbok (TS)	1110	24	6	178.5	142.1	*191.3*	183.7	165.3	94.4	**83.6**
		48	5	*351.2*	210.1	305.8	250.9	263.1	162.9	**106.2**
		72	4	*547.8*	350.7	492.2	349.1	388.4	212.6	**203.5**
Nanmado (T)	1111	24	7	86.6	71.3	*115.6*	78.8	75.9	109.2	**45.1**
		48	6	152.8	138.9	*223.0*	176.0	148.0	163.8	**84.3**
		72	5	150.5	230.5	*340.9*	313.2	231.5	244.0	**117.1**
Talas (T)	1112	24	8	80.4	62.9	*95.9*	86.5	73.2	70.7	**53.2**
		48	7	111.0	77.3	99.1	*125.2*	93.0	79.6	**70.3**
		72	7	169.6	123.0	*198.4*	182.2	145.9	132.3	**108.3**
Kulap (TS)	1114	24	2	*205.5*	115.5	191.3	168.1	151.9	**82.4**	121.1
		48	2	*374.2*	**145.7**	186.0	208.2	189.1	150.2	181.2
Roke (T)	1115	24	8	130.6	119.7	*145.6*	127.0	120.9	117.1	**68.7**
		48	6	219.0	165.4	215.8	*225.9*	196.8	126.9	**86.4**
		72	6	*384.5*	304.4	350.3	367.0	343.3	202.0	**130.6**
Sonca (T)	1116	24	4	*261.6*	225.7	255.1	217.5	235.7	**193.7**	196.1
		48	3	255.0	241.9	237.5	*274.3*	224.1	231.1	**193.4**
Nesat (T)	1117	24	6	203.1	140.9	190.5	194.3	176.8	146.2	**67.4**
		48	5	262.5	161.8	272.1	*329.3*	253.4	170.5	**90.5**
		72	4	425.0	243.7	397.5	*492.0*	384.1	257.2	**111.2**
Haitang (TS)	1118	24	2	*177.7*	69.1	48.9	37.9	52.4	128.7	**28.8**
Nalgae (T)	1119	24	7	147.6	145.2	159.2	*168.4*	139.0	75.0	**67.5**
		48	6	261.2	203.5	201.5	*275.0*	204.6	117.5	**128.6**
		72	5	374.0	281.6	265.3	*455.4*	295.9	192.9	**178.9**

Values in bold (italics) are minimum (maximum) errors

TS tropical storm, *T* typhoon

especially for the 24- and 48-h forecasts. For the 72-h forecast, all the other models (CMA, JMA and NCEP are not shown in the figure, but are reflected by the tracks predicted by the ensemble mean) except the Kalman Filter predict frequent landfall in the provinces of Fujian, Zhejiang, Jiangsu, Shandong and Liaoning. Any false alarms would affect people in these regions in that they would be told to evacuate their homes. However, even though the Kalman Filter forecasts are more reliable in terms of landfall position, one of the notable problems during the 72-h forecast is that the after-landfall typhoon forecasts show an opposing tendency of moving track on the last day. Much effort has
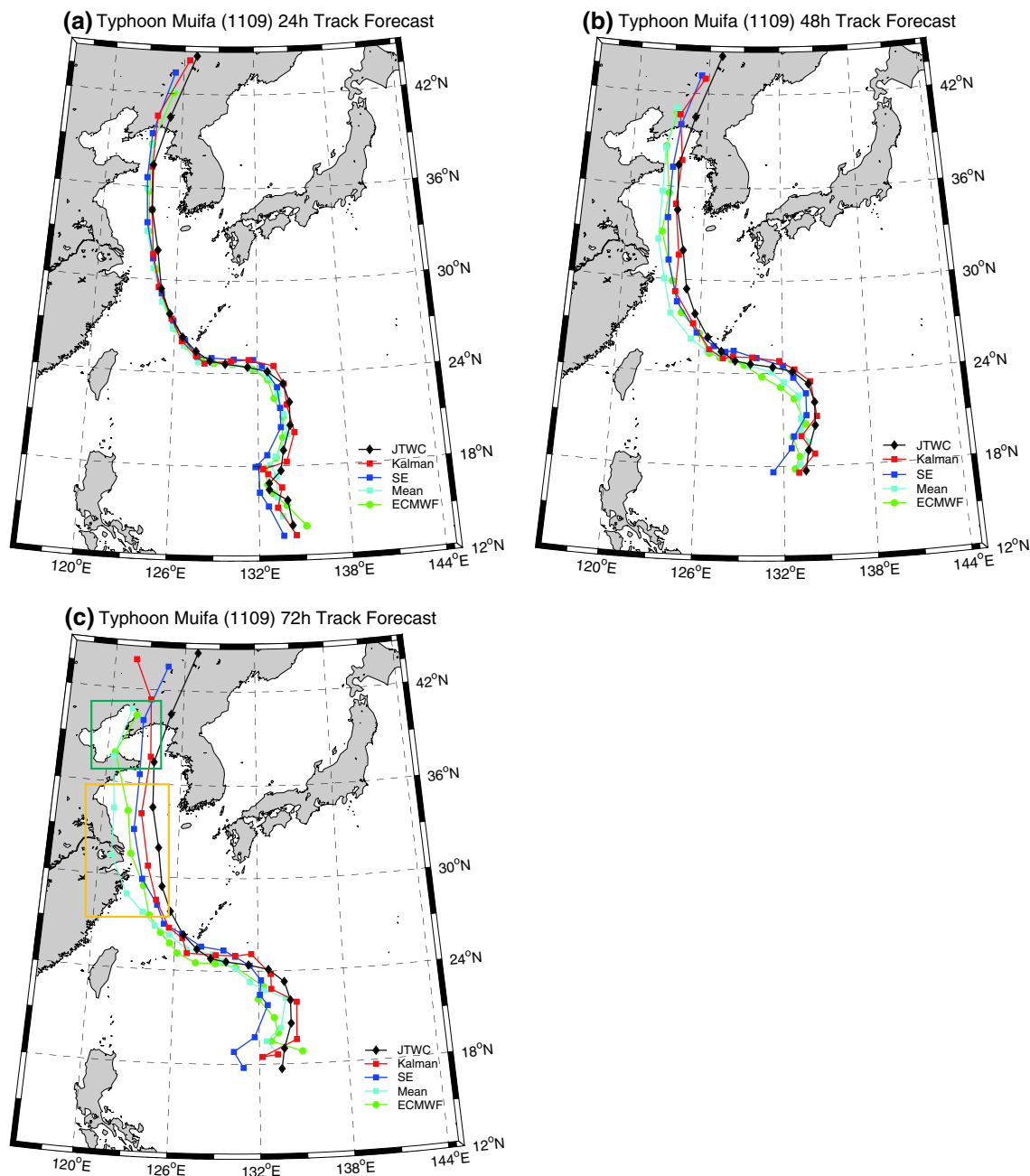
Fig. 3 The **a** 24-, **b** 48- and **c** 72-h track forecasts for typhoon Muifa from the best individual model (that of the European Centre for Medium-Range Weather Forecasts, ECMWF), the ensemble mean (mean), super-ensemble (SE), and Kalman Filter (Kalman). Initial conditions: 1200 UTC 29 Jul 2011, 1200 UTC 31 Jul 2011 and 1200 UTC 31 Jul 2011 for the 24-, 48- and 72-h forecasts, respectively

been invested to identify the reason. Even though the track forecasts from other member models were over land (not shown in Fig. 3c, but apparent from track of the ensemble mean in the yellow box) while the typhoon was still over the ocean, the forecasts from ECMWF are much more accurate than the others. Therefore, the Kalman Filter could take advantage of the forecasts from ECMWF (its weight is dominant in the algorithm) and improve the prediction. However, when the track reaches above 36°N, all forecasts converge to Liaoning Province, far from the actual landfall position of Muifa (not shown, but apparent from the track of the ensemble mean in the green box). That is, small variations of longitude forecast between the models result in the Kalman Filter forecasts moving westward, which can intercept the opposite tendency of the observed track. Nevertheless, the precision of the latitudinal forecast being much more accurate than all the other models should be emphasized. The forecasts from all multi-models are far from the
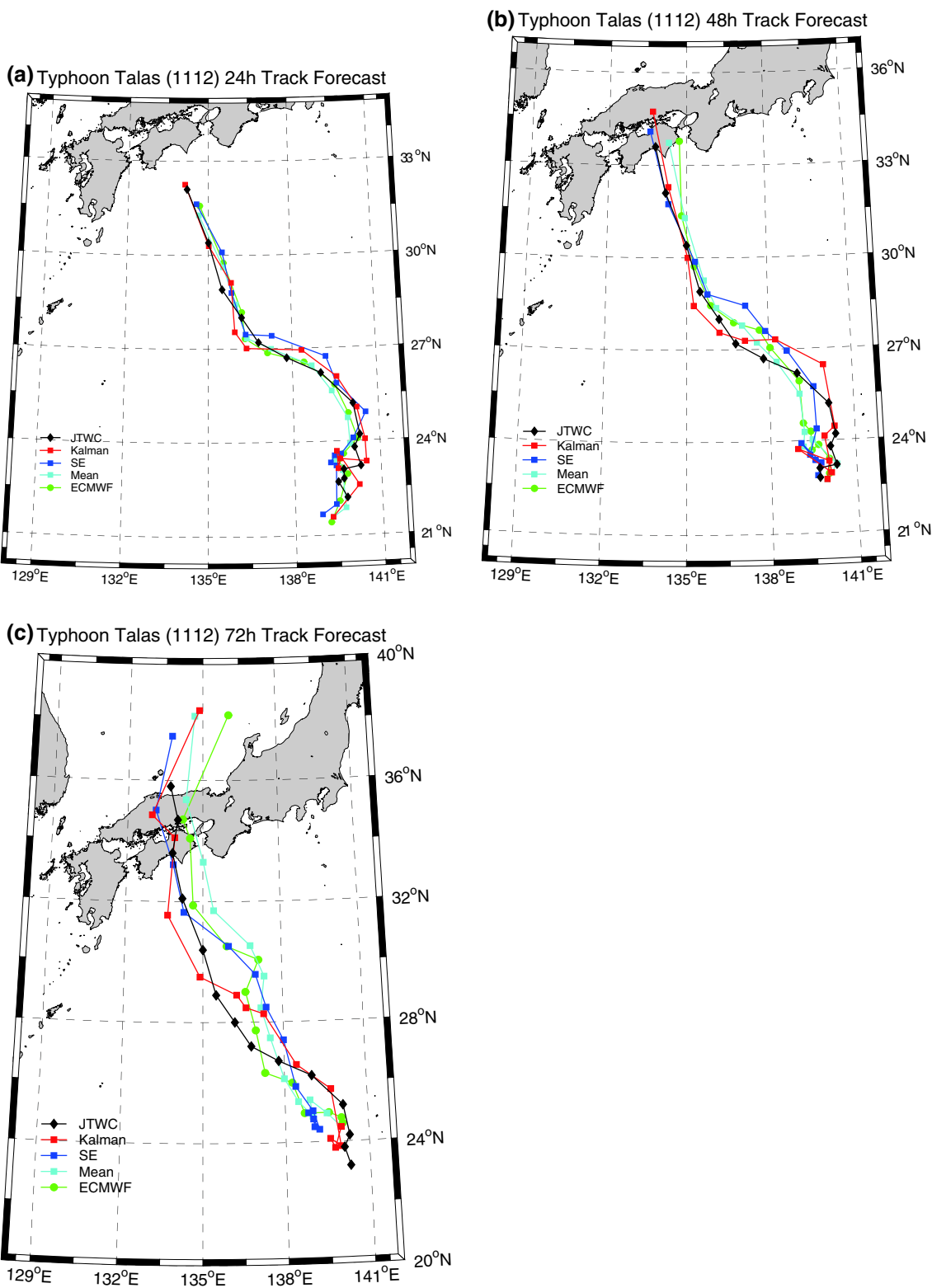
Fig. 4 As in Fig. 3, but for typhoon Talas and initial conditions of 1200 UTC 25 Aug 2011, 1200 UTC 27 Aug 2011 and 1200 UTC 28 Aug 2011 for the 24-, 48- and 72-h forecast, respectively
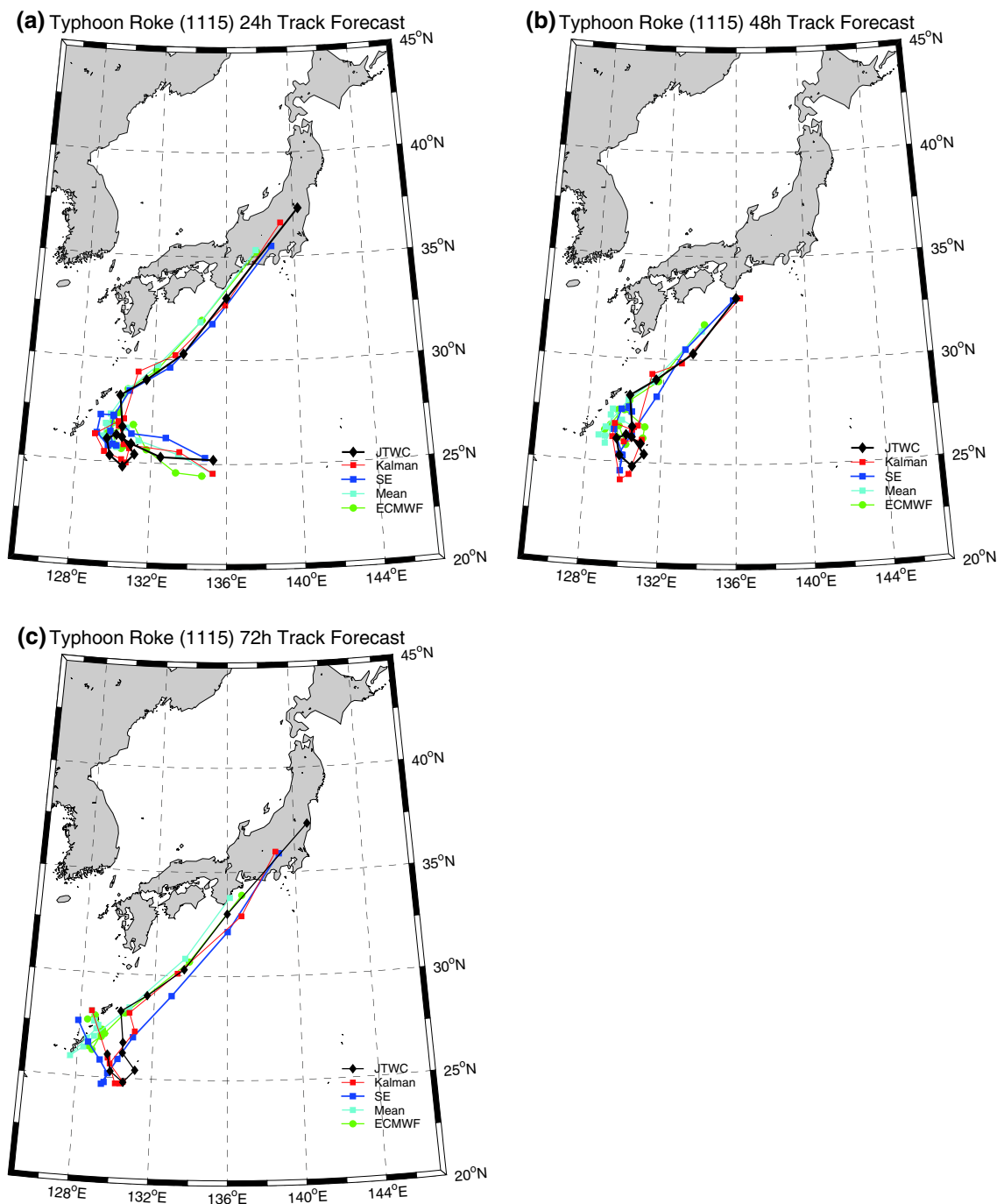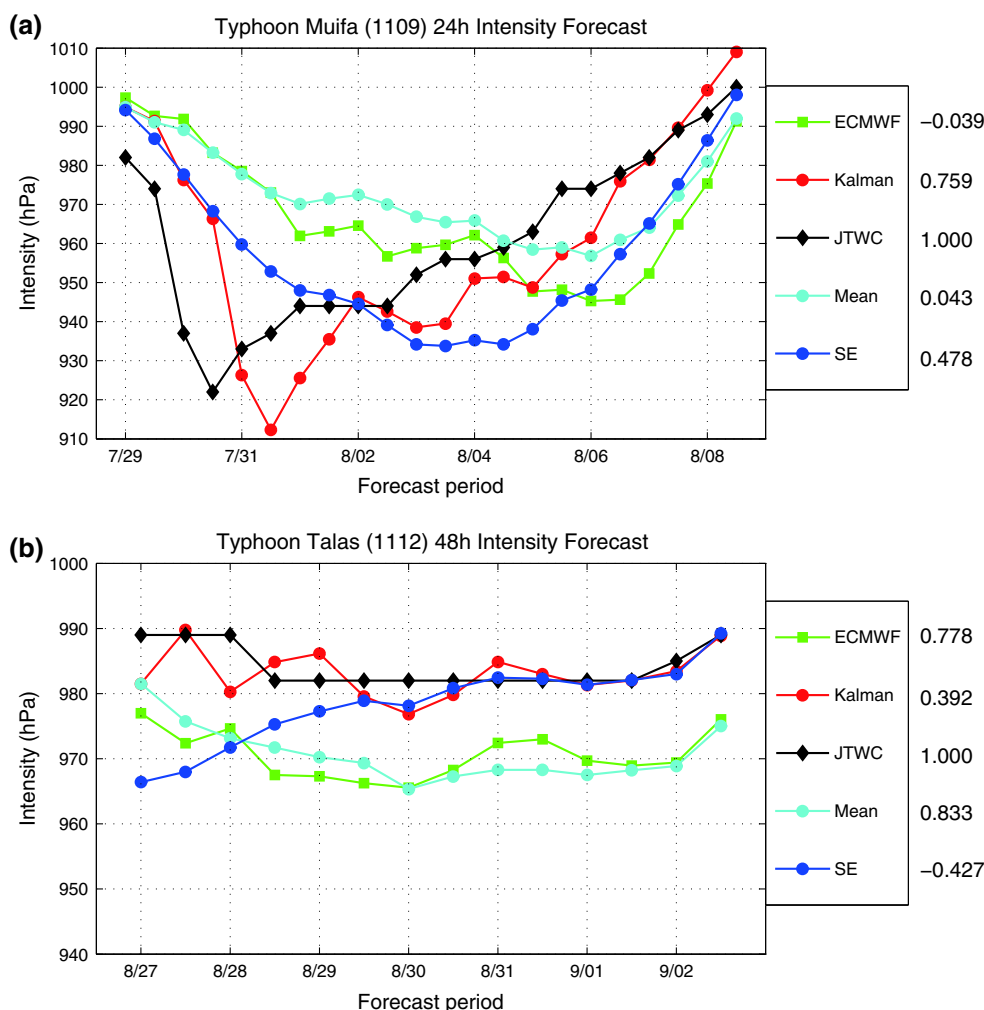
Fig. 5 As in Fig. 3, but for typhoon Roke and initial conditions of 1200 UTC 14 Sep 2011, 1200 UTC 15 Sep 2011 and 1200 UTC 16 Sep 2011 for the 24-, 48- and 72-h forecasts, respectively

landfall position of Muifa on the last day, but the super-ensemble still produces a correct forecast. This is because the super-ensemble takes advantage of the mean observed value $\overline{O_t}$ during the training period to adjust the track forecast, while the Kalman Filter only makes use of the latest observation and forecasts, which are unreliable up to 36°N.

The predicted and reanalysis best tracks for typhoon Talas and Roke are illustrated in Figs. 4 and 5. Again, we note the overall superior performance of the Kalman Filter. However, the 72-h forecast for Talas is not impressive for any of the models or the multi-model techniques. Given that it has a circular track, and all models predict tracks far behind the reanalysis best track,

(a) Typhoon Muifa (1109) 24h Intensity Forecast

ECMWF	−0.039
Kalman	0.759
JTWC	1.000
Mean	0.043
SE	0.478

(b) Typhoon Talas (1112) 48h Intensity Forecast

ECMWF	0.778
Kalman	0.392
JTWC	1.000
Mean	0.833
SE	−0.427

forecasting Roke offers a major challenge for all participating models. As the comparison in Fig. 5 shows, the forecasts of Roke by the Kalman Filter demonstrate superiority compared to the ECMWF- and ensemble mean-predicted tracks—a result that is especially noteworthy in the 72-h forecasts (Fig. 5c).

Figures 3, 4 and 5 also indicate that the early typhoon forecasts from the Kalman Filter are relatively less accurate compared to its later forecasts. Initially, feeding new TC, TS or typhoon information derived from individual models as inputs into the Kalman Filter, a new iteration cycle begins. However, the algorithm cannot converge within two or three steps, so the early forecasts are relatively less accurate, as demonstrated by the forecasts of Talas and Roke. When the typhoon moving track is approximately a straight line or in a stable state, the Kalman Filter is able to accurately predict the typhoon moving process. Even when the typhoon turns sharply, the Kalman Filter is still superior to the other approaches (Fig. 5).

4.2 Intensity forecasts for key typhoons

Intensity prediction has always been a difficult issue for tropical NWP (Williford et al. 2003; Kumar et al. 2003; DeMaria et al. 2014). As Fig. 1a–c show, a slight improvement is apparent in the 72-h forecast from the Kalman Filter. The biases are reduced by about 3–5 hPa compared to the member models, and the mean correlation coefficients increase by about 0.2. The intensity forecasts of Muifa and Talas are shown in Fig. 6. Similar to the track prediction, the central mean sea level pressure in the typhoon forecasts by the Kalman Filter produces relatively smaller errors in comparison to the member models and other multi-model techniques. Figure 6a shows that the Kalman Filter forecasts are highly correlated with the re-analysis data (see also Fig. 1a–c), but with small fluctuations during the forecast period. In contrast to Fig. 6a, the forecasts from both the super-ensemble and Kalman Filter are poorly correlated with JTWC data in Fig. 6b. Although the forecasts from the member models and

ensemble mean exhibit higher correlation, they all over-estimate the intensity of Talas by almost 15 hPa.

Overall, the improvement in the intensity forecast is less than that of the track forecast. Since global NWPs are too coarse to resolve the inner core of a mature tropical cyclone, intensity forecasts should be carried out more directly using higher resolution models.

5 Concluding remarks and future outlook

Three approaches (the ensemble mean, super-ensemble and Kalman Filter) were used to predict TCs' tracks and intensities in the western Pacific region based on the TIGGE data during 2010 and 2011. The differences between forecasts and observations showed that the Kalman Filter technique, as a dynamically adapting multi-model method, performed best, considerably reducing the forecast error of TC track and intensity. Quantitatively, compared to the best single model (ECMWF), the MAEs of the 24-, 48- and 72-h track forecasts were reduced by roughly 50, 80 and 100 km, respectively, and the AIDs were also reduced, by 3–5 hPa. In addition, in comparison to the other two methods, the track and intensity forecasts measured by central mean sea pressure also improved.

Given the low number of TC samples, and the fact that multi-models often undergo significant updates, a cross-validation approach in which all other TCs are used to determine the weighting coefficients, except the one being predicted, is frequently used in super-ensemble TC forecasting. However, the cross-validation technique disorders the chronology of the typhoons and, therefore, this method is difficult to apply operationally. Several measures can be employed to solve this problem to a certain extent, such as using principal component analysis and stepwise regression to select participating models more objectively.

However, the Kalman Filter is easy to apply operationally, even though its early forecasts may be relatively inaccurate compared to its later predictive skill. More investments in participating models and the observation-error covariance matrix may help to speed up the convergence rate of the algorithm and thus improve the forecast performance in the early stages of its application. Although the Kalman Filter technique shows considerable advantages for the whole 72-h forecast, data shortages make it impossible to provide forecasts for the 120-h and longer lead times. Using the central mean sea pressure as an indicator of TC or typhoon strength is also not entirely satisfying. Still, it is a promising multi-model technique for improving operational TC track and intensity forecasting, and this method can also be applied to other meteorological variables or fields, such as temperature, precipitation and winds.

References

DeMaria M, Sampson CR, Knaff JA, Musgrave KD (2014) Is tropical cyclone intensity guidance improving? Bull Am Meteorol Soc 95:387–398

Evensen G (1997) Advanced data assimilation for strongly nonlinear dynamics. Mon Weather Rev 125:1342–1354

Evensen G, van Leeuwen PJ (2000) An ensemble Kalman smoother for nonlinear dynamics. Mon Weather Rev 128:1852–1867

Feng JM et al (2011) Comparison of four ensemble methods combining regional climate simulations over Asia. Meteorol Atmos Phy 111:41–53

Franklin JL (2012) National Hurricane Center forecast verification report. http://www.nhc.noaa.gov/verification/pdfs/Verification_2012.pdf

Goerss JS (2000) Tropical cyclone track forecasts using an ensemble of dynamical models. Mon Weather Rev 128:1187–1193

Goh SL, Mandic DP (2007) An augmented extended Kalman filter algorithm for complex-valued recurrent neural networks. Neural Comput 19:1039–1055

Grijn GVD (2002) Tropical cyclone forecasting at ECMWF: new products and validation. ECMWF Tech, Mem **386**

Krishnamurti TN et al (1999) Improved weather and seasonal climate forecasts from multimodel superensemble. Science 285:1548–1550

Krishnamurti TN, Kishtawal CM, Shin DW, Williford CE (2000) Improving tropical precipitation forecasts from a multianalysis superensemble. J Clim 13:4217–4227

Kumar T, Krishnamurti TN, Fiorino M, Nagata M (2003) Multimodel superensemble forecasting of tropical cyclones in the Pacific. Mon Weather Rev 131:574–583

Lenartz F, Mourre B, Barth A, Beckers JM, Vandenbulcke L, Rixen M (2010) Enhanced ocean temperature forecast skills through 3-D super-ensemble multi-model fusion. Geophy Res Lett 37:5

Leslie LM, Fraedrich K (1990) Reduction of tropical cyclone position errors using an optimal combination of independent forecasts. Weather Forecast 5:158–161

Libonati R, Trigo I, DaCamara CC (2008) Correction of 2 m-temperature forecasts using Kalman Filtering technique. Atmos Res 87:183–197

Marchok TP (2002) How the NCEP tropical cyclone tracker works. AMS conference on hurricanes and tropical meteorology

McCollor D, Stull R (2008) Hydrometeorological accuracy enhancement via postprocessing of numerical weather forecasts in complex terrain. Weather Forecast 23:131–144

Qi LB, Yu H, Chen PY (2014) Selective ensemble-mean technique for tropical cyclone track forecast by using ensemble prediction systems. Q J R Meteorol Soc 140:805–813

Rixen M et al (2009) Improved ocean prediction skill and reduced uncertainty in the coastal region from multi-model super-ensembles. J Marine Syst 78:S282–S289

Shin DW (2001) Short- to medium-range superensemble precipitation forecasts using satellite products. Florida State University, p 35

Shin DW, Krishnamurti TN (2003) Short- to medium-range superensemble precipitation forecasts using satellite products: 1 deterministic forecasting. J Geophys Res Atmos 108:18

Tien TT, Thanh C, Van HT, Chanh KQ (2012) Two-dimensional retrieval of typhoon tracks from an ensemble of multimodel outputs. Weather Forecast 27:451–461

Vandenbulcke L et al (2009) Super-ensemble techniques: application to surface drift prediction. Prog Oceanogr 82:149–167

Williford CE, Krishnamurti TN, Torres RC, Cocke S, Christidis Z, Kumar TSV (2003) Real-time multimodel superensemble forecasts of Atlantic tropical systems of 1999. Mon Weather Rev 131:1878–1894

WMO (2008a) Guidelines on communicating forecast uncertainty. http://www.wmo.int/pages/prog/amp/pwsp/documents/TD-1422.pdf

Yamaguchi M (2013) Tropical cyclone predictability. http://www.ecmwf.int/newsevents/training/meteorological_presentations/2013/PR2013/Yamaguchi/Yamaguchi.pdf

Zhang L, Zhi XF (2015) Multimodel consensus forecasting of low temperature and icy weather over central and southern China in early 2008. J Trop Meteorol 121(1):67–75

Zhi XF, Zhang L, Bai Y (2011) Application of the multi-model ensemble forecast in the QPF. In: Information Science and Technology (ICIST), 2011 international conference on, IEEE, pp 657–660

Zhi XF, Qi HX, Bai YQ, Lin CZ (2012) A comparison of three kinds of multimodel ensemble forecast techniques based on the TIGGE data. Acta Meteorologica Sinica 26:41–51