Dependence of Probabilistic Quantitative Precipitation Forecast Performance on Typhoon Characteristics and Forecast Track Error in Taiwan

HSU-FENG TENG AND JAMES M. DONE

National Center for Atmospheric Research, Boulder, Colorado

CHENG-SHANG LEE

Department of Atmospheric Sciences, National Taiwan University, Taipei, Taiwan

YING-HWA KUO

National Center for Atmospheric Research, and University Corporation for Atmospheric Research, Boulder, Colorado

(Manuscript received 15 August 2019, in final form 7 January 2020)

ABSTRACT

This study investigates the probabilistic quantitative precipitation forecast (PQPF) performance of typhoons that affected Taiwan during 2011–16. In this period, a total of 19 typhoons with a land warning issued by the Central Weather Bureau (CWB) are analyzed. The PQPF is calculated using the ensemble precipitation forecast data from the Taiwan Cooperative Precipitation Ensemble Forecast Experiment (TAPEX), and the verification data, verification thresholds, and typhoon characteristics are obtained from the CWB. The overall PQPF performance of TAPEX has an acceptable reliability and discrimination ability, and the higher probability error is distributed at the mountainous area of Taiwan. The PQPF performance is significantly influenced by typhoon characteristics (e.g., typhoon tracks, sizes, and forward speeds). The PQPFs for westward-moving, large, or slow typhoons have higher reliability and discrimination ability, and lower-probability error than those for northward-moving, small, or fast typhoons, except for similar reliability between fast and slow typhoons. Because northward-moving or small typhoons have larger forecast track error, and their PQPF performance is sensitive to the accuracy of the forecast track, a higher probability error occurs than that for westward-moving or large typhoons. Furthermore, because there is no difference in track error between fast and slow typhoons, the larger track spread for slow typhoons increases the rainfall forecast spread and reduces the probability error. The orientation of Taiwan's topography and the topographic effect also influence and increase the distribution and value of probability error for northward-moving, small, or fast typhoons. In summary, forecast track characteristics are influenced by typhoon characteristics and further affect the PQPF performance.

1. Introduction

Typhoon forecast techniques are critical to rainfall forecast and disaster management in real-time operations. Forecasts of typhoon rainfall spatial distribution, duration, and intensity are important for saving lives and avoiding economic loss. Currently, most operational forecast models can provide typhoon rainfall forecasts with some skill. However, to improve the forecast product without model upgrades, the bias correction, uncertainty assessment, and forecast postprocess guidance deserve to be further studied. This study uses the probability perspective to characterize and understand rainfall forecast characteristics, error distributions, and mechanisms, for different types of typhoons.

Taiwan is a relatively narrow island (400 km in length and 150 km in width) off the East Asian coast with a high mountain range (higher than 3000 m). On average, 3–4 typhoons affect Taiwan each year, typically bringing 600–800 mm of rainfall (accounting for 27.6% of annual rainfall) and a direct loss of more than 570 million U.S. dollars (Wu and Kuo 1999; Lee et al. 2006, 2012a; Huang and Wang 2015). Therefore, improving the quantitative precipitation forecast (QPF) skill for typhoons is essential for Taiwan and the eastern Asian coast. With the consideration of physical processes and convective

Corresponding author: Hsu-Feng Teng, tenghsufeng@gmail.com

DOI: 10.1175/WAF-D-19-0175.1

features, many previous studies used numerical models to simulate typhoon rainfall, to understand the mechanisms of heavy rainfall from typhoons (e.g., Kuo and Wang 1997; Wu et al. 2002, 2009; Yang and Tung 2003; Chien et al. 2008; Yang et al. 2008; Lee et al. 2008; Wang et al. 2013; Wang 2015). Due to the interaction between typhoon circulation and the topography of Taiwan, the phase-locked effects of typhoon rainfall were used to construct statistical rainfall models and improve the rainfall forecasts (Chang et al. 1993; Lin et al. 2001; Chiao and Lin 2003; Lee et al. 2006, 2013). Data assimilation has also been used to improve QPF skill (Jian et al. 2003; Huang et al. 2005; Jian and McGinley 2005; Hsiao et al. 2010; Yen et al. 2011; Chang et al. 2014; Tsai et al. 2014; Hendricks et al. 2016). QPF using ensemble numerical modeling was developed specifically to account for the stochastic nature of these moist convective precipitation processes. Multiple models, physics schemes, initial times, and initial data have been used to create ensemble members (Fang et al. 2011; Hsiao et al. 2013; Chen and Wu 2016; Wang et al. 2016; Yang et al. 2016). Ensemble members were further classified based on specific typhoon features to improve the QPF performance of typhoons (Hong et al. 2015).

The probabilistic quantitative precipitation forecast (PQPF) has been used as an objective parameter to assess and understand the overall performance of ensemble systems and also to develop postprocess bias corrections. For example, Ruiz et al. (2009) calculated PQPF over South America using several ensemble forecast systems, and improved PQPF skill using static and dynamic calibrations. Similarly, Liu and Xie (2014) established PQPF over the Huaihe basin, China using the Observing System Research and Predictability Experiment Interactive Grand Global Ensemble and calibrated the raw PQPF using a Bayesian model averaging method. In addition, Chang et al. (2012) developed a short-range PQPF in Taiwan using the Local Analysis and Prediction System and showed that a clear wet bias in the ensemble system could be corrected by a linear regression calibration procedure. Fang and Kuo (2013) applied PQPF verification results to correct bias in topographically enhanced typhoon heavy rainfall in Taiwan. Yuan et al. (2008) explored the performance of QPF and PQPF for intensive operation periods in 2006 using multiple models and corrected the probabilistic bias using an artificial neural network and linear regression methods. Furthermore, Buizza et al. (1998, 1999) and Yuan et al. (2009) showed that a larger ensemble size and more varied models would partly improve PQPF. PQPF has also been shown to add potential economic value over deterministic forecasts (Chang et al. 2015).

Previous studies suggested that typhoon track, size, and forward speed all play critical roles for typhoon rainfall in Taiwan. Because typhoons with different track types produce significantly different rainfall patterns caused by the topographic effect, the typhoon track forecast is therefore a primary factor for operational rainfall forecast in Taiwan (Lin et al. 2002; Lee et al. 2006, 2013; Su et al. 2012; Hong et al. 2015; Wu et al. 2016). Typhoon size also influences rainfall distribution (Lin and Jeng 2000; Lee et al. 2006, 2013; Ge et al. 2010; Hong et al. 2015), and the structural changes of typhoons also present challenges for QPF (Lee et al. 2012b; Wang 2015; Chen et al. 2016; Huang et al. 2016a). Typhoon forward speed is also a factor in rainfall totals (Lonfat et al. 2004; Ge et al. 2010; Yen et al. 2011; Su et al. 2012; Chang et al. 2013; Hsu et al. 2013). Conversely, although typhoon intensity affects the local damage (Emanuel 2005), it does not appear to be correlated with typhoon rainfall for Taiwan and other areas (Lonfat et al. 2004; Jiang and Halverson 2008; Su et al. 2012; Chang et al. 2013; Yu et al. 2017). We therefore expect probabilistic rainfall forecasts to be influenced strongly by typhoon characteristics, but less so by typhoon intensity.

However, no previous study systematically analyzed the PQPF for typhoons with different characteristics and evaluated the impact of each characteristic on probability error distribution. Given that Taiwan Island is affected by typhoons with a broad range of track orientations, sizes, and forward speeds and has high-resolution rainfall observations, the purpose of this study is to construct the first guidance for typhoon probabilistic rainfall forecasting, quantify the distribution of probabilistic forecast bias, and understand the role of different typhoon characteristics on PQPF skill.

The ensemble forecast, observation data, PQPF method, and verification methods are introduced in section 2. PQPF reliability and discrimination ability between different types of typhoons are compared in section 3. Then the distributions of probability forecast error between different types of typhoons are compared in section 4. The relationships between PQPF performance and each typhoon characteristic and the underlying mechanisms are discussed in section 5. Finally, discussion and conclusions are provided in section 6.

2. Data and methodology

a. Data

Three datasets are used in this study. The ensemble precipitation forecast from mid-August 2011 to the

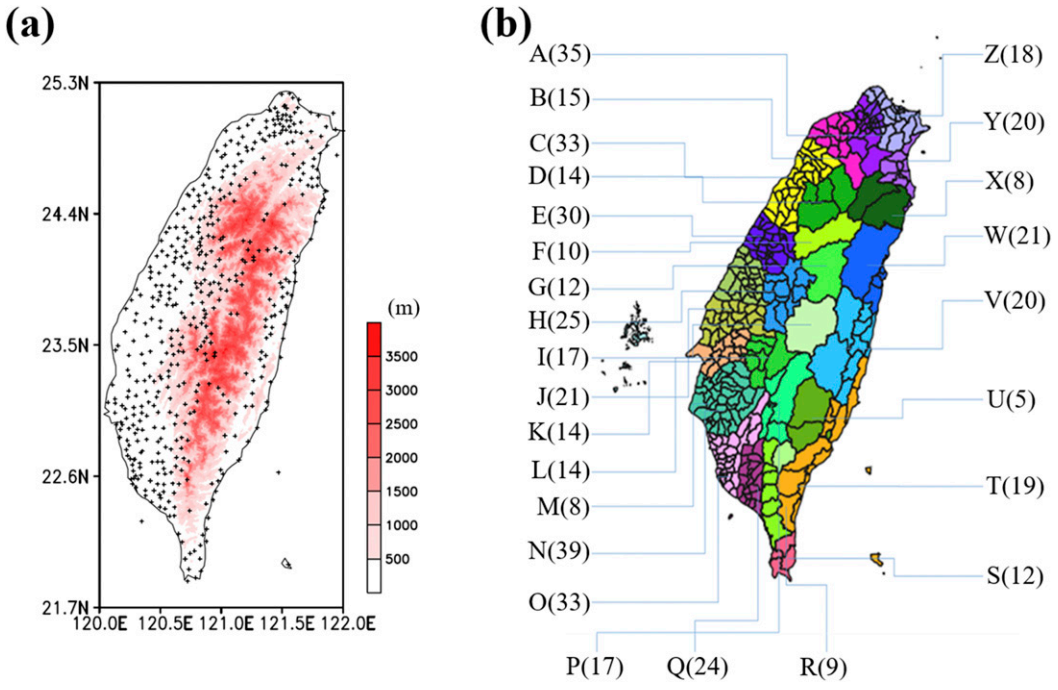

FIG. 1. (a) The topography of Taiwan (colored shading; m) and the location of rain gauge stations used in this study (black plus signs). (b) The 26 regions (labeled A–Z) used to analyze the spatial distribution of probability error. The number in parentheses is the number of rain gauge stations in each region.

end of 2016 is obtained from the Taiwan Cooperative Precipitation Ensemble Forecast Experiment (TAPEX; Hsiao et al. 2013) supported by the Taiwan Typhoon and Flood Research Institute (TTFRI). A total of 495 ensemble forecasts for 19 typhoons are analyzed. There are 20–30 ensemble members in TAPEX, which comprise multiple numerical models [the fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model (MM5) and Weather Research and Forecasting (WRF) Model; Grell et al. 1994; Skamarock et al. 2008], cumulus physics schemes, microphysics schemes, boundary layer schemes, and assimilation systems and strategies. Each ensemble member initializes at 0000, 0600, 1200, and 1800 UTC and provides a 72-h forecast. When the typhoon approaches Taiwan, more ensemble members are added in, which leads to the members to increase from about 20 to 30. Because each member is approximately added in at a similar time relative to the typhoon warning announcement (i.e., similar member number at the relative time among typhoons), this does not impact the comparison of verifications between different types of typhoons. The forecast rainfall from each ensemble member is hourly at 5-km resolution. The 24-h accumulated rainfall forecast is the sum of the hourly forecast rainfall for 24 h, which is calculated every hour from 0 to 48 h in each forecast. The detailed settings of TAPEX are described in Lee et al. (2012a) and Hsiao et al. (2013).

The observed precipitation data are hourly precipitation measurements from the ground rain gauge network supported by the Central Weather Bureau (CWB; https://e-service.cwb.gov.tw/HistoryDataQuery). To verify the ensemble forecast, observations from 493 rain gauge stations are considered (Fig. 1a). To show the spatial gridded distribution of precipitation, all rain gauge station data are interpolated based on the Cressman objective analysis system (Cressman 1959; Lee et al. 2006). Finally, typhoon track, size, and forward speed information is obtained from the CWB best track (https://rdc28.cwb.gov.tw). Moreover, to enable statistical significance testing of PQPF 19 typhoons for which CWB issued a land warning during 2011–16 and made landfall or brought significant rainfall over Taiwan are selected (Table 1 and Fig. 2a). Note that Typhoon Tembin (2012) affected Taiwan twice in its lifetime; therefore, 20 land warning cases are analyzed.

b. PQPF and verification period

In this study, the ensemble forecast of 24-h accumulated rainfall in TAPEX is translated into PQPF for six

TABLE 1. List of typhoon characteristics with a land warning issued by CWB between August 2011 and December 2016.

No.	Year	Name	Land warning begin	Land warning end	Track type	R14 (km)	Forward speed (km h^{-1})	Maximum wind speed (m s^{-1})
1	2011	Nanmadol	1230 UTC 27 Aug 2011	0030 UTC 31 Aug 2011	W	180	9.2	53
2	2012	Talim	1230 UTC 19 Jun 2012	1830 UTC 20 Jun 2012	N	150	27.9	25
3	2012	Saola	1230 UTC 31 Jul 2012	0630 UTC 3 Aug 2012	W	220	13.7	38
4	2012	Tembin	2130 UTC 21 Aug 2012	0030 UTC 25 Aug 2012	W	180	10.4	45
5	2012	Tembin	1830 UTC 26 Aug 2012	1230 UTC 28 Aug 2012	N	180	17.9	35
6	2013	Soulik	1230 UTC 11 Jul 2013	1530 UTC 13 Jul 2013	W	280	24.1	51
7	2013	Trami	1230 UTC 20 Aug 2013	0030 UTC 22 Aug 2013	W	180	27.1	30
8	2013	Kong-Rey	0330 UTC 28 Aug 2013	0930 UTC 29 Aug 2013	N	120	17.1	25
9	2013	Usagi	0030 UTC 20 Sep 2013	0030 UTC 22 Sep 2013	W	280	18.8	55
10	2013	Fitow	0630 UTC 5 Oct 2013	0030 UTC 7 Oct 2013	W	250	20.2	38
11	2014	Matmo	1830 UTC 21 Jul 2014	1530 UTC 23 Jul 2014	W	200	21.9	38
12	2014	Fung-Wong	1230 UTC 19 Sep 2014	2130 UTC 21 Sep 2014	N	150	20.8	25
13	2015	Linfa	1830 UTC 6 Jul 2015	0630 UTC 7 Jul 2015	N	120	8.4	30
14	2015	Chan-Hom	1230 UTC 9 Jul 2015	1530 UTC 10 Jul 2015	W	280	20.9	48
15	2015	Soudelor	1230 UTC 6 Aug 2015	0030 UTC 9 Aug 2015	W	300	22.2	48
16	2015	Dujuan	0930 UTC 27 Sep 2015	0930 UTC 29 Sep 2015	W	220	20.1	51
17	2016	Nepartak	1230 UTC 6 Jul 2016	0630 UTC 9 Jul 2016	W	200	17.5	58
18	2016	Meranti	0030 UTC 13 Sep 2016	0330 UTC 15 Sep 2016	N	220	23.0	60
19	2016	Malakas	0030 UTC 16 Sep 2016	1830 UTC 17 Sep 2016	N	180	18.5	45
20	2016	Megi	0330 UTC 26 Sep 2016	0930 UTC 28 Sep 2016	W	250	22.7	45

magnitudes: 10, 25, 50, 130, 200, and 350 mm, which are the rainfall categories and disaster prevention standards of the CWB. The PQPF is calculated for each grid point and is defined as follows:

$$\text{PQPF} = \frac{\text{number of members reaching the specific rainfall threshold}}{\text{number of total members}}. \quad (1)$$

Figure 3 shows an example of noninterpolated PQPF (5-km resolution) initialized at 0000 UTC 11 July for Typhoon Soulik (2013), from 0000 UTC 12 July to 0000 UTC 13 July. This result presents more complete information from the ensemble system [e.g., an extreme member can be presented with a low probability rather than only showing the ensemble mean or probability of precipitation (POP)]. Relative to the traditional QPF, PQPF is not a single deterministic forecast; rather, it quantifies the uncertainty using probability, which is more favorable for risk and economic evaluation (Chang et al. 2015).

For the verification, the gridded data of PQPF are interpolated to the location of each rain gauge station using the cubic Bessel interpolation method (Abramowitz and Stegun 1964). The verification period is from 48 h prior to the start of the land warning to 48 h after the end of the land warning. To construct smooth probability distributions for every threshold, the verification sample considers the pairs of observations and PQPFs for each station in the verification area, each typhoon with the specific characteristic, and each successive 24-h time window in the verification period. In contrast, due to the finite number of samples, the verification of the specific period (e.g., typhoon approaching, landing, and leaving period) is not considered in this study.

c. *Verification methods*

To objectively compare the PQPF performance between typhoons with different characteristics, four probability verification methods are used (WWRP/WGNE 2015). First, the reliability diagram (RD; Hsu and Murphy 1986) compares the observed frequency against the forecast probability, which indicates the reliability level for each probability interval. RD is used to evaluate the overestimation and underestimation for selected forecast probabilities and rainfall thresholds. The probability interval of RD should be equal to the number of ensemble members (Chang et al. 2012), thus it is set to 5% (20 intervals, 21 nodes) in this study for which TAPEX provides at least 20 members. Second, the relative operating characteristic (ROC; Mason and Graham 1999) curve compares the hit rate and false alarm rate in the area higher than a selected forecast probability for a specific rainfall threshold. The ROC curve is not sensitive to probability bias and indicates the ability of the ensemble system to discriminate the actual events reaching the threshold from

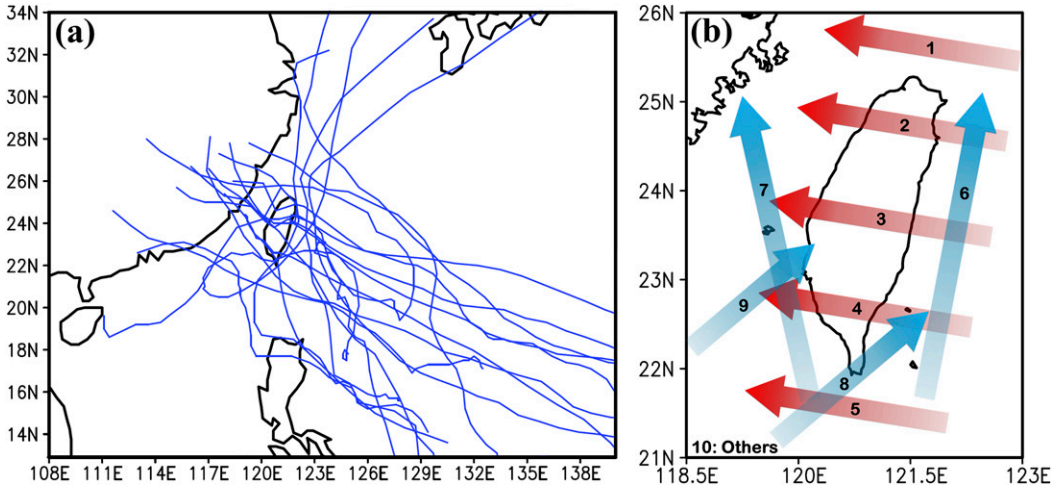

FIG. 2. (a) Tracks of 19 typhoons with a land warning issued by CWB during 2011–16. (b) Schematic diagram of the CWB typhoon track category, which is based on the figure on the CWB website (http://rdc28.cwb.gov.tw/).

the false events. The probability interval of ROC is also set to 5% (21 nodes) in this study. Relative to RD, ROC considers the forecast performance above each node, rather than at each node. When the area under the ROC curve is greater than 0.7, the forecast system has an acceptable discrimination ability (Hosmer and Lemeshow 2000, 160–164). Note that because the reliability and discrimination ability are different verification targets, the patterns of RD and ROC based on the same samples may differ.

Third, the Brier score (BrS; Brier 1950) shows the mean squared probability error for a specific threshold, as follows:

$$\mathrm{BrS} = \frac{1}{n}\sum_{j=1}^{n}(p_j - o_j)^2, \tag{2}$$

where n is the number of verification samples, p_j is the forecast probability of the jth verification sample, and o_j is the observation probability (0 or 1) of the jth

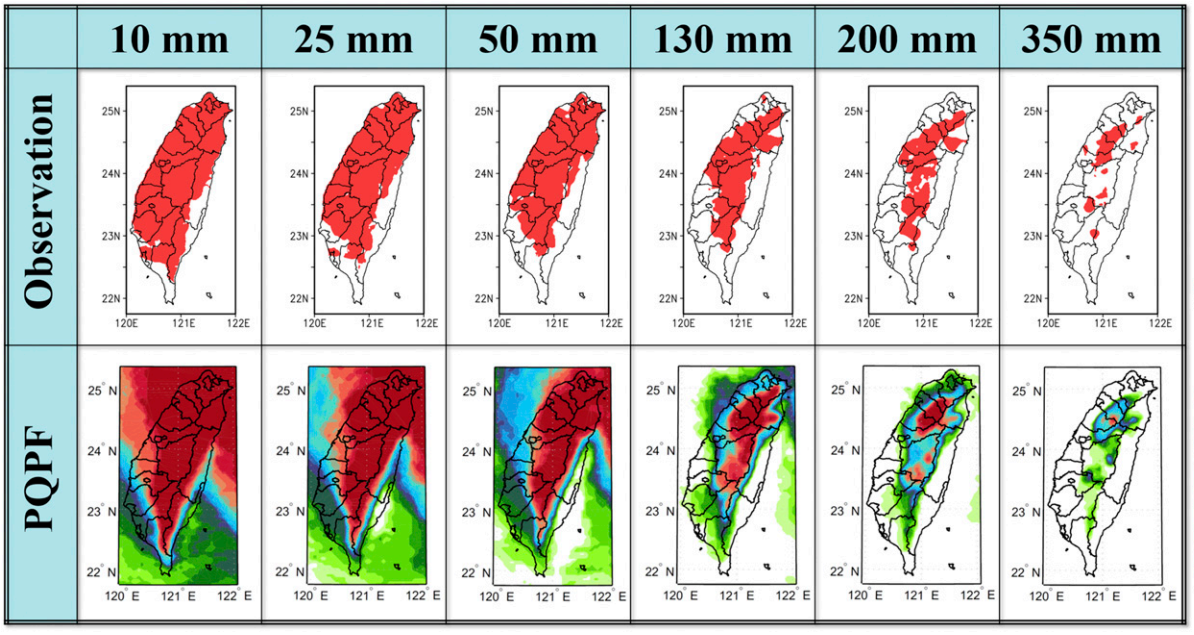

FIG. 3. An example of (bottom) PQPF and (top) observation for each 24-h rainfall threshold for Typhoon Soulik (2013). The initial time is 0000 UTC 11 Jul and the forecast period is from 0000 UTC 12 Jul to 0000 UTC 13 Jul. No interpolation is performed in the PQPF product, while a Cressman objective analysis is performed in the observation.

verification sample. Fourth, the rank probability score (RPS; Epstein 1969; Hersbach 2000) shows the sum of squared differences in cumulative probability space for a multithreshold probabilistic forecast, as follows:

$$\mathrm{RPS} = \frac{1}{M-1}\sum_{m=1}^{M}\left(\sum_{k=1}^{m} p_k - \sum_{k=1}^{m} o_k\right)^2, \qquad (3)$$

where M is the number of verification intervals, p_k is the forecast probability at the kth verification interval, and o_k is the observation probability (0 or 1) at the kth verification interval. Relative to BrS, RPS represents the error of the overall forecast probability distribution rather than that of a specific threshold. However, the error source from each threshold interval to RPS can be qualitatively interpreted using the magnitude and pattern of the BrS in the corresponding threshold. Furthermore, because the RPS and BrS are compared directly between different types of typhoons, the skill score (e.g., ranked probability skill score) is not considered in this study. Skill scores show the score change relative to a reference in a forecast verification but cannot be used in the comparison of two forecast verifications with different references.

RD and ROC, used in section 3, represent the verification features of the overall probability forecast distribution by a profile, which indicate the deviation of PQPF at specific forecast probabilities (all nodes). Their verification considers all stations over the entire Taiwan Island rather than at a specific region. Conversely, BrS and RPS represent the features of the probability forecast error by a single value, which indicate the absolute PQPF error. Their verification area can cover a small region or the entire Taiwan Island. In section 4, BrS and RPS are used to analyze the spatial distribution of PQPF error. Because the rain gauge distribution is nonuniform in Taiwan, to calculate the distribution of BrS and RPS, locations with similar topography are regarded as individual verification regions. As Fig. 1b shows, 26 verification regions are defined, including 17 city and 9 mountain regions. Each verification region has one RPS and BrS, which constructs the spatial distribution of PQPF error. In section 5, to show the relationship between the overall probability error of typhoons with a specific characteristic and forecast track error, RPS is calculated based on all stations in Taiwan. This RPS is not the average of 26 verification regions.

In addition, to compare the difference in PQPF between any two different typhoon types and show the distribution of the statistically significant area, the standard paired t test is used in the analysis of RD and ROC. Also, the block bootstrap test (block length is 24 h; Wilks 1997; Feldmann et al. 2019) is performed for BrS and RPS. Moreover, to quantify the uncertainty the bootstrap method is also used to calculate the standard deviation of each verification for RD and ROC (Hamill 1997).

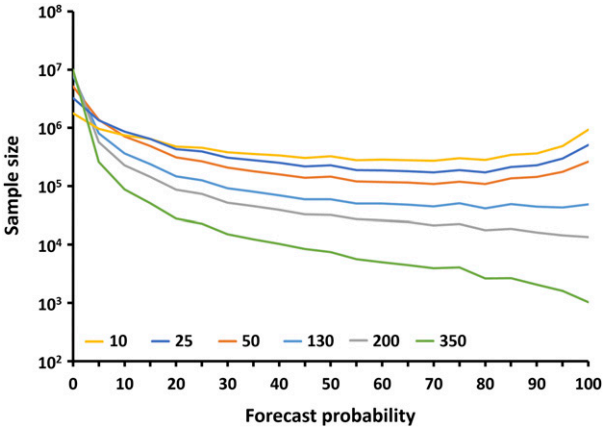

FIG. 4. Sample size at each probability node (every 5%) of different rainfall thresholds (mm) for all forecasts analyzed in this study.

3. Reliability and discrimination ability

The reliability and discrimination ability by rainfall threshold are quantified in the section, for all cases overall, and then subsetted by typhoon type. The total sample size analyzed in this study at each probability node for RD and ROC is shown in Fig. 4. For the overall performance of TAPEX, PQPF exhibits a consistent wet bias for all thresholds, according to the RD for all samples (Fig. 5). Specifically, this overestimation is more apparent in the higher rainfall thresholds, along the windward side of Taiwan, for cases that move parallel to Taiwan's north–south-oriented topography, and the bias also varies across the members using different models or cumulus parameterizations. The evidence suggests the bias mainly arises from a misrepresentation of precipitation produced through orographic lifting. This variability of the bias is similar to the findings of Hsiao et al. (2013) and Hong et al. (2015).

In general, because Taiwan Island has a steep topography and narrow plain and is surrounded by the sea, the topographic lifting and convergence tend to be overforecasted in the model when typhoons affect Taiwan. This causes the typhoon rainfall in TAPEX (i.e., magnitude or probability) to be substantially overestimated, especially over eastern Taiwan, which is the main windward side for all typhoons selected in this study. This relationship between PQPF wet bias and topography has been found by others and used to correct PQPF in previous studies (Yuan et al. 2005, 2008;

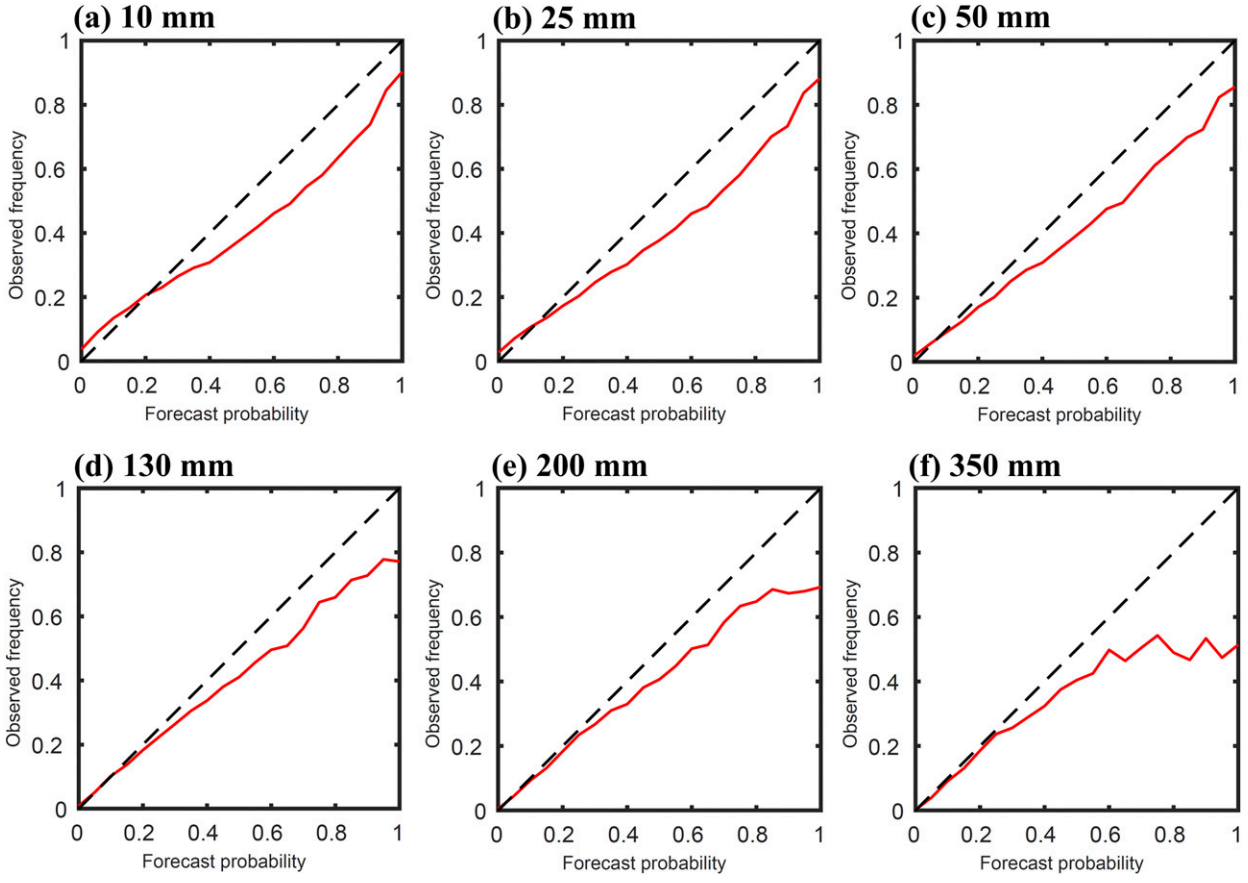

FIG. 5. RD for each rainfall threshold for all samples (red curve). The black dashed line shows the perfect reliability.

Chang et al. 2012, 2015). In addition, the ensemble members using the MM5 model or the Betts–Miller–Janjić scheme (Betts and Miller 1986) have a stronger overestimation than the others, which may be attributable to the water overproduced in the model (Yang and Tung 2003). Besides, the lower grid resolution and longer forecast time may also increase the total water produced in the model and cause a rainfall overestimation (Yang et al. 2008; Chen et al. 2014). Therefore, the 5-km grid resolution and the 72-h forecast time of TAPEX are also the possible reasons causing the overall PQPF performance with a wet bias. More work is needed to identify the specific cause of the misrepresented precipitation caused by the overforecasted topographic lifting and the model setting.

Furthermore, the TAPEX system has an acceptable discrimination ability for all thresholds, as their ROC areas are all greater than 0.7 (Fig. 6 and Table 2). The discrimination ability is relatively lower in the lowest and highest rainfall thresholds. Because the higher rainfall thresholds occur rarely, the hit rates and false alarm rates are also the lowest over all thresholds. However, the reliability and discrimination ability in each threshold may differ between different types of typhoons.

The difference in RD and ROC between typhoons with different moving tracks, sizes, and forward speeds are discussed next. The rainfall pattern and PQPF performance are not sensitive to the intensity of typhoon in Taiwan (not shown). Table 1 shows each characteristic of typhoons. For typhoon track, following the CWB typhoon track category, typhoons with track types 1–5 (Fig. 2b) or those with a larger westward than northward component to their moving direction during the warning period are defined as westward-moving (W) typhoons. The others are defined as northward-moving (N) typhoons.

For typhoon size, because Taiwan Island is approximately 400 km long, a radius of 200 km is chosen to distinguish large and small typhoons. Specifically, the average radius of 14 m s^{-1} (R14) during the typhoon warning period exceeding 200 km is used to define a large typhoon, while the rest are defined as small typhoons. R14 is a disaster prevention reference in Taiwan. For typhoon forward speed, previous studies showed that the mean forward speed of typhoons during the landfall period was approximately 22 km h^{-1},

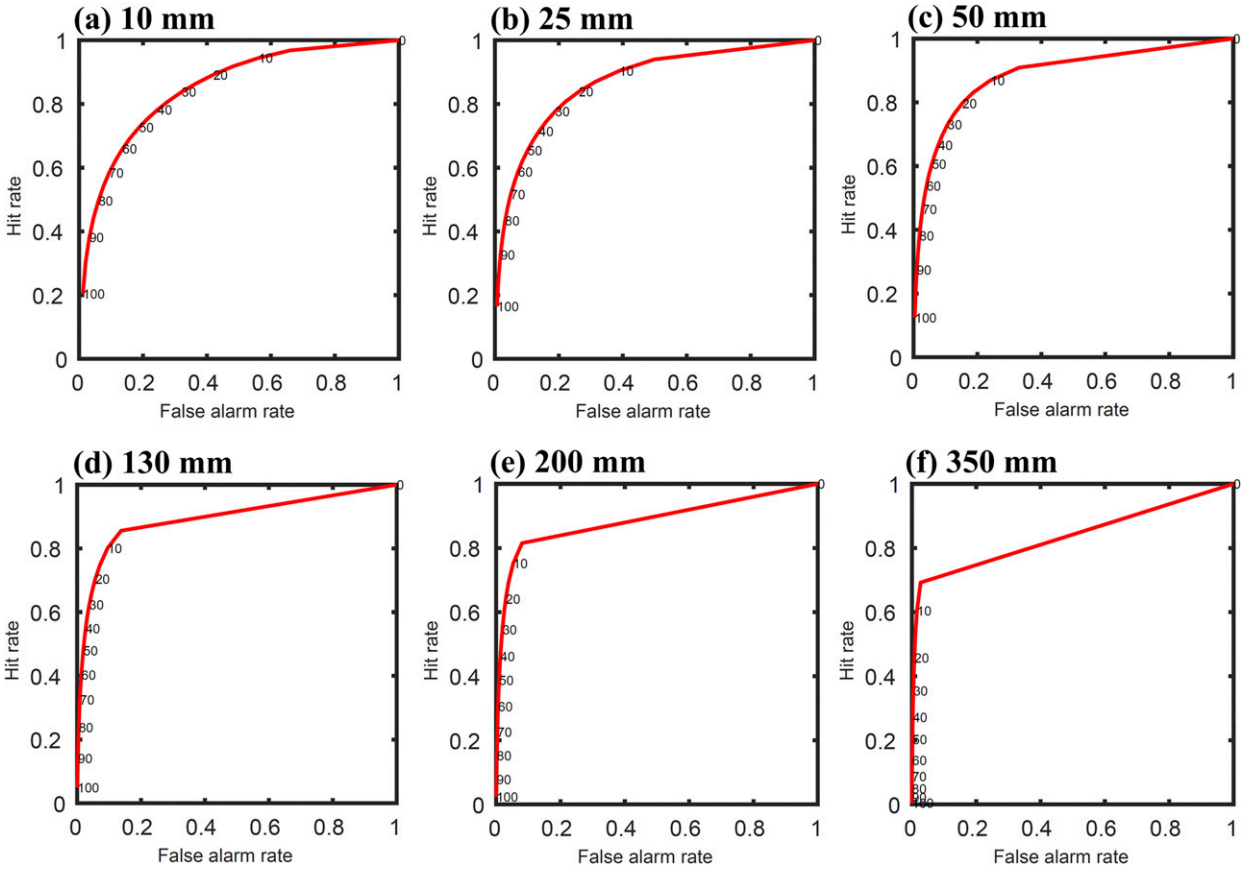

FIG. 6. ROC for each rainfall threshold for all samples. The numbers near the curve are the forecast probabilities (only 11 nodes are labeled).

and that different forward speeds significantly influenced rainfall totals over Taiwan (Hsu et al. 2013). In this study, 22 km h^{-1} is regarded as the criterion to distinguish between fast- and slow-moving typhoons. The forward speed of the typhoon is calculated as the average motion speed during the warning period based on the CWB best track. Note that because of the significant difference in rainfall patterns between northward-moving and westward-moving typhoons, the comparisons of PQPF performance between typhoons with different sizes and forward speeds only consider the westward-moving typhoons, which have many more case numbers. In summary, the number of forecasts for each typhoon type is shown in Table 2.

a. Track

As Fig. 7 shows, although the westward-moving and northward-moving typhoons are both overestimated in the mid- and high-probability areas for all thresholds, this overestimation is more significant in the northward-moving

TABLE 2. The number of forecasts for each type of typhoon and the area under the ROC curves for the different types of typhoons for each rainfall threshold (mm). Single asterisks (*) indicate that the difference between two groups within a typhoon characteristic passes the standard paired *t* test at a 95% confidence level.

	Number	10 mm	25 mm	50 mm	130 mm	200 mm	350 mm
All	495	0.85	0.87	0.89	0.89	0.89	0.84
Westward-moving	342	0.86*	0.88*	0.90*	0.90*	0.89*	0.84*
Northward-moving	153	0.81*	0.83*	0.85*	0.87*	0.86*	0.83*
Large westward-moving	208	0.87*	0.89*	0.90*	0.90	0.89	0.84*
Small westward-moving	134	0.84*	0.87*	0.89*	0.89	0.89	0.82*
Fast westward-moving	100	0.87*	0.89*	0.89	0.88*	0.87*	0.82*
Slow westward-moving	242	0.86*	0.88*	0.90	0.91*	0.90*	0.85*

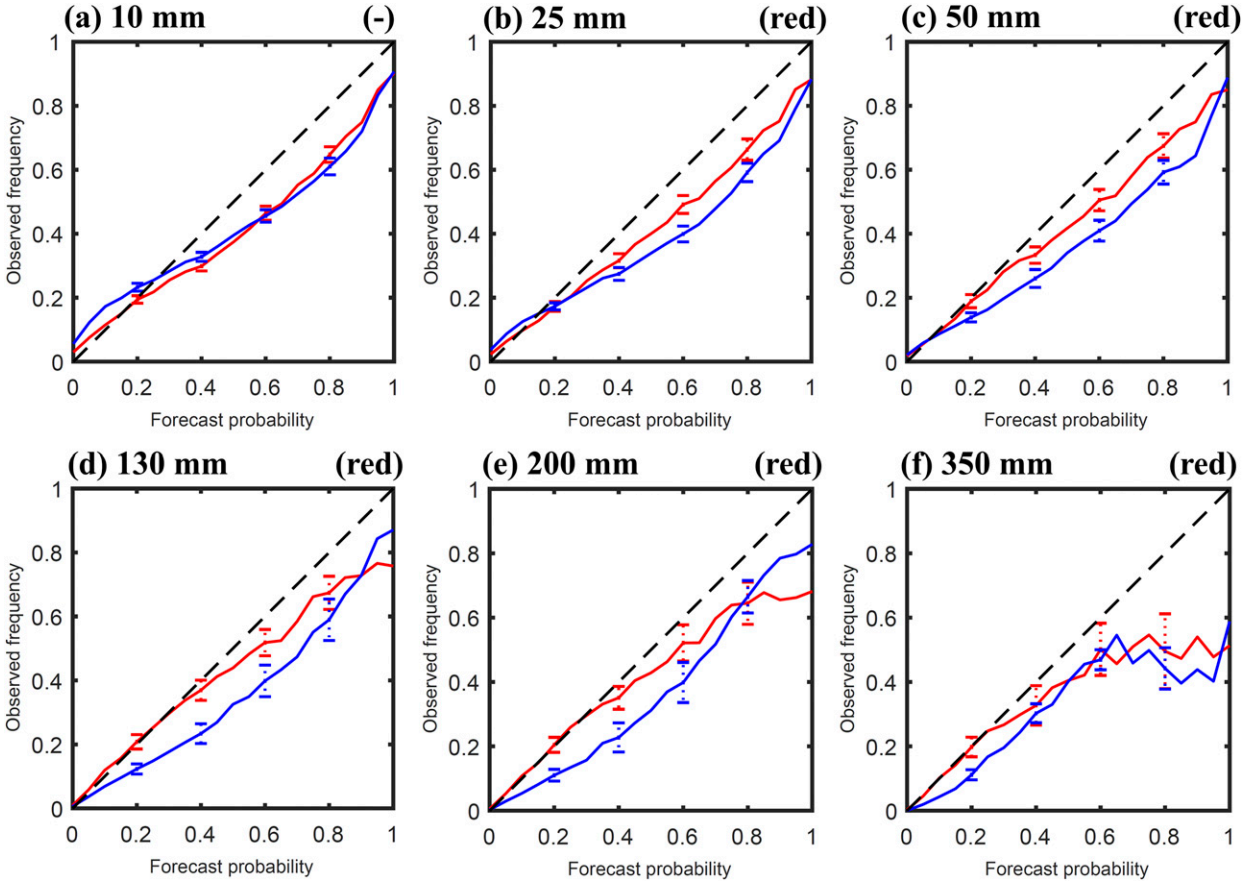

FIG. 7. RD for each rainfall threshold for westward- (red) and northward-moving (blue) typhoons. The black dashed line shows the perfect reliability. The label (red and blue) at the upper-right corner indicates which color line is closest to the perfect reliability based on the standard paired *t* test at a 95% confidence level, and a dash (—) indicates there is no significant difference. The dotted bar shows one standard deviation of observed frequency at forecasts of 20%, 40%, 60%, and 80% calculated by the bootstrap method.

typhoons. This is mainly attributed to the larger forecast track error and the larger windward region (the whole eastern Taiwan) for the northward-moving typhoons. Overall, the reliability of PQPF for westward-moving typhoons is higher than that for northward-moving typhoons, except in the 10-mm threshold. Meanwhile, the ROC curve shows that the ratios of hit rate to false alarm rate are all higher in westward-moving typhoons than in northward-moving typhoons for all thresholds (Table 2), which indicates that the PQPF for westward-moving typhoons has a higher discrimination ability.

b. *Size*

As Fig. 8 shows, except in the 130-mm threshold, PQPFs have higher reliability for large typhoons. PQPFs for small typhoons are more frequently overestimated than for large typhoons, especially at forecast probabilities between 0.3 and 0.8, perhaps due to the difficulty in simulating small typhoons in the models. In addition, the ROC curve shows that the discrimination ability is significantly higher for large typhoons in the 10-, 25-, 50-, and 350-mm thresholds (Table 2).

c. *Forward speed*

As Fig. 9 indicates, fast-moving typhoons have higher reliability than slow-moving typhoons only in the smaller thresholds (10 and 25 mm). There is no significant difference in reliability between fast- and slow-moving typhoons for the other thresholds. The ROC curve shows that the discrimination ability for fast typhoons is also higher in the smaller thresholds (10 and 25 mm) but lower in the larger thresholds (130, 200, and 350 mm) than that for slow typhoons (Table 2). However, for the overall performance PQPF of slow typhoons performs better than fast typhoons, except in the small rainfall events.

4. Distribution of probability error

Our attention now turns to the spatial distribution of the PQPF error. Figure 10a shows the RPS distribution

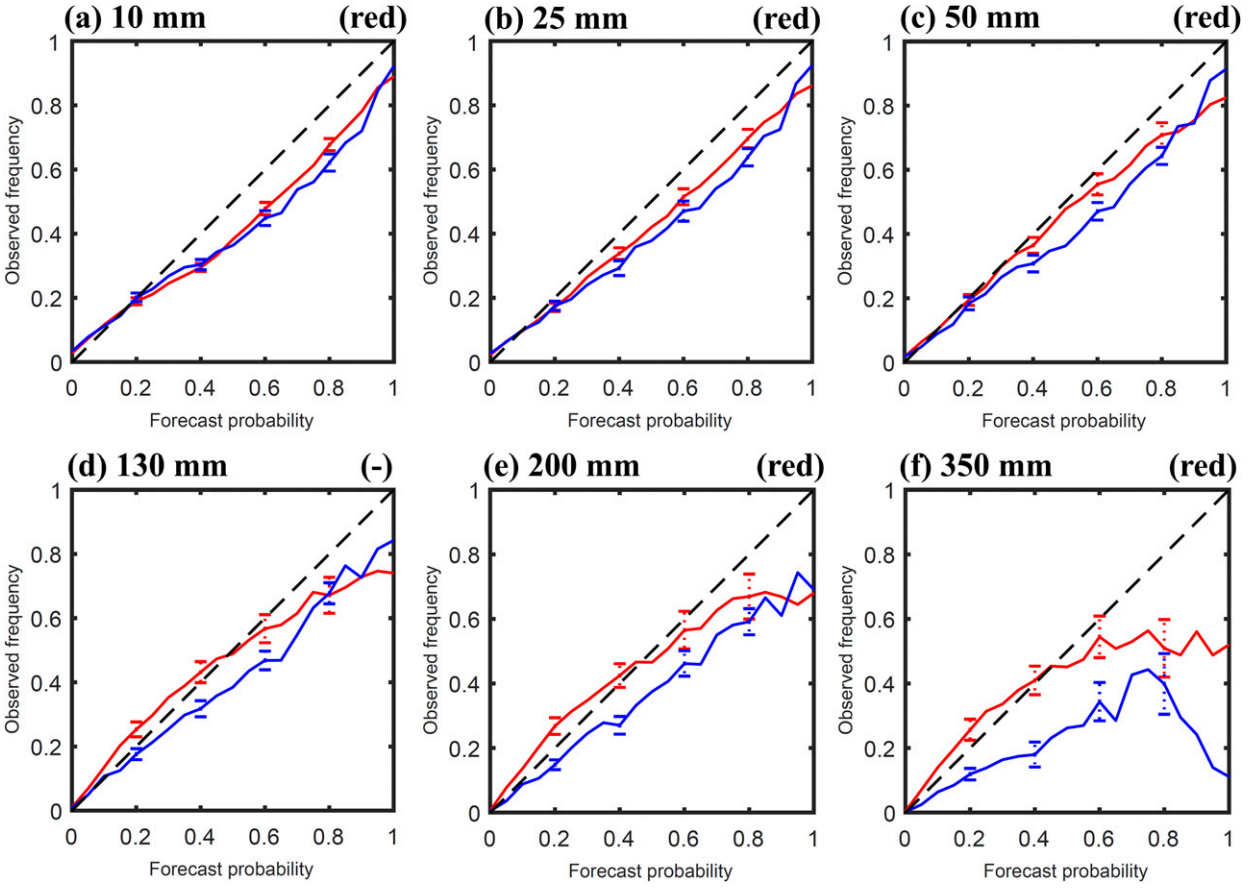

FIG. 8. As in Fig. 7, but for large westward-moving (red) and small westward-moving (blue) typhoons.

for all samples, and indicates that the multithreshold probability error is higher in eastern and southern Taiwan than in northwestern Taiwan. Further, separating the probability error in each threshold (Fig. 11) shows that the BrS gradually decreases from the lowest rainfall threshold to the highest rainfall threshold, and the highest errors are located in central-eastern and southern Taiwan. This suggests a role for high topography in enhancing probability error, given that central to eastern Taiwan and southern Taiwan is mountainous, while the western coast is characterized by plains (Fig. 1a). Moreover, the distribution of BrS in each threshold also suggests that the main source of RPS value is from the probability error of the smaller rainfall thresholds (10, 25, and 50 mm). In this section, the PQPF error distribution and source for different types of typhoons are analyzed by comparing their RPS and BrS distributions.

a. Track

Figure 10b shows the difference in RPS between westward-moving and northward-moving typhoons. The RPS is smaller in westward-moving typhoons than for northward-moving typhoons, especially in southeastern Taiwan (the main windward side). On average, the deviation of overall probability distribution over all Taiwan is smaller for westward-moving typhoons. Furthermore, analyzing the probability error for each threshold (Fig. 12) shows that the PQPF performance of westward-moving typhoons is better than that of northward-moving typhoons over all Taiwan for smaller thresholds (10, 25, and 50 mm), but the PQPF of northward-moving typhoons has a smaller error around the central mountain region in larger thresholds (130, 200, and 350 mm). Corresponding to RD results (Figs. 7d–f), this lower BrS of northward-moving typhoons in central Taiwan is attributed to the reduced forecast overestimation for high-probability events.

b. Size

As Fig. 10c shows, the multithreshold probability error for large typhoons is smaller than that for small typhoons in central-western Taiwan. This feature is mainly attributed to smaller thresholds (10, 25, and

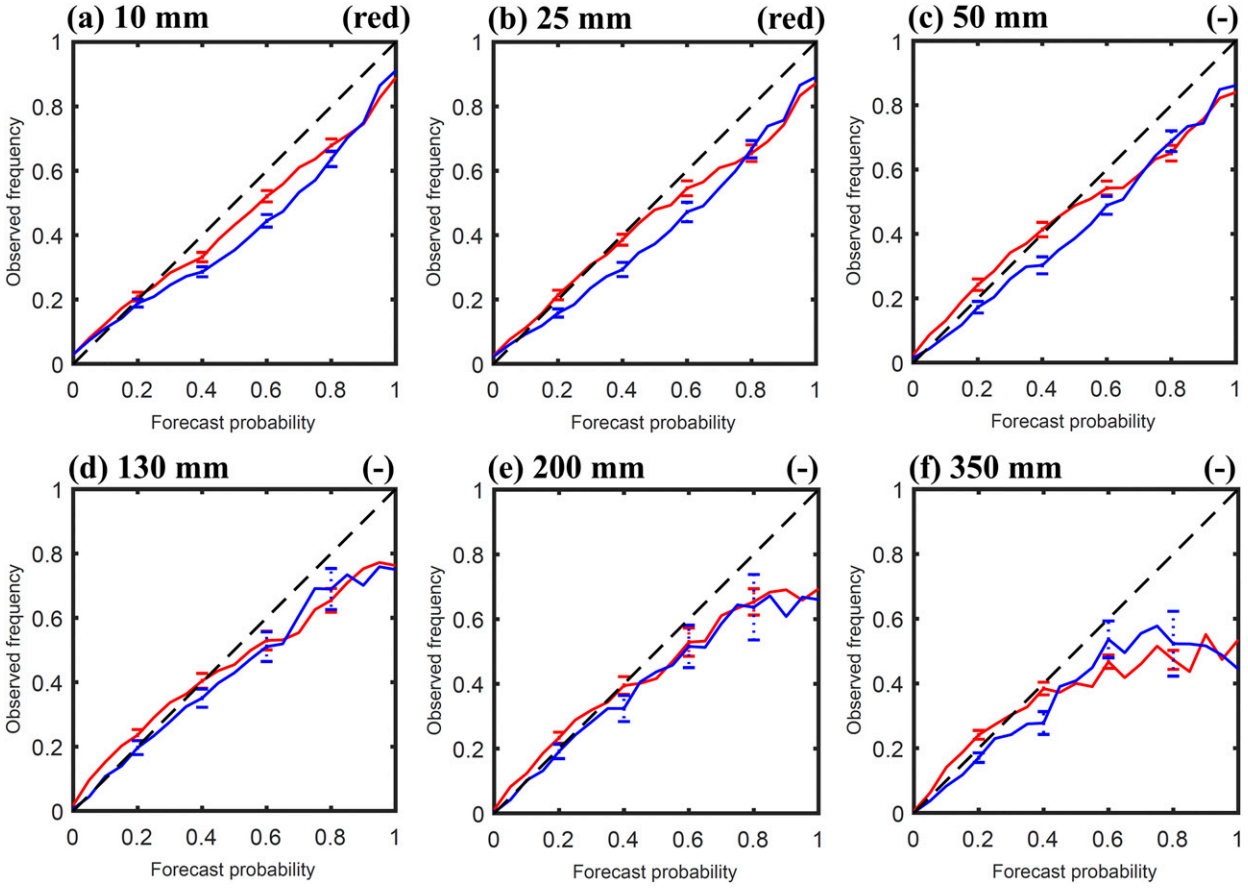

FIG. 9. As in Fig. 7, but for fast westward-moving (red) and slow westward-moving (blue) typhoons.

50 mm, Fig. 13). Moreover, although the differences in BrS exist in northern and southern Taiwan in five different thresholds (except 350 mm), RPS does not exhibit the same behavior because of the compensating effect between different thresholds. For example, BrS is smaller at smaller thresholds (10 and 25 mm) but is larger at larger thresholds (50, 130, and 200 mm) in northern Taiwan for large typhoons. In general, the overall PQPF error over Taiwan is smaller for large typhoons.

c. *Forward speed*

The RPS of slow-moving typhoons is lower than that of fast-moving typhoons in central to western Taiwan (Fig. 10d). This feature can be attributed to the thresholds exceeding 25 mm (Fig. 14). Although slow typhoons have longer impact duration, which leads to higher rainfall totals (Chang et al. 2013; Hsu et al. 2013), their probability errors of PQPF are smaller than those of fast typhoons. Furthermore, the RPS pattern also indicates that the difference in the PQPF error between different typhoon forward speeds mainly occurs when the typhoons leave Taiwan. The ensemble system captures PQPF better for slow than fast typhoons after they cross the Central Mountain Range of Taiwan.

5. Relationship between PQPF and forecast track error

Previous studies suggested that forecasts of rainfall spatial distribution and amounts are influenced by characteristics of the typhoon track (e.g., track orientation and forecast track error) (Lee et al. 2013; Hong et al. 2015; Wu et al. 2016). However, the relationship between the characteristics of the typhoon track and the PQPF for different typhoon types is still unknown. To quantify the possible impact of the typhoon forecast track on PQPF, the correlation coefficients between the RPS over all Taiwan and the track error mean and ensemble spread of each forecast for different types of typhoons are analyzed in this section. In the past decade, the average typhoon forecast track error over the western North Pacific is around 100 km in the first 24-h forecast (Peng et al. 2017). The 100 km

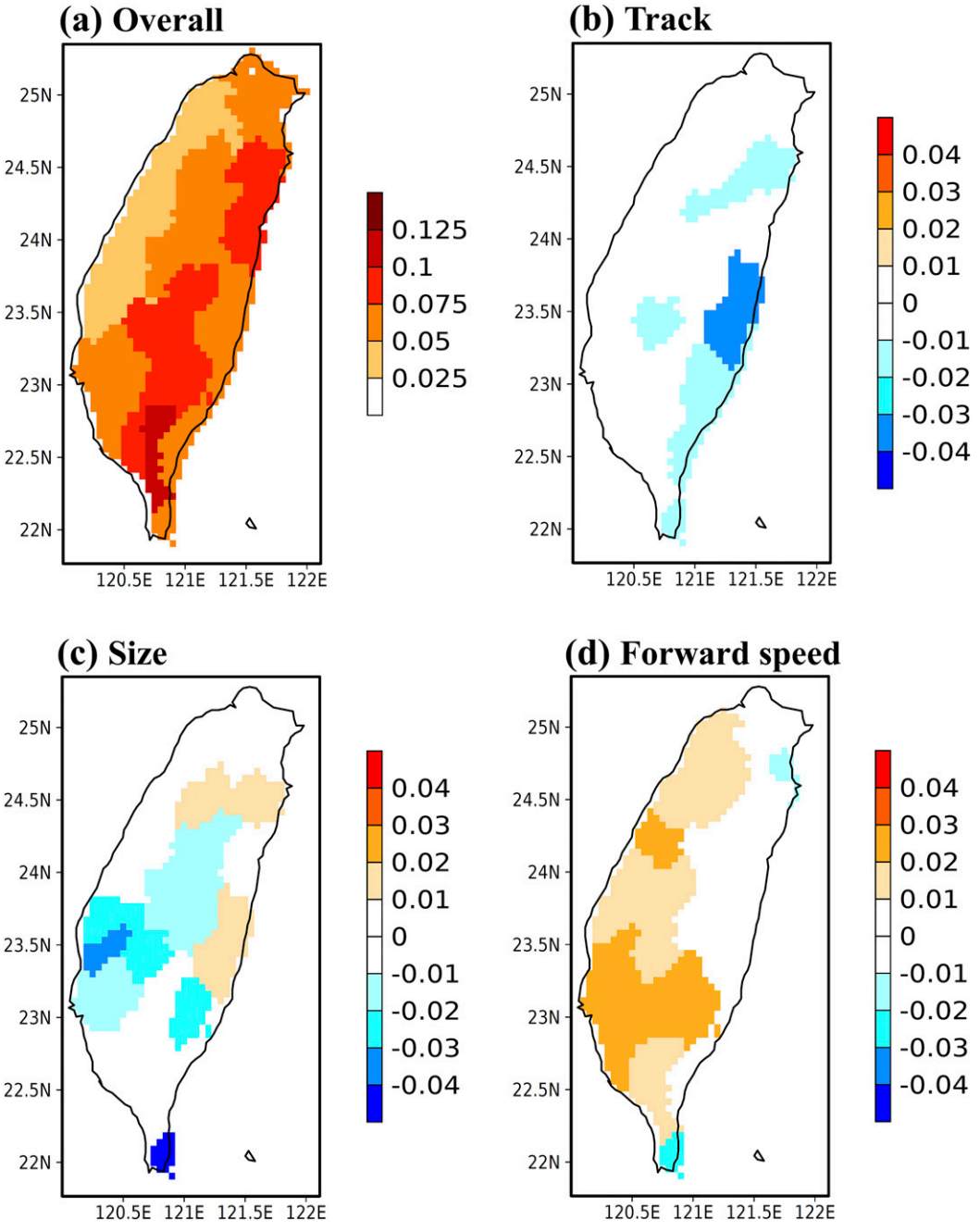

FIG. 10. (a) RPS for all samples and difference in RPS between (b) westward- and northward-moving typhoons, (c) large westward-moving and small westward-moving typhoons, and (d) fast westward-moving and slow westward-moving typhoons. For (b)–(d), all differences shown in colored shading pass the block bootstrap test at a 95% confidence level.

in 24-h forecast can therefore be regarded as a reference value to evaluate the track forecast. Thus, in this section, the track error and spread, RPS, and their correlation are all calculated for the first 24-h forecast in each ensemble forecast for the different types of typhoons.

The track error mean is the average of the difference between the TC center of each ensemble member and the observed TC location. The ensemble spread is the average of the difference between the TC center of each ensemble member and the location of all ensemble member mean. The mean of the track error indicates the accuracy

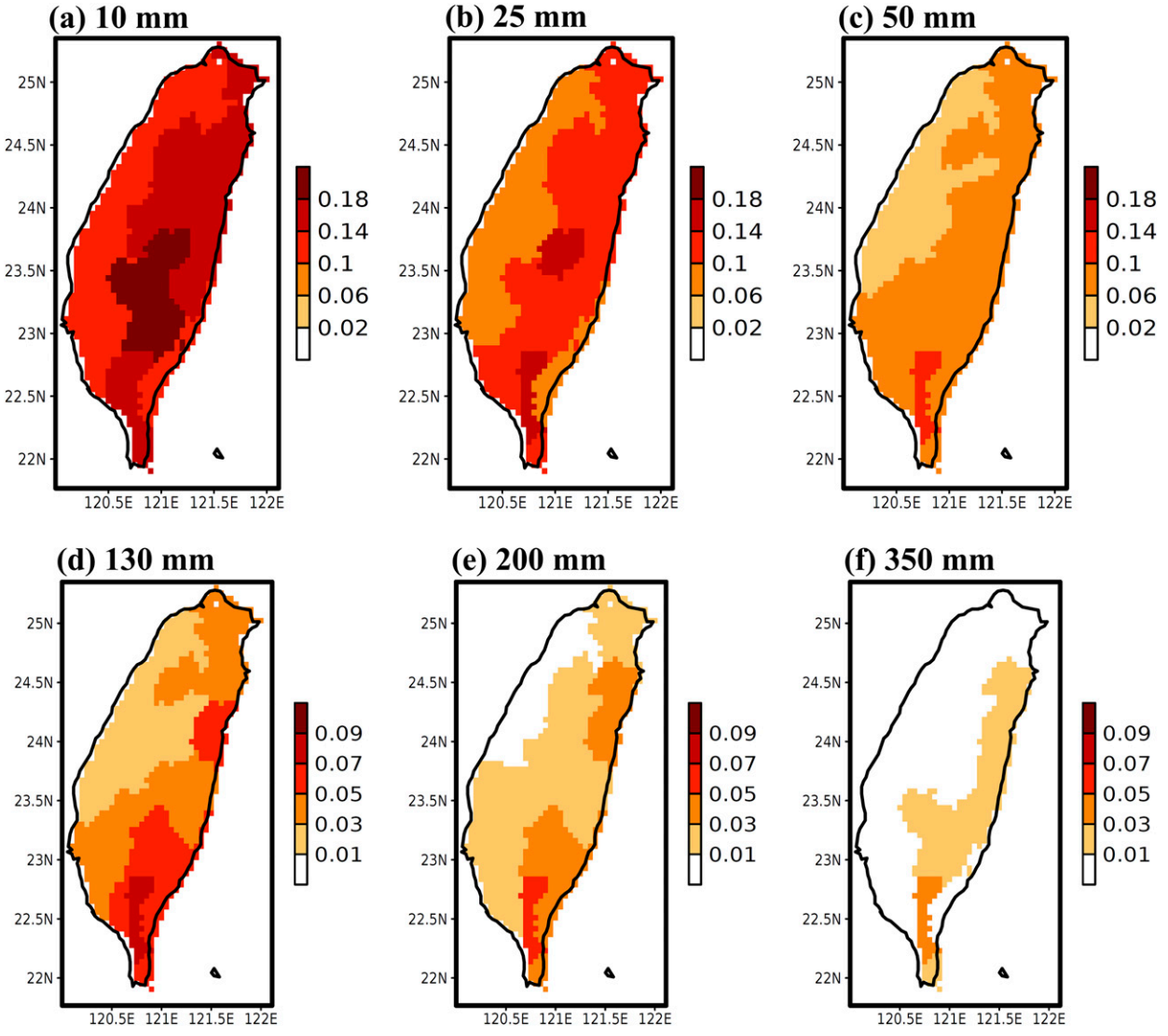

FIG. 11. BrS for each rainfall threshold for all samples. Note the difference in color scale between (a)–(c) and (d)–(f).

of the forecast track, and the ensemble spread indicates the variance of the forecast track. Table 3 shows that the track error and spread for westward-moving typhoons are smaller than those for northward-moving typhoons with statistical significance. Northward-moving typhoons have the largest track error and spread in the six types of typhoons. For the westward-moving but different-size typhoons, the large typhoons have smaller track error and spread than the small typhoons, and the main difference is attributed to the east–west direction. For the westward-moving but different-forward-speed typhoons, there is no significant difference in the mean track error. But slow typhoons have a greater track spread than fast typhoons.

Although the track error and spread are significantly different among the different forecasts and typhoons, only the correlations between track spread and RPS for northward-moving or fast westward-moving typhoons are significant (Table 4). On closer inspection we find that typhoons with larger and smaller forecast track error have different relationships with RPSs, and so the correlation between larger and smaller track error and the RPS must be calculated separately. For example, as Fig. 15 (red curve) shows, the RPS increases with track error for typhoons with smaller track error but decreases for typhoons with larger track error. Samples in each category (i.e., track, size, and forward speed) are classified into the groups with larger and smaller track error according to whether the error is larger or smaller than the average of the track error in each category. Because the forecast track error is so different in each category, separating based on the average track error of all samples is not considered.

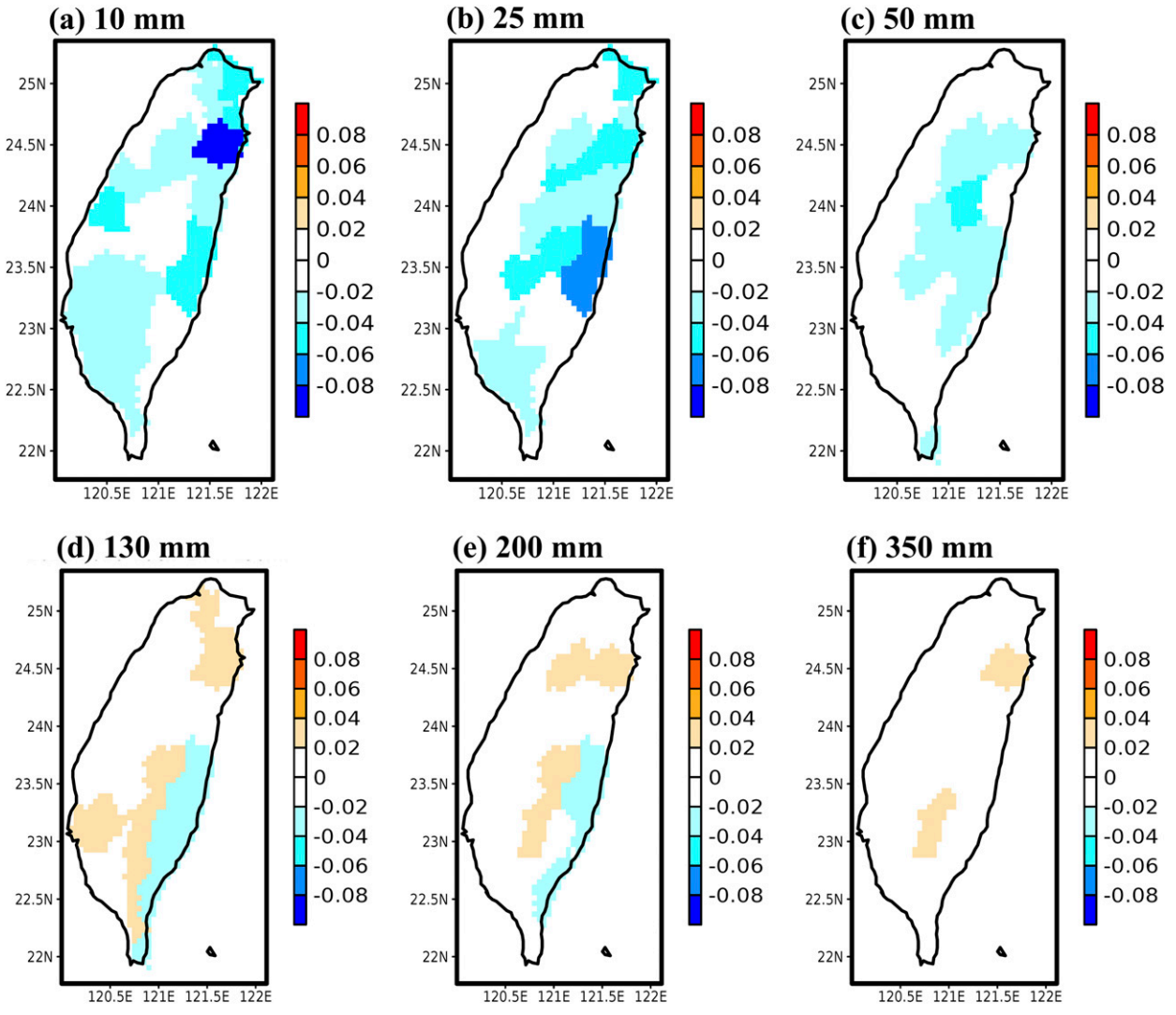

FIG. 12. Difference in BrS between westward- and northward-moving typhoons for each rainfall threshold. All differences shown in colored shading pass the block bootstrap test at a 95% confidence level.

For typhoons with a larger track error (Tables 3 and 4), RPS is only slightly correlated with the total forecast track error. Because the track error is much larger than the operational forecast average (~100 km in 24-h forecast) already, the increase in track spread can increase the variance of rainfall pattern forecast and reduce the RPS value, that is, for small or fast westward-moving typhoons. However, for northward-moving typhoons, because the typhoon heading and the orientation of the Central Mountain Range are consistent, the larger track spread includes some members that cross the mountain range and produce extremely high probability errors, leading to a positive correlation with RPS. Although the correlation signs are opposite, these three categories that have a significant correlation between RPS and track spread are the three with the largest probability error.

On the other hand, for typhoons with a smaller track error (Tables 3 and 4), because no forecast with excessive track error is considered in the correlation, the different relationships of the track error mean and ensemble spread among the different types of typhoons can be clearly presented. The RPS is positively correlated and statistically significant with the mean of the track error for northward-moving typhoons, especially the south–north track error, but no significant correlation for westward-moving typhoons is found. This feature suggests that the PQPF performance of northward-moving typhoons is more sensitive to the accuracy of the forecast track than that of westward-moving typhoons, if other characteristics of typhoon are not considered. Because Taiwan is a south–north long and east–west narrow island with

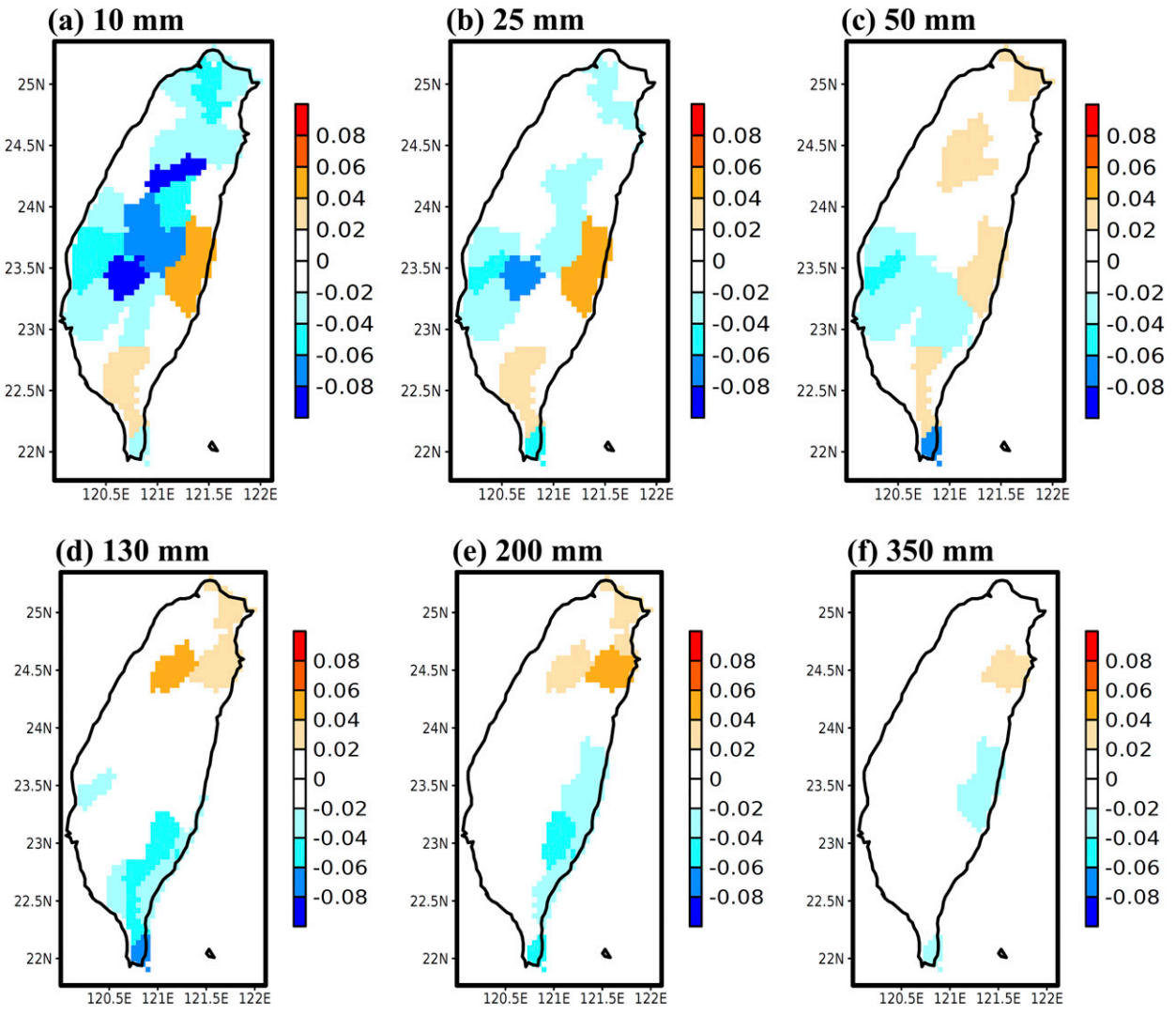

FIG. 13. As in Fig. 12, but between large westward-moving and small westward-moving typhoons.

high mountainous topography, the overall change of rainfall pattern caused by track shift (e.g., topography-locked rainfall) is more sensitive for northward-moving typhoons. For instance, the east–west track shift may cause the northward-moving typhoon to pass through Taiwan from the eastern or western coast, while the south–north track shift may cause the total, location, and timing of topography-locked rainfall originating from the northward-moving typhoon to change significantly. These all lead to higher RPS in northward-moving typhoons than in westward-moving typhoons even if they had a similar forecast track error.

Furthermore, Table 4 indicates that the RPSs of large westward-moving and fast westward-moving typhoons with smaller track error are negatively correlated and statistically significant with the ensemble spread, while not correlated with the mean track error. These negative correlations with the track spread are mainly derived from the east–west component. However, the components of east–west and south–north track forecast are positively correlated with the RPS for small westward-moving and slow westward-moving typhoons, respectively. Relative to large or fast typhoons, the importance of the typhoon center shift on the pattern of forecast rainfall over Taiwan is greater for small or slow typhoons. On average, because the small typhoons have a smaller radius with stronger wind and heavy rainfall (Matyas 2006) and the slow typhoons have a longer duration impacting Taiwan (Hsu et al. 2013), their track error will produce more significant differences in the distribution and accumulation of forecast rainfall, leading to higher PQPF error. In other words, the PQPFs of large or fast

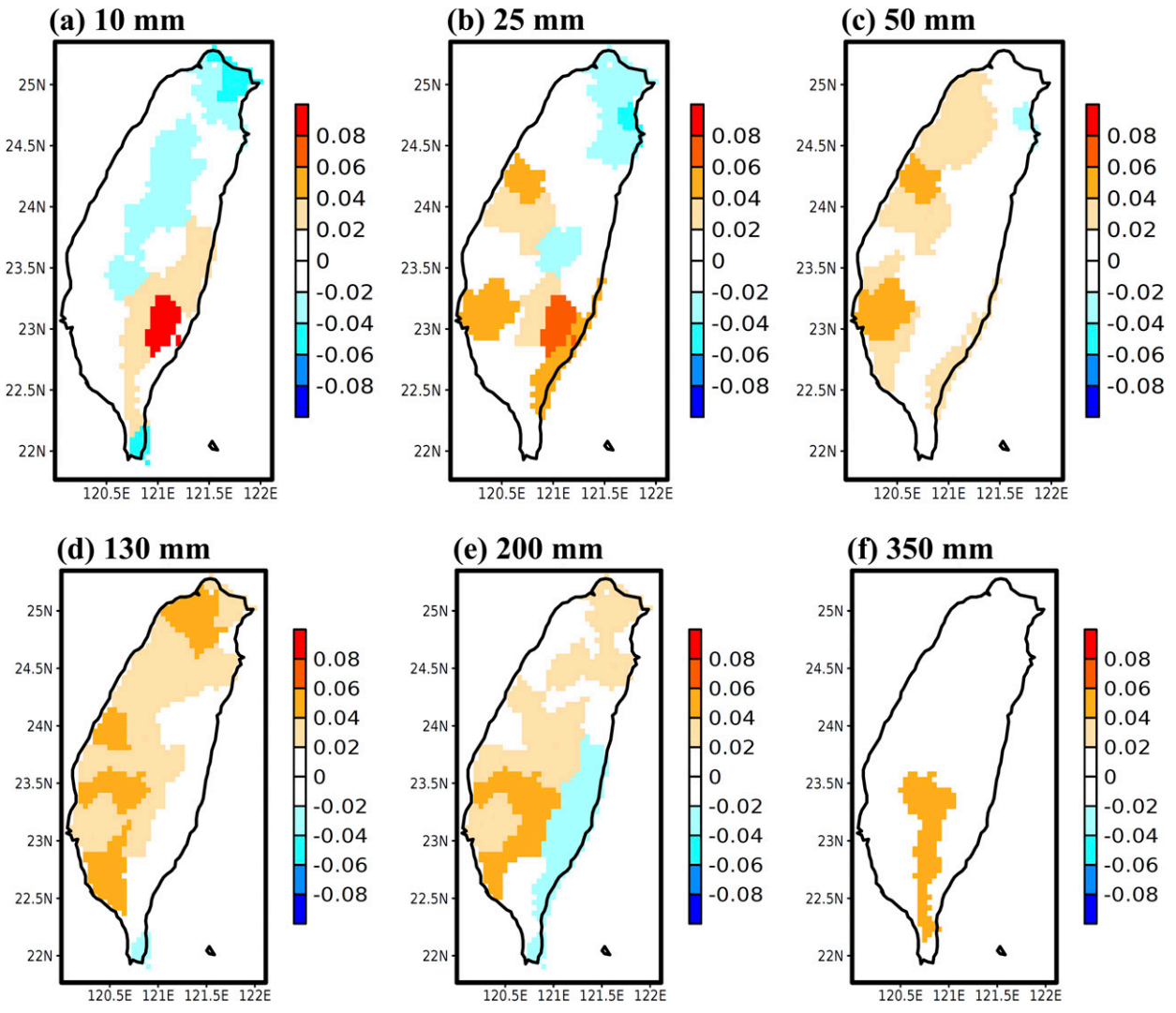

FIG. 14. As in Fig. 12, but between fast westward-moving and slow westward-moving typhoons.

typhoons have a higher tolerance to their track error than those of small or slow typhoons. The larger spread of typhoon track in the ensemble system for large or fast typhoons, which increases the tolerance of PQPF to the mean track error, has smaller RPS. This feature is not affected by the average RPS in each category.

For northward-moving and small westward-moving typhoons, although typhoons with larger track error and smaller track error have an opposite correlation sign between track error and RPS, the main influence on the average RPS in these two categories comes from the samples with smaller track error. Figure 15 (blue curve) shows that the peak of probability distribution function (PDF) of typhoon track error is on the left side of the mean of track error (i.e., the errors of most samples are smaller than the mean). Generally, the distribution of typhoon forecast track error is positively skewed (Peng et al. 2017), which is attributed to a few forecasts with extremely large track errors (e.g., 300 km in 24-h forecast), increasing the total track error mean. This indicates the average RPS in these two categories is more influenced by samples with smaller track error (positive correlation with RPS) than those with larger track error (negative correlation with RPS). Therefore, the larger mean track error for northward-moving and small westward-moving typhoons would cause them to have a higher average RPS than westward-moving and large westward-moving typhoons. Conversely, although the PQPF of slow typhoons is more sensitive to track error than for fast typhoons, there is no significant difference in the mean track error. For fast typhoons, the ensemble spread is so small that the rainfall

TABLE 3. Mean track error (km) and ensemble spread (km) of the first 24-h forecasts for the different types of typhoons. Values from left to right in each column indicate the value for all samples, samples with larger track error, and samples with smaller track error in each category, respectively. Single asterisks (*) indicate that the difference between two groups within a typhoon characteristic passes the standard *t* test at a 95% confidence level.

	Total				East–west				South–north			
	Error mean		Ensemble spread		Error mean		Ensemble spread		Error mean		Ensemble spread	
Westward-moving	95.0*	132.7*	61.5*	71.8*	53.9	67.0	95.7*	45.8	40.1*	45.4*	36.1	55.3*	76.2*	39.9*	38.2*	46.2*	32.4*
Northward-moving	108.7*	159.3*	73.5*	86.6*	115.2*	66.7	75.0	113.9*	47.9	60.1*	82.1*	44.7	66.0*	94.1*	46.4*	53.1*	70.0*	41.4*
Large westward-moving	89.3*	125.5*	62.2*	58.2*	69.9	49.5*	61.2*	86.1*	42.5*	37.3*	42.9*	33.0*	54.2	77.0	37.1*	36.8	46.2	29.7*
Small westward-moving	108.4*	149.4*	81.1*	69.4*	73.6	66.6*	80.9*	118.5*	55.8*	46.8*	51.6*	43.6*	57.8	74.4	46.7*	41.7	42.3	41.3*
Fast westward-moving	92.2	146.5	62.1*	51.3*	60.8*	46.0*	66.3	108.1	43.0	33.9*	38.2*	31.5*	53.3	83.4	36.6*	31.6*	39.6	27.1*
Slow westward-moving	96.1	128.5	70.0*	65.7*	74.9*	58.3*	67.3	92.6	46.9	42.6*	47.5*	38.7*	56.1	73.5	42.1*	41.0*	48.0	35.3*

forecast spread is reduced based on the same track error, which leads to higher RPS than slow typhoons.

6. Discussion and conclusions

a. Discussion

PQPF performance can be substantially impacted by the characteristics of ensemble forecast systems. In this study, some settings in the TAPEX ensemble system and the verification targets may have enhanced the overall PQPF reliability and discrimination ability but diminished RPS and BrS.

There are a number of reasons for the high reliability and discrimination ability and the low RPS and BrS. First, typhoons are large synoptic systems strongly steered by the large-scale environment. Relative to the heavy rainfall caused by mei-yu fronts, thunderstorms, and other mesoscale systems, the forecast capacity of typhoon rainfall is higher (Chen et al. 2014). Second, the forecast target of 24-h accumulated rainfall is longer than that is used by some previous studies (e.g., 6- and 3-h). This longer time window reduces the likelihood that the forecast time window entirely misses the rainfall event, thereby reducing PQPF error (Xu et al. 2010). Third, there is a strong topographic effect for typhoons affecting Taiwan. The overall rainfall pattern is substantially correlated with the relative location between typhoon centers and the topography of Taiwan (Chang et al. 1993; Lin et al. 2001; Chiao and Lin 2003; Lee et al. 2006, 2013), which also increases the numerical model forecast capacity for typhoon rainfall. This is similar to the finding of Chang et al. (2015). Fourth, TAPEX uses the final analysis of the National Centers for Environmental Prediction Global Forecast System (NCEP-GFS) as the initial condition ($t = 0$ h) and performs bogus assimilation of the typhoon vortex and assimilation of other data around Taiwan (Hsiao et al. 2013). This may improve the forecast performance of TAPEX over other system using the previous forecast field (e.g., $t = -6$ h) as the initial condition. However, note that all these settings do not disturb the comparison between the different types of typhoons under the same ensemble system.

In addition, as BrS figures show (e.g., Fig. 11), larger rainfall thresholds or lower altitude areas have smaller BrS than smaller rainfall thresholds or higher altitude areas. This is attributed to details of the statistical methods. Because the actual accumulated rainfall reaching the large rainfall threshold (e.g., 350 mm in 24 h) or the higher rainfall total happening over plain areas (relative to mountainous areas) has lower probability, the absolute value of probability error from observation probability (0 or 1) is small. Note that the minimum BrS is 0 and the distribution of rainfall

TABLE 4. Correlation coefficient between mean track error and ensemble spread and RPS for the first 24-h forecasts for the different types of typhoons. Values from left to right in each column indicate the coefficient for all samples, samples with larger track error, and samples with smaller track error in each category, respectively. Single asterisks (*) and plus signs (+) indicate that the correlation passes the Pearson correlation test at a 95% and 90% confidence level, respectively.

	Total					East–west					South–north				
	Error mean			Ensemble spread		Error mean			Ensemble spread		Error mean			Ensemble spread	
Westward-moving	0.00	−0.05	0.01	−0.04	−0.08	−0.01	−0.04	−0.04	−0.06	0.00	−0.17*	0.00	−0.04	0.06	−0.02	−0.07	0.04
Northward-moving	−0.05	−0.13	0.35*	0.30*	0.57*	0.09	−0.17+	−0.32*	0.18	0.31*	0.57*	0.10	0.10	0.16	0.32*	0.29*	0.58*	0.06
Large westward-moving	−0.01	0.00	−0.10	−0.06	−0.01	−0.26*	−0.01	0.04	−0.14+	−0.09	0.06	−0.35*	−0.02	−0.04	0.02	−0.04	−0.06	−0.03
Small westward-moving	−0.07	−0.24	0.29*	−0.08	−0.49*	0.23+	−0.11	−0.39*	0.24+	−0.11	−0.44*	0.19	0.05	0.11	0.13	0.05	−0.14	0.15
Fast westward-moving	0.00	−0.27+	−0.05	−0.21*	−0.41*	−0.28*	−0.03	−0.33+	−0.06	−0.35*	−0.61*	−0.34*	0.05	−0.04	−0.03	0.01	−0.14	−0.01
Slow westward-moving	0.00	−0.03	0.07	0.05	0.03	0.12	0.01	0.03	−0.08	0.09	0.13	−0.01	−0.02	−0.08	0.17+	0.00	−0.05	0.16+

probability for multithreshold almost follows a chi-square distribution (Wilson and Hilferty 1931). Moreover, the value of probability error is likely to be smoothed by the number of verification samples [n in Eq. (2)]. Thus, it is easy for the BrS to be lower for low-probability events than for high-probability events, as discussed in Richardson (2001), Jewson (2004), and Roulston (2007). The impacts of small absolute value of probability error and being smoothed by sample numbers are also found in RPS calculation. When typhoons have not made a landfall and produce small rainfall totals, the final RPS is smoothed by a large number of samples with very low probability error even though the forecasts have larger track error. This feature occurs in some typhoons with larger track error but smaller RPS.

Furthermore, comparing RD and BrS shows that, although PQPF for the large rainfall threshold has a lower reliability than that for the small threshold, it has smaller probability error. These two verifications need careful interpretation. RD results show that the observed probability at the points that are forecast to reach the large threshold in specific probability over the entire Taiwan area is lower than the forecast probability (i.e., overestimation), which represents the overall reliability of TAPEX. In contrast, BrS results indicate that probability error for the large threshold at its actual location is small, which represents the average error for all verification samples in the specific area. Summarizing these two verifications shows that the rainfall reaching the larger thresholds seldom happened; when it happened, the PQPF of TAPEX system was often overestimated. These also explain that PQPF tends to higher BrS but higher reliability in the mountainous areas.

On the other hand, some previous studies had similar findings to the results of Table 3. For northward-moving typhoons, the interaction with environmental circulation results a more unstable track than for westward-moving typhoons, thus the forecast track error is larger (Chan and Gray 1982; Holland 1983). For small westward-moving typhoons, the track deflection caused by topographic effect occurs closer to Taiwan than for large typhoons (Tang and Chan 2016), and the amplitude of this deflection is negatively correlated with typhoon size (Lin et al. 2002, 2005; Huang et al. 2016b). This sudden track deflection leads to larger forecast track error for small typhoons. For slow westward-moving typhoons, higher rainfall totals and longer interaction with topography increase the diabatic heating and perturbation around the typhoon (Wang et al. 2012; Hsu et al. 2013). This increases the spread of forecast track in the ensemble system, rather than changing the accuracy of forecast track (Puri et al. 2001). In other words, these typhoon characteristics can

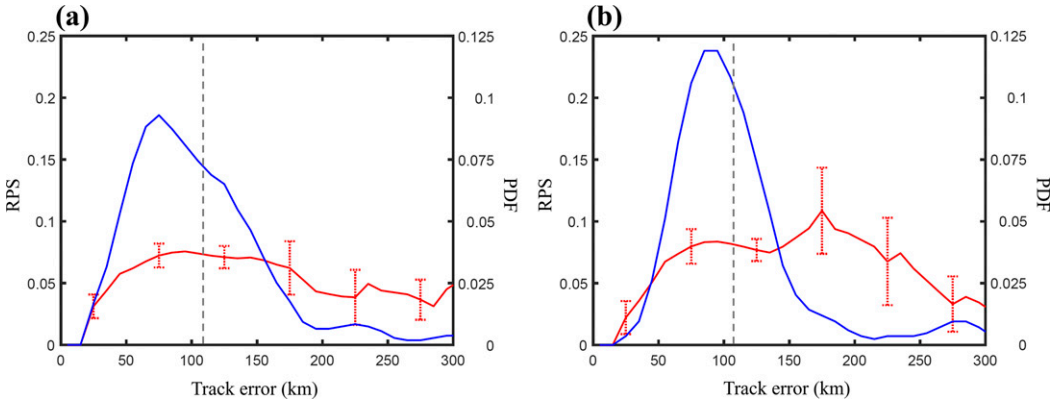

FIG. 15. Average RPS (red curve; left axis) and PDF of sample number (blue curve; right axis) at each value of the first 24-h forecast track error for (a) northward-moving and (b) small westward-moving typhoons. The gray dashed line indicates the average track error in each category. The dotted bar shows one standard deviation of RPS at track errors of 25, 75, 125, 175, 225, and 275 km, which is calculated by the bootstrap method.

influence the forecast track characteristics in an ensemble system, and further affect the PQPF performance.

b. Conclusions

Previous studies suggested that rainfall patterns of typhoons affecting Taiwan and elsewhere are related to characteristics of the typhoon (Yen et al. 2011; Su et al. 2012; Chang et al. 2013; Hsu et al. 2013; Lee et al. 2013; Hong et al. 2015). However, no previous study used the perspective of probability forecast to analyze features of precipitation forecasting and error distribution for different types of typhoons. The potential mechanisms of forecast errors are also largely unexplored. The results of this study show that the PQPF also exhibits a significant difference among typhoons with different characteristics. On average, the PQPFs for westward-moving, large-size, or slow-moving typhoons have higher reliability, higher discrimination ability, and lower probability error for most rainfall thresholds than those for northward-moving, small-size, or fast-moving typhoons, except the reliabilities are similar between fast- and slow-moving typhoons.

The relationship between PQPF error and forecast track error is different among the different types of typhoons. Excluding the forecast with excessive track error, the PQPF error for northward-moving, small, or slow typhoons is more sensitive to forecast track error than for westward-moving, large, or fast typhoons. The probability error for northward-moving typhoons is higher than that for westward-moving typhoons. This is because northward-moving typhoons have a larger track error, which is positively correlated with their RPS. Furthermore, most northward-moving typhoons pass through Taiwan from the eastern coast, and the topography of Taiwan is oriented south–north, which leads to a higher RPS distributed in eastern Taiwan. For the different-size typhoons, because small westward-moving typhoons may have a large track deflection caused by the topographic effect, the mean track error and ensemble spread are larger than those for large typhoons and are positively correlated with their RPS. The higher RPS is distributed in central Taiwan for small typhoons, which is attributed to the overlap of probability errors for all thresholds and the stronger topographic effect over the Central Mountain Range. For typhoons with different forward speeds, the mean track error is not different between fast and slow typhoons, but the track spread is significantly higher for slow typhoons. Thus, different rainfall patterns can be more completely considered for slow typhoons, especially as they leaving Taiwan after the structure of typhoon is destroyed by the topography, which causes slow typhoons to have lower probability error in western Taiwan. In summary, PQPF error is dominated by the track error for typhoons with different headings but the track spread for typhoons with different forward speeds, and is sensitive to both the track error and spread for typhoons with different sizes.

In general, regardless of typhoon characteristics, typhoons with smaller mean track error but substantial track spread can produce the PQPF with smaller probability error, which is similar to the finding of Buizza et al. (1999). However, once the forecast track error is excessive, increases in the ensemble spread can partly reduce the probability error, especially for typhoons with a heading different to the orientation of major topographic feautures (e.g., westward-moving typhoons and Taiwan's topography). The role of the track spread on PQPF performance depends on the typhoon mean track error and track orientation.

This is the first study to systematically analyze PQPF performance for different types of typhoons and summarize the relationships among probability errors, typhoon characteristics, and forecast track errors. Although only 19 typhoons were analyzed, the results of this study can provide guidance for potential modification of the operational PQPF and a reference to improve the evaluation policy for PQPF users and applications when different types of typhoons affect an island with steep topography (e.g., Taiwan, Japan, and the Philippines). Relative to QPF, the PQPF is not a deterministic forecast and is less influenced by any single extreme event (e.g., mesoscale convective system). Rather, it considers the accuracy of the ensemble mean and the substantial ensemble spread. Therefore, the overall PQPF performance is difficult to be improved by the single member or event improvement. According to the results in this study, statistical corrections that consider typhoon characteristics (e.g., ensemble typhoon QPF model; Hong et al. 2015) should play an important role to improve PQPF.

Acknowledgments. The authors thank the anonymous reviewers who helped to improve this paper, and also thank the TTFRI and CWB for providing the ensemble forecast and observation data. This study is supported by the National Space Organization (NSPO) of the Republic of China (Taiwan) via the American Institute in Taiwan (AIT) under NSPO-UCAR (University Corporation for Atmospheric Research) AIT-TECRO (Taipei Economic and Cultural Representative Office) Agreement Implementing Arrangement 5, the Ministry of Science and Technology under Grant 106-2917-I-564-047, the National Taiwan University under Grant G1003, and CWB under Grants MOTC-CWB-103-M-04 and MOTC-CWB-104-M-06. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

REFERENCES

Abramowitz, M., and I. A. Stegun, 1964: *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* National Bureau of Standards Applied Mathematics Series, Vol. 55, Dover Publications, 1046 pp.

Betts, A. K., and M. J. Miller, 1986: A new convective adjustment scheme. Part II: Single column tests using GATE wave, BOMEX, ATEX and arctic air-mass data sets. *Quart. J. Roy. Meteor. Soc.*, **112**, 693–709, https://doi.org/10.1002/QJ.49711247308.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3, https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.

Buizza, R., T. Petroliagis, T. Palmer, J. Barkmeijer, M. Hamrud, A. Hollingsworth, A. Simmons, and N. Wedi, 1998: Impact of model resolution and ensemble size on the performance of an ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **124**, 1935–1960, https://doi.org/10.1002/qj.49712455008.

——, M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, https://doi.org/10.1002/qj.49712556006.

Chan, J. C. L., and W. M. Gray, 1982: Tropical cyclone movement and surrounding flow relationships. *Mon. Wea. Rev.*, **110**, 1354–1374, https://doi.org/10.1175/1520-0493(1982)110<1354:TCMASF>2.0.CO;2.

Chang, C. P., T. C. Yeh, and J. M. Chen, 1993: Effects of terrain on the surface-structure of typhoons over Taiwan. *Mon. Wea. Rev.*, **121**, 734–752, https://doi.org/10.1175/1520-0493(1993)121<0734:EOTOTS>2.0.CO;2.

——, Y. T. Yang, and H. C. Kuo, 2013: Large increasing trend of tropical cyclone rainfall in Taiwan and the roles of terrain. *J. Climate*, **26**, 4138–4147, https://doi.org/10.1175/JCLI-D-12-00463.1.

Chang, F. J., Y. M. Chiang, M. J. Tsai, M. C. Shieh, K. L. Hsu, and S. Sorooshian, 2014: Watershed rainfall forecasting using neuro-fuzzy networks with the assimilation of multi-sensor information. *J. Hydrol.*, **508**, 374–384, https://doi.org/10.1016/j.jhydrol.2013.11.011.

Chang, H. L., H. L. Yuan, and P. L. Lin, 2012: Short-range (0-12 h) PQPFs from time-lagged multimodel ensembles using LAPS. *Mon. Wea. Rev.*, **140**, 1496–1516, https://doi.org/10.1175/MWR-D-11-00085.1.

——, S. C. Yang, H. L. Yuan, P. L. Lin, and Y. C. Liou, 2015: Analysis of the relative operating characteristic and economic value using the LAPS ensemble prediction system in Taiwan. *Mon. Wea. Rev.*, **143**, 1833–1848, https://doi.org/10.1175/MWR-D-14-00189.1.

Chen, B. F., R. Elsberry, and C. S. Lee, 2016: Synoptic controls of outer mesoscale convective systems with high impact rainfall in western North Pacific tropical cyclones. *Asia-Pac. J. Atmos. Sci.*, **52**, 11–23, https://doi.org/10.1007/s13143-015-0085-2.

Chen, D. Y. C., C. C. Chiang, L. R. Hwang, M. C. Wu, and L. Feng, 2014: The development of quantitative precipitation forecast techniques. Tech. Rep., Taiwan Typhoon and Flood Research Institute of National Applied Research Laboratories, 68 pp.

Chen, T. C., and C. C. Wu, 2016: The remote effect of Typhoon Megi (2010) on the heavy rainfall over Northeastern Taiwan. *Mon. Wea. Rev.*, **144**, 3109–3131, https://doi.org/10.1175/MWR-D-15-0269.1.

Chiao, S., and Y. L. Lin, 2003: Numerical modeling of an orographically enhanced precipitation event associated with Tropical Storm Rachel over Taiwan. *Wea. Forecasting*, **18**, 325–344, https://doi.org/10.1175/1520-0434(2003)018<0325:NMOAOE>2.0.CO;2.

Chien, F. C., Y. C. Liu, and C. S. Lee, 2008: Heavy rainfall and southwesterly flow after the leaving of Typhoon Mindulle (2004) from Taiwan. *J. Meteor. Soc. Japan*, **86**, 17–41, https://doi.org/10.2151/jmsj.86.17.

Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367–374, https://doi.org/10.1175/1520-0493(1959)087<0367:AOOAS>2.0.CO;2.

Emanuel, K., 2005: Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, **436**, 686–688, https://doi.org/10.1038/nature03906.

Epstein, E. S., 1969: A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor.*, **8**, 985–987, https://doi.org/10.1175/1520-0450(1969)008<0985:ASSFPF>2.0.CO;2.

Fang, X. Q., and Y. H. Kuo, 2013: Improving ensemble-based quantitative precipitation forecasts for topography-enhanced typhoon heavy rainfall over Taiwan with a modified

probability-matching technique. *Mon. Wea. Rev.*, **141**, 3908–3932, https://doi.org/10.1175/MWR-D-13-00012.1.

——, ——, and A. Y. Wang, 2011: The impacts of Taiwan topography on the predictability of Typhoon Morakot's record-breaking rainfall: A high-resolution ensemble simulation. *Wea. Forecasting*, **26**, 613–633, https://doi.org/10.1175/WAF-D-10-05020.1.

Feldmann, K., D. S. Richardson, and T. Gneiting, 2019: Grid- versus station-based postprocessing of ensemble temperature forecasts. *Geophys. Res. Lett.*, **46**, 7744–7751, https://doi.org/10.1029/2019GL083189.

Ge, X. Y., T. Li, S. J. Zhang, and M. Peng, 2010: What causes the extremely heavy rainfall in Taiwan during Typhoon Morakot (2009)? *Atmos. Sci. Lett.*, **11**, 46–50, https://doi.org/10.1002/ASL.255.

Grell, G., J. Dudhia, and D. Stauffer, 1994: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). NCAR Tech. Note NCAR/TN-398+STR, 117 pp.

Hamill, T. M., 1997: Reliability diagrams for multicategory probabilistic forecasts. *Wea. Forecasting*, **12**, 736–741, https://doi.org/10.1175/1520-0434(1997)012<0736:RDFMPF>2.0.CO;2.

Hendricks, E. A., Y. Jin, J. R. Moskaitis, J. D. Doyle, M. S. Peng, C. C. Wu, and H. C. Kuo, 2016: Numerical simulations of Typhoon Morakot (2009) using a multiply nested tropical cyclone prediction model. *Wea. Forecasting*, **31**, 627–645, https://doi.org/10.1175/WAF-D-15-0016.1.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570, https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Holland, G. J., 1983: Tropical cyclone motion: Environmental interaction plus a beta-effect. *J. Atmos. Sci.*, **40**, 328–342, https://doi.org/10.1175/1520-0469(1983)040<0328:TCMEIP>2.0.CO;2.

Hong, J. S., C. T. Fong, L. F. Hsiao, Y. C. Yu, and C. Y. Tzeng, 2015: Ensemble typhoon quantitative precipitation forecasts model in Taiwan. *Wea. Forecasting*, **30**, 217–237, https://doi.org/10.1175/WAF-D-14-00037.1.

Hosmer, D., and S. Lemeshow, 2000: *Applied Logistic Regression*. 2nd ed. John Wiley and Sons, 392 pp.

Hsiao, L.-F., C.-S. Liou, T.-C. Yeh, Y.-R. Guo, D.-S. Chen, K.-N. Huang, C.-T. Terng, and J.-H. Chen, 2010: A vortex relocation scheme for tropical cyclone initialization in advanced research WRF. *Mon. Wea. Rev.*, **138**, 3298–3315, https://doi.org/10.1175/2010MWR3275.1.

——, and Coauthors, 2013: Ensemble forecasting of typhoon rainfall and floods over a mountainous watershed in Taiwan. *J. Hydrol.*, **506**, 55–68, https://doi.org/10.1016/j.jhydrol.2013.08.046.

Hsu, L. H., H. C. Kuo, and R. G. Fovell, 2013: On the geographic asymmetry of typhoon translation speed across the mountainous island of Taiwan. *J. Atmos. Sci.*, **70**, 1006–1022, https://doi.org/10.1175/JAS-D-12-0173.1.

Hsu, W.-R., and A. H. Murphy, 1986: The attributes diagram: A geometrical framework for assessing the quality of probability forecasts. *Int. J. Forecasting*, **2**, 285–293, https://doi.org/10.1016/0169-2070(86)90048-8.

Huang, C. Y., Y. H. Kuo, S. H. Chen, and F. Vandenberghe, 2005: Improvements in typhoon forecasts with assimilated GPS occultation refractivity. *Wea. Forecasting*, **20**, 931–953, https://doi.org/10.1175/WAF874.1.

——, I. H. Wu, and L. Feng, 2016a: A numerical investigation of the convective systems in the vicinity of southern Taiwan associated with Typhoon Fanapi (2010): Formation mechanism of double rainfall peaks. *J. Geophys. Res. Atmos.*, **121**, 12 647–12 676, https://doi.org/10.1002/2016JD025589.

——, C. A. Chen, S. H. Chen, and D. S. Nolan, 2016b: On the upstream track deflection of tropical cyclones past a mountain range: Idealized experiments. *J. Atmos. Sci.*, **73**, 3157–3180, https://doi.org/10.1175/JAS-D-15-0218.1.

Huang, W. K., and J. J. Wang, 2015: Typhoon damage assessment model and analysis in Taiwan. *Nat. Hazards*, **79**, 497–510, https://doi.org/10.1007/s11069-015-1858-8.

Jewson, S., 2004: The problem with the Brier score. arXiv:physics/0401046v1.

Jian, G. J., and J. A. McGinley, 2005: Evaluation of a short-range forecast system on quantitative precipitation forecasts associated with tropical cyclones of 2003 near Taiwan. *J. Meteor. Soc. Japan*, **83**, 657–681, https://doi.org/10.2151/jmsj.83.657.

——, S. L. Shieh, and J. A. McGinley, 2003: Precipitation simulation associated with Typhoon Sinlaku (2002) in Taiwan area using the LAPS diabatic initialization for MM5. *Terr. Atmos. Oceanic Sci.*, **14**, 261–288, https://doi.org/10.3319/TAO.2003.14.3.261(A).

Jiang, H. Y., and J. B. Halverson, 2008: On the differences in storm rainfall from Hurricanes Isidore and Lili. Part I: Satellite observations and rain potential. *Wea. Forecasting*, **23**, 29–43, https://doi.org/10.1175/2007WAF2005096.1.

Kuo, Y.-H., and W. Wang, 1997: Rainfall prediction of Typhoon Herb with a mesoscale model. Preprints, *Workshop on Typhoon Research in the Taiwan Area*, Boulder, CO, National Science Council, 35–45.

Lee, C. S., L. R. Huang, H. S. Shen, and S. T. Wang, 2006: A climatology model for forecasting typhoon rainfall in Taiwan. *Nat. Hazards*, **37**, 87–105, https://doi.org/10.1007/s11069-005-4658-8.

——, Y. C. Liu, and F. C. Chien, 2008: The secondary low and heavy rainfall associated with Typhoon Mindulle (2004). *Mon. Wea. Rev.*, **136**, 1260–1283, https://doi.org/10.1175/2007MWR2069.1.

——, L. R. Huang, L. F. Hsiao, D. Y. C. Chen, and L. Y. Chang, 2012a: A study on the typhoon precipitation forecast in Taiwan. *30th Conf. on Hurricanes and Tropical Meteorology*, Ponte Vedra Beach, FL, Amer. Meteor. Soc., 5C.2, https://ams.confex.com/ams/30Hurricane/webprogram/Paper205087.html.

——, B. F. Chen, and R. L. Elsberry, 2012b: Long-lasting convective systems in the outer region of tropical cyclones in the western North Pacific. *Geophys. Res. Lett.*, **39**, L21812, https://doi.org/10.1029/2012GL053685.

——, L. R. Huang, and D. Y. C. Chen, 2013: The modification of the typhoon rainfall climatology model in Taiwan. *Nat. Hazards Earth Syst. Sci.*, **13**, 65–74, https://doi.org/10.5194/nhess-13-65-2013.

Lin, M. L., and F. S. Jeng, 2000: Characteristics of hazards induced by extremely heavy rainfall in Central Taiwan–Typhoon Herb. *Eng. Geol.*, **58**, 191–207, https://doi.org/10.1016/S0013-7952(00)00058-2.

Lin, Y. L., S. Chiao, T. A. Wang, M. L. Kaplan, and R. P. Weglarz, 2001: Some common ingredients for heavy orographic rainfall. *Wea. Forecasting*, **16**, 633–660, https://doi.org/10.1175/1520-0434(2001)016<0633:SCIFHO>2.0.CO;2.

——, D. B. Ensley, S. Chiao, and C. Y. Huang, 2002: Orographic influences on rainfall and track deflection associated with the passage of a tropical cyclone. *Mon. Wea. Rev.*, **130**, 2929–2950, https://doi.org/10.1175/1520-0493(2002)130<2929:OIORAT>2.0.CO;2.

——, S. Y. Chen, C. M. Hill, and C. Y. Huang, 2005: Control parameters for the influence of a mesoscale mountain range on

cyclone track continuity and deflection. *J. Atmos. Sci.*, **62**, 1849–1866, https://doi.org/10.1175/JAS3439.1.

Liu, J. G., and Z. H. Xie, 2014: BMA probabilistic quantitative precipitation forecasting over the Huaihe Basin using TIGGE multimodel ensemble forecasts. *Mon. Wea. Rev.*, **142**, 1542–1555, https://doi.org/10.1175/MWR-D-13-00031.1.

Lonfat, M., F. D. Marks, and S. Y. S. Chen, 2004: Precipitation distribution in tropical cyclones using the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager: A global perspective. *Mon. Wea. Rev.*, **132**, 1645–1660, https://doi.org/10.1175/1520-0493(2004)132<1645:PDITCU>2.0.CO;2.

Mason, S. J., and N. E. Graham, 1999: Conditional probabilities, relative operating characteristics, and relative operating levels. *Wea. Forecasting*, **14**, 713–725, https://doi.org/10.1175/1520-0434(1999)014<0713:CPROCA>2.0.CO;2.

Matyas, C. J., 2006: Relating tropical cyclone rainfall patterns to storm size. *27th Conf. on Hurricanes and Tropical Meteorology*, Monterey, CA, Amer. Meteor. Soc., 2B.6.

Peng, X. D., J. F. Fei, X. G. Huang, and X. P. Cheng, 2017: Evaluation and error analysis of official forecasts of tropical cyclones during 2005-14 over the western North Pacific. Part I: Storm tracks. *Wea. Forecasting*, **32**, 689–712, https://doi.org/10.1175/WAF-D-16-0043.1.

Puri, K., J. Barkmeijer, and T. N. Palmer, 2001: Ensemble prediction of tropical cyclones using targeted diabatic singular vectors. *Quart. J. Roy. Meteor. Soc.*, **127**, 709–731, https://doi.org/10.1002/qj.49712757222.

Richardson, D. S., 2001: Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quart. J. Roy. Meteor. Soc.*, **127**, 2473–2489, https://doi.org/10.1002/qj.49712757715.

Roulston, M. S., 2007: Performance targets and the Brier score. *Meteor. Appl.*, **14**, 185–194, https://doi.org/10.1002/met.21.

Ruiz, J., C. Saulo, and E. Kalnay, 2009: Comparison of methods used to generate probabilistic quantitative precipitation forecasts over South America. *Wea. Forecasting*, **24**, 319–336, https://doi.org/10.1175/2008WAF2007098.1.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., https://doi.org/10.5065/D68S4MVH.

Su, S. H., H. C. Kuo, L. H. Hsu, and Y. T. Yang, 2012: Temporal and spatial characteristics of typhoon extreme rainfall in Taiwan. *J. Meteor. Soc. Japan*, **90**, 721–736, https://doi.org/10.2151/jmsj.2012-510.

Tang, C. K., and J. C. L. Chan, 2016: Idealized simulations of the effect of Taiwan topography on the tracks of tropical cyclones with different sizes. *Quart. J. Roy. Meteor. Soc.*, **142**, 793–804, https://doi.org/10.1002/qj.2681.

Tsai, C. C., S. C. Yang, and Y. C. Liou, 2014: Improving quantitative precipitation nowcasting with a local ensemble transform Kalman filter radar data assimilation system: Observing system simulation experiments. *Tellus*, **66A**, 21804, https://doi.org/10.3402/TELLUSA.V66.21804.

Wang, C. C., 2015: The more rain, the better the model performs—The dependency of quantitative precipitation forecast skill on rainfall amount for typhoons in Taiwan. *Mon. Wea. Rev.*, **143**, 1723–1748, https://doi.org/10.1175/MWR-D-14-00137.1.

——, H. C. Kuo, Y. H. Chen, H. L. Huang, C. H. Chung, and K. Tsuboki, 2012: Effects of asymmetric latent heating on typhoon movement crossing Taiwan: The case of Morakot (2009) with extreme rainfall. *J. Atmos. Sci.*, **69**, 3172–3196, https://doi.org/10.1175/JAS-D-11-0346.1.

——, ——, T.-C. Yeh, C.-H. Chung, Y.-H. Chen, S.-Y. Huang, Y.-W. Wang, and C.-H. Liu, 2013: High-resolution quantitative precipitation forecasts and simulations by the Cloud-Resolving Storm Simulator (CReSS) for Typhoon Morakot (2009). *J. Hydrol.*, **506**, 26–41, https://doi.org/10.1016/j.jhydrol.2013.02.018.

——, S. Y. Huang, S. H. Chen, C. S. Chang, and K. Tsuboki, 2016: Cloud-resolving typhoon rainfall ensemble forecasts for Taiwan with large domain and extended range through time-lagged approach. *Wea. Forecasting*, **31**, 151–172, https://doi.org/10.1175/WAF-D-15-0045.1.

Wilks, D. S., 1997: Resampling hypothesis tests for autocorrelated fields. *J. Climate*, **10**, 65–82, https://doi.org/10.1175/1520-0442(1997)010<0065:RHTFAF>2.0.CO;2.

Wilson, E. B., and M. M. Hilferty, 1931: The distribution of chi square. *Proc. Natl. Acad. Sci. USA*, **17**, 684–688, https://doi.org/10.1073/pnas.17.12.684.

Wu, C. C., and Y. H. Kuo, 1999: Typhoons affecting Taiwan: Current understanding and future challenges. *Bull. Amer. Meteor. Soc.*, **80**, 67–80, https://doi.org/10.1175/1520-0477(1999)080<0067:TATCUA>2.0.CO;2.

——, T. H. Yen, Y. H. Kuo, and W. Wang, 2002: Rainfall simulation associated with typhoon herb (1996) near Taiwan. Part I: The topographic effect. *Wea. Forecasting*, **17**, 1001–1015, https://doi.org/10.1175/1520-0434(2003)017<1001:RSAWTH>2.0.CO;2.

——, K. K. W. Cheung, and Y. Y. Lo, 2009: Numerical study of the rainfall event due to the interaction of Typhoon Babs (1998) and the northeasterly monsoon. *Mon. Wea. Rev.*, **137**, 2049–2064, https://doi.org/10.1175/2009MWR2757.1.

——, T. H. Yen, Y. H. Huang, C. K. Yu, and S. G. Chen, 2016: Statistical characteristic of heavy rainfall associated with typhoons near Taiwan based on high-density automatic rain gauge data. *Bull. Amer. Meteor. Soc.*, **97**, 1363–1375, https://doi.org/10.1175/BAMS-D-15-00076.1.

WWRP/WGNE, 2015: Joint working group on forecast verification research. WWRP/WCRP, accessed 6 June 2019, http://www.cawcr.gov.au/projects/verification/.

Xu, T., J. Li, X. F. Wang, H. Yu, B. Chen, and Y. H. Yang, 2010: The estimation of TC quantitative precipitation forecast affecting Shanghai in 2009. *Third WMO Int. Conf. on Quantitative Precipitation Estimation and Quantitative Precipitation Forecasting and Hydrology*, Nanjing, China, WMO, A2, https://ral.ucar.edu/jnt/events/qpe_qpf_III/Abstract_Doc_Final_7Oct10.pdf.

Yang, M. J., and Q. C. Tung, 2003: Evaluation of rainfall forecasts over Taiwan by four cumulus parameterization schemes. *J. Meteor. Soc. Japan*, **81**, 1163–1183, https://doi.org/10.2151/jmsj.81.1163.

——, D. L. Zhang, and H. L. Huang, 2008: A modeling study of Typhoon Nari (2001) at landfall. Part I: Topographic effects. *J. Atmos. Sci.*, **65**, 3095–3115, https://doi.org/10.1175/2008JAS2453.1.

Yang, T. H., G. D. Hwang, C. C. Tsai, and J. Y. Ho, 2016: Using rainfall thresholds and ensemble precipitation forecasts to issue and improve urban inundation alerts. *Hydrol. Earth Syst. Sci.*, **20**, 4731–4745, https://doi.org/10.5194/hess-20-4731-2016.

Yen, T. H., C. C. Wu, and G. Y. Lien, 2011: Rainfall simulations of Typhoon Morakot with controlled translation speed based on EnKF data assimilation. *Terr. Atmos. Oceanic Sci.*, **22**, 647–660, https://doi.org/10.3319/TAO.2011.07.05.01(TM).

Yu, Z. F., Y. Q. Wang, H. M. Xu, N. Davidson, Y. D. Chen, Y. M. Chen, and H. Yu, 2017: On the relationship between intensity and rainfall distribution in tropical cyclones making landfall over China. *J. Appl. Meteor. Climatol.*, **56**, 2883–2901, https://doi.org/10.1175/JAMC-D-16-0334.1.

Yuan, H. L., S. L. Mullen, X. Gao, S. Sorooshian, J. Du, and H. M. H. Juang, 2005: Verification of probabilistic quantitative precipitation forecasts over the southwest United States during winter 2002/03 by the RSM ensemble system. *Mon. Wea. Rev.*, **133**, 279–294, https://doi.org/10.1175/MWR-2858.1.

——, J. A. McGinley, P. J. Schultz, C. J. Anderson, and C. G. Lu, 2008: Short-range precipitation forecasts from time-lagged multimodel ensembles during the HMT-West-2006 campaign. *J. Hydrometeor.*, **9**, 477–491, https://doi.org/10.1175/2007JHM879.1.

——, C. G. Lu, J. A. McGinley, P. J. Schultz, B. D. Jamison, L. Wharton, and C. J. Anderson, 2009: Evaluation of short-range quantitative precipitation forecasts from a time-lagged multimodel ensemble. *Wea. Forecasting*, **24**, 18–38, https://doi.org/10.1175/2008WAF2007053.1.