Evaluation, Calibration and Economic Value Analysis of the Probabilistic Quantitative Precipitation Forecasts (PQPFs) from WRF Ensemble Prediction System in Taiwan Area

Hui-Ling Chang[1], Kuan-Ju Chen[1], Hochin Chang[1], Chia-Jung Wu[1], Yu-Jhen Liu[1], Tsorung Wang[1], Jing-Shan Hong[1], and Shu-Chih Yang[2]

1. Central Weather Bureau, Taipei, Taiwan
2. Department of Atmospheric Sciences, National Central University, Jhong-Li, Taiwan
Yu-Jhen Liou1

Abstract

This study aims to develop the probabilistic quantitative precipitation forecasts (PQPFs) associated with typhoons by utilizing the WRF ensemble prediction system (WEPS) developed by Central Weather Bureau (CWB). The ultimate goal is to provide more valuable precipitation forecast products associated with typhoons based on the WEPS and help users optimize their decision making and obtain the maximum economic value by using the PQPFs. The research results include evaluation, calibration, and economic value (EV) analysis of the PQPFs from WEPS.

Forecast evaluation results showed that the ensemble spread can well represent the forecast uncertainties. The PQPFs have good discriminating ability but also obvious biases. The reliability and discrimination decrease with increasing precipitation thresholds. Verification results from different areas show that the reliability over plain areas is better than that over mountain areas, but the discrimination over mountain areas excels that over plain areas. Taken as a whole, the forecast skill over mountain areas surpasses that over plain areas. This is because the terrain-locking effect leads to higher predictability of typhoon rainfall over windward slopes than over plain areas. Therefore, the discrimination and forecast skill of WEPS over mountain areas are better. The non-parametric Mann–Whitney test indicated that the greater forecast ability over mountain areas than over plain areas is statistically significant.

Because WEPS has obvious wet bias, we adopted the Linear Regression (LR) and Artificial Neural Network (ANN) techniques to calibrate the PQPFs. Calibration results show that the calibration effect of LR and ANN methods are quite similar; but the ANN needs more computing time and training samples to establish a stable calibration relationship. Economic value analysis shows that compared with deterministic ensemble forecast products, the probabilistic ensemble forecast can obviously offer more EV to a wider range of users.

Keywords: probabilistic quantitative precipitation forecasts (PQPFs), forecast bias, calibration, economic value (EV)

1. Introduction

This study aims to develop PQPFs associated with typhoons using WRF ensemble prediction system (WEPS) developed by Central Weather Bureau (CWB). PQPF means probabilistic quantitative precipitation forecasts. The ultimate goal is to provide more reliable and valuable precipitation forecasts for typhoons.

Three scientific issues were addressed in this study: 1) what's the impact of calibration methods on calibration results? 2) What's the impact of terrain-locking effect on the difference of forecast ability between mountain and plain areas? 3) What benefit can be obtained if users make decisions based on ensemble probabilistic forecasts instead of deterministic forecast products derived from ensembles.

This paper is organized as follows. The WEPS, study data, and PQPF products are introduced in Section 2. The methodology of economic value analysis, and the deterministic forecast products derived from ensembles are described in Section 3. Section 4 presents the evaluation and calibration results from the WEPS PQPFs. A summary of this study is provided in Section 5.

2. **Model, Data, and PQPF Products**

WRF ensemble prediction system is called WEPS, which has 20 members and is perturbed by different initial states, boundary conditions, and physical parameterizations. Because WEPS has obvious wet bias, we adopted Linear Regression (LR) and Artificial Neural Network (ANN) techniques to calibrate the PQPFs. The data used in this study included 12 typhoon cases during 2013 to 2015 in Taiwan area (Table 1).

Figure 1 shows the 0-24h PQPFs at different thresholds and the corresponding observed probabilities for Typhoon Usage in 2013. We use radar QPEs as observations in this study. It's clear from Fig. 1 that the rainfall patterns of forecasts and observations have good correspondence.

3. *Methodology*

a. Economic value (EV)

When we say that a forecast system can provide economic value to users, it means that users can benefit from making decisions based on the forecast information. For example, the forecast information can help users to lower the cost of preventive action or decrease their losses.

The EV of a forecast system (Richardson 2000) is defined as:

$$EV = \frac{E_{climate} - E_{forecast}}{E_{climate} - E_{perfect}} . \quad (1)$$

where $E_{climate}$, $E_{forecast}$ and $E_{perfect}$ are a user's expected expense. He or she makes a decision based on climatological information, a forecast system, and a perfect forecast system respectively. A perfect forecast system means that it always provides accurate predictions for the occurrence and non-occurrence of a weather event. According to this definition, *EV* can be interpreted as: relative to the use of climatological information, if a perfect forecast system can save a user 100 dollars, then a forecast system with economic value *EV* will save the user 100**EV* dollars. Richardson (2000) further showed that EV can be expressed as:

$$EV = \frac{\min[\bar{o}, r] - FARr(1-\bar{o}) + HR\bar{o}(1-r) - \bar{o}}{\min[\bar{o}, r] - \bar{o}r} . \quad (2)$$

Equation (2) shows that *EV* is related not only to the forecast performance [hit rate (*HR*) and false alarm rate (*FAR*)], but also to the climatological frequency ($\bar{o}$) of a weather event and the cost-loss ratio (*r*) of a user.

b. Deterministic forecasts derived from ensembles

CWB developed some ensemble deterministic forecast products because of the requirements from the general public and the hydrology and water resources agency. For the general public, they usually have difficulty in making decisions based on probabilistic forecasts since they're not sure a forecast probability, say 70%, indicates the event will happen or not. For the hydrological people, they can only use a deterministic forecast as the initial condition for their hydrological model. Therefore, we generated three kinds of ensemble deterministic forecasts products, including ensemble mean (ESM), probability-matched ensemble mean (PM) and QPF percentile (QPFP).

ESM means the rainfall average of 20 members at each grid point. ESM tends to smooth out the rainfall extremes due to the averaging process. For PM (Ebert 2001), we assume the rainfall pattern of ensemble mean is reasonable, and try to keep the rainfall extremes of ensemble members. The PM product is generated as follows. First, we rank the gridded rainfalls from the ensemble mean from largest to smallest. Second, we rank the gridded rainfalls from all members from largest to smallest, and then keep every median rainfall of each group, which has the same number of ensemble members. Finally, match the two histograms to generate the PM product. That is, mapping the rain rates from second step onto the locations from first step. The PM has the same rainfall pattern as ensemble mean and the same frequency distribution of rain rate as ensembles.

QPFP is the rainfall amount by percentile. For example, the 20th percentile is the rainfall value below which 20% of the rainfall forecasts may be found. According to previous studies, the observed rainfall is close to the 80th percentile for typhoon cases in Taiwan area. Therefore, we use the QPFP_80th for comparison of EV in this study.

4. Evaluation, calibration, and EV analysis of the PQPFs from WEPS

a. Spread-skill relationship

Figure 2 shows the spread-skill relationship, which is a critical measure of the quality of ensemble prediction system. We can see from this scatter plot that ensemble spread and RMSE of ensemble mean are highly correlated with a correlation of 0.9, which indicates a good spread-skill relationship. That is, the ensemble spread can well represent the forecast uncertainties or forecast errors.

b. Discrimination and Reliability

Figure 3 shows the area under the relative operating characteristic curve (ROC area) from WEPS 0-24 h PQPFs at different thresholds for different areas, including total, land, mountain, and plain areas. The ROC areas are all greater than 0.7 at different thresholds for different areas, which indicates skillful discrimination.

Figure 4 shows the reliability diagrams for different areas. The blue and red curves show the results from before and after calibration respectively. Except for Fig. 4(e) using Artificial Neural Network (ANN) technique, all adopted Linear Regression calibration method. We can see from the blue curves that before calibration the PQPFs for different areas display varying degree of wet biases. In fact, we also compared the reliability diagrams at different thresholds and found that the bias grows with increasing threshold. The red curves indicate that the biases were corrected after calibration. The calibration effects of LR and ANN are quite similar [Fig. 4(d) vs. (e)], but the ANN needs more computing time and training samples to establish a stable calibration relationship (章等2018).

c. Comparison of Forecast Ability over Mountain and Plain Areas

From the analysis of reliability and discrimination, we can see that the reliability over plain is better than that over mountain, but the result of discrimination is just opposite. This result reflects that the PQPFs over mountain have better rainfall pattern but also more obvious wet bias than that over plain. Fig. 5 shows the Brier skill scores (BrSS) at different thresholds over mountain and plain areas. Blue and red lines show the results from before and after LR calibration respectively. The forecast skill over mountain surpasses that over plain. That is because the terrain-locking effect leads to higher predictability of typhoon rainfall over windward slopes than over plain area.

We also calculated the median correlation difference between mountain and plain areas, and its 95% confidence interval (CI) for 20 members (1~20) and ensemble mean (E) (Fig. 6). The median correlation difference and its 95% CI are all greater than zero, which indicates that the rainfall pattern over mountain is better than that over plain.

Having seen the difference in median correlation between these two areas, we wonder whether the difference is statistically significant or not; therefore, the Mann–Whitney test was conducted at a specified level of significance ($\alpha=0.1$) with the following hypotheses:

H_0: median $correlation_1$ = median $correlation_2$

H_A: median $correlation_1$ = median $correlation_2$

where the subscripts 1 and 2 refer to mountain and plain area respectively.

Based on the samples of typhoon cases in this study, 95% of members have the p-value < 0.1. That is, the probability of no difference in median correlation between both areas is < 10%. In other words, the better rainfall patter over mountain than over plain is statistically significant at the 10% test level (張等 2018).

d. Comparison of EV between probabilistic and deterministic ensemble forecast

Figure 7 shows the EV distributions at different thresholds over mountain and plain areas. The black curve is the max EV curve from PQPF and the three colorful ones are from the three deterministic forecast products derived from WEPS, including ESM, PM, and QPFP products. Fig. 7 indicates that PQPF can obviously offer more EV to a wider range of users than the three deterministic forecasts products. In addition, users with a very small (< 0.1) or a larger (> 0.7) cost-loss ratio can only benefit from making decisions based on PQPF.

5. Conclusions

This study developed the PQPFs associated with typhoons using WEPS developed by CWB, as well as evaluated, calibrated, and performed EV analysis of the PQPFs. The ultimate goal is to provide more valuable typhoon precipitation forecast products and help users optimize their decision making to obtain their maximum economic benefit.

The analysis of spread-skill relationship shows that the ensemble spread of WEPS can well represent the forecast uncertainties, and the PQPFs have good discriminating ability but also obvious forecast bias. The reliability over plain is better than that over mountain, but the result of discrimination is just opposite. Taken as a whole, forecast skill over mountain surpasses that over plain. Mann–Whitney test indicated that the better rainfall pattern over mountain than over plain is statistically significant at the 10% test level. In addition, calibration effect of LR and ANN methods are quite similar; but the ANN needs more computing time and training samples to establish a stable calibration relationship.

Economic value analysis showed that PQPF can obviously offer more EV to a wider range of users than the deterministic forecasts products derived from ensembles. In addition, users with a very small or a larger cost-loss ratio can only benefit from making decisions based on PQPF.

Reference

張惠玲、陳冠儒、吳佳蓉、汪琮、洪景山、楊舒芝，2018：臺灣地區WRF颱風系集降雨機率預報之評估、校正與經濟價值分析-第一部分：預報評估。大氣科學，**46**，71–106。

章鶴群、 陳冠儒、劉豫臻、張惠玲、洪景山、楊舒芝，2018：臺灣地區WRF颱風系集降雨機率預報之評估、校正與經濟價值分析-第二部分：校正。大氣科學，**46**，71–106。

陳冠儒、張惠玲、楊舒芝、洪景山、吳佳蓉、汪琮，2018：臺灣地區 WRF 颱風系集降雨機率預報之評估、校正與經濟價值分析-第三部分：經濟價值分析。大氣科學 (審稿中)。

Chang, H. L., H. Yuan, P. L. Lin, 2012: Short-Range (0-12h) PQPFs from Time-Lagged Multimodel Ensembles Using LAPS. *Mon. Wea. Rev.*, **140**, 1496–1516.

___, S.-C. Yang, H. Yuan, P. L. Lin and Y. C. Liou, 2015:Analysis of relative operating characteristic and economic value using the LAPS ensemble prediction system in Taiwan area. Mon. Wea. Rev., 143, 1833–1848.Lakshmanan, V., J. Crockett, K. Sperow, M. Ba, and L. Xin, 2012: Tuning AutoNowcaster automatically. *Wea. Forecasting*, **27**, 1568–1579.

Ebert, E. Elizabeth, 2001: Ability of a Poor Man's Ensemble to Predict the Probability and Distribution of Precipitation. *Mon. Wea. Rev.*, **129**, 2461–2480.

Richardson, D. S., 2000: Skill and relative economic value of the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **126**, 649–667.

TABLE 1. Typhoon cases in this study.

	Typhoon	Start - end
2013	Soulik (TY 01)	0000 UTC 11 Jul – 0000 UTC 14 Jul
	Trami (TY 02)	0000 UTC 20 Aug – 0000 UTC 22 Aug
	Kong-Rey (TY 03)	0000 UTC 27 Aug – 0000 UTC 30 Aug
	Usage (TY 04)	1800 UTC 19 Sep – 1200 UTC 22 Sep
	Fitow (TY 05)	1800 UTC 04 Oct – 0600 UTC 07 Oct
2014	Matmo (TY 06)	0000 UTC 21 Jul – 1800 UTC 23 Jul
	Fung-Wong (TY 07)	1800 UTC 18 Sep – 1200 UTC 22 Sep
2015	Linfa (TY 08)	1800 UTC 05 Jul – 0600 UTC 08 Jul
	Chan-Hom (TY 09)	1200 UTC 08 Jul – 0600 UTC 11 Jul
	Soudelor (TY 10)	1800 UTC 05 Aug – 0600 UTC 09 Aug
	Goni (TY 11)	0600 UTC 20 Aug – 1800 UTC 23 Aug
	Dujuan (TY 12)	1800UTC 26 Sep – 1800 UTC 29 Sep

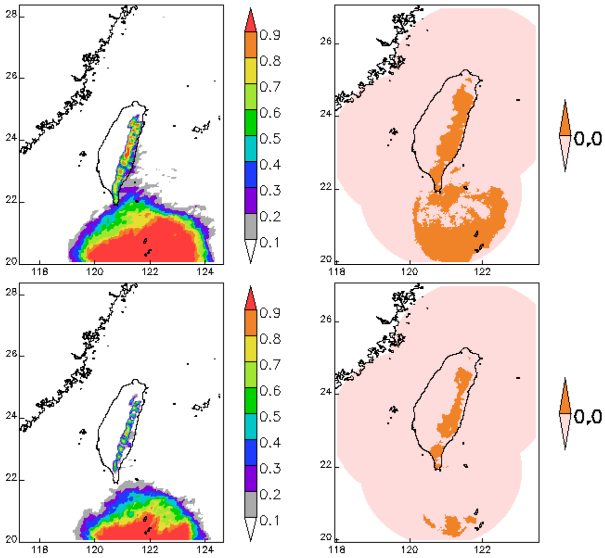

FIG. 1 Distribution of (left) WEPS 0–24-h PQPFs and (right) radar QPE (used as truth) probabilities at thresholds (a) 80, and (b) 130 mm (24 h)$^{-1}$ ending at 0600 UTC 21 Sep 2013.

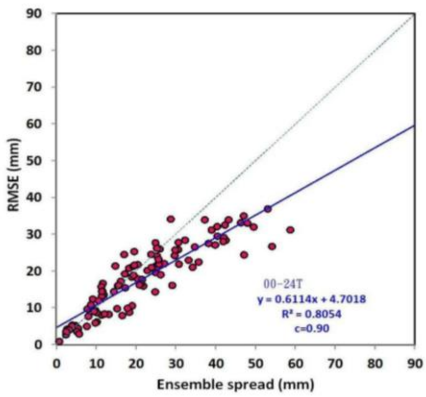

FIG. 2 Scatter plots of the RMSE of ensemble mean against the ensemble spread for all typhoon cases using WEPS 0–24-h QPFs.

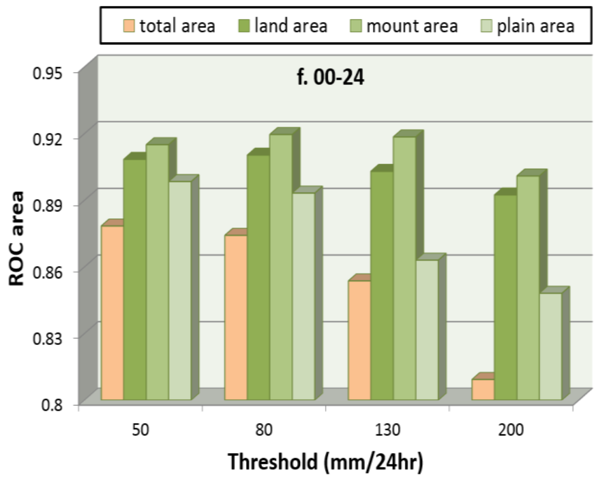

FIG. 3 ROC area at different thresholds for WEPS 0–24-h PQPFs over different areas.

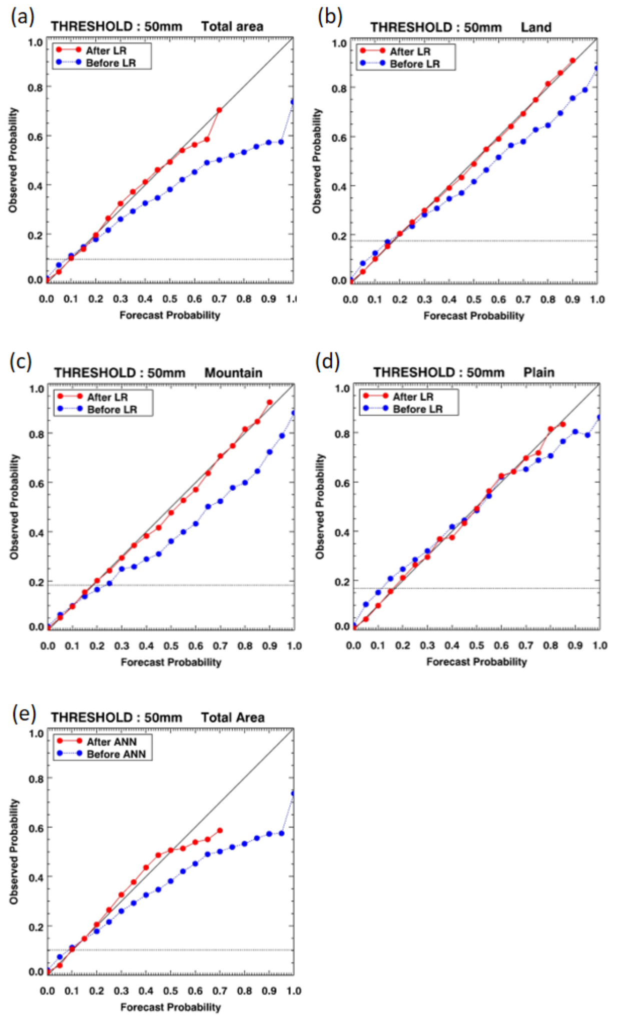

FIG. 4 Reliability diagrams for WEPS 0–24-h PQPFs over (a) total, (b) land, (c) mountain, (d) plain, and (e) total areas. Except for (e) using the ANN technique, all adopted the LR calibration method. The horizontal dashed line indicates the sample climatology frequency. Blue and red lines show the results from before and after calibration.

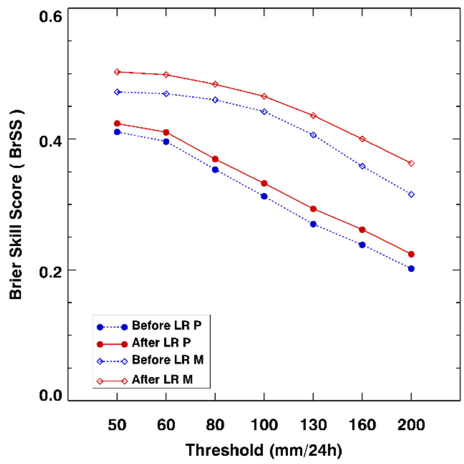

FIG. 5 Brier skill scores at different thresholds for WEPS 0–24-h PQPFs over mountain (M) and plain (P) areas. Blue and red lines show the results from before and after LR calibration.

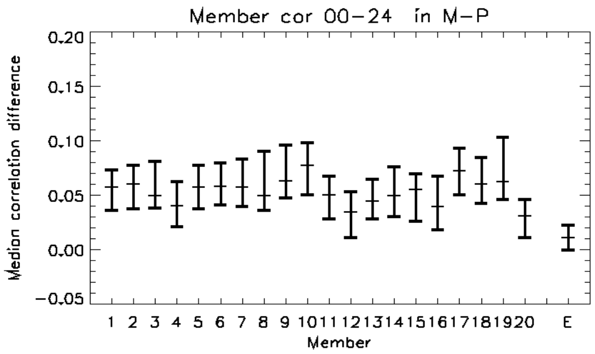

FIG. 6 Median correlation difference between mountain and plain areas, and its 95% confidence interval (CI) using bootstrap method (resampling 10,000 times) for WEPS 20 members (1~20) and ensemble mean (E).

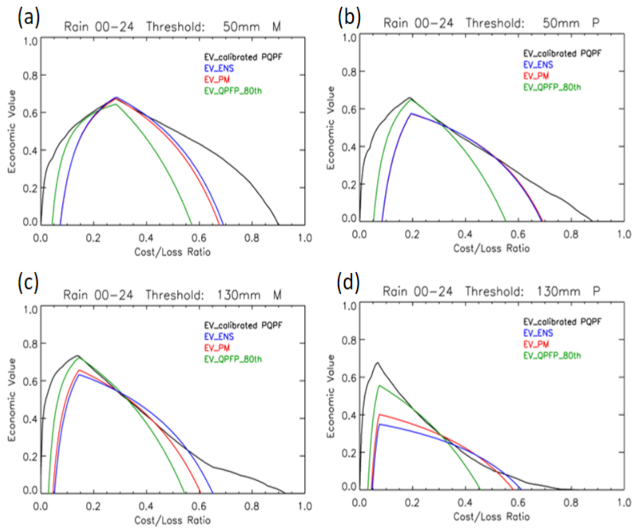

FIG. 7 Distribution of economic value for WEPS 0–24-h calibrated PQPFs and three ensemble deterministic forecasts, including ensemble mean (ENS), PM, and QPFP_80th, at different thresholds [50 and 130mm $(24h)^{-1}$] over mountain (M) and plain (P) areas.