Analysis of relative operating characteristic and economic value using the LAPS ensemble prediction system in Taiwan area

Hui-Ling Chang[1,2], Shu-Chih Yang[2], Huiling Yuan[3,4], Pay-Liam Lin[2] ,Yu-Chieng Liou [2],
Chia-Rong Chen[1], and Wen-Ho Wang[1]

1. Meteorological Satellite Center, Central Weather Bureau, Taipei, Taiwan;
2. Department of Atmospheric Sciences, National Central University, Jhong-Li, Taiwan
3. School of Atmospheric Sciences, and Key Laboratory of Mesoscale Severe Weather
/ Ministry of Education, Nanjing University, Nanjing, Jiangsu, China;
4. Jiangsu Collaborative Innovation Center for Climate Change, China

Abstract

Measurement of the usefulness of numerical weather prediction considers not only the forecast quality but also the economic value (EV) of their actual use in the daily decision-making process of users. Discrimination ability, which can be assessed by the relative operating characteristic (ROC), is one way to evaluate the quality of forecasts and is closely related to the EV provided by the same forecast system.

Focusing on short-range probabilistic quantitative precipitation forecasts (PQPFs) for typhoons, this study demonstrates the consistent and strongly related characteristics of ROC and EV based on the Local Analysis and Prediction System (LAPS) operated at the Central Weather Bureau (CWB) in Taiwan. Sensitivity experiments on ROC and EV show that the potential EV provided by a forecast system is mainly determined by the ROC of the same system. The ROC and maximum EV (EV_{max}) of a forecast system are insensitive to calibration, but the optimal probability threshold to achieve the EV_{max} becomes more reliable after calibration. In addition, ensemble probabilistic forecasts (EPFs) have the advantage over deterministic forecasts (DFs) in respect of both ROC and EV, and such an advantage grows with increasing precipitation intensity.

Keywords: relative operating characteristic (ROC), economic value, calibration, ensemble probabilistic forecasts (EPFs)

1. Introduction

To measure the usefulness of weather forecasts, one should consider not only the forecast quality but also the economic benefit (or economic value) associated with their actual use in the daily decision-making process of users. Relative operating characteristic (ROC) is one verification method to evaluate the discrimination ability of a forecast system, which is one kind of characteristic for forecast quality. Economic value (EV) is used to evaluate whether users can benefit from making decisions based on the forecast information, that is, if the forecast information can help users to lower the cost of preventive action or decrease their losses.

Several past studies indicated that ROC is closely related to the EV provided by the same forecast system. However, because of the complexity of the EV formula, their relationship has rarely been discussed in detail. This study (Chang et al. 2014) investigates the consistent and related characteristics between the ROC and EV based on the Local Analysis and Prediction System (LAPS) ensemble prediction system (EPS), and focuses on the short-range (0–6 h) probabilistic forecasts for typhoon heavy rainfall. Two scientific issues are explored in this study: (1) How does the ROC of a forecast system influence the maximum EV (EV_{max}) provided by the same system? (2) Will the EV_{max} obtained by users be decreased when adopting biased forecasts?

This paper is organized as follows: Section 2 introduces the LAPS EPS and data. Section 3 describes the methodology for computing ROC and EV. Section 4 presents the results from the sensitivity experiment, including the effects of terrain, calibration, and forecast uncertainties on ROC and EV, to answer the scientific issues. Finally, a summary is provided in Section 5.

2. LAPS EPS and Data

The 0-6 h probabilistic quantitative precipitation forecasts (PQPFs) used in this study are generated from ensemble forecasts based on the LAPS. By adopting diabatic data assimilation, the LAPS mitigates the spin-up problem and perform reasonable precipitation forecasts during the early stage of a forecast period. The LAPS PQPFs are operationally generated every 3 h at the Central Weather Bureau (CWB) in Taiwan. To produce four basic multi-model ensemble members, the LAPS EPS adopts two types of background states, including the forecasts of the Global Forecast System (GFS) at the National Centers for Environmental Prediction (NECP) and the non-hydrostatic forecast system (NFS) at the CWB, to construct two sets of analysis fields with a horizontal resolution of 9 km, and then initializes two mesoscale models, including the MM5 and WRF/ARW models. For each of these basic members, the EPS adopts two more members initialized from the analyses generated three and six hours earlier. Therefore, in total, 12 time-lagged multi-model members are available and the PQPFs used in this study are derived from this 12-member LAPS EPS. Chang et al. (2012) showed that the LAPS EPS has a good spread-skill relationship and skillful discrimination ability, and thus can be regarded as an EPS with good quality and predictive capability. The data used in this study (same as in Chang et al. 2012) for evaluating the ROC and EV include a total of 148 cases of 0–6 h PQPFs based on all typhoon cases in 2008 and 2009.

A calibration method based on linear regression (Yuan et al. 2008) has been used to calibrate the wet-biased PQPFs. Chang et al. (2012) show that this calibration method successfully corrects the wet bias and improves the post-processing forecast skill.

3. Methodology

a. Relative operating characteristic (ROC)

The ROC is a plot of hit rate (HR) against false alarm rate (FAR) for a set of probability thresholds (P_t). The P_t means that an event is regarded as “will occur” when the forecast probability (P_f) is more than or equal to this threshold. The choice of a P_t converts the probabilistic forecast to a deterministic one. By varying P_t, a sequence of values of HR [= h/(h + m)] and FAR [=f/(f + c)] can be derived using several 2x2 contingency tables (Table 1) and the ROC can be obtained. The area under the ROC curve is called ROC area, which can be used to evaluate the discrimination ability of a forecast system.

b. Economic value (EV)

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 the expected expenses of a user who takes preventive action based on the climatological information, a forecast system, and a perfect deterministic forecast system, respectively. According to the above definition, the EV can be interpreted as the relative performance taking the climatological information as a baseline. For example, if a perfect forecast can save the 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 FAR and HR of a forecast system but also to the climatological frequency ($\bar{o}$) of a weather event and the cost-loss ratio (*r*) of a user. Since ROC is defined by FAR and HR, eq. (2) also indicates that EV and ROC are associated (Zhu et al. 2002) .

4. Sensitivity of ROC and EV

a. Effect of Terrain

Due to the terrain-locking effect of Central Mountain Range (CMR) on typhoon rainfall, the predictability of typhoon rainfall over mountain areas is usually higher than that over plain areas. The LAPS PQPFs over most mountain areas are much more skillful than that over the plain areas.

In this study, the climatology frequency ($\bar{o}$) is computed according to all observations in the LAPS EPS domain and the EV values are computed based on same $\bar{o}$ without consideration of its spatial distribution. To investigate the impact of the spatial distribution of $\bar{o}$, we divide the land area into two areas, the mountain (terrain height ≥ 500 m) and plain areas (terrain height < 500 m), and compare the ROC and EV over both areas to explore how the terrain affects the ROC and EV of LAPS PQPFs.

The ROC areas from the LAPS 0–6 h PQPFs over the mountain and plain areas at various thresholds (Fig. 1a) shows that the better discrimination ability (i.e., larger ROC area) is obtained over the mountain area than that over the plain area. The EV_{max} curves over the mountain and plain areas (Fig. 1b) are significantly different for the users with larger r. For example, for the users with $r = 8/12$, the LAPS PQPFs has no economic benefit over the plain area, but it can still offer the EV value of 30% to the users over the mountain area. In Fig. 1b, dots denote the real optimal P_t that users must adopt to achieve their EV_{max} when making decisions based on the LAPS PQPFs. The theoretical P_t is not shown since the theoretical value should locate on the diagonal because the optimal P_t for a particular user is equal to his/her r if adopting perfectly reliable forecasts (Murphy 1977). Comparing the real optimal P_t and theoretical value, we can infer the forecast bias of LAPS PQPFs over the plain and mountain areas. As the real optimal P_t over the plain area clearly deviate from the diagonal and locate above it, we can infer that there is wet bias over the plain area. In comparison, there is almost no forecast bias (near perfectly reliable) over the mountain area.

The EV provided by a forecast system is related not only to the discrimination ability (i.e., HR and FAR) of this system, but also to the $\overline{o}$ of a particular weather event and the r of a user (Eq. (2)). We wonder whether the difference in EV_{max} between the mountain and plain areas comes from the difference in the discrimination ability of LAPS PQPFs or the $\overline{o}$ for a particular rainfall event. For this purpose, two sensitivity experiments are conducted: (1) the $\overline{o}$ over the mountain area is replaced by that over the plain area; (2) the discrimination ability (i.e., the values of HR and FAR) over the mountain area is replaced by that over the plain area. The first experiment (Fig. 2a) shows that changing the $\overline{o}$ does not have much impact on the potential EV (i.e., the area under the EV_{max} curve), but mainly shifts the EV_{max} curve along the abscissa axis. The second experiment (Fig. 2b) indicates that as the discrimination ability declines (the discrimination ability over the plain area is lower), the potential EV provided by the LAPS over the mountain area is also reduced, especially for larger r. Results from both experiments suggest that the role of $\overline{o}$ is more like to move the EV_{max} curve horizontally, and redistribute the potential EV to different users. In general, the potential EV does not change significantly. By contrast, the potential EV provided by a forecast system is mainly determined by the discrimination ability of the same forecast system. This implies that the potential EV of a forecast system can only be increased by improving its discrimination ability.

b. Effect of calibration

Most forecast systems possess systematic biases because of the incompleteness of model physics and dynamics. The LAPS PQPFs have an obvious wet bias, and the bias becomes more apparent with the increasing precipitation intensity (Chang et al. 2012). This section explores the effect of calibration on the ROC and EV of a forecast system.

Figure 3a shows that the ROC curves before and after calibration at the 20 mm $(6\,\mathrm{h})^{-1}$ precipitation threshold overlap, which indicates that with or without the calibration, the LAPS PQPFs possess the same discrimination ability.

Figure 3b compares the distribution of EV_{max} from the LAPS PQPFs with and without calibration and also the optimal P_t that users with different r must adopt to achieve their EV_{max}. If an appropriate P_t is adopted, the EV_{max} from the LAPS PQPFs before and after calibration almost did not change. However, if adopting un-calibrated PQPFs, the chosen optimal P_t is higher than the theoretical value due to the wet bias (Section 4a). This will prevent users from achieving their EV_{max} if they adopt the theoretical P_t as their optimal P_t. For example, the theoretical P_t for users with $r = 5/12$ should be 5/12; however, when un-calibrated PQPFs are adopted, prevention is only necessary when $P_f \geq 9/12$ because of the wet bias. Therefore, taking action at $P_f = 5/12$ (over confidence in the forecast) will result in over-prevention, which wastes money and reduces EV. Therefore, when un-calibrated PQPFs are adopted, the optimal P_t must be determined based on the past long-term statistics of economic value.

The analyses of ROC and EV also confirm that reliability and discrimination are two independent statistical characteristics to describe the performance of a forecast system. Although a forecast system may achieve perfect reliability through calibration, its discrimination, which is reflected in the ROC area and EV_{max}, is not affected by calibration. This is expectable since calibration to improve the reliability of a forecast system only corrects the precipitation amount instead of modifying the rainfall pattern, which is controlled by the model physics/dynamics processes and is more associated with the discrimination of the forecasts. Although ROC and EV_{max} are both insensitive to forecast bias, a systematic bias causes the real optimal P_t to deviate from the theoretical one. Therefore, directly using a theoretical optimal P_t for the biased (i.e., not perfectly reliable) ensemble probabilistic forecasts (EPFs) will implicitly increase the $E_{forecast}$, and thus prevent the users to reach their EV_{max}.

c. Effect of forecast uncertainty

Richardson (2000) has demonstrated that the EPFs has better discrimination and can provide higher EV to a wider range of users than the deterministic forecasts (DFs) by using the 6-day forecasts of 850 hPa temperature anomalies. Zhu et al. (2002) also showed that the discrimination of the EPFs derived at lower-resolution grids is even superior to that of DFs performed at high-resolution grids. We investigate whether this is valid for short-range PQPFs by comparing EPFs and DFs. The ROC (Fig. 4a-c) and EV (Fig. 4d-f) curves of the LAPS EPFs and four DFs are compared for various precipitation thresholds. The four DFs are the four LAPS multi-model members without time-lagged configuration, including the LAPS-MM5 (GFS), LAPS-MM5 (NFS), LAPS-WRF (GFS), and LAPS-WRF (NFS) models. The EV curve for the EPFs is the EV_{max} curve. It can be seen that compared with the DFs, the EPFs have better discrimination ability (i.e., the ROC area is greater) and can provide higher EV for a wider range of users (more cost-loss ratios). This is because the EPFs provide forecast uncertainties by using multiple P_t and enable users to choose the optimal P_t for decision-making. In comparison, DFs have only one value of P_t (= 100%) and its uncertainty has been ignored (a DF is either completely correct or completely incorrect).

In addition, the advantages of EPFs in discrimination ability and EV grow with the increasing precipitation threshold, i.e., the differences of ROC area and the potential EV increase with increasing precipitation intensity (Fig. 4a vs. 4c and Fig. 4d vs. 4f). This result illustrates the use and development of EPFs is very important for the forecasts of typhoon heavy rainfall.

7. Summary

The usefulness of weather forecasts should be evaluated not only based on the forecast quality but also on the economic value associated with their actual use in the daily decision-making process of users. Discrimination ability, which can be assessed by the ROC, is one of the characteristics for evaluating the quality of a forecast and is closely related to the EV provided by the same forecast system. In contrast to the previous studies focusing on the ROC and EV of global forecast systems, this study investigates the consistent and related characteristics between ROC and EV based on a mesoscale EPS, LAPS, and focuses on the short-range (0–6 h) probabilistic forecasts for typhoon heavy rainfall. Most important at all, this study suggests how to improve the forecast system in order to increase the economic value for the general users.

The EV formula indicates that the EV is related not only to the HR and FAR from the ROC, but also to the $\overline{o}$ of an event and the r of a user. With the terrain-locking effect on typhoon rainfall, significant differences are shown in the climatology frequency of heavy rainfall and forecast skill over the mountain and plain areas, which allow us to explore the sensitivities of EV with respect to $\overline{o}$ and ROC. Results show that the potential EV is mainly determined by the ROC (i.e., discrimination ability) of the same forecast system. Without changing the potential EV, the role of $\overline{o}$ is more like to move the EV_{max} curve horizontally, and redistribute the potential EV to different users. This implies that the potential EV of a forecast system can only be increased by improving its discrimination ability. Results also show that the LAPS PQPFs have the advantage of providing more reliable forecast information and higher EV to the users with relatively larger r over the mountain area, where severe disasters frequently occur.

In addition, previous studies did not particularly discuss whether, in reality, the biased forecast with imperfect reliability will prevent users from achieving their EV_{max}. Due to the dominant wet bias in the LAPS EPS, results show that the real optimal P_t for users to achieve their EV_{max} deviates from the theoretical value, which is equal to his/her r. Such real optimal P_t largely deviates from the theoretical value as the rainfall intensity increases. Nevertheless, if the user can adopt the real optimal P_t, the EPS either with or without bias correction can offer the same EV_{max}. However, when the uncalibrated EPFs are adopted, the optimal P_t must be determined based on the past long-term statistics of economic value in order to achieve EV_{max}. Our result also shows that the ROC and potential EV are insensitive to the calibration, implying that improving the reliability via calibration cannot increase the discrimination ability and potential EV of a forecast system.

Focusing on the severe weather prediction, this study further emphasizes the advantage of using forecast information from EPS. Comparing the results derived from the EPFs with the DFs, we have confirmed that the discrimination ability and EV of the EPFs is superior to those of the DFs. We also found that such an advantage grows with increasing precipitation intensity, which indicates the use and development of EPFs is very important for the forecasts of typhoon heavy rainfall.

Reference

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, 2014: Analysis of relative operating characteristic and economic value using the LAPS ensemble prediction system in Taiwan area. *Mon. Wea. Rev.* (submitted).

Murphy, A.H., 1977: The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. *Mon. Wea. Rev.*, **105**, 803–816.

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

Yuan, H., J. A. McGinley, P. J. Schultz, C. J. Anderson, and C. Lu, 2008: Short-range precipitation forecasts from time-lagged multimodel ensembles during the HMT-West-2006 campaign. *J. Hydrometeor.*, **9**, 477–491.

Zhu, Y., Z. Toth, R. Wobus, D. S., Richardson, and K. Mylne, 2002：The economic value of ensemble-based weather forecasts. *Bull. Amer. Meteor. Soc.*, 83, 73-93.

TABLE 1. Contingency table for forecasts and observations of a binary event.

		Forecast / action	
		Yes	No
Observation	Yes	Hit (*h*) Mitigated loss ($C+L_u$)	Miss (*m*) Loss (L_p+L_u)
	No	False alarm (*f*) Cost (*C*)	Correct rejection (*c*) No cost (*N*)

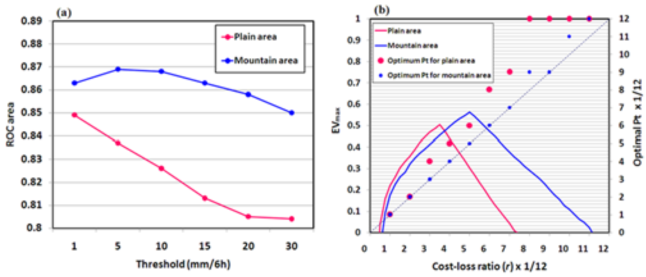

FIG. 1. (a) The area under the ROC curve (ROC area) at different thresholds and (b) maximum economic value (EV_{max}) curves and optimal probability threshold (P_t) against the cost-loss ratio (*r*) at the 20 mm (6 h)$^{-1}$ threshold from the LAPS 0–6 h PQPFs over the plain (red) and mountain area (blue).

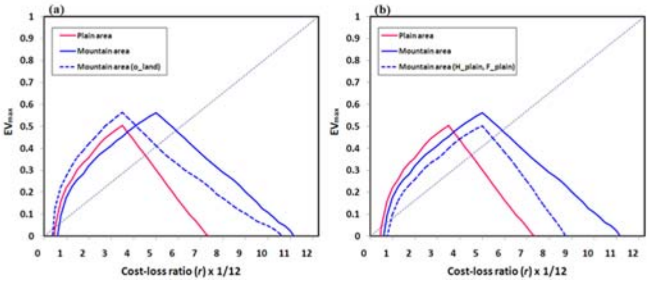

FIG. 2. Solid curves are the maximum economic value (EV_{max}) curves against the cost-loss ratio (*r*) at the 20 mm (6 h)$^{-1}$ threshold from the LAPS 0–6 h PQPFs over the plain (red) and mountain area (blue), and plain areas (red) and the dashed curve is the EV_{max} curve over the mountain area with (a) the climatological frequency ($\bar{o}$) and (b) the values of HR and FAR replaced by those over the plain area.

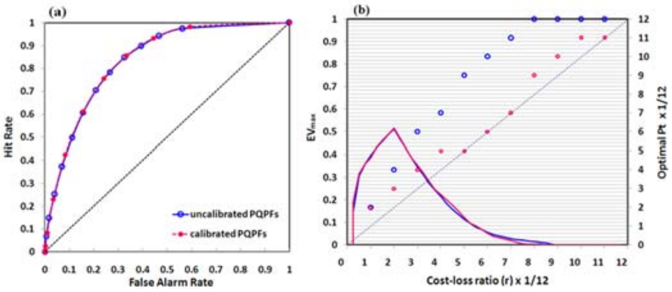

FIG. 3. (a) The ROC curves and (b) maximum economic value (EV_{max}) and optimal probability threshold (P_t) against the cost-loss ratio (*r*) before and after calibration at the 20 mm (6 h)$^{-1}$ precipitation threshold.

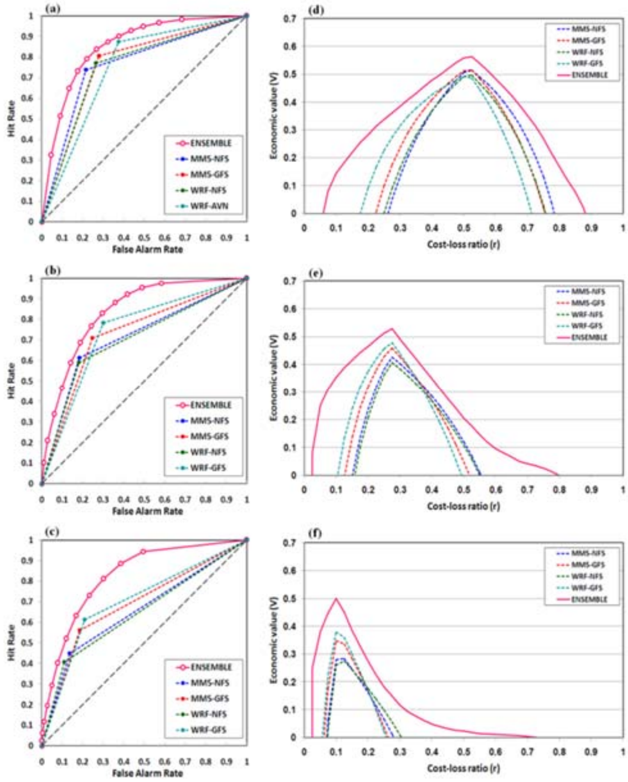

FIG. 4. Distribution of the ROC (left column) and economic value (right column) for LAPS 0–6 h PQPFs and four deterministic forecasts (0–6 h QPFs) at the (a)(d) 1, (b)(e) 10 and (c)(f) 30 mm (6 h)$^{-1}$.