Are the discrimination ability and maximum economic value of a forecast system affected by forecast bias?

Hui-Ling Chang[1] and Shu-Chih Yang[2]

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

e-mail: lingo@msc.cwb.gov.tw (Hui-Ling Chang)

Motivation

Most forecast systems possess systematic biases because of the crude representation of model physics and dynamics to the real atmosphere. Are the discrimination ability (i.e., the ability to discriminate between events and non-events) of a forecast system affected by forecast bias? If users make decisions based on imperfectly calibrated or uncalibrated forecasts, will their economic benefit be reduced?

Focusing on short-range probabilistic quantitative precipitation forecasts (PQPFs) for typhoons, this study explores the effect of calibration on discrimination ability and maximum economic value (EV_{max}) based on Local Analysis and Prediction System (LAPS) ensemble prediction system (EPS) operated at the Central Weather Bureau (CWB) in Taiwan.

Ensemble configuration and data

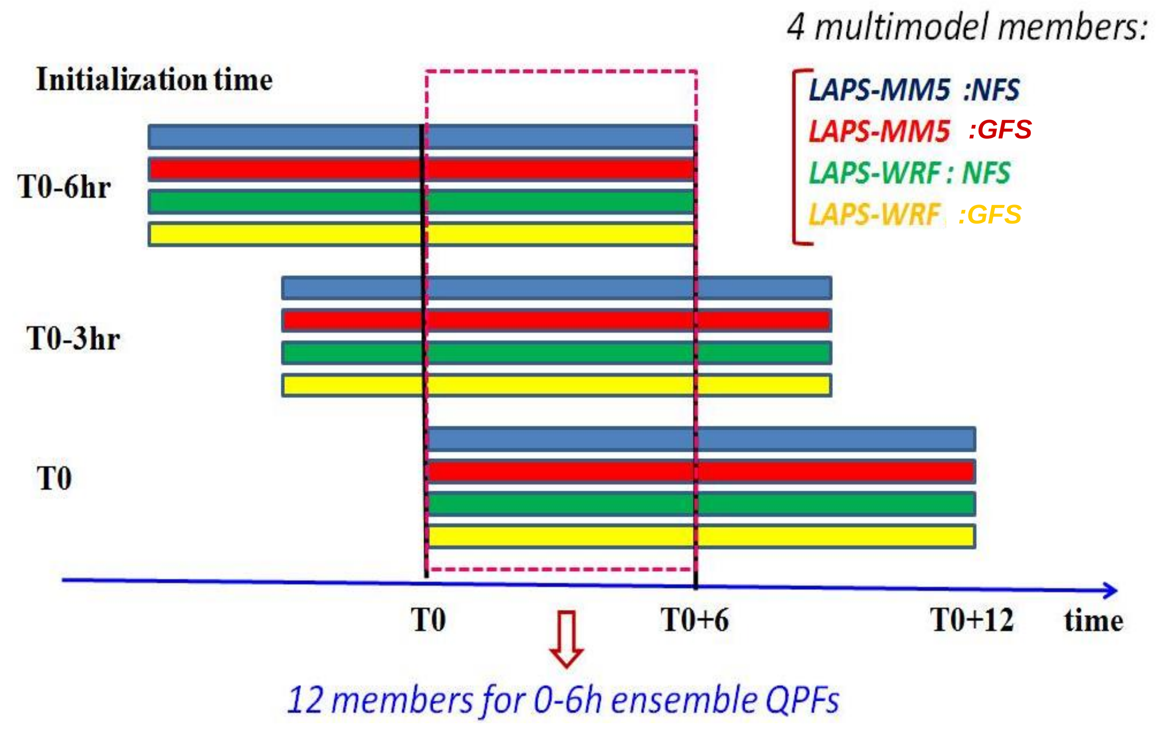


Figure 1. Schematic diagram of LAPS time-lagged multimodel ensemble.

The 15-km NFS and 0.5° GFS provide boundary conditions for four LAPS initialized models (9-km resolution). For the 0-6 h ensemble precipitation forecasts, a total of 12 members are available for the LAPS EPS, in which four models are adopted and each of them has three time-lagged members (the 0–6 h, 3–9 h, and 6–12 h QPFs). The data used to evaluate discrimination ability and economic value (EV) includes a total of 148 cases of 0–6 h PQPFs based on all typhoon cases in 2008 and 2009 in Taiwan area.

Methodology

Linear Regression Calibration

Because the LAPS EPS is wet-biased (Chang et al. 2012), a calibration method based on linear regression (LR) is used to calibrate the PQPFs. The LR equation is expressed as:

$$P(x,t) = a + \sum_{i=1}^{M} b_i f_i(x,t) \qquad (1)$$

Where $M=7,\ f_i(x,t),\ i=1,2,\cdots,7$ are the seven ordinal ensemble precipitation probabilities centered at the calibration threshold, $P(x,t)$ is the corresponding observed precipitation probability, and a is a constant interpreted as error residual. Chang et al. (2012) showed that this calibration method can successfully correct the wet bias and improve the forecast skill.

Relative operating characteristic (ROC)

The ROC is a plot of hit rate (H) against false alarm rate (F) 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 H [$= h/(h+m)$] and F [$=f/(f+c)$] can be derived using several 2X2 contingency tables (Table 1). The area under the ROC curve is called ROC area, which reflects the discrimination ability of a forecast system.

TABLE 1. Contingency table for forecast and observation of a binary event.

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

Economic value (EV)

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

$$EV = \frac{E_{climate} - E_{forecast}}{E_{climate} - E_{perfect}}. \qquad (2)$$

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. The EV can be interpreted as the relative performance taking the climatological information as a baseline. Richardson (2000) showed that the EV can be expressed as

$$EV = \frac{\min(\bar{o},r) - Fr(1-\bar{o}) + H\bar{o}(1-r) - \bar{o}}{\min(\bar{o},r) - \bar{o}r}. \qquad (3)$$

Equation (3) shows that EV is related to the F and H of a forecast system as well as the climatological frequency ($\bar{o}$) of a weather event and the r of a user.

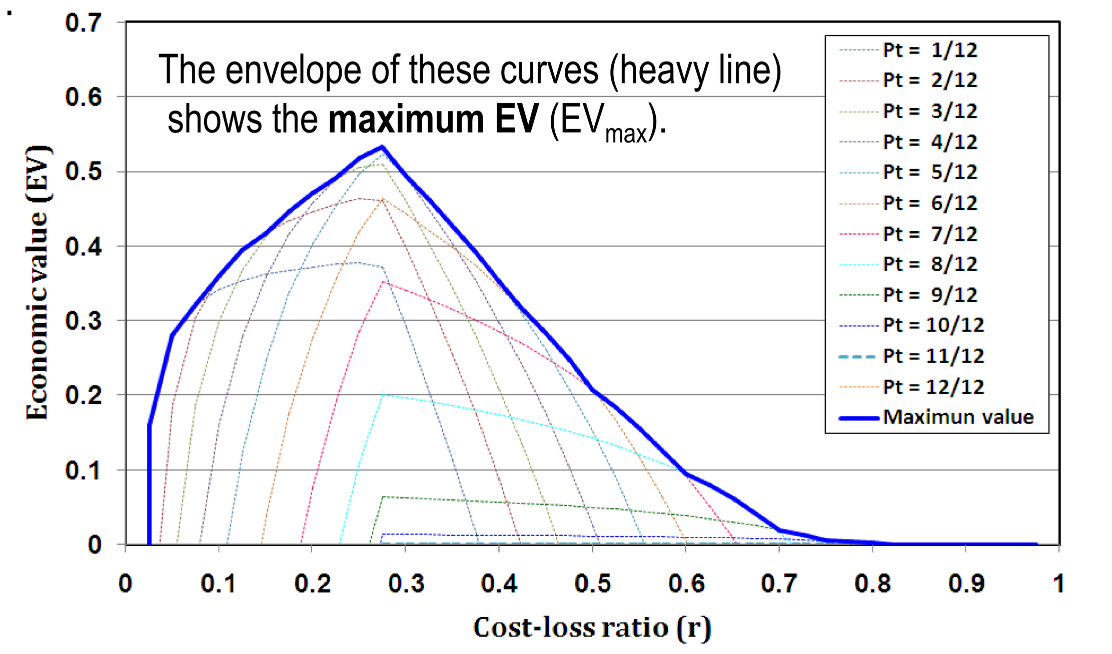


Figure 2. Thin curves show the EV against the cost-loss ratio (r) at different P_t for LAPS 0-6 h PQPFs at the 10 mm (6 h)$^{-1}$ threshold.

Effect of calibration on ROC and maximum EV

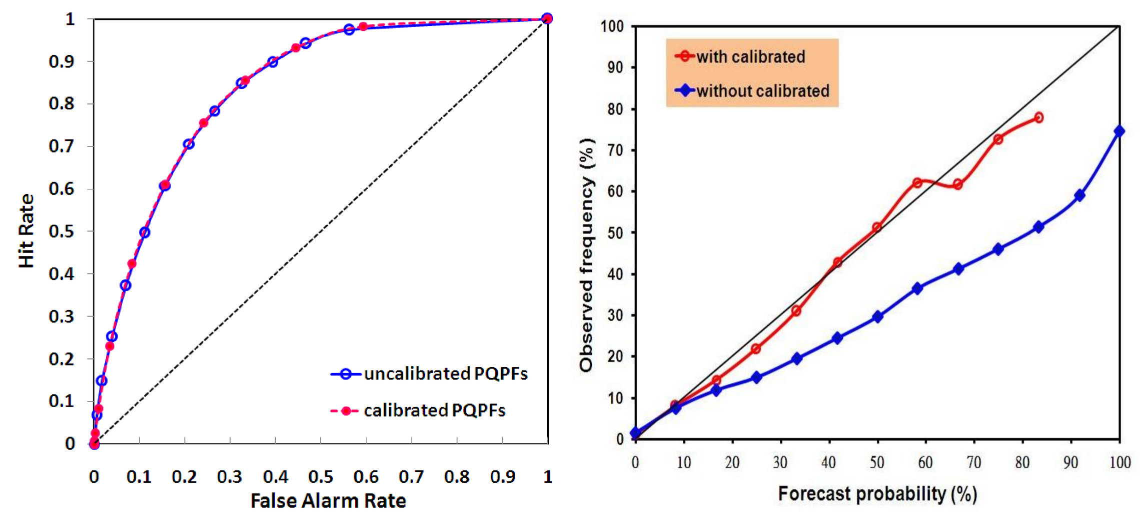


Figure 3. The ROC curves (left) and reliability diagram (right) before (blue) and after (red) calibration at the 20 mm (6 h)$^{-1}$ precipitation threshold.

- The discrimination ability of an EPS is insensitive to forecast bias.
- Improving the reliability via calibration cannot increase the discrimination of a forecast system (Chang et al. 2015).

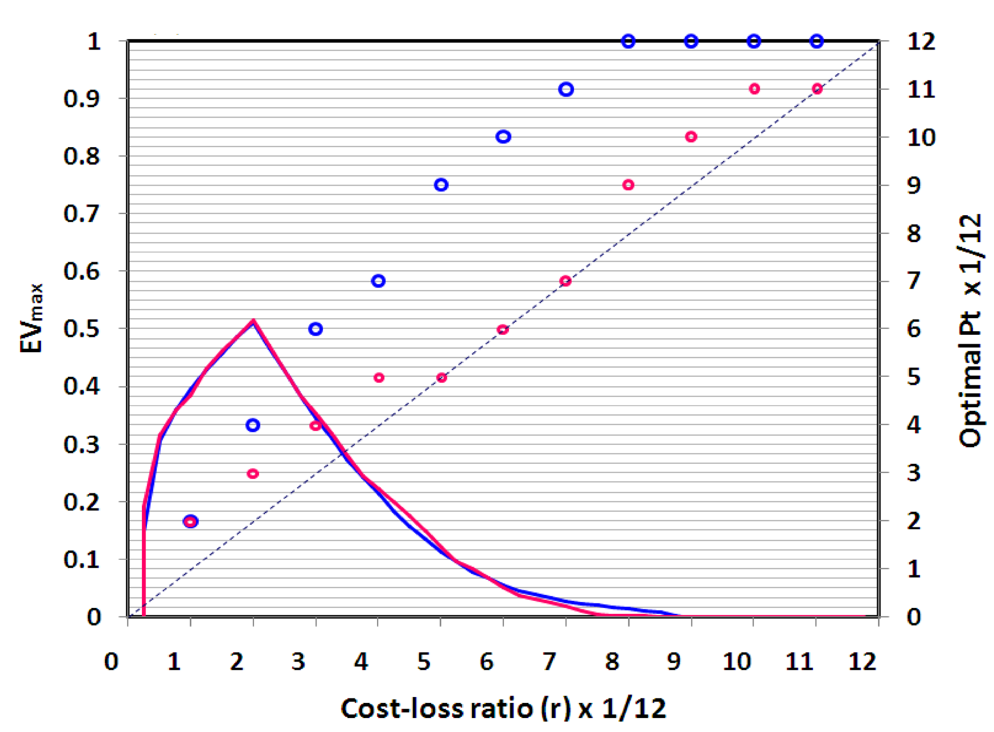
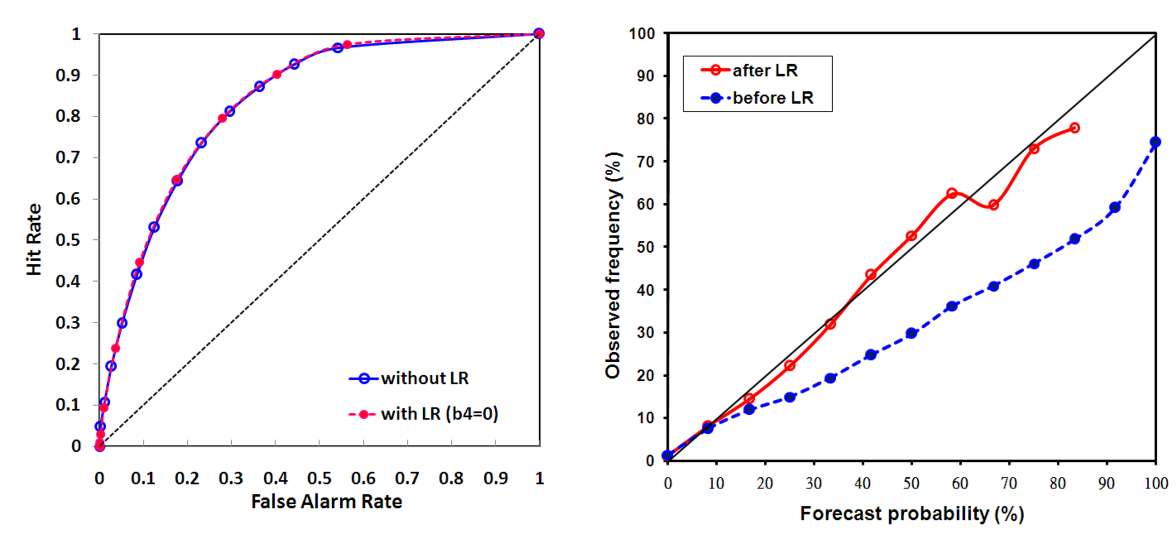


Figure 4. The EV_{max} curves and optimal P_t against the cost-loss ratio (r) before (blue) and after (red) calibration at the 20 mm (6 h)$^{-1}$ precipitation threshold.

- The EV_{max} provided by a forecast system is insensitive to forecast bias.
- The optimal P_t for users to obtain their EV_{max} is different while adopting calibrated or uncalibrated forecasts.
- When adopting an appropriate P_t, users can still obtain their EV_{max} with biased forecasts (Chang et al. 2015).

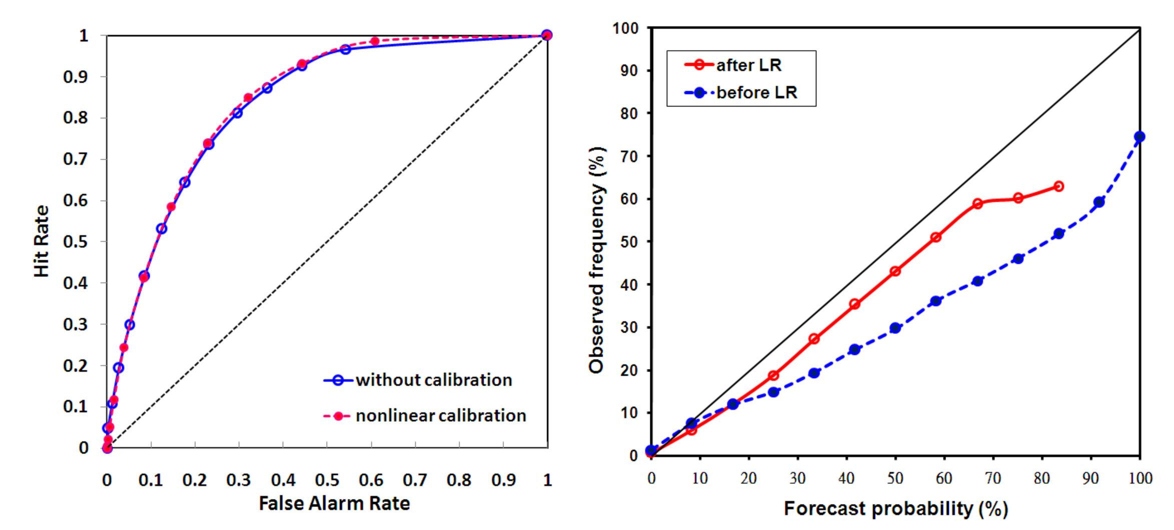
Sensitivity of ROC curve to calibration

- The coefficient of x_4 is reset to zero after the coefficients in Eq. (1) are derived.



- Adopting a nonlinear calibration equation：

$$p(x,t) = a + b_1[f_1(x,t)]^2 + \sum_{i=2}^{6} b_i f_i(x,t) + b_7[f_7(x,t)]^2$$



- The error residual term is reset to 0 (i.e., a = 0) after the coefficients in Eq. (1) are derived

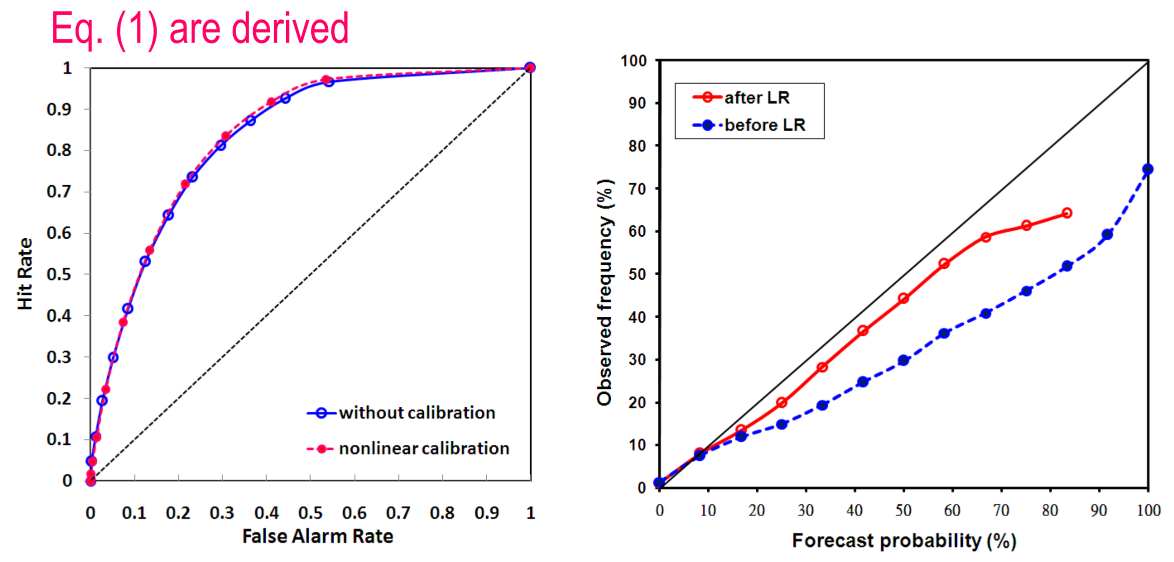


Figure 5. The ROC and reliability curves before (blue) and after (red) calibration at the 20 mm (6 h)$^{-1}$ threshold.

- Discrimination ability almost remains the same after a linear or nonlinear calibration, even though forecast bias cannot be properly corrected during the calibration process.

Conclusions

- Discrimination ability and EV_{max} of a forecast system are insensitive to forecast bias. Calibration, though improving the reliability, has no effect on increasing the discrimination and EV_{max} of a forecast system. In addition, both the linear property of calibration methodology and the accuracy of calibration results do not have the ability to modify discrimination.
- The optimal P_t for users to obtain their EV_{max} is different while adopting calibrated or uncalibrated forecasts.
- When adopting an appropriate P_t, users can still obtain their EV_{max} with biased forecasts.

Acknowledgement and references

We greatly appreciate valuable discussions with Dr. Zoltan Toth at NOAA/ESRL/GSD and Prof. Pao-Shin Chu at the University of Hawaii-*Manoa*. H.-L. Chang is sponsored by the Ministry of Science and Technology, R.O.C. (MOST 104-2625-M-052-004 and 105-2625-M-052-001).

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

Chang, H. L., 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.

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