7.1 ON THE ECONOMIC VALUE OF ENSEMBLE BASED WEATHER FORECASTS

Zoltan Toth[1], Yuejian Zhu[1], and Richard Wobus[1]
Environmental Modeling Center, NCEP, NWS/NOAA
Washington DC 20233

1. INTRODUCTION

In earlier studies we presented a detailed analysis of the quality of probabilistic forecasts generated based on the NCEP ensemble forecasting system (Toth and Kalnay, 1997). The performance of the NCEP ensemble forecasts was also compared to that of the ECMWF ensemble prediction system (Zhu et al., 1996), and a single higher resolution MRF control forecast (Toth et al., 1998).

These earlier studies give valuable insight into the behavior of the different forecast systems, thus providing feedback to the developers. Nevertheless, the ultimate measure of the utility of weather forecasts is arguably the economic benefit associated with their actual use in the daily decision making process of individuals or different organizations. Simplistically, users of weather forecasts either do, or do not take action (e. g., introduce protective action to prevent/reduce weather–related loss), depending on whether a particular weather event is forecast or not. Cost–loss analysis of different complexity can be applied to evaluate the economic impact of the use of weather forecasts on the users (Katz and Murphy, 1997). In this paper we evaluate the economic value associated with the use of an ensemble of forecasts, vs. a higher resolution control forecast, using a relatively simple cost–loss model discussed previously by Richardson (2000a) and Mylne (1999) that, after some simplifications, can generally be applied in most cases.

2. COST/LOSS ANALYSIS

A decision maker becomes a user of weather forecasts if he/she alters his/her actions based on forecast information. If, based upon a particular weather forecast, a user takes a preventive action, and the predicted harmful event does not occur (*false alarm, FA*), the user incurs a *cost (C)* associated with his/her action (Table 1). In a case

		FORECAST	
		YES	NO
OBSERVATION	YES	H(its) ***Mitigated Loss***	M(isses) ***Loss***
	NO	FA (False Alarm) ***Cost***	CR (Correct Rejection) ***No Cost***

Table 1. Contingency table indicating the costs/losses accrued by the use of weather forecasts, depending on forecast and observed values.

where the event is not forecast the user does not take action, and if the event does not occur (*correct rejection, CR*), there is *no cost (N)* on the part of the user. If an event is not forecast but occurs (*missed event, M*), the user is not protected and suffers a loss ($L>C$). When, following the forecast of a harmful event a user takes protective action and the event occurs (*hit, H*), the user has to pay the cost of his action *(C)*, plus may still incur some reduced loss, with a total cost called *mitigated loss* (ML, $C \le ML < L$). For simplicity, in the following calculations we will assume that the mitigated loss is equal to C, the cost of the protective action, i. e., the users, by their protective action, can perfectly shield themselves from the adverse effect of weather.

3. MEAN EXPENSE

If the relative frequency of the four different outcomes in Table 1 (*H, FA, CR*, and *M*) is known and marked by *h, fa, cr*, and *m*, one can assess, in a statistical sense, the mean expense (ME) of a user of a forecast system:

$$ME_{fc}=hML+mL+faC\ (+crN). \qquad (1)$$

Furthermore, one can determine the mean expense associated with using climatological information only:

$$ME_{cl}=min[oL,\ oML+(1-o)C], \qquad (2)$$

where o is the climatological frequency of the event. Based on the climatological frequency of the event and on the user's associated costs and losses, the user will either *always* or *never* take protective action. A decision maker will choose to use a forecast system if his/her mean expense associated with the forecast system will be lower than that associated with using only climatological information.

4. ECONOMIC VALUE

The minimum expense for a user, given a perfect forecast system that provides accurate predictions for the occurence and non–occurence of a particular event, can be written as:

$$ME_{perf}=oML. \qquad (3)$$

In this ideal situation, the user takes protective action if and only if a harmful event actually occurs. Using Eqs. (1–3) the definition of the economic value *(V)* of a forecast system can be given as

$$V=\frac{ME_{cl}-ME_{fc}}{ME_{cl}-ME_{perf}}. \qquad (4)$$

Using a forecast system that is perfect will result in an economic value of 1 (maximum value), while a forecast system associated with the mean expense equal to (larger than) that attainable using climatological information only will have zero (negative) economic value.

The economic value defined above is related to the Relative Operating Characteristics (ROC, see, e.g., Stanski et al., 1989), a signal detection theory based measure of forecast performance. Substituting Eqs. (1–3) into Eq. 4 we arrive at

$$V=\frac{min(o,C/L)-(h+fa)C/L-m}{min(o,C/L)-oC/L}. \qquad (5)$$

It can be shown (see, e. g., Mylne, 1999) that V depends only on the hit rate ($HR=H/(H+M)$) and false alarm rate ($FR=FA/(FA+CR)$), that also defines ROC, as forecast performance parameters. Not surprisingly the overall eco-

1 GSC (Beltsville, MD) at NCEP. Corresponding author address: Z. Toth, NCEP/EMC, 5200 Auth Rd., Room 207, Camp Springs, MD 20746. Zoltan.Toth@noaa.gov

nomic value of a forecast system and the ROC–area and Brier Skill Score measures (BSS, which is a measure related to ROC–area for systems with forecast probabilities exactly matching observed frequencies, Talagrand et al., 1998) are closely related (Richardson, 2000b). For example, BSS measures the overall economic value associated with a particular forecast system, given that we assume all cost–loss ratio values are equally important.

Beyond the parameters describing the forecast system, *V* also depends on *o*, the climatological frequency of the event, and on *C/L*, the cost–loss ratio that depends on the particular user of a forecast system. The fact that all users can be characterized in this framework by a single variable, *C/L*, offers a convenient way to evaluate the economic value of any forecast system for all potential users on a two–dimensional, *V* vs. *C/L* or *L/C* plot.

5. SINGLE VS. ENSEMBLE FORECAST SYSTEMS

In the following section we compare the economic value of the MRF T126 resolution control forecast to that of a 14–member set of the T62 horizontal resolution NCEP ensemble for the April – June 1999 period. These two forecast systems use comparable computational resources. In the example below, weather events are defined as the 500 hPa geopotential height at gridpoints over the Northern Hamisphere extratropics being in any of 10 climatologically equally likely bins.

Deterministic guidance from a single forecast can be unambiguosly interpreted by a user. If a particular adverse weather event is forecast, the user can take protective action, and do nothing otherwise. In case of an ensemble of N forecasts, the user has N options. He/she can choose to take action only if all N forecasts predict the adverse weather, act if N–1, N–2, ..., or even if only 1 member predicts the adverse weather. Each of these decision criteria corresponds with a different economic value. Based on their *C/L* ratio, users can choose the decision criterion that offers the most value to them. In fact, it can be shown that the best decision level *p*, corresponding to the predicted probability of the weather event, is equal to *C/L* (Murphy, 1977). The higher the cost of the protective action relative to the potential loss, the more certainty the user requires about the forecast before he/she takes action. One of the potential advantages of using an ensemble forecast system is that it naturally provides a multitude of such decision criteria. Different users can then tailor their use of the forecast information to their particular application, characterized by their cost–loss ratio.

Relative frequency values based on counting how many ensemble members predict a certain event usually provide probabilistic forecasts that are biased in a sense that they do not necessarily match corresponding observed frequency values. This is because of deficiencies in model and ensemble formulation. For example, when half of the ensemble members predict a weather event, that event may, over a long verification period, verify only 40% of the time. Such biases in ensemble–based probabilities are generally consisitent in time and can be easily eliminated (see, e.g., Zhu et al., 1996). The calibrated forecast that would be issued based on the past verification statistics in the above case, where half of the ensemble members predict an event, for example, is 40%. The April – June 1999 ensemble based probabilistic forecasts evaluated in this paper have been calibrated using independent data from February 1999 verification statistics. For each loss–cost ratio shown in Figs. 1–4 the decision criterion for the ensemble is based on the calibrated probability forecasts. In particular, it is assumed that a user will take protective action if the calibrated probability forecast value is greater or equal to the cost–loss ratio *($p \geq C/L$)*. For the extremely high (and low) probablity values where the finite ensemble cannot provide optimum guidance, the best available guidance was used, i. e., the highest (lowest) probability values associated with all (only one) members predicting the weather event. The above decision making algorithm, based on the users' cost–loss ratio and the calibrated probability forecasts, represents an operationally feasible optimum strategy for the use of ensemble guidance.

6. RESULTS

In Fig. 1 we show the economic value of the control

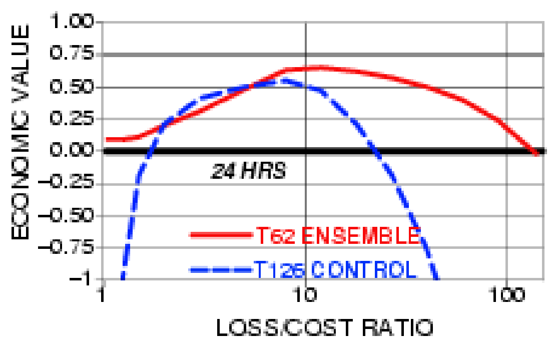

Fig. 1. Economic value of 24–hour MRF T126 control (dashed), and 14–member T62 ensemble forecasts (solid) in predicting events defined in terms of 10 climatologically equally likely bins for 500 hPa height over the NH extratropics, for April–June 1999, for users characterized by different loss/cost ratios (horizontal axis, logarithmic scale). For the ensemble, the optimum decision strategy evaluated here is based on the probabilistic forecasts, calibrated using February 1999 verification statistics, being greater or equal to C/L ($p > C/L$).

forecast vs. an ensemble of forecasts at 24–hour lead time, as a function of the *L/C* ratio, as discussed above. The economic value comparison results indicate that even at this short, 24–hour lead time, for this well predictable variable, most potential users, except those with loss–cost ratios in a relatively narrow band between 2 and 5, can realize more economic value when using the ensemble forecasts. At and beyond 72 hours lead time (Figs. 2–4) virtually all users are better off using the ensemble system than the high resolution control forecasts. Furthermore, the range of loss–cost ratios for which the forecasts exhibit value, compared to using climatological information only, is substantially widened, indicating that a much larger group of users can benefit from the ensemble forecasts as compared to the high resolution control forecasts. Note that on each of the figures the largest economic benefit is, as expected theoretically (see, e.g., Richardson, 2000a), attained by users whose C/L ratio is approximately equal to *o*, the climatological frequency of the weather event,

Contribution to the 15th AMS Conference on Probability and Statistics in the Atmospheric Sciences, 8–11 May 2000, Ashville, NC

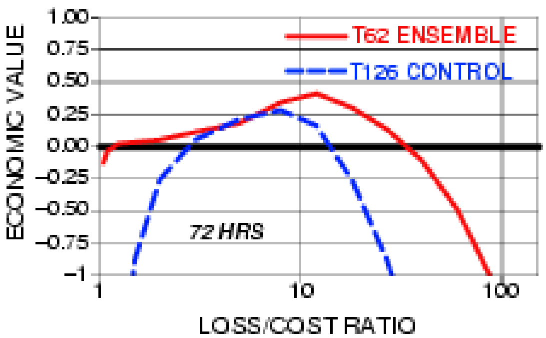

Fig. 2. Same as Fig. 1, except for 72–hour forecast lead time.

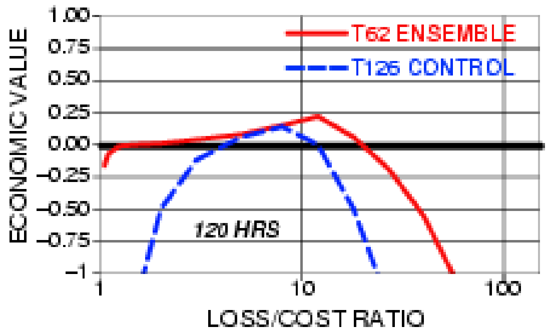

Fig. 3. Same as Fig. 1, except for 120–hour forecast lead time.

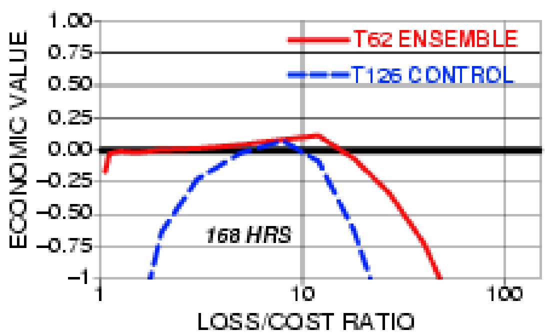

Fig. 4. Same as Fig. 1, except for 168–hour forecast lead time.

which in our case is 0.1. Note also that with increasing lead time the economic value, as compared to using perfect forecasts, just as the forecast information content (see Fig. 8 of Toth et al., 1998), is reduced.

To summarize the results, in Fig. 5 we show the ROC–area scores for the T126 control, and the T62 ensemble forecast systems, for different lead times. Recall that the ROC–area can be considered as a summary measure of overall economic value of the forecasts (Richardson, 2000b). The exact implicite assumption in the ROC–area calcuations about the distribution of users with respect to their cost–loss ratio is not known. Unfortunately, little if any information is available on most users' cost–loss ratio either. Nevertheless Fig. 5 can provide an indication for the

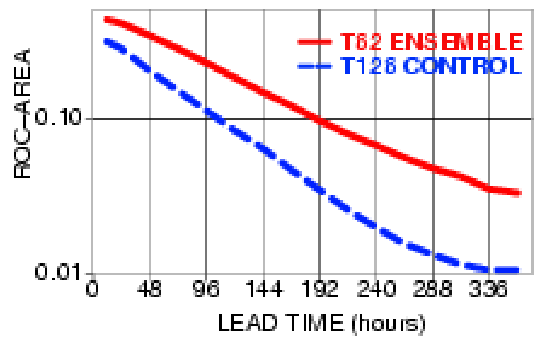

Fig. 5. ROC (Relative Operating Characteristics) area for T126 control and 14–member T62 ensemble forecasts for the 500 hPa height, NH extratropics, for April–June 1999.

overall utility of the two forecast systems. Perfect (climatological) forecasts correspond to a value of 0.5 (0) ROC area, while negative values indicate economic loss compared to using climatological guidance.

The ensemble forecast system is found to outperform the higher resolution control forecast at all lead times. For example, at day 2 (6) lead time the use of the ensemble forecast system provides close to 70% (130%) more overall economic benefit than the control forecast; to put it in another way, a 4–day (10–day) ensemble forecast offers as much value as a 2–day (6–day) single higher resolution control forecast. Similarly, the use of a 6–day ensemble forecast offers 130% more economic value than a 6–day control forecast; and the same level of value can be attained by the use of a 10–day ensemble forecast. These results are in good agreement with Figs. 1–4, and at later lead time with those of Mylne (1999) and Richardson (2000a).

7. CONCLUSIONS

One can draw the following conclusion from the above results. At and beyond 3 days lead time, the direct model output from the ensemble forecasts offers more economic value, and for a wider range of users, than that from the higher resolution control forecast. For a wide range of users this also holds true for shorter lead times. These findings confirm earlier results by ECMWF (Richardson, 2000a) and UK Met. Office (Mylne, 1999) scientists. We should keep in mind that because the horizontal model resolution for the ensemble forecasts is degraded, the cost of their generation is comparable to that of the higher resolution control forecast. Therefore the use of an ensemble forecast system can significantly increase the overall economic benefit weather predictions can deliver to society.

8. DISCUSSION

The 500 hPa height values studied here over the Northern Hemisphere extratropics is considered as one of the most predictable atmospheric variables. It is expected that forecasts for other variables, related more closely to sensible weather, would benefit more from the ensemble approach. The superior performance of the ensemble forecast system is due to two factors. First, the ensemble can distinguish between forecasts with higher and lower than average reliability at the time the forecasts are issued. As Toth et al. (1998) showed by using ROC, Brier score, Ranked Probability Skill Score, and information content as measures of forecast performance, the ensemble provides important extra information to the users through its case dependent reliability estimates.

The ensemble technique's second advantage is that it generates probabilistic forecasts with a multitude of probability values, as compared to dichotomous probability values provided by a single control forecast[2]. This again, as Toth et al. (1998) showed by using different probabilistic verification measures, makes a significant difference. Multiple–value probability forecasts can of course be constructed based on a single deterministic forecast, using past verification statistics. Such a forecast system, however, was still found deficient by Talagrand and Candille (1999, personal communication) when compared to the performance of ensemble forecast systems. The ensemble's superior performance in their comparison, since the control forecast was also expressed in the form of a full probability distribution, can only be due to its ability to do well capturing day–to–day variations in the expected reliability of the forecasts.

There is a third aspect in which the control and ensemble forecasts differ, and that is their overall accuracy, measured, e. g., by the hit rate of the control vs. ensemble mode forecast. At short lead times, the control has an advantage due to its higher resolution (see, e. g., Fig. 3 of Toth et al., 1998), while at longer lead times the ensemble has an advantage, due to its nonlinear error filtering capability (Toth and Kalnay, 1997). While these differences may be significant, comparing Figs. 3 and 7 of Toth et al. (1998) suggests that they play a secondary role compared to the influence of flow dependent full probability distributions, discussed above.

As the above results and discussion indicate, it is critical that the users have access to multiple–value probabilistic information that capture the large day–to–day variations in the expected reliability of the forecasts. Such information facilitates the use, and increases the economic value of weather forecasts. It is not surprising that companies selling weather derivatives[3] are among the core users of ensemble forecasts.

A weather forecast is in fact not complete unless it is expressed in the form of full and joint probability distributions. And in case of appreciable uncertainty, the goal of weather forecasting, including statistical postprocessing, such as MOS and other methods, should not be the provision of a best estimate of the state of the atmosphere but rather of a full probability distribution (Murphy, 1977). The former can only (best) serve users with one particular cost–loss ratio (associated with the hit rate or reliability of the single deterministic forecast system) while the latter can serve users characterized by a wide range of cost–loss ratios.

As an example, let us consider the use of minimum temperature forecasts by two farmers in the same geographical area who grow different crops that are all sensitive to freezing temperature. Let us assume that the cost of protecting their crops is the same but their potential loss differs dramatically due to differences in the vulnerability and value of their crops. The farmer with less to lose (C/L=0.9, high cost–loss ratio) will only spend on protection if the frost is almost a certainty (p=0.9, or higher forecast probabilities), whereas the farmer who can suffer large losses (C/L=0.05) will want to take protective action even if the forecast probability values are low (p=0.05, or higher). Note that in this example the farmers translate the probabilistic weather forecast into their "protect – do not protect", yes–no decision, using very different decision criteria (low vs. high probability values). If a forecaster provides only his/her best estimate on whether the minimum temperature will be above or below freezing, this forecast will likely be useless for either farmer (cf. Fig. 2). Such a forecast, with an intermediate average hit rate of say 60%, will be suitable only for users with a cost–loss ratio around 0.6. Neither the low, nor the high cost–loss ratio customer can benefit from such a product. Instead, they will use climatological information and the former farmer will *always*, while the latter *never* protect his/her crop. To be of any use for them, the forecasts would need to be issued in the form of multiple probability values; and as we saw such guidance can be readily derived from an ensemble of forecasts.

9. ACKNOWLEDGEMENTS

We benefitted from discussions with David Richardson of ECMWF, and Ken Mylne of the UK Met. Office. We acknowledge the support and encouragement of Stephen Lord, Director of EMC.

10. REFERENCES

Katz, R. W., and A. H. Murphy, 1997: Economic value of weather and climate forecasts. Eds., Cambridge University Press, 222pp.

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

Mylne, K. R., 1999 The use of forecast value calculations for optimal decision making using probability forecasts. Preprints of the 17th AMS Conference on Weather Analysis and Forecasting, 13–17 September 1999, Denver, Colorado, 235–239.

Richardson, D. S., 2000a: Skill and economic value of the ECMWF ensemble prediction system, Q.J.R.Meteorol. Soc., 126, 649–668.

Richardson, D. S., 2000b: The application of cost–loss models to forecast verification. Proceedings of the Seventh ECMWF Workshop on Meteorological Operational Systems. November 15–19, 1999, Reading, England, in print.

Stanski, H. R., L. J. Wilson, and W. R. Burrows, 1989: Survey of common verification methods in meteorology. WMO World Weather Watch Technical Report No. 8, WMO/TD. No. 358.

Talagrand, O., R. Vautard, and B. Strauss, 1998: Evaluation of probabilistic prediction systems. Proceedings of ECMWF Workshop on Predictability, 20–22 October 1997, 1–25.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev*, 125, 3297–3319.

Toth, Z., Y. Zhu, T. Marchok, S. Tracton, and E. Kalnay, 1998: Verification of the NCEP global ensemble forecasts. Preprints of the 12th Conference on Numerical Weather Prediction, 11–16 January 1998, Phoenix, Arizona, 286–289.

Zhu, Y, G. Iyengar, Z. Toth, M. S. Tracton, and T. Marchok, 1996: Objective evaluation of the NCEP global ensemble forecasting system. Preprints of the 15th AMS Conference on Weather Analysis and Forecasting, 19–23 August 1996, Norfolk, Virginia, p. J79–J82.

2. The yes–no forecast of a deterministic system, based on past verification statistics, can be converted to dichotomous probabilistic forecasts just as the ensemble–based probabilistic forecasts can be calibrated, see Toth et al., 1998.

3. Weather derivatives are insurance policies offered for premiums that depend on expected forecast reliability. The insured receives a payment in case a weather forecast fails.