Obtaining economic value from the EPS

The performance of the EPS is routinely monitored using a range of verification measures (see the article on EPS verification in ECMWF Newsletter no. 72). This assessment demonstrates that the EPS is a skilful prediction system and has been used to illustrate the improvement of the enhanced EPS introduced in December 1996 (ECMWF Newsletter no. 74). However, these measures do not explicitly address the question which is perhaps of most concern to potential users, namely "Is the EPS worth paying for?"

Providing an answer to such a question is not straightforward. To benefit from a forecast, a potential user must have alternative courses of action available, the consequences of which will depend on the weather that occurs. If, by using forecasts, the user decides on actions which he would not otherwise take, and benefits economically from these alternative actions, then the forecasts have been of value to the user. Thus, a proper evaluation of the benefits of a forecast system to a particular user will involve not only the intrinsic skill of the forecasts, but also detailed knowledge of the exact weather-sensitivity and decision-making process of the user.

Although specific cases may be complex, the general concept of forecast value can be demonstrated using a simple model of the decision process. Results indicate at least qualitatively the value of the forecasts and the framework can be extended if more information is available for a particular user. This approach expresses the performance of the EPS in a way perhaps more directly relevant to end users than the traditional skill measures.

For further discussion on the issue of forecast value together with a summary of recent research see the recent book "Economic Value of Weather and Climate Forecasts", eds. Katz and Murphy (CUP, 1997).

The cost-loss ratio decision model

Consider a decision maker who has just two alternatives, to take action or to do nothing, the choice of which depends exclusively on his belief that a given weather event *E* will occur or not. Taking action incurs a cost *C* irrespective of the outcome. If the event does occur and no action has been taken then the decision maker incurs a loss *L*. For example, the weather event could be the occurrence of ice on roads and the action "to grit the roads"; C would be the cost of the gritting procedure while *L* would be the economic loss due to traffic delays and accidents on icy roads. The expense associated with each combination of action and occurrence of *E* is shown in table 1 (the expense matrix).

The decision maker wishes to pursue a strategy which will minimise any losses over a large number of cases. If only climatological information is available there are just two options: either always take protective action or never protect. Always taking action incurs a cost *C* on each occasion (irrespective of whether the event occurs or not), while if action is never taken the loss *L* occurs only on that proportion $\bar{o}$ of occasions when the event occurs, hence the average expense is $\bar{o}L$. Thus in the absence of information other than climatology, the optimal course of action is always act if $C < \bar{o}L$ and never act otherwise.

It is convenient to consider the expense of the various courses of action in terms of the "cost-loss" ratio *C/L*. If the cost of protection is greater than the potential loss there is no benefit to be obtained from taking any protective action. Thus *C/L* need only be considered to be in the range 0 to 1. The mean expense per unit loss (*ME*) can be plotted as a function of *C/L* on an expense diagram (figure 1). The minimum ME given climate information is shown by the solid curve. The dashed curve shows the minimum expense which could be obtained given perfect knowledge of the future weather - the decision maker would only need to take action when the event was going to occur and would never incur a loss, so the *ME* would be $\bar{o}(C/L)$. Of course, given the chaotic nature of the atmosphere and our inevitably uncertain knowledge of the exact initial state of the atmosphere, such perfect forecasts are not likely to be achieved in practice.

However, there is clearly potential for reduction of the *ME* from that of climatological information towards the perfect-forecast limit.

		Weather event occurs	
		No	Yes
Take action	No	0	L
	Yes	C	C

Table 1: *Expense matrix: cost and loss for different outcomes*

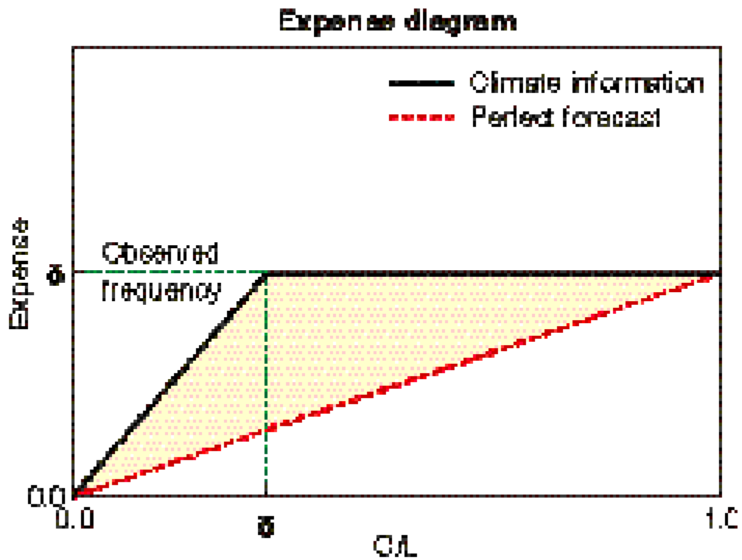

Figure 1: *Mean expense per unit loss for decisions based on climate information or made with perfect knowledge of future weather.Both depend on cost-loss ratio and observed frequency of the event.*

Event	Jan–Feb 1998			
	FAR	HR	KS	$\bar{o}$
T < -8K	0.039	0.445	0.406	0.058
T < -4K	0.144	0.611	0.468	0.228
T > +4K	0.091	0.548	0.457	0.179
T > +8K	0.027	0.393	0.367	0.043

Table 2: *Skill of the EPS control forecasts of temperature over Europe*

The provision of additional information in the form of forecasts may allow the decision maker to revise his strategy and reduce his expected expense. The extent by which the expense is reduced is a measure of the value of the forecasts to the decision maker. We define the value *V* of a forecast system as the reduction in ME as a proportion of that which would be achieved by a perfect forecast. Thus maximum value $V=1$ will be obtained from a perfect forecast system, while $V=0$ for a climate forecast. If $V > 0$ then the user will benefit from the system.

Skill and value for a deterministic forecast system

Consider first a deterministic forecast system, that is each forecast is a simple statement either that a given weather event will occur or that it will not occur. The value of the system depends on the hit rate (HR) and false alarm rate (FAR) of the forecasts, on the observed frequency of the event, and on the user-specific cost-loss ratio (see appendix).

The performance of the EPS control forecast for the prediction of T+144 850 hPa temperature anomalies exceeding certain thresholds is shown in table 2 for January and February 1998 over Europe. The skill of the forecasts is measured using the Kuipers skill score (KS, see appendix) - a perfect forecast will score 1, random or constant forecasts score 0, so a positive score is indicative of skill. The control forecast has substantial skill for all thresholds. Forecasts for the smaller thresholds are more skilful than for the more extreme events, and positive anomalies appear more difficult to predict than negative anomalies. However the question for potential users is how does this skill relate to the economic value of the forecasts?

For a given weather event and forecast system, $\bar{o}$, *HR* and *FAR* are given and the economic value *V* of the forecast system depends only on the cost-loss ratio. *V* is shown in figure 2 as a function of *C/L* for the forecasts of the four events. Although the model is skilful according to the scores in table 2, it is clear that the usefulness to a decision maker depends greatly on his particular cost-loss ratio. For *C/L* greater than about 0.6 none of the event forecasts are useful; for *C/L* between 0.1 and 0.5 forecasts of the ±4K events are useful, while for *C/L* less than 0.1 it is only the forecasts of larger anomalies which have value.

Maximum value always occurs for $C/L = \bar{o}$; at this point the expense of taking either climatological option (always or never protect) is the same: climatology does not help the decision maker and the forecast has the greatest benefit. As the cost approaches the limits of 0 and 1, the climatological options become harder to beat - high expense resulting from even occasional incorrect forecasts outweighs the low expenditure of the default action.

The maximum value itself is equal to the Kuipers score given in table 2. Thus the skill is related to the usefulness of the forecasts: *KS* is the maximum value that can be obtained from the system. Whether this potential maximum value will be achieved depends on the cost-loss ratio of the user; the closer *C/L* is to $\bar{o}$ the higher will be the value. Note that this maximum value is independent of *C/L* and $\bar{o}$; if two systems predicting different events (with quite different $\bar{o}$) have the same *KS* then the potential maximum value will be the same, but it will occur for different values of *C/L*, equal to the respective observed frequencies.

Probability forecasts

If forecasts are supplied to the decision maker as probabilities then the question facing the user is at what probability threshold should action be taken. Should the user take action if the event is forecast with a probability of, say, 50% or should he wait until the forecast is more certain (perhaps 80%)? Is there an optimum probability above which action should be taken?

In effect this choice of a threshold probability *p** converts the probability forecast to a deterministic one - consider those forecasts with higher probability for the event as forecasts that the event will occur and those with lower probability as forecasts the event will not occur. For a given *p**, the value of the system can then be determined in the same way as for a deterministic system. By varying *p** from 0 to 1 a sequence of values for *HR* and *FAR* and hence *V* can be derived; the user can then choose that value of *p** which results in the largest value. Note that since *V* also depends on $\bar{o}$ and *C/L* the appropriate value of *p** will be different for different users and different weather events.

Probability forecasts of the temperature events considered in the previous section are produced using the EPS. The relative operating characteristic *(ROC)* is a plot of *HR* against *FAR* for a set of threshold probabilities *p** between 0 and 1 (figure 3). The endpoints of the *ROC* (1,1 and 0,0) result from the baseline actions of always forecasting or never forecasting the event respectively. A perfect forecast system, with *HR = 1* and *FAR = 0*, would give a point at the top left corner of the graph, so the closer the *ROC* is to the top left corner the better. If a forecast system had no ability to discriminate the occurrence of an event from nonoccurrence, then *HR* and *FAR* will always be equal and the *ROC* for the system would lie along the diagonal line *HR* = *FAR*. The area (*A*) under the *ROC* is used as an index of the quality of the forecast system. A perfect system would have *A = 1.0*, while the no-skill system (*HR* = *FAR*) would have *A = 0.5*. The areas under the *ROCs* of figure 3 are shown in the legend.

Also plotted in figure 3 are the *HR* and *FAR* for the control forecast (table 2). One of the benefits of the ROC is that it allows direct comparison of deterministic and probabilistic forecast systems. In this case the points for the control lie below the EPS *ROC*. These forecasts are less useful than the EPS for this event, since for the same *FAR* a higher hit rate is obtained using the EPS probabilities.

For each probability threshold *p**, the corresponding *HR* and *FAR* can be used to generate a value curve, just as in the deterministic case. The set of curves for the *T > +4K* event are shown in figure 4. The EPS forecasts have value for most users, although the benefit varies substantially for users with different cost-loss ratios. The most important feature of figure 4 is that the value depends crucially on the appropriate choice of threshold probability *p**. Users with small cost-loss ratios, i.e. relatively large potential losses, will gain maximum benefit by taking action even when the forecast probability is low, while for users with high cost, value is obtained by taking action only if there is high forecast probability for the event. An inappropriate choice of p^* can result in substantial reduction in forecast value. For example, a decision maker with a cost-loss ratio of 0.1 will receive 40% value by acting when the EPS probability is 10% or more, but would gain no value at all from the EPS if action was not taken until the forecast probability was greater than 50%.

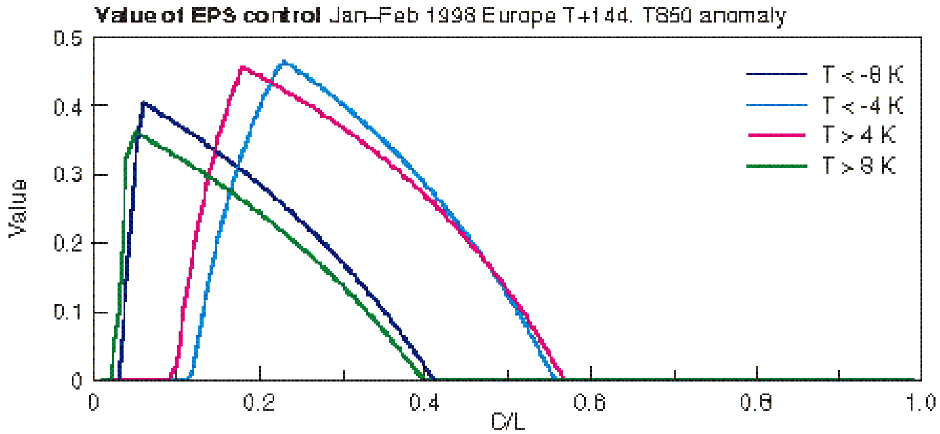

Figure 2: *Value of EPS control forecasts of 850 hPa temperature anomalies exceeding different thresholds.*

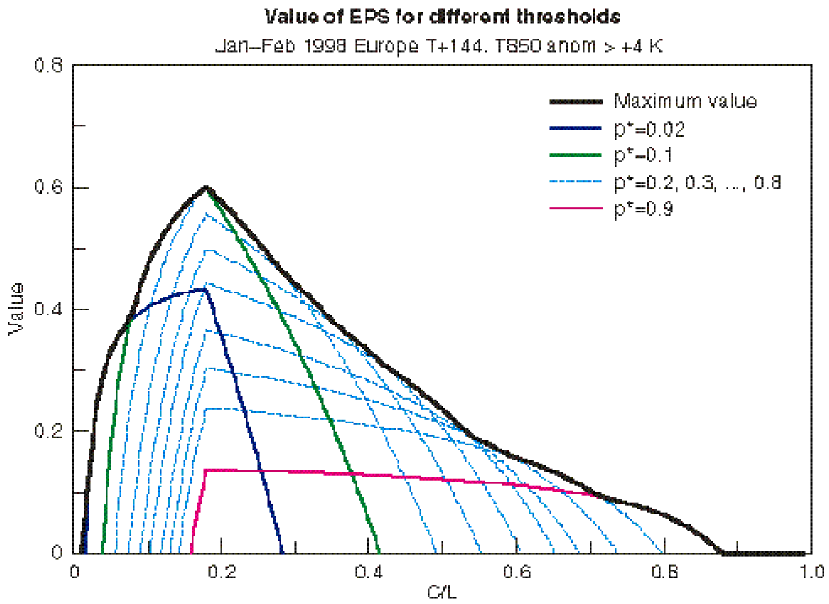

Figure 3: *ROCs for EPS forecasts of 850 hPa temperature anomalies*

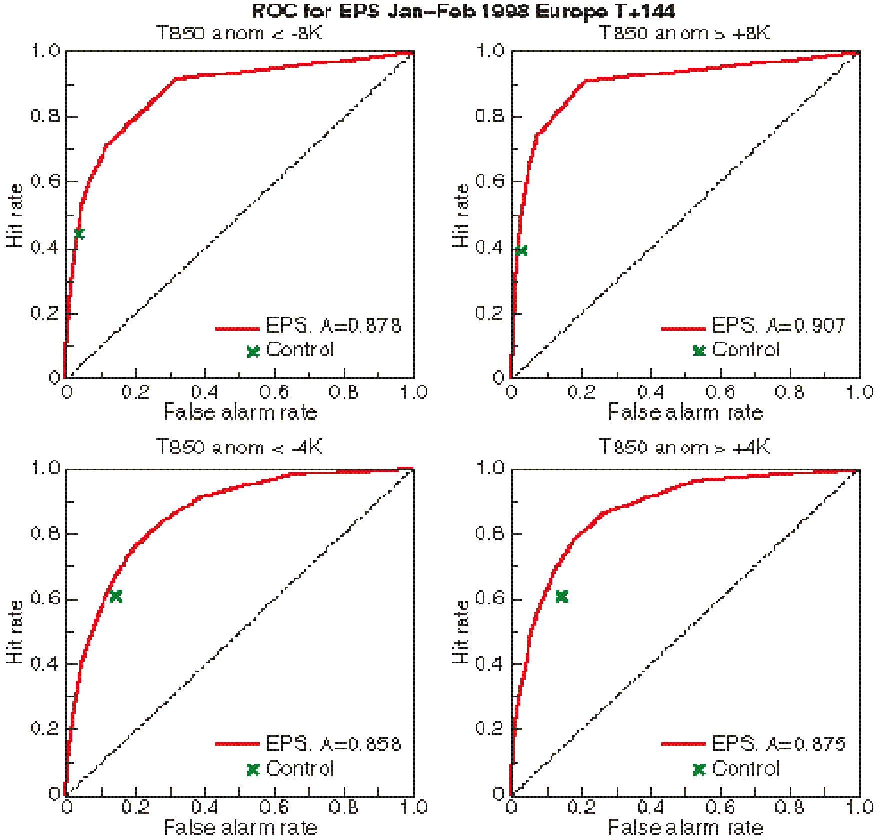

Figure 4: *Value of EPS forecasts for different choices of threshold probability p*.*

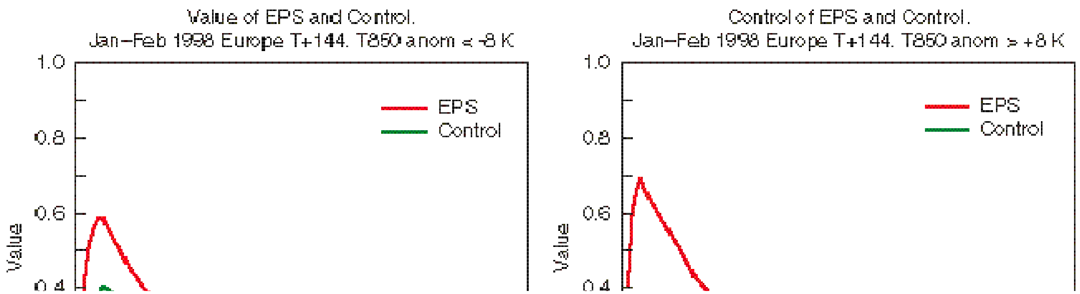

Figure 5: *Value of EPS probability forecasts and EPS control.*

This example illustrates the important advantage of providing probability information to users: the value of the EPS forecasts depends significantly on the choice of probability threshold *p** and on the user's cost-loss ratio. There is no single threshold for which the EPS has value for all users - different users must use different thresholds to benefit from the forecasts. If the forecast is reduced to a single deterministic one for all users, for instance by using the ensemble mean or by choosing an arbitrary threshold, the value to some users will be reduced and may even be eliminated completely.

Comparison of the value curves for the EPS probability forecasts and the control deterministic forecast highlights the advantage of the probability forecasts (figure 5). The flexibility of being able to choose the threshold probability greatly increases the range of users who will benefit from the forecasts. Even though the deterministic forecasts appear close to the EPS curves on the *ROCs*, the extra value of the probability forecasts can be substantial.

Conclusions

There is no simple relationship between the skill of a forecasting system and the value of that system to users. A simple cost-loss model of economic value can be used to give an indication of the potential benefit to a user in a more relevant way. While a system with no skill will not have value, it is not necessarily the case that a skilful system will be beneficial to a given user. The value of the system depends not only on the performance of the system (as measured by hit rate and false alarm rate) but also on the observed frequency of the event and, importantly, on the relevant costs of the user.

Probability forecasts are generally more useful than deterministic forecasts of comparable quality because of the facility for the user to select a probability threshold appropriate to his needs. The arbitrary determination of such a threshold without knowledge of the particular user's requirements can severely reduce the value of the system.

Although it may be difficult to determine the costs and losses for a particular user (users themselves may not readily have this information) the simple value curves presented here do present the forecast verification in a form relevant to the user's needs. The EPS will indeed have economic value to many users, providing at day six perhaps 60% of the savings which would be obtained with perfect knowledge of future weather. That surely is worth paying for.

		Observed		
		No	Yes	
Forecast	No	a	b	a+b
	Yes	c	d	c+d
		a+c	b+d	

Table 3: *Contingency table for forecast and occurrence of binary event*

Appendix

For a deterministic forecast system a contingency table can be constructed showing the proportion of correct and incorrect forecasts of a weather event occurring or not occurring (table 3). The hit rate (*HR*) is defined as the proportion of occurrences of the event which were correctly forecast, while the false alarm rate (*FAR*) is the proportion of nonoccurrences for which the event was (incorrectly) forecast. Note that both *HR* and *FAR* are expressed in terms of the observed relative frequency of the event $\bar{o}$; it is assumed that $\bar{o} > 0$, i.e. that the event does occur in the sample.

$$HR = \frac{d}{b+d} = \frac{d}{\bar{o}} \qquad (1)$$

$$FAR = \frac{c}{a+c} = \frac{c}{(1-\bar{o})} \qquad (2)$$

From table 3 and the expense matrix (table 1), the expected mean expense (ME) for the forecast system is:

$$ME = \frac{bL+(c+d)C}{L} = b+(c+d)\frac{C}{L} \qquad (3)$$

This can be written in terms of HR and FAR using equations (1) and (2) as

$$ME = FAR\frac{C}{L}(1-\bar{o}) - HR\bar{o}\left(1-\frac{C}{L}\right) + \bar{o} \qquad (4)$$

The value of a forecast system is a measure of the improvement over the climatological ME. Here value is defined relative to the maximum possible improvement given by a hypothetical perfect forecast system

$$V = \frac{ME(climate) - ME(forecast)}{ME(climate) - ME(perfect)} \qquad (5)$$

$$= \frac{\min\left(\frac{C}{L},\bar{o}\right) - FAR\frac{C}{L}(1-\bar{o}) + HR\bar{o}\left(1-\frac{C}{L}\right) - \bar{o}}{\min\left(\frac{C}{L},\bar{o}\right) - \bar{o}\frac{C}{L}}$$

So the value of a particular forecast system depends on the external (to the system) parameters *C/L* and *o¯*, and the internal parameters *HR* and *FAR*. It can be shown that the maximum value *V* occurs for *C/L* = *o¯*, at which point *V = HR-FAR.*

Skill of the forecasts is measured using the Kuipers skill score (*KS*). In the notation of

$$KS = \frac{ad - bc}{(a+c)(b+d)} \qquad (6)$$

table 3, this can be written as:
The KS has the desirable characteristics that random or constant forecasts will score zero; perfect forecasts will have a score of 1. The *KS* can be rewritten in terms of the hit rate and false alarm rate as *KS = HR-FAR*. Thus, the Kuipers score is equal to the maximum value that can be obtained from the system.

David Richardson

27.11.2001