Predictability and economic value

David S. Richardson

ECMWF

1. Introduction

To many people, probability forecasts are still much less familiar than traditional deterministic forecasts. Two issues are often raised as practical problems for the use of probabilities. First, there is a common perception that probability forecasts have no place in the real world, where users need to make hard yes/no decisions. Secondly there is the feeling that probability forecasts are difficult to assess – 'probability forecasts are never wrong', the scores are complicated, and different scores tend to show different 'skill'. As an illustration of this last point, Fig. 1 shows two examples of the evaluation of probabilistic skill for the ECMWF Ensemble Prediction System (EPS). The ROC skill score (based on the area under the ROC curve; Richardson, 2000, 2003) shows substantial skill, remaining above 40% throughout the 10-day forecast range. However, the Brier Skill Score (BSS; Wilks, 1995) decreases quickly so that there is no skill at all beyond day 8. Clearly the two skill measures present contrasting perceptions of the performance of the EPS. This raise the obvious question of whether the forecasts are skilful or not and, perhaps more importantly, are the forecasts useful or not? It should be noted that these questions are not restricted to probability forecasts but are equally relevant to the more traditional deterministic forecasts. It is perhaps just unfamiliarity with using probability forecasts and with the scores used to evaluate them that makes the issues more apparent.

To understand how probability forecasts can be used, whether they give benefits to the user and how this relates to the various skill measures, it is necessary to consider the decision-making process of individual users. In general this is a complex process which may not easily be modelled, but to illustrate some of the important concepts it is useful to study a simple decision-making model. This provides a user-perspective on the use and evaluation of probability (and deterministic) forecasts.

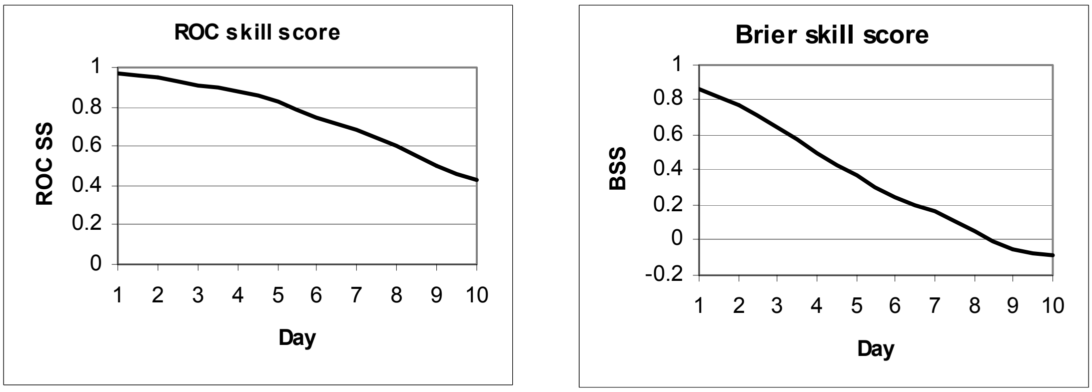

Figure 1 Skill of ECMWF EPS probability forecasts of 500 hPa height anomalies more than 2 standard deviations above normal over the extra-tropical Northern Hemisphere for spring 2002. a) ROC area skill score (ROCSS) and b) Brier (BSS) skill score.

First the simple cost-loss decision model is introduced. This provides a useful introduction and framework for understanding the concept of forecast value and how probability forecasts can be used in making yes/no decisions. The model is used to explore how the benefit of forecast information can vary between users, the value of probability forecasts compared to deterministic forecasts, and also the difference between Brier and

ROC skill scores. Despite the benefits of the cost-loss model, there are a number of obvious limitations. Two apparent deficiencies of the model are that it assumes that the consequences of the user's actions can be expressed numerically (for example as financial costs and losses) and that the user is assumed to be risk-neutral (they are only concerned with long-term average expense). In the final section, the concept of 'utility' is introduced to generalise the decision model and show that the results from the earlier sections can have wider application.

2. The cost-loss decision model

The simple cost-loss decision model has a history dating back to the early twentieth century (Ångström 1922, Liljas and Murphy 1994). For a recent comprehensive introduction see Richardson (2003). Consider a user, or decision-maker, who is sensitive to a specific adverse weather event X. For example, X might be the occurrence of ice on the road, or more than a certain amount of precipitation in a given period. If this event occurs and the user has not taken any preventative action then they suffer a financial loss L. Alternatively, the user could take action at a cost C that would protect against this potential loss. The costs and losses of the various combinations of action and outcome are shown in Table 1a. The aim of the user is to minimise their overall expense by deciding on each occasion whether to protect or not.

a.

		Event occurs	
		Yes	No
Action taken	Yes	C	C
	No	L	0

b.

		Event occurs	
		Yes	No
Event forecast	Yes	a	b
	No	c	d
		$\overline{o}$	1-$\overline{o}$

Table 1. a. Costs and losses associated with different actions and outcomes in the cost-loss model. b. Contingency table for deterministic forecast of specified event over a set of cases, showing fraction of occasions for each combination of forecast and outcome.

Over a large number of cases, let $\overline{o}$ be the fraction of occasions when X occurs. If the user always protects then the cost will be C on every occasion and the average expense (per case) will be

$$E_A = C$$

Alternatively, if the user never take action, the loss L will only be incurred when X happens, so the average expense will be

$$E_N = \overline{o}L\,.$$

Assuming the user knows $\overline{o}$ but has no additional forecast information, then the optimal strategy is either always or never protect, depending on which gives the lower overall expense. This gives a baseline against which improvements from using forecast information can be judged, which we will call the climatological expense, E_C, where

$$E_C = \min(C, \overline{o}L)$$

Another useful reference point is provided by the expense associated with perfect forecast information: the user would only protect if X was going to occur and the average expense would be

$$E_P = \overline{o}C$$

A deterministic forecasting system gives a simple yes/no prediction for X to occur. The performance of the forecast system over a large set of cases can be summarized in a contingency table as shown in Table 1b.

Note that a, b, c and d are the fraction of occasions on which the various combinations of forecast and observation occurred, so they sum to 1. The average expense of using the deterministic forecast is obtained by multiplying the corresponding cells of Tables 1a and 1b

$$E_F = aC + bC + cL$$

The difference in expense between E_F and E_C is a measure of the saving the user can make by using the forecast, compared to having only climatological information. We can define the relative value of the forecasts by comparing this saving with the maximum possible saving that could be made from perfect deterministic forecasts

$$V = \frac{E_C - E_F}{E_C - E_P}$$

From now on we will refer to V simply as the 'value' of the forecasts.

In Table 1b, the entries a,b,c, and d together describe the quality of the forecasts. Various measures of quality can be defined using these terms. Two such measures are the hit rate $H = a/(a+c)$ and the false alarm rate $F = b/(b+d)$. It is often convenient to write the expression for V in terms of H and F. Using the various expressions for the different expenses, this gives

$$V = \frac{\min(\alpha,\bar{o}) - F(1-\bar{o})\alpha + H\bar{o}(1-\alpha) - \bar{o}}{\min(\alpha,\bar{o}) - \bar{o}\alpha}$$

where $\alpha \equiv C/L$ is known as the cost-loss ratio of the user. A number of points should be noted from this expression for V. Value depends on two independent measures of forecast quality (H and F here, but you could choose other measures). This is key to differences noted previously between different skill scores – a single score is not enough to give information relevant to a range of different users. Value also depends on the event through the climatological frequency $\bar{o}$, and on the individual user through the cost-loss ratio.

For a given weather event and forecast system, $\bar{o}$, H and F are fixed and V is then only a function of C/L. Figure 2 shows the variation of V with C/L for precipitation forecasts from the deterministic control forecast of the EPS. It is clear from Figure 2 that forecast value varies considerably between users with different cost-loss ratios. While users with C/L~0.3 may receive up to 40% of the value they would obtain from perfect deterministic forecasts, others (C/L above 0.6 or below 0.2 will get no benefit from these forecasts.

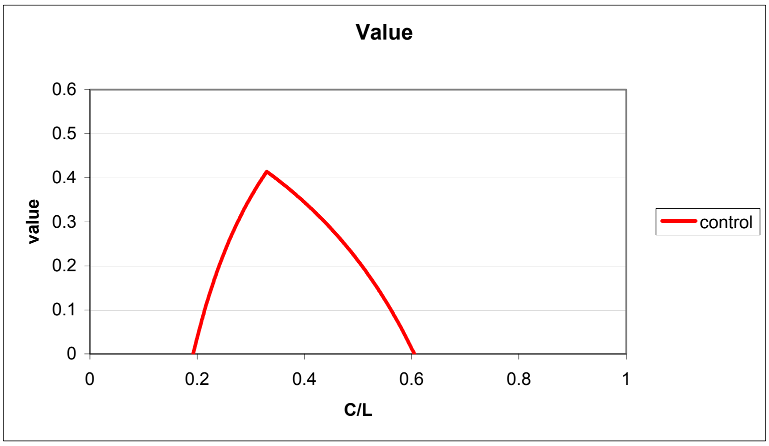

Figure 2 Value V of ECMWF deterministic control forecasts of more than 1 mm precipitation in 24 hours at day 5 over Northern hemisphere extra-tropics for Winter 2001-02.

The shape of the value curve is typical. It is straightforward to show that maximum value is obtained for users with $C/L = \overline{o}$ and that the maximum value is given by

$$V_{\max} = H - F \equiv P$$

where P is the Peirce skill score, also known as the Kuipers score or true skill statistic (see Richardson 2000, 2003 for details). This provides a first link between skill and value. P can be interpreted as a measure of potential forecast value as well as being a skill score. But P only indicates the maximum possible value and does not show how the forecasts will benefit users with C/L different from $\overline{o}$. It can also be shown that the range of C/L for which the forecasts give positive value is related to a different skill measure, the Clayton skill score (Wandishin and Brooks, 2002). This is just one example illustrating that different skill measures can be related to different aspects of forecast value. It may therefore be important when evaluating forecast performance to bear in mind the purpose of the evaluation and to use a range of performance measures.

3. Probability forecasts

The use of deterministic forecasts in the cost-loss decision framework is straightforward: take action whenever the event is forecast. In contrast, probability forecasts may at first sight appear inappropriate when a hard yes/no decision needs to be made. The cost-loss model gives a simple illustration of the way that probability forecasts can be used in such situations and the benefits that may be obtained.

Given forecast information as a probability, the user is faced with the decision of whether the probability of adverse weather is high enough for protective action to be needed. The user needs to set a probability threshold p_t and take action if the forecast probability exceeds that threshold. This choice of p_t effectively converts the probability forecast to a deterministic one: forecasts with probability higher than p_t become 'yes' forecasts while the remainder become 'no' forecasts. The resulting deterministic forecast can be evaluated as in the previous section to obtain $H(p_t)$, $F(p_t)$ and $V(p_t)$. Different choices of p_t will result in different values for H, F and V.

Figure 3 shows V(C/L) for three choices of p_t for EPS probability forecast corresponding to the control forecast of figure 2. Different users will benefit more from different choices of p_t. In general users with lower C/L will benefit more from acting at lower probability thresholds while those with high C/L will gain more from acting only when there is greater certainty about the event. For example a user with C/L=0.2 will

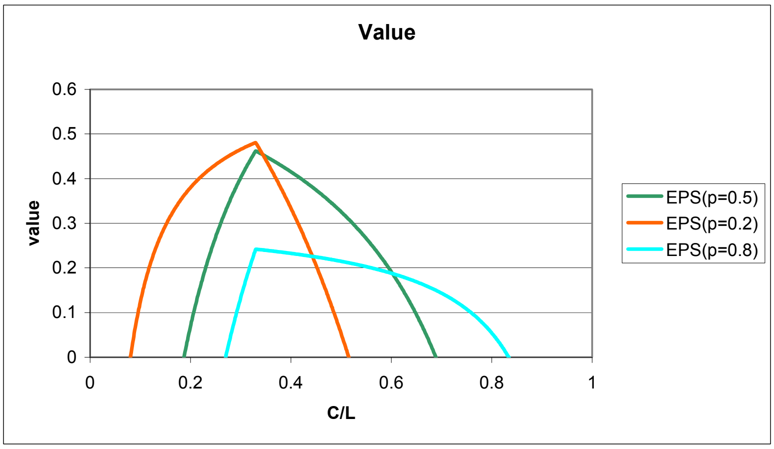

Figure 3 Value of the ECMWF probability forecast for more than 1 mm precipitation in 24 hours at day 5 over the northern hemisphere extra-tropics for Winter 2001-02. The curves show the variation of V with C/L for probability thresholds p_t = 0.2 (orange), 0.5 (green) and 0.8 (cyan).

have a relatively high potential loss and benefit substantially be taking action at $p_t = 0.2$. Users with much higher relative costs would lose out by taking action when the chance of the event is so low, but they would benefit from using a higher threshold probability.

It is, then, important that probability forecasts are used with care. An inappropriate or arbitrary choice of p_t may substantially reduce the benefit of the forecasts to a user. Moreover, it is not possible to choose a suitable p_t without knowing the cost-loss ratio of the user. It is straightforward to show that for reliable probability forecasts (Wilks 1995) the optimal choice is $p_t = C/L$ (Richardson 2000, 2003). In other words, each user should take action when the probability of the event exceeds their own cost-loss ratio. Consider just those occasions when the probability of the event is a specific value q. Then taking action incurs the constant cost C. The alternative is not to act and accept the consequences: if the event occurs the loss is L, while if it does not occur there is no expense. The average expense of not acting is then Lq. So to minimise the average expense the user should act if $C/L < q$, but accept the risk otherwise. In this way, $p_t = C/L$ can be seen to be the appropriate choice for threshold probability. If the probability is higher than p_t the user should act; if not it is better to risk the loss.

Probability forecast that are not completely reliable can generally be corrected for these biases by calibrating, or re-labelling, the forecast probabilities (Zhu et al. 1996, 2002). In this way the users can treat the forecast probabilities at face value and gain maximum benefit. In general, users should be presented with reliable, calibrated forecasts. Each user should take action when the probability of the event exceeds their own cost-loss ratio.

The optimal value that would be obtained when each user chooses the most appropriate p_t is shown in Fig 4, together with the corresponding curve for the deterministic control forecast. All users will gain more from using the probabilities than from the control forecast, and by allowing different users to take action at different probability thresholds the ensemble probabilities benefit a much wider range of users. This benefit should also be compared to the limited benefits provided by any single choice of p_t (Fig 3).

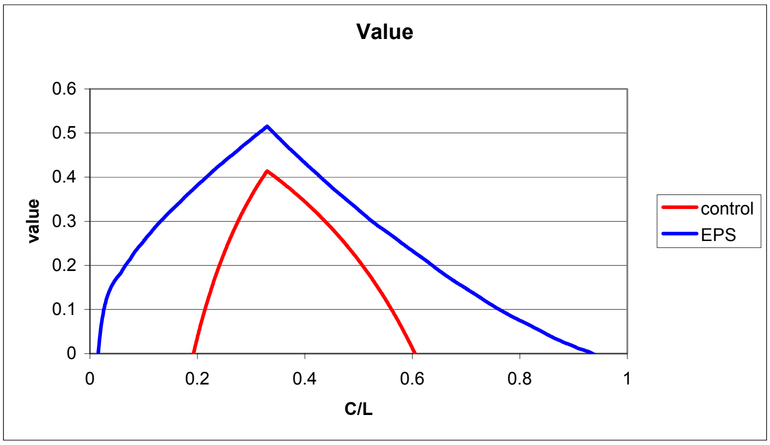

Figure 4 Value of the ECMWF deterministic control forecast (red) and EPS probability forecast (blue) for more than 1 mm precipitation in 24 hours at day 5 over the northern hemisphere extra-tropics for Winter 2001-02. The probability forecast curve shows the optimal value, obtained when each user chooses the most appropriate p_t.

3.1. Skill and value

We have already noted the relationship between value and certain skill measures for deterministic forecasts. In this section we explore the link between skill and value for probability forecasts. This will for instance help us understand the striking difference, shown in Fig 1, between the perception of EPS skill that would obtained when using the Brier Skill Score (BSS) or the ROC area skill score (ROCSS).

The skill scores are single overall summary measures of forecast performance. Yet we have seen that forecast value varies greatly between users depending on their particular costs and losses. There is no simple relationship between either of the skill scores and the value to individual users. However we can imagine that it could be useful to have some summary measure of value that reflects the overall benefit to the range of users of the forecast data. For example, if we knew the appropriate C and L for each user we could work out the total saving that the group of users would make from the forecasts. In general little is known about the circumstances of individual users and of course the cost-loss model will not be appropriate in many cases. It turns out that the Brier skill score is equal to the overall value that would be obtained for a set of users distributed uniformly through the range of possible C/L (see Richardson 2001, 2003 for details).

Figure 5 shows EPS value for a heavy precipitation event. This is a rather rare event ($\bar{o} = 0.02$) and the value is concentrated around low C/L users. The Brier skill score, being a measure of the overall value for all possible users, is low: $BSS = 0.06$. However, $ROCSS = 0.65$, similar to the maximum value of around 0.6. In fact, ROCSS is closely linked to maximum value. For deterministic forecasts, the relationship is exact ($ROCSS = V_{\max}$). For probability forecasts ROCSS is always greater than $V_{\max}$ (see Richardson 1999, 2000 for more discussion on this relationship).

a)

b)

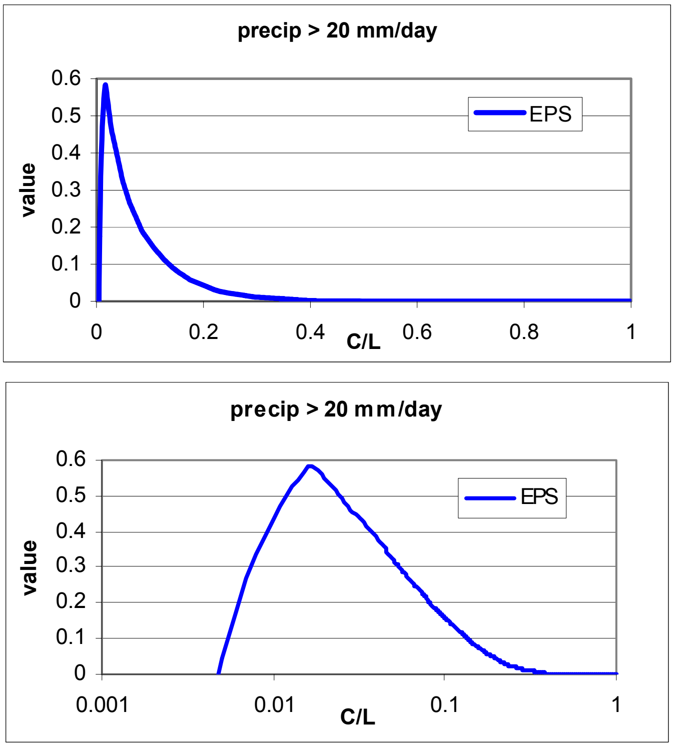

Figure 5 Value of the ECMWF EPS probability forecast (blue) for more than 20 mm precipitation in 24 hours at day 5 over the northern hemisphere extra-tropics for Winter 2001-02. The probability forecast curve shows the optimal value, obtained when each user chooses the most appropriate p_t .

a) plotted on regular axis for C/L; b) logarithmic axis for C/L

Which skill measure is more appropriate depends on the distribution of users. Although little is known about real-world costs and losses, general economic considerations tend to suggest that lower values of C/L are more likely than higher values (Roebber and Bosart, 1996). The few studies that have applied the simple cost-loss model to financial decisions seem to support this. Examples include C/L of 0.03 for raisin drying (Kolb and Rapp, 1962), 0.02-0.05 for orchardists (Murphy 1977), 0.01-0.12 for fuel-loading of aircraft (Leigh 1995), 0.125 for winter road-gritting (Thornes and Stephenson 2001). With this in mind, the BSS and ROCSS can be seen as indicating lower and upper bounds respectively for the value of the forecasts. The important factor to remember is that value is very much user-dependent. Any summary measure condenses information and will not reflect the benefits to be gained by individual users. Unless details of the decision making and associated costs of specific users or groups of users are known, caution is needed in interpreting summary skill measures. In general, a range of measures, including for example both BSS and ROCSS, should be used.

4. Effect of resolution and ensemble size

Just as different users will benefit differently from the EPS, so the benefits of changes to the EPS configuration will also be different. Figure 6 shows an example of the improvement for different users for the changes in resolution for the EPS (Buizza et al., 2003). This is shown as a percentage improvement in value RI(V). This is defined so that RI=0 indicates no change and RI=100% indicates an improvement equivalent to a 1-day gain in skill compared to the previous system (Buizza et al. 2003). In this example while all users benefit substantially, the gains are substantially higher for users with low C/L.

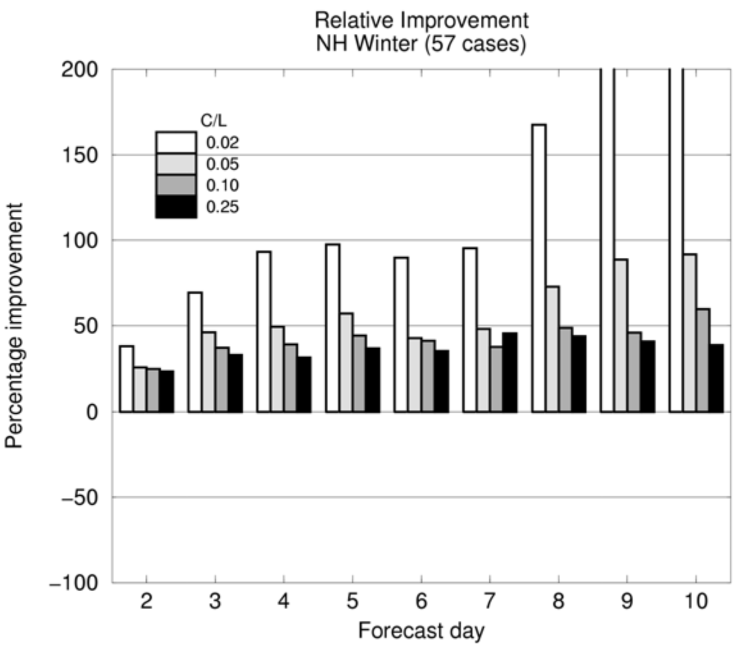

Figure 6 Improvement in value for high-resolution EPS compared to previous system for moderate 850 temperature anomalies over 57 winter cases. The benefit of the new system is shown using a Relative Improvement index (RI, where RI=100% is equivalent to a 1-day gain in skill) for selected cost/loss ratios: C/L=0.02 (white), C/L=0.05 (light grey), C/L=0.10 (dark grey) and C/L=0.25 (black).

While Fig. 6 considered the impact of resolution changes, Fig. 7 shows the effect of ensemble size on value. This is taken from an idealised study (Richardson 2001) and shows the effect of ensemble size on the value of an otherwise perfectly specified (completely reliable) ensemble system. The left-hand panel shows the value of 10-member (solid line) and 50-member (dashed line) ensembles for a common event (obar=0.5); the dash-dotted curve shows the maximum potential value for the large ensemble limit. Increasing ensemble size from 10 to 50 members benefits all users. There is less to be gained from further increases in ensemble size. The right-hand figure shows corresponding curves for a less common event (climate frequency 0.05; note the logarithmic x-axis). Here the benefit of increasing ensemble size depends considerably on the cost-loss ratio of the users. While increasing from 10 to 50 members gives general benefit, the main gains for further increases in size are for low C/L. The Brier skill scores for the ensembles in both cases is the same (1/7). Less predictable events (lower Brier skill score) will show more sensitivity to ensemble size; for more predictable events, the importance of ensemble size is less. More examples are shown in Richardson (2001).

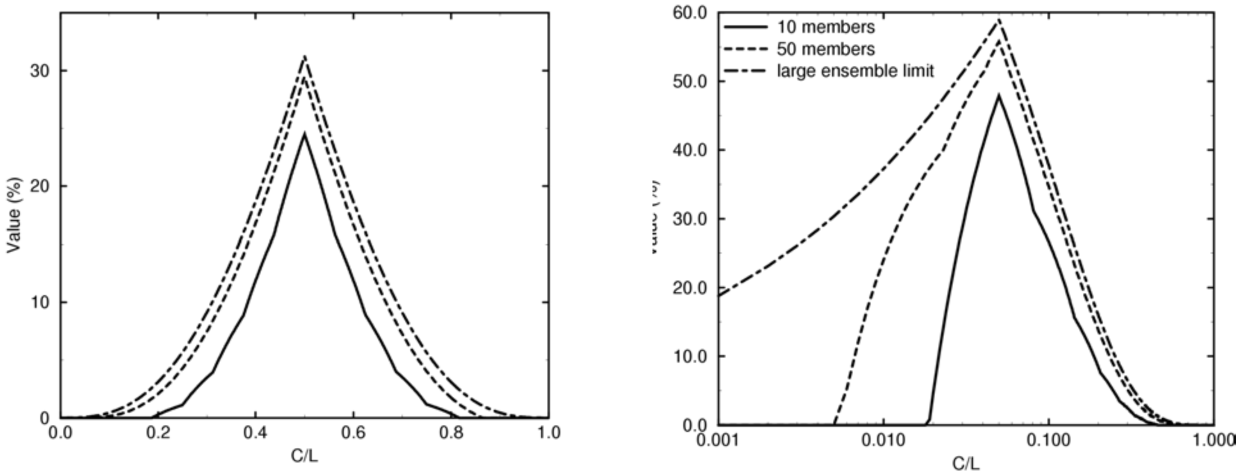

Figure 7 Variation of value with ensemble size. Curves show V as a function of C/L for 10-member (solid) and 50-member (dashed) ensembles compared with the potential value for the underlying distribution (large-ensemble limit) for an idealised perfectly reliable probability forecast system. The left-hand panel is for a common event ($\overline{o}$ = 0.5) and the right-hand panel for a rarer event ($\overline{o}$ = 0.05); the underlying predictability, given by BSS for the underlying probability distribution is the same in both cases.

Figure 7 shows that that the effect of ensemble size on value depends very much on the individual user. The impact on skill, as an overall measure of performance, is shown in Fig 8. The two panels show contrasting examples of the variation of skill with ensemble size (also from Richardson 2001). Each curve shows how skill increases as the ensemble size is increased. The different curves represent different levels of predictability of the underlying distribution; lower curves correspond to lower predictability. The left-hand panel (Fig 8a) is for the Brier Skill Score. If predictability is relatively low then the Brier skill score for a small ensemble may well be negative. However, there is little to be gained from increasing ensemble size beyond 30-50. For more predictable events, even small ensembles will score highly.

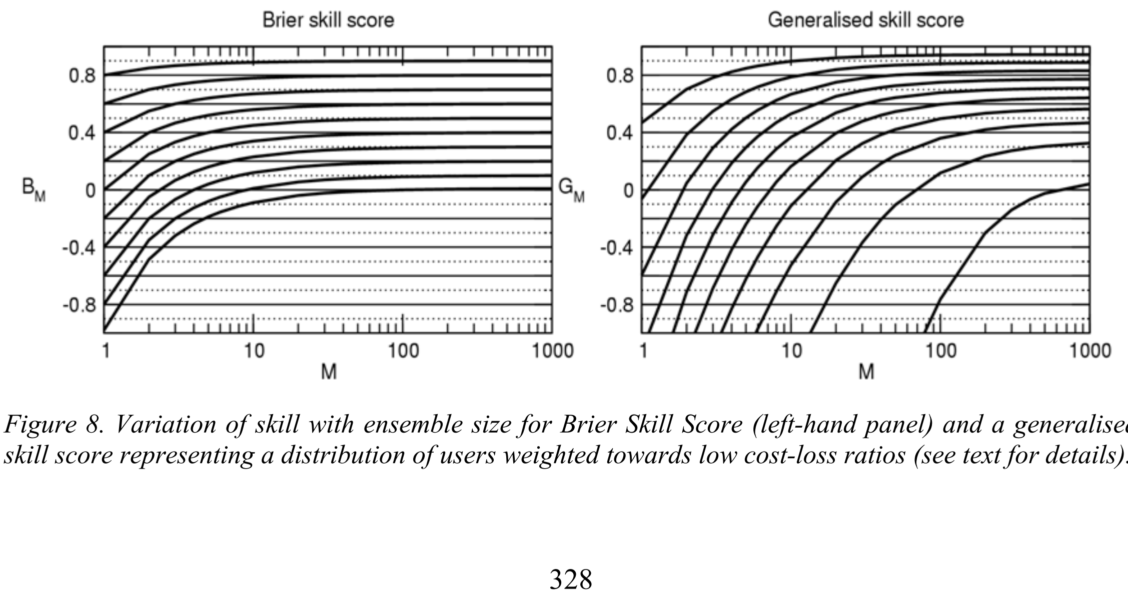

Figure 8. Variation of skill with ensemble size for Brier Skill Score (left-hand panel) and a generalised skill score representing a distribution of users weighted towards low cost-loss ratios (see text for details).

As noted in the previous section, the Brier skill score is equal to the overall value for users with cost-loss ratios C/L distributed uniformly throughout (0,1) (Murphy, 1966; Richardson 2001). Looking back at Fig 7 confirms that for much of this interval the benefit of more than around 50 members will be slight. However, Fig 7 also showed that for low C/L, larger ensembles may give significant gains. While Fig 8a shows overall value for uniformly distributed users, Fig 8b shows the variation of an equivalent measure of overall value, but this time for a non-uniform distribution of users, concentrated towards low C/L. In general, overall value also depends (unlike the Brier skill score) on the climatological frequency of the event (Richardson 2001). Fig 8b shows the variation of this general skill score with ensemble size for a rare event (climatological frequency $\overline{o} = 0.01$). Here, the sensitivity to ensemble size is much greater, with potential benefits for ensembles of several hundred members. Which figure is most appropriate depends of course on the users of the product, but the properties of the scores being used must always be borne in mind.

5. Utility

We have used the cost-loss model as a simple example to illustrate the decision-making process of a user. In this model we assumed that the consequences of the actions can be expresses directly as financial costs and losses, that the cost C gives complete protection against the loss L, and that the user is only concerned with minimising the long-term average expense. In this section we show that the results presented so far are also valid in the more general situation where each of these assumptions is relaxed.

		Event occurs	
		Yes	No
Action taken	Yes	C_{11}	C_{12}
	No	C_{21}	C_{22}

Table 2. Costs and losses associated with different actions and outcomes in the general cost-loss model.

First, consider the general case where each combination of action and outcome results in a different expense (Table 2). For a fixed forecast probability q, the mean expense of taking action is

$$E_A = C_{11}q + C_{12}(1-q)$$

and the mean expense of not acting is

$$E_N = C_{21}q + C_{22}(1-q)$$

As before, these expressions can be used to find the probability threshold p_t above which the user should take action

$$p_t = \frac{C_{12} - C_{22}}{C_{21} - C_{22} + C_{12} - C_{11}} = \beta .$$

β is a generalised version of the simple cost-loss ratio $\alpha = C/L$. Each term in the above expression can be written as a cost relative to the expense of the 'default' outcome where no action is taken and the adverse weather does not occur, so there is not loss of generality in setting this term C_{22} to zero as in Table 1. The remaining terms can be interpreted as follows. C_{12} is simply the cost of taking the protective action and C_{21} is the total loss that would be experienced if the adverse event occurred and no action was taken. In this more general situation we allow for the possibility that the action does not fully protect against the loss, so that C_{11} is greater than C_{12}. The denominator of β is the part of the loss which can be protected by taking action, so β still has the interpretation of the ratio of the cost of taking action to the protectable loss.

The expression for forecast value can be derived in this more general situation (Richardson 2000). The expression for V is the same, just with β replacing α. All the subsequent results apply equally to this general situation.

So far we have considered the different outcomes in terms of the financial consequences and have assumed the user acts to minimise the long-term mean expense. A further generalisation is to move away from the strict financial interpretation of the outcomes and introduce the concept of the utility of the outcomes. Rather than working directly with the costs C_{ij} in Table 2, consider a function $U(C_{ij})$ that represents in a more general sense the usefulness of this consequence to the user. All the previous derivations can again be made using the utilities $U(C_{ij})$ instead of the costs C_{ij}. In this case, the user aims to maximise the expected utility of his decisions (utility is a positive concept) rather than minimise the expected expense. Again, the same expression is found for value, with α now representing a ratio of the utilities instead of costs.

In terms of utility, the threshold probability above which the user should take action is

$$p_t = \frac{U(C_{12}) - U(C_{22})}{U(C_{21}) - U(C_{22}) + U(C_{12}) - U(C_{11})} = \gamma$$

If U is a linear function of the costs then $\beta = \gamma$, this utility ratio is the same as the cost-loss ratio: the threshold probability for maximising utility is the same as for minimising the expense. Users with linear utility functions are described as risk-neutral.

The relationship between utility and threshold probability is easiest to understand in the original cost-loss situation. On each occasion the user has to choose between the certainty of paying the cost C for protection and the risk of a greater potential loss L. p_t is the probability at which the user is indifferent between these two alternatives. For the risk-neutral user, this threshold was measured simply in terms of the mean expense of the two options and the threshold probability was shown to be $p_t = C/L$.

Now consider the threshold probability in terms of utility

$$p_t = \frac{U(C) - U(0)}{U(L) - U(0)} = \gamma$$

Consider the utilities of the four possible outcomes. The best outcome for the user is where there is no expense (no action is taken and the event does not occur). This then has maximum utility and we can put $U(0) = 1$. The worst outcome for the user is when the expense is greatest (loss L). This outcome has minimum utility and we can put $U(L) = 0$. These choices for maximum and minimum utility allow utility to be interpreted as a probability: the threshold probability and utility ratio become

$$p_t = 1 - U(C) = \gamma .$$

So the utility of taking action is equivalent to the threshold probability at which the user is prepared to act. For a linear utility function (risk-neutral user) our choices for $U(0)$ and $U(L)$ mean that $U(C) = 1 - C/L$, so that $p_t = C/L$. However, a user may be concerned about the immediate impact of a substantial loss L and may be prepared to pay C at lower probabilities to reduce the occurrence of this loss. This concern can be described by giving the user a higher utility for taking action $U(C)$. This will lower the threshold probability, so the user will take action at lower probabilities than would be indicated by a simple cost-loss analysis. Such a user is described as risk-averse: they are unwilling to risk the loss simply on the basis of minimising long term expense.

In this section we have introduced a number of generalisations of the cost-loss model. The expression for value from the cost-loss model is equally valid for all two-state, two-action decisions (as long as the consequences or utilities do not change with time). All the results presented in earlier sections are equally valid for the general case. Users who are averse to risk (probably the majority) will have utility ratios than would be suggested by the cost-loss ratios in the original model.

From a practical perspective the use of utility has several advantages. The user does not need to be able to express the consequences of his actions in financial terms. They do not even need to know precisely what the consequences are. The user does need to be able to express their preference for different outcomes. It is assumed that the user can choose the best and worst consequences. These are given utility 1 and 0 respectively. The utility of intermediate consequences is then deduced by preferences expressed by the user when they are offered a (hypothetical) choice between the certainty of a particular outcome or the possibility of either the best or worst outcome with a given probability. This is perhaps easier to obtain from the user than a precise evaluation of the costs and losses associated with each action.

6. References

Angstrom, A, 1922. On the effectivity of weather warnings. *Nordisk Statistisk Tidskrift*, **1,** 394-408.

Buizza, R., Richardson, D. S., and Palmer, T. N., 2003. Benefits of increased resolution in the ECMWF ensemble system and comparison with poor-man's ensembles. *Q. J. R. Meteorol. Soc.*, **129,** 1269-1288.

Kolb, L.L. and Rapp, R.R., 1962. The utility of weather forecasts to the raisin industry. *J. Applied Meteorology*, **1,** 8-12.

Liljas, E. and Murphy, A., H., 1994. Anders Angstrom and his early papers on probability forecasting and the use/value of weather forecasts. *Bull. Amer. Met. Soc.*, **75**, 1227-1236.

Leigh, R. J., 1995. Economic benefits of Terminal Aerodrome Forecasts (TAFs) for Sydney Airport, Australia. *Met. Apps.*, **2**, 239-47.

Murphy, A., H., 1966. A note on the utility of probabilistic predictions and the probability score in the cost-loss ratio decision situation. *J. Appl. Meteor.*, **5**, 534-537.

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., 1999. The application of cost-loss models to forecast verification. *Proceedings of the 7th Workshop on Meteorological Systems*. ECMWF, Shinfield Park, Reading, RG2 9AX, UK.

Richardson, D. S., 2000. Skill and relative economic value of the ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.*, **126,** 649-668.

Richardson, D. S., 2001. Measures of skill and value of Ensemble Prediction Systems, their interrelationship and the effect of ensemble size. *Q. J. R. Meteorol. Soc.*, **127**, 2473-2489.

Richardson, D. S., 2003. Economic value and skill. In *Forecast verification: a practitioner's guide in atmospheric science, Jolliffe, I. T. and Stephenson, D. B., Eds., Wiley*, 240pp.

Roebber, P. J. and Bosart, L. F., 1996. The complex relationship between forecast skill and forecast value: a real-world analysis. *Weather and Forecasting*, **11**, 544-559.

Thornes, J. E. and Stephenson, D. B., 2001. How to judge the quality and value of weather forecast products. *Met. Apps.*,**8**, 307-314.

Wandishin, M. S. and Brooks, H. E., 2002. On the relationship between Clayton's skill score and expected value for forecasts of binary events. *Meteorol. Appl*., **9,** 455-459.

Wilks, D. S., and Hamill, T. M., 1995. Potential economic value of ensemble forecasts. *Monthly Weather Review*, **123**, 3565-3575.

Zhu, Y., Iyengar, G., Toth, Z., Tracton, M. S., and Marchok, T., 1996. Objective evaluation of the NCEP global ensemble forecasting system. Preprints, *15th Conf. on Weather Analysis and Forecasting*, Norfolk, VA, Amer. Met. Soc., J79-J82.

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