Displaying Economic Value

Caren Marzban[1,2*]

[1] Applied Physics Laboratory, University of Washington, Seattle, WA 98195

[2] Department of Statistics, University of Washington, Seattle, WA 98195

Abstract

The distinction between forecast quality and economic value in a cost-loss formulation is well known. Also well-known is their complex relationship, even with some instances of a reversal between the two, where higher quality is associated with lower economic value, and vice versa. It is reasonable to expect such counter-intuitive results when forecast quality and economic value - both, multi-faceted quantities - are summarized by single scalar measures. Diagrams are often used to display forecast quality in order to better represent the multi-dimensional nature of forecast quality. Here, it is proposed that economic value be displayed as a *region* on a plot of hit rate versus false alarm rate. Such a display obviates any need to summarize economic value by a scalar measure. The choice of the axes is motivated by the ROC diagram, and so, this manner of displaying economic value is useful for deterministic as well as probabilistic forecasts.

*Corresponding Author: marzban@stat.washington.edu

1 Introduction

The general situation involving a binary event, and binary or probabilistic forecasts of the event, is thoroughly studied (Jolliffe and Stephenson 2003; Wilks 2006). The quality of such forecasts is a multifaceted quantity (Murphy and Winkler 1987, 1992), and therefore, information regarding forecast quality is lost when it is summarized by a single, scalar measure. For example, the probability of detection (or hit rate), and the false alarm rate, individually provide an incomplete assessment of forecast quality. A high hit rate (suggesting high quality) may be accompanied by a high false alarm rate (suggesting low quality), or vice versa. Focusing on any single measure can lead to a completely false assessment of the true quality of forecasts.

The multi-faceted nature of forecast quality calls for a paradigm where diagrams take the place of a single measure of quality. For probabilistic forecasts, attributes diagrams, refinement diagrams, discrimination plots, and Relative Operating Characteristic (ROC) curves, are used to display different facets of forecast quality. For categorical (deterministic) forecasts, a recent proposal includes the performance diagram (Roebber 2009; Taylor 2001), where multiple scalar measures are plotted on a single diagram. Although there are differences between these diagrams, what is common to them is that they acknowledge the importance of displaying forecast quality in a multidimensional fashion, i.e., via a diagram.

Another well-studied problem involves the situation when a binary decision or action is to be based on forecasts (Doswell and Brooks 1998; Katz and Murphy

1997; Mason 2004; Richardson 2000; Wandishin and Brooks 2002; Wilks 2001, 2006). The concept that arises from considering such problems is the economic value of the forecasts. The economic value and the quality of forecasts are different facets of forecast goodness, often with a complex relationship between them (Murphy and Ehrendorfer 1987; Roebber and Bosart 1996).

Assessing economic value (henceforth, value) requires a specification of the costs and losses incurred in taking, or not taking, an action based on the forecasts. However, not all decision problems lend themselves to a cost-loss analysis. For example, Stewart, Pielke, and Nath (2004) discuss the value of precipitation forecasts with respect to snow removal. They show that the decision making process in that problem is too complicated for a simple cost-loss analysis. And even when the problem is simplified, they find that the unavailability of appropriate data can preclude a proper assessment of value. In the following, it is assumed that the decision process does allow for a cost-loss analysis.

The importance of examining the value of forecasts has been highlighted in a wide range of practical problems. Palmer (2002) compares three different precipitation forecasting systems, and shows that ensemble systems have higher value than deterministic forecasts. In the context of Terminal Aerodrome Forecasts (TAFs), Keith (2003) has shown that airlines can benefit, through lower fuel usage, when the value of TAFs is taken into account. He shows that for some flights, even moderate-quality forecasts can provide most of the economic savings gained by perfectly reliable TAFs. Keith and Leyton (2007) show that the value of probabilistic forecasts at airports can be used to determine the optimal amount of fuel to be carried by an airplane. The value of hurricane forecasts is taken into account by Letson, Sutter, and Lazo (2007) when they compare the benefits from multiple actions, e.g., "improved forecast provision and dissemination vs. alternative public investments such as infrastructure or forecasts of other hazards." The value of wind forecasts has been examined for the Utilities (Milligan, Miller, and Chapman 1995), and Teisberg, Weiher, and Khotanzad have analyzed the value of temperature forecasts in electricity generation. All of these studies demonstrate the usefulness of examining the value of forecasts in conjunction with their quality.

In its simplest realization, value, like quality, is summarized by a single measure, which is plotted as a function of a quantity depending on the costs and losses. Also like quality, value is a multifaceted notion, and so information is lost when it is summarized by a single measure. Thus, neglecting the multi-dimensional nature of value can lead to false conclusions. At worse, it can lead one to believe that forecasts have high value, when in fact, they do not. Or, the forecasts may be declared to have little value, when in reality they have high value. Neglecting the multi-faceted nature of value can also lead to counter-intuitive or apparently contradicting conclusions. Indeed, a reversal of quality and value has been noted in the literature, where higher quality is associated with lower value, or vice versa (Murphy and Ehrendorfer 1987). This type of counter-intuitive result can be explained by a number of arguments, including one proposed by Mason (2004) where the culprit behind the unexpected result is attributed to a non-optimal probability threshold. For deterministic forecasts, an alternative explanation is that the aforementioned reversal occurs because multifaceted quantities (i.e., quality and value) are summarized by single, scalar measures. Avoiding a scalar measure of value not only precludes such counter-intuitive conclusions, but also provides a more complete, and therefore more useful, representation of the value of forecasts.

In this note it is proposed that the value of binary (deterministic) forecasts be displayed as a *region* on a plot of the hit rate versus the false alarm rate, without the need for a scalar summary measure for value at all. Such a plot is of course the "background" upon which the ROC curve is drawn, and so, the proposed method of displaying value is useful for probabilistic and deterministic forecasts alike.

2 Forecast Quality and Economic Value

Denote the occurrence or nonoccurrence of an event with a 1 or 0, respectively, and similarly, for taking an action (1), or not taking an action (0). Let H and F, represent the hit rate and the false alarm rate, respectively. A commonly used (scalar) measure of forecast quality, specifically of discrimination, is the True Skill Score (TSS), also known by many other names (Wilks 2006):

$$TSS = H - F. \tag{1}$$

In the cost-loss formulation, an often-used toy-model for the loss-matrix is (Berger

1985; Katz and Murphy 1997)

$$
\begin{array}{cc} & \text{Action} \\ & \begin{array}{cc} 0 & 1 \end{array} \\ \text{Event} \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 0 & C \\ L & L_m \end{pmatrix} \end{array} . \tag{2}
$$

In this model, if an event does not occur, and no action is taken (Action = 0), then no cost or loss is incurred. If an event does not occur, but action is taken (Action = 1), then a cost C is incurred. If an event does occur, and no action is taken, then the user loses an amount L. Finally, if an event occurs, and an action is taken, then a loss of L_m is incurred. Note that the choice of the loss matrix is completely independent of forecasts or their quality.

The following quantities are central to ascertaining the value of forecasts (Katz and Murphy 1997; Wilks 2006):

$$
\begin{aligned}
E_0 = [\text{Expected cost if Action} = 0] &= p\,L \ , \\
E_1 = [\text{Expected cost if Action} = 1] &= (1-p)\,C + p\,L_m \ , \\
E_f = [\text{Expected cost if Action} = \text{forecast}] &= (1-p)\,F\,C + p\,(1-H)\,L + p\,H\,L_m \ , \\
E_p = [\text{Expected cost if Action} = \text{perfect forecasts}] &= p\,L_m \ , \\
E_r = [\text{Expected cost if Action} = \text{random forecasts}] &= (1-p)\,F_1\,C + p\,F_0\,L + p\,F_1\,L_m \ ,
\end{aligned} \tag{3}
$$

where p is the prior (climatological) probability of the occurrence of an event, and F_0 and F_1 are the proportion of forecasts of “0” and “1”, respectively. In the absence of forecasts, it is self-evident that an action should be taken if it leads to a lower expected cost. Defining the ratio,[1]

$$Cl = \frac{C}{C + L - L_m} \ , \tag{4}$$

then the first two equations in Eq. 3 imply

$$\text{Take action if } \ p > Cl \ \ ; \ \ \text{otherwise, do not act.} \tag{5}$$

In other words, from an economic point of view, and without any forecasts, it is beneficial to always take action if the probability of an event exceeds the ratio of cost and loss appearing in Cl.[2]

A popular scalar measure of forecast value is (Richardson 2000, Wilks 2001)

$$V \ = \ \frac{\text{Expected savings from forecasts}}{\text{Expected savings from perfect forecasts}} = \frac{min(E_0, E_1) - E_f}{min(E_0, E_1) - E_p} \ , \tag{6}$$

$$= \ \frac{min(p, \frac{C}{L} - p(\frac{C - L_m}{L})) - p - (1-p)(F)\frac{C}{L} + p(H)(1 - \frac{L_m}{L})}{min(p, \frac{C}{L} - p(\frac{C-L_m}{L})) - p\frac{L_m}{L}}$$

Although this is normally the way V is written, it simplifies considerably when written for $p < Cl$ and $p > Cl$, separately:

$$V = \begin{cases} H - RF & \text{if } \ p < Cl \ \ (i.e., R > 1) \\ (1-F) - \frac{1}{R}(1-H) & \text{if } \ p > Cl \ \ (i.e., R < 1) \end{cases} \ , \tag{7}$$

where the quantity R is defined as[3]

$$R = (\frac{1-p}{p})(\frac{Cl}{1-Cl}) \ . \tag{8}$$

---

[1]Richardson (2000) denotes this quantity α.

[2]If $Cl > 1$, then no action should be taken, independently of p, because the expected cost associated with no action is always lower than that associated with action. But if $0 < Cl < 1$, then the optimal decision depends on the value of p. For this reason, only $0 < Cl < 1$ is examined here.

[3]If both p and Cl are very small, i.e., $p << 1$, and $Cl << 1$, then $R \sim Cl/p$. With $L_m = L$, Cl

Note that

$$p < Cl \Longleftrightarrow R > 1 \quad \text{and} \quad p > Cl \Longleftrightarrow R < 1 \ . \qquad (9)$$

If $p = Cl$, i.e., $R = 1$, then $V = H - F = TSS$. In other words, if $p = Cl$, then V, i.e., a measure of value, reduces to a measure of quality.

3 Reversal of Quality and Value

The above presentation allows for a simple demonstration of the aforementioned reversal phenomenon noted in the literature (Mason 2004; Murphy and Ehrendorfer 1987). TSS and V both depend on H and F, and so are best displayed on a plot of H vs. F. This choice of the variables across the x- and y-axes is the same as that of the Relative Operating Characteristic (ROC) diagram (Fawcett 2006, Marzban 2004). On a plot of H vs. F, both TSS = constant and V = constant are straight lines. Figure 1 shows the (solid) lines $TSS = 0.3$, $TSS = 0.4$, and the (dashed) lines $V = 0.1$, $V = 0.2$. The former have slope 1, while the slope of the latter is $R = 1.6$.[4] Consider two forecasting systems corresponding to the filled and open circles in Figure 1. One has higher quality (TSS) than the other, but lower value (V). This type of reversal may seem concocted, but it is actually quite natural and is dictated by the geometry of two parallel lines intersecting another pair of parallel lines. Note that if

becomes C/L. Some reported C/L ranges are as follows: For orchardists, 0.02-0.05 (Murphy 1977); loading of fuel for airplanes, 0.01-0.12 (Leigh 1995); and winter road-gritting, 0.125 (Thornes and Stephenson 2001).

[4]For small p and Cl, an R of 1.6 corresponds to $p/Cl = 1/R = 0.625$.

$p = Cl$ (i.e., $R = 1$), then all lines have slope = 1, and so no intersection between the lines can occur. In other words, this type of reversal occurs when $p \neq Cl$.

4 Value Region

Such a reversal of quality and value arises because they are summarized by single, scalar measures. The reason is that by virtue of being a scalar quantity, V does not fully capture the multiple facets of value. But one can convey a more complete representation of value by returning to the expected costs in Eq. 3, from which V is constructed. For example, it is reasonable to define valuable forecasts as those satisfying

$$E_p < E_f < E_0,\ E_1,\ \text{and}\ E_r \quad , \tag{10}$$

In other words, the expected cost from using forecasts ought to be greater than that from perfect forecasts, but lower than either E_0, E_1, and the cost associated with acting according to random forecasts. Here, Eq. 10 *defines* economically valuable forecasts.[5]

In light of the expected costs in Eq. (3), the inequalities in Eq. (10) become

$$L_m < L \ , \ \frac{H}{F} > R \ , \ \frac{1-H}{1-F} < R \ , \ H > F \ . \tag{11}$$

The left-most constraint is trivial in that it is satisfied on physical grounds. It is easy to show that the right most constraint is implied by the middle two constraints,

[5]The $<$ in Eq. 10 may be changed to $\leq$, but not much is gained from that revision.

which together place a severe restriction on possible values of H, F, and R. These two constraints can be written as

$$H > RF \;\;,\;\; H > (1 - R) + RF \;, \tag{12}$$

and their linearity in H and F again allows for a simple representation on a diagram of H vs. F. For a given p/Cl, only one of these constraints is nontrivial, depending on $p < CL$, (i.e., $R > 1$), or $p > Cl$, (i.e., $R < 1$), respectively. Therefore, the region corresponding to valuable forecasts - termed "value region" - is a triangular region bounded below by the lines $H = RF$ or $H = (1 - R) + RF$. To make a connection with the scalar measure of value, V, these lines correspond to $V = 0$, depending on whether $p < CL$ or $p > Cl$. The connection between the value region and $V =$ constant lines is entirely expected. The important point, however, is that value can be displayed without a scalar measure.

For a given H and F, it is possible to solve Eq. 12 for the the critical value of R separating forecasts with value from those without value. R-values corresponding to valuable forecasts are

$$\frac{1 - H}{1 - F} < R < \frac{H}{F} \;. \tag{13}$$

These constraints on R can be translated to constraints on Cl, by virtue of Eq. 6. Forecasts with value must have

$$\frac{1}{1 + (\frac{1-p}{p})(\frac{1-F}{1-H})} < Cl < \frac{1}{1 + (\frac{1-p}{p})(\frac{F}{H})} \;. \tag{14}$$

Again, in order to make a connection with the scalar measure of value, V, the interval specified in Eq. (14) is where V is non-negative. It is worth pointing out that Richardson (2000) shows that the quantities appearing in Eq. 14 are the conditional probability of the occurrence of an event, given a forecast of "yes" or "no", respectively. The particular form in Eq. 14 (not written in terms of these conditional probabilities) is intended to distinguish F and H, which assess forecast quality, from p which is determined by climatology.

The value region for different values of p and Cl is shaded in gray in Figure 2. The specific values of p and Cl selected here are 0.001, 0.008, 0.018, 0.279, and 0.99 (motivated by the "Finley data," described in the next paragraph.) Only forecasts whose hit rate and false alarm rate fall in the shaded region are economically valuable. Also, note that the extent of the value region is maximum when p and Cl are comparable. For situations when they differ significantly, the value region is relatively small. This is a reflection of the fact that taking action, and not taking action, are the optimal decisions when $p >> Cl$ and $p << Cl$, respectively.

As a concrete example, consider the Finley data (Murphy 1996) whose contingency table is shown in Table 1. One has $H = 0.549$, $F = 0.026$, $TSS = 0.523$, and $p = 0.018$. The filled circle in the $p = 0.018$ panels in Figure 2 corresponds to this data. The manner in which it falls inside, on the boundary, or outside the value region, depending on Cl, is evident. Indeed, according to Eqs. (13) and (14), Finley forecasts have value only if $0.463 < R < 21.11$, or $0.0084 < Cl < 0.279$. For users whose Cl falls outside of this range, the forecasts have no value at all, regardless of their quality. As such, the user should ignore the forecasts and simply either act, or not act, according to the prescription in Eq. (5). Again, note that no scalar measure of value has been summoned in this representation of value.

This display of value is useful even when forecasts are probabilistic, or of a type for which ROC curves can be generated. The panels in Figure 2 also show an ROC curve based on a binormal model (Marzban 2004) going through the “Finley point.”[6] It does not correspond to any “real” probabilistic forecasts of tornados. Its purpose is only to demonstrate the interplay between the ROC curve and the value region. For example, if $p \neq Cl$, i.e., away from the diagonal panels in Figure 2, the ROC curve has segments that do not fall in the value region. In other words, even though the underlying probabilistic forecasts clearly have high quality, reflected in the high arc of the ROC curve, or the large area under it, for some probability thresholds the resulting binary forecasts have no value. This is consistent with the arguments of Mason (2004) for being careful in selecting optimal thresholds when value is summarized by V.

5 Uncertainty

The value region is bounded below by a straight line whose slope and intercept are determined by R only; see Eq. 12. As such, the only source of uncertainty with respect to the extent of the value region is R itself. Given that R is determined by p and Cl, the uncertainty in R, denoted δR, can be computed from uncertainty in the latter, denoted δp, and δCl, respectively. It is reasonable to assume that uncertainty

[6]The means of the two normal distributions are $-1.035, +1.035$, and both standard deviations are 1.

in p is independent of uncertainty in Cl, because the former is estimated from data, while the latter depends on the specifics of a user. If this assumption is valid, then Eq. 8 implies

$$(\frac{\delta R}{R})^2 = \frac{1}{(1-p)^2}(\frac{\delta p}{p})^2 + \frac{1}{(1-Cl)^2}(\frac{\delta Cl}{Cl})^2 \quad . \tag{15}$$

This equation allows one to compute the uncertainty in R from uncertainties in p and Cl. To simplify further, one may assume that events and nonevents occur independently (a poor assumption), in which case $(\delta p)^2$ can be measured by the variance of p, i.e., $\frac{p(1-p)}{N}$, where N is the total number of events and nonevents in the sample. Under this assumption, the first term in Eq. 15 is inversely proportion to N, and so it will be negligible relative to the second term, for a sufficiently large sample size. In that case, Eq. 15 simplifies to $\frac{\delta R}{R} = \frac{1}{1-Cl}\frac{\delta Cl}{Cl}$. So, for sufficiently large sample, this simple formula can readily compute the uncertainty in R from the uncertainty in Cl itself.

The visual effect of uncertainty is to broaden the lower boundary line of the value region into a triangular region. This is shown in Figure 3, when $p < Cl$ (left panel) and $p > Cl$ (right panel). In these figures, and for purely visual purposes, the specific uncertainty in p is $\pm.05$. Also shown in these figures, only as a point of reference, is the "Finley point" along with its uncertainty in both the x and y directions. Given that HR and FR are proportions, and assuming independence of daily tornadic activity (again, a poor assumption), their uncertainty is proportional to the standard deviation of a proportion, based on a sample of size n, i.e., $\sqrt{p(1-p)/n}$. The "cross" atop the Finley point in Figure 3 is ± 2 times this standard deviation. The horizontal segment of the cross is much smaller than the vertical segment, because, as evidenced in the Finley data set (Table 1), the sample size from which FR is computed (i.e., 2680+72) is much larger than that upon which HR is based (23+28). Still, these figures illustrate how uncertainty plays into the value region: The HR and FR (as well as their uncertainty) of forecasts will generally be represented by a cross, while the uncertainty in p and Cl will lead to a broadening of the lower boundary of the value region. If a significant portion of the cross falls within the boundary region's "significant portion," then one may conclude that the value attributed to the forecasts is unlikely to be due to chance. Given that these displays are intended to provide information in a visual manner, and more information than that provided by a single number, it is unnecessary to quantify that (un)likelihood by a statistical test. In the cases displayed in Figure 3, for example, one would be justified in concluding that the value is not due to chance. On the other hand, if a significant portion of the cross falls outside of the value region, then one can only conclude that the data do not provide sufficient information in support of any claim regarding the population/true value of economic value.

6 Summary and Discussion

It is proposed that the (economic) value associated with a set of forecasts not be measured by a scalar quantity. It is argued that forecasts can be said to have value if the expected cost associated with actions linked to forecast satisfies some very general and reasonable inequalities. For example, actions linked to valuable forecasts ought to have a lower expected cost than actions based on random forecasts, or actions not based on forecasts at all. The inequalities define a *value region* which is most naturally displayed on a diagram of the hit rate versus the false alarm rate, i.e., the "background" upon which the ROC curve is drawn. As such, quality and value can be displayed on a single diagram, without a summary measure for either. If a point on the ROC diagram falls within the value region, then the deterministic forecasting system can be said to have value. A consequence of using the value region is that one can no longer rank different forecasting systems, because ranking requires a scalar measure; all systems within the value region must be treated as equal in terms of their value. For probabilistic forecasts, or other forecasts for which an ROC curve can be produced, the portion of the ROC curve which falls within the value region is said to have value. In this way, some probability thresholds lead to valuable forecasts, some do not. In addition to offering a more complete picture of value, use of the value region (as opposed to a scalar measure) also precludes counter-intuitive conclusions such as the reversal of the relationship between quality and value. The value region is defined essentially by the equation of a straight line, and is therefore extremely easy to compute without any sophisticated computer code. The formulas for computing the uncertainty in the value region are also simple to implement.

The connection between the value region and the scalar measure, V, is simple: The former corresponds to all points on the ROC diagram for which the latter is non-negative. This connection is not surprising because both concepts are based on the same set of expected costs. However, the value region carries more information by virtue of being a 2-dimensional quantity. And displaying it on an ROC diagram in particular, makes it especially useful given the ubiquity of ROC diagrams.

In addition to the works mentioned in the Introduction, the connection between the ROC curve and expected cost has also been examined in fields outside of meteorology; two of these works are worth discussing here because of their close connection to the notion of the value region. Provost and Fawcett (1997) consider the slope of an "iso-performance line," which is the locus of points on an ROC diagram with equal expected cost, E_f. Although they examine the situation where $L_m = 0$, it is easy to show that the slope of their iso-performance line is exactly the slope of the constant value line, with value defined by the scalar measure V in Eq. (7).

Drummond and Holte (2006) begin with the connection between cost (or loss) and the slope of a line on the ROC diagram, but argue that the comparison of different forecasting systems is hampered by the visual effort to compare slopes of lines (tangent to the ROC curve, for example). Instead, they consider an alternative diagram involving what they call "cost curves," defined as straight lines in "cost space," i.e., a plot of expected cost versus $\frac{pL}{pL+(1-p)C}$.[7] ROC space and cost space are described as having a "dual" relationship in that the intercept and slope of a cost curve are determined by the coordinates of a point on the ROC diagram; similarly a line on the ROC diagram translates to a point in cost space. The main advantage of cost curves over

[7]The analysis in Drummond and Holte (2006) is based on a cost model where $L_m = 0$, in which case the x-axis of the cost curve is equal to $\frac{1}{1+R}$.

ROC curves is that the former make it easier to compare two different forecasting systems in terms of the expected cost associated with the forecasts. Another advantage is that confidence intervals for cost curves can be more easily displayed. The notion of a cost curve has many similarities to the notion of a value region proposed here; but there are some important differences. For example, the value region is displayed on an ROC diagram. In spite of all the advantages of cost curves over ROC curves, the latter are still useful and commonly employed. As such, displaying the value region adds useful information to the ROC diagram. Also, whereas cost curves are conducive to the comparison of two forecasting systems, the value region is useful even when examining a single forecasting system, i.e., a single point on the ROC diagram (e.g., for deterministic forecasts), or a single ROC curve (e.g., for probabilistic forecasts).

Acknowledgments We are grateful to Harold Brooks and Daniel S. Wilks for valuable comments. David Jones is acknowledged for providing the initial impetus for this work.

References

Doswell, C. A., H. Brooks, 1998: Budget Cutting and the value of weather services, *Wea. Forecasting*, **13**, 206-212.

Fawcett, T., 2006: An introduction to ROC analysis. *Pattern Recognition Letters*, **27**, 861874.

Jolliffe, I.T., and D.B. Stephenson, 2003: *Forecast verification: a practitioner's guide in atmospheric science.* John Wiley and Sons.

Katz, R.W., and A.H. Murphy, 1997: *Economic Value of Weather and Climate forecasts.* Cambridge University Press, 225 pp.

Keith, R. 2003: Optimization of value of aerodrome forecasts. *Wea. Forecasting*, **808-824**

Keith, R., S. M. Leyton, 2007: An experiment to measure the value of statistical probability forecasts for airports. *Wea. Forecasting*, **22**, 928-935.

Letson, D., D. S. Sutter, and J. K. Lazo, 2007: Economic value of hurricane forecasts: An overview and research needs. *Natural Hazards Review*, **8:3**, 78-86.

Mason, I. 2004: The cost of uncertainty in weather prediction: modelling quality-value relationships for yes/no forecasts. *Aus. Met. Mag.*, **53**, 111-122.

Marzban, C. 1998: Scalar measures of performance in rare-event situations. *Weather and Forecasting*, **13**, 753-763.

Marzban, C. 2004: The ROC Curve and the Area Under it as a Performance Measure. *Wea. Forecasting*, **19**, 1106-1114.

Milligan, M. R., A. H. Miller, and F. Chapman, 1995: Estimating the economic value of wind forecasting to utilities. *Proceedings of Windpower '95*, March 27-30, Washington, DC.

Murphy, A.H., 1991: Forecast verification: Its Complexity and Dimensionality. *Monthly Weather Review*, **119**, 15901601.

Murphy, A.H., 1993: What Is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting*, **8**, 281-293.

Murphy, A. H., 1996: The Finley Affair: A signal event in the history of forecast verification. *Wea. Forecasting*, **11(1)**, 3-20.

Murphy, A.H., and M. Ehrendorfer, 1987: On the relationship between the accuracy and value of forecasts in the Cost-Loss ratio situation. *Wea. Forecasting*, **2**, 243-251.

Murphy, A.H., and R.L. Winkler, 1987: A general framework for forecast verification. *Monthly Weather Review*, **115**, 1330-1338.

Murphy, A.H., and R.L. Winkler, 1992: Diagnostic verification of probability forecasts. *International Journal of Forecasting*, **7**, 435-455.

Palmer, T. N., 2002: The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Q. J. R. Meteorol. Soc.*, **128** , 747-774.

Richardson, D.S., 2000: Skill and relative economic value of the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, **126**, 649-667.

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

Roebber, P.J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601-608.

Stewart, T. R., Pielke, R. Jr., Nath, R., 2004: Understanding user decision making and value of improved precipitation forecasts: Lessons from a case study. *Bull. Amer. Meteor. Soc.*, **85**, 223-235.

Taylor, K.E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106 (D7)**, 7183-7192.

Teisberg, T. J., R.F., Weiher, and A. Khotanzad, 2005: The economic value of temperature forecasts in electricity generation. *Bull. Amer. Meteor. Soc.*, **86**, 1765-1771.

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

Wandishin, M.S., and H. E. Brooks, 2002: On the relationship between Clayton's skill score and expected value for forecasts of binary events.

Wilks, D.S., 2001: A skill score based on economic value for probability forecasts. *Meteorological Applications*, **8**, 209-219.

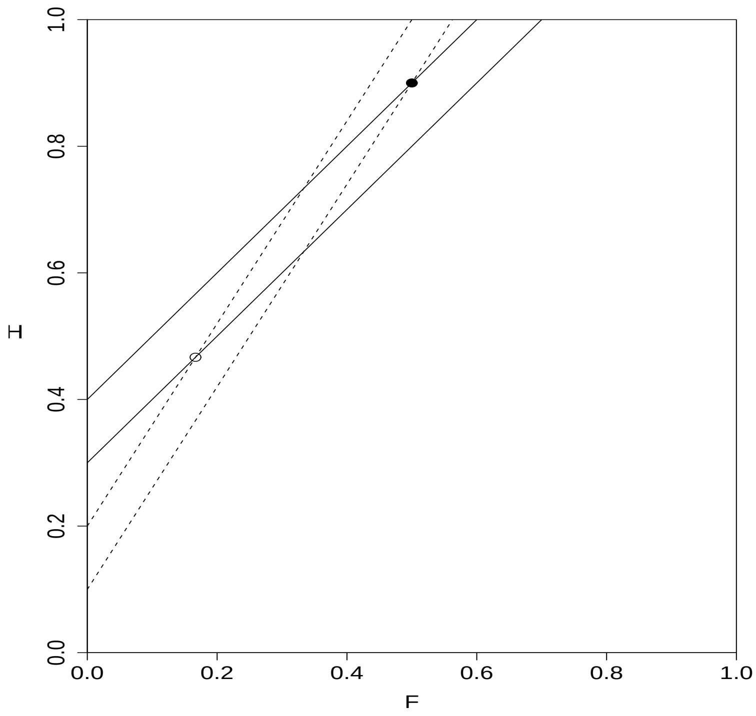
Wilks, D.S., 2006: *Statistical Methods in the Atmospheric Sciences*. Academic Press, San Diego, CA, 627 pp.

Figure Captions

Figure 1. A demonstration of the reversal of quality and value. The solid parallel lines correspond to $TSS = 0.3$ and 0.4, and the two dashed lines represent forecasts with $V = 0.1$, and 0.2. The forecasts corresponding to the filled circle have higher quality, but lower value than the forecasts associated with the open circle.

Figure 2. The value region (shaded) for different values of p and Cl. The filled circle (in the middle row) corresponds to the Finley data set. The arched curve is an example of an ROC curve based on a binormal model whose parameters have been chosen so that the ROC curve goes through the "Finley point."

Figure 3. The effect of uncertainty on the value region, when $p < CL$ (left) and $p > Cl$ (right). The "cross" on the left side of the graph shows the "Finley point" and its uncertainty.



401 Figure 1. A demonstration of the reversal of quality and value. The solid parallel lines correspond to $TSS = 0.3$ and 0.4, and the two dashed lines represent forecasts with $V = 0.1$, and 0.2. The forecasts corresponding to the filled circle have higher quality, but lower value than the forecasts associated with the open circle.

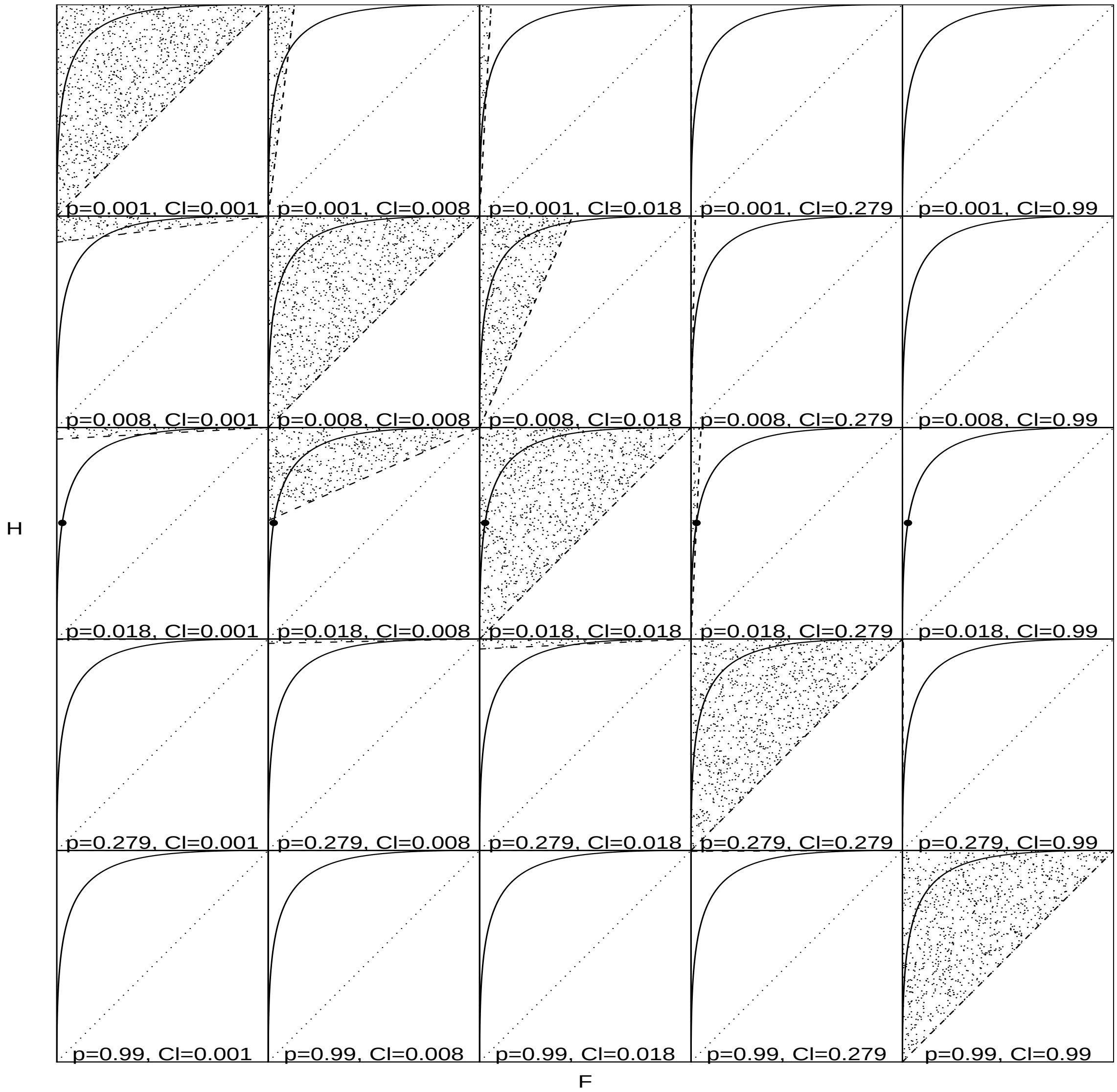


Figure 2. The value region (shaded) for different values of p and Cl. The filled circle (in the middle row) corresponds to the Finley data set. The arched curve is an example of an ROC curve based on a binormal model whose parameters have been chosen so that the ROC curve goes through the "Finley point."

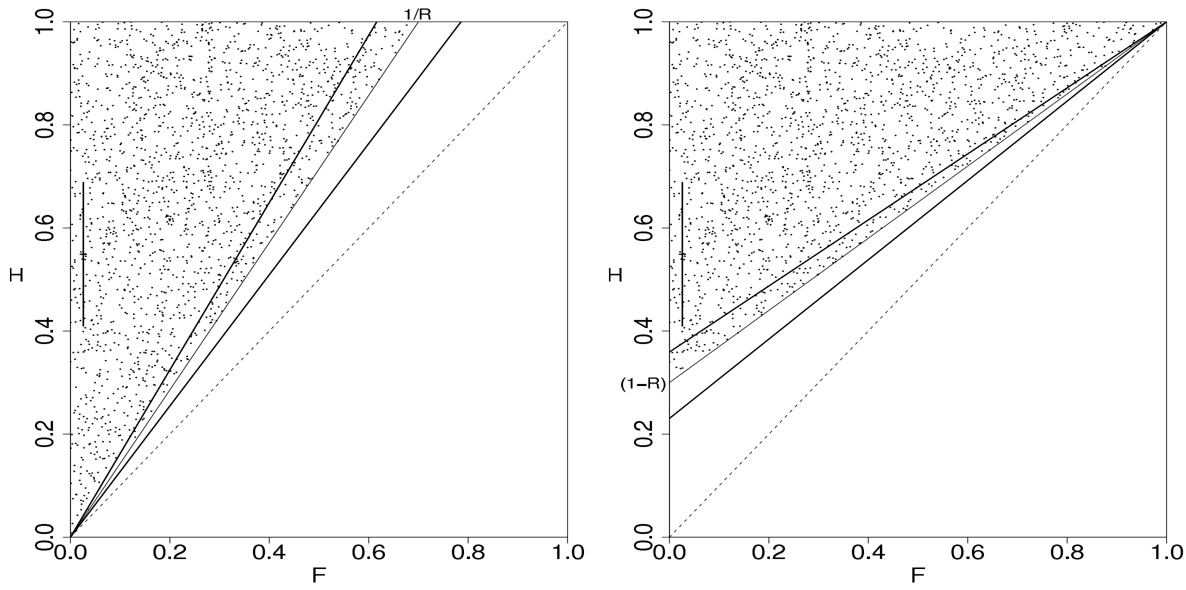


Figure 3. The effect of uncertainty on the value region, when $p < CL$ (left) and $p > Cl$ (right). The "cross" on the left side of the graph shows the "Finley point" and its uncertainty.

.	no tornado forecast	tornado forecast
no tornado observed	2680	72
tornado observed	23	28

Table 1. The Finley data set.