The Boy Who Cried Wolf Revisited: The Impact of False Alarm Intolerance on Cost–Loss Scenarios

M. S. ROULSTON AND L. A. SMITH

Pembroke College, Oxford University, Oxford, and Centre for the Analysis of Time Series, London School of Economics, London, United Kingdom

(Manuscript received 31 October 2002, in final form 10 July 2003)

ABSTRACT

Meteorologists often interpret the value of a probabilistic weather forecast using the binary cost–loss scenario. The socioeconomic benefit of such a forecast will depend on the compliance rate of users and, hence, the number of warnings that are not followed by the corresponding high-impact weather. A modified version of the canonical binary cost–loss problem in which the compliance rate of users is a function of the warning probability threshold, and hence of the ‘‘false alarm rate’’, is presented. In this version of the problem, the value of the forecast can be enhanced by choosing a probability warning threshold that is higher than the cost–loss ratio. It is found that the advantage of modifying the probability warning threshold is greatest when the frequency of highly confident forecasts of an event is relatively high and when users are moderately intolerant of false alarms. Using this simple example it is illustrated that forecasters who issue nonprobabilistic, or ‘‘unequivocal,’’ forecasts are making implicit assumptions about the false alarm intolerance of users, as well as assumptions about their cost–loss ratios. Adopting a probabilistic approach to forecasting avoids these assumptions and separates the activity of forecasting from the activity of decision making.

1. Introduction

Aesop’s fable, the ‘‘The Boy who Cried Wolf,’’ is about a young shepherd boy guarding the village flock who cries that there is a wolf worrying the sheep. The villagers rush out to protect the sheep, but there is no wolf. This event is repeated two or three times before a wolf actually does show up on the hillside. The boy cries ‘‘Wolf!,’’ but to no avail: the villagers, no longer regarding the warning as credible, fail to act and the wolf decimates the flock. The traditional moral of this tale is that liars are not believed, even when telling the truth. But, are we being too harsh on the shepherd boy? Perhaps his mistake was to overestimate the rationality of the villagers. This traditional tale can be expressed in terms of a cost–loss problem, used in the analysis of the utility of weather forecasts (Ångström 1922; Murphy 1966; Liljas and Murphy 1994; Richardson 2000). The villagers’ cost–loss matrix is

	No wolf	wolf
Villagers respond	C	C
Villagers do not respond	0	L,

where C is the price the villagers paid when they ran to aid the shepherd boy. This price is largely an opportunity cost, the value of the goods and services they could have been producing had they not been running about the hills. The loss associated with losing part of the flock to a lupine predator is represented by L. If the shepherd boy believed the probability that there was a wolf about is p, then the expected cost to the village is[1]

$$E[\mathrm{COST}] = \begin{cases} pL & \text{if you do not cry wolf} \\ C & \text{if you cry wolf.} \end{cases} \quad (1)$$

To minimize the expected cost, the boy should have cried wolf if $p > C/L$. If we assume that the flock of sheep was one of the village’s most important assets, and that a wolf attack would result in the loss of several sheep, it seems reasonable to say that the villagers’ cost–loss ratio, C/L, was quite modest. An estimate of C/L using contemporary prices gives

$$\frac{C}{L} \approx \frac{\text{six villagers for 1 h at \$10 h}^{-1}}{\text{three sheep at \$200 each}} = \frac{\$60}{\$600} = 0.1, \quad (2)$$

where three has been taken as a reasonable numerical value of ‘‘several.’’ The cost–loss analysis implies that

Corresponding author address: Mark S. Roulston, Pembroke College, Oxford OX1 1DW, United Kingdom.
E-mail: roulston@maths.ox.ac.uk

[1] Whether or not the probability p was a reliable estimate of the probability would not affect the shepherd boy’s behavior, as long as the shepherd boy *believed* p to be a reliable estimate.

the shepherd boy should have issued a warning even if he believed there was only a 10% probability of lupine activity. While this is a rational strategy, it would obviously have resulted in a high rate of ''false alarms''—possibly as high as 90%, if the probabilities underlying the forecasts were reliable.[2] The fact that the villagers were apparently unprepared to tolerate a false alarm rate of around 67% seems to suggest that they had not performed a cost–benefit analysis of responding to the cries of ''Wolf!'' Unfortunately, such behavior is not confined to the characters in stories written over two millenia ago. Psychological factors influence people's perception of risk, and this perception affects the way that they interpret meteorological forecasts. In this paper, we examine how the public's compliance with warnings might affect the value of the forecasts, and whether the forecast value can be increased by taking the user's response into account. We do this in the context of the familiar cost–loss scenario. The standard cost–loss scenario is a *prescriptive* decision model, which specifies how a user *should* respond. In contrast, *descriptive* decision models describe how users actually do respond (Stewart 1997). The modified cost–loss model we present is an idealized descriptive model, but it provides a quantitative illustration of the descriptive–prescriptive distinction. We hope that a better appreciation of the importance of this distinction will motivate further study in the area of descriptive modeling. The classical cost–loss problem illustrates that any forecast that is nonprobabilistic includes an implicit assumption about the cost–loss ratios of users. The extension presented in this paper demonstrates that such nonprobabilistic forecasts also include an implicit assumption concerning the tolerance of users to false alarms. Forecasters who issue strictly probabilistic forecasts do not make any such tacit assumptions. With probabilistic forecasts, the process of forecasting can be explictly separated from the process of decision making in a way that is not possible with unequivocal (nonprobabilistic) forecasts.

2. Cost–loss scenario with imperfect compliance

To model imperfect compliance with warnings in the cost–loss scenario we introduce a compliance rate, q. This is the probability that action will be taken if a warning is given. The expected cost, E[COST], if the forecast probability is p is thus

[2] In this paper we use the phrase ''false alarm rate'' to mean the proportion of times the warning is issued but the event does not occur. The term ''false alarm rate'' is unfortunate in that is perpertuates the belief that one can evaluate a single so-called probability forecast: perfect probability forecasts would still generate ''false'' alarms or nonevent warnings. We prefer not to introduce new jargon here but instead suggest that compliance will improve when decision makers understand this point.

$$E[\mathrm{COST}] = \begin{cases} pL & \text{when no warning is given} \\ qC + (1-q)pL & \text{when a warning is given.} \end{cases} \quad (3)$$

The first term on the lower line of the right-hand side (rhs) of Eq. (3) denotes the expected cost of protective action being taken, while the second term is the expected cost of a loss if no protective action is taken. If warnings are issued when the forecast probability exceeds some threshold, p_w, then the expected cost averaged over time is given by

$$\langle E[\mathrm{COST}]\rangle = \int_0^{p_w} \rho(p)pL\,dp + \int_{p_w}^1 \rho(p)[qC + (1-q)pL]\,dp, \quad (4)$$

where $\rho(p)$ is the frequency distribution of forecasted probabilities. The first term on the rhs of Eq. (4) represents days on which no warning is issued, and the second term is days on which a warning is issued. Equation (4) can be rewritten as

$$\langle E[\mathrm{COST}]\rangle = L\int_0^1 \rho(p)p\,dp + qC\int_{p_w}^1 \rho(p)\,dp - qL\int_{p_w}^1 \rho(p)p\,dp. \quad (5)$$

The problem becomes analytically tractable if we assume that the forecast probability, p, is drawn from a uniform distribution; that is, $\rho(p) = 1$ on the interval [0, 1]. In this case, Eq. (5) can be rewritten

$$J = \frac{\langle E[\mathrm{COST}]\rangle}{L} = \frac{1}{2} + qc(1 - p_w) - q\left(\frac{1}{2} - \frac{p_w^2}{2}\right), \quad (6)$$

where J is the expected cost per unit loss and $c = C/L$ is the cost–loss ratio. If the compliance rate, q, is independent of the threshold at which warnings are issued, p_w, then the minimum of J occurs when $p_w = c$; this is the standard cost–loss scenario.

A more interesting situation arises if we assume that the compliance rate is a function of the false alarm rate. The definition of the false alarm rate we are using is

$$\mathrm{FAR} = \frac{\text{number of times the warning is issued but event does not occur}}{\text{number of times the warning is issued}} \quad (7)$$

$$\mathrm{FAR} = \frac{\int_{p_w}^1 (1-p)\rho(p)\,dp}{\int_{p_w}^1 \rho(p)\,dp}. \quad (8)$$

Assuming that the forecasts are reliable, and that $\rho(p) = 1$, Eq. (8) gives

$$\mathrm{FAR} = \frac{1 - p_w}{2}. \tag{9}$$

Note that the false alarm rate is *not*, in general, equal to $(1 - p_w)$. The expected false alarm rate is a monotonically nonincreasing function of the warning threshold, p_w, so we can write $q = q(p_w)$. As p_w increases the false alarm rate decreases, and we can expect the compliance rate to rise. Thus, a reasonable form of $q(p_w)$ is a monotonically increasing function of p_w. The simplest function of this form is $q = p_w$. Assuming that relationship and differentiating J with respect to p_w gives

$$\frac{dJ}{dp_w} = c - 2p_w c - \frac{1}{2} + \frac{3}{2}p_w^2. \tag{10}$$

Thus, the value of the warning threshold, p_w, that minimizes the cost per unit loss is

$$p_w^* = \frac{4}{3}c \pm \sqrt{\frac{4c^2 - 6c + 3}{9}}. \tag{11}$$

If the cost of acting is small relative to the potential loss if no action is taken then $c \ll 1$ and Eq. (11) gives

$$p_w^* \approx \frac{1}{\sqrt{3}} \approx 58\%. \tag{12}$$

This threshold is close to the 60% threshold that the U.K. Met Office's National Severe Weather Warning Service (NSWWS) uses for its ''early warnings'' (Mylne and Legg 2002). Given the low cost–loss ratios often associated with severe weather, this threshold would appear too high with respect to the standard cost–loss analysis. Imperfect compliance on the part of the public provides a possible justification for the high value of the operational threshold.

3. Imperfect compliance when warnings are rare

While analytically tractable, the assumption of a uniform distribution of forecasted probabilities is rarely appropriate when considering extreme and rare events. Murphy and Wilks (1998) used a beta distribution to model the frequency of forecasted probabilities in a precipitation forecasting system. This type of distribution can represent the bimodality that an ideal forecasting system would have, with most forecasts being close to 0% or 100%. In this study an exponential distribution of forecasted probabilities is used. The frequency of forecast probabilities generated by a severe weather forecasting system based on the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble forecasts (Mylne and Legg 2002) generally declines with increasing probability—a pattern that can be represented by an exponential distribution described by a single parameter. The analysis presented could be applied to other forms of $\rho(p)$, including estimates of $\rho(p)$ made using actual forecasts. The exponential distribution is given by

$$\rho(p) = \frac{\alpha e^{-\alpha p}}{1 - e^{-\alpha}} \qquad \alpha > 0, \tag{13}$$

where α parameterizes the rarity of higher forecast probabilities. The normalization of $\rho(p)$ is such that $\int_0^1 \rho(p)\,dp = 1$. When the distribution of forecast probabilities is given by Eq. (13) then the frequency with which the forecast probability is equal to or exceeds r is

$$\mathrm{Prob}(p \geq r) = \frac{e^{-\alpha r} - e^{-\alpha}}{1 - e^{-\alpha}}. \tag{14}$$

As expected, as α increases the frequency of probability forecasts exceeding r decreases. Figure 1 shows the form of $\rho(p)$ for three different values of α. Values of α close to zero correspond to an almost uniform distribution on the interval [0, 1]. If the forecasted probabilities are reliable, then the climatological probability of the event is the expected value of p; that is,

$$\int_0^1 \rho(p)p\,dp = \frac{1}{\alpha} - \frac{e^{-\alpha}}{1 - e^{-\alpha}}. \tag{15}$$

Given that the warning threshold is p_w, and the compliance rate is q, the time-averaged expected cost is given by substitution of Eq. (13) into Eq. (5). Integration gives

$$\begin{aligned} J &= \frac{\langle E[\mathrm{COST}]\rangle}{L} \\ &= \frac{1}{\alpha} - \frac{\beta}{\gamma} + qc\frac{\beta^{p_w} - \beta}{\gamma} + q\frac{\beta}{\gamma}\left(\frac{1}{\alpha} + 1\right) \\ &\quad - q\frac{\beta^{p_w}}{\gamma}\left(\frac{1}{\alpha} + p_w\right), \end{aligned} \tag{16}$$

where $\beta = e^{-\alpha}$ and $\gamma = 1 - e^{-\alpha}$. We seek the value of p_w that minimizes J. If the compliance rate, q, is independent of p_w then setting the derivative of dJ/dp_w equal to zero gives the standard cost–loss scenario result, $p_w^* = c$. We can now quantify how user intolerance to false alarms impacts the forecast value.

In the remainder of this section we perform the same analysis as for the case of a uniform distribution of forecast probabilities, but, in addition to using an exponential distribution of forecast probabilities, we introduce a more general dependence of the compliance rate on the warning threshold. We do this to show how the value of unequivocal forecasts varies with the frequency of the event, the cost–loss ratio of the user, and the intolerance of the user to false alarms.

We parameterize the intolerance of the user to false alarms with the following model:

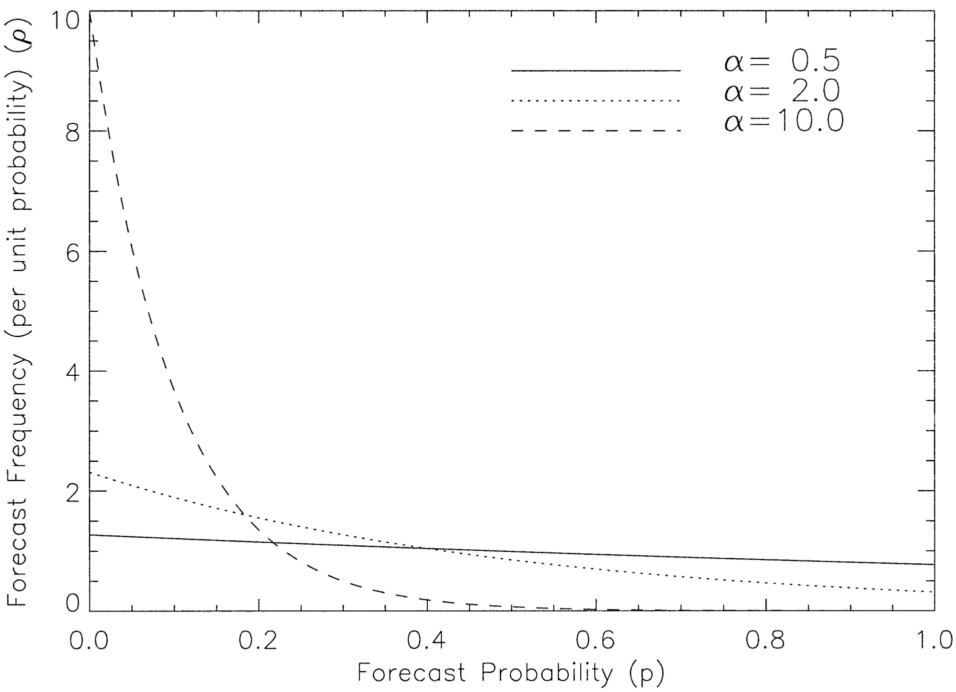

FIG. 1. The frequency with which probabilistic forecasts of an event with a particular probability are issued, as a function of the parameter α. Low values of α correspond to events for which high-probability forecasts are relatively common. High values of α correspond to events for which high-probability forecasts are rare.

$$q = p_w^\lambda \qquad \lambda \geq 0. \tag{17}$$

The parameter λ is a quantification of the strength of the ''cry wolf effect.'' It will also be assumed that the probability forecasts are reliable. If $\lambda = 0$ then the compliance rate is 1 and the standard cost–loss scenario is recovered. If λ is small then the compliance rate remains quite high until the warning threshold, p_w, is set very low (the case of frequent warnings). In the case of $p_w = 0$, warnings occur all the time, and the compliance rate falls to zero.

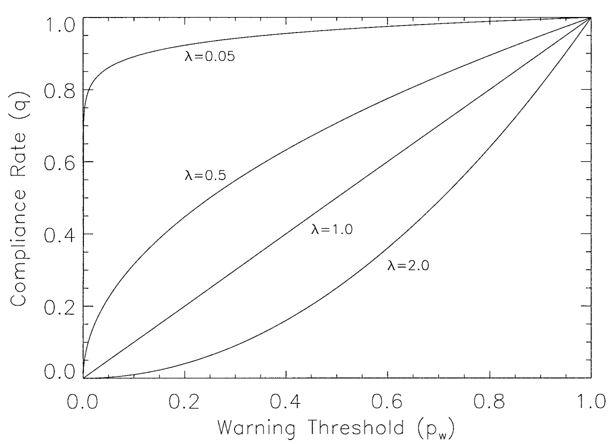

FIG. 2. The compliance rate, q, as a function of the probability threshold at which warnings are issued: four different values of the parameter λ are shown. The model of compliance, q, as a function of warning threshold, p_w, is $q = p_w^\lambda$. Thus, a value of $\lambda = 0$ corresponds to full compliance ($q = 1$) regardless of the warning threshold.

In this case, there is a major difference between ideal rational users of the traditional cost–loss scenario who will always take protective action if $c = 0$ and the intolerant users of the modified scenario who will never take action. This unwillingless to protect against improbable, but potentially catastrophic, events has been studied in the context of attitudes to insurance (Slovic et al. 1977). If p_w is set to one, warnings are only issued when there is certainty, and total compliance is achieved. The extreme cases of $p_w = 0$, 1 are the same for all values of $\lambda > 0$, but higher values of λ correspond to less tolerance of false alarms. The higher the value of λ the more rapidly the compliance rate falls as the warning threshold is reduced from near certainty. Figure 2 shows the compliance rate as a function of the warning threshold for four different values of λ.

In this more general case, finding the global minimum value of the function $J(p_w)$ is not analytically tractable. The minimum value of J, and the corresponding optimum value of p_w that gives this minimum value, can be found numerically. For cost–loss ratios of $c = 0.01$, $c = 0.1$, and $c = 0.5$ the results are shown in Figs. 3, 4, and 5, respectively. Figure 3a shows the optimum value of p_w as a function of α and λ. Figure 3b gives the percent reduction in the average expected cost when this optimum is used, rather than a value of $p_w = c = 0.01$, which is what the conventional cost–loss analysis would suggest. The maximum cost reduction is over 50% and is obtained for $\alpha \ll 1$ and $\lambda \approx 0.2$. This corresponds to an almost uniform frequency distribution of forecast probabilities

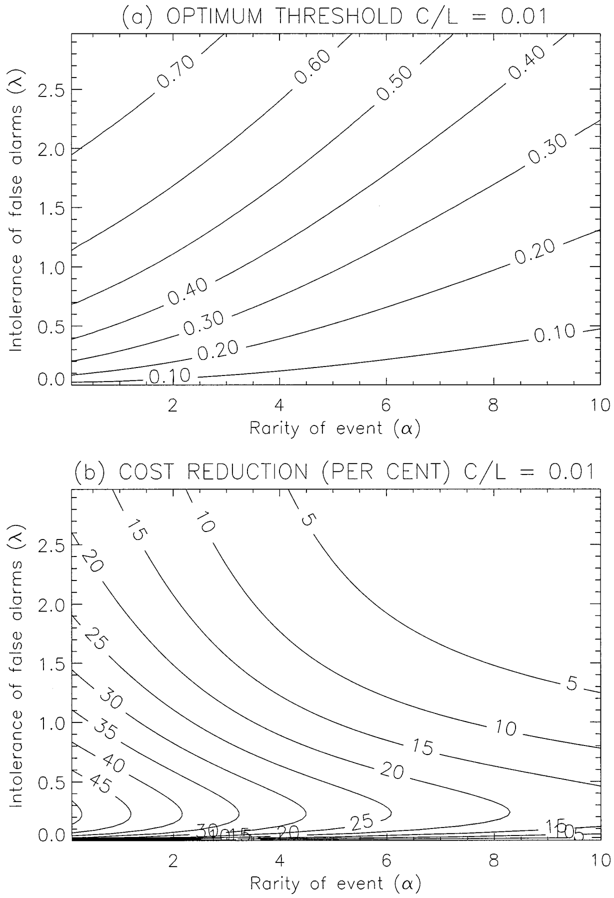

FIG. 3. (a) The optimum warning probability threshold as a function of event rarity parameterized by α and intolerance of false alarms parameterized by λ when the cost–loss ratio is 0.01. (b) The reduction (%) in the expected cost when the optimum threshold is used rather than the cost–loss ratio.

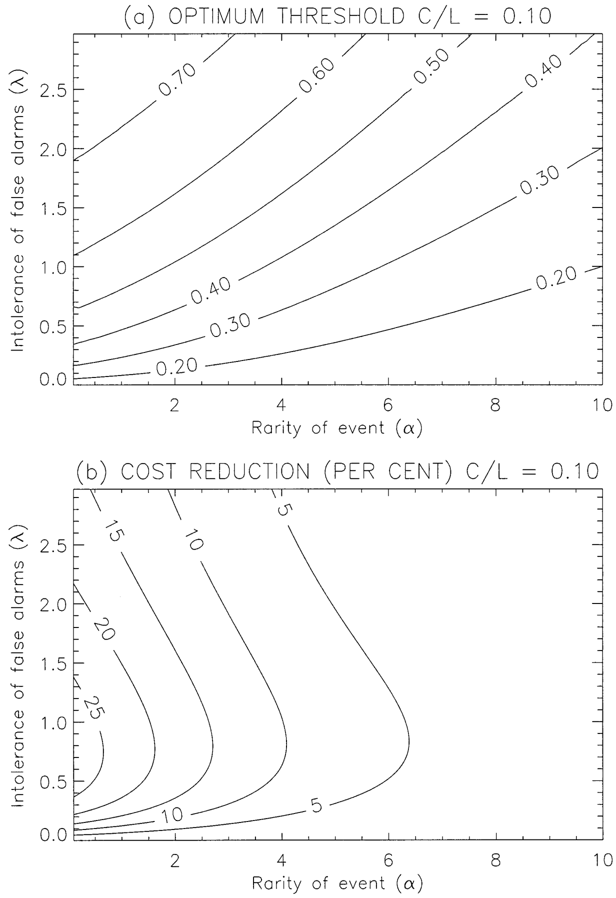

FIG. 4. As in Fig. 3, except for a cost–loss ratio of 0.1. Note that with this higher cost–loss ratio the maximum cost reduction is reduced to about 30%.

and a relatively, but not perfectly, compliant user. The optimal warning threshold in this case is $p_w \approx 0.30$, which is considerably higher than the cost–loss ratio of 0.01. Inspection of Fig. 3a suggests that the value of 60% used by the NSWWS suggests a value of $\lambda > 1$, and since the effective value of α for extreme events is likely to be at the high end of the range shown, the implied value of λ may be more than 3. Such a value would correspond to a user that is quite intolerant of false alarms. Reference to Fig. 3b indicates that in this case the reduction in the average expected cost obtained by inflating the warning threshold is quite low, less than 5%. Figure 4 shows the same analysis as Fig. 3, except for the case of a cost–loss ratio of $c = 0.1$. Again, the 60% contour corresponds to relatively high values of λ, and again the cost reductions obtained for high values of α and λ are below 5%. In this region of the α–λ plane, the high value of λ means that the user tends to only respond to almost certain forecasts, but the high value of α means that such forecasts are highly uncommon; hence, the overall impact of accounting for imperfect compliance is small. Figure 5b indicates that the cost reductions possible by accounting for imperfect compliance when the cost–loss ratio is 0.5 are not very large for any combination of α and λ. In Figs. 3 and 4 it can be seen that the biggest reductions in expected cost, for modest values of α, are obtained for values of $0 < \lambda < 1$. In this range, the user deviates from the perfect compliance enough so that taking this deviation into account matters, yet they do not deviate so much that it is infeasible to compensate for their intolerance of false alarms.

The introduction of imperfect compliance increases the average cost incurred relative to the standard cost–loss scenario. The value of the forecast is the difference between the average cost incurred when the forecasts are used compared to the average cost incurred in the absence of forecasts. The ratio of the values of the forecasts in the imperfect compliance case, with an optimal warning threshold, and in the standard perfect compliance case is given by the following expression:

$$\frac{\langle \mathrm{COST}(\text{zero compliance})\rangle - \langle \mathrm{COST}(\text{imperfect compliance, optimal } p_w)\rangle}{\langle \mathrm{COST}(\text{zero compliance})\rangle - \langle \mathrm{COST}(\text{perfect compliance, } p_w = c)\rangle}. \quad (18)$$

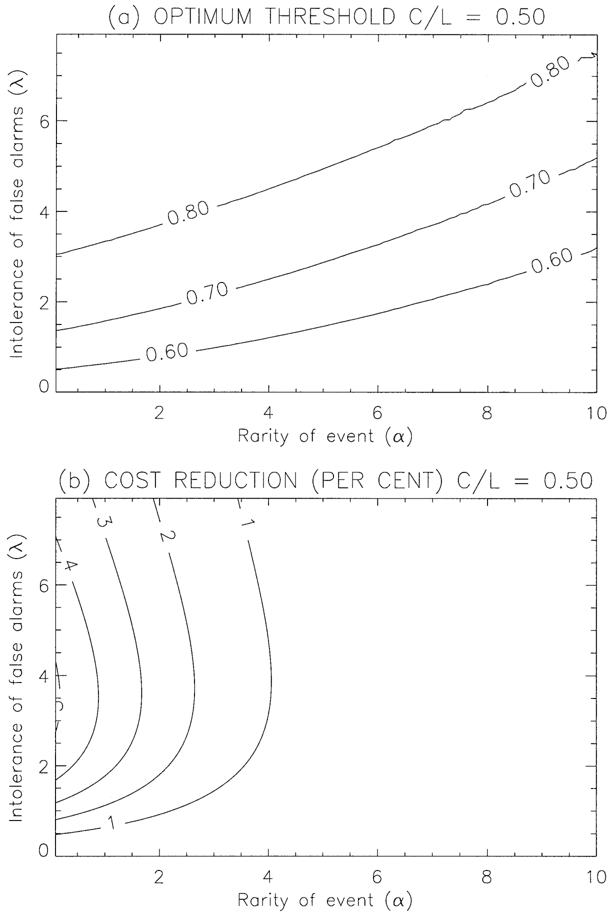

FIG. 5. As in Fig. 3, except for a cost–loss ratio of 0.5. Note that with this cost–loss ratio there is very little cost reduction and the optimum warning thresholds do not deviate from the cost–loss ratio as much as they did in Figs. 3 and 4.

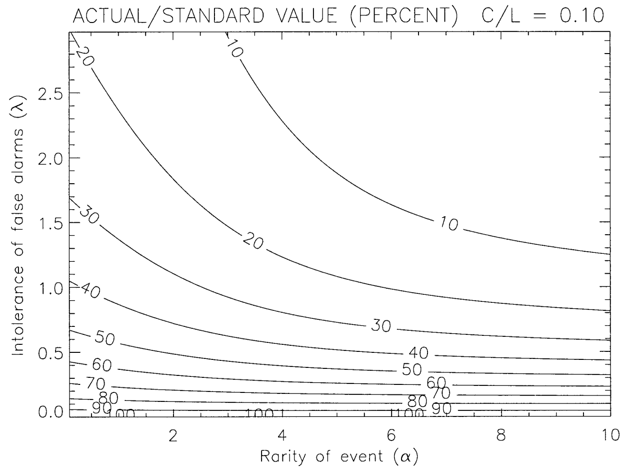

FIG. 6. The ratio (%) of the value (average cost reduction) of the forecasts when there is imperfect compliance to the value when there is perfect compliance. A value of 100% indicates that incompliance has no impact on forecast value, while a value of 10% indicates that the actual value of the forecast is only 10% of the value that would be calculated using the conventional cost–loss method. The optimum warning threshold for the parameter values was chosen for the imperfect compliance, and a threshold of $p_w = C/L = 0.1$ was used for the standard cost–loss comparison.

Figure 6 shows the value of the ratio in Eq. (18) for different values of λ and α. The cost–loss ratio used was $c = 0.1$. From Fig. 6 it can be seen that as the rarity of events increases, and the user's intolerance for false alarms increases, the actual potential value of the forecasts falls in relation to the value that the standard cost–loss analysis predicts. Replacing COST (zero compliance) in Eq. (18) with the average cost when decisions are based on climatological probabilities, COST (climatology) will reduce the value of the forecasts and reduce the ratio defined by Eq. (18). That is, if we assume that users, in the absence of a forecast, make better decisions than zero compliance, then the value of forecasts with imperfect compliance will be *further* reduced relative to the forecasts with perfect compliance.

4. Discussion and implications

We have shown that introducing a compliance rate that is a function of the false alarm rate in the cost–loss model can have a substantial impact on the optimal choice of warning threshold and the value of forecasts. The extent to which false alarm intolerance modifies the results of the cost–loss analysis depends upon the frequency of forecasted probabilities (α in our model), the cost–loss ratio (c), and the intolerance of the users to false alarms (λ in our model). The modification is most pronounced for low cost–loss ratios ($c \ll 0.1$), relatively high frequency events ($\alpha < 1$), and users who are moderately intolerant of false alarms ($0 < \lambda < 1$). For such situations the optimal warning threshold can be many times the cost–loss ratio, and the savings obtained by changing the warning threshold can exceed 25%.

Establishing the value of α, or a more appropriate form of $\rho(p)$, for a particular type of forecast is a problem that lies comfortably inside the domain of meteorologists. Past forecasts can be used to estimate $\rho(p)$. Determination of the cost–loss ratio, c, is more a problem of economics than meteorology. Nevertheless, the cost–loss type of decision model is now well established in the field of weather forecasting (Katz and Murphy 1997; Richardson 2000). The value of the parameter, λ, indeed the functional relationship $q(p_w)$, is poorly understood. At its simplest, q can be interpreted as the probability that a random individual will comply with forecast warning. The compliance rate can also be interpreted as the fraction of individuals who will comply with a warning. These two interpretations are equivalent if each individual's decision is independent of the decisions made by other individuals, but in practice this is unlikely to be the case. For example, an individual's decision on whether to evacuate their home may depend on what his or her neighbors are doing. This means that the form of $q(p_w)$ will be an emergent property of a system of interacting individual choices. In addition, q may depend on the false alarm rate during a finite period in the recent past. In such a situation the optimum warning threshold will be time dependent.

The simplified problem addressed in this paper illustrates the importance of including the actual user response in models of forecast value. This descriptive component cannot be neglected if estimates of the true value of forecasting systems to the economy and society are required. The results in this paper also indicate that consideration of users' response by forecasters can increase the realized value of their forecasts. Many forecasters will feel uncomfortable about modifying their forecasts in an attempt to compensate for imperfections in user responses. If forecasters make only *probabilistic* forecasts then they need not, and should not, report forecasts that deviate from their best estimate of the probabilities of future weather outcomes. Whenever forecasters issue nonprobabilistic forecasts, however, they are making implicit assumptions about both the cost–loss ratios of users and their tolerance of false alarms. Unless forecasters are prepared to make these assumptions explicit, and to justify them, then they should exclusively issue probabilistic forecasts and thus untangle themselves from the process of decision making.

Acknowledgments. The authors would like to thank C. Bishop, H. Brooks, and K. Mylne, discussions with whom inspired this paper, and an anonymous reviewer whose suggestions improved it. This work was supported by ONR DRI Grant N00014-99-1-0056.

REFERENCES

Ångström, A. K., 1922: On the effectivity of weather warnings. *Nord. Stat. Tidskr.,* **1,** 394–408.

Katz, R. W., and A. H. Murphy, 1997: Forecast value: Prototype decision-making models. *Economic Value of Weather and Climate Forecasts,* R. W. Katz and A. H. Murphy, Eds., Cambridge University Press, 183–217.

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

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.

——, and D. S. Wilks, 1998: A case study of the use of statistical models in forecast verification: Precipitation probability forecasts. *Wea. Forecasting,* **13,** 795–810.

Mylne, K. R., and T. P. Legg, 2002: Early warnings of severe weather from the ECMWF ensemble prediction system. Preprints, *16th Conf. on Probability and Statistics in the Atmospheric Sciences,* Orlando, FL, Amer. Meteor. Soc., J254.

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

Slovic, P., B. Fischhoff, S. Lichtenstein, B. Corrigan, and B. Combs, 1977: Preference for insuring against probable small losses: Insurance implications. *J. Risk Insur.,* **XLIV,** 237–257.

Stewart, T. R., 1997: Forecast value: Descriptive decision studies. *Economic Value of Weather and Climate Forecasts,* R. W. Katz, and A. H. Murphy, Eds., Cambridge University Press, 147–181.