Probabilistic Weather Forecasts for Aviation Operations

Could it Work?

At this workshop, Ian Mason demonstrates that the Expected Cost (EC) of the errors of any weather forecast, i.e. with respect to a perfect forecast, is given by:

$EC = (1-p_c).f.(\text{False Alarm Cost}) + p_c.(1-h). (\text{Miss Cost})$,

Where p_c = the climatological base rate of the event,
f = false alarm rate,
h = hit rate.

False Alarm Cost (FAC) is the cost of incorrectly forecasting an event, i.e. the cost of taking protective action unnecessarily. Miss Cost (MC) is the cost of not forecasting an event *over and above any cost incurred by a hit or correct forecast of an event.*

Ian further demonstrates, in a signal detection framework, that the optimal decision threshold that minimises EC is given by:

$p(opt) = R / (1 + R)$,

where R = False Alarm Cost / Miss Cost.

Description and Results of Australian Experiment

For two years meteorologists at three offices, Townsville, Sydney and Melbourne input probability estimates that the weather at their local airport would be below the alternate minimum. This is the level of visibility, cloud base or weather below which regulations require aircraft operating into that aerodrome must carry extra fuel to proceed to a suitable alternate aerodrome. These probability estimates were done for lead times of 1, 3, 6, 12 and 18 hours, four times daily at the same time as the routine issue of the terminal aerodrome forecast (TAF). The forecast probability were at intervals of 10%. Note that insufficient events have been captured at Sydney for complete analysis, because the climatological rate of occurrence of below minimum weather there is very low.

The aim of the experiment was twofold. Firstly it was necessary to test the applicability of the normally distributed weight of evidence, as assumed in signal detection theory, to forecaster’s thought processes when formulating TAFs. Once this was confirmed, the effect of variations in decision thresholds on Expected Cost can be modelled. The difference in Expected Cost between actual and optimal decision thresholds can then be determined.

Fig. 1 shows (h, f) plots for individual forecasters at each office. These were derived from the Australian Bureau of Meteorology's automated TAF verification scheme, and are a composite of values for the first 6 hours of the TAF. The sole purpose of this figure is to demonstrate that there are large differences in attitude to risk between individuals. Put another way, decision thresholds vary significantly from one forecaster to another. Forecasters A, C and E exhibit a less cautious attitude (higher decision threshold) than B, D and F respectively. Differences in skill between A and B, C and D, and E and F as measured by the signal detection parameter d' are quite small.

Fig. 2 is a plot of the normal deviates of h and f for each lead time for Townsville and Melbourne. The linearity of these plots is strongly suggestive of approximate normality of the underlying probability distribution.

Figs. 3a, 3b and 3c show reliability diagrams for each of the three offices. Melbourne exhibits mild over-forecasting and Townsville a little more over-forecasting. The reliability at Sydney, where events are less frequent than the others, is quite poor.

Cost Analysis

QANTAS Airways supplied, inter alia, costs for the operations of two flights. For a long flight from Singapore to Melbourne, duration about 7 hours, the FAC is $1,390 and the MC is $10.535. This produces an R value of 0.132 and p(opt) = 0.12. From *h* and *f* values measured in the experiment at Melbourne at 6 hour lead time, the average decision threshold is 0.02. This extreme degree of conservatism is caused by the forecasters' perception of the consequences of a missed event.

Fig. 4a is a graph of EC vs decision threshold for that flight, using measured d' and p_c of 0.02 which is the climatological base rate at Melbourne. Fig. 4b is a graph of EC vs d' for the same flight.

Figs. 5a and 5b are the same graphs for a shorter flight, from Brisbane to Townsville, which takes a little under 2 hours. For this flight p(opt) = 0.04, and the average decision threshold measured in the experiment at 3 hour lead time is 0.01.

Discussion

From the equation behind fig. 4a, at the measured average decision threshold of 0.02, the cost of the errors in the aerodrome forecast is $231 per flight. If the forecast was reliably made at the optimum decision threshold of 0.12, the cost would be $128. So a perfectly reliable forecast would save about 45% of the total cost of the errors. Using the reliability diagram in fig. 3b, if a forecaster in the Melbourne office was asked to use 0.12 as his or her decision threshold, the effective decision threshold would be about 0.07. If this is used, the cost is $135. So, for the Singapore to Melbourne flight, even the moderately reliable probability forecasts, as currently produced, would provide most of the savings (41%) gained by the perfectly reliable forecasts (45%).

Fig 4b shows how skill affects cost. The two plots are for the optimum decision threshold of 0.12 and for the measured average decision threshold of 0.02. The most significant feature to note is that if the forecasts are done at the optimum decision threshold, a decrease in skill does not matter all that much. So it could be suggested that the issue of financial risk management can be more important than raw forecaster skill.

The same graphs in fig. 5a and 5b for the shorter flight show similar results. The dollar values are of course smaller, and the actual decision threshold of 0.01 is not quite as far away from the optimum decision threshold of 0.04 as it is for the Singapore to Melbourne flight. So the costs in fig. 5b for the two different decision thresholds don't diverge so strongly as for fig. 4b.

The CCFP and Probabilities

The CCFP is produced as a forecast area of low (<40%)), medium (40-69%) or high (>=70%) confidence of ranges of areal coverage. The ranges of areal coverage are low (<25%), medium (50-74%) and high (75-100%). These forecasts are therefore an overall probability forecast, the end result being the product of the confidence and coverage percentages. For example, if a forecaster outputs a 60% chance of 50% coverage, he is saying that he thinks there will be a 30% chance of convection at any one point in the area, which is the same as an overall 30% coverage.

If one assumes that this forecast is reliable, i.e. it matches the observed frequency at all points over a reasonably long period, then whether or not to take protective action based on this forecasts depends on the economic and safety consequences of the decision. Using Ian Mason's formula for p(opt) = R / (1 + R), one can make an estimate about how definite we must be with the forecast probability before it is economically warranted to take protective action, i.e. close or reduce traffic on routes. R = FAC / MC. The FAC for the CCFP is actually the same as a Hit, or correct forecast. The forecast causes the effect on airspace regardless of whether it happens or not. MC is defined as the cost of a missed event MINUS the cost incurred by a hit or correct forecast. What we need to know is -: what is the extra cost of having to make changes to airspace and traffic in a reactionary, unplanned way, having not forecast the event, over and above the cost of doing it in a planned way a few hours earlier from the CCFP. For the sake of this exercise, let's say the extra is 0.5 times the cost of a 'hit'. So the total cost of the unforecast event is 1.5 times the hit cost.

So we arrive at R = FAC / MC = Hit Cost / (1.5 * Hit Cost – Hit Cost) = 2 and p(opt) = 2 / (1 + 2) = 0.67.

This means that it is only commercially warranted to take protective action if forecast confidence of the convection is > 67%.