EASTERN REGION TECHNICAL ATTACHMENT
NO. 2004-03
APRIL 2004

ON THE BEHAVIOR OF THE CRITICAL SUCCESS INDEX

Harry Gerapetritis
Joseph M. Pelissier
NOAA/National Weather Service
Greer, South Carolina

1. INTRODUCTION

The "Winter Weather" section of the "Best Practices" homepage (hereafter referred to as WWBP) on the National Weather Service (NWS) Eastern Region Intranet Site has spurred discussion over ways to improve the Critical Success Index (CSI) for winter storm warnings. The Strategic Problem section (II) of the page concludes the following with regard to maximization of the CSI:

An analysis of historical verification statistics (attachment 3) indicated that the False Alarm Ratio (FAR) has the greatest impact on CSI scores. Thus it appears that any training, decision making or verification efforts focused on reducing FARs will have the largest positive impact on improving CSIs.

We will show that while this is generally true for Eastern Region as a whole, it may not be the correct approach for every WFO to follow. Furthermore, we will point out the many difficulties inherent in translating the understanding of CSI behavior into a workable warning strategy. It is important to note that this analysis will focus solely on the statistical implications of the warning decision-making process. Customer service, which should occupy a central role in any WFO warning strategy, is not taken into account for the purposes of this manuscript.

2. ANALYSIS

WWBP attachment 3 (not reproduced here) consists of 77 seasonal Winter Storm Warning scores taken from the legacy Eastern Region Weather Service Forecast Offices (WSFO's) during the 1994-95 to 2000-01 period. The study comprises a statistical analysis containing plots of CSI versus POD (Probability of Detection) and CSI versus FAR (False Alarm Ratio). A linear best-fit is applied to each scatter-plot, and a multiple R-squared statistic is used to quantify the linear variance in each case. The results of this statistical analysis led to the WWBP conclusion regarding the impact of FAR scores on the maximization of CSI.
However, it is not necessary to employ statistical analysis to determine whether FAR or POD is more strongly correlated to CSI. We know CSI to be explicitly defined as a function of FAR and POD (Schaefer 1990) by:

$$\mathbf{CSI} = \frac{\mathbf{1}}{\mathbf{1/(1\text{-}FAR) + (1/POD) - 1}} \qquad \mathbf{(1)}$$

This relation is shown graphically by a contour plot of the CSI surface in Fig. 1, which also contains a plot of the data from WWBP. The two-dimensional relationships between CSI and POD for fixed values of FAR, and between CSI and FAR for fixed values of POD are shown graphically in Figs. 2 and 3 respectively. It can be seen that CSI is a non-linear function of both FAR and POD individually.

Further insight into the behavior of CSI may be gained by examination of the slope, or gradient, of the relationship over the domain of interest. Each (FAR, POD) data pair occupies a unique position on the CSI surface (Fig. 1). The magnitude of the component of CSI's slope with respect to FAR and POD determines the rate at which CSI varies with respect to each quantity. The value of this variation with respect to each component is given analytically by partial derivatives of CSI with respect to FAR and POD.
From equation (1) it can be shown that:

$$\frac{\partial \mathbf{CSI}}{\partial \mathbf{FAR}} = \frac{\mathbf{-1}}{\mathbf{(1\text{-}FAR)^2[1/(1\text{-}FAR) + (1/POD) - 1]^2}} \qquad \mathbf{(2)}$$

Likewise:

$$\frac{\partial \mathbf{CSI}}{\partial \mathbf{POD}} = \frac{\mathbf{1}}{\mathbf{(POD)^2[1/(1\text{-}FAR) + (1/POD) - 1)]^2}} \qquad \mathbf{(3)}$$

So, for any combination of FAR and POD, the variation of CSI with respect to FAR and POD may be determined by substituting values in (2) and (3), respectively. Then, the relative effects of changes in POD and FAR on CSI can be determined by comparing the magnitudes of these slopes. FAR and POD will affect CSI equally when the magnitudes of their respective slopes are equal, that is:

$$\frac{|\partial \mathbf{CSI}|}{|\partial \mathbf{FAR}|} = \frac{|\partial \mathbf{CSI}|}{|\partial \mathbf{POD}|}$$

or:

$$\left| \frac{\mathbf{-1}}{\mathbf{(1\text{-}FAR)^2[1/(1\text{-}FAR) + (1/POD) - 1]^2}} \right| = \left| \frac{\mathbf{1}}{\mathbf{(POD)^2[1/(1\text{-}FAR) + (1/POD) - 1]^2}} \right|$$

which yields

$$\mathbf{POD = 1 - FAR} \qquad \mathbf{(4)}$$

So, when POD is greater (less) than 1-FAR, changes in FAR have a greater (lesser) influence on CSI than changes in POD. It can be seen in Fig. 1 that POD > 1-FAR for about 85 percent of the cases in the WWBP sample. For this subset, the recommended strategy of reducing FAR to optimize CSI would follow, but for the remaining 15 percent one could argue that effort should be primarily directed toward improving POD in order to improve CSI. Thus, by simply comparing the POD to 1-FAR, any forecast office can quickly determine which component's improvement would produce a larger improvement in CSI. However, the question would still remain as to how this knowledge would translate into a warning strategy.

3. DISCUSSION

The best warning strategy is to always be right. That is to say, always choose to warn when the event will occur, and always choose not to warn when the event will not occur. This perfect forecaster will thus have a POD of 1, an FAR of 0, and a perfect CSI of 1. Of course a forecaster can never know with certainty beforehand whether an event will occur or not. Therefore, all forecasters implicitly affix a confidence level to the chance that an event will occur. The accuracy with which forecasters assign this probability may be termed "reliability." The success of the forecast then, as measured by CSI, is the result of two factors: (1) the forecaster's skill in assessing this confidence level, and (2) the strategic selection of a *warning threshold* such that CSI has the greatest likelihood of being maximized. The second factor is truly the one that is at the heart of a warning strategy. Once a confidence level has been determined by the forecaster, what threshold should trigger the issuance of a warning? The answer depends on whether it is desirable to improve the POD by warning aggressively or to improve the FAR by warning conservatively.

NWS Eastern Region Supplement 02-2003 to NWSI 10-513OML E-7-01 sets the required confidence level for issuing a Winter Storm Warning at 80 percent. This value appears to be motivated by the goal of reducing the FAR. For forecasters with perfect reliability, however, this threshold would likely be too high. A forecaster might assess that an event's chances of occurrence are between 50 and 80 percent, but be prevented from issuing a warning even though the event is more likely to happen than not. Such a strategy could be detrimental to the CSI, since the forecaster could be penalized with a missed event while correctly believing the event will occur.

The historical "over-warning" noted in WWBP shows that forecaster reliability has not been perfect. The high warning threshold, then, is apparently an attempt to reduce the Regional bias toward over-warning. There are several problems with this approach. First, a minority of the offices have under-warned over the course of the study period. Strict application of the policy could end up hurting the CSI at such offices. The second concern is that past warning biases may not prove to be good predictors of future warning biases. The expected result of the continued infusion of science and technology into the forecast and warning process is an improvement in forecaster reliability. It will be difficult to assess true reliability improvements if the warning threshold is set artificially high to compensate for past performance. Third, adopting an 80 percent confidence threshold for warnings is tantamount to saying that POD will be sacrificed in order to improve FAR. This is an important consideration since POD strongly affects lead-time. Every time an event is missed, a zero-hour lead-time is assigned to the event, and the average lead-time suffers correspondingly.

4. SUMMARY

WWBP states that "...reducing FAR's will have the largest positive impact on improving CSI's." This statement was shown to be generally true for the Region as a whole over the period of interest since the sample set examined was one in which POD was generally greater than 1 - FAR. This condition delineates that portion of the domain of the CSI function in which changes in FAR affect CSI more strongly than equivalent changes in POD (upper right portion of Fig. 1). However, for about 15 percent of the WWBP sample, POD was less than 1 – FAR depicted in the lower left portion of Fig. 1. For those offices, increases in POD would have the greater influence on improving their CSI.

It is suggested here that the CSI achieved is the result of two separate factors in the warning decision making process: (1) the forecaster's skill at assessing the confidence level that a given event will occur, and (2) the strategic selection of a warning threshold such that CSI has the greatest likelihood of being maximized. Whenever the FAR and POD are both relatively high, it is safe to assume that over-warning has occurred.[1] Over-warning may result when either (1) a forecaster routinely assigns too high a level of confidence to the chance of the event occurring, or (2) the confidence threshold for issuing a warning is set too low. Whenever FAR and POD are relatively low, under-warning has typically occurred. Under-warning may result when (1) the forecaster routinely assigns too low a confidence level for the event, or (2) the threshold for warning is set too high.

A successful warning program must also clearly take into account the needs of the users. Issuing warnings with greater accuracy would lead to both increased customer satisfaction and improved CSI scores, but it is not at all clear that a "false alarm" and a "hit" affect our customers in the same proportion that they affect the value of CSI. An ideal verification system would be one in which the only way verification scores can be improved is by issuing warnings with greater "utility", i.e. the warnings do a better job of fulfilling their basic purpose, the protection of lives and property. Although CSI may be the best available surrogate to measure the utility of our warnings, it should be acknowledged that optimizing CSI may not necessarily lead to more effective warnings.

ACKNOWLEDGMENTS

The authors would like to thank Steve Burrus for his useful discussion of these topics, and Larry Lee and Heather Hauser for their helpful reviews of this paper.

REFERENCES

Schaefer, Joseph T., 1990: The Critical Success Index as an Indicator of Warning Skill. *Weather and Forecasting*, **5**, 570-575.

National Weather Service Eastern Region, cited 2004: Supplement 20-2003 Applicable to NWSI 10-513, October 31, 2003. [Available online at: http://www.nws.noaa.gov/directives/010/pd01005013e022003a.pdf]

Watling et. al., cited 2004: Eastern Region Winter Weather Best Practices Team Report (2001). 13 pp.[Available online at https://bestpractices.nws.noaa.gov/contents/erh/winter/tg_II.htm]

---

[1] Over-warning is addressed only from a statistical sense. Customer service is not taken into account.

Figures

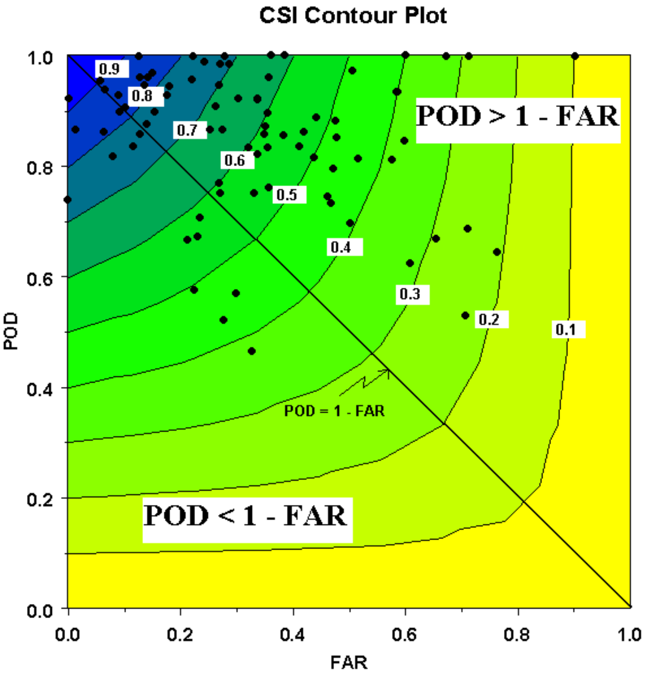

Figure 1. CSI contours as a function of FAR and POD. The data points are from WWBP with the POD = 1 - FAR line drawn in.

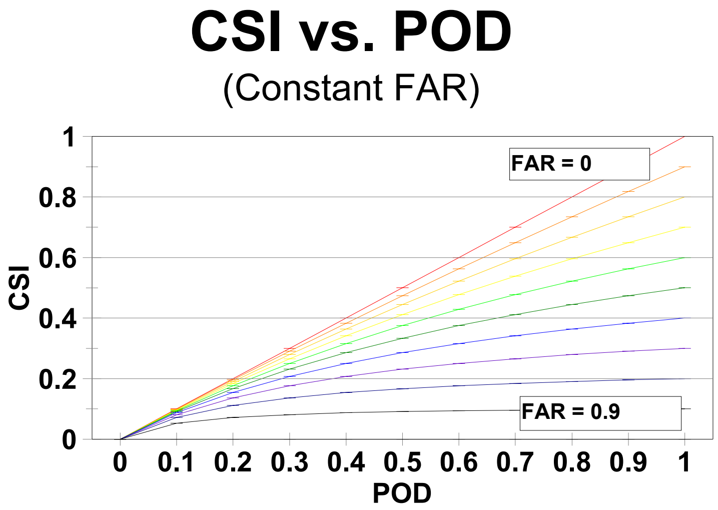

Figure 2. CSI as a function of POD for fixed values of FAR

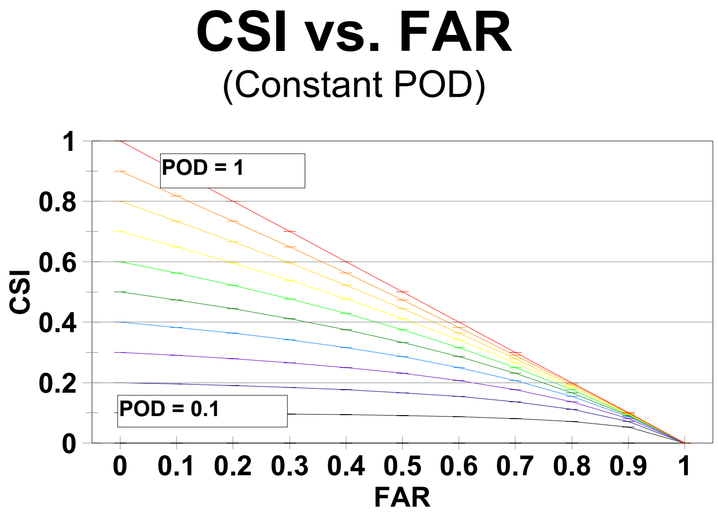

Figure 3. CSI as a function of FAR for fixed values of POD