NOTES AND CORRESPONDENCE

Applying a General Analytic Method for Assessing Bias Sensitivity to Bias-Adjusted Threat and Equitable Threat Scores

KEITH F. BRILL

National Centers for Environmental Prediction, Hydrometeorological Prediction Center, Camp Springs, Maryland

FEDOR MESINGER

National Centers for Environmental Prediction, Environmental Modeling Center, Camp Springs, and Earth System Science Interdisciplinary Center, University of Maryland, College Park, College Park, Maryland

(Manuscript received 6 February 2009, in final form 14 May 2009)

ABSTRACT

Bias-adjusted threat and equitable threat scores were designed to account for the effects of placement errors in assessing the performance of under- or overbiased forecasts. These bias-adjusted performance measures exhibit bias sensitivity. The critical performance ratio (CPR) is the minimum fraction of added forecasts that are correct for a performance measure to indicate improvement if bias is increased. In the opposite case, the CPR is the maximum fraction of removed forecasts that are correct for a performance measure to indicate improvement if bias is decreased. The CPR is derived here for the bias-adjusted threat and equitable threat scores to quantify bias sensitivity relative to several other measures of performance including conventional threat and equitable threat scores. The CPR for a bias-adjusted equitable threat score may indicate the likelihood of preserving or increasing the conventional equitable threat score if forecasts are bias corrected based on past performance.

1. Introduction

This note extends the analysis of Brill (2009) to include new performance measures, bias-adjusted threat and equitable threat scores, derived by Mesinger (2008). The work of Mesinger (2008) was motivated by heuristic evidence of frequently misleading bias sensitivities for the threat score (TS) and equitable threat score (ETS), such as shown by Baldwin and Kain (2006) using a geometrical model. Specifically, this sensitivity was demonstrated to undermine the presumably desired property of the scores to indicate improvement as the bias approaches unity for a given placement accuracy. The objective of the Mesinger (2008) bias adjustment was to remove to the extent possible the bias sensitivity, aiming to arrive at scores that reflect the placement ac-

curacy, as opposed to representing performance measured in overall terms.

Brill (2009) derives a critical performance ratio (CPR) that quantifies the bias sensitivity for measures of performance of dichotomous forecasts. The CPR specifies the fraction of added forecast area or points observing the event that must be surpassed for the performance measure to indicate improvement if the bias is increased. Conversely, if the bias is decreased, the CPR value specifies the fraction of the removed forecast area or points observing the event that must not be exceeded for the performance measure to indicate improvement.

Section 2 introduces the 2×2 contingency table in the form needed for this analysis and states the formulation of the CPR following Brill (2009). Section 3 describes the bias-adjusted threat and equitable threat scores and presents the results of the CPR analysis. Section 4 discusses the implications for anticipating the performance of forecasts corrected to remove systematic error (bias). Section 5 presents a brief summary.

Corresponding author address: Keith F. Brill, NCEP/HPC, W/NP32, NOAA Science Center, Rm. 410B-2, 5200 Auth Rd., Camp Springs, MD 20746-4304.
E-mail: keith.brill@noaa.gov

TABLE 1. Contingency table in terms of H, F, O, and N.

Events	Observed	Not observed	Total
Forecast	H	$F - H$	F
Not forecast	$O - H$	$N - F - O + H$	$N - F$
Total	O	$N - O$	N

TABLE 2. Normalized contingency table in terms of B, P, and α.

Events	Observed	Not observed	Total
Forecast	αP	$\alpha(B - P)$	αB
Not forecast	$\alpha(1 - P)$	$1 - \alpha(B + 1 - P)$	$1 - \alpha B$
Total	α	$1 - \alpha$	1

2. The CPR formulation

Table 1 is the 2×2 contingency table used in verifying dichotomous forecasts, expressed in terms of the number of hits (points at which the event is both forecast and observed, H), the number of forecasts of the event (F), the number of observations of the event (O), and the total number of points examined for the occurrence or non-occurrence of the event (N). Table 2 is a transformation of Table 1 obtained by multiplying all F and H values by $(O/O) = 1$ and making the following replacements:

$$B = \frac{F}{O} \qquad (1)$$

and

$$P = \frac{H}{O}, \qquad (2)$$

where B is the frequency bias and P is the probability of detection (POD), respectively, as defined in Wilks (2006) and elsewhere. The cells are normalized by dividing by N, and (O/N) is replaced with the event frequency α. It is assumed that N is sufficiently large by combining events over space and/or time to populate the contingency table adequately.

Using Table 2, any performance measure and its CPR may be written in terms of B, P, and α as described in Brill (2009). For a particular sample of verification data, the event frequency, α, is independent of the forecasts and is treated as a constant so that performance measures and CPR expressions may be written as functions of two variables. Results thereby obtained may then be evaluated for different event frequencies.

With α as a constant, any performance measure, S, is expressed as

$$S = S(B, P). \qquad (3)$$

With reference to Brill (2009), the CPR (ρ) is defined as follows:

$$\rho = -\frac{\left(\dfrac{\partial S}{\partial B}\right)}{\left(\dfrac{\partial S}{\partial P}\right)}, \qquad (4)$$

where the partial derivatives are taken in the usual way, with the other independent variable assumed constant. The performance measures evaluated in this note increase as performance improves (P increases) and, therefore, are said to be positively oriented (Wilks 2006). This requires $(\partial S/\partial P) > 0$ as discussed in Brill (2009). With α constant, Brill (2009) shows that a performance measure S will indicate improvement for increasing bias if

$$\frac{\Delta H}{\Delta F} = \frac{\Delta P}{\Delta B} > \rho. \qquad (5)$$

If the ratio of new hits to the number or area of added forecasts is sufficiently large, S will indicate improvement, even if the bias increases beyond unity. On the other hand, for a decrease in bias, S indicates improvement if

$$\frac{\Delta H}{\Delta F} = \frac{\Delta P}{\Delta B} < \rho. \qquad (6)$$

For this case, the ratio of relinquished hits to the number or area of removed forecasts must be sufficiently small for S to indicate improvement. Thus, it is possible to "hedge" a performance measure by either increasing or decreasing the bias, although the former is most commonly done in practice.

In the next section, following the procedure outlined in Brill (2009), the performance measures considered here are expressed according to Eq. (3). The partial derivatives are derived and substituted into Eq. (4); then, the results are simplified to obtain the CPR formulas for graphical analysis.

3. Analysis of performance measures

The formulas in Table 3 for bias-adjusted threat scores and equitable threat scores are based on those derived in Mesinger (2008). Table 3 shows the results of the derivations for all of the performance measures considered for comparison in this note: threat score (TS), equitable threat score (ETS), Clayton skill score (CSS), odds ratio skill score (ORSS), and the bias-adjusted threat and equitable threat scores. Thorough descriptions of TS, ETS, CSS, and ORSS can be found in Wilks (2006) and elsewhere. For complete generality in the case of each performance measure, the partial derivatives with respect to B and P are shown in Table 3 along with the

TABLE 3. Table of performance measure (PM) formulas, partial derivatives with respect to B and P, CPRs, and CPRs for bias equal to one. Here, $u = W(u)e^{W(u)}$, $W(u)$ is the Lambert W function, such that $u = W(u)e^{W(u)}$. All of these have a positive orientation.

PM	$S(B, P)$	$\partial S/\partial B$	$\partial S/\partial P$	CPR	CPR for $B = 1$
TS	$\dfrac{P}{(B+1-P)}$	$\dfrac{-P}{(B+1-P)^2}$	$\dfrac{B+1}{(B+1-P)^2}$	$\dfrac{P}{(B+1)}$	$\dfrac{P}{2}$
ETS	$\dfrac{P-\alpha B}{(B+1-P-\alpha B)}$	$\dfrac{P(2\alpha-1)-\alpha}{(B+1-P-\alpha B)^2}$	$\dfrac{B+1-2\alpha B}{(B+1-P-\alpha B)^2}$	$\dfrac{P+\alpha-2\alpha P}{(B+1-2\alpha B)}$	$\dfrac{P+\alpha-2\alpha P}{2(1-\alpha)}$
CSS	$\dfrac{P-\alpha B}{B(1-\alpha B)}$	$\dfrac{2\alpha PB-P-\alpha^2 B^2}{B^2(1-\alpha B)^2}$	$\dfrac{1}{B(1-\alpha B)}$	$\dfrac{P+\alpha^2 B^2-2\alpha PB}{B(1-\alpha B)}$	$\dfrac{P+\alpha^2-2\alpha P}{(1-\alpha)}$
ORSS	$\dfrac{P-\alpha B}{D}$, $D = P - 2\alpha PB - 2\alpha P + 2\alpha P^2 + \alpha B$	$\dfrac{2\alpha P(P-1)(1-\alpha)}{D^2}$	$\dfrac{2\alpha Y}{D^2}$, $Y = B - P^2 - \alpha B^2 - \alpha B + 2\alpha BP$	$\dfrac{P(1-P)(1-\alpha)}{Y}$	$\dfrac{P(1-\alpha)}{(1+P-2\alpha)}$
TSA dHdF	$\dfrac{1-(1-P)^{1/B}}{1+(1-P)^{1/B}}$	$\dfrac{2(1-P)^{1/B}\ln(1-P)}{B^2[1+(1-P)^{1/B}]^2}$	$\dfrac{2(1-P)^{(1/B)-1}}{B[1+(1-P)^{1/B}]^2}$	$\dfrac{(P-1)\ln(1-P)}{B}$	$(P-1)\ln(1-P)$
ETSA dHdF	$\dfrac{1-\alpha-(1-P)^{1/B}}{1-\alpha+(1-P)^{1/B}}$	$\dfrac{2(1-\alpha)(1-P)^{1/B}\ln(1-P)}{B^2[1+(1-P)^{1/B}]^2}$	$\dfrac{2(1-\alpha)(1-P)^{(1/B)-1}}{B[1+(1-P)^{1/B}]^2}$	$\dfrac{(P-1)\ln(1-P)}{B}$	$(P-1)\ln(1-P)$
TSA dHdA	$\dfrac{1-u^{-1}W(u)}{1+u^{-1}W(u)}$, $u = \dfrac{-\ln(1-P)}{(B-P)}$	$\dfrac{2[W(u)]^2[1+W(u)]^{-1}}{\ln(1-P)[1+u^{-1}W(u)]^2}$	$-\dfrac{\partial S}{\partial B}\left[\dfrac{(1-P)u+1}{(1-P)u}\right]$	$\dfrac{(P-1)\ln(1-P)}{B-P+(P-1)\ln(1-P)}$	$\dfrac{\ln(1-P)}{\ln(1-P)-1}$
ETSA dHdA	$\dfrac{(1-\alpha)-u^{-1}W(u)}{(1-\alpha)+u^{-1}W(u)}$, $u = \dfrac{-\ln(1-P)}{(B-P)}$	$\dfrac{2(1-\alpha)[W(u)]^2[1+W(u)]^2}{\ln(1-P)[1+u^{-1}W(u)]^2}$	$-\dfrac{\partial S}{\partial B}\left[\dfrac{(1-P)u+1}{(1-P)u}\right]$	$\dfrac{(P-1)\ln(1-P)}{B-P+(P-1)\ln(1-P)}$	$\dfrac{\ln(1-P)}{\ln(1-P)-1}$

CPR formulas before and after substituting $B = 1$. Each performance measure is discussed below with reference to its row in Table 3.

Two methods of bias adjustment derived by Mesinger (2008) are designated as the dHdF method (also discussed in Baldwin and Kain 2006) and the dHdA method. Bias adjustment of the scores is intended to remove the impacts of under- or overforecasting on scores, to arrive at values that reflect the placement accuracy alone. This is done by expanding the area of underbiased forecasts or contracting the area of overbiased forecasts, thereby adding or subtracting hit area, hypothesizing that the forecast and observed areas are proximate. In so doing, an assumption is needed to recompute the hit area so that scores can be recomputed with bias adjusted to unity. The nature of this assumption defines the two different methods. These methods calculate an interpolated or extrapolated value of the hit area to the condition of bias equal to one, performing this interpolation or extrapolation along a function, $H(F)$, resulting from two different assumptions. The dHdF method assumes that as the forecast area expands, the hit area increase per unit increase in the forecast area is proportional to the observed area not hit, $O - H$. The dHdA method (Mesinger 2008) refines the approach by assuming that as the forecast area reduced by hit area, $A = F - H$, expands, the hit area increase per unit increase in the false alarmed area, A, is proportional to the observed area not hit. The dHdA method arrives at a hit area dependence on forecast area, $H(F)$, having better asymptotic properties than that resulting from the dHdF method. The resulting bias-adjusted TSs and ETSs are positively oriented performance measures for both methods of adjustment. The dHdF method is simpler than the dHdA method, which involves the Lambert W function. The Lambert W function is described in the appendix, where the first derivative with respect to its argument is expressed [Eq. (A4)].

The results from Mesinger (2008) and Eq. (A4) are used to derive the entries in Table 3 for the bias-adjusted scores. The CPR for the conventional ETS exhibits explicit dependence on event frequency. In contrast, for both the dHdF and dHdA methods, the CPR is not explicitly dependent on event frequency, because the event frequency factor appears in both the numerator and the denominator of the CPR and is eliminated by division (see Table 3). An explicit dependence on bias exists; therefore, an implicit dependence on event frequency remains, hidden in the bias itself (see appendix A of Brill 2009). The loss of explicit dependence on event frequency is consistent with the effects of transforming to $B = 1$, which renders the information supplied by the event frequency redundant and unnecessary
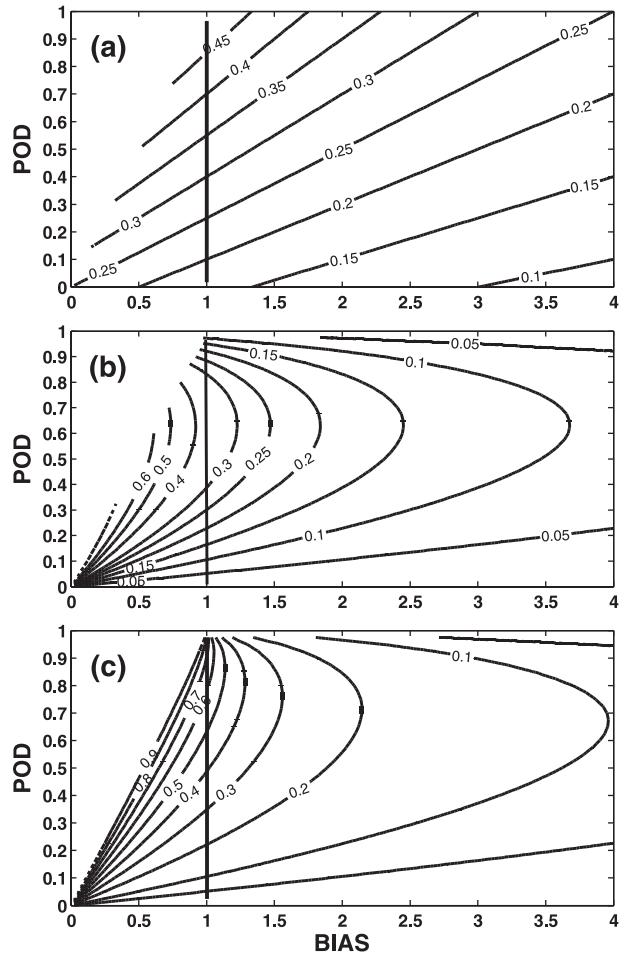

FIG. 1. CPR contours in the bias–POD plane for (a) ETS with event frequency 0.25, (b) dHdF bias-adjusted ETS, and (c) dHdA bias-adjusted ETS. Vertical line denotes a bias equal to one. The POD is impossibly greater than the bias in the blank areas.

for determining what fraction of added or removed forecasts must be hits for an improved score if the bias of the forecasts is changed. Consequently, the bias-adjusted TSs and ETSs have identical CPR expressions, but the CPR formulas are different for the two different bias adjustment methods (dHdF and dHdA). For both methods, the CPR exhibits a singularity at POD = 1.

To serve as a comparison, Fig. 1a displays the contours of the ETS CPR without bias adjustment for an event frequency of 0.25, which is appropriate to relatively common events (see Baldwin and Kain 2006). Over all of the admissible domain of the bias–POD plane, ETS CPR values reduce with decreasing event frequency, as one can verify using the ETS CPR formula in Table 3. Figure 1b displays contours of the ETS CPR for the dHdF method on the bias–POD plane. For a given value of POD, the CPR decreases with increasing bias. For

fixed values of bias greater than about 0.7, the dHdF ETS CPR increases with increasing POD until about 0.6; above POD ~0.6, the dHdF ETS CPR begins to decrease. Contours of the CPR for the dHdA method on the bias–POD plane are shown in Fig. 1c. The behavior of the CPR is similar to that seen in Fig. 1b, except that the value of POD above which CPR decreases (the turning point of the contours) is nonexistent for a bias of 1 or less and then begins to decrease toward lower POD values as the bias increases above one. The existence of the turning point is best understood by considering overbiased forecasts (the region to the right of the unit bias line in Fig. 1), since this will apply to both the dHdF and dHdA methods as well as address the considerable difference between the asymptotic values of the CPR as POD approaches 1 in Figs. 1b and 1c compared with Fig. 1a. This is caused by the definitions of the measures involved: the conventional ETS does not include an "awareness" of the limit to the number of new hits possible as the POD is increasing. Thus, in the limit of POD equal to 1, the plot in Fig. 1a displays CPR values such as 0.4, etc., even though the ETS cannot be improved further by increasing the bias since there is no observed area left to hit (although those CPR values are relevant for a decrease in the bias). The definitions of dHdF and dHdA, on the other hand, are based on interpolations along functions, $H(F)$, such as the one shown in Fig. 1 of Mesinger (2008), that for overbiased forecasts ($F > O$) approach the limiting line $H = O$ as POD approaches 1. Therefore, as this limit is approached, the interpolation functions become flatter and flatter, requiring for improved bias-adjusted ETSs larger and larger increases in bias for a given small increase in the hit area. This is reflected by the decreasing values of the CPR, and by it approaching 0 as the POD approaches 1. For this to be shown in the CPR contours, they need to display turning points as seen in Figs. 1b and 1c. As to the difference between Figs. 1b and 1c, the CPR values in Fig. 1c are greater than those of Fig. 1b; thus, the dHdA method results in bias-adjusted TS or ETS measures that are more difficult to hedge by increasing the bias than those resulting from the dHdF method.

Figure 2 displays the CPR for $B = 1$ versus POD for the two bias adjustment methods applied to either TS or ETS compared to the TS, the ORSS, the CSS, and the ETS. The latter three are evaluated for event frequencies of 0.05 (Fig. 2a) and 0.25 (Fig. 2b), the approximate event frequencies for relatively rare and common events, respectively (see Baldwin and Kain 2006). Although sensitivity to event frequency is an important matter to consider when evaluating performance measures, especially in verifying rare events (Stephenson et al. 2008),
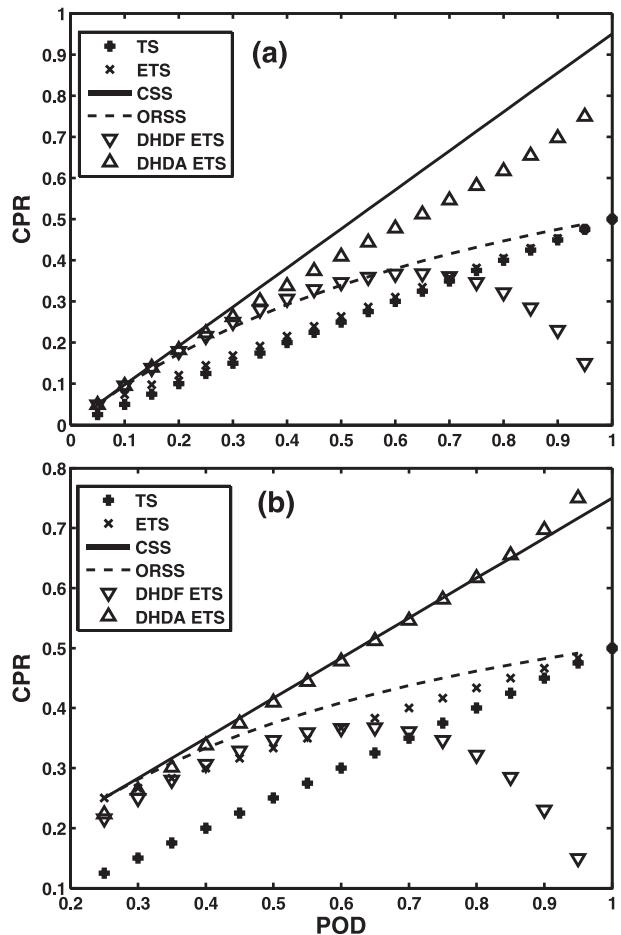

FIG. 2. The CPR vs POD for both the dHdF and dHdA methods of bias adjustment applied to either TS or ETS at $B = 1$ shown along with the TS CPR at $B = 1$. The ORSS, CSS, and ETS CPRs for $B = 1$ and event frequencies of (a) 0.05 and (b) 0.25 are included. See annotations in the figure.

a detailed analysis is beyond the scope of this note. In Fig. 2, only results for unbiased forecasts having better than random skill are displayed; hence, the initial value on the abscissa is $P = \alpha$. This follows readily from Table 2: the probability of a random hit, αP, must be equal to the product of the forecast frequency, αB, with $B = 1$, and the observed frequency, α. Figure 2b includes curves showing the values of CPR along the vertical line corresponding to $B = 1$ in Fig. 1.

Figure 2a reveals the CSS to be most resistant to hedging by inflating the bias beyond 1 for all values of POD, with the dHdA ETS falling into second place. For more commonly occurring events, the dHdA ETS overtakes the CSS at values of POD higher than about 0.8 (Fig. 2b). The TS and ETS CPR values are nearly the same for the 0.05-event frequency (Fig. 2a), but the TS becomes easier to hedge relative to ETS for the higher

event frequency (Fig. 2b). The ORSS CPR is always below that for the CSS and is less than the dHdA ETS CPR except at low POD for the higher event frequency (Fig. 2b). The dHdF ETS reaches a maximum CPR at a POD value near 0.6 and then begins to decrease, indicating easier ability to hedge this performance measure by increasing the bias when the POD is greater than 0.6. When considering hedging by increasing the bias, the fact that the dHdA ETS increases monotonically and exceeds the dHdF ETS for all POD values in Fig. 2 indicates the dHdA ETS as the better performance measure.

4. Considerations for bias correction

The CPR for the bias-adjusted ETS may indicate the extent of changes to be expected in the conventional ETS if forecasts are bias corrected. Here, *bias correction* refers to using past performance to remove bias (systematic error) from forecast values prior to verification such as performed by McCollor and Stull (2008) for quantitative precipitation forecasts (QPFs) and is not related to the bias adjustment of scores discussed heretofore. In fact, bias correction may not be specifically intended to improve scores for dichotomous forecasts but, rather, to reduce the mean absolute error (MAE); however, improving the MAE is likely to improve TS and ETS as well. In the case of bias reduction, the CPR poses an upper bound for the fraction of removed forecasts that are hits below which the performance measure indicates improvement. In other words, if the fraction of hits for the forecast points or area removed in reducing bias exceeds the CPR for a given performance measure, the performance measure will indicate a degraded forecast. Therefore, the lower the ETS CPR value for an overbiased forecast, the less likely it may be to correct (reduce) the bias and not degrade the ETS. Of course, the CPR value changes as the bias and POD change. If the dHdA method approximately simulates the effects of a bias correction, then the resulting dHdA ETS for the uncorrected biased forecasts may estimate the conventional ETS for bias-corrected forecasts and the dHdA ETS CPR may indicate the degree of difficulty involved in achieving such bias-corrected forecasts. This assertion is speculative, requiring that the dHdA method approximate the effects of the systematic error removal algorithm used to perform the bias correction.

Figure 3a shows the ETS (histogram bars), ETS CPR (lines), and bias (symbols) for the National Centers for Environmental Prediction (NCEP) North American Mesoscale (NAM) model along with the same for the NCEP Global Forecast System (GFS) for 24-h QPFs at the 24-h projection time. The performance measures shown are cumulative results over the continental United
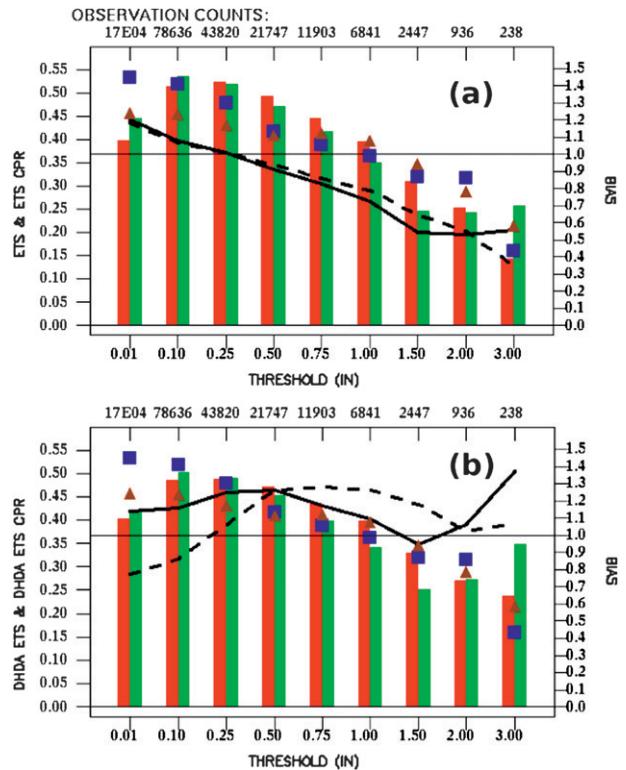

FIG. 3. Histogram comparison of (a) ETS with (b) dHdA ETS for GFS (red left bar) vs NAM (green right bar) at each threshold (in.) along the abscissa. In each panel, overlaid lines show corresponding CPR values for GFS (dashed) and NAM (solid). The GFS and NAM biases (blue squares and brown triangles, respectively) have values on the right axis. See text for additional details.

States at 40-km resolution for 1 December 2007–29 February 2008 computed from summary statistics generated by the NCEP/Environmental Modeling Center (EMC). The model QPF grids and the verifying analyses from the NCEP/Climate Prediction Center (Shi et al. 2003) are all remapped to the 40-km verification grid. This is done by a procedure in which precipitation is considered constant over the model or analysis grid box. With this scheme, the total precipitation is intrinsically conserved to a desired degree of accuracy. (More details on the precipitation verification system are available online at http://www.emc.ncep.noaa.gov/mmb/ylin/pcpverif/scores/.) The lowest three thresholds in Fig. 3a show the highest biases for both the GFS and the NAM models, with the GFS bias noticeably greater. Both models have nearly the same ETS for the 0.25-in. threshold, but the NAM model ETS exceeds that for GFS at the two lowest thresholds. The ETS CPR values are nearly the same for both models for these three lowest thresholds, indicating nearly equal likelihood of preserving or improving the ETS if the bias is decreased by a small incremental change.

For the ongoing discussion, quantities associated with the dHdA adjustment are denominated as ''adjusted'' quantities. Figure 3b plots the adjusted ETS and its CPR for the GFS and NAM models and is comparable to Fig. 3a. Unlike for the ETS CPR, the adjusted CPR values are much different at the lowest three thresholds where the bias is large. The CPR for the GFS-adjusted ETS is markedly lower than that for the NAM model. This indicates that it would be more difficult to improve the adjusted ETS by decreasing the bias for GFS compared to the NAM model. For example, at the 0.10-in. threshold, the GFS-adjusted ETS would improve only if no more than about 32% of the removed forecasts are hits; while, for the NAM model, one could have up to 43% of the removed forecasts be hits without degrading the adjusted ETS. Using the adjusted ETS CPR values in Fig. 3b to assess the chances of preserving the conventional ETS for a bias correction yields a result consistent with the intuitive impression that the model with the higher bias (GFS) may pose a greater challenge for preserving the conventional ETS after a bias correction. Therefore, the dHdA-adjusted ETS CPR for uncorrected forecasts may be a useful indicator of the likelihood of preserving or improving the conventional ETS for subsequent bias-corrected forecasts, especially when the required bias change is substantial.

5. Summary

This note derives the critical performance ratio (CPR) for bias-adjusted threat scores and equitable threat scores. Bias adjustment cannot remove explicit bias dependence from either the scores themselves or their CPR functions as formulated using Table 2. In qualitative terms, this is partly because bias measures the degree of adjustment needed; therefore, it is reasonable to expect to retain the explicit bias dependency. The dHdA method with its consistently increasing resistance to hedging by increasing the bias beyond one as the POD increases is superior to the dHdF method, and the dHdA bias-adjusted ETS is superior in this regard to all but the CSS at low event frequencies. To the extent that the dHdA bias adjustment simulates a bias correction, the CPR for the dHdA ETS may indicate the extent of changes to expect in the conventional ETS for bias-corrected forecasts.

APPENDIX

The Lambert *W* Function

The Lambert *W* function dates from the eighteenth century and is described in Mesinger (2008) with reference to the Internet. A useful Internet source is Weisstein (2005). The Lambert *W* function, $W(z)$, is defined as follows:

$$z = W(z)e^{W(z)}. \tag{A1}$$

The first derivative is obtained in a straightforward approach by differentiating Eq. (A1) with respect to *z*. This yields

$$1 = W(z)\frac{dW(z)}{dz}e^{W(z)} + \frac{dW(z)}{dz}e^{W(z)}. \tag{A2}$$

Using Eq. (A1) to make substitutions in Eq. (A2) allows Eq. (A2) to be written as

$$1 = \frac{dW(z)}{dz}\left[z + \frac{z}{W(z)}\right]. \tag{A3}$$

Solving (A3) for the derivative of *W* with respect to its argument *z* gives the desired result:

$$\frac{dW(z)}{dz} = \frac{W(z)}{z[1 + W(z)]}. \tag{A4}$$

REFERENCES

Baldwin, M. E., and J. S. Kain, 2006: Sensitivity of several performance measures to displacement error, bias, and event frequency. *Wea. Forecasting,* **21,** 636–648.

Brill, K. F., 2009: A general analytic method for assessing sensitivity to bias of performance measures for dichotomous forecasts. *Wea. Forecasting,* **24,** 307–318.

McCollor, D., and R. Stull, 2008: Hydrometeorological accuracy enhancement via postprocessing of numerical weather forecasts in complex terrain. *Wea. Forecasting,* **23,** 131–144.

Mesinger, F., 2008: Bias adjusted precipitation threat scores. *Adv. Geosci.,* **16,** 137–142.

Shi, W., E. Yarosh, R. W. Higgins, and R. Joyce, 2003: Processing daily rain-gauge precipitation data for the Americas at the NOAA Climate Prediction Center. Preprints, *19th Conf. on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology,* Long Beach, CA, Amer. Meteor. Soc., P1.6. [Available online at http://ams.confex.com/ams/pdfpapers/56719.pdf.]

Stephenson, D. B., B. Casati, C. A. T. Ferro, and C. A. Wilson, 2008: The extreme dependency score: A non-vanishing measure for forecasts of rare events. *Meteor. Appl.,* **15,** 41–50.

Weisstein, E. W., 2005: Lambert W-function. *MathWorld*—A Wolfram Web resource. [Available online at http://mathworld.wolfram.com/LambertW-Function.html.]

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences.* 2nd ed. Academic Press, 630 pp.