SUPPLEMENT TO
THE WGNE ASSESSMENT OF SHORT-TERM QUANTITATIVE PRECIPITATION FORECASTS

BY ELIZABETH E. EBERT, ULRICH DAMRATH, WERNER WERGEN, AND MICHAEL E. BALDWIN

Categorical, or conditional, statistics quantify the skill in the prediction of the *occurrence* of rain, and are based on the familiar 2 × 2 (yes/no) contingency table. Given a set of matched rain forecasts and observations, the contingency table is a matrix giving the frequencies of predicted and observed rain occurrence and nonoccurrence. "Hits," H, are correct predictions of rain occurrence. "Misses," M, indicate rain occurrences that were not predicted, while "false alarms," F, denote predictions of rain for which no rain occurred. Finally, the frequency of correct forecasts of "no rain" can be denoted by Z (for "zero"). The total number of forecasts is $N = Z + F + M + H$. Although several categorical statistics can be computed from the entries in the contingency tables for brevity we will concentrate on the two statistics that are used to evaluate model quantified precipitation forecasts (QPFs) at the National Centers for Environmental Prediction (NCEP), namely the bias score and the equitable threat score.

The bias score,

$$\mathrm{BIAS} = \frac{F+H}{M+H}, \tag{S1}$$

measures the ratio of the predicted rain area (or frequency) of predicted to the observed area (frequency), without regard to forecast accuracy. The bias score is used to assess the model's tendency to under- or overpredict rain occurrence. If BIAS = 1.0, then the predicted rainfall area (frequency) is the same as was observed, although it may not be located in the same place (time).

The threat score,

$$\mathrm{TS} = \frac{H}{H+F+M}, \tag{S2}$$

has been used for several decades to measure the correspondence between the forecast and observed rain occurrence. It takes into account both false alarms and misses and has the advantage of not being dominated by the no-rain events, as occurs in most regimes for the simple accuracy, $(Z + H)\, N^{-1}$. The threat score ranges from 0 to 1, with perfect score of 1 indicating that all rain events were perfectly forecast, and all rainfall forecasts corresponded to observed events. Schaefer (1990) showed that the threat score is biased toward sample sets with higher rain frequencies, making it a misleading statistic to use when comparing forecast skill across different regimes. He proposed modifying the threat score to take into account the number of correct rain forecasts (hits) that would be expected purely due to chance. This "equitable threat score" is calculated as

$$\mathrm{ETS} = \frac{H - H_{\mathrm{random}}}{H+F+M-H_{\mathrm{random}}}. \tag{S3}$$

The number of random hits expected due to chance is given by

$$H_{\text{random}} = \frac{(H+M)(H+F)}{N}. \quad \text{(S4)}$$

The ETS ranges from –1/3 to 1, with a value of 1 again indicating perfect correspondence between predicted and observed rain occurrence. The ETS is now used in preference to the threat score for rainfall verification at most operational centers.

Continuous statistics such as the mean error, mean absolute error, and root mean squared (rms) error are often used to verify rain amount. An alternative approach is to compute categorical statistics using a variety of rain thresholds, T. Setting T close to 0 produces the usual rain/no rain contingency table. Increasing the value of T gives information on the ability of the model to correctly predict increasingly greater rainfall amounts. Many of the results shown in Ebert et al. 2003 (se section titled "Results for Current Operational NWP Models") make use of threshold-dependent categorical statistics.

The statistics described thus far are generally applied to a set of matched forecast and observed (or analyzed) points in space and/or time. A somewhat different approach treats the forecast and observed rain systems as contiguous rain areas (CRAs) in the gridded data, and verifies the properties of these *entities* (Ebert and McBride 2000). Pattern matching is used to estimate the displacement of the forecast rain systems to the nearest grid point. This is done by horizontally translating the forecast until the total squared error within the domain defined by the overlap of the entities is minimized. The scheme does not consider rotation or distortion of the forecast. The areal extent, rain volume, and intensity of the rain systems are also verified.

An advantage of this technique is that the mean squared error (mse) of the forecast may be decomposed into relative contributions from displacement, volume, and pattern errors, that is,

$$\text{MSE}_{\text{total}} = \text{MSE}_{\text{displacement}} + \text{MSE}_{\text{volume}} + \text{MSE}_{\text{pattern}}. \quad \text{(S5)}$$

The terms on the right-side of (S5) are estimated as follows. The mean squared error of the original forecast is given by

$$\text{MSE}_{\text{total}} = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2 , \quad \text{(S6)}$$

where f_i and o_i are respectively the forecast and observed rainfall at grid point i and N is the number of grid points in the verification domain. The verification domain in this case is the set of grid points in the forecast and observed rain entities before and after shifting the forecast to a position of zero displacement. The mse after shifting the position of the rain forecast to the correct location is similarly calculated as

$$\text{MSE}_{\text{shift}} = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2 , \quad \text{(S7)}$$

where f_i' is the shifted forecast at grid point i. The contribution to total error due to displacement is simply the difference between the mean square errors before and after the shift,

$$\text{MSE}_{\text{displacement}} = \text{MSE}_{\text{total}} - \text{MSE}_{\text{shift}} . \quad \text{(S8)}$$

The remaining error can be further separated into components due to volume and pattern errors,

$$\text{MSE}_{\text{volume}} = (\bar{f}' - \bar{o})^2 \quad \text{(S9)}$$

$$\text{MSE}_{\text{pattern}} = \text{MSE}_{\text{shift}} - \text{MSE}_{\text{volume}} , \quad \text{(S10)}$$

where the overbar denotes the mean value over the verification domain. The volume error accounts for the difference between the mean forecast and observed rain intensities, while the residual pattern error accounts for finescale differences in the shape and structure of the rain fields. This error decomposition can help model developers and users to separate dynamical and physical sources of QPF errors.

REFERENCES

Ebert, E. E., and J. L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrol.,* **239,** 179–202.

——, U. Damrath, W. Wergen, and M. E. Baldwin, 2003: The WGNE assessment of short-term quantitative precipitation forecasts (QPFs) from operational numerical weather prediction models. *Bull. Amer. Meteor. Soc.,* **84,** ???–???.

Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting,* **5,** 570–575.