RECALIBRATING MRF RE-FORECASTS:
A LOGISTIC REGRESSION APPROACH

Thomas M. Hamill, Jeffrey S. Whitaker, and Xue Wei

NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado

contact: hamill@cdc.noaa.gov, (303) 497-3060

This note describes a method for generating calibrated (highly reliable) tercile probability forecasts generated from NOAA-CIRES CDC's "re-forecast" data set. Tercile probabilities here will refer to the probability that the subsequent observed anomaly will be higher than the 67th percentile of the distribution of climatological observed values (the "upper" tercile) and lower than the 33rd percentile (the "lower" tercile). Our data set for calibrating these probabilities will be ensemble forecasts from 1979 to 2001 and the associated NCEP-NCAR reanalyses. The newly designed approach is more "MOS" in flavor than "perfect-prog." That is, we develop here a method for forecasting probabilities that inherently takes into account model systematic errors rather than setting the probabilities simply based on relative frequency from the ensemble. Evaluation of the benefits of this approach are provided after a brief description of the algorithm.

Our approach applies a logistic regression technique (e.g., Wilks 1995) separately at each model grid point using ensemble mean forecast data as the only predictor. Model grid points are separated by 2.5 degrees in both latitude and longitude. The reader may question both the use of only ensemble mean data as a predictor and the wisdom of treating each grid point separately. Concerning the former, we continue to search for other valuable predictors, but some obvious candidates such as ensemble spread have not improved forecast accuracy. Concerning the latter, we have found that undesirable small and unpredictable spatial scales are generally smoothed out by the averaging over time and over ensemble members.

The logistic regression model sets the probability that the observed anomaly V will exceed the upper or lower anomaly terciles $T_{0.67}$ and $T_{0.33}$ according to the equation (here, for upper tercile)

$$P(V > T_{0.67}) = 1 - \frac{1}{1 + exp\ (\hat{\beta}_0 + \hat{\beta}_1 x)} \quad (1)$$

where x is the ensemble mean forecast anomaly and $\hat{\beta}_0$ and $\hat{\beta}_1$ are fitted regression coefficients.

The process for determining coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ for the upper tercile probabilities are described here for week 2 (days 8-14) forecasts; lower tercile probabilities are handled in an analogous fashion. The regression parameters are determined from a data set of predictors (here, ensemble mean week 2 anomalies from climatology) and the associated binary verification data (was the observed anomaly above the upper tercile ($P = 1$) or below ($P = 0$). The process for generating the regression coefficients for a given day of the year is as follows:

(1) Calculate a daily running mean climatology of the week 2 forecast and week 2 observed values (from NCEP-NCAR reanalyses). For a given day of the year, the climatology is the week 2 value averaged over the 23 years and the 31 days (15 before, 15 after) centered on the date of interest.

(2) Subtract the forecast and observed climatology from the week 2 forecasts and observations, respectively. This creates a week 2 forecast anomaly (relative to the model) and week 2 observed anomaly (relative to the reanalysis).

(3) Set the upper tercile as the 67th percentile of the sorted observed anomaly data.

(4) Create the binary verification data using a data set of 23 years * 31 days (again, 15 before and 15 after the date of interest). Each sample verification is categorized as being above the upper tercile (P = 1) or below it (P = 0).

(5) Determine $\hat{\beta}_0$ and $\hat{\beta}_1$ through logistic regression using the ensemble mean anomaly as the only predictor. Code and documentation can be found at http://users.bigpond.net.au/amiller/

.

Figure 1 illustrates the process for determining tercile probabilities for 850 temperatures, here for a December 16 along the Oregon coast. A scatterplot of the ensemble mean week 2 forecast anomaly is plotted against the corresponding week 2 observed anomaly using the 23×31 samples. From the observed data, the upper and lower terciles are calculated (horizontal dashed lines). If one were to set the upper tercile probabilities just using the relative frequencies of observed values in a bin around a forecast value (the bin limits denoted by the vertical lines), then the average bin probabilities would be denoted by the horizontal solid lines. For example, counting all the forecasts with an anomaly between -6 and -4 C and tallying how often the observed anomaly exceeds the upper tercile, the probability is approximately 12 percent. If all the data is fed into the logistic regression, probabilities are determined as a function of the forecast anomaly according to the dotted curve.

The resulting probabilities are highly reliable and somewhat skillful. Figure 2 shows reliability diagrams for December forecasts over the 23 years, with probabilities determined both by the approach outlined above (probabilities were cross validated, e.g., when setting probabilities for year 1 data, year 1 was excluded from the training data set) and by a simpler process, whereby probabilities are set by ensemble relative frequency (e.g., if 10 of 15 members exceed the upper tercile, set the probability to 67 %. Note that using the raw ensemble data in this manner produces unskillful and unreliable probabilistic forecasts, whereas with the logistic regression method, the forecasts are both reliable and slightly skillful.

The above methodology is used for all fields except precipitation, where a slightly different method is used. For precipitation, we use ensemble mean data without subtracting off the climatology. Also, because precipitation forecast and observation data tends to be non-normally distributed, we power transform the data before applying the logistic regression. Specifically, if x denotes the ensemble mean forecast, we generate a transformed forecast $\widetilde{x}$ according to $\widetilde{x} = x^{0.25}$. The process is illustrated in Fig. 3, and the reliability of subsequent forecasts is demonstrated in Fig. 4.

REFERENCES

Wilks, D. S., 1995: *Statistical methods in the atmospheric sciences: an introduction.* Academic Press. 467 pp.

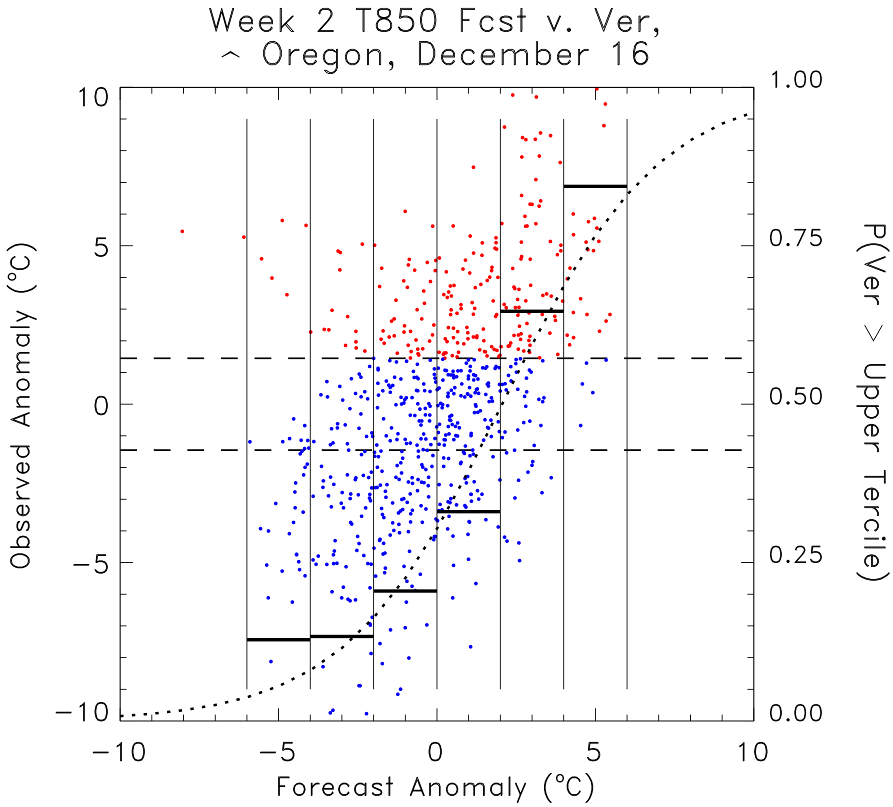

Figure 1. Scatterplot of the ensemble mean week 2 forecast anomaly and corresponding week 2 observed anomaly. Upper and lower terciles are denoted by dashed lines. Red dots are samples with observed anomalies above the upper tercile; blue dots below. Vertical lines denote bin thresholds for setting tercile probabilities based on the relative frequencies of observed values above the upper tercile. Thick horizontal lines denote the probabilities associated with each bin (refer to probabilities labeled on the right side of the plot). Dotted curve denotes the upper tercile probabilities determined by logistic regression.

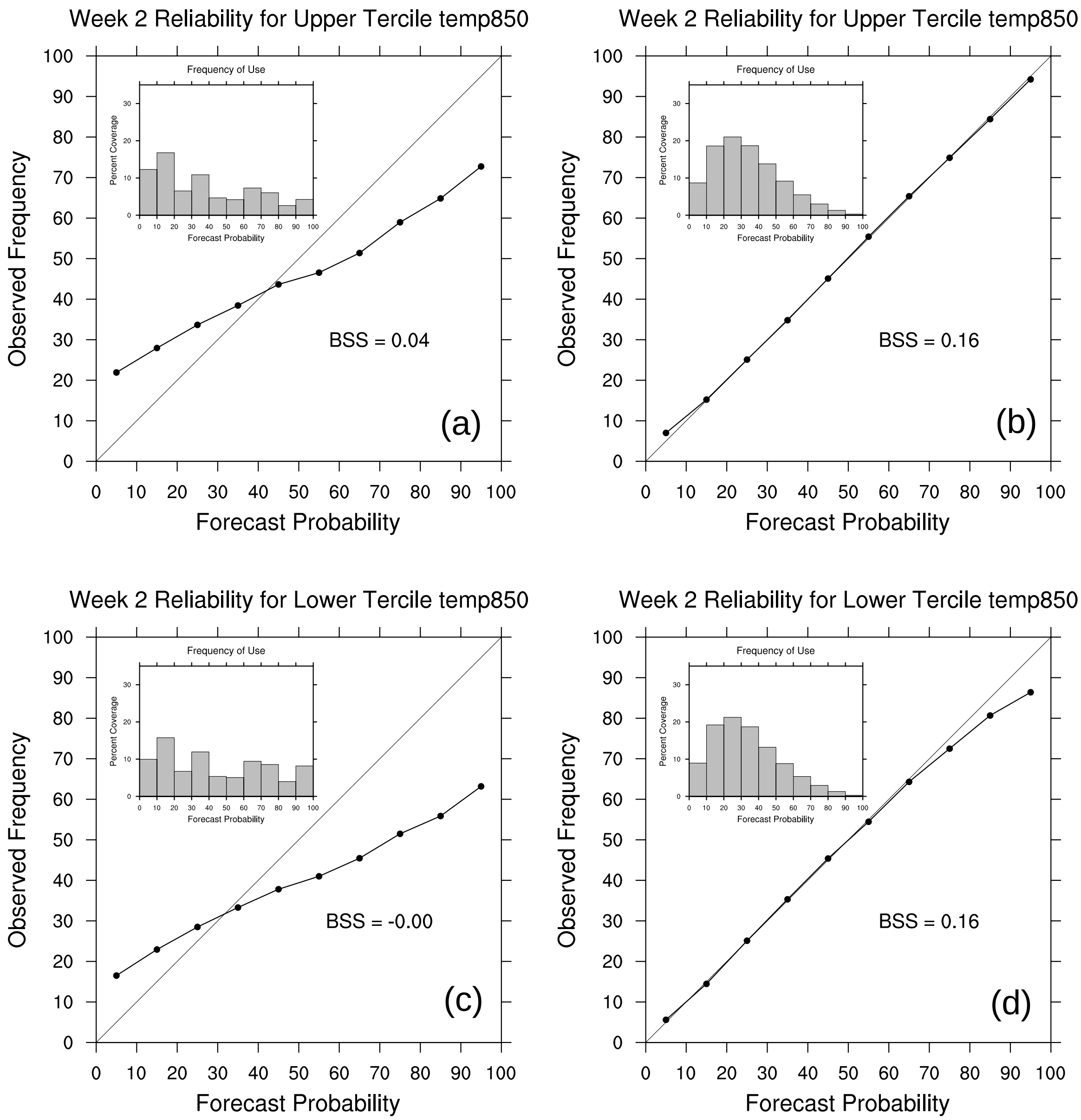

Figure 2. Reliability of week 2 forecasts of 850 hPa temperature for the month of December. Heavy solid line denotes reliability curve (relative frequency of observed exceeding tercile as a function of forecast probability) Inset histogram indicates relative frequency of issuing various probabilities. Brier skill score relative to climatology is denoted in upper right. Reliabilities determined from Northern Hemisphere forecast data north of 20^{o}. (a) Upper tercile, with probabilities determined from the ensemble relative frequency. (b) Upper tercile, probabilities determined by logistic regression with ensemble mean anomaly as predictor. (c) As in (a), but for lower tercile. (d) As in (b), but for lower tercile.

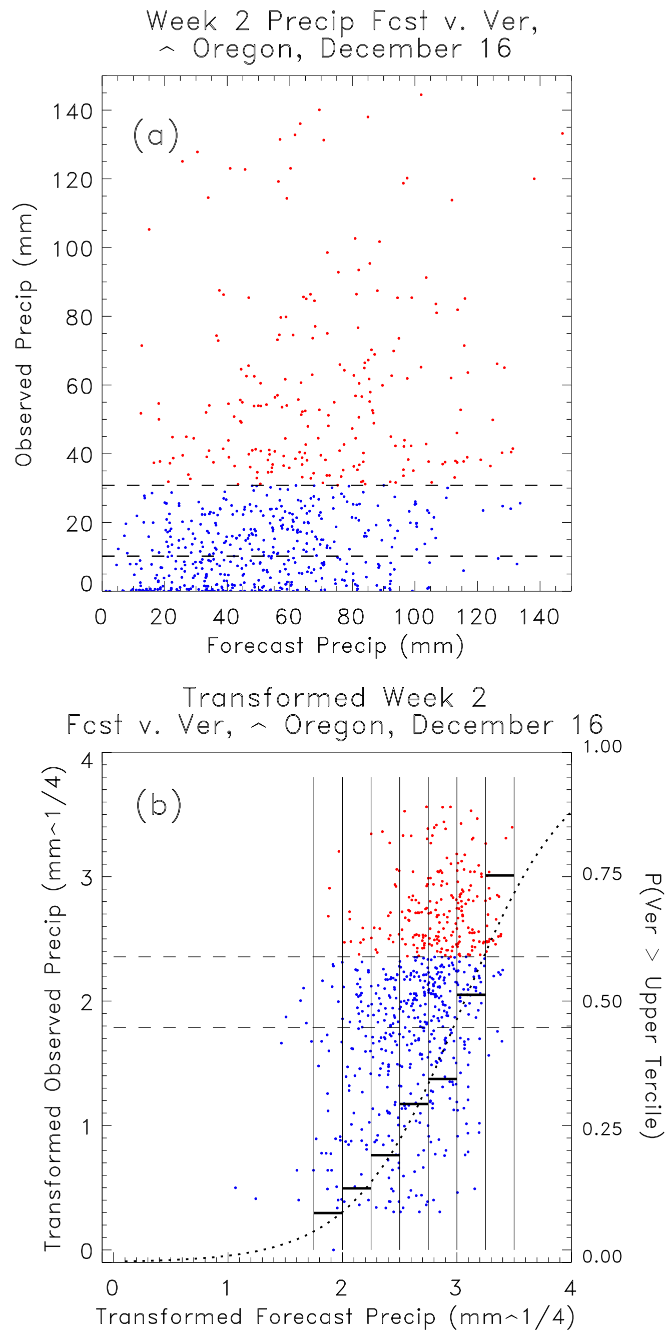

Figure 3. Illustration of logistic regression process for week 2 precipitation. Data is for 16 December at grid point near Oregon. (a) Scatterplot of ensemble mean forecast precipitation amount vs. observed amount (determined from an average of all observations inside the $2.5^o \times 2.5^o$ grid box). Horizontal dashed lines denote lower and upper terciles. (b) As in (a), but after power transformation. Additionally, vertical black lines denote bin thresholds as in Fig. 1, and thick horizontal solid black lines denote estimated probabilities determined by relative frequency. Dotted curve denotes the upper tercile probabilities determined by logistic regression.

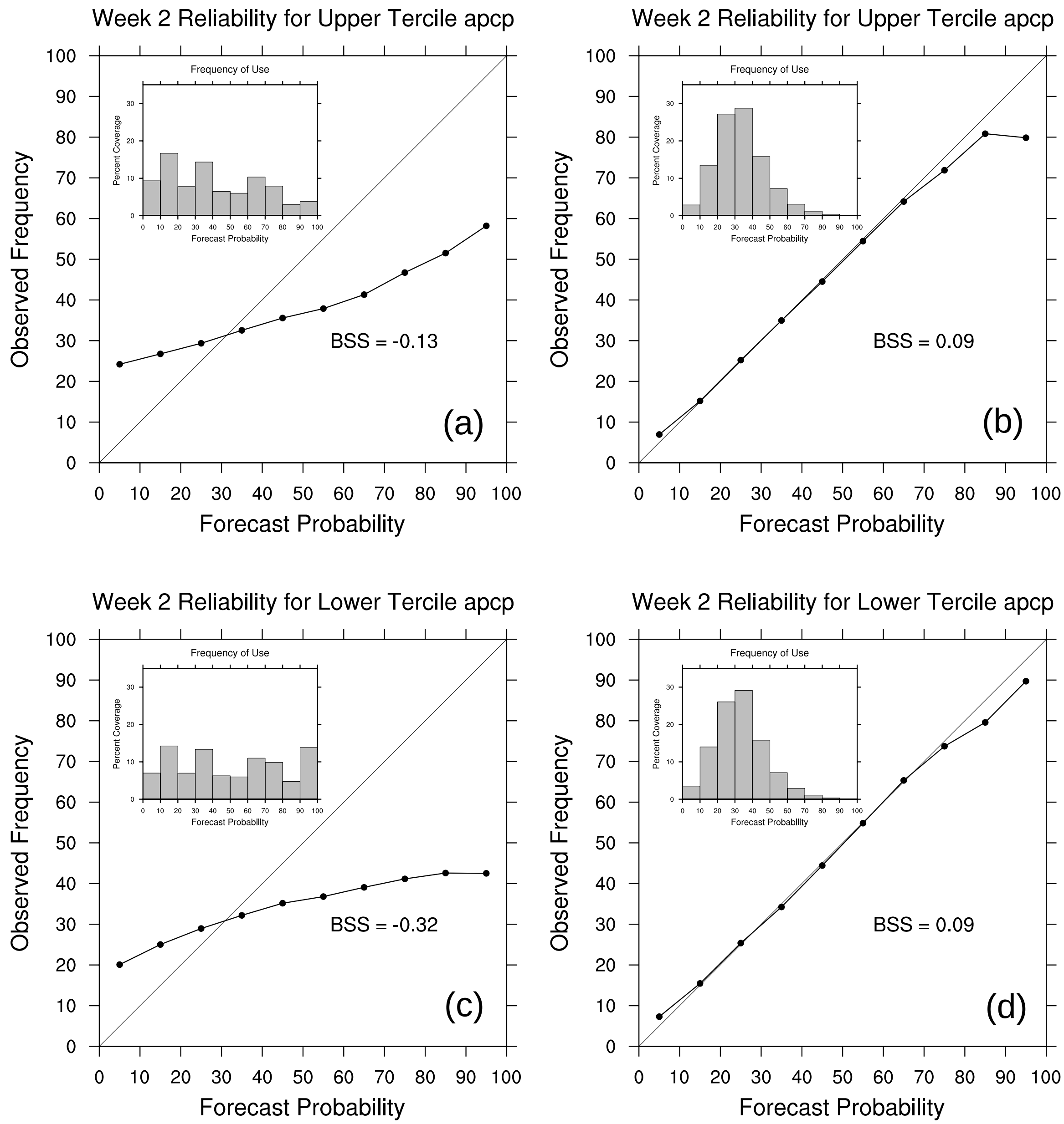

Figure 4. As in Fig. 2, but for reanalysis precipitation.