6.3 Short Range Ensemble Forecast (SREF) Calibrated Thunderstorm Probability Forecasts: 2007-2008 Verification and Recent Enhancements

David R. Bright* and Jeremy S. Grams

NOAA/NWS/NCEP Storm Prediction Center
Norman, Oklahoma

1. INTRODUCTION

The Storm Prediction Center (SPC) is the National Weather Service's (NWS) center of expertise for forecasting thunderstorms, including high-impact and dangerous severe thunderstorm events such as tornadoes, large hail, damaging winds, and heavy rainfall. In addition to severe thunderstorm products such as watches and outlooks, the SPC also issues general thunderstorm outlooks for the conterminous United States and its adjacent coastal waters. For the purposes of this paper, a thunderstorm is defined as deep moist convection producing at least one cloud-to-ground (CG) lightning strike within 10 miles of a point. The SPC thunderstorm/no thunderstorm outlook demarcation is, by definition, a line separating areas with a ≥ 10% probability of one or more thunderstorms. The operational thunderstorm outlook for the day one period (i.e., today) is issued initially at 0600 UTC and covers the 24 hour period ending at 1200 UTC the following day. As such, the operational outlook provides minimal temporal, spatial, or probabilistic details concerning the evolution of thunderstorms.

The experimental Enhanced Resolution Thunderstorm (ENHT) outlook was designed to address these shortcomings by providing additional thunderstorm forecast information for use by NWS Weather Forecast Offices (WFO), emergency managers, media, and the general public. The ENHT outlook breaks the convective day into two periods with the division at 0300 UTC, which normally separates diurnally-driven thunderstorms from nocturn events. The initial ENHT forecast is issued at 0600 UTC and is valid for the periods of 1200 to 0300 UTC and 0300 to 1200 UTC. It is subsequently updated at 1300 UTC, 1630 UTC, 2000 UTC, and 0100 UTC. The ENHT outlooks include three stepwise probability thresholds of 10%, 40%, and 70%. These thunderstorm hazard probabilities provide SPC customers and partners with additional information about the likelihood and location of thunderstorms, and can help increase public safety. An example of an ENHT outlook from 10 June 2008 is shown in Fig. 1.

As a source of guidance to address the national mission of the SPC, ensemble forecast output has been incorporated into all SPC program areas (e.g., Bright et al. 2007). Because almost 90% of all SPC products are for forecast periods three days or less, the National Centers for Environmental Prediction (NCEP) Short Range Ensemble Forecast (SREF; Du et al. 2006) system is particularly well suited to meet the operational needs of the SPC. Specialized post-processing of the SREF by the SPC has been performed since 2003. This post-processing is designed to extract specific information relevant to the SPC mission, including innovative applications toward the convective forecast problem (Bright et al. 2008) and the development of calibrated probabilistic thunderstorm and severe thunderstorm guidance (Bright et al. 2005; Bright and Wandishin 2006). Here, calibration refers to additional statistical post-processing to provide reliable, unbiased, and skillful probabilistic guidance for the phenomena of interest.

The SPC SREF post-processing includes all 21 members of the NCEP SREF plus the 3-hour time lagged, operational WRF-NAM (for a total of 22 members) every 6 hours (0300, 0900, 1500, and 2100 UTC). Output is available at 3 hour intervals through 87 hours. The SPC SREF post-processing focuses on diagnostics relevant to the prediction of SPC mission-critical high-impact,

**Corresponding author address*: David R. Bright, Storm Prediction Center, 120 David L. Boren Blvd, Norman, OK 73072; e-mail: david.bright@noaa.gov

mesoscale weather. All SPC SREF guidance products shown herein are available in real-time on the SPC website at http://www.spc.noaa.gov/exper/sref/.

2. VERIFICATION OF SREF CALIBRATED THUNDERSTORM PROBABILITIES AS ENHANCED THUNDERSTORM OUTLOOK GUIDANCE

The SREF calibrated thunderstorm guidance as described in Bright et al . (2005) is computed on the 40 km AWIPS 212 grid (which is roughly equivalent to the probability of a thunderstorm within ~10 miles of a point) covering 3-hour time periods. Separate calibration tables are maintained for each run of the SREF (0300, 0900, 1500, and 2100 UTC) and for each 3-hour forecast interval beginning at forecast hour 03 through the entire SREF run (i.e., forecast hour 87). Data from the National Lightning Detection Network (NLDN) serve as the predictand, where at least one CG lightning strike within a 40 km grid box defines the occurrence of a thunderstorm. The SREF predictors are the cloud physics thunder parameter (CPTP; Bright et al. 2005) and the total precipitation during the previous 3-hour period. The system employs historical data from the past year to produce reliable, 3-hour thunderstorm forecasts for the continental U.S. and near-shore waters. As a demonstration of its reliability, an attributes diagram for all 3-hour forecasts valid between forecast hours 39 and 45 during most of the 2008 warm season (25 April to 22 October) shows the 3-hour guidance to be both reliable and skillful (Fig. 2).

To meet the requirements of the longer ENHT valid periods (e.g., 1200 UTC to 0300 UTC), the 3-hourly calibrated forecasts are combined at the SPC into 12-hour and 24-hour forecast periods. To illustrate the extension of the probabilistic guidance from 3-hour to 12-hour periods, a frequency correction identical to the approach in Bright et al. (2005) is used. The consecutive 3-hour probabilities are rounded into 5% increments, weighted spatially, and matched to the historical frequency of occurrence over the past year. For example, at one randomly chosen grid point, the 12-hour probability given four consecutive 3-hour calibrated probabilities of [0%, 10%, 50%, 5%] was 53%. Although the 12-hour periods do not match perfectly the ENHT valid periods, they are currently the closest available SREF-based calibrated thunderstorm guidance.

The SREF 12-hour calibrated thunderstorm guidance was verified for the one-year period beginning 1 December 2007 and ending 30 November 2008. The verification domain covers the contiguous U.S. and adjacent waters as shown in Fig. 1. Although higher probabilistic resolution is available internally and on the SPC website, for verification purposes the SREF guidance is binned into the three ENHT stepwise probabilities of 10%, 40%, and 70%. Thus, any calibrated value less than 10% is assigned 0%; from 10% to 39.99% is assigned 10%, from 40% to 69.99% is assigned 40%; and at or above 70% is assigned 70%. Since the objective is to evaluate the guidance available to SPC forecasters, the SREF cycle chosen is the run available to SPC forecasters during ENHT preparation (Table 1 - columns 1 and 2 show the SREF cycle and ENHT issue time, respectively). For example, the first issuance of the ENHT outlook is 0600 UTC, and the SREF guidance available during forecast composition is the 2100 UTC run.

Results for the first 12-hour forecast period are now discussed (1500 UTC to 0300 UTC). Table 1 contains a compilation of verification scores computed daily over the entire verification domain, and then averaged over all available days during the year. Thus, in Table 1 all days are equally weighted regardless of actual or predicted thunderstorm coverage. The total number of forecasts verified is contained in the third column of Table 1; ideally 366 days would be available, but random data outages, archive issues, etc. resulted in about 50 missing days. No attempt was made to recover data that was not archived in real-time. The reliability scores (REL10, REL40, REL70 - columns 4 to 6, respectively) are a measure of forecast bias. It is calculated by examining the reliability of the 10%, 40%, and 70% thresholds daily, and assigning a 0, 1, or 2 if the predicted value was an over-forecast, reliable forecast, or under-forecast, respectively. The reliability scores were then averaged over all available days, such that a reliability of 1 is a perfect (unbiased) forecast, and values < 1 or > 1 are indicators of over-forecasting (high bias) or under-forecasting (low bias), respectively. Again, the probability thresholds have been truncated to step functions. For example, a SREF prediction of 10% is considered reliable if the verifying frequency of occurrence is ≥ 10% and < 40%. The SREF reliability at 10% (REL10) suffers from a small, over-forecast (high bias). At 40% (REL40) the results are approaching perfect reliability (0.94 to 0.98), while at 70% (REL70) the score indicates a tendency for the SREF to over-forecast thunderstorms (high bias). Interestingly, the reliability does not appear to improve with time, such that more recent SREF runs are not necessarily less biased than older runs.

The Brier Score (BS; Wilks 1995) is often used to verify probabilistic forecasts with lower scores indicating better forecasts, and a perfect score of 0.0. The BS tends to show improvement as new runs become available (0.062 to 0.058; Table 1, column 7). The Relative Operating Characteristic (ROC; Mason 1982) was also examined. The ROC is often integrated such that values of 1.0 represent a perfect forecast system and values ≤ 0.5 are indicative of no skill. The ROC scores show a moderate-to-high ability to discriminate thunderstorms from non-thunderstorm locations (0.849 to 0.876; Table 1, column 8). The score also improves as new SREF guidance becomes available, indicating better discrimination of thunderstorm potential than in older runs. The last two metrics in Table 1 are the HIT and MISS (columns 9 and 10, respectively). The HIT represents the average predicted probability of a thunderstorm given that a CG strike occurred; the MISS is the average predicted probability at all points without CG lightning. The HIT shows a positive trend with time, starting at 21% from the 21 UTC SREF and reaching 26% by the 1500 UTC cycle. The MISS held steady at 3% over the first period.

For the second forecast period (0300 UTC to 1500 UTC), the reliability of the 10% and 40% thresholds (REL10 and REL40, respectively) generally improved when compared to the first period, although the 70% (REL70) threshold showed a slight degradation with an increasing tendency to over-forecast. Similarly, the BS improved during the second period, and like the first period results, the SREF guidance tended to improve with time (decreasing from 0.050 to 0.047). But, the ROC area decreased slightly when compared to the first period. The fact that the BS shows an improvement is probably due to a larger sample of correct 0% forecasts, while the decrease in the ROC seems to infer that the nocturnal events are more difficult to discriminate. Similar to the ROC, the HIT and MISS averages (~19% and 2%, respectively) also show less separation at night. Nevertheless, the ROC, HIT, and MISS scores are respectable, and still indicate an overall ability to discriminate events from non-events.

As previously mentioned, the values in Table 1 reflect the long-term average of the daily, domain values. This weights days with little or no CG lightning (predicted or expected) equal to days with considerable thunderstorm activity. The reliability of the entire sample collectively tells a slightly different story for both the first and second period (Figs 3 and 4, respectively). In the interest of space, only the 0900 UTC SREF guidance is shown, which corresponds to the bold line in Table 1. From the reliability diagrams, both forecasts are extremely reliable and well within the stepwise boundaries of the ENHT probability thresholds (i.e., the region between the solid lines). Because Table 1 inferred a tendency for the SREF guidance to over-forecast the 10% and 70% values (REL10 and REL70, respectively), and Figs. 3 and 4 do not seem to bear this out, the SREF guidance seemed to pay a penalty by weighting each day equally and averaging.

Another consideration is how well the guidance performs spatially across the entire ENHT domain. If the reliability at 10% is calculated at each grid point for the 0900 UTC SREF over entire year (and then evaluated as done in the reliability of REL10 described above), most of the nation receives reliable guidance for the 1500 UTC to 0300 UTC period (Fig. 5). The notable exceptions are Florida and along the West Coast where the system over-forecasts thunderstorms (high bias), and portions of the Intermountain West where an under-forecast (low bias) occurs. The under-forecast in the West appears reasonably well correlated to terrain and is likely related to unresolved terrain features in the SREF and its 40 km post-processed grid (note the Southeast Highlands and Mogollon Rim in Arizona, and the spine of the Rockies in Fig. 5). The second period (0300 UTC to 1500 UTC) again shows most of the nation with reliable, unbiased guidance, except for a tendency to over-forecast nocturnal storms in Florida, in some areas along the Gulf Coast and south Texas, and over much of the far West near and along the coast (Fig. 6). The second period summer season (01 June - 31 Aug 2008) shows a particularly notable over-forecast (high-bias) in SREF calibrated guidance over Florida and inland areas along the Gulf Coast through south Texas (Fig. 7; other seasons not shown). A similar story is gleaned from the 40 percent threshold (Figs. 8 and 9). Prior to 0300 UTC, a low-bias was found over much of the interior West with a high bias off the coast (while just inland a 40% forecast never occurred; Fig. 8). Most of the remainder of the nation, including Florida, is well calibrated. The second period shows the lack of nocturnal activity in the West, where many locations never experienced a 40% forecast. A pronouced over-forecast (high bias) continues around Florida and south Texas, and also over the southern Rockies (Fig. 9). And this high bias is particularly pronounced at night in the summer in Florida, south Texas, and along the Gulf Coast (Fig. 10).

The Brier Skill Score (BSS) is commonly used to verify probabilistic forecasts relative to climatology. Here, the skill is assessed relative to a 3-hour thunderstorm climatology for each week of the year, computed on the identical grid from eight years (1995-2002) of NLDN data (Bothwell 2005). Again, the 0900 UTC SREF over the entire year is used, with the BSS computed at each grid point. For the first period (1500 UTC to 0300 UTC), most grid points across the ENHT domain show a positive BSS (Fig. 11). In fact, the SREF guidance of the majority of the country is at least a 10% improvement over climatology (IOC), with the greatest skill between 30% to 40% IOC from the upper Midwest to New England, and over parts of the Ozarks into Alabama and Georgia. The only area of substantial negative skill is off the West Coast, with neutral or near-neutral skill in the Southwest (gray shading). As seen in the daily average values (Table 1), skill decreases during the second period, although a large portion of the country is still a 10% IOC particularly east of the Rockies (Fig. 12). Other than a few isolated "hot spots" in the West, the BSS is maximized from the lower Mississippi Valley to the upper Midwest (20% to 30% IOC).

3. SUMMARY AND DISCUSSION

An examination of SPC calibrated SREF guidance specifically for the ENHT outlook process was examined. Overall, the system provides reliable and skillful guidance for the ENHT outlook process. Examining the guidance seasonally and spatially confirmed some of the biases noted subjectively by SPC forecasters. First, the SREF system over-forecasts convection over the Southeast and along the Gulf Coast, particularly at night. The system also over-forecasts convection over the eastern Pacific Ocean, but under-forecasts convection in the Intermountain West during the day. This latter low-bias is likely due, at least in part, to the relatively coarse resolution of the SREF. The bias over the eastern Pacific and along portions of the Gulf Coast is believed to be due, at least in part, to the current calibration method. Briefly, the calibration relies on the SPC cloud physics thunder parameter (CPTP) and grid scale (i.e., total) precipitation from each member of the SREF (see Bright et al. 2005 for details). Over the ocean, it is not unusual to have a CPTP indicative of lightning potential but no convective precipitation. In this case, the calibration may be "fooled" by grid scale precipitation originating from low clouds in the model's marine boundary layer. To address this problem, a modified calibration scheme is currently under development that appears to improve the guidance near and over the oceans.

4. REFERENCES

Bothwell, P.D., 2005: Development of an operational statistical scheme to predict the location and intensity of lightning. Preprints, *Conf. on Meteorological Applications of Lightning Data,* San Diego, CA, Amer. Meteor. Soc., CD-ROM, 4.2.

Bright, D.R., S.J. Weiss, J.J. Levit, and R.S. Schneider, 2008: The evolution of multi-scale ensemble guidance in the prediction of convective and severe convective storms at the Storm Prediction Center. Preprints, *24th Conf. on Severe Local Storms*, Savannah GA, Amer. Meteor. Soc., P10.7.

Bright, D. R., M. S. Wandishin, S. J. Weiss, R. S. Schneider, and J. T. Schaefer, 2007: The Application of Climate Data in Calibrating Ensemble Guidance for the Prediction of Hazardous Weather. Preprints. *16th Conf. on Applied Climatology*, San Antonio, TX, Amer. Meteor. Soc., CD-ROM, 4.2.

Bright, D.R. and M.S. Wandishin, 2006: Post processed short range ensemble forecasts of severe convective storms. Preprints, *18th Conf. on Probability and Statistics in the Atmospheric Sciences,* Atlanta, GA, Amer. Meteor. Soc., CD-ROM, 5.5.

Bright, D.R., M.S. Wandishin, R.E. Jewell, and S.J. Weiss, 2005: A physically based parameter for lightning prediction and its calibration in ensemble forecasts. Preprints, *Conf. on Meteorological Applications of Lightning Data,* San Diego, CA, Amer. Meteor. Soc., CD-ROM, 4.3.

Du, J., J. McQueen, G. DiMego, Z. Toth, D. Jovic, B. Zhou, and H. Chuang, 2006: New Dimension of NCEP Short-Range Ensemble Forecasting (SREF) System: Inclusion of WRF Members, Preprints,*WMO Expert Team Meeting on Ensemble Prediction System*, Exeter, UK, Feb. 6-10, 2006, 5 pages [available online http://wwwt.emc.ncep.noaa.gov/mmb/SREF/reference.html or http://www.wmo.int/web/www/DPFS/Meetings/ET-EPS_Exeter2006/DocPlan.html]

Wilks, D.S., 1995: Statistical Methods in the Atmospheric Sciences. Academic Press, San Diego, CA, 467 pp.

5. FIGURES AND TABLE

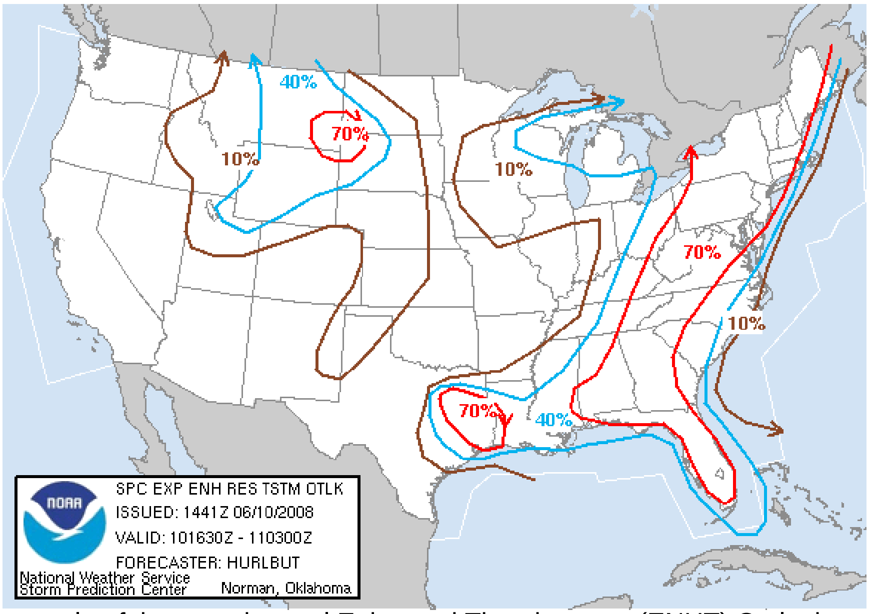

FIG. 1. An example of the experimental Enhanced Thunderstorm (ENHT) Outlook produced by the Storm Prediction Center. This forecast was issued at 1441 UTC 10 June 2008 for the period between 1630 UTC through 0300 UTC 11 June 2008. The forecast and verification area covers the contiguous U.S., and extends to the thin white lines over the western Atlantic, northern Gulf of Mexico, and eastern Pacific.

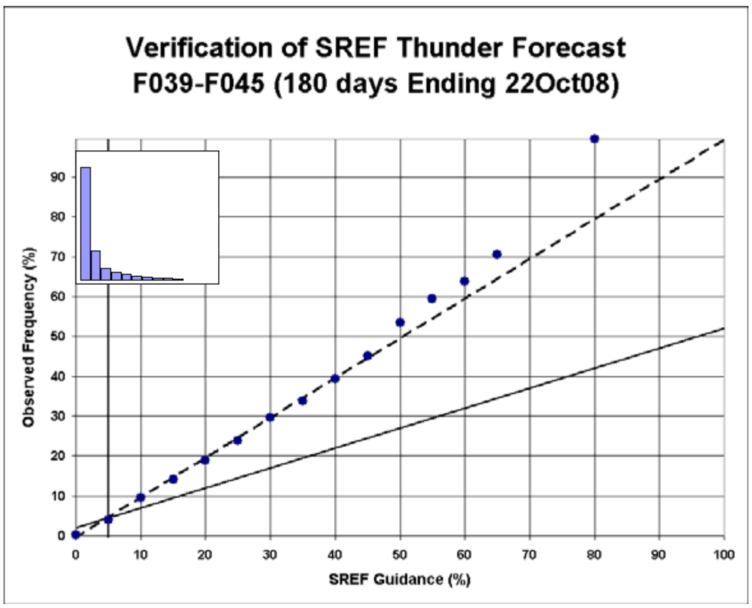

FIG. 2. Attributes diagram for all 3-hour SREF forecast valid between forecast hours 39 and 45 for the period 25 April 2008 to 22 October 2008. The dashed line represents a perfectly reliable forecast where the predicted probability (along the abscissa) exactly matches the observed frequency of occurrence (along the ordinate). The thick, solid lines in the figure mark the boundaries of no skill, such that points between the two lines contribute positively to forecast skill as measured by the Brier Skill Score (relative to the sample climatology, which in this case is about 5%). The points plotted in the diagram denote the performance of the SREF 3-hour calibrated thunderstorm guidance. The SREF guidance is skillful at all predicted probabilities. The graph in the upper left corner denotes the relative frequency distribution of predicted probabilities at 5% increments beginning at 0%.

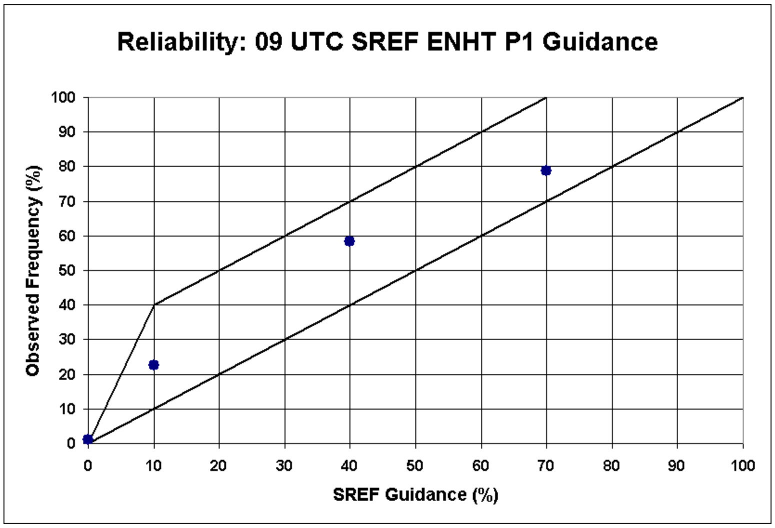

FIG. 3. Reliability diagram for the 0900 UTC SREF calibrated ENHT thunderstorm guidance for the 1500 UTC to 0300 UTC period (1 Dec 2007 through 30 Nov 2008). The solid lines are the reliability bounds for the ENHT stepwise probabilities.

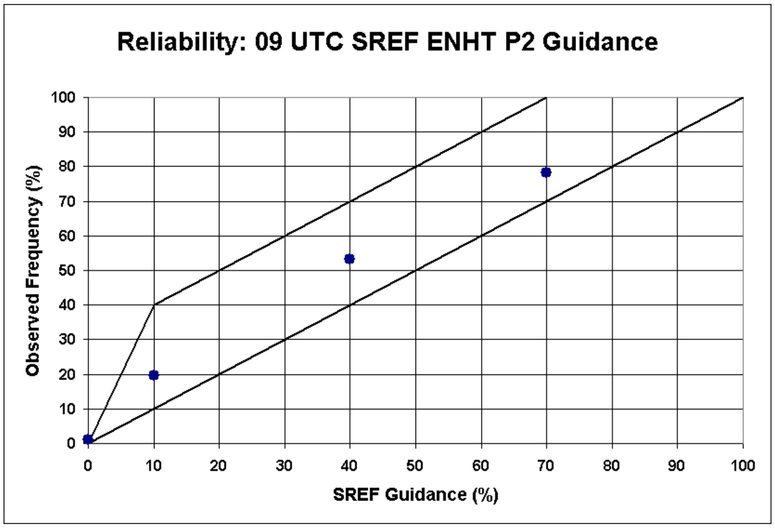

FIG. 4. As in Fig. 3, except for the 0300 UTC to 1500 UTC period.

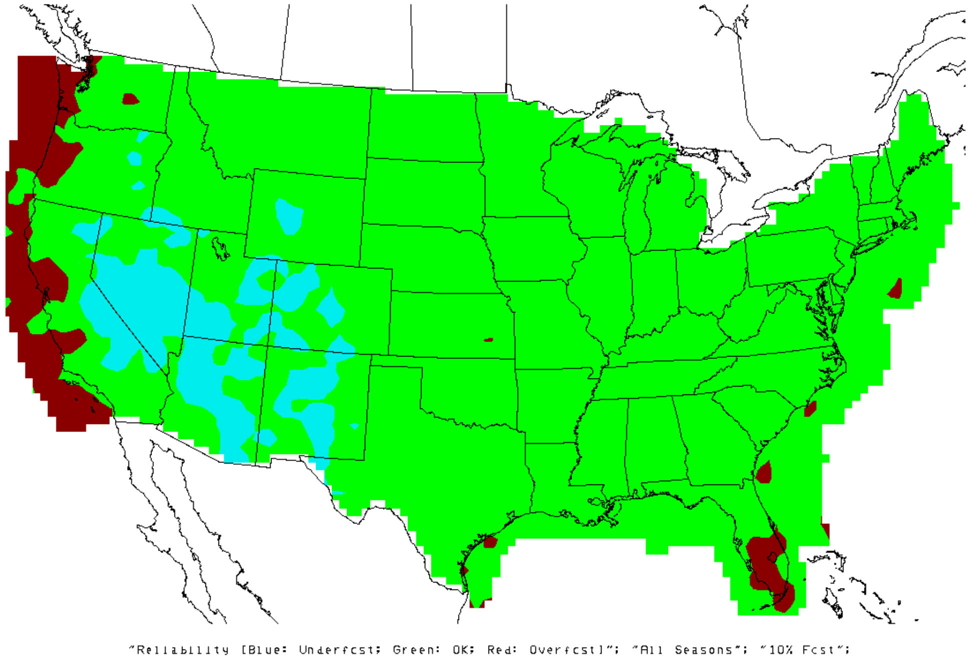

FIG. 5. The reliability of the SREF 12-hour calibrated thunderstorm guidance at 10% from the 0900 UTC run for the 1500 UTC to 0300 UTC forecast period (1 Dec 2007 through 30 Nov 2008). Green areas indicate locations where the reliability is within the expected bounds of ≥ 10% and < 40%. Red areas indicate a SREF over-forecast (high SREF bias) and blue areas an under-forecast (low SREF bias).

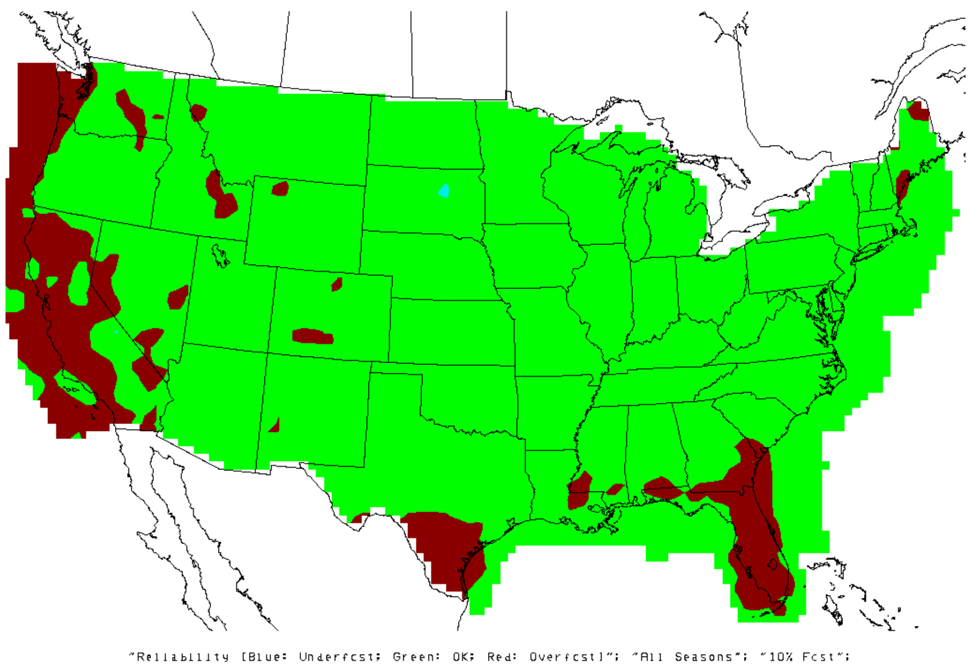

FIG. 6. As in Fig. 5, except for the 0300 UTC to 1500 UTC forecast period.

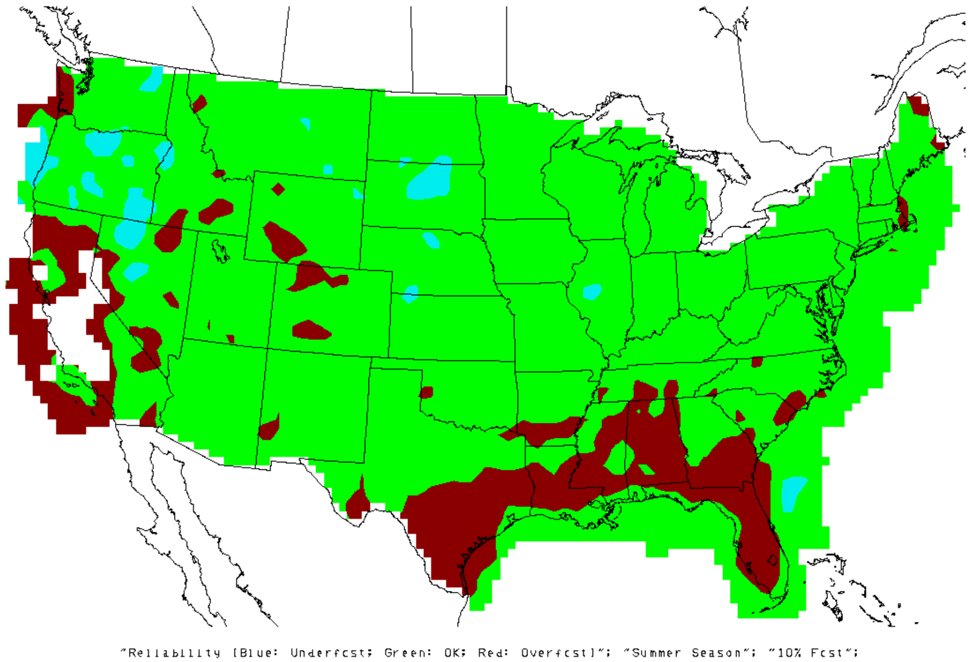

FIG. 7. As in Fig. 6, except for the summer season only (01 June 2008 through 31 August 2008).

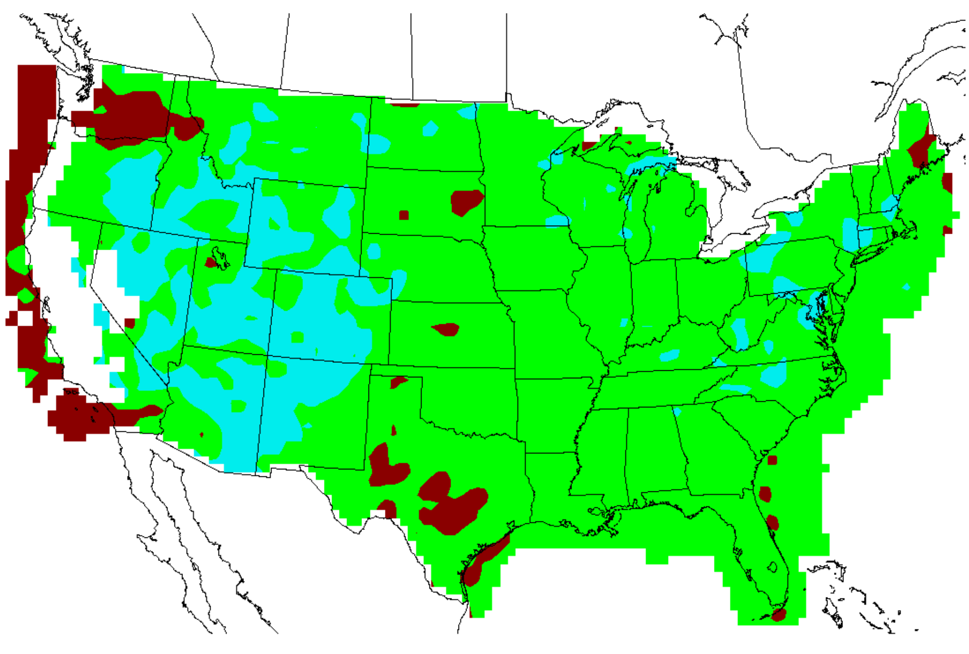

FIG. 8. As in Fig. 5, except the reliability of the 40% SREF 12-hour calibrated thunderstorm guidance.

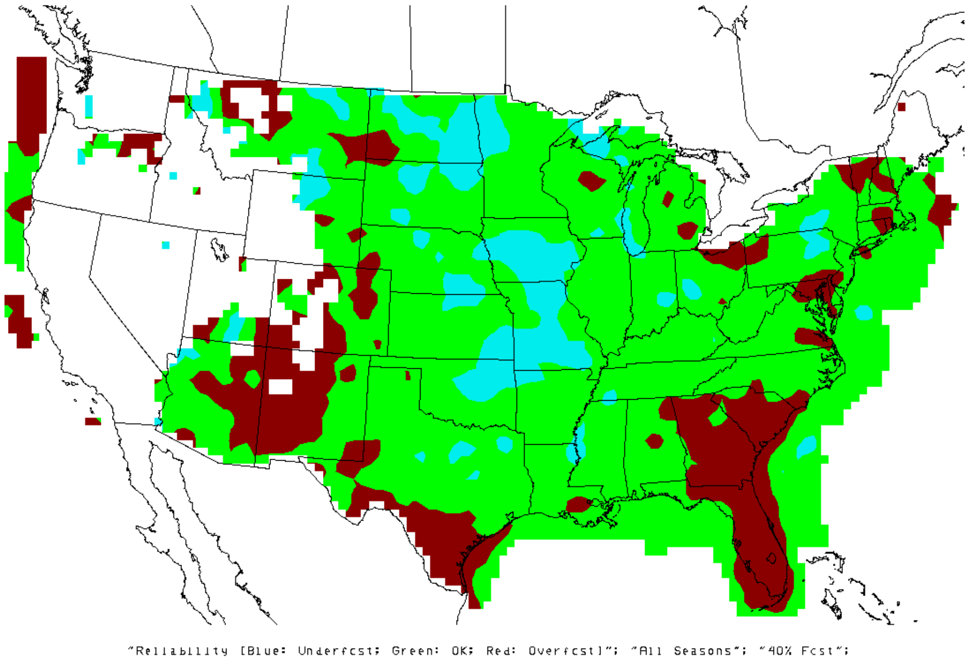

FIG. 9. As in Fig. 6, except the reliability of the 40% SREF 12-hour calibrated thunderstorm guidance.

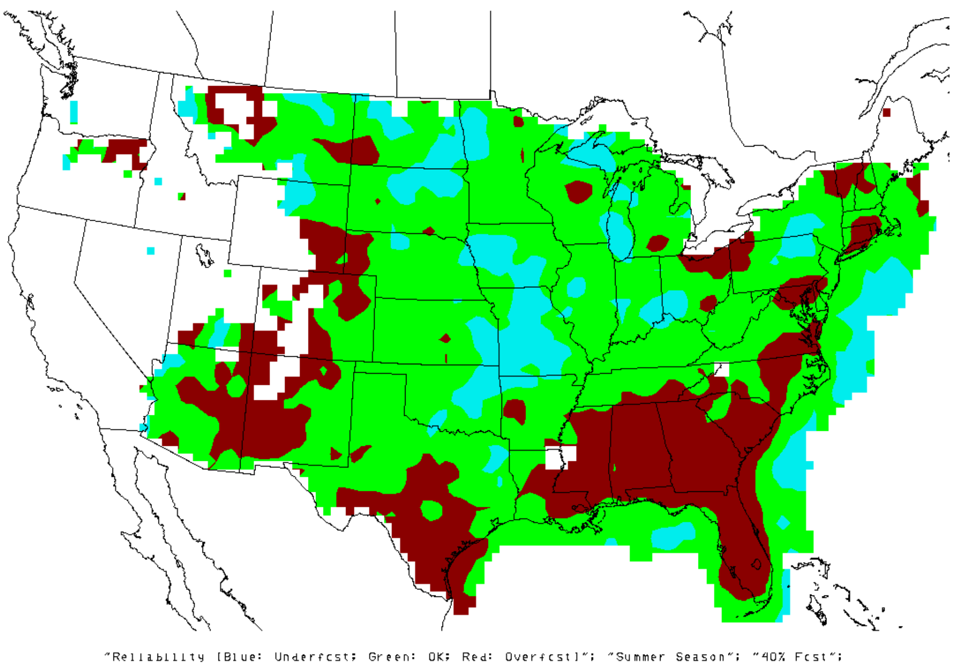

FIG. 10. As in Fig. 7, except the reliability of the 40% SREF 12-hour calibrated thunderstorm guidance.

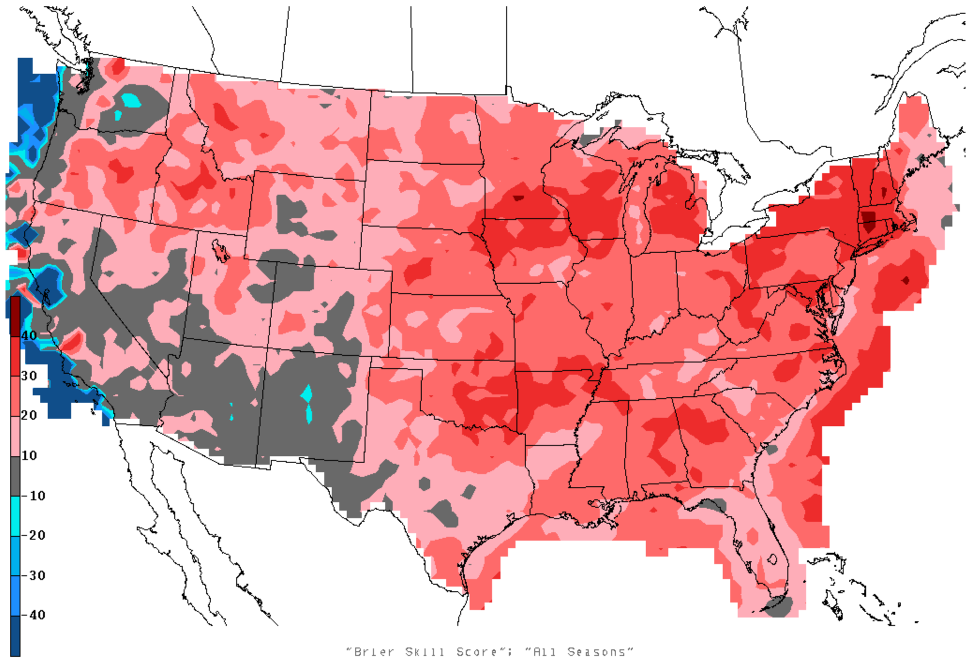

FIG. 11. The Brier Skill Score (BSS) improvement over climatology (%) of the SREF 12-hour calibrated thunderstorm guidance from the 0900 UTC run for the 1500 UTC to 0300 UTC forecast period (1 Dec 2007 through 30 Nov 2008). Warm colors are positive skill, cool colors negative skill, and gray neutral and near-neutral skill.

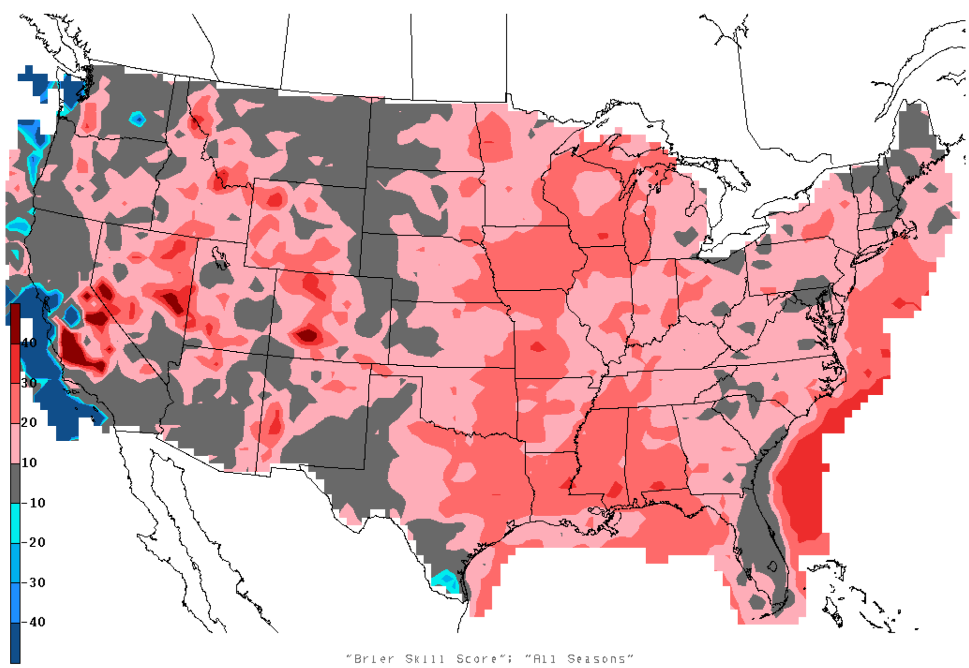

FIG. 12. As in Fig. 11, except for the 0300 UTC to 1500 UTC forecast period.

TABLE 1. Verification scores computed daily and averaged over the annual period. The bold line emphasizes the 0900 UTC SREF run as discussed in the text.

First Period (Day: 1500 UTC to 0300 UTC)

RUN	ISSUE TIME	# FCSTS	REL10	REL40	REL70	BS	ROC	HIT	MISS
21Z	0600 UTC	317	.86	.94	.78	.062	.849	21%	3%
03Z	1300 UTC	315	.83	.94	.77	.060	.855	23%	3%
09Z	**1630 UTC**	**306**	**.86**	**.97**	**.72**	**.061**	**.862**	**24%**	**3%**
15Z	2000 UTC	314	.84	.98	.76	.058	.876	26%	3%
ALL		1252	.85	.96	.75	.060	.860	23%	3%

Second Period (Night: 0300 UTC to 1500 UTC)

RUN	ISSUE TIME	# FCSTS	REL10	REL40	REL70	BS	ROC	HIT	MISS
21Z	0600 UTC	310	.93	1.03	.67	.050	.828	17%	2%
03Z	1300 UTC	305	.90	.97	.67	.049	.839	18%	2%
09Z	**1630 UTC**	**300**	**.92**	**1.01**	**.73**	**.049**	**.842**	**19%**	**2%**
15Z	2000 UTC	309	.89	1.05	.70	.047	.852	20%	2%
15Z	0100 UTC	308	.89	1.05	.70	.047	.852	20%	2%
ALL		1532	.91	1.03	.70	.049	.843	19%	2%