10.2 Investigation of an Automated Temporal Disaggregation Technique for Convective Outlooks during the 2012 Hazardous Weather Testbed Spring Forecasting Experiment

Israel L. Jirak[1*], Christopher J. Melick[1,2], Andrew R. Dean[1], Steven J. Weiss[1], and James Correia, Jr.[1,2]

[1]NOAA/NWS/NCEP/Storm Prediction Center
[2]CIMMS, University of Oklahoma
Norman, OK

1. INTRODUCTION

Storm Prediction Center (SPC) forecasters have demonstrated skill over the years in creating severe weather outlooks for the convective day (i.e., 12-12 UTC). Although increasing the temporal resolution of convective outlooks is desirable, the question arises about how to accomplish this task without requiring a significant increase in workload for the forecaster. An automated approach to address this topic was developed for the 2012 NOAA Hazardous Weather Testbed Spring Forecasting Experiment (SFE2012). The temporal disaggregation technique takes a forecaster-generated, longer-period total severe storm probability outlook and breaks it into shorter individual time periods using storm-scale ensemble model guidance as the input. The end result is an automated higher temporal resolution forecast of total severe weather probability consistent with the forecaster-generated, longer-period outlook.

The objective of this study is to investigate the feasibility and utility of a temporal disaggregation approach to the full-period human convective outlook. Section 2 will discuss the data and methodology used in the temporal disaggregation technique. A statistical comparison of the resultant automated forecasts to human-generated forecasts during the SFE2012 will be shown in Section 3. The last section will provide a summary of the findings.

2. DATA AND METHODOLOGY

2.1 Forecasts

Forecasts for the probability of severe convective weather (i.e., hail, wind, and tornado) within 25 miles of a point were made daily (M-F) for a movable mesoscale area of interest during the five-week SFE2012 (7 May – 8 June). In the morning, forecasts were manually drawn for the full period outlook (i.e., 16-12 UTC) and for three individual sub-periods (i.e., 20-00, 00-04, and 04-12 UTC). These shorter time periods match the operational thunderstorm outlook periods produced by SPC. The manual forecasts were made using traditional forecasting techniques by examining observational and model data to assess whether the mesoscale environment was favorable for the development of severe thunderstorms. That assessment was supplemented by examining the mode (e.g., Done et al. 2004; Weisman et al. 2008) and characteristics [e.g., hourly maximum fields (HMFs); Kain et al. 2010] of simulated storms produced by operational and experimental convection-allowing models (CAMs) and ensembles to arrive at a probabilistic forecast of severe weather.

The automated forecasts produced by the temporal disaggregation technique were constrained by the manual full-period forecast. This technique utilized calibrated severe guidance from the SPC storm-scale ensemble of opportunity (SSEO; Jirak et al. 2012) to break down the manual forecast into higher temporal resolution periods. Analogous to calibrated severe guidance from the SREF (Bright and Wandishin 2006), the calibrated severe guidance from the SSEO utilized a frequency-adjustment technique to arrive at the final calibrated probability. The unique aspect of the SSEO calibrated severe probabilities, however, is that neighborhood probabilities (Harless et al. 2010) of storm-attribute HMFs [e.g., updraft helicity (UH; Kain et al. 2008) and updraft speed] are utilized rather than environmental parameters (e.g., CAPE and shear), as is done with the SREF. Ultimately, the algorithm generated automated forecasts valid from 20-00, 00-04, and 04-12 UTC that were compared to independently created manual forecasts for the same periods.

2.2 Verification Data

The manual and automated forecasts were verified against the "practically perfect" technique (Brooks et al. 1998) applied to the preliminary hail, wind, and tornado reports. This approach involves using a 40-km radius of influence to each report and employing a 2-D Gaussian kernel estimation with a 120-km smoothing parameter. This results in a verification field that has similar appearance and characteristics of a typical severe weather outlook.

With the forecast and verification data on a common 40-km grid over the mesoscale domain of interest, a 2x2 contingency table was tallied to calculate several metrics including probability of detection (POD), frequency of hits (FOH), bias, and critical success index (CSI) for each forecast period at all probability thresholds (i.e., 5,15,30,45,60%). These metrics are concisely displayed on a performance diagram (Roebber 2009) to compare the manual and automated

* *Corresponding author address*: Israel L. Jirak, NOAA/NWS/NCEP/Storm Prediction Center, 120 David L. Boren Blvd., Norman, OK 73072; e-mail: Israel.Jirak@noaa.gov

forecasts. Similarly, the fractions skill score (FSS; Roberts and Lean 2008; Schwartz et al. 2010) was calculated and accumulated for each forecast during the SFE2012.

2.3 Methodology

There are several steps and checks involved in the temporal disaggregation process. The first step involves scaling the full-period calibrated severe SSEO guidance to the human forecast. This initial step ensures that the automated forecasts will be consistent with the manual forecast in terms of magnitude and location. A scaling factor is calculated at every grid point. The example shown in Fig. 1 reveals a smoother field in the manual forecast (Fig. 1a) with a mix of higher probabilities in some locations and lower probabilities in other locations than the SSEO calibrated severe forecast (Fig. 1b).

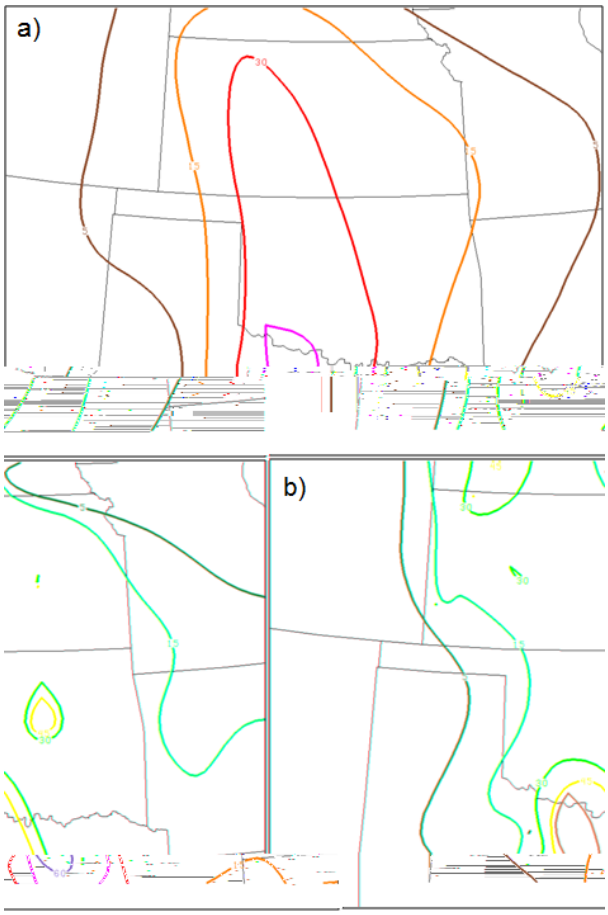

Figure 1. The full-period a) manual and b) SSEO calibrated severe forecasts valid from 1600 UTC on 30 May 2012 to 1200 UTC on 31 May 2012.

The next step is to apply this full-period, grid-dependent scaling factor to the SSEO calibrated probabilities for each sub-period (Fig. 2; 20-00 UTC). The scaled SSEO calibrated probability (Fig. 2b) adjusts the original SSEO calibrated probability (Fig. 2a) to more closely resemble the spatial distribution and magnitude of the full-period manual forecast (Fig. 1a). For example, the scaled probabilities in southwestern Oklahoma are higher than the original probabilities to more closely match the manual forecast.

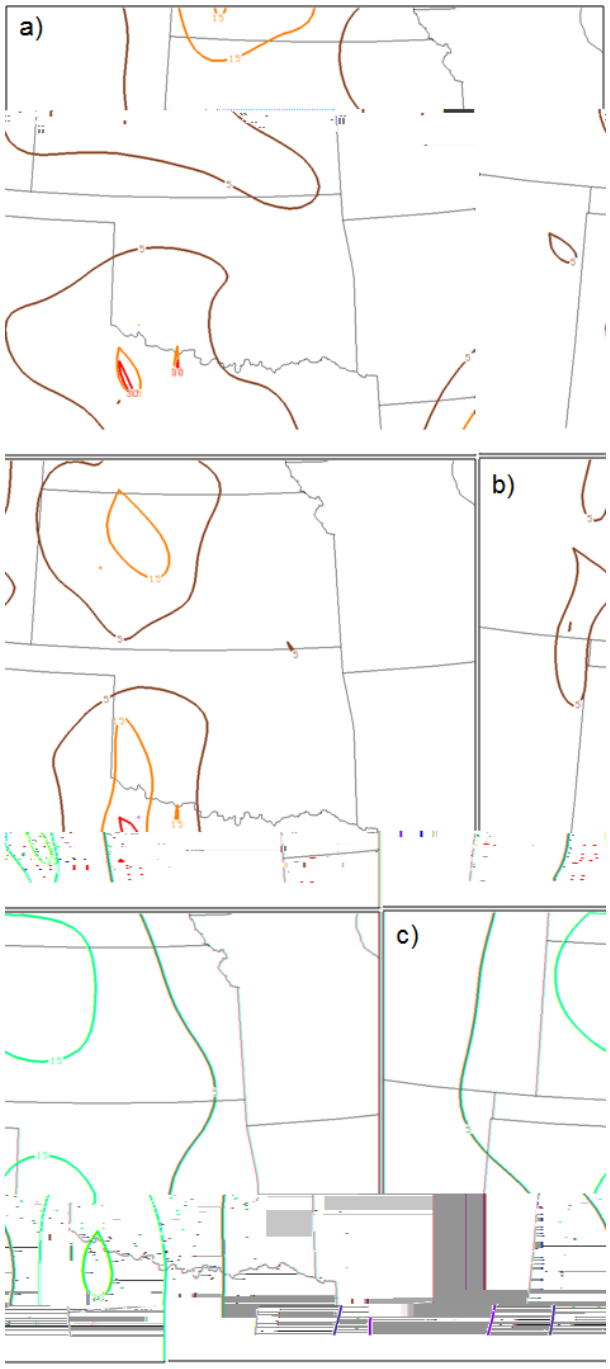

Figure 2. The sub-period a) SSEO calibrated severe, b) scaled SSEO, and c) automated forecasts valid from 2000 UTC on 30 May 2012 to 0000 UTC on 31 May 2012.

The scaled forecast could be used directly as the automated forecast; however, there are many instances where the scaled forecast is noisy (e.g., very small-scale features that are not representative of what a forecaster would draw). Thus, a couple of additional steps are

performed to smooth the automated forecast and ensure probabilistic consistency. Firstly, a 2-D Gaussian smoother with a smoothing parameter of 80 km was applied to all scaled forecasts to remove the small-scale artifacts that arose as part of the scaling process. Finally, a common multiplier is applied across all grid points and forecasts based on the maximum ratio of the full-period human forecast to the probability of severe in one or more of the periods. This step restores the peak magnitudes that were damped in the smoothing process. It is worth noting that other methods were explored in determining the value of this multiplier, and this method provided the best results. Finally, an additional check is performed to ensure that an individual period probability does not exceed the full-period manual probability to arrive at the final automated forecast for each period (e.g., Fig. 2c; 20-00 UTC).

3. RESULTS FROM SFE2012

The resultant automated forecasts generated by the temporal disaggregation procedure outlined above were compared to the independently generated manual forecasts for 20-00, 00-04, and 04-12 UTC on each day during the SFE2012. Figure 3 shows the verification results for the case discussed in the previous section. Interestingly, the automated forecast for this period (Fig. 3b) correctly picks up on two separate clusters of severe reports and better captures the eastward extent of the reports in Kansas, resulting in a higher FSS than the manual forecast (Fig. 3a). Notice that the CSI at the 15% probability threshold is lower for the automated forecast, as the practically perfect verification (Fig. 3c) is ≥15% between the gap in reports in northern Oklahoma and southern Kansas. These types of displays were available to participants in the SFE2012 to compare the automated and manual forecasts in next-day evaluations.

The overall performance of the manual, automated, and SSEO calibrated forecasts were compared during the SFE2012 (Fig. 4). One common trend among all forecasts was that the lowest CSI occurred during the overnight period (i.e., 04-12 UTC) when severe convective activity was relatively infrequent. At the 5% probability threshold (Fig. 4a), the manual forecasts had the highest POD and the calibrated SSEO had the lowest POD. While the automated forecasts had a lower POD than the manual forecasts, the FOH was slightly higher (i.e., lower FAR), resulting in similar values of CSI. The manual and automated forecasts had very similar performance characteristics at the 15% probability threshold (Fig. 4b) with a maximum in CSI during the 00-04 UTC period. At this threshold, the human is adding considerable value to the calibrated model output, which is lagging well behind in POD and CSI values, though it does have a bias near one. Although the sample size is smaller at the 30% probability threshold (Fig. 4c), the automated forecasts had noticeably higher POD than the manual forecasts. Again, the poorer performance from the calibrated SSEO indicates the importance of human input to the forecast.

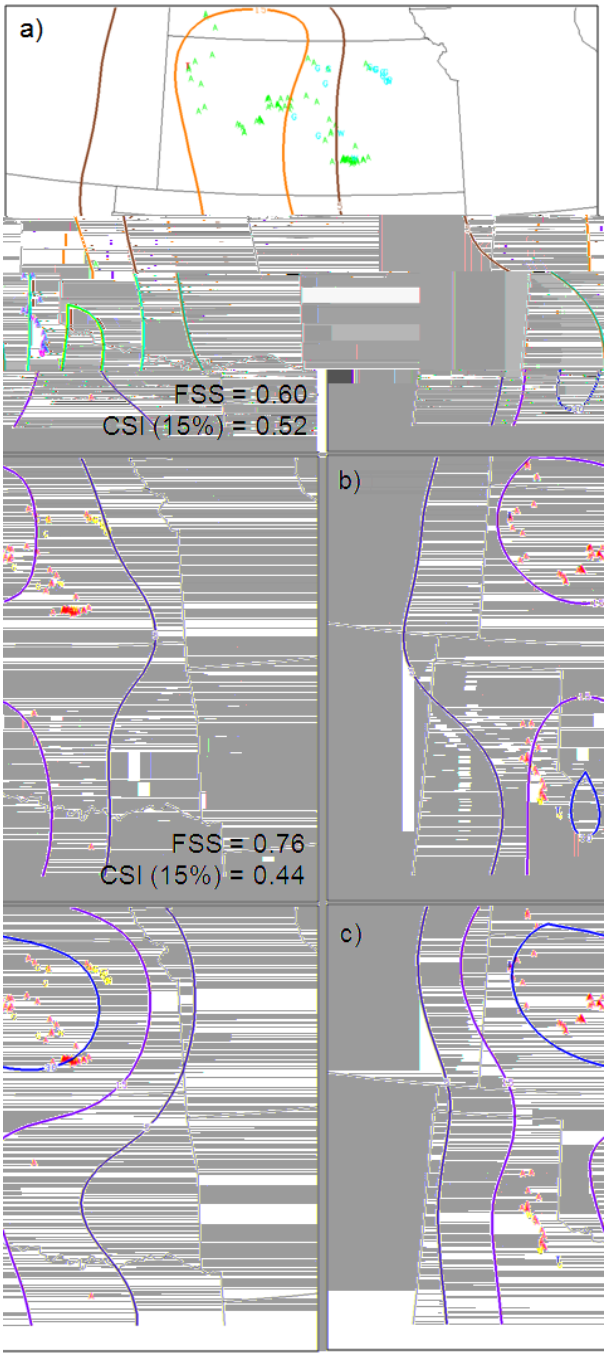

Figure 3. The sub-period a) manual and b) automated forecasts and c) practically perfect verification valid from 2000 UTC on 30 May 2012 to 0000 UTC on 31 May 2012. For the manual and automated forecasts, the FSS and CSI (at 15%) are indicated in the lower right portion of the respective panels. The preliminary storm reports from this 4-h period are shown (tornado – red, hail – green, and wind – blue).

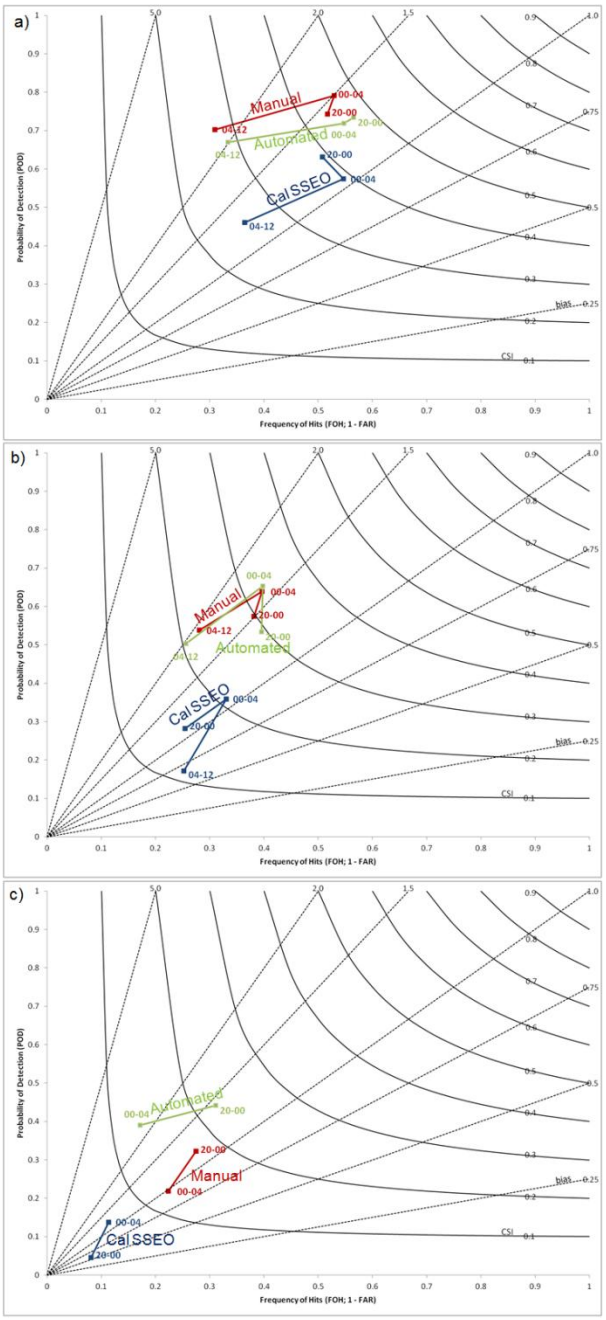

Figure 4. Performance diagram (Roebber 2009) at a) 5%, b) 15%, and c) 30% probability thresholds for manual (red), automated (green), and SSEO calibrated (blue) forecasts at the specified time periods (i.e., 20-00, 00-04, and 04-12 UTC) during the SFE2012.

In addition to calculating verification metrics from a 2x2 contingency table, the FSS was also calculated for all forecasts during the SFE2012 (Fig. 5). Overall, the FSS results are similar to those seen in the performance diagrams. The FSSs indicated that the automated forecasts were as good as, or even slightly better than, the manual forecasts for the 20-00 and 00-04 UTC periods. The overnight period (i.e., 04-12 UTC) was the only period during which the manual forecasts had a higher FSS than the automated forecasts. The automated and manual forecasts had higher FSSs than the calibrated SSEO forecasts for all periods, again revealing the value added by the forecaster to the process.

Figure 5. Accumulated FSS for manual (red), automated (green), and SSEO calibrated (blue) forecasts at 20-00, 00-04, and 04-12 UTC during the SFE2012.

The manual and automated forecasts were subjectively compared and rated by the SFE2012 participants (Fig. 6) prior to seeing the objective skill scores. The majority of automated forecasts were rated as "about the same" as the manual forecasts with approximately the same number of forecasts falling into the "better" and "worse" ratings bins. Thus, the subjective impressions of the participants were in agreement with objective verification results, which indicated similar overall skill between the automated and manual forecasts.

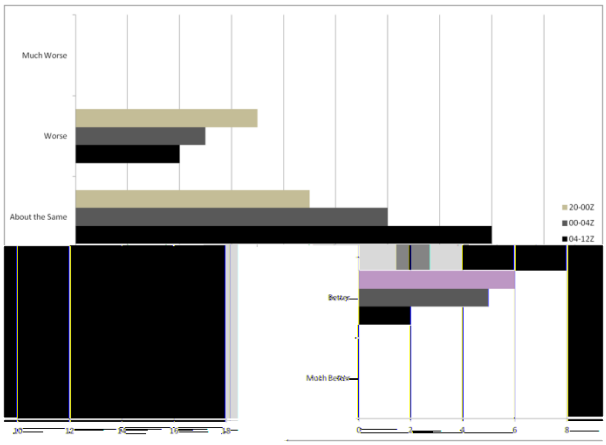

Figure 6: Subjective SFE2012 participant ratings of automated forecasts compared to manual forecasts from much worse to much better for each period: 20-00 UTC (brown), 00-04 UTC (gray), and 04-12 UTC (black).

4. SUMMARY AND CONCLUSIONS

Manually created full-period convective outlooks were temporally disaggregated using storm-scale ensemble guidance during the SFE2012 to generate higher temporal resolution (i.e. 4- & 8-h) convective outlooks. When these automated forecasts were compared to independently created manual forecasts for the same time periods, they fared favorably both in terms of objective metrics (e.g., CSI, FSS) and subjective impressions from participants during SFE2012. The fact that both the manual forecasts and the automated forecasts (constrained by the manual full-period forecast) performed better than the calibrated model guidance by itself clearly indicates the important role that the forecaster plays in this process. Given the promising results from the SFE2012, the SPC is planning to experimentally apply this technique to thunderstorm outlooks in the spring of 2013.

Acknowledgements. We would like to thank SPC forecasters and other participants from the SFE2012 who provided useful feedback for this evaluation.

REFERENCES

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

Brooks, H.E., M.P. Kay, and J.A. Hart, 1998: Objective limits on forecasting skill of rare events. *Preprints*, 19th Conf. Severe Local Storms. Minneapolis, MN, Amer. Meteor. Soc., 552-555.

Done, J., C. Davis, and M. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecasting (WRF) model. *Atmos. Sci. Lett.*, **5** (6), 110–117.

Harless, A.R., S.J. Weiss, R.S. Schneider, M. Xue, and F. Kong, 2010: A report and feature-based verification study of the CAPS 2008 storm-scale ensemble forecasts for severe convective weather. *Preprints*, 25th Conf. Severe Local Storms, Denver CO. Amer. Meteor. Soc., 13.2.

Jirak, I. L., S. J. Weiss, and C. J. Melick, 2012: The SPC storm-scale ensemble of opportunity: Overview and results from the 2012 Hazardous Weather Testbed Spring Forecasting Experiment. *Preprints*, 26th Conf. Severe Local Storms, Nashville, TN. Amer. Meteor. Soc., P9.137.

Kain, J. S., S. J. Weiss, D. R. Bright, M. E. Baldwin, J. J. Levit, G. W. Carbin, C. S. Schwartz, M. L. Weisman, K. K. Droegemeier, D. B. Weber, K. W. Thomas, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931-952.

Kain, J. S., S. R. Dembek, S. J. Weiss, J. L. Case, J. J. Levit, and R. A. Sobash, 2010: Extracting unique information from high resolution forecast models: Monitoring selected fields and phenomena every time step. *Wea. Forecasting*, **25**, 1536–1542.

Roberts, N. M. and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608.

Schwartz, C. S., and Coauthors, 2010: Toward improved convection-allowing ensembles: model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280.

Weisman, M. L., C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model. *Wea. Forecasting*, **23**, 407–437.