

Neil I. Fox, George L. Limpert, William T. Gilmore, Jose Miranda, Amy Schnetzler, Neville Miller, Ali Koleiny and Steven A. Lack

University of Missouri-Columbia, Columbia, Missouri

1. INTRODUCTION

Most radar-based nowcast systems delineate areas of precipitation, or cells, track these from image to image and generate a prediction of motion based on the observed path. However, the forecast motion is dependent upon the choice of precipitation area that is delineated. In particular, in a radar context, the choice of reflectivity threshold will impact the area and shape of the cell and its diagnosed track. Two issues arise from this, one positive and one negative. Firstly there is clearly a forecast sensitivity to threshold selection and this sensitivity needs quantification. Secondly, by running the same forecast sequences with different thresholds one can generate a simple ensemble and the ensemble forecast has the potential to provide a better forecast that also gives an indication of forecast confidence. A preliminary study of this type was presented by Lack et al. (2006) which showed that for a widespread stratiform rainfall event the ensemble nowcast comprising a suite of members, generated from varying threshold ranges within the Warning Decision Support System – Intelligent Information (WDSS-II), outperformed the individual ensemble members. This study extends that work to investigate if similar results were found for a variety of storm types. In particular nowcast storm tracks for convective storms are examined to see whether different threshold ranges may be appropriate for forecasting the motion of different types of storm.

Through variation of thresholds and calculation of categorical skill scores, it is possible to evaluate which thresholds provide the best forecast of storms. This may be useful to a forecaster in evaluating which thresholds to use when running the WDSS-II w2segmotion algorithm to provide the most useful information. There are a number of hypotheses that can be investigated using varying thresholds. Firstly we can see if the default threshold range is suitable for all situations. Second we can see if there is a lead-time dependence whereby different ranges have better performance at different forecast times. Additionally, we can test the hypothesis that the ensemble mean of the WDSS-II w2segmotion output will outperform the deterministic WDSS-II w2segmotion algorithm. Each of these can be achieved by comparing forecast performance by calculating categorical skill scores for the forecasts produced at each reflectivity threshold and comparing the results against the results for the ensemble mean.

2. METHODOLOGY

For this work the Warning Decision Support System – Integrated Information (WDSS-II) was used. Its core nowcast scheme is the WDSS-II w2segmotion algorithm which employs a K-means clustering technique (Lakshmanan *et al.* 2003) that generates pseudo-radar reflectivity forecast fields. Not only does WDSS-II w2segmotion take into account the motion of cells, but additionally it accounts for strengthening and weakening of quantities. Typically the parameter of greatest interest is reflectivity, but w2segmotion can be used to forecast a variety of quantities including vertically integrated liquid and cloud top temperatures from infrared satellite imagery. WDSS-II w2segmotion estimates wind fields by tracking clusters in successive images and diagnoses translation of features on multiple scales.

The WDSS-II w2segmotion also provides a tunable parameter which permits adjustment of the thresholds at which it estimates motion vectors. The allows one to select the range of reflectivity values which determine the cluster and diagnoses a motion field. Although all values of the quantity are translated, only values within the threshold range are used to determine the motion field. Variation of this parameter allows a user to decide whether higher or lower values of the quantity are most interesting when determining motion. When applied to reflectivity, meteorological features of stronger reflectivity tend to be smaller than features of weaker reflectivity. Smaller features often move more rapidly when compared to larger features.

By varying the thresholds, different motion vectors are generated because the algorithm detects the motion of different structures and therefore different portions of the radar image. Therefore, varying the range may be analogous to performing some kind of spatial cascade method. Statistics such as the mean and standard deviation can then be calculated on the output of the WDSS-II w2segmotion algorithm. This provides not only a mean outcome but also a measure of the uncertainty of the forecast.

By varying the choice of reflectivity range in the algorithm one can generate a limited ensemble of forecasts which can provide a further indication of forecast uncertainty or sensitivity and could generate outliers which would not be produced using a deterministic forecast.

The WDSS-II w2segmotion nowcasting system was used to examine the cases. Forecasts were created for a 60-minute lead time at 10 minute intervals. For each initial time, there are six forecasts produced, each with a 10-minute increment in lead time.

Several performance metrics were calculated on each of the forecasts. These include probability of detection (POD), false alarm rate (FAR), and critical

* Corresponding author address: Neil I. Fox, 302 ABNR Building, Univ. of Missouri-Columbia, Dept. of Soil, Environmental, and Atmospheric Science, Columbia, MO, 65211; email: foxn@missouri.edu

success index (CSI), which are presented here. While these metrics do not provide an in-depth view of forecast performance they do allow a basic assessment of which forecast run is most accurate. These metrics were plotted to demonstrate how forecast uncertainty increases with longer lead times and how algorithm performance decreases as lead times become longer.

3. CASE STUDIES

Six cases were examined for this paper, representing six different meteorological situations such that a preliminary conclusion could be reached regarding the importance of changing reflectivity threshold to the quality of forecasts in a variety of circumstances. In each case 1-km resolution radar data was used on either a 256 x 256 km or 512 x 512 km grid.

3.1 St. Louis, MO – 19th July 2006

Case 1 was a bowing MCS that passed through the St. Louis area on July 19th 2006. For this we used a number of times between 2300 UTC and 0100 UTC. This interesting MCS event occurred in the midst of a heatwave on July 19th. Earlier in the day, a well defined moisture gradient with much drier air to the west, existed over east-central MO stretching north into southeastern Iowa. The MCS formed along this pronounced moisture gradient in western Iowa and pushed across northern sections of Illinois during the early and mid-afternoon hours. The MCS intensified as it propagated southwest through west-central Illinois and eastern Missouri in the early evening.

However, the MCS began to bow out and intensify as it pushes into the St Louis metropolitan area. This was partially due to the presence of a strong rear inflow jet pushing south along the MCS. The main outflow boundary along with the associated MCS resulted in straight-line winds (on the order of 40 m/s) and downdraft.

A critical period of this event was studied as the storm passed through the St. Louis area and produced damaging downburst winds. In this case similar results were produced by each reflectivity range nowcast. However, as shown in Figure 1, the precise advection and structure of the MCS varied somewhat from forecast to forecast. At 40-minutes lead time one can see variations in the shape and location of the western part of the storm and the development on the northwest flank. The nowcasts captured the motion of the boundary well, although this is not verified as the reflectivity values are too low. This does provide confidence that a forecaster would assess potential development. By 60 minutes the forecast fields showed the bowing of the leading edge of the MCS indicative of the likelihood of downburst.

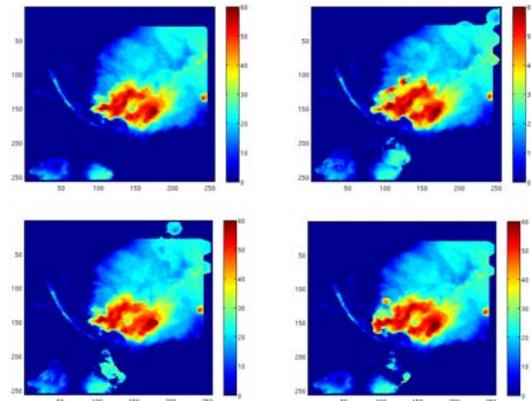


Figure 1: 30-minute forecasts from July 19th, 2006, forecast time 2330 UTC, for four runs using different reflectivity ranges. Top left is for a range of 30-60dBZ, top right is 20-40dBZ, bottom left is 20-60dBZ, and bottom right is 30-50dBZ.

3.2 San Marcos, TX – 13th January 2007

Case 2 was a multicellular training case that occurred in the San Marcos, TX region on January 13th 2007 and the times of interest were 1200, 1300 and 1400 UTC. The 13 January 2007 event produced more than 127 mm over the Austin/San Antonio region. The cells trained slowly over south-central Texas beginning around 1000 UTC and continued to intensify as the morning progressed. Around 1500 UTC portions of Austin experienced flash flooding, roads were shut down, and eight water rescues were made. Southwest of Austin, strong straight-line winds caused damage in San Marcos around 1300 UTC.

Due to the slow and, in some instances, retrograde propagation of the cells this was considered a difficult case. Figure 2 shows two complete 60-minute forecast runs using different reflectivity ranges. It is observed that the two forecasts produce significantly different locations of significant precipitation over the course of the forecast period. The nowcast based on the motion of the 30-60-dBZ range has the storm significantly further north than that of the 40-70-dBZ range. This is due to the prior motion of structures in the 30-60 dBZ range capturing a more representative motion of the storm system as a whole. This is to be expected in such a case when severe weather like large hail, which would produce higher reflectivities is not a significant factor.

The categorical statistics shown in Figure 3 show that the forecast based on the 30-60-dBZ range produced a much better and consistent product than that based on the 40-70-dBZ range.

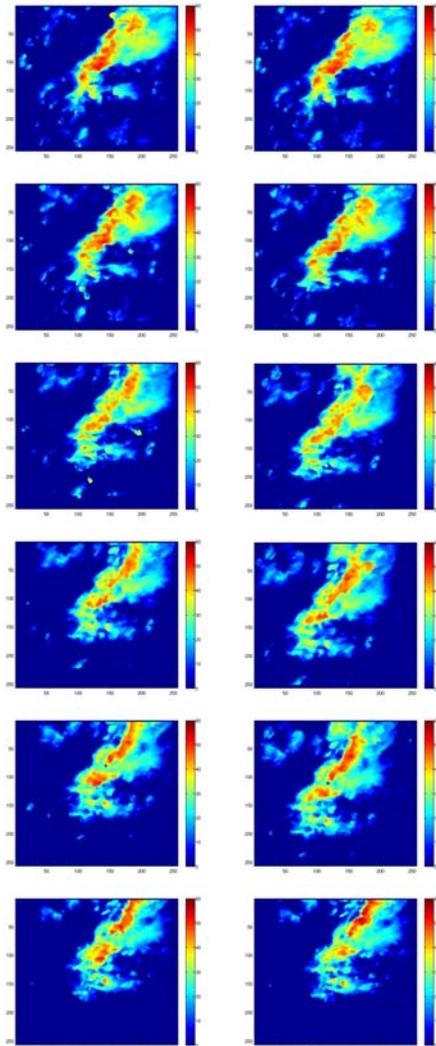


Figure 2: Forecast sequences from 1300UTC 13 January 2007. Left column are 10-minute interval forecasts out to T+60 Minutes using the 30-60-dBZ reflectivity range. The right column is the same except using the 40-70-dBZ range.

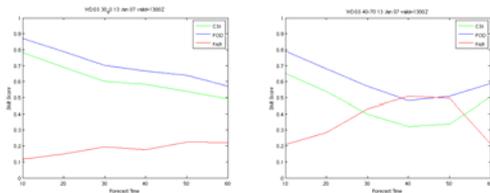


Figure 3: Categorical verification statistics for the forecast sequences shown in Figure 2. The left-hand figure shows the performance of the 30-60-dBZ range forecast and the right-hand figure shows that for the 40-70-dBZ range.

3.3 Florida – 2nd February 2007

Case 3 involved a line of supercells that passed through central and northern Florida on February 2nd 2007. A line of low-topped severe thunderstorms developed along a cold front in northern and central

Florida between 0800 UTC and 1000 UTC. A supercell developed in the southern portion of the linear structure of cells just off the westward, Gulf-facing coast of Florida.

For this case the ensemble mean forecast generally outperforms the individual members, but, as appears common, a single member does the best. As can be seen in figure 4, there are significant structural differences in the forecast reflectivity fields. The mean forecast appears to smooth some of the detailed structure and does not retain the high-intensity reflectivity cores visible in the individual forecast runs. This, however, results in better categorical statistics than the single runs. This seems to be a similar effect to that observed in using the Spectral Prognosis nowcast scheme (Seed 2003) which employs a spatial cascade technique.

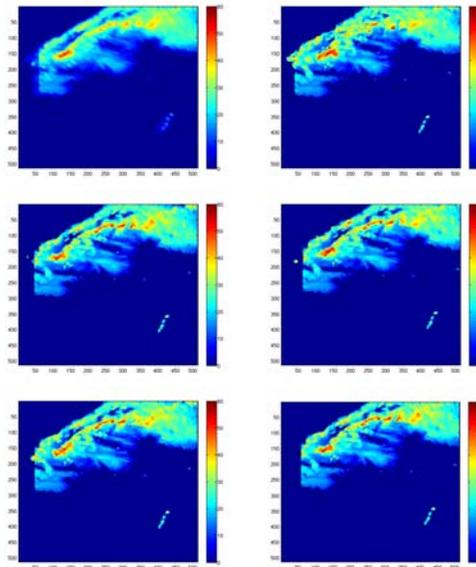


Figure 4: 60-minute forecasts from 2 February 2007 showing the variation between all the different forecast runs. Top left uses the ensemble mean field, while top right is 20-40 dBZ, middle left is 20-60 dBZ, middle right is 30-60 dBZ, bottom left is 30-50 dBZ and bottom right is 40-70dBZ.

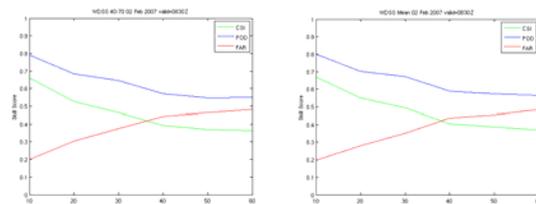


Figure 5: A comparison of the categorical statistics for the mean forecast to the 40-70-dBZ forecast.

3.4 SW Missouri – 4th May 2003

Case 4 involved a supercell tornado outbreak in northeast Oklahoma and southwest Missouri on May 4th 2003. The data from the three sites, Springfield, MO (KSGF), Fort Smith, AR (KSRX), and Tulsa, OK (KINX),

were merged, and processed into 10-minute time steps. This event encompasses multiple storm types, linear and single-cell convection, along with splitting and merging, and the decay and growth of convective cells. In figure 6 there are shown a range of 40-minute lead time forecasts from 2240 UTC. It can be seen that each nowcast run produces a slightly different location for each isolated supercell. This provides a range of possible paths of severe weather and could allow a forecaster to determine the uncertainty in a projected storm motion.

The accompanying verification statistics (shown in figure 7) indicate that although two superficially similar reflectivity ranges, in this case 20-40dBZ and 30-50dBZ) can produce significantly different forecasts with different performance levels.

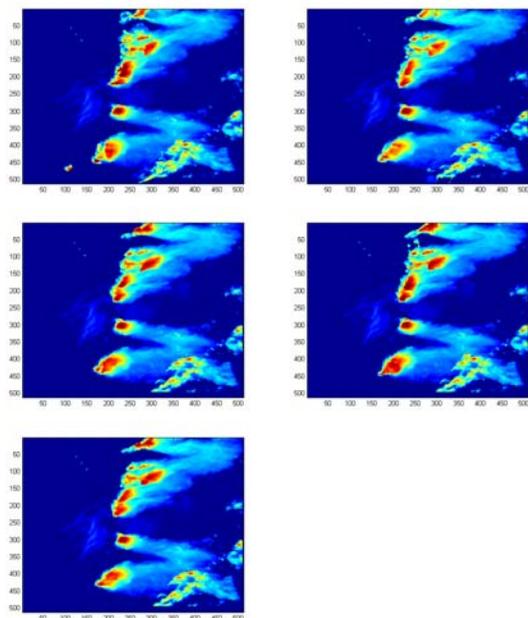


Figure 6: Five versions of the 40-minute forecast from 2250 UTC on 4th May, 2003. Top left is 20-40 dBZ, top right is 20-60 dBZ, middle right is 30-60 dBZ, middle left is 30-50 dBZ, bottom left is 40-70 dBZ.

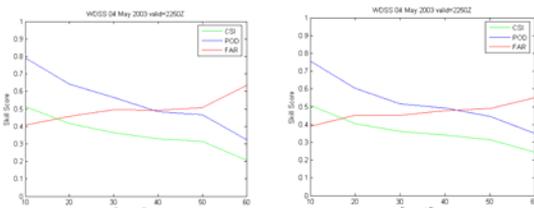


Figure 7: Verification statistics for 20-40 dBZ and 30-50 dBZ.

3.5 Eastern Missouri – 10th June 2006

On 10 June 2006 a series of convective boundaries propagated east across sections of Kansas, Missouri, and Illinois, producing various reports of tornadoes, hail, flooding, and wind damage. Between 2000 UTC 10

June 2006 and 1100 UTC 11 June 2006, at least three separate severe convective episodes raked the region. The first convective episodes featured mostly discrete supercells moving east across eastern sections of Missouri and western sections of Illinois from 2000 UTC 10 June to 0300 UTC 11 June which produced a possible tornado over Interstate 70 in Eastern Missouri. The second convective episode from 0300 UTC 11 June to 0800 UTC 11 June featured a south-southeasterly moving squall-line with bow echoes which fed off the rear-flank of the supercells of the first episode, causing widespread damage and reports of tennis-ball sized hail in Pattonsburg, MO. The third convective episode from 0800 UTC 11 June to 1100 UTC 11 June featured multicellular storms originating from a convective cold-pool from the second episode across central Missouri. 24-hour rainfall reports of 50 mm to 125 mm were common across central and eastern Missouri, with reports of flash flooding in Callaway, Boone, Macon, Audrain, and Monroe counties in Central Missouri.

Again, in this case there were differences between the forecast tracks produced by the different threshold ranges. The ensemble produced better performance statistics than most of the individual members.

3.6 Hurricane Andrew – 24th August 1992

Case 6 involved a period from the landfall of Hurricane Andrew on August 24th 1992. The period used was 0500 and 1200 UTC.

Although the forecasts produced generally are not poor forecasts, there are some notable issues. In particular, there is often poor continuity between forecasts from one initial time to the next. Not only is the positioning of important features affected, but also the intensity. The intensity of storms appears to be extrapolated, perhaps with some limit on the maximum intensity. If one forecast extrapolates a general strengthening trend while the next extrapolates a weakening trend, a significant difference may be evident for long lead times. Additionally, many important features of the storm are often deformed if they are still present.

Generally, POD tended to be high because of the large area of precipitation with a slow motion around the eye. Typical hurricane motion is often around 20 kilometers per hour, and so although individual storms and rain bands may move within the rain field, the motion of the overall rain field tends to be slow. So, although the motion vectors may be inaccurate, it is possible that categorical skill scores will still be high. This also explains the little difference in POD and CSI, even for the ensemble members which performed the worst in the qualitative analysis of the nowcast. Indeed, the motion of the eye and eyewall was often faster or slower than was observed for some ensemble members, leading to sometimes significant errors at longer lead times. Although it is not included in the quantitative analysis, later forecasts around 11 UTC projected a track of the eye and eyewall to the west-northwest while the observed motion was almost due west. This is poorly captured by the standard categorical

skill scores but is certainly an important consideration in evaluating the quality of the nowcast produced.

The ensemble mean forecast generated from the five member ensembles performed well. Its performance quantitatively was slightly better than the performance of the best of the ensemble members based on CSI. Figure 8 illustrates the performance of the ensemble mean as assessed by categorical skill scores. In a qualitative analysis, the ensemble mean outperformed the individual ensemble members and provided a forecast which would be more useful to a forecaster interested in the track of the center of circulation.

For most of the lead times out to 60 minutes, the ensemble mean forecast maintained an eye structure which generally showed little deformation and little filling. A well defined eyewall was present in the ensemble mean, even at a lead time of 60 minutes. The position was reasonably close to the observed position of the eye at 0900 UTC, even at the longest lead times.

However, the ensemble mean was less accurate in some areas than the deterministic forecasts produced by each ensemble member. For example, many of the higher reflectivity values within the large area of rain around the center of the storm have been smoothed out in the ensemble mean. Additionally, the reflectivity values of the more isolated storms in the outer bands have been greatly diminished and do not correspond with what was observed. Although the ensemble mean does provide a good forecast of location, it does not provide a good forecast of intensity.

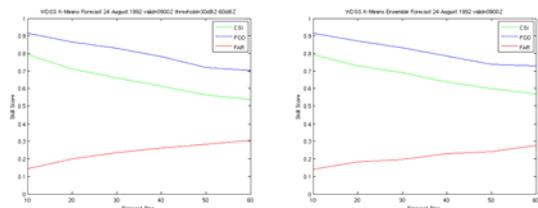


Figure 8: Verification statistics for one sample forecast and the ensemble mean for Hurricane Andrew.

4. DISCUSSION

The primary finding of this limited study is that in each case selection of the reflectivity range on which to base a forecast of storm location can impact the projected path of the storm. This is most clearly observed when the storms are isolated. The different

In most cases the ensemble mean forecast performs better than the individual members. However, there is generally a single member that provides the best forecast as measured by the simple categorical measures. On the other hand there is little consistency regarding which member will be the best choice. This may be a function of the particular case, or it may be a product of the storm type and reflectivity values exhibited by the areas of the storm that dominate the observed motion. If the latter is the case, as is plausible, then it is possible that one can find the optimum reflectivity range on which to base motion estimates, based on an objective determination of storm type. This

will be the focus of further work as this study has only limited cases of a small variety of storms.

The ensemble mean, naturally produces smoother forecast fields with generally lower intensities. The resultant field is similar in appearance to those produced by spatial cascade methods. In effect the varying of the threshold range can produce a spatial cascade as higher reflectivity areas tend to be smaller. But in this case the cascade is performed in an object-oriented fashion, rather than on a fixed grid, which might produce a more natural cell size response. Meanwhile, the entire ensemble, if performed, can provide a measure of forecast sensitivity and divergence, thereby giving a forecaster knowledge of the confidence that should be placed in a particular forecast.

Only a small, number of possible reflectivity ranges were chosen for this study. It would be possible to select any number of combinations of upper and lower limits to the range of reflectivities and produce a much larger ensemble in this manner. Alternatively it may be possible to discover, through repeated trials, those threshold ranges that capture the important aspects of the cell motion for particular storm types.

In the future we plan to look at more cases to provide robust measures of threshold sensitivity and ensemble performance. We also plan to use more descriptive verification measures to assess the performance.

5. REFERENCES

- Lack, S.A., G.L. Limpert and N.I. Fox, 2006: Alternative approaches to current nowcasting schemes. Preprints, 23rd Conference on Severe Local Storms. P5.1. CDROM, Amer. Met. Soc., Boston, MA.
- Lakshmanan, V., R. Rabin, and V. DeBrunner, 2003: Multiscale storm identification and forecast. *J. Atm. Res.*, **67-68**, 367-380.
- Lakshmanan, V., T. Smith, G. J. Stumpf, and K. Hondl, 2006a (In Press): The warning decision support system - integrated information (WDSS-II). *Weather and Forecasting*.
- Seed, A.W., 2003: A dynamic and spatial scaling approach to advection forecasting. *J. Appl. Met.*, **42(3)**, 381-388.

