Verification of Severe Weather Proxies from the NSSL-WRF for Hail Forecasting

NATHAN A. WENDT*

Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma and NOAA/National Weather Service/Storm Prediction Center, Norman, Oklahoma

ISRAEL L. JIRAK

NOAA/National Weather Service/Storm Prediction Center, Norman, Oklahoma

CHRISTOPHER J. MELICK

Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma and NOAA/National Weather Service/Storm Prediction Center, Norman, Oklahoma

ABSTRACT

With convection allowing models (CAMs) becoming more prevalent and finer in resolution, expansive datasets exist for information extraction of simulated storm attributes. Because convective processes are becoming increasingly resolved in CAMs, the question is whether useful forecast information can be extracted to improve upon existing severe weather proxies. First of all, however, the skill of existing severe weather proxies for hail forecasting must be established. The NSSL-WRF is a 4 km CAM that is used operationally at the Storm Prediction Center (SPC). While some forecast verification studies have been performed with the NSSL-WRF, none of the efforts have focused specifically on severe hail prediction. The current work will focus on providing a long-term (2012-2015) neighborhood forecast verification of the NSSL-WRF for severe hail using local storm reports and radar-derived maximum estimated size of hail (MESH) as the observational datasets. Overall, verification using MESH performs better than with LSRs. Long-term verification statistics show that maximum UH provides the most skillful severe hail forecast. The utility of all severe hail proxies is greater in the spring than in the summer; however, maximum UH is the most consistent scoring proxy across seasons. Maximum UH also matches observed severe hail probabilities better than other proxies using the fractions skill score metric. Modifications to the NSSL-WRF to include maximum vertical velocity below 100 mb and minimum UH have produced skillful forecasts in some case studies.

1. Introduction

Hail contributes to a substantial portion of the total insurance damages for crops and other property in a given year (Changnon et al. 2009). Because of the potential societal impact of severe hail ($\geq$ 25.4 mm, or 1 in.), the ability to make skillful forecasts for the timing and location of severe hail is of great importance to operational forecasters. Extracting information from forecast models to help improve forecasts has been done in many contexts (Kain et al. 2010; Jirak et al. 2014; Melick et al. 2014). With the increase of higher resolution convection-allowing models (CAMs), there exists an even greater need to quantify the skill of the information extracted and to develop new ways to extract more useful information.

The NSSL-WRF is a 4-km CAM that is used at the Storm Prediction Center (SPC). While some effort has been made to extract information from this model, much of it has focused on ways to improve forecasts of *all* severe hazards. No specific efforts have been made, thus far, to quantify the skill of the existing severe hail proxies in the NSSL-WRF. The goal of this study is to provide a long-term, objective, neighborhood forecast verification of severe hail proxies from the NSSL-WRF and to determine the best overall proxy for severe hail. This information should allow for optimum, targeted information extraction from other existing CAMs as well. Methods and data are described in section 2, results of the verification are discussed in section 3, and concluding remarks in section 4.

2. Methods

a. Data

The severe weather proxy data for this study was extracted from the NSSL-WRF model from 2012–2015. The NSSL-WRF is a short term (36-hour) convection-allowing forecast model that is initialized twice daily at 0000 UTC

*Corresponding author address: CIMMS, University of Oklahoma, 120 David L. Boren Blvd., Norman, OK.
E-mail: nathan.wendt@noaa.gov

and 1200 UTC and is used in operations at the Storm Prediction Center. For this work, only data from the 0000-UTC initialization are considered. The NSSL-WRF domain spans the entirety of the CONUS at a grid spacing of 4 km. Some pertinent configuration options of the NSSL-WRF can be seen in Table 1. Verification was done for each *convective day* (1200 UTC – 1200 UTC) which means data from forecast hour 12 (F12) through forecast hour 36 (F36) are aggregated. The aggregation consists of taking the maximum value at each grid point throughout the F12 to F36 period and generating a daily maximum field.

A majority of the verification in this study uses the Maximum Estimated Size of Hail (MESH; Witt et al. 1998) product from the NSSL Multi-radar/Multi-sensor (MRMS; Smith et al. 2016) system as the observational "truth". The data are on a 1 km grid that spans the entirety of the CONUS and are from 2012–2015. As with the NSSL-WRF data, the MESH are also aggregated into daily maximum fields after being quality controlled, as described in section 2b. Local storm reports (LSRs) from the same time period were also used as another observational data set for comparing with MESH.

b. MESH quality control

To ensure that MESH data were legitimate, a similar quality control procedure as in Melick et al. (2014) was used. Hourly MESH data were used in this step to ensure that, in particular, the lightning quality control step would not match flashes with storms from a different period in the day—which would be the case if daily maximum fields were used. First, a Gaussian smoother with a σ value of 3 grid cells is applied. This smoothed MESH field is then used as a mask on the raw MESH field to eliminate isolated pixels. The next step involves using quality-controlled National Lightning Detection Network (NLDN) data to determine if MESH pixels are associated with a storm. To do this, those MESH pixels that fall within 40 km of a detected flash are deemed legitimate. MESH values are then removed below 29 mm as this magnitude has shown the best skill in coinciding with severe hail reports (Cintineo et al. 2012). MESH values above 127 mm (5 in.) were also removed as there is evidence to support those values being very rare and likely spurious (Cintineo, 2016, personal communication; Blair et al. 2011). The final step involved interpolating the quality-controlled MESH data onto the 4-km NSSL-WRF grid. This was done using a bi-linear interpolation algorithm in the ESMPy Python package (v7.0.0; Earth System CoG 2016). Daily maximum MESH fields were created using the quality-controlled hourly fields.

c. Verification

For this study, the NSSL-WRF variables used as proxies for severe hail are 1-km AGL simulated reflectivity, HAILCAST (Adams-Selin and Ziegler 2016), maximum 2–5 km AGL updraft helicity (UH), maximum vertically-integrated graupel (VIG), and maximum updraft speed (below 400 mb), and 250 mb updraft speed. It is important to note here that the 250 mb updraft speed field is not an hourly maximum field; however, the other selected variables are hourly maximum fields. The reason for this will be discussed in the results.

The verification metrics used in this study are all neighborhood statistics (Gilleland et al. 2009), as this is more appropriate for high resolution grids. The neighborhood radius of influence (ROI) can be selected by the user, but has traditionally been chosen to be 40 km to remain consistent with the SPC probabilistic forecast definitions. However, one goal of this study is to maximize the forecast skill of the severe hail proxies. To do this, a change in the forecast field ROI is tested to investigate the optimum ROI for severe hail proxies in the NSSL-WRF. For this study the ROI ranges from 0 km (i.e., grid-point-to-grid-point) to 200 km by an interval of 20 km, similar to Harless et al. (2010). It should be noted that the ROI used for the observations (i.e., MESH and LSRs) always remains at 40 km. With these given ROIs, neighborhood maximum values are calculated and used to build contingency tables and calculate performance metrics (Roebber 2009). In order to create the contingency table, a range of appropriate threshold values of the severe hail proxy variables need to be chosen. All values above the thresholds signify that a severe hail event is forecasted at that grid point. Care was taken to choose thresholds that spanned the range of data as well as POD values on the performance diagram. Where possible, this was achieved by selecting roughly equally-spaced percentile values within the data.

Another metric used is the fractions skill score (FSS; Roberts and Lean 2008). The use of this score better incorporates spatial density of the forecasted and observed severe hail events. Given the emphasis on spatial density, FSS scores in this study use the traditional 40 km forecast neighborhood to remain consistent with the SPC probabilistic forecast definitions (i.e., severe weather occurring within 25 mi. of a point). In this way, a particular proxy variable at some threshold for severe hail may better capture the spatial extent and magnitude of an event even without scoring the highest on the performance metrics. To calculate these scores, the practically perfect (PP) methodology (Hitchens et al. 2013) was applied to the forecast and the observation fields as was done in Melick et al. (2014). Using the FSS algorithm, these PP fields from forecasts and observations are used to calculate a skill score.

3. Results & discussion

The results of varying the forecast ROI can be seen in Fig. 1. Here, 2–5 km AGL maximum UH is used to calculate the performance metrics. UH was chosen as it was found to be the most skillful severe hail proxy through this work (discussed later). What is evident is that results are quite similar at ROI values at or above 40 km, especially for lower UH threshold values. The highest CSI values occurred for the 80 km ROI at the 70 m^2s^{-2} threshold. Because of this finding, the performance diagrams shown in this study use an 80-km ROI for the forecast neighborhoods.

As mentioned in section 2c, the maximum updraft speed output by the NSSL-WRF is set to only be below 400 mb—this is also true for all other experimental and operational CAMs. The NSSL-WRF initialized at 0000 UTC on 25 June 2015 was used as a sample case to see the vertical structure of the vertical velocity. Grid points were chosen if the 2–5 km AGL UH was in exceedance of 70 m^2s^{-2} (a value chosen based on results that will be discussed in section 3a). For each grid point chosen, the vertical profiles were combined to create the vertical distribution of vertical velocity. Those results are seen in Fig. 2. Interestingly, average maximum vertical velocity occurs above 400 mb for the storms in this simulation. This indicates that the maximum updraft speed in simulated storms often occurs above 400 mb and that the NSSL-WRF, as currently configured, will miss the true maximum value in many cases.

a. Performance diagrams

When you compare MESH-based and LSR-based performance metrics, seen in Figs. 3 and 4, respectively, it is evident that MESH-based verification results in higher scores. However, the differences are not incredibly large. This increase in MESH-based performance metrics is due to the increased spatial coverage of MESH data over that of LSRs. The subsequent verification calculations in this study are done using MESH as the baseline for that reason.

Some interesting results appear when the analysis is broken down by season. The two seasons focused on here are meteorological spring (MAM) and summer (JJA). Figs. 5 and 6 show the same MESH-based performance analysis, but broken into spring and summer periods, respectively. Seasonally speaking, severe hail proxies do better in the spring than in the summer. The only exception being that hourly maximum 1 km AGL simulated reflectivity does just as poorly in both seasons. The other notable feature in these plots is that 2–5 km AGL UH is the most consistent proxy, in terms of forecast skill, across both seasons. During the fall (not shown) performance is similar to that of summer. Performance during the winter (not shown) is the worst of all seasons, but this may be an artifact of the sample size; i.e., proxy variables do not reach severe hail thresholds nearly as often in the cold season.

From these performance diagrams we can gain information about what proxy variable, and at what threshold, is the most skillful for predicting severe hail in the NSSL-WRF. The obvious variable that stands out is 2–5 km AGL UH. More specifically, the 70 m^2s^{-2} threshold level gives the highest CSI values overall.

b. Fractions skill score

FSS plots for the full 2012–2015 period, spring season, and summer season are shown in Figs. 7, 8, and 9, respectively. For FSS, a score above 0.5 is generally considered skillful. Once again, 2–5 km AGL UH is the best performer both overall and across seasons. The 70 m^2s^{-2} threshold also performs better than other UH thresholds as well. Even though it is not an hourly maximum field, 250 mb updraft speed still is able to reasonably match observed spatial probabilities and verfies as the second best hail proxy in terms of FSS.

c. NSSL-WRF modifications and case study

As was shown in Fig. 2, the maximum vertical velocities in simulated storms frequently occur above 400 mb. With this in mind, modifications were made to the NSSL-WRF code base to extract the maximum vertical velocities up to 100 mb. Furthermore, the NSSL-WRF only calculates positive 2–5 km AGL UH which will neglect all anticyclonic mesocyclones from left-moving supercell thunderstorms. Given that anticyclonic, left-moving storms are known to produce severe hail (Edwards and Hodanish 2004), another modification was made to the NSSL-WRF to calculate negative 2–5 km AGL UH values.

To test how well these modifications would perform, a case study was run using the NSSL-WRF initialized at 0000 UTC for 08 May 2016. This day was selected as several strong left-movers were observed and it allows for the investigation of the impact of adding negative UH values. For 2–5 km AGL UH, a maximum, minimum, and full field was produced and compared to other proxies. Full 2–5 km UH was defined as taking the absolute value of the maximum and minimum fields and then taking the maximum of that resultant field. The resulting field is the maximum *magnitude* at every grid cell. Performance results using a MESH baseline for this case study can be seen in Fig. 10.

What stands out immediately from Fig. 10 is that the modified variables are the most skillful on this particular day. Including maximum vertical velocity from below 100 mb allows for this field to beat all other proxies. Comparing all three 2–5 km AGL UH variables shows that adding minimum UH added a great deal of forecast skill on this day. This reveals that the impact of negelecting such information out of the model could hinder forecasts

on days when left-movers are common. To emphasize this point, Fig. 11 shows where each individual UH variable had an impact on the performance metrics. There is a large area of minimum UH (magenta) contribution within the MESH neighborhood (red circles) where severe hail was observed in northwest Oklahoma. At present time, this negative UH information is currently unavailable to forecasters.

4. Conclusions

A long-term, objective verification was done on the current severe hail proxies in the NSSL-WRF. Analyses show that 2–5 km AGL maximum UH is the most skillful proxy in matching observed severe hail probabilities, particularly when using 70 m^2s^{-2} as the threshold. All severe hail proxies are more skillful in the spring versus summer, though maximum UH has much less of a drop off in skill among the seasons. Performance of severe hail proxies in the NSSL-WRF were maximized when using an 80-km forecast ROI in neighborhood verification statistics.

Out of this verification work it was found that only capturing the maximum updraft speed below 400 mb is not appropriate in many cases. Searching for the maximum updraft below 100 mb is more likely to capture the maximum and can improve scores significantly. The extraction of minimum UH within the NSSL-WRF also can have a significant impact on performance metrics, particularly when anticyclonic, left-moving storms are prolific.

Future work will involve understanding how often the addition of minimum UH improves verification scores. Other efforts will be made to develop new hail proxies that may be able to improve on the current options. Finally, further work should be done to include these variables in other deterministic and ensemble models to see how consistent these results are under varying configurations.

Acknowledgments. Funding for this work was provided by NOAA grant #NA15OAR4590192. Some of the computing for this project was performed at the OU Supercomputing Center for Education & Research (OSCER) at the University of Oklahoma (OU) as well as at NSSL. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the authors and do not necessarily reflect the views of NOAA or the Department of Commerce.

References

Adams-Selin, R. D. and C. L. Ziegler, 2016: Forecasting hail using a one-dimensional hail growth model within WRF. *Mon. Wea. Rev.* doi:10.1175/MWR-D-16-0027.1, in press.

Blair, S. F., D. R. Deroche, J. M. Boustead, J. W. Leighton, B. L. Barjenbruch, and W. P. Gargan, 2011: A radar-based assessment of the detectability of giant hail. *Electron. J. Severe Storms Meteor.*, **6** (7).

Changnon, S. A., D. Changnon, and S. D. Hilberg, 2009: Hailstorms Across the Nation: An Atlas about Hail and Its Damages. Illinois State Water Survey Contract Rep. 2009-12, 92 pp. [Available online at http://www.isws.illinois.edu/pubdoc/CR/ISWSCR2009-12.pdf]

Cintineo, J. L., T. M. Smith, V. Lakshmanan, H. E. Brooks, and K. L. Ortega, 2012: An Objective High-Resolution Hail Climatology of the Contiguous United States. **27** (5), 1235–1248. doi:10.1175/WAF-D-11-00151.1.

Earth System CoG, 2016: ESMPy v7.0.0. [Available online at https://www.earthsystemcog.org/projects/esmpy/]

Edwards, R. and S. J. Hodanish, 2004: Environmental analysis and photographic documentation of an intense, left-moving supercell on the Colorado plains. *22nd Conference on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., 12.2. [Available online at http://www.spc.noaa.gov/publications/edwards/acycsls.pdf]

Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of Spatial Forecast Verification Methods. **24** (5), 1416–1430. doi:10.1175/2009WAF2222269.1.

Harless, A. R., I. Jirak, R. Schneider, S. Weiss, 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. *25th Conference on Severe Local Storms*, Denver, CO, Amer. Meteor. Soc., 13.2. [Available online at https://ams.confex.com/ams/25SLS/techprogram/paper_175883.htm]

Hitchens, N. M., H. E. Brooks, and M. P. Kay, 2013: Objective Limits on Forecasting Skill of Rare Events. **28** (2), 525–534. doi:10.1175/WAF-D-12-00113.1.

Jirak, I. L., C. J. Melick, and S. J. Weiss, 2014: Combining Probabilistic Ensemble Information from the Environment with Simulated Storm Attributes to Generate Calibrated Probabilities of Severe Weather Hazards. *27th Conference on Severe Local Storms*, Madison, WI, Amer. Meteor. Soc., 2.5. [Available online at https://ams.confex.com/ams/27SLS/webprogram/Paper254649.html]

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. **25** (5), 1536–1542. doi:10.1175/2010WAF2222430.1.

Melick, C. J., I. L. Jirak, J. Correia Jr., A. R. Dean, and S. J. Weiss, 2014: Exploration of the NSSL Maximum Expected Size of Hail MESH Product for Verifying Experimental Hail Forecasts in the 2014 Spring Forecast Experiment. *27th Conference on Severe Local Storms*, Madison, WI, Amer. Meteor. Soc., 76. [Available online at https://ams.confex.com/ams/27SLS/webprogram/Paper254292.html]

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** (1), 78–97. doi:10.1175/2007MWR2123.1.

Roebber, P. J., 2009: Visualizing Multiple Measures of Forecast Quality. **24** (2), 601–608. doi:10.1175/2008WAF2222159.1.

Smith, T. M., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) Severe Weather and Aviation Products: Initial Operating Capabilities. *Bull. Amer. Meteor. Soc.*, **97** (9), 1617–1630. doi:10.1175/BAMS-D-14-00173.1.

Witt, A., M. D. Eilts, G. J. Stumpf, J. T. Johnson, E. D. W. Mitchell, and K. W. Thomas, 1998: An Enhanced Hail Detection Algorithm for the WSR-88D. **13** (2), 286–303. doi:10.1175/1520-0434(1998)013<0286:AEHDAF>2.0.CO;2.

Tables & Figures

Model Parameter	Treatment
Microphysics	WSM6
Radiation (Shortwave	Dudhia
Radiation (Longwave)	RRTM
Surface Layer Scheme	Monin-Obukhov
Land Surface Scheme	Noah
PBL Scheme	MYJ

TABLE 1. Pertinent NSSL-WRF configuration settings.

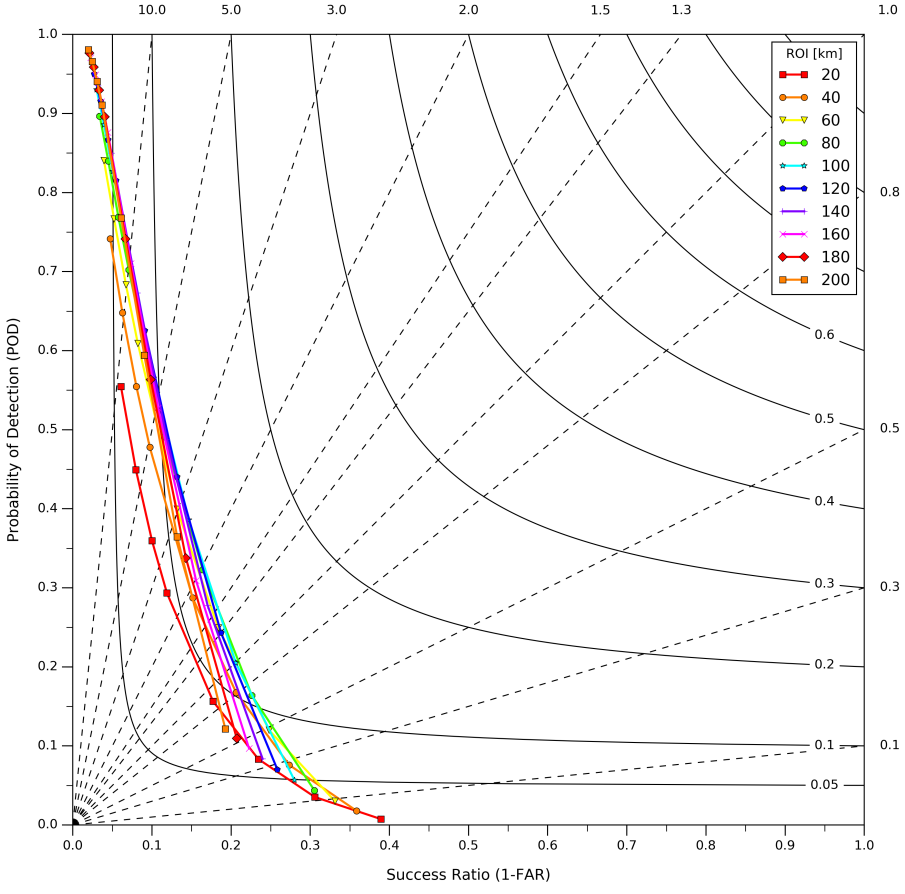

FIG. 1. Performance diagram with varying ROI (in km) for maximum 2–5 km AGL UH ($m^2 s^{-2}$) for the full period 2012–2015.

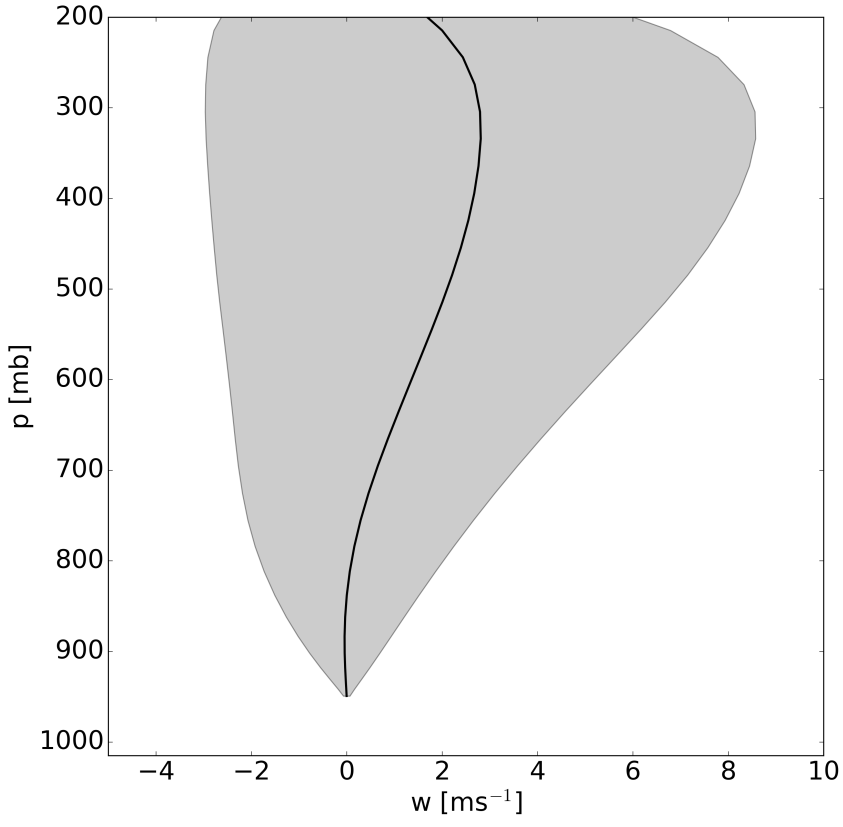

FIG. 2. Vertical profile of vertical velocity for NSSL-WRF simulation initialized 0000 UTC on 25 June, 2015. Data are from grid points where 2–5 km AGL UH values $\geq$ 70 m^{2}s^{-2}. Black line represents the mean value while gray shading represents $\pm$ 1 standard deviation.

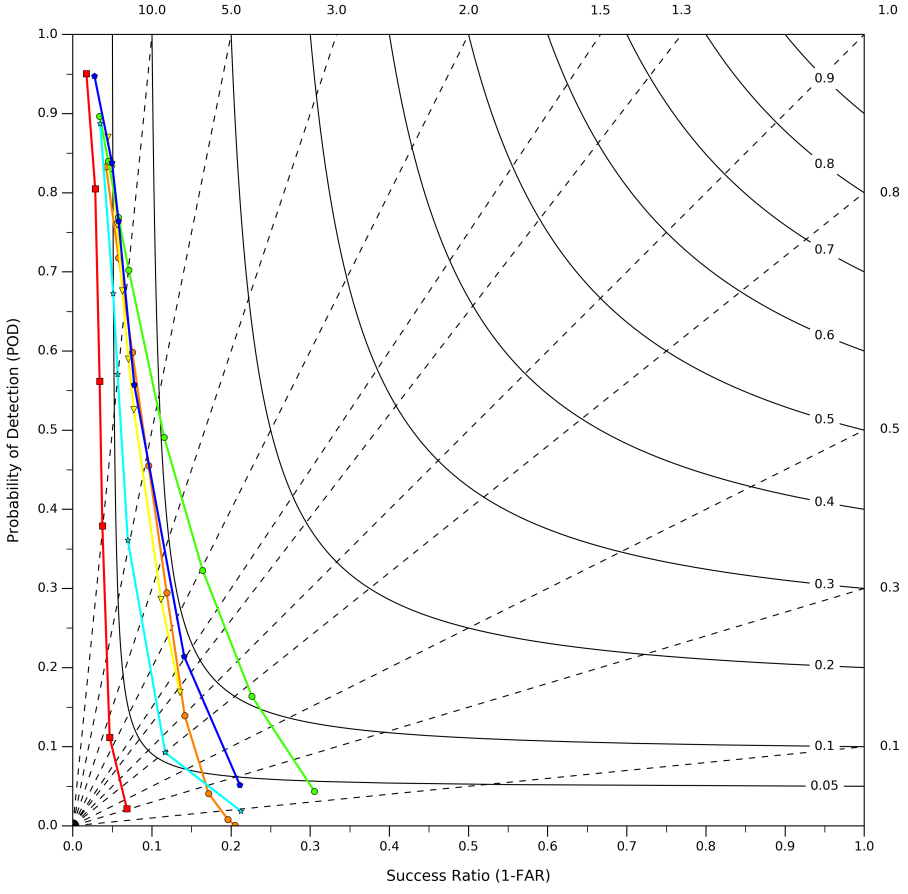

FIG. 3. 2012–2015 performance diagram for NSSL-WRF hail proxies using MESH as the verification data set.The variables are 1 km simulated reflectivity (dBZ; red squares), 250 mb updraft speed (ms^{-1}; orange circles), HAILCAST (mm; yellow triangles), 2–5 km maximum UH (m^2s^{-2}; green octagons), maximum vertically integrated graupel (kgm^{-2}; cyan stars), and maximum updraft speed (ms^{-1}; blue pentagons). All fields are hourly maximum fields except for 250 mb updraft speed.

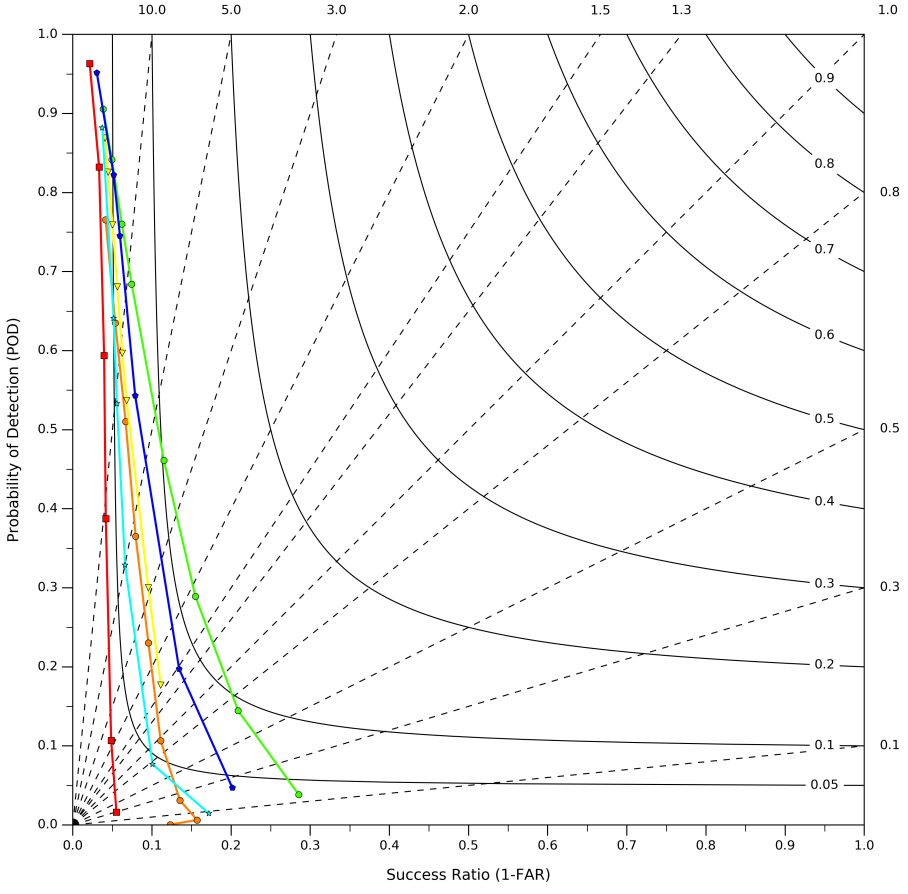

FIG. 4. As in Fig. 3, but using LSRs as the verification data set.

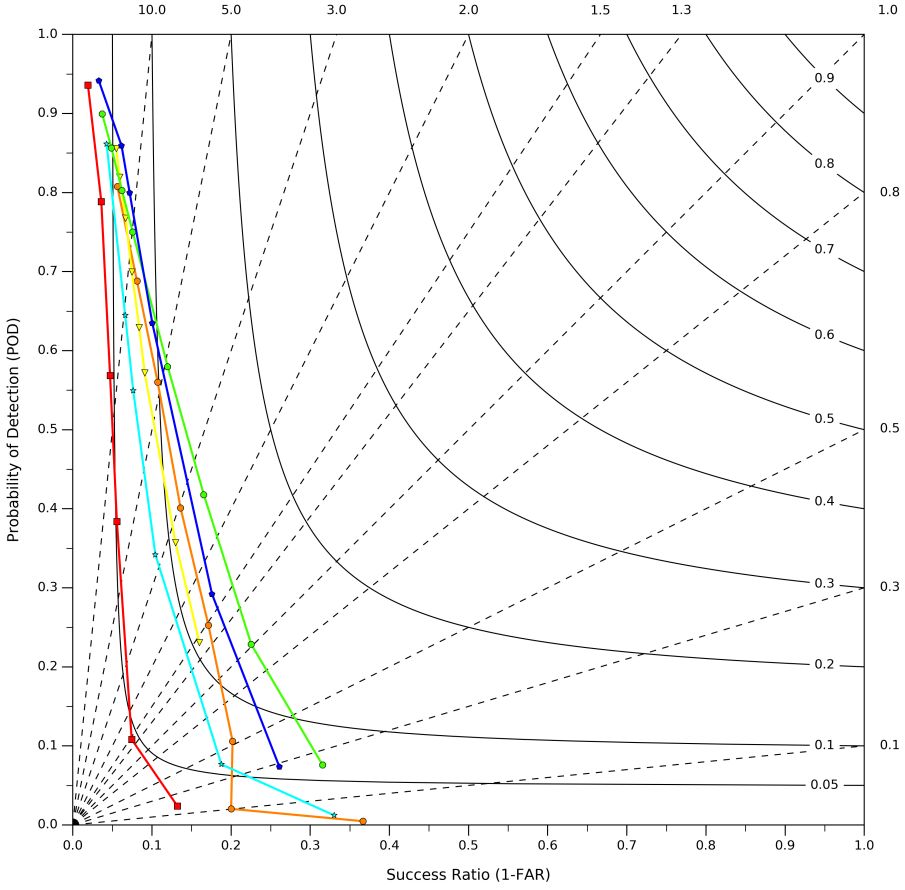

FIG. 5. As in Fig. 3, but for spring (MAM) months only.

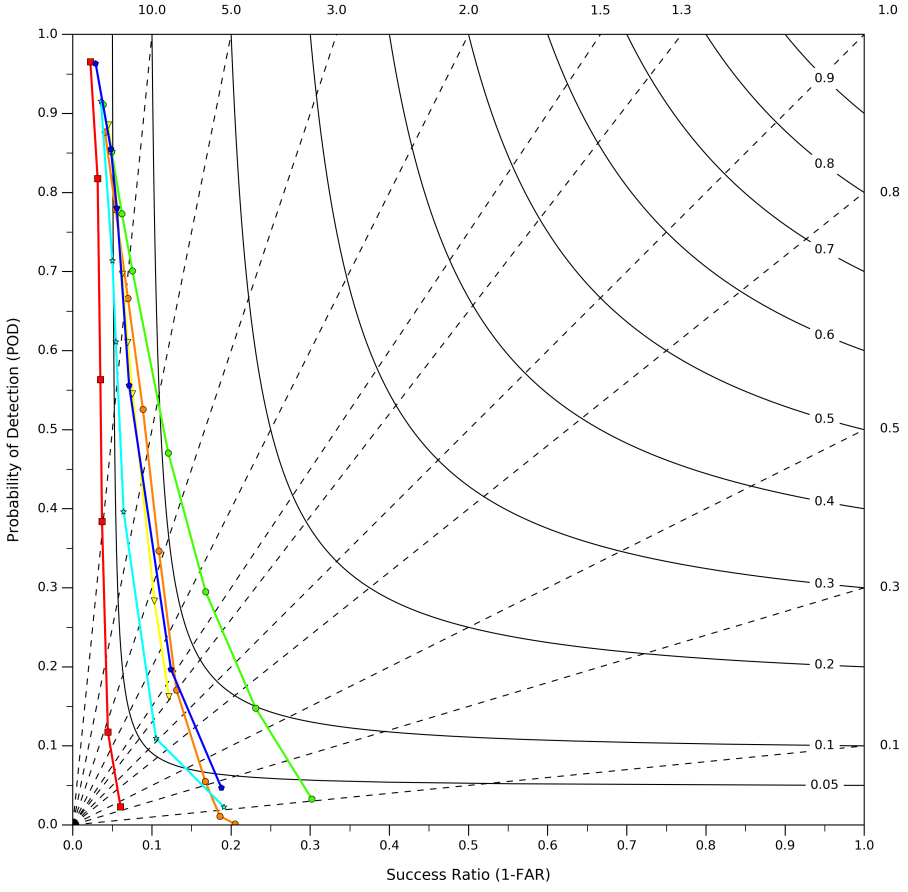

FIG. 6. As in Fig. 3, but for summer (JJA) months only.

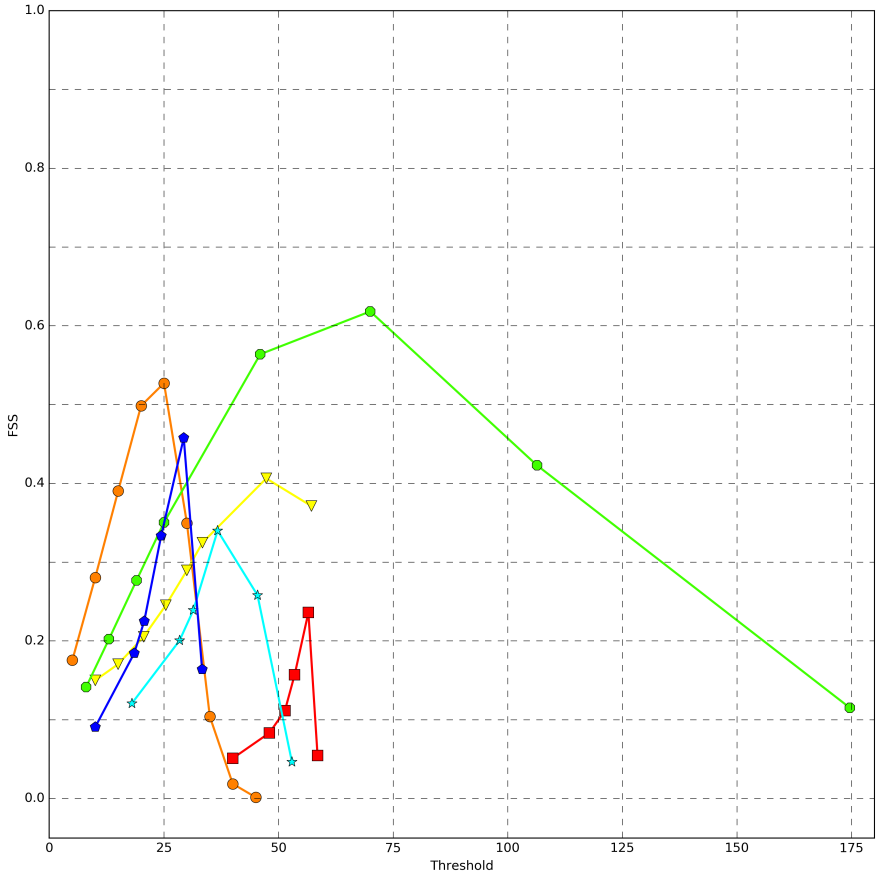

FIG. 7. FSS for 2012–2015 for NSSL-WRF hail proxies using MESH as the verification data set.The variables are 1 km AGL simulated reflectivity (dBZ; red squares), 250 mb updraft speed (ms^{-1}; orange circles), HAILCAST (mm; yellow triangles), 2–5 km AGL maximum UH (m^2s^{-2}; green octagons), maximum vertically integrated graupel (kgm^{-2}; cyan stars), and maximum updraft speed (ms^{-1}; blue pentagons). All fields are hourly maximum fields except for 250 mb updraft speed.

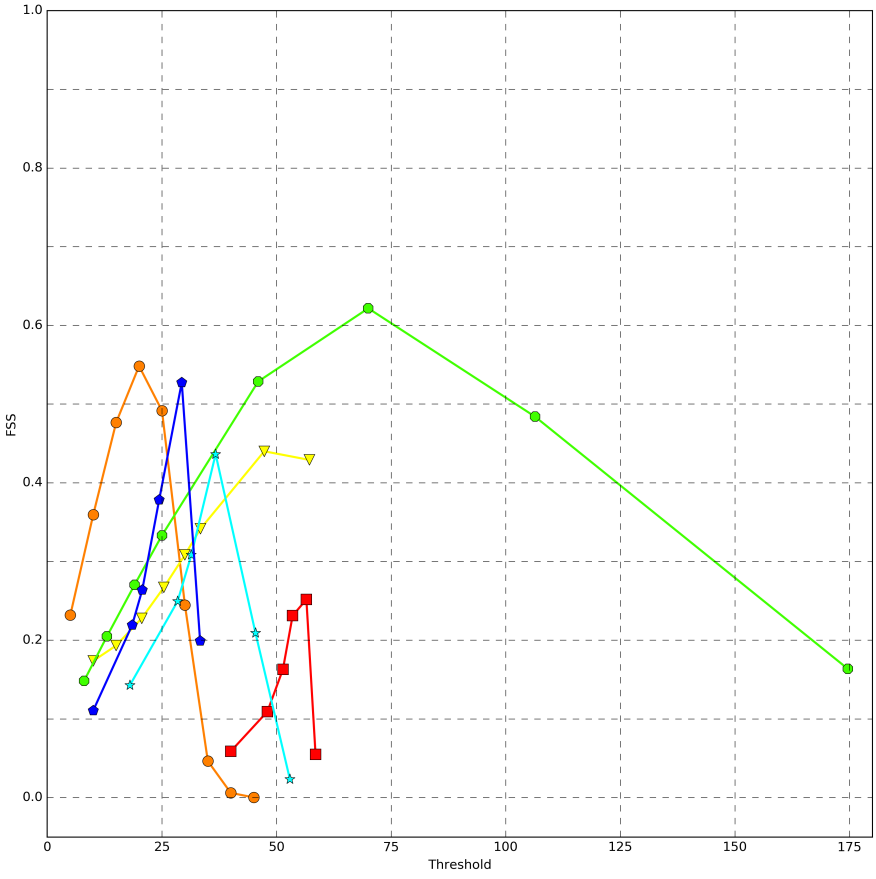

FIG. 8. As in Fig. 7, but for sping (MAM) months only.

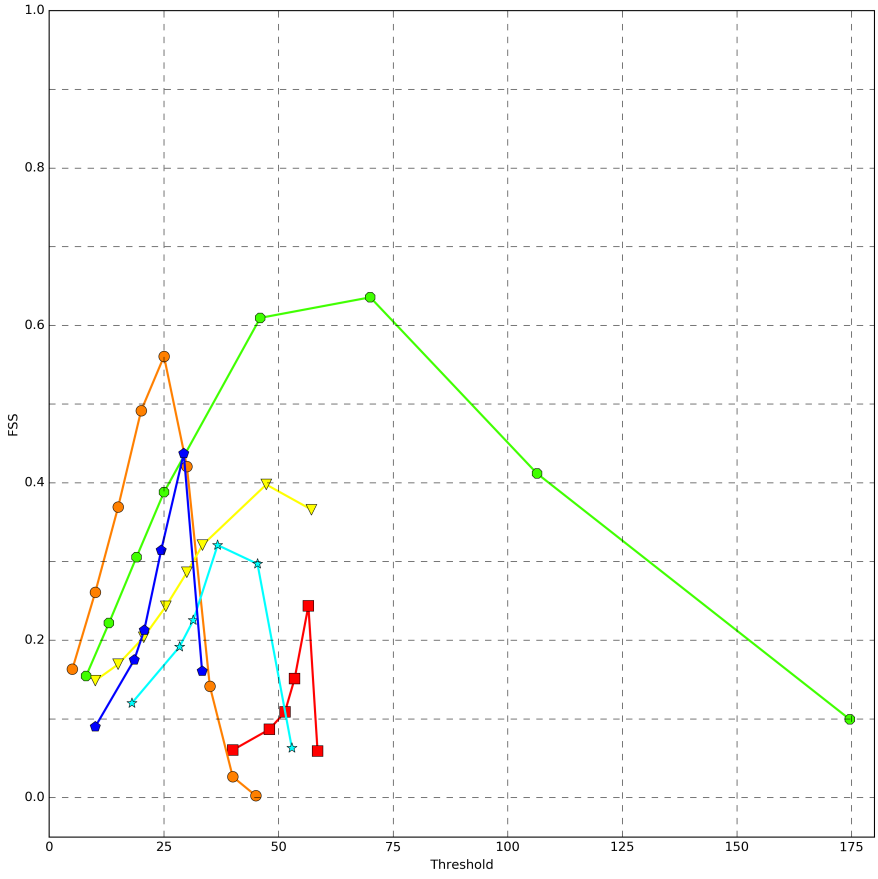

FIG. 9. As in Fig. 7, but for summer (JJA) months only.

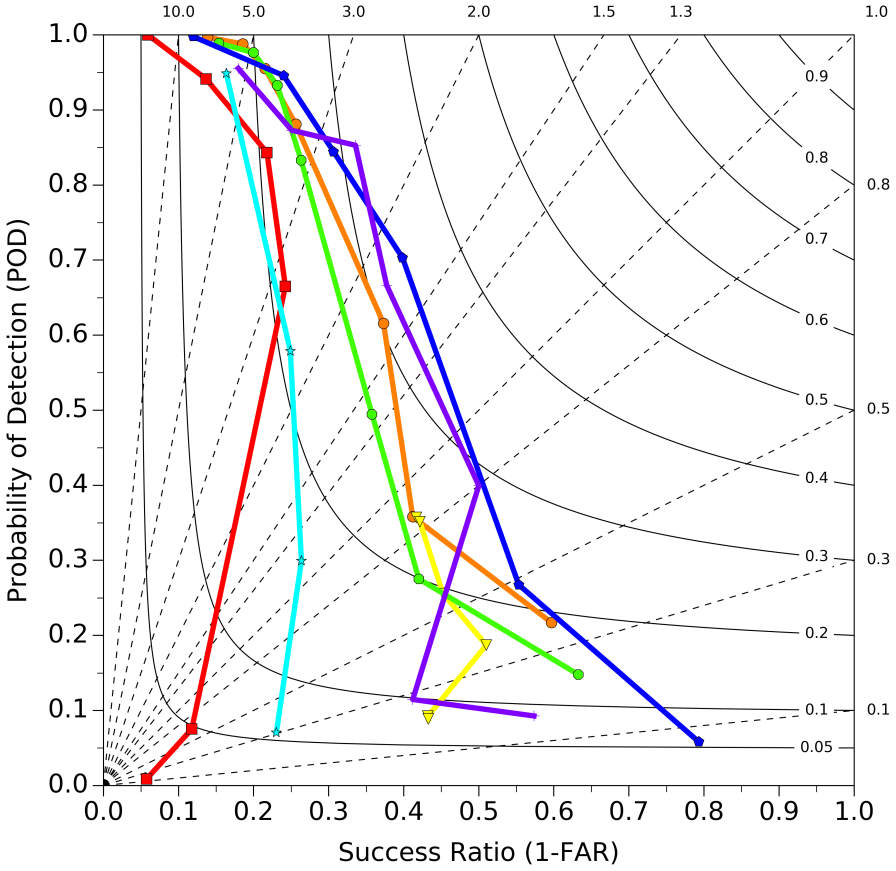

FIG. 10. Performance diagram using a MESH baseline for the modified NSSL-WRF case stud for the 08 May 2016 0000-UTC initialized forecast. The variables are 1 km AGL simulated reflectivity (dBZ; red squares), full 2–5 km AGL UH (m^2s^{-2}; orange circles), HAILCAST (mm; yellow triangles), 2–5 km AGL maximum UH (m^2s^{-2}; green octagons), maximum vertically integrated graupel (kgm^{-2}; cyan stars), maximum updraft speed **below 100 mb** (ms^{-1}; blue pentagons), and minimum 2–5 km AGL UH (m^2s^{-2}; purple crosses).

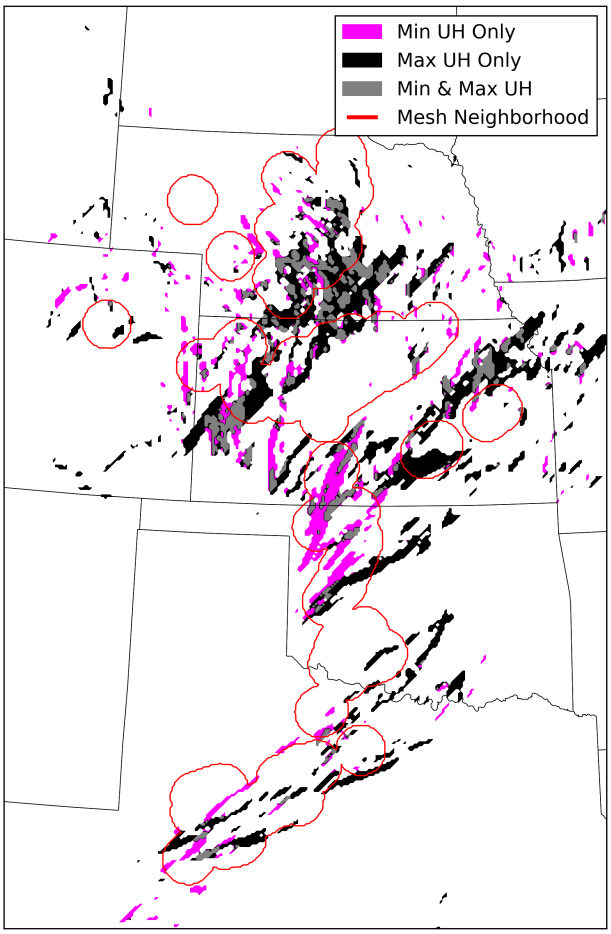

FIG. 11. Spatial contribution of 2–5 km AGL minimum (magenta), maximum (black), and full (gray) UH to severe ($\geq$ 29 mm) hail MESH 40 km neighborhoods (red circles). Data from 0000-UTC initialized NSSL-WRF on 08 May 2016.