Anti-Goldilocks Forecasting: Modeling Forecast Probability Going From Small to Large Scales

JOSEPH DALE*

National Weather Center Research Experiences for Undergraduates Program
Norman, OK
University of Northern Colorado
Greeley, Colorado

BURKELY TWIEST GALLO

NOAA/NWS/NCEP/Storm Prediction Center
Cooperative Institute for Severe and High-Impact Weather Research and Operations
Norman, OK

HAROLD BROOKS

NOAA/OAR/National Severe Storms Laboratory
School of Meteorology, University of Oklahoma
Norman, OK

ABSTRACT

The Forecasting a Continuum of Environmental Threats (FACETs) project aims to span the range of convective forecast scales with consistent probabilistic forecasts. To investigate how areal coverage probabilities behave across a continuum of space and time, this study analyzes simulations from an idealized model in time and space. Two grids are created. On the first, "events" are randomly placed throughout the grid. On the second, reports are placed with the same overall coverage but organized in time and space (e.g., lines). Aggregation is done over time and space scales on the grids to calculate the coverage probability as the size of the aggregation changes. Dividing the "organized" coverage probability by the "random" yields a u-shaped curve as a function of aggregation size. The depth and location of the minimum of the u-curve is related to the organization of the threat and its underlying coverage on the finest grid. Experiments with this framework - plotting synthetic data and numerical model proxies as organized events - indicate that the location where the u-curve reaches max depth is between the watch and warning time and space scales. These results show that forecast interpretation is different between long and short scales, and that organization of storms strongly influences forecasts in the watch-to-warning space. These results characterize a challenge in creating consistent probabilities across the spectrum of scales, a goal of FACETs.

1. Introduction

Currently in the United States, the National Weather Service (NWS) issues convective forecasts ranging from the warning time and space scale up to Storm Prediction Center's (SPC) convective outlook scale. To convey information regarding convective hazards between these scales, the NWS and SPC issue products such as mesoscale discussions and convective watches. However, a dichotomy exists between the scales: at a small spatial scale, the warning scale, the probabilities are binary – either a threat is imminent, or it is not – while at a large spatial scale, the convective outlook scale, percent probabilities are used to describe the likelihood of a hazard within 25 miles of a point. Under recommendations to bridge the gaps between these scales (National Research Council 2006; National Institute of Standards and Technology 2013), the Forecasting a Continuum of Environmental Threats (FACETs) project proposes a method to shift environmental threat communication into a paradigm of continuous probabilistic hazard information (Rothfusz et al. 2018). This study aims to aid forecasters and decision-makers by providing a framework by which forecast probability is understood across numerous time and space scales.

*Corresponding author address: Joseph Dale, OU CAPS, 120 David L. Boren Blvd, Suite 2500, Norman, Oklahoma 73072.
E-mail: josephddale0@gmail.com

Prior research has investigated probabilistic hazards at a variety of scales. Climatological studies have examined daily tornado and severe storm probability (Brooks et al. 2003; Doswell et al. 2005, respectively) as a way to understand SPC's daily forecast regions of probability. On smaller time scales, Krocak and Brooks (2018) looked at hourly tornado probability and determined the time of day that tornadoes are most likely to occur across the United States, while Krocak and Brooks (2020) focused on severe weather probabilities between the watch and convective outlook scale, and found that SPC's convective outlooks can confidently be interpreted as valid for a particular 4 hour period of the day. As forecasts go to smaller and smaller scales, however, making useful probabilities becomes challenging. As the spatial and temporal scales approach zero, the raw probabilities become vanishingly small. The framework developed herein will investigate coverage probabilities over a continuum of time and space and resolve forecast interpretation across that spectrum.

2. Data and Methods

To model coverage probabilities over a continuum of time and space, we create 3D grids modeling two space dimensions and one time dimension. Two synthetic experiments are then carried out: on one grid, 100 random points are plotted (Fig. 1a as a scaled-down example of a "random experiment"), while on the other grid, lines of points are plotted to represent organized storm tracks (Fig. 1b as a scaled-down example of an "organized experiment"). In both cases, 100 total points are plotted. Therefore, the lines of points (organized experiment) are plotted with equal coverage as the random experiment. In order to calculate areal coverage, an algorithm is developed to pass over the grids multiple times and determine the probabilities associated with the corresponding forecast size. To represent different time and space forecast scales, we iterate through the grid with different aggregation sizes (where aggregation size is the volume given by the space dimensions multiplied by the time dimension). The areal coverage is calculated as the number of aggregations containing an event divided by the total number of aggregations. For the random experiment, an analytic solution exists:

$$P = 1 - (1 - p)^n \tag{1}$$

Where P is the areal coverage, p is the probability of success, and n is the number of trials. The cases examined herein do not always reflect the perfect analytic solution due to the single randomized instance used per trial and the limited domain, but the probability that we obtain is sufficiently close enough to the true solution. For the organized case, however, the probability of an event occurring in one spot strongly influences the probability of an event occurring in another spot. Therefore, Eq. 1 cannot be used to describe probabilities in the organized case, and the areal coverage of the random case is always greater than or equal to the areal coverage of the organized case.

Figure 1 outlines the aggregation process for a simplified case in two dimensions. The entire range of time and space scales is subsequently iterated through to understand how areal coverage changes spatiotemporally. The specific ranges examined in this study will be detailed below.

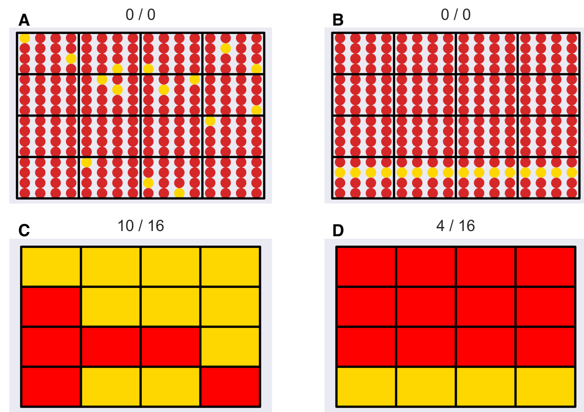

FIG. 1. A 2D example case of the 4x4 aggregation size of the random (left) and organized (right) scenarios. (A) and (B) outline the location of the events, (C) and (D) show hits and misses after the aggregating has run. If an event was in the 4x4 box, the box turns yellow, otherwise it turns red. Areal coverage is given as the fraction on top of each plot.

In order to further investigate how organization affects areal coverage, the relationship between areal coverage and various organization scenarios are examined. To create these various organization scenarios, lines of varying lengths are plotted on a grid. To represent storms with shorter tracks, more lines (with fewer consecutive points) are plotted, keeping the overall coverage the same. Last, the total number of points plotted on the grids is changed from 100 to 500 in order to analyze how the amount of coverage affects probability across the different scales.

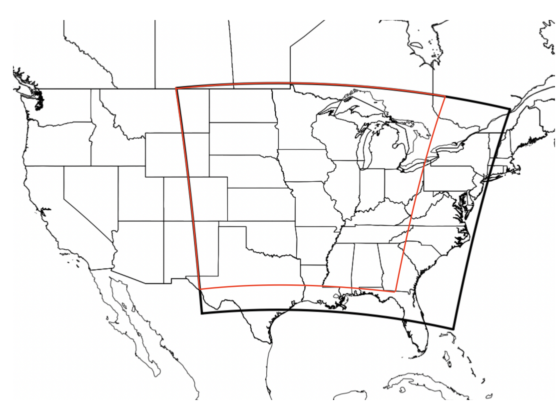

FIG. 2. Spatial Domain of the NSSL-WRF (black box) and the model used in study (red box)

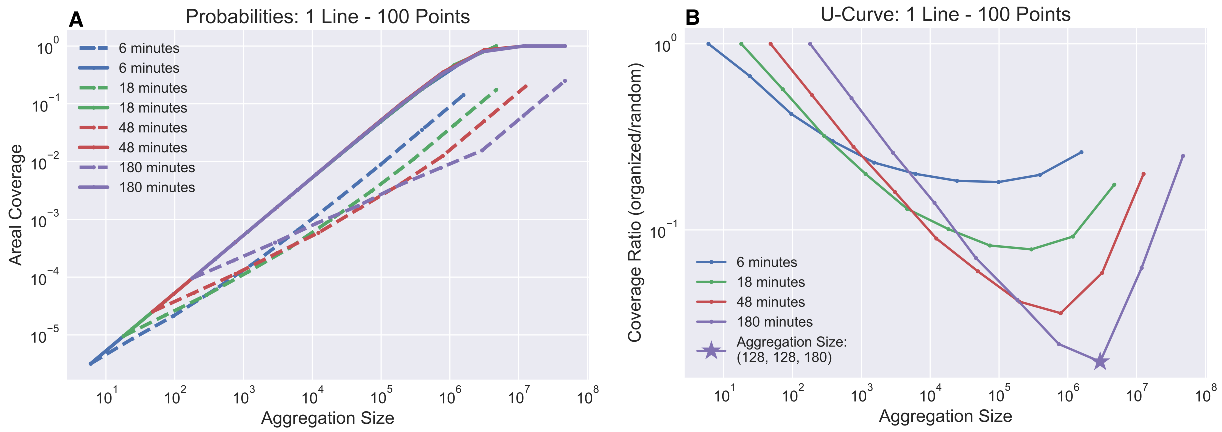

FIG. 3. (A) Areal coverage as a function of aggregation size for a 3D Grid. Colors are separated by time aggregation size. Solid lines represent areal coverages for the random scenario, dashed lines represent areal coverages for the organized scenario. (B) Coverage ratios are calculated as: organized scenario areal coverage divided by the random scenario areal coverage. Minimum point of all ratios is indicated by a star.

Once the model is investigated for synthetic organization scenarios, convective-allowing model (CAM) data is used to simulate a real-world environment. Hourly maximum updraft helicity (UH; Kain et al. 2008) values are acquired from the National Severe Storms Laboratory (NSSL) Weather Research and Forecasting (WRF; Skamarock et al. 2008) model (NSSL-WRF) with 1-km horizontal grid spacing. Because of their proven usefulness as a CAM indicator of mesocyclone formation at low-levels (Sobash et al. 2016) and mid-levels (Kain et al. 2008; Clark et al. 2012), the 0-3 km and the 2-5 km hourly maximum UH are examined. UH values found to be at the climatological 99.85 percentile were used as a threshold (Clark et al. 2019). At each grid point, if the UH value exceeds the percentile threshold for the respective UH height, then the point is plotted onto the grid as an event. The aforementioned aggregation method is then used to calculate areal coverage across the spatiotemporal scales, thus allowing for interpretation of how storm organization affects areal coverage on real-world scales. A few notable differences exist between this "real-world environment" and the synthetic data experiment. The domain size changes (specifics noted below), overall coverage is significantly greater in the real-world environment ($\sim$ 1 event per 2,500 grid points) as opposed to our synthetic environment ($\sim$ 1 event per 100,000 grid points), and our CAM data is hourly, rather than by minute.

Two case studies are used to examine real-world scenarios. Case 1 is the December 10, 2021 severe outbreak that spawned numerous long-track supercells and swept through the lower Mississippi valley and into the Ohio river valley through the early evening and into the night. Case 2 arrived five days later and with a more linear storm mode on December 15, 2021, when a derecho swept through the Corn Belt region from the late afternoon into the late evening. While the two cases were characterized by different storm modes, they were both significant tornado outbreaks. For these two cases, we use a domain size of 2048x2048-km over an 18 hour forecast period. The spatial domain is cut down from that of the NSSL-WRF to be a perfect square and thus fit in our model. Figure 2 outlines the 2048x2048-km spatial domains used in the cases, along with the difference between ours and that of the NSSL-WRF.

3. Results and Discussion

After running the model in the 3D synthetic scenario with a grid size of 512x512x720 (which can be interpreted as km x km x min), probabilities are calculated for individual time aggregates and graphed as a function of aggregation volume (Fig 3a). At equal aggregation sizes, the areal coverage for the random case (solid lines) stay the same across different time aggregations, while they change for the organized case (dashed lines). Notice that the larger time aggregations begin up and to the right, as the larger aggregation shifts right along the x-axis and shifts up due to a higher areal coverage. When graphed with logarithmic scales, the random scenario is characterized by a straight line until the areal coverages begin to saturate as the aggregations become too large for the grid. On an infinitely large grid, saturation would never occur and the random scenario would remain a straight line.

To understand how the organized scenario differs from the random, the ratio of their areal coverages is graphed (Fig 3b). Notice the general shape of the lines is a "u-curve"; as the aggregation sizes increase, the organized areal coverages become increasingly less than the random areal coverages up until the random scenario aggregations become too large for grid and the areal coverage begins to saturate, at which point the ratio begins to return to one.

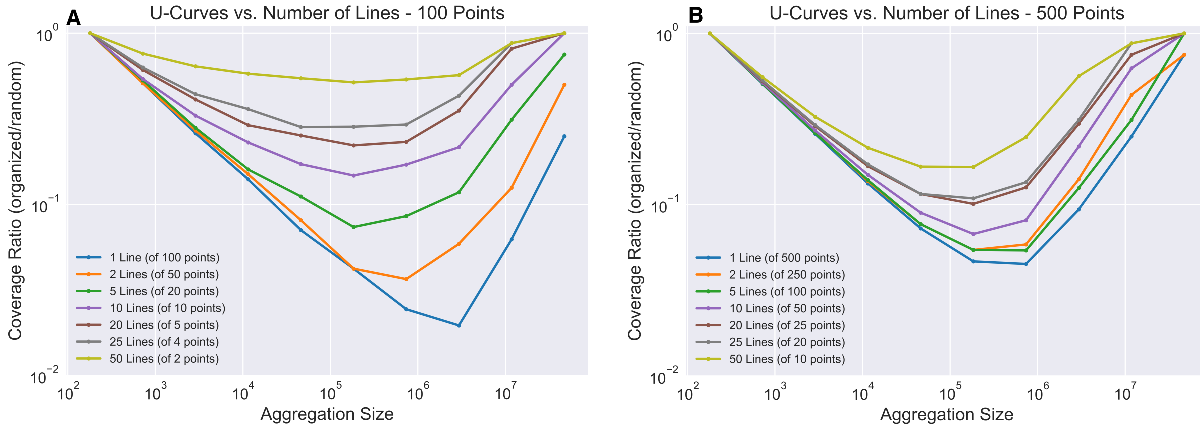

FIG. 4. The 180-minute time aggregation u-curves from (A) 100 point coverage and (B) 500 point coverage.

Date	0-3 km UH Events	2-5 km UH Events
12-10-21	17844	10056
12-15-21	29939	2410

TABLE 1. A table showing the coverage of each model data scenario.

The depth of the u-curve and the location where it reaches its minimum value is related to the organization of the severe threat and its underlying coverage on the finest grid. If the underlying grid is interpreted as 1-km x 1-km x 1-min, then the maximum depth of the u-curve exists in a space between the watch-to-warning scale. This means that the organization of storms most strongly affects forecasts in between the watch-to-warning scale.

The synthetic scenario with different organization scenarios (Fig. 4a) reveals that organization affects the shape of the u-curve. As the storms become more organized (meaning fewer lines with longer tracks), the depth of the curve increases and the location of its minimum shifts. In addition, when coverage is increased from 100 points to 500 points (Fig. 4b), the maximum depth of the u-curve decreases and the location of its minimum shifts back left. We hypothesize this is a result of how the synthetic scenario has been created. Since the points plotted on the organized grid move one space step per time step, higher coverages overpopulate the time dimension, causing more aggregations in the organized scenario to contain an event, increasing its areal coverage and thus decreasing depth of the curve.

Investigation into the behavior of scenarios with real-world data is shown by plotting the cases' corresponding u-curves side by side in Figure 5. It is important to note that case 1 (December 10th, 2021) had more overall coverage for the 2-5 km UH, while case 2 had more overall coverage for the 0-3 km UH (Table 1). Notice that the 0-3 km u-curves are strikingly similar (Fig. 5a), perhaps due to the fact that both events were a significant tornado outbreak, as low-level UH measurements are a skillful indica-

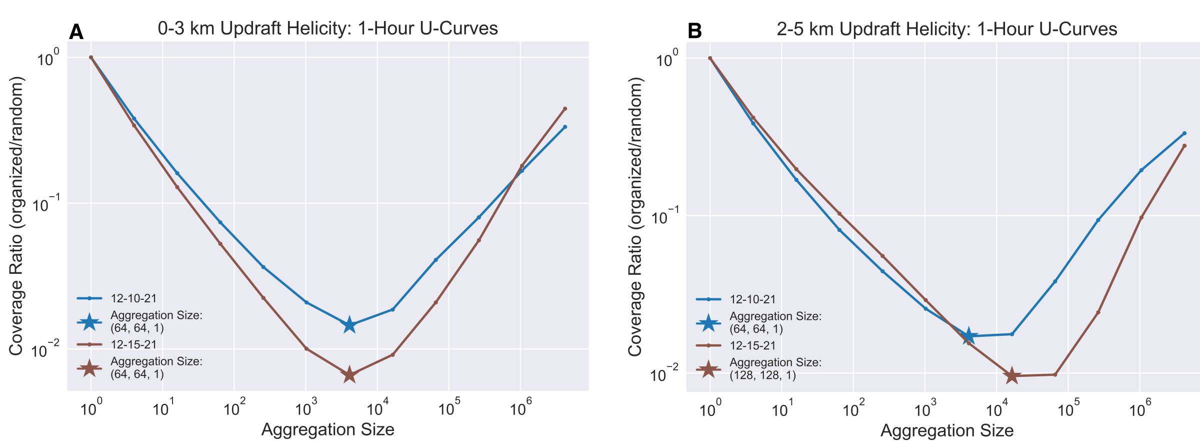

FIG. 5. A graph contrasting u-curves of the 1-hour time aggregation for the (A) 0-3 km UH and the (B) 2-5 km UH.

tor in forecasting tornadoes (Sobash et al. 2019). Notice that the depth and the location of the u-curves differ for the 2-5 km cases (Fig. 5b). This is likely due to the differing storm modes for the two cases (supercellular on 12/10, derecho on 12/15).

4. Conclusion

This study highlights a method that allows for inspection of forecast areal coverage probabilities going from small to large scales. Two different scenarios were explored. In the first, a synthetic scenario involved plotting lines on a grid to investigate how the probabilities might change in a variety of cases. In the second, UH values exceeding a threshold from the 1-km NSSL-WRF were plotted onto the grid to see how the probabilities behaved in a real-world environment. An algorithm was then created that aggregated over a broad range of time and space scales and calculated the probabilities of those aggregations containing an event as a means to interpret forecast probability on those scales. Out of these probabilities, a "u-curve" graph was created, representing the difference in the probabilities of organized scenarios versus random scenarios. The purpose of the study was to determine how forecast probabilities changed over a broad continuum of time and space. The results aid forecasters and decision-makers by detailing the most challenging spatiotemporal scales associated with particular weather regimes and coverages, and explain a difficulty in the FACETs vision.

Because of the organized nature of weather, forecast interpretation changes based on the time and space scales being used. This is represented in the u-curves for a variety of different weather regimes which are simulated in the synthetic storm scenarios. It is also shown that the depth and location of the u-curve changes based on the characteristics of the weather pattern that day, as shown by the model UH scenarios. This means that forecast interpretation also changes based on the type of weather that occurs. Results indicate that the u-curve minima occurs somewhere in the 64 km x 64 km x 1 hour and 128 km x 128 km x 2 hour time scales, indicating that the organization of storms strongly influences forecasts in the watch-to-warning space. These results characterize a challenge in creating consistent probabilities across the spectrum of scales, a goal of FACETs.

Future research can expand upon these findings by adding more cases examining 0-3 km and 2-5 km UH. Particular attention might be given to days with tornado outbreaks as opposed to days that produce storms, but few tornadoes. We hypothesize that the u-curves would exhibit notably different characteristics for such days. In addition, more cases investigating days of different storm modes would help confirm the hypothesis that a difference in the u-curves for the 2-5 km level can be attributed to the dominant storm mode.

Acknowledgments. We want to express our gratitude to Alex Marmo, Dr. Daphne LaDue, and the entire National Weather Center Research Experience for Undergraduates Program for making this work possible. We would also like to give Allison Brannan a huge thanks for her tremendous help in the drafting of this paper.

This work was prepared by the authors with funding provided by National Science Foundation Grant No. AGS-1560419, and NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA11OAR4320072, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, NOAA, or the U.S. Department of Commerce.

References

Brooks, H. E., C. A. Doswell III, and M. P. Kay, 2003: Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting*, **18**, 626–640, doi:10.1175/1520-0434(2003)018⟨0626:CEOLDT⟩2.0.CO;2.

Clark, A., J. S. Kain, P. T. Marsh, J. Correia, M. Xue, and F. Kong, 2012: Forecasting tornado pathlengths using a three- dimensional object identification algorithm applied to convection-allowing forecasts. *Wea. Forecasting*, **27**, 1090–1113, doi:10.1175/WAF-D-11-00147.1.

Clark, A., and Coauthors, 2019: Spring Forecasting Experiment 2019 Preliminary Findings and Results. [Available online at https://hwt.nssl.noaa.gov/sfe/2019/docs/HWT_SFE_2019_Prelim_Findings_FINAL.pdf.].

Doswell, C. A., H. E. Brooks, and M. P. Kay, 2005: Climatological estimates of daily local nontornadic severe thunderstorm probability for the United States. *Wea. Forecasting*, **20**, 577–590, doi:10.1175/WAF866.1.

Kain, J. S., and Coauthors, 2008: Some practical considerations regarding horizontal resolution in the first generation of operational convection-allowing NWP. *Wea. Forecasting*, **23**, 931–952, doi:10.1175/WAF2007106.1.

Krocak, M. J., and H. E. Brooks, 2018: Climatological Estimates of Hourly Tornado Probability for the United States. *Wea. Forecasting*, **33**, 59–69, doi:10.1175/WAF-D-17-0123.1.

Krocak, M. J., and H. E. Brooks, 2020: An Analysis of Subdaily Severe Thunderstorm Probabilities for the United States. *Wea. Forecasting*, **35**, 107–112, doi:10.1175/WAF-D-19-0145.1.

National Institute of Standards and Technology, 2013: Technical investigation of the May 22, 2011 tornado in Joplin, Missouri. National Institute of Standards and Technology Final Rep.,. 428 pp., doi:10.6028/NIST.NCSTAR.3.

National Research Council, 2006: Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts. National Academics Press, 124 pp., doi:10.17226/11699.

Rothfusz, L. P., R. Schneider, D. Novak, K. Klockow, A. E. Gerard, C. Karstens, G. J. Stumpf, and T. M. Smith, 2018: A proposed next-generation paradigm for high-impact weather forecasting. *Bull. Amer. Meteor. Soc.*, **99**, 2025–2043, doi:10.1175/BAMS-D-16-0100.1.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. 113, [Available online at http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf].

Sobash, R. A., G. S. Romine, C. S. Schwartz, D. J. G. II, and M. L. Weisman, 2016: Explicit forecasts of low-level rotation from convection-allowing models for next-day tornado prediction. *Wea. Forecasting*, **31**, 1591–1614, doi:10.1175/WAF-D-19-0044.1.