2.3: Creating spatio-temporal tornado probability forecasts using fuzzy logic and motion variability

Valliappa Lakshmanan[1,2], Kiel Ortega[1,3], Travis Smith[1,2*]

Abstract

In this paper, we describe our approach to addressing the problem of creating good probabilistic forecasts when the entity to be forecast can move and morph. We formulate the tornado prediction problem to be one of estimating the probability of an event at a particular spatial location within a given time window. The technique involves clustering Doppler radar-derived fields such as low-level shear and reflectivity to form candidate regions. Assuming stationarity, the spatial probability distribution of this region T minutes ahead is estimated and combined with the probability that the candidate region becomes tornadic T minutes later. Using these two probabilities and the variability of the motion estimates, a spatio-temporal probability field is derived.

The neural network training required to correctly estimate the probabilities has not yet been developed. Therefore, this paper illustrates the underlying idea using fuzzy logic, storm half-life and motion variability.

1. Motivation

A principled probabilistic prediction can enable users of the information to calibrate their risk and can aid decision making beyond what simple binary approaches yield (Murphy 1977). Techniques to create good probabilistic forecasts are well understood, but only in situations where the predictive model is a direct input-output relationship. If the threats in consideration move and change shape, as with short-term weather forecasts, the well-understood techniques can not be used directly. For forecasts in earth-centered domains to be useful, the forecasts have to be clearly demarcated in space and time. This paper presents a data mining approach to address the problem of creating principled probabilistic forecasts when the entity to be forecast can move and change shape.

2. Method

For fast moving threats, no principled approach to estimating probability fields exists. This is because the input features at a point in space and time affect the threat potential at a different point N minutes later. Thus, a simple input/output mapping is insufficient because which location will be af-

*Corresponding author: lakshman@ou.edu [1]The Cooperative Institute of Mesoscale Meteorological Studies (CIMMS), University of Oklahoma, [2]The National Severe Storms Laboratory, Norman, OK. [3]School of Meteorology, U. Oklahoma

fected, and when it will be affected needs to be known. Yet, in practical situations, the time and location that will be affected is not known with certainty. Consequently, it is necessary to assume a spatial probability distribution associated with the locations that will be affected at a given time instant. Once this second probability is introduced, prior work on principled probability estimates is no longer applicable. A new formulation is needed in the development of a spatiotemporal framework to estimate the probability of occurrence of moving or spreading threats to account for the dynamics of spatial probability field of area at risk over time.

One can develop the spatiotemporal formulation first by identifying threat "precursors" and computing their features. Each of the features signals a probability distribution of threats in space and time. Threat probabilities can be combined from multiple features and dynamics among these features to estimate the probability of a threat occurring at a particular location within a specific time window. One factor that needs to be considered is that even if the spatial and temporal distribution of the locations that will be affected by a particular feature is estimated, whether the feature will lead to the threat still needs to be estimated. This, of course, is a problem that has been thoroughly addressed in the literature on data mining algorithms.

If motion estimates of a moving potential threat are available, the variability of the motion estimates themselves can be used to gauge the probability distribution of the motion estimates.

The probability of a tornado, $P(Tornado|v)_{xyT}$, at a particular location (xy) T minutes into the future, given the set of current measurements (v) taken from a variety of radar fields, can be estimated using these steps:

1. Find $P(Tornado|v)_{t=T}$ through traditional data mining methods. Create a training set of radar measurements, associate each cluster of measurements with 0 or 1 depending on a whether a tornado was observed with that storm T minutes into the future. Then train a neural network to provide $P(Tornado|v)_{t=T}$ at every point xy of radar domain.

2. Estimate movement using a technique such as (Lakshmanan et al. 2003). Find the temporal variability of that movement and assume that the temporal variability and the spatial variability are identical. Using this variability, form the set of points impacted by every point xy and associate that point with the probability that xy moves to that location.

3. Compute $P(Tornado|v)_{xyT}$ by numerical integration of the probability distribution corresponding to motion vectors that impact xy will yield a tornado at this location based on where the threats are currently present.

4. At every point xy of radar domain, the probability is the maximum of $P(Tornado|v)_{xyt}$ where t ranges from 0 to T.

3. Demonstration

We are currently in the process of creating the training data set in order to be able to estimate the probability using sound data-driven principles. In the mean time however, the proof of our concept shown in Figure 1 was created by approximating the probabilities using fuzzy logic (i.e. with nothing more than an educated guess).

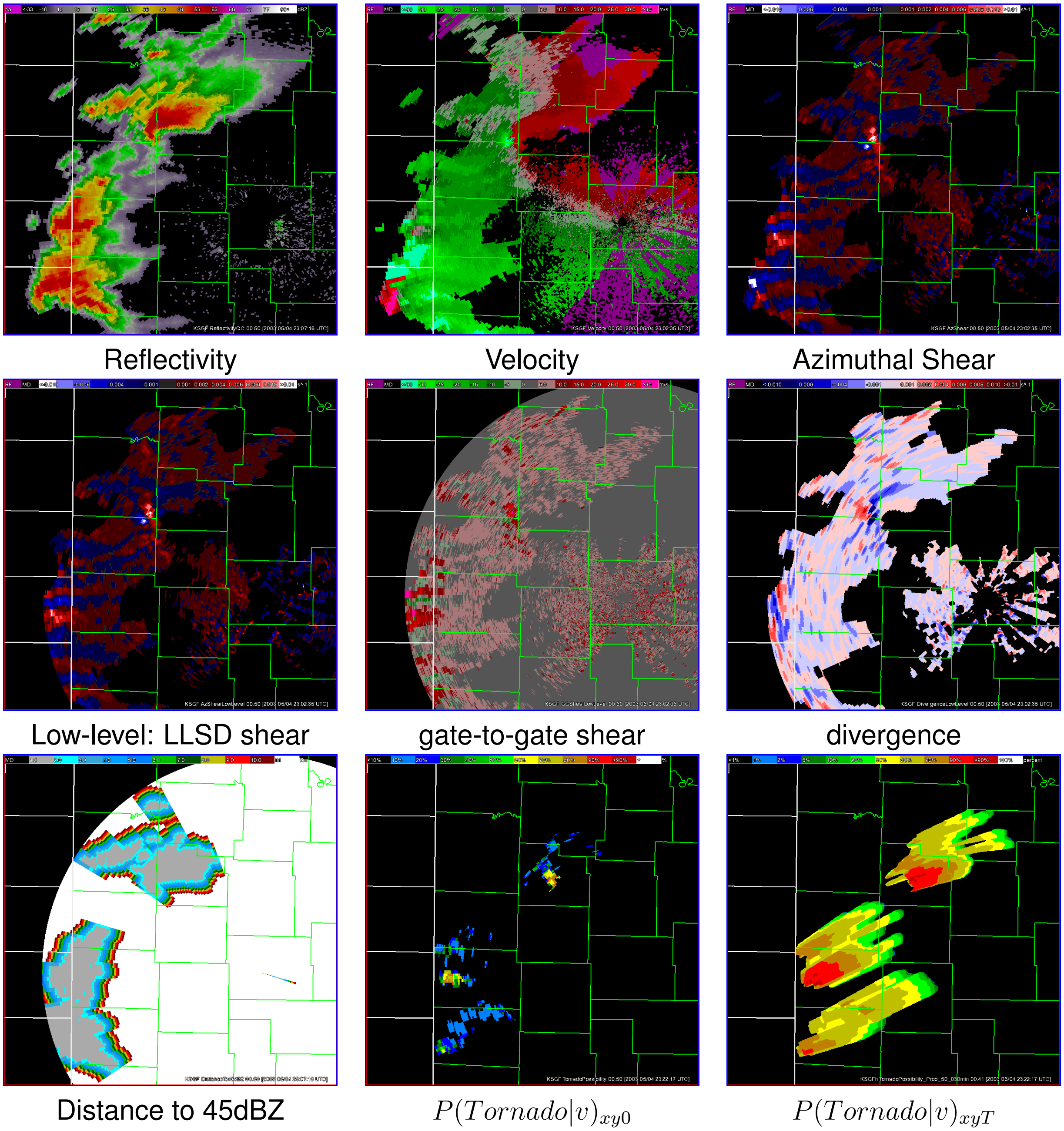

Figure 1: Some of the input vectors v that are to be used to train the neural network and (last two figures), the output tornado possibility at a single time and over a time period of 30 minutes.

It should be noted that the last two images in Figure 1 are the result of educated guesses, not of neural network training. The tornado possiblity was obtained by computing a fuzzy average of fuzzy membership functions derived from the magnitudes of azimuthal shear, distance to 45 dBZ and divergence. The shear was computed following the linear least squares derivatives (LLSD) approach of Smith (2002).

The tornado probability over a time period was derived by simply assuming a storm half-life of 40 minutes, i.e. that the probability at a time 40 minutes later would be half the probability that it was now.

These assumptions will be made unnecessary once the required probabilities are estimated from historical data sets.

Acknowledgments

Funding for this research was provided under NOAA-OU Cooperative Agreement NA17RJ1227. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of the National Severe Storms Laboratory (NSSL) or the U.S. Department of Commerce.

References

Lakshmanan, V., R. Rabin, and V. DeBrunner, 2003: Multiscale storm identification and forecast. *J. Atm. Res.*, 367–380.

Murphy, A. H., 1977: The value of climatological categories and probabilistic forecast in the cost-loss ratio situations. *Monthly Weather Review*, 803–816.

Smith, T., 2002: A two-dimensional, local, linear, least-squares method of derivative estimates from Doppler radial velocity. *21st Conference on Severe Local Storms*, Amer. Meteor. Soc., San Antonio, TX.