

A technique for creating probabilistic spatio-temporal forecasts

V Lakshmanan

*University of Oklahoma and
National Severe Storms Laboratory
lakshman@ou.edu*

Kiel Ortega

*Sch. of Meteorology
University of Oklahoma*

Keywords: Probability; Atmospheric Science

Probabilistic forecasts can capture uncertainty better and provide significant economic benefits because the users of the information can calibrate risk. For forecasts in earth-centered domains to be useful, the forecasts have to be clearly demarcated in space and time. We discuss the characteristics of a good probability forecast – reliability and sharpness – and describe well-understood techniques of generating good probability forecasts in the literature. We then describe our approach to addressing the problem of creating good probabilistic forecasts when the entity to be forecast can move and morph.

In this paper, we apply the technique to severe weather prediction by formulating the weather 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 the regions is estimated, conditioned based on the level of organization within the regions and combined with the probability that a candidate region becomes severe. For example, the probability that a candidate region becomes tornadic is estimated using a neural network with a sigmoid output node and trained on historical cases.

1. Motivation

A principled estimate of the probability that a threat will materialize can be more useful than a binary yes/no prediction because a binary prediction hides the uncertainty inherent in the data and predictive model from users who will make decisions based on that prediction. 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.¹ Probabilistic forecasts can capture uncertainty better and provide significant economic benefits because the users of the information can calibrate risk.

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.

The rest of the paper is organized as follows. The characteristics of good probabilistic forecasts, and standard data mining approaches to create such probabilistic forecasts are described in Section 2. The limitations of the standard data mining approaches in creating clearly demarcated forecasts in space and time are explained, and the first part of our predictive model (to create principled spatial forecasts) is explained in Section 3. The second part of model, to create principled temporal forecasts that can be tied with the spatial forecast, is explained in Section 4. The way of tying together the two probabilities is explained in Section 5. Results of applying this approach to predicting liquid content are described in Section 6.

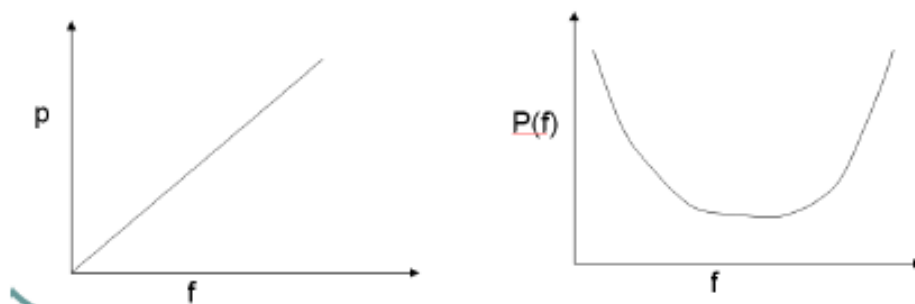


Fig. 1. A good probability forecast needs to be both reliable (left) and sharp (right).

2. Probabilistic Forecasts

A good probability forecast has two characteristics (See Figure 1): (a) it is reliable. For example, 30% of all the times that the threat is forecast as 30% likely to occur, the threat should occur and (b) it is sharp, i.e. the probability distribution function of the forecast probabilities should not be clustered around the a priori probability. Instead, there should be lots of low and high-probability events and relatively few mid-probability events.

In many cases, there are three probabilities of interest: (a) the probability that an event will occur, (b) the probability that the event will occur at a particular location (c) the probability that the event will occur at a particular location within a specified time window. When stated like this, most decision makers will aver that it is the third probability that is of interest to them,² but that is not what they are commonly provided. Even though research studies implicitly place spatial and temporal bounds on their training set, it has long been unclear how to explicitly form the spatial and temporal variability of the probability field. Thus, if probabilities are presented to decision maker, those probabilities are usually of the first type. The probability that an event will occur is commonly estimated through the formation of a mapping between input features and the desired result on a set of historical cases.³ This mapping function may take the form of a neural network, a support vector machine or a decision tree. In the case of a neural network, the choice of a sigmoid function as the output node is sufficient to ensure that the output number is a probability, provided that the training set captures the a priori probabilities of the

true input feature space.⁴ In the case of maximum marginal methods such as support vector machines or bagged, boosted decision trees, scaling the output using a sigmoid function yields the desirable property that the output is a probability.⁵

3. Spatial probability forecast

Thus, techniques to estimate the probability of an event are well understood. The estimation of spatial probabilities when the threat under consideration is stationary can be performed using the standard formalisms, for example,⁶ on soil variability. If the threat is stationary, then the value of each of the input fields at a pixel or statistical or morphological values computed in the neighborhood of a pixel may be used as the input features to a classifier. If the classifier is a neural network with a sigmoid output node or a Platt-scaled maximum marginal technique, then the output of the classifier yields a spatial probability.

One approach to probabilistic forecasts when the threat areas are not stationary is to use kriging,⁷ a geospatial interpolation approach. Naturally, such an approach is limited to studies where the kriging resolution can be finer than the rate of transformation. This is a condition that exists in slow-moving systems such as diseases (studied by⁸), but not in fast moving systems such as thunder storms.

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 affected,

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 building an ontology of threat precursors with identified 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.

The formulation can be easily described by comparing to an analogue of a set of billiards balls (See Figure 2) and asking: what is the future position of any ball on the table? Each grid point in the set of continuous fields will be considered akin to a ball, and the driving force of a set of points, such as a storm cell boundary will be considered akin to the ring.

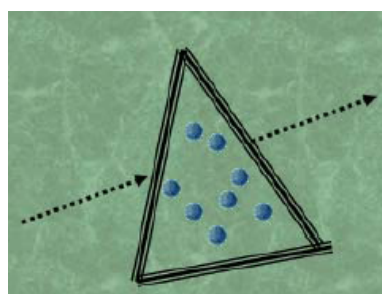


Fig. 2. The probabilistic formulation assumes rigid bodies operated upon by a driving force that permits individual degrees of motion.

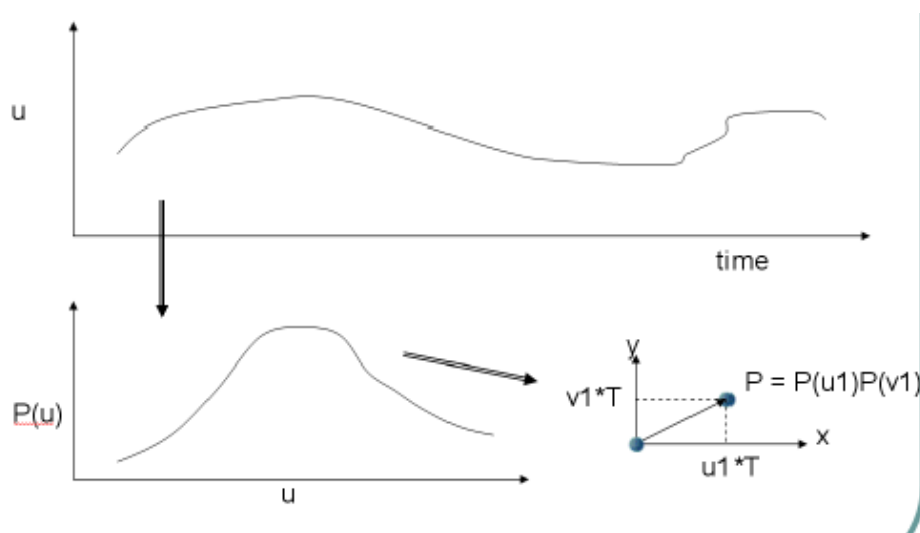


Fig. 3. By assuming that the probability distribution in space is identical to the probability distribution in time (stationarity), the estimated probability distribution in time from historical data can be used to create a spatial probability field. The figure shows how the data values of one component of the velocity vector in time (top) is used to estimate a probability distribution of that component (bottom left). This probability distribution, though estimated from the time field, is considered the spatial distribution in order to formulate the probabilistic location of a grid cell (bottom right) at time T into the future.

We will make the simplified assumption of space-time stationarity: the probability distribution of a particle in space is identical to the probability distribution of the particle in time (See Figure 3). Loosely, this assumption is similar to assuming that the probability distribution obtained by tossing 50 coins is the same as the probability distribution obtained by tossing the same coin 50 times. This assumption has to be made on faith in situations where one does not have 50 coins; we will never have 50 identical thunderstorms or hazardous events.

By making the assumption of space-time stationarity and from the motion estimated by the hybrid technique described in the next section, one can formulate the spatial probability distribution of a ball on the billiards table.

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 (Figure 3). Under stationarity assumptions, the historical probability distribution can be used to estimate the future spatial probability distribution of this current billiard ball.

There is one further complication, however. The future expected position of a billiard ball depends on how tightly the balls are packed or on how much control the forcing function exerts. If the balls are loosely packed, the movements of different balls are independent. If the balls are tightly packed, all the balls move together. The actual movement of the balls will be somewhere in between the two extremes. The actual balance depends on the problem at hand and will likely be estimated from the data.

If X is the event of interest (such as the probability that there will be a tornado at a particular location within the next 30 minutes), then simple Bayesian analysis yields:

$$P(X) = P(X|packed) \times P(packed) + P(X|notpacked) \times (1 - P(packed)) \quad (1)$$

The first quantity, $P(X|packed)$: the probability of lightning given that all the grid points associated with a thunderstorm move together – can be estimated quite readily using numerical integration of the probability distribution corresponding to motion vectors that will yield a lightning strike at this location based on where the threats are currently present.

The third quantity, $P(X|notpacked)$: the probability of lightning given that the grid points associated with a thunderstorm move independently (in other words, the state transition from a cell to its neighbor is independent of other cells in a thunderstorm) can be computed through probability analysis:

$$P(X|notpacked) = 1 - ((1 - P(x_1)) \times (1 - P(x_2)) \times \vdots (1 - P(x_N))) \quad (2)$$

where $P(x_1)$, $P(x_2)$, etc. are the spatial probabilities that the individual grid points will end up impacting this location. The weighting factor $P(packed)$ needs to be estimated from historical data. We anticipate that the $P(packed)$ will depend upon the type of event under consideration and needs to be estimated from the data – a single number may not work for all problems.

4. Estimating Movement

There are two broad methods of estimating movement: (a) optical flow methods including cross-correlation and spectral methods that minimize a displacement error with sub-grids of two frames of a sequence of images^{9,10} (b) object tracking methods that identify objects in the frames of a sequence and correlate these objects across time.¹¹ The object tracking methods provide the ability to track small-scale features and extract trends associated with those features but miss large-scale transformations. In comparison, the optical flow methods yield more accurate estimates over large scales, but miss fine-scale movements and do not provide the ability to extract trends.¹²

Nevertheless, trends are very important in a data-mining predictive approach. A measured value may bear no correlation with whether a threat will materialize, but sharp increases or decreases of some measurements are often reliable indicators of coming threats. Thus, even though optical flow methods can yield more accurate estimates of movement than object-tracking methods, optical flow methods cannot be directly applied to problem domains in which object-specific motion and trend information is critical.

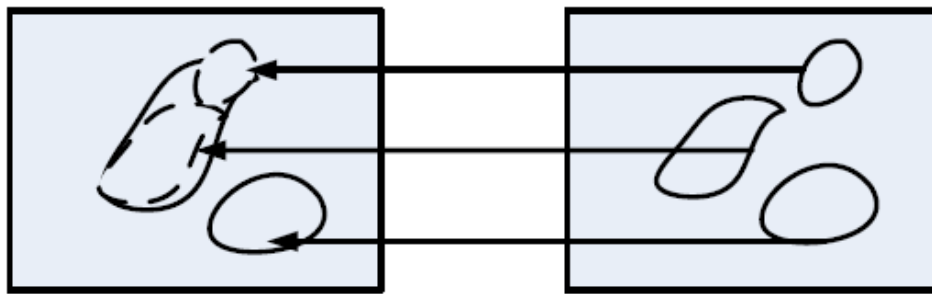


Fig. 4. A hybrid tracking technique of correlating objects in the current frame to pixels in the previous frame can yield trend information as well as the high accuracy associated with optical flow methods.

One way to achieve the high accuracy of optical flow methods while retaining the trending capability of object-based tracking is to create a hybrid technique. It is possible to find objects by clustering the input features using a multi-scale hierarchical approach.¹² This hybrid technique will not correlate objects across frames as in a typical object-tracking scenario. Instead, the objects should be correlated in the current frame with the images in the previous frame using the current shape of the object itself as the domain in which displacement error will be

minimized. To extract trends, the object domain can then be simply displaced by the temporally adjusted amount in previous frames and the change in statistical properties of the object computed and used for prediction. Such a hybrid technique, as Figure 4 illustrates, neatly sidesteps the problems associated with splitting and merging that are commonly associated with object-tracking methods. Motion estimates from pairs of frames are smoothed across time using a Kalman filter.¹³

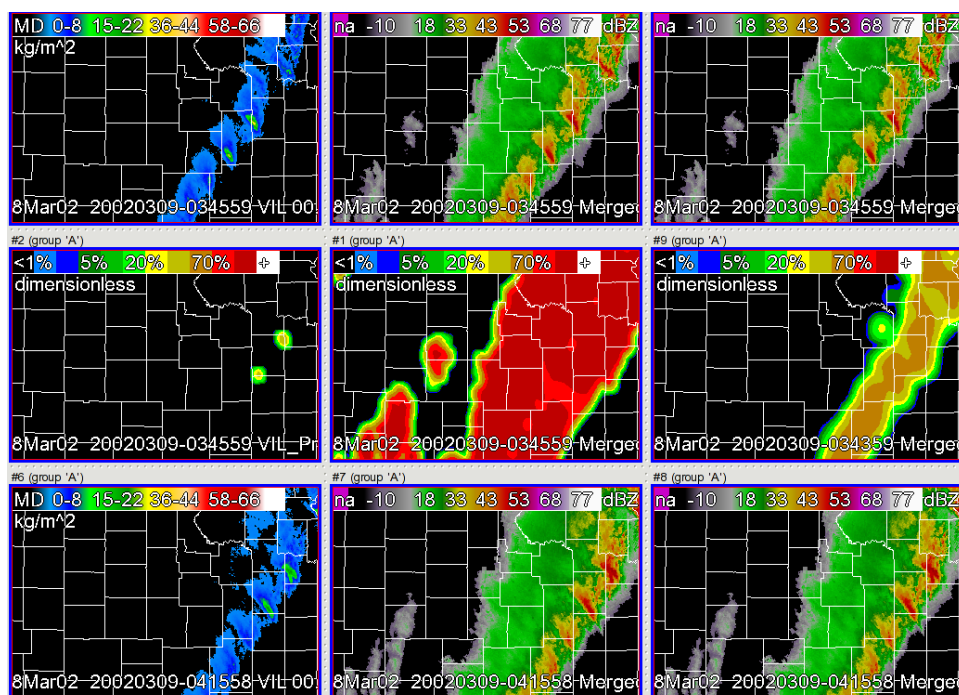


Fig. 5. Top row: Values of Vertical Integrated Liquid (VIL) in kg/m^2 and reflectivity (Z) in dBZ at $T=t_0$. Middle row: Probabilistic forecast of VIL i_{20} , Z_{i0} , Z_{i30} at $T=t_0+30$. Bottom row: Actual values of VIL and Z at $T=t_0+30$.

5. Creating a spatio-temporal probability forecast

In the short term (under 60 minutes), extrapolation forecasts of storm location often provide reasonable results. Therefore, a data mining approach to probabilistic forecasts of severe weather hazards holds promise of being able to predict where already-occurring severe weather is likely to be. At the same time, research studies have shown that it is possible to predict, with a high degree of skill, which storms are likely to initiate lightning (one measure of storm severity) and where new storms are likely to form. What has been missing is a formal mechanism of combining these two disparate sets of studies. Our spatiotemporal formalism provides the framework to combine the two probabilities to yield, for example, a lightning probability forecast.

The suggested approach is to apply the various components described above to train the system as follows:

- Cluster the input remotely sensed data into candidate regions
- Using the hybrid motion estimation technique, associate current severe weather to threat signals that occurred N minutes ago.
- Train the associated threat signals to the severe weather activity using a neural network capable of providing predictive probabilities.
- Use the spatiotemporal framework and the motion estimates derived from the hybrid technique to estimate spatial probabilities.
- Estimate the weighting factor P (packed) to create optimal (in terms of reliability and sharpness) probability factors.

The trained model and pattern recognition techniques can be applied in real-time on routinely arriving satellite, radar and model data to predict up to N minutes into the future. The resulting probabilistic forecasts can be evaluated using three metrics: (a) reliability (b) sharpness (c) Radar operating characteristics (ROC) curves, created by setting risk thresholds that vary from 0 to 1 and determining the probability of detection and false alarm rate on a grid-point by grid-point basis.

6. Results and Conclusion

The technique of the paper was applied to predicting the spatial distribution of vertical integrated liquid (VIL;¹⁴) and radar reflectivity 30 minutes into the future. Examples of the original, forecast and actual images are shown in Figure 5. The reliability and sharpness diagrams when evaluated over a 6 hour period is shown in Figure 6.

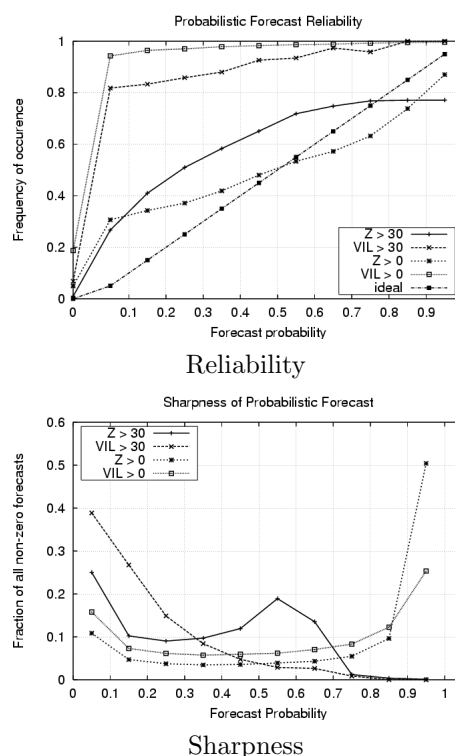


Fig. 6. Reliability and sharpness of Vertical Integrated Liquid (VIL in kg/m^2) and radar reflectivity (Z in dBZ) forecasts for a Mar 8, 2002 case in the central United States.

Reflectivity forecasts were sharp, but not reliable. The reflectivity forecasts were usually underestimates of the true probability. The VIL probability forecasts were both reliable and sharp, so a probabilistic forecast of VIL would be of high utility to users of weather information. The lead time, of 30 minutes, is higher than possible using deterministic forecast techniques. Thus, the technique described in this paper works for severe weather situations (such as VIL), but not for any weather (as highlighted by the poor performance on reflectivity). Future work will involve testing against more useful diagnostics of severe weather – the initiation of lightning and

the potential for a tornado.

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

1. A. H. Murphy, The value of climatological categories and probabilistic forecast in the cost-loss ratio situations, *Monthly Weather Review*, 803 (1977).
2. I. Adrianto, T. M. Smith, K.A.Scharfenberg and T. Trafalis, Evaluation of various algorithms and display concepts for weather forecasting, in *21st Int'l Conf. on Inter. Inf. Proc. Sys. (IIPS) for Meteor., Ocean., and Hydr.*, (Amer. Meteor. Soc., San Diego, CA, Jan. 2005).
3. M. Richard and R. P. Lippman, Neural network classifiers estimate bayesian a posteriori probabilities, *Neural Computation* **3**, 461 (1991).
4. C. Bishop, *Neural Networks for Pattern Recognition* (Oxford, 1995).
5. J. Platt, *Advances in Large Margin Classifiers* (MIT Press, 1999), ch. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods.
6. J. Prevost and R. Popescu, Constitutive relations for soil materials, *Electronic Journal of Geotechnical Engineering* (1996).
7. M. Oliver and R. Webster, Kriging: a method of interpolation for geographical information systems, *Int. J. Geographical Information Systems* **4**, 313 (1990).
8. P. Goovaerts, Geostatistical analysis of disease data: visualization and propagation of spatial uncertainty in cancer mortality risk using poisson kriging and p-field simulation, *Int. J. Health Geography* **5** (2006).
9. J. Barron, D. Fleet and S. Beauchemin, Performance of optical flow techniques, *Int'l J. Comp. Vis.* **12**, 43 (1994).
10. J. Tuttle and R. Gall, A single-radar technique for estimating the winds in tropical cyclones, *Bull. Amer. Met. Soc.* **80**(Apr. 1999).
11. M. Dixon, Automated storm identification, tracking and forecasting – a radar-based method, PhD thesis, University of Colorado and National Center for Atmospheric Research 1994.
12. V. Lakshmanan, R. Rabin and V. DeBrunner, Multi-scale storm identification and forecast, *J. Atm. Res.*, 367(July 2003).
13. R. Kalman, A new approach to linear filtering and prediction problems, *Trans. ASME – J. Basic Engr.*, 35(March 1960).
14. D. R. Greene and R. A. Clark, Vertically integrated liquid water – A new analysis tool, *Mon. Wea. Rev.* **100**, 548 (1972).