1.1 An Experimental Severe Weather Nowcast Algorithm Based on a Back-propagation Neural Network and A Comparison With An Algorithm Based on Multiple Linear Regression

Yerong Feng*
Guangdong Provincial Meteorological Observatory
Guangzhou, 510080, China

David H. Kitzmiller**
Meteorological Development Laboratory
Office of Science and technology
National Weather Service, NOAA
Silver Spring, MD 20910,USA

1. INTRODUCTION

The information processing system of the Weather Surveillance Radar 1988 (Doppler) (WSR-88D) incorporates a number of automated algorithms for interpreting reflectivity and velocity data in terms of the threat of severe weather. These include the Severe Weather Potential (SWP) product (Kitzmiller et al. 1995), which is a numerical index proportional to the probability of large hail, damaging convective wind gusts, and/or tornadoes. The SWP is based solely on volumetric radar reflectivity information. As WSR-88D information was integrated with other data streams in the Advanced Weather Interactive Processing System (AWIPS), it became possible to refine the radar-only algorithms through the addition of information on the near-storm environment, such as static stability and upper-level winds (Kitzmiller and Breidenbach, 1993). Versions of this refined algorithm, referred to here as the Severe Weather Threat Index (SWTI), have been incorporated with the AWIPS System for Convective Analysis and Nowcasting (SCAN, Smith et al. 1998).

Both SWP and SWTI were developed through statistical regression, by correlating a combination of radar-observed and environmental indices to severe storm occurrence. The development sample was rather limited, including data from ~7000 storms (both severe and nonsevere) observed during the 1980's and early 1990's. Much of the radar data was collected from 12 earlier-generation NWS conventional radars that were specially adapted for automatic volumetric scanning and digital reflectivity processing.

The goals of the current study are to obtain a general algorithm that (1) is based on a much larger data sample collected from most radars within the conterminous U.S., (2) incorporates still more candidate predictors, including some derived from Doppler velocity information, and (3) utilizes statistical prediction methods other than LR, if such methods prove to be superior. In particular, we have tested artificial neural networks (NN) as an alternative to LR. We were motivated by previous findings that many candidate predictors had a highly nonlinear relationship to severe weather relative frequency, and by a desire to incorporate some categorical predictors (e.g., mesocyclone indications) that might not be adequately handled by linear correlation. Though it is possible to adapt linear predictor combinations for nonlinear predictand response (see for example Charba 1979), most available methods are subjective and time-consuming.

2. DEVELOPMENT DATA

2.1. Radar data

The radar data were taken from the WSR-88D network within the conterminous United States (CONUS) during the period April-July, 2000-2001. This corresponds to the spring-early summer peak in severe weather activity. Radar umbrellas containing thunderstorm events were identified through automated monitoring of a national radar reflectivity mosaic, and detailed radar information for sites of interest was obtained from a centralized data server. This information included gridded vertically-integrated liquid (VIL), cell-based (storm updraft) VIL, 30-dBZ storm top height, maximum reflectivity, and mesocyclone and tornado vortex signature (TVS) indications. All radar-based

* Corresponding author: Yerong Feng, Guangdong Provincial Meteorological Observatory, 6 Fujin St., Guangzhou, 510080, Peoples Republic of China, <yerong_feng@yahoo.com>
** Current affiliation: Hydrology Laboratory, Office of Hydrologic Development, National Weather Service, NOAA.

predictors were objectively interpolated to a 4 km square horizontal grid extending to a radius of 230 km from the radar. Following the initial WSR-88D SWP convention, an individual storm cell was defined as a square region 28 km on a side (7 X 7 grid boxes) centered on a local maximum in the VIL field. This size approximates the upper limit of the average spatial scale for individual thunderstorm cells (10 to 30 km) as defined Byers and Braham (1949). The probability of severe weather is very low for cells with VIL less than 10 kg m^{-2} (Kitzmiller et al., 1995), and therefore only cells having at least two grid boxes with VIL of 10 kg m^{-2} or more were considered for inclusion in the dataset. For cells that are far from radar site (>=200km) the VIL could be overestimated, so we considered only those cells that were within 90 nautical miles (~167km).

All storm cells meeting the above criteria were initially included. To insure some measure of statistical independence and eliminate repeated sampling of individual cells, the initial sample was culled so that any two cells in the final dataset were separated from each other by at least 40 km in space, and/or 30 minutes in time. As explained below, these cells were defined as nonsevere or severe based on their proximity to severe local storm reports.

2.2 Upper-air data from numerical models

Another type of predictors (known as environmental predictors) was collected from numerical weather prediction (NWP) model products. The environment predictors were derived from analyses and 6-h or 12-h forecasts of the National Centers for Environmental Prediction (NCEP) Eta model, as archived by the Meteorological Development Laboratory. The Eta model is run four times per day (0000,0600,1200,1800 UTC). To objectively assign environmental variables to individual storm cells, the available ETA model's initial-time analyses and 6-h or 12-h forecasts at 0000 and 1200 UTC were projected to the most close storm cell valid times. For instance, for cell times between 0000 and 02000 UTC, the model data of 12-h forecast initialized at 1200 UTC were applied; for cell times between 0300 and 0400 UTC, the initial-time analyses at 0000 UTC were used; And for cell times between 0500 and 0900 UTC, the 6-h forecast initial at 0000 UTC were used, etc.

2.3 Severe local storm reports

Predictand data come from the reports of severe local storms related to tornadoes, surface wind gusts causing damage or $\geq$ 25 m s^{-1}, and/or large hail ($\geq$ 2cm diameter). These severe weather reports are routinely collected from various sources by NWS Weather Forecast Offices and later processed by the Office of Climate, Water, and Weather Services. These events were assigned to a cell if they occurred within 24km of the cell centroid and were collected from 10 min before to 30 min after the nominal radar observation time. This convention should account for both events in progress and those about to develop. Then the number of associated severe weather reports were summed up and recorded.

To create the predictor-predictand dataset needed in our algorithm development, it was necessary to keep the sampling procedure statistically reliable. We wished to insure that there was a reasonable chance that severe storms were observed as such. A reporting bias toward populated areas is well known (Smith 1999). Therefore only cells passing through populated areas that had yielded multiple reports over the period 1973-1999 were considered in the final statistical dataset. This procedure yielded a data sample of 115,594 storm cells, of which 4.4% were judged to be severe.

3. CANDIDATE SEVERE WEATHER PREDICTORS

A total of 27 predictors, including both radar-based and NWP-based indices, were selected for the SWTI development. The set of predictors included some of VIL-based indices used in the development of the original SWP and SWTI algorithms, additional reflectivity-based predictors yielding information on maximum reflectivity and vertical extent, and Doppler-based predictors. Storm-environment predictors include a number of commonly-used stability indices, as well as others found to be useful in predicting conditional severe storm probability in Model Output Statistics applications (Reap and Foster 1979, Charba 1978). These predictors are among those most strongly correlated with severe weather events within a population of severe and nonsevere thunderstorm cases. A complete list appears in the Table.

The Doppler velocity predictors MESO and TVS indicate mesoscale or microscale rotation within a convective cell, respectively (Robert and White, 1998, Mitchell 1995). For MESO, possible indications are None, UNCO (for horizontal velocity shear with no detectable vertical extent), 3DCO (for horizontal shear with limited vertical extent), and MESO, the last state indicating a mesocyclone with significant vertical extent. The TVS indications ETVS and TVS represent intense horizontal shear

aloft and near the surface, respectively (TVS values other than None are rare).

The VIL indicators GVIL and CVIL are estimates of the mass of water per unit area within a vertical column (for gridded VIL, GVIL), and within the storm updraft column as approximated by the maximum reflectivity at each vertical level (cell-based VIL, CVIL). This mass is estimated from reflectivity through the Marshall-Palmer Z-M relationship. The horizontal area predictors SVG10, SVG20, and SVG30 are simply the number of 4-km grid boxes within the 7x7 box cell region that have GVIL $\geq$10, $\geq$ 20, and $\geq$ 30 kg m^{-2} (Kitzmiller et al. 1995).

Most of the listed storm environment predictors are well-known or self explanatory. The thermal advection indices are based on the wind speed, direction, and vector change within a vertical layer of the atmosphere (Kitzmiller and McGovern 1990). The thermal advection is positive for wind vectors veering (turning clockwise with height) from the bottom to the top of the indicated layer. The "best" lifted index is the one indicating the greatest instability from any parcel within the lowest 4 layers of the Eta model vertical domain.

These predictors were then used as candidate predictors for the development of a forward-selection multiple linear screening regression equation, referred to henceforth LR (Draper and Smith 1985), and as the input nodes in a back-propagation (BP) neural network (NN) algorithm. The severe weather statistical predictand was taken to be 0 for a nonsevere cell and 1 for a severe cell.

4. THE PREDICTION ALGORITHMS

The dataset was separated into two subsets (known as dependent and independent dataset) in a randomly mixed order. The dependent subset contains 86,678 cases, which were used for algorithm development through NN and LR approaches. Both neural network and regression equation were used to relate severe storm occurrence (the predictand) to radar and environmental indices (the predictors). The independent subset contains 28,916 cases, which was applied later for verifying the performances of these two algorithms.

Neural network (NN) techniques have been widely used in many meteorological forecasting problems, such as tornadoes (Marzban and Stumpf, 1996), damaging winds (Marzban and Stumpf, 1998), thunderstorms (McCann, 1992) and quantitative precipitation forecast (Kuligowski and Barros, 1998; Feng and Kitzmiller, 2002). In this experimental work, a three-layer BP neural network was constructed so as to describe the nonlinear statistical relationship that might exist between one-dimension predictand space and multiple-dimension predictor space. As note in the work cited above, this nonlinear effect seemed significant in severe weather nowcasting. Though it is possible to adapt predictors for nonlinear relationships to the predictand, and for nonlinear interactions among predictors (Breidenbach et al. 1995), this approach is partly subjective and can be time-consuming.

The BP has been the most commonly used network architecture in many meteorological problems and its detailed network structure is documented in many references (Bishop, 1995; Hall et al., 1999; Feng and Kitzmiller 2002). The configuration of the BP network used here includes an input layer with 27 input nodes, 1 hidden layer with 10 nodes, and a single output node that produces severe weather potential or threat index in terms of probabilities. The 27 input nodes are all predictor variables discussed in section 3, normalized to the range 0.01-0.99 (instead of 0-1, which can produce singularity problems in computation). The number of output nodes (the predictand) is generally determined by the prediction problem itself, which aims at producing probabilistic forecasts of the severe weather. The number of nodes in the hidden layer is usually determined by trial and error. We chose to use 10 based on the number of input nodes and earlier experimentation (Aviolat et al., 1998; Feng and Kitzmiller 2002).

The set of values in the successive layer H were determined from the values in the previous layer X by:

$$H^T=F(W.X^T+Q^T) \quad (1)$$

where W is a weight matrix logically relates two adjacent layers, Q is a vector containing activation threshold values, F is a sigmoid function, which nonlinearly projects a linear combination of nodes to each node in the follow-up layer, and the superscript T denotes the transpose of a matrix. The NN methodology involved a training process to obtain knowledge from the training data set by determining the network's values in two weight matrices and two activation threshold vectors. This process was done by designing a gradient descent logic and applying the sample data for the purpose of gaining a minimum error between the network output (probability) and severe weather observed (1 for severe and 0 for nonsevere).

For comparison, the same 27 candidate predictors were applied to a linear screening regression process to obtain a linear expression of

severe weather probabilities. The method used to develop a linear regression probabilistic equation of severe weather was similar to that employed in the development of the previous SWP algorithms. The selection procedure yielded the following algebraic relationships between the available predictors and event probability:

SWTI = -15.514 + 2.681 SVG30 + 0.303 SVG10 + 0.109 GVIL + 0.153 MAXREFHGT + 0.281 TT + 0.141 WSPD500 + 2.248 MESO - 0.002 FRZLVL. (2)

Here, the predictors and their units are as shown in the Table. The categorical MESO variable was numerically encoded as 0 for None and as 1 for UNCO, 3DCO, or MESO. This expression explained 15.7% of the predictand variance. The predictors based on gridded VIL had the highest linear correlation to severe storm relative frequency. The presence of storm rotation as indicated by a nonzero MESO is often an indicator of high winds or hail as well as tornadoes, which are less common than the other severe storm phenomena. Finally, storms are more likely to be severe in situations with relatively high static instability, strong mid-tropospheric winds, and generally cold conditions, as indicated by the selection of Total Totals index, 500-hPa wind speed, and freezing level.

A sample severe weather probability analysis is shown in Fig. 1, where probabilities in per cent are plotted next to storm cells within a VIL analysis. The probabilities are based on (2). An alternative presentation consists of text entries in a cell lookup table.

5. RESULTS

The output of each algorithm is probability or potential of severe weather occurrence for an individual cell. To test the reliability of the algorithms, evaluations were made with both the LR and NN algorithm on an independent data sample. The Plains (central United States area) AWIPS SWP algorithm was also evaluated within the same data sample. The observations of severe weather were verified over each 10% probability interval to see how closely the average forecasted probability approximated the actual event relative frequency. In Fig. 2, the observed relative frequency is plotted as a function of the mean forecasted probability. The dotted line represents perfect reliability. Overall, the LR and AWIPS algorithms show some departures from reliability in the probability range 20-40%. The NN curve is closer to the reliability criterion for most probability ranges. All of the algorithms tend to depart from perfect reliability and probabilities above 80%, which are rarely forecasted.

Though SWP algorithms provide probabilistic guidance, their performance is most easily evaluated by examining categorical (severe/nonsevere) forecasts based on the probabilities. Categorical forecasts are generally derived by setting some fixed threshold probability values, and forecasting all storm cells with probabilities at or above the threshold to be severe. The performance of these forecasts may be described by three commonly-used measures, the probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI) (Donaldson et al., 1975; Schaefer, 1990). Let x be the number of severe events correctly forecasted to be severe, y the number of severe events incorrectly forecasted to be nonsevere, and z the number of nonsevere events incorrectly forecasted to be severe. Then the scores are defined by:

POD = x / (x + y) (3)
FAR = z / (x + z) (4)
CSI = x / (x + y + z) (5)

The performance of the neural network and multiple regression algorithms in terms of POD and CSI are shown in Fig. 3. These scores were from a sample of cases outside the development dataset. For both NN and LR algorithms, the CSI curve reaches a peak value between 16-32%. Note that, for probability thresholds within this range, the POD and CSI of the NN were higher than those of LR. These results may confirm our previous expectation that a non-linear approach like NN should be better than a linear approach for the severe weather nowcast problem. For probability thresholds less than 10%, the score differences were smaller.

As has been shown, the yes/no forecast skill is highly dependent on probability threshold. If the threshold SWP value is set very high, not enough severe storms would be detected, while if the threshold is set too low, too many false alarms would be issued. The optimum choice may be setting the threshold as close as possible to the peak CSI value while achieving an acceptable POD. For example, when using 16% as threshold probability, about 50% of the severe cells were detected (POD = 0.50); 70% of the "yes" forecasts were false alarms (FAR = 0.70, not shown). The CSI at the 16% threshold was 0.23.

Another way to evaluate the relative performance of the algorithms is to plot the FAR curves with respect to POD. We determined the false alarm ratio that would be issued by both NN, LR, and AWIPS SWP in the process of achieving the same POD value. As shown in Fig. 4, the NN and LR

algorithms consistently yielded fewer false alarms than did the SWP algorithm. For yes/no probability thresholds yielding lower probabilities of detection (where there are fewer false alarms in general) the NN gave fewer false alarms than did the LR.

6. SUMMARY AND CONCLUSIONS

Two kinds of statistical severe weather potential algorithms were used in this experimental work: the neural network approach and the multiple linear regression approach. The algorithms utilize WSR-88D Doppler Radar data and numerical weather prediction model outputs to provide probabilities of severe weather centered on a convective storm cell, 44 km on a side, 30 min after radar observation.

Although both techniques do not possess high absolute accuracy in identifying severe storms, that is, they do appear to improve on the operational AWIPS SWP algorithm, which was developed from a much smaller sample of observational data. Validation of the NN and LR algorithms on independent data showed that NN approach, with its nonlinear and collective expression of all the candidate predictors, provided higher forecast scores and greater statistical reliability. Given that the NN approach is only minimally more complicated to implement than the LR, it seems that NN is a good candidate for operational use.

Acknowledgements

The completion of this work was possible only with the joint effort of the authors and the cooperation of China Meteorological Administration and National Weather Service under the terms of the China-U.S. Atmospheric Science and Technology Protocol. Data storage and computer resources at the Meteorological Development Laboratory greatly facilitated the work.

REFERENCES

Aviolat, F., T. Cornu, and D. Cattani, 1998: Automatic cloud observations improved by an artificial neural network. *J. Atmos. Oceanic Technol.*, **15**, 114-126

Breidenbach J. P., D. H. Kitzmiller, W. E. McGovern, and R. E. Saffle, 1995: The use of volumetric radar reflectivity predictors in the development of a second-generation severe weather potential algorithm. *Wea. Forecasting*, **10**, 369-379.

Bishop, M. A., 1995: *Neural network for pattern recognition*. Clarendon Press, 482pp.

Byers, H. R., and R. R. Braham Jr., 1949: *The thunderstorm*. U.S. Department of Commerce, Weather Bureau, 287 pp.

Charba, J. P., 1979: Two to six hour severe local storm probabilities: An operational forecasting system. *Mon. Wea. Rev.*, **107**, 268–282.

Donaldson, R. J. Jr., R. M. Dyer, and J. J. Kraus, 1975: An objective evaluator of techniques for predicting severe weather events. *Preprints Ninth Conference on Severe Local Storms*, Norman, Amer. Meteor. Soc., 321-326.

Draper, N. R., and H. smith, 1981: *Applied regression analysis*. John Wiley and Sons,709 pp.

Feng, Y. R., and D. H. Kitzmiller,2002: A comparison between neural network and linear regression approaches to a short-range quantitative precipitation forecasting problem. *Preprints of the 16th Conference on Probability and Statistics in the Atmospheric Sciences*, Orlando, Florida, Amer. Meteor. Soc., J220-J226.

Hall, T., H. E. Brooks, and C. A. Doswell III,1999: Precipitation forecasting using a neural network. *Wea. Forecasting*, **14**, 338-345.

Kitzmiller, D. H., and Mcgovern, W. E. 1990: Wind profiler observations preceding outbreaks of large hail over northeastern Colorado. *Wea. Forecasting*, **5**, 78–88

_____, _____, and R. E. Saffle, 1995: The WSR-88D Severe Weather Potential Algorithm. *Wea. Forecasting*, **10**, 141-159.

_____, and Breidenbach, J. P., 1995: Detection of severe local storm phenomena by automated interpretation of radar and storm environment data. *NOAA Technical Memorandum* NWS TDL 82, National Weather Service, NOAA, U.S. Department of Commerce, 33 pp. [Available from Techniques Development Laboratory, W/OSD2, National Weather Service, 1325 East West Highway, Silver Spring, Md.]

Kuligowski, R. J., and A. P. Barros, 1998: Localized precipitation forecasts from a numerical weather prediction model using artificial neural network. *Wea. Forecasting*, **13**, 1194-1204.

Marzban, C., and G. J. Stumpf, 1996: A neural network for tornado prediction based on Doppler Radar-derived attributes. *J. Appl. Meteor.*, **35,** 617-626.

Marzban, C., and G. J. Stumpf, 1998: A neural network for damaging wind prediction. *Wea. Forecasting*, **13**, 151-163.

McCann, D. W., 1992: A neural network short-term forecast of significant thunderstorms. *Wea. Forecasting*, **7**, 525-534.

Mitchell, E.D., 1995: An enhanced NSSL tornado detection algorithm. *Preprints 27th Conference on Radar Meteorology*, Vail, Amer. Meteor. Soc., 406-408.

Reap, R. M., and D. S. Foster, 1979: Automated 12–36 hour probability forecasts of thunderstorms and severe local storms. *J. Appl. Meteor.*, **18**, 1304–1315.

Robert, R. L., and A. White, 1998: Improvement of the WSR-88D mesocyclone algorithm, *Wea. Forecasting*, **13**, 341-351.

Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570-575.

Smith, P. L., 1999: Effects of imperfect storm reporting on the verification of weather warnings. *Bull. Amer. Meteor. Soc.,* **80**, 1099-1105.

Smith, S. B., J. T. Johnson, R. D. Roberts, S. M. Zubrick, and S. J. Weiss, 1998: The System for Convection Analysis and Nowcasting (SCAN) 1997-1998 field test. *Preprints, 19th Conference on SevereLocal Storms*, Minneapolis, Amer. Meteor. Soc., 790-793.

Table. Candidate predictors of severe local storms.

Abbreviation	Description
Radar-based predictors:	
TVS	Tornado vortex signature (NONE, TVS or ETVS, reduced to 0/1)
MESO	Mesocyclone feature (NONE, UNCO, 3DCO, or MESO, reduced to 0/1)
CVIL	Cell's updraft-integrated liquid, kg m^{-2} (Cell-based VIL)
REFLECTIVITY	Maximum equivalent reflectivity (dBZ)
MAXREFHGT	Height of maximum reflectivity (kft MSL)
STOP	Echo top with reflectivity of 30 dBZ or higher, kft MSL
GVIL	Maximum vertically integrated liquid VIL, kg m^{-2} (Gridded VIL)
SVG10	Number of cell grid boxes with VIL $\geq$ 10 kg m^{-2}
SVG20	Number of cell grid boxes with VIL $\geq$ 20 kg m^{-2}
SVG30	Number of cell grid boxes with VIL $\geq$ 30 kg m^{-2}
Storm environment predictors from Eta model:	
WSPD85	Wind speed at 850mb, m s^{-1}
WSPD70	Wind speed at 700 mb
WSPD50	Wind speed at 500 mb
WSPD30	Wind speed at 300 mb
TOTALS	Total totals index (T_{850} + TD_{850} - $2T_{500}$), ° C
TADV850-500	Thermal advection between 850 and 500 mb, s^{-1}
TADV950-700	Thermal advection between 950 and 700 mb
LPRATE7-5	Lapse rate between 700 and 500mb, °C km^{-1}
MEAN RH	Surface to 500mb mean relative humidity, %
FRZLVL	Freezing level height, m MSL
CAPE	Convective available potential energy (J m^{-2}
BESTLI	Best lifted index (°C)
LI500	Surface/500 mb lifted index, °C
BRN	Bulk Richardson number (dimensionless)
SFC TOTALS	Surface total totals index (T_{1000} + TD_{1000} - $2T_{500}$), °C
U500	U-component of wind at 500 mb, m s^{-1}
THICK1000-500	Thickness between 1000 and 500 mb, m

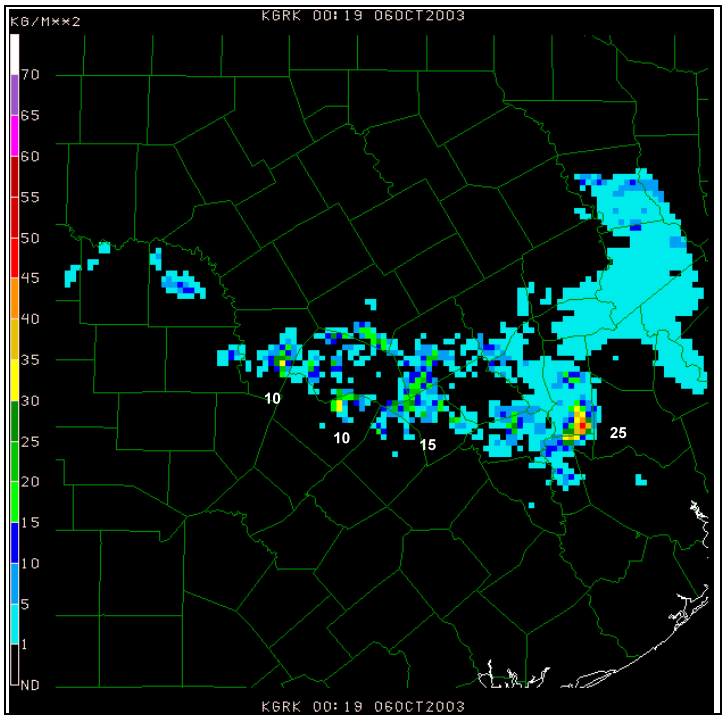

Figure 1. Vertically-integrated liquid (VIL) analysis over southeastern Texas with severe local storm probabilities plotted next to storm cells.

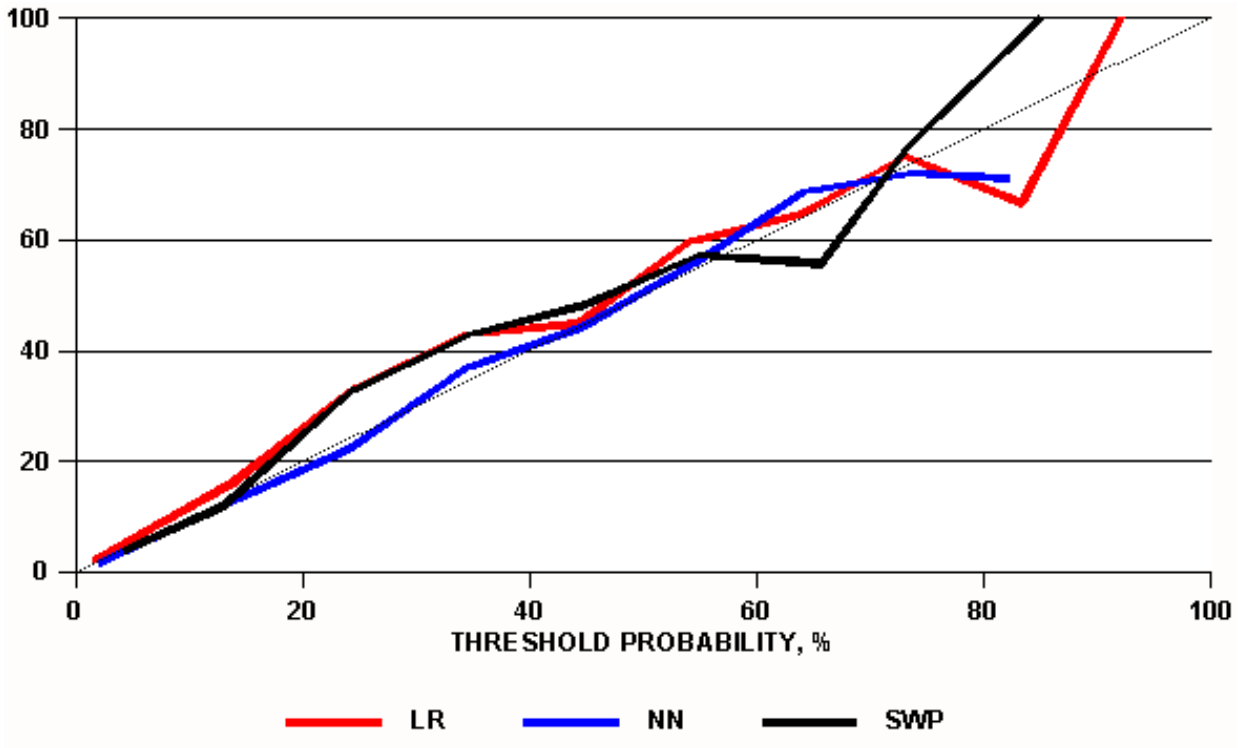

Figure 2. Observed relative frequency of severe weather as a function of probability forecasts from LR (red), NN (blue), and operational AWIPS SWP algorithm (black).

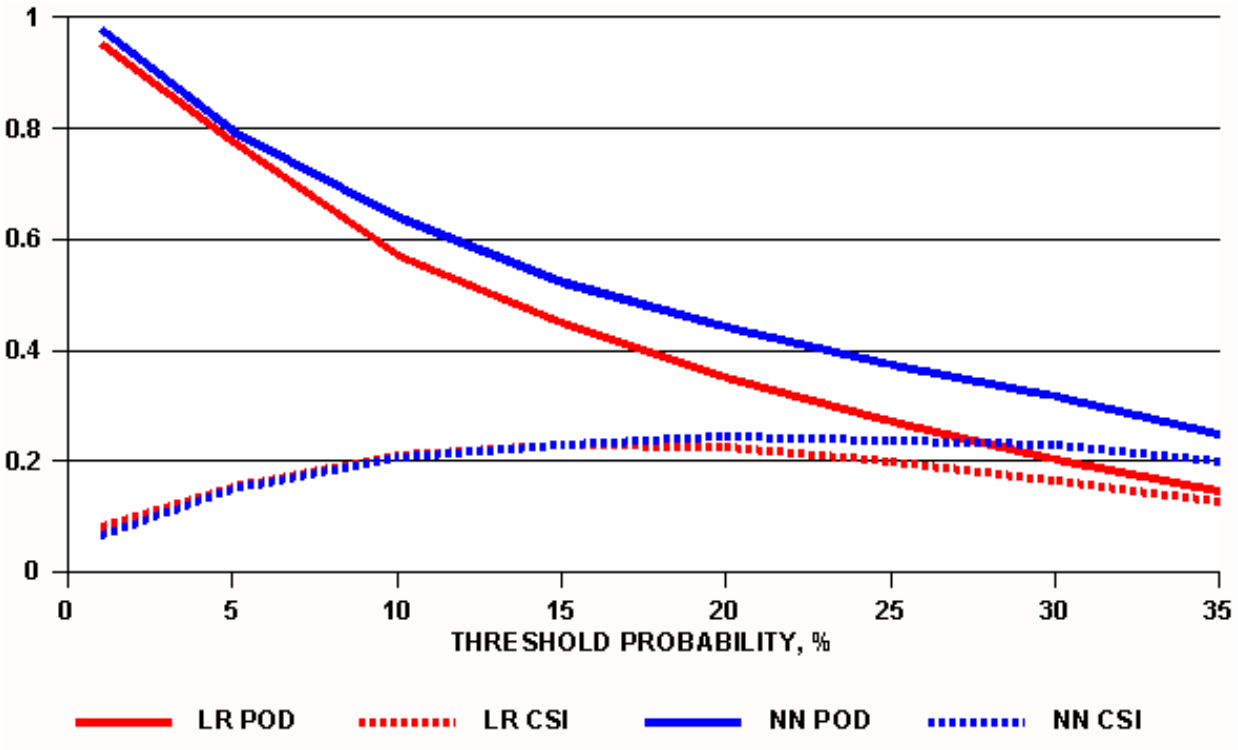

Figure 3. Probability of detection (POD) and critical success index (CSI) as functions of regression and neural network probability forecasts, independent data sample.

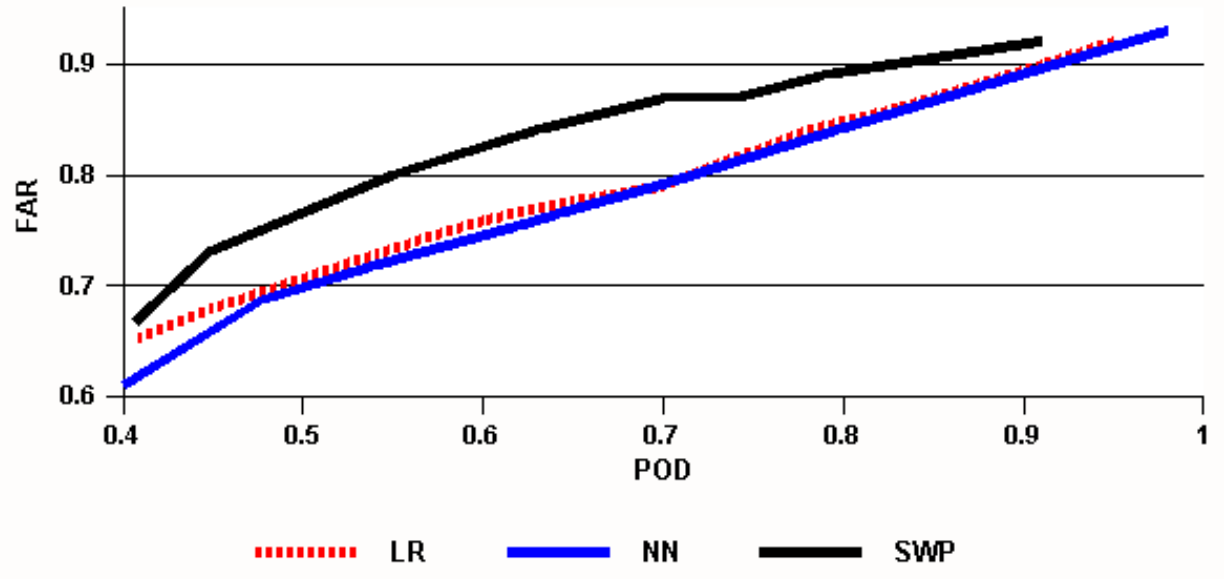

Figure 4. False alarm ration (FAR) as a function of probability of detection (POD) for AWIPS SWP, neural net (NN), and linear regression (REG) algorithms, within the independent dataset.