USING A KNOWLEDGE-BASED SYSTEM TO PREDICT THUNDERSTORMS

Harvey Stern*
Victorian Regional Office, Bureau of Meteorology, Australia

1. Introduction

Treloar and Stern (1993) developed a climatology of Victorian severe thunderstorms, stratifying the data according to synoptic type. The basis for the synoptic types was the direction, strength and curvature of the surface flow. Some broad generalisations were derived. These were that:

- Severe local wind damage, including that caused by tornadoes, is most frequent during the months of November, December, and January, between the hours of 1400 and 1800, and in association with strong cyclonic NNW or NNE flow.
- Large hail is most frequent during the months of November and December, between the hours of 1400 and 1800, and in association with strong cyclonic NNW flow.
- Flash flooding is most frequent during the months of November and December, between the hours of 1400 and 1600, and in association with strong cyclonic NNE flow and moderate cyclonic NNW flow.

2 Determining the Synoptic Types

The synoptic types over the region (refer to Figure 1) were determined as follows:

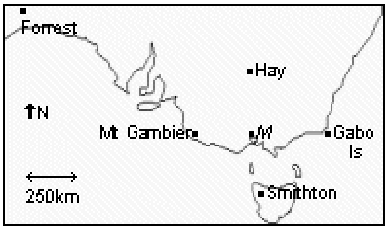


Figure 1 Diagram depicting the grid of locations used to determine synoptic characteristics around Melbourne (*M*).

The strength of the flow is divided into four categories. Defining a and b, respectively, as the 0900 hours EST pressure differences:
a: Smithton [41°S 145°E]-Hay [35°S 145°E]; and,
b: Gabo Is [38°S 150°E]- Mt Gambier [38°S 141°E],
which reflect easterly and northerly gradient wind components, the categories are:

(1) light L, where $a^2+b^2(\text{hPa})^2 \leq 1$;
(2) weak W, where $16 \geq a^2+b^2(\text{hPa})^2 > 1$;
(3) moderate M, where $81 \geq a^2+b^2(\text{hPa})^2 > 16$; and,
(4) strong S, where $a^2+b^2(\text{hPa})^2 > 81$.

Where the strength is *L,* the direction is said to be variable (V); otherwise:

(a) where $a > 0$ the direction of the surface flow is $(\pi/2)$-atan(b/a) divided into 4 octants (NNE,ENE, ESE,SSE); otherwise,

**Corresponding author address:* Harvey Stern, Victorian Regional Office, Bureau of Meteorology, Box 1636M, Melbourne, 3001, Australia; e-mail: H.Stern@bom.gov.au

(b) where a≤0, the direction of the surface flow is (3π/2)-atan(b/a) divided into 4 octants (SSW, WSW,WNW,NNW); unless,
(c) the direction of the surface flow places it on the boundary between two octants, in which case the direction is the first of the series: NNW,NNE,WNW,ENE,WSW,ESE,SSW,SSE.

The cyclonicity of the flow is determined by whether or not the pressure at Melbourne [38°S 145°E] is greater than that at Forrest [31°S 128°E]. If (Melbourne pressure)>(Forrest pressure) then the flow is anticyclonic (*A*); otherwise, it is cyclonic (*C*).

3. Re-Deriving the Synoptic Climatology

The climatology has since been re-derived (Table 1, Figure 2, Figure 3) specifically for the Melbourne metropolitan area, drawing upon data from the Melbourne Central Business District (CBD), Melbourne Aerodrome, Moorabbin Aerodrome and Laverton. It encompasses all thunderstorm occurrences (not just severe thunderstorms). Cyclonic flow from the ENE, NNE or NNW, is most likely to be associated with thunderstorms, while thunderstorms are unlikely to be associated with anticyclonic flow.

It is pertinent to state that a thunderstorm climatology derived from data over an area will over-estimate the frequency of occurrence of thunderstorms for a point. This happens in much the same manner as return periods of extreme rainfall events derived for a point are greater than corresponding return periods for an area. For example, the return period for a 100 mm fall over 24 hours is approximately 50 years for occurring *at a single point* in the Melbourne CBD, but is only about 5-10 years for occurring *somewhere* over the entire Melbourne Metropolitan area (Figure 4).

Table 1 The synoptic types: L, W, M, and S correspond to light, weak, moderate and strong flow; V, NNW, WNW, etc. correspond to variable flow, and flow from the eight octants; and, *C* and *A* correspond to cyclonic and anticyclonic flow. Frequency of thunderstorms over the Melbourne Metropolitan Area for each type is given in brackets.

Synoptic Type	Flow Strength	Flow Direction	Flow Cyclonicity
1(18%)	*L*	V	*C*
2(3%)	*L*	V	*A*
3(29%)	*W*	NNW	*C*
4(5%)	*W*	NNW	*A*
5(14%)	*W*	WNW	*C*
6(1%)	*W*	WNW	*A*
7(9%)	*W*	WSW	*C*
8(<1%)	*W*	WSW	*A*
9(5%)	*W*	SSW	*C*
10(1%)	*W*	SSW	*A*
11(4%)	*W*	SSE	*C*
12(<1%)	*W*	SSE	*A*
13(8%)	*W*	ESE	*C*
14(1%)	*W*	ESE	*A*
15(23%)	*W*	ENE	*C*
16(1%)	*W*	ENE	*A*
17(26%)	*W*	NNE	*C*
18(5%)	*W*	NNE	*A*
19(22%)	*M*	NNW	*C*
20(5%)	*M*	NNW	*A*
21(14%)	*M*	WNW	*C*
22(1%)	*M*	WNW	*A*
23(4%)	*M*	WSW	*C*
24(<1%)	*M*	WSW	*A*
25(5%)	*M*	SSW	*C*

Synoptic Type	Flow Strength	Flow Direction	Flow Cyclonicity
26(<1%)	*M*	SSW	*A*
27(4%)	*M*	SSE	*C*
28(<1%)	*M*	SSE	*A*
29(3%)	*M*	ESE	*C*
30(<1%)	*M*	ESE	*A*
31(16%)	*M*	ENE	*C*
32(3%)	*M*	ENE	*A*
33(32%)	*M*	NNE	*C*
34(6%)	*M*	NNE	*A*
35(16%)	*S*	NNW	*C*
36(6%)	*S*	NNW	*A*
37(5%)	*S*	WNW	*C*
38(<1%)	*S*	WNW	*A*
39(4%)	*S*	WSW	*C*
40(<1%)	*S*	WSW	*A*
41(6%)	*S*	SSW	*C*
42(<1%)	*S*	SSW	*A*
43(7%)	*S*	SSE	*C*
44(<1%)	*S*	SSE	*A*
45(3%)	*S*	ESE	*C*
46(2%)	*S*	ESE	*A*
47(22%)	*S*	ENE	*C*
48(7%)	*S*	ENE	*A*
49(28%)	*S*	NNE	*C*
50(9%)	*S*	NNE	*A*

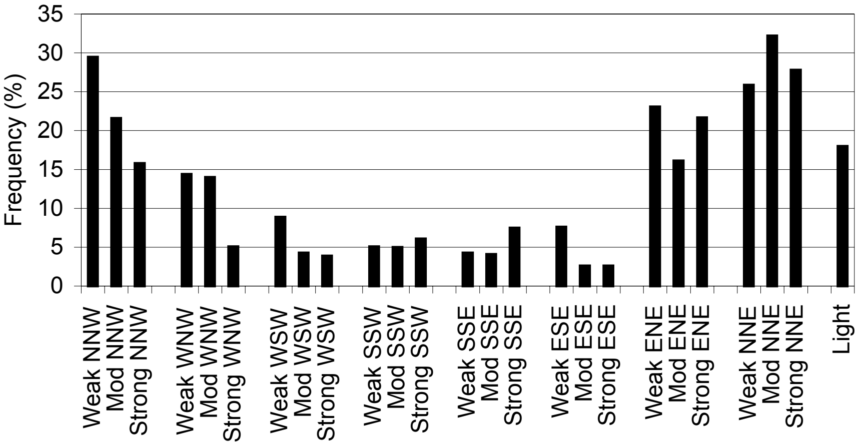


Figure 2 Frequency (%) of thunderstorms associated with each direction for weak (left column), moderate (middle column), and strong (right column) cyclonic synoptic flow, and also for light and variable cyclonic flow.

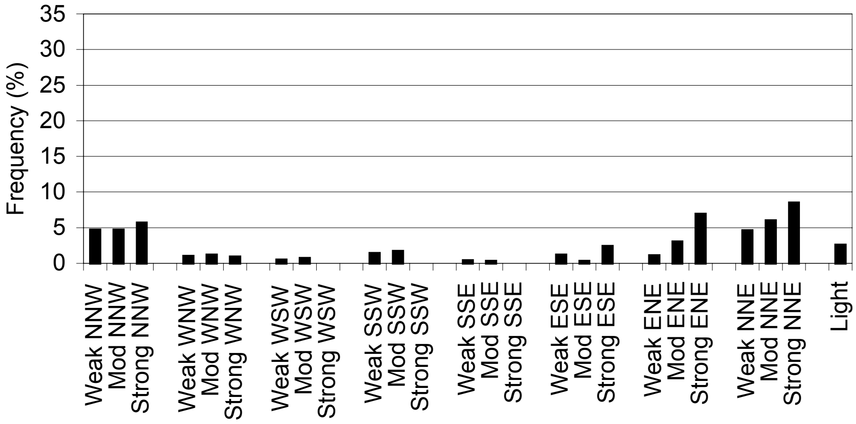


Figure 3 Frequency (%) of thunderstorms associated with each direction for weak (left column), moderate (middle column), and strong (right column) anticyclonic synoptic flow, and also for light and variable anticyclonic flow.

4. Linear Relationship

The results of an analysis of the linear relationship between the predictand (Probability of Thunderstorms - *PoTS*), and the predictors (various combinations of *Sine Day of Year, Cosine Day of Year, Precipitation Amount, Occurrence of Precipitation,* and *Strength of Flow*), carried out on 40 years of unstratified data (1961-2000), are now discussed.

The most significant predictor is √*Precipitation Amount*, which suggests that the greater the amount of precipitation, the higher is the likelihood of thunderstorms. Both the partial correlation coefficients between *PoTS* and the combinations *(√Precipitation Amount x Sine Day of Year)* and *(√Precipitation Amount x Cosine Day of Year)* are positive, suggesting that this relationship is strongest during the late summer/early autumn.

Also highly significant is the predictor, *Occurrence of Precipitation*, suggesting that the greater the likelihood of precipitation, the higher is the likelihood of thunderstorms. Both the partial correlation coefficients between *PoTS* and the combinations (*Occurrence of Precipitation x Sine Day of Year*) and (*Occurrence of Precipitation x Cosine Day of Year*) are positive, suggesting that this relationship (also) is strongest during the late summer/early autumn.

The partial correlation coefficient between *PoTS* and *Strength of Flow* is negative, identifying how weaker synoptic flow is more likely to be associated with thunderstorms.

Table 2 Partial (Linear) Correlation Coefficients for the Predictors (in order of significance) - Unstratified Development Data (1961-2000).

Significance Data: Predictor:	Partial Correlation Coefficient	Regression Coefficient	t (14600)	Significance
√Precipitation Amount	+0.189	+0.267	23.26	<0.01%
√Precipitation Amount X Cosine Day of Year	+0.0921	+0.147	11.17	<0.01%
Occurrence of Precipitation	+0.0407	+0.0578	4.92	<0.01%
Occurrence of Precipitation X Cosine Day of Year	+0.0358	+0.0629	4.32	<0.01%
Strength of Flow	-0.0292	-0.0284	-3.53	0.04%
Cosine Day of Year	+0.0287	+0.0342	3.46	0.05%
√Precipitation Amount X Sine Day of Year	+0.0260	+0.0423	3.14	0.17%
Occurrence of Precipitation X Sine Day of Year	+0.0168	+0.0302	2.03	4.21%
Sine day of Year	+0.0109	+0.0130	1.31	18.90%

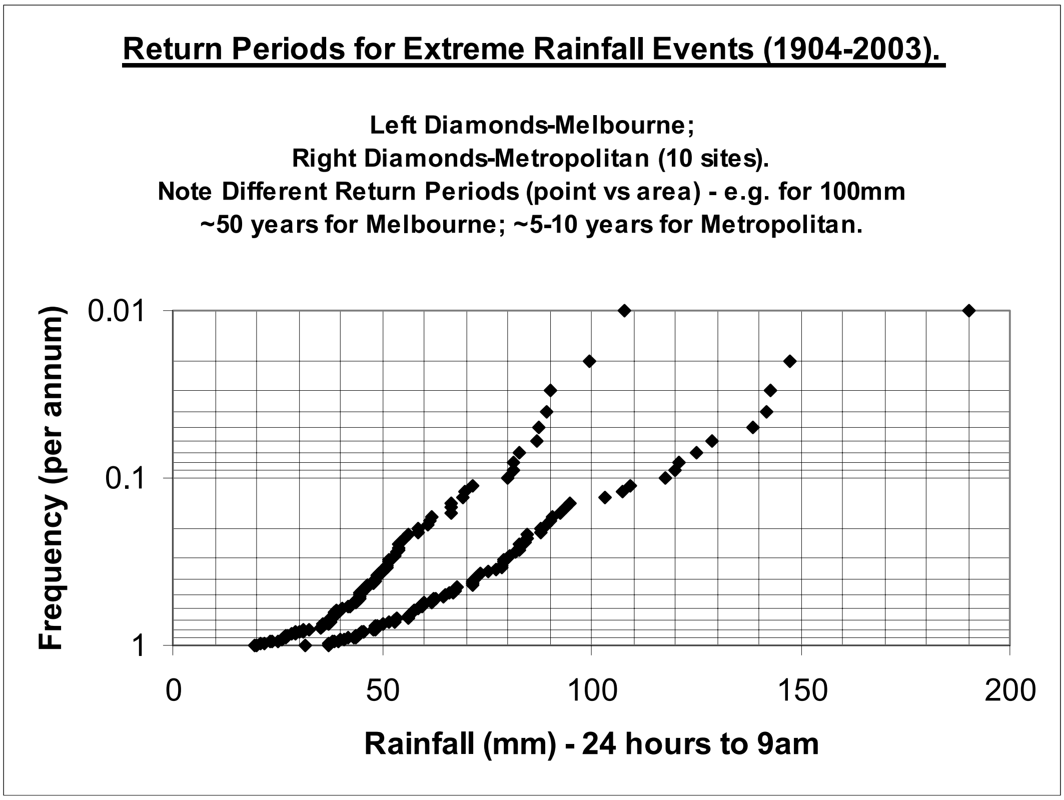


Figure 4 Return Periods for Extreme Rainfall Events.

5. Logistic Regression

Utilising the Treloar and Stern (1993) system for synoptic typing, Stern and Parkyn (1998, 1999, 2000, 2001) derived techniques for the prediction of fog and low cloud by applying logistic regression to synoptically stratified data. The application of logistic regression is appropriate for estimating the probability of occurrence of a particular weather element because the predicted values for the dependent variable will never be less than or equal to 0, nor greater than or equal to 1, regardless of the values of the independent variables. This is accomplished by applying the following regression equation

$$y=(\exp(a+\Sigma b_i x_i))/(1+(\exp(a+\Sigma b_i x_i)))$$

where 'y' is the dependent variable, the x_i are the independent variables, and a and the b_i are constants. In operation, where 'y' is a yes/no variable, the equation yields the probability of occurrence of a particular phenomenon.

6. Predicting Thunderstorm Likelihood

Over recent years, Stern (2002, 2003, 2004) has been involved in developing a knowledge based weather forecasting system (http://www.weather-climate.com/knowledge.html). This system generates forecasts for a range of weather elements, including the likelihood of thunderstorms. A logistic model is used by the system to predict the likelihood of thunderstorms. In operation, the system's *Quantitative Precipitation Forecast (QPF),* and its *Probability of Precipitation (PoP)* estimate, are fed into a set of prediction equations, developed by applying logistic regression to sets of synoptically stratified data (1961-2000), to yield an estimate of the *PoTS*.

The output of both the worded and the Terminal Aerodrome Forecast (TAF) components of the knowledge based system (Figure 5) depend upon whether or not preset cut-off values of *PoTS* have been exceeded. Critical Success Index (CSI) values, derived using the development data (Figure 6), suggest a cut-off in the vicinity of 20% (where there is a maximum CSI of 27%).

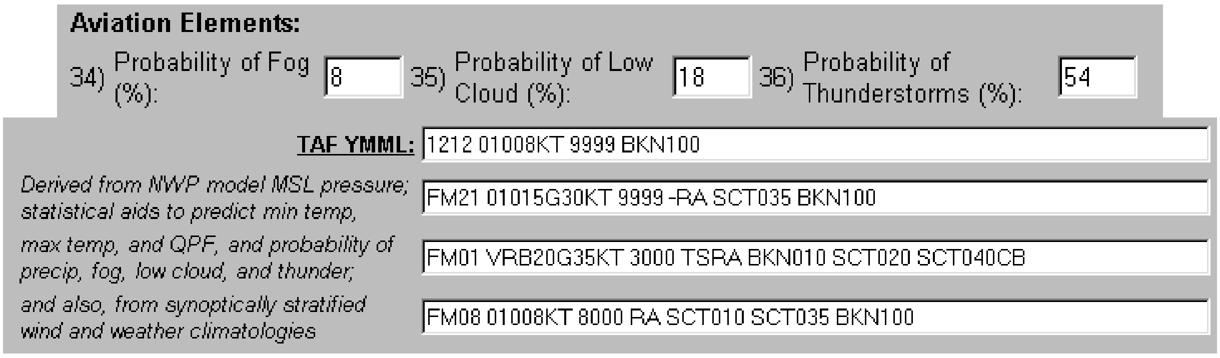


Figure 5 Extracts from the output of the knowledge-based system (Web Reference: http://www.weather-climate.com/knowledge.html).

A comparison was made between Probability of Detections (PODs) and False Alarm Ratios (FARs) for different thunderstorm probability cut-off criteria using the development data (Figure 7). This comparison shows that, at 20%, a POD of 49% and a FAR of 63% would result. These figures are comparable with the overall performance figures achieved by the official TAFs issued by the Bureau of Meteorology (BoM) Victorian Regional Forecasting Centre (RFC) during 2003 (CSI=29%; POD=54%; FAR=61%).

The attributes diagram (Figure 8), which depicts the relationship between observed thunderstorm frequency and the corresponding frequency distribution of probability of thunderstorm estimates, shows that the relationship is linear.

7. Preliminary Evaluation

A one-year (2003) preliminary test of the system's performance using independent data was conducted. These independent data were obtained from the output of the BoM Global Numerical Weather Prediction (NWP) Model and they were used to generate TAFs for Melbourne Aerodrome, which were evaluated. The performance, as tested using these independent data, proved to be inferior to that carried out using the development data. For example, applying a cut-off of 20% to the independent data, the POD was 31%, somewhat lower than the 49% achieved using the development data, and also lower than the 54% achieved by the official TAFs issued by the BoM Victorian RFC.

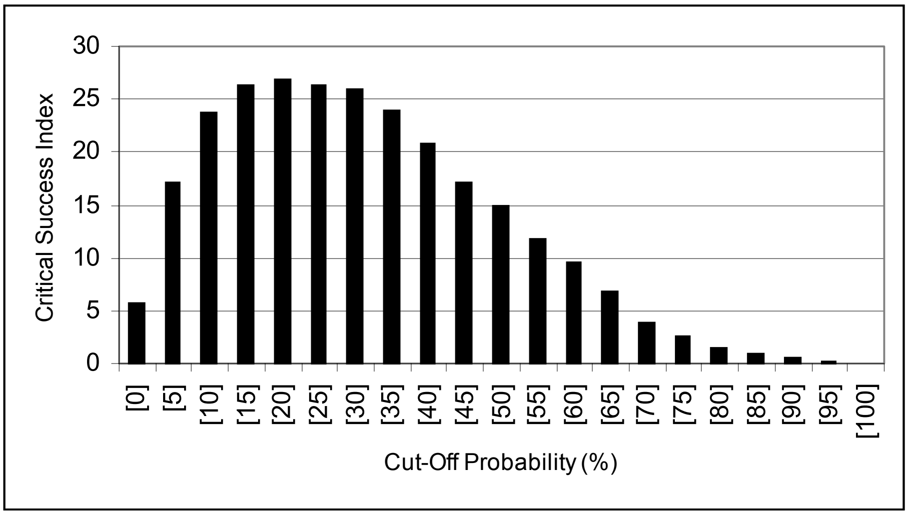


Figure 6 Critical Success Index (CSI) for various cut-off probabilities (Brier Skill Score +0.21).

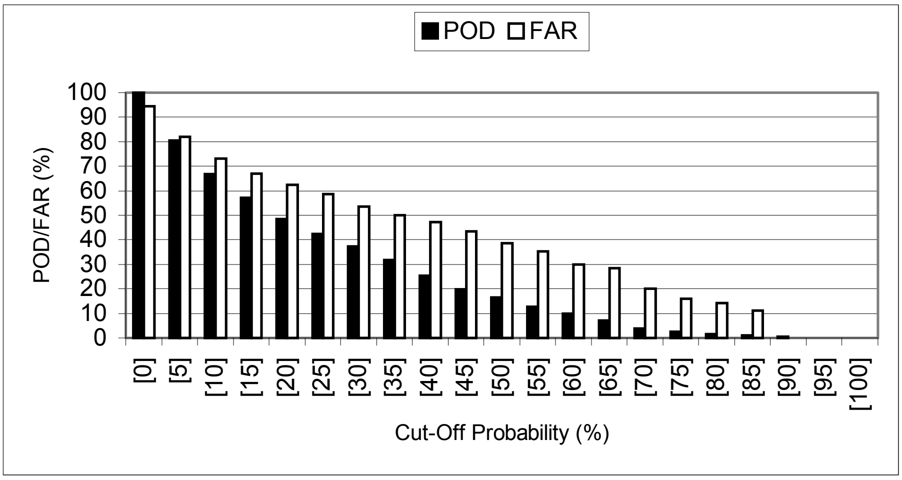
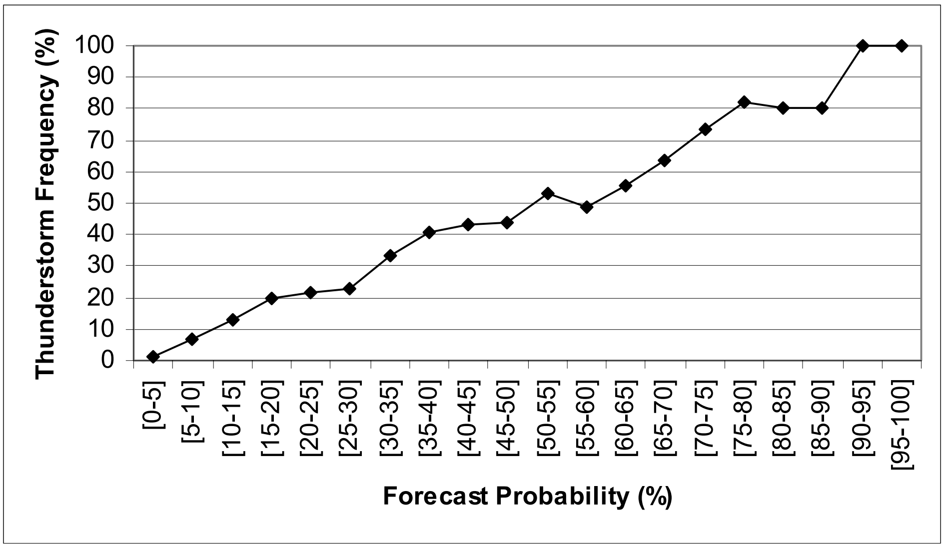


Figure 7 A comparison between Probability of Detections (PODs) and False Alarm Ratios (FARs) for different thunderstorm probability cut-off criteria.

8. Planned Future Work

Firstly, further evaluation is planned, this time operating the knowledge based system under the assumption of the "perfect prog", and using actual 2003 observational data as input. This should provide a measure of the stability of the prediction equations, and also highlight any inadequacies in the NWP model output.

Secondly, Hall et al. (1997) previously have achieved considerable success with their neural network developed for PoP estimates and for QPFs over the Dallas-Fort Worth (Texas) area. In the context of the present work, preliminary results from an exercise involving the application of Artificial Neural Networks (ANNs) to thunderstorm prediction (employing the Software Package Statistica 6 (http://www.statsoft.com)) show that non-linear models do not always outperform linear models. Nevertheless, in one experiment, of the ANNs developed on types with NNE cyclonic flow (NNE is the direction associated with the highest frequency of thunderstorms), the "best" model proved to be a 4-Layer Perceptron with 5 inputs (*Cosine Day of Year, QPF, QPF x Cosine Day, PoP, PoP x Cosine Day*), 12 nodes at Layer 2, and 8 nodes at Layer 3 (Figure 9). Its predictions recorded a relatively high Brier Skill Score of +0.37, only slightly below the Logistic Model's +0.38. It is planned to further investigate the potential application of ANNs to thunderstorm prediction.



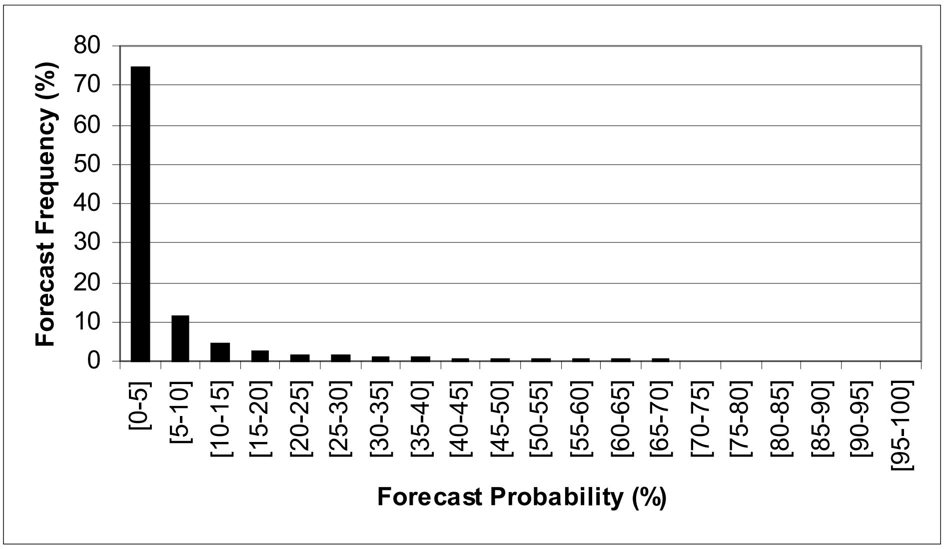


Figure 8 Attributes diagram (observed thunderstorm frequency vs forecast thunderstorm probability and predictive distribution - frequency of forecasts).

9. References

Hall T, Brooks HE, and Doswell III CA (1997) Precipitation forecasting using a neural network. http://www.nssl.noaa.gov/users/brooks/public_html/hall/neural.html

Stern H (2002) A knowledge-based system to generate internet weather forecasts. *18th Conference on Interactive Information and Processing Systems, Orlando, Florida, USA 13-17 Jan., 2002.*

Stern H (2003) Progress on a knowledge-based internet forecasting system. *19th Conference on Interactive Information and Processing Systems, Long Beach, California, USA 9-13 Feb., 2003.*

Stern H (2004) Incorporating an ensemble forecasting proxy into a knowledge based system. *20th Conference on Interactive Information and Processing Systems, Seattle, Washington, USA 11-15 Jan., 2004.*

Stern H and Parkyn K (1998) Synoptic climatology of fog at Melbourne Airport. *Abstracts, ANZ Climate Forum, Perth, 30 Nov.-2 Dec., 1998.*

Stern H and Parkyn K (1999) Predicting the likelihood of fog at Melbourne Airport. *8th Conference on Aviation, Range and Aerospace Meteorology, Amer. Meteor. Soc., Dallas, Texas, 10-15 Jan., 1999.*

Stern H and Parkyn K (2000) Low cloud at Melbourne Airport: A synoptic climatology leading to a forecasting technique. *AMOS Annual Conference, Melbourne, 7-9 Feb., 2000.*

Stern H and Parkyn K (2001) A web-based Melbourne Airport fog and low cloud forecasting technique. *2nd Conference on Fog and Fog Collection, St John's, New Foundland, Canada 15-20 Jul.,2001.*

Treloar ABA and Stern H (1993) A climatology and synoptic classification of Victorian severe thunderstorms. *4th International Conference on Southern Hemisphere Meteorology and Oceanography, March 29 to April 2, 1993, Hobart, Australia, American Meteorological Society.*

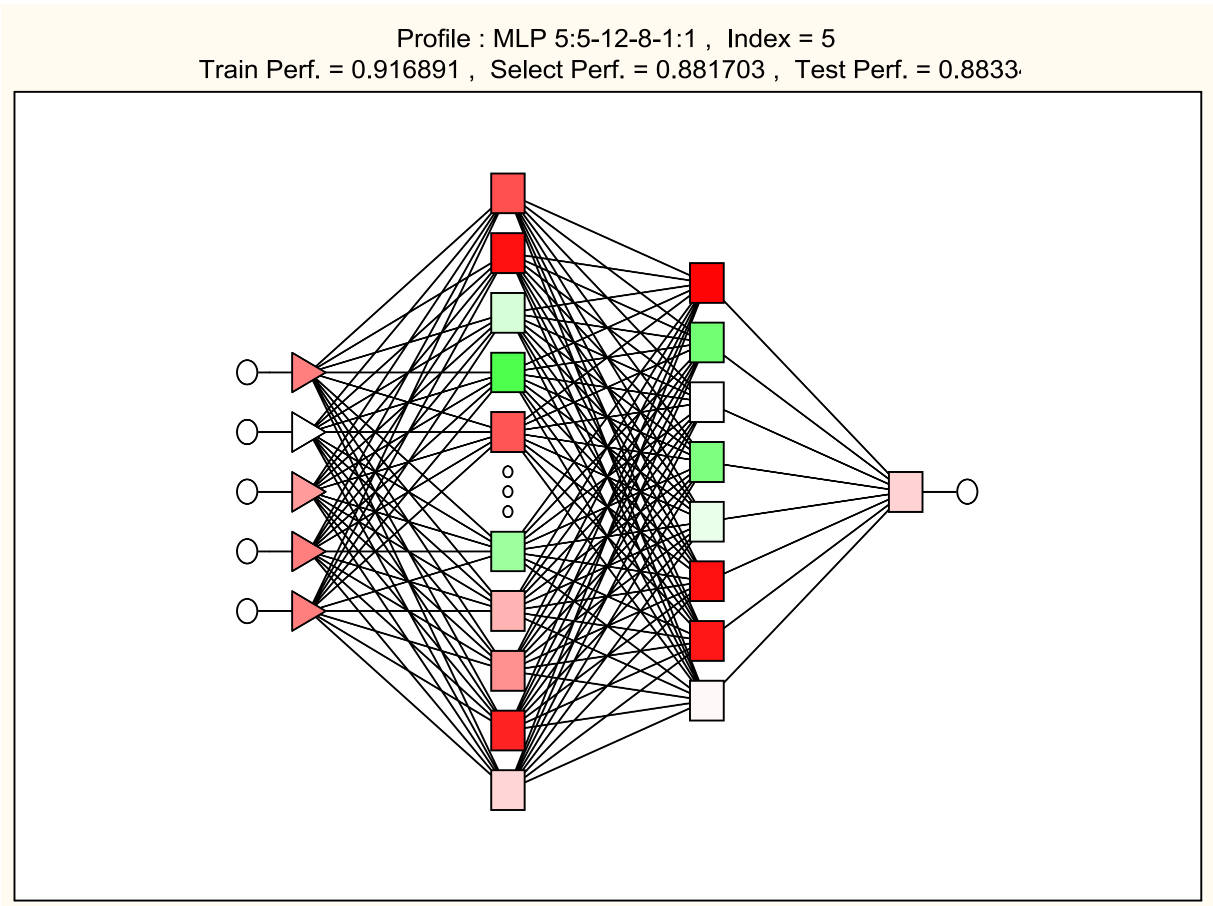


Figure 9 The best ANN developed for types with cyclonic NNE flow, a 4-Layer Perceptron with 5 inputs (*Cosine Day of Year, QPF, QPF x Cosine Day, PoP, PoP x Cosine Day*), 12 nodes at Layer 2, and 8 nodes at Layer 3 (activation levels are displayed in color - red for positive activation levels, green for negative).