18

Validation of a Weather Forecasting Expert System

S. Zubrick*
Radian Corporation,
Austin, Texas, USA

Abstract

This paper compares two sources of advice for forecasting of severe thunderstorms: an expert system (WILLARD) and government-issued severe weather outlooks. WILLARD was constructed by a meteorologist using the RuleMaster expert system building facility, which features rule induction from examples of expert decision-making. The validation period spans two months during the peak central United States thunderstorm season for 1984. The forecast comparisons are presented in terms of statistical properties: the Probability of Detection, the False Alarm Rate, and the Critical Skill Index. Even though WILLARD was developed as a demonstration system, its forecasting accuracy on major severe weather days is comparable to government-issued forecasts for the validation period. By examining the results of the comparison, deficiencies in WILLARD were identified that can be rectified in future versions, thereby increasing WILLARD's store of weather knowledge.

1. INTRODUCTION

This paper describes the results obtained in comparing two sources of advice for severe thunderstorm forecasting for the central United States (US): one, the standard convective outlook issued by forecasters of the Severe Local Storm Unit (SELS) of the National Weather Service's (NWS's) National Severe Storms Forecast Center (NSSFC); the other, a similar outlook made by a prototype expert system called WILLARD.

A thunderstorm is considered severe if any one of the following phenomena accompanies the thunderstorm (and is reported):

- tornadoes (intense, small-scale cyclones);
- hailstones ≥2 cm ($\frac{3}{4}$ in.) in diameter;
- surface wind gusts in excess of $93\,\mathrm{km\,h^{-1}}$ (50 knots) and/or significant wind damage.

* Current address: National Oceanic and Atmospheric Administration, National Weather Service, Silver Spring, Maryland, USA.

To understand better the nature of severe thunderstorm forecasting. Section 2 discusses some fundamental reasoning processes used by a meteorologist in forecasting severe thunderstorms. An overview of the method used by SELS forecasters at NSSFC will be provided, followed by a description of the WILLARD expert system.

Testing WILLARD using actual weather data is the subject of Section 3, which includes a discussion of the verification methods used. Also included are definitions for and the significance of three statistical parameters important in assessing a severe thunderstorm forecast: the Probability of Detection (*PoD*), the Critical Skill Index (*CSI*), and the False Alarm Ratio (*FAR*). These will be given for WILLARD's and SELS' outlooks. Other statistical parameters useful in verifying forecasts are also presented, together with comparison and discussion of forecasts made by SELS and WILLARD for selected days during the springtime 1984 central US severe thunderstorm season. The statistical parameters introduced in Section 3 for WILLARD's outlooks are shown to be comparable to NSSFC forecast outlooks in many respects. WILLARD's forecast reasoning is compared with that of SELS for a cross-section of severe, marginally severe, and non-severe weather days. An appendix for each test case study day highlights relevant meteorological factors recognized by each forecaster (SELS and WILLARD).

2. PROBLEM DEFINITION

'*Successful tornado and severe-thunderstorm forecasting is largely dependent upon the forecaster's ability to carefully analyze, coordinate, and assess the relative values of a multitude of meteorological variables, and mentally integrate and project these variables three-dimensionally in space and time.*' (Miller, 1972, p. 1.)

2.1. Weather data and forecasting

An important forecasting task is evaluating the basic accuracy and reliability of meteorological data. The forecaster must have physical access to weather data with enough time to analyse them and make his forecast. Much weather data are contained in graphical format, such as maps, charts, and satellite images. Most time spent by the forecaster in preparing his forecast is used in carefully examining and analysing these graphical data. Pattern recognition is an important technique the forecaster uses in extracting features important in severe thunderstorm forecasting. In fact, recent advances in severe thunderstorm forecasting accuracy have been attributed to use of sophisticated graphic display workstations that allow forecasters to examine quickly a large amount of meteorological data (Suomi *et al.*, 1983; Reynolds, 1983; Kerr, 1984; Mandics and Brown, 1985).

A fundamental factor affecting severe thunderstorm forecast accuracy involves the temporal and spatial resolving power of the current data-gathering network in the central US. This network was originally designed to observe large-scale weather systems: the major cyclones and anticyclones that govern regional weather and the fronts between air masses of different properties. In the central US, stations taking surface observations are spaced roughly 100–250 km apart and measure surface conditions hourly. Stations which measure winds, temperatures, and moisture through the depth of the atmosphere are spaced even further apart; averaging 300–500 km apart and taking observations only twice per day.

As such, this network does a good job providing information concerning these large-scale systems. However, it provides only sporadic information about individual thunderstorms and tornados which occur on the 'storm-scale' (i.e. between 2–200 km; 0.5–6 hours) between observing stations, and very little information about the specific location and time of occurrence of these severe events. This affects severe thunderstorm forecasting because there is insufficient resolution of weather features that play major roles in delineating severe areas.

Since the data network is not equipped to observe storm-scale weather phenomena, the forecaster must squeeze out small details from data sets to make his forecast of severe weather, often basing his forecast on somewhat shaky evidence obtained from the data. Solutions have been offered by respected institutions which would effectively increase the amount of data available from the network 10-fold (UCAR, 1982). One wonders if the forecasters will be able to assimilate this increase effectively. However, it is hoped that expert systems for weather data assimilation and forecast guidance support might go some way towards solving this problem in the future.

2.2. Human forecaster skills

Severe thunderstorm forecasting requires the ability to integrate and process an enormous amount of weather information contained in the data to produce a forecast within a short time period. In many instances a parameter overlooked by a forecaster because he is too busy could significantly affect the placement of a severe weather threat area.

Another factor affecting forecast accuracy is the forecaster's ability to recall his knowledge accurately and consistently. Forecasts have often been wrong simply because the forecaster forgot a general heuristic rule governing a particular severe weather situation. This occurs more frequently at the beginning of the peak severe thunderstorm season, when forecasters tend to be a little rusty in applying their rules. In the central US, the peak severe thunderstorm season runs from March through July.

WILLARD was designed as an expert system to aid severe thunderstorm forecasters to improve their ability to forecast severe weather in a more accurate and timely manner, by: (i) providing a consistent expert-forecast knowledge base to the forecaster; and (ii) routinely applying this knowledge base to incoming data to yield initial guidance available in real time.

Before discussing how WILLARD operates, this paper will discuss the basic ingredients for producing severe thunderstorms, and introduce some basic terminology. This is necessary for understanding the discussion of comparisons between WILLARD's forecast advice and advice given in government-issued forecasts.

2.3. Severe thunderstorm forecasting methods

It is beyond the scope of this paper to explain all details involved with severe thunderstorm forecasting. There are a few 'cookbook-style' texts that offer fairly explicit methods (see Miller, 1972; Crisp, 1979). The meteorological literature contains a wealth of information (see, e.g., Foster and Bates, 1956; Maddox and Doswell, 1982; Porter *et al.*, 1955). This paper provides a cursory introduction to severe thunderstorm forecasting to familiarize the reader with common parameter names. It will briefly discuss three ingredients necessary for severe thunderstorms: moisture (convective instability), lifting (triggering) mechanisms, and venting mechanisms (which also act as trigger mechanisms).

2.3.1. *Moisture (convective instability)*

It is not surprising that moisture is a key parameter analysed by severe thunderstorm forecasters. When moist air ascends it cools and allows condensation of its water vapour to form clouds. But without a mechanism that allows moist air to be rapidly carried upwards, causing an explosive release of an air parcel's latent heat of vaporization to the surrounding environment, there would be no severe thunderstorms.

Most severe thunderstorms are associated with areas where the convergence, or focusing, of moisture in the lowest few kilometers of the atmosphere is concentrated over a relatively small area (several thousand square kilometers) (Hudson, 1971). Moisture convergence zones tend to be found along thermal boundaries, like warm and cold fronts. These zones of concentrated low-level moisture are the favoured breeding grounds for severe thunderstorms.

In the central US, this moisture usually originates from the Gulf of Mexico and is normally fairly warm. This warmth also aids in thunderstorm formation. In severe situations, moisture is often found in a distinct tongue of high moisture that is rapidly carried northward into the interior of the central US. An unstable air mass is one that is both warm and moist. If lifted, air parcels within an unstable air mass easily become buoyant and aid the growth of thunderstorms. This latter concept is referred to as *convective instability*. There are various measures of air mass instability. Some common ones are the Lifted Index (Galway, 1956), K-Index, and Total–Totals Index.

2.3.2. *Lifting mechanisms*

Once an area of low-level moisture convergence is found, one looks for a suitable mechanism that allows the moist air mass to be lifted aloft. These mechanisms are generally characteristic of atmospheric features called upper-level, low-pressure centres or troughs. In fact, a low-pressure trough is akin to a trough in a water wave. Forecasters look for severe weather in front of an approaching low-pressure trough because this region provides an environment of rising air motions.

Movement of upper-level, low-pressure troughs are monitored by examining a parameter that measures the spin of the atmosphere known as vorticity. Regions where upper-level vorticity is a maximum are highly correlated with maximum rising air currents.

Other mechanisms that allow for rapidly rising air currents include strong surface daytime heating (causing moist air parcels to be intensely buoyant), orographic flow (as in flow rising over a mountain), frontal boundaries (which provide mechanical lift from cold (dense) air wedging underneath warm, moist (less dense) air, and small scale surface circulation features known as meso-lows.

2.3.3. *Venting mechanisms*

After the moisture field and lifting mechanisms are identified, the forecaster then determines if there is present an upper-level feature that essentially acts as a 'venting mechanism'. This mechanism allows rising air currents to be carried up and out of the lower atmosphere. If strong enough, this sets up a vertical circulation that intensifies the thunderstorm (McNulty, 1978).

Upper-level venting mechanisms are found when the forecaster spots an extremely fast and narrow upper level wind current (or jet streak) flowing at about 10 km in altitude. This high-speed wind current tends to draw air upwards through the storm's centre—akin to a high wind drawing air up a chimney. This results in increased amounts of warm moist air being drawn into the storm by the low-level winds converging into the storm system. The higher this jet streak's speed, the more destructive are the severe storms. Speeds observed in violent severe storm systems range anywhere from 250 to over 400 km h^{-1}.

In the central US, it is common in the springtime for this high-speed wind current to be associated with the subtropical jet stream, which is frequently associated with severe thunderstorm outbreaks (Whitney, 1977). Certain regions surrounding the upper-level jet max core are conducive to enhancing severe weather, especially when the exit region of the jet streak interacts with lower-level wind features to create differential temperature and moisture transports (Uccellini and Johnson, 1979). The presence of a strong upper-level jet streak is highly correlated with the formation of strong tornadoes.

Generally, for severe thunderstorms to occur there needs to be a 'phasing in' of maximum low-level moisture convergence, lifting mechanisms, and venting mechanisms over a small region. The major forecast objective is to identify these small regions and define the times of severe weather onset and cessation.

2.4. Severe thunderstorm outlooks

2.4.1. *Government-issued severe thunderstorm outlooks*

The SELS of the NSSFC issues medium-range, severe local storm outlooks three times daily during the US severe thunderstorm season. An early outlook is issued at 08.00 Universal Coordinated Time (UCT or Z), a morning outlook is issued at 15.00Z and an afternoon update is issued at 19.30Z. These outlooks are disseminated to various government and private agencies to provide preliminary guidance on expected severe local thunderstorm development in an 18–24 time period covering the entire contiguous US. For example, the outlooks are used by NWS Regional Forecast Offices in preparing state, zone, local, and aviation forecasts (Otsby, 1979).

A severe thunderstorm outlook contains a phrase specifying the expected areal density of severe weather coverage occurring within the valid period of the forecast area. A typical SELS outlook covers an area approximately 337,000 square kilometres [130,000 square statute miles (sq sm)]. The areal density/risk categories as specified in the NWS Operations Manual Chapter C-40 (1979) that are followed by SELS forecasters are:

- Slight risk: 2–5 per cent areal coverage or 4–10 MDR (Manually Digitized Radar, see below) blocks with severe thunderstorms per 100,000 sq sm of outlook;
- Moderate risk: 6–10 per cent areal coverage or 11–21 MDR blocks with severe thunderstorms per 100,000 sq sm of outlook;
- High risk: greater than 10 per cent areal coverage or more than 21 MDR blocks with severe thunderstorms per 100,000 sq sm of outlook.

An MDR (Manually Digitized Radar) block measures approximately 41 km (22 sm) on a side, occupying an area of 1681 sq km (484 sq sm).

Note that only a small fraction of the outlook area is expected to experience severe weather. Other thunderstorm categories are given in a SELS outlook. They include an approaching density/risk and a general non-severe thunderstorm category. However, both of these categories are classified as non-severe and are not verified at SELS so they will not be considered in this paper.

Each SELS outlook contains forecast reasoning in the form of a narrative text. This reasoning points out the major factors that influenced the forecaster's selection of a threat area. It basically relates how present weather conditions will evolve to allow the formation of severe thunderstorms. This reasoning is used by forecasters in the field either to accept or to modify the outlook area. Graphic maps are also disseminated over facsimile circuits for SELS outlook areas.

In producing the outlook, SELS forecasters have considerable prognostic and diagnostic guidance available (Pearson and Weiss, 1979). Prognostic guidance is derived in part from an operational numerical primitive equation model of the atmosphere called the Limited Fine Mesh (LFM) II Model.

From LFM data, contour maps are generated and disseminated to SELS forecasters who examine these maps for clues of impending severe weather. The same LFM data used to generate these maps is used in its gridded form by a host of FORTRAN analysis routines callable by WILLARD. These routines are used to extract information on features necessary in producing the severe weather outlook.

Diagnostic aids used at SELS include computer-plotted surface weather maps, upper-air soundings, and numerous derived objective analyses, such as 500 millibar absolute vorticity, upper tropospheric mean divergence, air mass stability, and low-level moisture convergence. In addition, the latest visible and infrared satellite imagery is available to forecasters for evaluating and updating the numerical guidance. Other members of the NSSFC staff assist SELS forecasters in analysis of local radar data for identifying severe thunderstorms. (Satellite data was not used in WILLARD.)

This paper will examine (for comparison purposes) the forecast outlooks issued by SELS at 08.00Z. This outlook relies heavily upon numerical guidance from the LFM model, especially the 24-h model forecasts made from the 00.00Z LFM model run cycle. The forecaster usually has at least 3–5 h available to examine model results before issuing the 08.00Z outlook.

2.4.2. *WILLARD: an expert system to produce severe outlooks*

The expert system WILLARD produces an outlook of severe thunderstorms that is similar to the SELS 08.00Z convective outlook. However, it differs from the SELS outlooks in that its areal extent only covers the central one-third of the US (Figure 1), and is valid for a 12-h time period rather than a 24-h period.

Figure 2 outlines the general information flow path while running WILLARD. The main source of input data used by WILLARD (viz., the LFM model output forecast data) is a major subset of input data used by SELS forecasters in generating their 08.00Z severe outlook.

To produce an outlook with WILLARD, gridded data from the LFM model is obtained and stored on the computer (WILLARD runs on a Sun 100). A host of FORTRAN analysis routines that compute various diagnostic parameters from the LFM data files are available to WILLARD. When WILLARD's knowledge-base is run, it obtains most of the necessary answers to its questions by requesting information from the FORTRAN analysis routines. Answers not available from the data base are requested from the meteorologist running the expert system.

A point forecast is made for a grid point, which coincides with the grid mesh of the LFM model output data. This grid mesh is roughly 200 km on a side. WILLARD is run repeatedly over a 14×10 grid mesh covering the central US. The result of this run is an array of 140 point forecasts stored in a disk file. Subjective contour analysis is used to delineate areas with the same density/risk categories (some smoothing of the categories is done for verification purposes). These areas are plotted on a base relief map of the central US.

WILLARD's knowledge base was developed over the course of several months by a meteorologist familiar with severe thunderstorm forecasting procedures. Discussions with former SELS forecasters and others identified main parameters to be examined. The meteorological literature was used to understand better the effects of various relationships between parameters pertaining to severe thunderstorms. Actual weather data during severe, marginally severe, and non-severe weather days were also examined.

It became apparent from reading the meteorological literature and discussing forecasting methods that no coherent system of rules covering all possible severe storm cases had yet been synthesized. The availability of an expert system building facility called RuleMaster (see Michie *et al.*, 1984), which could build classification rules by rule induction, was thought to provide a potential solution. Using this system, classification rules are induced by generalization over examples of expert decision-making. An example is expressed as a vector of values pertaining to attributes of the decision, together with the expert's classification (Michie *et al.*, 1984; see also Quinlan, 1979).

For purposes of rapid development, subjectively selected examples were used to build the prototype expert system. Cases of real weather data were subsequently applied in the ongoing refinement of WILLARD.

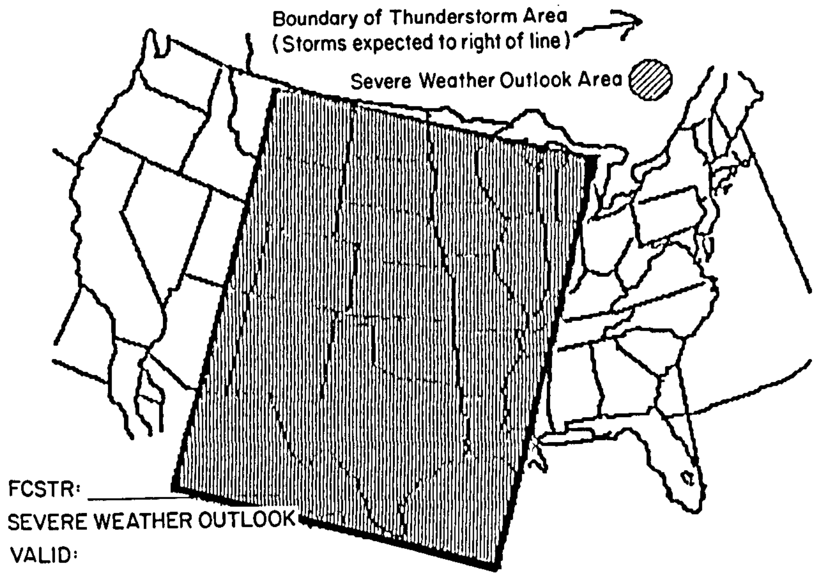

Figure 1. Map of central United States representing WILLARD's forecast domain area (hatched area).

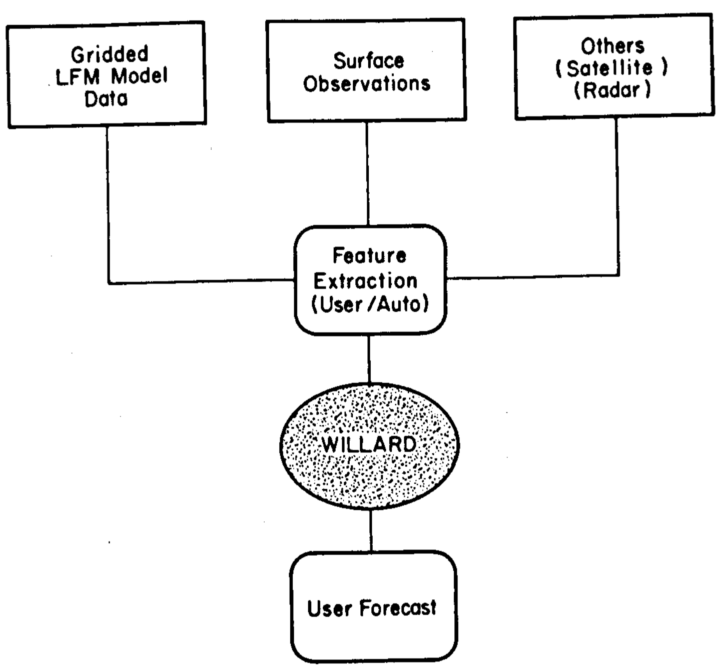

Figure 2. Information flow diagram for WILLARD.

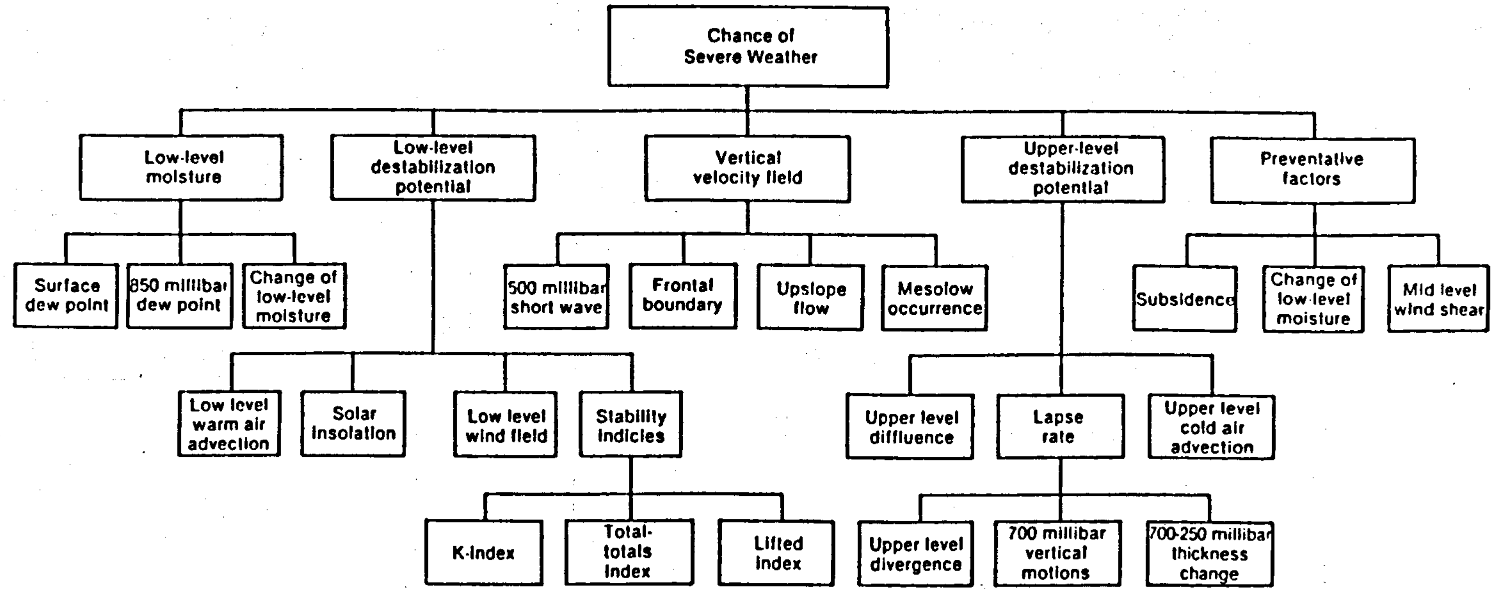

Figure 3. Hierarchical structure of WILLARD's knowledge-base.

The WILLARD expert system is composed of a hierarchy of 30 modules (Figure 3), each containing a single decision rule. This hierarchy is on average four levels deep. All decision rules within each module were developed using inductive generalization (except for some looping control for executing over the grid of data). About 140 examples out of a possible nine million situations were used in building WILLARD.

For the top-level module (Chance of Severe Weather in Figure 3), inductive generalization was able to order the critical meteorological factors in a manner consistent with the way forecasters perform their analysis. For example, if the key factors were all totally unfavourable, a rapid decision could be made: otherwise, more parameters were investigated until a decision could be reached.

WILLARD was designed to operate in either manual or automatic forecast mode. In manual mode, the system asks the meteorologist about pertinent weather conditions for the forecast area and produces a complete, reasoned forecast. In automatic mode, WILLARD obtains necessary information from National Meteorological Center data files (viz., the LFM gridded data), with some information obtained from a meteorologist interactively.

The form of a typical decision rule along with the attribute value set is shown in Figure 4. This rule is used in determining whether the low-level

EXAMPLE SET

solar insolation	low-level jet	indices		
strong	present	marginal	=>	(favor, GOAL)
strong	absent	weak	=>	(unfav, GOAL)
weak	present	strong	=>	(margin, GOAL)
strong	absent	strong	=>	(favor, GOAL)
strong	absent	marginal	=>	(margin, GOAL)
weak	absent	marginal	=>	(unfav, GOAL)
weak	present	marginal	=>	(margin, GOAL)

INDUCED RULE FROM 'LL_DSTAB.IND'

```
[indices]
        weak : => ( unfav, GOAL )
    marginal : [solar_insol]
               strong : [low_level_jet]
                      present: => ( favor, GOAL )
                       absent : => ( margin, GOAL )
                 weak : [low_level_jet]
                      present: => ( margin, GOAL )
                       absent : => ( unfav, GOAL )
      strong  : [solar_insol]
                 strong: => ( favor, GOAL )
                   weak: => (margin, GOAL )
```

Figure 4. Example of induction set and corresponding decision rule.

destabilization is unfavourable, marginal, or favourable for severe thunderstorm formation.

The decision rules for all of the modules shown in Figure 3 were examined by a meteorologist for correctness and consistency by applying the rules individually and collectively using actual severe weather data. The meteorologist could then change attribute values and induce a new set of decision rules until finally he was satisfied with the rules produced.

Another utility of the RuleMaster system, Radial, was used to execute the complete WILLARD expert system. Initial testing uncovered numerous errors in both the knowledge-base and also in the FORTRAN analysis routines. In mid-1984, after these errors were rectified, it was decided that the WILLARD knowledge-base must remain static while a series of validation runs were completed and analysed. The results of this validation and analysis are presented in the next section. These results should be viewed as the first verification data of a prototype weather forcasting expert system. They have already provided direction for future improvements and refinements in both the WILLARD knowledge-base and analysis package.

3. VERIFICATION

The methodology used in verifying WILLARD's and SELS' forecasts was the same as that used by NSSFC researchers (Weiss *et al.*, 1980). Each severe thunderstorm outlook area is examined to produce statistical data useful in evaluating the forecasts. Since WILLARD produces a severe thunderstorm outlook similar to those produced by NSSFC forecasters, this was a reasonable verification method.

Three test case study days are discussed in detail and included as an Appendix. The study days were 29 April 1984, 25 May 1984, and 7 June 1984. Two of these days (29 April and 7 June) were chosen to highlight forecasting abilities during major severe weather outbreaks. The other day (25 May 1984) was chosen as a day representing a minor outbreak day.

Actual SELS outlooks contain forecast reasoning on why an area could experience severe thunderstorms. This reasoning was compared with WILLARD's reasoning for each of the test case study days. Actual weather data for these test days were consulted to aid in interpreting the reasoning behind each forecast. The following discussion of verification statistics follows closely the discussion found in Weiss *et al.* (1980).

3.1. Methodology for verification statistics

3.1.1. *Definitions of major verification statistics*

Verification of severe thunderstorm outlooks is based upon the critical skill index (*CSI*) (Donaldson *et al.*, 1975) applied over a large area. It is the ratio of successful predictions of severe weather to the number of severe events that occurred or were forecast to occur. *CSI* scores over 0.5 are considered good by SELS forecasters.

For purposes of severe weather verification, the *CSI* is first computed by dividing all weather events for a given outlook into four groups:

x—severe storm reports correctly predicted (i.e. those reports found within a severe risk outlook area);

y—severe storm reports not predicted (i.e. those lying outside the severe risk outlook area);

z—non-severe weather predicted as severe; and

w—non-severe weather correctly predicted.

The probability of detection (*PoD*) is the proportion of severe weather events correctly forecast:

$$PoD = x/(x+y) \tag{1}$$

An outlook area that contains all of the severe weather reports will have a *PoD* of unity. The *PoD* is normally expressed in per cent, so that its range is 0–100 per cent. A *PoD* of 100 per cent is the best value for an outlook.

The false alarm ratio (*FAR*) is the proportion of predictions that fail to verify:

$$FAR = z/(z+x). \tag{2}$$

The *FAR* ranges between 0 and 1. A *FAR* of 0 indicates a perfect forecast. The *FAR* is modified by the use of an areal distribution term (Weiss *et al.*, 1980), which quantitatively determines the amount of over-forecasting from either the outlook area being too large or insufficient density of severe reports. Thus while a high *PoD* is obtained when a large percentage of severe weather events occurs within a forecaster's outlook area, he is discouraged from forecasting excessively large areas which would increase the *FAR* and decrease the *CSI*.

The *CSI* can now be expressed in terms of *PoD* and *FAR*.

$$CSI = x/(x+y+z) = \{(1/PoD) + [1/(1-FAR)] - 1\}^{-1}. \tag{3}$$

The *CSI* ranges from zero to unity, with higher numbers indicating better forecasts. The *CSI* is also known as the Threat score. Some outlooks may forecast severe weather in several unconnected regions. Here, separate *FARs* are calculated for each area, and an area-weighted average is computed for the entire outlook. This average *FAR* is then used with the total percentage of all severe events within the forecast area (*PoD*) to calculate a single *CSI* via equation (3) for the entire outlook for that day (Weiss *et al.*, 1980).

3.1.2. *Other statistical parameters*

The extent of areal coverage (CA) is that portion of the outlook area in which severe weather events occur. It is defined as:

$$CA = [(\text{no. of MDR blocks with severe events}) \times K]/(\text{outlook area}) \tag{4}$$

where a MDR block can only be counted once no matter how many severe events might be clustered within a single MDR block (by definition at NSSFC). The constant K is equal to 1681 sq km (484 sq sm), the area covered by a single MDR block.

The coverage bias ($CBIAS$) is defined as the ratio of the actual areal coverage to the forecast areal coverage. The forecast areal coverage is determined from the outlook risk category of the outlook. If the actual areal coverage lies within the range of the forecast risk category, then no coverage bias exists (i.e. $CBIAS = 1.0$). If the actual areal coverage is outside the range of the forecast risk category, the forecast coverage is taken as the category extreme closest to the actual coverage.

The good area is that portion of the outlook area affected by severe weather. This statistic incorporates both the forecast areal coverage and the areal distribution of severe reports within the outlook area. In particular, for a Slight risk each event is assumed to affect a 6×6 MDR grid array surrounding the event (2.77 per cent areal coverage); for a Moderate risk each event affects a 4×4 array (6.25 per cent areal coverage); and for a High risk each event affects a 3×3 array (11.11 per cent areal coverage). The total number of MDR blocks determined in this manner is summed to compute the good area. Each MDR block can be counted only once using this method. If the good area is the same as the original outlook area, then this is considered a representative forecast (although, this outlook could still miss severe events outside its area).

A FAR can also be defined as one minus the good area percentage (the proportion of the outlook area affected by severe weather), or

$$FAR = 1 - (\text{affected area/area of ourlook}). \tag{5}$$

The bad area is that portion of the outlook area not affected by severe weather, and is defined as:

$$\text{bad area} = (\text{area of outlook}) - (\text{good area}). \tag{6}$$

A bad area equal to zero would mean that the good area is equal to the original area of the outlook (which is what one desires). The sum of the good and bad areas equals the original outlook area.

3.2. Overall verification statistics

This section discusses the verification results for both WILLARD and SELS on the selected days during the spring 1984 central US peak thunderstorm season. There were a total of 30 WILLARD severe thunderstorm outlooks generated spanning a period from 22 April through 11 June 1984. Since the NSSFC verification scheme is only applicable on days when severe weather was outlooked (i.e. a category of Slight, Moderate, or High risk), the actual number of WILLARD outlooks verified by NSSFC was 24, because WILLARD generated six no severe outlooks. Statistics on the 24 WILLARD outlooks verified by NSSFC are given in Table 1.

The overall *PoD* for WILLARD's 24 outlooks was 37 per cent. The *FAR* was 0.628 for these outlooks. These two parameters combined yielded an average *CSI* of about 0.20. Each of these experienced a wide range of daily values. For WILLARD's *PoD*, the range was 0–100 per cent; for the *FAR*, the range was 0.116–1.000; and for the *CSI*, the range was 0.000–0.691.

The average size area for WILLARD outlooks was about 260,000 sq km (100,000 sq sm). This was almost one half the average size for SELS outlooks during this period and a similar springtime period (Weiss and Reap, 1984). The average good area for WILLARD was over 93,000 sq km (36,000 sq sm). The average bad area for these outlooks was about 166,500 sq km (64,300 sq sm). June 7 had the largest good area of all days, with an area of over 114,500 sq sm. June 4 had the largest bad area of 430,399 sq sm. There were two days on which the good area equalled the original outlook area (26 and 29 April) while there were six days on which the bad area equalled the original outlook area (28 April; 2, 22, 23 May; and 1, 2 June).

For the 24 WILLARD outlook forecasts, there were 1001 severe weather reports, 190 of these being tornadoes. WILLARD captured 369 of 1001 reports, including 82 of the tornadoes within its outlook areas. The areal coverage for the test period was 5.5 per cent, for which a forecast Slight risk category would give a coverage bias of unity. WILLARD's category tended to over-forecast slightly as indicated by an average *CBIAS* of 0.830. There were five days when the coverage bias was near unity.

Since WILLARD used fairly large-scale data, its outlooks forecast areas of widespread severe weather rather than isolated severe thunderstorms. When days were chosen that had more than 10 tornadoes, the WILLARD *CSI* became 0.33, with the *PoD* being 40 per cent and the *FAR* being 0.442. In addition, on these days the average WILLARD outlook area became 206,200 sq km (79,610 sq sm). The average good area became 169,600 sq km (65,480 sq sm) with the average bad area being 36,600 sq km (14,130 sq sm). The areal coverage of severe weather on these days was 14.4 per cent, which implies a High risk category being the proper outlook category. The coverage bias on these days was 1.92, implying that WILLARD outlooks tended to under-forecast on these days.

Table 2 contains verification data for SELS 08.00Z convective outlooks for the same 24 days of WILLARD outlooks (P. W. Leftwich, NSSFC,

Table 1. Verification statistics for WILLARD outlooks on selected days in 1984.

Year	Month	Day	Area	Hits	Severe reports	*PoD* %	Tornado hits	Tornados	%	Areal coverage	*CBIAS*	Far	Good area	Bad area	*CSI*
84	4	25	44356	11	19	57	3	6	60	0.065	1.091	0.225	34364	9932	0.489
84	4	26	132775	44	220	20	6	29	20	0.102	1.580	0.409	89477	43298	0.176
84	4	27	21359	10	55	18	2	13	15	0.204	3.399	0.484	21359	0	0.154
84	4	28	42001	0	1	0	0	0	0	0.000	0.000	1.000	0	42001	0.000
84	4	29	93899	78	142	54	12	26	46	0.258	3.127	0.335	93899	0	0.424
84	5	2	8507	0	121	0	0	19	0	0.000	0.000	1.000	0	8507	0.000
84	5	3	32900	4	13	30	0	1	0	0.059	0.981	0.655	15725	17175	0.191
84	5	22	61391	0	0	0	0	0	0	0.000	0.000	1.000	0	61391	0.000
84	5	23	13507	0	2	0	0	1	0	0.000	0.000	1.000	0	13507	0.000
84	5	24	91148	8	25	32	1	4	25	0.037	1.000	0.124	79860	11288	0.306
84	5	25	77742	15	22	68	1	2	50	0.068	1.000	0.477	40656	37086	0.420
84	5	26	182077	10	11	90	0	0	0	0.019	0.310	0.902	17908	164169	0.097
84	5	27	80496	16	60	26	0	2	0	0.060	1.002	0.212	63404	17092	0.243
84	5	28	167512	1	1	100	0	0	0	0.003	0.029	0.974	4356	163256	0.026
84	5	31	20347	1	8	12	1	6	16	0.024	1.000	0.144	17424	2923	0.119
84	6	1	25783	0	12	0	0	0	0	0.000	0.000	1.000	0	25783	0.000
84	6	2	50828	0	3	0	0	0	0	0.000	0.000	1.000	0	50828	0.000
84	6	4	480251	11	22	50	4	7	57	0.009	0.151	0.896	49852	430399	0.094
84	6	5	112626	7	38	18	2	8	25	0.021	0.519	0.553	50336	62290	0.147
84	6	6	166468	3	23	13	0	1	0	0.006	0.291	0.878	20328	146140	0.067
84	6	7	137934	86	120	71	30	39	76	0.186	1.430	0.309	114574	23360	0.539
84	6	9	175378	13	14	92	2	2	100	0.028	0.275	0.838	28556	147322	0.160
84	6	10	104630	16	23	69	4	6	66	0.051	0.753	0.547	47432	57198	0.377
84	6	11	83185	35	46	76	14	19	73	0.116	1.939	0.116	73568	9617	0.691

FCSTR WILLARD	Number outlooks	Average area	Hits	Reports	*PoD*	Tornado hits	Tornados	%	Areal coverage	*CBIAS*	*FAR*	Good area	Bad area	*CSI*
49	24	100321	369	1001	37	32	190	43	0.035	0.830	0.628	35962	64339	0.197

Table 2. Verification of SELS convective outlooks issued at 0800Z on selected days in 1984 (Leftwich, NSSFC, personal communication, 1987)

Date	Area	*PoD*	*CBIAS*	*FAR*	*Good area*	*Bad area*	*CSI*
4/25	126180	0.80	1.00	0.44	70664	55516	0.49
4/26	307095	0.89	1.85	0.44	218621	88474	0.52
4/27	476273	0.77	1.04	0.53	225060	251213	0.41
4/28	65578	0.00	0.00	1.00	0	66578	0.00
4/29	366116	0.83	1.00	0.49	185856	180260	0.46
5/2	366354	0.66	1.34	0.50	182468	183886	0.40
5/3	399755	0.97	1.00	0.36	257972	141783	0.63
5/22	261119	0.66	0.19	0.91	23232	237887	0.09
5/23	135950	0.55	1.00	0.46	73568	62382	0.38
5/24	303966	0.82	1.00	0.50	152944	151022	0.45
5/25	214510	0.93	1.00	0.52	103092	111418	0.46
5/26	181739	0.16	0.40	0.76	43076	138663	0.11
5/27	214855	0.90	1.28	0.30	150524	64331	0.65
5/28	94522	0.00	0.00	1.00	0	94522	0.00
5/31	No severe forecast						
6/1	69430	0.41	1.00	0.33	46464	22966	0.34
6/2	67122	0.16	0.36	0.74	17424	49698	0.11
6/4	245426	0.63	1.00	0.55	110352	135074	0.36
6/5	387216	0.81	1.00	0.56	169884	217332	0.40
6/6	493657	0.73	0.59	0.79	105028	388629	0.20
6/7	413460	0.94	1.45	0.47	241457	172003	0.51
6/9	323633	0.68	0.67	0.71	92928	230705	0.25
6/10	413827	0.87	0.67	0.64	147136	266691	0.34
6/11	284051	0.84	1.00	0.57	123420	160631	0.40

Number outlooks	Average area	*PoD*		*FAR*	Average good area	Average bad area	*CSI*
24	270080	0.81	—	0.56	119181	150942	0.40

personal communication, 1987). During this period, the *PoD* for SELS was 81 per cent. The *FAR* was 0.56. These two parameters combined yielded a *CSI* of 0.40—which is double WILLARD's *CSI*. SELS' outlooks overall were better than WILLARD's for this period. Figure 5 gives a daily breakdown comparing SELS' and WILLARD's outlooks. Of note is that WILLARD's *CSI* scores tended to be relatively in phase with the trend of SELS' *CSI* scores.

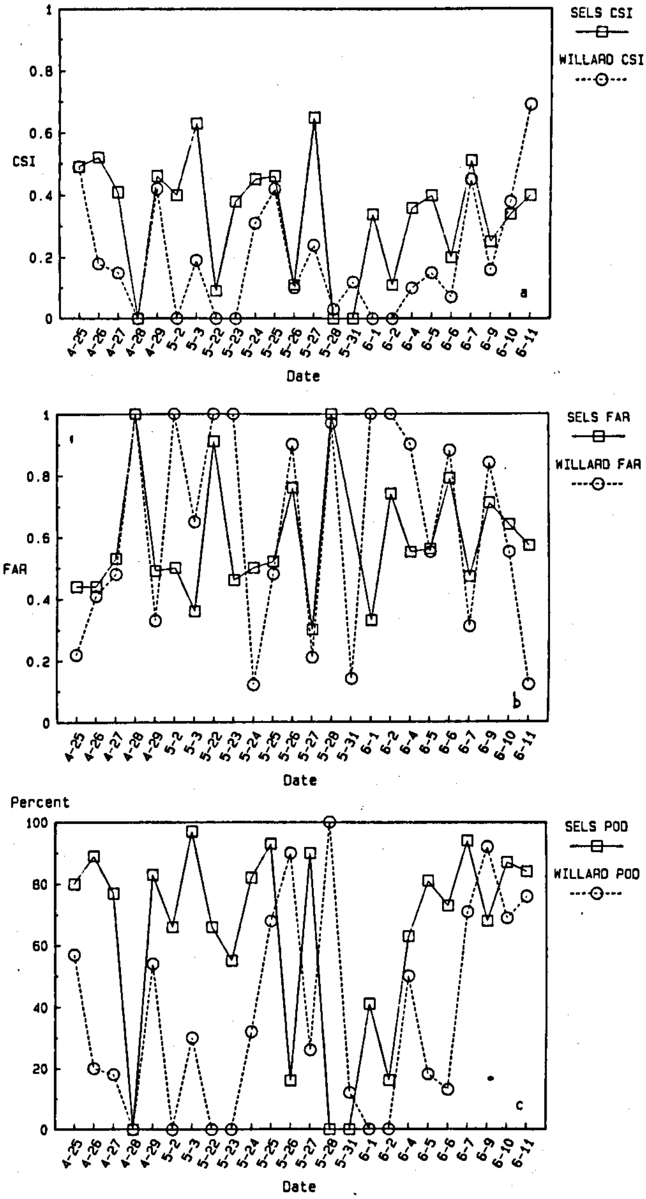

Figure 5. Comparative verification statistics for SELS and WILLARD 08Z Convective Outlooks for Selected Days in Spring 1984: a) *CSI*, b) *FAR*, c) *PoD*.

Examining the six days (viz., 22, 23, 24 April; 1, 29, 30 May 1984) when no severe thunderstorms were forecast by WILLARD, it was found that SELS forecast no severe thunderstorms on those same days with one exception. On 1 May 1984 SELS issued a Slight risk area that covered northeastern Texas, southeastern Oklahoma and western Louisiana. There were two dozen severe weather reports within SELS' outlook area 1 May. The reason why WILLARD did not forecast severe weather on this day was because of insufficient moisture in the lower levels of the LFM model data. Examining the forecast reasoning for SELS and WILLARD on the other non-severe outlook days showed that both agreed that lack of moisture was the primary cause for the non-severe outlook forecast. A detailed analysis of the reasoning given by WILLARD and SELS for the test case study days is given in the Appendix.

4. SUMMARY AND CONCLUSIONS

This paper examined two sources of advice for severe thunderstorm forecasting: an outlook produced by a government agency (SELS/NSSFC) and an outlook given by an expert system (WILLARD). Overall the two sets of advice were comparable in critical skill index, and forecast reasoning but different in several other statistical parameters. The WILLARD forecast is comparable to the actual SELS forecast in overall skill on major and most minor outbreak days.

The valid time period for the 08.00 SELS outlook covers a 24-h period whereas the 08.00Z WILLARD outlook spans a 12-h period. This could affect the results presented in this paper. The forecasting strategy behind WILLARD's production of a 08.00Z outlook is to use only the 24-h LFM Model forecast from the 00.00Z LFM model run to prepare the outlook. This was mainly done because the 36-h gridded LFM model forecast data were unavailable. However, it was felt that ±6 h from the time of the 24-h forecast was valid (viz., 00.00Z the next day) and consistent with the fact that severe thunderstorms generally reach their peak in number and intensity within a few hours of 00.00Z. Examination of the NSSFC severe storm log showed this to be true in most cases used in this study.

Therefore, it is estimated that, although some verification values might change for WILLARD's outlook if a full 24 hours' worth of severe weather reports were used, the change should be insignificant. Conversely, the statisitics for the SELS group should also not differ greatly.

On 2 May 1984, WILLARD's 08.00Z outlook forecast relatively little severe weather to occur anywhere within the forecast domain. However, this day turned out to be a major severe weather outbreak day (NOAA, 1984). There were over 120 severe weather reports in northern Texas, Oklahoma, Arkansas, Louisiana, and western Mississippi including 19 confirmed tornadoes. A script of the original WILLARD run was carefully examined, as were many weather maps and data. The SELS 08.00Z outlook had a Moderate risk area centred over the affected area mentioned above.

After examining relevant data and forecasts for 2 May, it appeared that WILLARD had difficulty in properly classifying the low-level moisture field. The LFM model data near the affected area indicated somewhat drier conditions at lower levels than those actual data showed. WILLARD concluded that low-level moisture was insufficient to support severe thunderstorms in this area. This contradicts data taken from vertical measurements of moisture near the threat area. Since WILLARD's moisture decision rule only examined a few vertical points from the LFM forecast data and did not examine actual sounding data, the effects of a slightly drier air mass as forecast by the LFM model were significant. This could be corrected by inserting more knowledge into the moisture module.

Additionally, on 2 May the LFM model was unable to handle the large number of short-wave, low-pressure troughs—as noted by meteorologists responsible for interpreting satellite imagery at the NWS Satellite Field Services Station located at NSSFC. A satellite interpretation message received near 00.00Z on 3 May 1984 indicated model guidance from earlier in the day was not resolving smaller scale features which were causing most of the severe activity. Examination of the LFM data confirmed this, showing that the model had lumped everything together into an a ill-defined low-pressure trough. Therefore identification of trigger mechanisms was clouded by the unrealistic model output. In the future, it might be possible to develop rules for predicting this condition and add them to the vertical velocity field module of WILLARD.

When there are errors in the LFM forecast data, it is likely that SELS forecasters are able to adjust the data to compensate. WILLARD did not apply any data adjustment of the LFM data nor did it attempt to recognize model errors (it did check for gross data range errors). Installation of rules for adjusting erroneous model forecast data was beyond the scope of this project. It is something which needs to be implemented in future weather forecasting expert systems which automatically provide guidance from numerical models. The rules governing these adjustments are complex and based on pattern recognition. However, an easily expandable system like WILLARD could accommodate addition of these rules.

Overall, the results of this paper are encouraging for pursuing the application of expert system technology to weather forecasting. Further research is underway to improve the knowledge-base to better handle smaller scale severe thunderstorm outbreaks.

Acknowledgements

Support for this study was provided by Radian Corporation and is gratefully appreciated. This study benefited from the assistance provided by Dr Preston Leftwich, of the National

Severe Storms Forecast Center, Techniques Development Unit, who graciously provided the necessary verification data on WILLARD and SELS. Drs Charles Doswell and Robert Maddox of the Environmental Research Laboratories, NOAA, supplied many stimulating ideas for this project. Steven Muggleton of the University of Edinburgh supplied much useful information in the structuring and explanation capability of WILLARD. Finally, Charlie Riese of Radian motivated the author to explore the application of expert systems technology to weather forecasting.

APPENDIX

29 April 1984—test case no. 1

This was a major and widespread severe thunderstorm outbreak day. Within the area and time of WILLARD's forecast domain there were 142 severe weather reports including 26 reports of tornadoes. One of these tornadoes caused one death and extensive property damage to the town of Mannford, in northeast Oklahoma (Ferguson *et al.*, 1985). Most of the severe weather reports came from a six state area covering all of Missouri, Iowa, and Illinois, and portions of eastern Oklahoma, Kansas, and Texas. Scattered reports were received from northern portions of Arkansas, Louisiana, and Mississippi, and southern Wisconsin.

The SELS 08.00Z convective outlook issued on 29 April 1984 recognized that meteorological conditions were favourable for a widespread severe weather outbreak in the southern Great Plains States (Oklahoma, Arizona, Missouri, and Kansas). They issued a High risk outlook area that covered most of Oklahoma, Missouri, Illinois, and Arkansas, portions of northern Texas, extreme northwest Louisiana, southern Iowa, western Tennesse, and western Indiana (Figure 6a).

High risk outlooks are rarely issued by SELS forecasters, and are only issued when the threat is clearly recognized as being substantial in severity and areal extent. Although there are usually less than 10 High risk outlooks issued annually by SELS forecasters, their greatest forecasting ability is exhibited when the threat of severe thunderstorms and tornadoes is the highest (Weiss and Reap, 1984). SELS' outlook area on this day covered over 360,000 sq sm.

WILLARD forecast a smaller outlook area evenly divided between Slight and Moderate risk areas covering northeastern Oklahoma, eastern Kansas, and southern two-thirds of Illinois, all of Missouri, and southeast Iowa (Figure 6b). This outlook area covered almost 94,000 sq sm.

In comparing statistical results from the two forecasts, the *CSI* was 0.49 for WILLARD and 0.46 for SELS, which are similar. The *PoD* was 54 per cent for WILLARD contrasted with 83 per cent for SELS. WILLARD had 78 severe reports within its outlook area out of 142, with 12 out of 26 tornadoes included. However, mostly due to the smaller size of WILLARD's outlook, the *FAR* was 0.34 for WILLARD vs. 0.49 for SELS. So even though the *CSI* was similar for both outlooks, the SELS outlook captured a majority of severe weather reported at the expense of an

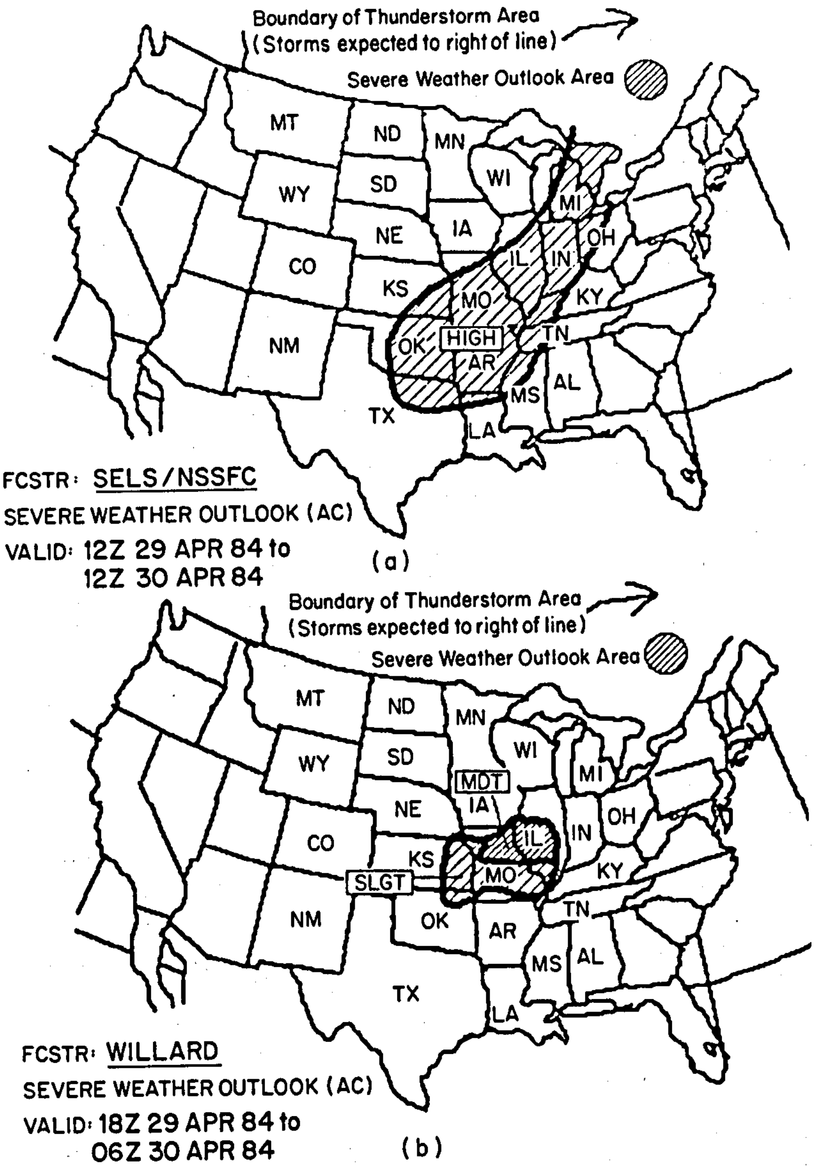

Figure 6. Map of SELS and WILLARD outlooks—29 April 1984.

increased false alarm ratio. This is reflected in the similarity of the *CSI* scores: a low *PoD* coupled with a low *FAR* can yield a *CSI* score nearly equivalent to a high *PoD* coupled with a high *FAR*. It is a matter of individual preference as to whether *PoD* or *FAR* is of greater importance, and is not an issue in this paper.

The good area for WILLARD's outlook equalled the original outlook area. The areal coverage of severe weather within WILLARD's outlook was almost 26 per cent, with a *CBIAS* of 3.13. This was the highest coverage bias encountered in this study. It indicates that WILLARD under-forecast the severe weather that occurred within its outlook. If WILLARD had forecast a High risk category it would have had a coverage bias of unity. The SELS coverage bias was unity, as they did forecast a High risk category.

In examining the meteorological reasoning behind each forecast it was apparent that both forecasts identified that the interaction of a strong upper level low pressure system with a very warm and moist unstable air mass near the surface, and a source of upper-level venting (or diffluence) would all phase in over eastern Oklahoma and Kansas, moving into Arkansas, Missouri, and later Illinois.

In determining the moisture field, WILLARD found high LFM forecast values of low-level moisture convergence over eastern Oklahoma and Kansas, and all of Missouri and the southern two-thirds of Illinois to occur at 00.00Z, 30 April 1984—the middle of WILLARD's valid time period. Examination of model forecast dew points at an average altitude of 1500 m (850 millibar pressure surface) and near the surface also confirmed this to WILLARD. Further, various stability indices were examined by WILLARD and found to be favourable for severe weather, especially because a triggering mechanism was present.

The SELS forecast reasoning stated that due to the strengthening of the low level wind field during the day (29th), there would be an attendant strong influx of low-level moisture into the threat region. High model forecast values of near-surface dew points were also mentioned as being within the threat area. Both forecast reasonings examined the low-level wind field for location of the maximum wind speeds and found favourable conditions for the formation of severe thunderstorms.

The trigger mechanism was the result of the strong upper-level, low-pressure system, as identified by the large values of vorticity advection approaching both threat areas from the southwest. WILLARD used maximum vorticity analysis from the LFM model data to locate intense activity. This is basically what SELS did, in that the SELS forecaster examined contour maps of LFM forecast vorticity and upper-level wind speed and identified areas experiencing maximum vorticity advection and strong upper-level winds.

While not explicitly mentioned by SELS on their 08.00Z forecast discussion, the presence of a warm frontal boundary lying across southern Missouri and Illinois was picked up by WILLARD as being an important triggering mechanism. It has been found (Maddox *et al.*, 1980) that severe storms often reach maximum intensity within the environment attending sharp thermal boundaries, as was the case on the 29 April 1984.

WILLARD checked for the presence of an upper-level venting mechanism at several grid points and found that there was such a mechanism present in the form of a core of high-speed winds in the upper levels of the atmosphere. This was clearly stated by SELS in their reasoning as being a major contributing factor.

A factor considered by SELS but not included in the WILLARD knowledge-base is the presence of a mid-level dry slot (punch) moving into the threat area. Investigations of the literature show this to be of the prime importance in defining the extent and severity of thunderstorms (Miller, 1972). This may account for part of WILLARD's under-forecasting on this day.

Overall, the two forecasts had similar *CSI*s, but were different in *PoD*, *CBIAS*, and *FAR*. The SELS forecast discussion of their lines of reasoning was generally similar to the explanation given by WILLARD, although the SELS reasoning contained much finer scale details that WILLARD did not have. But when it is considered that this was a widespread severe weather day, most parameters were easily identified from the model forecast data. Additionally, comparison of the LFM forecast maps with later maps that show how conditions actually turned out indicate that the LFM model had a reasonable handle on the major features contributing to severe weather on this day.

25 May 1984—test case no. 2

On this day, a slow moving cold front touched off severe thunderstorms along its leading boundary as it moved southeastward across the southern Great Plains region during the afternoon. Because severe thunderstorms formed in a narrow line along the front, severe weather reports were mainly restricted to this narrow frontal zone, and consequently were not very widespread. There were only 22 severe weather reports within the forecast domain consisting of mostly large hail and high winds, with only a few minor tornadoes being observed (NOAA, 1984). Missouri and Illinois were the only states to report severe weather—all of it occurring in thunderstorms along the cold front. This day would be considered a minor severe thunderstorm day by meteorologists.

Even though there were a relatively small number of severe reports, the meteorological conditions favourable for severe weather along the front were present and easily recognized by both SELS and WILLARD (Figure 7). The SELS outlook had a Slight risk threat area that covered northeastern Oklahoma, eastern Kansas, most of Missouri, most of Illinois, and portions of extreme southeast Iowa and northwest Arkansas. WILLARD had a moderate risk outlook for virtually the same area that SELS outlooked, with a narrower major axis than SELS' area. The forecast area for SELS was about 214,000 sq sm, while for WILLARD the area was about 78,000 sq sm.

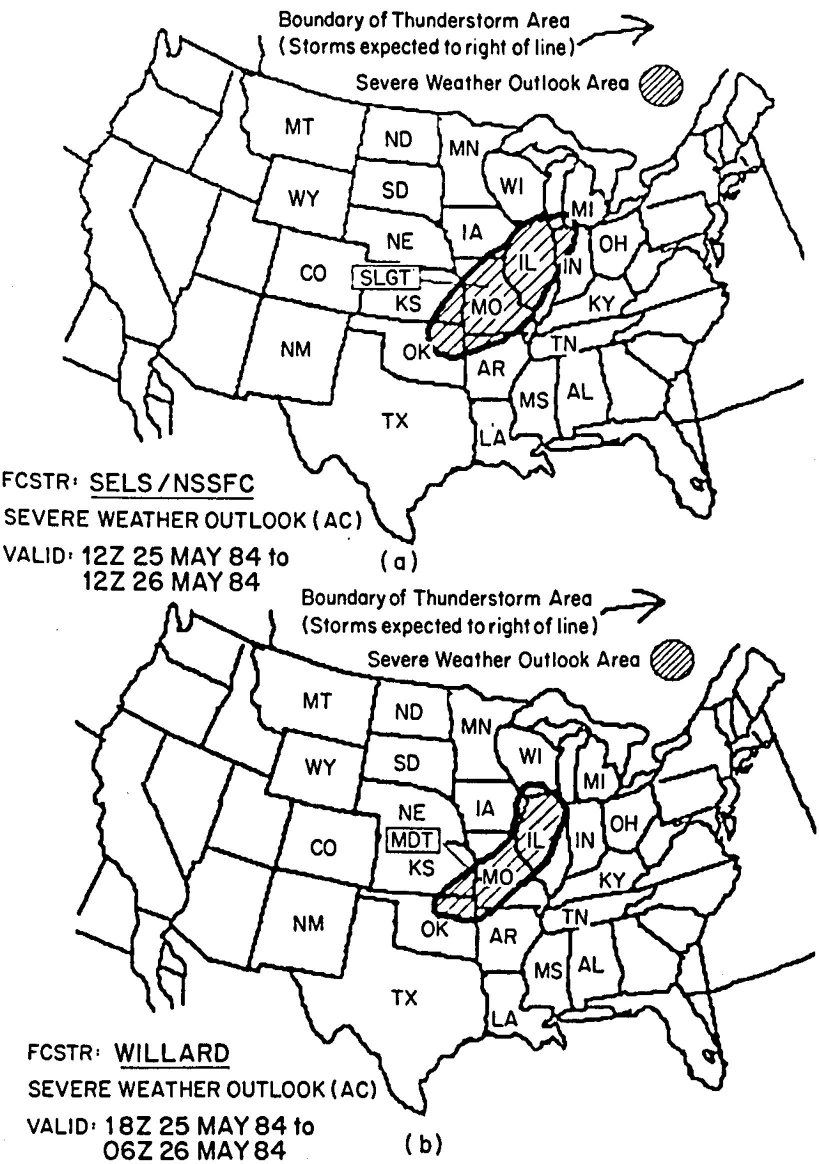

Figure 7. Map of SELS and WILLARD outlooks—25 May 1984.

In terms of the statistical measures of forecast ability, the *CSI* for each forecast were similar. Again, the probability of detection for the SELS outlook was high; 93 per cent in this case, but the *FAR* was also high; being 0.52, resulting in a *CSI* for SELS of 0.46. In contrast, WILLARD had a relatively low *FAR* of 0.48, coupled with a relatively low probability of detection of 68 per cent, to yield a *CSI* of 0.42. So because of the slightly smaller area forecast by WILLARD the *CSI* scores were very close. WILLARD had 15 out of 22 severe reports within its outlook, with one out of two tornadoes also included.

The coverage bias was unity for WILLARD, since the areal coverage was 6.8 per cent. The coverage bias for SELS' outlook slightly under-forecast this day. The good area for WILLARD was one half of the forecast outlook area; almost 41,000 sq sm, while the bad area was about 37,000 sq sm.

Comparing the two forecasts, there was little difference in forecast reasoning behind the two outlook areas: a cold front, in concert with an approaching short wave of low pressure, would provide the primary lifting mechanism of a moist and unstable air mass over Missouri and Illinois, with a strong flow of upper-level winds providing a suitable venting mechanism.

Both forecasts pointed out the strong focusing of low-level moisture convergence along the frontal boundary. In examining the LFM model output, it appeared that the model was a little slow in moving the front southeastward, which may have affected WILLARD's *PoD* because it did not extend the outlook area further into southern sections of Missouri and Illinois (where there were some severe reports).

Both SELS and WILLARD acknowledged the movement of the cold front across the area accompanied by a strong low-level wind field. WILLARD also indicated that if surface heating occurred, any thunderstorms occurring could be severe. Examination of surface data indicated periods of sunshine in the vicinity of the severe weather reports in Missouri and Illinois. WILLARD found the absence of preventive factors was also favourable for severe thunderstorms.

7 June 1984—test case no. 3

During an 11-h period from mid afternoon on 7 June to the early morning of 8 June, a massive severe weather outbreak struck many states in the central US unleashing a violent torrent of killer tornadoes covering four states. There were over 120 severe weather reports within WILLARD's forecast domain (NOAA, 1984).

There were 39 confirmed tornadoes occurring primarily in Iowa and Wisconsin, with some tornadoes reported in northern Missouri and southern Minnesota (Figure 8). One of these tornadoes tracked on the ground for a incredible distance of 204 km (127 miles) from northern Missouri across east central Iowa. This was the longest tornado track observed in 1984. It was rated as a devastating tornado for most of its existence and was responsible for three deaths and extensive property and agricultural damage (NOAA, 1984). It completely levelled or extensively damaged all buildings in the towns of Wright and Delta, Iowa resulting in property damage exceeding 30 million dollars.

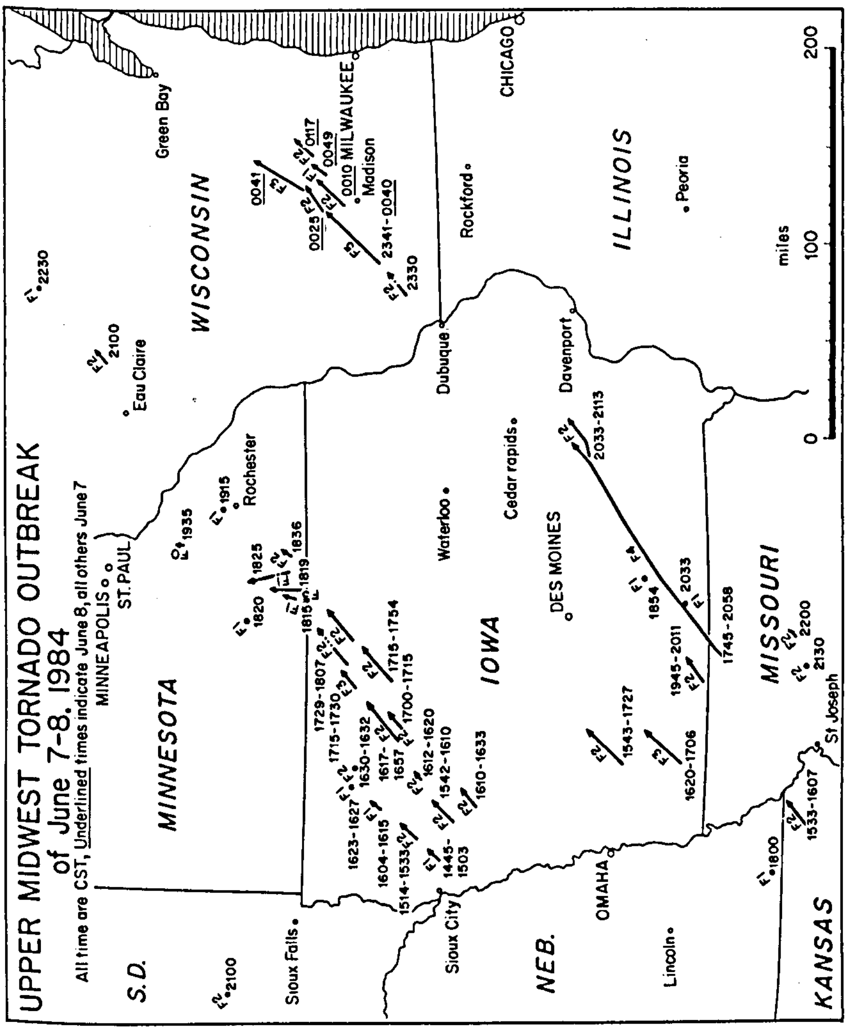

Figure 8. Upper midwest tornado outbreak of 7–8 June 1984 (from Storm Data—NOAA, 1984).

On this day the most violent tornado of the 1984 season demolished the town of Barneveld in south central Wisconsin and killed nine people. It was the most powerful tornado of its class since 1982 and only the second since 1977 to reach such great intensity (Ferguson *et al.*, 1985). Total property damage from this storm that tracked over 58 km (36 miles) on the ground exceeded 40 million dollars. Winds were estimated in excess of 420 km h^{-1} (261 miles h^{-1}) near Barneveld.

The SELS 08.00Z outlook called for a Moderate risk of severe thunderstorms over portions of northeastern Kansas, northwestern Missouri, the western two-thirds of Iowa, extreme western Wisconsin, southern Minnesota, southeastern South Dakota, and eastern Nebraska (Figure 9a). This Moderate risk area was enclosed within a large Slight risk area as shown in Figure 9a, which included large portions of Wisconsin, Minnesota, and Kansas. The Moderate risk area covered about 190,000 sq sm, while the entire outlook spanned roughly 413,000 sq sm. The SELS forecasters noted that the stage appeared set for a more active day than the past few days. All of the killer tornadoes that occurred in Iowa were contained in SELS' Moderate risk outlook, while the tornado that struck Barneveld was under a Slight risk.

WILLARD had a complex outlook, with two Slight risk areas flanking a fairly large High risk area (Figure 9b). The High risk area covered most of the Iowa and southern Minnesota, and the northwest portion of Missouri. All of the killer tornadoes that occurred in Iowa were contained in WILLARD's High risk outlook area. The western Slight risk area covered central and northeast Kansas, south central and southeast Nebraska, and portions of extreme northwest Missouri and southwest Iowa. The eastern Slight risk area from WILLARD covered northeastern Missouri, central Illinois, and southeastern Iowa. Notice that no severe weather was outlooked by WILLARD for south central Wisconsin, where the Barneveld killer tornado occurred. The total WILLARD outlook area encompassed almost 138,000 sq sm, with the High risk area covering about 80,000 sq sm.

The verification statistics show that, again, the *CSI* for each forecast was similar; SELS had a *CSI* of 0.51 while WILLARD's *CSI* was 0.54; a forecast is considered a 'good' one by members of SELS if the *CSI* is above 0.5 (Leftwich, NSSFC, personal communication, 1985). The same trend discussed in the previous test study day (no. 2) continued here. The probability of detection for the SELS outlook was 94 per cent for all severe reports and 87 per cent for all reports of tornadoes. WILLARD's outlook had a *PoD* of 71 per cent for all severe weather reports and 76 per cent for all reports of tornadoes.

The *FAR* was again lower for WILLARD, with a value of 0.31 compared to SELS' *FAR* value of 0.47. In this case, the primary threat areas denoted by SELS as a Moderate risk area and by WILLARD as a High risk area were very similar in areal extent and region, as an examination of Figure 9 shows. However, the Slight areas were different: SELS had more areal coverage of Slight risk than did WILLARD, and this might have accounted for the slightly higher *FAR* for SELS.

Examination of forecast reasoning given in the 08.00Z SELS outlook and the explanation given by WILLARD show that most of the major

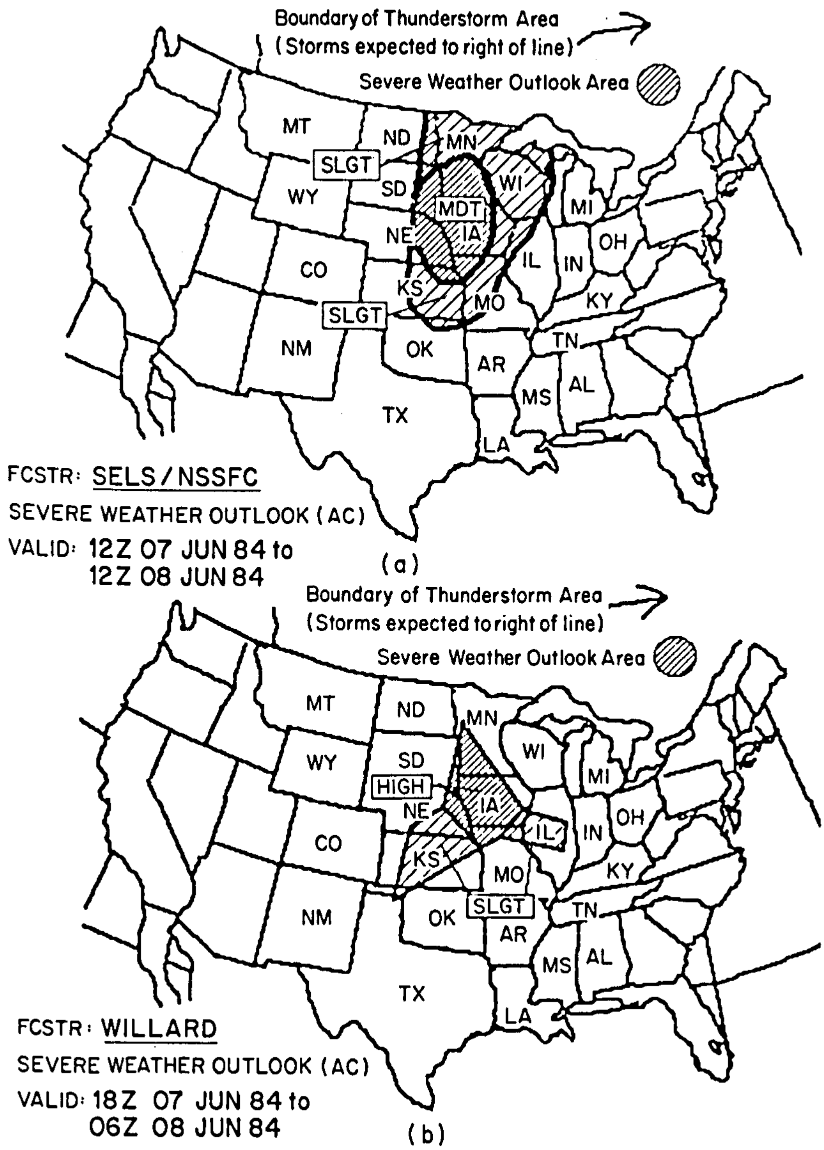

Figure 9. Map of SELS and WILLARD outlooks—7 June 1984.

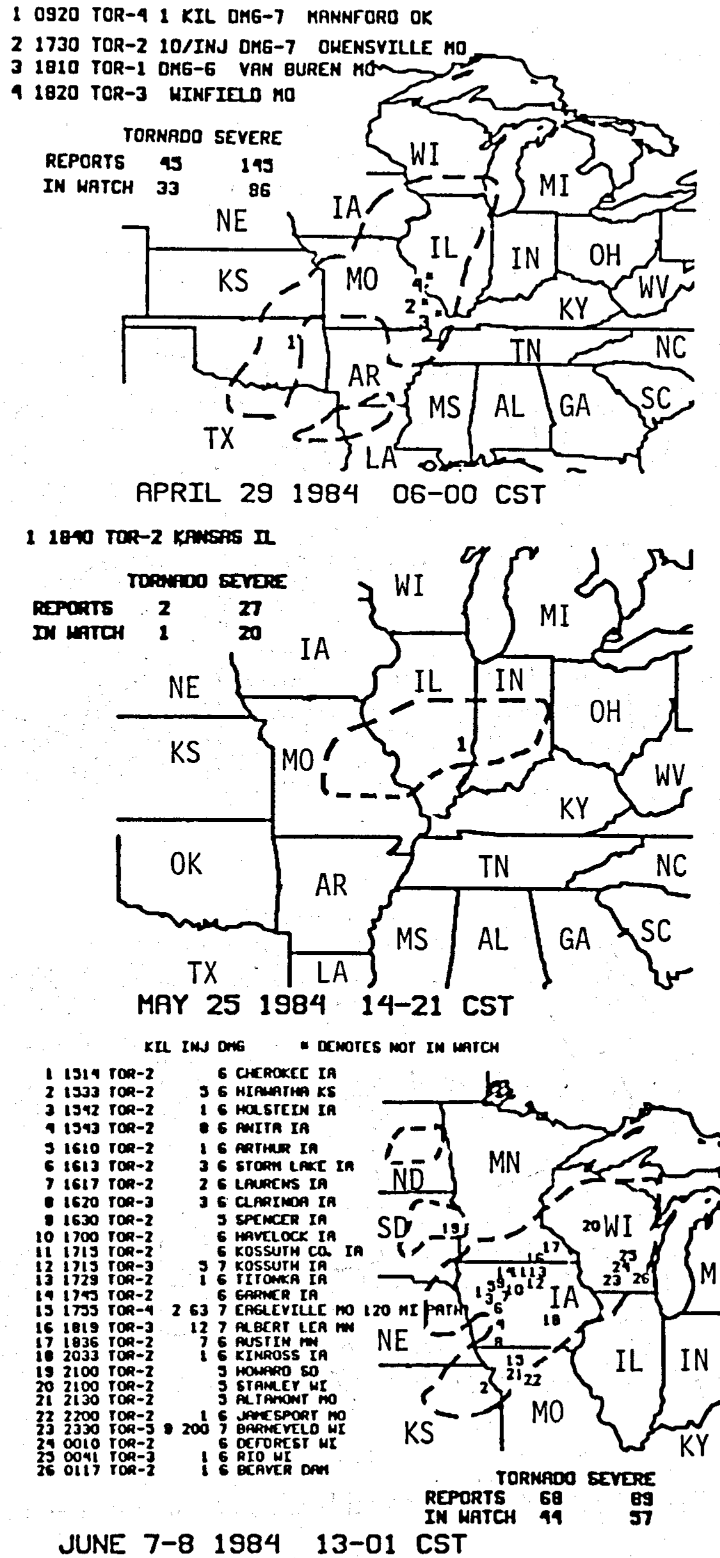

Figure 10. Significant severe weather events for the three days discussed in Appendix (dashed line encloses area of days' organized severe activity), modified from Hales and Crowther (1985).

factors were identified in each. Both determined that a significant upper-level trigger mechanism was present in the form of short-wave, low-pressure trough moving out of Nebraska into southern Minnesota during the period. Both found areas of strong low-level moisture convergence across Iowa and southern Minnesota. Both identified a suitable upper-level venting mechanism occurring over the region.

There was mention in the SELS forecast reasoning of a 'dry punch' (a narrow tongue of dry air at mid levels) extending from northwest Missouri into southern Minnesota and southwestern Wisconsin. Examination of Figure 9 shows that this feature was possibly a key parameter to defining the general direction and movement of the tornadoes (Miller, 1972). WILLARD did not perform any checks for this type of feature.

The coverage bias for WILLARD's outlook was 1.5 compared with SELS' outlook bias of 1.5. Both of these values indicate that each outlook under-forecast the actual areal extent of severe weather, which was about 19 per cent of the MDR blocks. However, WILLARD did appear to have the proper category in the major Iowa tornado outbreak area.

Overall, both forecasts were similar in the critical score index, category, and areal coverage, but different in *PoD* and *FAR*. The WILLARD forecast strategy appeared consistent with the relevant features indicated in the 08.00Z SELS outlook discussion. An examination of the LFM model output with model verification data showed the 24-h model forecast used by both SELS and WILLARD to be fairly reasonable in handling the major feature associated with this severe outbreak day. Therefore, since the LFM model input data into WILLARD was reasonable, it appears that the procedural rules used by WILLARD in producing the outlook areas were comparable to SELS' line(s) of reasoning.

REFERENCES

Crisp, C. A. (1979) Training guide for severe weather forecasters. *Air Weather Service, Tech. Note 79/002*. (Available from NTIS, Springfield, Va., No. ADA083125).

Donaldson, R. J., Dyer, R. M., and Kraus, M. J. (1975) An objective evaluator of techniques for predicting severe weather events. *Preprints 9th Conf. on Severe Local Storms,* Norman, OK, American Meteorological Society, pp. 321–6,

Ferguson, E. W., Otsby, F. P., and Leftwich, P. W., Jr (1985) Tornadoes cause record number of fatalities. *Weatherwise* **38,** 20–25, 48–49.

Foster, D. S. and Bates, F. C. (1956) A hail size forecasting technique. *Bull. Am. Meteor. Soc.* **37,** 135–41.

Galway, J. G. (1956) The lifted index as a predictor of latent instability. *Bull. Am. Meteor. Soc.* **37,** 528–9.

Hales, J. E. and Crowther, H. G. (1985) Severe thunderstorm cases of 1984. *NOAA Tech. Memo NWS NSSFC-7* (available from NTIS, Springfield, Va., No. PB85-210748/AS).

Hudson, H. R. (1971) On the relationship between horizontal moisture convergence and convective cloud formation. *J. Appl. Meteor.* **10,** 755–62.

Kerr, R. A. (1984) Forecasting of severe storms improved. *Science* **223** (4635), 477.

Leftwich, P. W. Jr. (1985) Verification of severe local storm forecasts issued by the

National Severe Storms Forecast Center, 1984). *NOAA Tech. Memo NWS NSSFC-9* (available from NTIS, Springfield, Va., No. PB86-128105/AS).
Maddox, R. A. and Doswell, C. A., III (1982) Forecasting severe thunderstorms: a brief consideration of some accepted techniques. *National Weather Digest* **7**, 26–30.
Maddox, R. A., Hoxit, L. R., and Chappell, C. F. (1980) A study of tornadic thunderstorm interaction with thermal boundaries. *Mon. Weather Rev.* **108**, 322–36.
Mandics, P. A. and Brown, R. C. (1985) When will the storm get here. *Research & Development* January, 72–7.
McNulty, R. P. (1978) On upper tropospheric kinematics and severe weather occurrence. *Mon. Weather Rev.* **106**, 662–72.
Michie, D., Muggleton, S., Riese, C. E., and Zubrick, S. M. (1984) RuleMaster: a second-generation knowledge engineering facility. *Preprints First Conf. on Artificial Intelligence Applications,* Denver, Colo., IEEE Computer Society and American Association for Artificial Intelligence, pp. 591–7.
Miller, R. C. (1972) Notes on analysis and severe-storm forecasting procedures of the Air Force Global Weather Central. Air Weather Service, *Tech. Report 200* (*Rev.*) (Available from NTIS, Sprintfield, Va., No. Ad-744-042.)
National Oceanic and Atmospheric Administration (1984) *Storm Data* **26,** US Department of Commerce. (Available from National Climatic Data Center, Ashville, N.C.).
Otsby, F. P., Jr (1979) The value of the convective outlook as a planning aid. *Preprints 11th Conf. Severe Local Storms,* Kansas City, Mo., Americal Meteorological Society, pp. 625–7.
Pearson, A. and Weiss, S. J. (1979) Some trends in forecast skill at the National Severe Storms Forecast Center. *Bull. Am. Meteor. Soc.* **60,** 319–26.
Porter, J. M., Means, L. L., Hovde, J. E., and Chappell, W. B. (1955) A synoptic study on the formation of squall lines in the north central United States. *Bull. Am. Meteor. Soc.* **36,** 390–6.
Quinlan, J. R. (1979) Discovering rules by induction from large collections of examples. In *Expert systems in the microelectronic age* (ed. D. Michie). Edinburgh University Press, Edinburgh.
Reynolds, D. W. (1983) Prototype workstation for mesoscale forecasting. *Bull. Am. Meteor. Soc.* **64,** 264–73.
Suomi, V. E., Fox, R., Limaye, S. S., and Smith, W. L. (1983) McIADAS III: A modern interactive data access and analysis system. *J. Climat. Appl. Meteor.* **22,** 766–778.
Uccellini, L. W. and Johnson, D. R. (1979). The coupling of upper and lower tropospheric jet streaks and implications for the development of severe convective storms. *Mon. Weather Rev.* **107,** 682–703.
University Corporation for Atmospheric Research (1982) *National STORM Program: Scientific and Technological Bases and Major Objectives.* Report Prepared for National Oceanic and Atmospheric Administration, under contract NA81RAC00123.
Weiss, S. J., Kelly, D. L., and Schaefer, J. T. (1980) New objective verification techniques at the National Severe Storms Forecast Center. *Preprints 8th Conf. Weather Forecasting and Analysis,* Denver, CO, American Meteorological Society, pp. 412–19.
Weiss, S. J. and Reap, R. M. (1984) Performance characteristics of the TDL automated severe storm outlook: a statistical evaluation. *Preprints 10th Conf. Weather Forecasting and Analysis,* Clearwater, FL, American Meteorological Society, pp. 181–8.
Whitney, L. F., Jr. (1977) Relationship of the subtropical jet stream to severe local storms. *Mon. Weather Rev.* **105,** 398–412.