A Simple Technique for Multiple-Parameter Pattern Recognition with an Example of Locating Fronts in Model Output

STEVEN S. FINE AND ALISTAIR B. FRASER

The Pennsylvania State University, State College, Pennsylvania

(Manuscript received 25 October 1989, in final form 19 July 1990)

ABSTRACT

A statistical pattern recognition technique called IREW is described. IREW has several strengths, such as fast execution, small storage requirements, and incremental learning, that may make the technique useful for many meteorological pattern recognition problems. A weakness of IREW is that it may not recognize complex patterns as well as more sophisticated pattern recognition techniques do.

IREW analyzes data by using a derivative of Bayes' rule to multiplicatively combine the influence of predictors. The weight assigned to each predictor is determined empirically from a training dataset containing the data and verifications for multiple cases. IREW uses several methods to select useful subsets of a large set of predictors.

The objective identification of surface fronts in Nested Grid Model forecasts is described to illustrate how IREW can be applied to a typical pattern recognition problem. The work consisted of identifying factors related to fronts and using some of those factors to make analyses. Given 27 000 predictors, IREW selected many that meteorologists associate with fronts. IREW's analyses were compared to subjective analyses for seven test cases. In this limited test, IREW performed similarly to meteorologists in terms of the number of grid points correctly classified as frontal or non-frontal.

1. Introduction

Pattern recognition techniques are likely to become an increasingly important component of meteorological analysis software. These techniques infer from data the existence of a feature or pattern that is not explicitly represented in the data. Locating a front on a chart containing station models is a common meteorological, pattern recognition problem. Analysis software with pattern recognition components could aid meteorologists by providing an overview of the data and a context for detailed queries. Fronts and pressure systems drawn on surface charts, for instance, describe some of the atmosphere's large-scale structure and trigger the recall of mental models that help an analyst interpret each station's observations.

This paper presents a simple statistical pattern recognition technique, IREW (IREW is a former acronym that was retained despite a change in the underlying terminology), which may be a useful tool for examining meteorological datasets. To provide a context for understanding the strengths and weaknesses of IREW, some popular pattern recognition techniques will be be discussed briefly. The general characteristics of IREW will then be described, and IREW will be compared to other techniques that use a similar weighting method. The implementation of IREW is presented, and an application of IREW to identifying fronts in model output is described. Finally, some of IREW's characteristics are summarized.

2. Automated pattern recognition techniques

A broad range of pattern recognition techniques have been applied to meteorological problems. The following brief descriptions of techniques are intended to be a short introduction to some popular methods that examine multiple predictors, not a comprehensive survey.

All of the following techniques except expert systems learn to recognize patterns from examples that are presented to the computer [some expert systems do learn from examples (Wilkins et al. 1986), but most do not]. Techniques that learn are often easier to apply to a new problem than techniques that do not learn because someone must determine the criteria that a nonlearning technique should use to recognize patterns.

Learning can be supervised or unsupervised. Pattern recognition techniques that depend on supervised learning must be given the correct classification of examples. Unsupervised learning techniques use just the structure observed within the examples to classify new sets of data. Unsupervised learning is harder to achieve than supervised learning (Duda and Hart 1973). The difference between the two types of learning is illustrated by the difference between learning to recognize prime numbers by examining a list of prime numbers

Corresponding author address: Steven S. Fine, 1436 Houserville Road, State College, PA 16801.

and a list of composite numbers, and learning to recognize prime numbers by examining a list of positive integers. Of the learning techniques described below, only cluster analysis learns without supervision.

Some techniques learn incrementally. This means that new learning cases can be presented to the computer without previous learning cases being examined again. Incremental learning is important if it is not possible to keep all of the learning examples online simultaneously. Incremental learning also makes refinement of pattern recognition programs more convenient; if the initial results are not satisfactory then more learning examples can be examined without reprocessing the original learning examples.

a. Decision trees

Decision trees are lists of questions that are organized such that answering one question leads to a decision or to another question. Identifying an object in the game 20 Questions is an example of the use of a decision tree. Decision trees can be designed by meteorologists and then traversed automatically (Dockus 1988; Wu et al. 1985).

There are also several techniques for generating decision trees automatically (Michie et al. 1984; Brieman et al. 1984; Quinlan 1986). The general process is to repeatedly extend the tree by choosing the variable, from a set of specified variables, with the most predictive power for a subset of examples. This variable is incorporated into a question. Each branch from the question is then considered. Branching stops when the answer to a question leads to a decision about the situation that is being analyzed.

The concepts embedded in decision trees that are automatically produced can often be understood by people, and decision trees can be applied to new data quickly. Generating the trees though, can require a large amount of a computer's time and memory because almost all predictors are evaluated each time a new question is added to the tree. Also, these techniques usually cannot support incremental learning. With some extra computations and storage, decision trees can be designed to tolerate missing data (Brieman et al. 1984; Quinlan 1987).

b. Expert systems

An expert system is a computer program that is designed to make expert judgments in a limited domain of interest, such as predicting hail. Conway (1989) discusses meteorological expert systems. An expert system is usually developed by interviewing a human expert to learn what guidelines he uses to make decisions in the domain of interest. Those guidelines are encoded into "rules" that the computer will apply to data to make a decision. The rules are often of the form "if proposition *X* is true then proposition *Y* is true." Simple expert systems may be similar to decision trees, but most expert systems have more complicated data structures and control algorithms than decision trees.

Since expert systems are usually based on a person's ideas that are translated into a program, expert systems can incorporate complex reasoning and models of phenomena (Campbell and Olson 1987; Elio et al. 1987; Peak and Tag 1989). The translation process can be slow and difficult though (Hoffman 1987; Hayes-Roth et al. 1983).

The capability of expert systems to contain complex representations of knowledge does not imply that expert systems will be more accurate than other techniques. Stewart et al. (1989) showed that a regression equation could predict the occurrence of hail almost as accurately as an expert system. Carroll (1987) discusses the advantages of simple linear models over expert systems, especially in medical domains, but the discussion is overly critical of expert systems. For instance, the author does not credit expert systems with the ability to make more complex decisions, such as suggesting combinations of drugs, than can be made by linear equations.

c. Regression equations

A common technique is to compute a regression equation that converts known values into an estimate of an unknown value. Regression equations are used for many purposes, such as forecasting (Carter et al. 1989) and data quality control (Elmore et al. 1989). Generally the time and memory required to compute regression equations grows nonlinearly as the number of predictors is increased. Regression techniques can support incremental learning, but missing data may not be easy to handle.

d. Cluster analysis

The general principle of cluster analysis is to partition samples into groups such that two samples in a group are similar and two samples in different groups are dissimilar. Many algorithms have been developed to accomplish the partitioning (Anderberg 1973). An example of a meteorological application of cluster analysis is clustering 500 mb pressure patterns to classify flow regimes over the Pacific Northwest (Yarnal 1985). Few generalizations can be made about clustering algorithms because the characteristics of the algorithms vary widely.

e. Neural nets

Neural nets consist of a number of simple processing units that work in parallel to recognize complex patterns (Fahlman and Hinton 1987). Neural nets learn to recognize patterns by adjusting the links between processing units in response to training examples of data and analyses. Neural nets can learn incrementally, but usually the learning process requires a relatively large amount of processing time because there are many links that must be adjusted iteratively in response to each learning case. Also, people often have difficulty translating the connections that are formed between processing units into easily recognizable concepts. After a neural net has learned to recognize a pattern, the user may have to consider the net as a black box that produces results through an unknown process.

3. Overview of IREW

a. Combining the influence of predictors

An important component of IREW is its technique for combining the influence of predictors. This technique uses knowledge about the existence or absence of a characteristic in the data, such as a pressure trough, by multiplying the odds that a feature, such as a front, exists by one of two odds factors.

The odds factors are derived from Bayes' rule in odds form. Odds are related to probabilities by probability = odds/(odds + 1). Following Moninger (1988),

$$O_2 = (\mathrm{OF})(O_1), \tag{1}$$

where OF is an odds factor, O_1 is the prior odds that the feature exists, and O_2 is the posterior odds conditional on the characteristic of the data related to OF. The quantity OF has one value, OF_t, when the characteristic is present and another value, OF_f, when the characteristic is not present. Here OF_t is given by

$$\mathrm{OF}_t = p1/p2, \tag{2}$$

where $p1$ is the probability that the characteristic will be present given that the feature exists and $p2$ is the probability that the characteristic will be present given that the feature does not exist. The quantity OF_f has a similar form except the complementary probabilities are used because OF_f applies when the characteristic is not present,

$$\mathrm{OF}_f = (1 - p1)(1 - p2)^{-1}. \tag{3}$$

Consider an example where the target feature is a frontal zone and the characteristic is that precipitation was observed at a station. If the fraction of stations in frontal zones observing precipitation were twice the fraction of stations outside frontal zones observing precipitation, then OF_t would be 2 and OF_f would be less than 1. However, if precipitation were equally likely in frontal and nonfrontal zones then $\mathrm{OF}_t = \mathrm{OF}_f = 1$. In this case, the presence or absence of precipitation would have no effect on the odds that a front exists.

This Bayesian technique for combining the influence of predictors was used in the expert systems ARCHER (Moninger 1988) and Prospector (Gaschnig 1982). Gaschnig called OF_t "measure of sufficiency" because it indicates the degree to which a characteristic leads to the belief that the feature of interest exists. Similarly, Gaschnig called OF_f "measure of necessity." Odds factors have also been called "likelihood ratios" (Duda et al. 1976). The knowledge for ARCHER and Prospector was adapted from heuristics specified by a human expert. Each expert's heuristics included estimates of OF_t and OF_f or $p1$ and $p2$. For instance, an expert might feel that the presence of a particular radar signature would double his confidence that a hypothesis was true. People though, are often reluctant or inaccurate estimators of the influence of predictors (Moninger 1988; Dawes et al. 1989).

Equations (2) and (3) may produce a discontinuity in a characteristic's effect on the existence odds. Consider testing "is a station's temperature above the mean of all the observed temperatures?" with $\mathrm{OF}_t = 2$ and $\mathrm{OF}_f = 0.5$. A temperature a little below the mean would halve the odds, and a temperature a little above the mean would double the odds. Usually there is little justification for such a large change in the odds based on a small difference in relation to the mean temperature.

ARCHER and Prospector use certainty values to represent the degree of belief in a piece of evidence. The magnitude of the certainties ranges from $-C_{\max}$ to $C_{\max}$. Certainties of $C_{\max}$, 0, and $-C_{\max}$ correspond to evidence that is certainly true, completely unreliable, and certainly false. A scaled odds factor, OFS, is calculated from the odds factor and the certainty, C, by

$$\mathrm{OFS} = 1 + (\mathrm{OF} - 1)(|C|/C_{\max}). \tag{4}$$

Equation (1) is replaced by

$$O_2 = (\mathrm{OFS})(O_1). \tag{5}$$

Thus, unreliable evidence has little influence on the odds. The discontinuity in the example with the mean temperature could be eliminated by reducing the certainty as the temperature approached the mean.

STAGGER, a pattern recognition program (but not an expert system), used the same basic Bayesian technique [given by (1)–(3)], but the program learned OF_t and OF_f empirically (Schlimmer and Granger 1986). In addition to computing predictor weights, STAGGER also combined predictors to create new predictors when a sample was misclassified. STAGGER performed well on tests described by Schlimmer and Granger. They also noted that the empirical Bayesian technique produced results similar to what has been observed in tests of learning in animals.

IREW uses the same basic Bayesian technique used by ARCHER, Prospector, and STAGGER. IREW also incorporates empirical learning and certainties.

Each of the four programs (ARCHER, Prospector, STAGGER and IREW) combines the influence of multiple predictors by multiplying the odds factors for each predictor. This calculation is only probabilistically correct if the predictors are statistically independent. This assumption is rarely satisfied in meteorology because atmospheric parameters are often related.

Neapolitan (1987), who showed that there is no general way to correctly combine odds, suggested that it may be worthwhile in some cases to make adjustments to a method for combining independent odds, instead of completely discarding the method. IREW follows this approach by combining the influence of predictors as if the predictors were independent while limiting the interdependence between the predictors and allowing the user to adjust the odds threshold that is applied to the odds that a feature exists.

b. Complexity versus cost

In addition to not considering interactions when computing probabilities, IREW does not form new predictors from the predictors that are specified by the user. For instance, given sky cover and wind speed as predictors IREW will not discover that both clear skies and weak winds are required for radiational fog to form. Other techniques, including cluster analysis, neural nets, and decision trees, would be much more likely to discover the relationship.

IREW's neglect of interactions between predictors does not imply that the technique will be ineffective for problems where predictors are related. Several studies have suggested that simple means of combining the influences of predictors may capture much of the skill of experts in some fields (Dawes et al. 1989; Stewart et al. 1989).

Since most of IREW's algorithms consider the predictors to be independent, much of IREW's processing for a new example consists of performing a fixed sequence of calculations for each predictor. Also, IREW requires long-term storage for only four integers for each predictor. Thus, IREW is not subject to nonlinear growth in resource requirements when the number of predictors increases. As a result, IREW can be used to consider a large number of predictors.

4. The characterizations

a. Structure of IREW's characterizations

IREW's predictors are called characterizations. The characterizations have a structure that can easily be applied to a wide range of problems. Each characterization consists of a variable, a test, and two odds factors. The variable represents a value derived from data. Temperature, the magnitude of the gradient of wind speed, and the change in pressure over three hours are examples of variables. Examples of tests include "is the value larger than the values at neighboring grid points?" and "is the value below zero?" The result of applying the test to the value of the variable is "true" or "false." The combination of a variable and a test specify which characteristic of the data the characterization examines. For instance, a variable and test consisting of "is the pressure change greater than zero?" specify that the characterization's test is true for rising pressures. The odds factors determine the effect a characterization has on the odds that the target feature (e.g., a front) exists. In other words, each odds factor is a weight that determines the influence of a characterization. A characterization would be expressed in if-then-else form as "if the result of applying the test to the variable is true, then multiply the odds of existence by OF_t, else multiply the odds by OF_f."

A characterization can be applied wherever or whenever the variable has a value. For a spatial analysis, characterizations could be evaluated at each station or at each point on a grid. For an analysis of a time series, characterizations could be evaluated whenever observations were made.

When characterizations are evaluated to analyze a case, the initial odds that the target feature exists can be based on any appropriate source, such as climatology. The initial odds are the prior odds for the first characterization. The posterior odds conditional on the first characterization are the prior odds for the second characterization, and so on. The final odds provide an estimate of the likelihood that the feature exists at that space/time location. A threshold can be applied to the odds to decide whether or not the feature is expected to be present. The use of a threshold is illustrated below. Note that if the data needed to evaluate a characterization are missing for a case, a decision can be made based on the characterizations for which data are available.

b. Determining the odds factors

To determine the appropriate odds factors a series of learning cases is examined. Each learning case contains both the observations and the verifications for a set of samples. If a case contains observations and verifications for points on a grid at a particular time, then each grid point represents a sample.

For each characterization the experience gained from the learning cases is contained in four counts of samples: the number of samples coincident with the target feature (F_t), the number of samples not coincident with the target feature (F_f), the number of samples coincident with the feature where the characterization's test was true ($T_t | F_t$) and the number of samples not coincident with the feature where the test was true ($T_t | F_f$). For frontal analysis, F_t would be the number of grid points in frontal zones, F_f would be the number of grid points not in frontal zones, $T_t | F_t$ would be the number of grid points in frontal zones where the characterization's test was true and $T_t | F_f$ would be the number of grid points not in frontal zones where the characterization's test was true.

A characterization learns from a case by comparing the verification with the results from the characterization's test and incrementing the counts appropriately. The counts are used to calculate $p1$ and $p2$ by

$$p1 = (T_t | F_t)/F_t \tag{6}$$

$$p2 = (T_t | F_f)/F_f. \quad (7)$$

The odds factors can be calculated from (2) and (3).

Since learning from a case consists of just incrementing four counts, new learning cases can be examined without reevaluating the cases that were examined previously. Also, since learning from each sample requires only two additions, learning can proceed quickly once the result of a characterization's test has been calculated. Another advantage of this learning mechanism is that if the information needed to evaluate some characterizations is missing for a case, the remaining characterizations can still be evaluated.

5. Finding a useful subset of the characterizations

a. Strategy

There are an infinite number of possible characterizations, each related to a characteristic of the data. Deciding which characterizations to include in the analysis is an important part of using IREW. By specifying which characterizations are examined, the user determines which characteristics of the data will be examined and how much storage space and computational time will be needed. For well-defined problems, using a few characterizations may produce adequate results.

For exploratory work though, many characterizations should be specified to cover a broad range of possible relationships between the target feature and the data. An efficient way to specify a large number of predictors is to use modular characterizations. For example, variables could be considered to be composed of an observed parameter and a gradient operation. Choosing three observed parameters, such as pressure, temperature, and humidity, and three gradient operations, such as no operation, gradient, and Laplacian, would produce nine characterizations. If two tests were chosen then there would be eighteen characterizations. In addition to providing an efficient way to specify many rules, modularizing the components also allows the software that evaluates the characterizations to be relatively compact. A more extensive example of using modular characterizations is described below.

To reduce the computational resources required when evaluating characterizations and to produce better results, IREW will select three nested subsets of the characterizations specified by the user. IREW chooses two of the subsets based on the correlation between the characterizations and the target feature. First, IREW chooses the "relevant" characterizations, which are not statistically independent of the target feature. Next, IREW selects from the relevant set the "interesting" characterizations, which are highly correlated to the target feature. Finally, if the user desires automated analyses, then IREW chooses from the interesting set the "analytic" characterizations, which should work in concert to produce a reasonable analysis.

There is no claim that IREW's several techniques for culling characterizations are optimum; they are techniques that have been found to produce acceptable results. Different techniques will probably be developed as work with IREW continues.

b. Removing characterizations while learning

To select the relevant characterizations, IREW applies a chi-square test (Walpole and Meyers 1978, p. 268–273) to determine if the hypothesis that a characterization is independent of the target feature can be rejected. If the hypothesis of independence cannot be rejected at the 0.01 level of significance then the characterization is considered to be independent of the target feature. Characterizations that are independent of the feature are never considered again for learning and will not be examined in later selection steps. Eliminating characterizations while learning is especially important when a large number of characterizations of unknown quality are being tested. Reducing the number of relevant characterizations will reduce the time needed to learn from new cases.

The chi-square test is based on the four quantities that represent the accumulated experience for each characterization. The Appendix contains the algorithm used. No decision is made until 1000 samples have been tested to reduce the influence of anomalous cases.

c. Choosing characterizations that merit further exploration

At any time, the user can have IREW select a set of interesting characterizations consisting of characterizations that may merit further exploration. IREW forms the set by selecting characterizations that have been highly correlated with the target feature while avoiding obvious redundancy in the selected characterizations.

The maximum of an odds factor and its reciprocal is called an effect. It provides an indication of how much difference the presence or absence of a characteristic will make to an analysis. Experience has shown that characterizations that have large effects when the test is both true and false produce better results than characterizations that have a large effect for only one value of the test. For instance, "is the temperature above 110°?" might be true 0.5% of the time outside of frontal zones and 0.1% of the time within frontal zones. The corresponding odds factors would be $\mathrm{OF}_t = 0.20$ and $\mathrm{OF}_f = 1.00$. The effect for the true case would be 5, but the characterization would have little influence on the outcome of an analysis because the test would be false most of the time.

The set of interesting characterizations is formed by finding characterizations with both effects larger than a user-specified value. If multiple characterizations with large effects are known to be closely related, then only one of the related characterizations will be put in the interesting set. The example below illustrates one criterion that might be used to determine if characterizations are closely related. If there are too many interesting characterizations, the set can be reduced to an arbitrary number of characterizations with the largest expected effects. An expected effect is calculated by weighting the effects derived from the two odds factors (effect_t and effect_f) by the probabilities of the two outcomes of a characterization's test:

$$p_t = \frac{T_t | F_t + T_t | F_f}{F_t + F_f} \tag{8}$$

$$\text{expected effect} = \text{effect}_t \times p_t + \text{effect}_f \times (1 - p_t). \tag{9}$$

If few cases have been examined, then the set of interesting characterizations may change as each new learning case is processed. As more cases are examined, the list of interesting characterizations should stabilize. Some of the interesting characterizations may be the result of random correlations, but most will represent characteristics of the data that are related to the target feature. Determining the characteristic represented by each characterization may lead to new insights about the target feature.

d. Choosing characterizations for automated analysis

To choose the analytic characterizations IREW applies a final selection technique to the interesting characterizations. Note that the size of the interesting set determines how many characterizations will be considered for the analytic set. The Appendix contains the algorithm for selecting the analytic set. The general approach is to find for each case in a series of new cases a subset of the interesting characterizations that produces a good analysis. The accuracy of an analysis is measured by a score that can be designed to reflect the importance of the different types of errors that will occur. The analytic set of characterizations is the union of the characterizations chosen for each case.

Characterizations are chosen to maximize performance for individual cases, rather than maximizing performance for all of the cases simultaneously, so characterizations will be chosen to recognize a variety of situations. For example, assume that four cases are to be used to choose the analytic characterizations and that the target feature is manifested in one way in three cases and in a different way in the fourth case. If the characterizations were chosen to maximize a score for the four cases simultaneously, characterizations to recognize the characteristics of the feature in the fourth case might not be chosen.

IREW limits the interdependence between the analytic characterizations by not selecting a characterization if more than 50% of the variance of the characterization's certainty over the new cases is accounted for by the analytic characterizations that have already been chosen. A lower limit on the variance contribution would cause the analytic characterizations to be less interdependent, which might improve the accuracy of the probability estimates. A lower limit might also overly restrict the characterizations that would be considered for the analytic set.

The variance calculations require that the data for all of the new cases be available simultaneously. This is the only step where data for more than one case are needed simultaneously. The number of interesting characterizations and the number of cases examined while selecting the analytic characterizations can be adjusted to limit the amount of storage and time required.

The analytic characterizations should produce useful results when applied to a variety of situations. The performance of these characterizations can be tested by applying them to several new cases. If the performance is not satisfactory then more cases can be used to learn odds factors or to select the analytic characterizations.

6. An example: Identifying fronts in model output

a. Overview

IREW has already been tested in two domains. IREW was applied to synoptic data and radar reports to examine synoptic-scale factors that are related to thunderstorms. In this work, IREW examined more than 15 000 characterizations and identified many of the factors recognized by mesoscale forecasters, such as static stability, moisture supply, and upper-level support. IREW has also been applied to the task of locating fronts in Nested Grid Model (NGM) output. The frontal location work, which was more extensive than the thunderstorm work, will be described to illustrate how IREW can be applied to a typical problem.

Frontal location was chosen as a test of IREW because finding fronts is a common meteorological problem that requires the consideration of multiple parameters. The goal of the work was to explore how well IREW could duplicate an expert's surface frontal analyses, not to produce results of operational quality. Objective frontal analysis is a complex problem that might require the application of multiple techniques to produce operationally acceptable results.

The use of forecast fields made for a better test of IREW than would have been possible if observations had been used. Some important sources of information for frontal analysis of observations, such as satellite images, were not available to IREW. IREW did have access though to much of the data that would be used for frontal analysis of forecast fields. The verification for the test of IREW was derived from fronts on weather depiction charts that were routinely produced by a consensus of several local meteorologists for the following day. NGM 30-hour forecast fields [received from the National Meteorological Center (NMC) via a computer link] were the primary—but not sole—source of information for the meteorologists. The same NGM fields provided the data for IREW. Note that IREW tried to match the subjective analysis derived from forecast data, not the frontal locations that were derived from the next day's observations.

Cases were selected randomly; they appeared to represent a reasonable sample of synoptic situations.

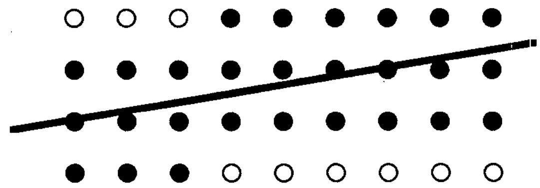

FIG. 2. An illustration of how the verification field was related to a subjectively analyzed front. The circles represent points on the grid. The line represents a section of a front. The black circles would be considered frontal, and the white circles would be considered nonfrontal.

b. Verification and characterizations

IREW was applied at 651 points on a grid that covered the continental United States and some surrounding regions, shown in Fig. 1. To simulate the frontal zone and to provide a cushion in case there were errors in translating a front's location to grid coordinates, grid points near a subjectively analyzed front were considered as frontal points in the verification, as shown in Fig. 2.

All of the types of fields that were available to IREW and some operators that had meteorological significance were incorporated into modular characterizations. Each characterization's variable had four components: the base value, a wind influence, a degree of smoothing and a gradient operation. The base values were surface pressure; lifted index; 1000–850 mb, 1000–700 mb, and 1000–500 mb thicknesses; vertical velocity (ω) at 850, 700, and 500 mb; relative humidity at 900, 850, 700, 500, 400 and 300 mb; and height, temperature, u- and v-wind components, vorticity, and divergence at 1000, 850, 700, 500, 400, 300, 250, 200, 150 and 100 mb. The wind influence was applied to the base value. Alternatives included no wind influence, advection ($-\mathbf{V} \cdot \nabla s$), and divergence of the wind multiplied by a scalar ($\nabla \cdot s\mathbf{V}$), where s was the base value. The degree of smoothing was expressed as the number of passes of a smoothing algorithm to be applied to the result of the wind influence operation. Zero, 8, and 16 passes of the smoother were tried. The gradient operation was applied to the output from the smoother. The possible gradient operations were none, the magnitude of the gradient ($\|\nabla s\|$), the Laplacian ($\nabla^2 s$), and the gradient operator used by Renard and Clarke (1965) to locate fronts:

$$GG(s) = -\frac{\nabla\|\nabla s\| \cdot \nabla s}{\|\nabla s\|}.$$

The characterizations also included 11 types of tests such as "is the variable's value above zero?," "is the variable's value above the mean for the grid?," and "is the variable's value bounded by the third and fourth quintiles of values for the grid?" The combinations of the components produced 29 304 characterizations. Of those combinations, 2112 were not testable. For instance, no wind information was available at 900 mb, so characterizations with a wind influence applied to the 900 mb relative humidity could not be tested. The remaining 27 192 characterizations were examined.

For most of the characterizations a certainty function was used that was -1 for roughly $\frac{1}{6}$ of the samples, 1 for roughly $\frac{1}{6}$ of the samples, and linearly varied from -1 to 1 for the remaining samples. The certainty function seemed to have a small influence on the results of the analyses.

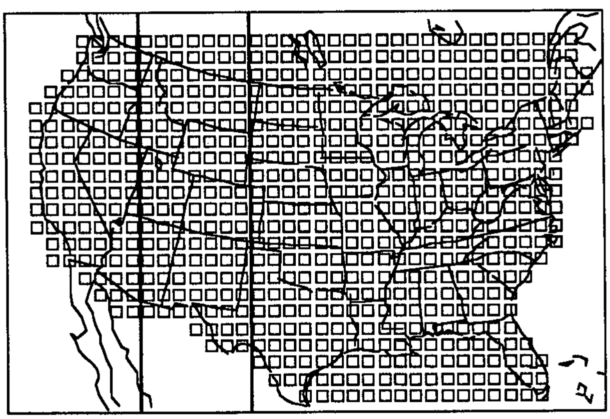

FIG. 1. Locations of the 651 points in the analysis area. The vertical lines enclose the area that was considered to be near the Rocky Mountains.

c. Characterization selection

Forty cases were provided in the learning phase. The learning cases were primarily from December, January, and February of the winter 1988/89. Frontal zones comprised 13% of the samples in the learning cases. At the end of the learning phase, 23 665 characterizations were found to be independent of the frontal zones, leaving 3527 relevant characterizations. Characterizations with both effects above 1.4 were considered for the interesting set. To limit redundancy, if two characterizations with the same variable (i.e., the two characterizations only differed in the tests they applied to the variable) had suitable effects, then the characterization with the lower expected effect was discarded.

To save time and storage space, only the top 75 of 110 interesting characterizations were retained. Some of the characteristics that were found to be related to fronts were

- Laplacian of 1000 mb height above zero
- Laplacian of 700–1000 mb thickness below zero
- 1000 mb height less than the mean
- 1000 mb temperature gradient above the mean
- 1000 mb relative vorticity above zero
- Laplacian of v-component (positive for southerly winds) of 500 mb wind below zero
- 900 mb relative humidity above the mean

The odds that a point in the learning cases was in a frontal zone, 0.149, was used as the prior frontal odds. While selecting the analytic characterizations, a point with odds above 0.209, 1.4 times the prior frontal odds, was considered to be classified as a frontal point. If the threshold had been higher, the algorithm for selecting the analytic characterizations would not have been able to select some characterizations as the first one to apply to a case (because the odds factors for some characterizations would have been insufficient to raise the initial odds above the threshold). After the analytic characterizations were chosen, the odds threshold could be adjusted to control the bias of the analytic characterizations. A higher threshold decreased the number of points that were classified as frontal.

The Gringorten skill score (Gringorten 1967), in the form given by Murphy and Daan (1985), was chosen as the measure of performance because it gives significantly greater weight to the identification of frontal points than to the identification of nonfrontal points. If identifying nonfrontal points were rewarded as well as identifying frontal points, then classifying all points as nonfrontal would produce a good score. The Gringorten skill scores for a perfect analysis and a random analysis should be near 1 and 0, respectively. Other skill scores that provided similar rewards for correctly identifying frontal points should have produced similar results.

To choose the analytic characterizations, IREW examined six new cases, from March, April, and May 1989, and selected fourteen characterizations. Testing of the fourteen characterizations showed that they performed the worst in the vicinity of the Rockies, the region where the local meteorologists had the least trust in the NGM output.

To compensate for the different guidelines used by the meteorologists near the Rockies, a second set of characterizations was evaluated for the region marked in Fig. 1. The same learning and test cases that were used for the country-wide characterization selection were used for evaluating characterizations over the Rockies. All of the characterizations selected for the country-wide work tested for a value above zero or above the mean, so the initial set of characterizations for the Rockies was limited to the 5328 characterizations that tested for a value above zero or for a value above the mean. From those characterizations, IREW formed an analytic set of ten characterizations. The analyses presented below were made with the Rocky Mountain analytic characterizations near the Rockies and the original analytic characterizations over the rest of the domain.

d. Testing the analytic characterizations

To test the performance of the analytic rules, seven cases were randomly selected from March and April 1990. The routine, consensus depiction charts were taken to be the verification. To estimate the variability that might exist between meteorologists' opinions, one of the meteorologists that contributed to the consensus analyses made an independent frontal analysis for each of the seven cases before contributing to the consensus analysis. The fronts in the individual analyses were treated by the same process, illustrated in Fig. 2, that was applied to the consensus analyses. If the differences between IREW's analyses and the consensus analyses were comparable to the differences between the individual analyses and the consensus analyses, then IREW could be considered to possess some skill at identifying fronts. This experiment probably underestimated the variability between meteorologists' analyses because the person making the individual analyses also contributed to the consensus analyses.

For the sake of comparison, IREW's odds threshold was adjusted so IREW's bias for the spring 1989 cases was close to the bias of the individual analyses for the 1990 cases. The resulting threshold, 4, was used by

TABLE 1. Performance of the individual analyst and IREW for the seven test cases when the consensus analyses were used as the verification.

	Case	Frontal points correctly identified (%)	Points correctly classified (%)	Bias	Gringorten skill score
Individual analyses	1	62	84		0.58
	2	53	89	1.30	0.59
	3	52	87	0.74	0.47
	4	54	88	0.96	0.55
	5	36	80	0.84	0.13
	6	67	86	1.90	0.74
	7	63	86	1.10	0.65
	All 7	57	86	1.10	0.53
IREW's				1.10	
analyses	1	42	87	0.67	0.41
	2	31	84	0.62	0.30
	3	70	81	1.80	0.57
	4	59	87	1.00	0.59
	5	73	88	1.80	0.44
	6	65	86	1.10	0.73
	7	46	87	0.73	0.48
	All 7	54	86	1.00	0.50

TABLE 2. Contingency tables for the individual analyses and IREW when the consensus analyses were used as the verification.

		Consensus Analyses		
		Frontal points	Nonfrontal points	Total
Individual analyses	Frontal	382	355	737
	Nonfrontal	292	3528	3820
	Total	674	3883	4557
IREW's analyses	Frontal	363	333	696
	Nonfrontal	311	3550	3861
	Total	674	3883	4557

IREW to analyze the cases from 1990. Note that IREW was not tuned with the 1990 data.

Table 1 shows some measures of performance for the individual analyses and IREW's analyses. The bias is the number of frontal points in the analysis divided by the number of frontal points in the consensus analysis. Table 2 contains contingency tables for the comparisons to the consensus analyses. Figures 3, 4, 5 show

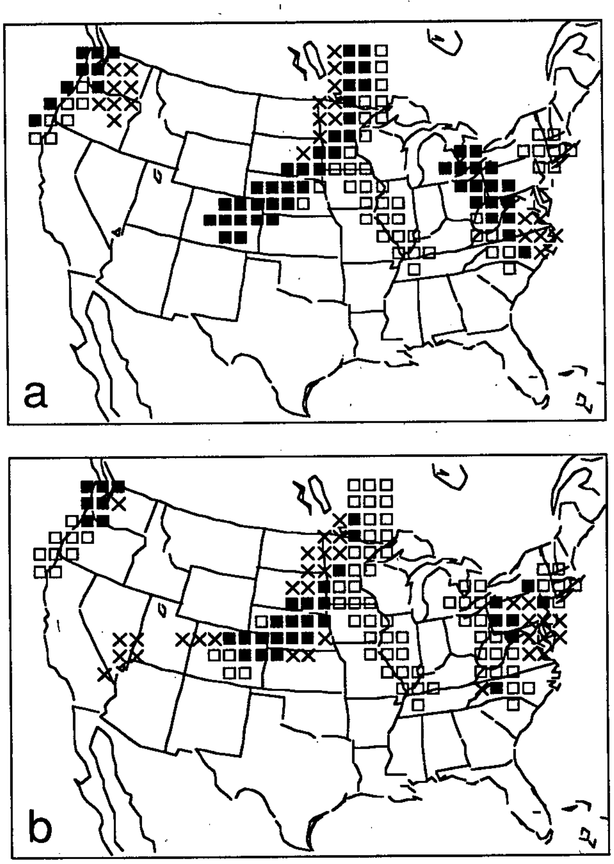

FIG. 3. Results of (a) the individual analysis for case 2 and (b) IREW's analysis for case 2. Squares represent points that were considered to be in frontal zones in the consensus analysis. Black squares were correctly classified; white squares were incorrectly classified. X's are nonfrontal points that were classified as frontal.

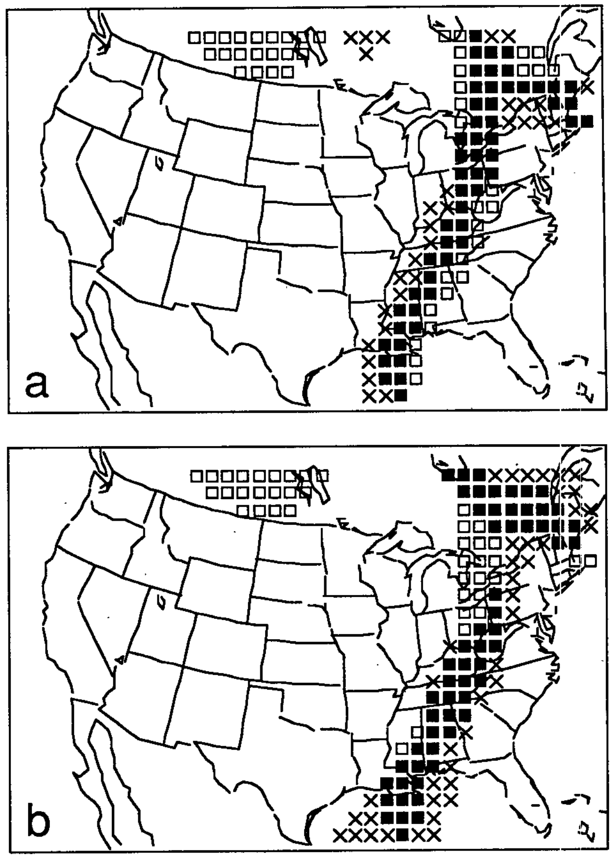

FIG. 4. Similar to Fig. 3 but for case 4.

the individual and objective analyses made for cases 2, 4, and 5.

e. Discussion of frontal analyses

The performance measures for IREW and the individual meteorologist varied considerably for the cases tested. The cumulative values for IREW and the individual meteorologist though, were almost the same.

Another encouraging result is that IREW's analyses and the individual analyses contained some of the same mistakes. For instance, in case 2 both analyses missed the front through Iowa and Illinois, and in case 5 both analyses showed a front passing over Lake Erie. One possible cause for the similar mistakes is that there were features in the data that misled both the individual meteorologist and IREW. Another possibility is that the consensus analyses were wrong.

Figure 6 shows the fronts that were analyzed by NMC in the observations that corresponded to cases 2, 4, and 5. For case 2, no analysis correctly predicted the location of the front through the Midwest. This suggests that the NGM output for case 2 was wrong or did not provide clear frontal signatures. In case 5 though, the front was placed over Lake Erie. This suggests that the consensus analysis for case 5 was wrong.

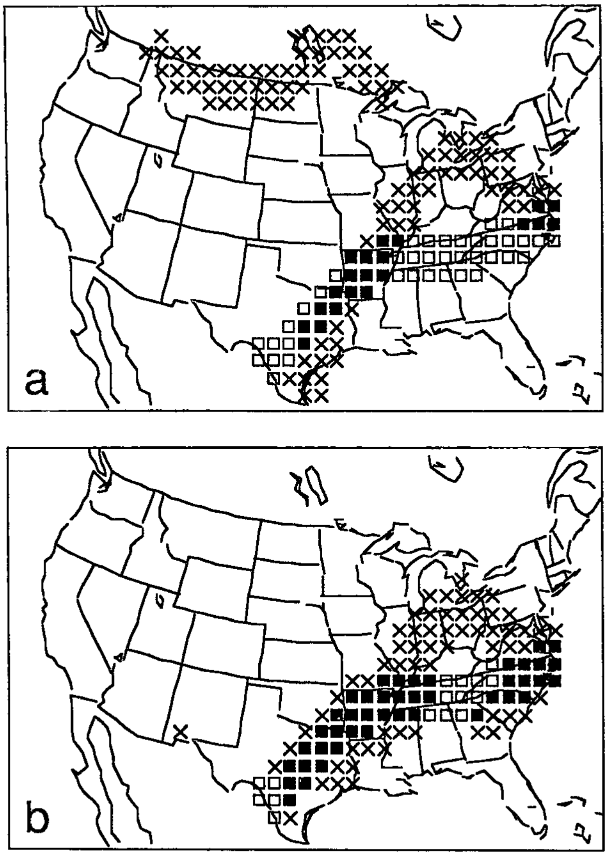

FIG. 5. Similar to Fig. 3 but for case 5.

Table 3 shows how the individual analyses, IREW's analyses, and the consensus analyses compared to NMC's analyses of observations. NMC's analyses were treated by the same process that was applied to the other subjective analyses. The scores for IREW's analyses were intermediate between the scores for the individual analyses and the scores for the consensus analyses. Table 4 contains contingency tables for the comparisons to the observed frontal locations. Model errors probably accounted for some of the differences between the forecast fronts and the observed fronts. Also, the subjective analyses of forecast fields were intended for a nontechnical audience. It is possible that, in meeting the needs of a nontechnical audience, the local meteorologists used a different definition of a front than was used by NMC's analysts.

The comparison of the analyses for the seven cases does show that IREW can make skillful frontal analyses. More extensive tests with more cases and different measures of performance would be required before a claim could be made that the objective analyses were as accurate as the subjective analyses.

The frontal odds produced by the characterizations overestimated the probability that a point was in a frontal zone. Only 26% of the points with a frontal odds of 1 were in a frontal zone. The percentage for an unbiased estimate would be near 50. IREW's bias could have several causes, such as the use of dependent analytic characterizations and the choice of skill score. Dependent characterizations exaggerate the influence of a predictor. Using the same characterization twice is an extreme example. The information that is being used for the analysis would not change, but the information's effect on the odds would be squared. The exaggerated influence would increase the odds of points that exhibited some frontal characteristics. The skill score contributed to IREW's bias by rewarding biased analyses, which would make the selection of a biased set of analytic characterizations more likely.

The bias is not a serious problem when a threshold is used to make a yes/no decision. The odds threshold for classifying samples can be adjusted to set the bias to a level that is appropriate for different analysis tasks.

IREW's analyses may have matched the consensus analyses better if the meteorologists had examined only the NGM output when drawing the weather depiction charts. IREW's analyses could also have been processed to produce frontal zones that match the prototypical frontal zone better. Such processing might have included eliminating isolated frontal regions, thinning wide frontal zones, and linking strong frontal areas that were separated by regions with weaker frontal characteristics. Post-processing was not performed because the purpose of the frontal identification work was to test IREW's effectiveness, not to produce operational frontal analyses.

7. Discussion

The frontal recognition work shows that IREW has the capability to recognize some patterns. The discussion of IREW's algorithms shows that the technique has several other characteristics that are desirable for meteorological applications.

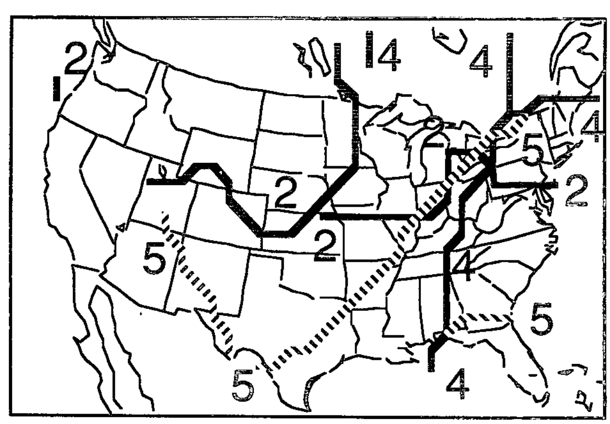

FIG. 6. Fronts analyzed in the observations corresponding to cases 2, 4, and 5. Each front is identified by a case number. The fronts for case 5 are dashed.

The algorithms for learning odds factors and recognizing relevant characterizations require computation time proportional to the product of the number of samples and the number of characterizations. The algorithm to select interesting characterizations could require time proportional to the square of the number of relevant rules; although if a bound is placed on the number of interesting rules then the time required is proportional to the number of relevant rules. Since IREW learns incrementally, the amount of storage space required is proportional to the number of characterizations, except when IREW is selecting analytic characterizations. The selection of the analytic rules requires more time and space as a function of the number of characterizations considered than the other steps, but the last step is applied to only a small subset of characterizations and samples. IREW's relatively small time and memory requirements allow IREW to examine many characterizations and large datasets faster and in less space than techniques—such as decision trees, regression techniques, and some clustering algorithms—that require resources proportional to the second or third power of the number of predictors or the number of samples.

Like regression techniques, IREW computes weights for predictors so the user can gain some insight into which of the specified predictors are related to the target feature. This is illustrated by IREW's selection of many of the temperature, moisture, and wind patterns at the surface and aloft that meteorologists associate with surface fronts. Another advantage IREW provides is that no extra work is required to adapt IREW to handle missing data.

IREW's disadvantages relative to other techniques include an inability to combine predictors and the absence of off-the-shelf software packages to implement the technique. IREW could be extended to combine predictors (perhaps producing a technique similar to STAGGER), but this would probably increase IREW's resource requirements.

TABLE 3. Performance of the individual analyst, IREW, and the group of meteorologists for the seven test cases when analyses of observations were used as the verification.

		Frontal points correctly identified (%)	Points correctly classified (%)	Bias	Gringorten skill score
Individual analyses	Max	74	90	1.30	0.75
	Min	27	78	0.63	0.25
	All 7	50	84	1.00	0.48
IREW's analyses	Max	76	91	1.80	0.66
	Min	25	80	0.53	0.23
	All 7	48	84	0.95	0.47
Consensus analyses	Max	59	87	1.30	0.72
	Min	27	80	0.56	0.24
	All 7	47	84	0.92	0.45

TABLE 4. Contingency tables for the individual analyses, IREW, and the consensus analyses when the analyses of observations were used as the verification.

		Analyses of observations: Frontal points	Nonfrontal points	Total
Individual analyses	Frontal	362	375	737
	Nonfrontal	369	3451	3820
	Total	731	3826	4557
IREW's analyses	Frontal	353	343	696
	Nonfrontal	378	3483	3861
	Total	731	3826	4557
Consensus analyses	Frontal	340	334	674
	Nonfrontal	391	3492	3883
	Total	731	3826	4557

8. Conclusions

As the volume of meteorological data continues to grow, pattern recognition techniques will become increasingly important data analysis tools. IREW is presented as another option for meteorologists that need such tools.

IREW may be a useful pattern recognition option because IREW has different characteristics than most of the pattern recognition techniques used by meteorologists. Most significantly, IREW sacrifices the consideration of interactions between predictors to reduce the resources required to learn to recognize a pattern. The effect of the sacrifice on IREW's ability to recognize patterns is an open question. Results so far are encouraging.

Many modifications could be made to IREW to try to improve its performance. For instance, the effect of using probabilistic instead of categorical skill scores to aid the selection of analytic rules could be examined. Also, the algorithm for selecting analytic rules could be changed.

The frontal identification work could be extended by selecting characterizations for every combination of frontal type, season, and region of the United States. Better analyses of model output might be produced if analyses of observations were used as the verification for learning cases.

Work is currently underway to apply IREW to analyzing profiler wind data. We hope that this work will lead to improved processing of profiler data, will help us understand better the types of problems for which IREW is best suited, and will suggest modifications that will improve IREW's performance.

Acknowledgment. Joe Lundberg and the Penn State Weather Communications Group graciously provided information for the frontal analysis work. George Young, Bill Moninger, and anonymous reviewers provided comments that contributed to a better description of IREW. Partial support for this project was supplied by NSF Grants ATM-8607577 and ATM-8917596.

APPENDIX

Algorithms

a. Algorithm to determine if a characterization is related to the target feature

```
/*
 * calculate the number of points examined by the
 * characterization
 */
```

TOTAL_POINTS = $F_t + F_f$;
If TOTAL_POINTS < 1000 then return NO_DECISION;

```
/*
 * Enough points have been tested to consider making
 * a decision
 */

/*
 * Calculate the number of points where the charac-
 * terization's test was true and where the test was
 * false
 */
```

$T_t = (T_t | F_t) + (T_t | F_f)$;
T_f = TOTAL_POINTS $- T_t$;

If ($T_t = 0$ or $T_f = 0$) and TOTAL_POINTS > 2000 then return INDEPENDENT;

```
/*
 * Compute the expected frequencies of the four pos-
 * sible outcomes
 */
```

$E\{T_t, F_t\} = T_t * F_t$ / TOTAL_POINTS;
$E\{T_f, F_t\} = T_f * F_t$ / TOTAL_POINTS;
$E\{T_t, F_f\} = T_t * F_f$ / TOTAL_POINTS;
$E\{T_f, F_f\} = T_f * F_f$ / TOTAL_POINTS;

```
/*
 * the statistical test is most accurate when each ex-
 pected frequency is above a minimum value
 */
```

If $\min(E\{T_t, F_t\}, E\{T_f, F_t\}, E\{T_t, F_f\}, E\{T_f, F_f\})$ < 15 then return NO_DECISION;

$$\text{CHI\_SQUARE} = (T_t | F_t - E\{T_t, F_t\})^2 / E\{T_t, F_t\} + (T_t | F_f - E\{T_t, F_f\})^2 / E\{T_t, F_f\} + [(F_t - T_tF_t) - E\{T_f, F_t\}]^2 / E\{T_f, F_t\} + [(F_f - T_tF_f) - E\{T_f, F_f\}]^2 / E\{T_f, F_f\};$$

If CHI_SQUARE > 6.635 then return DEPENDENT; else return INDEPENDENT;

b. Algorithm to select a set of characterizations suitable for making analyses:

Let INTERESTING_CHARACTERIZATIONS = the interesting characterizations;

Let ANALYTIC_CHARACTERIZATIONS = an empty list of characterizations;

Let CASES = the cases available for the final selection;

Let CASE_CHARACTERIZATIONS = an array of empty lists with one list for each case;

Repeat until none of the cases change ANALYTIC_CHARACTERIZATIONS;

  If this is the first iteration then CASE = the first case;
  else if CASE = the last case then CASE = the first case;
  else CASE = the case in CASES that follows CASE;
  Choose the characterization in INTERESTING_CHARACTERIZATIONS
    that when applied to CASE has both effects > the effect cutoff for the interesting characterizations,
    that maximizes the measure of performance when applied with CASE_CHARACTERIZATIONS (CASE), and
    that has a squared multiple correlation with ANALYTIC_CHARACTERIZATIONS over CASES < 0.5;

  If a characterization was chosen, put it into CASE_CHARACTERIZATIONS and into ANALYTIC_ CHARACTERIZATIONS;

  End repeat;

Return ANALYTIC_CHARACTERIZATIONS;

REFERENCES

Anderberg, M. R., 1973: *Cluster Analysis for Applications.* Academic Press, 359 pp.

Brieman, L., J. H. Friedman, R. A. Olshen and C. J. Stone, 1984: *Classification and Regression Trees.* Wadsworth International Group, 358 pp.

Campbell, S. D., and S. H. Olson, 1987: Recognizing low-altitude wind shear hazards from Doppler weather radar: an artificial intelligence approach. *J. Atmos. Oceanic Technol.,* **4,** 5–18.

Carroll, B., 1987: Artificial intelligence, expert systems for clinical diagnosis: Are they worth the effort? *Behav. Sci.,* **32,** 274–292.

Carter, G. M., J. P. Dallavalle and H. R. Glahn, 1989: Statistical forecasts based on the National Meteorological Center's numerical weather prediction system. *Wea. Forecasting,* **4,** 401–412.

Conway, B. J., 1989: Expert systems and weather forecasting. *Meteor. Mag.,* **118,** 23–30.

Dawes, R. M., D. Faust and P. E. Meehl, 1989: Clinical versus actuarial judgment. *Science,* **243,** 1668–1674.

Dockus, D. A., 1988: DDT II: Computerized lake-effect snow forecasts. *Nat. Wea. Dig.,* **13,** 18–26.

Duda, R. O., and P. E. Hart, 1973: *Pattern Classification and Scene Analysis.* John Wiley & Sons, 482 pp.

——, —— and N. J. Nilsson, 1976: Subjective Bayesian methods for rule-based inference systems, *Proc. of the National Computer Conference,* AFIPS, 1075–1082.

Elio, R., J. de Haan and G. S. Strong, 1987: METEOR: an artificial intelligence system for convective storm forecasting. *J. Atmos. Oceanic Technol.,* **4** 19–28.

Elmore, K. L., F. W. Wilson, Jr. and M. J. Carpenter, 1989: Recovering "lost" data: an adaptive, least-squares outlier detection technique. *J. Atmos. Oceanic Technol.,* **6,** 537–541.

Fahlman, S. E., and G. E. Hinton, 1987: Connectionist architectures for artificial intelligence. *Computer,* **20,** 100–109.

Gaschnig, J., 1982: Prospector: an expert system for mineral exploration. *Introductory readings in expert systems.* D. Michie, Ed., Gordon and Breach, 47–64.

Gringorten, I. I., 1967: Verification to determine and measure forecasting skill. *J. Appl. Meteor.,* **6,** 742–747.

Hayes-Roth, F., D. A. Waterman and D. B. Lenat, 1983: *Building Expert Systems.* Addison-Wesley, 444 pp.

Hoffman, R. R., 1987: The problem of extracting the knowledge of experts from the perspective of experimental psychology. *AI Applications in Natural Resource Management,* **1,** 35–48.

Michie, D., S. Muggleton, C. Riese and S. Zubrick, 1984: RuleMaster: a second-generation knowledge-engineering facility. *Proc. of the First Conf. on Artificial Intelligence Applications,* The Institute of Electrical and Electronic Engineers, 591–597.

Moninger, W. R., 1988: ARCHER: a prototype expert system for identifying some meteorological phenomena. *J. Atmos. Oceanic Technol.,* **5,** 144–148.

Murphy, A. H., and H. Daan, 1985: Forecast evaluation. *Probability, Statistics, and Decision Making in the Atmospheric Sciences.* A. H. Murphy and R. W. Katz, Eds., Westview Press, 379–437.

Neapolitan, R. E., 1987: A note of caution on combining certainties. *Int. J. Pattern Recognition and Artificial Intelligence,* **1,** 427–433.

Peak, J. E., and P. M. Tag, 1989: An expert system approach for prediction of maritime visibility obscuration. *Mon. Wea. Rev.,* **117,** 2641–2653.

Quinlan, J. R., 1986: Induction of decision trees. *Mach. Learning,* **1,** 81–106.

——, 1987: Decision trees as probabilistic classifiers. *Proc. of the Fourth International Workshop on Machine Learning,* Irvine, California, Morgan Kaufmann, 31–37.

Renard, R. J., and L. C. Clarke, 1965: Experiments in numerical frontal analysis. *Mon. Wea. Rev.,* **93,** 547–556.

Schlimmer, J. C., and R. H. Granger, Jr., 1986: Incremental learning from noisy data. *Mach. Learning,* **1,** 317–354.

Stewart, T. R., W. R. Moninger, J. Grassia, R. H. Brady and F. H. Merrem, 1989: Analysis of expert judgment in a hail forecasting experiment. *Wea. Forecasting,* **4,** 24–34.

Walpole, R. E., and R. H. Myers, 1978: *Probability and Statistics for Engineers and Scientists.* 2nd ed. MacMillan Publishing Company, 580 pp.

Wilkins, D. C., W. J. Clancey and B. G. Buchanan, 1986: Overview of the Odysseus learning apprentice. *Machine Learning: A Guide to Current Research,* T. M. Mitchell, J. G. Carbonell and R. S. Michalski, Eds., Kluwer Academic Publishers, 369–373.

Wu, R., J. A. Weinman and R. T. Chin, 1985: Determination of rainfall rates from GOES satellite images by a pattern recognition technique. *J. Atmos. Oceanic Technol.,* **2,** 314–330.

Yarnal, B., 1985: A 500 mb synoptic climatology of Pacific Northwest coast winters in relation to climatic variability, 1948–1949 to 1977–1978. *J. Climatol.,* **5,** 237–252.