Classification of Radar Clutter in an Air Traffic Control Environment

SIMON HAYKIN, FELLOW, IEEE, WOLFGANG STEHWIEN, CONG DENG, PETER WEBER, AND RICHARD MANN

This paper may be viewed as the culmination of a radar clutter classification research program that has spanned over 15 years. In particular, we present the results of an experimental study aimed at the classification of radar clutter encountered on ground-based coherent scanning radar systems used for air traffic control. The clutter signals of interest are primarily those due to birds as well as clouds and weather systems. These two sources of radar returns represent potential hazards to aircraft safety. The aim of the clutter classifier is therefore to vector aircraft around such areas.

The clutter classification presented herein is based on a set of features derived from a sequence of reflection coefficients computed using Burg's multisegment algorithm. These features contain two types of information, namely, signal strength, and Doppler.

Two feature classifiers were evaluated experimentally:

1. *A parametric Bayes classifier, assuming that the features are distributed according to the multivariate Gaussian distribution.*
2. *A neural network classifier that makes no such assumptions. Training of the neural network classifier is achieved using the popular back-propagation algorithm.*

The classification is performed using real-life radar data contained in a single resolution cell and collected during the course of a single antenna scan.

I. Introduction

The two most common aviation hazards to aircraft are weather and birds. The most dangerous forms of weather are heavy precipitation (in which airframe icing frequently occurs) and wind-shear related phenomena (including thunderstorms). Aircraft collisions with birds can also result in serious damage, crashes, and fatalities. Indeed, periods of heavy night-time migration during the spring and fall (at airports located on bird-migration corridors) present a major problem for night flying jet training operations and may require a complete halt to such activities. Traditional radar detection schemes consider all nonaircraft targets as undesirable "clutter," which must be suppressed to enhance the visibility of aircraft within the clutter areas. The *moving target indicator* (MTI), used for such a purpose, is essentially a Doppler filter designed to remove all nonmoving echoes such as ground clutter. Fast-moving rain and storm systems, however, break through the filter and appear on the processed displays as targets, often overloading the digital target processing units. The so-called *moving target detector* (MTD) consists of a bank of Doppler filters, each tuned to a different frequency. Separation of targets and clutter, which generally travel at different velocities, can thus be achieved and the probability of detection increased considerably.

In this paper, we take a different approach in that an attempt is made to classify the different forms of radar signals, rather than merely detect moving targets. It can be said that detection is simply classification with two classes, and thus this work is a generalization of the target detection problem.

A. The Radar Clutter Classification Problem

A generic block diagram of the classifier is shown in Fig. 1. The measurements available from the surveillance radar are the amplitude and phase of the received echo pulses. Pulses are transmitted at regular intervals, and even though the radar antenna is continuously rotating, the transmitted beam is wide enough to provide several echoes from the same source before the beam has scanned past it. This time series of echoes thus contains information on the size and velocity of the scatterers, as well as their internal motion. Since not all of this information is unique to any one clutter type, those features that can reliably identify the clutter need to be *extracted* and passed on to the classifier.

A central question for the classification of clutter concerns the *selection* of features, that is, which features (if

Manuscript received May 18, 1990; revised November 26, 1990. This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Canadian DND at different stages of the program.

S. Haykin and P. Weber are with the Communications Research Laboratory, McMaster University, Hamilton, Ont., Canada L8S 4K1.

W. Stehwien is with Litton Systems Canada Limited 404-1-C, Etobicoke, Ont., Canada M9W 5A7.

C. Deng is with Bell-Northern Research Ltd. Station C, Ottawa, Ont., Canada K1Y 4H7.

R. Mann is with the Laboratory of Computer and Information Science, Helsinki University of Technology, Rakentajanaukio 2 C, SF-02150 Espoo Finland.

IEEE Log Number 9100177.

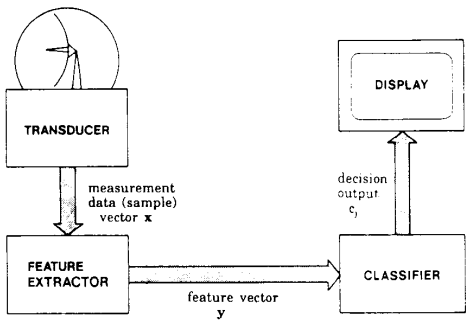

Fig. 1. Block diagram of the classifier.

any) are present in coherent radar returns and able to accurately identify each clutter type. The amplitude of the returns is related to the radar cross-section (RCS) of the target and, in the case of distributed clutter, the radar reflectivity per unit volume. It is, however, also a function of the angular position of the target within the antenna beam pattern, which cannot be determined without knowledge of the altitude of the target. This knowledge is generally not available in surveillance radars and as a result, no definitive statement can be made about the target size, regardless of the amplitude of the returns.

The phase difference between successive returns (or, more precisely, the rate of change of the phase or Doppler frequency) is a measure of the radial velocity (or range rate) of the clutter or target. As such, it does not reveal much about the clutter either, except in the case of ground clutter which is never expected to deviate much from zero velocity. Finding zero radial velocity does not, however, prove that the clutter is ground-based; it simply makes it more likely. Targets moving tangentially to the radar will also show a zero range rate. In addition, a phase shift of integer multiples of 2π radians between successive pulses will appear as zero Doppler frequency and hence zero range rate. Such aliasing will further reduce any uniqueness that might have been contained in the radial velocity, especially for low pulse repetition frequencies (PRF).

Besides radial velocity, computing the spectrum of the sample series will also reveal the "purity" of the velocity. The spectral spread is an indication of the presence of sidebands to the Doppler frequency itself. There are a number of causes for spectral spread. First, there is the modulation of the time series by the azimuth gain pattern of the main scanning beam, which is usually approximately Gaussian or $\sin(x)/x$. A reduction in the number of samples available from the target would be the result of a reduced number of target hits per beamwidth, which in turn causes a widening of the spectral spread. Another cause is amplitude modulation resulting from pulse-to-pulse scintillation, which results from rapid changes in the observed radar cross-section or reflectivity (target glint). Internal motion, or instantaneous deviation from the average velocity will manifest itself as frequency modulation with the potential of large spectral spreads. The presence of multiple scatterers and hence multiple frequencies leads to both frequency and amplitude modulation. This is particularly true for windshear conditions in weather systems, and large numbers of birds within a radar resolution cell moving in different directions and at different speeds. Such internal motion has most certainly different characteristics for the various clutter types and should provide the basis for any feature set, provided of course that it is not completely obscured by the scanning modulation.

Given the feature vector $\boldsymbol{y}$, the classifier must then decide which radar clutter class c_j the signal vector $\boldsymbol{x}$ belongs to. The traditional method for the design of the classifier is to use a Bayes formulation of the classification problem. The solution so obtained usually assumes a multivariate Gaussian distribution [1]. Alternatively, we may use a neural network classifier, the design of which is derived without any such assumption. Accordingly, a neural network classifier may outperform a Bayes classifier for non-Gaussian features. Both classifier types are considered in the paper.

The body of the paper is organized as follows. Section II presents a historical perspective of the radar clutter classification problem and related issues. In Section III we describe the important features of radar as a sensor in an air traffic control environment. This is followed by a discussion of physical phenomena in radar clutter and targets in Section IV, with emphasis on the spectral contents of radar returns. That material provides the physical basis for the discrimination between the different radar clutter classes. Section V highlights several issues relating to the collection of radar data used in subsequent experimental studies. In Section VI, we describe the feature selection/extraction procedure, with emphasis on the multisegment Burg algorithm. In Section VII we present the performance evaluation of a classical Bayes classifier. For the neural network approach, the popular back-propagation algorithm is used. The performance evaluation for this nonlinear classifier is presented in Section VIII. The paper concludes with a discussion of results in Section IX. Two appendixes are included at the end of the paper, one summarizing the specifications of the radar used for data collection, and the other dealing with experimental evidence for the presence of bird clutter.

II. Historical Perspective

This section presents a brief review of the developments leading up to the current state of research into radar clutter classification, particularly as it concerns air traffic surveillance radars. The presentation is essentially chronological although, in the interest of cohesiveness, independent disciplines are discussed under their own section headings. The history of radar itself is not addressed here. Skolnik [2] has prepared a brief history and discussion of the current state of radar.

A. The Desire to Classify Radar Targets and Clutter

The desire to classify dates back to the earliest days of radar experimentation, when it was found that "spurious"

or clutter echoes from airborne scatterers such as clouds could be sufficiently strong to obscure the real targets of interest: aircraft. There was also the war-time need for the identification of friendly and enemy aircraft and, to a lesser extent, the need to identify the type of aircraft. A solution to the latter problem was quickly found by installing transponders on board aircraft as part of a system called *Identification, Friend or Foe* (IFF). This system is still in use today and has found application in air traffic control (ATC) systems as a secondary surveillance radar (SSR) or, more formally, the *Air Traffic Control Radar Beacon Systems* (ATCRBS) [3]. While this system provided aircraft visibility in all kinds of clutter (and still does), it did nothing for the identification of clutter or aircraft without transponders, and radar operators were still required to use their judgment and experience with unprocessed radar data displays [4].

After the Second World War, when radars became more powerful and used higher and higher frequencies, clutter problems also became greater and various clutter-suppression techniques, including MTI were devised. At the same time, researchers devoted more time to the study of "angels," that is, unknown and unidentifiable radar echoes. Many meteorological conditions were relatively easy to correlate with radar observations, but the realization that many of these "angels" were birds and even insects, came more slowly [4], [5]. With that realization also came an interest by ornithologists to study birds using radar, which grew throughout the 1960's and continued into the early 1970's [6]. Aviators also became interested in separating birds from aircraft, particularly with the advent of high-speed jet aircraft and the phenomenal increase in air traffic [7]. The use of radar to help avoid bird strikes seemed logical, especially since most busy airports use radar for air traffic control [8].

Most of the clutter identification techniques in current use are based on visual examination of PPI displays by trained radar operators. If the unprocessed radar video is available, an "educated guess" may be made with a reasonable probability of success. More reliable are photographic records, either in the form of time-lapse motion pictures or single time-exposures, where bird tracks become readily apparent. Unfortunately, the time delay involved in processing the photographs limits their use for bird-strike avoidance. They can, however, be very useful for ornithologists in migration studies, and for migration forecast used in the planning of flight operations [9], [10].

Bird-strike avoidance programs will undoubtedly benefit from migration studies and forecasts, but immediate information will always be required. While acute problems at airports can be managed, to some extent, with visual observations and bird-scaring techniques, en-route avoidance would be much more successful if up-to-date radar warnings can be issued [8]. Unfortunately, modern target detection and clutter suppression techniques remove, if not all of the clutter, much of the identifying information and make manual clutter classification quite unreliable. Synthetic radar displays are designed to eliminate clutter and leave the radar operator with little or no data on airborne hazards, apart from what is automatically detected and displayed with special symbols. Research into automatic clutter classification thus arose from this need to present airborne hazard information to the radar operator in real time.

B. Automatic Classification Attempts Using Noncoherent Radar Data

Most early radars provided only signal strength (signal amplitude) with which a detection decision could be made. Resolution was usually too low to form true "images" of the target, and the only information that could be obtained about the target was a lower bound on its maximum length [3]. The actual value was clearly a function of orientation and only broad categories of targets could be identified. Even today, the resolution of surveillance radars is low enough that practically no information about target size is available. What is available, however, are signal amplitude and its spatial distribution, amplitude modulation, polarization dependence, scan-to-scan motion, and fluctuation characteristics. While probably insufficient for the identification of targets, these parameters provide much information about clutter.

Much work has been done to find amplitude signatures of bird echoes and several good reviews have been published on this subject [5], [6], [11]. These studies found wingbeat modulation frequencies of a few Hertz to tens of Hertz, which have been observed over periods of several seconds. The difficulty with a scanning surveillance radar, however, is that the target dwell time (the time the beam spends directed at the target) is quite short and often insufficient for the formation of a reliable estimate of the signature. With dwell times measured in milliseconds and scan times of several seconds, surveillance radars will interpret such signatures as scan-to-scan fluctuations with little possibility for a positive identification. Flock and Green [5] recognized this problem and proposed a combination of surveillance and signature analysis radar, the latter using pulse-Doppler techniques.

Much less work has been done on the spatial distribution and polarization properties of bird clutter. Spatial amplitude distributions have been examimned by Barry *et al.* [12]; and Dill and Major [13] actually studied the distributions of bird flocks in three dimensions. Some references to the work on polarization properties were provided by Vaughn [6].

Meteorological clutter has long been the subject of study and much is known about the characteristics or rain and storm clutter (see Nathanson and Reilly [14]; Smith *et al.*, [15]; Doviak and Zrnic [16]). Long-term pulse-to-pulse correlation is much greater than it is for bird clutter [17], as is the spatial uniformity in amplitude distribution and area movement. In the context of this statement, "long-term" refers to periods greater than one dwell period, which in our case is somewhere around 25 ms [17]. Many modern radars make use of a weather channel to derive weather contours for presentation to the operator, and attempt to filter weather clutter out completely from the target detection channel.

These contours are simply derived from amplitude thresholding [18]. Weather radars represent a further evolution in the desire to learn more about meteorology, and are used extensively as an aid in forecasting and to provide weather reports to the aviation community. These radars derive rain intensities from calibrated reflectivity measurements which are then plotted in color on PPI displays.

Ground clutter is, perhaps, the most complex of all commonly occurring clutter types. Its lack of homogeneity makes it difficult to characterize, although many attempts have been made to do so [19]. Even today, ground clutter characterization cannot be considered complete [20]. Air traffic control radars generally combat clutter through the use of MTI and MTD techniques and ground clutter maps [18], [21], [22].

Some attempts to classify clutter based on this knowledge have been made, but they demonstrated only marginal success. Perhaps one of the simplest techniques to automatically detect the presence of birds is to determine the spatial density of targets and to relate this density to the various clutter types. Hunt [23] describes automatic equipment that computes this density by counting targets, and relates it to the probability of a bird strike. Unfortunately, since this device counts all detections and not just birds, it would likely become rather unreliable when weather clutter is encountered.

The desire to add as little new equipment as possible to existing surveillance radars lead to several studies into the spatial amplitude characteristics of clutter at the *Communications Research Laboratory* (CRL) at McMaster University. Haykin and Carter [24] reviewed the then current literature and theory, and proposed a processor to measure the distribution of amplitudes and Doppler spreads. Their report also contains an appendix describing the work on weather and bird clutter carried out by other Canadian and United States agencies. Currie *et al.* [25] continued this work by testing the processor algorithms using recorded radar data. They tested amplitude histograms, adjacent sample variabilities, and clutter area movement (scan-to-scan). The most promising of the three appeared to be the adjacent sample variability measure, although the authors state that the area movement algorithm may be of some use. Amplitude histograms of bird and weather clutter were found to be too similar to be of use. It appears that this work has not been continued beyond this point, with the emphasis shifting to the study of Doppler spectra.

C. Evolution of the Use of Spectral Parameter

The principles of Doppler radar have been known since the 1930's, but it did not see widespread exploitation until the development of digital technology in the 1960's (Skolnik, [2]). Spectral analysis is the Fourier transformed version of correlation analysis and thus can be carried out on noncoherent radar data to find spatial and temporal correlation parameters. Temporal correlation will clearly detect the wingbeat characteristics of bird clutter and could serve well to distinguish it from weather clutter and aircraft targets. Due to their short dwell times, however, surveillance radars cannot effectively measure such temporal correlation, and any "discovered" spatial correlation must be corrected for the effects of the scanning antenna beam.

Analysis of the Doppler shift, which arises from the motion of the target, had the promise of providing additional information for discrimination. Haykin and Carter [24] proposed, and Currie *et al.* [25] tested the viability of using the spread of the Doppler spectrum as a discriminant. Chan and Haykin [26] compared the performance of the *Fast Fourier Transform* (FFT) and the *Maximum Entropy Method* (MEM) as spectral estimators of the clutter spectrum, given that only a few samples were available. Kesler [27] extended this work and concluded that, for clutter which can be modeled as an autoregressive process, the MEM is a superior estimator than FFT-based and other linear methods. Using recorded radar data, it was found that significant differences appear to exist in the spectral spreads of birds, weather, and ground clutter [27], [28]. Currie and Haykin [29]; Haykin *et al.* [30] actually constructed a working "classifier" prototype which computed "turbulence indexes" to warn of the dangers of the airborne hazards. These indexes were then related to individual clutter types for identification.

The CRL (McMaster) work did not, however, relate the observed characteristics to any of the physical phenomena at work in the scatterers that cause the clutter; once the theory was developed, it proceeded largely on an empirical basis. On the other hand, studies which correlated the physical motion of clutter with their spectra used special-purpose radars. Bird clutter analyses were mostly done using nonscanning radars that allowed continuous measurements for several seconds [5], [6]. From these analyses, researchers were able to define many of the amplitude and Doppler signatures of individual species of birds. The amplitude modulation introduced by the scanning beam modulation, though acknowledged, was not generally discussed.

Weather radars supplied much of the knowledge on weather clutter [14], [15]. These radars do not, in general, use fan beams such as those used by air traffic surveillance radars. Pencil beams are used to enable a more detailed breakdown of the scanned volume. As a consequence, resolution cells are not columns and their smaller vertical extent increases the homogeneity of the measured characteristics. These characteristics include, in addition to reflectivity, Doppler, Doppler velocity profiles and gradients with height and range, and Doppler spectral spread to detect turbulence [16].

D. The Application of Pattern Recognition Techniques

The work discussed up to this point has dealt essentially with finding characteristics of clutter, on the basis of which several categories could be distinguished. No actual classification had been carried out; in fact, the term *classification* was often used rather loosely where the term *feature selection* would have been more appropriate. Several excellent books on pattern recognition techniques were published in the early 1970's [1], [31], [32]. Comprehensive summary

papers on the state of the research at that time were written by Ho and Agrawala [33] and Kanal [34]. Most of this work is still valid; and more recent research results have been compiled by Fu [35] and Jain [36].

In regard to radar, Skolnik [3] presents several techniques for automatic classification of radar targets, which is cited as an example of M-ary detection. Two candidates are given as examples: filtering (a matched filter and detector per class) and cross-correlation (similar to template matching). The difference between these two classifiers is in fact only one of practicality; fundamentally, they are substantially the same and belong to the category of *linear machines* (see Duda and Hart [1]). Barry *et al.* [12] used such a linear machine in an attempt to separate "angels" from aircraft targets. They applied as many as four hyperplanes to feature vectors consisting of 20 successive amplitude samples, amplitude distribution statistics and correlation measures. They claimed success rates of up to 95% angel rejection.

Most work reported in the literature on *statistical* pattern recognition takes a more direct approach based on Bayes decision theory. As a result, the decision rules tend to be much more complex than those described in [3], but also much more effective. While this work is now considered relatively mature [36], it has not been applied broadly to radar signal analysis. Nonradar applications appear to have received considerably more attention. Electrocardiogram (ECG) waveforms have frequently been used as examples of time series classification [32], and "optimal" rules have been devised for speaker waveform recognition [37]. Chen [38] used the nearest neighbor algorithm to classify seismic waveforms based on parameters derived directly from *maximum-entropy* spectra.

Work with radar clutter is much more recent. Agnel [39], [40] tested the effectiveness of the parametric Bayes classifier on three classes of recorded noncoherent ground clutter data (plowed fields, trees without leaves, and snow-covered ground), and achieved up to 90% correct classification. He also compared the performances of the so-called *minimum intraclass distance* (MICD) and *nearest neighbor* classifiers, but concluded that they were not satisfactory. Several feature-vectors were used, but most of them were derived from autoregressive analysis; that is, they ostensively measured the spatial correlation of the clutter. Stehwien [41], [42], and Stehwien and Haykin [43], [44], appear to have been the first to apply the Bayes classifier to parameters arising out of the estimation of the Doppler spectrum.

A final mention is in order, for the sake of completeness, of the application of *syntactic* pattern recognition techniques to the analysis of radar clutter. These techniques approach the problem from the perspective of image or scene analysis and attempt to describe the spatial properties of the clutter along with scan-to-scan changes in shape and position. Algorithm inputs are usually in the form of processed and scan-converted data. As such, syntactic techniques are ideal companions to follow statistical classifiers. Duda and Blackmer [46] applied such techniques to weather radar data to trace echo contours and track weather systems. Blackmer *et al.* [47] extended the work and used the echo descriptions to forecast weather system movements.

III. The Radar As A Sensor

A surveillance radar consists of a high power pulse transmitter, a rotating antenna, and a receiver (Skolnik [3], [48]). The antenna focuses the pulse energy into a beam of some finite width, and scans the surrounding volume of space with it. Any reflector within this volume will return some energy back to the antenna. The number of pulse echoes received by the radar is a function of the beamwidth, the PRF, and the rate of antenna rotation, given by

$$N_\theta = f_s t_d = f_s \frac{\theta}{\omega_a} \tag{1}$$

where N_θ is the number of pulses, f_s is the PRF in Hz, t_d is the dwell time on the target in seconds, θ is the antenna beamwidth in radians, and ω_a is the antenna rotational velocity in radians/second. If the receiver operates coherently with the transmitter, then both amplitude and phase of the echo pulses can be measured. The phase is related to the distance, hence a change of phase from pulse to pulse (Doppler frequency shift) implies a radial velocity. The phase ϕ, measured at a distance x from the transmitting radar t seconds after transmission, is

$$\varphi(x,t) = \varphi(0,0) + \omega_c t - \omega_c \frac{x}{c} \tag{2}$$

where $\phi(0,0)$ is the starting phase of the pulse, ω_c is the transmitter carrier frequency, and c is the speed of light. The pulse frequency of a moving receiver is thus the rate of change of the phase

$$\frac{d\varphi}{dt} = \omega_c - \omega_c \frac{1}{c}\frac{dx}{dt} = \omega_c + \omega_d \tag{3}$$

which can be split into the original carrier frequency and the Doppler shift ω_d:

$$\omega_d = -\omega_c \frac{1}{c}\frac{dx}{dt} = -\omega_c \frac{2v}{c}. \tag{4}$$

This shift arises from the relative motion of the receiver or, in the case of a monostatic radar, the change in the round-trip distance $2\times$ due the velocity of the target v. Equation (4) can be written in terms of the wavelength λ and the Doppler frequency expressed in Hertz:

$$f_d = -f_c \frac{2v}{c} = -\frac{2v}{\lambda}. \tag{5}$$

The carrier frequency will have been removed in a coherent receiver, reducing (3) to $d\phi/dt = \omega_d$. A pulsed Doppler radar is inherently a sampled system and therefore

$$2\pi f_d = \omega_d = \frac{\Delta\varphi}{\Delta t} = f_s \Delta\varphi = f_s \phi \tag{6}$$

where $\phi = \Delta\varphi$ is the relative Doppler frequency. Combining (5) and (6) then provides the relationship between target velocity and the pulse-to-pulse phase change

$$v = -f_s \frac{\lambda}{2}\frac{\Delta\varphi}{2\pi} \tag{7}$$

where the velocity v is expressed with the same distance units as the wavelength λ. Clearly, the largest unambiguously measurable Doppler frequency is the Nyquist frequency of $\pm f_s/2$. Larger frequencies will be folded back in the spectrum onto lower requencies about the Nyquist rate (aliasing). To resolve the sign of ϕ, both in-phase and quadrature demodulation is necessary, resulting in complex-valued video data.

The coherent video signal is sampled at fixed intervals, whose length depends on the range resolution. To ensure a low probability of missing targets, the video signal is typically oversampled; that is, the sampling interval is smaller than the width of the transmitted pulse often less than half. The entire sampled data set then consists of a number of time series, one per range cell. The limited dwell time on any one target limits the number of samples that can be regarded as having originated from the same target to N_θ. Because of the oversampling, adjacent range samples will usually contain returns from the same target. The number of adjacent time series that may be used to extract information from any one target will depend on the length of the sampling interval relative to the pulse width.

IV. Targets, Clutter, and Their Spectra

Radar echoes arise from any reflecting surface or object within the line of sight of the radar. These objects include, besides aircraft, a variety of more broadly distributed features such as mountains, hillsides, trees and other vegetation, buildings, rain clouds, birds, and even insects. Only a few of these objects are truly static; even ground-based structures such as trees and towers sway and change their position continuously, albeit very slightly. This change in position leads to both the observed scintillation in net reflectivity and the measured relative velocity. The nature of these changes provides a clue to the type of target or clutter observed.

The discussions in this section are illustrated with MEM spectra of the respective clutter types. These spectra of model order 9 were computed using the multisegment Burg formula and the procedure outlined below in Section VI. To allow comparisons, similar signal strengths (30 dB signal-to-noise ratio (SNR)) were selected except where not available (rain clutter at 20 dB). Mean Doppler shifts were also removed.

A. Target and Clutter Types

1) Targets and moving vehicles: The irregular shape of aircraft (from the radar's point of view) can cause significant variation in RCS depending on the angle, or aspect, from which the aircraft is seen by the radar (Nathanson [49]; Skolnik [3]). Any maneuvering or vibrations will change the aspect, and thus the observed RCS. Turbulence in the air will modify the flight path and consequently modulate the velocity, particularly of light aircraft. Another and potentially more significant source of velocity and amplitude modulation are rotating propellers and jet engine compressor blades [3]. In general, however, the body of the aircraft is expected to be the dominant reflector in all but the nose and tail aspects. Any propeller or compressor blade modulation is thus likely to be found at low levels, especially for large jet aircraft which dominate today's air traffic environment. Figure 2 presents sample aircraft spectra, displaying some of these characteristics.

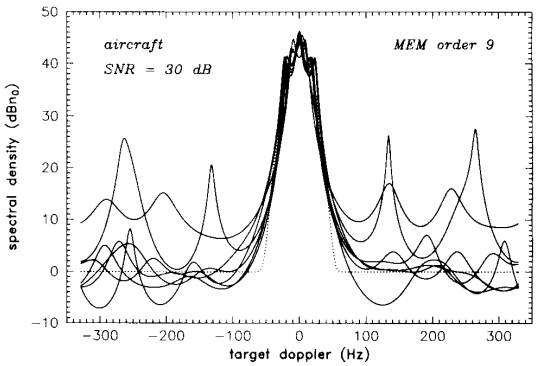

Fig. 2. Maximum entropy spectra of aircraft targets.

Ground-based moving vehicles are often mistaken by radar operators as light aircraft following a highway due to their similar speeds. However, there are a few significant differences between their radar returns. The RCS of cars is likely to be small compared to aircraft whose wings present a large reflecting surface. Trucks are therefore the most likely ground-based vehicles to be observed on a radar screen, with modulation present due to tire motion and vibration. Other vehicles such as tractors and construction machinery are usually very slow-moving and are unlikely to be mistaken for aircraft by radar operators. The absence of wings and the reduced maneuverability may result in lower RCS fluctuations. Neither cars nor trucks have propellers; thus less velocity modulation is expected. The proximity to the ground does, however, imply that if the vehicle is visible to the radar, so will be other nearby ground-based objects and, perhaps, the ground itself. This will introduce an additional spectral line at zero frequency. The spectra shown in Fig. 3 have been normalized to move the primary peak to zero frequency; hence, the ground clutter peak is elsewhere.

2) Ground clutter: This type of clutter consists of returns from mountains, hillsides, buildings, towers, power lines, and vegetation such as trees, bushes, and cultivated fields (particularly when wet). Building reflections are particularly annoying to radar operators since it tends to be the strongest of all clutter types and dominates in radars located near large cities. By its very nature, ground clutter is fixed in position and does not exhibit any overall velocity relative to the radar. Some internal motion is usually present, however, except in the case of low and massive buildings. Trees, transmitter towers, and tall buildings all tend to sway by varying amounts, depending on wind conditions. Higher frequency vibrations, depending on strength, may also be visible to the radar.

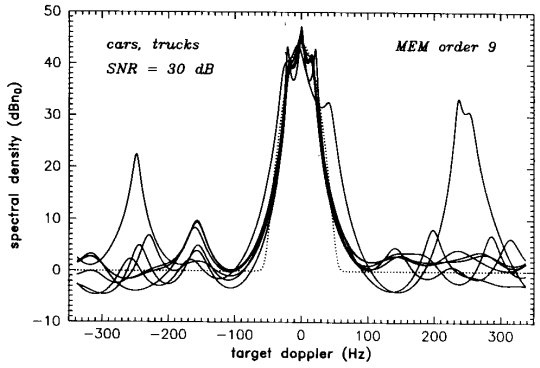

Fig. 3. Maximum entropy spectra of car and truck targets.

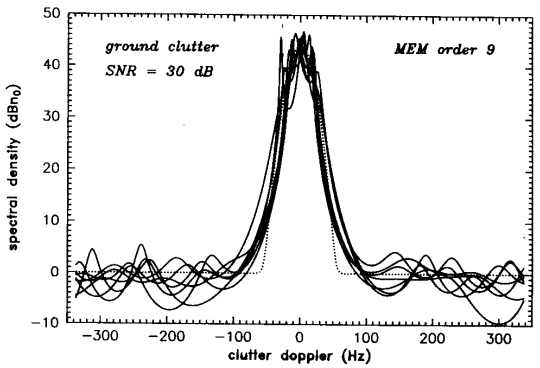

Fig. 4. Maximum entropy spectra of ground clutter.

Distributed features such as forests and cultivated fields consist of many individual "targets" or scatterers within any one radar resolution cell, and the effects of the relative motion of each sum coherently. This summation can either enhance or reduce the net return strength and thus the apparent reflectivity. Scintillation or amplitude modulation results from the constantly changing positions of the individual scatterers relative to each other.

The character of the clutter does change somewhat with range as the resolution cell of the radar widens and more scatterers are included within it. Long-range returns are also subject to changes in the refractive index of the atmosphere. Atmospheric inversions can lead to ducting phenomena that make ground clutter visible well beyond the line of sight horizon. Ranges to 80 nmi are not uncommon. The total return path is a function of the refractive index. At L-band (wavelength $\lambda = 23$ cm) a one-way path length change of 6 cm (0.4 parts per million at 80 nmi) will cause a 180° phase shift of the echo return. Rapid fluctuations in the refractive index will thus modulate the phase of the echo, resulting in apparent nonzero instantaneous velocities. Nevertheless, the intrinsic spectrum of ground clutter is expected to be very narrow, much narrower than the effects caused by the antenna scanning modulation. An example of ground clutter spectra is shown in Fig. 4. More details on ground clutter may be found in (Long [19]; Skolnik [3]).

3) Rain and storm systems: Water is a good reflector of radar energy, which makes clouds (consisting of suspended water droplets or ice crystals) a common source of clutter. Clouds, unlike ground clutter, can show a significant velocity component relative to the radar depending, of course, on wind speed and direction. While all water droplets within the cloud move in generally the same direction, there are several mechanisms that cause some internal motion as well. First, wind and gravity effects on the individual velocities of the droplets will vary with their size. The larger droplets will be more resistant to changes in wind velocity and also tend to fall toward the ground faster in the absence of any updrafts. Considerable downward motion of this type will itself cause downdrafts, as well as updrafts to replace the downward moving air. Thunderstorms are examples of cloud systems with particularly violent turbulence of this type. The second mechanism for internal motion is mechanical turbulence, present in low-altitude clouds and caused by the uneven drag of the earth's surface and obstacles such as mountains and buildings. This turbulence, while not generally as severe as that found in thunderstorms, can still result in significant variations in droplet velocities relative to each other. The third mechanism is windshear, or radically different wind speeds and directions at different altitudes.

Turbulence, or the (on a macro level) random internal motion of the water droplets will modulate the measured velocity of the cloud system somewhat. Obviously, the greater the turbulence, the greater will be the modulation. Scintillation is not usually a consequence due to the very large number of individual scatterers, although it may be observed in windshear conditions when the overlaying of two or more clutter areas, each with its own distinctive overall velocities, will result in a type of spatial interference pattern.

Yet another mechanism that affects the clutter spectrum is *beam broadening*. The width of the beam, both in elevation and azimuth, is such that particles traveling through the beam with a constant velocity will exhibit different Doppler frequencies at different points in the beam. This is a direct consequence of the changing geometry and leads to a minimum spectral width even if no internal motion is present. This applies to both vertical (gravity-induced) and horizontal (wind-induced) motion.

Examples of rain clutter spectra are shown in Figs. 5 and 6. While the main spectral peaks are similar in both cases, the presence of a windshear component is visible in Fig. 6. The data for these spectra have been taken from different areas at different distances from the radar, but within the same time period. More information on rain and weather clutter may be found in the comprehensive reviews published by Nathanson and Reilly [14], Smith *et al.* [15], and Zrnic and Doviak [16].

4) Birds and bird flocks: Birds have long perplexed radar operators who refer to their echoes as "angels" (Eastwood

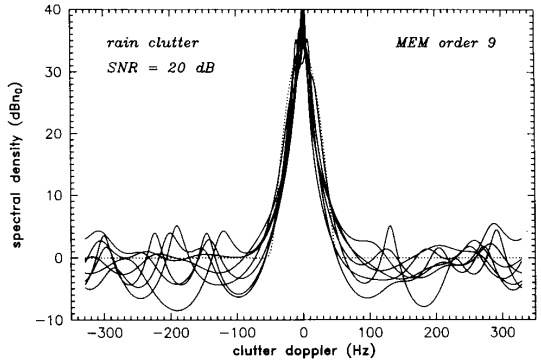

Fig. 5. Maximum entropy spectra of rain clutter.

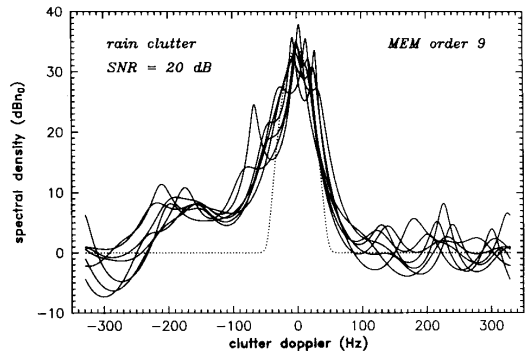

Fig. 6. Maximum entropy spectra of rain clutter with a windshear component.

[4]). It is very difficult to correlate bird echoes with visual observations (see Appendix B), and aside from the often large numbers of echoes present, each individual echo resembles that of a small aircraft both in strength and velocity. It is now known that echoes of single birds are usually too small to be seen on long-range surveillance radars, and it is whole bird flocks which, depending on the number of birds present within a resolution cell, make up single strong echoes. Unlike water droplets, which passively move with the wind, birds are individually powered scatterers who do not move in exactly the same direction and at the same speed. Birds are continuously shifting their position within the flock, even those that fly in a particular formation. This causes large fluctuations in net echo strength as individual echoes sum constructively at one moment and destructively at the next moment. Clearly the variation in net velocities, and consequently the spectrum, will be quite large for the same reason.

Bird echoes are well known to exhibit wingbeat modulation, that is a variation in RCS as a function of the wing motion [5], [11]. This wing motion also causes instantaneous changes in the relative velocity as the bird

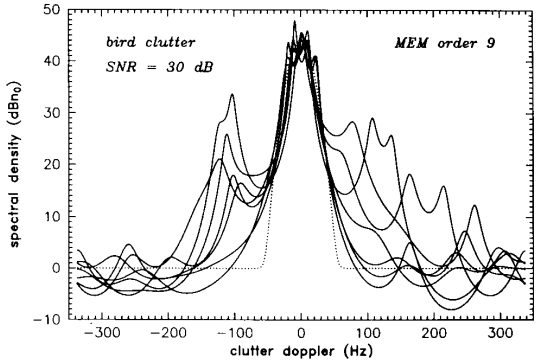

Fig. 7. Maximum entropy spectra of bird clutter.

moves up and down. It is unlikely that the wings themselves contribute much to the echo itself; their mass is quite small and the feathers do not reflect radar energy very well. Both wingbeat and velocity modulation will vary considerably with the species. In general, bird echoes exhibit the largest variation in the spectrum of all clutter types. This is dramatically demonstrated with the spectra shown in Fig. 7. An excellent and more complete summary of the current knowledge on birds as radar targets has recently been published by Vaughn [6].

B. Spectral Spread and Shape

Changes in RCS, such as wingbeat modulation in bird echoes, modulate the entire pulse echo in amplitude. This will further contribute sidebands to the spectrum. Their strength is proportional to the range of the RCS fluctuations, and their spread is a function of the frequency of the changes. Figure 8 illustrates the effects of both amplitude and frequency modulation on the MEM spectra of a complex sinusoid in noise. The modulating signal was a single 20 Hz sinusoid, adjusted such that it would result in a modulation index of 50% for the AM case, and a peak deviation of 20 Hz for the FM case. Both were also used simultaneously for the spectrum labeled AM and FM. It is interesting to note that frequency modulation appears to dominate and widen the spectrum much more than the particular level of amplitude modulation chosen. A 20-Hz frequency deviation corresponds to approximately 4.5 kn and is not an unreasonable level of internal motion. A 20-Hz modulation is typical and representative of the wingbeat for bird echoes [6] and the Doppler spectrum deviations for rain clutter [14].

The presence of more than one scatterer within a resolution cell leads to more complex and variable spectra. This can occur quite frequently in surveillance radars since their fan beam is designed to detect targets at various heights. The resolution cell is thus a column extending from near the ground to altitudes of over 50 000 feet. Each target or group of targets has its own average velocity and internal motion, and thus leads to individual spectral

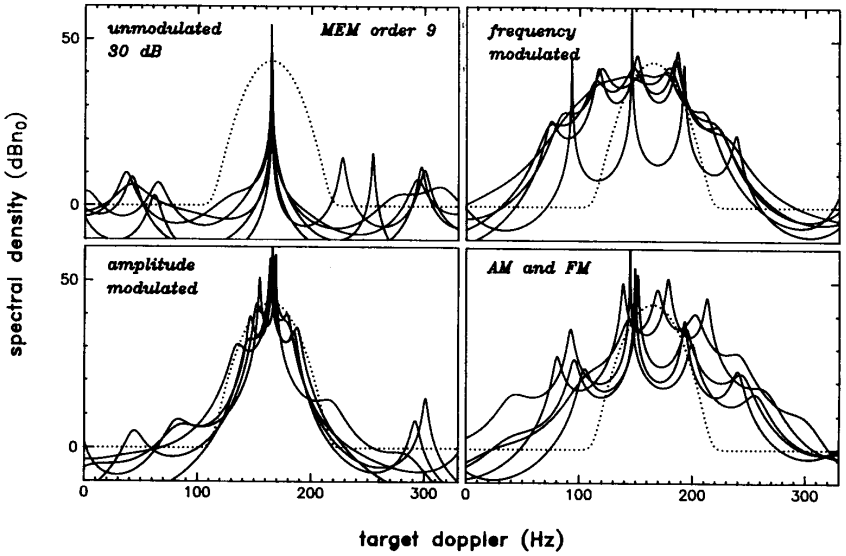

Fig. 8. Maximum entropy spectra of amplitude and frequency modulated sinusoids-in-noise.

peaks. Superposition of these spectra leads to variable, and often unpredictable composite spectra. Destructive as well as constructive superposition will distort the spectrum significantly, and expected adjacent spectral peaks will either disappear or merge into one. Little can be done about this with target-like clutter such as birds. Rain, on the other hand, has a larger spatial extent and it may well be possible to take in sufficient data to derive accurate spectra. Still, the effects of beam broadening will always be present, especially in distributed clutter and for time series extending over large scan angles. Pure line spectra are thus never expected, even in calm weather conditions. As turbulence increases, however, the spectral spread will increase further and two (or more) distinct peaks may then become discernible in windshear conditions.

C. Effects of the Antenna Scanning Motion

The presence of unique characteristics within the Doppler spectrum of the radar returns is a necessary, but not a sufficient condition for classification to be possible. These characteristics must also be present in the spectral estimate, or at least the feature set arising from the estimate. Unfortunately, a number of measurement limitations severely restrict their visibility. The first limitation is quite simply the small number of samples per target, N, as the beam scans past it. This will lead to a relatively large variance in the spectral estimate and, conversely, any feature set arising from it. The increase in variance with low values of N has the effect of obscuring small details and effectively reducing the resolution capabilities of spectral estimators. This again suggests that as much data as possible should be included in the estimate, particularly if the true variance is already known to be large.

Another limitation is the shape of the antenna beam itself. As the beam scans past the target, the echo pulses will be modulated accordingly. This amplitude modulation, usually approximating a Gaussian or $\sin(x)/x$ pulse shape [50], will convolve with the basic spectrum, introduce additional spreading, and further reduce the maximum available resolution. For distributed clutter this may not appear to be a problem since new scatterers are continuously entering the beam, but these new scatterers are moving independently which, in most cases, can lead to significantly different spectra. These will superimpose and also spread the resulting spectrum; hence, the resolution of individual clutter spectra is not necessarily any better.

The third limitation is the introduction of an apparent Doppler shift due to the size and orientation of reflecting objects. A building that presents a reflecting surface wider than the beamwidth, but with a nontangential orientation, will result in successive returns of similar magnitude but continuously changing phase as the center of the beam scans past it. This will result in a fairly sharp spectral peak at a nonzero frequency, quite unlike that of a nonmoving point target with its Gaussian spectrum.

V. Data Preparation

In order to carry out the experimental studies on the classification of radar clutter reported in this paper, approximately 260 min (20 gigasamples) of radar data were collected from two operational L-band ATC radars. Some understanding of the radar is clearly essential to prepare the data for classification. Preparation itself includes selection

from the indexed tape library, correction of tape errors, and compensation for alignment errors in the radar demodulator. Once such preprocessing is complete, feature extraction (described in Section VI) can take place.

A. The Radar System

The choice of radar to use for this research was based on several requirements, and it had to be representative of the new generation of ATC radars being installed. A coherent digital MTI or MTD radar was preferred to allow high integrity data capture and recording. Difficulties with maintaining coherence using a two-channel analog video recorder made digital recording an absolute requirement. The choice of operating frequency of the radar was less important, and past analyses have been made with both L- and S-band radars [27].

The site of the radar was perhaps the most important consideration. Since one of the main motivations for the original research was bird clutter identification, the radar had to be situated below major bird-migration flyways. In cooperation with the Canadian Department of National Defence (DND), Ottawa, the Moose Jaw (Saskatchewan) TRACS-ASR was made available for the research during the migration season in the spring of 1985. CFB Moose Jaw is Canada's jet training base and is plagued by heavy bird migration both during the spring and the fall. For safety reasons, the policy of the base is to shut down night flying during the migration season. Since little itinerant traffic passes through Moose Jaw, the radar was normally not required for traffic control during hours of darkness. It could therefore be relatively freely configured to the research parameters, which made it an ideal choice. An identical radar at CFB Trenton (Ontario) was also made available during the fall of 1984 for early trials and testing and debugging of the recording equipment. This radar was, however, dedicated to air traffic control on a 24-hour basis and its parameters (PRF, beam gating, and STC) could not be adjusted as freely.

The TRACS-ASR is a derivative of the Westinghouse ARSR-3 long-range air route surveillance radar. It was specially modified to meet Canadian DND requirements for a terminal radar as part of their ATC modernization program. The two-channel radar transmits an uncompressed 2-μs pulse with a peak power of 1.5 MW at a repetition frequency of about 650 Hz. Both channels can operate simultaneously, each on a different frequency and polarization. This enhances the detection capability of the radar by providing frequency and polarization diversity. Unfortunately, only one channel was operative at the time of recording and use could not be made of this feature. The use of polarization as a discriminant thus remains untried, but should not be overlooked in future classifier designs. It was fortuitous, however, that the operating channel in Moose Jaw was horizontally polarized, since bird reflectivity for this polarization has been reported to be as much as 4 dB greater than that for vertical polarization [6].

The radar uses a dual-feed antenna (high beam is receive only), which forms a fan shaped beam with a modified cosecant-squared elevation pattern and a narrow 1.5° azimuth beam width. The scan rate of 12 rpm results in approximately 14 hits per beamwidth which, when processed coherently, can provide up to 10 dB of integration gain above a single hit detection. The TRACS signal processor is fully digital, and features MTI for both I and Q channels, log-CFAR, noncoherent pulse-to-pulse integration, and adaptive threshold detection to maintain control over the dynamic range. Staggered PRF's are used to kill MTI blind speeds and tear up second-time-around echoes. Detected targets are correlated with SSR transponder returns and sent to the main computer system for target tracking, display, and logging. The radar operator sees a fully synthesized display complete with target identification and track history, as well as weather contours. Wide-band log video is also available as an overlay and is usually fed from the weather channel. More detail on the radars is provided in Appendix A.

In Section III it was seen that the data of interest are the coherently demodulated in-phase (I) and quadrature (Q) channels. Both are available from the analog-to-digital converters to a precision of 9 bits plus sign. The digital data are linear in amplitude and the 9-bit mantissa thus results in a 46-dB dynamic range. Conventional spectral analysis requires not only linear data, but also fixed intervals between samples. The radar has four choices of fixed PRF's, and two of these were used.

The coherent demodulation process in the radar receivers is subject to alignment errors, as is the analog-to-digital conversion. Unequal gains in each of the channels and the lack of true orthogonality must be compensated for if the errors are sufficiently large. Given a signal with amplitude S and phase ϕ, the correct in-phase and quadrature components are defined as

$$I = S\cos\phi, \qquad Q = S\sin\phi. \tag{8}$$

The corrections that must be applied to the measured values I' and Q' are thus

$$I = k_1 I' \quad \text{and} \quad Q = k_2\,(Q' + k_3 I) \tag{9}$$

where

$$k_1 = \frac{1}{k_c}; \quad k_2 = \frac{1}{\cos\delta}; \quad k_3 = \sin\delta. \tag{10}$$

The quantities k_c and δ are the gain ratio between the I and Q channels and the quadrature error (deviation from orthogonality), respectively. A complete discussion on these errors and how to find the required correction factors is given in Stehwien [41]. Before this correction can be applied, any offset due to A/D misadjustment must be removed. I' and Q' are known to be zero-mean parameters and the sample mean of thermal noise data is thus a good (low variance) estimate of this offset. Thermal noise, which has the lowest variance of any radar signal and is white relative to the PRF, can be found in abundance in the region beyond 80 nmi where the radar is still receiving, but where few targets and even less clutter would be found. From this region the thermal noise powers W_I and W_Q can also

be estimated. They are required for both the procedure to find the above correction factors and the computation of the SNR-based feature as discussed later.

A final preprocessing consideration is the integrity of the data on tape. All data were stored on 1″ high-density tape using an Ampex AHBR-1728 (airborne high bit-rate, 28 track) 24-channel digital recording system. High recording rates (up to 4 Mbits/s for each channel) are achieved using an analog recording scheme with proprietary encoding to limit bit errors. Bit densities used were 26.67 and 23.33 kbits/in, and all 24 bits (10 each for I and Q, 4 for control) were written twice to detect the presence of errors. Isolated bit and burst errors were in fact encountered, and some error correction was necessary before processing the data. Simulated PPI displays reveal such errors quite readily and, since they tended to be confined to a single track on the tape, some form of interactively directed median filtering was able to find and correct them. Small errors, which are within the dynamic range of the local data itself, can safely be ignored on the assumption that their effects are negligible when compared to the large volumes of correct data.

B. Data Library

The complete library consists of 15 indexed tapes, each containing 15 to 18 min (approximately 200 scans) of continuous radar data. Each scan is numbered and retrievable singly or as a sequence of consecutive scans. Notes made while recording refer to the tape in general and the scan number in particular. As a result, individually identified targets and clutter areas can be found quickly and recovered. Supporting data include the time and date of the recordings, the relevant radar parameters, locations of clutter and aircraft targets, video tapes of the PPI display, time exposure photographs, a daily record of the meteorological conditions including surface and upper wind forecasts and hourly weather reports, and bird migration predictions. In some cases comparisons were made with bird migration predictions and findings from CFB Cold Lake, which is situated 300 nmi north-west of Moose Jaw, and under the same general migration path.

Bird clutter, regardless of any expectations of heavy migration, proved to be the most elusive of all the clutter types, or perhaps only appeared so at first. However, since it was impossible to positively identify bird and bird-flock echoes, recordings were made whenever bird clutter was suspected or whenever unusual clutter was observed. Later analysis showed that most of the suspected clutter was indeed caused by birds (see Appendix B); and consequently 10 of the 12 tapes recorded in Moose Jaw contain at least some bird clutter. Seven of these are considered "bird tapes," with bird clutter as a main component of the radar data. Two contain "heavy" bird clutter with the number of completed video counts (MTI-detected targets within one scan) exceeding 700. Weather clutter is present on 8 tapes, and three of these were specifically recorded to capture periods of heavy precipitation, including thunderstorms.

Not surprisingly, some aircraft targets are present on virtually every tape, but only two were recorded during peak flying hours. One recording made at Trenton during a late afternoon peak air traffic period captured many aircraft traveling the Ottawa-Toronto-Montreal routes and some within New York State. Compared to clutter, however, aircraft targets are very sparse and therefore cannot be characterized with the same reliability and confidence. The same is generally true for ground-based vehicular traffic, including cars and trucks. Most of this type of clutter was found on one tape recorded in Moose Jaw when a slight atmospheric inversion made a significant stretch (more than 20 miles) of the Trans-Canada Highway visible to the radar. While expressway traffic near Trenton also appeared on time exposure photographs, it was much more spotty due to the hilly terrain and locating a sufficient number of vehicle echoes was not possible.

Ground clutter is obviously present on all tapes; however, the atmospheric inversions experienced quite frequently and regularly in the Province of Saskatchewan, Canada, can increase the maximum range of the extent of ground clutter to over 80 nmi. This proved to be a significant problem for the radar, since the available STC curves were not designed for these ranges and the sometimes extremely powerful clutter returns. Occasional MTI breakthroughs were noted, although it was not clear whether these were caused by dynamic range overload, vehicular traffic, or some other inversion related phenomenon. Two tapes were recorded during such conditions.

One Moose Jaw tape was recorded during the daytime, at the end of a regularly scheduled radar preventive maintenance period.This tape contains several radar parameter variations and injected reference targets. It also contains jet training activity and localized bird clutter. More detail on the recorded data and the data library may be found in [43].

VI. Feature Selection and Extraction

Before classification can proceed, the radar measurements must be transformed into a set of features in order to enhance class differences that contribute to their separability. This is not an easy task, especially if the measurements are contaminated with dominant information which is not unique to any one class.

Most feature selection and extraction techniques proceed on a rather *ad hoc* basis. While several authors claim "optimality," their procedures are, in fact, based on assumptions that apply only to a specific group of problems [37]. These procedures may be optimal if the assumptions made are correct, but when they are suspect one necessarily returns to *ad hoc* rules.

What, then, are the most likely features of radar clutter that may lead to successful classification? As pointed out in Section III, given that classification is to be accomplished on a single-scan basis, the most important features relate to the shape and variability of the Doppler spectrum as it reflects the internal motion of the clutter. Spatial distribution and the relative size of the scatterers can best be identified on a scan-to-scan basis using image processing and syntactic recognition techniques, although aspects of

the relative distribution within one resolution cell can be estimated and included as a feature. No other features are assumed to be available.

The overall Doppler frequency itself is not unique to any one clutter type and depends much on the relative direction of travel of the clutter with respect to the radar. In addition, aliasing due to sampling at the PRF will effectively generate uniform probability densities for the aliased center frequencies of most moving clutter types.

Signal strength can clearly only be a feature if the radar is calibrated, which is generally not the case for surveillance radars. Many factors affect the strength of the echo, including target position within the beam pattern and multipath propagation. Realistically, neither of these effects can be quantified since they are markedly affected by anomalous propagation. Target glint and the distribution of clutter scatterers within one resolution cell can be considered random variables and are the source of significant scintillation.

A. Selection of Separable Features

Previous work (Stehwien [45]) established the utility of Burg's *reflection coefficients*, arising from the MEM of spectral analysis (Burg [51]), for the extraction of spectral features. The coefficients arise out of the lattice implementation of the prediction error frequency (PEF), which attempts to minimize the prediction error power at each stage. This minimization results in a whitening filter, and as such the reflection coefficients represent the incremental "predictable" information extracted from the time series at each stage. In his original work, Burg dealt only with real-valued coefficients derived from real valued data; the complex form of the algorithm is given by [52]

$$\rho_m = -2\frac{\sum_{n=m+1}^{N} f_n^{(m-1)} b_{n-1}^{*(m-1)}}{\sum_{n=m+1}^{N}\left[\left|f_n^{(m-1)}\right|^2 + \left|b_{n-1}^{(m-1)}\right|^2\right]} \tag{11}$$

where $f_n^{(m)}$ and $b_n^{(m)}$ are, respectively, the forward and backward prediction errors of the prediction error filter of order m. They are computed using the lattice filter, described by the following pair of equations:

$$f_n^{(m)} = f_n^{(m-1)} + \rho_m b_{n-1}^{(m-1)} \tag{12}$$

$$b_n^{(m)} = b_{n-1}^{(m-1)} + \rho_m^* f_n^{(m-1)}. \tag{13}$$

The first stage of the lattice uses the data values themselves as input

$$f_n^{(0)} = b_n^{(0)} = x_n. \tag{14}$$

The magnitude of the reflection coefficient ρ_m depends on the incoming signal strength, which suggests that the average signal power itself should be measured. Such a measurement also helps to bring the comparisons onto an equal basis. The average signal power is simply (assuming ergodicity in the mean)

$$P_0 = E[x_n x_n^*] \approx \frac{1}{N}\sum_{n=1}^{N} x_n x_n^* \tag{15}$$

where E is the expectation operator, and N is the total number of data samples.

The signal power P_0 is an average value computed over the entire measurement window. The size of this window is limited by the length of time the beam of the radar spends on any one scatterer. In a scanning radar this time is determined by the antenna rotation rate, the PRF, and the antenna beamwidth. The echoes from the scatterer are amplitude-modulated by the azimuth beamshape itself, which can usually be approximated by a Gaussian function. Single scatterers such as aircraft targets will thus generate, to a first approximation, Gaussian amplitude-modulated complex sinusoids. Echoes from multiple, widely distributed scatterers will not exhibit such modulation as the density of scatterers which the radar sees, will remain relatively constant. This distinction may not become apparent in spectral analysis and a measure of amplitude distribution should thus be explicitly computed. The normalized variance of the individual sample amplitudes $s = \mid x \mid$ is such a measure, as shown by

$$P_{var} = \frac{E\left[(s-\bar{s})^2\right]}{E(s)^2} = \frac{P_0 - \bar{s}^2}{P_0} = 1 - \frac{\bar{s}^2}{P_0} \tag{16}$$

where $\bar{s}$ is the mean value of s. A feature to aid the separation of distributed clutter from point targets is thus introduced. Unfortunately, this feature does not yield complete information regarding the nature of the signal variability within the observation window. Comparing the power levels in the center of the window (P_I) with those outside the center (P_E) will yield a feature that can distinguish between convex and concave signal shapes:

$$P_{dif} = \frac{P_I - P_E}{P_0} = \frac{N}{N_E}\left[\frac{P_I}{P_0} - 1\right] \tag{17}$$

where $NP_0 = N_I P_I + N_E P_E$ and $N = N_I + N_E$, the total number of samples in each computation. The ratio N/N_E is simply a constant and can be ignored when the number of samples is fixed. A computationally simpler feature is thus

$$P_{dif} = \frac{P_I}{P_0} - 1. \tag{18}$$

This completes the separable feature set that may be extracted from the radar data.

B. Dimensionality

One question that remains to be answered regarding the feature set is: how many of the reflection coefficients ρ_m are required to achieve sufficient separability between classes to make classification practical? The obvious answer might be as many as possible within the computational constraints provided by technology. While it is generally

true that including additional independent features will increase separability, there is a very real problem with adding many features arbitrarily, even if there are no computational constraints. This problem has been recognized in the literature (Hughes [53]; Duda and Hart [1]), and can generally be attributed to a lack of sufficient data for the estimation of the class prototypes.

Increasing the number of features frequently increases the *intrinsic dimensionality*, or complexity of the data. The notion that this increase leads automatically to a refinement of the available information is cautioned against by van Campenhout [54] who shows that, unless the statistical models are comparable, predicting a decrease in the Bayes risk with increasing complexity is not justified. Such an increase in complexity is expected to occur when adding higher-order reflection coefficients, but it is not clear whether this new information is a help or a hindrance. The following heuristic argument should provide insight into the problem.

The transformation from the radar measurements (in the form of time series) to the reflection coefficients is highly nonlinear and concentrates the dominant information in the low-order coefficients. While this is a highly desirable property from the point of view of reducing the dimensionality, it also displays nonlinear behavior toward measurement errors and interfering noise. In general, high-order coefficients contain not only less information but they are estimated using fewer data samples. Hence, their variance is greater and their individual separability is therefore quite low. They are also based on the residuals of the earlier stages, and therefore tend to be more susceptible to measurement noise and quadrature errors (Stehwien [42]). Depending on the relative strength of these errors, residual and helpful information may be completely obscured to the point of reducing classifier performance. At which filter order this problem becomes significant will undoubtedly depend on the data set itself, the number of samples used in the coefficient calculation, and the size of the measurement errors.

Another consideration in the choice of the number of coefficients to use is problem specific, that is, the number of classes to be defined. The clutter processes represent a continuum in the feature space, and the class boundaries are not likely to be described with great accuracy. The definition of a few large and coarsely defined classes will most certainly require only a few features for acceptable classification, whereas the successful identification of many refined classes might well require more. Clearly, a large dimensionality allows more regions of "equal" size to be defined, which may be necessary if the feature set for any one class is not a hyperellipsoidal cluster (using the multivariate normal assumption), but rather some complex shape stringing its way through the feature space. In such a case a contiguous set of small hyperellipsoidal regions can be defined along the irregular cluster, with each region belonging to the same class.

C. Multisegment Burg Algorithm for Computing the Reflection Coefficients

The importance of low variance in selecting the number of measurement samples is well recognized. If the number of samples, N, is restricted to N_θ, the number of hits per beamwidth (which is less than 15 for the TRACS radar), then only a second-order lattice predictor may be used if a low error rate is to be maintained. Similarly, about $N = 100$ samples would be needed for a fifth-order lattice predictor. For even longer lattice predictors, the required number of samples N increases rather dramatically (Stehwien [42]).

However, the danger in increasing N arbitrarily is the inclusion of clutter from outside the resolution cell of interest. While this may be useful in cases of widespread and homogeneous clutter, composite spectra would result in all other cases. This must clearly be avoided for successful classification. A reasonable scenario is to take samples in both range and azimuth out to some level of attenuation as determined by the receiver bandwidth and the azimuthal beamwidth of the antenna. Within the 3-dB (one-way) width, the radar resolution cell includes up to 15 samples in azimuth and 3 in range for a total of 45. Not all of these can be considered independent, however, and it may be appropriate to expand the extent in azimuth somewhat.

Including several "parallel" time series raises the question of how to combine them to form a single coefficient estimate. Haykin *et al.* [28] propose the use of the multisegment Burg formula—first described by Moorcroft [55], who applied it to UHF radar data to obtain radio-auroral spectra—for the analysis of radar clutter. This formula accomplishes the stage-by-stage minimization of the average prediction error power of all lattice filters using the same reflection coefficient. Averaging the error powers is equivalent to separately averaging the numerator and denominator of the Burg formula given by (11). We may thus modify this formula by writing

$$\rho_m = \frac{-2\sum_{k=1}^{K}\sum_{n=m+1}^{N} f_{n,k}^{(m-1)} b_{n-1,k}^{*(m-1)}}{\sum_{k=1}^{K}\sum_{n=m+1}^{N}\left[\left|f_{n,k}^{(m-1)}\right|^2 + \left|b_{n-1,k}^{(m-1)}\right|^2\right]} \tag{19}$$

where ${f^{(m)}}_{n,k}$ and ${b^{(m)}}_{n,k}$ are, respectively, the forward and backward prediction errors from the m^{th} stage of the k^{th} lattice PEF:

$$f_{n,k}^{(m)} = f_{n,k}^{(m-1)} + \rho_m b_{n-1,k}^{(m-1)} \tag{20}$$

$$b_{n,k}^{(m)} = b_{n-1,k}^{(m-1)} + \rho_m^* f_{n,k}^{(m-1)}. \tag{21}$$

Note the common use of the same reflection coefficient. Each of the K lattice filters is initialized with data values from the k^{th} time series

$$f_{n,k}^{(0)} = b_{n,k}^{(0)} = x_{n,k} \tag{22}$$

The value of K thus might be set to 3 to include the three parallel time series, one per range sample cell. An alternate solution is to include $3 \times 15 = 45$ time series, each centered around one sample within the 3-dB resolution cell. The series are then clearly overlapping in azimuth and

thus a form of data weighting is introduced. The length of the series determines the form of weighting employed and the number of samples accessed outside the resolution cell of interest. A length of 15 leads to an initial triangular weighting with the emphasis linearly decreasing from a maximum at the center of the resolution cell to a minimum at the 10-dB (one-way) point, using a total of 29 samples in azimuth. The benefits of such data weighting are a reduced impact of scatterers from adjacent cells, while maintaining a relatively large sample count. Another benefit is the ability to pretest the individual time series based on their respective ρ_1 and to prevent their inclusion in (19) if it is clear that they belong to a different cell.

A similar form of data weighting is applied to the computation of P_0 in (15); specifically, we may write

$$P_0 = \frac{1}{NK} \sum_{k=1}^{N} \sum_{n=1}^{N} x_{n,k} x_{n,k}^{*}. \tag{23}$$

This prevents the power from being unduly reduced by edge and beamshape effects. In fact, it can be shown that, in the case of a Gaussian amplitude shape for both range and azimuth similar to that found in the TRACS radar data, P_0 is 4.6 dB below what would be expected if the peak amplitude were extended uniformly over all samples. This is less of a reduction than would be experienced with uniform weighting of the same number of samples (6.3 dB).

This form of the *multisegment Burg formula* is used throughout the research reported in this paper. The effective number of independent samples used was experimentally determined (by application to pure thermal noise data) to be approximately 61. This number is reasonable since it is greater than 45 that should arise from the use of only those samples inside the resolution cell, and less than 87 that would result from weighting all 3×29 samples equally. A fourth-order filter may thus be used with a clear conscience.

Once computed, the reflection coefficients must be normalized to the center frequency by a rotation according to

$$\rho'_m = \rho_m \exp(-jm\phi) \tag{24}$$

where

$$\phi = arg(-\rho_1). \tag{25}$$

Rotating ρ_1 by its own argument simply leaves its magnitude. The magnitude $r = | rho_1 |$ is related to n_0, the SNR, as follows

$$r = |\rho_1| = \frac{n_0}{1 + n_0}. \tag{26}$$

This leads to the normality transformation

$$u_0 = 10 \log(n_0) = 10 \log\left(\frac{\mathrm{r}}{1 - \mathrm{r}}\right). \tag{27}$$

The need for a similar transformation also applies to the signal strength feature P_0, conveniently expressed in decibels relative to the thermal noise power W as follows

$$U_0 = 10 \log(P_0/W). \tag{28}$$

Two additional features remain: The normalized variance of the sample amplitudes, P_{var}, and the relative difference in sample amplitudes between the center and the edges of the resolution cell, P_{dif}. Their basic definitions are given in (16) and (18), respectively. For the multisegment architecture of the feature extractor, they are computed as

$$P_{var} = 1 - \frac{\bar{s}^2}{P_I} \quad \text{and} \quad P_{dif} = \frac{P_I}{P_0} - 1 \tag{29}$$

where

$$\bar{s} = \frac{1}{K} \sum_{k=1}^{K} s_k \quad \text{and} \quad P_I = \frac{1}{K} \sum_{k=1}^{K} s_k^2. \tag{30}$$

Only the centers of each of the K time series are used, and s_k is defined as

$$s_k = \left| x_{(N+1)/2,k} \right|. \tag{31}$$

The parameter P_{var} is thus the normalized variance of the amplitudes inside the 3-dB resolution cell without the application of any data weighting; and P_{dif} is related to the difference between the unweighted average power inside the resolution cell and that outside of it, insofar that it is included in P_0. For the Gaussian shaped ideal target centered inside the window, P_I is nominally 0.75 dB above P_0, and P_{dif} takes on a value of 0.19. The variance of the sample amplitudes is also a fixed fraction for Gaussian shapes for which P_{var} is then expected to be 0.13. Values of zero for both features indicate evenly distributed clutter, whereas negative values for P_{dif} can occur along clutter edges or in cases of destructive interference of multiple clutter returns inside the resolution cell.

To summarize, the feature set is based on the following parameters:

- ϕ mean Doppler frequency,
- P_0 mean signal power relative to the known measurement noise power W,
- ρ'_{m} (zero mean Doppler) heterodyned reflection coefficients,
- P_{var} normalized variance of the amplitude distribution within the measurement window,
- P_{dif} normalized power difference between window center and edges.

D. Variations of Reflection Coefficients with Signal-to-Noise Ratios (SNR's)

Signal strength, and therefore SNR, is a function of the measurement itself; it therefore does *not* contain useful information about the source of radar returns. Only absolute RCS measurements may be viewed as a feature, but these are generally not derivable from surveillance radar data. In fact, in the context of the radar clutter classification studies reported herein, SNR information has proved to be detrimental rather than helpful (Stehwien [42]).

A quantitative presentation of the relationship of the reflection coefficients with SNR is given in Fig. 9. For these plots, all sample feature vectors were separated into 1-dB

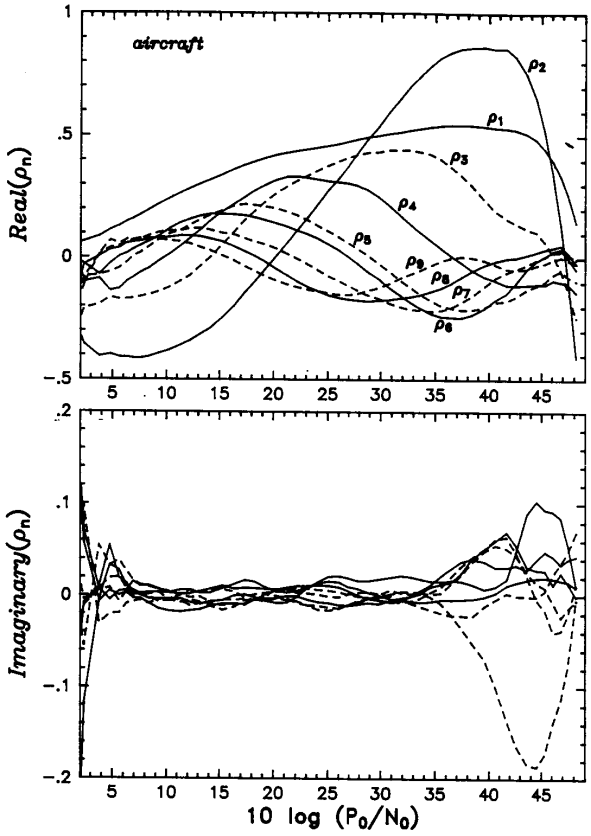

Fig. 9(a). Mean feature values for aircraft target.

bins according to the value of U_0 and then averaged. The averages of the real and imaginary parts of the reflection coefficients were plotted against the average values of U_0 for each bin. Two important observations may be made from Figs. 9(a)–9(e):

1. The "symmetric" nature of the spectra of aircraft echoes and ground clutter is evident from the absence of meaningful information in the imaginary parts of the reflection coefficients; see Figs. 9(a) and 9(b). Estimation errors due to the low number of samples contribute to the deviation from zero at both the low and high ends of the plot.
2. The difference in the high-order coefficients of the bird and weather clutter classes is dramatically visible in the plots presented in Figs. 9(c)–9(e). Bird clutter appears to have little meaningful information in the real part past ρ_3, although the increase in the values of the imaginary parts indicates the greater likelihood for asymmetrical spectra. Weather clutter, which includes both rain and storm systems, has a less defined cutoff point in the order, although ρ_7' to ρ_9' do not appear very meaningful. It is interesting to note the effects of the secondary peak due to windshear. The difference in the real parts of the reflection coefficients is slight, whereas it is significant in the imaginary parts, particularly for ρ_2' and ρ_3'. This establishes the importance of retaining the imaginary parts in the feature vector.

E. Class Prototypes

The question to be resolved then is: how can class prototypes be defined in order to mitigate the dependence of the reflection coefficients on the signal strength as evidenced in the plots of Fig. 9? Short of finding some nonlinear transformation that removes this dependence entirely, comparisons among classes may clearly proceed only for equal values of the parameter U_0. That, of course, requires a prototype definition for every possible value of U_0, which is an unworkable solution. It is possible, however, to define prototypes for a range of signal strengths, with the ranges defined such that the nonlinear behavior within them is limited and manageable using the multivariate normal assumption.

A definition of this kind is necessarily somewhat arbitrary

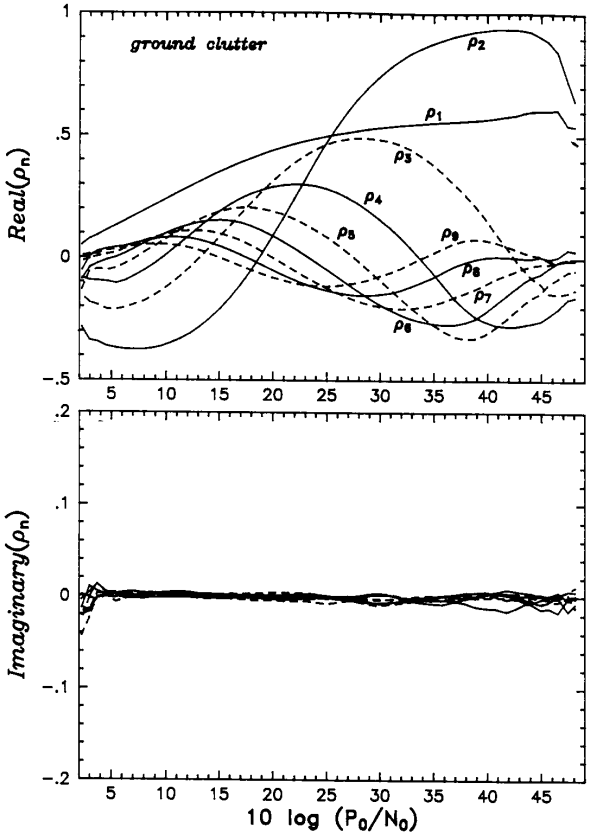

Fig. 9(b). Mean feature values for ground clutter.

since nature rarely provides logical breakpoints for such efforts. This case is no exception, and Fig. 9 merely emphasizes this point. Ultimately, the capacity for the number of classifiers which can be reasonably constructed and the requirement for a large number of training samples must be traded off against the strength of the residual nonlinearities. While a global minimization of error rates would lead to a theoretically optimum tradeoff, this minimization is cumbersome to carry out and quite likely impossible to achieve with the limited data volumes in research data bases. It is not at all clear what, if any, improvement would result from such minimization. The ranges used in this paper are therefore not likely to be those that would be used in a real-time implementation. They did, nevertheless, lead to quite acceptable results. The range breakpoints for U_0 were chosen as follows:

$$\underset{\min}{|}\ A\ \underset{4\,dB}{|}\ B\ \underset{10\,dB}{|}\ C\ \underset{17\,dB}{|}\ D\ \underset{24\,dB}{|}\ E\ \underset{32\,dB}{|}\ F\ \underset{\max}{|}\ . \quad (32)$$

Most of the data sets fell into ranges B to D, with E and F mostly appearing for point-target classes and some of the bird clutter. The latter became available only when turning the radar's STC off. Unfortunately, this was not done for the weather clutter data, ,and its strongest data sets turned out to fall around 23 to 25 dB. Range A was defined effectively only to allow the identification of weak moving clutter, since little discriminating information was expected at such low levels.

VII. Bayes Classification

Given the set of features extracted from the radar data in the manner described in the previous section, we are ready to undertake the task of classification. In this section, we consider the classical Bayes decision theory that provides optimum decision rules when the statistics of the problem are known. As is customary, we assume a multivariate Gaussian distribution for the features. The classifier performance so obtained provides a frame of reference for the neural network classifier studied in the next section.

A. Bayes Decision Theory

Designing the classes of interest as c_i, and the probability that the d-dimensional feature vector $\mathbf{y}$ belongs to any one of them as $P(c_i \mid \mathbf{y})$, then a simple rule for assignment of $\mathbf{y}$ to c_j is one that chooses the class with the largest a

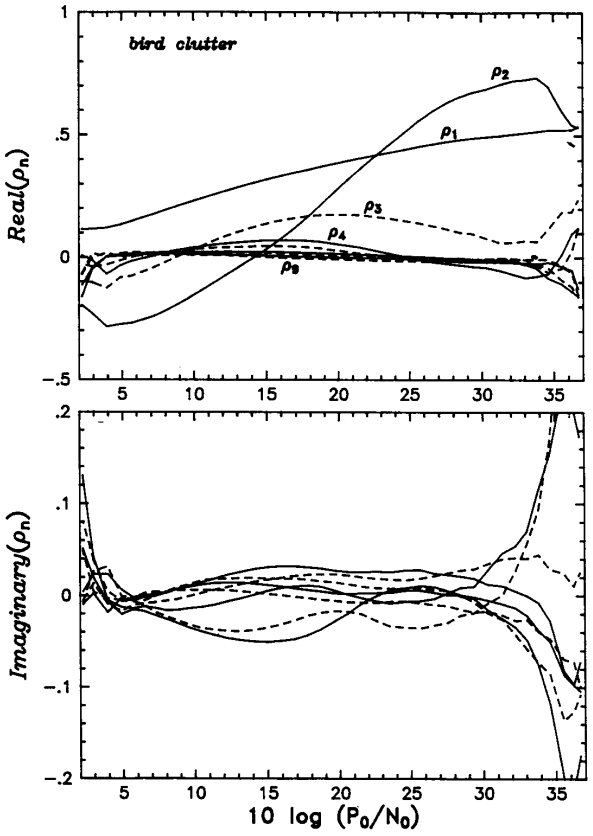

Fig. 9(c). Mean feature values for bird clutter.

posteriori probability

$$P(c_j \mid \mathbf{y}) > P(c_i \mid \mathbf{y}), \qquad \text{all } i \neq j. \tag{33}$$

These probabilities are, however, seldom known or directly measurable. It is more likely that the *a priori* probabilities $P(c_i)$ are known, along with the *class-conditional* probability densities $p(\mathbf{y} \mid c_i)$, which describe the statistics of the feature vector $\mathbf{y}$, given the underlying stochastic process belonging to class c_i. Then we may use Bayes rule:

$$P(c_i \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid c_i)P(c_i)}{p(\mathbf{y})} \tag{34}$$

where

$$p(\mathbf{y}) = \sum_{i=1}^{L} p(\mathbf{y} \mid c_i)P(c_i) \tag{35}$$

and L is the number of classes under consideration. Note that the mixture density $p(\mathbf{y})$ is independent of the index i, and therefore acts only as a scale factor that may be ignored when searching for the maximum per (33).

It may well be that deciding simply on the basis of probabilities is insufficient. For example, the cost associated with erroneously classifying a flock of birds as rain could be extremely high, as numerous bird strikes in the past have demonstrated. The reverse is much more tolerable. Let the loss incurred by taking action α_i, whenever class c_j is the correct class, be $\lambda(\alpha_i|c_j)$. Then the *conditional risk* associated with taking action α_i, given the feature vector $\mathbf{y}$, is

$$R(\alpha_i \mid \mathbf{y}) = \sum_{j=1}^{L} \lambda(\alpha_i \mid c_j)P(c_j \mid \mathbf{y}). \tag{36}$$

The optimal decision rule *(Bayes classifier)* is then to choose that action α_i which presents the smallest risk. It can be shown (Duda and Hart [1]) that this rule minimizes the overall Bayes risk R. Combining (34) and (36) yields

$$R(\alpha_i \mid \mathbf{y}) = \sum_{j=1}^{L} \lambda(\alpha_i \mid c_j)\frac{p(\mathbf{y} \mid c_j)P(c_j)}{p(\mathbf{y})}. \tag{37}$$

Hence,

$$R(\alpha_i \mid \mathbf{y})p(\mathbf{y}) = \sum_{j=1}^{L} \lambda(\alpha_i \mid c_j)P(c_j)p(\mathbf{y} \mid c_j). \tag{38}$$

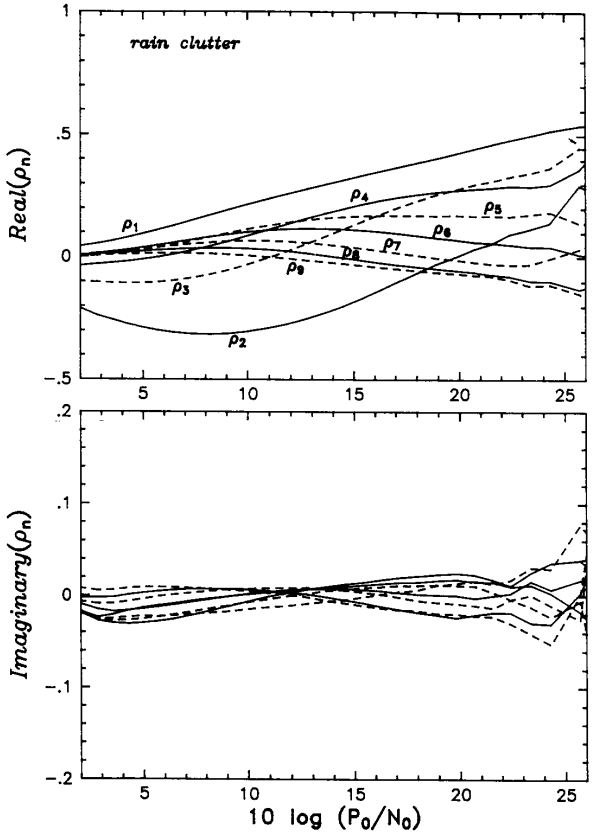

Fig. 9(d). Mean feature values for rain clutter.

The quantity $R(\alpha_i \mid \mathbf{y})p(\mathbf{y})$ is a scaled risk and may be tested instead of the risk $R(\alpha_i \mid \mathbf{y})$ without changing the outcome of the decision. The quantities $\lambda(\alpha_i \mid c_j)P(c_j)$ present, in effect, a bias toward the most likely classes and the least costly decisions and are determined by the nature of the problem itself. In the clutter classification case, they describe the probabilities of seeing the various clutter types and the safety consequences resulting from misclassification.

In the absence of any information indicating otherwise, the most prudent action may well be to assume that each clutter type is equally likely to occur, and that all misclassifications are equally costly. The latter leads to the *zero-one* loss function

$$\lambda(\alpha_i \mid c_j) = 1 - \delta_{ij} = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases} \qquad \text{all } i, j \quad (39)$$

which, when substituted in (36) yields

$$R(\alpha_i \mid \mathbf{y}) = \sum_{j \neq i}^{L} P(c_j \mid \mathbf{y}) = 1 - P(c_i \mid \mathbf{y}). \quad (40)$$

This risk is now effectively an error rate, and minimizing this error rate is seen to be equivalent to maximizing the a posteriori probability as in (33) *(maximum a posteriori probability* (MAP) *classifier)*. Further simplifying by setting $P(c_i) = 1/L$ for all i in (34) gives

$$(R(\alpha_i \mid \mathbf{y}) - 1)p(\mathbf{y})L = -p(\mathbf{y} \mid c_i) \quad (41)$$

and provides a classification rule based solely on the class-conditional probability density function. Equation (41) leads logically to the concept of discriminant functions, which may be used instead of the risk function itself. The classification rule then assigns feature vector $\mathbf{y}$ to class c_j if

$$g_j(\mathbf{y}) > g_i(\mathbf{y}), \qquad \text{all } i \neq j. \quad (42)$$

In the most general case $g_i(\mathbf{y}) = -R(\alpha_i \mid \mathbf{y})$ and, for the case of equally likely classes and equally costly misclassifications, $g_i(\mathbf{y}) = p(\mathbf{y} \mid c_i)$. In either case, the structure of the classifier is determined by the conditional densities $p(\mathbf{y} \mid c_i)$. For computational purposes a simple structure is desirable, especially in multidimensional, multiclass problems.

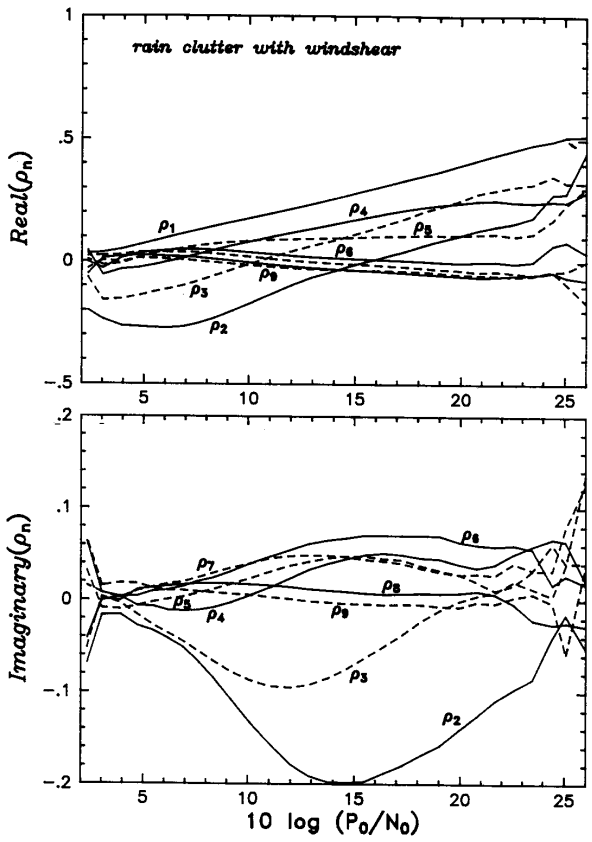

Fig. 9(e). Mean feature values for rain clutter with windshear.

A simple, and for many situations a fairly realistic density function is the normal, or Gaussian, density. Its multivariate form is written as follows:

$$p(\mathbf{y}) = \frac{1}{(2\pi)^{d/2} \mid \sum \mid^{1/2}} \cdot \exp\left[-(1/2)(\mathbf{y}-\mu)^t \sum^{-1}(\mathbf{y}-\mu)\right] \quad (43)$$

where

$$\mu = E[\mathbf{y}] \quad (44)$$

$$\sum = E\left[(\mathbf{y}-\mu)(\mathbf{y}-\mu)^t\right]. \quad (45)$$

Note that the form of this probability density function is completely described by the mean vector μ and the covariance matrix Σ. Defining the discriminant function

$$g_i(\mathbf{y}) = \ln[p(\mathbf{y} \mid c_i)] \quad (46)$$

$$g_i(\mathbf{y}) = -(1/2)(\mathbf{y}-\mu_i)^t \sum_i^{-1} (\mathbf{y}-\mu_i) - (d/2)\ln(2\pi) - (1/2)\ln \mid \sum_i \mid \quad (47)$$

and removing constants yields a new discriminant function

$$g_i'(\mathbf{y}) = (\mathbf{y}-\mu_i)^t \sum_i^{-1} (\mathbf{y}-\mu_i) + \ln\left|\sum_i\right|. \quad (48)$$

The change of sign requires $g_i(\mathbf{y})$ to be minimized; that is, $\mathbf{y}$ is assigned to class c_j whenever

$$g_j'(\mathbf{y}) < g_i'(\mathbf{y}), \qquad \text{all } i \neq j. \quad (49)$$

The quantity $D_i^2(\mathbf{y}) = (\mathbf{y}-\boldsymbol{\mu}_i)^t \sum_i^{-1}(\mathbf{y}-\boldsymbol{\mu}_i)$ is also known as the squared *Mahalanobis distance* from $\mathbf{y}$ to $\boldsymbol{\mu}_i$. The determinant $|\Sigma_i|$ is the product of the eigenvalues of Σ_i, and as such a measure of the combined variances of the marginal distributions of $\mathbf{y}$, or the "volume" inside the multivariate distribution's equiprobability contours. Adding the term $\ln|\Sigma_i|$ to $D_i^2(\mathbf{y})$ thus biases the discriminant toward the denser, lower variance classes.

B. Experimental Results

The number of possible classification and testing scenarios is extremely large and cannot possibly be fully explored in this paper. Hence, only a representative sample of results

Table 1 Experimental 6-Class Classification Results

labeled samples:	birds	thunder storms	rain	aircraft	cars,trucks	targets	total
prototype name	(58 281)	(6766)	(19 168)	(5685)	(2483)	(10 385)	(102 768)
rejected	3.7	0.0	0.1	5.1	3.1	0.1	12.1
birds	68.3	2.9	4.6	0.5	5.4	0.1	81.7
thunder storms	9.9	75.6	27.1	5.5	7.8	0.7	126.6
rain	6.6	12.5	64.0	0.2	1.6	0.1	84.9
aircraft	4.2	4.5	1.7	26.7	14.8	2.5	54.5
cars and trucks	6.0	1.0	1.6	4.6	47.1	2.1	62.4
synthetic targets	1.3	3.5	0.9	57.3	20.3	94.4	177.7
total	100.0	100.0	100.0	100.0	100.0	100.0	600.0

labeled samples:	birds	weather	targets	total
prototype name	(58 281)	(25 934)	(8168)	(92 383)
rejected	3.7	0.1	4.5	8.3
birds	68.3	4.2	2.0	74.4
weather	16.5	90.3	6.8	113.6
targets	11.6	5.4	86.7	103.7
total	100.0	100.0	100.0	300.0

is presented. One of the difficulties is the division of classes into subclasses. For example, a weather clutter class may be subdivided into rain (little wind or turbulence) and storms (thunderstorms, systems with considerable turbulence), and perhaps even rain with a detectable windshear component. Clearly, there will be significant overlap between those classes, but misclassification among them is not serious. In terms of the ability to discriminate between major classes, such as bird clutter, weather clutter, and point targets (aircraft, cars and trucks), it is difficult to predict which scenario is preferable. If each subclass provides a "tight" prototype, then the combined decision boundaries might be expected to conform to the true cluster shape more closely and result in a better overall performance. On the other hand, the greater number of classes also provides more opportunities for errors to occur and the classifier may become biased toward that major class for which the greatest number of subclasses has been defined. A greater number of classes also requires more training samples to maintain the quality of each of the estimates. Subdividing a fixed number of training samples among more class prototypes will reduce the quality of each.

Table 1 shows the percentage of labeled samples (columns) assigned to each class (rows). The sum of the row percentages is a measure of the bias toward the class at the expense of another. The individual class assignments are listed in the upper table. The results of combining these assignments (excluding those for synthetic targets) into major classes is shown in the lower table. The mean classification accuracy is then 81.8%, the standard deviation of the biases is 16.6%. Table 1 presents classification results using samples from an SNR range of 17 to 24 dB. The set of eleven features used in the classification were as follows: u_0, $Re[\rho'_2]$, $Im[\rho'_2]$, $Re[\rho'_3]$, $Im[\rho'_3]$, $Re[\rho'_4]$, $Im[\rho'_4]$, $Re[\rho'_5]$, $Im[\rho'_5]$, P_{var}, P_{dif}. Six classes were defined and tested against over 100 000 labeled samples. In the case of only a small number of samples being available (such as for the car and truck class), all samples were used for both training and testing and the results must therefore be examined with caution. This type of testing the training samples has been shown to be always optimistic (Foley [56]), and researchers in pattern recognition have long suggested the use of the so-called *leave-one-out* method of error rate estimation [34]. This method requires as many

classifiers to be trained as there are samples to be tested and is, for obvious reasons, not practical for the large volumes of radar data. In addition, Glick [57] pointed out that the leave-one-out method, while being an unbiased estimator, has a rather large variance. Despite the optimistic results, testing the training data is useful when the number of samples is large. Testing a separate test data set, which was not used for classifier training, provides pessimistic results and can be used when the sample set is large enough to allow such subdivision without sacrificing the quality of the class prototype. Foley [56] showed that both methods converge for large sample sizes, and for sample size-to-dimensionality ratios greater than 10 the difference is small. Ratios used for these experiments exceeded this value by a large margin. The relatively small number of samples available for the target classes (approximately 700 aircraft and 200 to 300 cars and trucks resulting in 21 384 and 7447 feature vectors, respectively) has been partially compensated for by the inclusion of the synthetic target class (53 999 feature vectors arising from approximately 2700 targets), and the assignment of erroneously classified samples to related classes enhances the confidence that the prototypes, as developed, are substantially correct. A better measure of the quality of the prototypes is to combine the classifier decisions into major classes. This results in an average percentage of correct classification of 81.8%, which should be considered quite satisfactory for a *single scan, single resolution cell decision.*

Another measure of the quality of the prototypes is the bias toward one class at the expense of another. If no one class is favored by the classifier, then every class is expected to receive an equal number of assignments, assuming of course that each labeled sample set also contains an equal number of samples. Of course, the samples must be distributed according to the multivariate normal density with which the prototype was constructed, and the *a priori* probabilities of each class occurring must be equal. Summing the percentage of assignments from all sample sets into a given class gives such a measure for that class, and would take on an equal value for each class if the classifier is unbiased (this value would be 100% only if no rejection occurred). To obtain a measure from the entire classifier, one might compute the standard deviation of all bias measures, which is desired to be as small as possible.

The results in Table 1 show a slight bias toward the weather class at the expense of the bird class. The bias toward synthetic targets in the upper table is of less concern since it stems mostly from the similar classes of aircraft and cars and trucks. The source of this bias is not easy to identify, since it is related to how the actual sample statistics deviate from the multivariate normal density and the relative "location" of the prototype in the feature space. While such a bias is undesirable, it is not clear what should be done about it. The final classifier design, which will be based on a complete risk analysis, must take such biases into account when considering the costs of misclassification. Minimizing the total cost with respect to the experimental performance will introduce its own biases to counteract the ones currently present. Accordingly, no further action toward eliminating these biases was taken in this research.

The table shows the percentage of labeled samples (columns) assigned to each class (rows). The sum of the row percentages is a measure of the bias toward the class at the expense of another. The individual class assignments are listed in the upper table. The results of combining these assignments (excluding those for synthetic targets) into major classes is shown in the bottom table. The mean classification accuracy is then 81.8%, the standard deviation of the biases is 16.6%. All data have been taken from an SNR range of 17 to 24 dB.

VIII. Neural Network Classifier Using The Back-Propagation Algorithm

In this section, we present results of ATC radar clutter classification using a multilayer artificial neural network. Artificial neural networks (or neurocomputers) are an alternate model of computation based on design principles derived from biological neural systems. In many respects, neurocomputers are superior to conventional (von Neumann type) computers (Lippmann [61]; Sejnowski and Rosenberg [62]). The motivation for taking the neural network approach to solve the radar clutter classification problem is based on three principal reasons: 1) a neural network has an intrinsic ability to generalize; 2) it makes weaker assumptions about the statistics of the input data than a parametric Bayes classifier; and 3) a neural network is capable of forming highly nonlinear decision boundaries in the feature space and therefore has the potential of outperforming a parametric Bayes classifier when the feature statistics deviate significantly from the assumed Gaussian statistics.

In this section, we first present an introduction to the popular back-propagation algorithm (Rumelhart *et al.* [63]) which is used to train the multilayer neural network in an off-line manner. Then we summarize the design considerations, and experimental results obtained with a neural network classifier based on this algorithm.

A. Multilayer Neural Network

A multilayer neural network consists of an input layer, one or more hidden layers, and an output layer. Each layer (except for the input layer) is comprised of a number of "neurons" or processing units, each one of which consists of a linear combiner and a nonlinear device as in Fig. 10(a). The linear combiner uses "synapses" of adjustable "strengths" or weights. A typical 4-layer neural network is depicted in Fig. 10(b).

In the study reported in this paper, two specific neural network configurations are considered: a 3-layer network (input layer, hidden layer, and output layer), and a 4-layer network (input layer, two hidden layers, and output layer).

Ordinarily, a sigmoidal logistic function is used to describe the input-output relation of the nonlinear device. Thus in mathematical terms, the behavior of neuron j is

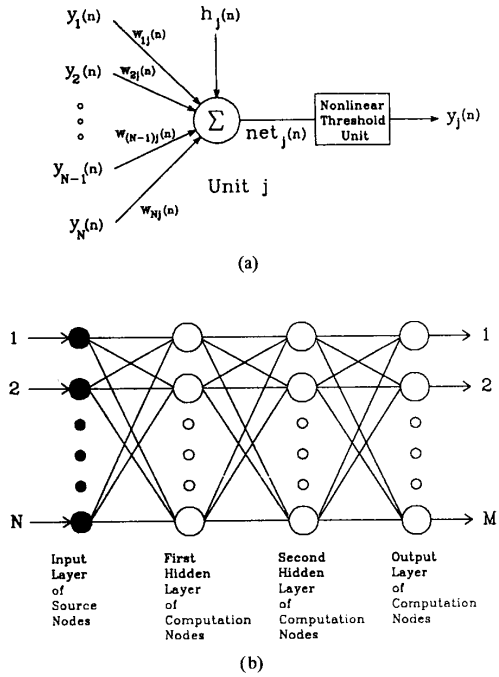

Fig. 10. Architecture of a layered neural network. (a) Neuron or processing unit in the network. (b) Four-layer neural network.

described by the pair of relations:

$$net_j(n) = \sum_i w_{ij}(n)y_i(n) + h_j(n) \tag{50}$$

and

$$y_j(n) = f_j(net_j(n)) = \frac{1}{1 + \exp\{-net_j(n)\}} \tag{51}$$

where n refers to presentation of the nth input pattern; w_{ij} is the weight from the i-th neuron to the jth neuron; y_i is the ith input to the jth neuron; h_j is the bias of the jth neuron; net$_j$ is a weighted sum obtained from the linear combiner; $f_j(\ )$ is the nonlinearity function; and y_j is the output from neuron j.

B. Training of the Neural Network

In order to apply neural network techniques to solve a pattern classification problem, an algorithm must be used to train the network. The *back-propagation algorithm* is one of the most popular and yet very powerful learning algorithms available for this function. This algorithm was first developed by Werbos [64], and later rediscovered by several other researchers independently (Parker [66]; Rumelhart *et al.*, [63]; LeCun, [65]). In the following, we briefly describe the back-propagation algorithm.

There are two phases involved in back-propagation learning. During the first phase, inputs are presented to the network, which propagate forward through the network to produce an actual output $y_j(n)$ for each neuron in the output layer. The activity of each neuron in the network is determined by (50) and (51). We then compare a desired response $d_j(n)$ with the corresponding output $y_j(n)$ if neuron j is in the output layer, to generate an error signal $e_j(n)$:

$$e_j(n) = d_j(n) - y_j(n). \tag{52}$$

In the case of pattern classification, the desired responses $\{d_j\}$ are defined to be the encoded class labels. For example, for a neural network with three output neurons representing the three different classes, we assign $\{d_j\} = \{1,0,0\}$ for the first class, $\{d_j\} = \{0,1,0\}$ for the second class, and $\{dj\} = \{0,0,1\}$ for the third class.

We may thus define an index of performance $\epsilon(n)$ (for nth input pattern):

$$\varepsilon(n) = \frac{1}{2}\sum_j [d_j(n) - y_j(n)]^2. \tag{53}$$

This index of performance is minimized by the weight update rule.

During the second phase, the error signals (computed in the output layer) propagate backward through the network to allow the recursive computation of the weight updates. The essence of the back-propagation algorithm is in the second phase. Applying the "chain rule" to differentiate (53) with respect to $w_{ij}(n)$, we get

$$\frac{\partial\varepsilon(n)}{\partial w_{ij}(n)} = -\delta_j(n)y_i(n) \tag{54}$$

where $\delta_j(n)$ is the "modulated" error signal. It can be shown (Haykin [68]) that for the *output neurons*, the "modulated" error signal is

$$\delta_j(n) = f_j'(net_j(n))e_j(n) \tag{55}$$

whereas for the *hidden neurons* we have

$$\delta_j(n) = f_j'(net_j(n))\sum_k \delta_k(n)w_{jk}(n). \tag{56}$$

The index k refers to the kth neuron in the layer above the one where neuron j is located.

For the network to learn, the generalized delta rule [63] is applied to update the network weight $w_{ij}(n+1)$, as shown by

$$w_{ij}(n+1) = w_{ij}(n) - \eta\frac{\partial\varepsilon(n)}{\partial w_{ij}(n)} = w_{ij}(n) + \eta\delta_j(n)y_i(n) \tag{57}$$

where η is the learning rate (or adaptation constant) that controls the step size of the weight updates. In general, a momentum term is also included in the above weight update rule to filter out high-frequency variations of the error surface in the weight parameter space [63]. Thus we have

$$w_{ij}(n+1) = w_{ij}(n) + \eta\delta_j(n)y_i(n) + \tau\Delta w_{ij}(n-1). \tag{58}$$

The back-propagation algorithm that uses the weight update rule in (58) is referred to as the "conventional" back-propagation method. To start the back-propagation algorithm, the weights and biases are initialized to small random values uniformly distributed between -0.5 and 0.5. The learning rate η and momentum factor τ are small constants whose values lie between 0 and 1.

Although the weight update rule in (58) has been widely used in training multilayer networks, it is often found that its rate of convergence is too slow to be useful for many practical applications. The *modified* form of the back-propagation algorithm applied in this paper combines the learning rate adaptation method (Jacobs [69]) with the gradient reuse method (Hush and Salas [70]). This modified method improves the rate of convergence of the back-propagation learning process, and yet reduces the computational complexity arising from using the learning rate adaptation method.

In the modified back-propagation method, every weight w_{ij} in the network is given its own learning rate η_{ij}, and the training data set is divided into a number of epochs each containing K training patterns. The weight w_{ij} and learning rate η_{ij} are updated every time after an entire training epoch has been presented to the network. The weight and learning rate updating rules of the modified back-propagation algorithm can be summarized as follows [69], [71]

$$w_{ij}(n+1) = w_{ij}(n) + \eta_{ij}(n+1) \sum_{k=1}^{K} \delta_{kj}(n) y_{ki}(n) + \tau \Delta w_{ij}(n-1) \tag{59}$$

$$\eta_{ij}(n+1) = \eta_{ij}(n) + \Delta \eta_{ij}(n) \tag{60}$$

$$\Delta \eta_{ij}(n) = \begin{cases} \Omega, & \text{if } S(n-1)D(n) > 0 \\ -\phi \eta_{ij}(n), & \text{if } S(n-1)D(n) < 0 \\ 0, & \text{otherwise} \end{cases} \tag{61}$$

$$D(n) = \sum_{k=1}^{K} \frac{\partial \varepsilon_k(n)}{\partial w_{ij}(n)} \tag{62}$$

$$S(n) = (1-\Theta)D(n) + \Theta S(n-1). \tag{63}$$

The index n refers to the nth epoch in the training data; the index k refers to the kth pattern in an epoch containing K patterns; δ_{kj} is the modulated error signal of neuron j with the kth pattern in an epoch; y_{ki} refers to the actual computed output of neuron i with the kth pattern in an epoch; ε_k is the index of performance to be minimized by the weight update rule with the kth input pattern; finally, Ω, Φ, and Θ (all of which have values between 0 and 1) are the control parameters.

C. Experimental Results

As indicated in Section VI, a prediction-error filter of order 5 is used as the feature extractor. Following the terminology described in that section, the computed features are summarized as: $[uo,\ Re\{\rho'_2\},\ Im\{\rho'_2\},\ Re\{\rho'_3\},\ Im\{\rho'_3\},\ Re\{\rho'_4\},\ Im\{\rho'_4\},\ Re\{\rho'_5\},\ Im\{\rho'_5\},\ P_{var},\ P_{dif}]$, ϕ, and U_o. The eleven features inside the square brackets are used to form the input vector of the neural classifier; hence the input layer of the multilayer network is composed of eleven units.

With three major categories of radar clutter of interest, namely, weather, birds, and ground, we observe that the first two classes are due to moving objects, whereas the last one is due to objects that are essentially stationary. Moreover, we note that the heterodyned reflection coefficients of ground clutter are of a similar nature to those of aircraft targets, as evidenced by the results presented in Section VI and additional results presented in (Stehwien [42]). Hence, no distinction is made between ground and aircraft targets in the neural classifier training phase. The design of the neural classifier using this prior information about the input data may thus be summarized as follows:

1. The neural classifier is trained to distinguish the three moving object classes, aircraft targets, weather, and birds, based on the output of the feature extractor. The computed Doppler frequencies (i.e., the values of ϕ) are disregarded during the training phase.
2. Testing data sets, which may contain the ground class, are fed into the feature extractor. The computed features are passed to the classifier, which will classify the input into a) birds, b) weather, or c) target class. The class label is determined by recognizing the output neuron which generates the largest output value among others.

We may identify the ground class of radar returns by recognizing the small value of the mean Doppler frequency ϕ of such returns. This observation is similar to the MTI principle. We recall that ϕ equals the phase of the first reflection coefficient ρ_1. Specifically, we may feed the classified results (encoded labels), together with their mean Doppler frequencies, into a post-processing unit. If the mean Doppler frequency ϕ exceeds a prescribed threshold ($\pm 0.025\ f_s$, say, where f_s is the PRF), the post-processing unit simply passes the input label to the output. Otherwise, it sends a ground label as output. Clearly, this form of post-processing may be used with a Bayes or neural network classifier.

1) Formation of training and testing data sets: There are two different ways of forming the training and testing data sets. One way is to select feature patterns randomly from available feature data base described in Table 1 in Section VII. We refer to this first way as the *shuffled data set*, which is summarized in Table 2. The other way is merely to split the available feature files into two parts, and to use the first part for training and the second part for testing. We refer to this second way as the *split data set*.

Clearly, these two ways of forming the training and

Table 2 The Training and Testing Data Sets (Shuffled)

Class	(subclass)	Training	Testing
BIRDS:		6000	6000
WEATHER:	(storm)	3000	3000
	(rain)	3000	3000
TARGET:	(aircraft)	2000	2000
	(cars)	2000	480
	(synthetic)	2000	2000
Total:		18 000	16 480

Table 3 Classification Results on Training Data using a Bayes Classifier

	bird	storm	rain	aircraft	car	synthetic
Original	68.3	75.6	64.0	26.7	47.1	94.4
Shuffled	70.6	74.9	62.2	29.4	49.7	95.2
Split	71.2	**64.9**	60.3	24.6	48.1	92.5

testing data will result in different statistical properties for the input into the neural network classifier, because of the limited number of feature patterns available. In order to demonstrate this statistical variation, a Bayes classifier similar to the one in Section VII was used to test the training data sets, with the results being summarized in Table 3. Here we observe that the shuffled data set exhibits almost an identical classification performance as the original data set of Table 1. On the other hand, because the split data set comes from an ordered subset of the original data base, it is unable to characterize some of the statistics of the original data set. In particular, a large variation in the storm subclass can be observed.

Therefore, in order to select a subset from the original data, which carries nearly identical statistical properties as in the original one, the shuffled data set should be used. Accordingly, the comparison between the neural network classifier and the Bayes classifier is based on this shuffled data set. Also, the training and test data sets used for this neural network study were all selected from the SNR range of 17–24 dB.

2) Network configurations: Neural network theory, in its present form, is inadequate when it comes to specifying the desired configuration of a neural network classifier for a prescribed set of input features. If the network is too small, it may be unable to capture the statistical properties of the input data. On the other hand, if the network is too large, it may learn details of the undesired noise background in the input data, and thus its ability to generalize is compromised. The common practice is to use an experimental approach for specifying a satisfactory network configuration. Indeed, it is for this reason that we try to reduce the time required for network training as far as possible, by using a parallel computer for training, an accelerated convergence technique, or both.

Table 4 Classification Results (11–24–12–3 Network) on Testing Data (Shuffled)

Labeled sub-class	No. of patterns in testing data	Classification rate (%)			
		Bird	Weather	Target	Average
Birds	6000	**84.4**	10.8	4.8	84.4
Storm	3000	5.6	**89.3**	5.1	
Rain	3000	5.5	**93.1**	1.4	91.2
Aircraft	2000	2.1	1.7	**96.2**	
Cars	480	6.5	3.7	**89.8**	94.9
Synthetic	2000	0.4	0.3	**99.3**	
					90.2

In any event, through a series of experiments involving different network sizes, it was discovered that the 4-layer network with size 11–24–12–3 delivers the best generalization performance. For the sake of comparison, a 3-layer network with size 11–42–3 was constructed, which has a similar memory capacity to the 11–24–12–3 network. Convergence was also achieved for this 3-layer network, but the classification accuracy was lower than the corresponding 4-layer network. Learning curves for these two networks with the conventional back-propagation method, and for the 11–24–12–3 network with the modified back-propagation method, are shown in Fig. 11. In this figure, the total squared sum of errors (TSSE) is plotted against the number of training sweeps, which is the number of passes through the training data set. These learning curves confirm that the modified method indeed converges faster than its conventional counterpart.

For the conventional back-propagation method, the learning rate η is chosen to be 0.01. The momentum factor τ is fixed at 0.1. For the modified back-propagation method, the initial learning rates $\eta_{ij}(0)$ are all set to be 0.05. The control parameters Ω, Φ, and Θ are chosen to be 0.002, 0.1, and 0.6, respectively.

In Table 4, we summarize the clutter classification performance of the 11–24–12–3 network. From this table, the misclassification of each clutter type can be observed. We feed the classifier with feature data from six subclasses, and classify them into one of three major classes; a) birds, b) weather, and c) target. An average classification rate of 90.2% is observed.

Comparing Table 4 to Table 1, we also observe an improvement in the classification accuracy of the birds and target classes. This improvement may be due to the fact that the feature vectors representing these two classes deviate markedly from the multivariate Gaussian assumption. Hence, the advantage of the neural network classifier, which is capable of forming highly nonlinear decision boundaries in the feature space, pays off in improving the accuracy of classifying radar returns from birds and targets. For

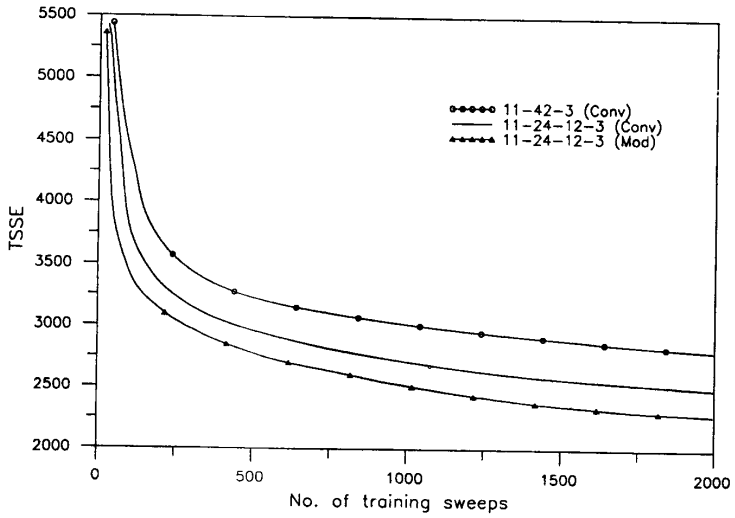

Fig. 11. Learning curves for multilayer networks with different sizes and different forms of the back-propagation algorithm.

the weather class, on the other hand, we observe that only a small gain is made by using the neural network classifier. This suggests that the feature vector representing the weather class may be modeled closely as a multivariate Gaussian distribution.

IX. Discussion And Conclusions

The motivation for radar clutter classification in an air traffic control environment is to identify hazardous areas of weather storms and flocks of migrating birds, and thereby vector aircraft around them. In this paper, we have presented an experimental study of the radar clutter classification problem in such an environment using a ground-based coherent radar system. The study is based on real-life data collected at operational radar sites. The important findings of the study may be summarized as follows:

1) The use of reflection coefficients, computed using the multisegment Burg algorithm, provides a practical method for the extraction of radar features based on second-order statistics.

2) A direct comparison of the performance of the Bayes classifier presented in Section VII-B with that of the neural network classifier presented in Section VIII-C is rather difficult to make for two reasons. First, in the case of the Bayes classifier, an algorithm was included in its design to reject data that clearly did not belong to any of the classes under consideration. Second, the neural network classifier included the use of synthetic targets, which had the effect of raising the classification rate of targets. Nevertheless, these two independent studies of radar clutter classification tend to support each other in two important aspects:

a) They both confirm the utility of the reflection coefficients in describing the second-order statistics of radar returns.

b) They demonstrate that there is sufficient information content in the different types of radar clutter to enable their classification. Moreover, the extensive PPI images reported in [42] and [71] show that both the Bayes classifier and the neural network classifier are sufficiently robust for operational use.

The superior performance of the neural network classifier over the parametric Bayes classifier reported herein should not be taken as a proof. Rather, it should be viewed as a practical demonstration of the potential value of neural networks as a new tool for the classification of radar clutter.

3) In training the neural network classifier, no distinction is made between radar echoes for aircraft targets and ground, because their heterodyned spectral parameters have similar characteristics. Discrimination between these two classes of radar returns is made on the basis of the mean Doppler frequency represented by the phase of the first reflection coefficient. This procedure is similar in philosophy to the well known MTI principle.

4) The results of classification performance, presented for both the Bayes classifier and the neural network classifier, were based on data collected by a coherent radar on a single antenna scan. Clearly, the performance of a classifier, be it of a Bayes or neural network type, can be improved by processing data collected on successive scans of the antenna.

The extraction of radar features, exemplified by the reflection coefficients and power-related parameters, are based entirely on second-order statistics of the radar data. Yet, useful information is contained in higher-order statistics. Here we recognize that radar echoes, particularly those from aircraft and ground, exhibit significant deviations from a Gaussian distribution. It is therefore anticipated that by adding features based on higher-order statistics, a further improvement in clutter classification performance may be

attainable. This observation has indeed been confirmed [72].

The coherent radar used to collect the database that was used to perform the clutter classifications reported herein was operated at L-band frequencies. Although the radar had provisions for both horizontal and vertical polarizations (see Appendix A), unfortunately (due to unforseen difficulties) only one channel was available for the data collection. It is quite likely that the radar classification theory described in this paper may yield better results if there were access to a weather radar with an operating frequency (e.g., C-band) better suited for the remote sensing of weather systems, or if there were access to different polarization channels. Indeed, it would be most interesting to apply our theory to data collected from the new generation of weather radars, NEXRAD [73] and TDWR [74], which may have more suitable parameters.

Finally, we should mention that the Bayes and neural network classifiers as described herein could be used on MTD radars that employ fixed PRF's during each coherent processing interval (CPI). Feature extractors designed for use with staggered PRF's are more complicated due to the use of nonuniform sampling [42]. Staggered PRF's are intended to achieve specific functions in conjunction with MTI processors.

Appendix A
TRACS-ASR Specifications

The radar data used throughout this research were recorded from two of the new L-band air traffic control area surveillance radars (ASR) which form part of the Terminal Radar and Control Systems (TRACS) installed by the DND at six Canadian Forces Bases.These two radars, which are derivatives of the Westinghouse ARSR-3, were located at CFB Trenton and CFB Moose Jaw. Other TRACS components are a SSR, a digital target extractor and correlator, a computer control system and an instrument flight rule control center (IFRCC) with fully synthesized digital displays at the controller stations. The system was designed to modernize air traffic control at Canada's busiest military airports and terminal areas. Since it was not meant for en route surveillance, which is handled by the Ministry of Transport, the original 200 nautical mile range of the ARSR-3 was reduced to 80 nmi, with a corresponding increase in the PRF range from 310–365 Hz to 650–675 Hz, an increase in antenna scan rate from 5 to 12 rpm, and a decrease in peak power from 5 to 1.5 MW.

The radar system is fully redundant, with each channel operating on a different frequency and polarization. Frequency and polarization diversity can thus be achieved during diplex operation. Circular polarization improves detection in rain, and the combination of both left and right hand polarized RF signals provides a weather channel. Target reports are correlated with SSR returns and sent independently to the central computer (a Data General Nova System), where they are combined and displayed on the controller's display. Unfortunately, only one channel was operational at the time of the recordings and it was not possible to take advantage of the diversity feature. The following characteristics of the radar therefore refer only to the channel from which the data was taken.

Site:	Moose Jaw	Trenton
Transmitter		
frequency of operation (L-band)	1343.784 MHz	1307.538 MHz
wavelength	22.305 cm	22.924 cm
transmitter output power	1.5 MW min	
transmitter type	tunable klystron	
pulse width	2 μs	
PRF (if fixed), selectable	657.8, 674.3 Hz	
Antenna		
polarization (if linear)	horizontal	vertical
polarization (if circular)	right hand	left hand
antenna feed	dual beam, switchable	
gain (low beam)	33 db min	
gain (high beam)	31 dB min	
scan rate (nominal)	12 to 12.1 rpm	
azimuth beamwidth (one-way, 3 dB)	1.5°	
hits per beamwidth (approximate)	13.5 to 14	
azimuth sidelobes	22 dB min	
elevation beamwidth (one-way, 3 dB)	4°	
elevation beampattern	modified $\csc^2$	
height above ground (electrical center)	22 m (73 ft)	
Receiver		
STC (programmable) maximum attenuation	63 dB	
receiver noise figure (low beam)	4 dB max	
receiver noise figure (high beam)	3.5 dB max	
bandwidth	500 kHz	
dynamic range	80 dB min	
video output	coherent I and Q	

Radar Video Data Parameters	
radar visible range	up to 120 nmi
instrumented range	80 nmi
A/D resolution	10 bits (9+sign)
digital video dynamic range	46.3 dB 49.6 dB
sampling frequency	1294.6 kHz
range sampling interval	1/16 nmi (116 m)
azimuth sampling interval	0.088°
azimuth accuracy	within 0.264°
Signal Processing Features	
MTI	4 pulse I and Q
MTI improvement factor	50 dB
integrated cancellation ratio	18 dB
LOG CFAR (post MTI)	8 sample (1.3 nmi)
pulse-to-pulse integrator	single pole feedback
detector	adaptive threshold

Appendix B
Evidence For Bird Clutter

It was found extremely difficult to positively identify "bird clutter" as having actually been caused by migrating birds. No direct visual sightings of any flock was made that could be correlated with a target on the PPI of the radar. This difficulty was compounded by the cone of silence above the radar, which caused any target tracking inbound to disappear from the radar at about 3 to 6 miles. Positive identification of bird echoes during radar studies has certainly never been easy (Barry *et al.* [12]; Blokpoel [75]), and repeated attempts to sight the flocks at the estimated time of overflight failed. A part of the problem may well have been the small size of the birds against the night sky, another the lack of night vision glasses.

There is, however, a considerable amount of other evidence available. With respect to the radar echoes, speed and direction of travel (40 to 80 kn, north-west) coincided with what may be reasonably expected from migrating birds. The characteristic fluctuations in return strength were always present making it in fact rather difficult to track any one flock for long distances. Since most large geese had already passed through the area by the time of the trip, only small birds such as ducks, coots, sandpipers, and passerines would have made up the majority of these flocks. Consequently, few single bird echoes would likely be detected, and no targets were specifically identified as such at the time of recording.

The strength of the returns together with their apparent flight altitudes also pointed to the clutter being caused by birds. Time exposure photographs taken following the recording sessions showed large numbers of tracks of

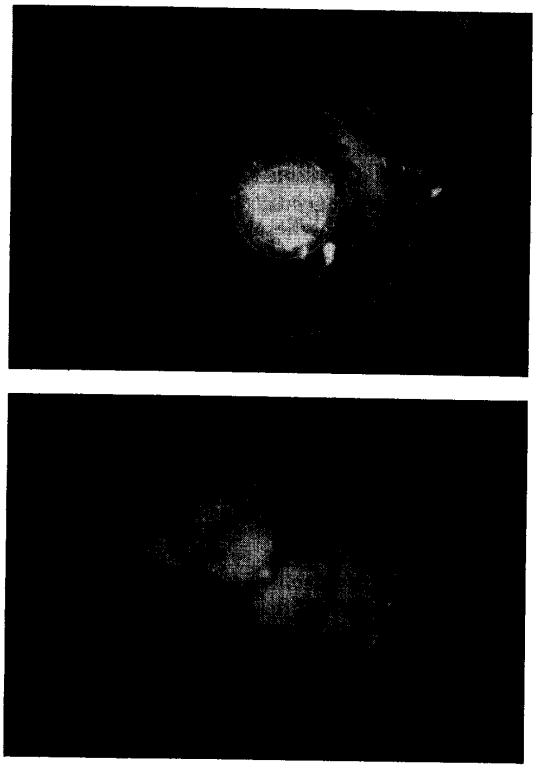

Fig. 12. PPI time exposure photographs of bird clutter.

targets traveling in the north-west direction (Fig. 12). The cone of silence is very apparent, as is the notch in the STC curve at 24 miles. The number of targets detected in these ranges relative to the number elsewhere (less than one half or even one third) indicates that the RCS of most targets was probably between 50 and 200 cm^2. This coincides well with the expected radar cross-sections for migrating birds and bird flocks of 10 to 500 cm^2 [6], [22].

Flight altitudes depend largely on the weather, particularly wind and cloud cover, but are expected to be mostly below 10 000 ft. (Richardson [76], [77]; Blokpoel [75]; Blokpoel and Burton [78]; Blokpoel [8]). Although south-east winds were sometimes forecast for high altitudes, such favorable winds were mostly found only at the lower altitudes. The prevailing winds in Moose Jaw are westerlies, and south-east winds at ground level frequently turn to become westerly at high altitudes. Hence, the largest concentration of birds were expected to fly within the first few thousand feet above the ground. The photographs show a rather high density of clutter inside 20 miles, and suggest that the flight altitudes were indeed below 8000 ft agl. The poor visibility of the targets to the west, where a marked radar shadow blocks low altitude targets, leads to the additional conclusion that the great majority of birds must have been traveling below 3000 ft agl and in large flocks (200 cm^2) at most times. On the other hand, the

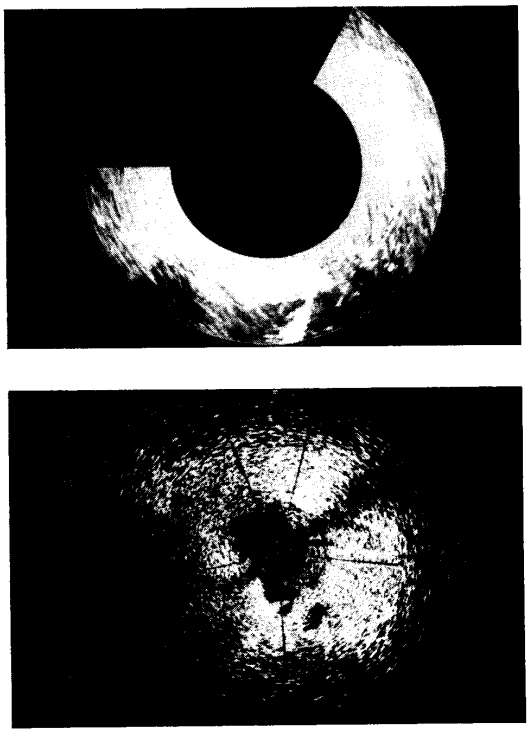

Fig. 13. Simulated PPI images of processed radar data including bird clutter.

high densities present beyond 30 nmi require flight altitudes of at least 1000 to 4000 ft agl depending on target size (unless anomalous propagation conditions exist). Indeed, some strong targets were seen out to 50 miles to the south, where high terrain produces a rather pronounced radar shadow. This implies that some flocks, at least, were traveling as high as 6000 to 8000 ft agl. In those cases, however, direction of travel tended to differ from those at lower altitudes because of the wind shift. At 55 to 60 miles most targets were found to disappear, with only the largest flocks (200 cm^2 and larger) visible beyond 65 miles.

Compelling evidence for the "bird clutter" is also the fact that it was never seen on unfavorable nights. Yet it was always visible on favorable nights, except when obscured by strong inversion related clutter. This clutter was at times so strong that considerable sidelobe "ringing" was generated, which looked very similar to bird clutter. The time exposure photographs did, however, subsequently show the presence of bird target tracks even under these conditions.

Analysis of data from the tapes also pointed toward bird clutter. Performing scan-to-scan integration over one to five minute segments (up to 64 scans) showed the same tracks seen on the time exposure photographs. From these, precise directions of travel could be determined. SNR's ranged from virtually undiscernible to 15 dB when STC was turned on. Clutter strengths of up to 35 dB were found when turned off. Doppler frequencies ranged from zero to up to 400 Hz, corresponding to an upper speed of travel of about 85 kn. Airspeeds are generally not this high (Blokpoel [75]; Richardson [79], [80]), but groundspeeds of this magnitude are possible during periods of favorable winds (Barry *et al.* [12]; Muehe *et al.* [21]). The average direction of travel was also discernible as the bearing from the radar where the minimum Doppler frequency occurred (most negative, that is the birds are flying away from the radar). This was generally around 325° true, and coincided well with what was expected. It is interesting to note that CFB Cold Lake (another TRACS site) is located 300 nmi to the north-west (at a bearing of 326.5° true) and in the same general direction as the birds were found to travel (in fact, the direction of travel evident from Fig. 13 is 326°!). Perhaps none of these arguments constitute a "proof" of the observed clutter being caused by birds, but it surely is strong evidence. An alternate explanation for its source is certainly not immediately apparent.

Nonradar evidence of migration activity came from several sources. A number of telephone conversations were held with biologists at the Canadian Wildlife Service office in Saskatoon. The CWS does not actively monitor the spring migration (unlike the fall migration, due to hunting activity), but information was obtained on the most likely species to have migrated at that time, including their flyways and staging areas. DND itself does, however, monitor and predict bird migration activity at CFB Cold Lake using a long-range high power *L*-band radar (not the TRACS-ASR). Predictions are made using a technique devised by the CWS (Blokpoel [9]; Blokpoel and Gauthier [81]), and is based on such factors as date, forecast upper wind directions, air temperatures, cloud cover and precipitation. Migration activity is confirmed with polaroid photographs of the PPI taken hourly and exposed for ten minutes. Experience with prediction success rates is then taken into account to refine new predictions. Periodic comparisons between the Cold Lake predictions and those made for Moose Jaw showed very similar conditions, which was not unexpected due to their relative proximity. Even though there are staging areas between these two air bases, some species of birds fly continuously for as long as 8 to 10 hours and could pass over both locations the same night. Polaroid photographs taken on two particularly heavy clutter nights for Moose Jaw confirmed the presence of birds at Cold Lake as well.

Finally, the presence of newly arrived birds was confirmed after a strong migration night, when numerous ducks and many other bird species were spotted in virtually every slough and pothole in the area south of Moose Jaw. While the ducks would not likely have migrated further, the large number of swans (approximately 1000) sighted on Old Wives Lake, a bird sanctuary and staging area about 25 miles south-west of Moose Jaw, would have moved on and may well be part of the recorded bird clutter.

ACKNOWLEDGMENT

This paper represents the culmination of a radar clutter classification research program that has spanned over 15 years. Many graduate students and members of the research staff of the CRL, McMaster University have contributed to this program in one form or another. We are indebted to them all.

Access to operational radar sites at CFB Trenton (Ontario), and CFB Moose Jaw (Saskatchewan), made it possible to collect an invaluable radar data library that has provided a solid base for the experimental study reported herein. We are all grateful to the Canadian Department of National defence for giving us the permission to do so.

REFERENCES

[1] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
[2] M. I. Skolnik, "Fifty years of radar," *Proc. IEEE*, vol. 73, pp. 182–197, Feb. 1985.
[3] M. I. Skolnik, Ed., *Radar Handbook*. New York: McGraw-Hill, 1970.
[4] E. Eastwood, *Radar Ornithology*. London, U.K: Methuen, 1967.
[5] W. L. Flock and J. L. Green, "The detection and identification of birds in flight, using coherent and noncoherent radars," Proc. IEEE, vol. 62, pp. 745–753, June 1974.
[6] C. R. Vaughn, "Birds and insects as radar target: A review," Proc. IEEE, vol. 73, pp. 205–227, Feb. 1985.
[7] E. R. Stables and N. D. New, "Birds and aircraft: The problems," in *The Problems of Birds as Pests*, R. K. Murton and E. N. Wright, Ed. London, U.K.: Academic, 1968, pp. 3–16.
[8] H. Blokpoel, *Bird Hazards to Aircraft*. Ottawa, B.C.: Clarke, Irwin & Company, 1976.
[9] H. Blokpoel, "Bird migration forecasts for military operations," Canadian Wildlife Service, Occasional Paper 16, Ottawa, B.C., 1973.
[10] W. W. H. Gunn and V. E. F. Solman, "A bird warning system for aircraft in flight," in *The Problems of Birds as Pests*, R. K. Murton and E. N. Wright, Eds. London, U.K.: Academic, 1968, pp. 87–96.
[11] G. W. Schaefer, "Bird recognition by radar: A study in quantitative radar ornithology," in *The Problems of Birds as Pests*, R. K. Murton and E. N. Wright, Eds. London, U.K.: Academic, 1968, pp. 53–86.
[12] J. R. Barry, B. K. Carter, R. J. Erdahl, R. L. Harris, J. T. Miller, G. D. Smith, and R. M. Barnes, "Angel clutter and the ASR air traffic control radar," Applied Physics Laboratory, John Hopkins University, Silver Spring, MD, Final Report under Federal Aviation Administration Contract DOT-FA72WA-2705, Feb. 1973.
[13] L. M. Dill and P. F. Major, "Spatial structure of bird flocks," Associate Committee on Bird Hazards to Aircraft, National Research Council, Ottawa, B.C., Field Note 76, June 1977.
[14] F. E. Nathanson and J. P. Reilly, "Radar precipitation echoes: Experiments on temporal, spatial, and frequency correlation," IEEE Trans. Aerosp. Electron. Syst., vol. AES-4, pp. 505–514, July 1968.
[15] P. L. Smith, K. R. Hardy, and K. M. Glover, "Applications of radar to meteorological operations and research," *Proc. IEEE*, vol. 62, pp. 724–745, June 1974.
[16] R. J. Doviak and D. S. Zrnic, *Doppler Radar and Weather Observations*. New York: Academic, 1984.
[17] I. I. Zawadzki, "Statistical properties of precipitation patterns," J. Applied Meteorology, vol. 12, pp. 459–472, Apr. 1973.
[18] J. W. Taylor and G. Brunins, "Design of a new airport surveillance radar (ASR-9)," *Proc. IEEE*, vol. 73, pp. 284–289, Feb. 1985.
[19] M. W. Long, *Radar Reflectivity of Land and Sea*. Lexington, MA: Lexington Books, 1975.
[20] D. K. Barton, "Land clutter models for radar designs and analysis," *Proc. IEEE*, vol. 73, pp. 198–204, Feb. 1985.
[21] C. E. Muehe, W. H. Drury, E. M. Hofstetter, M. Labitt, P. B. McCorison, and V. J. Sferrino, "New techniques applied to air-traffic control radars," *Proc. IEEE*, vol. 62, pp. 716–723, June 1974.
[22] S. J. Rabinowitz, C. H. Gager, E. Brookner, C. E. Muehe, and C. M. Johnson, "Applications of digital technology to radar," *Proc. IEEE*, vol. 73, pp. 325–339, Feb. 1985.
[23] F. R. Hunt, "Automatic radar equipment to determine bird strike probability PART II. Migrating water-fowl flocks," Associate Committee on Bird Hazards to Aircraft, National Research Council, Ottawa, B.C., Field Note 75, Mar. 1977.
[24] S. Haykin and C. R. Carter, "Radar separation of aircraft from birds and weather," Communications Research Laboratory, McMaster University, Hamilton, Ont., CRL Internal Rep. Series CRL-28, Apr. 1975.
[25] B. W. Currie, C. R. Carter, and S. Haykin, "Signal processing applied to the separation of aircraft from birds and weather using air traffic control radar," Communications Research Laboratory, McMaster University, Hamilton, Ont., CRL Internal Rep. Series CRL-50, Nov. 1977.
[26] H. C. Chan and S. Haykin, "An adaptive MTI processor using FIR and IIR filters," Communications Research Laboratory, McMaster University, Hamilton, Ont., CRL Internal Rep. Series CRL-38, May 1976.
[27] S. B. Kesler, "Nonlinear spectral analysis of radar clutter," Communications Research Laboratory, McMaster University, Hamilton, Ont., CRL Internal Rep. Series CRL-51, Jan. 1978.
[28] S. Haykin, S. B. Kesler, and B. W. Currie, "An experimental classification of radar clutter," *Proc. IEEE*, vol. 67, pp. 332–333, Feb. 1979.
[29] B. W. Currie and S. Haykin, "Radar clutter classifier prototype for use with the TRACS airport surveillance radar, Part I," Communications Research Laboratory, McMaster University, Hamilton, Ont., CRL Internal Rep. Series CRL-94, Dec. 1981.
[30] S. Haykin, B. W. Currie, and S. B. Kesler, "Maximum-entropy spectral analysis of radar clutter," *Proc. IEEE*, vol. 70, pp. 953–962, Sept. 1982.
[31] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic, 1972.
[32] T. Y. Young and T. W. Calvert, *Classification, Estimation and Pattern Recognition*. New York: American Elsevier, 1974.
[33] Y. C. Ho and A. K. Agrawala, "On Pattern Classification Algorithms - Introduction and Survey," *Proc. IEEE*, vol. 56, pp. 2101–2114, Dec. 1968.
[34] L. Kanal, "Patterns in pattern recognition: 1968–1974," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 697–722, Nov. 1974.
[35] K. S. Fu, "Recent developments in pattern recognition," *IEEE Trans. Comput.*, vol. C-29, pp. 845–854, 1980.
[36] A. K. Jain, "Advances in statistical pattern recognition," in *Pattern Recognition Theory and Applications*, P. A. Devijver and J. Kittler, Eds. Berlin, Germany: Springer-Verlag, 1987, pp. 1–19.
[37] R. L. Kashyap, "Optimal feature selection and decision rules in classification problems with time series," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 281–288, May 1978.
[38] C. H. Chen, *Nonlinear Maximum Entropy Spectral Analysis Methods for Signal Recognition*. Chichester, U.K.: Research Studies Press, 1982.
[39] J. F. Agnel and E. N. S. T., "Modélisation autorgréssive et classification de fouillis de sols. Application à la reconnaissance," in *Proc. Int. Conf. Radar*, Versailles, France, May 1984, pp. 109–114.
[40] J. F. Agnel, "Elimination de fouillis en radar et détection séquentielle des cibles: une approche par la reconnaissance des formes et le filtrage en trellis," These de Docteur Ingenieur, Ecole Nationale Superieure des Telecommunications, Paris, France, Mar. 1985.
[41] W. Stehwien, "A clutter classifier—Preliminary results," Communications Research Laboratory, McMaster University, Hamilton, Ont., CRL Internal Rep. Series CRL-135, Dec. 1984.
[42] W. Stehwien, "Radar clutter classification," Ph.D. dissertation, McMaster University, Hamilton, Ont., 1989.
[43] W. Stehwien and S. Haykin, "Experimental radar clutter data evaluation for clutter classification research," Communications Research Laboratory, McMaster University, Hamilton, Ont., CRL Internal Rep. Series CRL-152, Jan. 1986.
[44] W. Stehwien and S. Haykin, "A statistical radar clutter classifier," in *Proc. 1989 IEEE National Radar Conf.*, Dallas, TX, Mar. 1989, pp. 164–169.

[45] W. Stehwien, "Parametric spectral analysis of radar clutter," Masters Eng. thesis, McMaster University, Hamilton, Ont., Sept. 1983.
[46] R. O. Duda and R. H. Blackmer, "Application of pattern recognition techniques to digitized weather radar data," Stanford Research Institute, Menlo Park, CA, Final Rep. (covering the period May 25, 1971 to March 31, 1972) under NOAA National Weather Service Contract NWS 1–36 072, May 1973.
[47] R. H. Blackmer, R. O. Duda, and R. Reboh, "Application of pattern recognition techniques to digitized weather radar data," Stanford Research Institute, Menlo Park, CA, Final Rep. (covering the period March 31, 1972 to April 30, 1973) under NOAA National Weather Service Contract NWS 1–36 072, May 1973.
[48] M. I. Skolnik, *Introduction to Radar Systems.* New York: McGraw-Hill, 1980.
[49] F. E. Nathanson, *Radar Design Principles.* New York: McGraw-Hill, 1969.
[50] L. V. Blake, *Radar Range-Performance Analysis.* Lexington, MA: Lexington Books, 1980.
[51] J. P. Burg, "A new analysis technique for time series data," presented at the NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, Enschede, The Netherlands, 1968.
[52] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1986.
[53] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 55–63, Jan. 1968.
[54] J. M. van Campenhout, "On the peaking of the hughes mean recognition accuracy: The resolution of an apparent paradox," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, pp. 390–395, May 1978.
[55] D. R. Moorcroft, "Maximum-entropy spectral analysis of radio-auroral signals," *Radio Science*, vol. 13, no. 4, pp. 649–660, July-Aug. 1978.
[56] D. H. Foley, "Considerations of sample and feature size," IEEE Trans. Inform. Theory, vol. IT-18, pp. 618–626, Sept. 1972.
[57] N. Glick, "Additive estimators for probabilities of correct classification," *Pattern Recognition*, vol. 10, pp. 211–222, 1978.
[58] M. B. Wilk and R. Gnanadesikan, "Probability plotting methods for the analysis of data," *Biometrika*, vol. 55, no. 1, pp. 1–17, 1968.
[59] T. Y. Young and K. S. Fu, Eds. *Handbook of Pattern Recognition and Image Processing.* Orlando, FL: Academic, 1986.
[60] W. Stehwien and S. Haykin, "Classification of radar clutter based on spectral parameters," in *Proc. Int.Conf. Radar*, Versailles, France, May 1984, pp. 126–131.
[61] R. P. Lippmann, "An introduction to computing with neural networks," *IEEE ASSP Magazine*, vol. 4, pp. 4–22, 1987.
[62] T. J. Sejnowksi and C. R. Rosenberg, "NETtalk: A parallel network that learns to read aloud," John Hopkins University, Baltimore, MD, Tech. Rep. JHU/EECS-86/01, 1986.
[63] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," Institute for Cognitive Science, University of California, San Diego, CA, Rep. 8506, 1985.
[64] P. J. Werbos, "Beyond regression: New tools for prediction and analysis in the behavioral sciences," Ph.D. dissertation (applied mathematics), Harvard University, Cambridge, MA, 1974.
[65] Y. LeCun, "Une procedure d'apprentissage pour reseauy a seuil assymetrique", (A learning procedure for asymmetric threshold network), in *Proc. Cognitiva 85*, Paris, France, 1985, pp. 599–604.
[66] D. B. Parker, "Learning-Logic," Office of Technology Licensing, Stanford University, Stanford, CA, Invention Rep., S81–64, File 1, 1982.
[67] D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations.* Cambridge, MA: MIT Press, 1986.
[68] S. Haykin, *Introduction to Neural Networks.* New York: MacMillan, to be published.
[69] R. A. Jacobs, "Increased rates of convergence through learning rate adaptation," *Neural Networks*, vol. 1, pp. 295–307, 1988.
[70] D. R. Hush and J. M. Salas, "Improving the learning rate of back-propagation with the gradient reuse algorithm," in *Proc. 2nd Int. Conf. Neural Networks*, vol. I, 1988, pp. 441–447.
[71] C. Deng, "A multilayer neural network classifier for radar clutter," Masters Eng. thesis, McMaster University, Hamilton, Ont., Canada, 1990.
[72] C. Deng and S. Haykin, "Classification of radar clutter using neural networks," *IEEE Trans. Neural Networks*, to be published.
[73] P. Mahapatra and D. S. Zrnic, "A physical basis for NEXRAD data update rates," *J. Aircraft*, pp. 840–850, 1984.
[74] D. R. Mann, "TDWR clutter residue map generation and usage," MIT Lincoln Laboratory, LL ATC-148, DOT/FAA/PM-87–26, 1989.
[75] H. Blokpoel, "Migration of lesser snow and blue geese in spring across southern Manitoba, Part 1: Distribution, chronology, directions, numbers, heights and speeds," Canadian Wildlife Service, Ottawa, B.C., Rep. Series 28, 1974.
[76] W. J. Richardson, "Spring migration and weather in eastern Canada: A radar study," *American Birds*, vol. 25, no. 4, pp. 684–690, Aug. 1971.
[77] W. J. Richardson, "Autumn migration and weather in eastern Canada: A radar study," *American Birds*, vol. 26, no. 1, pp. 10–17, Feb. 1972.
[78] H. Blokpoel and J. Burton, "Weather and height of nocturnal migration in eastcentral Alberta: A radar study," *Bird-Banding*, vol. 46, no. 4, pp. 311–328, 1975.
[79] W. J. Richardson, "Spring Migration over Puerto Rico and the Western Atlantic: A radar study," *Ibis*, vol. 116, pp. 172–193, 1974.
[80] W. J. Richardson, "Southeastward shorebird migration over Nova Scotia and New Brunswick in Autumn: A radar study," *Canadian Journal of Zoology*, vol. 57, pp. 107–124, 1979.
[81] H. Blokpoel and M. C. Gauthier, "Migration of lesser snow and blue geese in spring across southern Manitoba, Part 2: Influence of the weather and prediction of major flights," Canadian Wildlife Service, Ottawa, B.C., Rep. Series 32, 1975.

Simon Haykin (Fellow, IEEE) received the B.Sc. degree (First-Class Honors) in 1953, the Ph.D. degree in 1956, and the D.Sc. degree in 1967, all in electrical engineering from the University of Birmingham, England.

He is the founding Director of the Communications Research Laboratory and Professor of Electrical and Computer Engineering at McMaster University, Hamilton, Ontario, Canada. His research interests include statistical signal processing, adaptive filters, neural networks, and chaotic theory with applications to radar. He is the author of several text books, including *Communications Systems*, Second Edition (Wiley), and *Adaptive Filter Theory*, Second Edition (Prentice-Hall).

Dr. Haykin was elected Fellow of the Royal Society of Canada in 1980. He is corecipient of the Ross Medal from the Engineering Institute of Canada and the J. J. Thomson Premium from the Institution of Electrical Engineers, London. He was awarded the McNaughton Gold Medal, IEEE (Region 7), in 1986.

Richard Mann received the B.Eng. degree in computer engineering and management and the M.Eng. degree in electrical engineering both from McMaster University in Hamilton, Ontario, Canada, in 1988 and 1990, respectively.

He is currently a visiting researcher at the Helsinki University of Technology in Espoo Finland. His interests are in the areas of neural networks, pattern recognition, computer vision, and parallel processing.

Cong Deng was born in Chengdu, China. He received the B.Eng. degree with first class honors in 1988, and the M.Eng. degree in 1990, both in electrical engineering from McMaster University, Hamilton, Ontario, Canada.

Currently, he is working at Bell-Northern Research Ltd. Ottawa, Canada. His research interests include high-speed digital signal processing and neural networks.

Wolfgang Stehwien received the B.Sc.E. E. degree (with Great Distinction) from the University of Saskatchewan in 1976. He worked in the field of communications and automatic control systems for Pan Canadian Petroleum Ltd. in Calgary and SED Systems Inc. in Saskatoon before returning to post-graduate studies at McMaster University in Hamilton, Ontario in 1981. He received the M.Eng. degree in 1983, and the Ph.D. degree in 1989, both in electrical engineering.

Since 1987, he has been employed as radar systems engineer at Litton Systems Canada, Ltd. He is currently working on the problem of detection and tracking of small targets in sea clutter with airborne maritime surveillance radars.

Peter Weber received the B.Eng. and M.Eng. degrees in electrical engineering from McMaster University, Hamilton, Ontario, Canada in 1980 and 1983, respectively.

Since 1982, he has worked as a research engineer for the Communications Research Laboratory, McMaster University, Hamilton, Ontario, Canada. His research has been on the design and evaluation of radar systems, concentrating on signal processing, detection and estimation.